King Saud University
College of Computer and Information Sciences
Department of Information Technology
IT362: Principles of Data Science
Project-Paper Draft
1st 2024/2025

Riyadh Metro in Focus: Comparative Analysis with Dubai and Madrid Metro Systems

Student's Name:	Student's ID:
Ghada Binoun	443200646
Raghad Alzkeri	443201025
Mashael Aljaad	443203034
Shahad Basalamah	442200841
Sara Alzayed	443203107
Alanoud Alamri	443200043

Contents

Authors

Raghad Alzkeri, Ghada Binoun, Mashael Aljaad, Shahad Basalamah, Alanoud Alamri,

Sara Alzayed, Khulood Alyahya

Emails: 443201025@student.ksu.edu.sa, 443200646@student.ksu.edu.sa, 443203034@student.ksu.edu.sa, 442200841@student.ksu.edu.sa, 443200043@student.ksu.edu.sa, 443203107@student.ksu.edu.sa, Kalyahy1@ksu.edu.sa

Abstract

This project aims to analyze metro station data from Riyadh, Dubai, and Madrid, focusing on data collection, bias evaluation, and exploratory data analysis (EDA). The team gathered datasets from various sources, including Google Maps, Wikipedia, and Kaggle, facing challenges with incomplete data and the need for manual interventions. Data processing steps included removing inconsistencies, filling in missing values, and visualizing data distributions and correlations to uncover patterns.

The EDA revealed insights such as the variability in walking distance and duration across stations, the seasonal temperature impact on station usage, and possible biases stemming from selective data sources.

In conclusion, the project highlighted the necessity of comprehensive and updated data for reliable analysis. Recommendations include improving data accuracy by incorporating additional sources and automating manual collection where feasible. These steps are suggested to improve the robustness of Riyadh's metro and bus data analyses by studying and comparing it with Dubai's and Madrid's metro and bus systems .

Keywords

- Data Collection
- Exploratory Data Analysis (EDA)
- Data Cleaning
- Metro Accessibility
- Geospatial Analysis
- Temperature Impact
- Walkability
- Nearby Attractions

Introduction

This study evaluates the Riyadh Metro system by comparing it with the established metro networks of Dubai and Madrid. As a recent addition to Riyadh's public transportation infrastructure, the metro has generated significant interest regarding its efficiency, accessibility, and potential to transform urban mobility. Given Riyadh's unique characteristics, including its climate and city layout, understanding how its metro system compares to other well-established networks is essential. The analysis of Dubai and Madrid's metro data serves as a baseline for identifying Riyadh's strengths and areas for improvement.

The study focuses on three key aspects: the system's structure and accessibility, which examines geographical distribution, station density, and the ability to serve diverse urban areas; the environmental impact, particularly how climate factors like temperature affect passenger comfort; and the integration with attractions and points of interest to assess how well each system supports tourism and urban engagement. This comparative analysis aims to highlight Riyadh's potential and provide insights to guide future improvements in its public transportation network.

The paper is organized to discuss related work, data collection and preprocessing methods, and the methodology used for analysis. It then presents the results and discussion, followed by the conclusion, and acknowledgements.

1. Related Work

Several studies have examined factors influencing the adoption and integration of metro systems in rapidly developing cities, including Riyadh. One area of focus is the factors driving Riyadh residents to transition from private vehicles to public transportation. Research indicates that variables such as fuel costs, parking fees, and walking time significantly influence commuters' likelihood of adopting the Riyadh Metro. For instance, a stated preference survey combined with binary logistic regression reveals that higher fuel and parking costs encourage metro usage, while high metro fares and long walking distances deter it. Another study employing a mixed logit model highlights demographic variations, showing that younger, educated individuals are more likely to adopt the metro system.

Geospatial accessibility has also been a critical area of research, particularly in evaluating Riyadh's metro network coverage. A 2024 study using GIS demonstrates that while the metro improves accessibility for car users, significant portions of the city remain underserved due to challenges in pedestrian connectivity. This finding aligns with a report by KAPSARC (2024), which emphasizes the importance of integrating diverse transportation modes to reduce car dependency and better support Riyadh's rapid urban expansion.

