# Urban Digital Memory

*Leveraging Social Media Data on Public Spaces to Inform Urban Design Processes*

Soroush Saffarzadeh



# Colophon


| | |
|---|---|
| TITLE | Urban digital memory<br>Leveraging social media data on public spaces to inform urban design processes |
| STATUS | A3 Defended |
| DATE | June9th, 2026 |
| AUTHOR | Soroush Saffarzadeh |
| STUDENT Nr. | 6227007 |
| INSTITUTE | Delft University of Technology,<br>Faculty of Architecture & the Built Environment (Bouwkunde) |
| DEGREE | Master of Science Architecture, Urbanism and Building Sciences |
| TRACK | Urbanism |
| CLUSTER | Data-Informed Strategies for Sustainable Urban Futures |
| 1ST MENTOR | Clémentine Cottineau-Mugadza |
| TITLE | Assistant Professor in Urban Studies |
| DEPARTMENT | Urbanism |
| 2nd MENTOR | Claudiu Forgaci |
| TITLE | Assistant Professor in Urban Design |
| DEPARTMENT | Urbanism |

# Urban Digital Memory

## Leveraging Social Media Data on Public Spaces to Inform Urban Design Processes

Keywords
*Digital Memory, Right to the city, Social Media, Public Engagement, Data-Driven Design, Tehran*

## Abstract

This study seeks to examine the capabilities of citizens' lived experiences shared on social media platforms as digital memories to value the spatial features of public squares in Tehran. The evaluation is based on emotions expressed by social media users about the public space, whether they have positive, negative, neutral, or mixed impacts. The data will be collected as text strings from X (formerly Twitter) using the Apify scraping app. The collected text data went through topic modeling and sentiment analysis to extract emotions in the form of negative, positive and neutral values. The classified emotions were associated with the spatial features addressed by tweet descriptions through a cognitive approach to spatial perception and the expert's interpretation. The valued associations shaped restrictive, reconsidered, or no-action mitigations. At the end, a design framework was proposed and applied to the case study. The project brought the concept of right to the city, not only as a physical or mental entity, but also as a digital memory, based on the fact that public space is validated by people in digital spaces, which grants them the right to the memory as well. The other aspect of the study would be the passive public information on design provided through digital inputs, giving an initial overview to designers, planners, and decision makers of an urban space.

# Acknowledgements

I'd like to thank my mentors, Clémentine and Claudiu, who kept giving their support and empathy to me through this journey. They did their best to help me with all the procedures and how to bring this project to life. I learnt a lot from them that will shape my path towards further research and relevant projects. I'm so glad that I had both of them. Clémentine always gave detailed and thoughtful feedback, and Claudiu helped to organize and make the whole process feasible, which really made the practical matters clearer and easier for me. All of this, and they made it happen as smoothly as possible. They were the perfect team.

To my family, who have always been there for me and never gave up on me or my work: I can never repay your kindness and sacrifice, so I'd keep thanking you anytime possible.

A big thanks to all my friends who helped me with their support, ideas, and comfort. Specifically, I want to acknowledge Mahdi, whose Bachelor's project was the inspiration for me to work on this topic. Thanks to Fatemeh, who showed her interest and helped me by providing images of recent changes in the study area. I'd like to thank Nima, who gave me very constructive feedback on my final output and ideations.

And thanks to Yasaman, who was my library pal in the last three months towards the end of this project; having her was a way to survive the hardships and challenges of the time, and I really appreciated the coffee/lunch breaks we had. I'd also like to thank Megha, who has always been supportive and showed up for me, and I enjoyed the times we spent in the Polis room working quietly. And, last but not least, a special thanks to Shreya and Adriano, who were actually the ones who encouraged me to pick this topic when we were drinking beers on a random summer evening and it sparked in my head: if not now and here in TU Delft, then when?

# Contents


Introduction ..... 1
City Theater ..... 5
(Teatré Shahr-e Tehran) ..... 5
Problem Statement ..... 11
Research Objectives ..... 17
Research Questions ..... 19
Literature Review ..... 21
Right to the memory as a right to city ..... 21
Digital era (archival memories) ..... 22
Public Engagement (digital participation) ..... 23
Research Background ..... 25
Theoretical Framework ..... 27
Conceptual Framework ..... 29
Framework ..... 32
Expected Outcomes ..... 35
Scale ..... 36
Scope ..... 36
Timeline ..... 36
1st Scale: City Theater ..... 37
2nd Scale: key public spaces across Tehran ..... 37
X (Formerly Twitter) ..... 43
Methodology ..... 44
Methodological Framework ..... 45
Co-occurrence Matrix ..... 48
Analytical Process ..... 48
Topic Modeling ..... 49
Word Embedding ..... 51
Sentiment Analysis ..... 51
Data Collection, Processing & Analysis ..... 52
Preprocessing ..... 53
Processing ..... 55
Visualisation ..... 56
General Topics ..... 58
Results ..... 58

Localised Topics ..... 59
Word Cloud ..... 61
Topic Distributions ..... 63
Spatial Localized Topics' Distributions ..... 65
Topic-Sentiment Overall Distribution ..... 67
City Theater Between 2022-2026 ..... 69
Combined Scales ..... 71
Tweet Frequency per Average Sentiment Value ..... 73
Chi-Square Test ..... 75
Chi-Square Test Results ..... 75
Over/Under Representation ..... 77
Representation based on Location Quotient ..... 77
Key Places ..... 79
Co-occurrence Matrices ..... 81
Synthesis ..... 87
Design Framework ..... 88
Analytical Process ..... 91
Design Process ..... 93
Examples of the Process in the Study Area ..... 97
Example of Mitigations Based on the Guiding Values ..... 105
Proof of Concept ..... 110
Conclusion ..... 115
Answering the Research Questions ..... 117
Answer to the Main Question ..... 118
Discussion ..... 119
Reflection ..... 122
AI Acknowledgments ..... 124
References ..... 125
Appendix ..... I
Data Management Checklist ..... XIII
Scripts and Workflows ..... XIV

*List of Figures*

Figure 1. Tehran Province in Iran and city divisions....2
Figure 2. The Urban Digital Memory....3
Figure 3. City Theater Square's location in Tehran....5
Source: Google Map Satellite, 2026.....5
Figure 4. Aerial photo of City Theater Plaza and Daneshjoo Park in 1975.Source: Ehsanrostamipo-ur, 2023. ....6
Figure 5. City Theater's Building and plaza in 1971....7
Source: Wikimedia, 2026. ....7
Figure 6. City Theater's building and square in 2023. ....8
Source: Hamshahri Online, 2026....8
*Figure 7. City Theater Square transformations between 2000 and 2025 ....9*
*Source: Google Earth Engine, 2026. ....9*
*Figure 8. Fence placement around City Theater for reconstruction.... 12*
*Sources: (Bottom) Wikimedia Commons , 2026; (Top) Rostamipour.... 12*
Figure 9. Demolishing the central space of City Theater's square for reconstruction....13
Source: Rokna, 2025. ....13
Figure 10. Satellite image of City Theater in its latest condition ....13
Source: Google Earth Engine, 2025....13
Figure 11. People share their opinions on X (Formerly Twitter) regarding the new interventions on the City Theater square....14
Sources: "...in recent years, the sanctity of the City Theater, 2025; "Stop fencing around the City Theater...", 2023; "The Shahr Theater building, vibrant yesterday, ruined today!...", 2025; "Despite widespread protests against the fencing plan...", 2025....14
Figure 12. Theoretical Framework....28
Figure 13. Space to place transformation....30
Figure 14. Emotions emergence followed by spatial interventions....31
Figure 15. Conceptual Framework....32
Figure 16. Lynch's cognitive urban elements from 'The Image of the City'....34
Source: Can, 2019....34
Figure 17. Expected Outcomes ....35
Figure 18. Timeline and scale definitions of the research scope ....36
Figure 19. Tehran's overview and City Theater's location within ....38
Source: Google Earth, 2024. ....38
Figure 20. Current state of the City Theater after removing the reconstruction dividing walls ....38
Source: Fatemeh Fallah, 2026....38
Figure 21. Key public spaces of Tehran....39
Figure 22. Key Public Places in Tehran, 1st zone. ....40
Figure 23. Key Public Places in Tehran, 2nd zone. ....41
Figure 24. Key Public Places in Tehran, 3rd zone. ....42
Figure 25. Methodological Framework. ....45
Figure 26. Topic-sentiment co-occurrence matrix in two timelines and Tehran's scale....47
Figure 28. Data collection, preprocessing and analysis ....51
Figure 27. Adaptation from Ekman's (1999) basic emotion categories ....52
Source: Ekman, 1999....52
Figure 29. WordCloud of Spatial Topics....61
Figure 30. General spatial topics distribution ....63
Figure 31. Localised spatial topics....65
Figure 32. Localised topic-sentiment distribution ....67
Figure 33. Topic-sentiments for City Theater before 2024....69
Figure 34. Topic-sentiments for City Theater after 2024 ....70
Figure 35. Overall topic-sentiments before 2024 ....72
Figure 36. Overall topic-sentiments after 2024....72
Figure 37. Localised spatial topics with an average sentiment score before 2024....74
Figure 38. Localised spatial topics with an average sentiment score after 2024 ....74
Figure 39. Over-under representation based on Location Quotient....79
Figure 40. General topic-place matrix....82

Figure 41. Localised topic-place matrix ..... 82
Figure 42. Topic-sentiment shifts between 2022-2026 ..... 84
Figure 43. Topic-sentiment shifts with weighted scores between 2022-2026..... 86
Figure 44. Discourse shifts with probability weight scores between 2022-2026..... 86
Figure 45. Generic elements in the design process..... 88
Source: Van Dooren et al, 2014. ..... 88
Figure 46. Moving between domains as the experimentation ..... 90
Source: Van Dooren et al, 2014. ..... 90
Figure 47. Proposed Analysis & Design Framework ..... 91
Figure 48. Proposed Analysis & Design Framework ..... 93
Figure 49. Zoom-in of framework's example on Vali'asr..... 98
Figure 50. Framework application example on Tajrish & Vali'asr ..... 99
Figure 51. Framework application example on City Theater & Vali'asr ..... 101
Figure 52. Framework application example on Toopkhaneh, HassanAbad and Grand Bazaar ..... 103
Figure 53. Policy / Design mitigations ..... 105
Figure 54. Community inputs on High Line design process ..... 112
Source: Lee (2026) ..... 112
Figure 55. Leaflet map of Gaza with stories told and spotted by citizens online..... 112
Source: (GazaPlaces, 2026) ..... 112
Figure 56. Vision of implemented design framework backed by urban digital memory .. 113
Figure 57. Research objectives met by the design framework..... 116

# Introduction

You might have a favorite spot around the corner in your neighborhood or within the city blocks. Somewhere you visit quite often and find it to be a place you feel safe and comfortable. Somewhere that you've made a lot of memories in. In other words, it's the place of your memories. The place is rooted in your heart and you may feel a connection with it. And one day, you decide to visit your haven, you go on a ride or simply on foot to get there. You're at the spot you'd surely recognize anytime that you stand on. Yet, the place doesn't look the same as it used to be. It's gone. You find yourself staring at a reconstruction site that used to be your precious spot in the city. You might ponder some questions. How did you use to feel about that place before, and how do you feel now? And... ?what happened to all of those memories

1

The spatial intervention always brings a lot of non-spatial consequences with it. Those can be the lived-experiences of citizens towards a former space or a location within the urban fabric. Places that are attractive and even valuable to their visitors. They store various types of human emotions that constantly change from place to place. But to what extent are firm- or municipal-led urban projects ?truly citizen-oriented

The urban transformations through planning and design have gone a long way. Starting from technical blueprint products to co-design processes with citizens (Arnstein 2019; Horelli and Wallin 2024; Mueller et al. 2018). In the past few years, participatory approaches have amplified the focus on engaging the public in urban interventions. However, not all countries have a democratic .structure for planning and design

Iran is a country in western Asia, often considered part of the Middle East. The city of Tehran is the capital, with a population of nearly 9.7 million in 2025 (MacroTrends, 2026). A dense urban area with hustle and bustle in its city center and various commercial, cultural, and official activities. The city is divided into 22 districts governed by the main municipality, and each district has its own subset of local government organised by the central municipality. Within each district, there are several neighborhoods, the smallest official divisions, also governed by the district municipality through community development offices in direct contact with the local communities. However, this connection .is not active



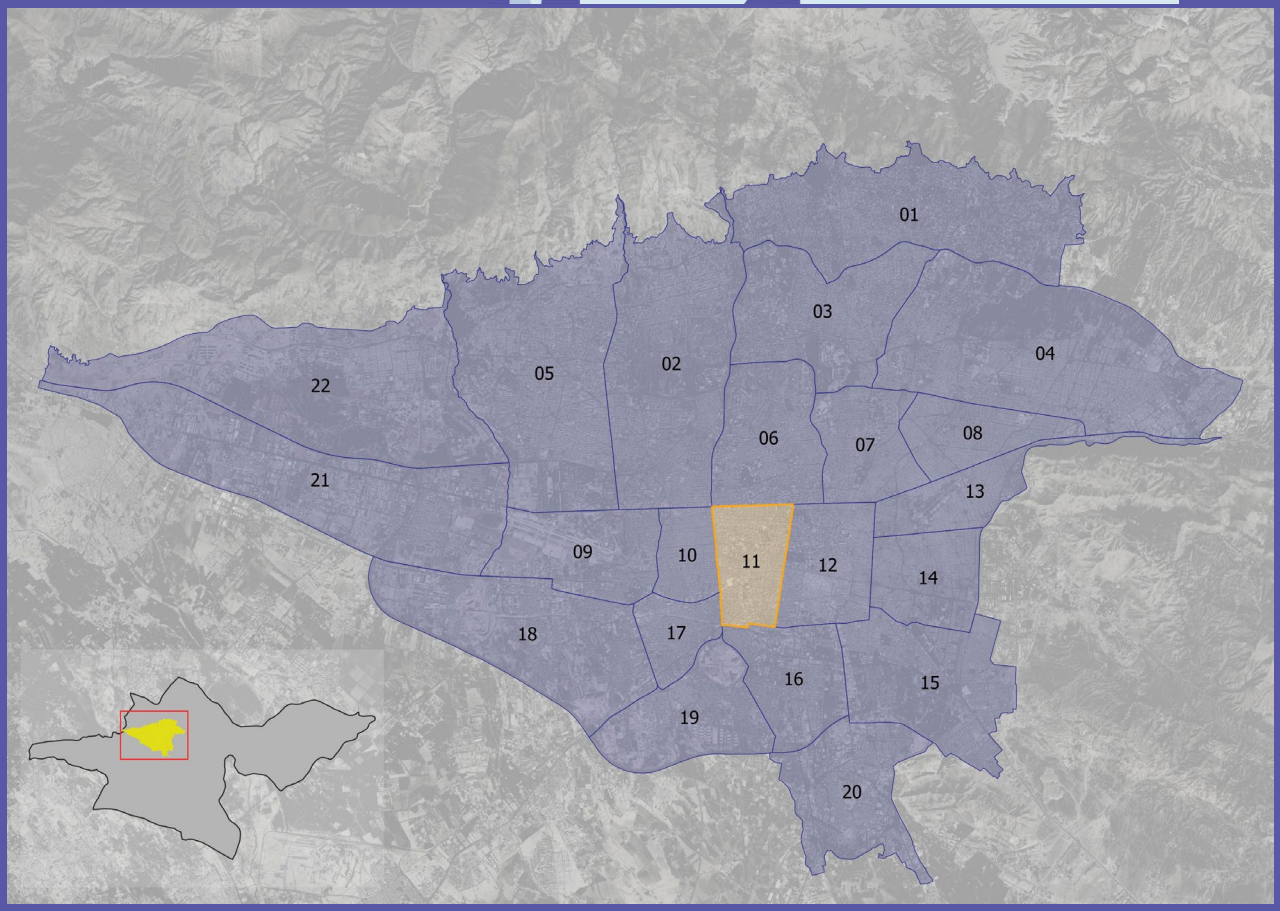




*Figure 1. Tehran Province in Iran and city divisions*

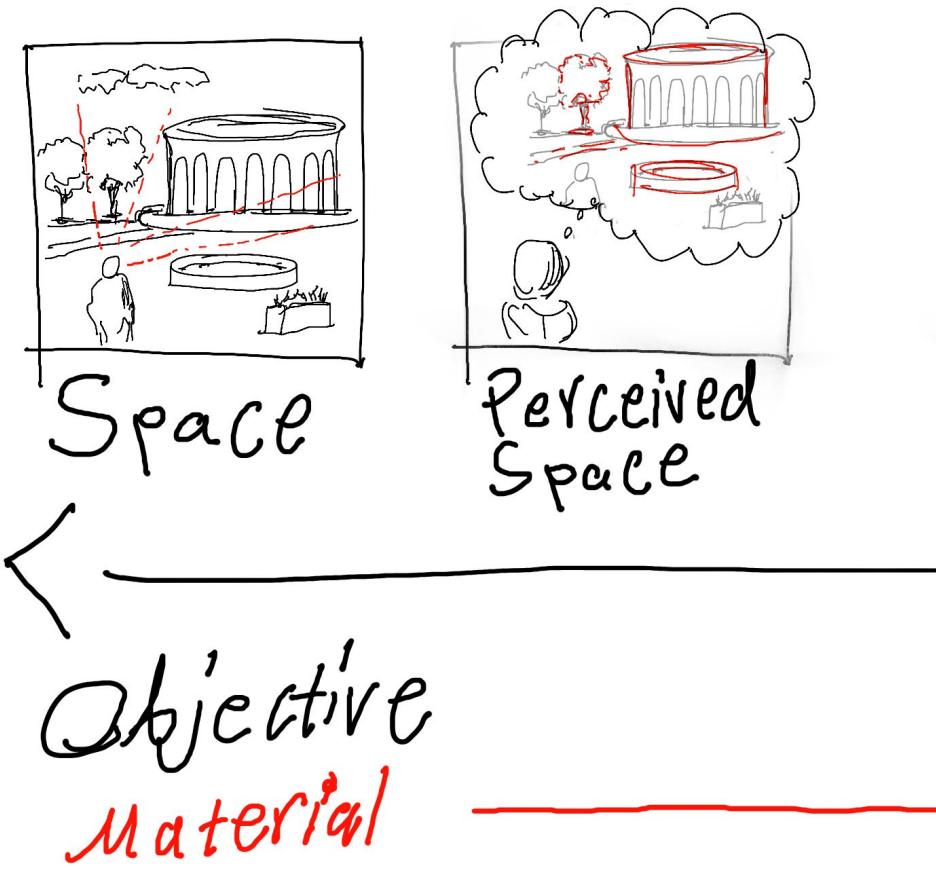


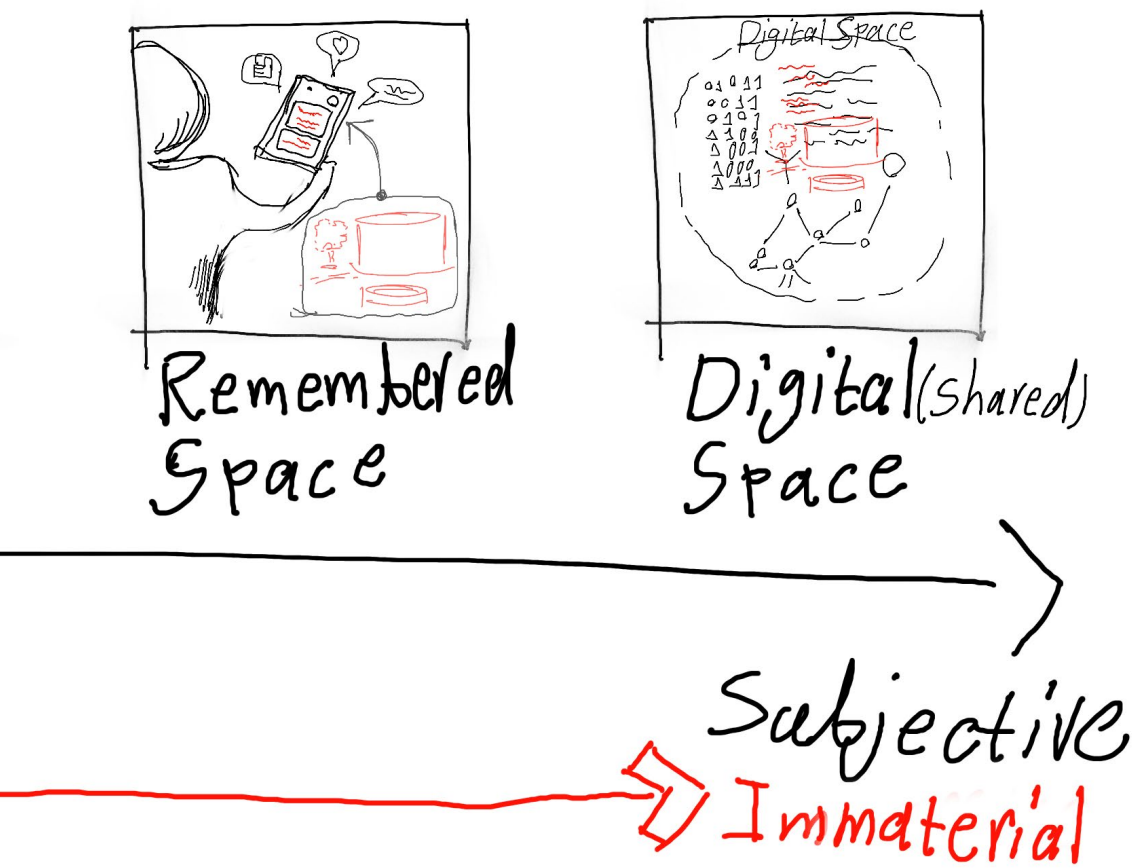

*Figure 2. The Urban Digital Memory*



In countries like Iran, despite following the theories and citizen engagement exercises in academia, the practice is far behind this mindset. Many urban development projects are implemented with a top-down approach, regardless of who is responsible. Whether a contractor, developer, designer, planner, or policy maker, they rarely ever approach the public for their opinions on an urban project. There are minimal practices that try to include citizens within their work by conducting surveys or interviews. Nevertheless, the number of participants has barely met the needs of the affected population, which looked more like manipulation (Arnstein, 2019). The planning and design approaches in countries with less democratic political systems would yield the least successful public engagement through their processes.

Although active participation has never been easy, tools and platforms have made the process smoother for both parties. In the digital era, many citizens have access to online social platforms where they can usually speak up for themselves, with the chance that their opinions could be heard nationally and even globally. In such an era, Iran is not an exception either. Although Iranians have filtered internet access, they use VPNs to bypass the restrictions and access the global internet, sharing their opinions, voices, cries, compliments, memories, and other experiences that are made inaccessible on a national scale.

This new accessibility can be considered a new step towards democratic approaches in public engagement and participation (Liang, 2020; Pereira, Célia, et al., 2012; Williamson & Ruming, 2020). Sharing






news and posts about urban spaces is an ongoing phenomenon that goes beyond a single media post with captions. It can be seen as acknowledging a new space, or, conversely, as downgrading a new intervention in the city. The background of a city can now be traced through online platforms run by individuals who have captured it over time. This doesn't end with the past of urban spaces, but also with how urban spaces change in real time, which would have its own responses. How citizens react to an urban project is evident on social media, and how they perceive the transformed urban space is evident as well. The urban space is not just a physical or mental entity; it's also digital.

This study aims to examine how citizens' lived experiences shared on social media form urban digital memories that can value and guide spatial interventions. Digital memory creates new concepts on top of the notion of the right to the city, which can be called the right to the memory.

**The project focuses on how the right to the memory can serve as a significant guiding value for urban interventions in spatial features through a digital citizen-informed data perspective.**

# City Theater

# (Teatré Shahr-e Tehran)

One of the public squares that has always been the site of intense tension between its visitors and those who want to change it is City Theater[1] Square[2]. As can be seen in Figure 2, the square is in the intersection of Vali'asr and Enghelab streets, which are the longest NS / EW armatures of Tehran. The area is located in District 11 of Tehran's municipality.

The City Theater Square holds one of the oldest theaters in Tehran, dating back to the 1970s. The City Theater of Tehran building was built as an art and cultural complex. City Theater was not only a venue but also a taste of contemporary design inspired by Iranian architecture (Mohit, 2026; wikipedia, 2026). The layout, presence of the building and its surroundings, formed a public space, as the complex was hosting cultural, artistic and social activities within.

The City Theater Square used to look like a complex with its specific plaza in front of the theater's main building (Figure 4). The current Daneshjoo park, neighboring the Theater's building, was larger than it is now, covering the northern side of the City Theater that touches Enghelab Street. Based on the documentation of a user who shared the aerial image on X (formerly Twitter), the spatial layout in the image is the earliest and most original spatial intervention in this space.



1.Teatré Shahr in Persian [Translated by the author]

2. The square holds the city theater building within its space, which could be considered a semi-open public space in Tehran, with all of the activities and events that take place downtown.

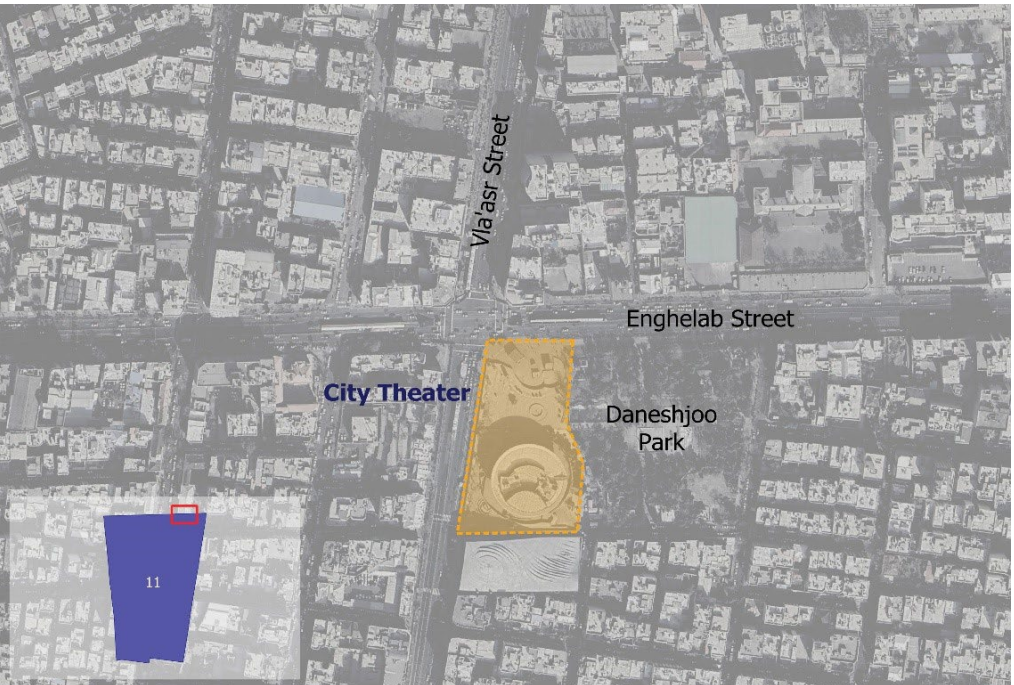



*Figure 3. City Theater Square's location in Tehran*

*Source: Google Map Satellite, 2026.*



*Figure 4. Aerial photo of City Theater Plaza and Daneshjoo Park in 1975. Source: Ehsanrostamipour, 2023.*

As shown in Figure 5 & Figure 6, the building and the plaza surrounding it have undergone many transformations, including changes in spatial composition, accessibility, mobility, vegetation, and urban furniture. The expansion in the open space of the area resulted in turning it more into a plaza/square rather than a plaza only. Based on that, I'd prefer to acknowledge it as a square or simply name it as "City Theater". City Theater Square has always been a central point in the daily urban life of Tehran. Moreover, the square has been the ground of various social and political events since its establishment.





*Figure 5. City Theater's Building and plaza in 1971.*

*Source: Wikimedia, 2026.*



*Figure 6. City Theater's building and square in 2023.*

*Source: Hamshahri Online, 2026.*

The City Theater has undergone several transformations over the decades (Figure 7). Various small- and medium-scale interventions have been implemented within its space, from removing statues to constructing the underground passage, pedestrian access, and the metro station entrance. Some of these transformations and spatial expansions are traceable through Google Earth's satellite imagery. However, satellite images have been captured since the 2000s. Yet they represent significant spatial interventions, whether inside the plaza or outside, around the square and on the main streets of Tehran.



*Figure 7. City Theater Square transformations between 2000 and 2025*

*Source: Google Earth Engine, 2026.*



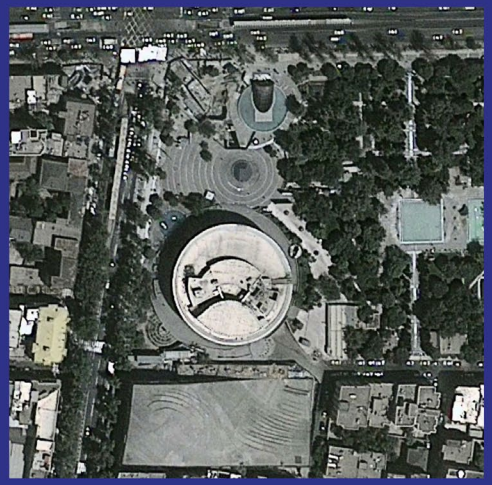
2000



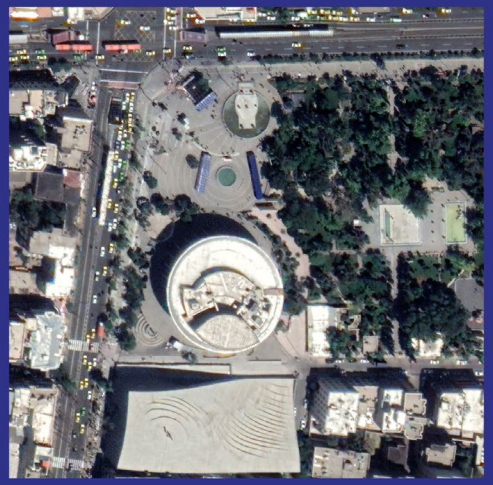
2005



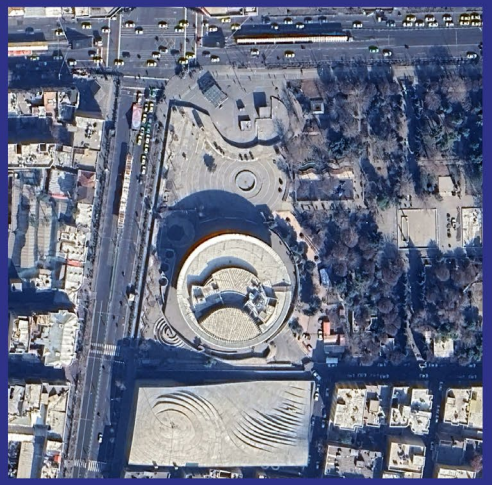
2009



2013



2018



2025

# Problem Statement

There are urban spaces, specifically, public spaces that hold a large pool of collective and individual memories within them, either tragic, joyful, or revolutionary (Solnit, 2009). City Theater Square is not an exception. However, the public apparently has no right or authority over space and its memories as common property. Some spaces are manipulated, transformed, or even demolished by governments, driven by economic or political interests (Paccoud, 2016; Solnit, 2001). An example could be the Haussmannian boulevards, which demolished existing urban streets, creating wide, long boulevards lined with luxury retail shops and police to oppress any citizen opposition in the most dominant way possible (Harvey, 2006). All the valuable memories and the space known to the public were lost in the blink of an eye.



The citizens don't have any agency over space or its corresponding memory to make decisions, or at least to be clearly informed about the interventions and changes made by the actors on higher ground. Not only to the actors but also to the decision-makers and designers; this applies as well. Right to the memory is a metaphor of collective/individual rights over space tied to memories that emerge within it. This represents a bond between the space and its dwellings and visitors, shaped by memories, emotions, presence, interactions, and activities that have shaped each person's lived experience in that space. Nevertheless, this valued connection is taken for granted by the actors, designers, policymakers, and governments holding the upper hand. Sometimes it's even ignored by them, or it conflicts with their main ambitions.

The memory is a valuable lived experience that holds evidence of a physical space, conceived and perceived by everyone (Shaw & Graham, 2017). The cognitive perception of space, recorded on an open online platform for sharing with others, has been discussed in the literature. As Solnit (2001) aesthetically showcases the typicality in urban daily life that may be only engraved in literature, it can emerge in the digital space as well. Social media can store individuals' encounters and interactions within urban spaces in their most typical daily routines, written by them, like a diary.

The spatial transformation of City Theater's Square was ongoing until late February 2026. Over the past two years, City Theater has undergone one of the most drastic transformations since its inception. It was when a long fence divided the adjacent pathways in the southern street, limiting pedestrian access to the square (Figure 8).






*Figure 8. Fence placement around City Theater for reconstruction*

*Sources: (Bottom) Wikimedia Commons , 2026; (Top) Rostamipour.*

A redesign project began to redevelop City Theater's square. As a matter of fact, no citizen was aware of this project until its design proposal was shared with the public. The design proposal depicted a completely new spatial layout for the square. And, as part of the reconstruction, a part of the central space, including an old fountain, was demolished (Figures 9 & 10).

As soon as the public found out about the intervention, they reacted immediately by writing Tweets, consisting of dissatisfied or even angry opinions mentioning how precious the fountain and the square were to them, and that was easily transformed without even asking about it.

The citizens expressed their concerns to the officials and the municipality regarding the fence placement and the City Theater's renovations on social media (Figure 11). However, the reconstruction project was ongoing, and its end date remained unknown to the public. The users shared some images of redesign proposals for the square and directly addressed the matter of fence placement around the City Theater.





*Figure 9. Demolishing the central space of City Theater's square for reconstruction*

*Source: Rokna, 2025.*

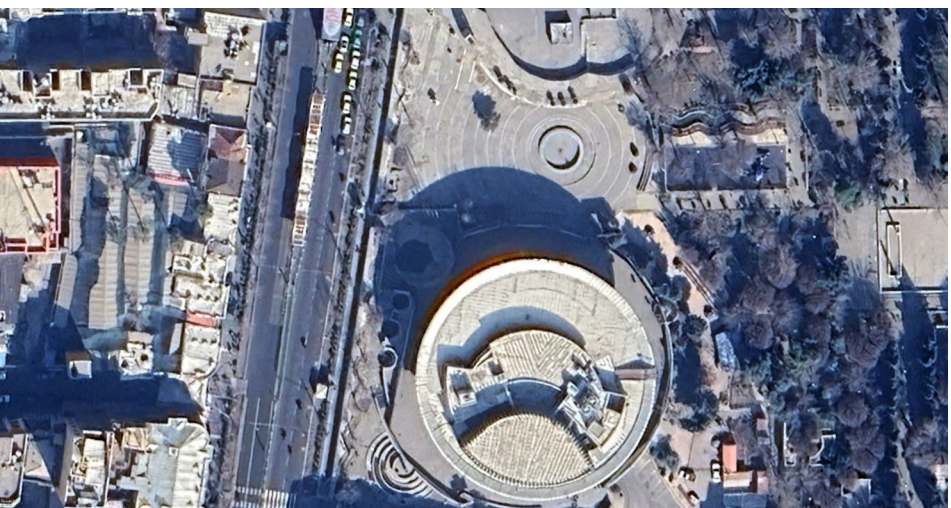


*Figure 10. Satellite image of City Theater in its latest condition*

*Source: Google Earth Engine, 2025.*

***"...in recent years, the sancity of the City Theater has repeatedly been violated..."***

The Tehran City Theater's pool has been [illegible] buried under soil.

The City Theater, which opened in 1972, is a symbol of modern Iranian culture, but in recent years, the sanctity of the City Theater has repeatedly been violated in various ways by the Tehran Municipality.

***"Stop fencing around the Ctiy Theater!"***

Stop fencing around the City Theater!

Following the unveiling of the fencing plan around the City Theater by the Minister of Culture and Islamic Guidance, a campaign has been launched to halt the restriction of the City Theater's premises by erecting walls around it.
To sign, take action via the link below:



***"This filthy mess... who knows when it's ever going to get cleaned up!?"***

This filthy mess that the sledgehammer-wielding minister from the previous government and his picture-frame buddies stirred up at the City Theater; who knows when it's ever going to get cleaned up!?

***"The Shahr Theater building, vibrant yesterday, ruined today! Is the #Tehran City Councel awake?"***

The [illegible] building, vibrant yesterday, has turned into ruins today!
Is the #Tehran City Council awake?
Is the #Tehran Cultural Heritage Administration awake?
Is the Cultural Commission of #Parliament awake?
What explanation exists for this #destruction, and what explanation for the silence and inaction of the Tehran Province Cultural Heritage Administration?

***"Despite widespread protests against the fencing plan... they've sealed off the entire complex to implement a plan whose details are still unclear."***

They've blocked off around Theater City for fencing!

Despite widespread protests against the fencing plan for Theater City, they've sealed off the entire complex to implement a plan whose details are still unclear.
-The images are from today. I saw them in a news group but couldn't find the photographer.

*Figure 11. People share their opinions on X (Formerly Twitter) regarding the new interventions on the City Theater square*

*Sources: "...in recent years, the sanctity of the City Theater, 2025; "Stop fencing around the City Theater...", 2023; "The Shahr Theater building, vibrant yesterday, ruined today!...", 2025; "Despite widespread protests against the fencing plan...", 2025.*

The citizens expressed their concerns to the officials and the municipality regarding the fence placement and the City Theater's renovations on social media (Figure 11). However, the reconstruction project was ongoing, and its end date remained unknown to the public. The users shared some images of redesign proposals for the square and directly addressed the matter of fence placement around the City Theater.

Even some users highlighted the artistic and cultural legacy that the City Theater has brought up since its foundation, with images of old newspapers in the 1970s announcing the famous plays being performed at the City Theater of Tehran (Check the relevant figures in Appendix). This reflects how prestigious the City Theater space is to the public, how its preservation is critical as a cultural and social heritage.



This was only one example of how the public encounters spatial interventions through digital spaces to archive, review, evaluate, value, or express, keeping themselves engaged with their city and urban life when their voices are not actively heard. This is not the City Theater's situation only; it is an issue for any other similar public space across Tehran. This brings up the question of how we can apply such citizen-driven digital inputs to the design/planning process.

## Problem as a short description

***Spatial memories are valuable insights into a space that are taken for granted by designers/decision-makers in their interventions, leading to the erasure of public space.***

# Research Objectives

In the era in which digital media has a vast impact on urban lives, it has even become an autonomous memorial archive (Haskins, 2007). Having such a large-scale, individual lived experience from citizens would help gain a better understanding of citizens' reviews of urban environments over time. Also, it's not only about opinions but about how they experienced space individually or collectively in varying temporal aspects. This is a helpful step towards developing a component for the planning and design process, which is defined by the public.

Also, within the right to the city concepts (Harvey, 2003; Lefebvre, 1968), it was discussed that memories of a space that holds dozens of lived experiences from various individuals or groups turn the space into a valuable environment, a place (Martin, 2025). Having a place is not limited to its physical existence, but also to how it's cloned in the memories of those who interacted with it. This would be the link between having the right to the city and having the right to the memory, and how to apply them in a new design approach.

Having already expressed opinions or archived memories on open, accessible platforms would aid designers and planners in studying a citizen perspective and eventually find efficient spatial strategies or design interventions based on public insights (Williamson and Ruming, 2020). The digital memory may not be the final solution, but it will be a key starting point for many urban projects that aim to be more inclusive and plural.

The objectives can be defined as follows:

### 1. Increasing public engagement through virtual spaces (social media)

Social media could be the starting point of digital participation and a digital memory base. Yet, it can be structured and flourish as an online platform in the form of a leaflet interactive map where citizens can put their emotionally valued memories about places in the city, specifically. It would be a direct and clear aggregation of citizens' active engagement in valuing and evaluating urban spaces in the past, present, and future.

### 2. Inclusive design and planning

Analyzing digital public insights could be an initial step in exploring the public opinion regarding a space or a development at the early stages of a project, proceeding to design/planning outcomes. The project can gain a bit of public legitimacy by having its indirect contribution as expressing their wishes, ideas, or experiences towards a space. It shapes a passive form of participation by providing inputs about a space that may be in the projects and contributors' focus, regardless of how democratic their approach will be.

### 3. Valuing urban space through digitally shared lived experiences and memories

There would be an organized and categorized recorded collection of either daily life experiences or present state expressions regarding the urban spaces. Citizens would share their first-hand encounters and interactions within a space by expressing how they felt, what actions they took, or what they observed at that moment. And, all of this would have a temporal component as well. It would make it simpler to trace the emotionally-valued memories through the city timeline. Overall, it will provide a basis for analyzing the spatial discourse, emotions, and their magnitude, addressing the spatial features within the site of study and intervention.

### 4. Design and planning with citizen-informed guiding values

The insights from the public regarding urban spaces would help the urbanist to locate the places of conflict, neutrality, or potential for enhancements. Having a collection of various descriptions and expressions regarding old or existing spatial features within an urban context would aid the experts in identifying the citizens' point of view, addressing the issues, neutral or positively valued spots and locations in the study area.

# Research Questions

## Main Research Question

**How can digital memories shared by citizens help identify spatial values as potential conditions for interventions in Tehran's public spaces and the City Theater square?**

## Sub-Questions

1. How do digital memories narrate the lived experiences of citizens in Tehran's public spaces on X (formerly Twitter)?

2. What emotions can be traced from digital memories of the City Theater?

3. What spatial qualities could be referred to from the shared emotions of Tehran's public spaces and City Theater?

4. What spatial values do citizens share about the City Theater's space transformation throughout time (before and after the latest intervention) as reflected on social media?

5. How can shared memories on social media inform design guides for Tehran's public spaces?

To clarify how each sub-question will be mitigated, the corresponding methodology is briefly described in Table 1. In terms of the type of analysis, questions from one to three are “Descriptive/Analytical”, and the rest are “Normative”.

| Question | Type of analysis | Methodology |
|---|---|---|
| 1. How do digital memories narrate the lived experiences of citizens in Tehran’s public spaces on X (formerly Twitter)? | Descriptive / Analytical | Literature review, Data collection using scraping apps as strings of text, Topic modelling |
| 2. What emotions can be traced from digital memories of City Theater? | | Sentiment analysis to recognize and classify emotions from texts extracted from Tweets and Instagram posts |
| 3. What spatial qualities could be referred to from the shared emotions of Tehran’s public spaces and City Theater? | | Locating the spatial features referred to by emotions through cognitive indicators |
| 4. What spatial values do citizens share about the City Theater’s space transformation throughout time (before and after the latest intervention) as reflected on social media? | Normative | Classifying the values of spatial features referred to by emotions into positive and negative |
| 5. How can shared memories on social media inform design guides for Tehran’s public spaces? | | Creating design guidelines based on spatial qualities and interventions corresponding to negative spatial values |

*Table 1. Sub-questions classification and methodology*

# Literature Review

## Right to the memory as a right to city

The Right to the City was coined by the French philosopher Henri Lefebvre in the era when social movements became viral among young intellectuals and labor activists seeking social justice in urban areas (Mitchell, 2003). After Modernism, the ownership of land in urban areas through capitalist monopoly expanded dramatically. At that time, politicians, social activists, urbanists, and philosophers realized the need for a just city and a healthy environment within the city blocks was something crucial and, in other words, could be considered a right for the citizens who live there and use its space (Habibi & Amiri, 2016; Harvey, 2015; Lefebvre, 1968).



Furthermore, the notion was developed by Harvey (2015), who argued that city dwellers can fight for their rights over space and, more generally, over the city. This act works against the manipulation and trespassing in the public domain by private developers and governments who attempt to impose capitalism across vacant spaces for profit. Harvey (2015) suggested that the public could prevent this through social activities and demonstrations by recognizing the Right to the City as their individual will to change their living space and ultimately themselves. Thus, the Right to the City is mostly interpreted as the right to public space. Who has access to it and who doesn't define this right (Harvey 2015; Habibi and Amiri 2016; Moroni & Chiodelli, 2014).

Suppose we consider the right over public space derived from the concept of the right to the city; Once people have lived experiences in a space, it forms their memories of that space. Since memory and physical space can be imagined as intertwined, memories are attached to physical space. In contrast, space holds the memories within its boundaries (notions of sense of place, etc.). Thus, having the right to a public space automatically grants the right to its memories at the same time. In other words, policy makers, designers, or any other entity that holds the authority or power to change the space must consider the sensitivity of a public space, not only based on general spatial values and qualities, but also on memories embedded in that space. It should be acknowledged that citizens have a right to space and to their memories of it as common goods, too.

Based on the same theoretical principles, Harvey (1993) has always discussed the right to the city as extending beyond mere access to resources (Harvey, 1993 as cited in Tedeschi, 2024). The abstract space described by Lefebvre (1991), limited to planning and design perspectives, has changed into a rapid representation in

various forms through digital information. The space is no longer an abstraction of a physical entity; it is regenerated and transformed into a digital domain based on rapidly changing information (Lefebvre, 1991 as cited in Shaw and Graham, 2017).

Technological development has brought new definitions of the right to the city concept. Tedeschi (2024) argues that algorithms have a right to space, as they rapidly produce and update information about physical space. This makes the accessibility and experience for its dwellers either disruptive or reinforcing, leading them to negotiate their social interactions and power over space with this non-human entity. This creates a different level of inequality in a digital dimension.

## Digital era (archival memories)

In the digital age, the media has become a tool for any individual to express their thoughts, share their feelings, raise ideas, share experiences and memories, or cry out for their rights as critical matters. One major feature of digital media is their archival potential. They are the database of memories of many individual and collective activities or events in a variety of contexts.

What archival memories can bring is not comparable to a curated, centrally organized museum set up exclusively for visitors, but to an independent, multi-voice platform that anyone with a digital device can share and publish to other users throughout the internet (Haskins, 2007). Even such an idea was suggested by Lefebvre (2003, as cited in Tedeschi, 2024): having a self-management feature on a digital platform should be a right, even though it was considered a right to urban resources rather than something higher.

The importance of the digital domain is not just about having horizontal archival memories created by numerous individuals, but about the digital space itself as a phenomenon that can trace the contents of a society in a physical space. Shaw and Graham (2017, pp.909) bring up the importance of data flows in urban studies:

***"Just as the urbanization of water is a processional notion which can uncover 'stories about the city's structure and development', so can a reading of the city's 'political, social, and economic conduits' through which information flows also "carry the potential for an improved, more just, and more equitable right to the city".***

The data flows of urban life have the potential to reflect many aspects of the city's dynamics, struggles, and synergies throughout its history. Not only the past, but also, thanks to real-time communication, the present state of the city. Solnit (2001, pp.91) also points out the historical importance of urban space, or more precisely, the streets, as "repositories of history" that are lived in. The lived experiences within these repositories can be shared and recorded on social networks as valued, recognized spaces that hold digital history.

# Public Engagement (digital participation)

From the right to the city concept, Mitchell (2003) highlights that the right to the city is an oeuvre , a piece of art that each citizen contributes to (Tedeschi 2024, pp.3). It is a production, like a tapestry, with which city dwellers can co-create by sharing their lived experiences and turning those experiences into action (Tedeschi, 2024, 6; Currie et al., 2022; Martin, 2024). Taking citizens' lived experiences and emotions into account in the planning and design process would lead to a valued space, or the so-called place (Martin, 2024). Moreover, public participation has always been better received in urban projects, regardless of their effectiveness (Martin, 2024).



Reframing the Right to the City leads to co-design from a governing perspective, which leads to shaping public space rather than having citizens gain more power and rights over their space solely (Tedeschi, 2024). This can be pivoted by autonomy in online data authorship.

One of the key significances of media in the digital era is the autonomous authorship that reflects the fact that each individual who holds access to a digital device and the open World Wide Web may be able to record and document anything that they find of their interest or desire (Haskins, 2007). Any citizen would have their way of storing and curating their daily life in the online digital sphere, making each of them an independent media broadcaster. (Esteves, 2025; Haskins, 2007; Martin, 2024; Pereira, Célia, et al., 2012).

Based on the concepts explained about archival memories, autonomous authorship would lead to democratized media (Haskins, 2007; Martin, 2024). This can be recognized as the first step towards public participation before engaging the stakeholders, mostly citizens or vulnerable groups, actively. To understand the extent of participation, Arnstein (2019) proposed the concept of "The Ladder of Participation," which describes the levels of public engagement and citizen participation spectrum as a ladder with progressively higher levels. The ladder starts with the least level as non-participation, starting at Manipulation and Therapy, and culminates in the ultimate form of full citizen participation in the design process.

The non-participation rungs of Arnstein's (2019) Manipulation and Therapy can be described concisely as follows:

- **Manipulation**

The first rung doesn't actively include the citizen's opinions but pretends to be doing so. The experts tend to manipulate the participants by giving the impression that they are engaged and can make an impact in the project, which is not the case.

- **Therapy**

The interaction with citizens, despite extensive efforts, often resembles a therapy session more than a participatory activity attempting to involve citizens with genuine involvement. At this stage, the experts tend to support their end's diagnosis, interpreted from citizen perspectives rather than the root causes of the symptoms.

Based on the participation ladder, involving citizens' shared lived experiences on digital platforms may be a bottom rung, or even considered as non-participation, yet a passively engaging position that requires improvement. Because this is not about the planner/designer's active interaction with citizens, but about reading or tracing the memories they shared autonomously on open online digital platforms (Esteves, 2025; Fathullah & Willis, 2018; Haskins, 2007; Martin, 2024; Pereira, Célia, et al., 2012).

One positive aspect of analyzing digital inputs is the large scale of data. Despite usually having a limited number of active participants in urban projects, digital platforms offer the opportunity to gather qualitative insights from a broader range of citizens in a shorter amount of time (Fathullah & Willis, 2018; Martin, 2024).

Having new definitions of social platforms accessible to the public can change the rung definitions of the participation ladder. Not being directly-engaged while the opinions and emotions are shared online and the planners attempt to take them as inputs, for example, tweet collection and analysis, it can redefine the rungs. It will be a combination of various components of participation rungs in different levels of engagement, activity and interactions.

Digital participation is also supported by various studies (Haskins, 2007; Martin, 2024; Pereira, Célia, et al., 2012; Zeile et al., 2012). Williamson and Ruming (2020) point out a further step: using social media data on a participatory scale. They suggested the term 'Frontloading', which refers to engaging stakeholders in the early stages of planning by accessing their shared emotions and experiences on social media. This is a concrete step in the analytical process. Yet, how to develop the analysis-design transition on such a basis is a question that I attempted to answer through this study.

# Research Background

To have a reliable framework validating the introduced methodology, a review of the scientific background is implemented in the study. This will help to gain a comparison of current research analysis approaches and recent approaches that are alike, i.e., sentiment analysis, crowdsourced data mapping, and data-driven design.

The research background regarding analyzing urban emotions, social media, and digital participation fields reflects a large variety of methodologies, planning, design, and strategy approaches towards the challenges.

The major methods used in the studies include using technologies such as sensory, navigation, and positioning tools, specialized participatory platforms by local governments or NGOs, image-sharing online platforms containing geotagged data, traditional civic engagement approaches (in-depth interviews, surveys, workshops, etc.) and social media as the basis or complementary analysis within some of the shortlisted studies.

Studies conclude that new technologies have shaped an evolving form of democracy that makes accessibility, clarity and equity in civic engagements provided by the digital space. However, the digital platform is not a pure cure. There are challenges to this societal innovation, too. Inaccessibility to digital platforms, misinformation spreading, hierarchical authority, and legitimacy on online networked platforms, along with data privacy and security of citizens, are concerns that should be addressed as well (Burgess et al. 2006; Esteves, 2025; Liang, 2020; Pereira et al., 2012; Williamson & Ruming, 2020; Zeile et al., 2012, 2015).

The matter is that having the social media text-based data may not be sufficient to address the spatial challenges or achieve active responses from the citizens towards planning and interventions. The current study focused on the memorial value of the public space driven by daily activities recorded and shared by citizens on social media (Burgess et al., 2006). Creation of an informal social and cultural heritage of a place would be the spatial factor in digital participation in evaluating plans and decisions over urban space (Arrigoni & Galani, 2019; Burgess et al., 2006).

The reason for choosing these research works is that they support the concepts of public engagement through emerging digital technologies that have transformed them. Social media offers new participatory platforms that ease the process of dialogue, feedback and decision-making in an interactive way between the citizens and policymakers.



*Table 2. Research background*

| .No | Research | Methodology | Results |
|---|---|---|---|
| 1 | Can social media support large-scale public participation in urban planning? The case of the #MySydney digital engagement campaign Williamson & Ruming (2020) | Qualitative analysis on quantitative data (tweets and comments classification) | The potential of social media, yet more engagement is needed from experts (decision makers) |
| 2 | Urban Emotions: Benefits and Risks in Using Human Sensory Assessment for the Extraction of Contextual Emotion Information in Urban Planning *Zeile et al. (2015)* | Human sensory, social media acquisition, Tagging, psychophysiological analysis | The interdisciplinary nature of Urban Emotions helps to receive multiple inputs about spatial interventions from human expressions. |
| 3 | Public Participation in Urban Planning: How Emotional Mapping Could Contribute to Building a Sense of Belonging *(Martin, 2025)* | Emotional mapping (surveys, in-depth interviews, workshops) | Shaping co-designed resonant urban space with communities that hold a sense of belonging as the core |
| 4 | E-Participation: Social media and the public space *Pereira, Rocha, et al. (2012)* | Review of social media activities of participation in the geospatial context | The importance and effectiveness of social media activity regarding public engagement and participatory planning |
| 5 | A New Urban Sensing and Monitoring Approach: Tagging the City with the RADAR SENSING App *Zeile et al. (2012)* | Radar sensing App | Using social media data is encouraged. However, the awareness regarding the potential and multifaceted challenges of new tools in planning is highly suggested. |
| 6 | Participatory Management for Cultural Heritage: Social Media and Chinese Urban Landscape *Liang (2020)* | Review article from public participatory projects using social media | The public and experts made constructive and effective interactions with the local government through social media platforms. |
| 7 | Reimagining Citizen Engagement in the Digital Age: Platforms, Participation, and Public Trust *Esteves (2025)* | Review of E-government, social media, crowdsourcing initiatives example projects | The high potential of digital tools and platforms for public participation and the challenges ahead of it, such as misinformation, the digital divide and data privacy |
| 8 | From Place-Memories to Active Citizenship: The Potential of Geotagged User-Generated Content for Memory Scholarship *Arrigoni & Galani (2019)* | Using Geostream to aggregate images and their metadata from Flickr, a photo-sharing platform, for analysis | Focusing on the place rather than individuals led to achieving everyday memorialization of places, shaping informal heritage in the digital sphere. |
| 9 | Citizen participation in digitised environments in Berlin: Visualizing spatial knowledge in urban planning *Singh & Christmann (2020)* | Online Platform 'meinBerlin' and e-Participation on Noise Protection (Quantitative analysis on text-based & visual data) | The Progressive tendency of visualizations using digital tools in planning processes and the significance of visual submissions of citizens addressing the perceptions, challenges and their desires for the future urban spaces. |
| 10 | Everyday Creativity as Civic Engagement: A Cultural Citizenship View of New Media *Burgess et al. (2006)* | Review of digital tools like Flickr, Urban village sharing stories, youth radio network | Significance of public everyday-life creativities taking place in digital space, which is a potential for planners to nurture society and make it flourish |

The variety of quantitative and qualitative methods used in these studies helped frame the project's methodology that aligns with the research objectives and tackles the problem of spatial interventions without citizen engagement, ending in public space erasure.

The results were mostly bound to suggestions promoting the new ways of digital participatory, strategy making, and policy design from social aspects that reflect the significance of spatial justice in a 21st century society. However, these suggestions lack practicality. Ending up with a data-driven design was a missing component among the chosen studies. They were mainly focused on introducing the relevance and necessity of digital participation backed by the new technologies. In terms of conclusions and outcomes, the studies mostly closed with general suggestions regarding revising participatory approaches in urban planning and governance.



## Theoretical Framework

All of the reviewed literature, theories and concepts helped me to shape the theoretical knowledge orienting this study. To gain an overall understanding of the main concepts derived from the literature, I structured the theoretical framework (Figure 12). The framework included the interconnectivities of concepts and ideas that co-create the theoretical core of this study, e.g., Digital memory, Right to the memory, social media, and data-driven design.

The framework represented the original theories and the corresponding adaptations. For example, the right to the memory was derived from the right to the city as an idea. To start from a physical aspect, Lynch's urban elements are the very first perceptions of the space. Then, the human and spatial components combine and the space turns into a place, followed by a sense of belonging. This attachment gets recorded and archived digitally. The ones who find themselves attached to a place express their connection, thus, their right to the place's physical and memorial properties. This shapes a passive form of public engagement in the digital era.




*Figure 12. Theoretical Framework*

# Conceptual Framework

The conceptual framework will help identify the necessary criteria for running relevant analysis on the collected text strings from the targeted platform. The conceptual framework is based on the principles derived from the literature review, considering the spatial qualities that shape how a space evokes memories and lived experiences for citizens when they interact with it. And how this can be identified and classified through analyzing textual data from social media, and then applied as values in the design phase.

The initial concept would be the transformation of space into place through people's interactions, which shape their sense of belonging to their environment based on several spatial and contextual factors (Martin, 2025).

The belonging factors that develop the space to place concept in Figure 13, represent a very initial draft of the spatial and contextual aspects that would have a direct and indirect impact on one's sense of belonging to the environment in which they have been and interacted with.

- **Ecology:** green & blue elements, living species
- **Accessibility:** sidewalk, pathways, entrances, gateways
- **Openness** open space, tiling, profile details
- **Sitting space:** edges, fountain, benches
- **Landmarks:** City Theater building, statues, and a fountain
- **Visibility:** Eye view sheds, landmarks, corridors,
- **Safety:** pedestrian priority, traffic control
- **Security:** active presence, observer eye (eyes on the street)
- **Geography:** topography, location
- **Climate:** temperature, humidity, sun angle
- **Identity:** symbols, culture, language, lifestyle

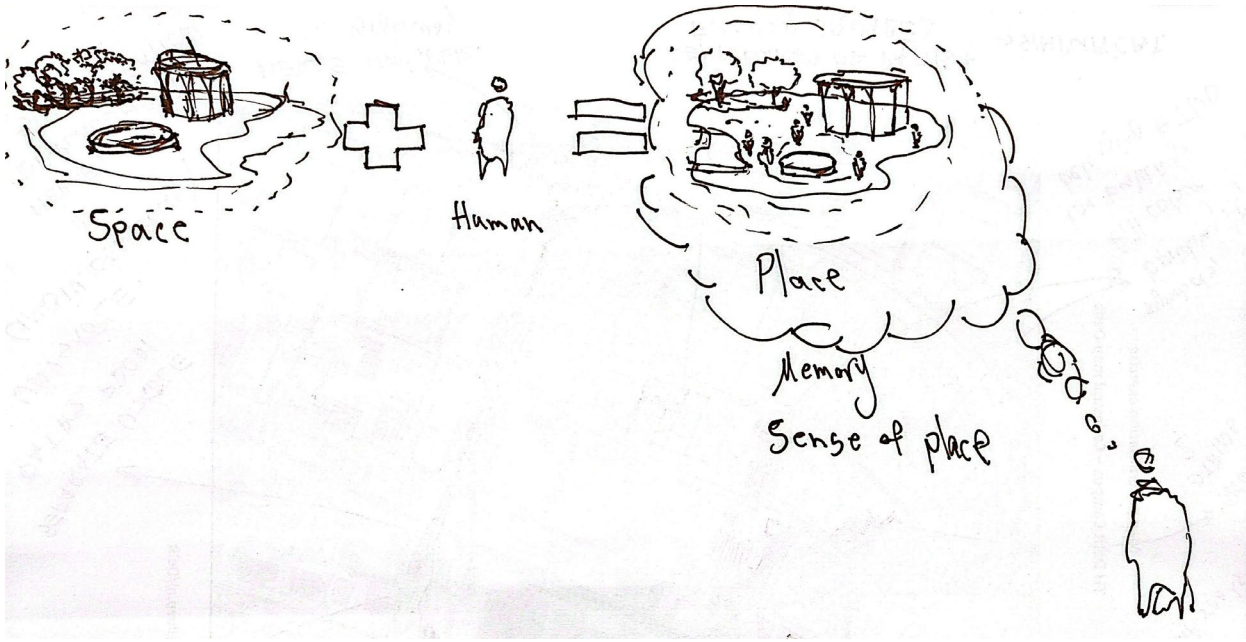



*Figure 13. Space to place transformation*



The memories of space or place that are turned into lived experience and expressed through emotions in digital space are useful for analyzing a space from a cognitive perspective.

Users share different experiences of their presence and interactions within the lived space on social media. As mentioned before, their posts and tweets would hold various emotions (Haskins, 2007). Based on the research background of relevant studies, the conceived gap is how to trace the spatial feature from the shared emotions. I considered addressing the spatial features associated with emotions embedded in tweets to tackle the intervention challenges as an outcome of non-participatory planning/design approaches. Moreover, this may also result in a data-driven design approach to draw and make spatial qualities out of quantified emotions.

Thanks to Lynch's (1964) cognitive analysis concepts and methods of the perceived space or city, the spatial features could be tracked and recognized by how the individuals described their situation or position in the space. And, their mentions could be classified into Lynch's cognitive features of nodes, lines, areas, edges, and landmarks. The subsequent workflow would focus on relating different sentiments to their corresponding spatial features in the context. For example, as shown in Figure 14, someone could have gotten angry due to the presence of recently installed walls at the entrance of an area in which they always used to walk by and enter the space through the former gateway.

The sentiment classification of each feature would help identify valuable places and how they were designed, with a corresponding design for the public that results in livability and placemaking. Also, spaces that need improvement or have no specific mitigations needed will be identified. Based on the feelings that may address a specific feature within, the association between physical space and emotions in memory will result in positive, negative, or neutral values. This would help with further steps of the design process and with developing spatial conditions for spaces that hold public-oriented values that can be recalled as a public good.



