Traffic Congestion Analysis

Kandy–Katugastota Urban Corridor

Data Analysis Report

Prepared as part of the study on urban traffic behavior in the Kandy District, Sri Lanka

Prepared By: D.H.A. GOONASEKERE

Table of Figures

List of Tables

Contents

1 Introduction

In this chapter, we examined traffic behaviors and patterns in the Kandy area, as well as major causes of traffic congestion during both peak and non-peak times. We had to mix simulation (and real-time data, such as Google Maps traffic features) with our lack of digitized traffic history logs (which were not available from the Road Development Authority [RDA] of Sri Lanka) - we tried to make as much meaning as possible regarding this analysis of traffic.

The purpose of the analysis is to identify traffic patterns and anomalies, understand peak congestion times, and to understand how vehicles interact with the bottlenecks or major congestion areas of the urban environment. The information learned in this analysis section is aimed at informing the design of a rule-based, intelligent traffic management and incident response system focused on the Kandy metropolitan area.

Beginning with the study of the most significant traffic bottlenecks in Kandy, we then document our data generation process prior to describing the exploratory data analysis (EDA) and other methods we used, which includes anomaly detection, cluster analysis, and time series trends in order to obtain usable information to inform the design of a useful system.

2 Identification of Key Bottlenecks

We identified the major entry bottlenecks into the Kandy city area before simulating and analyzing traffic flows. The Kandy city region has several routes which lead into the Kandy city area from several bridges and highway designations. We note that, added with many conflicts associated with the roads used during peak hours, these roadways are subject to vehicular congestion.

Key Bottlenecks Identified:

1. Peradeniya (Peradeniya Junction) via Colombo–Kandy Road (A1)
    - A crucial entry from Colombo.
    - Heavy amounts of vehicles along the route and constant traffic congestion during peak hours.
2. Katugastota Bridge via Kurunegala–Kandy Road (A10)

- Primary major northern entry into Kandy.
- Heavy flow of vehicles from the direction of Kurunegala and Matale.

3. Tennekumbura Bridge via Mahiyanganaya–Digana–Ampitiya Road (A26)
   - Secondary northeast direction entry into Kandy from agricultural regions.
   - Heavy use of bus routes and school zone traffic in that area.
4. Southern access via Watapuluwa–Kundasale route
   - Largely undefined and is a newer access to Kandy but a growing corridor.
   - Growing suburbia housing and school traffic bundles travel.

During our preliminary site visits as part of site logging, we observed that Katugastota Bridge and Peradeniya Junction were the most photographed along with evidence of most traffic congestion. Based upon these combinations of characteristics we narrowed our focus for detailed traffic flow simulation and traffic performance analysis to Katugastota Bridge, which is evidence of complex merging and strategic bottlenecks location.

Figure 1: Satellite view of the Katugasthota region highlighting the key traffic bottleneck

Figure 2: Normal view of the Katugasthota region highlighting the key traffic bottleneck

3 Bottleneck Overview

3.1 Bottleneck 1: Katugastota Entry (North-East Confluence)

- Main Entry Point A
- Routes:
    - ❖ Kandy–Matale–Jaffna Road (two-way, high volume)
    - ❖ Katugastota–Madawela–Bambarella–Digana Road (two-way, high volume)
- Sub-entry Routes:
    - ❖ Katugastota–Madawala Road (low traffic except peak hours)
    - ❖ Katugastota–Market Road
- Junction Type: Four-lane merge (2 entry, 2 exit)
- Complexity: Sub-road vehicles must navigate exit lanes to reach the main entry route

3.2 Bottleneck 2: Southern Merge Zone

- Routes:

    - ❖ Kurunegala–Kandy Main Road (two-way, high volume)
    - ❖ Peradeniya–Haloluwa Road (two-way, high volume)
- Junction Type: Four-lane structure (2 entry, 2 exit)
- Complexity: Vehicles from different directions merge into narrow central lanes near the town center

3.3 Main Bottleneck: Katugastota Junction

- Function: Central merging point of both Bottlenecks 1 and 2
- Features:
    - ❖ High-volume vehicle interaction zone
    - ❖ Complex merging and exit pattern
    - ❖ Exit vehicles from Katugastota combine with outbound traffic from both bottlenecks
- Impact: Primary source of congestion, especially during peak hours

4 Sample of Simulated Traffic Dataset

The traffic analysis for this study is based on a comprehensive simulated dataset generated for three years, from January 1, 2022, to December 30, 2024. This dataset forms the foundation for identifying and analyzing traffic bottlenecks around the Katugasthota region.

Dataset Overview:

- Total Records: 338,400
- Features (Columns): 21
- Date Range: 2022-01-01 to 2024-12-30
- Interval: Hourly data for selected days (2–3 days per week)
- Granularity: Captures route-level data, vehicle types, speeds, multipliers, and contextual flags (e.g., holidays, school days)
- Detected Anomalies (Z-score method): 5,814

Key Attributes:

The dataset includes the following critical fields:

- Timestamp: Date and time of the entry
- Bottleneck: Identifier of the bottleneck area (e.g., Bottleneck 1, Bottleneck 2)
- Route: Road segment or approach name
- Vehicle Type: Car, bus, lorry, motorcycle, etc.
- Volume: Number of vehicles recorded in that segment and time
- Traffic Intensity: Categorized as Low, Moderate, or High
- Average Speed: Calculated in km/h
- Speed Status: Slow, Normal, Fast
- Modifiers: Flags for school days, holidays, Sundays, and working days

Timestamp	Bottleneck	Entry Point	Route	Vehicle Type	Traffic Multiplier
2022-01-01 07:00:00	Bottleneck 1	Main Entry A	Kandy-Matale-Jaffna Road	Car	1.6
2022-01-01 07:00:00	Bottleneck 1	Main Entry A	Kandy-Matale-Jaffna Road	Truck	1.6
2022-01-01 07:00:00	Bottleneck 1	Main Entry A	Katugastota-Madawela-Bambarella	Car	1.6
2022-01-01 07:00:00	Bottleneck 1	Main Entry A	Katugastota-Madawela-Bambarella	Bus	1.6
2022-01-01 07:00:00	Bottleneck 1	Main Entry A	Katugastota-Madawela-Bambarella	Truck	1.6
2022-01-01 07:00:00	Bottleneck 1	Main Entry A	Katugastota-Madawela-Bambarella	Three-wheeler	1.6
2022-01-01 07:00:00	Bottleneck 2	Main Entry B	Kurunegala-Kandy Main Road	Motorcycle	1.6
2022-01-01 08:00:00	Bottleneck 1	Main Entry A	Kandy-Matale-Jaffna Road	Three-wheeler	1.5
2022-01-01 08:00:00	Bottleneck 1	Main Entry A	Katugastota-Madawela-Bambarella	Truck	1.5
2022-01-01 08:00:00	Bottleneck 1	Sub Entry A	Katugastota-Madawala Road	Three-wheeler	1.5
2022-01-01 08:00:00	Bottleneck 1	Sub Entry B	Katugastota-Market Road	Three-wheeler	1.5
2022-01-01 08:00:00	Bottleneck 2	Main Entry B	Kurunegala-Kandy Main Road	Motorcycle	1.5

Figure 3:Sample of Simulated Traffic Dataset 1

Traffic Multiplier	Vehicle Volume	Traffic Intensity	Avg Speed (km/h)	Speed Status	Is School Day
1.6	182	Moderate	27.2	Sluggish	TRUE
1.6	176	Moderate	12.2	Heavy Congestion	TRUE
1.6	180	Moderate	34.6	Smooth	TRUE
1.6	176	Moderate	15.4	Heavy Congestion	TRUE
1.6	176	Moderate	12.2	Heavy Congestion	TRUE
1.6	174	Moderate	28.3	Sluggish	TRUE
1.6	190	Moderate	43.6	Smooth	TRUE
1.5	177	Moderate	23.3	Sluggish	TRUE
1.5	174	Moderate	14	Heavy Congestion	TRUE
1.5	177	Moderate	23.6	Sluggish	TRUE
1.5	180	Moderate	22.5	Sluggish	TRUE
1.5	171	Moderate	44.2	Smooth	TRUE

Figure 4:Sample of Simulated Traffic Dataset (2)

