Congestion Analysis and Route Optimization for Buses

Presented by: Emaan Bashir (20594616)
Supervised by Colin Johnson

Agenda

1. Introduction
2. Problem Statement
3. Research Objectives
4. Significance of the Study
5. Literature Review
6. Data Collection and Preprocessing
7. Identification of Congested Routes
8. Analysis of Congestion Patterns
9. Route Optimization
10. Evaluation of Alternative Routes
11. Conclusion and Future Work

Introduction

- Urban traffic congestion is a persistent issue worldwide, particularly affecting the efficiency of public transportation systems.

- As cities grow, increased traffic volume leads to delays, reduced public transport reliability and environmental pollution.

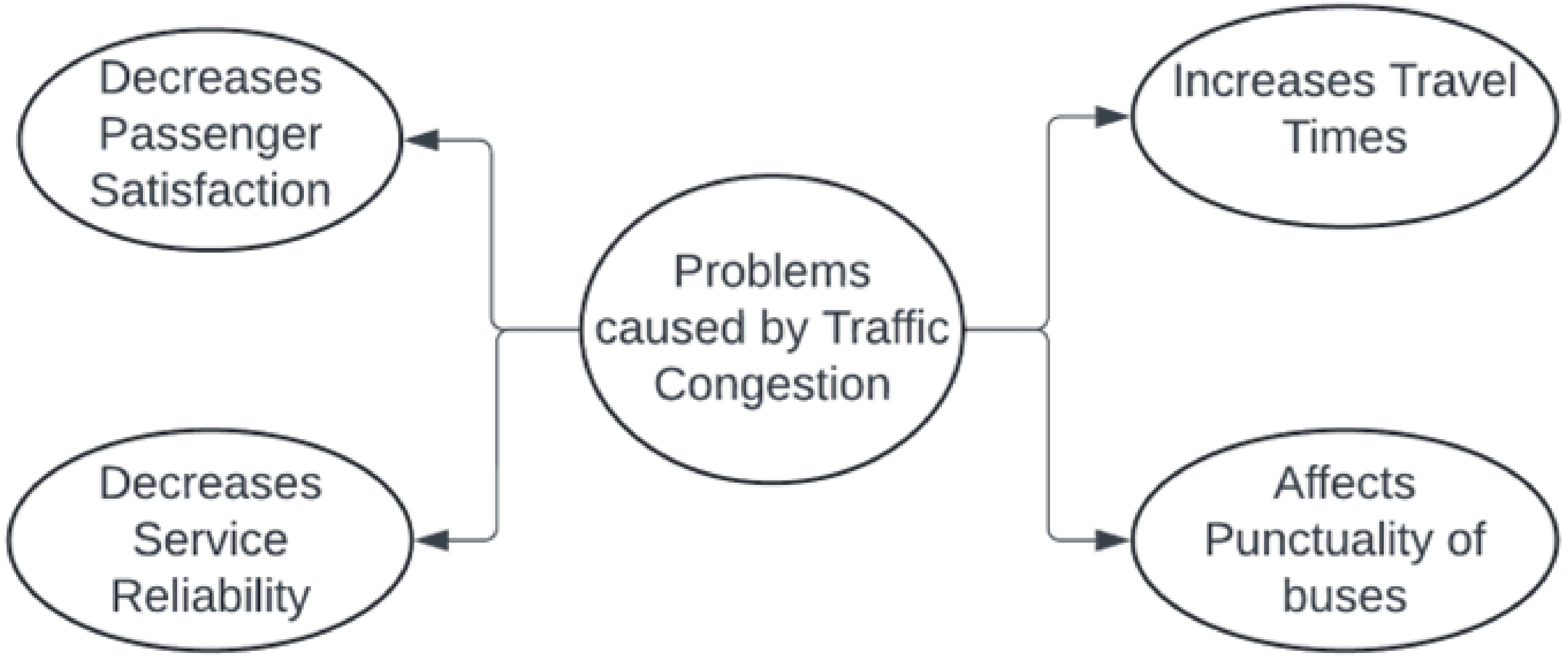
- Nottingham City Transport (NCT) operates around 300 buses across Nottingham, which are the backbone of the city’s public transportation network.

- Traffic congestion, especially during peak hours, disrupts the service quality, affecting punctuality and passenger satisfaction.

- Optimizing bus routes can reduce congestion and improve efficiency, leading to better public transportation services and decreased reliance on private vehicles.

Problem Statement

- Traffic congestion on NCT bus routes results in delays, non-adherence to schedules, and decreased passenger satisfaction.
- Congestion patterns vary across days and times, making it harder to manage.
- Limited application of advanced data analytics and machine learning techniques to predict and mitigate congestion.
- There is a dire need for an approach that leverages bus location data to study the traffic congestion patterns and uses data analytics to identify and optimize congested routes



Research Objectives

The main objective of this research is to develop and evaluate a congestion analysis and route optimization system for buses operating in Nottingham.

- **Identification of Congested Routes:** Locate routes in Nottingham experiencing high levels of traffic congestion.

- **Analysis of Congestion Patterns:** Determine peak congestion periods and hotspots.

- **Proposal of Alternative Routes:** Explore alternative, more efficient routes using data-driven methods.

- **Evaluation of the Proposed Routes:** Assess the performance of alternative routes in terms of travel time reduction.

Significance of the Study

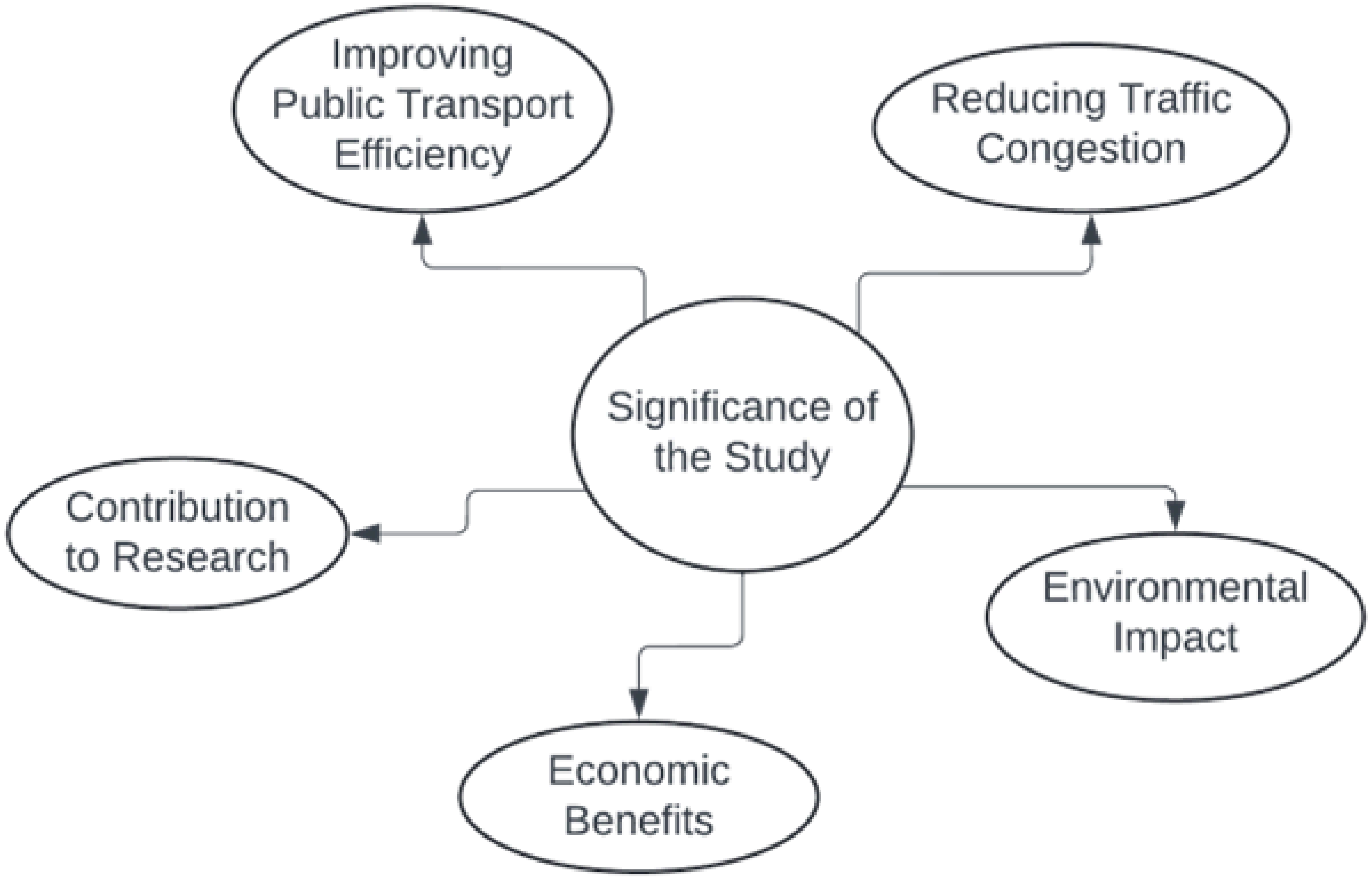
- **Improved Public Transport Efficiency:** Reduced travel times and enhanced punctuality, leading to a more reliable bus service.