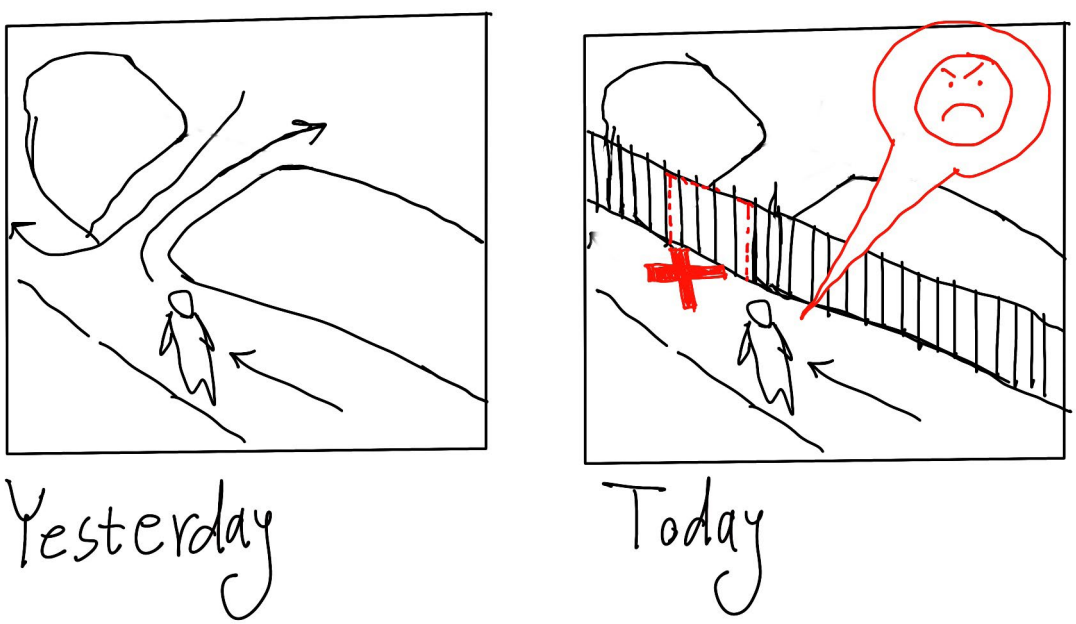



*Figure 14. Emotions emergence followed by spatial interventions*

# Framework

Furthermore, to integrate the adapted concepts from the literature review, as a guiding theme for the research and design process, a conceptual framework was illustrated in the following diagram.

The diagram in Figure 15 depicts the main concepts of the study, social media, digital memory, and space, and how they are related to each other based on the defined methods and processes.

# Definitions and Processes

## Space

The existing physical space that, in this case, can be considered as the public space, or to be precise, City Theater and Tehran's public spaces, which hold various activities, spatial elements, and mobility within. Users of social media interact with space. The result of this interaction is expressed in another sphere by them.

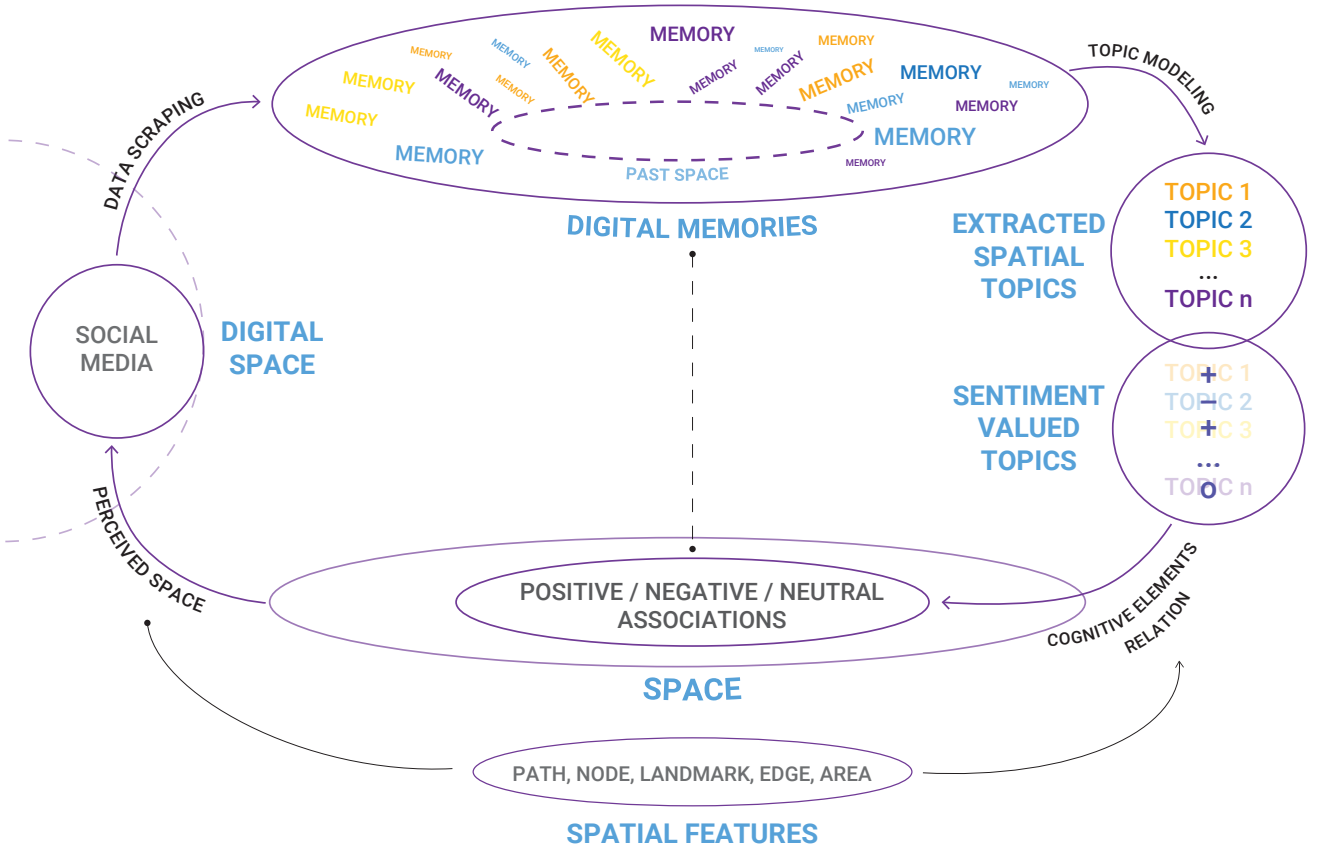



*Figure 15. Conceptual Framework*

## Social Media

Social media is a fragment of a larger sphere, the digital space, that includes many other domains with various activities, characteristics, and users. The study's focus is on a social media platform (i.e., X) where users share their expressions, opinions, attachments, and, in general, their lived experience of the space as a memory. These memories are shared as Tweets with footage, images, or plain descriptions. These data will be scraped from the platform as text strings (only the textual contents) that address the space with the keywords of City Theater area and similar public spaces in Tehran.

## Digital Memories

The scraped and collected data are literally the memories of the space in the past (or, as it is called, the past space). There are both spatial and non-spatial memories that will be filtered. Any of the spatial memories refers to space. But that's not all the input. Each of them holds a specific expression, for it is a personal experience. The various memories can be classified by how they were expressed, along with one or more emotions.

## Extracted Spatial Topics and Emotions

The emotions and descriptions within digital memories can be extracted and classified using topic modeling and sentiment analysis. The topic modeling will help to understand what the memories are referring to or describing. The group of memories with similar discourse will be grouped into a single topic. The emotions within the grouped memories will be identified from textual expressions. Each word or phrase addresses a feeling or experience that can be classified into the following items:

1. **Anger**
2. **Sadness**
3. **Comfort**
4. **Joy**

Having these classes, emotions can be categorized as negative or positive feelings. However, there can also be expressions with no specific sentimental classification that can be classified into neutral (Table 3).

| Emotion Categories | Value Classification |
|---|---|
| Anger | Negative |
| Sadness | Negative |
| Comfort | Positive |
| Joy | Positive |

*Table 3. Categories and Classifications adapted from Ekman's basic emotions*
*Source: Ekman, 1999.*

Spatial Features: Based on Lynch's (1964) perceptual urban elements, each shared emotion would have one or more corresponding spatial features that affected the person's cognition and co-created their lived experience. Based on Figure 16 features are as follows:

1. **Path**
2. **Node**
3. **Edge**
4. **District (areas with common character)**
5. **Landmark**

The urban elements are addressed, directly or indirectly, by users in their digital memories. This addressing comes with an evaluation of their perceptions of the spatial features. Based on this and the extracted emotions, the negative, neutral, and positive ratings can be linked to the corresponding urban elements recognized mentally by the valuers. This relational process helps create positive, negative, or neutral associations within space.

**Positive/Negative/Neutral Associations**

The associations made through the relational process will reveal spatial features with negative, positive and neutral values. Features with negative ratings require improvements and interventions. While the ones with positive ratings might have design qualities to be recognized, preserved, or reused through interventions within the space. Or, they may be left untouched for having neutral ratings.

Overall, as shown in Figure 15, the cycle can continue in the original space after the interventions based on positive/negative/neutral associations are applied. The transformed spatial features within the space can make an impression on their valuers, shaping their lived experience of the space again. Therefore, as impressions accumulate, a sense of place builds as it moves through space and its elements. The impressions emerge in different forms, such as Tweets, regarding the physical space sphere. In that way, the impressions of the physical space are transferred to the digital space (i.e., social media) and create a clone of the existing space.

The past, the transformation, and the existing space are not shaped only mentally. But also, that mental recognition of space is represented in social media, shared within the digital space.



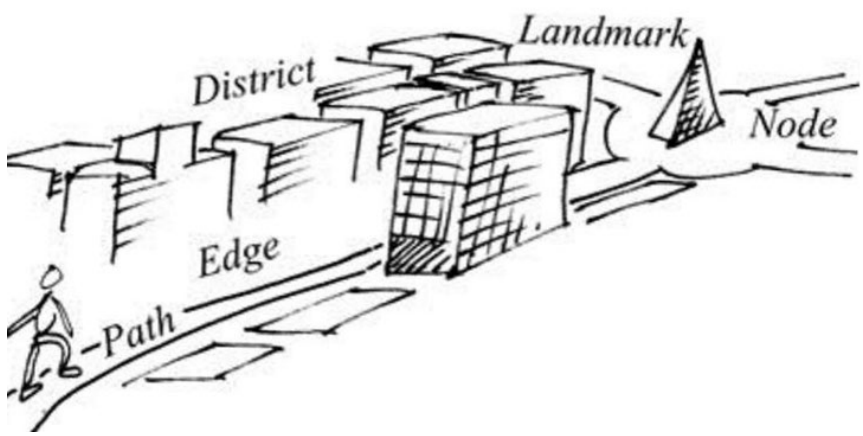



*Figure 16. Lynch's cognitive urban elements from 'The Image of the City'.*

*Source: Can, 2019.*

# Expected Outcomes

The expected outcomes are either the results of data analysis, spatial design guidelines, or interventions on the case study (Figure 17), also known as City Theater Square.

The expected outcomes (Figure 17) would be categorized into results of topic modeling and sentiment analysis of text strings collected from tweets, serving as an archive of emotions, along with their linked spatial features, matched based on cognitive indicators of urban space perception. This leads to valuing spatial features based on emotions as either negative, positive, or neutral elements within the study scope. These values will be localized based on their in-text descriptions of the space in the form of hand drawings, map drawings and street profiles. The drawings would be examples of how to link the analysis into the design phase based on a proposed design framework.

On the spatial category, the outcomes are followed up on the analysis of processed data that shapes the guiding values. Having the guiding values would be the first step in locating and interpreting the challenges, creating concise, straightforward mitigations. These can be applied to the area that serves as a policy or design intervention. After that, some detailed interventions, following the structured framework, will be demonstrated as examples. Using hand drawings, plans, sections, or bird-eye-view drawings would represent the possible outcome based on the given framework.



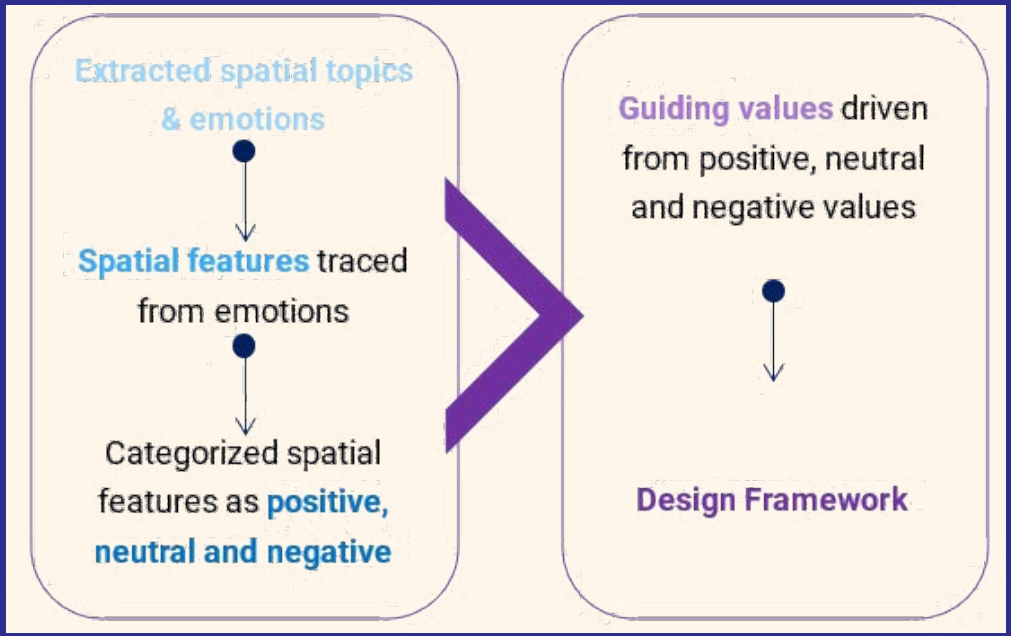



*Figure 17. Expected Outcomes*

# Scope

## Scale

To gain a better understanding of case studies followed by the conceptual framework, the analysis will be applied on two scales:

1. **The City Theater is the focus of this study**

2. **The public spaces with the same popularity and characteristics as the City Theater within the city of Tehran**

This helped to understand how citizens recognize public spaces both on a city scale and on a public space scale. The discussions and emotions regarding the existing public spaces were collected and processed as text-based data from the X platform. This double scale accompanied a temporal factor as well.

## Timeline

A timeline was set for analysis on both scales of studying what users talked about on social media and what their expressions were toward them (Figure 18). The keyframe of the timeline was based on the date on which the latest intervention was commenced at the City Theater by placing the construction walls around the square in 2024. This was considered the basis for the spatial transformation of the focus of the study area. A two-year buffer was set as before and after timelines to study the topics› shift and the change of emotional values towards the City Theater and its peers across Tehran.



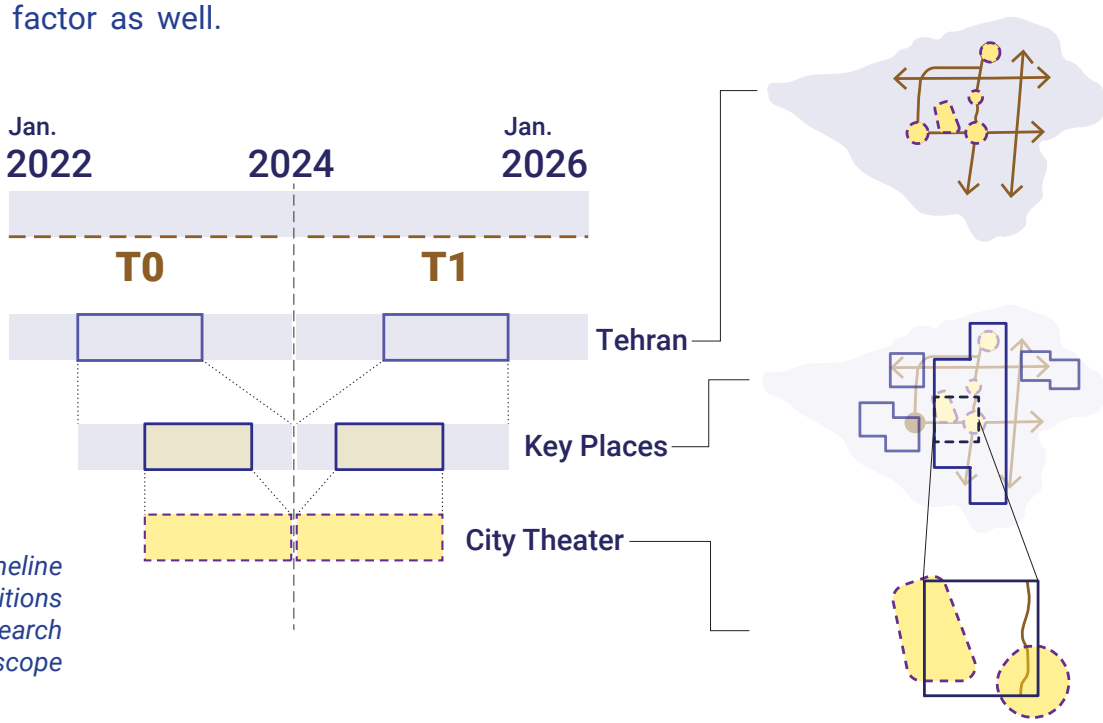



*Figure 18. Timeline and scale definitions of the research scope*

Another reason for choosing a two-year buffer was to make the search scope sufficient and efficient. Due to limitations in scraping the tweets from X platform using Apify, it was better to focus on a short time span, as the actors would operate efficiently, collecting as much data as possible within the given range. Otherwise, the actor would only collect a selected number of tweets each year so that it doesn't exceed the limit specified in the query.

## 1st Scale: City Theater

The reason for choosing this specific public space was not only its architectural quality, but also its historical legacy, which holds many urban events, activities, and daily life patterns. A public space at the heart of the Iranian capital, Tehran, has always been a focal point for its dwellers (Figure 19 & 20).

Thus, it holds a large scale of digital memories within itself to collect and study. Also, because there was no explicit online platform for citizens to share their digital memories of a specific urban space, the current study focused on the most popular and recognized public space. This increased the likelihood of relevant data availability and collectability.

## 2nd Scale: key public spaces across Tehran

The public spaces that had the same popularity and recognition as the City Theater also had high potential of being mentioned, discussed, and expressed on social media. Also, in terms of space form and functions, the places identical to the focus of this study were highlighted across Tehran[1]. These spaces were identifiable as Lynch's urban elements as well. The places that are either a path, an area or a node that have a spatial proximity to each other within the urban context were mapped. The following map depicts the key public spaces in Tehran (Figure 21). Moreover, the key places were briefly introduced in three smaller zones reflecting the typological similarities between the key places and the focused area, i.e., City Theater (Figure 22, 23 & 24).



1. A quick note regarding the square notion in the Iranian urban context should be mentioned. Speaking of which, the places entitled "square" in Tehran are technically roundabouts or traffic junctions that mainly serve for mobility with activities surrounding the center. The pedestrians are not the priority and can only occupy the sidewalk spaces around. The only locations that have both morphological and functional similarities to known open space squares are City Theater and Sabzeh Meydan Plaza.




*Figure 19. Tehran's overview and City Theater's location within*

*Source: Google Earth, 2024.*



*Figure 20. Current state of the City Theater after removing the reconstruction dividing walls*

*Source: Fatemeh Fallah, 2026.*



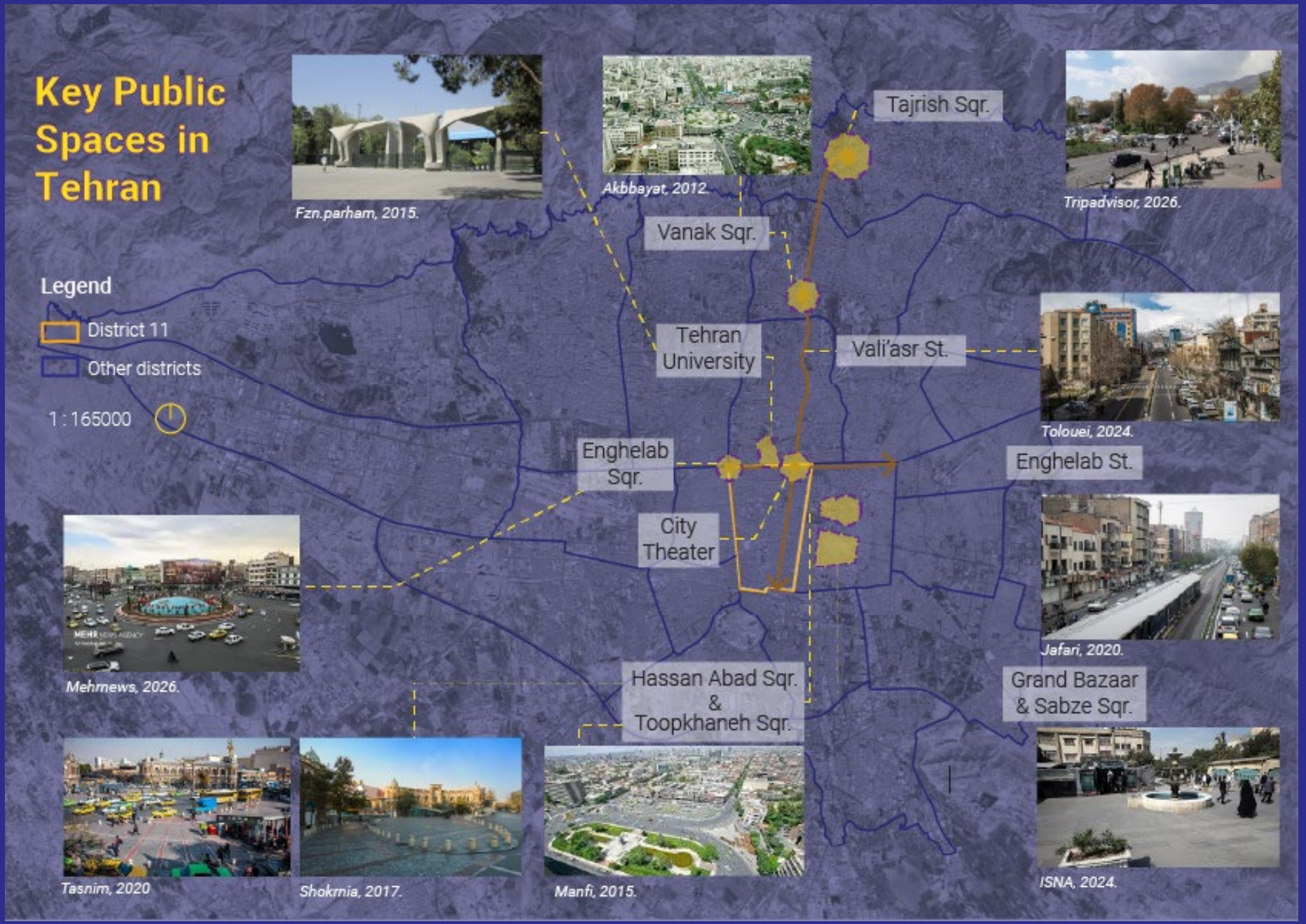



*Figure 21. Key public spaces of Tehran.*

# Key Public Spaces in Tehran

*Tajrish, Vali'asr & Vanak*

## Tajrish Sqr.



*Google Earth, 2025.*

**Tajrish** square is one of the most famous squares in Tehran that not only operates as a traffic node, but also, it's a socio-economic focal point in the city. The square was built upon the old mosque as a pilgrimage location and Tajrish's market, also known as Tajrish's Bazaar. There's one of the head stations of Tehran's metro network as well which connects the south to the northern part of the city.

## Vali'asr St.

**Vali'asr** is the longest North-South street in the country that starts from Tajrish sqaure in the north and ends to the railway station in the south, a true arterial street. Speaking of which, the street holds diverse types of activities and spatial typologies that vary from spot to spot along its way that creates a multifaceted identity of it as an urban space that can be considered a public space.



*Google Earth, 2025.*



*Shidfar, 2024.*



*Tripadvisor, 2026.*



*Wikipedia, 2026.*

## Vanak Sqr.



*Google Earth, 2025.*

**Vanak** square is the other important square that acts as the middle node of North-South main mobility network. This square can be considered a transportation hub for bus, BRT and pool taxis, along with mixed-use activities on its edges that have turned the square into a business and retail district through decades.



*Akbbayat, 2012.*

*Figure 22. Key Public Places in Tehran, 1st zone.*

# Key Public Spaces in Tehran

## *Vali'asr & City Theater area*



*Tolouei, 2024.*

### Vali'asr St.



*Google Earth, 2025.*

Pass the Vanak square, the street is divided by highway into northern and southern segments. The southern **Vali'asr** street leads to the city center. The vegetation density and slope descends. More mixed activities emerge and so do more urban complexities.



*Jafari, 2020.*




*Mehrnews, 2026.*

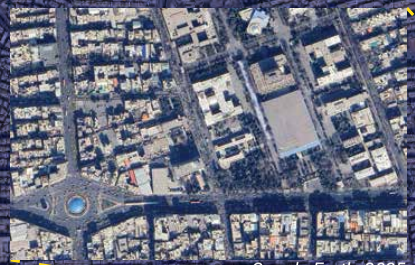
**Enghelab** square is one of the most significant roundabouts in the city center. This place is in which concentrates the transportation, social, cultural and educational activities in a single node.



*Fzn.parham, 2015.*

**Tehran University** is one of the largest education/research hubs at the heart of the city. The place where a large number of citizens, including students, professors and researchers would commute to and spend time during weekdays. The main entrance of UT is one of Tehran's iconic landmarks and spaces along the Enghelab street.

### Enghelab Sqr. & University of Tehran



*Google Earth, 2025.*

The head of the Enghelab square is stretched by the **Enghelab** street that connects the other important spaces attached to one segment from western to eastern side of the city. Places like the Enghelab square, University of Tehran, Vali'asr intersection, City Theater square and other important places towards the east of Tehran.

### City Theater



*Google Earth, 2025.*

### Enghelab St.



*Google Earth, 2025.*

*Figure 23. Key Public Places in Tehran, 2nd zone.*

# Key Public Spaces in Tehran

*Hassan Abad, Toopkhaneh and Bazaar*

## Toopkhaneh Sqr.



*Google Earth, 2025.*

The **Toopkhaneh** square is one of the oldest squares located at the edge of Tehran's historical district by being one of the main gateways leading to Tehran's grand bazaar. This square is also a retail center and public transportation node approximating the political centers as well.



*Safdarian, 2015.*



*Manfi, 2015.*

## Hassan Abad Sqr.



*Google Earth, 2025.*



*Shokrnia, 2017.*

**Hassan Abad** square is famous for its buildings with curved buildings with European style facades facing the square along with the traffic limitations, making the space look more like a plaza than a node.

## Grand Bazaar

Tehran's **grand bazaar** is one of the largest components of Tehran's spatial organization. It operates as an economic hub while holding a significant cultural and historical core narrative of the old Tehran.

## Sabzeh Meydan Plaza



*Google Earth, 2025.*



*Manfi, 2015.*



*Wikimedia, 2005.*



*Tasnim, 2020*

Within Tehran's grand bazaar, there lies one of the open squares in the historical business district of the city. **Sabzeh Meydan** can be considered as a plaza due to its small scale and having less importance compared to the main square yet having the spatial characteristics of a public open space within the grand bazaar area.



*ISNA, 2024.*

*Figure 24. Key Public Places in Tehran, 3rd zone.*

## X (Formerly Twitter)

Multiple social media platforms hold vast amounts of data about digital memories. In the Iranian case, users are mostly active on the two platforms, X[1] (formerly Twitter) and Meta's Instagram. The X has always been one of the most dynamic platforms for Iranians to share their opinions on the state of the country and their daily lives. Instagram is the most popular social media platform among Iranians, with a 28.33% share, and X holds 5.4% over the last decade (Dec 2015 - Dec 2025) (statcounter, 2015). X may not be the most popular platform, but it's where many users share their opinions on critical or general topics more straightforwardly and transparently. While other platforms like Instagram, Facebook, or Pinterest focus on sharing media and general descriptions, X would offer more textual content that conveys a stronger sense of individual or professional positionality regarding a topic.



Technically speaking, scraping is much easier to perform on textual data. After running a test scraping on both Instagram and X platforms, there were more relevant data were collected within the corpus compared to Instagram captions. The complexity with Instagram data was that I had to first locate the relevant posts out of a larger corpus. Then, the comments section of the relevant post had to be scraped using a different actor specialized for comment scraping. Due to time limits, this was neither convenient nor efficient.

Based on these reasons, I picked X as the target social media platform to collect the digital memories of general and local spatial discourses and to analyze the texts in tweets and mentions.

1. Since the mentioned social media platforms are restricted in Iran, there aren't precise data or information regarding the user accounts in Iran. The current descriptions are based on the author's experiences who was born and raised in Iran using the mentioned social media platforms that are popular among the existing internet users who access them via VPNs that change their IP addresses in Iran.

# Methodology

The conceptual framework depicted the goals of this study to structure design guides based on shared emotions on social media regarding the City Theater and Tehran's public space transformations. In this section, the "How" factor is discussed. I, as the responsible researcher, indicated the methods, approaches, tools, and research scope and elaborated on each part.

For the analysis, I collected the data by scraping texts from tweets using Apify's scraping actors, which are introduced in the following section. The scraping happened within two timeframes and two scales for comparison purposes.

The timeline focuses on periods before and after the recent intervention on the City Theater in 2024. To obtain sufficient data, I considered a two-year buffer based on the key time of intervention. The data was collected from 2022 to 2024 and from 2024 to 2026 in two defined scales of study.

The scales helped to make a comparison of emotions towards the intervention that happened at the City Theater and how its relation with other key public spaces was affected over time. First, I collected the tweets referring to City Theater before and after the interventions. Next, on a larger scale, I scraped the tweets addressing the other similar key public spaces across Tehran through the same timeline as the City Theater's to have expressions towards each of these places in similar time periods.

The main target was to collect data from tweets shared by public profiles on X (formerly Twitter). The analysis focused on the latest interventions in City Theater Square. There were pre-intervention (2022 - 2024) and post-intervention (2024 - 2026) phases in which the public shared their opinions, expressions, and impressions as lived experiences, both when the square was intact and when it was transformed through redesign procedures over the past two years of a time-stretched project (Aleksandar et al., 2025).

The collected variables are the text strings within tweets on X using a scraping actor specialized for this platform[1]. Which aligned with the legal and regulatory procedures of online scraping. The data collection followed the specified time frame mentioned previously and used defined keywords to narrow the search query as specifically as possible. Names like "Teatré Shahr", "Teatré Shahr square", or "Vali'asr intersection" were used as keywords to identify and collect any related text data to the chosen case study (Williamson and Ruming, 2020).

1. Tweet scraper unlimited for Twitter

Once the data collection was completed, the data needed to be cleaned of unnecessary details and irrelevant expressions so the analysis could be applied appropriately. The unnecessary details include personal information and opinions on highly sensitive topics that might cause risk to data subjects. The necessary data can include sentiments, spatial features, and memorial expressions related to the space (e.g., City Theater and its surroundings), all in string format. After that, the cleaned data was classified based on topics using the topic model. Then, the tweet texts received sentiment values using the sentiment analysis method. The sentiment values that were based on the defined components of the conceptual framework, as a qualitative content analysis (Williamson & Ruming, 2020; Zeile et al., 2015). By overlapping the driven criteria with sentimentally scored data, the spatial qualities and design guidelines were created, framing an overview of what was valuable and should be preserved, and what might be the alternatives of design for such public spaces. All of the syntheses were applied to developing the design framework.

To learn more about the methodologies and find the best approaches and tools in topic modeling and sentiment analysis, a thorough review of research projects and papers focused on the study's methodology was conducted. The synthesized intakes were discussed in the following sections. A table of methodology and approaches was structured to provide an overview of the literature in a concise way. Check the Appendix 2 to view the table.



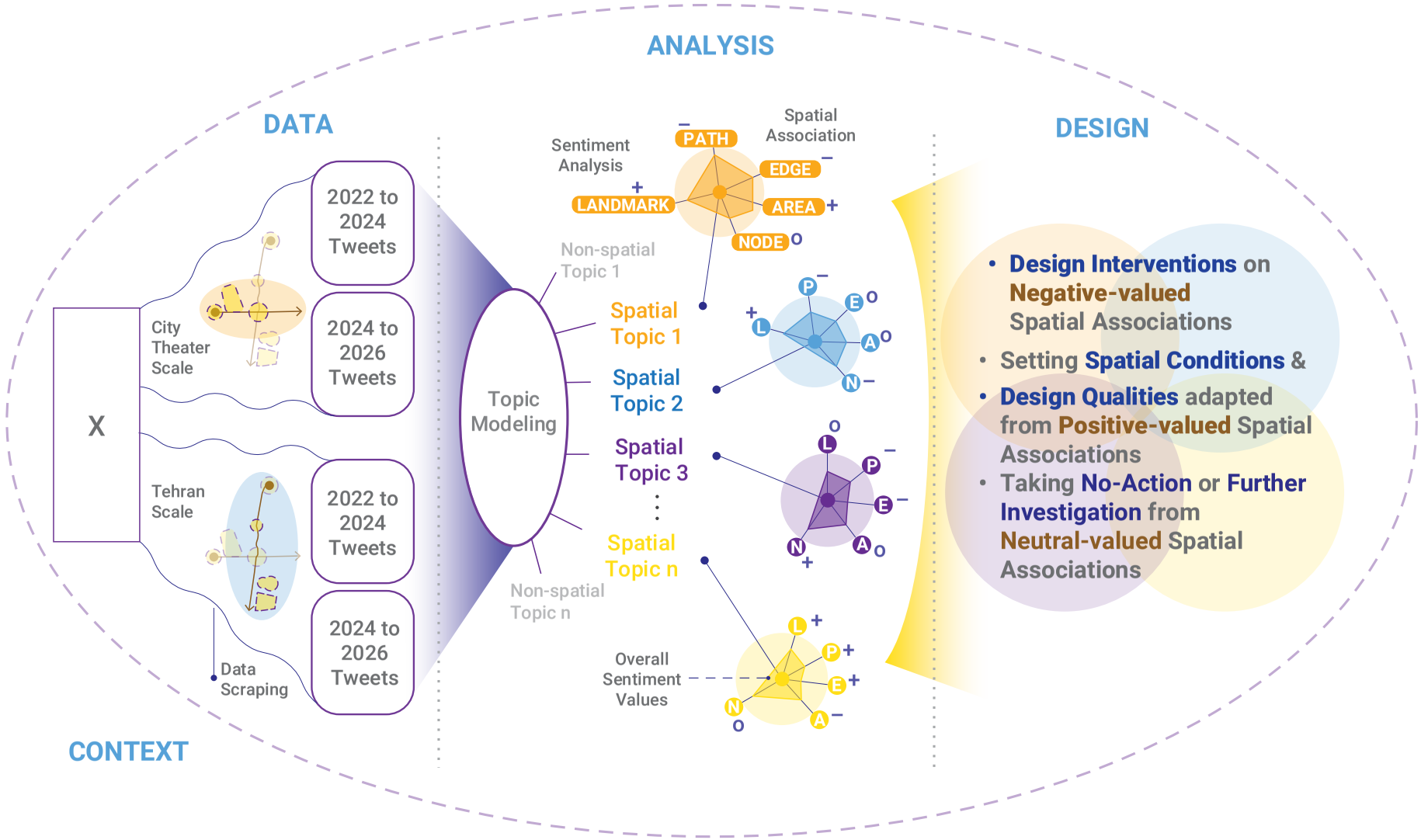
# Methodological Framework

Based on the above-mentioned concepts and theoretical backgrounds, the conceptual framework was structured, serving as a theoretical lens for the methodology and subsequent steps of this study. The details and connections of each notion are illustrated in the following figure.

The process based on the illustrated framework (Figure 25) is divided into three steps: data, analysis and design. Starting with data, the text-based data in Tweets was rooted in the virtual digital space and could be traced back to digital memories. That is the inner sphere, which holds the lived experiences within those tweets. There were different topics mentioned by users in the corpus of tweets. But not all of them were spatial. I decided to filter the tweets that express daily life experiences while referring to spatial features within the context, i.e., City Theater and similar public spaces in Tehran. To do so, the data had to be categorized into groups based on their topics.

The analysis stage started here, where I needed to filter the tweets based on their topics that address the spatial features. The topic model method was used to categorize the tweets based on the most relevant topics. Then, the spatial topics were picked before analyzing the emotions within them.

*Figure 25. Methodological Framework.*


ANALYSIS
DATA
DESIGN
CONTEXT
X
City Theater Scale
Tehran Scale
2022 to 2024 Tweets
2024 to 2026 Tweets
2022 to 2024 Tweets
2024 to 2026 Tweets
Data Scraping
Topic Modeling
Non-spatial Topic 1
Spatial Topic 1
Spatial Topic 2
Spatial Topic 3
Spatial Topic n
Non-spatial Topic n
Sentiment Analysis
Spatial Association
PATH
EDGE
AREA
NODE
LANDMARK
Overall Sentiment Values
• Design Interventions on Negative-valued Spatial Associations
• Setting Spatial Conditions &
• Design Qualities adapted from Positive-valued Spatial Associations
• Taking No-Action or Further Investigation from Neutral-valued Spatial Associations

To specify the emotions within a tweet, a sentiment analysis was needed. Four main emotions have negative, positive, neutral, or a combination of all. The sentiment analysis helped me to identify the emotional values of each tweet. The advantage of filtering the spatial topics was that I could trace the valued sentiments to the spatial feature. Each emotion referred to spatial features.

For simplification, the emotionally valued spatial features were categorized using cognitive approaches derived from Lynch's (1964) urban elements. Space is perceived as simple objects such as lines, nodes, and landmarks. Not only do the emotions address each spatial feature, but they also value it. There are physical elements that may have negative effects (Angry or Sad) or positive effects (Happy or Comfortable) or no effects at all on people experiencing their interaction with space, and they rate them by expressing emotions stored as digital memories formed in Tweets.



After associating each value with its corresponding physical elements within the context and timeline, the design stage began. The first approach was to indicate spaces that mostly held negative sentiment values. Emotions like anger or sadness towards a spatial element could be considered as a space that did not provide a welcoming space properly. Therefore, spatial improvements or interventions were required.

The second approach was focused on places that evoked happiness or comfort among social media users in X, expressing their thoughts about City Theater before and after the interventions. Those spots contribute to the space's livability. They were considered valuable spaces, while their spatial characteristics were processed and applied as design qualities in spaces that did not possess the same value within the study scope. All this process happened within its context, and each step had interrelations with the public spaces chosen as the study area.

There were also spots within the study area that held mixed feelings or received neutral values. In that case, the no-action or further investigation approach was applied as the public opinion was not clear or not straightforward enough.

I proposed spatial conditions, interventions, or no actions, based on planning/design principles, technical aspects, or systemic considerations, that could be applied to cases with such complexities.

Overall, based on the proposed conceptual framework, through all steps, the relationship was circular and interactive rather than being one-way. The intakes can be updated on a long-term basis as new topics may emerge on social media and the planner/designer would receive various sentimental values that may change the design qualities, spatial conditions and potential interventions.

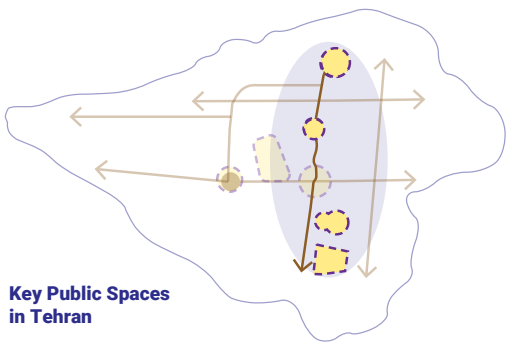
*Figure 26. Topic-sentiment co-occurrence matrix in two timelines and Tehran's scale*

# Analytical Process

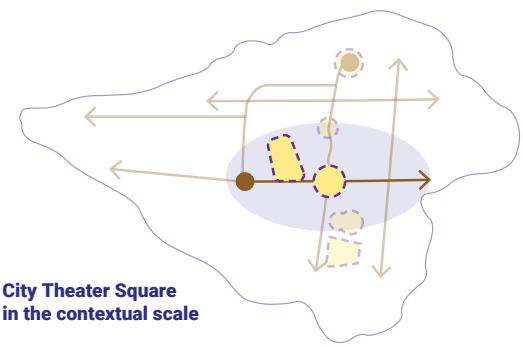
## Co-occurrence Matrix

Once the data was processed and the topic modeling and sentiment analysis were applied to it, they were aggregated into a matrix of topics and sentiment values to show the possible relations between topics and their relevant sentiment values of the study area and its peers across Tehran.

There were two matrices considered for timelines before and after the intervention. Based on that, each scale would have its double matrix, resulting in four tables of the co-occurrence, which highlighted the topic emergence and sentiment shifts within Tehran (Figure 26) and City Theater's scales.



**Key Public Spaces in Tehran**



**City Theater Square in the contextual scale**



Processed text-based data (Tweets)

| Spatial Topics \ Sentiment Values per Topic | Negative − | Neutral 0 | Positive + |
|---|---|---|---|
| Topic 1 | | | |
| Topic 2 | | | |
| Topic 3 | | | |
| Topic 4 | | | |
| .... | | | |
| Topic n | | | |

**T0: 2022-2024**

Processed text-based data (Tweets)

| Spatial Topics \ Sentiment Values per Topic | Negative − | Neutral 0 | Positive + |
|---|---|---|---|
| Topic 1 | | | |
| Topic 2 | | | |
| Topic 3 | | | |
| Topic 4 | | | |
| .... | | | |
| Topic n | | | |

**T0 & T1: 2024-2026**

## Topic Modeling

Topic modeling was used to quantify the topics' occurrences in documents as an unsupervised machine learning technique. With this method, it was possible to monitor the probability of co-occurrence of words in a corpus of documents, which in this case were the tweets (AlAgha, 2021). The topics were picked based on the frequency of terms that appeared per document that proportionally formed the topics (Alvarez-Melis & Saveski, 2016 as cited in AlAgha, 2021).

One of the common algorithms used in topic modeling is the Latent Dirichlet Allocation (LDA), to train the topic model. In general, the LDA takes bags of words from the documents and a vocabulary set, along with the k as a scalar value to determine the number of topics that the machine should learn (AlAgha, 2021; Blei et al., 2003; Kovács-Győri et al., 2018). This would give freedom to explore different scopes of learning to find the best proportion of topics that make sense.



The topic model on social media data helps to automate the process of finding the major topics mentioned or discussed by users based on the described algorithm. It helps to identify the major opinions regarding the case study to understand the public recognition of the study scale.

Despite the straightforward workflow in working with the LDA algorithm, I faced some technical issues regarding the output topics based on translated data (from Farsi to English). The topics in each group of topics were not very relevant and hard to comprehend. The lack of coherence would make it difficult to even choose a name for each topic group. The complexity was in the top terms in each topic that were not quite spatial and difficult to interpret the word chunks into a general name or topic. That was a result of using the translated data for the LDA algorithm. The nuance and coherence were lost through the translation process as well. Moreover, I realized the LDA model was not available to interpret the contextuality of the Farsi language in general, and in a translated format, it gets downgraded even more. Based on that, I decided to run the analysis on the Farsi data instead. Therefore, a model developed for advanced Natural Language Processing (NLP) of Farsi text-based data was needed.

## BERTopic

The BERT model is a newer transformer-based method for NLP that overcomes the limits in LDA model, which couldn't interpret the nuance of text-based data. Also, the BERTopic has the advantage of interpreting the context of the text, which is a significant item in the Farsi language and tweets. Moreover, the BERT package has the density-based clustering methods that help to categorize topics based on real-world data on a large scale (BERTopic Team, 2025). The topic modeling done by BERTopic is ahead of its peers, for it also runs word embedding.

## ParsBERT

The ParsBERT model is a subset of BERT as the base model, which was developed to process data in the Farsi language specifically. The transformer-based model for NLP in Farsi outperformed its peers even in the multilingual model of BERT itself, making ParsBERT the best fit for monolingual topic modeling and sentiment analysis. ParsBERT even outperformed the BERT multilingual models through assessments of language processing (Farahani et al., 2021). Also, what the developer group[1] added to this model package was to create a vocabulary set based on reviews of users on local websites for food delivery[2] and online shopping[3] as a provided context for the machine to interpret slang, sarcasm, and nuances in Farsi. The model is pre-trained based on a large corpus of text-based Farsi data, approximately 2 million documents, from various subjects (Farahani et al., 2021; HooshvareLab, 2020).

The model is developed as a Python library, and the complete package is accessible through the Huggingface website as an open-source platform open to developers and researchers. Another advantage of ParsBERT is its context-based analytical method, also known as word embedding.

Another advantage of the BERT model is that it is compatible with emoticons and emojis within the data that skips the emoji conversion through the data normalization steps. The model provides outputs as follows:

- **Final topic number (best fitting topic allocation) per tweet**
- **Tweet count per topic**
- **Topic general name**
- **Top terms per topic**
- **Probability weight per document, i.e., tweet (0.0 to 1.0)**
- **Representative tweets per topic**

The model also has the advantage of automatically defining the threshold for the topic numbers and adjusting the topics based on how close a document is to a topic. The model handles the UMAP[4] and HDBSCAN[5] for the dimension reduction and clustering, respectively.



1. Hooshvare Lab Group
2. Snappfood: An Iranian online food delivery platform
3. Digikala: An Iranian online shopping and delivery platform
4. Uniform Manifold Approximation and Projection
5. Hierarchical Density-Based Spatial Clustering of Applications with Noise

## Word Embedding

The word embedding is a method that doesn't count the terms in a tokenized format, but interprets each word regarding its context. The model goes through the terms and processes them based on the phrase in which they exist. This would give a better understanding of the genuine meaning and value of each term inside a document. Especially, having the text-based data as tweets, and considering that each data document is a single tweet, the model analyzes the terms in their own tweet context.

The word embedding is an alternative to LDA among NLP techniques. It turns the words into numerical vectors in a multi-dimensional space that helps the machine to interpret word densities within the corpus and link the terms among which are semantically relevant. Based on this, the model tries to position the words that have similar meanings near each other (IBM, 2023).



## Sentiment Analysis

Sentiment analysis utilizes the data to identify the emotions beneath each tweet or string of text by valuing the overall sentence based on the words that hold the emotions. The emotions can be negative, positive, or mixed in varying proportions. Based on Ekman's basic emotions (see Figure 27), feelings like joy or beauty, and grief or anger can be categorized as positive and negative, respectively, or even mixed (Ekman, 1999; Roberts et al., 2018).

The common and highly recognized model specifically developed for social media sentiment analysis is the Valence Aware Dictionary and Sentiment Reasoner (VADER), which overtakes its peers (Hutto & Gilbert, 2014, as cited in AlAgha, 2021). The VADER creates a sentiment score range starting from -1 to +1, where indicate the most negative to most positive sentiment value, respectively (AlAgha, 2021). However, due to constraints in translating the tweets into English, the VADER was not used. Instead, the ParsBERT-Sentiment model provided sentiment analysis within its module for NLP in Farsi. The model's outputs are as follows:

- **Sentiment label (positive, negative, neutral)**
- **Sentiment score (0.0 to 1.0)**

The ParsBERT-Sentiment model outputs are categorized as sentiment labels, including positive, negative and neutral emotions, along with a numerical score representing the emotional magnitude. The score varies between 0.0 and 1.0, showing the probability confidence, given by the machine, of the extent to which an emotion is present per tweet. For example, a tweet was labeled as positive and the score told us how intense that feeling was within the tweet.

*Figure 28. Data collection, preprocessing and analysis*

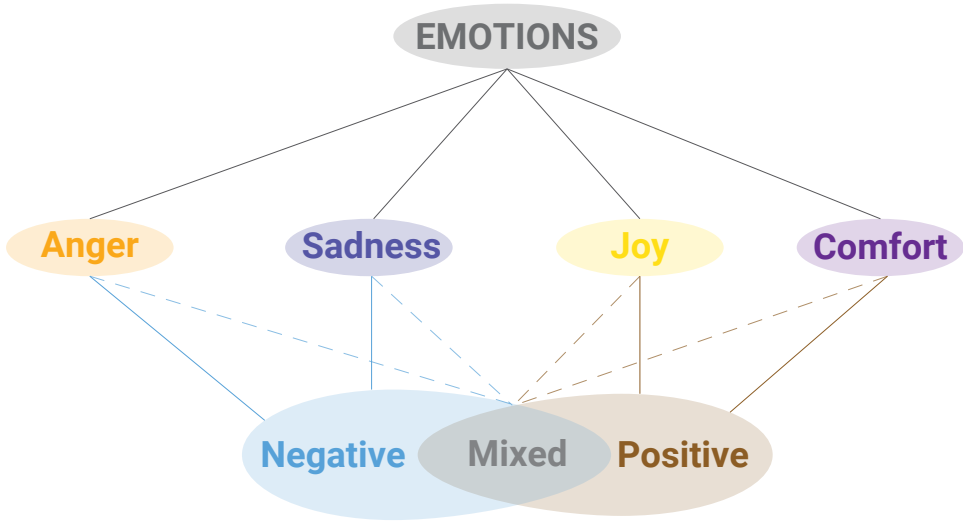



*Figure 27. Adaptation from Ekman's (1999) basic emotion categories*

*Source: Ekman, 1999*

# Data Collection, Processing & Analysis

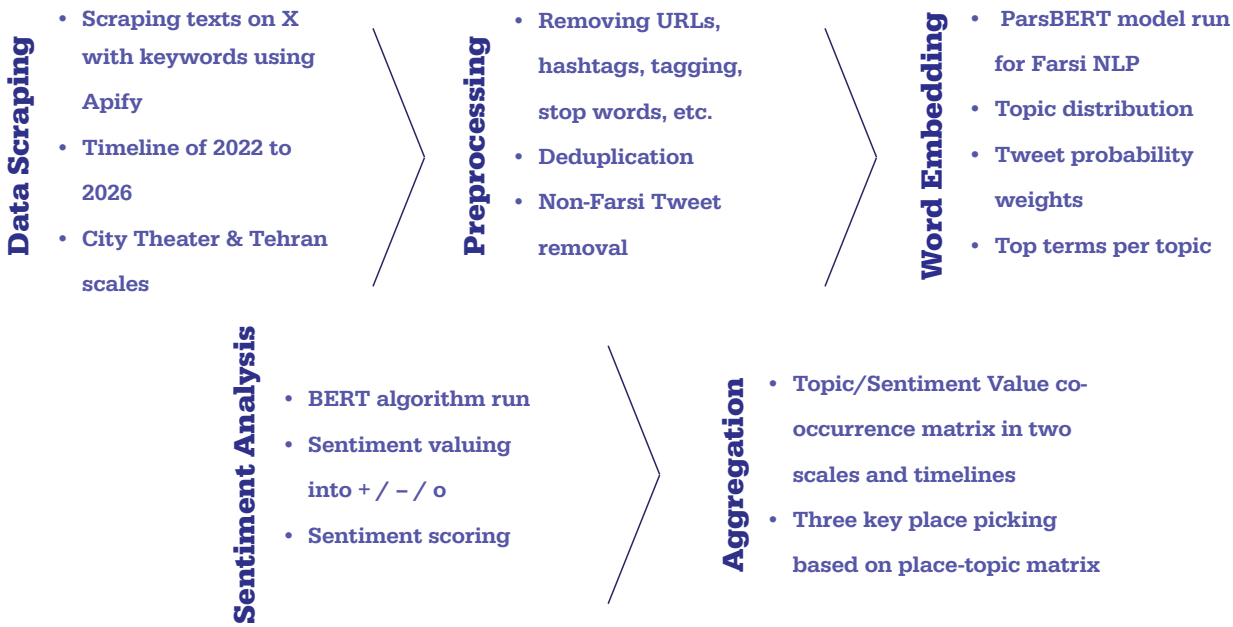
The main steps, tools, and methods that were used to collect, organize, preprocess, analyze, and represent the data are illustrated in the diagram in Figure 28. The steps were changed or modified over the process and the visualization below shows the very latest updated workflow of the analytical process.

# Data Collection

## Data Scraping

Data from X was scraped using the Apify scraping agents. The Twitter (X.com) Scraper Unlimited: No Limits, developed by Apidojo as a community member of the Apify network, was utilized to scrape tweets based on the given keywords of the study area, followed by an "OR" logical expression query.





The given timeframe in queries for both actors was a timeline divided by the approximate date of interventions, starting with placing the construction fence around the square. The timeframe was based on a before-and-after of the specified intervention. Overall, to include a sufficient amount of data regarding the City Theater and Tehran public spaces, containing daily experiences recorded on the digital sphere, a timeline starting from 2022 to 2026 was considered for the data collection timeframe.

The keywords used in the query builder were the specific names of the City Theater square located in Tehran, with keywords of the chosen key spaces that may contain relevant information of the study area, as the supporting context scale was defined. The contextual areas were Enghelab Square, Enghelab Street, Vali'asr Street, and Vali'asr Intersection. The given keywords in the query were given in Farsi, so the data collection could be done as accurately as possible. There were different forms that the name can be written, both in scripture and conversational way (Table 4). All possible ways were considered to increase the scope and precision of the search query. A sample of this query is represented in the table below. The rest of the query buildings for Tehran's scale were made based on the same approach (check the Appendix 3).



# Preprocessing

There were initial specific steps that needed to be taken before starting the data analysis. This was the preprocessing of the data. The steps are described below, and the corresponding code blocks, detailed workflows are provided in the end of this report.

### Taggings, URLs, Hashtags, White Space Removals

To make the processing as accurate as possible, there was a need to filter out the interrupting components inside tweets that contained taggings, URLs, Hashtags, and potential white spaces. The aim was to remove any irrelevant textual elements that would affect data processing, and the aforementioned items were the case. AlAgha, 2021; Ghahramani et al., 2021). There were specific functions that could be defined inside the RStudio environment to remove the items based on the given instructions. The relevant R scripts for text cleaning are shown in the code blocks in the Appendix. The library used for defining the functions was from the "stringr" package in the Tidyverse ecosystem.

### Deduplicating

There were some duplicate tweets inside the scraped corpus of the data. To make sure of the consistency and better processing of the data, the duplicates were identified and removed.

### Arabic tweets

A problem that I encountered was the unwanted tweets in a foreign language. There were several scraped tweets in Arabic, approx. 45, due to alphabetical similarities in both languages. In order to identify the tweets of different languages, the specialized codes were run in Rstudio environment. This helped to clean the data corpus of the foreign language tweets that coincidentally mentioned places in an Iranian context.

### Replacing Emojis

One of the encounters with social media data reflected that emotions were not just expressed by words, but also by emoticons as well. Using emojis to convey the message as short as possible is a common way in social media communications. An approach that AlAgha (2021) recommended to use the emojis as sentiment inputs as

| Keyword in Farsi | Keyword in English |
|---|---|
| تئاتر شهر | City Theater |
| تئاترشهر | CityTheater |
| تئاتر_شهر | City_Theater |
| تئاتر شهر تهران | City Theater Tehran |
| تئاترشهرتهران | CityTheaterTehran |
| تیاتر شهر | CityTeatre |
| متروتئاترشهر | CityTheaterMetro |
| چهارراه‌ولیعصر | ValiasrIntersection |
| چهارراه_ولیعصر | Valiasr_Intersection |
| خیابان انقلاب | Enghleb Street |
| خیابان‌انقلاب | EnghlebStreet |
| خیابان_انقلاب | Enghelab_Street |
| خیابون انقلاب | (Enghelab Street (conversational |

*Table 4. Keywords used for X scraping query*

well by converting them into single-word emotional expressions that align with the visual characteristics of the emoji. For example, the smiley face emoji can represent the "joy" or " happy" emotion, and this could be replaced with the emoticon inside the text data. As mentioned before, the BERT model was already compatible with processing emojis and emoticons as well. Therefore, there was no need to normalize the tweets based on an emoji dictionary before the processing.

### Stop Words Removal

One last step before the word embedding was to remove the stop words from the documents. The stop words diverted the word embedding and topic creation process, resulting in topics consisting only of stop words, due to their prominence and repetition over the data. Therefore, it was necessary to define and remove the stop words from the tweets.

### Half Space Removal

Another Python library was used for any half-space in Persian script, and alphabetical conversions from Arabic to Farsi was the Parsivar model. This helped to fix any inconsistencies in the possible cases through the data.

### Tokenization

Tokenization is a procedure to turn the sentences into single words or phrases that makes the word processing easier for the machine in Topic Modeling (AlAgha, 2021). To apply this, the process was already taken as a pre-step in a Python script for the ParsBERT model run.

### Translation

The collected data, which was in the text format, was translated from Farsi to English using Python's library of Argos, which supported Persian language for offline translation. The reason behind choosing a Python library to run the translation was to use an offline automated approach to follow privacy considerations regarding data exposure to foreign servers. In this way, not only was the labor-intensive work mitigated, but the raw data was not shared with any external online servers such as Google Translate.

The Argos library is an open-source Python library under MIT's license and accessible

via a GitHub repository. The translation in this case was only for visualization purposes. In making the word clouds and representative tweets from chosen topics, and general publishing means, the translation was applied to the data.

# Processing

### Word Embedding

To perform the word embedding, a Python script was written to run the BERT model with tokenization, word embedding and topic naming and save the run model for further applications in the PyCharm software environment. To run the model, you'd need to install the library package on your Python interface. The installation steps and code blocks are provided in the Appendix.



### ParsBERT Model Run

The first analysis was meant for word embedding to run through the large scope of data while making clusters of terms shaping the tweets. The model created the topics based on the group of terms that were relevant enough. In that sense, the model generated as many topics as possible. And, it positioned the terms as numerical vectors in the closest existing topic with the highest weight inside the multi-dimensional space.

1. The preprocessed data was loaded and assigned as a new value.

2. The Normalizer from the Parsivar library was used for one last preparation before the analysis.

3. The ParsBERT model was called and run within the BERT module, creating the word embeddings and finding the topics automatically.

4. A specific outlier reduction was set to gain the optimum number of topics.

5. The topics were updated with the stop words.

6. A topic probability weight matrix was converted into a data frame and joined with the main data structure, and the necessary info was called.

7. At last, the info data frame was saved as a comma-separated values (.csv) file.

8. The run model was also saved for further reuse, such as generating the aggregated probability weights of tweets belonging to each topic per document, i.e., tweet, and for sentiment analysis.

### Sentiment Analysis

Since the model was already run and saved, there was no need to rerun the whole word-embedding. The sentiment analysis could be run on the frozen model instead, making the data processing efficient and traceable.

1. The ParsBERT-Sentiment model was called in the script, and the necessary pipeline was run.

2. The process was done in batches to reduce the pressure from the machine to handle a large scope of processing while making it faster.

3. The model trained by the

1. The pre-trained model also outputs labels called "Recommended" followed by its given training set, which was based on user reviews on online shopping and food delivery platforms, where some of them may give feedback consisting of recommending a product or food.

given vocabulary set went through the documents and created the corresponding sentiment label[1].

4. Sentiment labels and the confidence probability scores were set to be represented as the results and saved to the existing processed data along with their topic modeling results in a .csv file.

## Visualisation

The output file went through multiple combinations of ordering, sorting, categorization, and visualizations to give out the most pleasing and clear results. The code blocks and workflows for modifying and visualizing the processed data are available in the Appendix. Also, some of the main steps and purposes of data analysis representation are discussed in the results chapter.

# Results

In this section, I discussed the results based on categorizing and sorting combinations showcased with visualizations to make the results as clear as possible both for the author and the audience.

## General Topics

The results of word embedding and sentiment analysis were a set of processed tweets with the following discrete, numerical and categorical values in Table 5.

There were 93 topics with semantic relevance within their tweets. Nevertheless, not all of them were spatial or related to a key public space, but more focused on economic, socio-political and other non-spatial topics. The model also created the final topic number zero; this topic may contain a large number of tweets, for the model recognized the tweets that were not a fit for a single topic but addressed something more general. Thus, the final topic zero consisted of a large scope of tweets that belonged to a wide range of small topics that weren't in a reasonable position to other defined topic clusters.



| Topic Modeling & Word Embedding | Sentiment Analysis |
|---|---|
| Final topic number (best fitting topic allocation) per tweet | Sentiment label (positive, negative, neutral) |
| Tweet count per topic | |
| Topic general name | |
| Top terms per topic | (Sentiment score (0.0 to 1.0 |
| Probability weight per document, i.e., (tweet (0.0 to 1.0 | |
| Representative tweets per topic | |

*Table 5. Analytical results*

# Localised Topics

There were 94 extracted topics from the BERT model using the word embedding technique. However, not all of the topics were spatial. So, I filtered the topics that were either spatial or linked to a key location across the city of Tehran.

Although the 27 topics were spatial, not all of them could be relevant to the study or address the issues or challenges that led to the study area, i.e., Tehran's key public spaces, such as the City Theater. This resulted in creating a separate column named contextuality to define whether a topic, and thus, the corresponding tweet, was explicit to this study or not. It was done by filtering out the narrowed topics.



At the end, there were nine topics considered explicit to the study scope that could be used for further steps of the analysis. Table 6 represents the topics, their contextuality, and the places the tweets referred to. The place name classification basis was either a key public space name or a general urban space within the city of Tehran. The tweets that addressed other spaces not related to any introduced key public spaces were considered as the general urban space in Tehran as well.