Route	Vehicle Type	Traffic Multiplier	Vehicle Volume	Traffic Intensity	Avg Speed (km/h)	Speed Status
Kandy-Matale-Jaffna Road	Car	1.6	182	Moderate	27.2	Sluggish
Kandy-Matale-Jaffna Road	Truck	1.6	176	Moderate	12.2	Heavy Congestion
Katugastota-Madawela-Bambarella	Car	1.6	180	Moderate	34.6	Smooth
Katugastota-Madawela-Bambarella	Bus	1.6	176	Moderate	15.4	Heavy Congestion
Katugastota-Madawela-Bambarella	Truck	1.6	176	Moderate	12.2	Heavy Congestion
Katugastota-Madawela-Bambarella	Three-wheeler	1.6	174	Moderate	28.3	Sluggish
Kurunegala-Kandy Main Road	Motorcycle	1.6	190	Moderate	43.6	Smooth
Kandy-Matale-Jaffna Road	Three-wheeler	1.5	177	Moderate	23.3	Sluggish
Katugastota-Madawela-Bambarella	Truck	1.5	174	Moderate	14	Heavy Congestion
Katugastota-Madawala Road	Three-wheeler	1.5	177	Moderate	23.6	Sluggish
Katugastota-Market Road	Three-wheeler	1.5	180	Moderate	22.5	Sluggish
Kurunegala-Kandy Main Road	Motorcycle	1.5	171	Moderate	44.2	Smooth
Kandy-Matale-Jaffna Road	Motorcycle	1.4	168	Moderate	41.7	Smooth
Kandy-Matale-Jaffna Road	Motorcycle	1.5	178	Moderate	32.5	Smooth
Katugastota-Madawala Road	Three-wheeler	1.5	175	Moderate	31.1	Smooth
Katugastota-Market Road	Bus	1.5	169	Moderate	15.8	Heavy Congestion
Katugastota-Market Road	Truck	1.5	178	Moderate	19.9	Heavy Congestion
Katugastota-Market Road	Three-wheeler	1.5	169	Moderate	20.4	Sluggish

Figure 5:Sample of Simulated Traffic Dataset (3)

h)	Speed Status	Is School Day	Is Holiday	Is Sunday	Is Working Day	Hour	Day	Month	Year	Volume Z-Score
27.2	Sluggish	TRUE	FALSE	FALSE	TRUE	7	Saturday	January	2022	3.466494771
12.2	Heavy Congestion	TRUE	FALSE	FALSE	TRUE	7	Saturday	January	2022	3.284148679
34.6	Smooth	TRUE	FALSE	FALSE	TRUE	7	Saturday	January	2022	3.40571274
15.4	Heavy Congestion	TRUE	FALSE	FALSE	TRUE	7	Saturday	January	2022	3.284148679
12.2	Heavy Congestion	TRUE	FALSE	FALSE	TRUE	7	Saturday	January	2022	3.284148679
28.3	Sluggish	TRUE	FALSE	FALSE	TRUE	7	Saturday	January	2022	3.223366648
43.6	Smooth	TRUE	FALSE	FALSE	TRUE	7	Saturday	January	2022	3.709622894
23.3	Sluggish	TRUE	FALSE	FALSE	TRUE	8	Saturday	January	2022	3.314539694
14	Heavy Congestion	TRUE	FALSE	FALSE	TRUE	8	Saturday	January	2022	3.223366648
23.6	Sluggish	TRUE	FALSE	FALSE	TRUE	8	Saturday	January	2022	3.314539694
22.5	Sluggish	TRUE	FALSE	FALSE	TRUE	8	Saturday	January	2022	3.40571274
44.2	Smooth	TRUE	FALSE	FALSE	TRUE	8	Saturday	January	2022	3.132193602
41.7	Smooth	TRUE	FALSE	FALSE	TRUE	13	Saturday	January	2022	3.041020556
32.5	Smooth	TRUE	FALSE	FALSE	TRUE	16	Saturday	January	2022	3.34493071
31.1	Smooth	TRUE	FALSE	FALSE	TRUE	16	Saturday	January	2022	3.253757663
15.8	Heavy Congestion	TRUE	FALSE	FALSE	TRUE	16	Saturday	January	2022	3.071411571
19.9	Heavy Congestion	TRUE	FALSE	FALSE	TRUE	16	Saturday	January	2022	3.34493071
20.4	Sluggish	TRUE	FALSE	FALSE	TRUE	16	Saturday	January	2022	3.071411571

Figure 6:Sample of Simulated Traffic Dataset (4)

5 Traffic Pattern Visualization

5.1 Heatmaps and Calendar Heatmaps

To better understand temporal traffic patterns, a calendar heatmap was generated to visualize total vehicle volume across days of the week and weeks of the year. This visualization helps identify recurring congestion trends and seasonal traffic fluctuations.

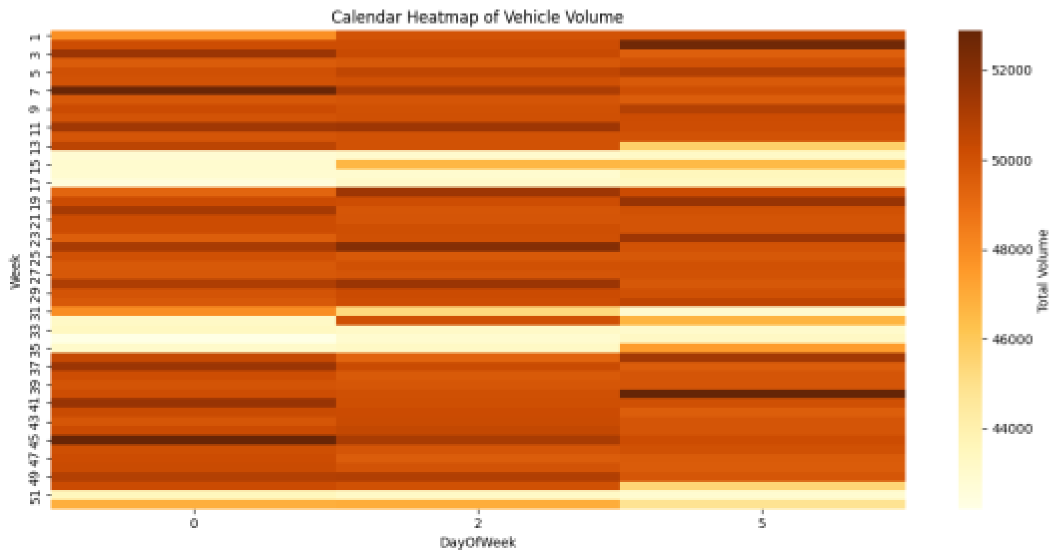

Figure 7: Calendar Heatmap of Vehicle Volume

- X-axis: Days of the week (0 = Monday, 6 = Sunday)
- Y-axis: Week numbers (1 to 52)
- Color Gradient: Represents total vehicle volume, ranging from light yellow (lower volume) to dark brown (higher volume)

This heatmap reveals:

- Consistent high traffic volumes during weekdays, especially mid-week.
- Noticeable dips in volume during weekends and public holidays.
- Seasonal peaks, likely corresponding to school terms or festive periods.

Such insights are crucial for planning dynamic traffic signal timings and scheduling road maintenance during low-traffic periods.

5.2 Monthly Calendar Heatmap

To also investigate intra-month traffic behaviors, a monthly calendar heatmap was developed. This heatmap presents vehicle volumes daily over the first 12 weeks of the year (see Figure below). It provides a detailed visualization of how traffic volume varies intra-monthly across the testing period.

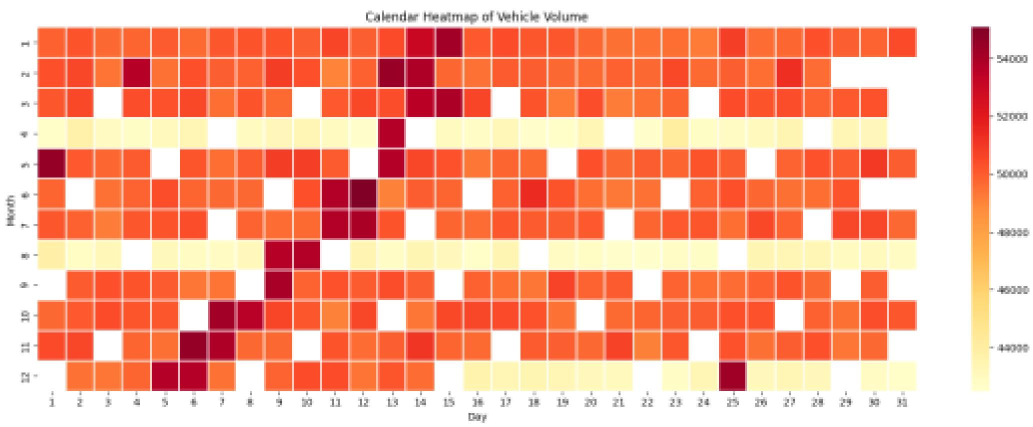

Figure 8:Monthly Calendar Heatmap of Vehicle Volume

- X-axis: Days of the month (1 to 31)
- Y-axis: Week numbers (1 to 12)
- Color Gradient: Ranges from light yellow (low volume) to dark red (high volume)

Insights:

- Traffic volumes appear to peak around the middle of the month. This may be connected to salary cycles, school timetables or market days.
- The amount of traffic appears less at the beginning and the end of the month. This corresponds to public holiday periods and also to periods with lower levels of activity.
- These visualizations can help in identifying suitable opportunities for road maintenance or traffic rerouting.