Urban growth and its implications for transportation have been extensively studied, particularly in relation to Riyadh's rapid expansion and the resulting traffic congestion. One study uses GIS and temporal data analysis to establish a correlation between urban sprawl and transportation inefficiencies. The research underscores the need for effective metro integration and strategic urban planning to alleviate congestion and enhance mobility within the city.

Innovative approaches, such as the use of machine learning in metro planning, offer promising solutions. A study focusing on Lanzhou, China, illustrates the potential of random forest models in optimizing metro station site selection based on spatial and socioeconomic data. The high accuracy of this model in predicting optimal station locations suggests that similar methodologies could be applied to Riyadh, aiding in the integration of metro stations with population density, road networks, and business hubs.

This study makes a unique contribution to the development of Riyadh's metro system by providing a comparative perspective and leveraging geospatial analysis. By comparing Riyadh's metro with those in Dubai and Madrid, the study identifies best practices in design, accessibility, and functionality, offering insights into structural and operational improvements that could enhance public transportation in Riyadh. Furthermore, the use of GIS tools enables a detailed mapping of metro accessibility, highlighting underserved areas and providing actionable recommendations for additional stations or services. These findings aim to improve ridership and extend the metro's reach, supporting Riyadh's broader transportation goals.

2. Data Collection and Preprocessing

Metro station identification was carried out using a Google Maps extractor to obtain the names of metro stations along with their geographic coordinates, including latitude and longitude. This step ensured precise mapping of the metro infrastructure across the cities under study.

Attractions data retrieval was conducted using the Foursquare API, which provided detailed information about nearby attractions. These attractions were categorized into key types such as entertainment, workplaces, healthcare, and education facilities, enabling a comprehensive understanding of the integration of metro stations with surrounding amenities.

Bus station mapping involved the use of Google Maps My Maps and Foursquare (specifically for Dubai) to identify and map bus stations located near metro stations. This process complemented the dataset with vital public transport connectivity information, enhancing the analysis of multi-modal transportation options in the studied cities.

Additionally, manual data collection was conducted where necessary, using Google Maps to fill in gaps and ensure the completeness of the dataset. This approach provided a robust foundation for analyzing metro and bus station accessibility and integration. Handling missing values involved visualizing data gaps using a heatmap to identify patterns and address inconsistencies. Uniform formats were ensured for critical attributes,

such as station names and geographic coordinates, to maintain data integrity and compatibility.

Feature engineering included calculating distances between metro stations and nearby attractions using geospatial formulas. New columns were added to the dataset to represent these calculated distances and estimated walking durations, providing additional dimensions for analysis.

Walkability analysis categorized metro stations as walkable ("Yes") or non-walkable ("No") based on predefined criteria. These criteria included average temperature, walking distances (in meters), and walking durations (in minutes), enabling a nuanced evaluation of pedestrian accessibility in varying urban conditions.

3. Methodology

The project focused on collecting data from metro and bus stations in Riyadh, Dubai, and Madrid to analyze and compare the transportation structures of each city, with a particular emphasis on their relevance to Riyadh's development. Data collection utilized a combination of web scraping, manual data gathering, and validation techniques to ensure comprehensive and accurate station coverage.

Python was the primary tool for data handling and cleaning, with the Pandas library employed to preprocess and organize datasets effectively. For visualization, interactive maps were developed using the Folium library. These dynamic maps displayed metro stations and nearby attractions with markers categorized by type, enhancing the analysis of station integration with urban features. Interactive elements such as pop-ups provided details on station names, walking distances, and nearby attractions.

Additionally, the maps included a visual representation of a 2,000-meter radius around each station, offering insights into coverage and accessibility. A slicer feature allowed filtering by attraction type, enabling users to focus on specific categories, such as education, healthcare, or entertainment. The analysis and visualization processes were carried out in a Jupyter Notebook environment, ensuring an interactive and iterative workflow for data processing and presentation.