- **Reduced Traffic Congestion:** Optimized routes encourage more people to use buses instead of cars.

- **Environmental Impact:** Decreased emissions as buses spend less time in traffic.

- **Economic Benefits:** Efficient transportation supports local economic activities by improving accessibility.

- **Contribution to Research:** The methodology and the insights generated, contribute to the academic field of transportation planning and data science.



Literature Review

Traffic Congestion:

- Defined as a condition where traffic flow is slower than expected. Major impacts include delays, increased emissions, and noise pollution. (Afrin & Yodo, 2020)
- Congestion can be recurrent (daily peak hours) or non-recurrent (accidents, roadwork) (Lomax & Levinson, 1997).

Real-Time Data and Congestion Analysis:

- Automatic Vehicle Location (AVL) systems provide real-time data, enabling the identification of congestion patterns (Almeida et al., 2023).
- Clustering techniques (k-Means, DBSCAN) are commonly used to classify congestion levels based on traffic flow and vehicle speed (Diker & Nasibov, 2012).

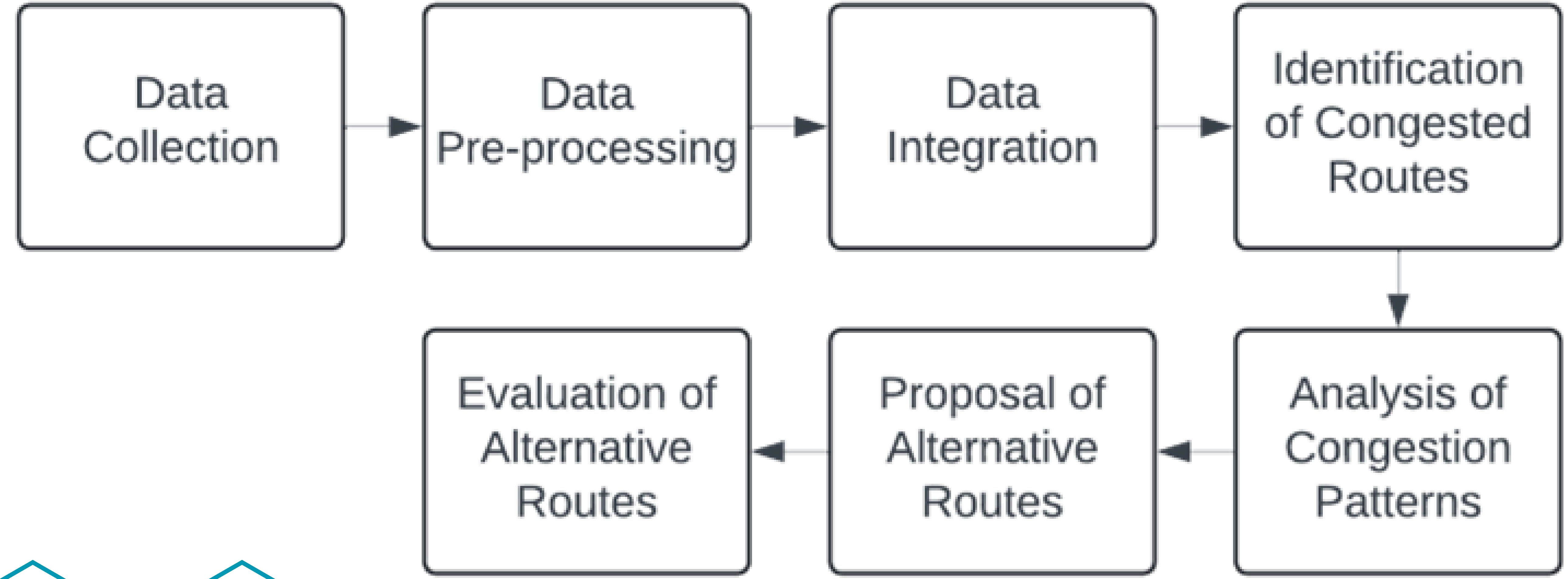
Predictive Modelling and Travel Time Estimation:

- Segment-based travel time estimation models provide more accurate predictions by dividing routes into segments based on factors like traffic signals and passenger boarding times (Ma et al., 2019).

Route Optimization:

- Traditional methods focus on minimizing travel time and maximizing coverage. Graph-based algorithms and real-time adaptive systems provide a more flexible, responsive approach (Huang et al., 2020).

Methodology



Data Collection

Bus Location Data:

- The primary source of bus location data was the Bus Open Data Service's public API, which provides real-time bus location data in XML format. This data contains details such as vehicle locations, journey references and timestamps.

- The data was fetched at 15 second intervals for two weeks. This frequency ensured the capture of high quality location data. The data was stored in the HDF5 format, to ensure efficient access to the data.

Bus Timetable Data:

- The timetable data was obtained from the Bus Open Data Service (BODS) in the TransXChange (TXC) format. TransXChange files are XML-based documents that contain detailed information about bus and coach services, including schedules, routes, stops, operators, and vehicle journeys.

- This data was utilized to map the location data of the buses to the planned schedule. The timetable data was pre-processed and stored in an HDF5 format, facilitating efficient querying and integration with the location data

Data Pre-processing

- **Parsing XML Data:** The XML data from the API was parsed using python library 'ElementTree' and transformed into pandas dataframes.
- **Data Cleaning:** This involved removing duplicates, data type conversion, removing inconsistent data and handling missing data.
- **Data Transformation:** This involved time zone handling, time-related features extraction, run time calculation, track distance calculation and computation of aimed departure and arrival times.
- **Data Integration:** The timetable and location datasets were merged to get a comprehensive view of bus operations.
- **Data Storage:** The data was stored as an HDF5 file, a format known for its efficiency in storing and retrieving large datasets with complex structures.

Identification of Congested Routes

Metrics Used:

- **Average Speed:** Calculated for each route link by dividing distance by run time.
- **Bus Density:** Number of buses per unit distance on each route link.

Exploratory Data Analysis:

- Histogram and box plot analysis of speed and density data.
- Identified outliers to clean and refine data.

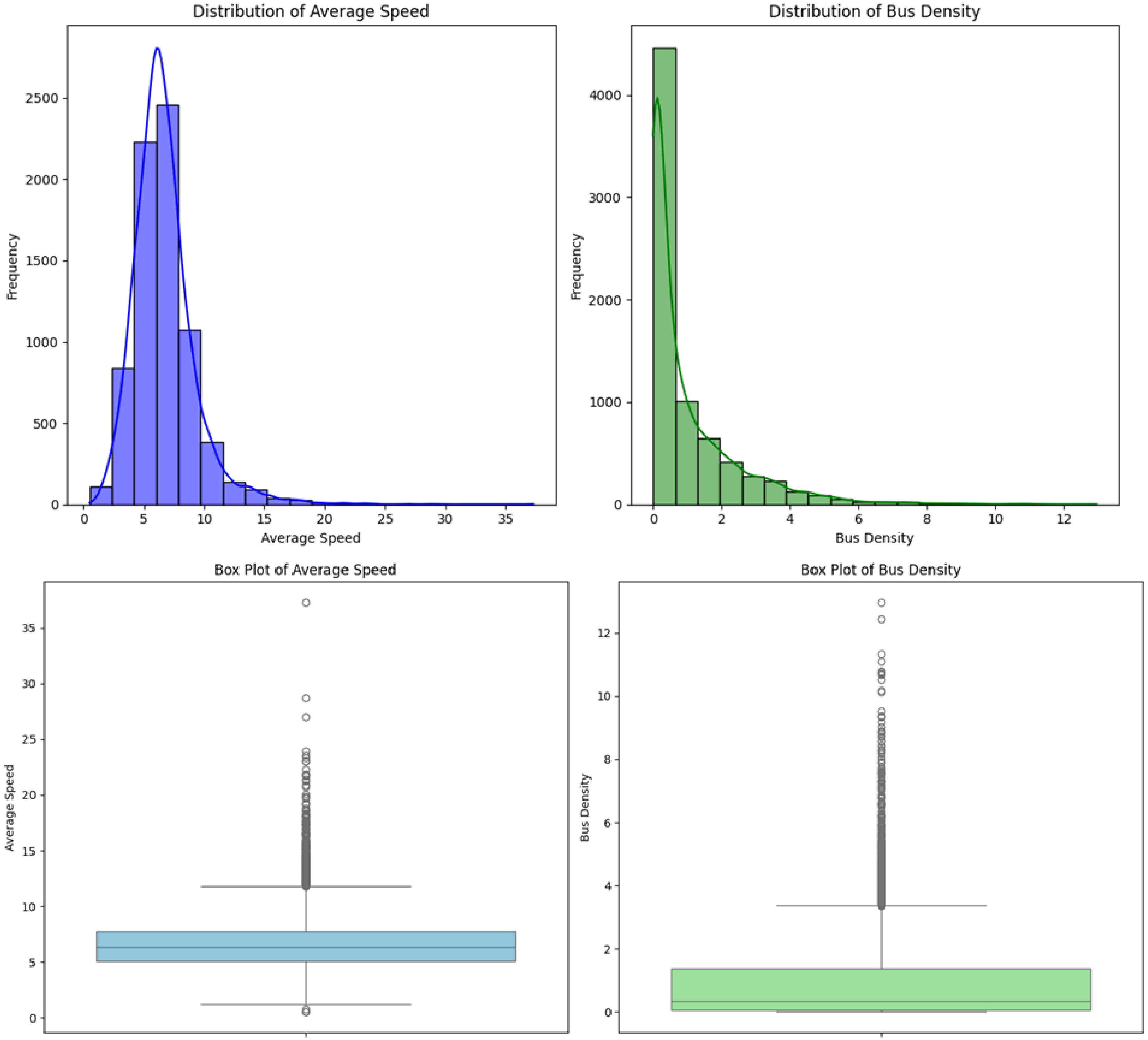
$$\boxed{Bus\ Density = \frac{Number\ of\ buses}{Track\ Distance}}$$