5.3 Correlation Heatmap of Contextual Features

In order to visualize the relationships between important traffic variables and contextual features a correlation heatmap was generated. The correlation heatmap visualizes the strength and direction of linear relationships between traffic variables like vehicle volume, average speed, traffic multipliers and contextual flags i.e. holidays or school days.

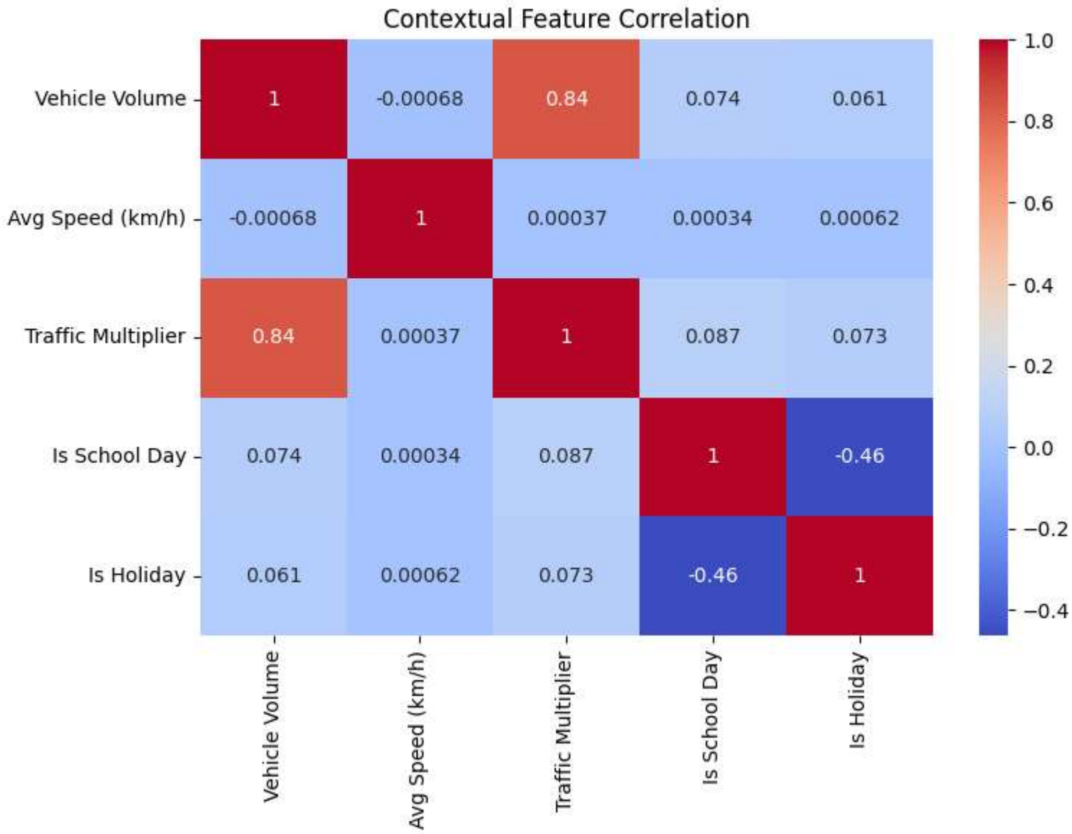

Figure 9:Contextual Feature Correlation Heatmap

- Color gradient: blue indicates a negative correlation, red indicates a positive correlation, and the intensity of the color indicates the amount of correlation from -1 to +1.
- Key findings:
  - ❖ Vehicle Volume has a strong positive correlation with Traffic Multiplier (0.84), meaning that as vehicle volume increases, volume will scale correctly and as expected based on contextual multipliers.
  - ❖ There are weak or negligible correlations between Vehicle Volume and Avg Speed as a whole (overall), suggesting that vehicle volume alone does not dictate speed.
  - ❖ School Days and Holidays each had small, positive correlations with Vehicle Volume and Traffic Multiplier, indicating possible shifts of human patterns during only school or holidays.

This heat map provides valuable information to improve traffic behaviour prediction models by demonstrating which contextual variables most strongly affect traffic behaviour.

5.4 Correlation Matrix of Core Traffic Variables

To investigate the potential relationships between core traffic metrics for the traffic patterned examined, a correlation matrix heatmap was created. A correlation matrix is a great first step to determine how variable such vehicle volume vs. average speed, traffic multipliers, and time of day/hour are linearly correlated and how strongly they are related.

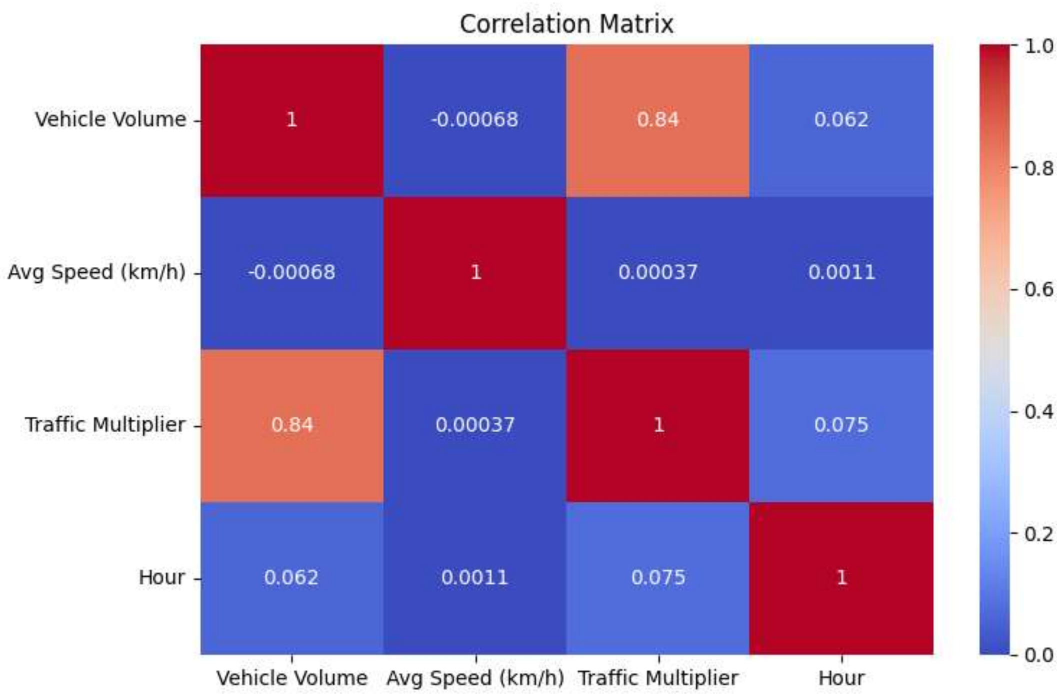

Figure 10:Correlation Matrix of Core Traffic Variables

- Colour Gradient: Ranging from blue (negative correlation) to red (positive correlation), with values from -1 to +1.
- Key Findings:
  - ❖ Vehicle Volume and Traffic Multiplier demonstrate a strong positive correlation (0.84), meaning that contextual multipliers (i.e. school days, holidays) greatly affect traffic volume.
  - ❖ A weak positive correlation (0.062) was revealed between Volume and Hour, indicating that hour does have a small effect on volume.
  - ❖ Average Speed has very little correlation with the other attributes. This suggests that average speed may be impacted by other more complex or external influences that are not included in this matrix.

5.5 Box Plot: Vehicle Volume on Holidays vs. Working Days

To assess the role of context in calendar policy around vehicle volume and traffic behaviour, a box plot was produced that identified comparatives between vehicle volumes on holidays and working days. This visualization provides a quantified difference in flow for amount of vehicular traffic for holidays and working days.