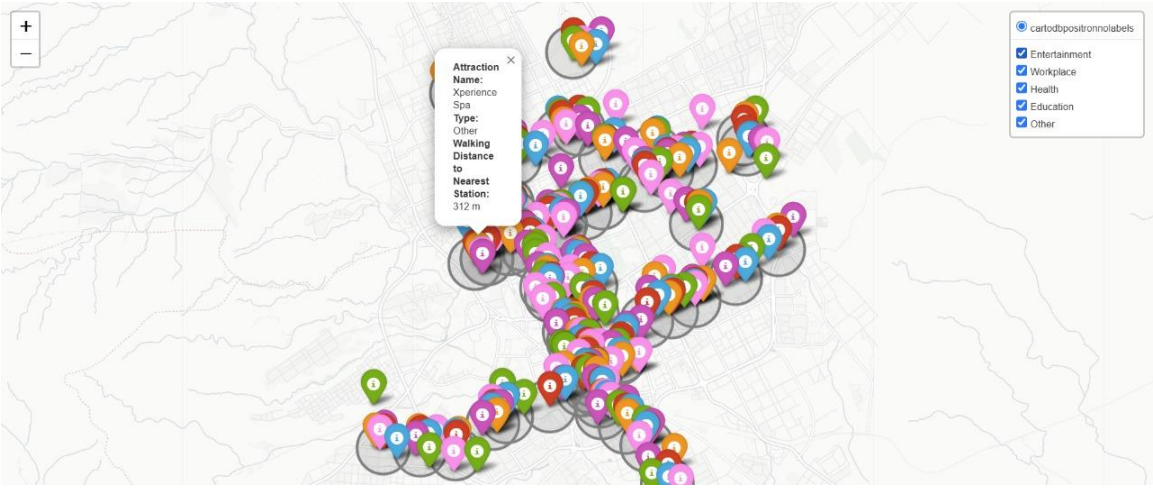


Figure1: Folium map for Metro Stations and Attractions.

4. Results

While the Riyadh Metro stations are considered well-distributed, particularly when compared to Dubai's metro system, our analysis reveals notable gaps in accessibility for Riyadh's bus stations. Many key attractions in Riyadh lack nearby bus stations, suggesting that the integration between metro and bus services is less effective. This issue is further compounded by the fact that many bus stations are not yet operational. In contrast, Dubai's bus system demonstrates exceptional walkability with direct connections to key tourist and commercial areas, making it highly accessible. However, Dubai's metro stations exhibit poor walkability overall, falling short when compared to Riyadh's metro system in terms of ease of access.

Similarly, Madrid's metro system excels due to its dense urban planning, with stations strategically placed close to major attractions, resulting in superior accessibility and integration. The walkability of Madrid's metro stations highlights its focus on urban efficiency, providing valuable insights for Riyadh to enhance its own network.

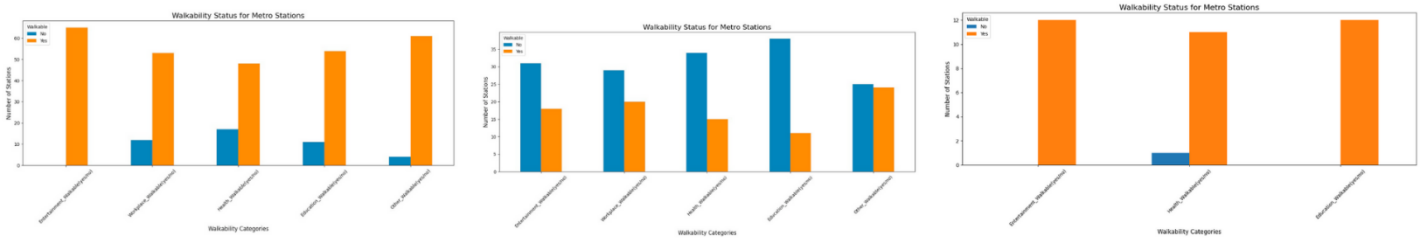


Figure 2: walkability bar chart for Riyadh, Dubai, and Madrid (respectively)

5. Discussion

This study has several limitations:

- **Data Scope**: Our research focused on structural and spatial factors, and we did not include socioeconomic data or predictive modelling techniques, which could further inform about potential usage patterns or preferences among different population segments.
- **Lack of Real-World Usage Data**: Since Riyadh's metro system is not yet operational, the analysis is based on planned station locations rather than real ridership data. This limits our ability to fully predict how the metro will perform in practice.
- **Comparative Analysis Limitations**: While Dubai and Madrid serve as useful benchmarks, each city's unique cultural, economic, and geographic context may limit the generalizability of our findings directly to Riyadh.