Cluster Analysis:

- Applied k-Means and DBSCAN algorithms to classify routes into low, medium, and high congestion levels.

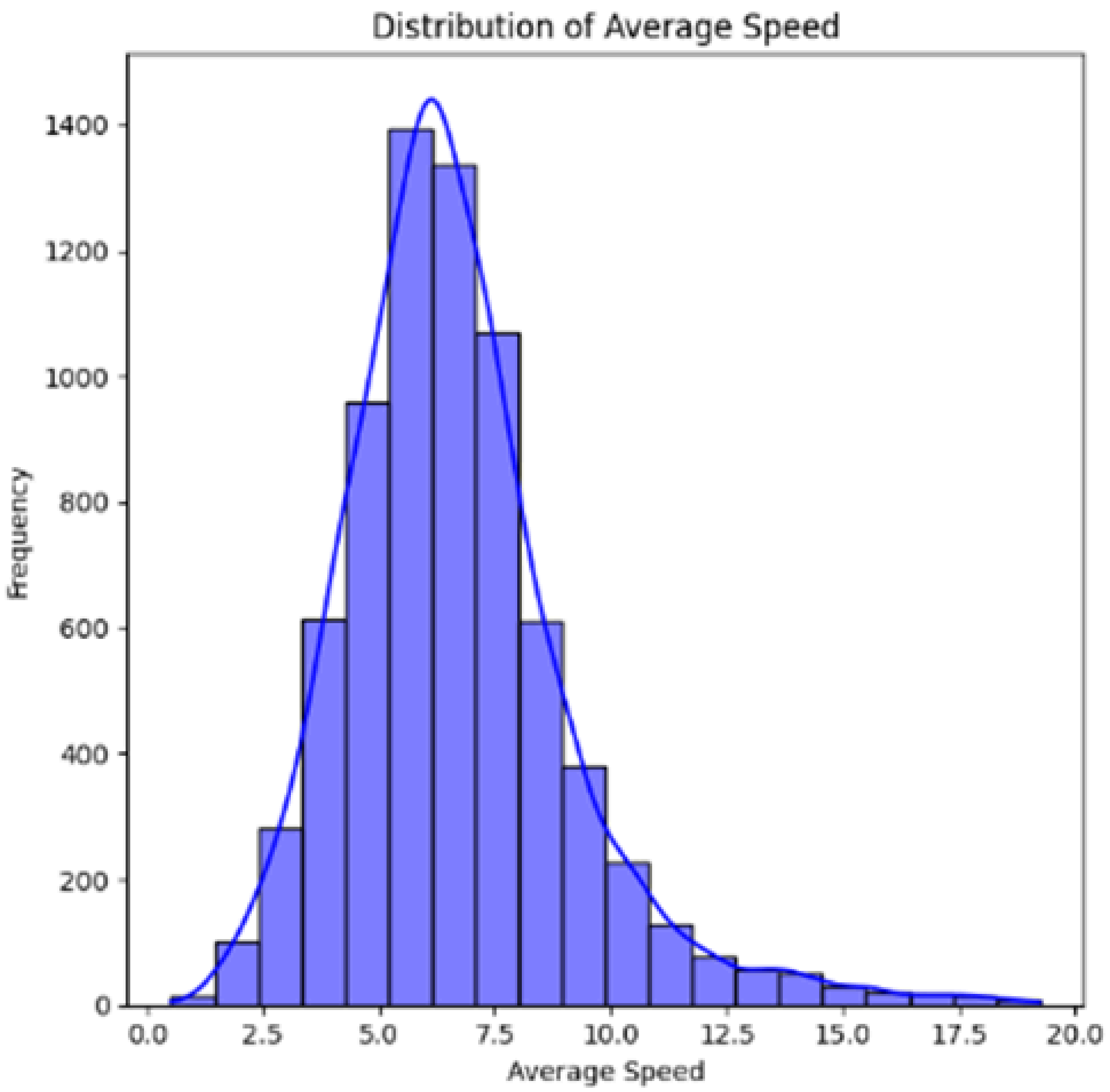
Exploratory Analysis

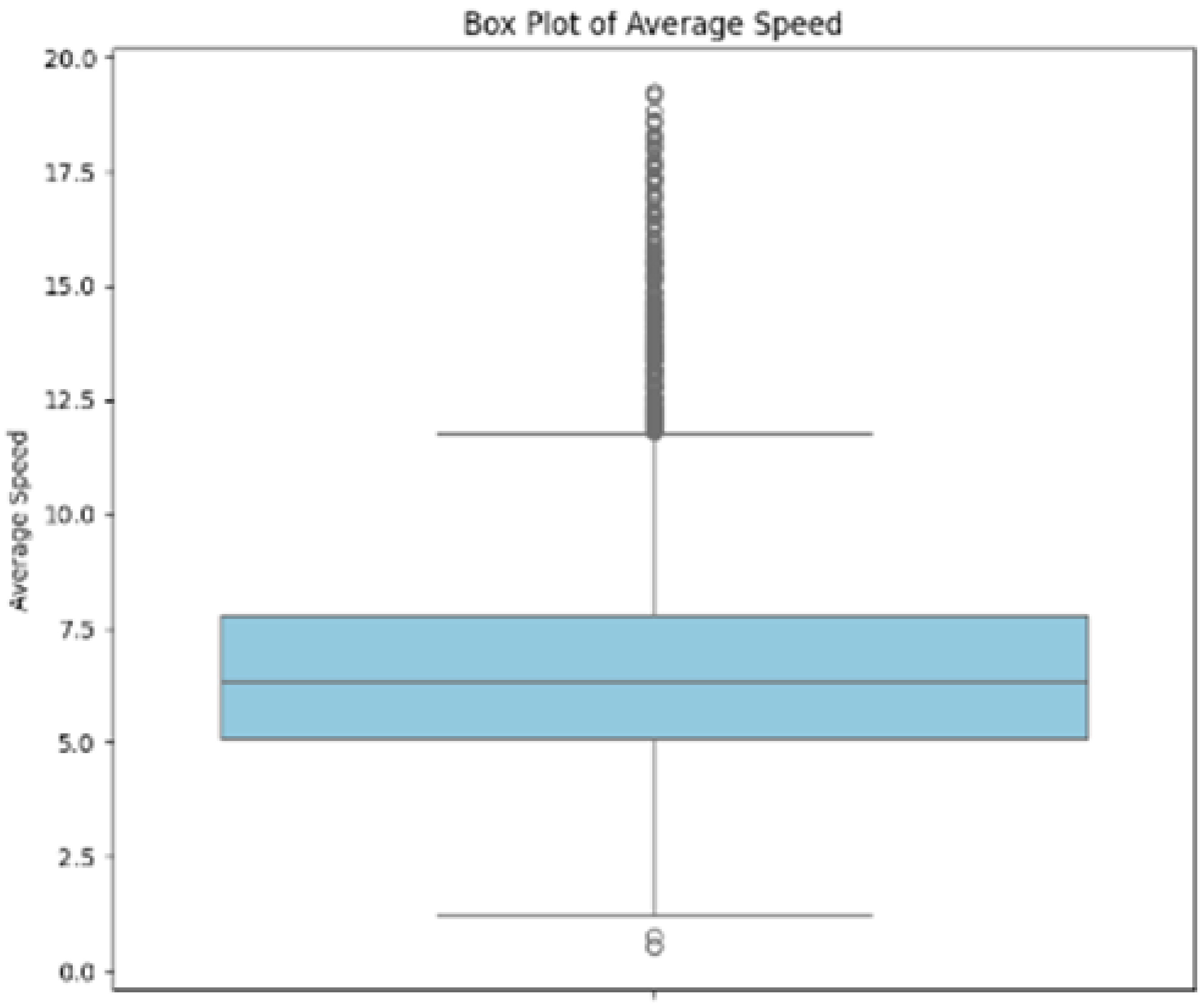
Distribution of Average Speed and Bus Density