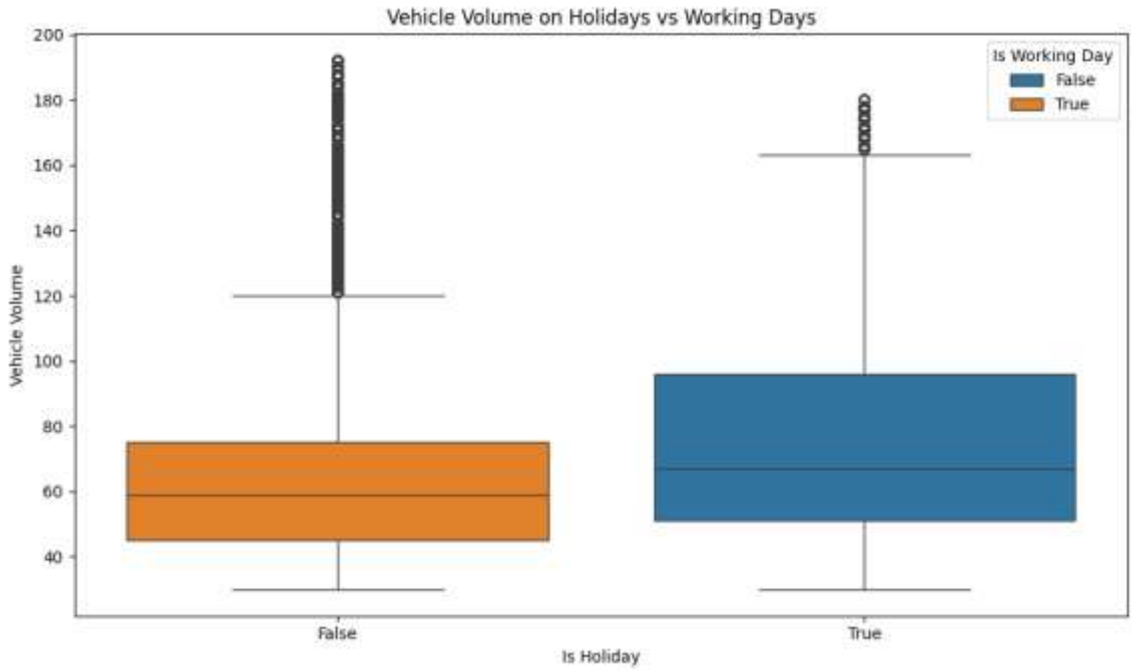

Figure 11:Box Plot of Vehicle Volume by Holiday Status

- X-axis: holiday status (True = Holiday, False = Non-Holiday)
- Y-axis: Vehicle Volume
- Color Legend: Working day status
  - ❖ Orange: Non-working days
  - ❖ Blue: Working days

Insights as follows:

- Working days have a higher median vehicle volume, a greater interquartile range, or more variation - meaning more consistent and heavier volumes.
- Holidays have lower median volumes, and less extreme values - thus suggests less travel activity.

- This comparison gives credit to the contextual flags/historical stats (holidays) as a potential factor for predictive modeling of traffic, as well as for the real-time traffic management logic.

5.6 K-Means Clustering: Volume vs. Speed

K-means clustering was used to discover natural clusters or groups within traffic behaviour, using vehicle volume and average speed as input features. K-means is an unsupervised learning technique and can help separate traffic conditions into differing congestion indicators which may help in determining whether these traffic patterns need to be handled differently.

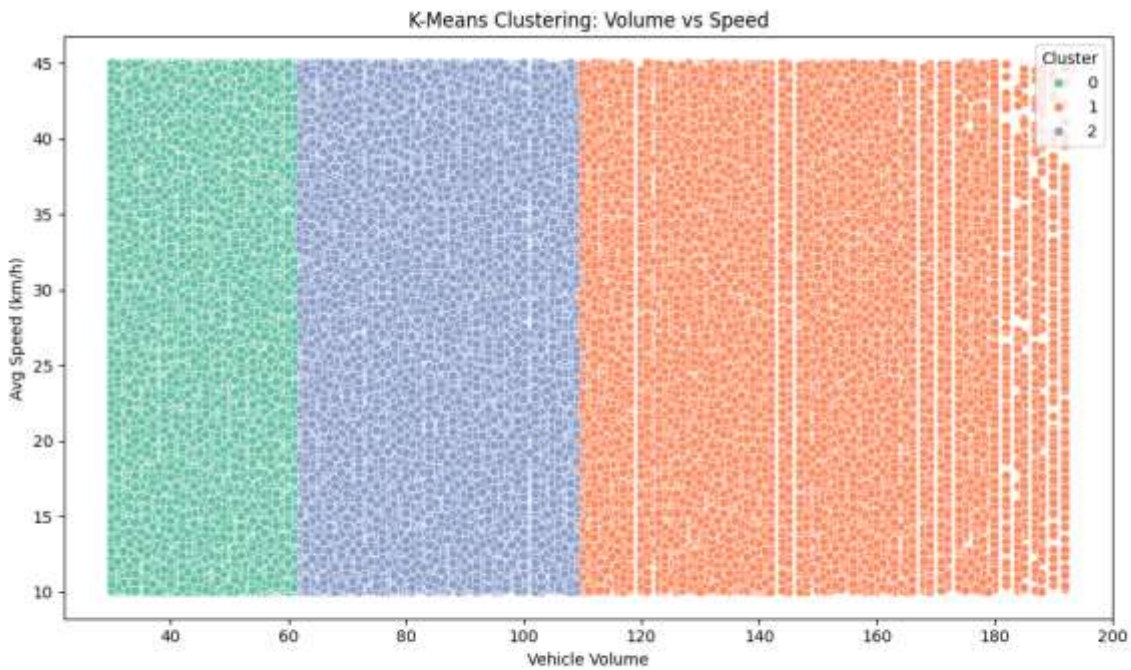

Figure 12:K-Means Clustering of Vehicle Volume vs. Average Speed

- X-axis: Vehicle Volume (0 to 200)
- Y-axis: Speed (km/h) (0 to 50)
- Cluster Labels:
  - ❖ Cluster 0 (Green): low-volume, high-speed; a freely flowing route or an off-peak period of driving.
  - ❖ Cluster 1 (Blue): moderate volume and speed; a likely transitional traffic condition.
  - ❖ Cluster 2 (Orange): indicative of high-volume, low-speed; may indicate congestion or peak-hour conditions.

Insights:

- The clustering provides a clear means of assessing behaviour in traffic flow that can support signal timings, route diversion, or delay estimation related to congestion.
- The cluster variables may also be used as tags for supervised model training or real-time traffic state classification.

5.7 Pair Plot: Multivariate Relationship Exploration

To further explore the pairwise relationships of key traffic variables, a pair plot (i.e. scatterplot matrix) was created. This format is useful as it provides a small sample of possible relationships between these variables while also looking for possible correlations or clusters.