6. Conclusion

This study has highlighted the strengths and weaknesses of Riyadh's metro and bus transit systems in comparison to those of Dubai and Madrid. While the Riyadh Metro stations are well-distributed and demonstrate strong walkability compared to Dubai's less accessible metro system, the city's bus stations show significant gaps in accessibility, with many key attractions lacking nearby stations. In contrast, Dubai excels in bus station walkability, offering direct connections to tourist and commercial areas, though its metro stations are comparatively less walkable. Madrid's transit system stands out for its integrated design, with metro stations strategically placed near major attractions and complemented by a well-connected bus network. These findings underscore the importance of improving Riyadh's pedestrian infrastructure and enhancing the integration between its metro and bus networks to maximize urban mobility and accessibility.

Future research should focus on collecting additional data for Madrid to provide a more comprehensive comparison and better insights into its transit system. Furthermore, additional data collection on bus station distribution and walkability for all three cities—Riyadh, Dubai, and Madrid—will help identify further opportunities for improvement. To refine the walkability criteria, future work will also focus on separating the temperature averages into morning and night temperatures, as well as distinguishing between colder and warmer seasons. Expanding the scope of the study to include socioeconomic and environmental factors could also provide deeper insights into the broader impacts of transit systems on urban development. These steps will not only enhance the current analysis but also guide policymakers in designing more sustainable and efficient transportation networks.

7. Acknowledgements

The authors would like to express their sincere gratitude to King Saud University (KSU) for providing the resources and support necessary to complete this research. We extend our heartfelt thanks to our supervisor, Dr. Khulood Alyahya, for her invaluable guidance and encouragement throughout the project. Additionally, we deeply appreciate the collaboration and dedication of our peers and teammates, whose efforts were instrumental in the success of this work.

8. References

1. Mandal, D., Deng, S., Jana, S., Wing, J., & Hsu, D. (2020). *Ensuring Fairness Beyond the Training Data*. In *Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020)*. Accessed: Oct. 03, 2024. [Online]. Available: https://proceedings.neurips.cc/paper/2020/file/d6539d3b57159babf6a72e106beb45bd-Paper.pdf
2. Tatineni, S. (2019). *Ethical Considerations in AI and Data Science: Bias, Fairness, and Accountability*. *International Journal of Information Technology and Management Information Systems (IJITMIS)*, 10(1), 11-20. [Online]. Available: https://www.researchgate.net/publication/377701616_Ethical_Considerations_in_AI_and_Data_Science_Bias_Fairness_and_Accountability
3. [Author(s)], "[Title of Paper]," in *Proceedings of [Conference Name]*, [Location], [Year]. [Online]. Available: http://iieng.org/images/proceedings_pdf/2526E0215016.pdf
4. Youssef, Z., Alshuwaikhat, H., & Reza, I. (2021). *Modeling the Modal Shift towards a More Sustainable Transport by Stated Preference in Riyadh, Saudi Arabia*. *Sustainability*, 13(1), 337. https://doi.org/10.3390/su13010337
5. Niu, Q., et al. (2023). *Selection and prediction of metro station sites based on spatial data and random forest: a study of Lanzhou, China*. *Scientific Reports*, 13(1), 22542. https://doi.org/10.1038/s41598-023-49877-6
6. Anwar, A. M., Oakil, A. T., Muhsen, A., & Arora, A. (2023). *What would it take for the people of Riyadh city to shift from their cars to the proposed metro? Case Studies on Transport Policy*, 12, 101008. https://doi.org/10.1016/j.cstp.2023.101008
7. Abdelrahman, "مشروع الملك عبدالعزيز للنقل العام بمدينة الرياض," *الهيئة الملكية لمدينة الرياض*, Nov. 29, 2022. https://www.rcrc.gov.sa/ar/projects/public-transport-project-riyadh
8. [1]"Riyadh Public Transportation," Riyadhbus.sa, 2024. https://www.riyadhbus.sa/ar/home.