Exploratory Analysis

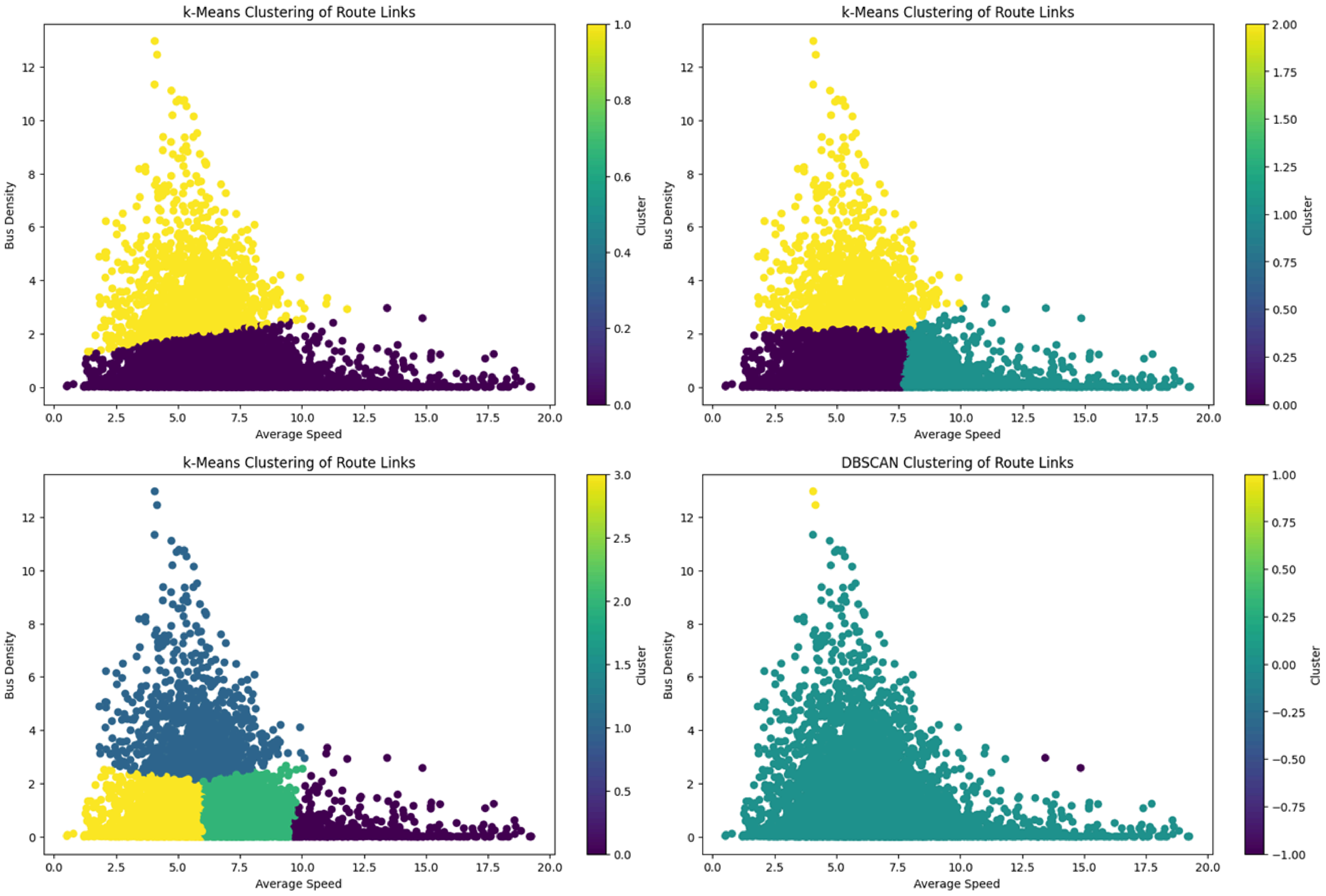
Outlier Removal based on z-scores





Cluster Analysis

k-Means and DBSCAN



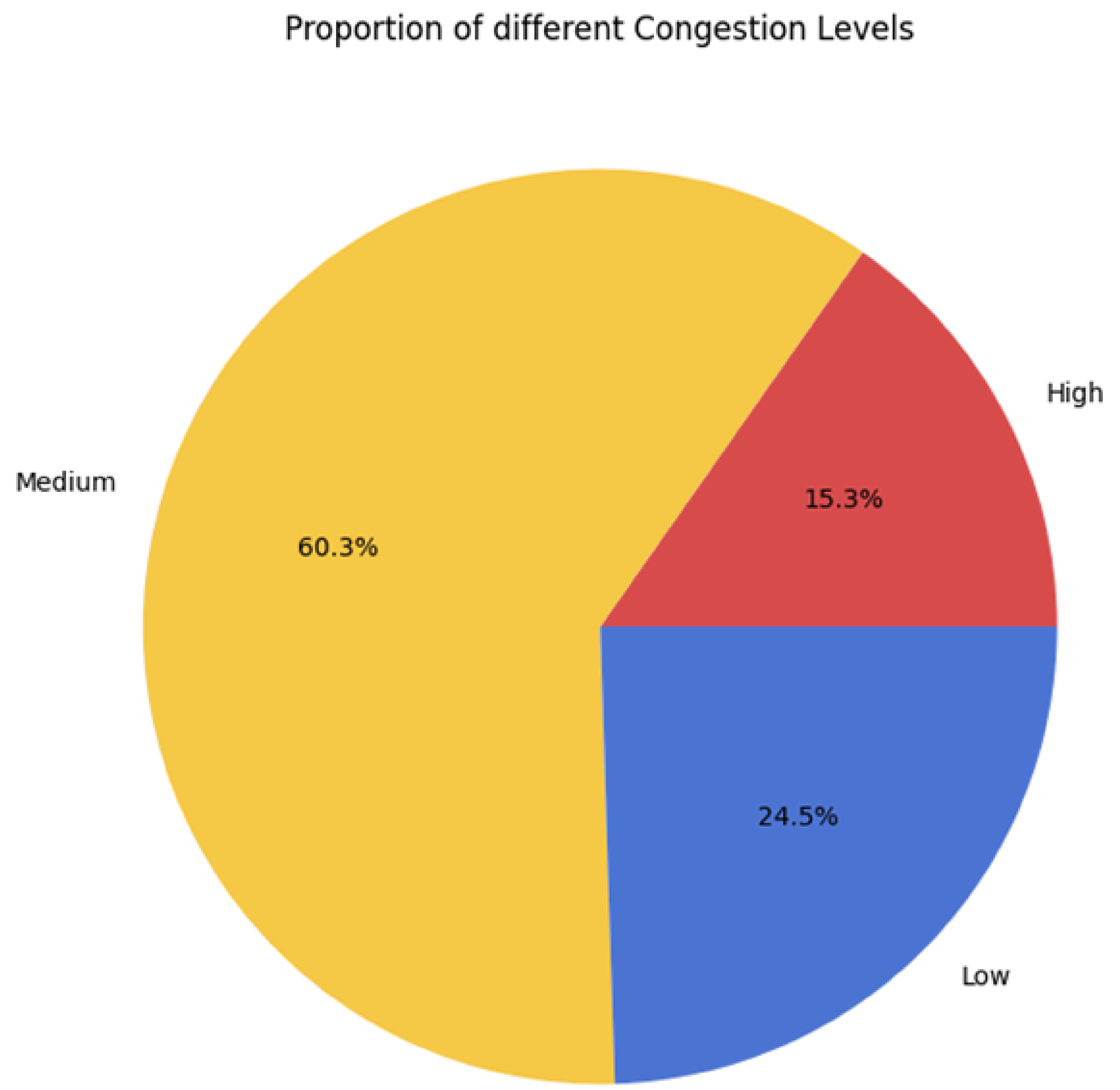
Analysis of Congested Routes

- Distribution of Congestion Levels
- Visualization of Congestion Levels on the Map
- Analysis of Congestion by Route Section
- Congestion Analysis by Day of the Week
- Congestion Analysis by Time Interval