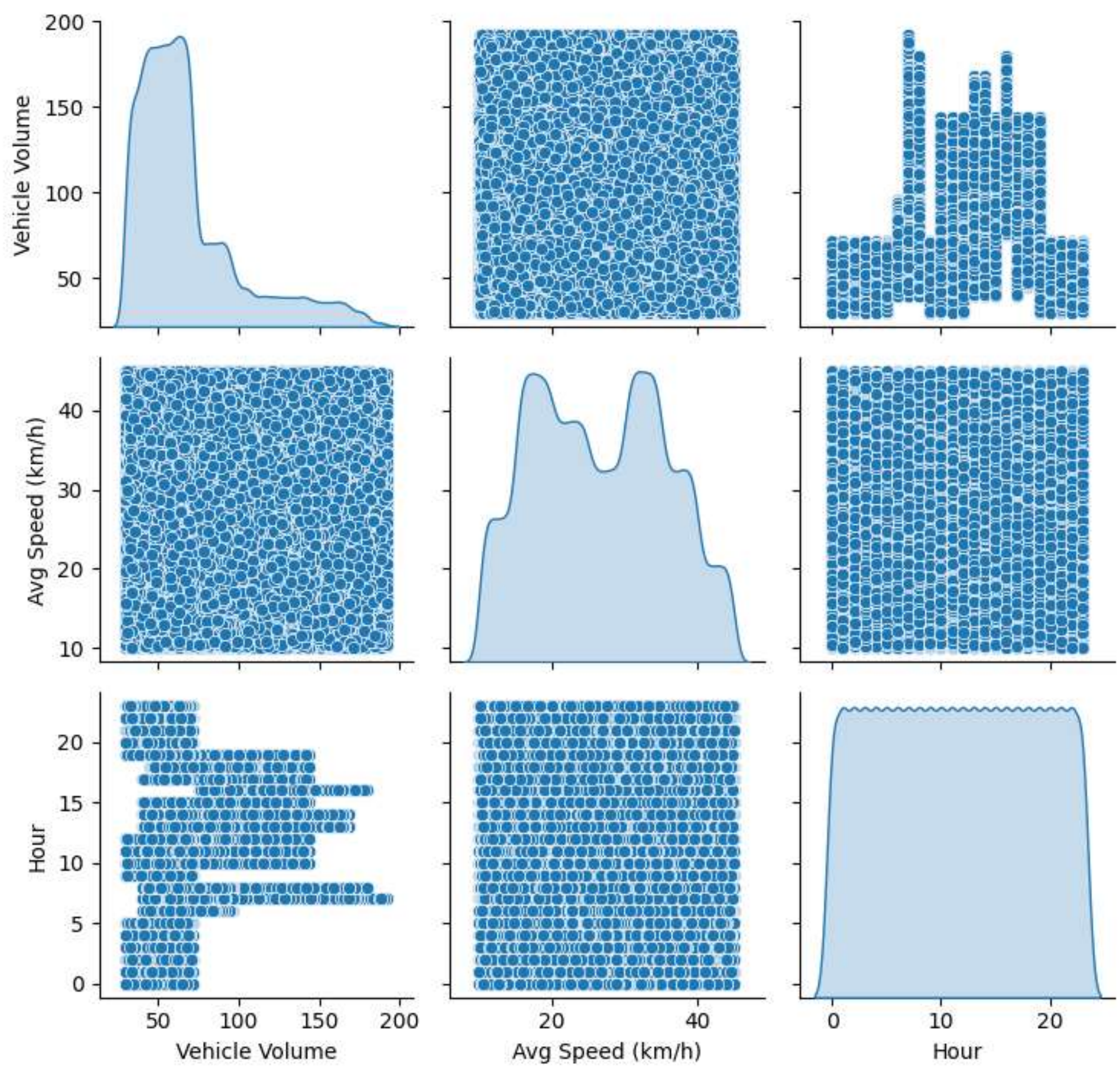

Figure 13:Pair Plot of Vehicle Volume, Average Speed, and Hour

- Diagonal Cells: Histograms showing the distribution of every variable.
- Off-Diagonal Cells: Scatter plots showing each of the pairwise relationships between:
    - ❖ Vehicle Volume and Avg Speed
    - ❖ Vehicle Volume and Hour
    - ❖ Avg Speed and Hour

Contrary to this, we gather the following insights:

- The Vehicle Volume vs Hour plot is shown that there is a slight level of concentration of traffic volume during distinctly specified hours showing peak periods.
- The Vehicle Volume vs Avg Speed plot shows a scattered distribution suggesting that speed is not linearly dependent on volume.
- The histograms illustrate that while both vehicle volume and hour have noticeable peaks, average speed is relatively uniform distributed.

Overall, this plot is useful for identifying trends, clusters, and maybe outliers in multivariate traffic data to support more advance statistical modeling.

5.8 Heatmap: Average Speed by Hour and Day

Heatmap: Average Speed by Hour by Day-A heatmap representing the average speed within each hour of every day was generated in order to evaluate any time based patterns in traffic speed. Heatmaps can help provide insights into the optimal times when traffic volume flowed quickly or slowly - which is important for optimizing signal timing and evaluating route planning.

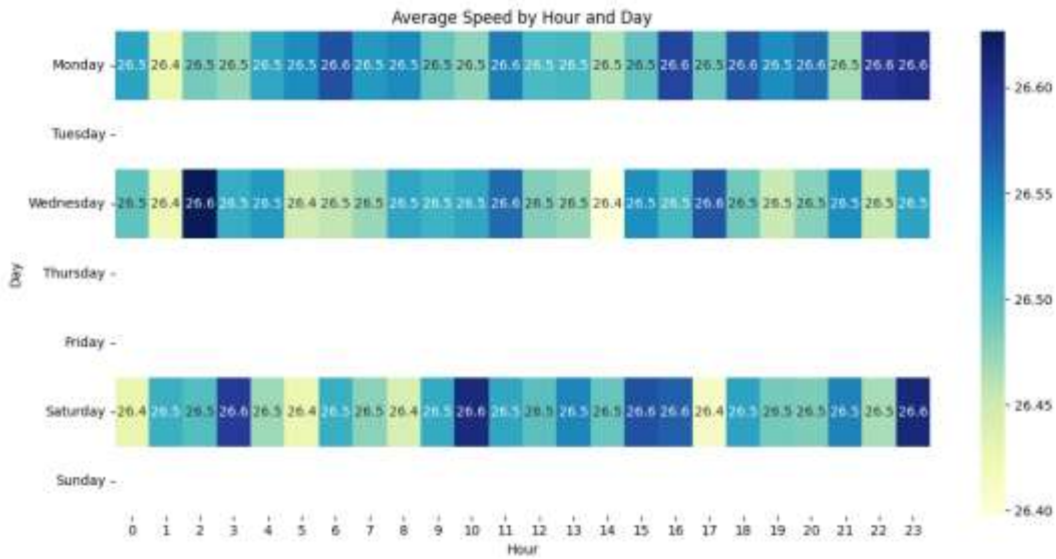

Figure 14:Heatmap of Average Speed by Hour and Day

This heatmap focused on:

- X-axis: Hours of the day (0-23)
- Y-axis: Weekday days (Monday-Sunday)
- Color Gradient
    - ❖ Light Yellow: Lower Average Speeds (~ 26.40 km/h)
    - ❖ Dark Blue: Higher Average Speeds (~ 26.60 km/h)

Conclusions:

- There are small differences in speed but they are consistent. Average speeds are slightly lower during morning and evening peak times of the weekday periods.
- Midday and late night hour have slightly higher average speed where traffic volume does not have a significant effect.
- Weekend days have a more uniform speed distribution, indicating lower amounts of congestion and traffic disruptions.

This heat map complements other congestion information, allowing one to identify higher periods of congestion, and manipulating change based on speed limits or traffic signal coordination.

5.9 Scatter Plot: Speed vs. Volume by Traffic Intensity

To identify how traffic volume impacted average speed, based on varying traffic conditions, a scatter plot is provided. The scatter plot contains traffic intensity information identified by category. The scatter plot provides insight into how congestion can impact traffic movement and average speeds.

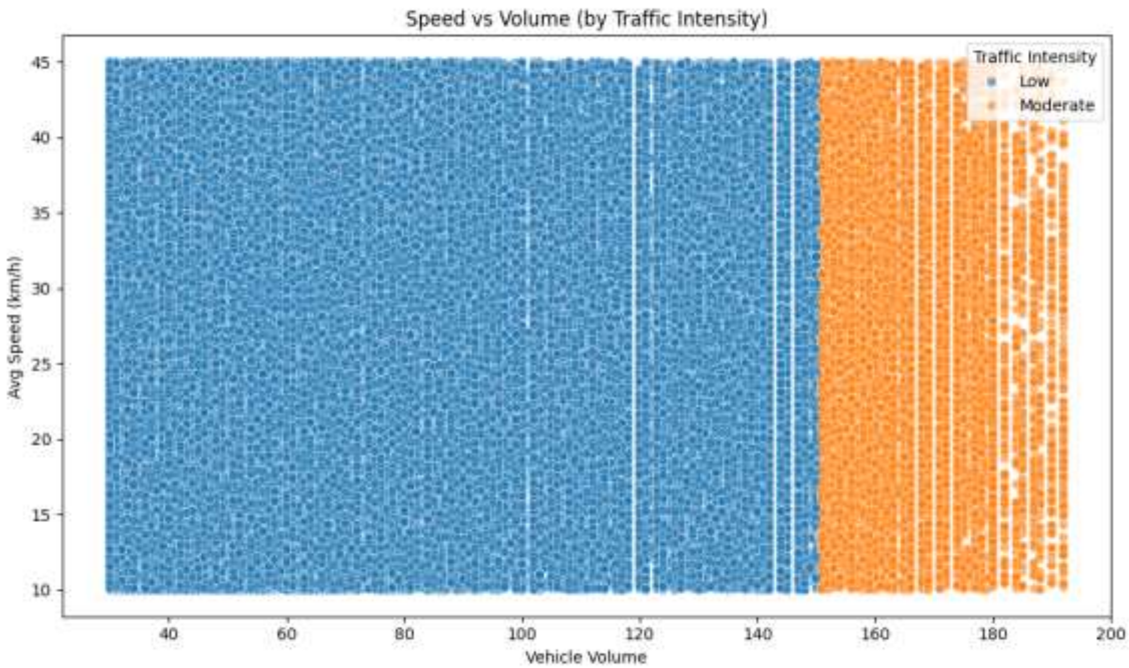

Figure 15:Scatter Plot of Speed vs. Volume by Traffic Intensity

- X-axis: Vehicle Volume (0–200)
- Y-axis: Average Speed (km/h) (0–50)
- Color Coding:
    - Blue: Low traffic intensity
    - Orange: Moderate traffic intensity

Insights:

- Under low traffic intensity, vehicles maintain a wide range of speeds, indicating free-flowing conditions.
- As volume increases, especially under moderate intensity, average speeds tend to cluster in the lower range, suggesting congestion effects.
- The inverse relationship between volume and speed becomes more apparent in moderate traffic, supporting the need for dynamic traffic control measures during peak periods.

This plot is instrumental in validating the system's logic for congestion detection and route optimization.

