

BESANT TECHNOLOGIES
DATA ANALYSIS PROJECT
Bangalore Traffic Analysis : Detailed
Report

SUBMITTED BY:

NAME:V.Barathkumar

PH.NO:9345273442

EMAIL:Barathkuma23@gmail.com

UNDER THE GUIDANCE OF TRAINER NAME:

PRIYANKA G

CONTENTS

1. Introduction
2. Objectives of the Analysis
3. Data Collection
4. Data Inspection/Initial Analysis
5. Data Cleaning and Transformation
6. Exploratory Data Analysis
7. Visualisation
8. Technologies Used.
9. Insights Generation
10. Conclusion

1.Introduction

This project focuses on the statistical and exploratory analysis of Bengaluru city traffic data.

The aim is to detect patterns and anomalies related to traffic volume, average speed, congestion levels, and environmental impact by leveraging a structured urban mobility dataset.

Such analysis helps in understanding traffic dynamics, identifying congestion-prone areas, and supporting smart city initiatives for efficient traffic management and urban planning.

Dataset: Bengaluru Traffic Data (minimum 1000 records, 15+ columns)

Environment: Jupyter Notebook, pandas, MySQL, seaborn, matplotlib

2.Objective of the Analysis

The objective of this study is to analyze traffic data from various areas in Bengaluru to understand congestion patterns, road usage, and environmental impact.

The analysis aims to extract meaningful insights from traffic, transport, and environmental features to support data-driven urban mobility planning and decision-making.

Specifically, the study aims to

- Identify key factors influencing traffic congestion and travel delays across different areas and intersections.
- Analyze distributions and trends in traffic volume, average speed, and road capacity utilization over time.
- Examine the impact of weather conditions, roadwork, and incidents on congestion and travel efficiency.
- Study correlations among traffic, environmental, and behavioral factors (e.g., public transport usage, parking, and compliance levels).
- Evaluate the relationship between environmental impact and vehicle activity levels to understand sustainability concerns.
- Provide data-driven insights for improving traffic management, reducing congestion, and promoting safer, more sustainable transportation systems in Bengaluru.

Questions to Ask

1.What are the areas has the highest traffic congestion level?

Based on the analysis, the areas that experience higher congestion levels in Bengaluru are:

1. Koramangala – Congestion Level: 94%
2. M.G. Road – Congestion Level: 91%
3. Indiranagar – Congestion Level: 88%

2.Why are certain areas like Koramangala, M.G. Road, and Indiranagar experiencing higher congestion levels?

These areas experience higher congestion levels mainly due to high traffic volume.

While public transport usage and parking usage remain nearly constant across all regions (around 45–46% and 75–76% respectively), traffic volume is significantly higher in these congested areas

1. Koramangala: 40,832 vehicles
2. M.G. Road: 35,300 vehicles
3. Indiranagar: 32,284 vehicles

This shows that increased vehicle density is the primary cause of congestion, rather than differences in public transport or parking usage.

3.How does congestion level affect the average vehicle speed in different areas of Bengaluru?

From the dataset, it is observed that areas with higher congestion levels have lower average speeds:

- Koramangala → Congestion: 94%, Speed: 36 km/h
- M.G. Road → Congestion: 91%, Speed: 38 km/h
- Indiranagar → Congestion: 88%, Speed: 39 km/h

In contrast, areas with lower congestion levels show higher speeds:

- Electronic City → Congestion: 54%, Speed: 44 km/h
- Yeshwanthpur → Congestion: 62%, Speed: 43 km/h

This clearly indicates a negative correlation between congestion and average speed — as congestion increases, vehicle speed decreases.

4.Does public transport usage help in reducing traffic congestion in Bengaluru areas?

Based on the dataset, public transport usage across areas (44%–46%) shows little variation, while congestion levels vary significantly from 54% to 94%.

This indicates that public transport usage does not have a noticeable impact on reducing congestion.

The likely reasons include:

- Limited public transport coverage in high-traffic zones.
- Preference for private vehicles despite availability.
- Inefficient scheduling or lack of capacity in public transport systems.

5.Does parking space usage contribute to traffic congestion in Bengaluru areas?

Based on the dataset, parking usage across all areas remains fairly consistent at around 75–76%, whereas congestion levels vary significantly — from 54% in Electronic City to 94% in Koramangala.

This shows that parking space usage does not have a direct correlation with traffic congestion.

Even though parking spaces are heavily utilized in every area, congestion is still high in zones like Koramangala, M.G. Road, and Indiranagar, indicating that other factors such as vehicle inflow, narrow roads, or high commercial activity are more influential in causing congestion.

6.Do areas with higher road capacity utilization experience higher traffic congestion levels in Bengaluru?

Based on the dataset, areas with higher road capacity utilization also show higher congestion levels, indicating a strong positive relationship between these two factors.

For instance:

- Koramangala → Road Utilization: 97%, Congestion: 94%
- M.G. Road → Road Utilization: 97%, Congestion: 91%
- Indiranagar → Road Utilization: 96%, Congestion: 88%

In contrast, areas like Electronic City and Yeshwanthpur have lower utilization (76–82%) and correspondingly lower congestion levels (54–62%).

3.Data Collection

a) Data Storage & Transfer

- The dataset ('bengalor_traffic_data.csv') was first loaded to MySQL for scalable storage and simulation of real-world ETL (Extract, Transform, Load) pipelines.
- Data was loaded back into a Jupyter Notebook for full analysis using SQLAlchemy and pandas read_sql().