Appendix

Algorithm for Walkability Evaluation

- evaluate walkability based on the following criteria:
  - Distance ≤ 2000 meters.
  - Walking duration ≤ 25 minutes.
  - Average temperature < 30°C.

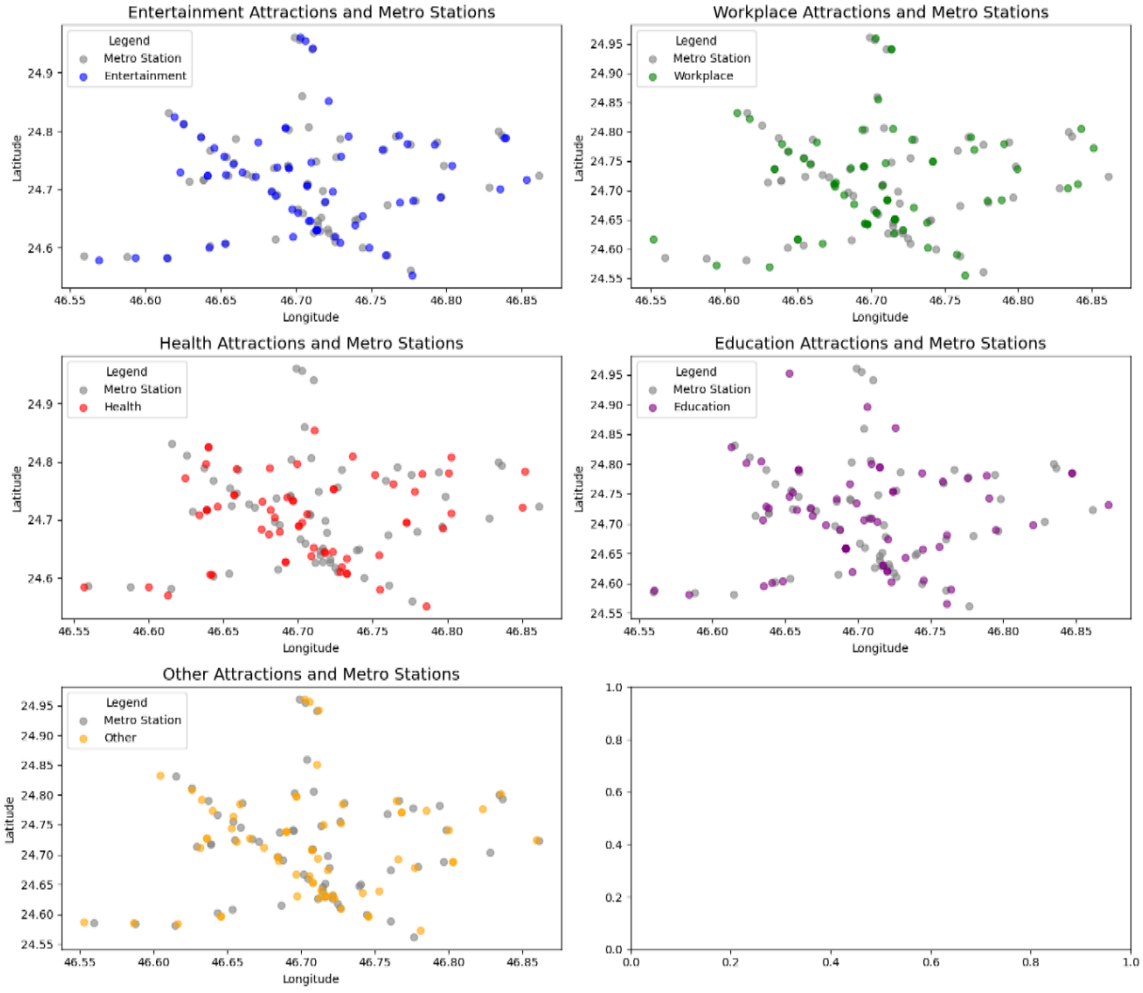
  If all conditions are met, return Yes; otherwise, return No.