5.10 Time Series Decomposition: Vehicle Volume

To uncover underlying patterns in traffic volume over time, a time series decomposition was performed. This technique separates the original signal into three components—trend, seasonality, and residuals providing a clearer understanding of long-term behavior and short-term fluctuations.

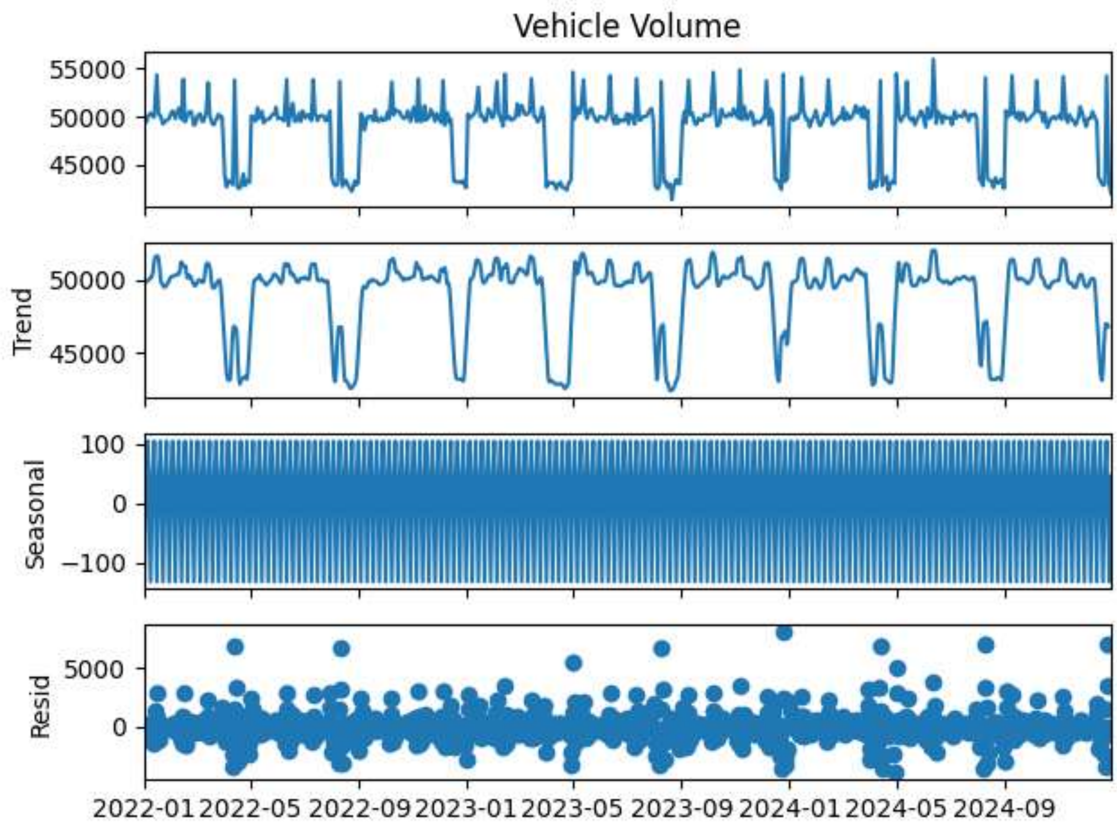

Figure 16: Time Series Decomposition of Vehicle Volume (2022–2024)

- Top Plot (Vehicle Volume): The original time series showing daily or weekly vehicle volume.
- Trend Component: A smoothed version of the data revealing the long-term direction of traffic volume. It shows a gradual increase over the years, indicating growing traffic demand.
- Seasonal Component: Captures repeating patterns within the data, such as weekly or monthly cycles. The seasonal variation is relatively small (±100), suggesting consistent but subtle periodic fluctuations.

- Residual Component: Represents irregularities or noise not explained by the trend or seasonality. These spikes may correspond to anomalies like road closures, events, or weather disruptions.

Insights:

- The upward trend supports the need for scalable traffic infrastructure.
- Seasonal patterns can inform scheduling of maintenance or public transport adjustments.
- Residuals highlight the importance of real-time monitoring for unexpected traffic surges.

5.11 Time Series Residual Analysis

To further investigate irregularities in traffic volume data, a residual plot was generated as part of the time series decomposition. Residuals represent the portion of the data not explained by the trend or seasonal components, highlighting anomalies and unpredictable fluctuations.

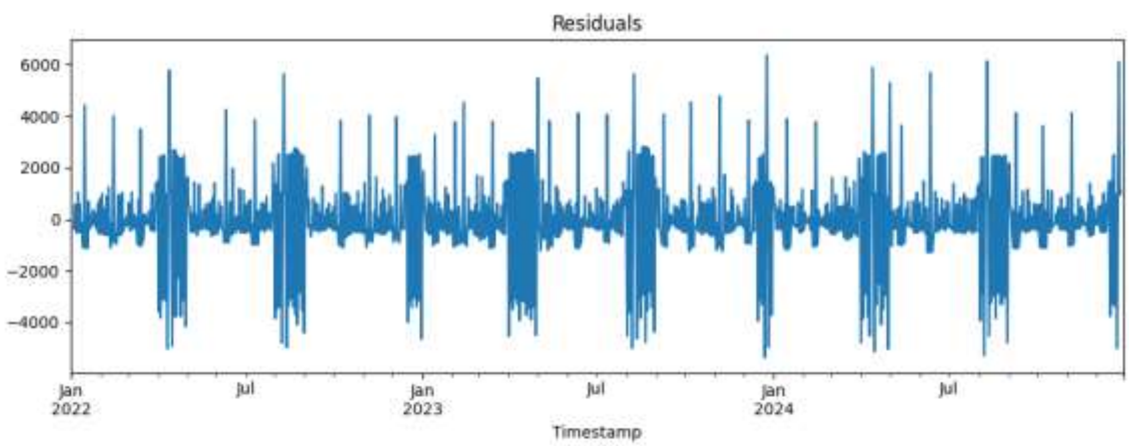

Figure 17:Time Series Plot of Residuals (2022–2025)

- X-axis: Timestamp (January 2022 to January 2025)
- Y-axis: Residual values (ranging from -6000 to +6000)
- Title: "Residuals"

Insights:

- The residuals fluctuate around zero, as expected, but several spikes indicate significant deviations from expected traffic patterns.

- These deviations may correspond to unplanned events such as road closures, accidents, weather disruptions, or public gatherings.
- Monitoring residuals in real time can enhance anomaly detection and support dynamic traffic response strategies.

This analysis reinforces the importance of integrating real-time data feeds and anomaly detection mechanisms into the traffic management system.

5.12 Seasonal Component Analysis

As part of the time series decomposition, the seasonal component was extracted to isolate recurring patterns in traffic volume across the observed period. This component helps identify periodic fluctuations that repeat over consistent intervals, such as weekly or monthly cycles.

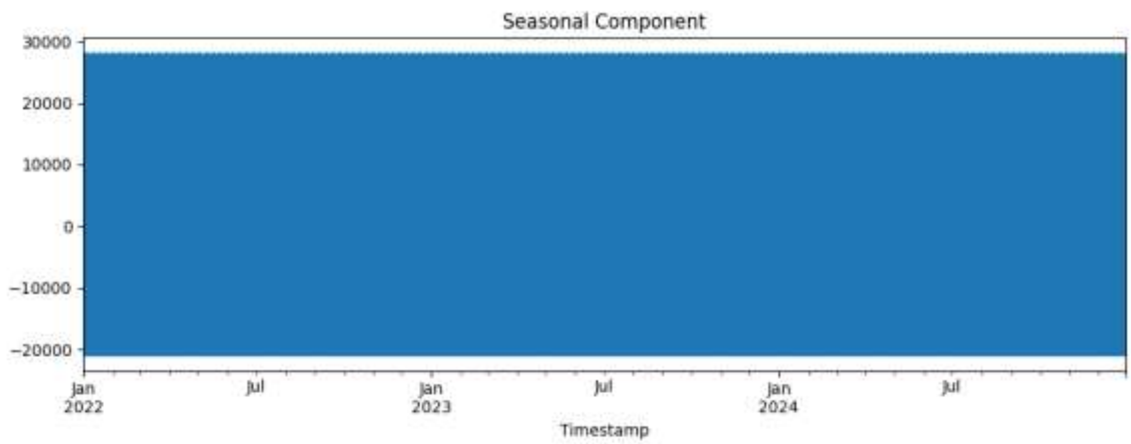

Figure 18:Seasonal Component of Vehicle Volume (2022–2025)

- X-axis: Timestamp (January 2022 to January 2025)
- Y-axis: Seasonal variation in vehicle volume (ranging from -20,000 to +30,000)
- Graph Style: Solid blue fill indicating dense, overlapping data points

Insights:

- The seasonal component reveals consistent cyclical behavior in traffic volume, likely corresponding to weekly work schedules, school terms, or monthly economic activity.
- Peaks and troughs repeat at regular intervals, validating the presence of predictable traffic surges and lulls.