*Table 6. Word Embedding results*

| No. | Count | Topic | Topic Name | Contextuality | Place Name |
|---|---|---|---|---|---|
| 1 | 558 | 0 | Tehran Valiasr Street Life & Nostalgia | Localised | Tehran Urban Spaces |
| 2 | 171 | 1 | Neighborhood Hubs & Squares | General | Vanak Sqr. |
| 3 | 252 | 2 | Grand Bazaar & Market Shopping | Localised | Tehran Urban Spaces |
| 4 | 236 | 4 | Toopkhaneh Square Naming & War | General | Toopkhaneh Sqr. |
| 5 | 185 | 5 | City Theater Metro & Underpass | Localised | City Theater |
| 6 | 145 | 11 | Metro Infrastructure & Tajrish Station | Localised | Tajrish Sqr |
| 7 | 119 | 12 | Avenues, Gardens & Boulevards | General | Tehran Urban Spaces |
| 8 | 118 | 13 | Performing Arts & Festivals (City Theater) | Localised | City Theater |
| 9 | 111 | 16 | Valiasr Street Connectivity & Length | Localised | Valiasr St |
| 10 | 78 | 18 | Valiasr-Tajrish Tree Preservation & Urban Ecology | Localised | Valiasr St. / Tajrish Sqr. |
| 11 | 99 | 19 | Urban Management & Heritage Protection (City Theater) | Localised | City Theater |
| 12 | 61 | 21 | Nostalgic Atmospheric Streets & Alleys | General | Tehran Urban Spaces |
| 13 | 130 | 22 | Urban Identity and Perception | General | Tehran Urban Spaces |
| 14 | 62 | 30 | Public Monuments | Localised | City Theater / Toopkhaneh Sqr. / Vanak Sqr. |
| 15 | 47 | 31 | Metro Station Entrances | Localised | City Theater / Tajrish Sqr. / Toopkhaneh Sqr./ Hassan Abad Sqr. |
| 16 | 61 | 32 | Seasonal Walks (Autumn in Valiasr) | Localised | Valiasr St. |
| 17 | 50 | 37 | Emotional Attachment to Tajrish | Localised | Tajrish Sqr |
| 18 | 125 | 41 | Emotional Vibe of Toopkhaneh | General | Toopkhaneh Sqr. |
| 19 | 101 | 46 | Hassan Abad Heritage Houses | General | Hassan Abad Sqr. |
| 20 | 52 | 52 | Historical Renaming of Squares | General | Toopkhaneh Sqr./ Hassan Abad Sqr. |
| 21 | 36 | 55 | Historical Squares (Ashura) | General | Toopkhaneh Sqr./ Hassan Abad Sqr. |
| 22 | 20 | 76 | Contemporary Urban History | General | Tehran Urban Spaces |
| 23 | 27 | 82 | Cafe Culture & Solitude | General | Tehran Urban Spaces |
| 24 | 18 | 84 | Homa Self-Immolation in Tajrish | General | Tajrish Sqr |
| 25 | 29 | 86 | Tajrish-Valiasr Walks | Localised | Valiasr St. / Tajrish Sqr. / Vanak .Sqr |
| 26 | 28 | 88 | Tehran's Historical Gates | General | Tehran Urban Spaces |
| 27 | 30 | 93 | Tehran Cafes | General | Tehran Urban Spaces |

# Word Cloud

The localised topics as the key representatives of 2022-2026 transformations among the key public spaces of Tehran and the City Theater, were illustrated as a word cloud. The top ten representative terms per topic are displayed in Figure 29. In general, the topics are grouped into two:

1. **City Theater Topics**
2. **Not-City-Theater Topics**

The second group was mainly focused on other defined key public spaces such as Tajrish square, Valiasr street, Vanak and Toopkhaneh squares. The word cloud representation (Figure 29) signifies the priority of these key public spaces compared to other defined places before. This was a valuable step in narrowing down the topics, orienting the key places based on their popularity and reputation, which was validated by term frequencies within the data corpus. It was evidence that proved the majority of tweets mentioning these key public spaces as discourses were of a public interest, or at least, attention.



*Figure 29. WordCloud of Spatial Topics*

## Spatial Topics of Tehran

Analysis of localised spatial discourse between 2022-2026

**City Theater Metro & Underpass**




**Emotional Attachment to Tajrish**




**Grand Bazaar & Market Shopping**




**Metro Infrastructure & Tajrish Station**




**Metro Station Entrances**




**Performing Arts & Festivals (City Theater)**




**Public Monuments**




**Seasonal Walks (Autumn in Valiasr)**




**Tajrish-Valiasr Walks**




**Tehran / Valiasr Street Life & Nostalgia**




**Urban Management & Heritage Protection of City Theater**




**Valiasr / Tajrish Tree Preservation & Urban Ecology**




**Valiasr Street Connectivity & Length**

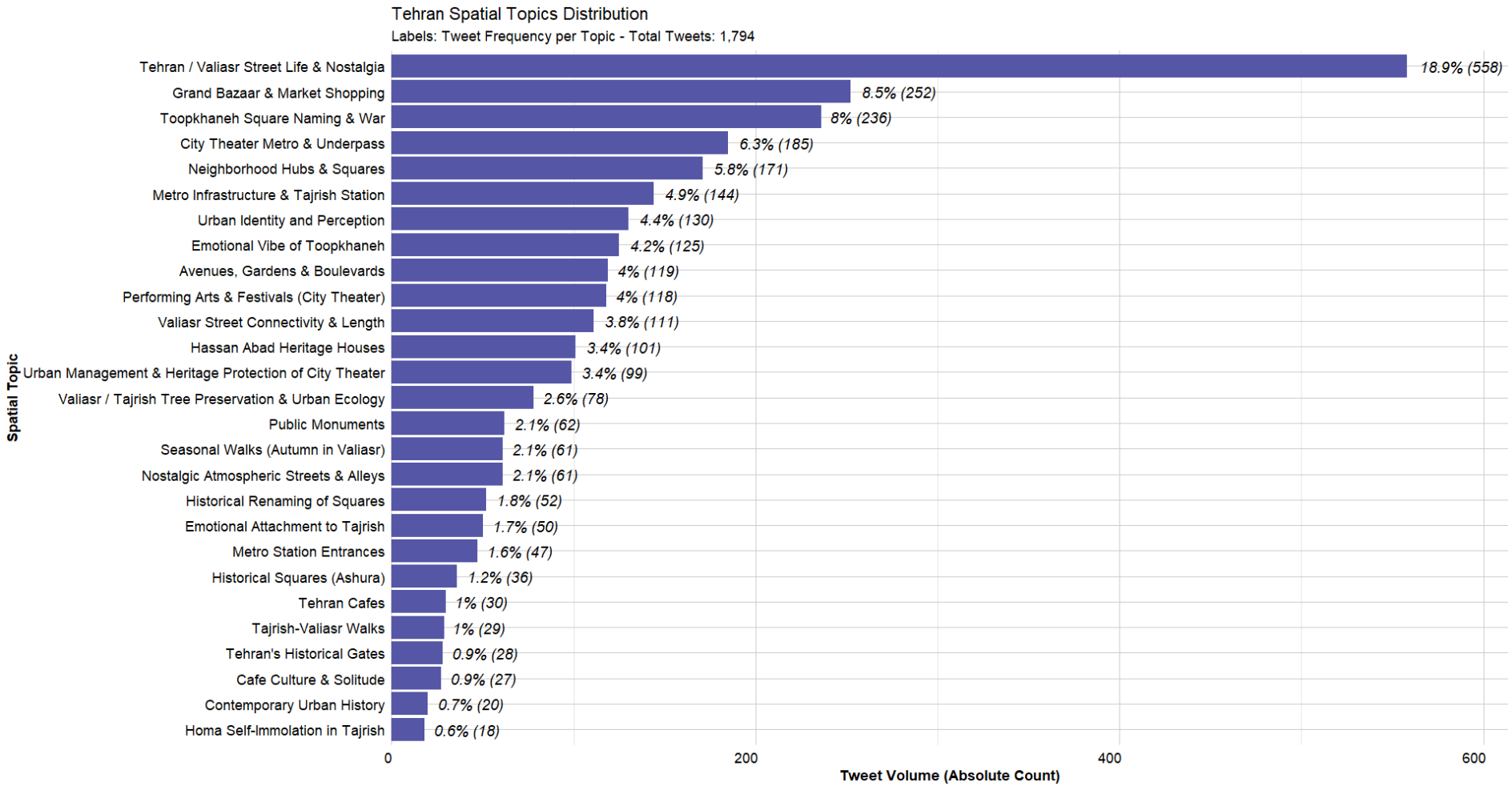


# Topic Distributions

All of the topics that represented the spatial discourses across Tehran were distributed in a plot based on quantity (Figure 30). The following visualization represents the overall topic distributions, which helped to gain a better understanding of each topic's plentitude. Valiasr Street and the grand bazaar have been a prominent topic within the general discourses already. But there was a need to view the topics addressing the study area more specifically.



*Figure 30. General spatial topics distribution*

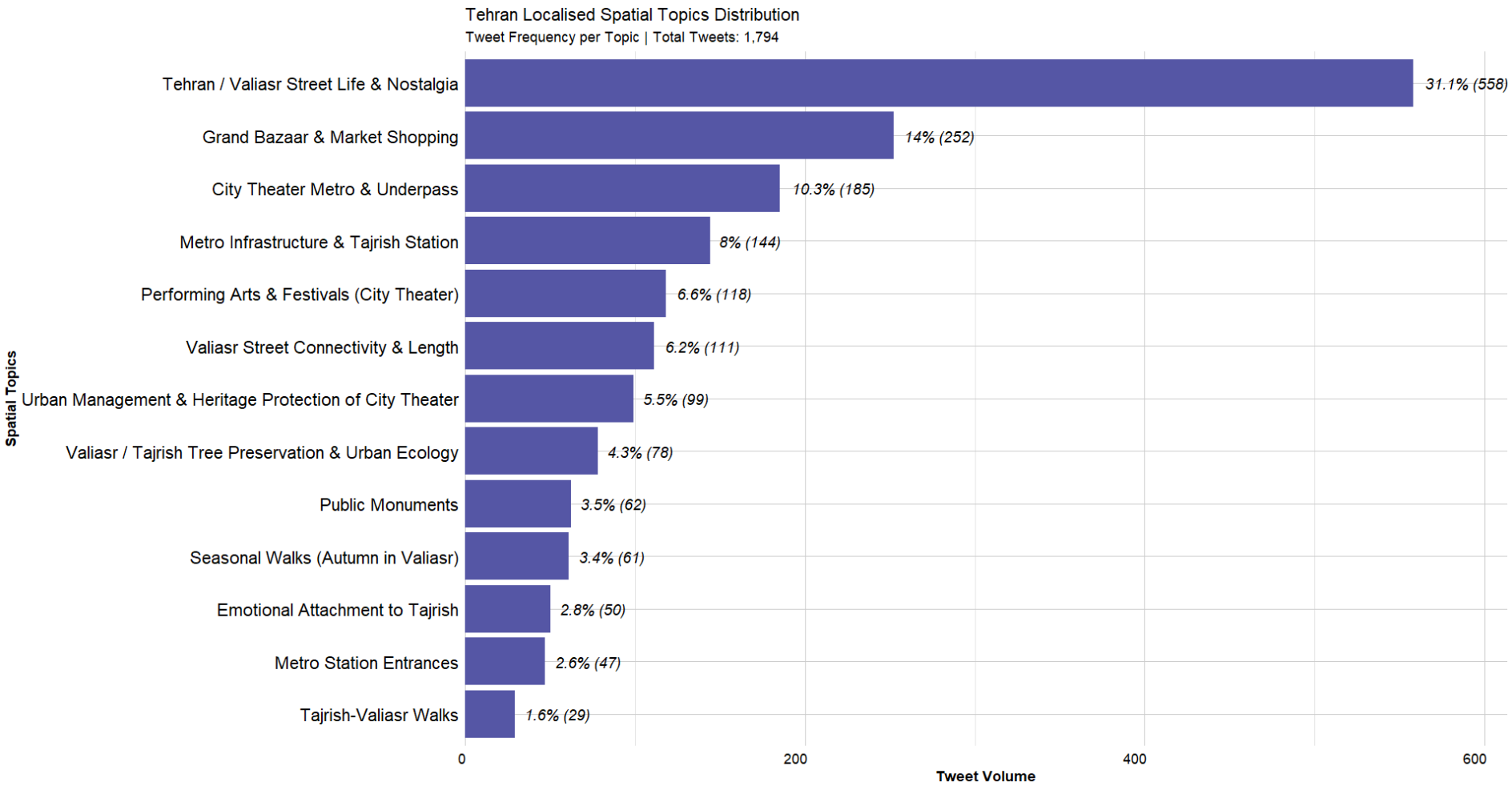

Tehran Spatial Topics Distribution
Labels: Tweet Frequency per Topic - Total Tweets: 1,794
Spatial Topic
Tehran / Valiasr Street Life & Nostalgia
Grand Bazaar & Market Shopping
Toopkhaneh Square Naming & War
City Theater Metro & Underpass
Neighborhood Hubs & Squares
Metro Infrastructure & Tajrish Station
Urban Identity and Perception
Emotional Vibe of Toopkhaneh
Avenues, Gardens & Boulevards
Performing Arts & Festivals (City Theater)
Valiasr Street Connectivity & Length
Hassan Abad Heritage Houses
Urban Management & Heritage Protection of City Theater
Valiasr / Tajrish Tree Preservation & Urban Ecology
Public Monuments
Seasonal Walks (Autumn in Valiasr)
Nostalgic Atmospheric Streets & Alleys
Historical Renaming of Squares
Emotional Attachment to Tajrish
Metro Station Entrances
Historical Squares (Ashura)
Tehran Cafes
Tajrish-Valiasr Walks
Tehran's Historical Gates
Cafe Culture & Solitude
Contemporary Urban History
Homa Self-Immolation in Tajrish
18.9% (558)
8.5% (252)
8% (236)
6.3% (185)
5.8% (171)
4.9% (144)
4.4% (130)
4.2% (125)
4% (119)
4% (118)
3.8% (111)
3.4% (101)
3.4% (99)
2.6% (78)
2.1% (62)
2.1% (61)
2.1% (61)
1.8% (52)
1.7% (50)
1.6% (47)
1.2% (36)
1% (30)
1% (29)
0.9% (28)
0.9% (27)
0.7% (20)
0.6% (18)
0
200
400
600
Tweet Volume (Absolute Count)

## Spatial Localized Topics' Distributions

The general spatial topics occupied a vast scope of topics across Tehran. Thus, a secondary filter was applied to the topics to create a more localized and interconnected set of spatial topics closer to the contextual scope of the study. In Figure 31 it is clearer to identify topics relatively positioned near the City Theater and related places across the city.



*Figure 31. Localised spatial topics*

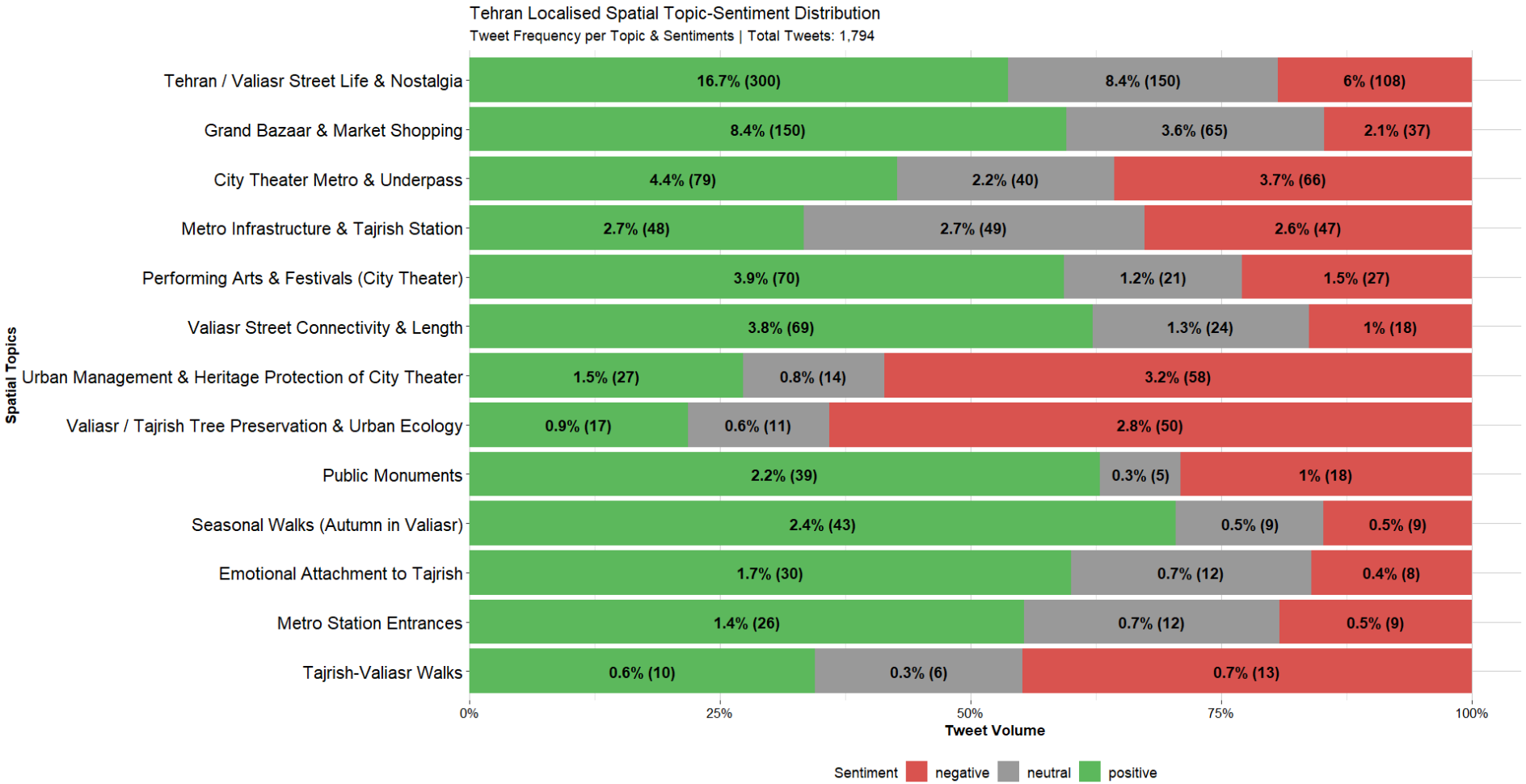

Tehran Localised Spatial Topics Distribution
Tweet Frequency per Topic | Total Tweets: 1,794
Tehran / Valiasr Street Life & Nostalgia
31.1% (558)
Grand Bazaar & Market Shopping
14% (252)
City Theater Metro & Underpass
10.3% (185)
Metro Infrastructure & Tajrish Station
8% (144)
Performing Arts & Festivals (City Theater)
6.6% (118)
Valiasr Street Connectivity & Length
6.2% (111)
Urban Management & Heritage Protection of City Theater
5.5% (99)
Valiasr / Tajrish Tree Preservation & Urban Ecology
4.3% (78)
Public Monuments
3.5% (62)
Seasonal Walks (Autumn in Valiasr)
3.4% (61)
Emotional Attachment to Tajrish
2.8% (50)
Metro Station Entrances
2.6% (47)
Tajrish-Valiasr Walks
1.6% (29)
Spatial Topics
0
200
400
600
Tweet Volume

## Topic-Sentiment Overall Distribution

Topic distribution based on sentiments was also a significant element in the analytical structure. The Figure 32 is an overall representation of all localised topic-sentiment tweet distributions. The elaborations took place in the further steps. The topic distributions are represented based on the percentage per topic.



*Figure 32. Localised topic-sentiment distribution*


Tehran Localised Spatial Topic-Sentiment Distribution
Tweet Frequency per Topic & Sentiments | Total Tweets: 1,794
Spatial Topics
Tehran / Valiasr Street Life & Nostalgia
Grand Bazaar & Market Shopping
City Theater Metro & Underpass
Metro Infrastructure & Tajrish Station
Performing Arts & Festivals (City Theater)
Valiasr Street Connectivity & Length
Urban Management & Heritage Protection of City Theater
Valiasr / Tajrish Tree Preservation & Urban Ecology
Public Monuments
Seasonal Walks (Autumn in Valiasr)
Emotional Attachment to Tajrish
Metro Station Entrances
Tajrish-Valiasr Walks
16.7% (300)
8.4% (150)
6% (108)
8.4% (150)
3.6% (65)
2.1% (37)
4.4% (79)
2.2% (40)
3.7% (66)
2.7% (48)
2.7% (49)
2.6% (47)
3.9% (70)
1.2% (21)
1.5% (27)
3.8% (69)
1.3% (24)
1% (18)
1.5% (27)
0.8% (14)
3.2% (58)
0.9% (17)
0.6% (11)
2.8% (50)
2.2% (39)
0.3% (5)
1% (18)
2.4% (43)
0.5% (9)
0.5% (9)
1.7% (30)
0.7% (12)
0.4% (8)
1.4% (26)
0.7% (12)
0.5% (9)
0.6% (10)
0.3% (6)
0.7% (13)
0%
25%
50%
75%
100%
Tweet Volume
Sentiment
negative
neutral
positive

# City Theater Between 2022-2026

As the first group of analytical results of the word embedding and sentiment analysis, the initial focus was on topics fitting in City Theater's scope. As explained before, the data collection went through the classification of tweets within the City Theater area with the type value of "ct". And, the data corpus belonging to external key public spaces was valued as "not_ct". This is why there were tweets that belong to topics regarding places other than the City Theater, but the content itself is linked to the City Theater as well.

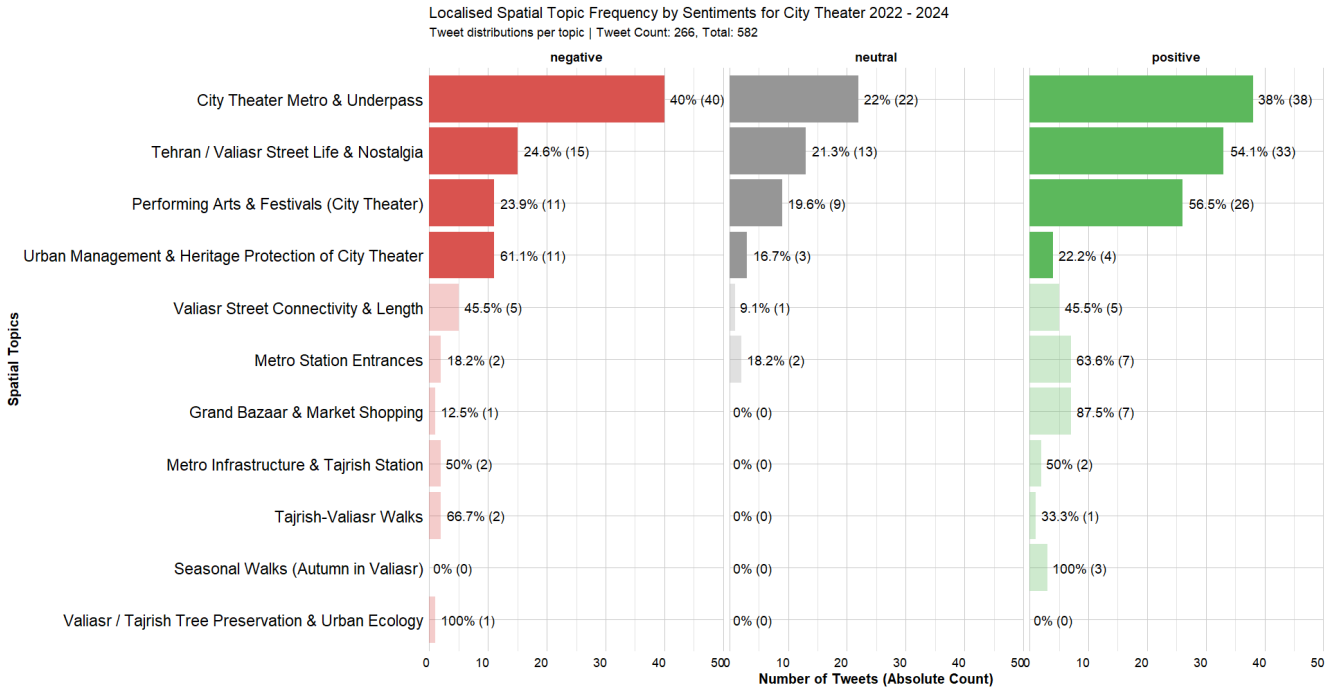
The "City Theater Metro & Underpass" was apparently the most recognized topic before "Tehran / Valiasr Street Life & Nostalgia" and "Performing Arts & Festivals (City Theater)" as the second and third top topics, respectively.

The first topic held the majority of tweets to itself. Making it an important aspect of the City Theater area. The underpass space at the Valiasr-Enghelab junction is quite a unique indoor pedestrian space as well. The underpass functions as a node and a connection to City Theater's metro station, NS-EW streets accesses and a center for retail shops, attracting commuters and visitors to the underground space.

As can be seen in Figure 33, the City Theater in the first period of the study timeline was neither a trend nor an outlier within the corpus. And there were a significant number of tweets with negative sentiments towards the "Urban Management & Heritage Protection of City Theater" topic. The topic was mainly about the main spatial intervention for which rumors were spread at that time. The results show that the reactions were not quite positive. Tweets with negative sentiments regarding the intervention took 61.1% (11) of its corpus, which makes it one of the top three topics with negative values.



*Figure 33. Topic-sentiments for City Theater before 2024*

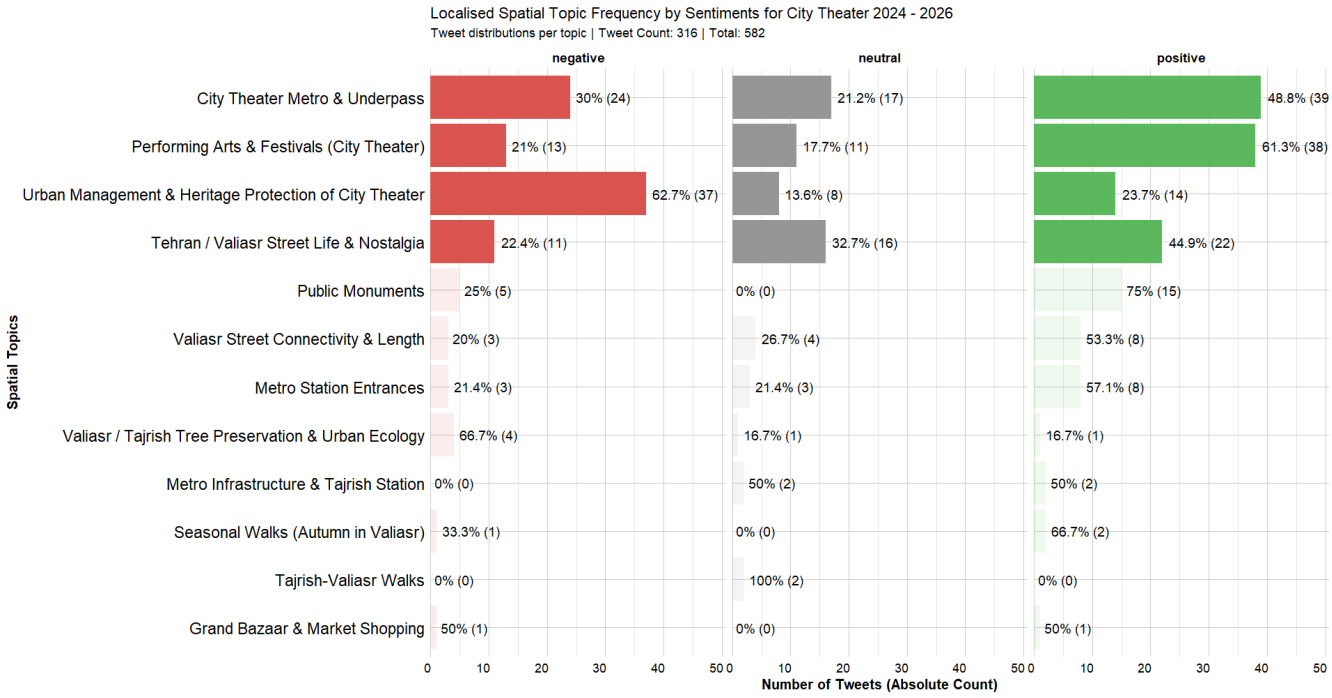


Moving to the second timeline (Figure 34), the topics' dominance had roughly remained the same. A major shift of negative tweets regarding the interventions of City Theater happened, which even overtook the number of negative tweets of the first two topics, consists 62.7% (37) of its corpus. At the same time, there weren't many tweets with neutral or positive values underlying the same topic. This signified how the intervention that took place in the City Theater was negatively valued through its long ongoing process.

The positive and neutral valued tweets are also concentrated in the underpass, Performing Arts and the general other classification of Tehran / Valiasr street. Tweets within the City Theater scale were well fitting, for even the topics not directly addressing the City Theater are about the linked places. This reassured the semantic integrity of tweets and topics that help to shape the overall understanding of topic frequencies by sentiments between 2022 and 2026 in City Theater.



*Figure 34. Topic-sentiments for City Theater after 2024*



## Combined Scales

The frequencies of topics plotted in separate scales illustrated the place of the issue specifically. But it didn't reflect the position of the study area among its peers. Spaces that share the same general spatial and non-spatial characteristics. Thus, another overview was to assess the results on a combined scale. By grouping the 1,794 tweets of "ct" and "not_ct" together, I saw the dynamics of topic-sentiment between the two timelines all at once. This helped to make a different, yet clear comparison of the transformation of sentiments towards each explicit spatial topic.

In Figure 35, the topic-sentiment dynamics are also significant. Also, it is clearer to see how each topic was of public interest between the two timelines on the overall scale of the study. The topic zero was the dominant discourse in both timelines, yet some topics faced a dramatic change in their priority values. Starting from the study's issue, the significant change of negative emotions towards the City Theater's redesign project was considerable here as well. It was a jump from 14 to 44 of the total corpuses between 2022 and 2026. Moreover, the proportion of negative tweets remained nearly consistent from 56% to 59.3%. The reason why I focused on these shifts in dynamics was that there had been some specific topics that acted as the baseline of the Twitter feed over the timeline. Topic zero, grand bazaar, and metro infrastructure were the dominant discourses in Tehran's Twitter 2022-2026. And based on the results, there are always discussions, arguments, expressions and opinions generated regarding these topics regardless of key events happening in other places. This hierarchy represented how some urban spaces not only played as driving factors of trends in Twitter, but also existed as core elements of Tehran's spatial organization.

Now that the baseline topics were defined, the dynamic shifts in other urban spaces were reviewed precisely. The City Theater reconstruction project was not the only turnover. Some topics gained reactions after 2024 (Figure 36), like the "Public Monuments" topic. It didn't have many emotionally valued discussions in the first timeline. However, the topic received majorly positive and negative reactions, with 62.5% (35) and 30.4% (17) of the topics, respectively. It emerged due to the installation of new statues in Vanak and Enghelab squares during the 2024-2026 period.

The topics like the public monuments and City Theater heritage violation are of emerging topics that were dependent on key events in a specific timeframe, while there were major topics that belonged to a longer timeframe. The Valiasr trees and ecology, metro infrastructure, Tajrish metro station and City Theater's underpass have been undergoing discourses along the whole chosen timeline. This argument resulted in defining topics that belonged to short-term changes and topics that were of a long-term transformation.

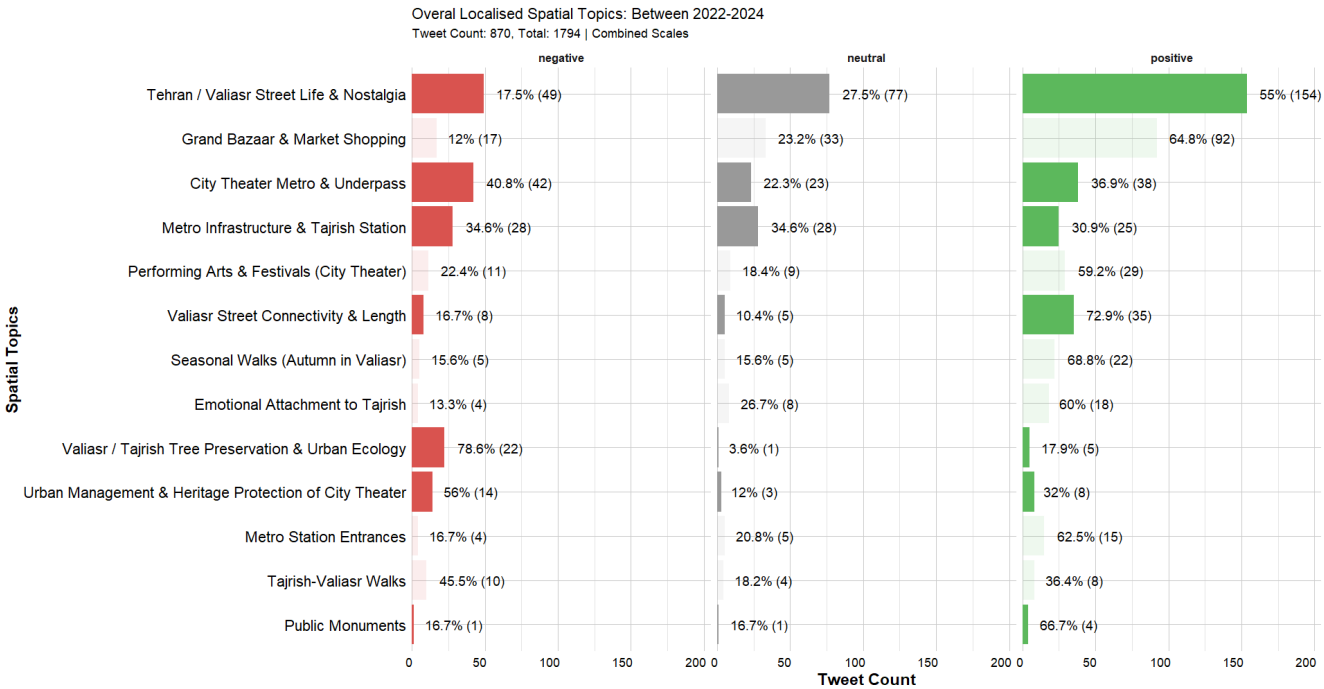



*Figure 35. Overall topic-sentiments before 2024*



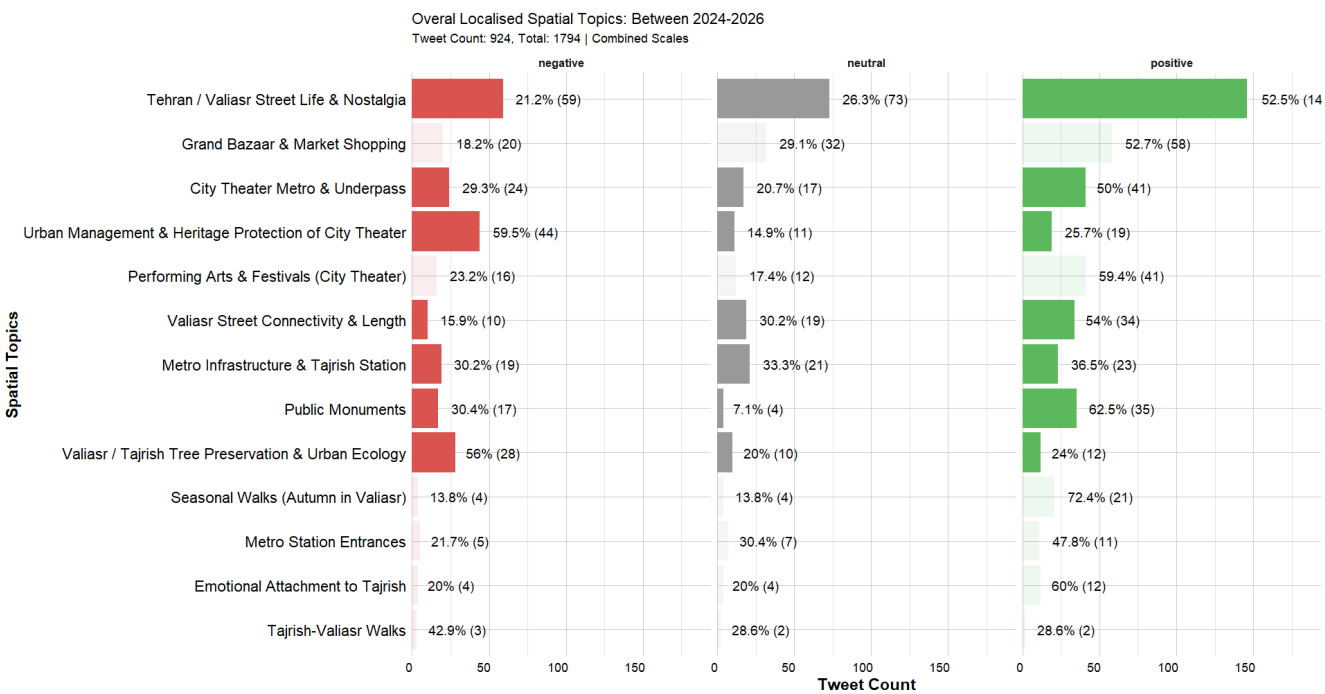



*Figure 36. Overall topic-sentiments after 2024*

# Tweet Frequency per Average Sentiment Value

The last representation of the results was the combined matrix of both scales and timelines with their average sentiment scores per topic (Figures 37 & 38). Having this type of visualization opened even newer insights into comparing the focused and large scales over the study timeline.

The average sentiment scores represented the intensity of each emotion towards the corresponding topic. In that way, I could understand how much negative, neutral, or positive a topic was on average. As can be seen in Figures 37 & 38 , there weren't many significant changes in the sentiment average scores, for the tweets were mainly valued as a whole through the BERT NLP processing. Therefore, the average scores were equally given to all of the localised topics on the Tehran scale. Maybe one of the tangible shifts in average sentiment scores belonged to the "Grand Bazaar & Market Shopping" topic that had lost its sentimental intensity in all three emotional categories after 2024.

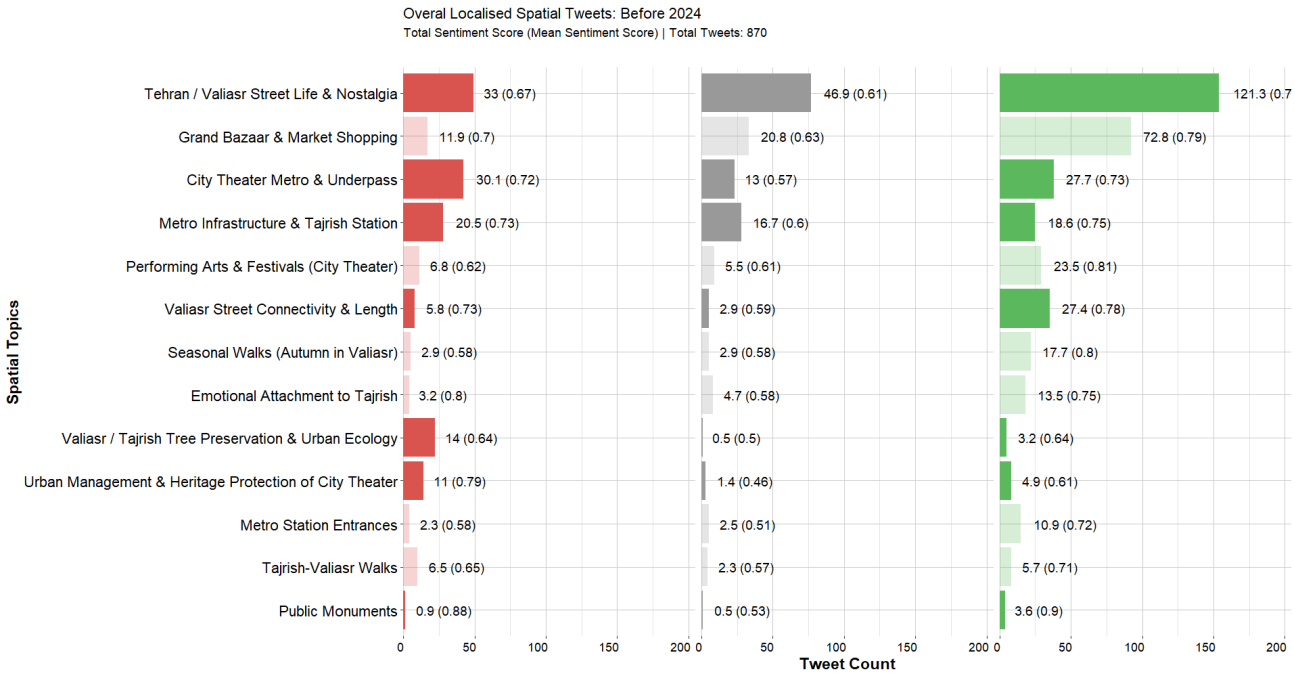



*Figure 37. Localised spatial topics with an average sentiment score before 2024*



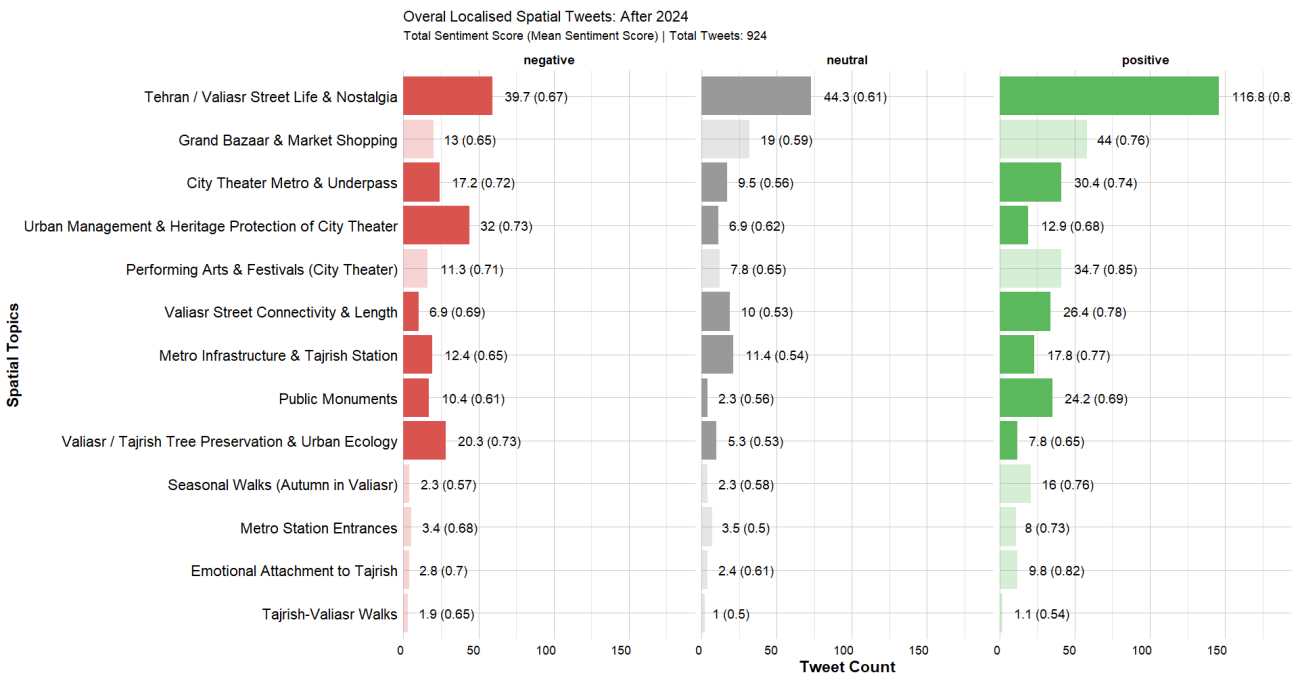



*Figure 38. Localised spatial topics with an average sentiment score after 2024*

# Chi-Square Test

As a part of proving the relevance of the main statement in this study, that the topics and sentiments towards the City Theater have changed after the key intervention in 2024, a statistical assessment of Chi-Square was applied to the results.

The Chi-Square assessment proves that the made argument regarding the relevance of observed and expected frequencies of two categorical values is statistically evident and reliable. In simpler terms, it helps to compare the expected and observed results of an analysis. The purpose of this test is to determine whether the difference between the two variables is due to chance or due to a meaningful relevance. This test depicted whether the two variables were statistically independent or associated with each other (Kent State University Library, 2026; University of Southampton, 2026).



Accordingly, this test cannot define any causation but an association between the two categorical variables (Kent State University Library, 2026). In the current study, the two categorical values were defined as sentiment labels regarding the city theater and not city theater topics within two timelines, before and after the 2024 key intervention.

This test was simply done in the R ecosystem without installing any additional packages. The R script provided the results that should be taken into account (check Appendix). On this test, two important indicators should be considered:

### Chi-Squared statistics ($x^2$)

The squared value determines the amount of variance between the observed and expected frequency if the variables were not associated (DataCamp, 2026).

### P - Value

The p-value indicates the probability of finding a chi-square result as high as or more than the existing one if there was no statistical association found (DataCamp, 2026).

# Chi-Square Test Results

Once the analytical results of topic modeling and sentiment analysis were ready, the chi-square assessment was done using the relevant R script. The results of the analysis were opened as a comma-separated value (.csv) file inside the R environment. The file was converted into a one-by-three matrix (Table 7) grouping the tweets regarding the City Theater within localised topics. The other row would be the total corresponding sentiment labels. Two matrices were made for the data before 2024 and after 2024, respectively. An additional overall matrix was also made to understand the significance over the whole timeline.

The matrices were technically a summary table of the processed data. The first two rows were the separated topics, timeline-wise. The test code was run on matrices separately, based on including CT and non-CT scales, to print the P-value and the X-squared results that can be seen in Table 8 & Table 9.

Ultimately, based on the results shown in Table 8 & Table 9, the hypothesis of sentiment shifts with discourses regarding the city theater transformations was significant. The p-value for the different timelines and sentiments for both CT and non-CT combinations was less than the 0.05 value limit.

| Timeline | Type | Negative | Neutral | Positive | Total |
|---|---|---|---|---|---|
| **Before '24** (2022-2024) | **CT** | 90 | 50 | 126 | 870 |
| | **Not CT** | 125 | 152 | 327 | |
| **After '24** (2024-2026) | **CT** | 102 | 64 | 150 | 924 |
| | **Not CT** | 151 | 152 | 305 | |
| Combined Timeline (2022-2026) | CT | 192 | 114 | 276 | 1794 |
| | Not CT | 276 | 304 | 632 | |

*Table 7. Summary table of City Theater and Non-City Theater tweets for chi-square*



| CT | X-squared | P-value |
|---|---|---|
| **Timeline-Sentiment Table** | 0.26266 | 0.8769 |

*Table 8. Chi-square results for CT matrices*

| Not CT | X-squared | P-value |
|---|---|---|
| **Timeline-Sentiment Table** | 3.2019 | 0.2017 |

*Table 9. Chi-square results for non-CT matrices*

## Over/Under Representation

It was a bit complex to tell whether the topics' proportions had changed significantly or not. If we take the City Theater topics as a sample corpus and Tehran topics as the reference, we can analyze what topics were overtaking the average discourse in the City Theater scale. This helped to identify the over/ under represented data (IBM Community, 2026). The overall localized topics that had directly mentioned the City Theater within their category were filtered out. The only topic that didn't have any relevant value regarding the City Theater was "Emotional Attachment to Tajrish". This also indicated - in terms of the study scope contextuality - how precise the localised topics are.



The under/over-representation highlighted the topics that were over-represented and those that were underrepresented within the overall corpus of tweets. The overrepresented topics were the ones that took over the trend from the average spatial topics of Tehran. The underrepresented topics were the ones with the least topic occupancy in City Theater over the Tehran discourse as the reference. There were different ways to calculate data representation, which I tried, including the volume difference and the Location Quotient. However, the second method was found to be more precise in terms of topic intensity and was picked as the ultimate method. The volume difference method description and development and the relevant scripts for both methods can be found in the Appendix.

## Representation based on Location Quotient

The over/under representation was analyzed per topic. I calculated the ratio of a topic in the CT sample over the Tehran reference. It indicated the intensity of a topic in the average discourse. Technically speaking, I calculated the Location Quotient (LQ) of that topic (U.S. Bureau of Economic Analysis, 2018; Xu et al., 2018). The formula for calculating the LQ is as follows:

$$\frac{\%\ of\ Topic\ A\ in\ City\ Theater}{\%\ of\ Topic\ A\ in\ all\ of\ Tehran} = \textbf{Location Quotient}$$

$$If\ \textbf{LQ} > 1 : Topic\ is\ specialized\ to\ CT$$

$$If\ \textbf{LQ} < 1 : Topic\ is\ deficient\ or\ not\ relevant\ to\ CT$$

The results based on the above formulation are represented in Table 10.

*Table 10. Representation Analysis based on LQ values*

| No. | Topic Name | CT Count | CT pct (%) | Global pct (%) | Local Quotient |
|---|---|---|---|---|---|
| 1 | City Theater Metro & Underpass | 180 | 30.92784 | 8.874459 | 3.485039 |
| 2 | Grand Bazaar & Market Shopping | 10 | 1.718213 | 11.90476 | 0.14433 |
| 3 | Metro Infrastructure & Tajrish Station | 8 | 1.37457 | 6.818182 | 0.201604 |
| 4 | Metro Station Entrances | 25 | 4.295533 | 2.489177 | 1.725684 |
| 5 | Performing Arts & Festivals (City Theater) | 108 | 18.5567 | 7.467532 | 2.484984 |
| 6 | Public Monuments | 20 | 3.436426 | 6.060606 | 0.56701 |
| 7 | Seasonal Walks (Autumn in Valiasr) | 6 | 1.030928 | 3.138528 | 0.328475 |
| 8 | Tajrish-Valiasr Walks | 5 | 0.859107 | 0.757576 | 1.134021 |
| 9 | Tehran / Valiasr Street Life & Nostalgia | 110 | 18.90034 | 30.08658 | 0.628198 |
| 10 | Urban Management & Heritage Protection of City Theater | 77 | 13.23024 | 8.008658 | 1.651992 |
| 11 | Valiasr / Tajrish Tree Preservation & Urban Ecology | 7 | 1.202749 | 5.411255 | 0.222268 |
| 12 | Valiasr Street Connectivity & Length | 26 | 4.467354 | 6.818182 | 0.655212 |

The results were visualized in a plot (Figure 39) with which values above one were specialized topics in City Theater, while the topics below one were not relevant or ignored within the scope. The examples of ignored topics were the Bazaar and Valiasr street general discourses. And, the CT Metro-Underpass, Performing Arts, and Intervention topics were highly specialized for the City Theater scale.

The interesting point was that, although some topics were not related to the City Theater directly, they were specialized to the scale. The "Metro Station Entrances" and "Tajrish-Valiasr Walks" were examples of these middle-ground topics. This showed the mental relation of these spaces through their physical connectivity, which was perceived and mentioned by users in X.



There was a credibility of the LQ method compared to a simple subtraction in its level of precision. The public monument topic was not considered a specialized topic for CT, even though it was nearly at the reference deviation center of the previous representations.

This representation helped me to emphasize the relativity of the topics to the City Theater scale and even highlighted the key spaces that were spatially and mentally interconnected based on the discourses on X.

These steps were taken to first indicate the significance of the sentiment value shifts with topics and second, how much. The chi-square results provided evidence for my initial hypothesis. The over/under representations helped to acknowledge how extreme tweets were towards some topics regarding the focus of the study area, i.e., City Theater. At the time that the argument was evident, I moved forward to pick the key places along with City Theater.

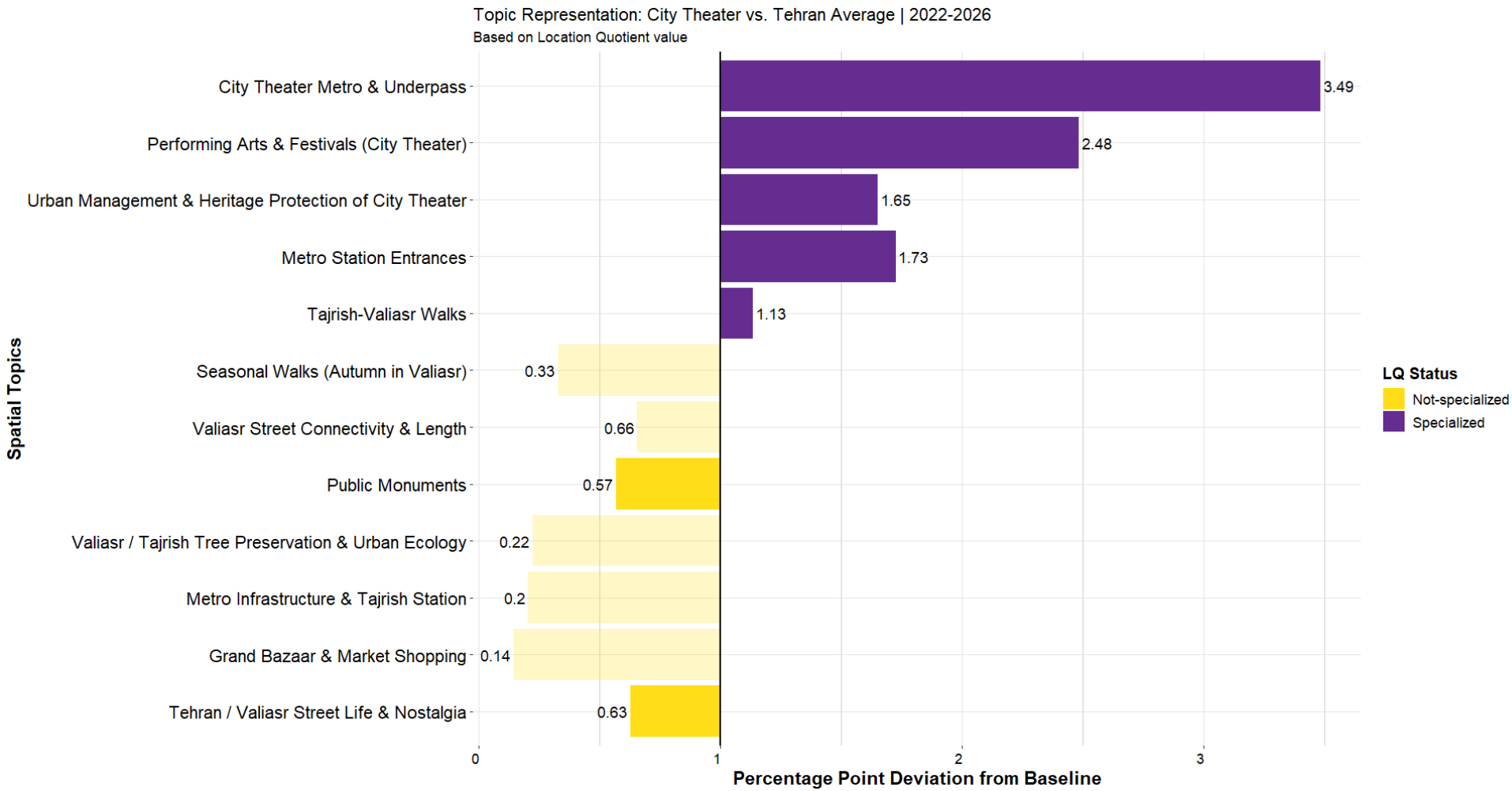
## Key Places

It was an idea to shift from the analysis into the design phase by picking the places that shared attention and attractions in common. Based on the analytical results of explicit spatial topics filtration, some key urban spaces gained the closest attention:

- **Valiasr St.**
- **Tajrish Sqr.**
- **Toopkhaneh Sqr.**
- **Hassan Abad Sqr.**
- **Vanak Sqr.**
- **City Theater**

The idea was to validate the applicability of the conceptual framework in every urban space that held the digital memory potential. It didn't have to be concentrated in one specific public space at the heart of the city. Therefore, I decided to pick the urban spaces that held the topic trends for themselves and check further which spaces had similarities with the focused study area, i.e., City Theater. This was the approach to introduce the conceptual framework of this study as a generalized component in urban design.

*Figure 39. Over-under representation based on Location Quotient*


Topic Representation: City Theater vs. Tehran Average | 2022-2026
Based on Location Quotient value
City Theater Metro & Underpass
Performing Arts & Festivals (City Theater)
Urban Management & Heritage Protection of City Theater
Metro Station Entrances
Tajrish-Valiasr Walks
Seasonal Walks (Autumn in Valiasr)
Valiasr Street Connectivity & Length
Public Monuments
Valiasr / Tajrish Tree Preservation & Urban Ecology
Metro Infrastructure & Tajrish Station
Grand Bazaar & Market Shopping
Tehran / Valiasr Street Life & Nostalgia
3.49
2.48
1.65
1.73
1.13
0.33
0.66
0.57
0.22
0.2
0.14
0.63
Spatial Topics
0
1
2
3
Percentage Point Deviation from Baseline
LQ Status
Not-specialized
Specialized

## Co-occurrence Matrices

There were also representations of the overall discourses in general and localised contexts, visualized in the following plots (Figure 40 & 41). The key spaces that had the cooccurrence with relevant topics and other places in both scales were highlighted.

This representation helped me to find the most relevant and coherent topic-space combinations that shaped the major presence across the general and local discourses of the study. I attempted to filter the key places using the two scales into a precise set of place-topics throughout the overall corpus.

The results were quite clear that Valiasr, Tajrish, and City Theater had the most topic-place relevance and coherence. While Toopkhaneh and Hassan Abad squares had a significant presence within the general focus of the study. The other three key spaces brought more diversity to topics and discourses that appeared on the localized scale as well.

The further it went, the clearer the results became. I attempted to connect the intersection points to each other, both vertically and horizontally. And based on the junctions of places and their proximity to adjacent topics, some topics were combined as they shared the same core. These relativities had real-world evidence, both physically and digitally. To conclude, this is where we can argue that the physical and the corresponding digital clone, i.e., space, overlap.

The key places that didn't have more than one corresponding topic were not shortlisted. The key place of the Grand Bazaar was considered a baseline topic that will always flow through urban daily lives and didn't face any spatial key transformation over the timeline, so, based on the previous experimentations, these sorts of topics were named as the global baseline.

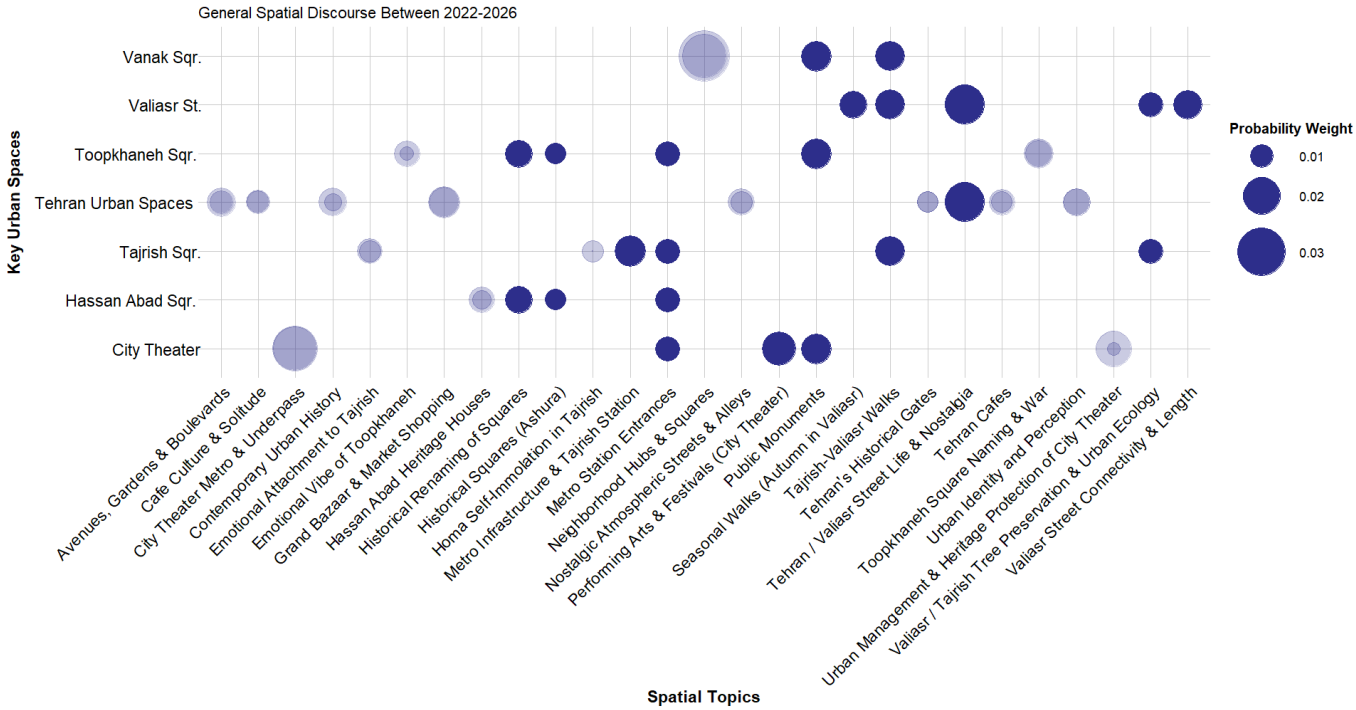
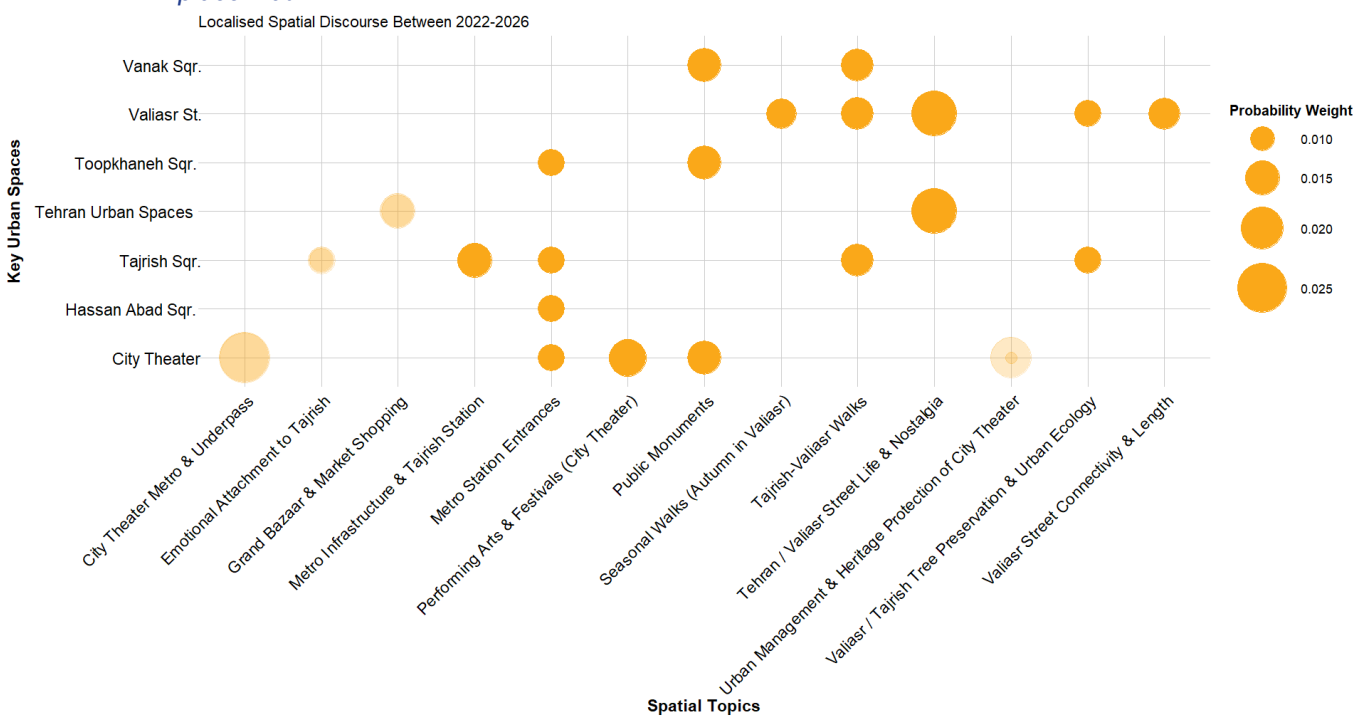



*Figure 40. General topic-place matrix*



*Figure 41. Localised topic-place matrix*



The matrix in Figure 42 was based on highlighting the cooccurrences of topics with the key spaces across Tehran. The topics literally shared the same name as the place, or overlapped with other places, too. The steps that were taken to recognize the spaces addressed in tweets per topic were to investigate them manually. This was due to constraints happened through translating the tweets from Farsi into English that some key names being lost. That was why I had to skim through the tweets to compile the most mentioned names of key places in the list. Some topics included several key places in their subset tweets. This highlighted the interconnectivity of places that were spatially and mentally connected, directly or indirectly. For clarification and better understanding of the context, the Valiasr street starts at Tajrish square and goes through Vanak and City Theater. Also, Toopkhaneh square is one station away from Hassan Abad square by metro.