Analysis of Congested Routes

Distribution of Congestion Levels

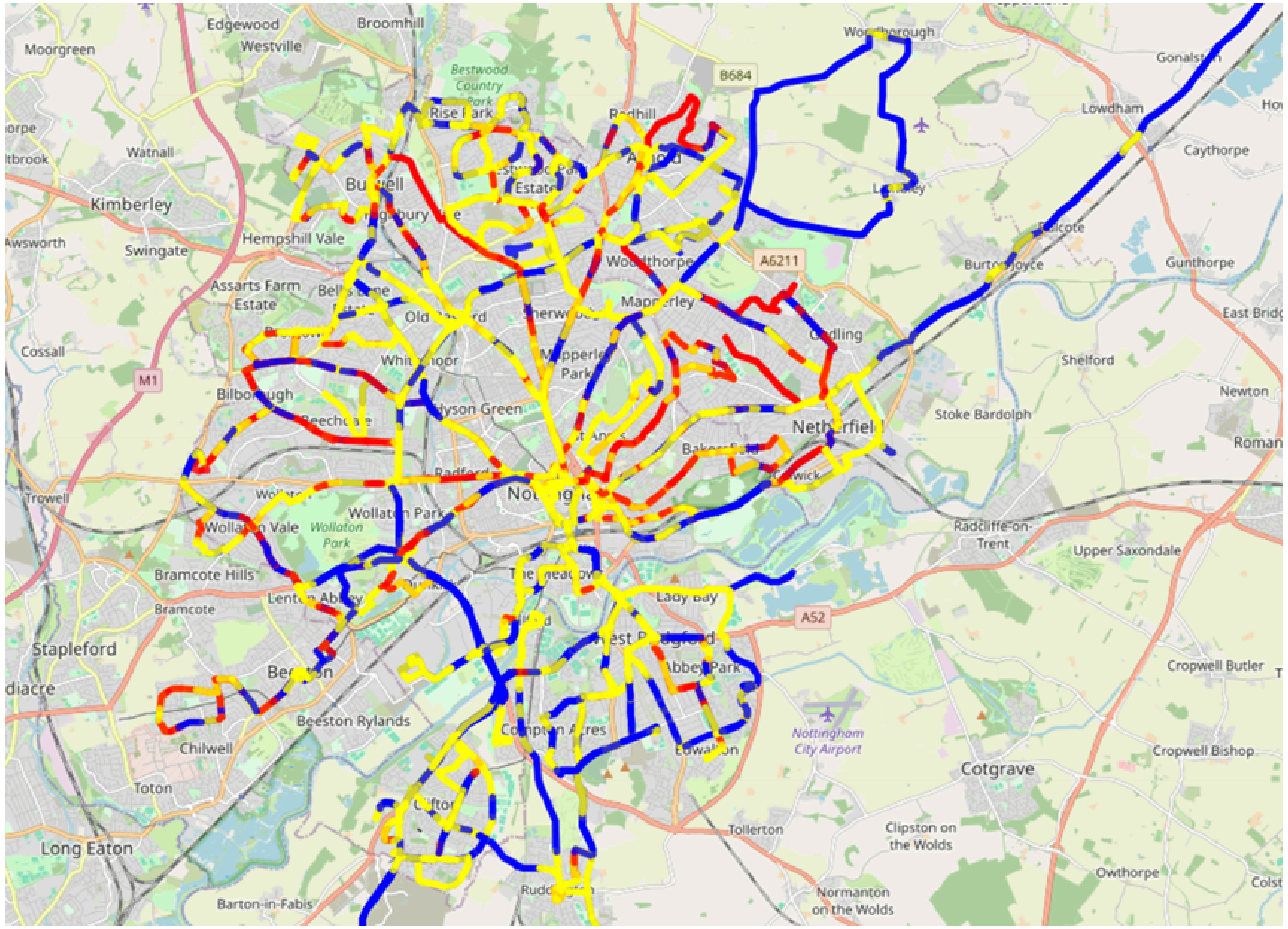
	Number of Route Links
High Congestion	1128
Medium Congestion	4454
Low Congestion	1810



Analysis of Congested Routes

Visualization of Congestion Levels

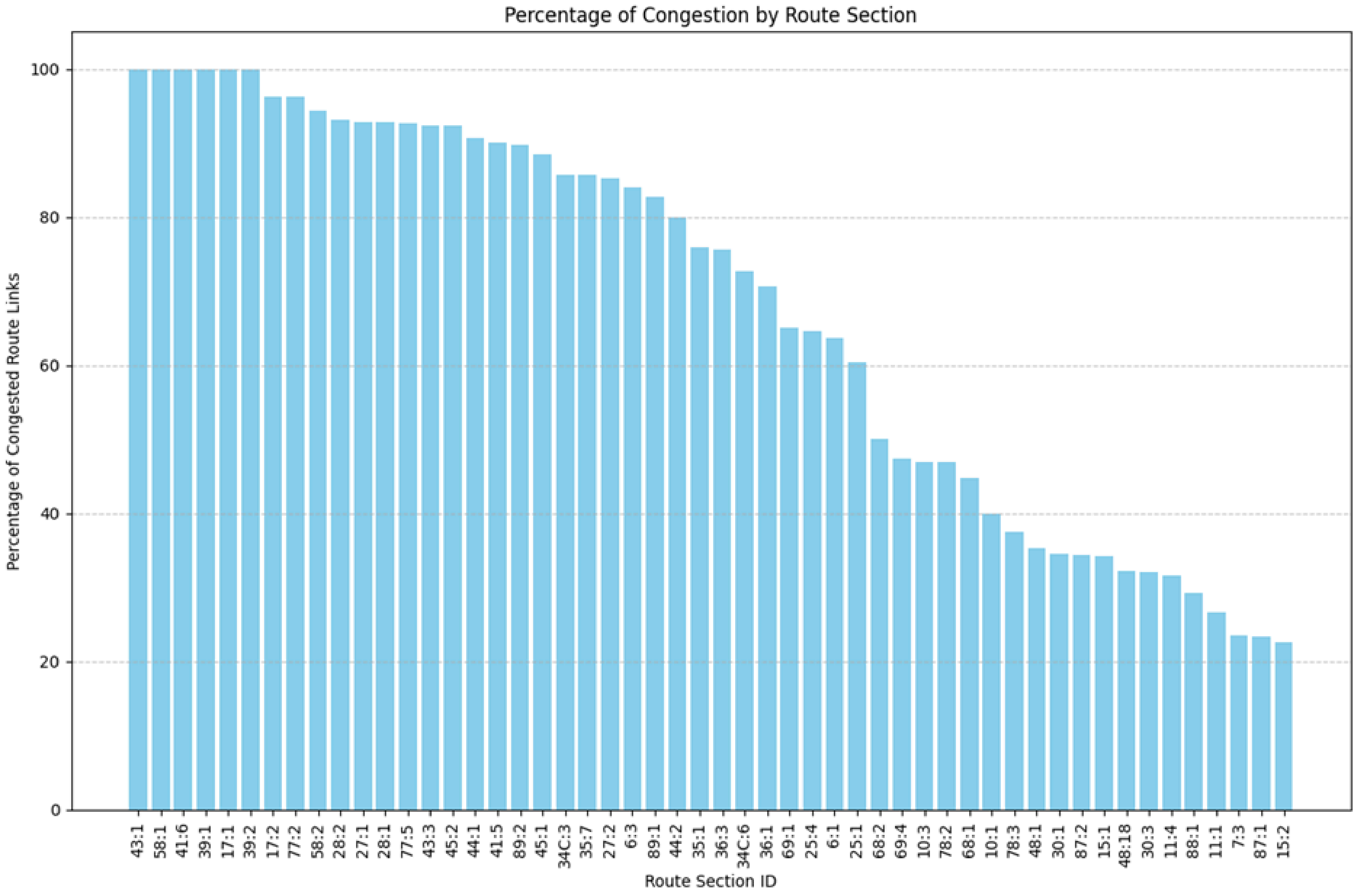
- A map was generated to show congestion levels on different routes:
  - Red: High congestion
  - Yellow: Medium congestion
  - Blue: Low congestion