- Understanding these patterns enables proactive traffic management, such as adjusting signal timings or deploying resources during expected high-volume periods.

This analysis supports the integration of seasonality-aware forecasting models into the traffic analytics system.

5.13 Trend Component Analysis

As part of the time series decomposition, the trend component was isolated to reveal the long-term progression of traffic volume over the three-year period. This component smooths out short-term fluctuations to highlight the underlying direction of change.

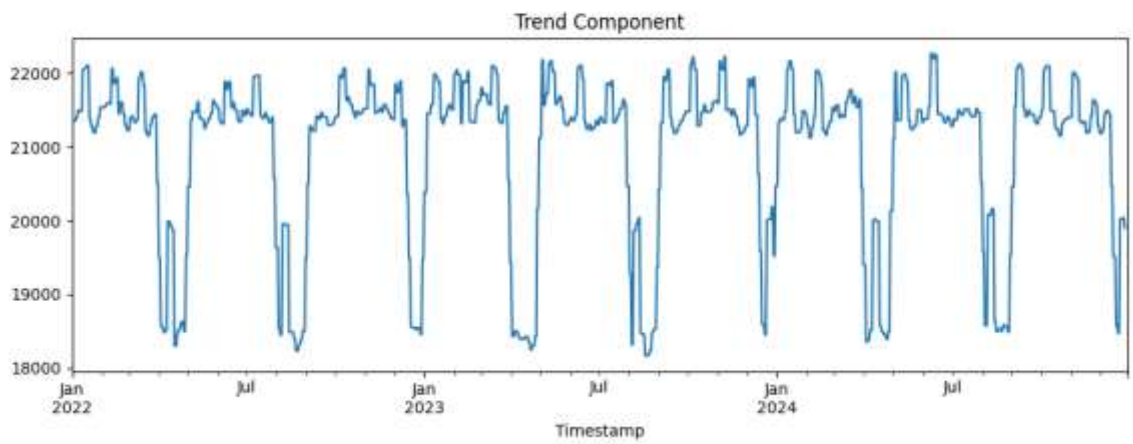

Figure 19:Trend Component of Vehicle Volume (2022–2025)

- X-axis: Timestamp (January 2022 to January 2025)
- Y-axis: Smoothed vehicle volume (ranging from 18,000 to 22,000)
- Graph Type: Line plot showing gradual changes over time

Insights:

- The trend line exhibits a cyclical rise and fall, with periodic peaks and troughs suggesting seasonal or policy-driven influences.
- A general upward trajectory is observed, indicating increasing traffic volume over the years—likely due to urban expansion, population growth, or increased vehicle ownership.
- These insights support the need for scalable infrastructure and adaptive traffic control systems that can accommodate long-term growth.

This trend analysis is essential for forecasting future traffic loads and planning sustainable urban mobility strategies.

5.14 Anomaly Detection in Vehicle Volume

To identify unusual traffic patterns, a time series anomaly detection analysis was conducted using statistical thresholds. This visualization highlights deviations in vehicle volume that fall significantly outside the expected range.

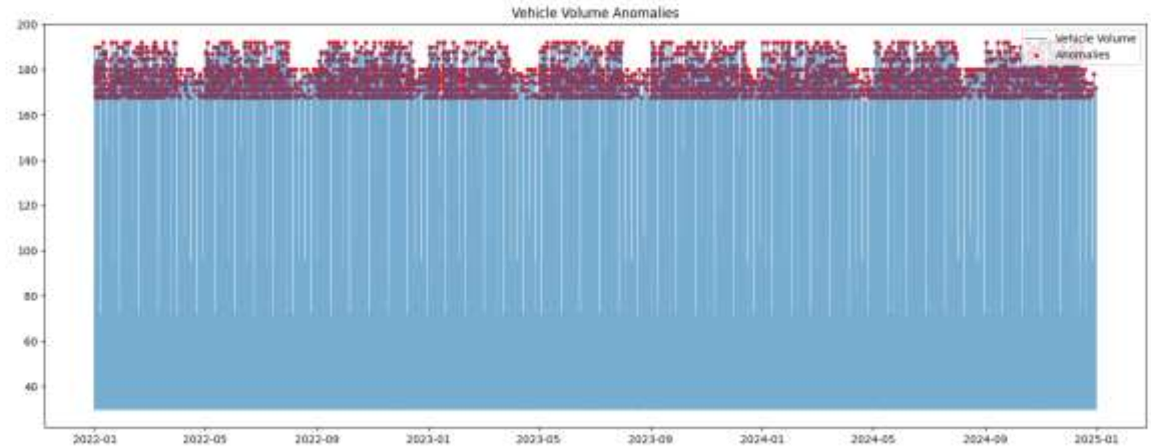

Figure 20:Vehicle Volume Anomalies (2022–2024)

- X-axis: Date range from January 2022 to January 2024
- Y-axis: Vehicle Volume (0–200)
- Data Series:
    - ❖ Blue Line: Normal vehicle volume over time
    - ❖ Red Dots: Detected anomalies based on Z-score thresholds

Insights:

- Anomalies are scattered across the timeline, with clusters appearing during certain months—possibly due to events, roadworks, or weather disruptions.
- These spikes represent instances where traffic volume was significantly higher or lower than the norm.
- Detecting such anomalies is crucial for real-time alert systems and for refining predictive models to account for exceptional conditions.

This analysis supports the development of a responsive traffic management system capable of adapting to unexpected changes in traffic flow.

5.15 Box Plot: Volume Distribution per Route per Hour

To understand how traffic volume varies across different routes and times of day, a multi-category box plot was generated. This visualization compares hourly vehicle volume distributions for each major route entering or exiting the Katugastota region.

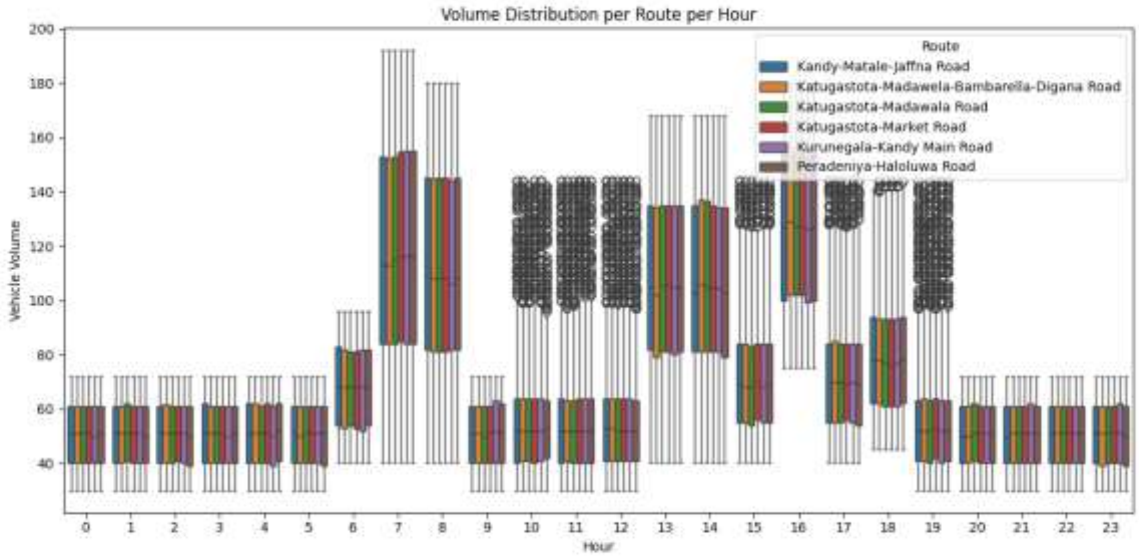

Figure 21:Volume Distribution per Route per Hour

- X-axis: Hour of the day (0 to 23)
- Y-axis: Vehicle Volume
- Color Legend: Represents seven key routes:
  - Matale–Jaffna Road
  - Katugastota–Madawala–Bambaralla–Digana Road
  - Katugastota–Madawala Road
  - Katugastota–Market Road
  - Kandy Man Road
  - Kurunegala–Kandy Main Road
  - Peradeniya–Halloluwa Road

Insights:

- Certain routes, such as Kurunegala–Kandy Main Road and Matale–Jaffna Road, show consistently higher volumes during morning and evening peak hours.
- Routes like Katugastota–Market Road exhibit more sporadic volume spikes, possibly due to localized commercial activity.
- The spread of the box plots indicates variability in traffic flow, with some routes experiencing wide fluctuations in volume throughout the day.