Algorithm for Generate an Interactive Map with Metro Stations and Attractions using Folium Library in Python

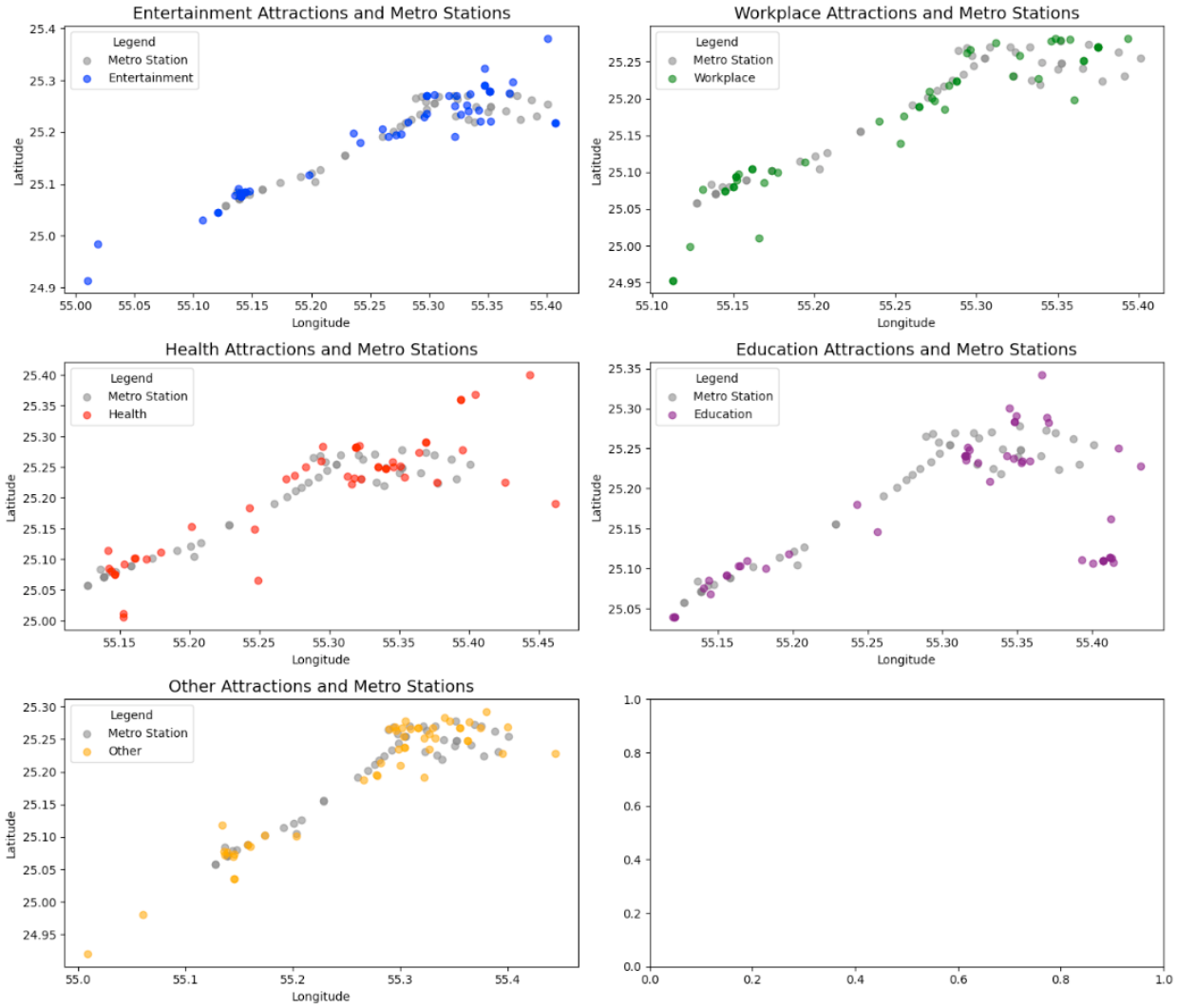
- Define the coverage radius for metro stations (2000 meters).
- Specify color codes for different attraction types (Entertainment, Workplace, etc.).
- Create a Folium map with a default zoom level and a tile style and add the metro stations and attractions to it.