The technicality behind scoring the places in the following matrices was based on the tweets' topic probability weights. The numbers achieved through BERT processing were averaged and linked to the shortlisted key places that were mentioned in the topics. The average intensities indicated how relevant a topic was to a place. The step-by-step process of this analysis is available in the Appendix.

The before-and-after 2024 timeline results were overlaid (Figure 42), made it easier to read the shift in topic intensity over time. However, the majority of the topics in the general discourse remained stable. One of the significant shifts was related to the public monument topic, which previously had little relevance, but gained more attention due to several new installations and transformations in the CT and Vanak squares. The other dramatic shift, or more precisely, the emerging topic, was City Theater's heritage protection.




*Figure 42. Topic-sentiment shifts between 2022-2026*

The topic-sentiment matrix was the main comparison aimed for this analysis, as it was represented in the methodological framework. In Figure 42, the points where the sentiment shifts happened were highlighted. Some of the shifts weren't visually significant, yet the topics' average sentiment scores had transformed between timelines. The average sentiment scores were not quite reliable, for there were few tweets about public monuments before 2024, but the current matrix shows they had a higher magnitude compared to 2024-2026. Nevertheless, in Figure 44, the place-topic matrix, it was significantly visible how the public monuments topic emerged in the 2024-26 timeline and was a faint discourse before that. To fix that error, I developed a weight-based normalization on the tweets per topic counts that represented the average sentiment values proportionally.



The matrix visualization based on the new weighted average value can be seen in Figure 43. The results were even clearer on the public monument and CT's space topics. The key shifts in overall sentiment scores happened in these topics.` The key shifts in overall sentiment scores happened in these topics. This again gives evidence of how tweets shift towards specific topics temporally. Following that, the sentiments' categorical values change.

The plotted matrix of place-topic and average probability weight shifts between 2022 and 2026 (Figure 44) depicts the value of each topic based on the changes in the average probability weights of their tweets in the corresponding places. This helped to locate where exactly topics have shifted and or emerged between the two timelines. It is visible that City Theater gained an increase in its topics after 2024 and the "public monuments" topic also emerged in places like Vanak and CT post 2024.

$$Tweet\ count\ per\ topic = Topic_C$$

$$Total\ tweet\ count\ per\ timeline = Topic_T$$

$$Average\ Sentiment\ Score = Score_{Avg}$$

$$\textbf{\textit{Proportionate Weight (Topic Share)}} = \frac{Topic_C}{Topic_T}$$

$$\textbf{\textit{Weighted Normalized Sentiment Score}} = Score_{Avg} \times Topic\ Share$$

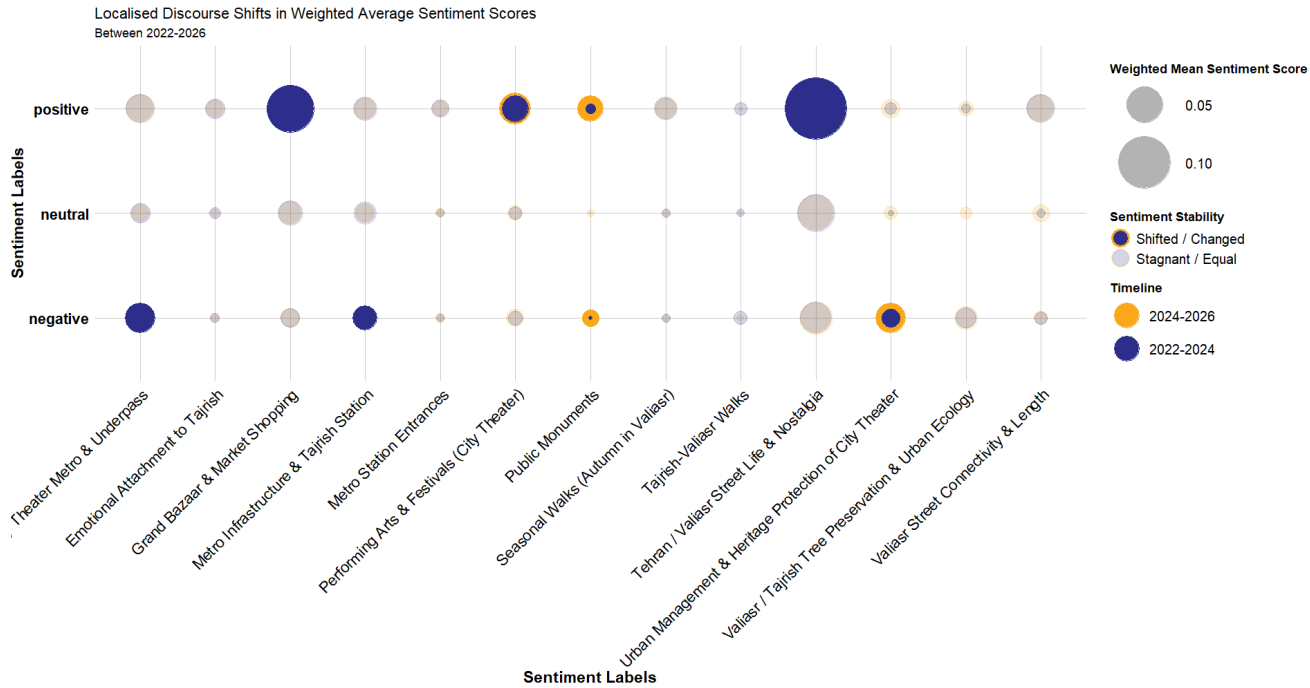
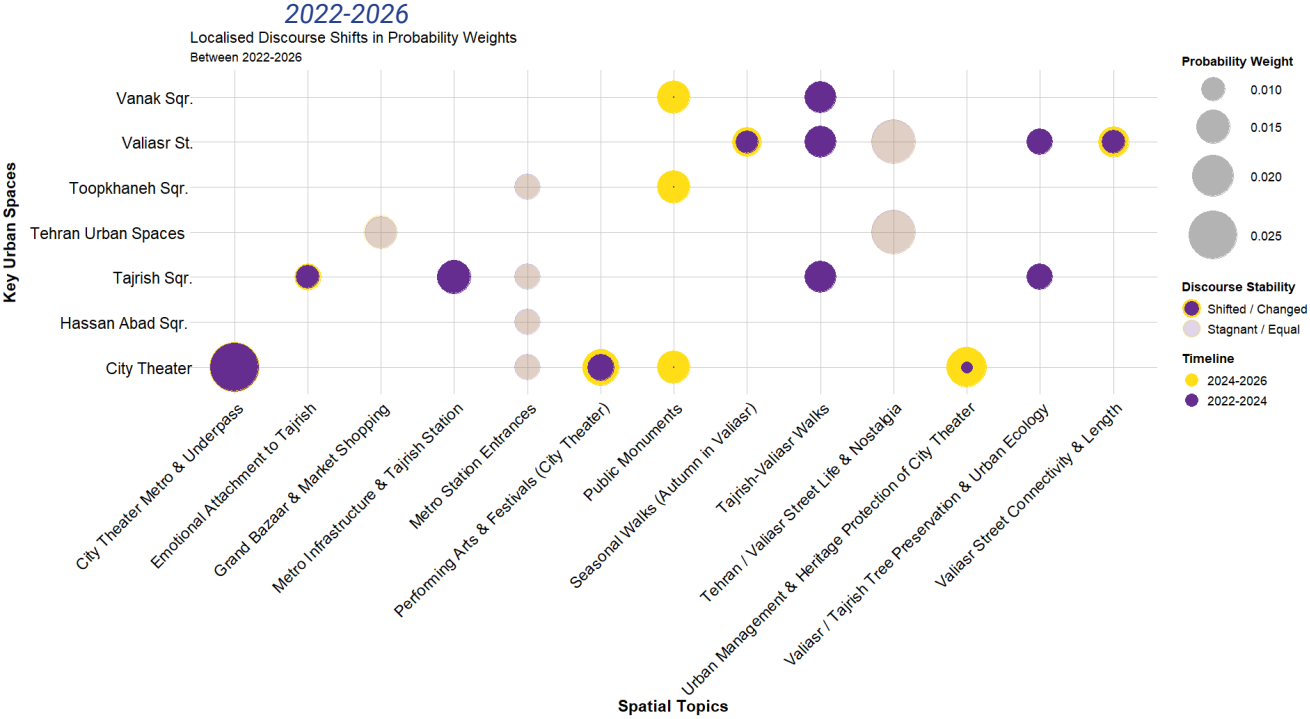



*Figure 43. Topic-sentiment shifts with weighted scores between 2022-2026*



*Figure 44. Discourse shifts with probability weight scores between 2022-2026*



# Synthesis

The impact that each space can have on the public interest in X (formerly Twitter) may vary based on scale, background and popularity. However, this only covers the mental aspect of these spaces. What about the spatial recognition? The significance of these spaces was defined by their position, function, connectivity, accessibility and livability. A square can be prominent because it creates a connection to main streets or arterials within the city. Sometimes, they may have specific functions that attract a large population of city dwellers, such as serving as a commercial hub or retail center. The chosen key places were located at a junction of the transportation network for both vehicle and public transport modes, like BRT[1] and metro stations.



This spatial impact on the general discourse varies from place to place and element to element. What I mean is that Lynch's cognitive urban elements carry different levels of discourse and attraction depending on the component. For example, Valiasr Street, as a path, holds a range of emotions. Its northern segments evoke more nostalgia and attachment than the southern ones. The more inclusive an element is, the more layered its emotional significance becomes. With monuments or landmarks, however, the discourse wasn't as nuanced; opinions tend to be either supportive or opposed, for they don't have the capacity for multidimensional interactions. But, imagine places like Tajrish or City Theater where activity, mobility and leisure meet. Also, add up the environmental elements on top of it, e.g., Valiasr plain trees that grew their way to the Tajrish and City Theater intersections.

The main outcome of the results was to highlight the topics and corresponding sentiment values, specifying the key locations and topics that were most relevant to urban design principles. The three key locations of City Theater, Valiasr street and Tajrish square within bigger urban contexts were picked as the main platforms of localised spatial topics that had a sufficient variety of spatial topics, descriptions and sentiment labels that were applied in the design process as policies, design interventions, or no actions.

The combination of prominent topics, their sentiment values, and corresponding key places was the basis to commence the design process. The localised spatial topics provided accuracy in where to focus, e.g., Tajrish-Valiasr-City Theater. Then, the given descriptions by the topic and overarching discourse of tweets gave the precision that we needed to distinguish which topic fits within which urban aspect. The given sentiment labels also highlighted the quality of those aspects, helping to identify problems and potentials of design quality and spatial conditions based on whether they're negative, positive, or neutral.

1. Bus Rapid Transit

# Design Framework

Once the analysis was complete and the key locations were outlined. The process of creating a design framework was developed within this section. Three main zones were picked, which included the previously introduced key places within, such as Vali'asr, Tajrish, City Theater, etc.

The framework proposes an additional component to the design process (Figure 45), defined by Van Dooren et al. (2014), which added guiding values that emerged from topic evaluations from the public discourse in X. There were visualizations made to explain how to approach a context based on spatial topics, descriptions, and sentiments, addressing the spatial features within each study scope. This helped me to approach the site from existing public discourses as an initial step towards design processes.



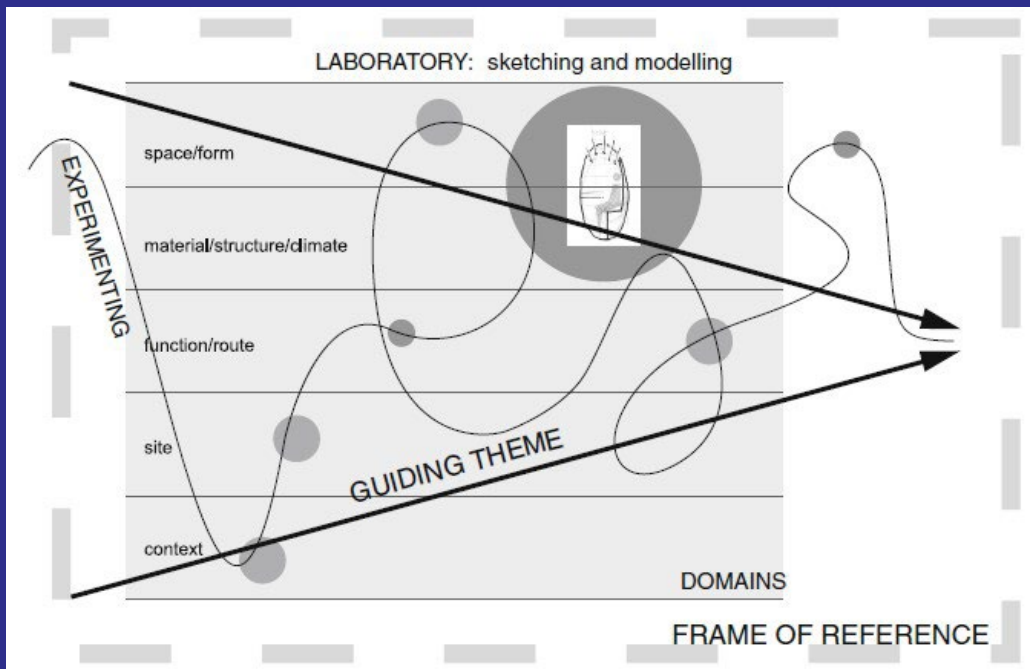



*Figure 45. Generic elements in the design process*

*Source: Van Dooren et al, 2014.*

Van Dooren et al. (2014) describe the design as a process, including several experimentations, traversing between specific domains with which you make an impact (Figure 46). Although you're free to move through domains, you'll narrow down your process with a guiding theme. The vertical and horizontal driving forces would lead the designer to a detailed, finalized and concrete output.

My proposal aligned with public sentiment values towards spatial features within the context. The inputs for data collection were driven out of the site/urban context. The locations were looked up on social media and the corresponding data was collected. After the preprocessing and processing steps, the topics with sentiment values were identified, the prominent key spaces within the overall discourse were framed.



Once the key spaces were evaluated sentimentally, the topics and descriptions were allocated to other design domains within the design framework. The allocations came with the designer/planner›s interpretation, as public topics may not be as accurate and precise in the urban studies field. The urbanist should acknowledge the valued descriptions as spatial conditions, design qualities, issues that require design intervention, policy, further investigation, or no action. Any mitigation depends on which values each topic-tweet synthesis holds.

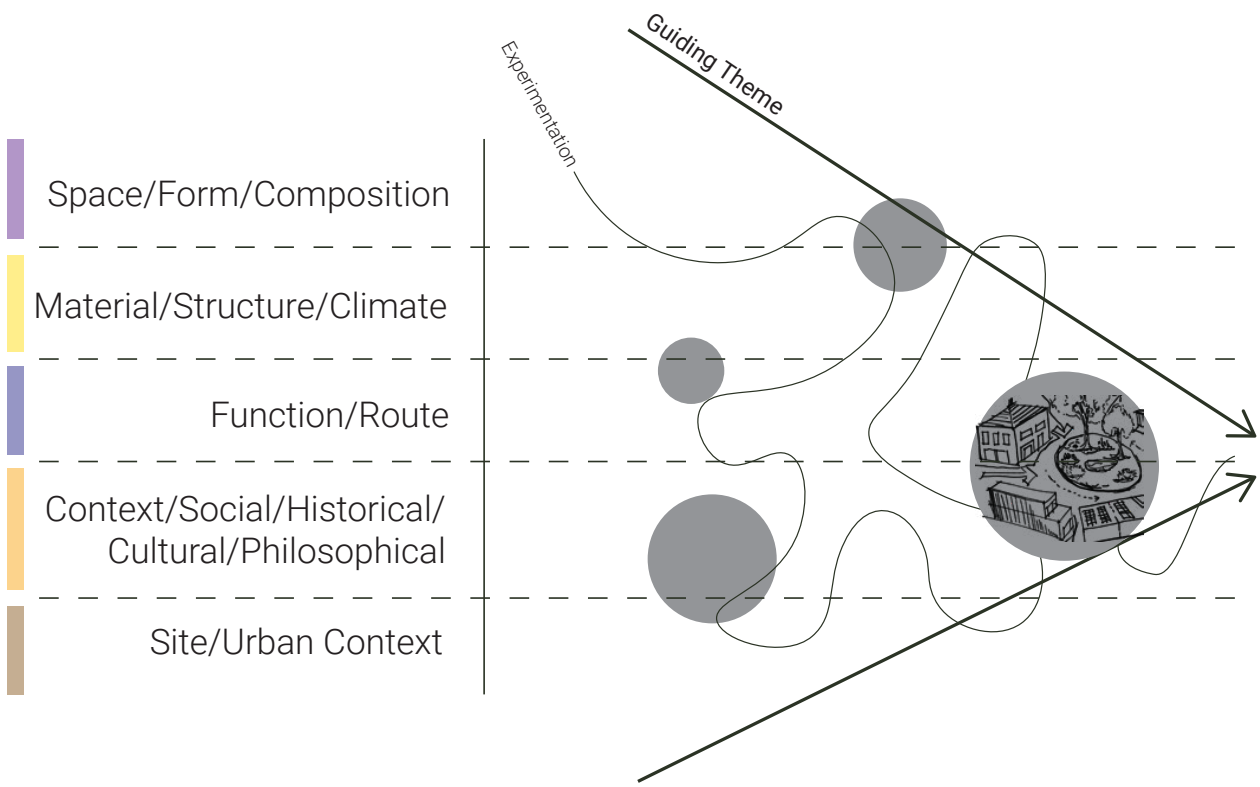
The designer has the authority to pick the guiding values that align with their guiding theme or the project's assignment and objectives. They can produce design qualities based on those values, or add some additional features to their major master plan, define preservative territories within the context, or simply take no action and suggest further exploration regarding neutrally valued features. Therefore, some guiding values may be used at any step through the process, and some may remain ignored or identified as unnecessary. Furthermore, the guiding values may result in readjusting the guiding theme. For example, in case the urbanist wants to strictly comply with citizen inputs, they may decide to readjust the scope of their guiding theme, which fits more public-driven guiding values within. This supports the idea to expand the social capacity in which the project is built, bringing an initial legitimacy towards active citizen engagement.

As an example of implementing the proposed framework, I picked the three main key places of City Theater, Vali'asr Street and Tajrish Square within the bigger urban frames, including other key places. These places had prominence over the topics and sentiments and were both mentally and spatially interrelated. Then, I allocated my chosen values to design domains based on their corresponding topics to specify which description belonged to which domain. After that, I made the policy-design table of design domains corresponding to each key location. The items were subsets of interpreted topics and consist of restrictive policies, design interventions, spatial qualities, organization and configurations.

The framework was based on two main paths: first, the analytical process, and second, the analytical output implementation on design process. The framework in Figure 47 was divided on this basis and the steps were elaborated in the following sections. For clearer instructions, in Figure 47, the "A" prefix stands for "Analysis", and the "D" stands for "Design".

*Figure 46. Moving between domains as the experimentation*

*Source: Van Dooren et al, 2014.*



# Analytical Process

### A1. Urban Context

We start with the urban context that can vary from district to city scale, consisting of all existing spatial features within the urban spaces. We'd also need the citizens' data as a complementary source to run the analysis on.

### A2. Accuracy gain with general spatial topics

Once you run the suitable analysis that gives out the public discourse topics, e.g., topic modelling on social media data, either spatial or non-spatial, which take place across the city in different locations. However, not all locations would have the same amount of public attraction. That's where you'll gain accuracy on the locations that should be focused on based on public discourse on a city scale.



### A3. Precision on places with localised spatial topics

Once you filter the spatial topics and categorize them into general and localised ones based on your project's study area, you'll gain precision on places corresponding to specific topics holding sentiments. For example, a main street is addressed in the topic. At this step, you'll notice where on the street is mentioned specifically.

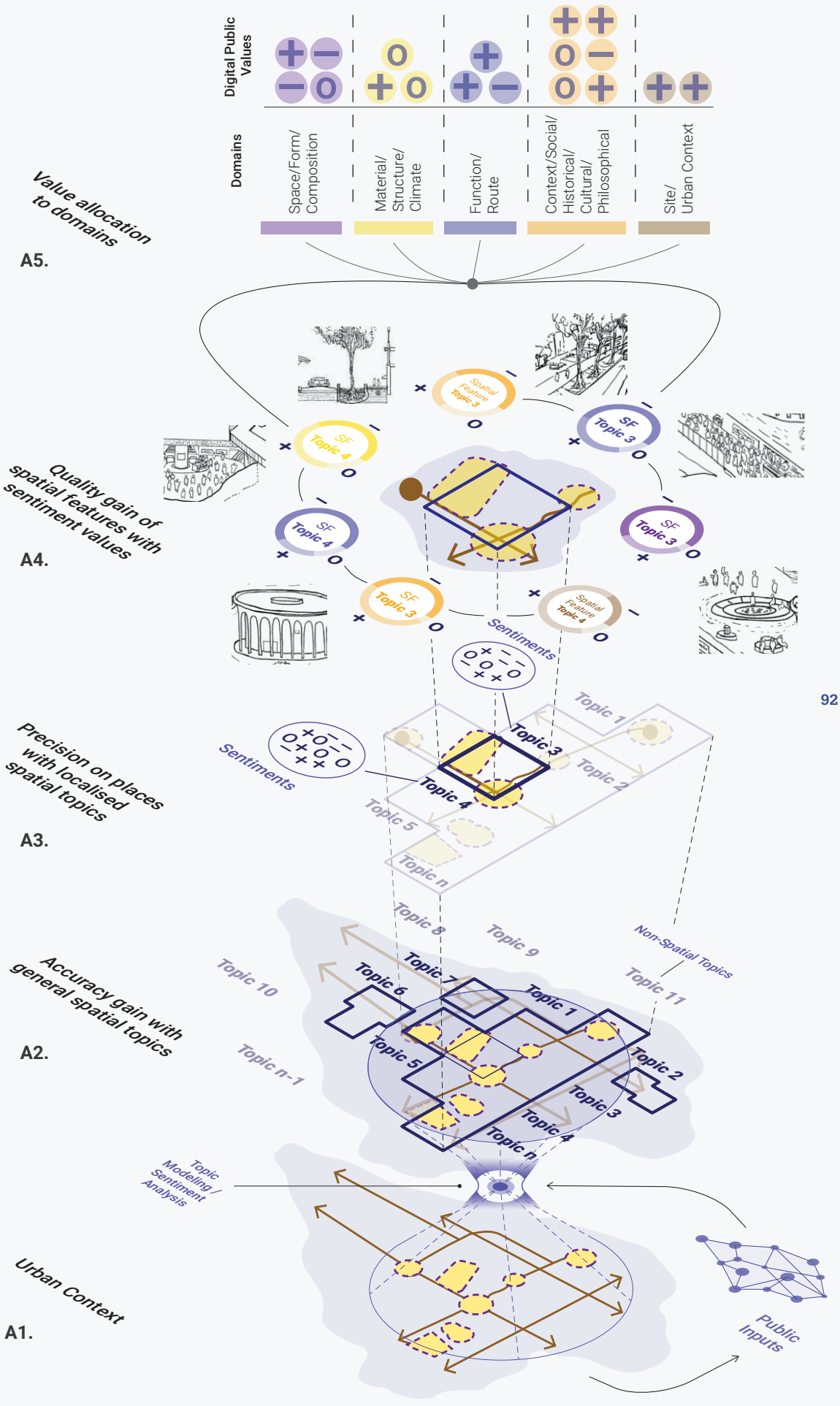
### A4. Quality gain of spatial features with sentiment values

When you land on a precise location, you'd gain the quality of spatial features graded by sentimentally-valued descriptions addressing the urban elements. This is when the spatial association happens. Also, the sentiment proportions differ from element to element. Based on the analysis, some features may contain highly negative values, and some may be highly positive or even moderately neutral. It's up to the designer to explore and pick the prominent sentimental value based on the descriptions given to each spatial feature. The designer can synthesize the topic and the representative descriptions to highlight the features as clearly as possible.

### A5. Value allocation to domains

At the final stage, all of the topics holding the valued descriptions should be interpreted and categorized into design domains. The spatially associated values should be allocated into domains; some topics would cover multiple domains, yet their described values are allocated into each domain more specifically. This builds up the basis as an additional component through the design process.

*Figure 47. Proposed Analysis & Design Framework*


Digital Public Values
Domains
Space/Form/ Composition
Material/ Structure/ Climate
Function/ Route
Context/Social/ Historical/ Cultural/ Philosophical
Site/ Urban Context
Value allocation to domains
A5.
Spatial Feature Topic 3
SF Topic 3
SF Topic 4
Spatial Feature Topic 4
Quality gain of spatial features with sentiment values
A4.
Sentiments
Topic 1
Topic 2
Topic 3
Topic 4
Topic 5
Topic n
Precision on places with localised spatial topics
A3.
Topic 6
Topic 7
Topic 8
Topic 9
Topic 10
Topic 11
Topic n-1
Non-Spatial Topics
Accuracy gain with general spatial topics
A2.
Topic Modeling / Sentiment Analysis
Urban Context
A1.
Public Inputs

# Design Process

### D1. Beginning of Design

At first, you'd begin the process with a set of features that are positively, negatively and neutrally valued among the domains. You'll set the guiding theme along with this component. These would be the guiding values through your design process. They're not blocks or limitations to the designer's procedure, but more like hints, signs, or blinking lights that would draw the designer's attention to the public view and interests.

### D2. Design under progress with Guiding Values

The designer starts the process and traverses across domains in their experimentation. The designer is to choose whether to utilize the guiding values in their process or not. The values are either a memorial element or an expression of the current state. The further the designer goes, the clearer the associated values become. They become more concrete and tangible as the designer produces work using the values they've chosen. The progress goes on while the designer examines whether the values fit in the project's guiding theme or not.

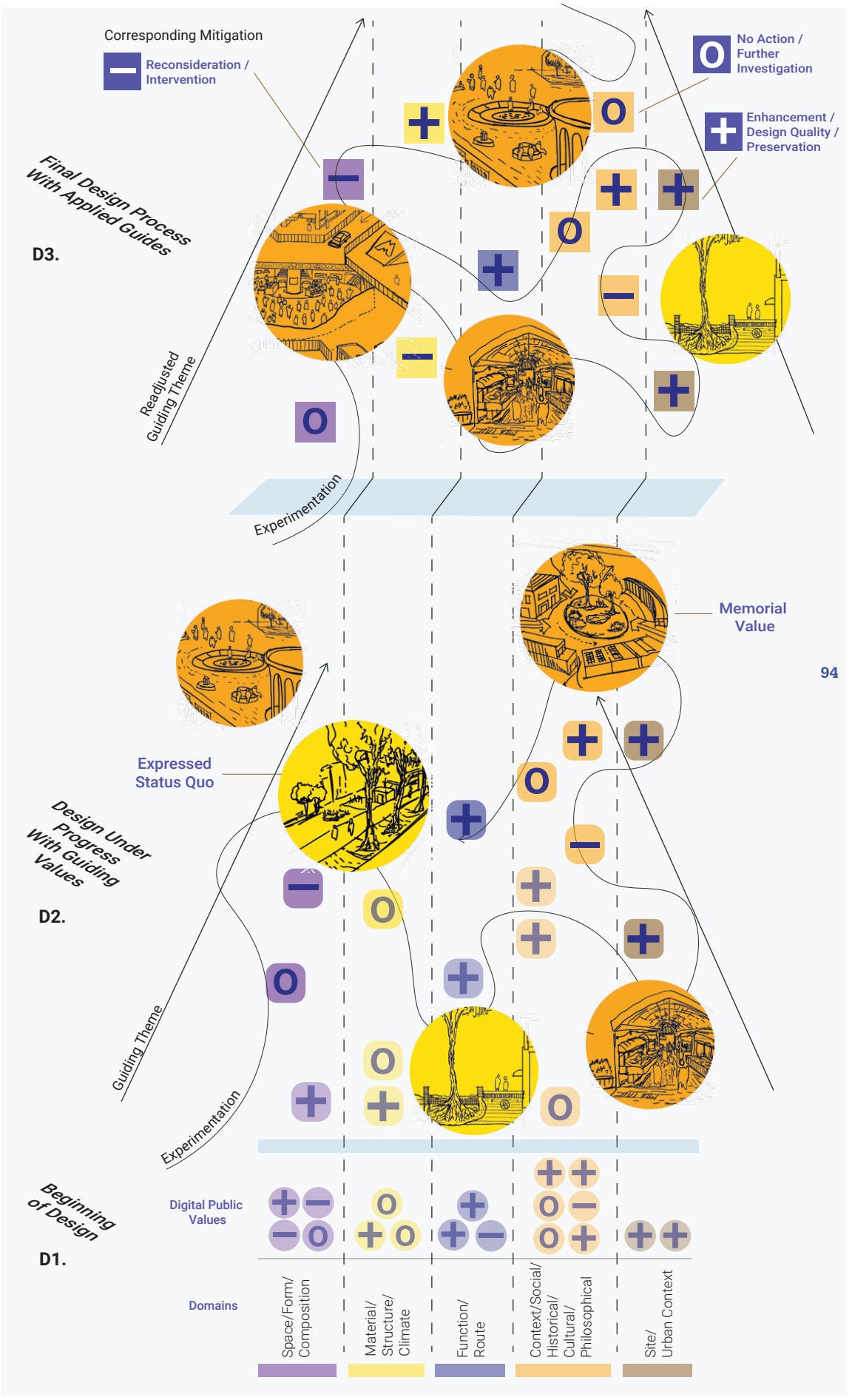


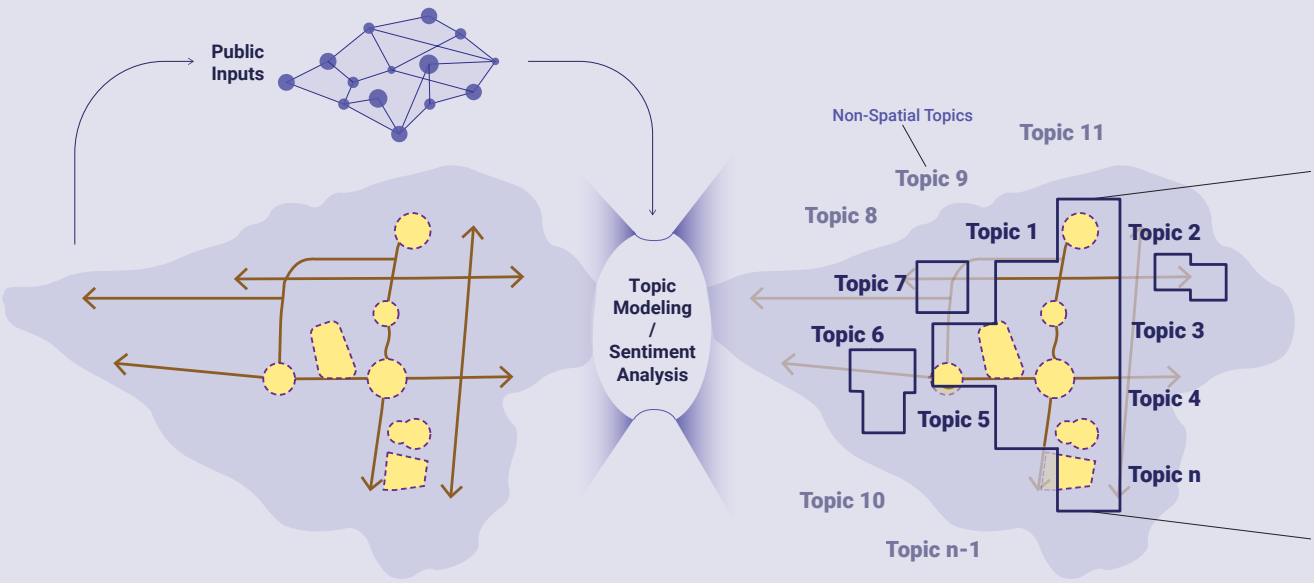
### D3. Final design process with applied Guides

The further the process approaches an acceptably finalized stage, the designer decides if they want to readjust the guiding theme to fit in more values in their scope or not. The production the designer will have varies from value to value. Each value requires a different approach to mitigating the issues and situations. As an example, for negatively valued features, reconsideration or intervention is recommended, while for neutral values, no action or further investigation is advised. At the end, it's up to the designer's decision through the process they go through, ending up in combining the mitigations for their design work, readjusting the guiding theme to fit in more public-driven values, or ignoring the values that are not quite relevant to their work.

In the next section, I elaborated on how the analytical and design process would be applied in the case study and how the mitigations would look, representing the framework's implementation in the chosen key places of Tajrish Square, Vali'asr Street, and City Theater.

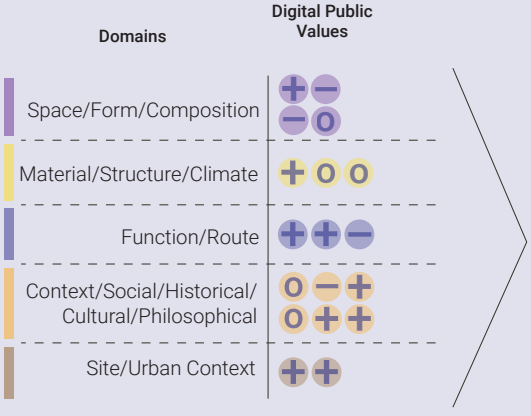
*Figure 48. Proposed Analysis & Design Framework*


Corresponding Mitigation
Reconsideration / Intervention
No Action / Further Investigation
Enhancement / Design Quality / Preservation
Final Design Process With Applied Guides
D3.
Readjusted Guiding Theme
Experimentation
Memorial Value
Expressed Status Quo
Design Under Progress With Guiding Values
D2.
Guiding Theme
Experimentation
Beginning of Design
D1.
Digital Public Values
Domains
Space/Form/ Composition
Material/ Structure/ Climate
Function/ Route
Context/Social/ Historical/ Cultural/ Philosophical
Site/ Urban Context

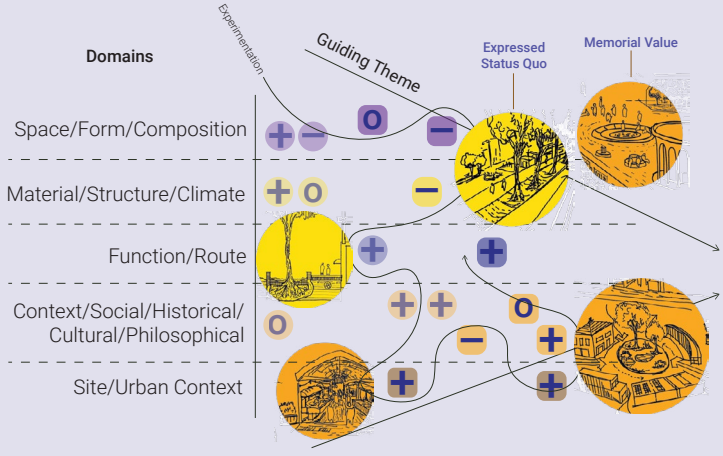



**A1. Urban Context**

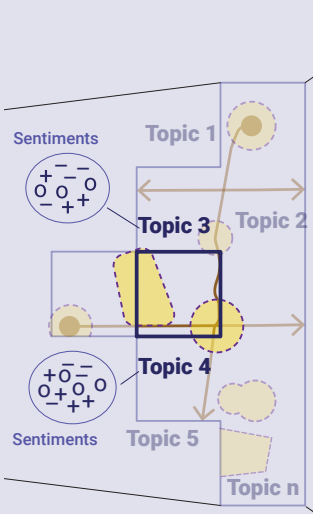
**A2. Accuracy gain with general spatial topics**



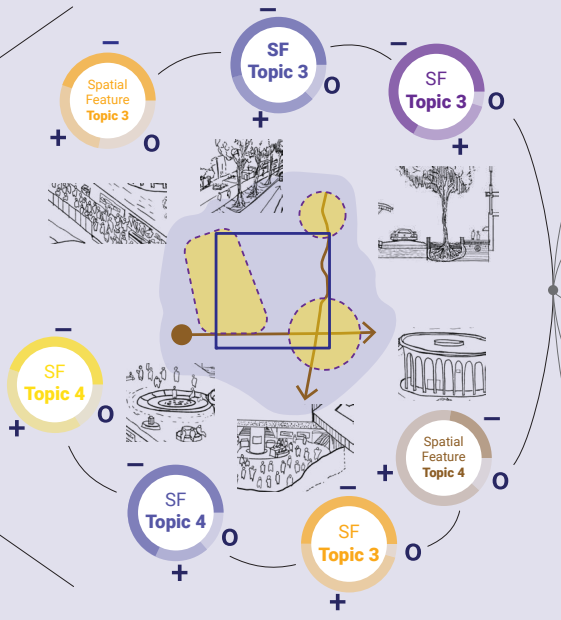

**D1. Beginning of Design**



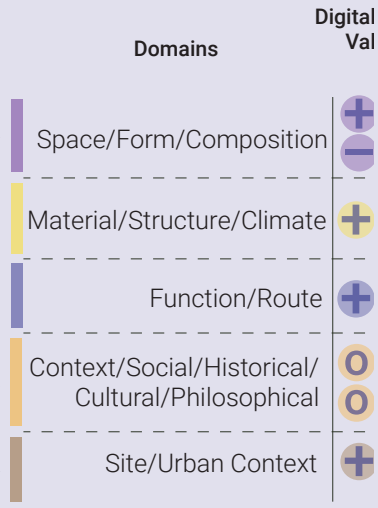

**D2. Design Under Progress With Guiding Values**



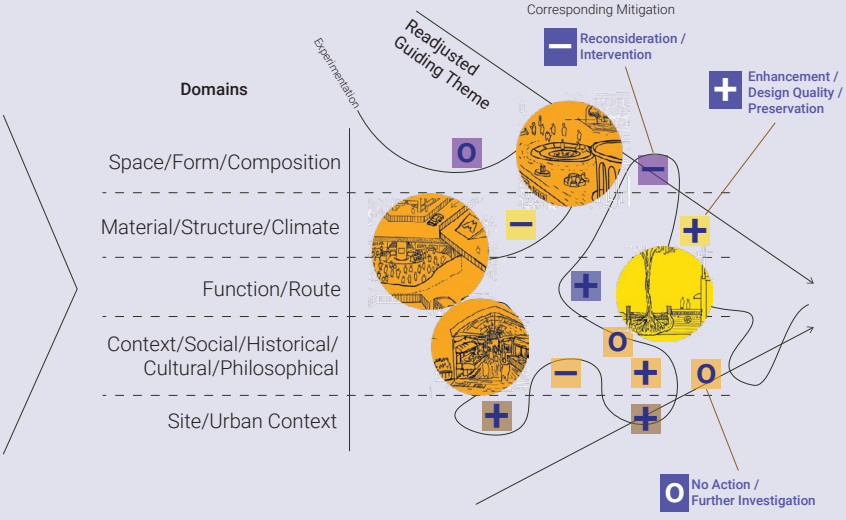

**A3. Precision on places with localised spatial topics**




**A4. Quality gain of spatial features with sentiment values**




**A5. Value allocation to domai**




**D3. Final Design Process With Applied Guides**

# Examples of the Process in the Study Area

In Figure 49, 50, and 51, the necessary steps after collecting and processing the data, resulting in spatially localised topics, towards transitioning from analysis to design, were visualized. A snippet of Figure 49's illustration was placed as Figure 48 and is used as a sample of describing the steps of applying the framework to the case study.

On Figure 48, the large-scale map on the right, there's **A1**'s urban context. The outlined places in yellow represent the **A2**'s accuracy I gained in identifying the locations based on topics. Then, there are zoomed-in illustrations of the key places' current state. The colored circles represent the category of the highlights. The orange is a memorial value and the yellow represents the status quo expression. This was backed by the given descriptions from the tweets regarding the key places, which gave me precision in locating the spatial features as described in **A3**.

The **A4** was taken as the next step in integrating two analytical components: a synthesis of the topic and the representative tweet, and the valued spatial feature based on its prominent sentiment. In Figure 48, the tree line and sidewalk profile were highlighted and annotated as "Tajrish-Valiasr Walks", which was a synthesis of the topic and its representative tweet. Also, the overall sentiment value regarding the trees and sidewalks was positive. Therefore, the described spatial feature was associated with a positive value.

The last step of transition was to translate the associations into design domains and allocate the values into each domain separately. This was the step that required the expert's urban knowledge. Not only interpreting each valued description into a specialized subject for mitigation, but also allocating them into design domains relied on the designer. As can be seen in Figure 48, the **A5** step is where "Valiasr-Tajrish Walks" was interpreted as "Sense of Place" and "Pedestrian Walkability and Accessibility" subjects. Also, they were allocated into the domains of "function/route" and "context/background", color-coded in blue and orange, respectively.

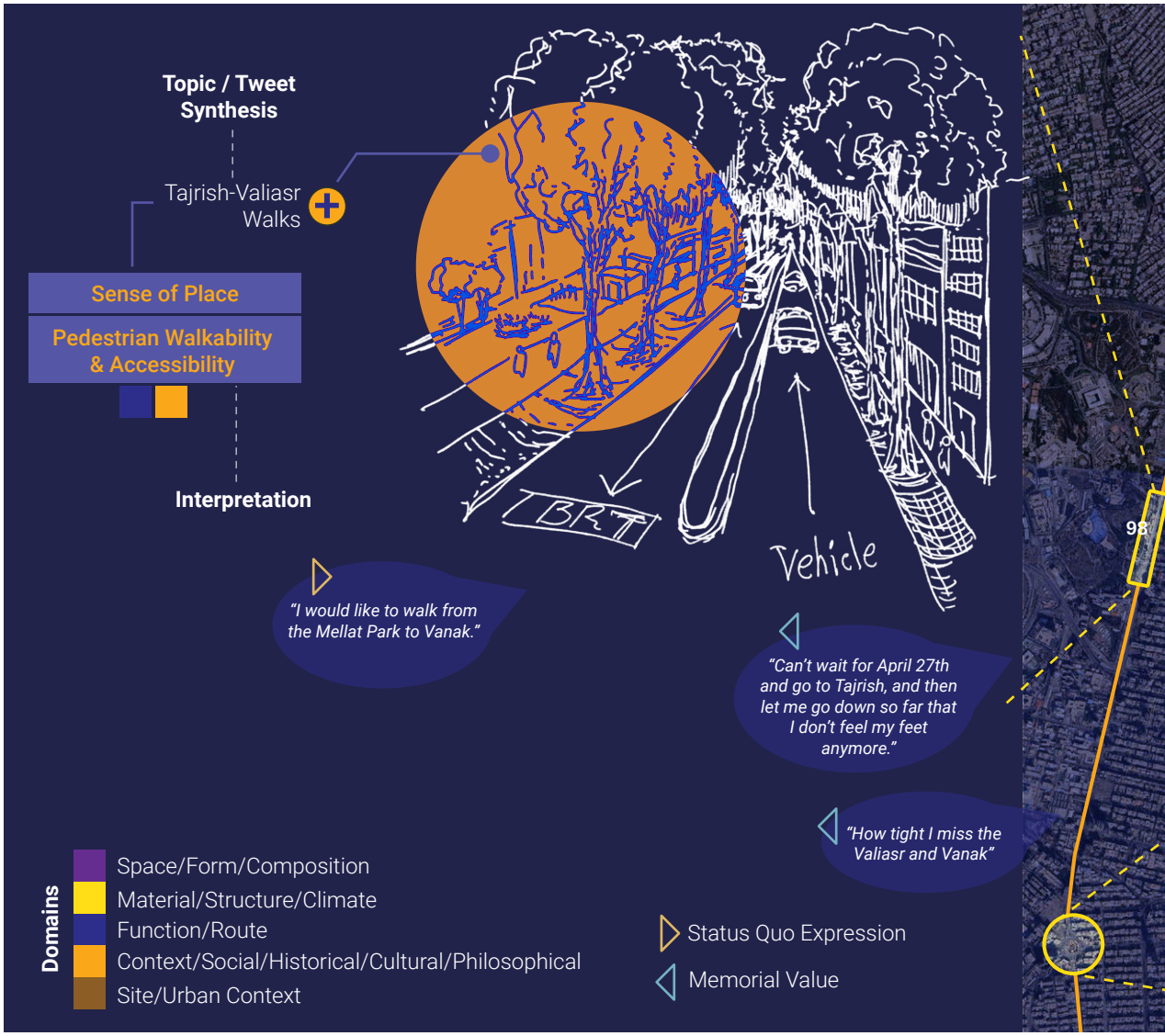



*Figure 49. Zoom-in of framework's example on Vali'asr*

# Tajrish - Vali'asr

## 1st Zone

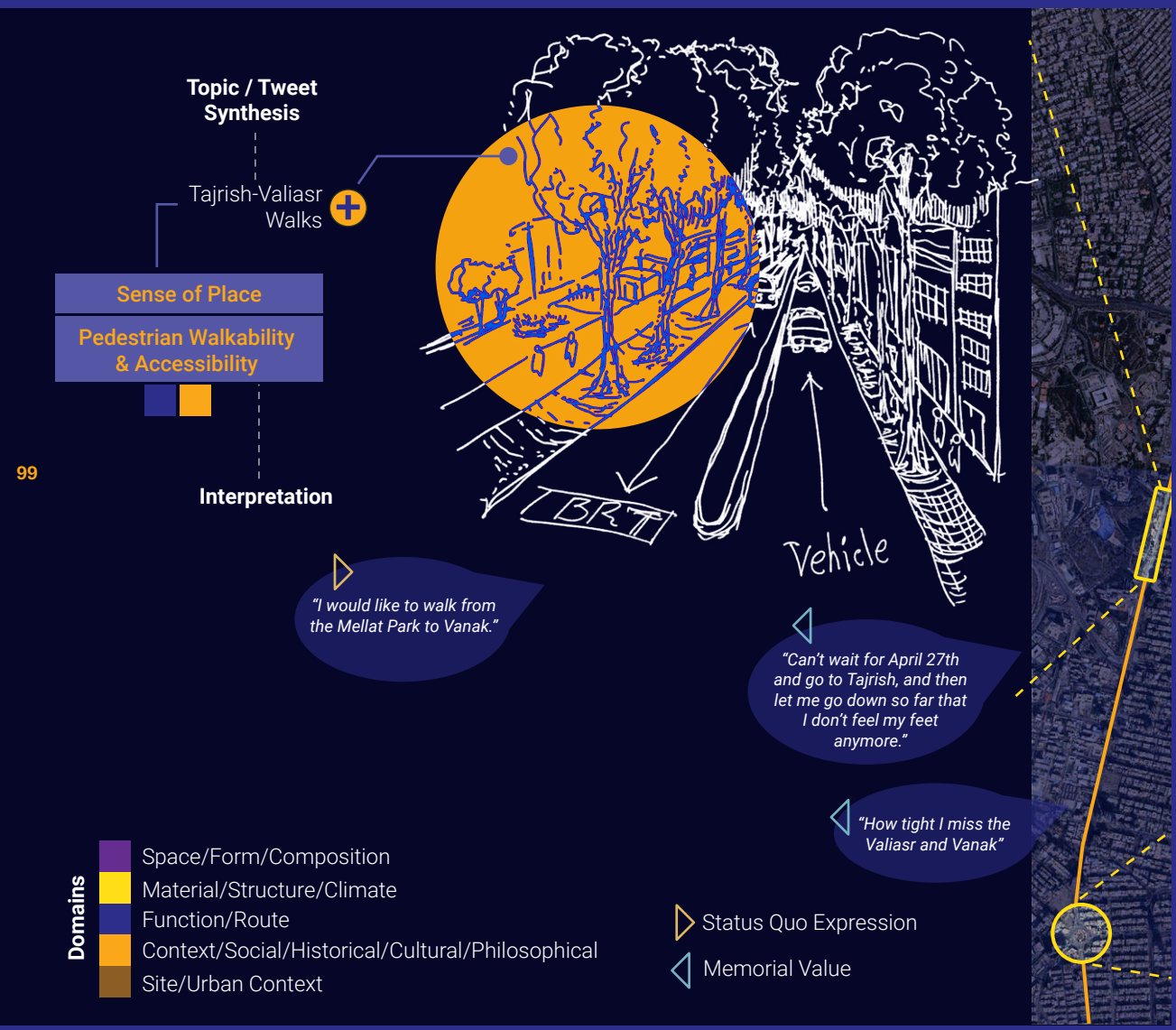
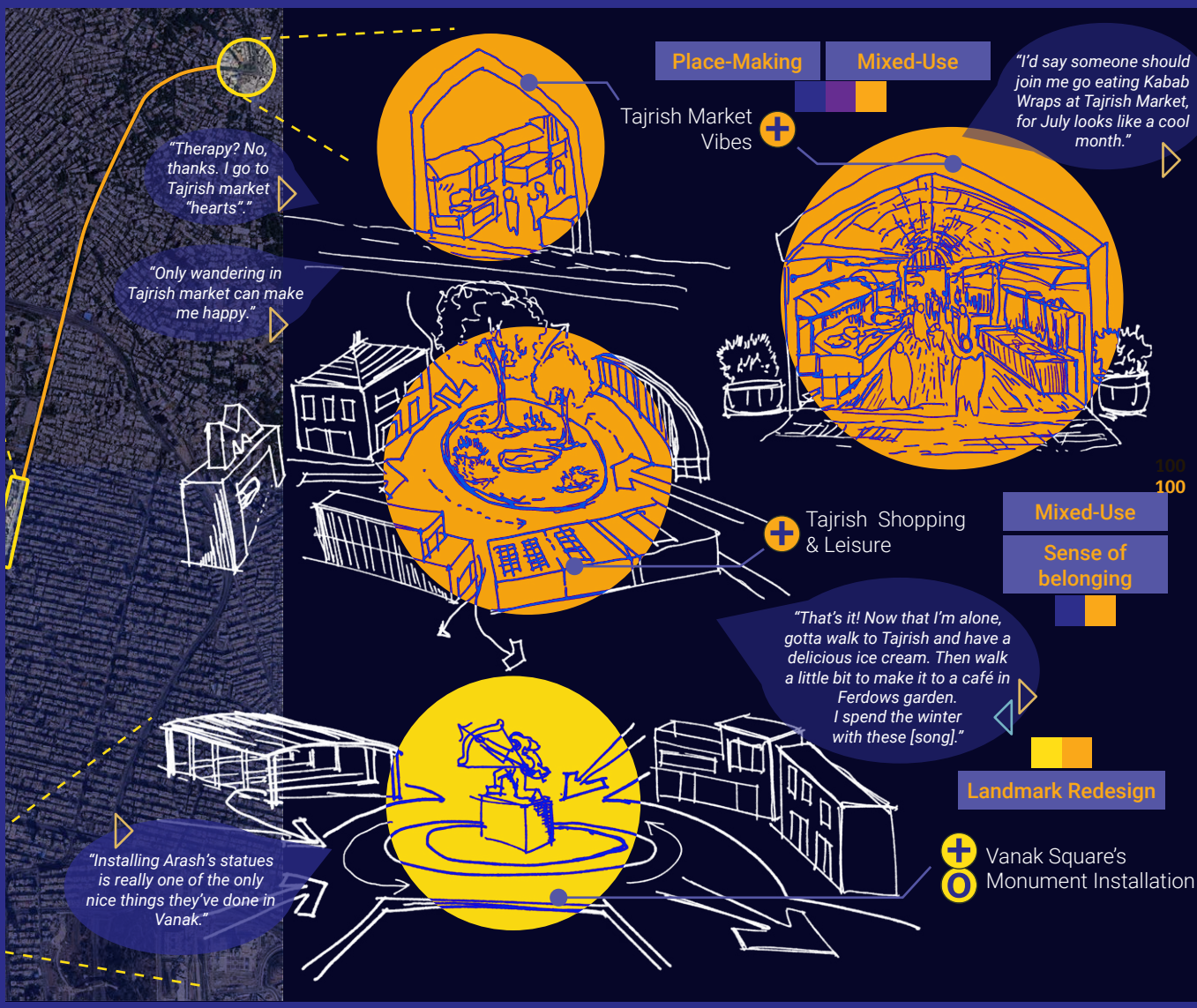





*Figure 50. Framework application example on Tajrish & Vali'asr*


Place-Making
Mixed-Use
Tajrish Market Vibes
"I'd say someone should join me go eating Kabab Wraps at Tajrish Market, for July looks like a cool month."
"Therapy? No, thanks. I go to Tajrish market "hearts"."
"Only wandering in Tajrish market can make me happy."
100
100
Mixed-Use
Sense of belonging
Tajrish Shopping & Leisure
"That's it! Now that I'm alone, gotta walk to Tajrish and have a delicious ice cream. Then walk a little bit to make it to a café in Ferdows garden. I spend the winter with these [song]."
Landmark Redesign
"Installing Arash's statues is really one of the only nice things they've done in Vanak."
Vanak Square's Monument Installation

# Vali'asr - City Theater

## 2nd Zone

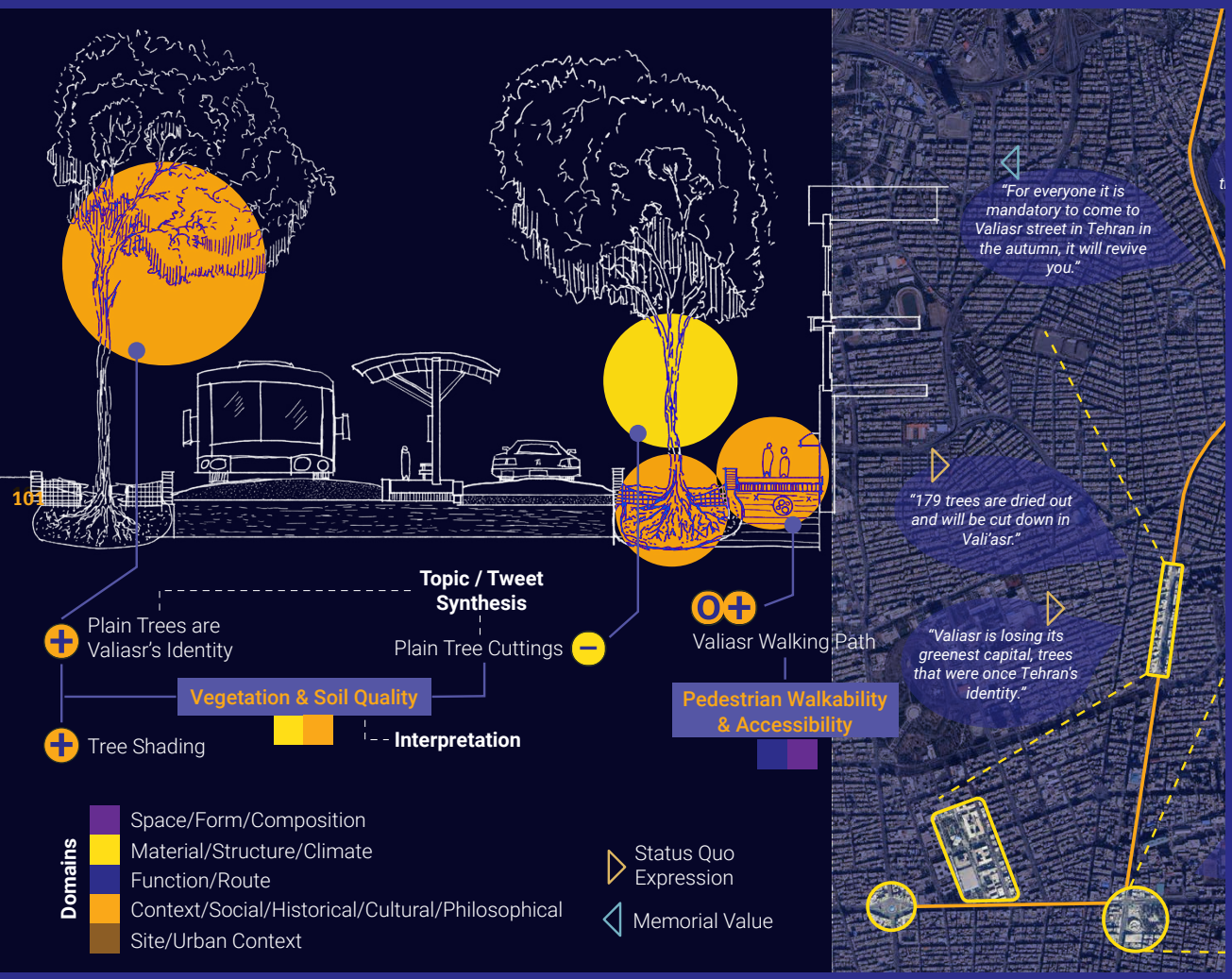
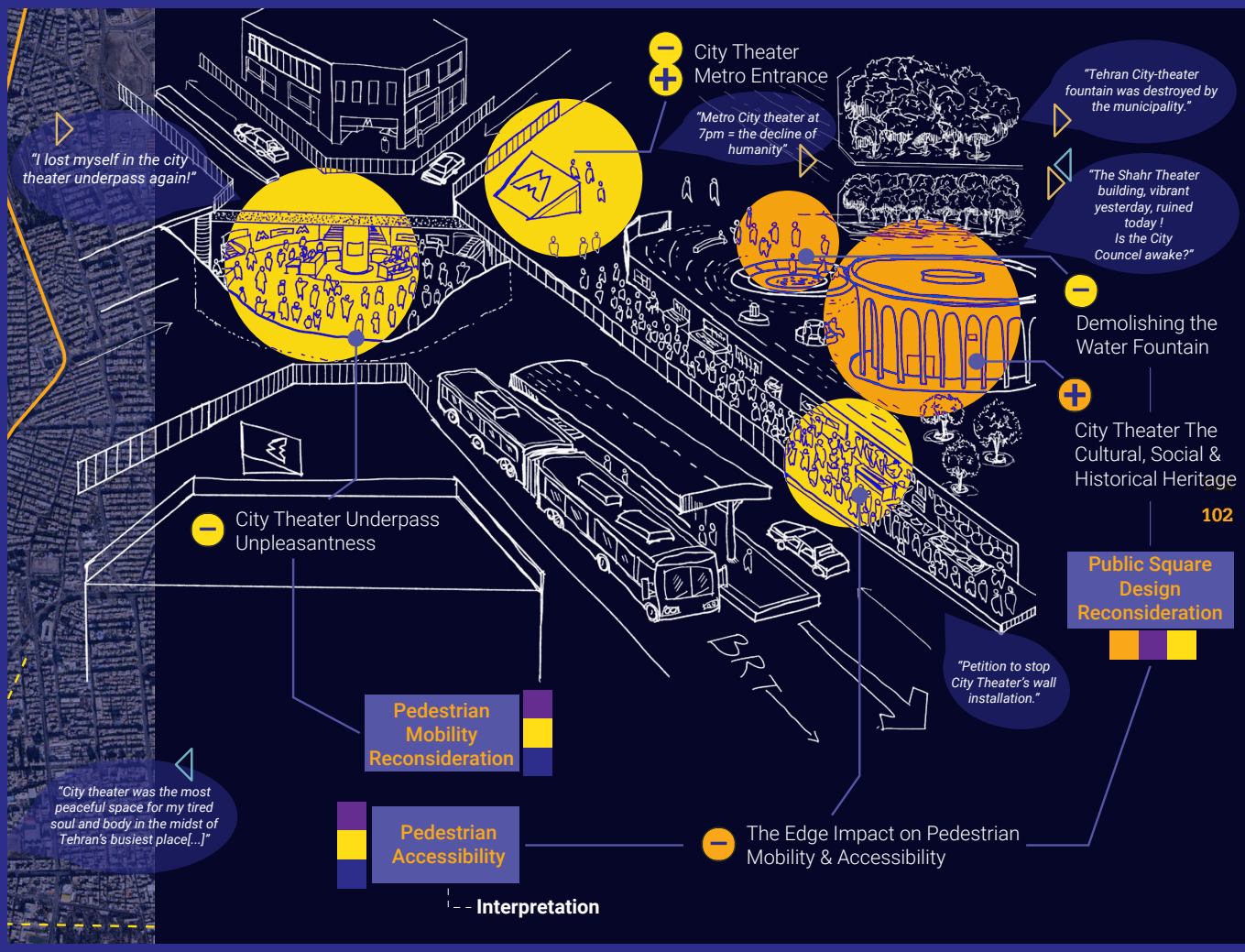