This analysis is essential for route-specific traffic signal optimization and for prioritizing infrastructure upgrades on high-volume corridors.

5.16 Trendline Visualization of Vehicle Volume

To observe the overall direction and consistency of traffic volume over time, a trendline visualization was created. This plot overlays a regression line on top of the raw vehicle volume data to highlight long-term movement and variability.

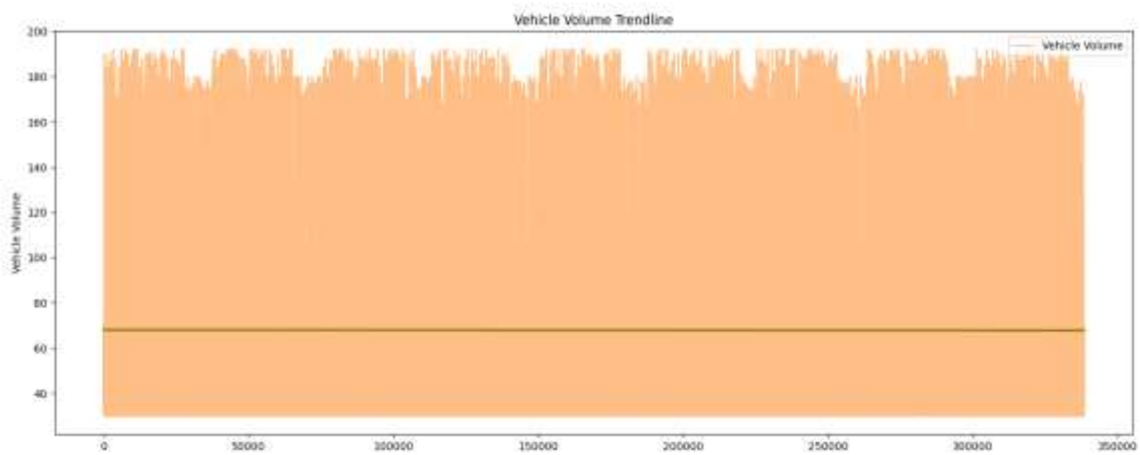

Figure 22:Vehicle Volume Trendline

- X-axis: Represents the timeline or data index (0 to ~3,000,000)
- Y-axis: Vehicle Volume (0 to 180)
- Visual Elements:
    - ❖ Orange Shaded Area: Raw vehicle volume data
    - ❖ Green Line: Fitted trendline indicating the general direction of volume changes

Insights:

- The trendline remains relatively flat, suggesting that while there are fluctuations in traffic volume, the overall average remains stable over the observed period.
- The orange area shows significant short-term variability, indicating the presence of daily or hourly traffic surges.
- This visualization is useful for confirming the stability of traffic flow and identifying whether interventions are needed to manage long-term growth.

This trendline supports the use of smoothing techniques in traffic forecasting and helps validate the consistency of the simulated dataset.

5.17 Z-Score Anomaly Detection in Vehicle Volume

To identify statistically significant deviations in traffic volume, a Z-score-based anomaly detection method was applied. This technique flags data points that deviate substantially from the mean, helping to isolate unusual traffic events.

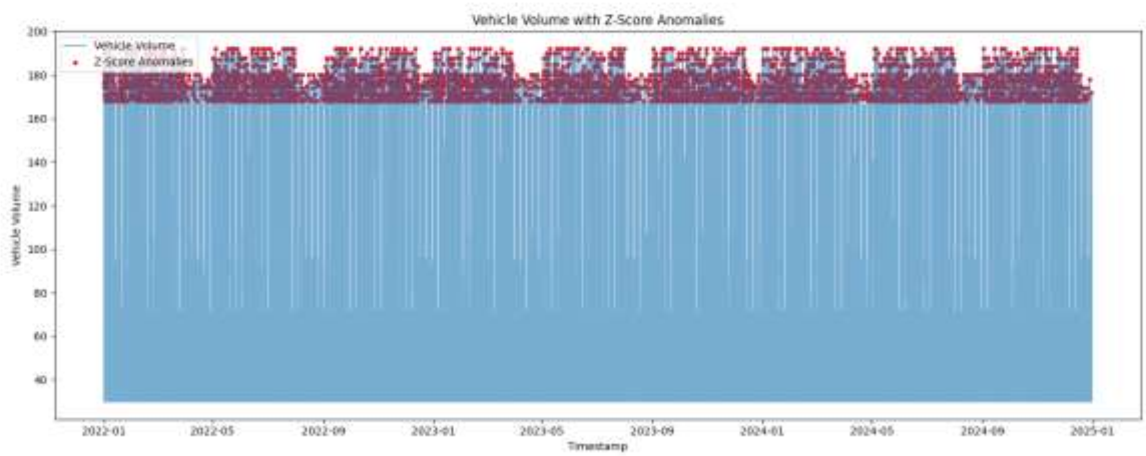

Figure 23:Vehicle Volume with Z-Score Anomalies (2021)

- X-axis: Timestamp (January 2021 to September 2021)
- Y-axis: Vehicle Volume (0 to 200)
- Data Points:
    - Blue: Normal vehicle volume
    - Red: Anomalies detected using Z-score thresholds

Insights:

- Several red points indicate traffic volumes that are significantly higher or lower than the norm, suggesting potential incidents, events, or data irregularities.
- The Z-score method is effective for real-time anomaly detection, enabling the system to trigger alerts or reroute traffic dynamically.
- This approach enhances the robustness of the traffic management system by accounting for unpredictable fluctuations.

This analysis supports the integration of statistical anomaly detection into the system's incident response module.

6 Exploratory Data Summary

This section presents a statistical overview of the simulated traffic dataset used for identifying congestion and incident-prone areas.

6.1 Null Value Inspection

```
EDA_Summary.md > # EDA Summary
# EDA Summary

**Dataset shape:** (338400, 18)

## Null Values:

Timestamp            0
Bottleneck           0
Entry Point          0
Route                0
Vehicle Type         0
Traffic Multiplier   0
Vehicle Volume       0
Traffic Intensity    0
Avg Speed (km/h)     0
Speed Status         0
Is School Day        0
Is Holiday           0
Is Sunday            0
Is Working Day       0
Hour                 0
Day                  0
Month                0
Year                 0
```

Figure 24:Null Value Inspection

6.2 Descriptive Statistics (Sample)

Metric	Traffic Multiplier	Vehicle Volume	Avg Speed (km/h)	Hour	Year
Count	338,400	338,400	338,400	338,400	338,400
Mean	0.80	67.94	26.51	11.5	2023
Min	0.60	30	10.0	0	2022
25%	0.60	45	18.8	5.75	2022
50%	0.60	60	26.3	11.5	2023

Table 1: Descriptive Statistics (Sample)

6.3 Anomaly & Clustering Highlights

- Volume Anomalies (Z-score > 3): 5,814 records
- Clustering Applied: K-Means (k=3)

7. Rule-Based Traffic Condition Detection

To derive actionable insights from the dataset, domain-specific rule logic was applied for detecting congestion conditions and potential incidents.

7.1 Rule Logic and Purpose

Rule Name	Condition Description
High Volume	Vehicle Volume > 1200 – identifies unusually high congestion load.
Low Speed	Avg Speed < 15 km/h – flags speed reductions or obstructions.
Potential Congestion	Combination of High Volume **and** Low Speed.
Incident Detection	Sudden variation in volume and speed across time intervals:

Table 2: Rule Logic and Purpose

7.2 Rule-Based Function

```
def apply_rules(df):
    df['Is High Volume'] = df['Vehicle Volume'] > 1200
    df['Is Low Speed'] = df['Avg Speed (km/h)'] < 15
    df['Potential Congestion'] = df['Is High Volume'] & df['Is Low Speed']

    df['Is Incident'] = (
        (df['Vehicle Volume'].diff().abs() > 500) &
        (df['Avg Speed (km/h)'].diff().abs() > 10)
    )

    summary = {
        "High Volume Records": df['Is High Volume'].sum(),
        "Low Speed Records": df['Is Low Speed'].sum(),
        "Detected Congestion": df['Potential Congestion'].sum(),
        "Potential Incidents": df['Is Incident'].sum(),
    }

    return df, summary
```

8 Summary

This chapter provided a structured analysis of traffic in Kandy, from bottleneck identification to detailed simulation and data-driven insights. Simulated data effectively captured real-world bottleneck behavior using structured rules, anomaly detection, and clustering. EDA revealed consistent trends, congestion triggers, and outliers. A rule-based model identified congestion zones and possible incidents, offering a foundation for building an intelligent traffic management and incident response framework tailored to Kandy's unique urban flow.