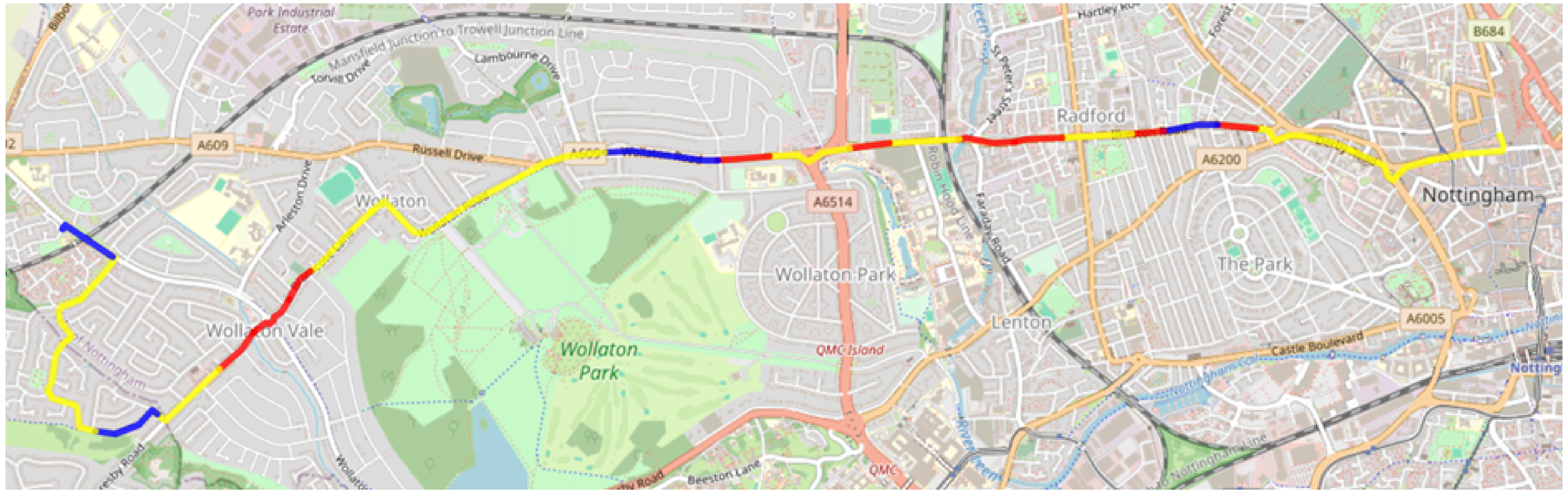


Analysis of Congested Routes

Analysis of Congestion by Route Section

$$\text{Congestion Percentage} = \frac{\text{Number of Congested Route}}{\text{Total Number of Route Links}}$$



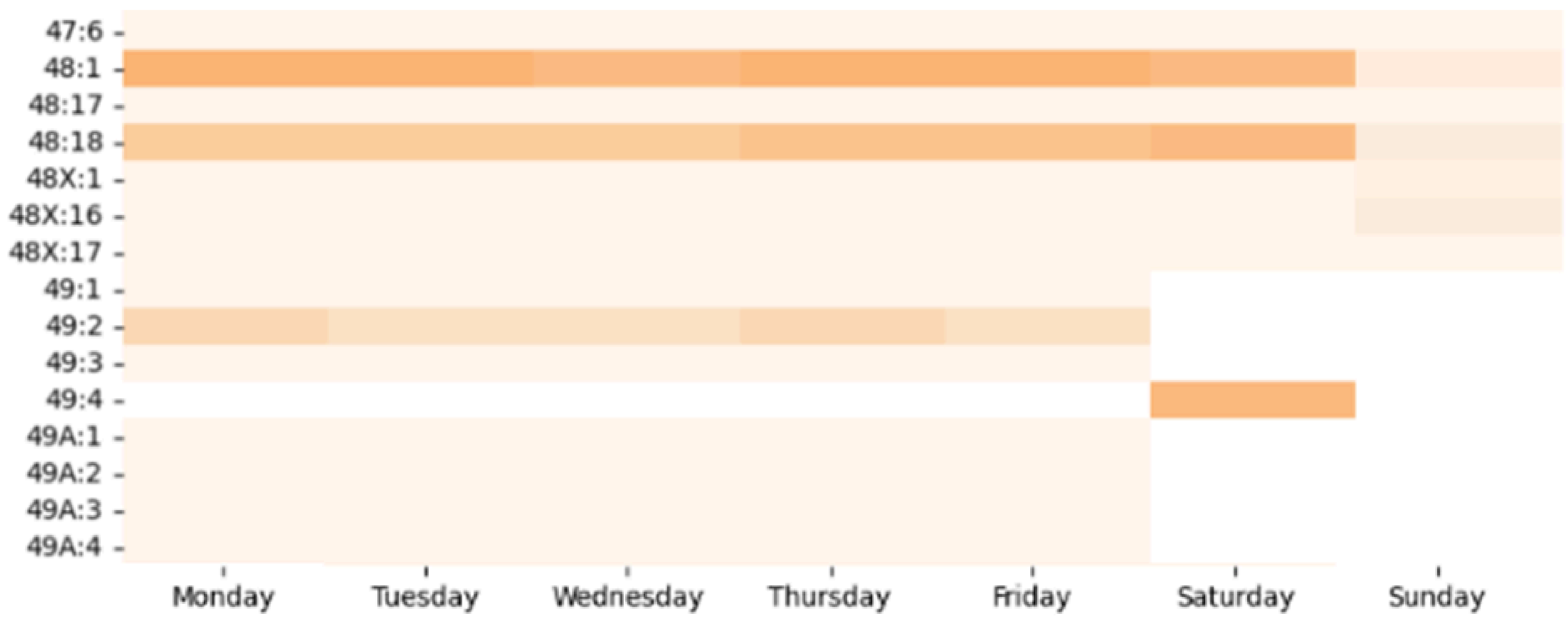


Visualization of congestion levels across Route 30:3

Analysis of Congested Routes

Congestion Analysis by Day of the Week

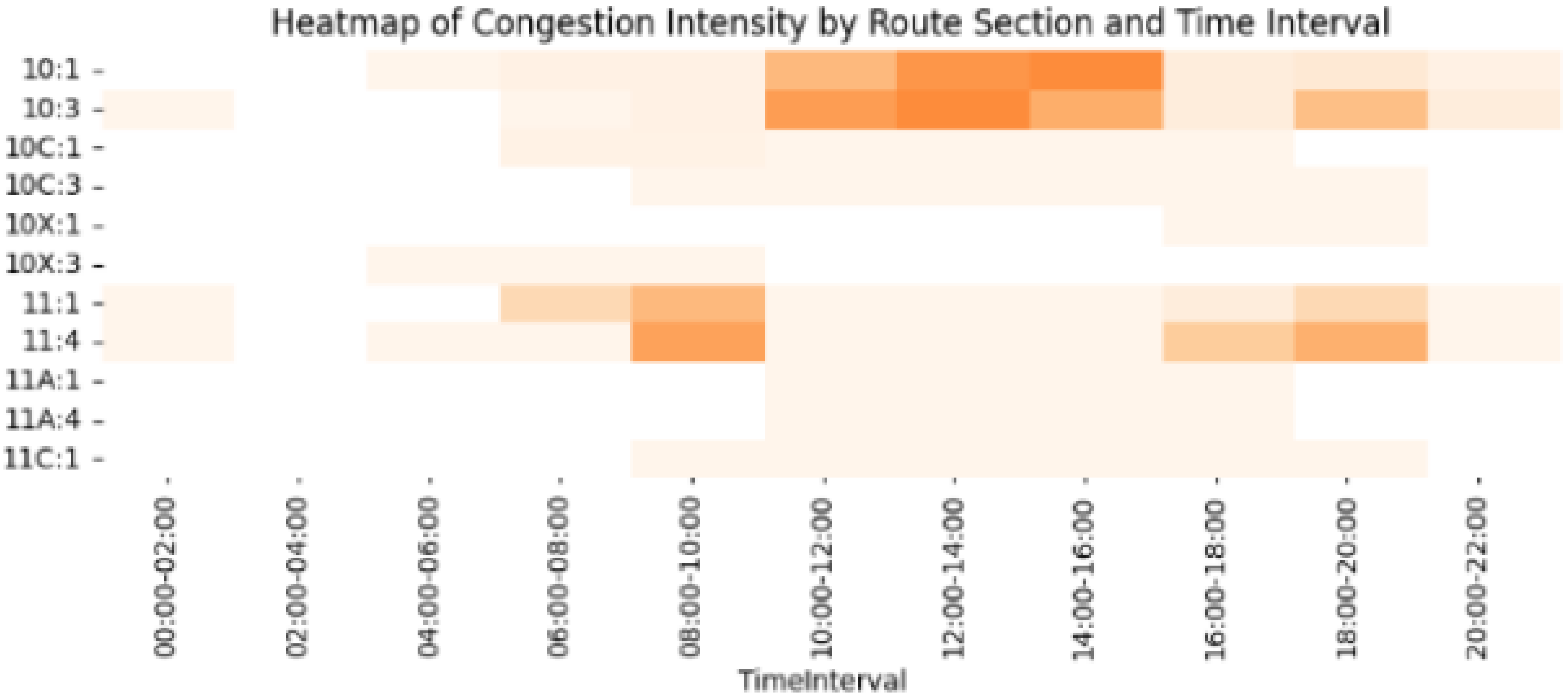
- Congestion levels vary across days of the week.
- Weekdays show higher congestion due to commuter traffic (e.g. offices, schools).
- Weekends generally have lower congestion, except for certain routes that maintain high levels consistently.
- Some routes, like Route 49:4, are operational only on Saturdays, while Routes 48:1 and 48:18 experience consistent congestion, with relief on Sundays,