*Figure 51. Framework application example on City Theater & Vali'asr*


City Theater
Metro Entrance
"I lost myself in the city theater underpass again!"
"Metro City theater at 7pm = the decline of humanity"
"Tehran City-theater fountain was destroyed by the municipality."
"The Shahr Theater building, vibrant yesterday, ruined today ! Is the City Councel awake?"
Demolishing the Water Fountain
City Theater The Cultural, Social & Historical Heritage
102
Public Square Design Reconsideration
City Theater Underpass Unpleasantness
BRT
"Petition to stop City Theater's wall installation."
Pedestrian Mobility Reconsideration
"City theater was the most peaceful space for my tired soul and body in the midst of Tehran's busiest place[...]"
Pedestrian Accessibility
The Edge Impact on Pedestrian Mobility & Accessibility
Interpretation

# Vali'asr - City Theater

## 3rd Zone

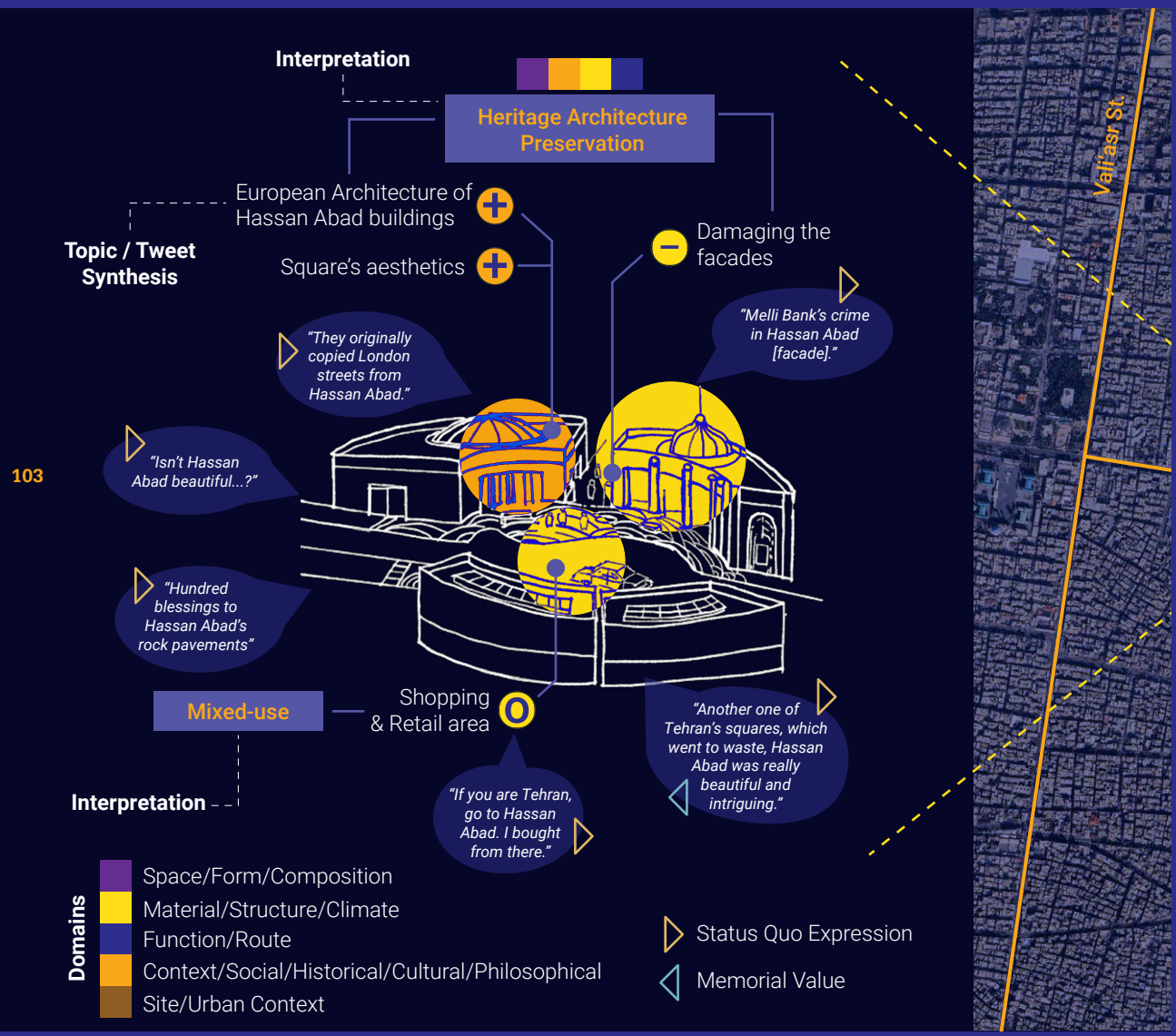
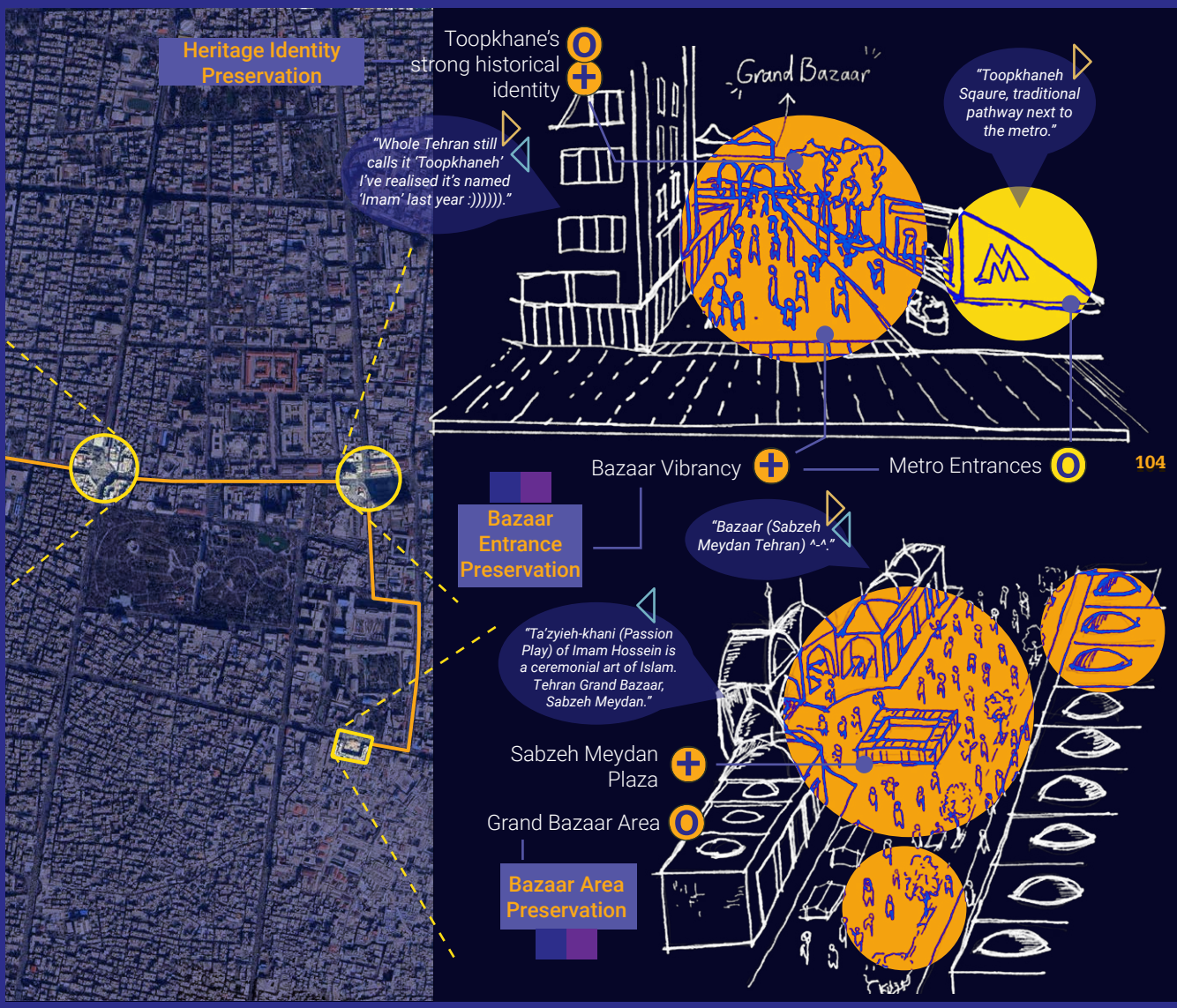





*Figure 52. Framework application example on Toopkhaneh, HassanAbad and Grand Bazaar*


Heritage Identity Preservation
Toopkhane's strong historical identity
"Whole Tehran still calls it 'Toopkhaneh' I've realised it's named 'Imam' last year :)))))."
Grand Bazaar
"Toopkhaneh Sqaure, traditional pathway next to the metro."
Bazaar Vibrancy
Metro Entrances
Bazaar Entrance Preservation
"Bazaar (Sabzeh Meydan Tehran) ^-^."
"Ta'zyieh-khani (Passion Play) of Imam Hossein is a ceremonial art of Islam. Tehran Grand Bazaar, Sabzeh Meydan."
Sabzeh Meydan Plaza
Grand Bazaar Area
Bazaar Area Preservation

# Example of Mitigations Based on the Guiding Values

A table of design/policy mitigations based on the outputs of the previous steps was made (Figure 52). This product was considered a textualized product that could be developed through the design process, tackling the challenges identified in the study area in a documented way.

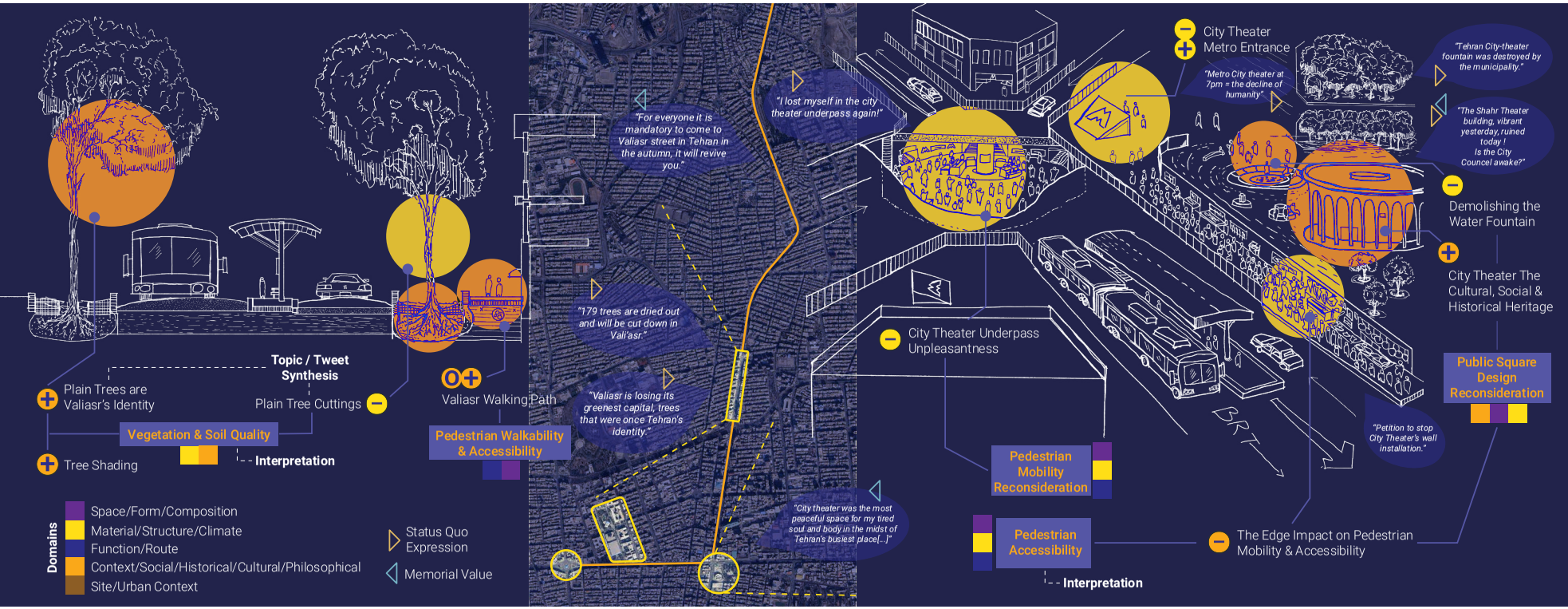
The mitigations were categorized based on the domains and relevant to their specialized subjects within the three key locations of the study. Each mitigation corresponded to its domain and complied with the subject in which it was specialized. The mitigations were defined based on their given values. They can be a restrictive policy, a spatial intervention such as installation and configuration, or a no-action advice.



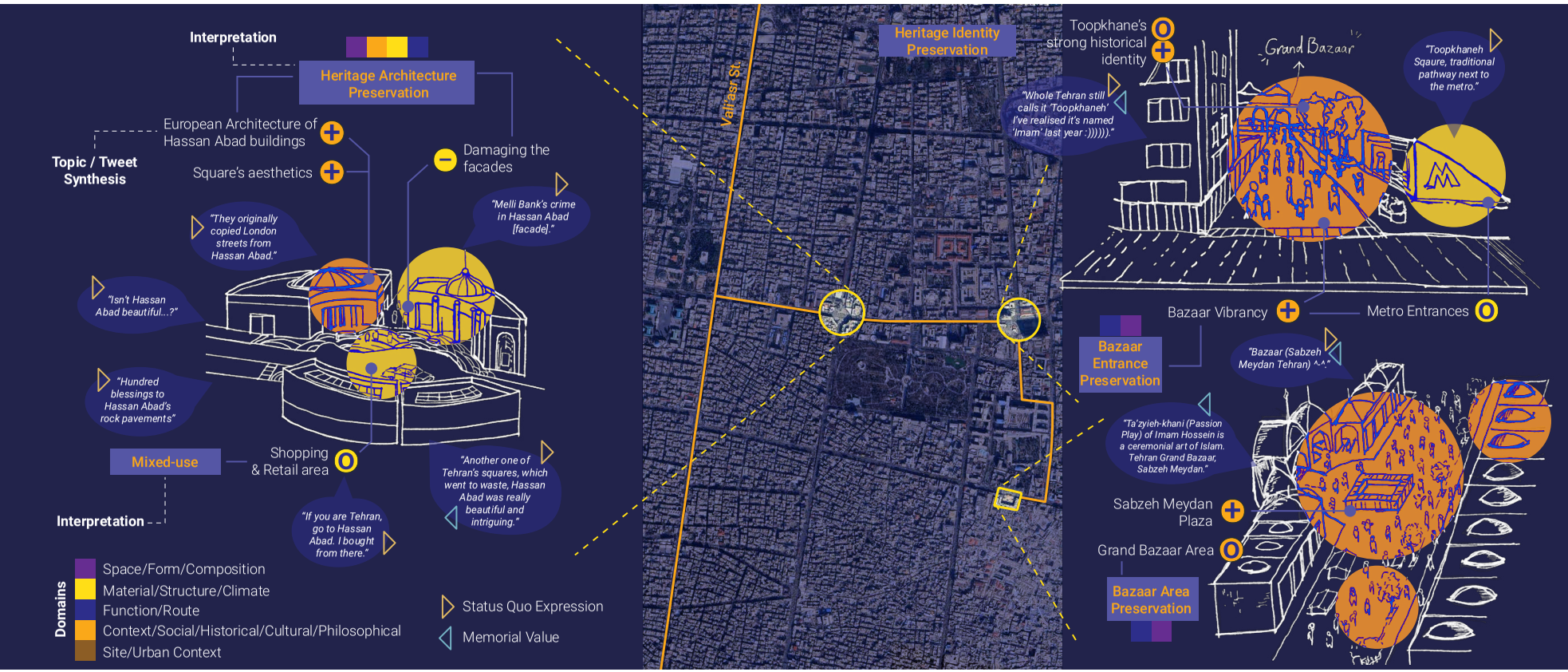
The given table was only an example of having a preliminary action plan before commencing the design, with developing an initial set of guiding values that will be used throughout the design process. The table elements can be modified, switched, or removed at any step of the experimentation through the process. Having the policies and interventions in a textual format would aid the experts in clarifying their mitigations and items they utilized through their design journey.

*Figure 53. Policy / Design mitigations*

| Design Domain | Policy / Design Mitigations | | |
|---|---|---|---|
| | Tajrish | Vali'asr | City Theater |
| Space/ Form/ Composition | Place-Making<br>• Sitting spaces within the market area<br>• Highlighting gathering spots around the retail and leisure functions.<br>Sense of Belonging<br>• Preserve the existing market space with historical architecture<br>• Preserve the identical territories of the market and square | Pedestrian Walkability & Accessibility<br>• Spatial connection of pedestrian paths all along the street with other connections<br>• Sidewalk space configuration based on street width and vegetations<br>• Urban furniture design and placement<br>• Visual and physical consistency of pavement<br>• Pedestrian flow | Pedestrian Walkability & Accessibility<br>• Sidewalk space configuration based on street width and vegetations<br>• Sidewalk & square design based on pedestrian flow<br>• Open square space organization & configuration |
| Material/ Structure/ Climate | – | Vegetation & Soil Quality<br>• Strict greenery preservation regulations<br>• Tree & soil profile re-engineering<br>• Water drainage and collection system<br>• Grass pocket design along the sidewalks<br>• Porous pavement (water drainage) | Pedestrian Walkability & Accessibility<br>• Pedestrian-priority elevated ramp<br>• Urban furniture design & placement<br>• Visual and physical consistency of pavement<br>• Underpass profile redesign |
| Function/ Route | Mixed-Use<br>• Placing diverse functions and activities around the square area<br>• Improving the activities around the metro station<br>Pedestrian Walkability & Accessibility<br>• Spatial connection of pedestrian paths all along the square with the surroundings<br>• Pedestrian priority path between the square, metro station and the market | Pedestrian Walkability & Accessibility<br>• Spatial connection of pedestrian paths all along the street with other connections<br>• Pedestrian flow<br>• High-demand activities within the walking distances at each street segment | Pedestrian Walkability & Accessibility<br>• Spatial connection of paths across the square and junctions<br>• Sidewalk & square design based on the pedestrian flow |
| Context/ Social/ Historical/ Cultural/ Philosophical | Place-Making<br>• Leverage of hotspots in the market and square<br>Sense of Belonging<br>• Historical value of Tajrish square<br>• Socio-economic identity of the Tajrish square in Tehran's cultural history | Ecological Heritage<br>• Highlighting the trees as the Valiasr's living identity within Tehran<br>• Acknowledging the historical significance of Valiasr's plain tree rows. | Public Square Design Reconsideration<br>• Acknowledge the historically/memorially values spatial features within the square<br>• Preserve the valued territories of the square based on daily activities and presence<br>• Leverage the existing identity with new interventions respecting the context |
| Site/ Urban Context | Emotional attachments to Tajrish | Vali'asr street plain trees and walking path | Art/culture intersection at the City Theater |


Topic / Tweet Synthesis
Plain Trees are Valiasr's Identity
Plain Tree Cuttings
Vegetation & Soil Quality
Tree Shading
Interpretation
Valiasr Walking Path
Pedestrian Walkability & Accessibility
"For everyone it is mandatory to come to Valiasr street in Tehran in the autumn, it will revive you."
"I lost myself in the city theater underpass again!"
"179 trees are dried out and will be cut down in Vali'asr."
"Valiasr is losing its greenest capital, trees that were once Tehran's identity."
"City theater was the most peaceful space for my tired soul and body in the midst of Tehran's busiest place[...]"
Domains
Space/Form/Composition
Material/Structure/Climate
Function/Route
Context/Social/Historical/Cultural/Philosophical
Site/Urban Context
Status Quo Expression
Memorial Value
City Theater Metro Entrance
"Metro City theater at 7pm = the decline of humanity"
"Tehran City-theater fountain was destroyed by the municipality."
"The Shahr Theater building, vibrant yesterday, ruined today ! Is the City Councel awake?"
Demolishing the Water Fountain
City Theater The Cultural, Social & Historical Heritage
Public Square Design Reconsideration
City Theater Underpass Unpleasantness
"Petition to stop City Theater's wall installation."
Pedestrian Mobility Reconsideration
Pedestrian Accessibility
The Edge Impact on Pedestrian Mobility & Accessibility
Interpretation


Topic / Tweet Synthesis
Tajrish-Valiasr Walks
Sense of Place
Pedestrian Walkability & Accessibility
Interpretation
BRT
Vehicle
"I would like to walk from the Mellat Park to Vanak."
"Can't wait for April 27th and go to Tajrish, and then let me go down so far that I don't feel my feet anymore."
"How tight I miss the Valiasr and Vanak"
"Therapy? No, thanks. I go to Tajrish market "hearts"."
"Only wandering in Tajrish market can make me happy."
"Installing Arash's statues is really one of the only nice things they've done in Vanak."
Place-Making
Mixed-Use
Tajrish Market Vibes
"I'd say someone should join me go eating Kabab Wraps at Tajrish Market, for July looks like a cool month."
Tajrish Shopping & Leisure
Mixed-Use
Sense of belonging
"That's it! Now that I'm alone, gotta walk to Tajrish and have a delicious ice cream. Then walk a little bit to make it to a café in Ferdows garden. I spend the winter with these [song]."
Landmark Redesign
Vanak Square's Monument Installation
Domains
Space/Form/Composition
Material/Structure/Climate
Function/Route
Context/Social/Historical/Cultural/Philosophical
Site/Urban Context
Status Quo Expression
Memorial Value


Interpretation
Heritage Architecture Preservation
Topic / Tweet Synthesis
European Architecture of Hassan Abad buildings
Square's aesthetics
Damaging the facades
"They originally copied London streets from Hassan Abad."
"Melli Bank's crime in Hassan Abad [facade]."
"Isn't Hassan Abad beautiful...?"
"Hundred blessings to Hassan Abad's rock pavements"
Mixed-use
Shopping & Retail area
Interpretation
"If you are Tehran, go to Hassan Abad. I bought from there."
"Another one of Tehran's squares, which went to waste, Hassan Abad was really beautiful and intriguing."
Domains
Space/Form/Composition
Material/Structure/Climate
Function/Route
Context/Social/Historical/Cultural/Philosophical
Site/Urban Context
Status Quo Expression
Memorial Value
Vali-asr St.
Heritage Identity Preservation
Toopkhane's strong historical identity
"Whole Tehran still calls it 'Toopkhaneh' I've realised it's named 'Imam' last year :)))))."
Grand Bazaar
"Toopkhaneh Sqaure, traditional pathway next to the metro."
Bazaar Vibrancy
Metro Entrances
Bazaar Entrance Preservation
"Bazaar (Sabzeh Meydan Tehran) ^-^"
"Ta'zyieh-khani (Passion Play) of Imam Hossein is a ceremonial art of Islam. Tehran Grand Bazaar, Sabzeh Meydan."
Sabzeh Meydan Plaza
Grand Bazaar Area
Bazaar Area Preservation

## Proof of Concept

The proposed framework was meant to be an additional component to the design process suggested by Van Dooren et al. (2014) helping the designer to have complementary public-driven intakes passively. It may not be easy or efficient to have the public values on the intervention proposals, especially in recognizing the places with deep roots in the memories of the citizens. Taking the data from social media and extracting citizens' emotional validations of places and spatial features would give the designer an initial overview of the public interests and stakes. Moreover, it will be a stepping stone to highlight the locations of priority within and around the project's site. Thus, they will have the assignment and partial public priorities in the first place.

The public insight can be guiding. There were some issues of focus brought up by users that were not my main concern at the beginning of my study. However, the public-driven topics acknowledged the underlying problems that the expert may overlook or miss while running the investigation. This is where the public digital input helps the designer/decision maker gain accuracy and precision in their studies. The people would inform you where the problem is, for they experienced it on the street in their daily lives. These are the kind of encounters that a designer may never notice on their own. Yet the citizens talk and express their feelings about it either as a memory or a current state expression.

There are some examples to elaborate on this argument. The Vali'asr plain tree cuts and ecology topic was not my focus. However, it reminded me, as a Tehranian, that it has been an ongoing problem within Vali'asr space. This was interpreted as environmental degradation within the study area. This helped me to take it further as a guiding value on a specific domain in the design process. The other example is the City Theater underpass. A conjunction for pedestrians to access one side of Vali'asr-Enghelab intersection from another. Moreover, it is an infrastructure where you can enter the metro station, a transition zone for commuters and passers. Despite the prominent functional effect of this underground space, it lacked the pedestrian-friendly qualities. All of this information is only the specifications made from the processed data through the analysis. My focus was not on that adjacent space, but it already had a dominant topic-emotion dynamic provoked by citizens. This topic brings up the possibilities of tackling many mobility and accessibility issues that overlap in the City Theater area.



In general, the public review factor helps to avoid crossing the red lines defined by citizens, too. However, the level of priority setting depends on the designer's and the project's ambitions. Whether the project is public-oriented or not, the prioritization can be completely based on citizen opinions and memorialized values, or only a general idea of what online users think about the spaces in the past and present within the study area, for the expert to be aware of.

The analytical output of this component held a temporal aspect. The digital memory analysis operated as an archive of public reviews regarding the past, existing, and ongoing urban projects. In the examples of analysis-design transitions, some findings indicated the public expressed their feelings towards ongoing projects on urban spaces, such as demolishing the fountain in City Theater as negative feedback, or promoting the new statue installation

in Vanak Square. In simple words, the approach was like reading through online public reviews about urban products, interventions and transformations over time. The aim is to understand how the citizens reacted to a change or an existing project to gain a wider comprehension of the general overview.

It is not easy to have everyone's direct opinion on a project. But having large-scale data can be complementary to co-design projects meant to engage community members actively. Having an initial overview would also help the planners and mitigators to readjust their methods towards the target communities before approaching them. Let's say an engaged participant can share their thoughts on a project in detail, but only a small group of people may have that opportunity. The social media data review opens the window to a broader scale of citizens, whether someone who was directly engaged with the project or was only passersby. The aggregation builds up an overall understanding of the population's general thoughts towards a place, a project, or the context. However, following the trend is not necessarily appropriate.



One of the downsides of this type of input is the potential bias developed within discourses. The expert should be cautious and aware of possible biases in what people share on social media platforms regarding urban projects and spaces. For example, part of the reactions to the City Theater redesign project was diverted based on doubts on the mayor's legitimacy in making decisions or operating any urban project, regardless of what the project's aim was and where it was leading. This is why the reviews are considered only as an initial input from the general public opinion. The inputs can be modified and synthesized with further investigations, active community engagements, surveys and interviews that complement the citizen engagement component.

The framework would allow gathering available public insights through different timelines. The inputs are public-informed and help to distinguish the civic society from other sectors' ambitions. Some valuable daily experiences would be gained for analysis that is usually not feasible without running person-to-person investigations. The public value guides can give accuracy in locating the hotspots and qualified precision in highlighting issues, constraints, qualities, and neutrality from spatial features within key places of the study area. The proposed framework creates the link between analyzing the large quantified textual data and design guiding values with a citizens' lived-experience essence.

There's a question about how to operationalize this proposal. There are several possibilities that I elaborated on a bit:

- The first possible implementation can be the manual way. The experts at the municipality or private firm can take the framework and follow the defined procedure based on their project on traditional on-hand design approach. The framework would contribute as a tool, a guide on socio-spatial mitigations backed by public input. Reference projects can be the #MySydney digital campaign (Williamson & Ruming, 2020) with collecting tweets and processing them as data in the project's context. Also, the input collection can work as surveys, interviews, or workshops (Figure 53). The High Line project (Martin, 2025), was focused on gathering and mapping input from the

community's emotions and applying them to design.

- Another approach is to develop the methodological tools further. By creating an interactive website or application, citizens can share their memories of spots in the city. The experts can collect and organize the inputs in a database format and process them and approach the challenges the same way suggested in the framework. A reference to methodological implementation is the Gaza Places website (Figure 54), an interactive map website for citizens to share their memories on vanished and damaged places due to war destruction in Gaza (GazaPlaces, 2026). The other reference project is the "meinBerlin" platform (Singh & Christmann, 2020). The "meinBerlin" platform was used to collect citizen participatory data regarding noise protection mitigations on a case study, which helped the experts to identify the issues, potentials and desires based on citizens' visual submissions.

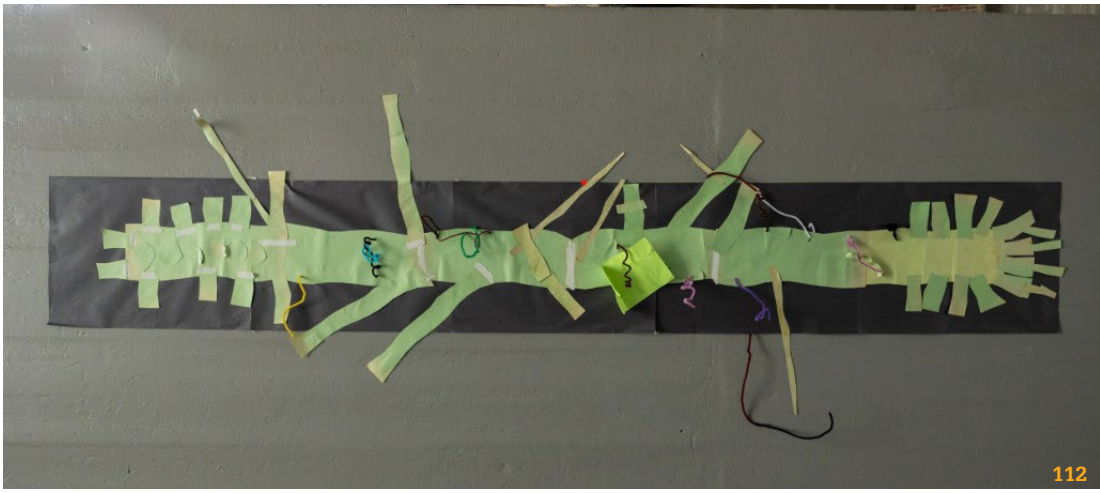


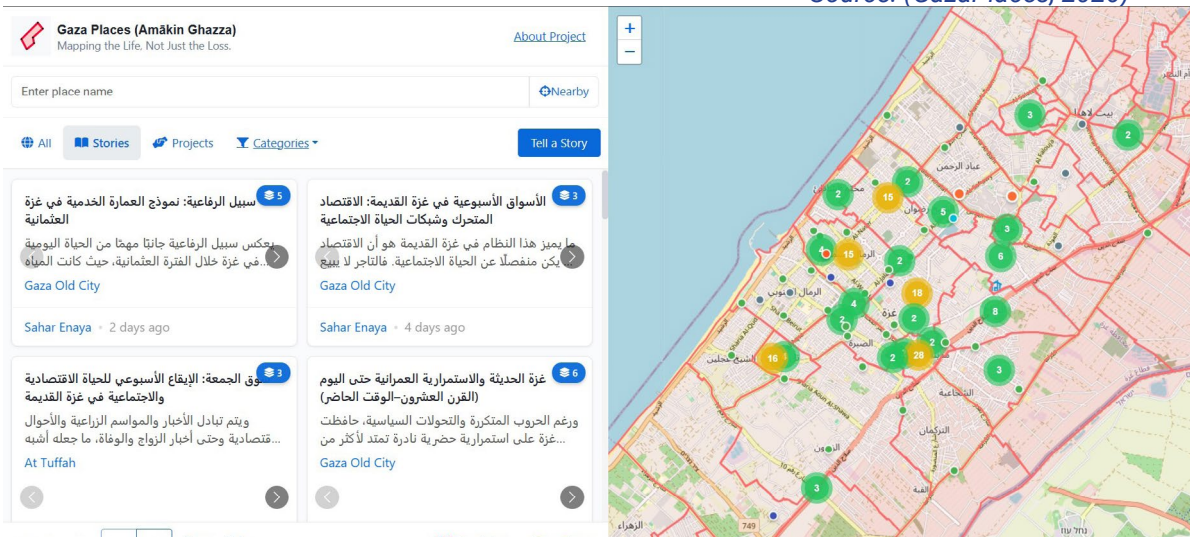
*Figure 54. Community inputs on High Line design process*

*Source: Lee (2026)*



*Figure 55. Leaflet map of Gaza with stories told and spotted by citizens online*

*Source: (GazaPlaces, 2026)*



• The third approach could be a stand-alone web-based platform where experts can run the same analytical workflow and data collection queries. Which means, for example, the scraping agents could be used to scrape the online data and all of the corpus could be poured into a structured database. The approach would be a fit if there's interest in analyzing the social media data, for they contain clearer, more straightforward and transparent expressions from the citizens, compared to inputs from a focused questionnaire. However, this approach requires strict considerations on privacy and ethical concerns regarding social media data management, which vary from context to context. The example project is the aggregation of geo-tagged images from Flickr[1] using the Geostream data gathering tool, specifically developed for collection and organization purposes (Arrigoni & Galani, 2019). The municipalities or firms can develop such specialized data-gathering tools to smooth the analysis-design workflows based on the framework within their project's scope.

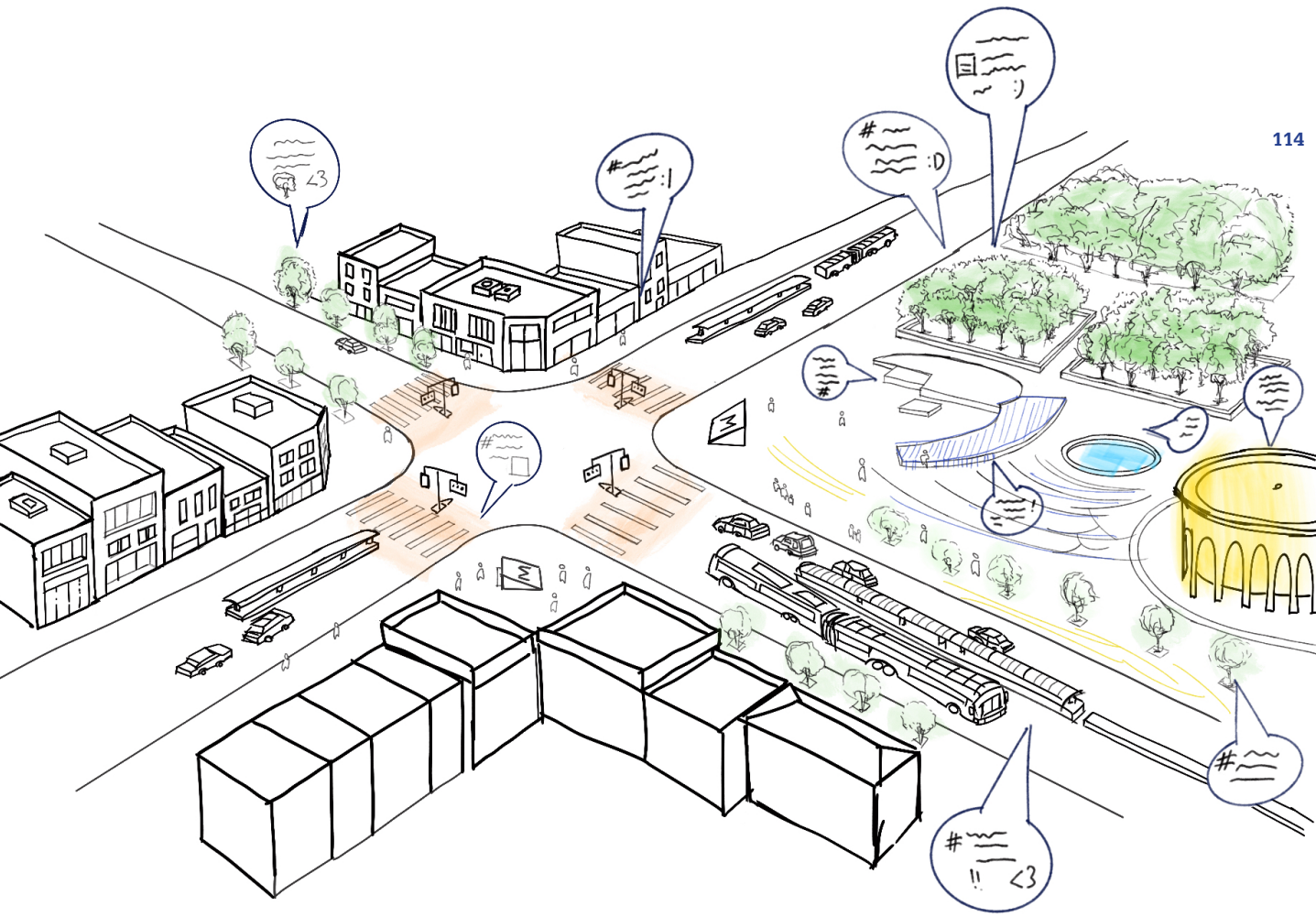
If we want to have a view of the framework's generalizability, I'd say this was the whole point of developing it. Having the framework implemented in other similar cases in different cities is the framework's potential capacity. Let's say there's a project in locating the valued places by the public across the city for development. The framework would help to locate any known and unknown spots within the urban fabric, recognized by its dwellers. The experts will be able to gather data and information regarding their study area smoothly by following the advised methodological approach, and they will gain an exceptional input from the citizens' point of view, who have already experienced living within the space and value it over time. The experts will end up with precise locations valued emotionally on a desired timeline based on past experiences or the present state feedback. Or even identifying locations not known on the city scale that are recognized by a group as it is valued by the local community. This highlights spaces that may have socio-cultural heritage, and thus can be acknowledged as places; spaces that hold memories within (Martin, 2025). Therefore, this ends towards developing preservative or enhancement mitigations based on the same spatial qualities acknowledged by citizens, interpreted by experts.

As a further step towards imagining the implementation of the framework from a text-based mitigation, to a possibility in the future, where the public inputs would inform the designer through their design works, an illustration was produced (Figure 55). The illustration depicts the possible design mitigations guided by the future tweets from citizens, showcasing the potential of urban digital memory that can be leveraged in urban design processes. It's the indirect inputs and expressions on public spaces by the citizens in a digital passive way, yet impactful evolving.

1. An online photo sharing platform



*Figure 56. Vision of implemented design framework backed by urban digital memory*


#
:)
#
:D
#
:|
#
<3
#
!!
<3

# Conclusion

The digital memory is a phenomenon of recording and archiving the urbanites' lived experiences within the space on a digital sphere over time. These are public inputs valuing the urban space, which they have the right to, based on what they feel about their encounters and interactions. An urbanist can trace these passive inputs as valued urban elements through various timelines. Resulting in identifying the places of discourse, accurate locations with precision on their qualified spatial features based on citizens' views. These analytical results were translated into guiding values through the analysis-design transition, leading to the development of the design framework proposal. The outcome of this study was to structure a design component derived from passive public engagement as an initial step in participatory approaches for urban planning and design.

The City Theater was a starting point for introducing the urban digital memory. As the study further went, the more generalized the study scope became. Finding a larger spatial structure on the city scale helped to identify the key locations similar to City Theater. In the analytical steps, the evidence proved the sentiment shifts on emergent discourses within the specific timeline. Classifying the spatial discourse into general and local helped to identify the intra- & interrelations between key places across the city, both mentally and spatially. Overall, the study scope expansion made the comparison of topic-sentiment transformations feasible, which structured the path towards generalizing the analysis and design frameworks.

The proposed framework, as a final output of this study, mostly met the research objectives that were defined at the beginning of the project (Figure 56). Therefore, the data-driven design framework was harnessed from social media, which framed a new type of passive citizen engagement in the planning/design process in the field of urbanism.

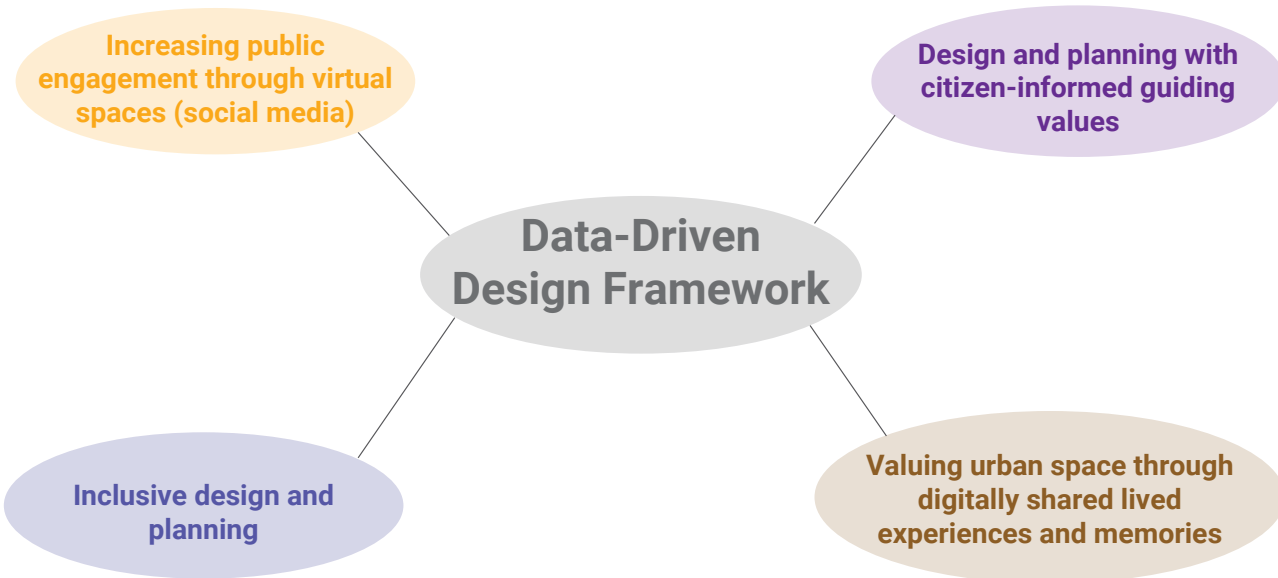



*Figure 57. Research objectives met by the design framework*

# Answering the Research Questions

1. **How do digital memories narrate the lived experiences of citizens in Tehran's public spaces on X (formerly Twitter)?**
2. **What emotions can be traced from digital memories of the City Theater?**
3. **What spatial qualities could be referred to from the shared emotions of Tehran's public spaces and City Theater square?**
4. **What spatial values do citizens share about the City Theater square space transformation throughout time (before and after the latest intervention) as reflected on social media?**
5. **How can shared memories on social media inform design guides for Tehran's public spaces?**



Digital memory is not only one's review of their memory in the past. But also, any daily life experience, encounters, activities and interactions within urban spaces that they shared on social media over time are counted. Citizens share how they experienced a space by walking through, standing on, or passing by it on a specific date and time in the past. The past can be a second or decades ago, depending on when the memory is shared online. Even the memory can belong to a specific time in the past, but be shared later. For example, some users even publish historical information and descriptions regarding a public space on social media and that becomes a discourse by itself, e.g., Hassan Abad & Toopkhaneh Historical Heritage topic.

In general, the emotions were categorized into positive, negative and neutral values. This type of categorization was applied because it helped to easily allocate multiple discrete values to spatial features. Any spatial feature holds a portion of sentiments from different topics. Most of the traced emotions were negative expressions due to spatial transformations in the sanctum of City Theater at the current condition. Moreover, the inconvenience with the underpass as an adjacent infrastructure had been a topic that wasn't recognized until the emotions and topics were produced. At the same time, there were strong positive emotions towards the City Theater as a place of nostalgia, art and cultural activities experienced by anyone who has been there at least once.

The main emotions were opposing any interventions on the open spaces or ecological elements, including some significant features such as the fountain, benches, statues, and old trees. Also, on the CT scale, there were multiple complaints regarding the inconvenience and limited accessibility for pedestrians, both in the sidewalk transformed by the fence placement and the underpass area. On a broader scale, in general, there were two types of emotional feedback:

1. **Conservative**
2. **Non-conservative**

The conservative supports the spatial stability, or enhancement at most, based on the positive/neutral experiences of living within. But the non-conservative is dissatisfied with the status quo and may support the intervention or improvement in the space. Based on that, the spatial qualities are the existing layout of CT and Tehran's spaces. Walkability,

openness, sitting spaces, plain tree lines and connectivity are the most relevant spatial qualities adapted from described emotions towards the study area. On the contrary, new design qualities should be implemented in the areas of conflict, e.g., the sidewalks of Valiasr and the City Theater's underpass.

Based on the descriptive emotions and overall topics revolving around the City Theater, the area holds a strong value in Tehran's memory. This includes the theater's building, fountain, benches, seating edges, and random and occasional performances. Along with that, the spatial and mental connection of CT with other key urban elements, such as the Vali'asr street, creates a spatial structure on the city scale. This connection is defined by mobility paths, spatial junction and therefore mental consistency. Thus, I'd like to argue that based on sentimentally-valued spatial topics, City Theater is both intra-and inter-related to its urban context.

It is now evident that shared memories can turn into design guides based on the proposed design framework. The corpus was preprocessed and went through topic modeling and sentiment analysis to identify the urban discourses within a defined timeline based on a key intervention. It was proved the topic emergence and sentiment shifts were driven by the intervention. Therefore, the public discourses follow the spatial transformations across the city. The shared memories were categorized into general and local topics, giving accuracy to the study. The sentimentally-valued descriptions addressed the spatial features within accurate locations, provided a qualified precision. The valued descriptions were allocated to design domains as the final step of the analysis-design transition. The final product was the guides with positive, negative and neutral values specified within each domain that could be utilized through the design process.

## Answer to the Main Question

**How can digital memories shared by citizens help identify spatial values as potential conditions for interventions in Tehran's public spaces and the City Theater square?**

The digital memories are of two types, the ones addressing the current state, and the ones holding memorial values. Both types should be traced through the city timeline. Upon that, the general and local spatial discourses are identified. It's the very step that frames the focus areas. The processed memories with sentiments within the framed locations provide a precise and emotionally-valued description addressing spatial features within. The allocation of the processed inputs with the design domains structures the additional component through the design process. The developed guiding values can be mitigated specifically based on their allocated domain, value category, and key location. The described workflow had been formed as a design framework that was applied to City Theater and Tehran's key public spaces as the case study of this project.

# Discussion

This study went through several specific workflows of data analysis, which may not be easy to replicate, but are not impossible. The study suggests a framework for linking a public-informed analysis with the designer's intervention. Precisely, it's a data-driven design. What matters is the data to be public-originated; in that sense, the data collection, preprocessing, processing and representation may differ depending on resources, tools, methods and legislations.

The project itself, as noted before, is a data-intensive undertaking. I first designed a query structure to collect qualitative data via web scraping actors, followed by data-cleaning and preprocessing steps. The topics and sentiments were generated using Python scripts and libraries. The translation, organization, sorting and visualizations were handled with scripts in the RStudio environment. All in all, it was a process of quantifying the qualitative data and translating it back into spatial qualities serving as guiding values through the design process based on the proposed design framework. Moreover, this study strongly aligns with the Data-Informed Strategies for Sustainable Urban Development (DISSUD) cluster principles, emphasizing data-driven design, open-science practices, and open-source software tools.



However, there were several difficulties, drawbacks and constraints I encountered through the process, which made me readjust my methods, study scope and approach.

First of all, the study scope. I commenced the project focusing only on one specific square in the city center, which brought up the question of how the conceptual, methodological and design frameworks could be generalized and applied to any other case, not only in Tehran, but also in any other city globally. Also, I realized the public perceptions of urban elements are not quite precise on a public space level, unless it's a landmark or any other Lynch element that is highly distinct from its context, but not any regular square element, such as the pavement, landscapes, or the open space. Therefore, I had to readjust the scale to a larger scope that matches the level of detail noticed and mentioned by the citizens on social media.

Apart from that, I aimed to have a generalized method and framework from this study that can be implemented on any other urban case. It was crucial to expand the case study to a city-wide range, consisting of other public spaces with the same characteristics and popularity level as the City Theater. Thus, I had a sufficient corpus of data relevant to Tehran's key places and the CT to run the analysis and make valid comparisons based on topic emergence and sentiment shifts between the two timelines. This helped to explore the discourse attractivity of the CT before and after the key intervention compared to the rest. This means that I managed to depict how the public discourse shifts towards the changes in the City Theater once such a spatial intervention happens, and how it pops up among the general baseline of Tehran topics.

There were also the timeline constraints. Due to the time limit in this project and limitations with data scraping queries on Apify , the timeline buffer could be larger. Given the historical background of the case study, there is a strong potential for further investigation in collecting, processing and organizing the digital memories from a long-term scope. It would bring more socio-spatial discourse shifted and transformed over Tehran's memory, which may bring more information regarding the significance of public spaces among the citizens. Even more valued spatial features lead to more accurate design qualities. Moreover, there's a strong potential in creating a digital archive of all memories towards public spaces in Tehran, shaping a base for digital archival reviews and research. Thus, a longer timeline would provide more possibilities towards a comprehensive study on urban digital memories and public-driven design framework development.

Another aspect of the study scope popped up to me when I did the topic modeling on the data. The overall discourse showed me the significance of some major urban elements in the city, such as Vali'asr Street and Tajrish Square, by the number of relevant topics they had. These are the most famous locations in Tehran, yet they also gain a lot of importance, for the impact they can make on the city. Also, there are not many comprehensive studies done on Vali'asr Street as a whole arterial structure in Tehran, so it can become a significant case on the agenda. What I consider is to move forward in running the digital memory study on Vali'asr street, as it holds a larger amount of data and various topics on each segment of it across the city, making it a constantly-changing phenomenon in urban studies.

There was also a limitation with the data; they were not geolocated. Having the geolocated data would ease mapping the tweets and descriptions as a result of the analysis. However, the descriptions were not precise enough to be placed in the probable spatial feature in a framed location. Mapping the tweets based on their specific locations can give a precise spatial analysis aspect to the study. It'd bring the analysis on GIS domain, where more geo-spatial analysis could be applied and developed through the study. The defect with mapping the data manually was the lack of precision in tweets. Not all tweets provide descriptions of specific spatial features within the locations. Therefore, I'd end up with a heatmap with highly dense spots strongly concentrated in the center points of each key public space in Tehran.



Due to very limited internet access in Iran and a lack of connection from my study location, I didn't have the chance to run the analysis through an active online survey. Therefore, social media was an existing archive of memorial data for this study. The analysis applied in this study helped to achieve legitimacy on public opinion regarding urban spaces. However, the data was not necessarily direct or sufficient in detailed descriptions regarding the spatial features within the space. Based on that, as a further step, developing an interactive map website is suggested. Where online users can pick their spots in the given areas and add details of their memories and emotions. Not in a positive, negative and neutral classification, but in a more varied set of sentiments results in having spatial features consisting of multifaceted emotion labels. The other advantage is that the data and information will already be geolocated, thus more accurately-

positioned emotions in the area for potential geo-spatial analysis.

Also, the interactive website leverages the public data characteristics from passive to active. There will be participants who will contribute to the study in an informed way and be actively engaged. This gives the study more legitimacy in terms of citizen engagement principles and way fewer risks concerning privacy and data exposure. However, there should be strict mitigations regarding the database collected from the interactive maps.

There can also be further investigations into the design process. More detailed interventions and policies could be developed based on the descriptive emotions and localised spatial topics to tackle the challenges, conserving the valued features, or taking no action. The results could be represented both textually and visually as final outputs to provide the designer or decision maker with clear and concrete guidance.

# Reflection

This project was an evolution towards the path I was unconsciously craving for at the beginning of my master's studies. I've been looking for ways to quantify the qualitative data on a spatial aspect. And it happened!

At the beginning, I went through several explorations in concepts, theories in the literature and previous studies over time. Coming up with the digital memory concept was an interesting topic for me to tackle, even though there's more room to elaborate on now that I look back. But, given the short amount of time. I believe the progress is acceptable, and I am glad that it came to a concrete output that can be presented. Furthermore, I'm happily looking forward to developing further research on better alternatives to methodological approaches in the urban digital memory context.

The first challenge was the data. At the beginning, I had picked textual data from Instagram and X (formerly Twitter) as potential platforms mostly and commonly used by Iranians. Nevertheless, Instagram data was difficult to collect relevantly. I ended up scraping the data only from X, for it was more straightforward to set up the query for. The data scraping was one challenge, while the greater difficulty was conducting the Data Management Plan (DMP) for the project. Because of human data involvement in the study, tweets are from real-world accounts, i.e., humans, I also had to apply for Human Research Ethical Committee (HREC) approval. Both of these procedures were highly demanding in providing legislation and risk mitigation considerations corresponding to my project. However, it was also an external motive to systematically plan my data management, handling, processing, and publication as concretely as possible. Although the process was ultimately a constructive experience, the bureaucratic and official steps were exhausting and, in some cases, mostly unnecessary.

But the biggest challenge and achievement at the same time, throughout my journey, was to find the link between the analysis and design. It was a key moment once I realized the valued spatial features are not necessarily design components, but elements that guide you through the design. I managed to make a transition from the analytical to the design process. I'm strongly delighted and excited that these steps structured the final proposed design framework, not as an exclusive product, but a data-driven output that can hopefully be implemented on any other case study. Frankly, I'm proud of myself for developing a framework that literally links the analysis and design on a data-driven approach; a gap that I found in many data-oriented projects. They lack the design mitigation at the end of their analytical process and usually close their work with strategic and policy suggestions.

I went through various types of events and news while running this project. Starting from the street protests in Iran and then the war and internet blackout came in. I was worried and anxious about events back home, especially for my family and friends. Because it wasn't easy to get in touch with

them. Surprisingly, this project helped me to distract myself from the uncontrollable situation back home. There weren't many things I could do from here, but working on my thesis. Keeping consistency with my thesis work, even though the progress was not mostly significant, was a driver to keep me motivated and sane through the hard times.

The very positive aspect of this project was that I really liked my topic. It sounded interesting, yet challenging and complex. However, none of the obstacles or hardships made me think of entirely changing it. I did think of readjusting the methods or parts of the concept, but not wiping out the whole idea. Also, having a personal connection with the case study was an addition. It made me think of different aspects of the project and the case study. I'm sure that my passion for urban digital memory will not end here. I'll try to develop this concept further. Either in the form of an open-source platform to share urban memories and emotions, or a large online digital archive of key public places, as a personal project.



This project helped me advance in many analytical and statistical skills, which I've always wanted to learn within the field of urban studies. This created a stepping stone towards the advanced spatial analysis approaches and methodologies that I can develop myself into, either as a researcher or a specialist. The experience I gained with pushing my limits over and over in refinements, paying attention to the details, making the productions as clear as possible, all in all, took me to a different level of my academic capacity. I believe the gains of a year and a quarter made the path to where I am now at the end of this master's thesis.

As an academic journey, I'd say the path towards my favorite specialization was gradually built up through this program. Firstly, I was not sure if I was on the right track. As the restrictions started to fade, I managed to tailor the specialization I'd been looking for. I realized how excited I get with the analytical concepts and methodologies where you can understand and measure social, spatial and economic indices within the natural and the built environment contexts. My intakes are not only about the analysis, though. I gained a lot more confidence in design as well. I learnt how important it is to keep a constant pace in producing ideas with drawings. And more importantly, they don't have to be perfect aesthetic drawings. As long as I find my path towards the final design, through producing rough hand sketches, it's totally fine. Conveying the message is far more crucial than the style or the medium. This was a great lesson I'll always carry with me onwards.

# AI Acknowledgments

As an acknowledgment, the use of generative AIs, Le Chat and Gemini, was only done on the analytical scripts and model runs, only for modification, refinement and efficiency purposes. The analytical steps, especially with debugging and trouble shooting was a demanding step that required a large labor work. Gemini was used as a script correction and modification tool through this process.

At the beginning of the study, with developing the topic and research questions, the relevant topics, projects and references were introduced by the generative AI to gain a better understanding of the topic and how it can be developed as an academic research project. As a percentage, I'd say, as an assumption, 85% of the items were discussed and asked by myself and the AI was basically an advisor that helped to narrow down the project's title scope. At the end, the research questions were revised and modified for simplification purposes. This was done solely by the author. No AI was involved in developing the finalized research questions and project topic.

The other use of AI was focused on rephrasing the writings made by the author themselves. I can say it was only few attempts, for the process of writing and transferring the text back and forth was time consuming. I can assure that the main text of the report, interpretations, design production, initial studies were done by myself without the help of any generative AI.

# References


AlAgha, I. (2021). Topic Modeling and Sentiment Analysis of Twitter Discussions on COVID19- from Spatial and Temporal Perspectives. Journal of Information Science Theory and Practice, 53–35 ,(1)9. https://doi.org/10.1633/JISTaP.2021.9.1.3

Aleksandar, A., Evgeny, ·, Ayse, P. ·, & Buckley, O. (2025). Sustainable Landscape Planning and Natural Resources Management IEREK Interdisciplinary Series for Sustainable Development Integrating Agriculture, Urban Systems, and Natural Resources A Pathway to Sustainable Landscapes.

Alvarez-Melis, D., & Saveski, M. (2016). Topic Modeling in Twitter: Aggregating Tweets by Conversations. www.aaai.org

Arnstein, S. R. (2019). A Ladder of Citizen Participation. Journal of the American Planning Association, 34–24 ,(1)85. https://doi.org/10.01944363.2018.1559388/1080





Arrigoni, G., & Galani, A. (2019). From Place-Memories to Active Citizenship: The Potential of Geotagged User-Generated Content for Memory Scholarship. In Doing Memory Research (pp. 168–145). Springer Singapore. https://doi.org/8_7-1411-13-981-978/10.1007

BERTopic Team. (2025). About Us - BERTopic. In BERTopic. https://bertopic.com/about-us/

Blei, D. M., Ng, A. Y., & Edu, J. B. (2003). Latent Dirichlet Allocation Michael I. Jordan. In Journal of Machine Learning Research (Vol. 3).

Burgess, J., Foth, M., & Klaebe, H. (2006). Everyday Creativity as Civic Engagement: A Cultural Citizenship View of New Media. http://eprints.qut.edu.au

Can, S. N. (2019). The City Image and Its Elements by Kevin Lynch. https://semanurcan.wordpress.com/27/10/2019/the-city-image-and-its-elements-by-kevin-lynch/

Currie, M., Knox, J., McGregor, C., & Editor, B. (2022). Book Title: Data Justice and the Right to the City. https://doi.org/10.3366/jj.7358679.8

DataCamp. (2026). Chi-Square Test in R. https://www.datacamp.com/tutorial/chi-square-test-r

Ekman, P. (1999). Basic Emotions. In T. Dalgleish & M. Power (Eds.), Handbook of Cognition and Emotion. John Wiley & Sons Ltd. https://www.paulekman.com/wp-content/uploads/07/2013/Basic-Emotions.pdf

Esteves, T. (2025). Reimagining Citizen Engagement in the Digital Age: Platforms, Participation, and Public Trust. In Public Administration, Governance and Globalization (Vol. 6, pp. 115–93). Springer. https://doi.org/6_4-00514-032-3-978/10.1007

Farahani, M., Gharachorloo, M., Farahani, M., & Manthouri, M. (2021). ParsBERT: Transformer-based Model for Persian Language Understanding. Neural Processing Letters, 3847–3831 ,(6)53. https://doi.org/10.1007/s4-10528-021-11063

Fathullah, A., & Willis, K. S. (2018). Engaging the Senses: The Potential of Emotional Data for Participation in Urban Planning. Urban Science, 98 ,(4)2. https://doi.org/10.3390/urbansci2040098

GazaPlaces. (2026). GazaPlaces. https://gazaplaces.com/?lang=en

Ghahramani, M., Galle, N. J., Ratti, C., & Pilla, F. (2021). Tales of a city: Sentiment analysis of urban green space in Dublin. Cities, 119. https://doi.org/10.1016/j.cities.2021.103395

Habibi, S. M., & Amiri, M. (2016). The Right to the City: The Existing City and the Desired City (What Is and What Should Be) [حق به شهر: شهر موجود و شهر مطلوب (آنچه هست و آنچه باید باشد)]. Iranian Anthropological Research, 30–9 ,(2)5.

Hamshahri Online. (2026). تغییرات جدید در چهارراه ولیعصر؛ تئاتر شهر محصور می‌شود. https://www.hamshahrionline.ir/news/817881/%D%8AA%D%8BA%DB8%C%DB8%C%D%8B%1D%8A%7D%8AA-%D%8AC%D%8AF%DB8%C%D%8AF-%D%8AF%D%8B-1%DA%86%D%87%9D%8A%7D%8B%1D%8B%1D%8A7%D%-87%9D%88%9D%84%9DB8%C%D%8B%9D%8B%5D%8B-1%D%8AA%D%87%9D%8B%1D%8A%7D-86%9%D%8AA%D%8A%6D%8A%7D%8AA%D%8B-1%D%8B%4D%87%9D%8B-1%D%85%9D%8AD%D%8B%5D%88%9D%8B%-1D%85%9DB-%8C-%D%8B%4D%88%9D%8AF

Harvey, D. (2003). The right to the city. International Journal of Urban and Regional Research, 941–939 ,(4)27. https://doi.org/10.1111/j.1317.2003.00492-0309.x

Harvey, D. (2006). The political economy of public space. In S. M. Low & N. Smith (Eds.), The politics of public space (pp. 34–17).