Data Visualization

Riyadh



Dubai


Entertainment Attractions and Metro Stations
Legend
Metro Station
Entertainment
Latitude
Longitude
Workplace Attractions and Metro Stations
Legend
Metro Station
Workplace
Latitude
Longitude
Health Attractions and Metro Stations
Legend
Metro Station
Health
Latitude
Longitude
Education Attractions and Metro Stations
Legend
Metro Station
Education
Latitude
Longitude
Other Attractions and Metro Stations
Legend
Metro Station
Other
Latitude
Longitude

Madrid

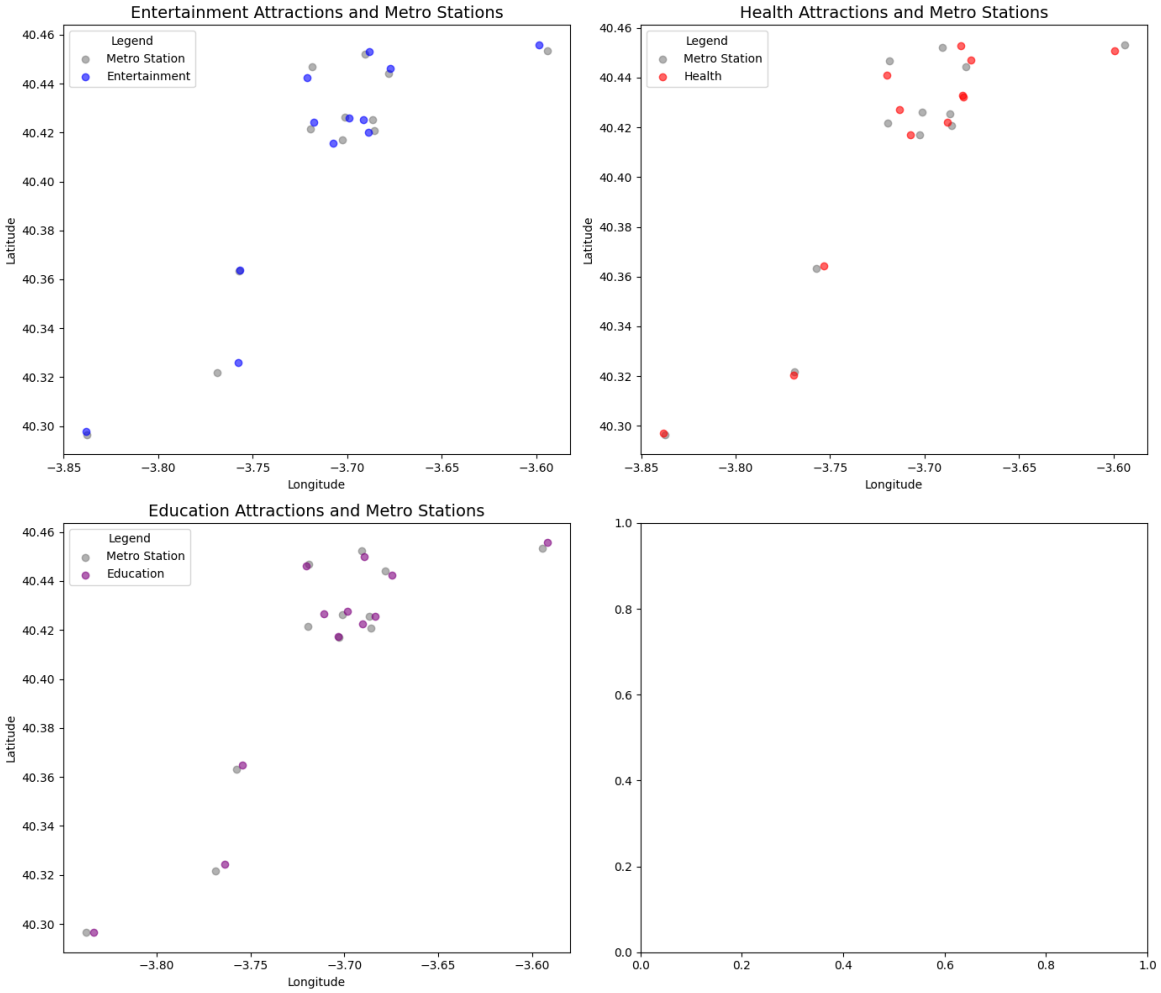
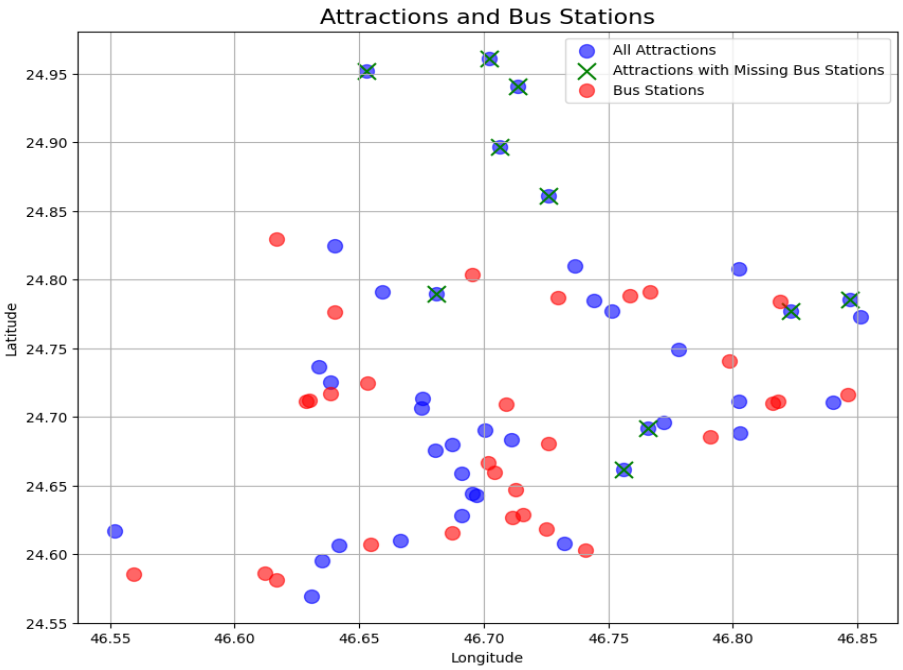