Analysis of Congested Routes

Congestion Analysis by Time Interval

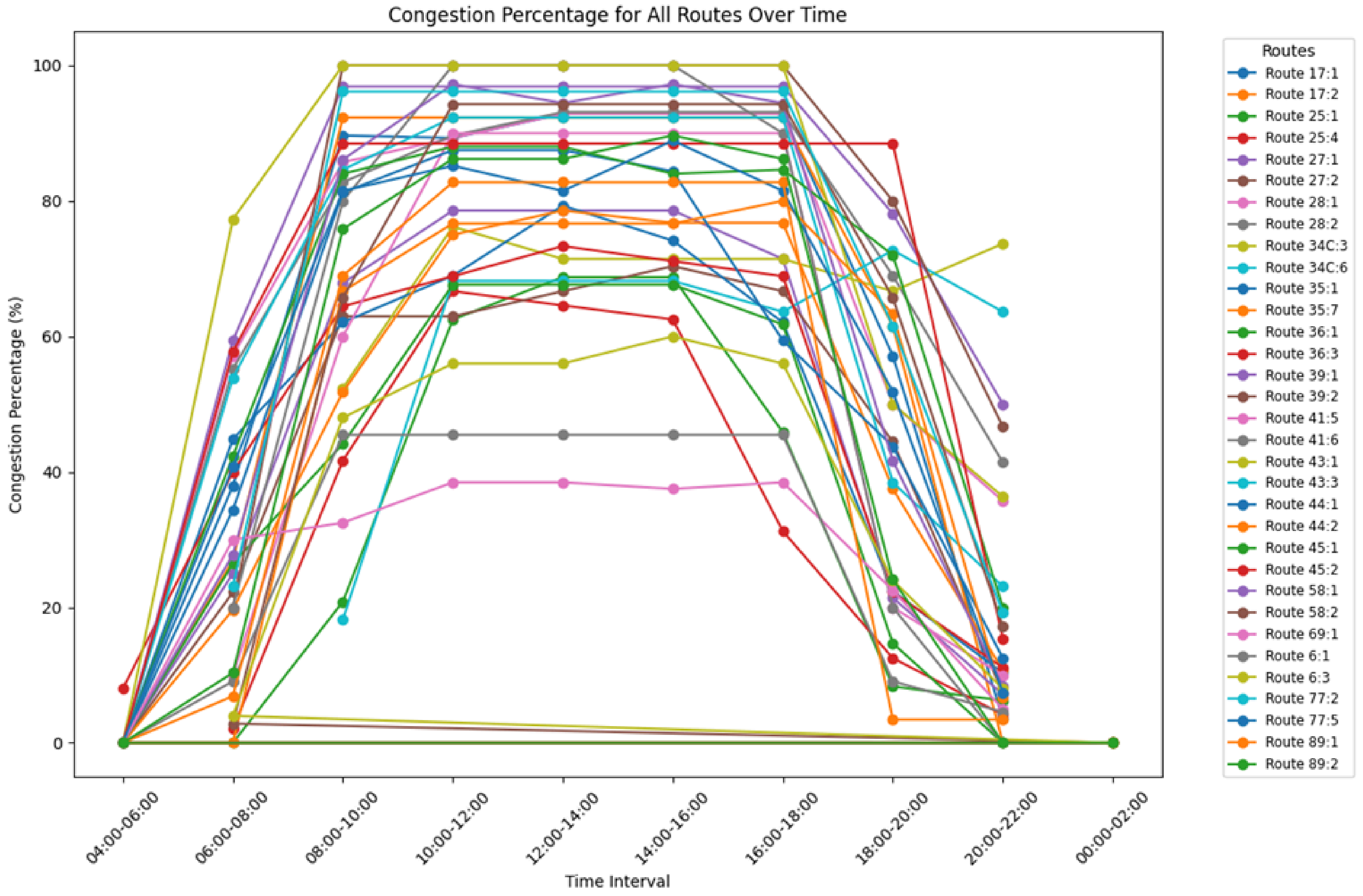
- Congestion levels vary across different time intervals during the day.
- Peak congestion occurs between 8:00 and 18:00, correlating with typical working hours.
- Off-peak hours (after 18:00) show significantly lower congestion.
- Routes like Route 11:4 show distinct congestion peaks during commuting hours (8:00 - 10:00 and 16:00 - 20:00).

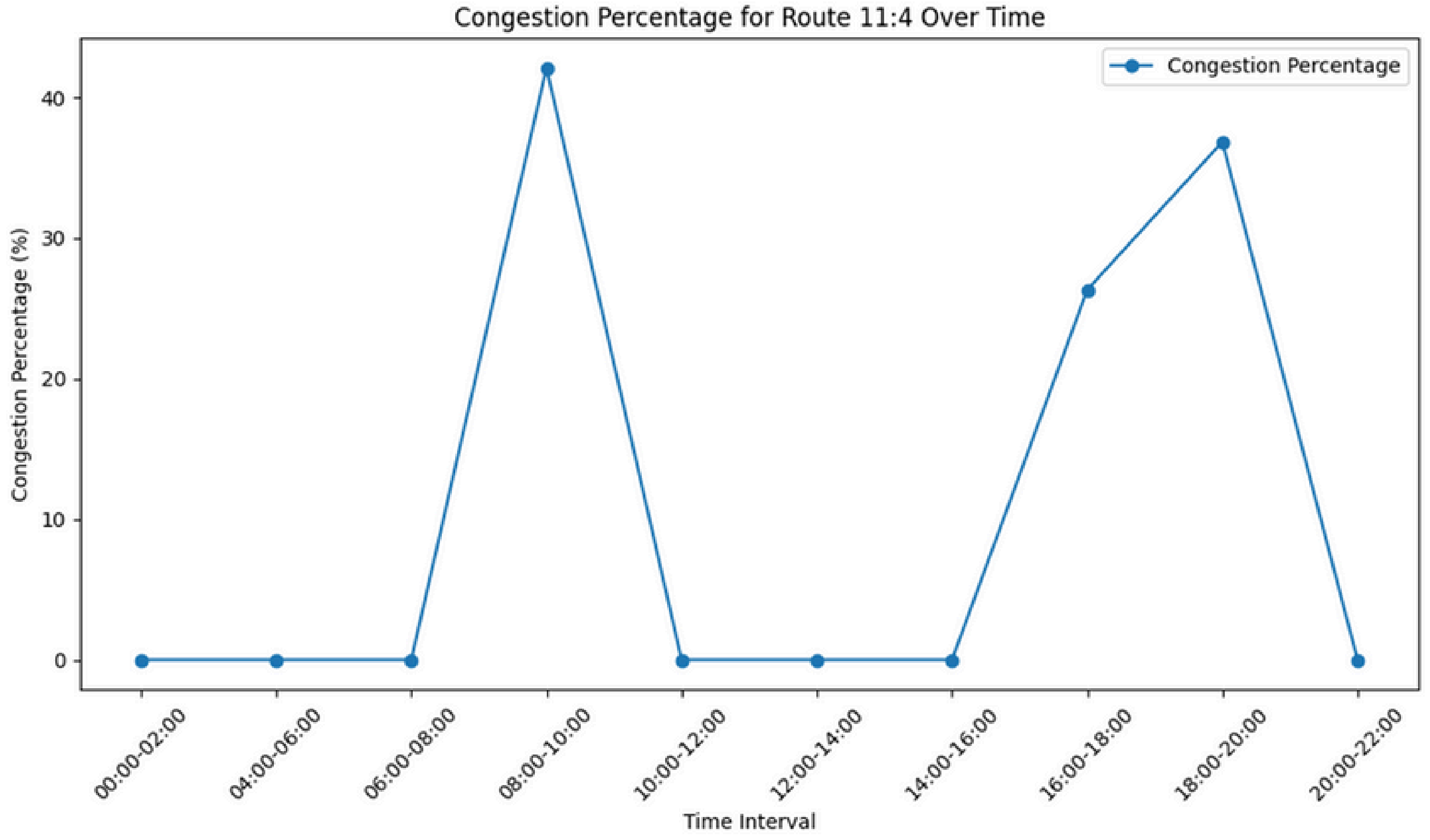


Analysis of Congested Routes

Congestion Analysis by Time Interval

- Most routes have similar congestion patterns during the day.
- Generally the routes are more congested during working hours.
- Route 11:4 experiences low congestion during most parts of the day, except during commuting hours when the congestion peaks significantly.