Routledge. https://doi.org/2-9780203390306/10.4324

Harvey, D., Hassan Khali, A., & Vaqefi, I. (2015). Political Economy of Public Space [اقتصاد سیاسی فضای عمومی]. Dialectical Space. http://dialecticalspace.com/the-political-economy-of-public-space/

Haskins, E. (2007). Between archive and participation: Public memory in a digital age. Rhetoric Society Quarterly, 422–401 ,(4)37. https://doi.org/02773940601086794/10.1080

HooshvareLab. (2020). HooshvareLab/bert-base-parsbert-uncased. https://huggingface.co/HooshvareLab/bert-base-parsbert-uncased

Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. http://sentic.net/

IBM. (2023). What Are Word Embeddings? https://www.ibm.com/think/topics/word-embeddings

IBM Community. (2026). SPSS Statistics: Quantifying Over/Underrepresentation of Sample Data vs. Population Data. https://community.ibm.com/community/user/discussion/spss-statistics-quantifying-overunderrepresentation-of-sample-data-vs-population-data

Kent State University Library. (2026). SPSS: Chi-Square Test. https://libguides.library.kent.edu/spss/chisquare

Kovács-Győri, A., Ristea, A., Havas, C., Resch, B., & Cabrera-Barona, P. (2018). #London2012: Towards citizen-contributed urban planning through sentiment analysis of twitter data. Urban Planning, ,(1)3 99–75. https://doi.org/10.17645/up.v3i1.1287

Lee, R. (2026). Programs & Community Photo. https://www.thehighline.org/photos/programs-community/?gallery=30859&media_item=30861

Lefebvre, H. (n.d.). The Right to the City.

Lefebvre, H. (1968). The Right to the City.

Liang, X. (2020). Participatory Management for Cultural Heritage: Social Media and Chinese Urban Landscape. Communications in Computer and Information Science, 1226 CCIS, 307–300. https://doi.org/40_9-50732-030-3-978/10.1007

Lynch, K. (1964). The Image of the City.

MacroTrends. (2026). Tehran, Iran Metro Area Population (–1950 2026). https://www.macrotrends.net/global-metrics/cities/21523/

tehran/population

Martin, M. (2024). Public Participation in Urban Planning: How Emotional Mapping Could Contribute to Building a Sense of Belonging. International Conference on Geo-Science and …. https://doi.org/8_0-81288-031-3-978/10.1007

Martin, M. (2025). Public Participation in Urban Planning: How Emotional Mapping Could Contribute to Building a Sense of Belonging. In A. Aleksandar, E. Panidi, & A. O. Buckley (Eds.), A Pathway to Sustainable Landscapes (pp. 85–77). Springer Nature Switzerland. https://doi.org/8_0-81288-031-3-978/10.1007

Mitchell, D. (2003). The Right to the City: Social Justice and the Fight for Public Space. The Guilford Press. https://www.google.nl/books/edition/The_Right_to_the_City/lw_oO3WyNTwC

Mohit. (2026). CITY THEATER. In Museum / Art Space. https://mohit.art/network/city-theatre/

Moroni, S., & Chiodelli, F. (2014). Public Spaces, Private Spaces, and the Right to the City. International Journal of E-Planning Research, 65–51 ,(1)3. https://doi.org/10.4018/ijepr.2014010105

Paccoud, A. (2016). Planning law, power, and practice: Haussmann in Paris (1870–1853). Planning Perspectives, 361–341 ,(3)31. https://doi.org/02665433.2015.1089414/10.1080

Pereira, G. C., Célia, M., Rocha, F., & Poplin, A. (2012). E-Participation: Social media and the public space. In LNCS (Vol. 7333, pp. –491 501). Springer. https://doi.org/37_3-31125-642-3-978/10.1007

Pereira, G. C., Rocha, M. C. F., & Poplin, A. (2012). E-Participation: Social media and the public space. International Conference on …. https://doi.org/37_3-31125-642-3-978/10.1007

Pierce, J., Williams, O. R., & Martin, D. G. (2016). Rights in places: An analytical extension of the right to the city. Geoforum, 88–79 ,70. https://doi.org/10.1016/j.geoforum.2016.02.006

Roberts, H., Resch, B., Sadler, J., Chapman, L., Petutschnig, A., & Zimmer, S. (2018). Investigating the emotional responses of individuals to urban green space using twitter data: A critical comparison of three different methods of sentiment analysis. Urban Planning, 33–21 ,(1)3. https://doi.org/10.17645/up.v3i1.1231

Rokna. (2025). City Theater's fountian was demolished. Municipality's progression to Tehran's heart of culture. [حوض تئاتر شهر با خاک یکسان شد / پیشروی تازه شهرداری در قلب فرهنگ

تهران. https://www.rokna.net/%D%8A%8D%8AE%D%8B-8
%D%8A%7D%8AE%D%8A%8D%8A%7D%8B%-1D%8A%7D%8
AC%D%8AA%D%85%9D%8A%7D%8B%9DB8%C-1179907/95-
%D%8AD%D%88%9D%8B%-6D%8AA%D%8A%6D%8A7-
%D%8AA%D%8B%-1D%8B%4D%87%9D%8B-1
%D%8A%8D%8A%-7D%8AE%D%8A%7DA%A-9
%DB8%C%DA%A%9D%8B%3D%8A%7D%-86%9D%8B%4D%8AF-%D9
%BE%DB8%C%D%8B%4D%8B%1D%88%9DB8%C-%D%8A-
A%D%8A%7D%8B%2D%-87%9D%8B%4D%87%9D%8B%1D%8AF%D8
%A%7D%8B%1DB8%C-%D%8AF%D%8B%-1D%82%9D%84%9D%8A-8
%D%81%9D%8B%1D%87%9D%86%9DA%AF-
%D%8AA%D%87%9D%8B%1D%8A%7D86%9

Rostamipour, E. (2024). Tweet by Ehsan Rostamipour. https://x.com/Ehsanrostamipor/status/1755203163483729936

Shaw, J., & Graham, M. (2017). An Informational Right to the City? Code, Content, Control, and the Urbanization of Information. Antipode, 927–907 ,(4)49. https://doi.org/10.1111/anti.12312

Singh, A., & Christmann, G. (2020). Citizen participation in digitised environments in Berlin: Visualising spatial knowledge in urban planning. Urban Planning, 83–71 ,(2)5. https://doi.org/10.17645/UP.V5I2.3030

Solnit, R. (2001). Wanderlust : a history of walking. Penguin Books, an imprint of Penguin Random House LLC.

Solnit, R. (2009). A field guide to getting lost. Canongate Books.

statcounter. (2015). Social Media Stats Islamic Republic Of Iran. https://gs.statcounter.com/social-media-stats/all/iran/#monthly202512-201512-

Tedeschi, M. (2024). Datafication and urban (in)justice: Towards a digital spatial justice. In Geography Compass (Vol. 18, Number 6). John Wiley and Sons Inc. https://doi.org/10.1111/gec3.12763

University of Southampton. (2026). Bivariate Analysis: Chi-Square Test. https://www.southampton.ac.uk/passs/full_time_education/bivariate_analysis/chi_square.page

U.S. Bureau of Economic Analysis. (2018). What are location quotients (LQs)? https://www.bea.gov/help/faq/478

Van Dooren, E., Boshuizen, E., Van Merriënboer, J., Asselbergs, T., & Van Dorst, M. (2014). Making explicit in design education: Generic elements in the design process. International Journal of Technology and Design Education, 71–53 ,(1)24. https://doi.org/10.1007/

s8-9246-013-10798

Wikimedia Commons. (2026a). Théâtre municipal de Téhéran (1). https://commons.wikimedia.org/wiki/File:Théâtre_municipal_de_Téhéran_(1).jpg

Wikimedia Commons. (2026b). چاپ خبر تاسیس تئاتر شهر - روزنامه اطلاعات. https://commons.wikimedia.org/wiki/File:%DA%86%D8%A7%D9%BE_%D8%AE%D8%A8%D8%B1_%D8%AA%D8%A7%D8%B3DB8%C%D8%B3_%D8%AA%D8%A6%D8%A7%D8%AA%D8%B1_%D8%B4D87%9%D8%B1_-_%D8%B1D88%9D8%B2D86%9D8%A7%D85%9D987%_%D8%A7%D8%B7%D84%9D8%A7%D8%B9%D8%A7%D8%AA.jpg

wikipedia. (2026). City Theater of Tehran. In Wikipedia. https://en.wikipedia.org/wiki/City_Theater_of_Tehran

Williamson, W., & Ruming, K. (2020). Can social media support large scale public participation in urban planning? The case of the #MySydney digital engagement campaign. International Planning Studies, 371–355 ,(4)25. https://doi.org/13563475.2019.16/10.108026221

Xu, N., Cheng, Y., & Xu, X. (2018). Using location quotients to determine public-natural space spatial patterns: A Zurich model. Sustainability (Switzerland), 10)10). https://doi.org/10.3390/su10103462

Zeile, P., Memmel, M., & Exner, J.-P. (2012). A New Urban Sensing and Monitoring Approach: Tagging the City with the RADAR SENSING App. In M. Schrenk (Ed.), REAL CORP 2012 : re-mixing the city : towards sustainability and resilience? : proceedings of 17th International Conference on Urban Planning, Regional Development and Information Society. CORP, Competence Center of Urban and Regional Planning.

Zeile, P., Resch, B., Exner, J. P., & Sagl, G. (2015). Urban emotions: Benefits and risks in using human sensory assessment for the extraction of contextual emotion information in urban planning. Lecture Notes in Geoinformation and Cartography, 225–209 ,213. https://doi.org/11_8-18368-319-3-978/10.1007

# Appendix

## Introduction

For the opening of the City Theater's building in the 1970s, the first play to be performed was announced in the newspapers.



صفحه ۱۹ ـ اطلاعات
یکشنبه ۸ بهمن ماه ۱۳۵۱ ـ شماره ۱۴۰۱۲

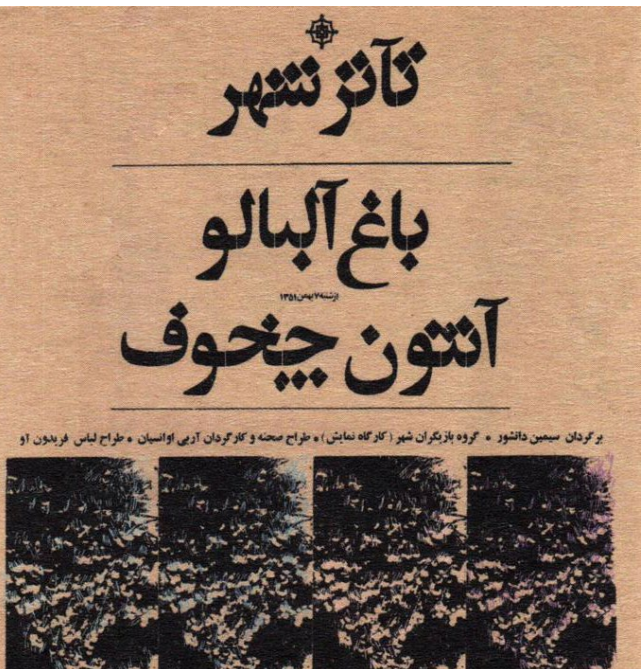
افتتاح تیاتر در حضور اعلیحضرتین

سقوط اشرافیت پیش از انقلاب

# باغ آلبالو در تیاتر شهر

در ورودی تیاتر شهر

آنتون چخوف و بنده باغ آلبالو



تئاتر شهر

باغ آلبالو

آنتون چخوف

برگردان سیمین دانشور • گروه بازیگران شهر (کارگاه نمایش) • طراح صحنه و کارگردان آربی اوانسیان • طراح لباس فریدون آو

*Apendix 1. (Top) Announcement of Anton Checkov's "Cherry Orchard" play to be performed at the City Theater on Newsletter dated January 1973; (Bottom) the play of "Cherry Orchard" poster at City Theater. Sources: (Top) Wikimedia Commons, 2026b.*

## Tehran scale key places

### Tajrish Square

Tajrish Square is one of the most famous squares in Tehran that not only operates as a traffic node, but also serves as a socio-economic focal point in the city. The square was built upon the old mosque as a pilgrimage location and Tajrish's market, also known as Tajrish's Bazaar. There's one of the head stations of Tehran's metro network as well, which connects the south to the northern part of the city.



*Figure 14. Tajrish Sqaure, Tehran. Source: Tripadvisor, 2026.*



*Figure 15. Tajrish Sqaure's market. Source: Shidfar, 2024*

### Vanak Square

Vanak square is the other important square that acts as the middle node of North-South main mobility network. This square can be considered a transportation hub for bus, BRT and pool taxis, along with mixed-use activities on its edges that have turned the square into a business and retail district through decades.





*Figure 16. Vanak Square, Tehran. Source: akbbayat, 2012*

### Vali'asr Street

Vali'asr is the longest North-South street in the country that starts from Tajrish sqaure in the north and ends to the railway station in the south, a true arterial street. Speaking of which, the street holds diverse types of activities and spatial typologies that vary from spot to spot along its way that creates a multifaceted identity of it as an urban space that can be considered a public space.



*Figure 17. Vali'asr Street facing North, Tehran. Source: Tolouei, 2024*

Enghelab Square

Enghelab square is one of the most significant roundabouts in the city center. This place is in which concentrates the transportation, social, cultural and educational activities in a single node.




*Figure 18. Enghelab Square, Tehran. Source: Mehrnews, 2026*

Enghelab Street

The head of the Enghleb square is stretched by the Enghelab street that connects the other important spaces attached to one segment from western to eastern side of the city. Places like the Enghelab square, University of Tehran, Vali'asr intersection, City Theater square and other important places towards the east of Tehran.

Imam Khomeini (Toopkhaneh) Square

The Toopkhaneh square is one of the oldest squares located at the edge of Tehran's historical district by being one of the main gateways leading to Tehran's grand bazaar. This square is also a retail center and public transportation node approximating the political centers as well.



Hassan Abad Square

Hassan Abad square is famous for its buildings with curved buildings with European style facades facing the square along with the traffic limitations, making the space look more like a plaza than a node.

Grand Bazaar

Tehran's grand bazaar is one of the largest components of Tehran's spatial organization. It operates as an economic hub while holding a significant cultural and historical core narrative of the old Tehran.

Sabzeh Meydan Plaza

Within Tehran's grand bazaar, there lies one of the open squares in the historical business district of the city. This place can be considered as a plaza due to its small scale and having less importance compared to the main square yet having the spatial characteristics of a public open space within the grand bazaar area.

## Methodology

### Overview table of reviewed methodologies and approaches.

| Title | Methodology | Tools / Approaches | Relation |
|---|---|---|---|
| GIS-based fuzzy sentiment analysis framework to classify urban elements according to the orientations of citizens and tourists expressed in social networks (Cardone et al., 2022) | Providing most frequent emotional categories towards the qualities of urban infrastructure and services per district in Bologna and spatializing them within the districts. | The framework of emotional classifications was used, GIS visualization, Python algorithm for classifying documents (unsupervised) | 1. Emotional framework<br>2. District categorizations<br>3. Degress of belonging |
| #London2012: Towards Citizen-Contributed Urban Planning Through Sentiment Analysis of Twitter Data (Kovács-Győri et al., 2018) | Assigning the emotions and expressions in tweets to urban events and transportation infrastructure via topic modeling in sentiment analysis | KDE (Kernel Density Estimation) for hotspot analysis, creating confidence intervals (Similarity Index) for comparing the areas between two different timings, LDA (Latent Dirichlet Allocation) for topic modelling | 1. KDE can be used for spatializing the digital memories<br>2. LDA usage approach on topic modelling |
| Social media as passive geo-participation in transportation planning – how effective are topic modeling & sentiment analysis in comparison with citizen surveys? (Lock & Pettit, 2020) | Sentiment analysis Vs. Survey on public transportation quality | Use of VADER model for sentiment analysis,<br>Use of LDA model for topic modeling,<br>Text normalization & cleansing for analysis purposes | 1. Use of VADER for social media sentiment analysis<br>2. Text normalization and data cleaning |
| Tales of a city: Sentiment analysis of urban green space in Dublin (Ghahramani et al., 2021) | | Data scraping using Python & Selenium, Text normalization & cleansing for analysis purposes, categorizing the text into specific topics, | 1. Valued associations can happen using topic modeling to categorize expressions based on topics (feelings) or spatial elements, memorial elements, and activity elements. |
| Investigating the Emotional Responses of Individuals to Urban Green Space Using Twitter Data: A Critical Comparison of Three Different Methods of Sentiment Analysis (Roberts et al., 2018) | Sentiment analysis | REST API of Tweeter for data harvesting, Ekman's emotion classification based on +/- or neutral into detailed more detailed emotions,<br>Use of three approaches of associating tweets to emotions of Automated, Manual and semi-automated analysis. | 1. Using Ekman's basic emotions classification for emotional valuing process |

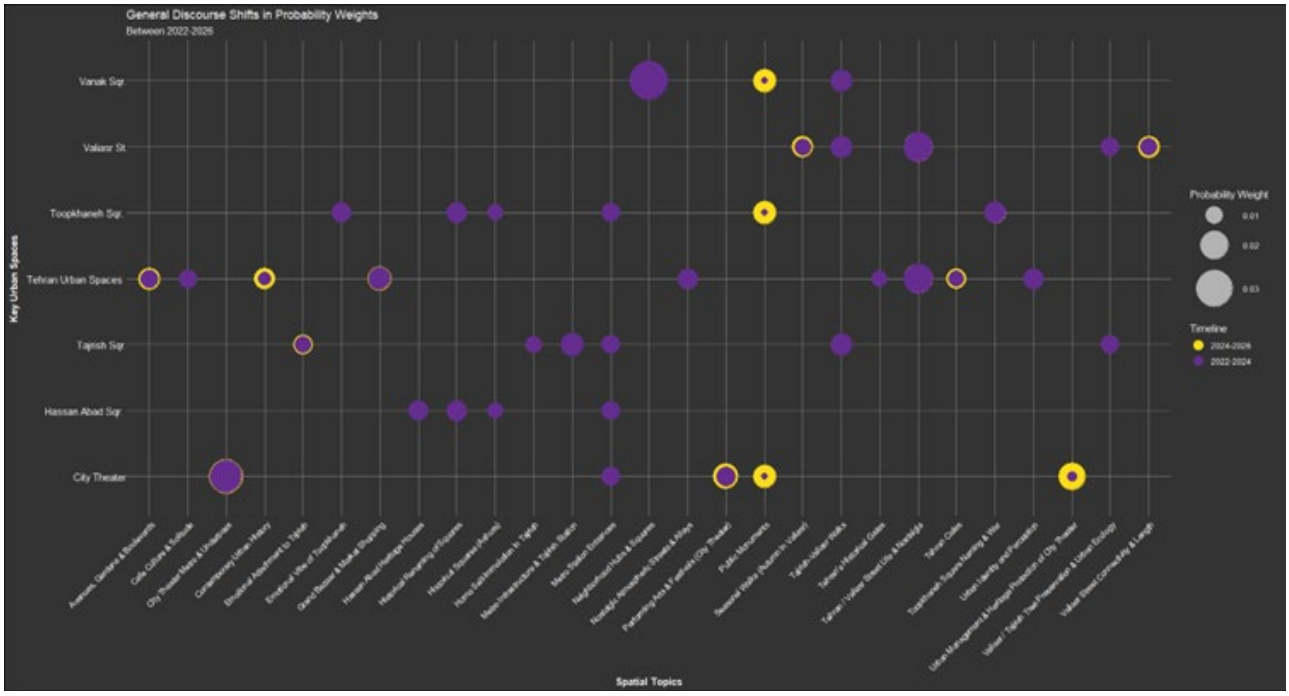
| Title | Methodology | Tools / Approaches | Relation |
|---|---|---|---|
| Topic Modeling and Sentiment Analysis of Twitter Discussions on COVID-19 from Spatial and Temporal Perspectives (AlAgha, 2021) | Topic modeling using LDA and sentiment analysis on tweets from different countries | Collecting data from Tweeter API. Cleansing the data for ,topic modeling<br><br>Tokenization and tagging ,with ArkTweetNLP library<br><br>Data normalization (texts) changing the emoticons to basic emotions using ,Dictionary<br><br>Removing stop words etc. As cleansing the texts before ,topic modeling<br><br>Use of Gensim library to im- ,prove LDA's limitation<br><br>Use of k means to find the optimal quantity of top- ,ics<br><br>Use of MLlib and multi-com- ,puter execution<br><br>Use of VADER model for sentiment analysis | .1 A comprehensive description of topic modeling and LDA algorithm to be used<br><br>.2 Data harvesting and pre-processing steps<br><br>.3 Use of text-based data to be merged and pooled as a single document to the LDA model<br><br>.4 Setting the k value for optimal number of topics<br><br>.5 Creating word clouds based on the topic modeling<br><br>.6 Use of VADER for sentiment analysis of social media data |

## Built-up query of key place keywords of Tehran's scale

| Vali'asr | | Tajrish | | Toopkhaneh | |
|---|---|---|---|---|---|
| Keyword in Farsi | Keyword in English | Keyword in Farsi | Keyword in English | Keyword in Farsi | Keyword in English |
| ولیعصر تهران | Valiasr Tehran | میدان تجریش | Tajrish Square | میدان توپخانه | Toopkhaneh Square |
| خیابون ولیعصر تهران | Valiasr Street Tehran | میدون تجریش | // | میدان توپخونه | Toopkhooneh Square |
| خیابان ولیعصر تهران | // | مترو تجریش | Tajrish Metro | میدون توپخونه | // |
| ولیعصر_تهران | Valiasr_Tehran | بازار تجریش | Tajrish Market | میدون توپخانه | Toopkhaneh Square |
| خیابان_ولیعصر_تهران | Valiasr_Street_Tehran | میدان تجریش تهران | Tajrish Square Tehran | میدان توپخانه | // |
| **Vanak** | | میدون تجریش تهران | // | میدون توپخانه | // |
| میدان ونک | Vanak Square | بازار تجریش تهران | Tajrish Market Tehran | میدون توپخونه | Toopkhooneh Square |
| میدون ونک | Vanak Square | میدان_تجریش_تهران | Tajrish_Square_Market | میدان توپخونه | // |
| میدان ونک تهران | Vanak Square Tehran | میدان_تجریش | Tajrish_Square | میدان توپ خانه | Toop Khaneh Square |
| میدون ونک تهران | Vanak Square Tehran | مترو_تجریش | Tajrish_Metro | میدان توپ خونه | Toop Khooneh Square |
| ونک | Vanak | بازار_تجریش | Tajrish_Market | میدون توپ خانه | Toop Khaneh Square |
| میدان_ونک_تهران | Vanak_Square_Tehran | بازار_تجریش_تهران | Tajrish_Market_Tehran | میدون توپ خونه | // |
| ونک_تهران | Vanak_Tehran | **Hassan Abad Square** | | میدان امام خمینی | Imam Khomeini Square |
| میدان_ونک | Vanak_Square | میدان حسن آباد تهران | Hassan Abad Square Tehran | میدون امام خمینی | // |
| **Sabzeh Meydan Square** | | میدون حسن آباد تهران | // | میدان توپخانه تهران | Toopkhaneh Square Tehran |
| سبزه میدان | Sabzeh Meydan | میدان حسن‌آباد تهران | HassanAbad Square Tehran | میدون توپخانه تهران | // |
| سبزه میدان تهران | Sabzeh Meydan Tehran | میدون حسن‌آباد تهران | // | میدان توپخونه تهران | Toopkhooneh Square Tehran |
| سبزه میدون تهران | Sabzeh Meydoon Tehran | میدان حسنآباد تهران | HassanAbad Square Tehran | میدون توپخونه تهران | // |
| سبزه میدون | Sabzeh Meydoon | میدون حسنآباد تهران | // | میدان توپ خانه تهران | Toopkhaneh Square Tehran |
| سبزه_میدان_تهران | Sabzeh_Meydan_Tehran | حسن آباد | Hassan Abad | میدون توپ خانه تهران | // |
| **Tehran Grand Bazaar** | | حسن آباد تهران | Hassan Abad Tehran | میدون توپ خونه تهران | // |
| بازار بزرگ تهران | Tehran Grand Bazaar | حسن‌آباد تهران | HassanAbad Tehran | میدان توپ خونه تهران | // |
| بازار تهران | Tehran Bazaar | حسنآباد تهران | // | میدان_توپخانه_تهران | Toopkhaneh_Square_Tehran |
| بازار بزرگ | Grand Bazaar | حسن‌آباد_تهران | HassanAbad_Tehran | میدان توپخانه | Toopkhaneh Square |
| بازار_تهران | Bazaar_Tehran | میدان_حسن‌آباد_تهران | HassanAbad_Square_Tehran | | |

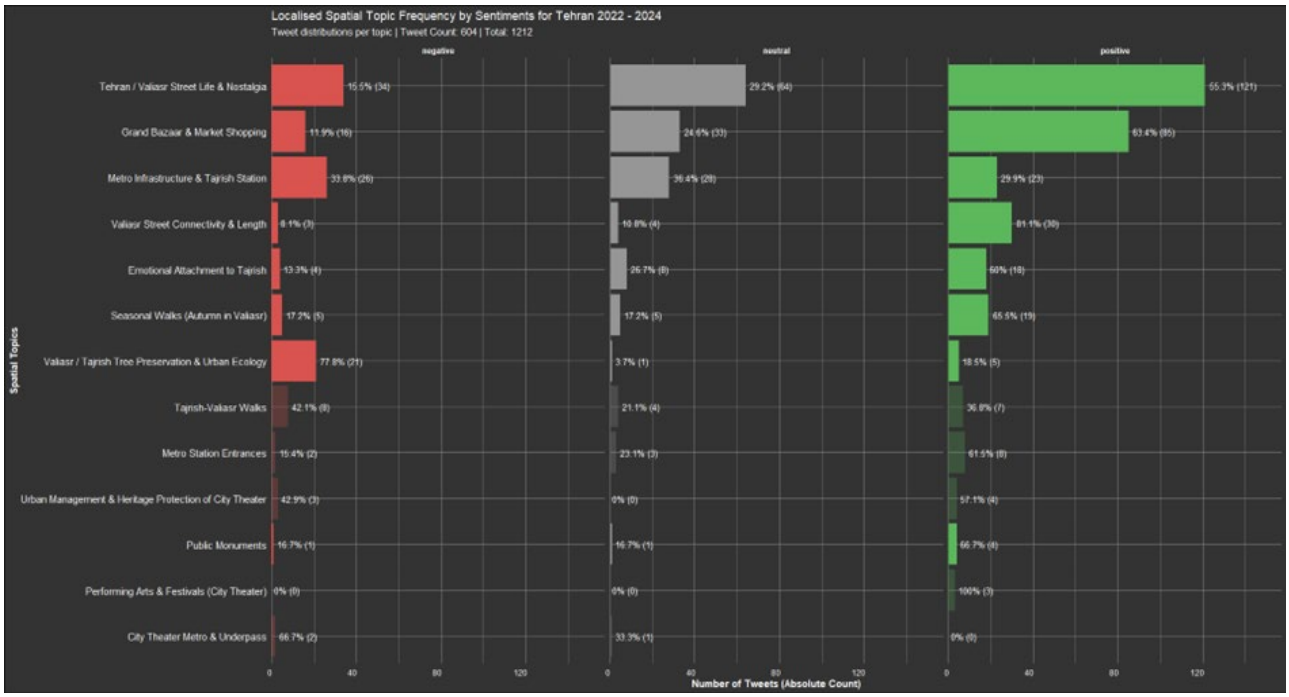
# Results

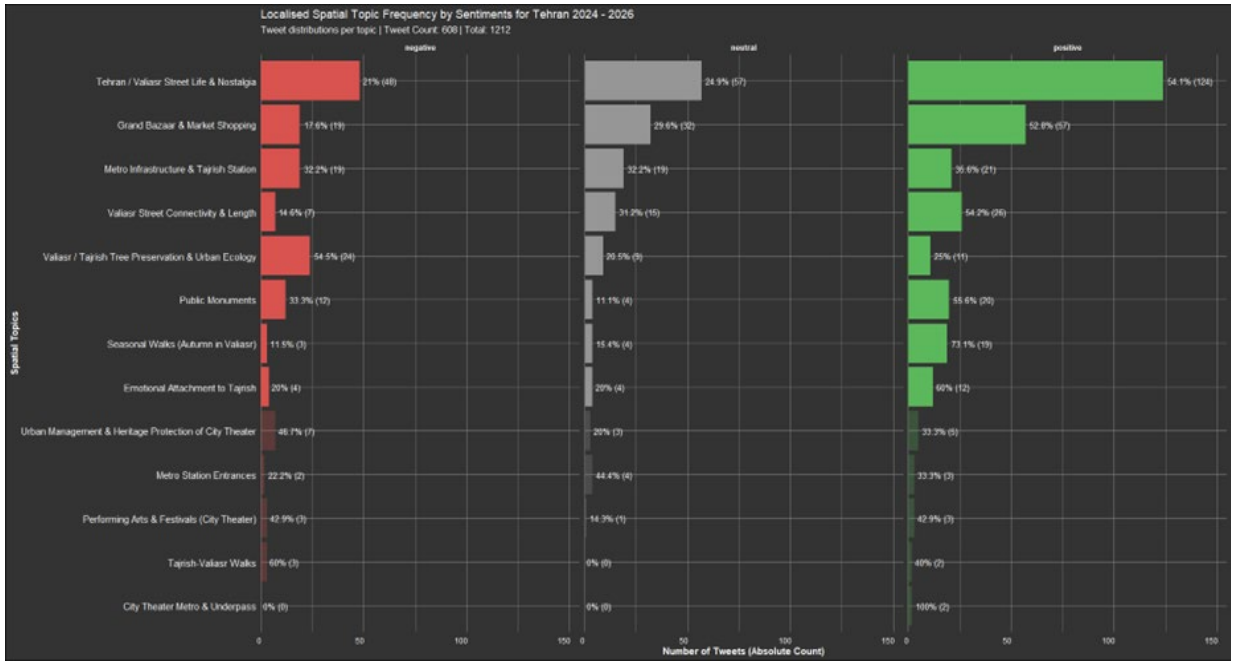
## Discourse shifts of Tehran






## Tehran scale 2022-2026

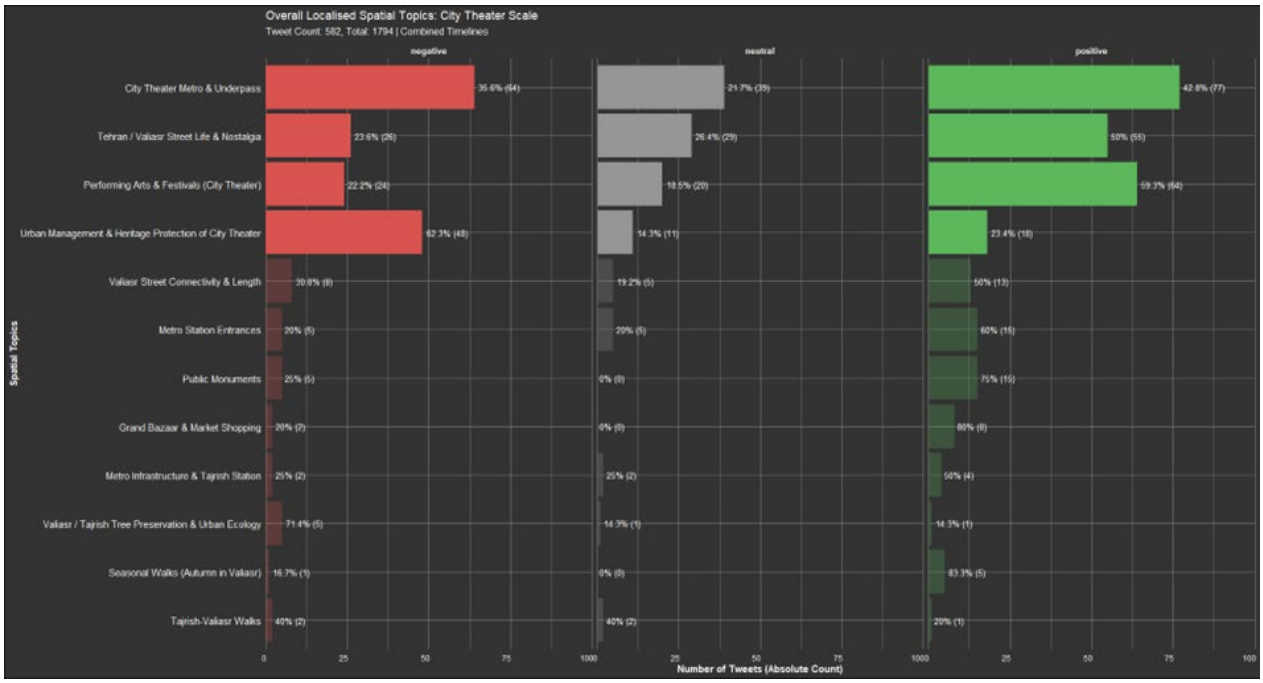




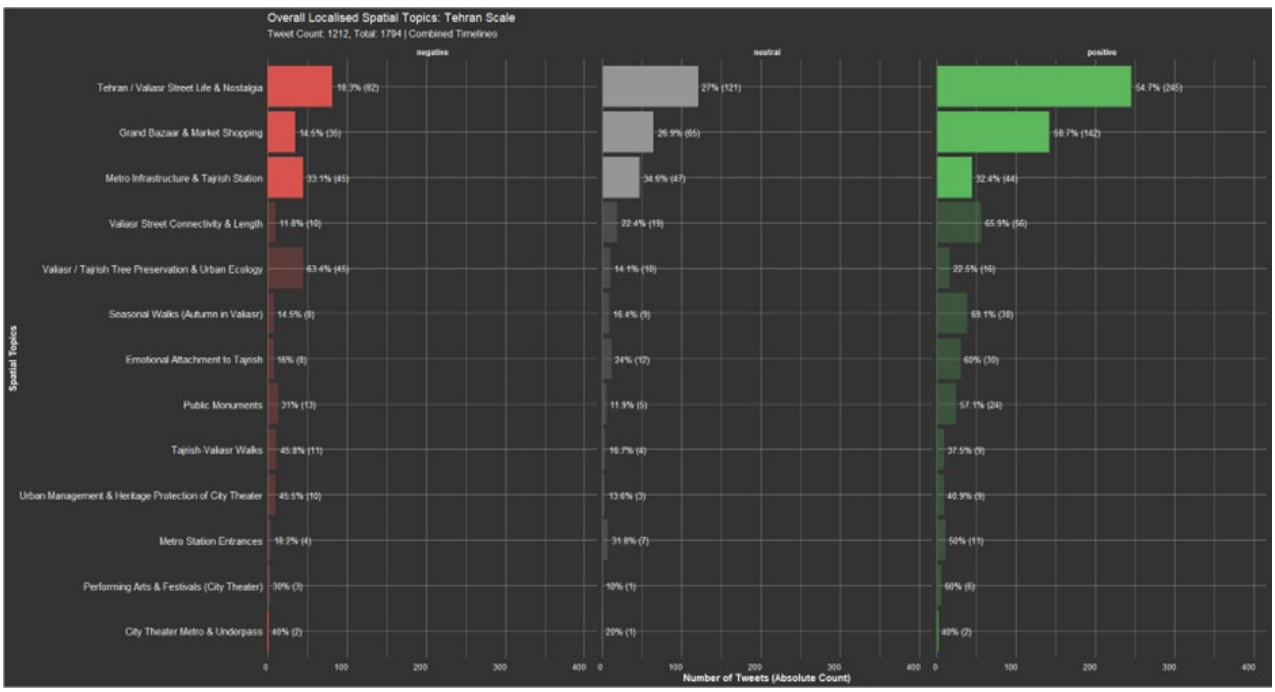

## Combined Timelines

Assessing the combined timelines is another way to review the results. This combination helps to gain a discrete classification of City Theater and non-city-theater scopes. The topics in the plot below significantly display how topics relevant to City Theater are fitted to the focused scale, while the other non-relevant topics are distributed on the larger scale. [VIII]

I've found this way of visualization quite helpful for distinguishing how truly a topic is sentimentally valued on both large and small scales. The discrete classification also helps to understand how significant each topic is on its defined scale. For example, we can understand the topic dominance percentage of the city theater underpass among the large-scale topics. And it can be compared with the other scale. Moreover, in this way, it is easier to identify the topics that have overlaps in terms of sentimental scores.


Overall Localised Spatial Topics: City Theater Scale
Tweet Count: 582, Total: 1794 | Combined Timelines
negative
neutral
positive
Spatial Topics
City Theater Metro & Underpass
Tehran / Valiasr Street Life & Nostalgia
Performing Arts & Festivals (City Theater)
Urban Management & Heritage Protection of City Theater
Valiasr Street Connectivity & Length
Metro Station Entrances
Public Monuments
Grand Bazaar & Market Shopping
Metro Infrastructure & Tajrish Station
Valiasr / Tajrish Tree Preservation & Urban Ecology
Seasonal Walks (Autumn in Valiasr)
Tajrish-Valiasr Walks
35.6% (64)
23.6% (26)
22.2% (24)
62.3% (48)
30.8% (8)
20% (5)
25% (5)
20% (2)
25% (2)
71.4% (5)
16.7% (1)
40% (2)
21.7% (39)
26.4% (29)
19.5% (20)
14.3% (11)
19.2% (5)
20% (5)
0% (0)
0% (0)
25% (2)
14.3% (1)
0% (0)
40% (2)
42.8% (77)
50% (55)
59.3% (64)
23.4% (18)
50% (13)
60% (15)
75% (15)
80% (8)
50% (4)
14.3% (1)
83.3% (5)
20% (1)
0
25
50
75
100
Number of Tweets (Absolute Count)

IX


Overall Localised Spatial Topics: Tehran Scale
Tweet Count: 1212, Total: 1794 | Combined Timelines
negative
neutral
positive
Spatial Topics
Tehran / Valiasr Street Life & Nostalgia
Grand Bazaar & Market Shopping
Metro Infrastructure & Tajrish Station
Valiasr Street Connectivity & Length
Valiasr / Tajrish Tree Preservation & Urban Ecology
Seasonal Walks (Autumn in Valiasr)
Emotional Attachment to Tajrish
Public Monuments
Tajrish-Valiasr Walks
Urban Management & Heritage Protection of City Theater
Metro Station Entrances
Performing Arts & Festivals (City Theater)
City Theater Metro & Underpass
18.3% (82)
14.5% (35)
33.1% (45)
11.8% (10)
63.4% (45)
14.5% (8)
16% (8)
31% (13)
45.8% (11)
45.5% (10)
18.2% (4)
30% (3)
40% (2)
27% (121)
26.9% (65)
34.6% (47)
22.4% (19)
14.1% (10)
16.4% (9)
24% (12)
11.9% (5)
16.7% (4)
13.6% (3)
31.8% (7)
10% (1)
20% (1)
54.7% (245)
58.7% (142)
32.4% (44)
65.9% (56)
22.5% (16)
69.1% (38)
60% (30)
57.1% (24)
37.5% (9)
40.9% (9)
50% (11)
60% (6)
40% (2)
0
100
200
300
400
Number of Tweets (Absolute Count)

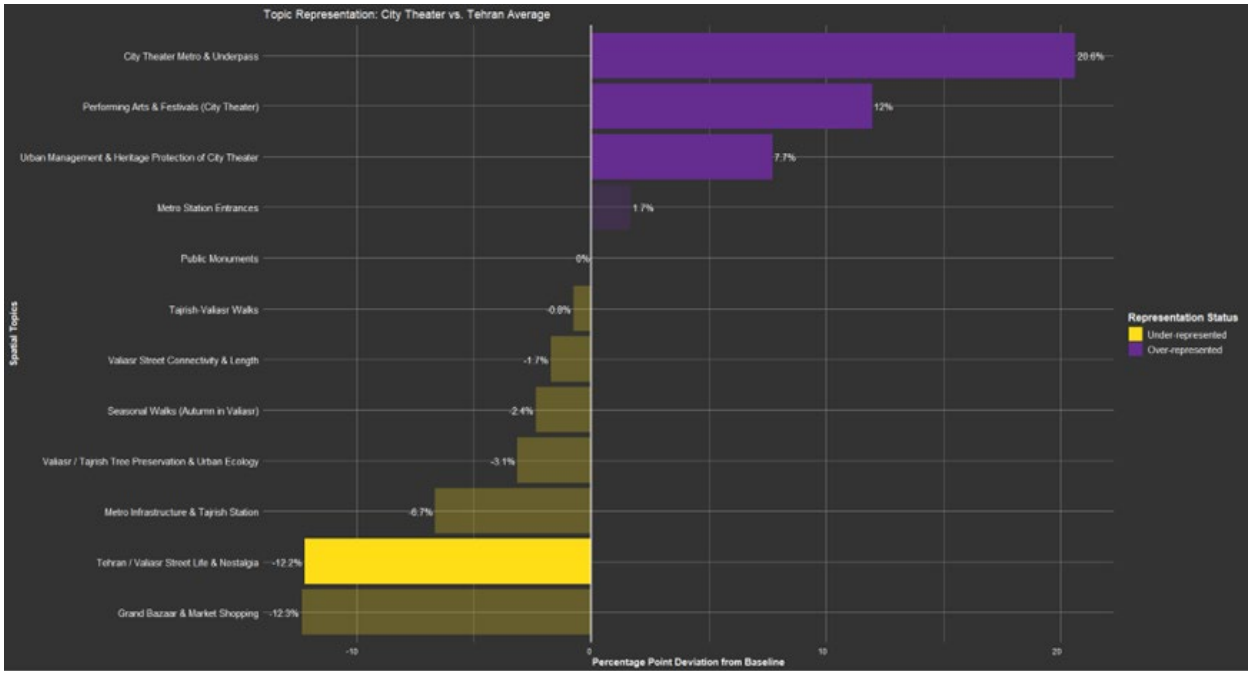
## Over/Under representation based on volume difference

The representation analysis is based on the difference between the City Theater (CT) and Global Baseline (Reference) volume percentages:

Representation Index = (ct_count / ct_total) * 100 – (global_count / global_total) * 100

The abovementioned equation was applied to the topics belonging to the City Theater scale. Since there was only one topic not belonging to the specified scale, “Emotional Attachment to Tajrish”, the values of both scales are integrated in the following table. At last, the outputs were visualized in a plot.

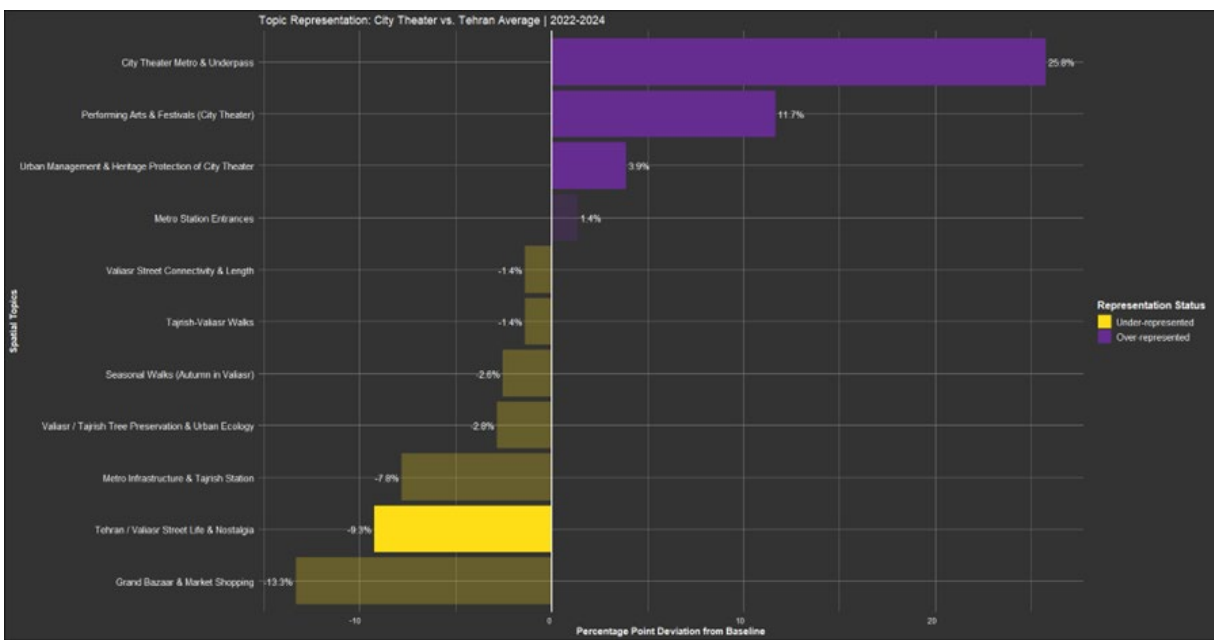
| .No | Topic Name | CT Count | CT pct (%) | Global pct (%) | Representation (%) index |
|---|---|---|---|---|---|
| 1 | City Theater Metro & Under-pass | 180 | 30.92784 | 10.31215 | 20.61568 |
| 2 | Performing Arts & Festivals ((City Theater | 108 | 18.5567 | 6.57748 | 11.97922 |
| 3 | Urban Management & Heri-tage Protection of City Theater | 77 | 13.23024 | 5.518395 | 7.711846 |
| 4 | Metro Station Entrances | 25 | 4.295533 | 2.619844 | 1.675689 |
| 5 | Metro Infrastructure & Tajrish Station | 8 | 1.37457 | 8.026756 | -6.65219 |
| 6 | Public Monuments | 20 | 3.436426 | 3.455964 | -0.01954 |
| 7 | Tajrish-Valiasr Walks | 5 | 0.859107 | 1.616499 | -0.75739 |
| 8 | Valiasr Street Connectivity & Length | 26 | 4.467354 | 6.187291 | -1.71994 |
| 9 | Seasonal Walks (Autumn in (Valiasr | 6 | 1.030928 | 3.400223 | -2.3693 |
| 10 | Valiasr / Tajrish Tree Preserva-tion & Urban Ecology | 7 | 1.202749 | 4.347826 | -3.14508 |
| 11 | Tehran / Valiasr Street Life & Nostalgia | 110 | 18.90034 | 31.10368 | -12.2033 |
| 12 | Grand Bazaar & Market Shop-ping | 10 | 1.718213 | 14.04682 | -12.3286 |
| 13 | Emotional Attachment to Tajrish | NA | NA | 2.787068 | NA |
| **Total** | | 582 | 100.0 | 100.0 excl. topic) (13 ~ 97.2 | |



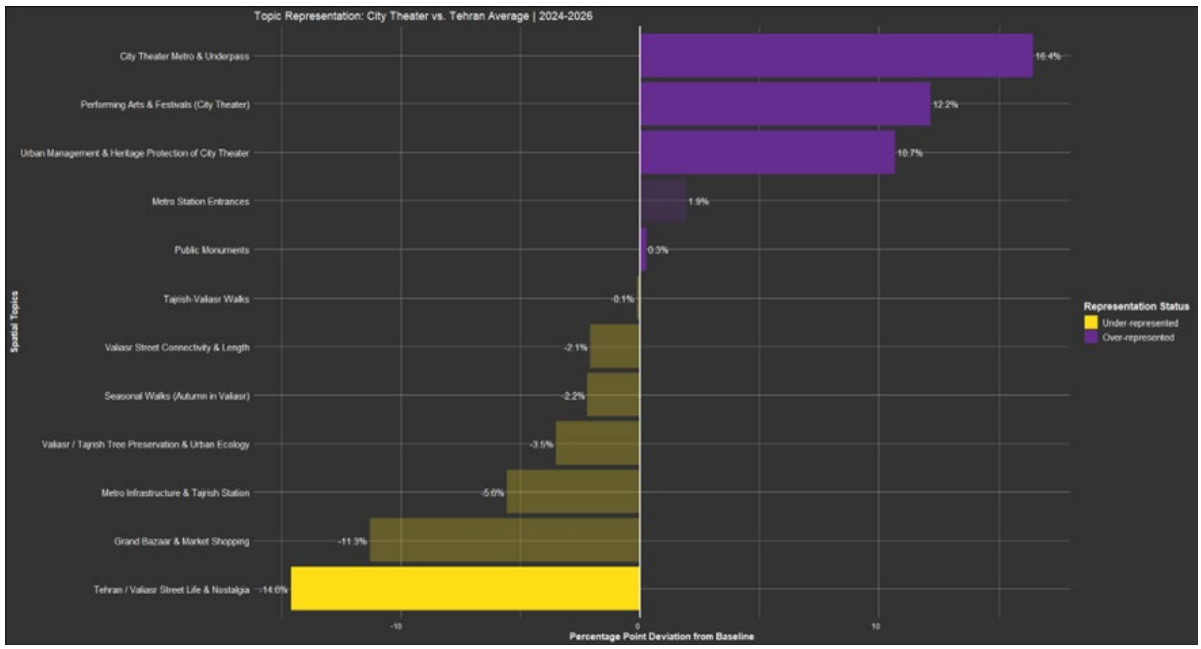

## Representations between 22 and 26

## Before 2024

xi



After 2024, it is interesting to see that the public monument shifted to being overrepresented, for it was not part of the City Theater discourse before 2024. The City Theater joined the topic group after the intervention.



# Data Management Checklist

# DATA MANAGEMENT CHECKLIST

**Instruction**

This checklist is relevant for all graduation projects of the Master AUBS. The form is intended to highlight common aspects of graduation projects that require particular attention with regard to planning the research and data management. Relevant information and supplementary sources regarding each question are provided below each question.

With this checklist, the faculty wants to avoid that students unexpectedly find themselves in complex and stressful situations, in which ethical or privacy matters and/or other laws and regulations become an issue. In projects involving humans, certain types of data processing increase the risks to the human participants: planning such projects requires additional evaluations and advice from university staff before ethical approval can be received and the project can begin. In the case of a graduation project, obtaining additional advice or permits may delay the project with an extra education period or semester. To avoid this, it is recommended that students set up a graduation project with a low level of risk. Therefore, all students have to check their risk, by completing this checklist before their A1.

The first section of the checklist (A) should be completed by all students, together with their supervisor, during the planning of the graduation project, before the A1. It does not need to be submitted to anyone for review or approval. Please consider questions 1 to 3 carefully in relation to the intended graduation project, and answer with 'yes' or 'no'.

The second section of the checklist (B) should only be completed if the graduation project involves working with data from human participants. In that case, the student and their supervisor must apply for and receive ethical approval from the Human Research Ethics Committee (HREC) before the project can begin (see the paragraph 'Explanation and follow-up' after the questions). The student can submit the application to the HREC, but the supervisor is responsible for making sure that the project is compliant with relevant privacy regulations and ethical policies.

| **Section A. General considerations** | yes | no |
|---|---|---|
| 1. Is the graduation project conducted as part of an internship (at a company), or as part of a research project at TU Delft?<br><br>If a student's graduation project is conducted at a company or as part of a research project at the university, questions of data ownership and intellectual property rights need to be addressed in a written graduation or internship agreement before the project begins. Students and their supervisor should consult the Intellectual Property Rights of Students webpage. Additional information can also be found in the Extended Personal Research Data Workflow. | | ✔ |
| 2. Does the project involve conducting (part of) the research outside the Netherlands?<br><br>Students who intend to travel abroad (even to other EU countries) for study, exchange, research, internship, or graduation project purposes need to follow the Travel Safety Protocol. This includes attending a mandatory Travel Safety Training Session: see the Disclaimer. | | ✔ |
| 3. Will the research involve processing data from humans, such as running a survey, conducting interviews or workshops, collecting data through social media or internet forums, or re-using existing datasets about humans provided by a third party? (If 'yes', see follow-up questions 4 to 13 in Checklist B.)<br><br>Students who work with data from human participants must complete the next section and apply for and receive ethical approval from the Human Research Ethics Committee (HREC) before conducting the research. | ✔ | |

| **Section B. Extended risk factors** (only if question 3 has been answered with 'yes'.) | yes | no |
|---|---|---|
| 4. Will the project involve participants who may be considered vulnerable, such as the elderly, refugees or asylum seekers, ethnic minorities, patients, or people with disabilities?<br><br>Participants who may suffer very adverse consequences (for instance, due to discrimination) if their personal data became publicly available can be considered vulnerable. | | ✔ |
| 5. Will the project involve participants who cannot themselves give informed consent for taking part in the project, but for whom consent must be obtained from a legal guardian?<br><br>Participants who cannot give informed consent can include, for instance, children or participants with intellectual disabilities, mental disorders, or dementia. Such participants are also considered vulnerable in the context of the General Data Protection Regulation (GDPR). | | ✔ |
| 6. Will the project involve processing any of the special categories of personal data below?<br>- Race<br>- Ethnicity<br>- Criminal offence data<br>- Political opinion<br>- Union membership<br>- Religious or philosophical beliefs<br>- Sex life and/or sexual orientation<br>- Health data (including measurements such as heart rate)<br>- Biometric or genetic data (including fingerprints, iris scanning, facial recognition)<br><br>The General Data Protection Regulation (GDPR) defines a stricter rules for processing special categories of personal data. If it is necessary to process these data in a project, it is it is important to provide additional safeguards. | | ✔ |
| 7. Will the project involve processing personal data that could be considered sensitive, such as the ones listed below?<br>- Information about a person's income, debts, or other payments<br>- Information about a person's (un-)employment status<br>- Information about a person's performance at school or work<br>- Information about relationship problems or (gambling) addiction<br>- Information about poverty, domestic violence, or youth welfare/social work involvement<br><br>Some types of personal data are considered sensitive, because they can have a high impact on the privacy of the data subject if other persons gain access to these data. Sensitive personal data should only be processed if necessary: in such cases, additional safeguards need to be put in place. | | ✔ |
| 8. Will the project involve processing video-recordings, or photographs of participants?<br><br>TU Delft considers photographic and video-materials of research participants to be sensitive personal data. If such data need to be processed, additional safeguards must be put in place. | | ✔ |

| **Section B. Extended risk factors** (only if question 3 has been answered with 'yes'.) | yes | no |
|---|---|---|
| 9. Will the project involve sharing or transferring personal data between multiple partners or collaborating organisations involved, such as between TU Delft and an internship company?<br><br>According to privacy law, sharing personal data between organisations requires a privacy agreement to be in place: setting this up takes time, and requires support from additional university staff. Furthermore, personal data sharing can potentially expose research participants to different types of risks: these risks must be considered in the ethical application. | | ✔ |
| 10. Will the project involve deception, or covert observation of participants?<br><br>In some types of research, obtaining informed consent for processing participants' personal data is not an option: for instance, if the research involves deception, or the research is covert (conducted without participants knowing about it). In such situations, the steps to mitigate risks to participants are important, and an alternative legal basis for processing the participant's data needs to be established with the help of additional support staff. | | ✔ |
| 11. Will the project involve working with social media data?<br><br>Social media data are personal data, but since it is usually not possible to ask for informed consent for processing social media data, another legal basis for processing the participant's data needs to be established. Processing of social media data also involves legal considerations related to terms of use of data from third-party platforms: therefore, research with social media data requires expert support on privacy, ethics, and legal matters. | ✔ | |
| 12. Will the project involve using learning algorithms or other AI to analyse, combine, or otherwise process data from participants?<br><br>The use of AI in research involves many considerations in terms of data protection, ethics, security, and intellectual property: for more information, see TU Delft's Instructions for use of Generative AI. | | ✔ |
| 13. Will the project involve participants who are based in a country or countries outside of the EU?<br><br>Students affiliated with TU Delft must comply with Dutch and EU regulations of personal data processing (GDPR). Furthermore, the student and their supervisor must make sure that the research complies with local (privacy) legislations of any foreign destinations. Additional support from an external (local) expert may be required. | ✔ | |

**Explanation and follow-up**

If you have answered 'no' to all questions 4 to 13, your project is likely to be considered low or minimal-risk: see the paragraph 'Projects with minimal or low-risk' on the next page.

If you have answered 'yes' to one or more of the questions 4 to 13, your research likely involves extended or high risks to participants, according to the General Data Protection Regulation (GDPR) and TU Delft's privacy and ethical policies: for information regarding such projects, see the paragraph 'Projects with extended or high-risk' on the next pages.

**Projects with minimal or low-risk**

If you have answered 'no' to questions 4 to 13, your project is likely to be considered low-risk. This does not mean that the project involves no risks at all, but suggests that these risks can likely be addressed by the student and supervisor in the application to the Human Research Ethics Committee (HREC) within the timeline for a graduation project and without need for additional support.

Compiling the HREC application:
An application to the HREC generally involves a Data Management Plan (DMP), a risk-identification and mitigation checklist, and informed consent materials. Master's students at ABE who intend to compile a HREC application are advised to make use of the following support documents:

- the student guide
- the Example Data Management Plan for MSc projects

The graduation supervisor is responsible for the student's project and ethical application, and must provide support for compiling the HREC application documents.

Additional support
For low-risk student graduation projects, compiling of the HREC application documents should be done by the student in consultation with the supervisor. The Faculty Data Steward can be contacted for individual questions at datasteward-BK@tudelft.nl: however, the Data Steward does not provide detailed feedback on student DMPs for low-risk HREC applications.

Additional resources
The HREC has guides available for completing the checklist and for compiling informed consent materials. Additionally, the Guide to the Extended Personal Research Data Workflow has been created to help researchers and students who work with human participants comply with both GDPR principles and TU Delft's policies on Data Management and Human Research Ethics.

Timeline
Minimal or low-risk HREC applications are generally processed faster than extended or high-risk applications (see the paragraph below). Nevertheless, the initial evaluation by the HREC usually takes approximately 2 weeks, and may take longer during busy periods or holiday: see the HREC website for up-to-date information. Additionally, the application may require revisions before final approval is granted. If you do not receive an initial response about your ethical application after 4 weeks from the time of submission, you may follow up with the HREC to enquire about an update.

**Projects with extended or high-risk**

If you have answered 'yes' to one or more of questions 4 to 13, there are potential increased risks related to how data from human participants will be processed in your project. These risks will need to be addressed in consultation with the Data Steward and other relevant support staff before submitting the ethical application to the Human Research Ethics Committee (HREC).

Compiling the HREC application
An application to the HREC generally involves a Data Management Plan (DMP), a risk-identification and mitigation checklist, and informed consent materials. Master's students at ABE who intend to compile a HREC application are advised to make use of the following support documents:

- the Ethical Approval & Data Management Planning Student Information
- the Example Data Management Plan for MSc projects

The graduation supervisor is responsible for the student's project and ethical application, and must provide support for compiling the HREC application documents.

Additional support
Once the DMP has been compiled and reviewed by the supervisor, feedback should be requested from the Data Steward via DMPonline. After this, any other necessary support staff will need to be contacted. Crucially, if the project involves one or multiple ways of personal data processing that could result in high-risk to the participants according to the GDPR, the TU Delft Privacy Team must be consulted to establish whether or not a Data Protection Impact Assessment (DPIA) is required.

Additional resources
The HREC has guides available for completing the checklist and for compiling informed consent materials. Additionally, the Guide to the Extended Personal Research Data Workflow has been created to help researchers and students who work with human participants comply with both GDPR principles and TU Delft's policies on Data Management and Human Research Ethics.

Timeline
It can take a long time to compile a complete research plan and HREC application for projects involving extended risks. DMP feedback from the Data Steward usually takes around 2 weeks, but can take longer during busy periods or holidays. Receiving additional support from other staff, such as the Privacy Team, can take anywhere from a few days to multiple weeks, depending on the project and capacity of university staff. If a DPIA is deemed necessary, it can take anywhere from 4 weeks to several months.

It is important to note that advice from the Privacy Team or other support staff, as well as any additional documents (such as necessary contracts, or a DPIA, if needed) must be in place before the application is submitted to the HREC. The initial evaluation by the HREC can be processed in 2 weeks, but may take longer during busy periods or holidays: see the HREC website for up-to-date information. Additionally, the application may require revisions before final approval is granted. If you do not receive an initial response about your ethical application after 4 weeks from the time of submission, you may follow up with the HREC to enquire about an update.

Considering the limited time available for students conducting their graduation projects, students working with data from human participants are strongly advised to prioritise low-risk research projects. If a student project necessitates processing data in ways that are considered extended or high-risk, both student and supervisor need to be aware of the extended processing times involved in obtaining ethical approval and beginning the graduation project.

# Scripts and Workflows

| | | | | | |
|---|---|---|---|---|---|
| سبزه میدون | Sabzeh Meydoon | میدون حسنآباد تهران | // | میدان توپ خانه تهران | Toopkhaneh Square Tehran |
| سبزه_میدان_تهران | Sabzeh_Meydan_Tehran | حسن آباد | Hassan Abad | میدون توپ خانه تهران | // |
| **Tehran Grand Bazaar** | | حسن آباد تهران | Hassan Abad Tehran | میدون توپ خونه تهران | // |
| بازار بزرگ تهران | Tehran Grand Bazaar | حسنآباد تهران | HassanAbad Tehran | میدان توپ خونه تهران | // |
| بازار تهران | Tehran Bazaar | حسنآباد تهران | // | میدان_توپخانه _تهران | Toopkhaneh_ Square_Tehran |
| بازار بزرگ | Grand Bazaar | حسنآباد_تهران | HassanAbad_Tehran | میدان توپخانه | Toopkhaneh Square |
| بازار_تهران | Bazaar_Tehran | میدان_حسنآباد_تهران | HassanAbad_Square_Tehran | | |

*Apendix 3. Keywords for Tehran's scale query*

## Data Analysis

## AI acknowledgment

As an acknowledgment, the use of generative AIs, Le Chat and Gemini, was only done on the analytical scripts and model runs, only for modification, refinement and efficiency purposes.