Figure2: Scatter Plot for Attractions location.

Riyadh



Dubai

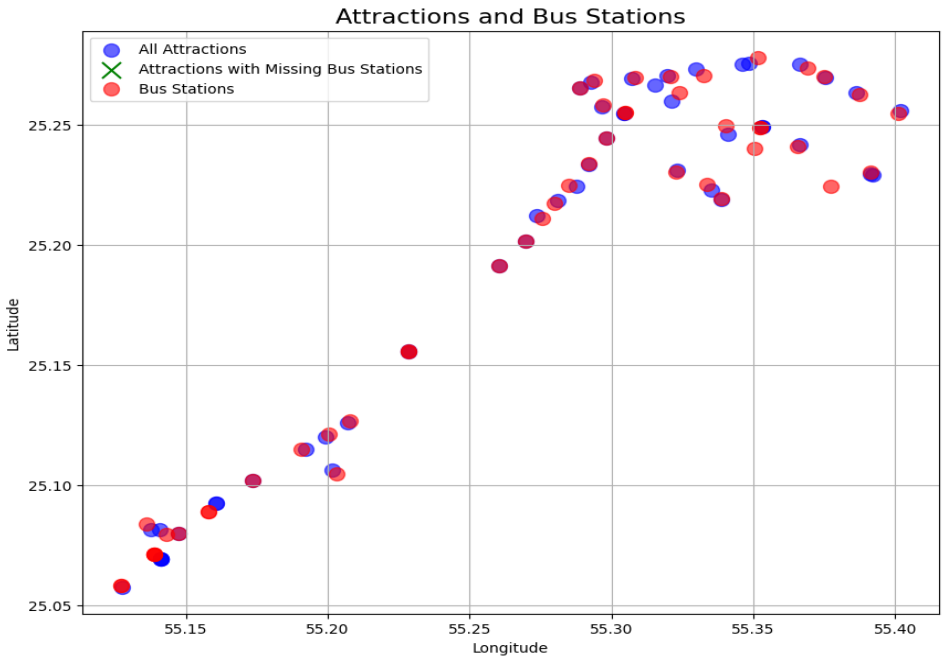


Figure3: Scatter Plot for Attractions and Bus Stations highlights attractions with missing bus station.