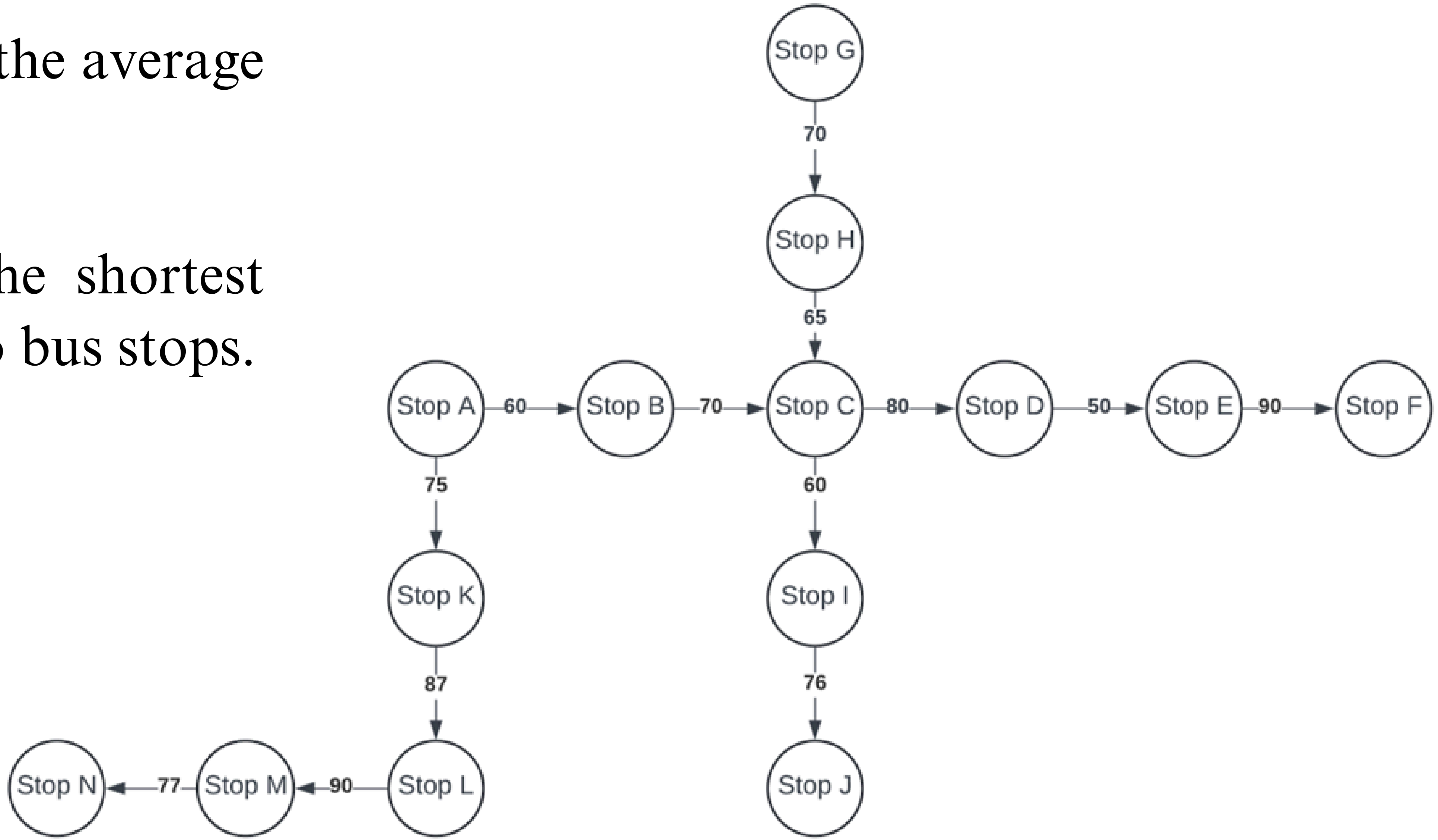




Route Optimization

Graph-Based Approach

- Created a direct graph with bus stops as nodes and route links as edges.
- The weight of each edge is determined by the average travel time between the respective stops.
- Used Dijkstra's Algorithm to identify the shortest paths (in terms of travel time) between two bus stops.



Route Optimization

Finding Alternative Routes

- Alternative routes were proposed for all the routes having congestion percentage greater than 70%.

- Route 36:1 optimization resulted in a significant travel time reduction, from 37.26 minutes to 20.4 minutes.

	Run Time (minutes)	Sequence of Stops
Main Route	37.2633	3390J4 → 3390E2 → 3390A4 → 3390Y4 → 3390CC04 → 3390LE07 → 3390LE08 → 3390LE09 → 3390LE10 → 3390LE11 → 3390LE12 → 3390QM04 → 3390UN15 → 3390UN16 → 3390UN17 → 3390UN18 → 3390UN19 → 3390UN20 → 3390UN21 → 3390UN22 → 3390UN23 → 3300BR0524 → 3300BR0523 → 3300BR0241 → 3300BR0213 → 3300BR0613 → 3300BR0095 → 3300BR0097 → 3300BR0080 → 3300BR0075 → 3300BR0503 → 3300BR0026 → 3300BR0028 → 3300BR0626 → 3300BR0197
Alternative Route	20.408	3390J4 → 3390E2 → 3390A3 → 3390Y4 → 3390CC04 → 3390LE07 → 3390LE08 → 3390LE09 → 3390LE10 → 3390LE11 → 3390LE12 → 3390QM04 → 3390UN43 → 3390UN46 → 3390UN44 → 3390UN45 → 3390UN33 → 3300BR0524 → 3300BR0523 → 3300BR0241 → 3300BR0213 → 3300BR0613 → 3300BR0095 → 3300BR0097 → 3300BR0080 → 3300BR0075 → 3300BR0503 → 3300BR0026 → 3300BR0028 → 3300BR0626 → 3300BR0197

Comparison of Main and Alternative Route for the Route 36:1

Evaluation of Alternative Routes

Run Time Comparison:

- The difference in run times between main and alternative routes serves as a critical metric for assessing the efficiency of the alternative routes.
- Percentage decrease in run time was calculated using the mean run time for both main and alternative routes.

Statistical Evaluation:

- Conducted a paired t-test to measure the significance of the improvements.
- Results: t-statistic = 3.463, p-value = 0.00174
- Conclusion: The improvements are statistically significant.

Main Route Mean Run Time(seconds)	Alternative Route Mean Run Time (seconds)	Percentage Decrease
1671.57	1456.06	12.89%

Conclusion

Congestion Analysis:

- k-Means clustering effectively classified route links into low (24.5%), medium (60.3%), and high (15.3%) congestion levels.
- Peak congestion identified during weekday working hours.
- Visual maps highlighted congestion hotspots.

Route Optimization:

- Dijkstra’s algorithm proposed alternative routes for highly congested sections (congestion > 70%).
- Resulted in a 12.89% reduction in average travel time.
- Paired t-test confirmed statistically significant improvements.

Future Work

- **Real-Time Data:** Incorporating real-time traffic data for dynamic route adjustments.
- **Expanded Factors:** Analyze weather, events, and roadworks to refine congestion insights.
- **Passenger Behavior:** Study boarding patterns to improve route efficiency.
- **New Bus Stops:** Explore introducing new stops for more optimized routes.

References

[1] Almeida, A., Brás, S., Sargento, S., and Oliveira, I. (2023) ‘Exploring bus tracking data to characterize urban traffic congestion’, Journal of Urban Mobility, 4, Art. no. 100065. Available at: https://doi.org/10.1016/j.urbmob.2023.100065.

[2] Ma, J., Chan, J., Ristanoski, G., Rajasegarar, S., and Leckie, C. (2019) ‘Bus travel time prediction with real-time traffic information’, Transportation Research Part C: Emerging Technologies, 105, pp. 536-549. Available at: https://doi.org/10.1016/j.trc.2019.06.008.

[3] Huang, K., Xu, L., Chen, Y., Cheng, Q., and An, K. (2020) ‘Customized Bus Route Optimization with Real-Time Data’, Journal of Advanced Transportation. Available at: https://www.researchgate.net/publication/343967029

[4] Afrin, T. and Yodo, N. (2020) ‘A Survey of Road Traffic Congestion Measures towards a Sustainable and Resilient Transportation System’, Sustainability, 12, p. 4660. Available at: https://doi.org/10.3390/su12114660.

[5] Lomax, T., Turner, S., Shunk, G., Levinson, H. S., Pratt, R. H., Bay, P. N., and Douglas, G. B. (1997) ‘Quantifying Congestion. Volume 1: Final Report’, NCHRP Report 398, Transportation Research Board, Washington, DC. Available at: http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_398.pdf.

[6] Diker, A. C. and Nasibov, E. (2012) ‘Estimation of traffic congestion level via FN-DBSCAN algorithm by using GPS data’, in 2012 IV International Conference "Problems of Cybernetics and Informatics" (PCI), Baku, Azerbaijan, 2012, pp. 1-4. Available at: https://doi.org/10.1109/ICPCI.2012.6486279.

Thank You