**Remove taggings, URLs, Hashtags etc.**

```
# In R, install and run the necessary libraries

library(tidyverse)
library(tidytext)
library(dplyr)

# Load your data
raw_data <- read.csv("data/your_data_file.csv", stringsAsFactors = FALSE)

# Select the Text column and rename it to 'text'
raw_data <- raw_data |>
  mutate(copytext = fullText)


# Customize and modify the column names

colnames(raw_data)[2] <- "textshrt"
colnames(raw_data)[1] <- "text"
```

```
# Initial Clean: Remove noise but keep Farsi text intact; Removing
# URLs, Hashtags, and tagging


clean1 <- raw_data |>
  mutate(
    # Remove URLs, Mentions, and Hashtags
    text = str_replace_all(text, "https?://\\S+|www\\.\\S+", ""),
    text = str_replace_all(text, "@\\w+", ""),
    text = str_replace_all(text, "#\\w+", ""),
    # Remove extra whitespace
  text = str_squish(text)
 )

# Remove rows(tweets) where the text is the same
deduplicated <- clean1 |>
  distinct(text, .keep_all = TRUE)

# Check how many tweets you lost
original_count <- nrow(clean1)
new_count <- nrow(deduplicated)
cat("Removed", original_count - new_count, "duplicate tweets.")

# Detect any foreign language in the dataframe
deduplicated <- deduplicated |>
  mutate(
    lang = detect_language(text)
  )

# Check how many Arabic/foreign language tweets you have
arabic_tweets <- deduplicated %>% filter(lang == "ar")
farsi_tweets <- deduplicated %>% filter(lang == "fa")
cat("Found", nrow(arabic_tweets), "Arabic tweets.")
unique(arabic_tweets$text)

# You may face some tweets that are actually Farsi, but some of
# them are detected as Arabic, you can identify the exact value
```

```
missed_tweets <- arabic_tweets|>
  filter(arabic_tweets$text == "من هروقت میرم مترو تئاترشهر:")
missed_tweets$lang <- "fa"


# Adding the missed fa tweet
farsi_tweets_filtered <- farsi_tweets|>
  add_row(missed_tweets)


# Organizing and choosing the desired columns to keep
cleaned_data_withkey <- farsi_tweets_filtered|>
  mutate(text2 = text, doc_id = row_number())|>
  select(text, text2, type, createdAt, doc_id)


# Export to Parquet (The "Memory-Safe" bridge) better for
# processing data in Python
write_parquet(cleaned_data_withkey,
"output/tweets_ready_withkey.parquet")
```

*Apendix 4. R Code block for Taggings, URLs, Hashtags and White space removal*

**Installing ParsBERT**

The ParsBERT should be downloaded as a package from the Huggingface website from the developer's repository. The preparation for running the ParsBERT model in Python is described in the following steps:

1. Inside a Python interface software, in this case I used the PyCharm app, install the BERTopic and ParsBERT libraries.
2. Load the pre-processed data in your code and run the model based on the written code

**Topic Modeling using Word Embedding**

Once the Python environment is set and other necessary libraries are installed (check the code block), you can start the topic modeling, using the word embedding with BERT and ParsBERT.

1. After that, the texts were tokenized into single words and removed the stop words. Then a counting function will calculate the number of words occurring in each document, i.e. tweet. Finally, the output will be cast to a document-term matrix also known as "dtm".
2. The LDA model will be applied next by defining the scalar value of *k*. After several trials and errors of *k* value configurations, the optimal with which

gives a reasonable number of topics was five. So, the *k* value was set to five as the finalized indicator.

3. To identify the top terms of each topic, another script was written to extract the top terms per topic along with their percentage of occurrence.
4. Also, to identify the most relevant topic in each document, or tweet per say, the topic distribution per document percentage was calculated using the specific script.
5. At last, the outputs were written in the comma-separated values (csv) as a compatible and simple file format for further processing and data merging.

The steps are depicted in the following code block:

```
# in Python, install & import the necessary libraries

import pandas as pd
import numpy as np
from parsivar import Normalizer
from bertopic import BERTopic
import torch
from transformers import AutoConfig, AutoTokenizer,
AutoModelForSequenceClassification, pipeline



# Load Data
df = pd.read_parquet("tweets_ready_withkey.parquet")
docs = df['text'].astype(str).tolist()


#  Normalization (Parsivar)
# This handles the half-space (ZWNJ) and Arabic-to-Farsi character
# conversion
normalizer = Normalizer()
docs = [normalizer.normalize(d) for d in docs]

# Word Embeddings & Topic Modeling (ParsBERT + BERTopic)
# We use BERTopic which can utilize ParsBERT under the hood.
# It handles UMAP (dimension reduction) and HDBSCAN (clustering).
topic_model = BERTopic(
    language="multilingual", # Supports Farsi context
    calculate_probabilities=True,
    verbose=True
)

# This creates the embeddings and finds topics automatically
topics, probs = topic_model.fit_transform(docs)


# Reduce Outliers
```

```
new_topics = topic_model.reduce_outliers(docs, topics,
strategy="c-tf-idf")

# Update Topics with Stopwords
from sklearn.feature_extraction.text import CountVectorizer
farsi_stopwords = ["و", "در", "به", "از", "که", "با", "این", "است", "را",
"آن"]
vectorizer_model = CountVectorizer(stop_words=farsi_stopwords)
topic_model.update_topics(docs, topics=new_topics,
vectorizer_model=vectorizer_model)

# Create a dataframe from the probability matrix
prob_df = pd.DataFrame(probs).add_prefix('prob_topic_')

# Join it with your main data
df = pd.concat([df, prob_df], axis=1)

# Get the info for topics
topic_info = topic_model.get_topic_info()
print(topic_info.head(20))

# Save the run model in your project's directory (So you never
# have to run the embedding again)
topic_info.to_csv("topic_definitions_withkey_prob.csv",
index=False)
df['final_topic'] = new_topics
df.to_csv("final_tweet_results_withkey_prob.csv", index=False)
topic_model.save("my_final_farsi_model_withkey_prob")
```

*Apendix 5. Python Code block for Word Embedding using BERT*

**Sentiment Analysis**

Now that the word embedding model has run, you can run the sentiment analysis based on it and merge both results into one dataframe.

1. Then a function was defined to run the VADER sentiment analysis model as a Python script. Then, the function was applied on each tweet's text to perform the sentiment analysis.
2. The output showed that each row of text was valued based on the overall sentiment scores between the –1 and +1 range of values.
3. The texts valued with VADER sentiment analysis were merged with the outputs from the topic model based on their ID fields as the final integration of the processed data.
4. The merged data frame was written into .csv file format as the final output of the analysis.

The mentioned steps are depicted in the code block below:

```
# In Python, import the necessary libraries

import pandas as pd
from transformers import pipeline
import torch

# Load the existing topic results
df = pd.read_csv("final_tweet_results_withkey_prob.csv")
docs = df['text'].astype(str).tolist()

# Setup the sentiment pipeline
# Point to the downloaded sentiment model
model_path = "ParsBERT-Sentiment/"
device = 0 if torch.cuda.is_available() else -1


# Run the sentiment analysis model as a pipeline
print("Running Sentiment Analysis")
sentiment_pipe = pipeline(
    "sentiment-analysis",
    model=model_path,
    tokenizer=model_path,
    device=device
)

# Execute in batches(For faster performance)
results = sentiment_pipe(docs, batch_size=32, truncation=True)

# Map the labels
# This fixes the 'recommended' value lable issue and standardizes
# your labels
def map_sentiment(output):
    label = output['label'].lower()
    # Adjust these keys based on what your specific model outputs
    if "not" in label or "neg" in label:
        return "negative"
    elif "recom" in label or "pos" in label:
        return "positive"
    else:
        return "neutral"

df['sentiment_label'] = [map_sentiment(res) for res in results]
df['sentiment_score'] = [res['score'] for res in results]

# Save the final merged file
df.to_csv("data_with_sentiment_key1_prob.csv", index=False)
print("Sentiment appended to your existing topics.")
```

*Apendix 6. Python Code block for Sentiment Analysis*

**Translation**

After finishing the data processing, we can now move forward with translating the data for readability and clearness. For data privacy and ethical concerns, the Argostranslate was used as an offline translation model that supports Farsi to English translations.

```
# In R, install and import necessary libraries

library(tidyverse)

library(tidytext)

library(dplyr)

library(reticulate)


# In R, install the Argostranslate model as a Python library

py_install("argostranslate")


# Run the argos model in Python inside the R script

py_run_string("

import argostranslate.package

argostranslate.package.update_package_index()

available_packages =
argostranslate.package.get_available_packages()

package_to_install = next(

    filter(

        lambda x: x.from_code == 'fa' and x.to_code == 'en',

        available_packages

    )

)

argostranslate.package.install_from_path(package_to_install.downlo
ad())

")

# Define a function to run the translation on the data

translate_text <- function(text) {

  py_run_string(glue::glue('

  import argostranslate.translate

  from_code = "fa"
```

```
  to_code = "en"

  translated_text = argostranslate.translate.translate("{text}",
from_code, to_code)

  '))

  return(py$translated_text)

}

# Define a function to remove any special characters that may
# affect the translation process in Argos

# Function to preprocess text (preserves Farsi text)

preprocess_text <- function(text) {

  # Define specific problematic characters to remove

  chars_to_remove <- "؛؟،«»\"'!()[]{}<>*"

  # Remove specific characters

  text <- remove_chars(text, chars_to_remove)

  # Collapse multiple spaces

  text <- str_replace_all(text, "\\s+", " ")

  # Trim leading/trailing whitespace

  text <- str_trim(text)

  return(text)

}

# Function for replacing keywords. The model will translate the
# place names too! You need to define a function to replace the
# farsi names into special nouns before running the whole
# translation. The function needs a keyword list to replace any
# matching Farsi names, with their English equivalents.


# Define the keywords list

keywords_replace <- c(

  "تئاتر شهر" = "City_theater",

  "تئاترشهر" = "City_theater",

  "تئاتر_شهر" = "City_theater",

  "تئاتر شهر تهران" = "City_theater",

  "تیاتر شهر" = "City_theater",
```

```
"تیاترشهر تهران" = "City_theater",
"تئاترشهر تهران" = "City_theater",
"تئاترشهر تهران" = "City_theater",
"تئاترشهر_تهران" = "City_theater",
"تئاتر_شهر_تهران" = "City_theater",
"تیاترشهر_تهران" = "City_theater",
"تئاترشهرتهران" = "City_theater",
"تئاترشهرتهران" = "City_theater",
"تیاترشهرتهران" = "City_theater",
"تیاترشهرتهران" = "City_theater",
"میدان ونک" = "Vanak",
"میدون ونک" = "Vanak",
"میدان ونک تهران" = "Vanak",
"میدون ونک تهران" = "Vanak",
"ونک" = "Vanak",
"میدان_ونک_تهران" = "Vanak",
"ونک_تهران" = "Vanak",
"میدان_ونک" = "Vanak",
"تجریش" = "Tajrish",
"میدان تجریش"= "Tajrish",
"میدون تجریش"= "Tajrish",
"مترو تجریش"= "Tajrish",
"بازار تجریش"= "Tajrish",
"میدان تجریش تهران"= "Tajrish",
"میدون تجریش تهران"= "Tajrish",
"بازار تجریش تهران"= "Tajrish",
"میدان_تجریش_تهران"= "Tajrish",
"میدان_تجریش"= "Tajrish",
"مترو_تجریش"= "Tajrish",
"بازار_تجریش" = "Tajrish",
"بازار_تجریش_تهران" = "Tajrish",
"خیابان ولیعصر" = "Valiasr",
```

```
"ولیعصر تهران" = "Valiasr",
"خیابون ولیعصر تهران" = "Valiasr",
"خیابان ولیعصر تهران" = "Valiasr",
"ولیعصر_تهران" = "Valiasr",
"خیابان_ولیعصر_تهران" = "Valiasr",
"میدان ولیعصر" = "Valiasr_sqr",
"میدون ولیعصر" = "Valiasr_sqr",
"خیابان_ولیعصر" = "Valiasr",
"چهارراه ولیعصر" = "City_theater",
"تقاطع ولیعصر" = "City_theater",
"مترو تئاترشهر" = "City_theater",
"متروی تئاترشهر" = "City_theater",
"ایستگاه تئاترشهر" = "City_theater",
"میدان توپخانه" = "Toopkhaneh",
"میدان توپخونه" = "Toopkhaneh",
"میدون توپخونه" = "Toopkhaneh",
"میدون توپخانه" = "Toopkhaneh",
"میدان توپ‌خانه" = "Toopkhaneh",
"میدون توپ‌خانه" = "Toopkhaneh",
"میدون توپ‌خونه" = "Toopkhaneh",
"میدان توپ‌خونه" = "Toopkhaneh",
"میدان توپ خانه" = "Toopkhaneh",
"میدان توپ خونه" = "Toopkhaneh",
"میدون توپ خانه" = "Toopkhaneh",
"میدون توپ خونه" = "Toopkhaneh",
"میدان امام خمینی" = "Toopkhaneh",
"میدون امام خمینی" = "Toopkhaneh",
"میدان توپخانه تهران" = "Toopkhaneh",
"میدون توپخانه تهران" = "Toopkhaneh",
"میدان توپخونه تهران" = "Toopkhaneh",
"میدون توپخونه تهران" = "Toopkhaneh",
"میدان توپ خانه تهران" = "Toopkhaneh",
```

```
  "میدون توپ خانه تهران" = "Toopkhaneh",
  "میدون توپ خونه تهران" = "Toopkhaneh",
  "میدان توپ خونه تهران" = "Toopkhaneh",
  "میدان_توپخانه_تهران" = "Toopkhaneh",
  "بازار بزرگ تهران" = "Bazaar",
  "بازار تهران" = "Bazaar",
  "بازار بزرگ" = "Bazaar",
  "سبزه میدان" = "Bazaar",
  "سبزه میدان تهران" = "Bazaar",
  "سبزه میدون تهران" = "Bazaar",
  "سبزه میدون" = "Bazaar",
  "بازار_تهران" = "Bazaar",
  "سبزه_میدان_تهران" = "Bazaar",
  "میدان حسن آباد تهران" = "HassanAbad",
  "میدون حسن آباد تهران" = "HassanAbad",
  "میدان حسن‌آباد تهران" = "HassanAbad",
  "میدون حسن‌آباد تهران" = "HassanAbad",
  "میدان حسنآباد تهران" = "HassanAbad",
  "میدون حسنآباد تهران" = "HassanAbad",
  "حسن آباد تهران" = "HassanAbad",
  "حسن‌آباد تهران" = "HassanAbad",
  "حسنآباد تهران" = "HassanAbad",
  "حسن‌آباد_تهران" = "HassanAbad",
  "میدان_حسن‌آباد_تهران" = "HassanAbad",
  "تهران" = "Tehran",
  "انقلاب" = "Enghelab",
  "خیابان انقلاب" = "Enghelab_st",
  "خیابون انقلاب" = "Enghelab_st",
  "میدان انقلاب" = "Enghelab_sqr",
  "میدون انقلاب" = "Enghelab_sqr"
)
# Define the function
```

```
replace_keywords <- function(text) {

  for (keyword in names(keywords_replace)) {

    text <- str_replace_all(text, fixed(keyword),
keywords_replace[keyword])

  }

  return(text)

}

# Load the processed tweets and topic definitions

processed_data <-
read.csv("output/Py_results/data_with_sentiment_key1_prob.csv")

topic_definitions <-
read.csv("output/Py_results/topic_definitions_withkey_prob.csv")


# Renaming the files into dataframes with selected columns

ready_data <- processed_data|>

  select(doc_id, createdAt, text,type, final_topic,

         sentiment_label, sentiment_score) |>

  mutate(txtenglish = text)


topic_lists <-  topic_definitions|>

  mutate(name_eng = Name)|>

  mutate(represent_eng = Representation)


# Preprocess text and removing special

# characters that affect the translation

ready_data$txtenglish <- sapply(ready_data$txtenglish,
preprocess_text)

topic_lists$name_eng <- sapply(topic_lists$name_eng,
preprocess_text)

topic_lists$represent_eng <- sapply(topic_lists$represent_eng,
preprocess_text)


# Replace keywords
```

```
ready_data$txtenglish <- sapply(ready_data$txtenglish,
replace_keywords)

topic_lists$name_eng <- sapply(topic_lists$name_eng,
replace_keywords)

topic_lists$represent_eng <- sapply(topic_lists$represent_eng,
replace_keywords)

# Apply translation to each text

translated <- ready_data

ready_topics <- topic_lists


# Translating the main data text into Eng

translated$txtenglish <- sapply(translated$txtenglish,
translate_text)


# translating the topic definitions and names into Eng

ready_topics$name_eng <- sapply(ready_topics$name_eng,
translate_text)

ready_topics$represent_eng <- sapply(ready_topics$represent_eng,
translate_text)


# Making the topic definitions ready for export

topic_translated <- ready_topics


# Making the topic probability weights as the joinery component

probs <- select(processed_data, -createdAt, -text,-type,
-final_topic, -sentiment_label,
-sentiment_score, -text2)


# adding the probs fields to the data

full_data <- translated|>

  left_join(probs, by = "doc_id" )


# Save the translated data into .csv for the next step
```

```
write.csv(topic_translated,"output/translatedtopic_definitions_keyprob.csv")

write.csv(full_data,
"output/full_data_translated_keyprob_sentiments.csv")
```

**Processed Data Visualisation**

```
# In R, install and run the necessary libraries

# install.packages(c("gghighlight", "tidyverse", "dplyr", "ggplot2") → remove the # sign to install the package for the first time

library(tidyverse)

library(dplyr)

library(ggplot2)

library(gghighlight)


# Define our Spatial General Topic IDs from processed final topic
# numbers in combined vector

spatial_ids <- c(0, 1, 2, 4, 5, 11, 12, 13, 16, 18, 19, 21, 22, 30, 31, 32, 37, 41, 46, 52, 55, 76, 82, 84, 86, 88, 93)


# Define the localised topics

spatial_ids_explicit <- c(0, 2, 5, 11, 13, 16, 18, 19, 30, 31, 32, 37, 86)

#  We'd better replace the topic names with the refined versions for clarity and
# consistency in the visualizations

spatialtopic_names <- c( "0" = "Tehran / Valiasr Street Life & Nostalgia",

            "1" = "Neighborhood Hubs & Squares",

            "2" = "Grand Bazaar & Market Shopping",

            "4" = "Toopkhaneh Square Naming & War",

            "5" = "City Theater Metro & Underpass",

            "11" = "Metro Infrastructure & Tajrish Station",

            "12" = "Avenues, Gardens & Boulevards",
```

```
            "13" = "Performing Arts & Festivals (City Theater)",
            "16" = "Valiasr Street Connectivity & Length",
            "18" = "Valiasr / Tajrish Tree Preservation & Urban Ecology",
            "19" = "Urban Management & Heritage Protection of City Theater",
            "21" = "Nostalgic Atmospheric Streets & Alleys",
            "22" = "Urban Identity and Perception",
            "30" = "Public Monuments",
            "31" = "Metro Station Entrances",
            "32" = "Seasonal Walks (Autumn in Valiasr)",
            "37" = "Emotional Attachment to Tajrish",
            "41" = "Emotional Vibe of Toopkhaneh",
            "46" = "Hassan Abad Heritage Houses",
            "52" = "Historical Renaming of Squares",
            "55" = "Historical Squares (Ashura)",
            "76" = "Contemporary Urban History",
            "82" = "Cafe Culture & Solitude",
            "84" = "Homa Self-Immolation in Tajrish",
            "86" = "Tajrish-Valiasr Walks",
            "88" = "Tehran's Historical Gates",
            "93" = "Tehran Cafes")
# We can already prepare the topics to create the word clouds
after refining the
# English topic titles

# filter and prepare spatial topics

spatial_topics <- topics |>
  filter(Topic %in% spatial_ids)


# exporting the classified topics to create wordclouds

topics_for_cloud <- spatial_topics|>
  mutate(eng_name = spatialtopic_names[as.character(Topic)])|>
```

```
  mutate(explicitness = case_when(Topic %in% spatial_ids_explicit
~ "Explicit",

                                  TRUE ~ "Non-explicit"))|>

  select(Topic, eng_name, explicitness, Score, Count,
English_Word,

         English_Representation)

# Write the output topic as .csv for wordcloud and further

# analytical steps

write.csv(topics_for_cloud ,
"output/Py_results/plot_topics_english.csv")

# Further processing with the analyzed tweets; temporal,
# categorical classifications and organization, topic name updates
# etc.

spatial_tweets <- tweets |>

  select(-X)|>

  filter(final_topic %in% spatial_ids)|>

  mutate(Topic = final_topic)|>

  left_join(counttopics, by = "Topic")|>

  mutate(eng_name = spatialtopic_names[as.character(Topic)]) |>

  mutate(explicitness = case_when(Topic %in% spatial_ids_explicit
~ "Explicit",

                                  TRUE ~ "Non-explicit"))|>

  mutate(year = year(parse_date_time(createdAt, "a b d H:M:S z
Y")))|>

  filter(year != "2021")|>

  mutate(timeline = if_else(year < 2024, "Before_24",
"After_24"))|>

  select(doc_id, text, txtenglish, Topic, eng_name, Count,

         sentiment_label, sentiment_score, explicitness, type,
year, timeline, createdAt,

         all_of(spatial_columns))

# You can save the processed data right away at this stage.
# Remove the Hash sign, so you can run the line.

write.csv(spatial_tweets,"output/Py_results/mostfinal_analysis_out
put.csv")
```

```
# Prepare the City Theater data for 2022-24 topic-sentiments for
# the plot
plot_data_ct24 <- spatial_tweets |>
  filter(explicitness == "Explicit") |>
  filter(type == "ct" & timeline == "Before_24") |>
  group_by(sentiment_label, eng_name) |>
  summarise(topic_count = n(), .groups = "drop") |>
  complete(sentiment_label, eng_name, fill = list(topic_count =
0)) |>


  # Row-wise percentage (Topic-based)
  group_by(eng_name) |>
  mutate(
    topic_total = sum(topic_count),
  # If the topic has at least one tweet, calculate %; otherwise, 0
    topic_pct = ifelse(topic_total > 0, (topic_count /
topic_total) * 100, 0),
    label_text = paste0(round(topic_pct, 1), "% (", topic_count,
")")
  ) |>
  ungroup() |>
  # Sort topics by total volume so the chart is organized
  mutate(eng_name = reorder(eng_name, topic_total))


# CT 24 plot in dark theme
ggplot(plot_data_ct24, aes(x = eng_name, y = topic_count,
                           fill = sentiment_label)) +
  geom_col(aes(alpha = eng_name %in% c(
    "City Theater Metro & Underpass",
    "Tehran / Valiasr Street Life & Nostalgia",
    "Urban Management & Heritage Protection of City Theater",
```

```
    "Performing Arts & Festivals (City Theater)"
  )),
  show.legend = FALSE) +
  scale_alpha_manual(values = c("TRUE" = 1.0, "FALSE" = 0.3),
                     guide = "none") +
  geom_text(aes(label = label_text),
            hjust = -0.1,
            size = 3.5,
            color = "white") +
  facet_wrap(~sentiment_label, scales = "fixed") +
  coord_flip() +
  theme_minimal() +
  scale_fill_manual(values = c(
    "neutral" = "#969696",
    "negative" = "#d9534f",
    "positive" = "#5cb85c"
  )) +
  labs(title = "Localised Spatial Topic Frequency by Sentiments
for City Theater 2022 - 2024",
       subtitle = "Tweet distributions per topic | Tweet Count:
266, Total: 582",
       ## nrows ct 24: 266 // Total ct: 582
       x = "Spatial Topics",
       y = "Number of Tweets (Absolute Count)") +
  theme(
    strip.text = element_text(size = 9, face = "bold", color =
"white"),
    ## panel & background
    panel.background = element_rect(fill = NA, color = NA),
    plot.background  = element_rect(fill = "gray20"),
    panel.grid.major  = element_line(color = "gray50"),
    panel.grid.minor  = element_line(color = "gray40"),
```

```
    ## axis
    axis.text = element_text(color = "white"),
    axis.text.y = element_text(size = 10, hjust = 1, color =
"white"),
    axis.text.x = element_text(hjust = 0.9, color = "white"),
    axis.title = element_text(color = "white", face = "bold"),
    ## title and subtitle
    plot.title = element_text(color = "white"),
    plot.subtitle = element_text(color = "white")
  ) +
  # Expand the horizontal axis to prevent the percentage labels
from being cut off
  scale_y_continuous(expand = expansion(mult = c(0, 0.2)))


# The same preparation for CT 2024-26
plot_data_ct26 <- spatial_tweets |>
  filter(explicitness == "Explicit") |>
  filter(type == "ct" & timeline == "After_24") |>
  group_by(sentiment_label, eng_name) |>
  summarise(topic_count = n(), .groups = "drop") |>
  complete(sentiment_label, eng_name, fill = list(topic_count =
0)) |>
  group_by(eng_name) |>
  mutate(
    topic_total = sum(topic_count),
    topic_pct = ifelse(topic_total > 0, (topic_count /
topic_total) * 100, 0),
    label_text = paste0(round(topic_pct, 1), "% (", topic_count,
")")
  ) |>
  ungroup() |>
  mutate(eng_name = reorder(eng_name, topic_total))
```

```
# CT 26
ggplot(plot_data_ct26, aes(x = eng_name, y = topic_count,
                           fill = sentiment_label)) +
  geom_col(aes(alpha = eng_name %in% c(
    "City Theater Metro & Underpass",
    "Tehran / Valiasr Street Life & Nostalgia",
    "Urban Management & Heritage Protection of City Theater",
    "Performing Arts & Festivals (City Theater)"
  )),
  show.legend = FALSE) +
  geom_text(aes(label = label_text),
            hjust = -0.1,
            size = 3.5,
            color = "white") +
  facet_wrap(~sentiment_label, scales = "fixed") +
  coord_flip() +
  theme_minimal() +
  scale_fill_manual(values = c(
    "neutral" = "#969696",
    "negative" = "#d9534f",
    "positive" = "#5cb85c"
  )) +
  theme(
    strip.text = element_text(size = 9, face = "bold", color =
"white"),
    ## panel & background
    panel.background = element_rect(fill = NA, color = NA),
    plot.background  = element_rect(fill = "gray20"),
    panel.grid.major  = element_line(color = "gray50"),
    panel.grid.minor  = element_line(color = "gray40"),
    ## axis
```

```
    axis.text = element_text(color = "white"),
    axis.text.y = element_text(size = 10, hjust = 1, color =
"white"),
    axis.text.x = element_text(hjust = 0.9, color = "white"),
    axis.title = element_text(color = "white", face = "bold"),
    ## title and subtitle
    plot.title = element_text(color = "white"),
    plot.subtitle = element_text(color = "white")
  ) +
  labs(title = "Localised Spatial Topic Frequency by Sentiments
for City Theater 2024 - 2026",
       subtitle = "Tweet distributions per topic | Tweet Count:
316 | Total: 582",
       # nrows ct 26: 316 // Total ct: 582
       x = "Spatial Topics",
       y = "Number of Tweets (Absolute Count)") +
  # Expand the horizontal axis to prevent the percentage labels
from being cut off
  scale_y_continuous(expand = expansion(mult = c(0, 0.2)))



# Preparing data for Tehran scale between 2022 and 2026


plot_data_teh24 <- spatial_tweets |>
  filter(explicitness == "Explicit") |>
  filter(type == "not_ct" & timeline == "Before_24") |>
  group_by(sentiment_label, eng_name) |>
  summarise(topic_count = n(), .groups = "drop") |>
  complete(sentiment_label, eng_name, fill = list(topic_count =
0)) |>
  group_by(eng_name) |>
  mutate(
    topic_total = sum(topic_count),
```

```
    topic_pct = ifelse(topic_total > 0, (topic_count /
topic_total) * 100, 0),
    label_text = paste0(round(topic_pct, 1), "% (", topic_count,
")")
  ) |>
  ungroup() |>
  mutate(eng_name = reorder(eng_name, topic_total))


plot_data_teh26 <- spatial_tweets |>
  filter(explicitness == "Explicit") |>
  filter(type == "not_ct" & timeline == "After_24") |>
  group_by(sentiment_label, eng_name) |>
  summarise(topic_count = n(), .groups = "drop") |>
  complete(sentiment_label, eng_name, fill = list(topic_count =
0)) |>
  group_by(eng_name) |>
  mutate(
    topic_total = sum(topic_count),
    topic_pct = ifelse(topic_total > 0, (topic_count /
topic_total) * 100, 0),
    label_text = paste0(round(topic_pct, 1), "% (", topic_count,
")")
  ) |>
  ungroup() |>
  mutate(eng_name = reorder(eng_name, topic_total))


# Plotting the data in dark theme


# Plot Teh 26 dark
ggplot(plot_data_teh26, aes(x = eng_name, y = topic_count,
                            fill = sentiment_label)) +
  geom_col(aes(alpha = eng_name %in% c(
    "Tehran / Valiasr Street Life & Nostalgia",
```

```
    "Valiasr Street Connectivity & Length",
    "Tehran/Valiasr Street Life & Nostalgia",
    "Emotional Attachment to Tajrish",
    "Seasonal Walks (Autumn in Valiasr)",
    "Public Monuments",
    "Valiasr / Tajrish Tree Preservation & Urban Ecology",
    "Grand Bazaar & Market Shopping",
    "Metro Infrastructure & Tajrish Station"
  )),
  show.legend = FALSE) +
  geom_text(aes(label = label_text),
            hjust = -0.1,
            size = 3.5,
            color = "white") +
  scale_alpha_manual(values = c("TRUE" = 1.0, "FALSE" = 0.25),
                     guide = "none") +
  facet_wrap(~sentiment_label, scales = "fixed") +
  coord_flip() +
  theme_minimal() +
  scale_fill_manual(values = c(
    "neutral" = "#969696",
    "negative" = "#d9534f",
    "positive" = "#5cb85c"
  )) +
  theme(
    strip.text = element_text(size = 9, face = "bold", color =
"white"),
    # panel & background
    panel.background = element_rect(fill = NA, color = NA),
    plot.background  = element_rect(fill = "gray20"),
    panel.grid.major  = element_line(color = "gray50"),
```

```
    panel.grid.minor  = element_line(color = "gray40"),
    ## axis
    axis.text = element_text(color = "white"),
    axis.text.y = element_text(size = 11, hjust = 1, color =
"white"),
    axis.text.x = element_text(hjust = 0.9, color = "white"),
    axis.title = element_text(color = "white", face = "bold"),
    ## title and subtitle
    plot.title = element_text(color = "white"),
    plot.subtitle = element_text(color = "white")
  ) +
  labs(title = "Localised Spatial Topic Frequency by Sentiments
for Tehran 2024 - 2026",
       subtitle = "Tweet distributions per topic | Tweet Count:
608 | Total: 1212",
       # nrows teh 26: 608 // Total ct: 1212
       x = "Spatial Topics",
       y = "Number of Tweets (Absolute Count)") +
  # Expand the horizontal axis to prevent the percentage labels
from being cut off
  scale_y_continuous(expand = expansion(mult = c(0, 0.25)))
```

**Topic Visualisation (Wordcloud)**

```
# In R, install and run the libraries
library(tidyverse)
library(ggwordcloud)
```

```r
# Load data

df_eng <- read.csv("output/Py_results/plot_topics_english.csv")


# Filter and Prepare

spatial_plot_data <- df_eng |>

  filter(explicitness == "Explicit")|>

  mutate(eng_name = str_wrap(eng_name, width = 30)) # Wrap long
names for better fit

# Create the Plot

p_spatial <- ggplot(spatial_plot_data, aes(label = English_Word,
size = Score, color = factor(Topic))) +

  geom_text_wordcloud_area(rm_outside = FALSE, shape = "circle") +

  scale_size_area(max_size = 25) +

  facet_wrap(~eng_name, ncol = 4) +

  theme_minimal() +

  theme(

    strip.text = element_text(size = 10, face = "bold", color =
"darkslategrey"),

    legend.position = "none",

    plot.title = element_text(size = 18, face = "bold", hjust =
0.5),

    plot.subtitle = element_text(size = 12, hjust = 0.5)

  ) +

  labs(

    title = "Spatial and Urban Topics of Tehran",

    subtitle = "Analysis of explicit urban spaces, heritage, and
daily street life"

  )

# Generate the plot by running this line

p_spatial
```

**Topic Distribution**

```
# Topic Distribution
topic_distribution <- spatial_tweets|>
  group_by(eng_name) |>
  summarise(topic_count = n(), .groups = "drop") |>
  complete(eng_name, fill = list(topic_count = 0)) |>
  mutate(
    topic_total = sum(topic_count),
    topic_pct = ifelse(topic_total > 0, (topic_count / topic_total) * 100, 0),
    label_text = paste0(round(topic_pct, 1), "% (", topic_count, ")")
  ) |>
  ungroup() |>
  mutate(eng_name = reorder(eng_name, topic_count, sum))


# General Topic Distribution
ggplot(topic_distribution, aes(x = eng_name, y = topic_count)) +
  geom_col(show.legend = FALSE, fill = "#5556a5") +
  # Adding the labels as "call-outs"
  geom_text(aes(label = label_text),
            hjust = -0.15,
            size = 3.5,
            # angle = 45,
            color = "white",
            fontface = "italic") +
  # facet_grid(type ~ timeline, scales = "free_y", space = "free_y") +
  coord_flip() +
  # Distinct Urban Sentiment Palette
  theme_minimal() +
  labs(title = "Tehran Spatial Topics Distribution",
```

```
       #Spatial or general spatial topics

       subtitle = "Labels: Tweet Frequency per Topic - Total
Tweets: 1,794",

       # general total 2,948 / spatial total 1,794

       y = "Tweet Volume (Absolute Count)",

       x = "Spatial Topic") +

  # Increase the expansion on the Y-axis to prevent labels from
being cut off

  scale_y_continuous(expand = expansion(mult = c(0, 0.1)))+

  theme(

    # strip.background = element_rect(fill = "whitesmoke", color =
"gray80"),

    # strip.text = element_text(size = 9, face = "bold"),

    panel.grid.minor = element_blank(),

    # strip.background = element_rect(fill = "whitesmoke", color =
"gray80"),

    ## panel & background

    panel.background = element_rect(fill = NA, color = NA),

    plot.background  = element_rect(fill = "gray20"),

    panel.grid.major  = element_line(color = "gray40"),

    ## axis

    axis.text = element_text(color = "white"),

    axis.text.y = element_text(size = 8, hjust = 1, color =
"white"),

    axis.text.x = element_text(hjust = 0.9, color = "white"),

    axis.title = element_text(color = "white", face = "bold"),

    ## title and subtitle

    plot.title = element_text(color = "white"),

    plot.subtitle = element_text(color = "white")

  )
```

```
topic_distribution_explicit <- spatial_tweets |>
  filter(explicitness == "Explicit") |>
  group_by(eng_name) |>
  summarise(topic_count = n(), .groups = "drop") |>
  complete(eng_name, fill = list(topic_count = 0)) |>
  mutate(
    topic_total = sum(topic_count),
    topic_pct = ifelse(topic_total > 0, (topic_count / topic_total) * 100, 0),
    label_text = paste0(round(topic_pct, 1), "% (", topic_count, ")")
  ) |>
  ungroup() |>
  mutate(eng_name = reorder(eng_name, topic_count, sum))


# Explicit topic distribution
ggplot(topic_distribution_explicit, aes(x = eng_name, y = topic_count)) +
  geom_col(show.legend = FALSE, fill = "#5556a5") +
  ## faa819 orange / ffde17 yellow / 2d2e8b blue / 662d91 purple
  ## 5556a5 light blue / 8c5d25 brown / d9534f red / 999999 grey / 5cb85c green
  # Adding the labels as "call-outs"
  geom_text(aes(label = label_text),
            hjust = -0.15,
            size = 3.5,
            # angle = 45,
            color = "white",
            fontface = "italic") +
  coord_flip() +
  theme_minimal() +
  labs(title = "Tehran Localised Spatial Topics Distribution",
```

```
       subtitle = "Tweet Frequency per Topic | Total Tweets:
1,794",
       x = "Spatial Topics",
       y = "Tweet Volume") +
  # Increase the expansion on the Y-axis to prevent labels from
being cut off
  scale_y_continuous(expand = expansion(mult = c(0, 0.1))) +
  theme(
    # strip.background = element_rect(fill = "whitesmoke", color =
"gray80"),
    strip.text = element_text(size = 9, face = "bold"),
    ## panel & background
    panel.background = element_rect(fill = NA, color = NA),
    plot.background  = element_rect(fill = "gray20"),
    panel.grid.major  = element_line(color = "gray40"),
    panel.grid.minor  = element_line(color = "gray50"),
    ## axis
    axis.text = element_text(color = "white"),
    axis.text.y = element_text(size = 8, hjust = 1, color =
"white"),
    axis.text.x = element_text(hjust = 0.9, color = "white"),
    axis.title = element_text(color = "white", face = "bold"),
    ## title and subtitle
    plot.title = element_text(color = "white"),
    plot.subtitle = element_text(color = "white")
  )


# Localised topic distribution
topic_distribution_explicit <- spatial_tweets |>
  filter(explicitness == "Explicit") |>
  group_by(eng_name) |>
  summarise(topic_count = n(), .groups = "drop") |>
```

```
  complete(eng_name, fill = list(topic_count = 0)) |>

  mutate(

    topic_total = sum(topic_count),

    topic_pct = ifelse(topic_total > 0, (topic_count /
topic_total) * 100, 0),

    label_text = paste0(round(topic_pct, 1), "% (", topic_count,
")")

  ) |>

  ungroup() |>

  mutate(eng_name = reorder(eng_name, topic_count, sum))


# Plot the localised topic distribution in dark theme

ggplot(topic_distribution_explicit, aes(x = eng_name, y =
topic_count)) +

  geom_col(show.legend = FALSE, fill = "#5556a5") +

  ## faa819 orange / ffde17 yellow / 2d2e8b blue / 662d91 purple

  ## 5556a5 light blue / 8c5d25 brown / d9534f red / 999999 grey /
5cb85c green

  # Adding the labels as "call-outs"

  geom_text(aes(label = label_text),

            hjust = -0.15,

            size = 3.5,

            # angle = 45,

            color = "white",

            fontface = "italic") +

  coord_flip() +

  theme_minimal() +

  labs(title = "Tehran Localised Spatial Topics Distribution",

       subtitle = "Tweet Frequency per Topic | Total Tweets:
1,794",

       x = "Spatial Topics",

       y = "Tweet Volume") +
```

```
  # Increase the expansion on the Y-axis to prevent labels from
being cut off
  scale_y_continuous(expand = expansion(mult = c(0, 0.1))) +
  theme(
    # strip.background = element_rect(fill = "whitesmoke", color =
"gray80"),
    strip.text = element_text(size = 9, face = "bold"),
    ## panel & background
    panel.background = element_rect(fill = NA, color = NA),
    plot.background  = element_rect(fill = "gray20"),
    panel.grid.major  = element_line(color = "gray40"),
    panel.grid.minor  = element_line(color = "gray50"),
    ## axis
    axis.text = element_text(color = "white"),
    axis.text.y = element_text(size = 8, hjust = 1, color =
"white"),
    axis.text.x = element_text(hjust = 0.9, color = "white"),
    axis.title = element_text(color = "white", face = "bold"),
    ## title and subtitle
    plot.title = element_text(color = "white"),
    plot.subtitle = element_text(color = "white")
  )


# Localised Topic-Sentiment distribution
topic_distribution_explicit_sentiment <- spatial_tweets |>
  filter(explicitness == "Explicit") |>
  group_by(sentiment_label, eng_name) |>
  summarise(topic_count = n(), .groups = "drop") |>
  complete(sentiment_label, eng_name, fill = list(topic_count =
0)) |>
  group_by(eng_name)|>
  mutate(
```

```
    topic_total = sum(topic_count),
    topic_pct = ifelse(topic_total > 0, (topic_count /
topic_total) * 100, 0),
    label_text = paste0(round(topic_pct, 1), "% (", topic_count,
")")
  ) |>
  ungroup() |>
  mutate(eng_name = reorder(eng_name, topic_total, sum))


# Topic Distribution per sentiments with fixed blocks dark themed
ggplot(topic_distribution_explicit_sentiment,
       aes(x = eng_name, y = topic_count, fill = sentiment_label))
+
  # Position = "fill" makes all bars the same length (0 to 1)
  geom_col(position = "fill", show.legend = TRUE) +
  # Adjust text to show proportions inside the blocks
  # Note: position_fill(vjust = 0.5) puts the text in the middle
  # of each color block
  geom_text(aes(label = label_text),
            position = position_fill(vjust = 0.5),
            size = 2.8,
            color = "white",
            fontface = "bold") +
  coord_flip() +
  # Distinct Urban Sentiment Palette
  scale_fill_manual(values = c("negative" = "#d9534f",
                               "neutral" = "#999999",
                               "positive" = "#5cb85c"),
                    name = "Sentiment") +
  # 3. Format Y axis as percentages (0% to 100%)
  scale_y_continuous(labels = scales::percent,
                     expand = expansion(mult = c(0, 0.05))) +
```

```
  theme(
    plot.background = element_rect(fill = "gray20", color = NA),
    panel.background = element_rect(fill = NA, color = NA),
    panel.grid.major = element_line(color = "gray40"),
    panel.grid.minor = element_line(color = "gray50"),
    text = element_text(color = "white"),
    axis.text = element_text(color = "white"),
    axis.text.y = element_text(size = 8),
    legend.position = "bottom",
    legend.background = element_rect(fill = "gray20", color = NA),
    legend.text = element_text(color = "white")
  ) +
  labs(title = "Tehran Localised Spatial Topic-Sentiment
Distribution",
       subtitle = "Tweet Frequency per Topic & Sentiments | Total
Tweets: 1,794",
       x = "Spatial Topics",
       y = "Tweet Volume")


## Combined scales between 2022-2026
# Combined scales before 2024
plot_data_timeline_before <- spatial_tweets |>
  filter(explicitness == "Explicit" & timeline == "Before_24") |>
  mutate(sentiment_label = case_when(
    sentiment_label == "" ~ "Neutral",
    TRUE ~ sentiment_label
  )) |>
  group_by(sentiment_label, eng_name, timeline) |>
  summarise(topic_count = n(), .groups = "drop") |>
  complete(sentiment_label, eng_name, fill = list(topic_count =
0)) |>
  group_by(eng_name) |>
```

```
  mutate(
    topic_total = sum(topic_count),
    topic_pct = ifelse(topic_total > 0, (topic_count / 
topic_total) * 100, 0),
    label_text = paste0(round(topic_pct, 1), "% (", topic_count, 
")")
  ) |>
  ungroup() |>
  mutate(eng_name = reorder(eng_name, topic_total, sum))


# Plot combined scales before ‘24
ggplot(plot_data_timeline_before, aes(x = eng_name, y = 
topic_count,
                                      fill = sentiment_label)) +
  geom_col(aes(alpha = eng_name %in% c("Public Monuments",
                                       "Tehran / Valiasr Street 
Life & Nostalgia",
                                       "City Theater Metro & 
Underpass",
                                       "Valiasr Street 
Connectivity & Length",
                                       "Urban Management & 
Heritage Protection of City Theater",
                                       "Valiasr / Tajrish Tree 
Preservation & Urban Ecology",
                                       "Metro Infrastructure & 
Tajrish Station"
  )), show.legend = FALSE) +
  geom_text(aes(label = label_text), hjust = -0.2, size = 3.5, 
color = "white") +
  # Split by Scale (Type) and Sentiment
  facet_grid(~sentiment_label, scales = "fixed") +
  scale_alpha_manual(values = c("TRUE" = 1.0, "FALSE" = 0.25), 
guide = "none") +
  coord_flip() +
```

```
  scale_fill_manual(values = c("negative" = "#d9534f", "neutral" =
"#999999", "positive" = "#5cb85c")) +

  theme_minimal() +

  labs(title = "Overal Localised Spatial Topics: Between 2022-
2024",

       subtitle = "Tweet Count: 870, Total: 1794 | Combined
Scales",

       ## nrow before: 870 // nrows after: 924

       x = "Spatial Topics",

       y = "Tweet Count") +

  theme(strip.text = element_text(face = "bold", size = 10, color
= "white"),

              strip.background = element_blank(),

              panel.background = element_rect(fill = NA, color =
NA),

              plot.background  = element_rect(fill = "gray20"),

              panel.grid.major  = element_line(color = "gray40"),

              panel.grid.minor = element_line(color = "gray30"),

              ## axis

              axis.text = element_text(color = "white"),

              axis.text.y = element_text(size = 11, hjust = 1,
color = "white"),

              axis.text.x = element_text(hjust = 0.9, color =
"white"),

              axis.title = element_text(color = "white", face =
"bold"),

              ## title and subtitle

              plot.title = element_text(color = "white"),

              plot.subtitle = element_text(color = "white")

        ) +

  scale_y_continuous(expand = expansion(mult = c(0, .2)))


# Combined scales after 2024

plot_data_timeline_after <- spatial_tweets |>
```

```
  filter(explicitness == "Explicit" & timeline == "After_24") |>
  mutate(sentiment_label = case_when(
    sentiment_label == "" ~ "Neutral",
    TRUE ~ sentiment_label
  )) |>
  group_by(sentiment_label, eng_name, timeline) |>
  summarise(topic_count = n(), .groups = "drop") |>
  complete(sentiment_label, eng_name, fill = list(topic_count =
0)) |>
  group_by(eng_name) |>
  mutate(
    topic_total = sum(topic_count),
    topic_pct = ifelse(topic_total > 0, (topic_count /
topic_total) * 100, 0),
    label_text = paste0(round(topic_pct, 1), "% (", topic_count,
")")
  ) |>
  ungroup() |>
  mutate(eng_name = reorder(eng_name, topic_total, sum))


# Plot the combined scales after ‘24
ggplot(plot_data_timeline_after, aes(x = eng_name, y =
topic_count, fill = sentiment_label)) +
  geom_col(aes(alpha = eng_name %in% c("Public Monuments",
                                       "City Theater Metro &
Underpass",
                                       "Tehran / Valiasr Street
Life & Nostalgia",
                                       "Valiasr Street
Connectivity & Length",
                                       "Urban Management &
Heritage Protection of City Theater",
                                       "Valiasr / Tajrish Tree
Preservation & Urban Ecology",
```

```
                                        "Metro Infrastructure &
Tajrish Station")),

           show.legend = FALSE) +

  geom_text(aes(label = label_text), hjust = -0.2, size = 3.5,
color = "white") +

  # Split by Scale (Type) and Sentiment

  facet_grid(~sentiment_label, scales = "fixed") +

  scale_alpha_manual(values = c("TRUE" = 1.0, "FALSE" = 0.25),
guide = "none") +

  coord_flip() +

  scale_fill_manual(values = c("negative" = "#d9534f", "neutral" =
"#999999", "positive" = "#5cb85c")) +

  labs(title = "Overal Localised Spatial Topics: Between 2024-
2026",

       subtitle = "Tweet Count: 924, Total: 1794 | Combined
Scales",

       ## nrow before: 870 // nrows after: 924

       x = "Spatial Topics",

       y = "Tweet Count") +

  theme(strip.text = element_text(face = "bold", size = 10, color
= "white"),

        strip.background = element_blank(),

        panel.background = element_rect(fill = NA, color = NA),

        plot.background  = element_rect(fill = "gray20"),

        panel.grid.major  = element_line(color = "gray40"),

        panel.grid.minor = element_line(color = "gray30"),

        ## axis

        axis.text = element_text(color = "white"),

        axis.text.y = element_text(size = 11, hjust = 1, color =
"white"),

        axis.text.x = element_text(hjust = 0.9, color = "white"),

        axis.title = element_text(color = "white", face = "bold"),

        ## title and subtitle

        plot.title = element_text(color = "white"),
```

```
        plot.subtitle = element_text(color = "white")
        ) +
  scale_y_continuous(expand = expansion(mult = c(0, .3)))
```

**Chi-square assessment**

```
# In R, install and run the necessary libraries
library(tidyverse)
library(dplyr)
# Chi-square assessment
# Filter and table
before_24_matrix <- spatial_tweets |>
  filter(explicitness == "Explicit")|>
  filter(timeline == "Before_24") |>
  select(type, sentiment_label) |>
  table()

print("BEFORE 2024")
print(before_24_matrix)
# Chi-Square
chisq.test(before_24_matrix)

# Run for the After 2024 data
after_24_matrix <- spatial_tweets |>
  filter(explicitness == "Explicit")|>
  filter(timeline == "After_24") |>
  select(type, sentiment_label) |>
  table()

print("AFTER 2024")
print(after_24_matrix)
chisq.test(after_24_matrix)
```

```
# Combined timeline with localised topics
all_time_matrix <- spatial_tweets|>
  filter(explicitness == "Explicit")|>
  select(type, sentiment_label)|>
  table()

print("SCALES Comparison")
print(all_time_matrix)
chisq.test(all_time_matrix)

### Chi-square test only on City Theater scale
ct_matrix24 <- spatial_tweets|>
  filter(type == "ct" & timeline == "Before_24" & explicitness ==
"Explicit")|>
  select(timeline, sentiment_label)|>
  table()
ct_matrix26 <- spatial_tweets|>
  filter(type == "ct" & timeline == "After_24" & explicitness ==
"Explicit")|>
  select(timeline, sentiment_label)|>
  table()

print("CT timeline-sentiment comparison_24")
print(ct_matrix24)
chisq.test(ct_matrix24)

print("CT timeline-sentiment comparison_26")
print(ct_matrix26)
chisq.test(ct_matrix26)
```

```
### Chi-square test only on Not City Theater scale

not_ct_matrix24 <- spatial_tweets|>

  filter(type == "not_ct" & timeline == "Before_24" & explicitness
== "Explicit")|>

  select(timeline, sentiment_label)|>

  table()


print("Not-CT timeline-sentiment comparison_24")

print(not_ct_matrix24)

chisq.test(not_ct_matrix24)




not_ct_matrix26 <- spatial_tweets|>

  filter(type == "not_ct" & timeline == "After_24" & explicitness
== "Explicit")|>

  select(timeline, sentiment_label)|>

  table()


print("Not-CT timeline-sentiment comparison_26")

print(not_ct_matrix26)

chisq.test(not_ct_matrix26)
```

*Apendix 7. Chi-square test code block*

**Cooccurrence matrices**

```
# In R, install and run the libraries

library(tidyverse)

library(dplyr)

library(ggplot2)

# Create a dataframe based on spatial topics and their
# corresponding key places

topic_to_place <- tribble(
```

```
  ~Topic, ~place_name,
  #  City Theater Specifics
  "5",    "City Theater",
  "13",   "City Theater",
  "19",   "City Theater",
  # Tajrish Specifics
  "11",   "Tajrish Sqr.",
  "37",   "Tajrish Sqr.",
  "84",   "Tajrish Sqr.",
  # Valiasr Specifics
  "16",   "Valiasr St.",
  "18",   "Valiasr St.",
  "18",   "Tajrish Sqr.",
  "32",   "Valiasr St.",
  "86",   "Valiasr St.",
  "86",   "Tajrish Sqr.",
  # Toopkhaneh Specifics
  "4",    "Toopkhaneh Sqr.",
  "41",   "Toopkhaneh Sqr.",
  # Hassan Abad
  "46",   "Hassan Abad Sqr.",
  # Vanak
  "1",    "Vanak Sqr.", # Based on 'Neighborhood Hubs' logic
  "86",   "Vanak Sqr.",
  # Topics that go to multiple tweets
  "0",    "Valiasr St.",
  "0",    "Tehran Urban Spaces",
  "31",   "City Theater",
  "31",   "Tajrish Sqr.",
  "31",   "Toopkhaneh Sqr.",
  "31",   "Hassan Abad Sqr.",
```

```
  "30",   "City Theater",
  "30",   "Toopkhaneh Sqr.",
  "30",   "Vanak Sqr.",
  "52",   "Toopkhaneh Sqr.",
  "52",   "Hassan Abad Sqr.",
  "55",   "Toopkhaneh Sqr.",
  "55",   "Hassan Abad Sqr."
)


# Load the data
cleaned_tweets <-
read.csv("output/Py_results/mostfinal_analysis_output.csv")
topics <- read.csv("output/Py_results/plot_topics_english.csv")


# Convert the topic names into string(text) format directly
topic_labels <- topics|>
  select(Topic, eng_name, explicitness)|>
  mutate(Topic = as.character(Topic))|>
  unique()


# Join the place names to topic names
place_links <- topic_labels|>
  left_join(topic_to_place, by = "Topic")|>
  mutate(place_name = case_when(!is.na(place_name) ~ place_name,
                                TRUE ~ "Tehran Urban Spaces"))
probs_df <- cleaned_tweets|>
  select(all_of(spatial_columns), Topic, explicitness, timeline)


### Place-Topic matrix for localised topics
## Pivot probs for the overall structure
probs_long <- probs_df|>
```

```
  pivot_longer(starts_with("prob_topic_"),
               names_to = "Topic_ID",
               values_to = "score") |>
  mutate(Topic_ID = gsub("prob_topic_", "", Topic_ID))


# Creating the spatial matrix with explicit(localised) data for
# both timelines
spatial_matrix_data24 <- probs_long|>
  inner_join(place_links, by = c("Topic_ID" = "Topic"))|>
  group_by(place_name, eng_name, timeline)|>
  summarise(mean_prob = mean(score, na.rm = FALSE), .groups =
"drop")|>
  filter(timeline == "Before_24")


spatial_matrix_data26 <- probs_long|>
  inner_join(place_links, by = c("Topic_ID" = "Topic"))|>
  group_by(place_name, eng_name, timeline)|>
  summarise(mean_prob = mean(score, na.rm = FALSE), .groups =
"drop")|>
  filter(timeline == "After_24")


# Build the Spatial DataFrame using an inner_join
# An inner_join will automatically drop any Topic_ID that isn't in
your explicit map
spatial_matrix_explicit_data <- probs_long |>
  inner_join(place_links_explicit, by = c("Topic_ID" = "Topic"))
|>
  # Summarize
  group_by(place_name, eng_name, timeline) |>
  summarise(mean_prob = mean(score, na.rm = FALSE), .groups =
"drop")


spatial_matrix_explicit_data24 <- probs_long |>
```

```
  inner_join(place_links_explicit, by = c("Topic_ID" = "Topic"))
|>

  group_by(place_name, eng_name, timeline) |>

  summarise(mean_prob = mean(score, na.rm = FALSE), .groups =
"drop")|>

  filter(timeline == "Before_24")


spatial_matrix_explicit_data26 <- probs_long |>

  inner_join(place_links_explicit, by = c("Topic_ID" = "Topic"))
|>

  group_by(place_name, eng_name, timeline) |>

  summarise(mean_prob = mean(score, na.rm = FALSE), .groups =
"drop")|>

  filter(timeline == "After_24")


# Create cooccurrence matrix plots

# General discourse matrix

spatial_matrix_data|>

  ggplot(aes(x = eng_name, y = place_name)) +

  geom_point(aes(size = mean_prob,

                 alpha = eng_name %in% c("Public Monuments",

                                         "Tehran / Valiasr Street
Life & Nostalgia",

                                         "Seasonal Walks (Autumn
in Valiasr)",

                                         # "City Theater Metro &
Underpass",

                                         "Valiasr Street
Connectivity & Length",

                                         # "Urban Management &
Heritage Protection of City Theater",

                                         "Valiasr / Tajrish Tree
Preservation & Urban Ecology",

                                         "Metro Station
Entrances",
```

```
                                          "Tajrish-Valiasr Walks",
                                          "Performing Arts &
Festivals (City Theater)",
                                          "Historical Renaming of
Squares",
                                          "Historical Squares
(Ashura)",
                                          "Metro Infrastructure &
Tajrish Station"),
                  show.legend = TRUE),
             color = "#faa819") +
  ## faa819 orange / ffde17 yellow / 2d2e8b blue / 662d91 purple
  ## 5556a5 light blue / 8c5d25 brown / d9534f red / 999999 grey /
5cb85c green
  scale_alpha_manual(values = c("TRUE" = 1.0, "FALSE" = 0.25),
                     guide = "none") +
  scale_size_continuous(trans = "exp", range = c(0, 15),
                        name = "Probability Weight", ) +
  theme_minimal() +
  theme(
        ## panel & background
        panel.background = element_rect(fill = NA, color = NA),
        plot.background  = element_rect(fill = "gray20"),
        panel.grid.major  = element_line(color = "gray50"),
        panel.grid.minor  = element_line(color = "gray60"),
        ## legend configurations
        legend.background = element_rect(fill = NA, color = NA),
        legend.key = element_rect(fill = NA, color = NA), # Box
behind the dots
        legend.text = element_text(color = "white"),
        legend.title = element_text(color = "white"),
        legend.key.size = unit(0.05, "cm"),
        ## axis
```

```
        axis.text = element_text(color = "white"),
        axis.text.x = element_text(size = 10, angle = 45, hjust =
1),
        axis.text.y = element_text(size = 11, hjust = 1.1, color =
"white"),
        axis.title = element_text(color = "white", face = "bold"),
        ## title and subtitle
        plot.title = element_text(color = "white"),
        plot.subtitle = element_text(color = "white"),
        ) +
  labs(title = "General Spatial Discourse Between 2022-2026",
       x = "Spatial Topics",
       y = "Key Urban Spaces"
       )
# Making the overlaying shifts for localised and general data in
# dark theme. You can switch the data in the first item in the
# bracket
ggplot(spatial_matrix_explicit_data,
aes(x = eng_name, y = place_name, color = timeline)) +
  geom_point(aes(size = mean_prob),
             alpha = 1,
             position = position_dodge(width = 0.0)) +
  # Making the differences look much more dramatic
  # scale_size_area(max_size = 15, name = "Probability Weight") +
  scale_size_continuous(trans = "exp", range = c(0, 20),
                        name = "Probability Weight", ) + # Giving
exponential size
  scale_color_manual(values = c("After_24" = "#ffde17",
"Before_24" = "#662d91"),
                     ## faa819 orange / ffde17 yellow / 2d2e8b
blue / 662d91 purple
                     ## 5556a5 light blue / 8c5d25 brown / d9534f
red / 999999 grey / 5cb85c green
                     labels = c("2024-2026","2022-2024")) +
```

```
  theme_minimal() +
  theme(
        ## panel & background
        panel.background = element_rect(fill = NA, color = NA),
        plot.background  = element_rect(fill = "gray20"),
        panel.grid.major  = element_line(color = "gray50"),
        panel.grid.minor  = element_line(color = "gray60"),
        ## legend configurations
        legend.background = element_rect(fill = NA, color = NA),
        legend.key = element_rect(fill = NA, color = NA), # Box
behind the dots
        legend.text = element_text(color = "white"),
        legend.title = element_text(color = "white"),
        ## axis
        axis.text = element_text(color = "white"),
        axis.text.x = element_text(size = 9, angle = 45, hjust =
1, color = "white"),
        axis.text.y = element_text(size = 11, hjust = 1.1, color =
"white"),
        axis.title = element_text(color = "white", face = "bold"),
        ## title and subtitle
        plot.title = element_text(color = "white"),
        plot.subtitle = element_text(color = "white"),


        # legend.box.spacing = unit(0.1, "cm"),
        legend.margin = margin(r = 15, l = 10)) + # adjusting
distances in legend box space
  guides(
    size = guide_legend(override.aes = list(color = "gray70")),
    color = guide_legend(override.aes = list(size = 5))
    ) +
  labs(title = "Localised Discourse Shifts in Probability
Weights",
```

```
       subtitle = "Between 2022-2026",
       x = "Spatial Topics",
       y = "Key Urban Spaces",
       color = "Timeline")
```

*Apendix 8. Coocurrence matrices*

**Sentiment shifts per topic between 2022-2026**

```
# For the average sentiment score and localised topic shifts
# matrix
spatial_tweet_local <- cleaned_tweets|>
  filter(explicitness == "Explicit")|>
  # Group only by topic and timeline to get the true average
  # sentiment
  group_by(eng_name, sentiment_label, timeline) |>
  summarise(
    topic_count = n(),
    avg_score = mean(sentiment_score, na.rm = TRUE),
  # True overall sentiment score
    .groups = "drop"
  ) |>
  group_by(timeline)|>
  mutate(
    timeline_total = sum(topic_count),
    topic_share = topic_count/timeline_total,
    normalized_score = avg_score * topic_share
  )|>
  ungroup()|>
  group_by(eng_name, sentiment_label)|>
  # To check if the avg scores are completely equal in both
timelines
  mutate(
    is_equal =
      max(normalized_score) - min(normalized_score) < 0.01)|>
```

```
  ungroup()


# Plot the shift matrix in dark theme
ggplot(spatial_tweet_local,
       aes(x = eng_name, y = sentiment_label, color = timeline)) +
  geom_point(aes(size = normalized_score, alpha = is_equal),
             position = position_dodge(width = 0.0)) +
  scale_alpha_manual(values = c("FALSE" = 1.0, "TRUE" = 0.2),
                     guide = "none") +
  scale_x_discrete(
    name = "Sentiment Labels", # Setting the Axis Title
    labels = c(
      "negative" = "Negative",
      "neutral"  = "Neutral",
      "positive" = "Positive"
    )
  ) +
  # Making the differences look much more dramatic
  scale_size_continuous(
    # range = c(2, 15), name = "Mean Sentiment Score"
    trans = "exp", range = c(0, 20),
                        name = "Weighted Mean Sentiment Score",
    ) + # Giving exponential size
  scale_color_manual(values = c("Before_24" = "#2d2e8b",
"After_24" = "#faa819"),
                     ## faa819 orange / ffde17 yellow / 2d2e8b
blue / 662d91 purple
                     ## 5556a5 light blue / 8c5d25 brown / d9534f
red / 999999 grey / 5cb85c green
                     labels = c("2024-2026", "2022-2024")) +
  # scale_fill_manual(values = NA) +
  theme_minimal() +
```

```
  theme(
    ## panel & background
    panel.background = element_rect(fill = NA, color = NA),
    plot.background  = element_rect(fill = "gray20"),
    panel.grid.major  = element_line(color = "gray50"),
    panel.grid.minor  = element_line(color = "gray60"),
    ## legend configurations
    legend.background = element_rect(fill = NA, color = NA),
    legend.key = element_rect(fill = NA, color = NA), # Box behind
the dots
    legend.text = element_text(color = "white"),
    legend.title = element_text(color = "white"),
    ## axis
    axis.text = element_text(color = "white"),
    axis.text.x = element_text(size = 9, angle = 45, hjust = 1,
                               face = "bold", color = "white"),
    axis.text.y = element_text(size = 11, hjust = 1.0, color =
"white"),
    axis.title = element_text(color = "white", face = "bold"),
    ## title and subtitle
    plot.title = element_text(color = "white"),
    plot.subtitle = element_text(color = "white"),


    # legend.box.spacing = unit(0.1, "cm"),
    legend.margin = margin(r = 15, l = 10)) + # adjusting
distances in legend box space
  guides(
    size = guide_legend(override.aes = list(color = "gray70")),
    color = guide_legend(override.aes = list(size = 5))
  ) +
  labs(title = "Localised Discourse Shifts in Weighted Average
Sentiment Scores",
```

```
       subtitle = "Between 2022-2026",
       x = "Spatial Topics",
       y = "Sentiment Labels",
       color = "Timeline")
```

**Over / under representation**

```
# Under/Over Representation
# Create the Global Baseline (The "Expected" frequency)
topic_baseline <- cleaned_tweets |>
  filter(explicitness == "Explicit") |>
  group_by(eng_name) |>
  summarise(global_count = n(), .groups = "drop") |>
  # Calculate the percentage of each topic relative to the WHOLE
city
  mutate(
    global_total = sum(global_count),
    global_pct = (global_count / global_total) * 100) |>
  select(eng_name, global_pct) # We only need these two for the
comparison


# Save the baseline and representation layers
write.csv(topic_baseline,
"output/Py_results/topic_baseline_reference.csv")
write.csv(representation_analysis,
"output/Py_results/representation_analysis.csv")
# Compare against the City Theater (The "Actual" frequency)
representation_analysis <- cleaned_tweets |>
  filter(explicitness == "Explicit" & type == "ct") |> # Target
the City Theater
  group_by(eng_name) |>
  summarise(ct_count = n(), .groups = "drop") |>
  mutate(
    ct_total = sum(ct_count),
```

```
    ct_pct = (ct_count / ct_total) * 100) |>

  # Join the baseline to calculate over/under representation

  left_join(topic_baseline, by = "eng_name") |>

  mutate(

    # The Indicator: Positive = Over-represented, Negative =
Under-represented

    representation_index = ct_pct - global_pct,

    label_text = paste0(round(representation_index, 1), "%")

    ) |>

  mutate(rep_label = ifelse(representation_index > 0, "Over-
represented",

                            "Under-represented"))|>

  mutate(eng_name = reorder(eng_name, representation_index))


# Plot overall representation between '22-26

ggplot(representation_analysis, aes(x = eng_name, y =
representation_index,

                                    fill = representation_index >
0)) +

  # Map alpha to the logic (TRUE/FALSE)

  geom_col(aes(alpha = eng_name %in% c(

    "City Theater Metro & Underpass",

    "Tehran / Valiasr Street Life & Nostalgia",

    "Urban Management & Heritage Protection of City Theater",

    "Performing Arts & Festivals (City Theater)")),

    show.legend = TRUE) + # This stays TRUE so the 'fill' legend
can appear


  geom_text(aes(label = label_text),

            hjust =
ifelse(representation_analysis$representation_index > 0, -0.1,
1.1),

            size = 4, color = "white") +
```

```
  coord_flip() +


  # highlight the baseline
  geom_hline(yintercept = 0, color = "white", linewidth = 0.8) +


  # Manually define the fill legend
  scale_fill_manual(values = c("TRUE" = "#662d91", "FALSE" =
"#ffde17"),
                    ## faa819 orange / ffde17 yellow / 2d2e8b blue
/ 662d91 purple
                    ## 5556a5 light blue / 8c5d25 brown / d9534f
red / 999999 grey / 5cb85c green
                    labels = c("Under-represented", "Over-
represented"),
                    name = "Representation Status") +
  # Manually define alpha and remove its legend specifically
  scale_alpha_manual(values = c("TRUE" = 1.0, "FALSE" = 0.25),
                     guide = "none") +
  labs(title = "Topic Representation: City Theater vs. Tehran
Average",
       y = "Percentage Point Deviation from Baseline",
       x = "Spatial Topics") +
  theme(# Change the size of the title (e.g., "Representation
Status")
    legend.title = element_text(size = 12, face = "bold", color =
"white"),
    # Change the size of the labels (e.g., "Over-represented")
    legend.text = element_text(size = 11, color = "white"),
    legend.background = element_blank(),
    strip.text = element_text(size = 11, face = "bold", color =
"white"),
    strip.background = element_blank(),
    ## panel & background
    panel.background = element_rect(fill = NA, color = NA),
```

```
    plot.background  = element_rect(fill = "gray20", color = NA),
    panel.grid.major  = element_line(color = "gray50"),
    panel.grid.minor  = element_line(color = "gray40"),
    ## axis
    axis.text = element_text(color = "white"),
    axis.text.y = element_text(size = 11, hjust = 1, color =
"white"),
    axis.text.x = element_text(hjust = 0.9, color = "white"),
    axis.title = element_text(color = "white", face = "bold"),
    ## title and subtitle
    plot.title = element_text(color = "white"),
    plot.subtitle = element_text(color = "white")
    )


# Representation for 24
representation_analysis24 <- cleaned_tweets |>
  filter(explicitness == "Explicit" & type == "ct" & timeline ==
"Before_24") |>
  group_by(eng_name) |>
  summarise(ct_count = n(), .groups = "drop") |>
  mutate(
    ct_total = sum(ct_count),
    ct_pct = (ct_count / ct_total) * 100) |>
  # 3. JOIN the baseline to calculate over/under representation
  left_join(topic_baseline24, by = "eng_name") |>
  mutate(
    # The Indicator: Positive = Over-represented, Negative =
Under-represented
    representation_index = ct_pct - global_pct,
    label_text = paste0(round(representation_index, 1), "%")
  ) |>
```

```
  mutate(rep_label = ifelse(representation_index > 0, "Over-
represented",

                            "Under-represented"))|>

  mutate(eng_name = reorder(eng_name, representation_index))




# Plot representation for '24

ggplot(representation_analysis24, aes(x = eng_name, y =
representation_index,

                                      fill = representation_index
> 0)) +

  # Map alpha to the logic (TRUE/FALSE)

  geom_col(aes(alpha = eng_name %in% c(

    "City Theater Metro & Underpass",

    "Tehran / Valiasr Street Life & Nostalgia",

    "Urban Management & Heritage Protection of City Theater",

    "Public Monuments",

    "Performing Arts & Festivals (City Theater)")),

    show.legend = TRUE) + # This stays TRUE so the 'fill' legend
can appear



  geom_text(aes(label = label_text),

            hjust =
ifelse(representation_analysis24$representation_index > 0, -0.1,
1.1),

            size = 4, color = "white") +

  coord_flip() +



  # Manually define the fill legend

  scale_fill_manual(values = c("TRUE" = "#662d91", "FALSE" =
"#ffde17"),

                    ## faa819 orange / ffde17 yellow / 2d2e8b blue
/ 662d91 purple
```

```
                    ## 5556a5 light blue / 8c5d25 brown / d9534f
red / 999999 grey / 5cb85c green
                    labels = c("Under-represented", "Over-
represented"),
                    name = "Representation Status") +


  # highlight the baseline
  geom_hline(yintercept = 0, color = "white", linewidth = 0.8) +


  # Manually define alpha and remove its legend specifically
  scale_alpha_manual(values = c("TRUE" = 1.0, "FALSE" = 0.25),
                     guide = "none") +


  labs(title = "Topic Representation: City Theater vs. Tehran
Average | 2022-2024",
       y = "Percentage Point Deviation from Baseline",
       x = "Spatial Topics") +
  theme_minimal() +
  theme(# Change the size of the title (e.g., "Representation
Status")
    legend.title = element_text(size = 12, face = "bold", color =
"white"),
    # Change the size of the labels (e.g., "Over-represented")
    legend.text = element_text(size = 11, color = "white"),
    legend.background = element_blank(),
    strip.text = element_text(size = 11, face = "bold", color =
"white"),
    strip.background = element_blank(),
    ## panel & background
    panel.background = element_rect(fill = NA, color = NA),
    plot.background  = element_rect(fill = "gray20"),
    panel.grid.major  = element_line(color = "gray50"),
    panel.grid.minor  = element_line(color = "gray40"),
```

```
    ## axis
    axis.text = element_text(color = "white"),
    axis.text.y = element_text(size = 11, hjust = 1, color =
"white"),
    axis.text.x = element_text(hjust = 0.9, color = "white"),
    axis.title = element_text(color = "white", face = "bold"),
    ## title and subtitle
    plot.title = element_text(color = "white"),
    plot.subtitle = element_text(color = "white")
    )



# Representation for 26
representation_analysis26 <- cleaned_tweets |>
  filter(explicitness == "Explicit" & type == "ct" & timeline ==
"After_24") |>
  group_by(eng_name) |>
  summarise(ct_count = n(), .groups = "drop") |>
  mutate(
    ct_total = sum(ct_count),
    ct_pct = (ct_count / ct_total) * 100) |>
  # Join the baseline to calculate over/under representation
  left_join(topic_baseline26, by = "eng_name") |>
  mutate(
    # The Indicator: Positive = Over-represented, Negative =
Under-represented
    representation_index = ct_pct - global_pct,
    label_text = paste0(round(representation_index, 1), "%")
  ) |>
  mutate(rep_label = ifelse(representation_index > 0, "Over-
represented",
                            "Under-represented"))|>
```

```
  mutate(eng_name = reorder(eng_name, representation_index))


# Plot the representation for ‘26

ggplot(representation_analysis26, aes(x = eng_name, y =
representation_index,

                                      fill = representation_index >
0)) +

  # Map alpha to the logic (TRUE/FALSE)

  geom_col(aes(alpha = eng_name %in% c(

    "City Theater Metro & Underpass",

    "Tehran / Valiasr Street Life & Nostalgia",

    "Urban Management & Heritage Protection of City Theater",

    "Public Monuments",

    "Performing Arts & Festivals (City Theater)")),

    show.legend = TRUE) + # This stays TRUE so the 'fill' legend
can appear


  geom_text(aes(label = label_text),

            hjust =
ifelse(representation_analysis26$representation_index > 0, -0.1,
1.1),

            size = 4, color = "white") +

  coord_flip() +


  # Manually define the fill legend

  scale_fill_manual(values = c("TRUE" = "#662d91", "FALSE" =
"#ffde17"),

                    ## faa819 orange / ffde17 yellow / 2d2e8b blue
/ 662d91 purple

                    ## 5556a5 light blue / 8c5d25 brown / d9534f
red / 999999 grey / 5cb85c green

                    labels = c("Under-represented", "Over-
represented"),

                    name = "Representation Status") +
```

```
  # Manually define alpha and remove its legend specifically
  scale_alpha_manual(values = c("TRUE" = 1.0, "FALSE" = 0.25),
                     guide = "none") +
  # Highlight the baseline
  geom_hline(yintercept = 0, color = "white", linewidth = 0.8) +


  labs(title = "Topic Representation: City Theater vs. Tehran
Average | 2024-2026",
       y = "Percentage Point Deviation from Baseline",
       x = "Spatial Topics") +
  theme_minimal() +
  theme(# Change the size of the title (e.g., "Representation
Status")
    legend.title = element_text(size = 12, face = "bold", color =
"white"),
    # Change the size of the labels (e.g., "Over-represented")
    legend.text = element_text(size = 11, color = "white"),
    legend.background = element_blank(),
    strip.text = element_text(size = 11, face = "bold", color =
"white"),
    strip.background = element_blank(),
    ## panel & background
    panel.background = element_rect(fill = NA, color = NA),
    plot.background  = element_rect(fill = "gray20"),
    panel.grid.major  = element_line(color = "gray50"),
    panel.grid.minor  = element_line(color = "gray40"),
    ## axis
    axis.text = element_text(color = "white"),
    axis.text.y = element_text(size = 11, hjust = 1, color =
"white"),
    axis.text.x = element_text(hjust = 0.9, color = "white"),
    axis.title = element_text(color = "white", face = "bold"),
```

```
    ## title and subtitle
    plot.title = element_text(color = "white"),
    plot.subtitle = element_text(color = "white")
  )

# Overall representation with Location Quotient (LQ)
representation_analysis_lq <- cleaned_tweets |>
  filter(explicitness == "Explicit" & type == "ct") |>
  group_by(eng_name) |>
  summarise(ct_count = n(), .groups = "drop") |>
  mutate(
    ct_total = sum(ct_count),
    ct_pct = (ct_count / ct_total) * 100) |>
  # Join the baseline to calculate over/under representation
  left_join(topic_baseline26, by = "eng_name") |>
  mutate(
    # The Indicator: Positive = Over-represented, Negative =
Under-represented
    representation_index = ct_pct - global_pct,
    label_text = paste0(round(representation_index, 1), "%")
  ) |>
  mutate(lq = ct_pct / global_pct,
         label_lq = paste0(round(lq, 2)))|>
  mutate(rep_label = ifelse(representation_index > 0, "Over-
represented",
                            "Under-represented"))|>
  mutate(lq_label = ifelse(lq > 1, "Specialized", "Not-
specialized"))|>
  mutate(eng_name = reorder(eng_name, representation_index))

write.csv(representation_analysis_lq,
"output/Py_results/under_over_lq.csv")
```

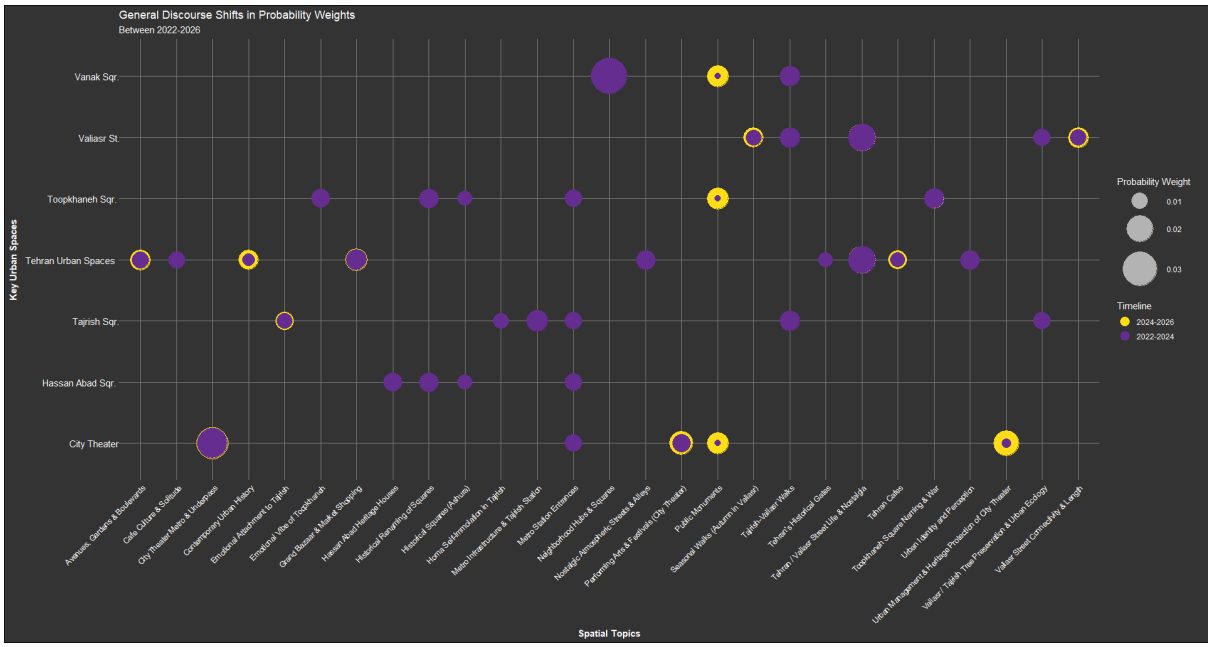
```
# Plot the LQ
ggplot(representation_analysis_lq, aes(x = eng_name, y = lq - 1,
                                       fill = lq > 1)) +
  # 1. Map alpha to the logic (TRUE/FALSE)
  geom_col(aes(alpha = eng_name %in% c(
    "City Theater Metro & Underpass",
    "Tehran / Valiasr Street Life & Nostalgia",
    "Urban Management & Heritage Protection of City Theater",
    "Public Monuments",
    "Tajrish-Valiasr Walks",
    "Metro Station Entrances",
    "Performing Arts & Festivals (City Theater)")),
    show.legend = TRUE) + # This stays TRUE so the 'fill' legend
can appear
  #Specify the baseline with value 1
  geom_hline(yintercept = 0, color = "white", linewidth = 0.8) +
  geom_text(aes(label = label_lq),
            hjust = ifelse(representation_analysis_lq$lq > 1, -
0.1, 1.1),
            size = 4, color = "white") +
  coord_flip() +
  # 2. MANUALLY DEFINE THE FILL LEGEND
  scale_fill_manual(values = c("TRUE" = "#662d91", "FALSE" =
"#ffde17"),
                    ## faa819 orange / ffde17 yellow / 2d2e8b blue
/ 662d91 purple
                    ## 5556a5 light blue / 8c5d25 brown / d9534f
red / 999999 grey / 5cb85c green
                    labels = c("Not-specialized", "Specialized"),
                    name = "LQ Status") +
  # shifting the lq values to their original number
  scale_y_continuous(labels = function(x) x + 1) +
```

```
  # 3. MANUALLY DEFINE ALPHA AND KILL ITS LEGEND SPECIFICALLY
  scale_alpha_manual(values = c("TRUE" = 1.0, "FALSE" = 0.25),
                     guide = "none") +
  labs(title = "Topic Representation: City Theater vs. Tehran
Average | 2022-2026",
       subtitle = "Based on Location Quotient value",
       y = "Percentage Point Deviation from Baseline",
       x = "Spatial Topics") +
  theme(# Change the size of the title (e.g., "Representation
Status")
    legend.title = element_text(size = 12, face = "bold", color =
"white"),
    # Change the size of the labels (e.g., "Over-represented")
    legend.text = element_text(size = 11, color = "white"),
    legend.background = element_blank(),
    strip.text = element_text(size = 11, face = "bold", color =
"white"),
    strip.background = element_blank(),
    ## panel & background
    panel.background = element_rect(fill = NA, color = NA),
    plot.background  = element_rect(fill = "gray20", color = NA),
    panel.grid.major  = element_line(color = "gray50"),
    panel.grid.minor  = element_line(color = "gray40"),
    ## axis
    axis.text = element_text(color = "white"),
    axis.text.y = element_text(size = 11, hjust = 1, color =
"white"),
    axis.text.x = element_text(hjust = 0.9, color = "white"),
    axis.title = element_text(color = "white", face = "bold"),
    ## title and subtitle
    plot.title = element_text(color = "white"),
    plot.subtitle = element_text(color = "white")
  )
```

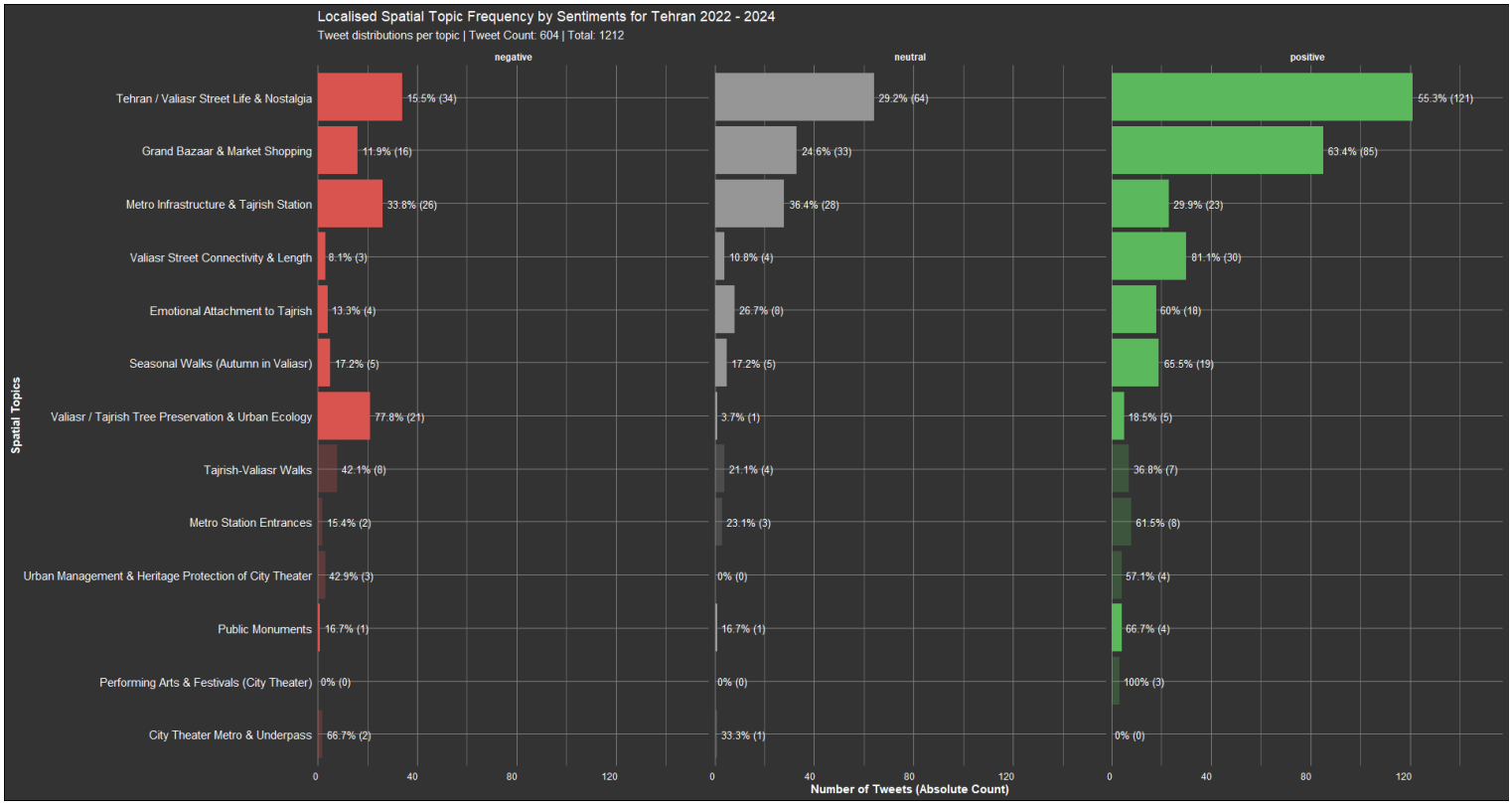
*Apendix 9. Over/under representation code block*

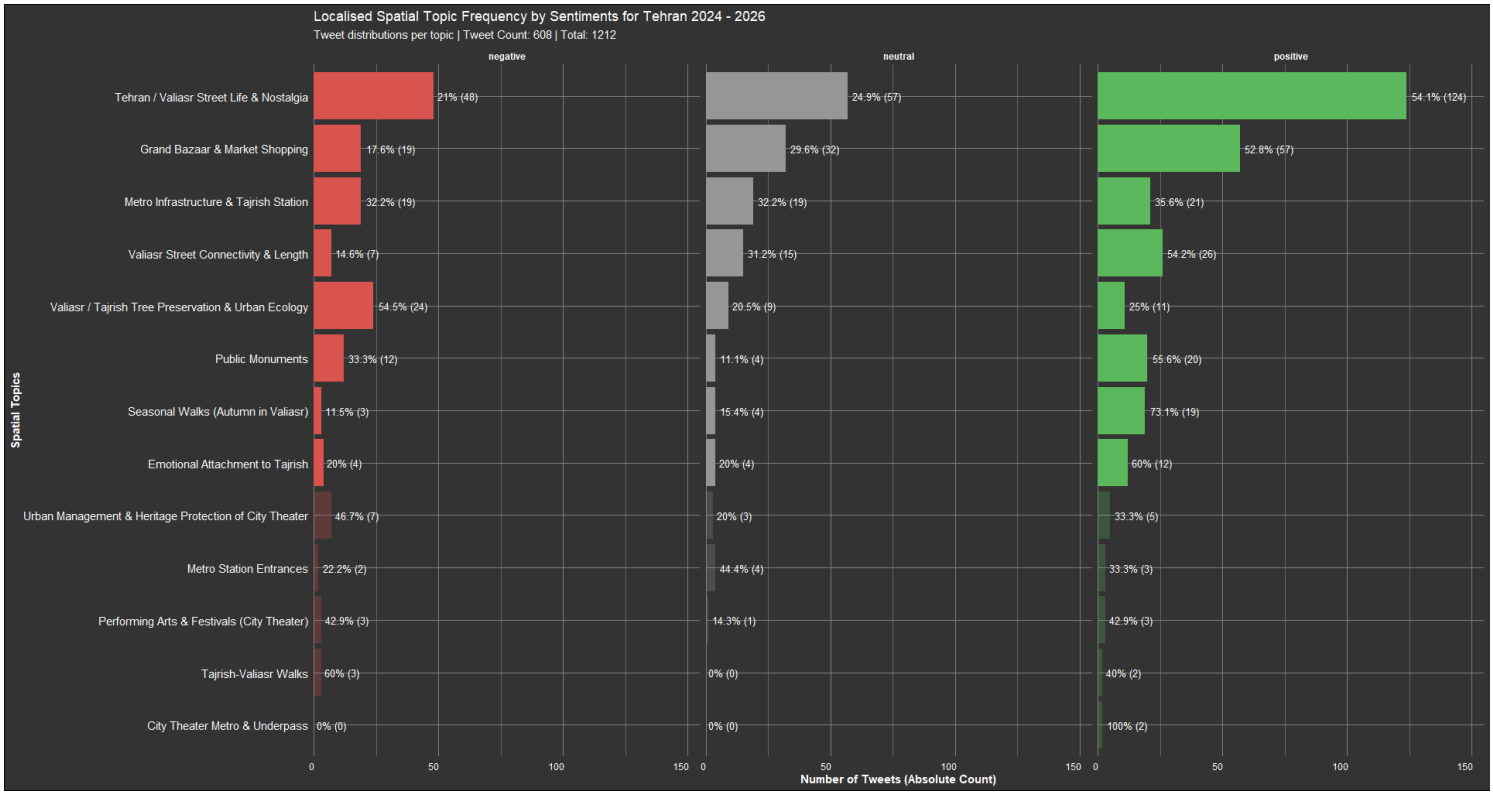
# Results

## Discourse shifts of Tehran




## Tehran scale 2022-2026

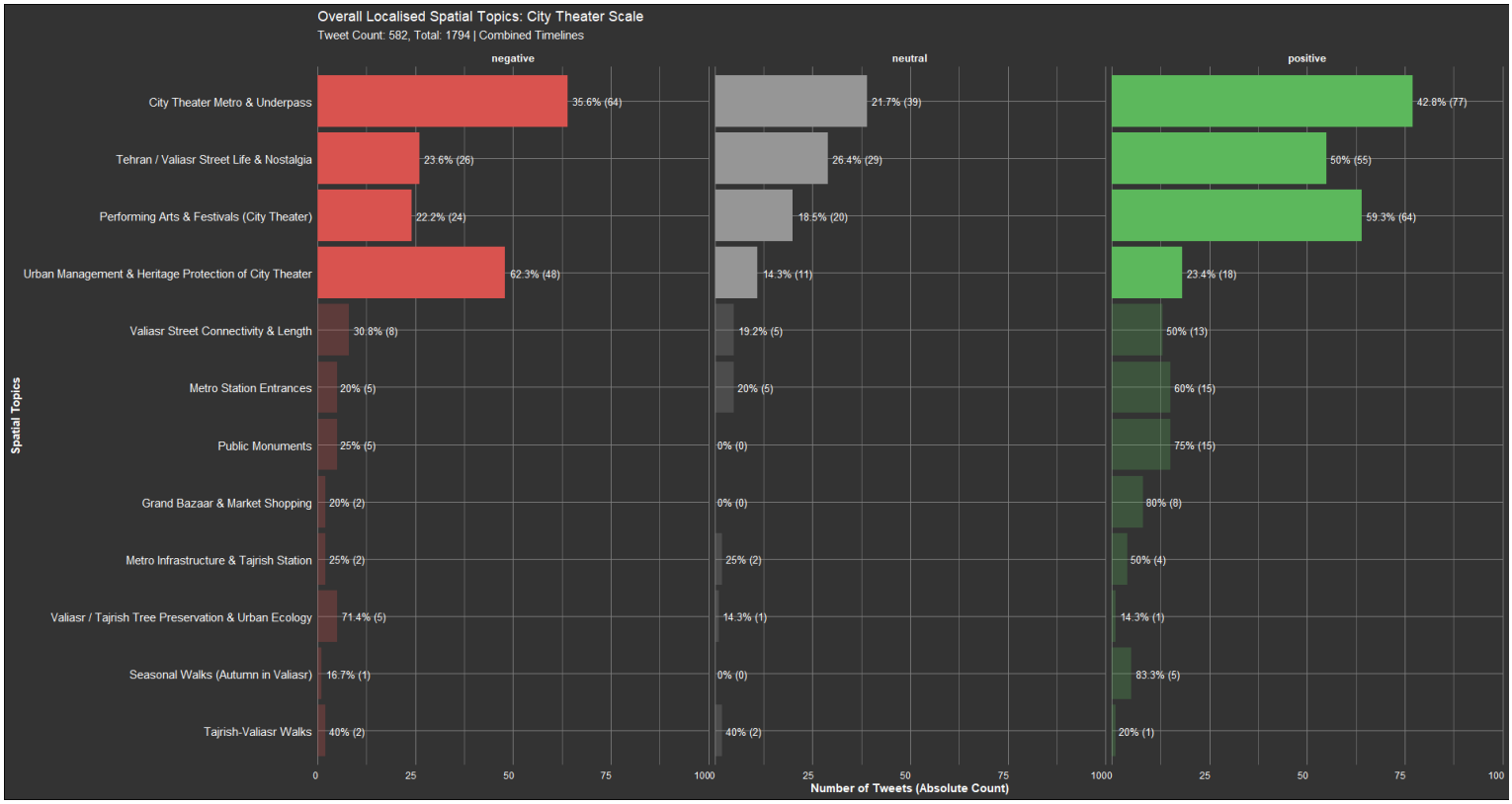


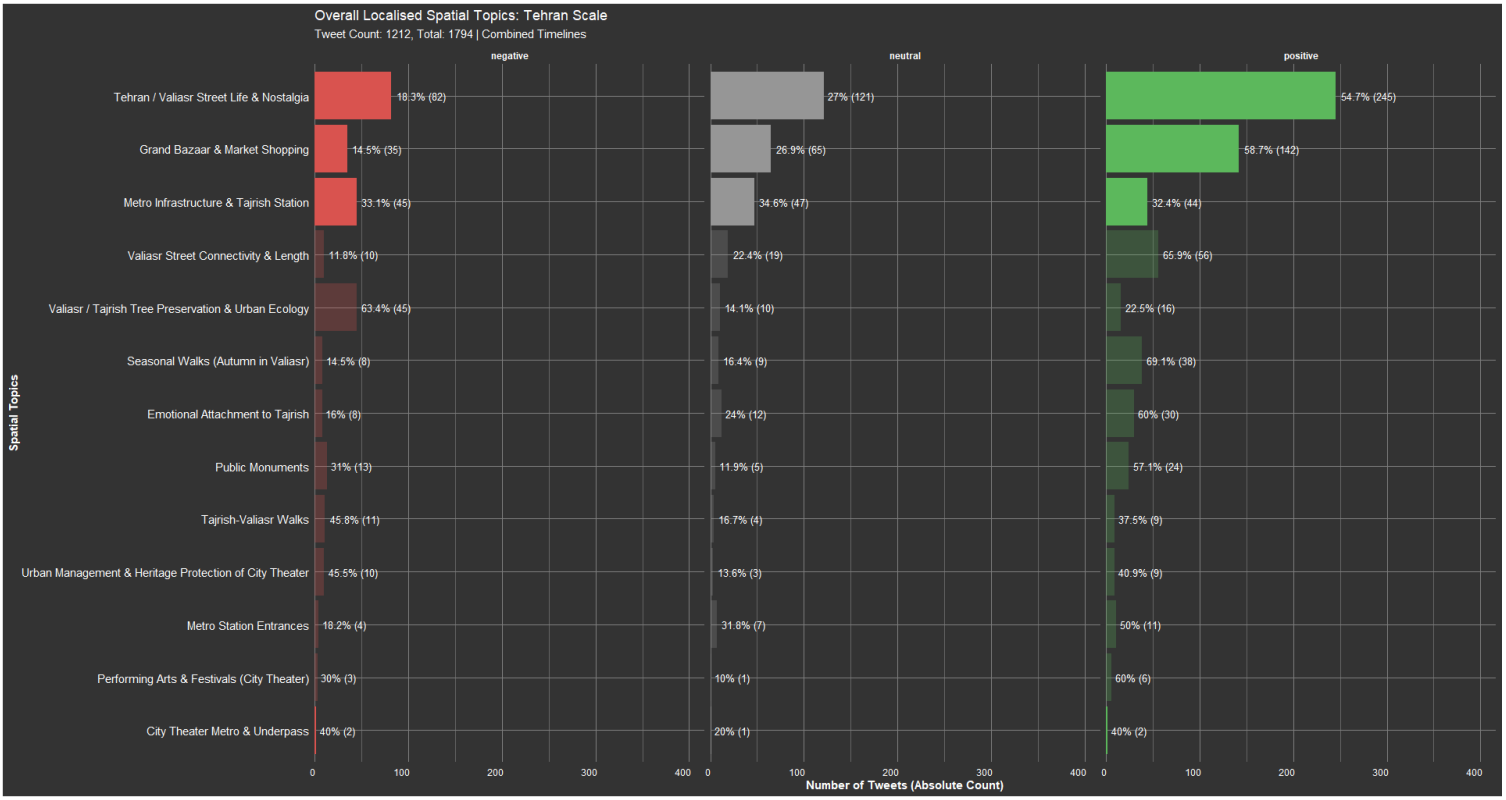



## Combined Timelines

Assessing the combined timelines is another way to review the results. This combination helps to gain a discrete classification of City Theater and non-city-theater scopes. The topics in the plot below significantly display how topics relevant to City Theater are fitted to the focused scale, while the other non-relevant topics are distributed on the larger scale.

I've found this way of visualization quite helpful for distinguishing how truly a topic is sentimentally valued on both large and small scales. The discrete classification also helps to understand how significant each topic is on its defined scale. For example, we can understand the topic dominance percentage of the city theater underpass among the large-scale topics. And it can be compared with the other scale. Moreover, in this way, it is easier to identify the topics that have overlaps in terms of sentimental scores.


Overall Localised Spatial Topics: City Theater Scale
Tweet Count: 582, Total: 1794 | Combined Timelines
negative
neutral
positive
Spatial Topics
City Theater Metro & Underpass
Tehran / Valiasr Street Life & Nostalgia
Performing Arts & Festivals (City Theater)
Urban Management & Heritage Protection of City Theater
Valiasr Street Connectivity & Length
Metro Station Entrances
Public Monuments
Grand Bazaar & Market Shopping
Metro Infrastructure & Tajrish Station
Valiasr / Tajrish Tree Preservation & Urban Ecology
Seasonal Walks (Autumn in Valiasr)
Tajrish-Valiasr Walks
35.6% (64)
23.6% (26)
22.2% (24)
62.3% (48)
30.8% (8)
20% (5)
25% (5)
20% (2)
25% (2)
71.4% (5)
16.7% (1)
40% (2)
21.7% (39)
26.4% (29)
18.5% (20)
14.3% (11)
19.2% (5)
20% (5)
0% (0)
0% (0)
25% (2)
14.3% (1)
0% (0)
40% (2)
42.8% (77)
50% (55)
59.3% (64)
23.4% (18)
50% (13)
60% (15)
75% (15)
80% (8)
50% (4)
14.3% (1)
83.3% (5)
20% (1)
0
25
50
75
100
Number of Tweets (Absolute Count)


Overall Localised Spatial Topics: Tehran Scale
Tweet Count: 1212, Total: 1794 | Combined Timelines
negative
neutral
positive
Spatial Topics
Tehran / Valiasr Street Life & Nostalgia
Grand Bazaar & Market Shopping
Metro Infrastructure & Tajrish Station
Valiasr Street Connectivity & Length
Valiasr / Tajrish Tree Preservation & Urban Ecology
Seasonal Walks (Autumn in Valiasr)
Emotional Attachment to Tajrish
Public Monuments
Tajrish-Valiasr Walks
Urban Management & Heritage Protection of City Theater
Metro Station Entrances
Performing Arts & Festivals (City Theater)
City Theater Metro & Underpass
18.3% (82)
14.5% (35)
33.1% (45)
11.8% (10)
63.4% (45)
14.5% (8)
16% (8)
31% (13)
45.8% (11)
45.5% (10)
18.2% (4)
30% (3)
40% (2)
27% (121)
26.9% (65)
34.6% (47)
22.4% (19)
14.1% (10)
16.4% (9)
24% (12)
11.9% (5)
16.7% (4)
13.6% (3)
31.8% (7)
10% (1)
20% (1)
54.7% (245)
58.7% (142)
32.4% (44)
65.9% (56)
22.5% (16)
69.1% (38)
60% (30)
57.1% (24)
37.5% (9)
40.9% (9)
50% (11)
60% (6)
40% (2)
0
100
200
300
400
Number of Tweets (Absolute Count)

## Over/Under representation based on volume difference

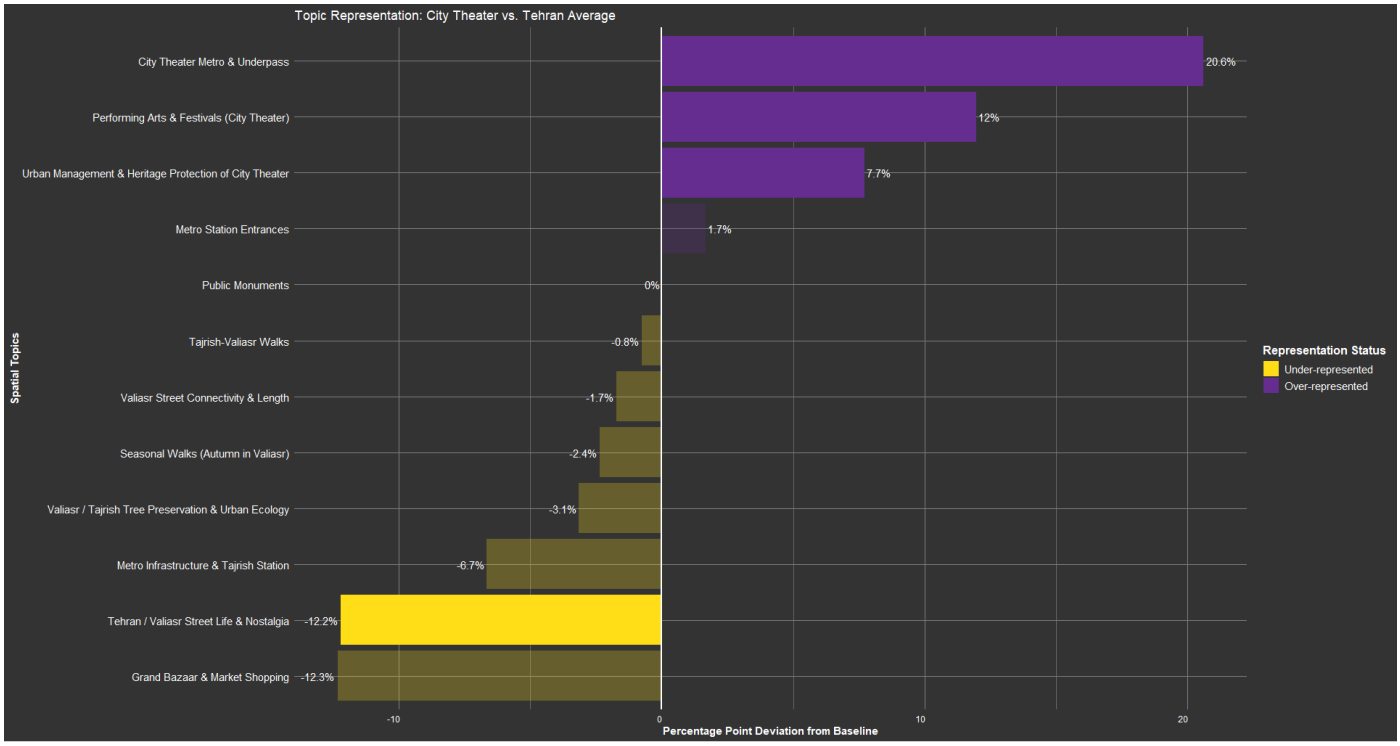
The representation analysis is based on the difference between the City Theater (CT) and Global Baseline (Reference) volume percentages:

$$\textbf{\textit{Representation Index}} = (ct_count / ct_total) * 100 - (global_count / global_total) * 100$$

The abovementioned equation was applied to the topics belonging to the City Theater scale. Since there was only one topic not belonging to the specified scale, “Emotional Attachment to Tajrish”, the values of both scales are integrated in the following table. At last, the outputs were visualized in a plot.

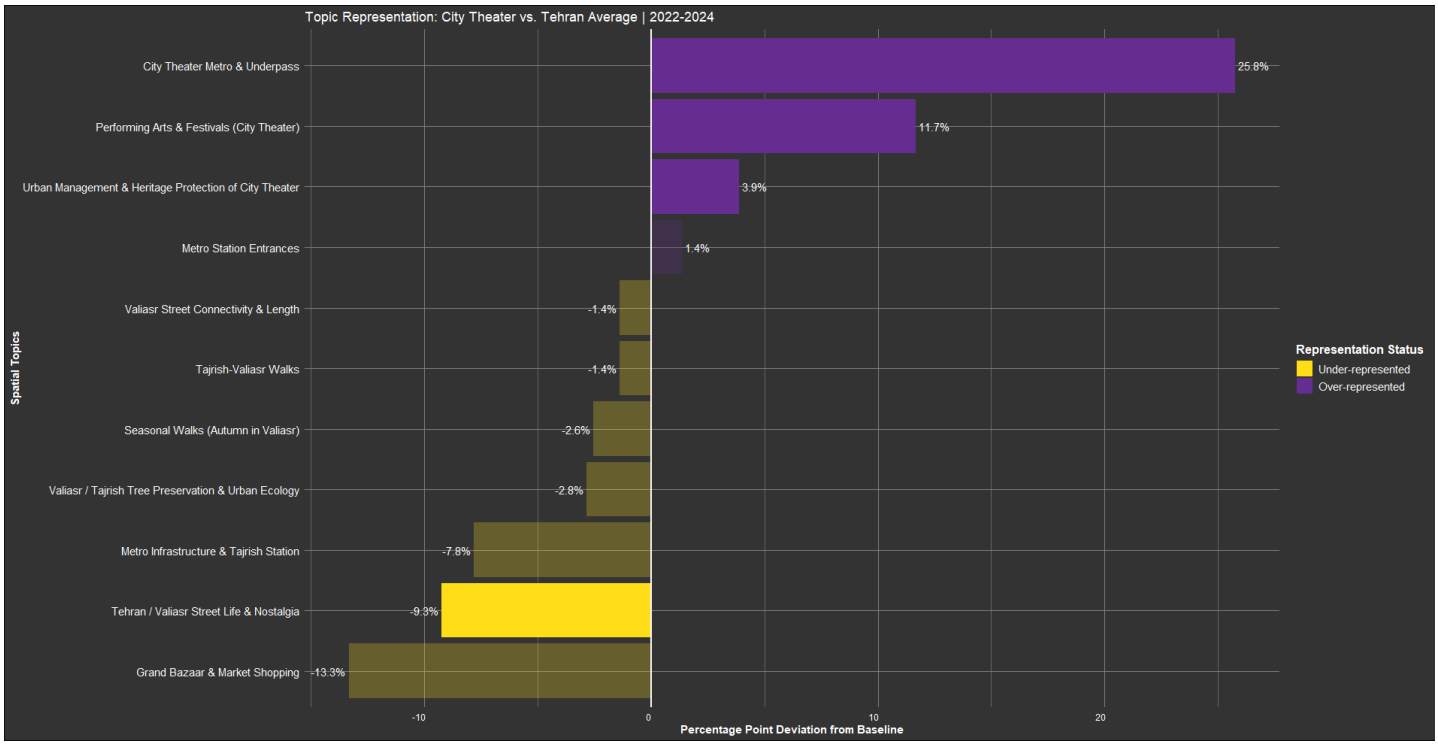
*Table X. Representation Analysis Values*

| No. | Topic Name | CT Count | CT pct (%) | Global pct (%) | Representation index (%) |
|---|---|---|---|---|---|
| 1 | City Theater Metro & Underpass | 180 | 30.92784 | 10.31215 | 20.61568 |
| 2 | Performing Arts & Festivals (City Theater) | 108 | 18.5567 | 6.57748 | 11.97922 |
| 3 | Urban Management & Heritage Protection of City Theater | 77 | 13.23024 | 5.518395 | 7.711846 |
| 4 | Metro Station Entrances | 25 | 4.295533 | 2.619844 | 1.675689 |
| 5 | Metro Infrastructure & Tajrish Station | 8 | 1.37457 | 8.026756 | -6.65219 |
| 6 | Public Monuments | 20 | 3.436426 | 3.455964 | -0.01954 |
| 7 | Tajrish-Valiasr Walks | 5 | 0.859107 | 1.616499 | -0.75739 |
| 8 | Valiasr Street Connectivity & Length | 26 | 4.467354 | 6.187291 | -1.71994 |
| 9 | Seasonal Walks (Autumn in Valiasr) | 6 | 1.030928 | 3.400223 | -2.3693 |
| 10 | Valiasr / Tajrish Tree Preservation & Urban Ecology | 7 | 1.202749 | 4.347826 | -3.14508 |
| 11 | Tehran / Valiasr Street Life & Nostalgia | 110 | 18.90034 | 31.10368 | -12.2033 |
| 12 | Grand Bazaar & Market Shopping | 10 | 1.718213 | 14.04682 | -12.3286 |
| 13 | Emotional Attachment to Tajrish | NA | NA | 2.787068 | NA |
| **Total** | | 582 | 100.0 | 100.0 (excl. topic 13 ~ 97.2) | |

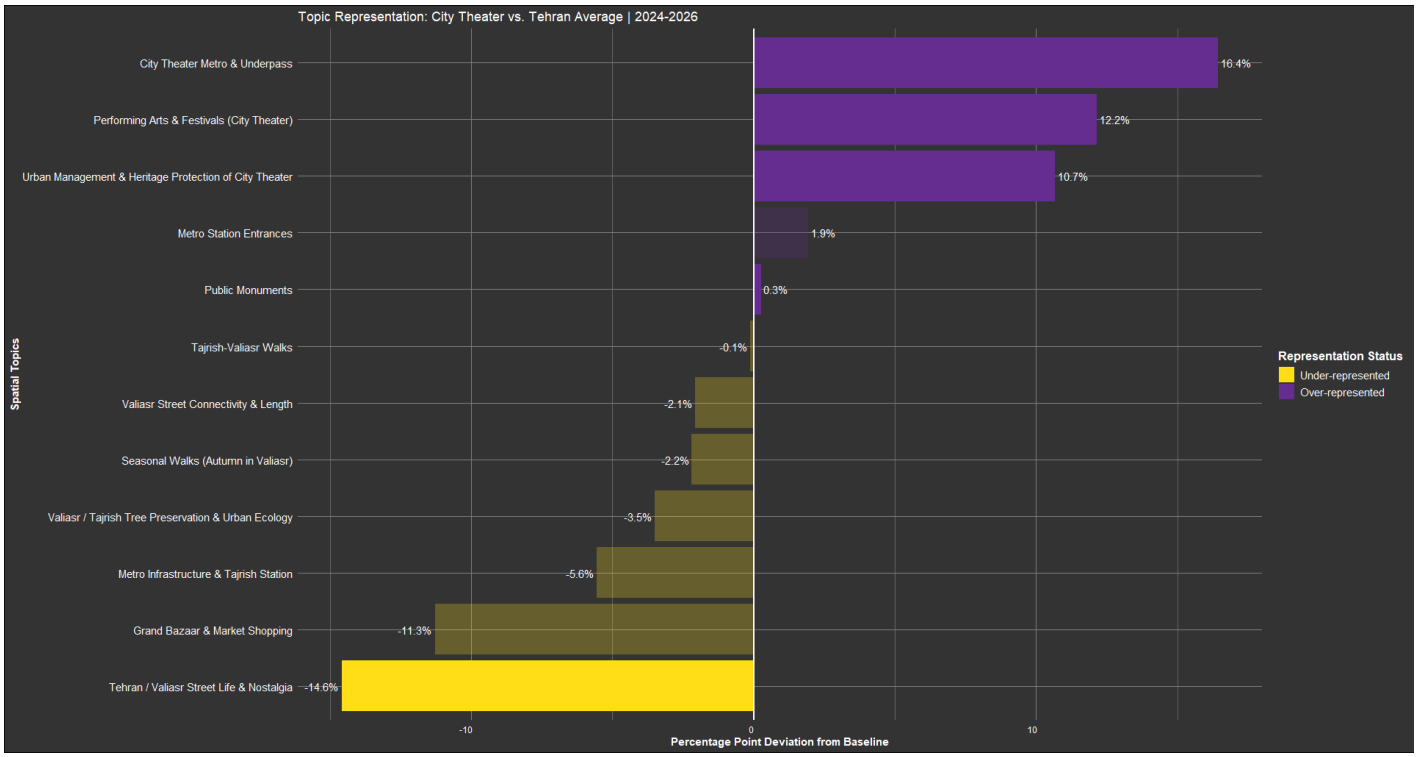

Topic Representation: City Theater vs. Tehran Average
City Theater Metro & Underpass
Performing Arts & Festivals (City Theater)
Urban Management & Heritage Protection of City Theater
Metro Station Entrances
Public Monuments
Tajrish-Valiasr Walks
Valiasr Street Connectivity & Length
Seasonal Walks (Autumn in Valiasr)
Valiasr / Tajrish Tree Preservation & Urban Ecology
Metro Infrastructure & Tajrish Station
Tehran / Valiasr Street Life & Nostalgia
Grand Bazaar & Market Shopping
20.6%
12%
7.7%
1.7%
0%
-0.8%
-1.7%
-2.4%
-3.1%
-6.7%
-12.2%
-12.3%
Spatial Topics
-10
0
10
20
Percentage Point Deviation from Baseline
Representation Status
Under-represented
Over-represented

# Representations between 22 and 26

## Before 2024




After 2024, it is interesting to see that the public monument shifted to being overrepresented, for it was not part of the City Theater discourse before 2024. The City Theater joined the topic group after the intervention.



# DATA MANAGEMENT CHECKLIST

**Instruction**

This checklist is relevant for all graduation projects of the Master AUBS. The form is intended to highlight common aspects of graduation projects that require particular attention with regard to planning the research and data management. Relevant information and supplementary sources regarding each question are provided below each question.

With this checklist, the faculty wants to avoid that students unexpectedly find themselves in complex and stressful situations, in which ethical or privacy matters and/or other laws and regulations become an issue. In projects involving humans, certain types of data processing increase the risks to the human participants: planning such projects requires additional evaluations and advice from university staff before ethical approval can be received and the project can begin. In the case of a graduation project, obtaining additional advice or permits may delay the project with an extra education period or semester. To avoid this, it is recommended that students set up a graduation project with a low level of risk. Therefore, all students have to check their risk, by completing this checklist before their A1.

The first section of the checklist (A) should be completed by all students, together with their supervisor, during the planning of the graduation project, before the A1. It does not need to be submitted to anyone for review or approval. Please consider questions 1 to 3 carefully in relation to the intended graduation project, and answer with 'yes' or 'no'.

The second section of the checklist (B) should only be completed if the graduation project involves working with data from human participants. In that case, the student and their supervisor must apply for and receive ethical approval from the Human Research Ethics Committee (HREC) before the project can begin (see the paragraph 'Explanation and follow-up' after the questions). The student can submit the application to the HREC, but the supervisor is responsible for making sure that the project is compliant with relevant privacy regulations and ethical policies.

| **Section A. General considerations** | yes | no |
|---|---|---|
| 1. Is the graduation project conducted as part of an internship (at a company), or as part of a research project at TU Delft?<br><br>If a student's graduation project is conducted at a company or as part of a research project at the university, questions of data ownership and intellectual property rights need to be addressed in a written graduation or internship agreement before the project begins. Students and their supervisor should consult the Intellectual Property Rights of Students webpage. Additional information can also be found in the Extended Personal Research Data Workflow. | | ✔ |
| 2. Does the project involve conducting (part of) the research outside the Netherlands?<br><br>Students who intend to travel abroad (even to other EU countries) for study, exchange, research, internship, or graduation project purposes need to follow the Travel Safety Protocol. This includes attending a mandatory Travel Safety Training Session: see the Disclaimer. | | ✔ |
| 3. Will the research involve processing data from humans, such as running a survey, conducting interviews or workshops, collecting data through social media or internet forums, or re-using existing datasets about humans provided by a third party? (If 'yes', see follow-up questions 4 to 13 in Checklist B.)<br><br>Students who work with data from human participants must complete the next section and apply for and receive ethical approval from the Human Research Ethics Committee (HREC) before conducting the research. | ✔ | |

| **Section B. Extended risk factors** (only if question 3 has been answered with 'yes'.) | yes | no |
|---|---|---|
| 4. Will the project involve participants who may be considered vulnerable, such as the elderly, refugees or asylum seekers, ethnic minorities, patients, or people with disabilities?<br><br>Participants who may suffer very adverse consequences (for instance, due to discrimination) if their personal data became publicly available can be considered vulnerable. | | ✔ |
| 5. Will the project involve participants who cannot themselves give informed consent for taking part in the project, but for whom consent must be obtained from a legal guardian?<br><br>Participants who cannot give informed consent can include, for instance, children or participants with intellectual disabilities, mental disorders, or dementia. Such participants are also considered vulnerable in the context of the General Data Protection Regulation (GDPR). | | ✔ |
| 6. Will the project involve processing any of the special categories of personal data below?<br>- Race<br>- Ethnicity<br>- Criminal offence data<br>- Political opinion<br>- Union membership<br>- Religious or philosophical beliefs<br>- Sex life and/or sexual orientation<br>- Health data (including measurements such as heart rate)<br>- Biometric or genetic data (including fingerprints, iris scanning, facial recognition)<br><br>The General Data Protection Regulation (GDPR) defines a stricter rules for processing special categories of personal data. If it is necessary to process these data in a project, it is it is important to provide additional safeguards. | | ✔ |
| 7. Will the project involve processing personal data that could be considered sensitive, such as the ones listed below?<br>- Information about a person's income, debts, or other payments<br>- Information about a person's (un-)employment status<br>- Information about a person's performance at school or work<br>- Information about relationship problems or (gambling) addiction<br>- Information about poverty, domestic violence, or youth welfare/social work involvement<br><br>Some types of personal data are considered sensitive, because they can have a high impact on the privacy of the data subject if other persons gain access to these data. Sensitive personal data should only be processed if necessary: in such cases, additional safeguards need to be put in place. | | ✔ |
| 8. Will the project involve processing video-recordings, or photographs of participants?<br><br>TU Delft considers photographic and video-materials of research participants to be sensitive personal data. If such data need to be processed, additional safeguards must be put in place. | | ✔ |

| **Section B. Extended risk factors** (only if question 3 has been answered with 'yes'.) | yes | no |
|---|---|---|
| 9. Will the project involve sharing or transferring personal data between multiple partners or collaborating organisations involved, such as between TU Delft and an internship company?<br><br>According to privacy law, sharing personal data between organisations requires a privacy agreement to be in place: setting this up takes time, and requires support from additional university staff. Furthermore, personal data sharing can potentially expose research participants to different types of risks: these risks must be considered in the ethical application. | | ✔ |
| 10. Will the project involve deception, or covert observation of participants?<br><br>In some types of research, obtaining informed consent for processing participants' personal data is not an option: for instance, if the research involves deception, or the research is covert (conducted without participants knowing about it). In such situations, the steps to mitigate risks to participants are important, and an alternative legal basis for processing the participant's data needs to be established with the help of additional support staff. | | ✔ |
| 11. Will the project involve working with social media data?<br><br>Social media data are personal data, but since it is usually not possible to ask for informed consent for processing social media data, another legal basis for processing the participant's data needs to be established. Processing of social media data also involves legal considerations related to terms of use of data from third-party platforms: therefore, research with social media data requires expert support on privacy, ethics, and legal matters. | ✔ | |
| 12. Will the project involve using learning algorithms or other AI to analyse, combine, or otherwise process data from participants?<br><br>The use of AI in research involves many considerations in terms of data protection, ethics, security, and intellectual property: for more information, see TU Delft's Instructions for use of Generative AI. | | ✔ |
| 13. Will the project involve participants who are based in a country or countries outside of the EU?<br><br>Students affiliated with TU Delft must comply with Dutch and EU regulations of personal data processing (GDPR). Furthermore, the student and their supervisor must make sure that the research complies with local (privacy) legislations of any foreign destinations. Additional support from an external (local) expert may be required. | ✔ | |

**Explanation and follow-up**

If you have answered 'no' to all questions 4 to 13, your project is likely to be considered low or minimal-risk: see the paragraph 'Projects with minimal or low-risk' on the next page.

If you have answered 'yes' to one or more of the questions 4 to 13, your research likely involves extended or high risks to participants, according to the General Data Protection Regulation (GDPR) and TU Delft's privacy and ethical policies: for information regarding such projects, see the paragraph 'Projects with extended or high-risk' on the next pages.

**Projects with minimal or low-risk**

If you have answered 'no' to questions 4 to 13, your project is likely to be considered low-risk. This does not mean that the project involves no risks at all, but suggests that these risks can likely be addressed by the student and supervisor in the application to the Human Research Ethics Committee (HREC) within the timeline for a graduation project and without need for additional support.

Compiling the HREC application:
An application to the HREC generally involves a Data Management Plan (DMP), a risk-identification and mitigation checklist, and informed consent materials. Master's students at ABE who intend to compile a HREC application are advised to make use of the following support documents:

- the student guide
- the Example Data Management Plan for MSc projects

The graduation supervisor is responsible for the student's project and ethical application, and must provide support for compiling the HREC application documents.

Additional support
For low-risk student graduation projects, compiling of the HREC application documents should be done by the student in consultation with the supervisor. The Faculty Data Steward can be contacted for individual questions at datasteward-BK@tudelft.nl: however, the Data Steward does not provide detailed feedback on student DMPs for low-risk HREC applications.

Additional resources
The HREC has guides available for completing the checklist and for compiling informed consent materials. Additionally, the Guide to the Extended Personal Research Data Workflow has been created to help researchers and students who work with human participants comply with both GDPR principles and TU Delft's policies on Data Management and Human Research Ethics.

Timeline
Minimal or low-risk HREC applications are generally processed faster than extended or high-risk applications (see the paragraph below). Nevertheless, the initial evaluation by the HREC usually takes approximately 2 weeks, and may take longer during busy periods or holiday: see the HREC website for up-to-date information. Additionally, the application may require revisions before final approval is granted. If you do not receive an initial response about your ethical application after 4 weeks from the time of submission, you may follow up with the HREC to enquire about an update.

**Projects with extended or high-risk**

If you have answered 'yes' to one or more of questions 4 to 13, there are potential increased risks related to how data from human participants will be processed in your project. These risks will need to be addressed in consultation with the Data Steward and other relevant support staff before submitting the ethical application to the Human Research Ethics Committee (HREC).

Compiling the HREC application
An application to the HREC generally involves a Data Management Plan (DMP), a risk-identification and mitigation checklist, and informed consent materials. Master's students at ABE who intend to compile a HREC application are advised to make use of the following support documents:

- the Ethical Approval & Data Management Planning Student Information
- the Example Data Management Plan for MSc projects

The graduation supervisor is responsible for the student's project and ethical application, and must provide support for compiling the HREC application documents.

Additional support
Once the DMP has been compiled and reviewed by the supervisor, feedback should be requested from the Data Steward via DMPonline. After this, any other necessary support staff will need to be contacted. Crucially, if the project involves one or multiple ways of personal data processing that could result in high-risk to the participants according to the GDPR, the TU Delft Privacy Team must be consulted to establish whether or not a Data Protection Impact Assessment (DPIA) is required.

Additional resources
The HREC has guides available for completing the checklist and for compiling informed consent materials. Additionally, the Guide to the Extended Personal Research Data Workflow has been created to help researchers and students who work with human participants comply with both GDPR principles and TU Delft's policies on Data Management and Human Research Ethics.

Timeline
It can take a long time to compile a complete research plan and HREC application for projects involving extended risks. DMP feedback from the Data Steward usually takes around 2 weeks, but can take longer during busy periods or holidays. Receiving additional support from other staff, such as the Privacy Team, can take anywhere from a few days to multiple weeks, depending on the project and capacity of university staff. If a DPIA is deemed necessary, it can take anywhere from 4 weeks to several months.

It is important to note that advice from the Privacy Team or other support staff, as well as any additional documents (such as necessary contracts, or a DPIA, if needed) must be in place before the application is submitted to the HREC. The initial evaluation by the HREC can be processed in 2 weeks, but may take longer during busy periods or holidays: see the HREC website for up-to-date information. Additionally, the application may require revisions before final approval is granted. If you do not receive an initial response about your ethical application after 4 weeks from the time of submission, you may follow up with the HREC to enquire about an update.

Considering the limited time available for students conducting their graduation projects, students working with data from human participants are strongly advised to prioritise low-risk research projects. If a student project necessitates processing data in ways that are considered extended or high-risk, both student and supervisor need to be aware of the extended processing times involved in obtaining ethical approval and beginning the graduation project.

Date 20-Oct-2026
Correspondence hrec@tudelft.nl

Human Research Ethics Committee TU Delft
(http://hrec.tudelft.nl)

Visiting address
Jaffalaan 5 (building 31)
2628 BX Delft

Postal address
P.O. Box 5015 2600 GA Delft
The Netherlands

*Ethics Approval Application: Urban Digital Memory: Leveraging Social Media Data on Public Spaces to Inform Urban Design Processes*
*Applicant: Saffarzadeh, Soroush*

Dear Soroush Saffarzadeh,

It is a pleasure to inform you that your application mentioned above has been approved.

Thank you very much for your submission to the HREC which has been approved.

Good luck with your research!

Sincerely,

Prof.dr.ir. Ibo van de Poel
Chair HREC
Faculty of Technology, Policy and Management