	Date	Area Name	Road/Intersection Name	Traffic Volume	Average Speed	Travel Time Index	Congestion Level	Road Capacity Utilization	Incident Reports	Environmental Impact	Public Transport Usage	Traffic Signal Compliance	Parking Usage	Pedes C
0	2022-01-01	Indiranagar	100 Feet Road	50590	50.230299	1.500000	100.000000	100.000000	0	151.180	70.632330	84.044600	85.403629	
1	2022-01-01	Indiranagar	CMH Road	30825	29.377125	1.500000	100.000000	100.000000	1	111.650	41.924899	91.407038	59.983689	
2	2022-01-01	Whitefield	Marathahalli Bridge	7399	54.474398	1.039069	28.347994	36.396525	0	64.798	44.662384	61.375541	95.466020	
3	2022-01-01	Koramangala	Sony World Junction	60874	43.817610	1.500000	100.000000	100.000000	1	171.748	32.773123	75.547092	63.567452	
4	2022-01-01	Koramangala	Sarjapur Road	57292	41.116763	1.500000	100.000000	100.000000	3	164.584	35.092601	64.634762	93.155171	
...	
8931	2024-08-09	Electronic City	Hosur Road	11387	23.440276	1.262384	35.871483	57.354487	1	72.774	21.523289	83.530352	97.898279	
8932	2024-08-09	M.G. Road	Trinity Circle	36477	45.168429	1.500000	100.000000	100.000000	3	122.954	29.822312	60.738488	60.355967	
8933	2024-08-09	M.G. Road	Anil Kumble Circle	42822	22.028609	1.500000	100.000000	100.000000	1	135.644	43.185905	85.321627	61.333731	
8934	2024-08-09	Jayanagar	South End Circle	20540	52.254798	1.020520	72.639152	97.845527	2	91.080	44.416043	89.586947	79.197198	
8935	2024-08-09	Yeshwanthpur	Yeshwanthpur Circle	14705	31.128967	1.048720	43.409821	77.734621	1	79.410	26.616725	80.778753	60.602672	

8936 rows × 16 columns

- Row and column count verified (>1000 records, 10+ columns).
- No missing data in the transfer, type integrity preserved.

4. Data Inspection / Initial Analysis

The dataset contains 8090 records and 15 features, each representing key traffic-related metrics recorded from different areas of Bengaluru.

Key attributes include:

- Area Name: The specific region or locality within Bengaluru (e.g., Koramangala, M.G. Road).
- Congestion Level (%): The percentage of road congestion, indicating how crowded an area is.
- Traffic Volume: The total number of vehicles passing through a particular area.
- Average Speed (km/h): The average speed of vehicles, inversely related to congestion levels.
- Road Capacity Utilization (%): The extent to which the available road capacity is being used.

a) Size & Description

- The dataset contains 8090 rows.
- The dataset contains 15 columns.
- Each row represents a specific area in Bengaluru.
- Each column represents a traffic-related feature such as congestion level, traffic volume, average speed, and road capacity utilization.

b) Null Value Analysis

- Null values checked per column—minimal or none.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8936 entries, 0 to 8935
Data columns (total 16 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Date                                8936 non-null   object
1   Area Name                          8936 non-null   object
2   Road/Intersection Name              8936 non-null   object
3   Traffic Volume                      8936 non-null   int64
4   Average Speed                       8936 non-null   float64
5   Travel Time Index                   8936 non-null   float64
6   Congestion Level                    8936 non-null   float64
7   Road Capacity Utilization           8936 non-null   float64
```

```
8 Incident Reports 8936 non-null int64
9 Environmental Impact 8936 non-null float64
10 Public Transport Usage 8936 non-null float64
11 Traffic Signal Compliance 8936 non-null float64
12 Parking Usage 8936 non-null float64
13 Pedestrian and Cyclist Count 8936 non-null int64
14 Weather Conditions 8936 non-null object
15 Roadwork and Construction Activity 8936 non-null object
dtypes: float64(8), int64(3), object(5)
memory usage: 1.1+ MB
```

b) Unique Value Analysis

Column:Area Name

- Unique Values: ['Indiranagar', 'Whitefield', 'Koramangala', 'M.G. Road', 'Jayanagar', 'Hebbal', 'Yeshwanthpur', 'Electronic City']
 - Count of Unique Values: 8
- Observation:**
- The dataset includes traffic data from eight major areas of Bengaluru.
 - These locations represent a mix of residential, commercial, and IT zones, which helps in analyzing traffic density and flow patterns across different parts of the city.

Column:Road/Intersection Name

- Unique Values:[100 Feet Road, CMH Road, Marathahalli Bridge, Sony World Junction, Sarjapur Road, Trinity Circle, Anil Kumble Circle, Jayanagar 4th Block, South End Circle, Hebbal Flyover, Ballari Road, Yeshwanthpur Circle, Tumkur Road, ITPL Main Road, Silk Board Junction, Hosur Road]
 - Count of Unique Values: 16
- Observation:**
- The dataset covers sixteen key traffic junctions and major roads across Bengaluru, including commercial hubs, residential zones, and IT corridors. This helps in identifying region-specific congestion trends and improving traffic management strategies.

Column: Weather Conditions

- Unique Values:[Clear, Overcast, Fog, Rain, Windy]
 - Count of Unique Values: 5
- Observation:**
- The dataset includes five distinct weather conditions observed during data collection. These variations help analyze how different weather types affect traffic flow and vehicle density across Bengaluru.

Column: Road Constructions

- Unique Values:[No, Yes]
 - Count of Unique Values: 2
- Observation:**
- This column indicates whether road construction activity was present at the given time and location. It helps in understanding how ongoing construction projects affect traffic flow and congestion levels across Bengaluru.

5. Data Cleaning and Transformation

a) Handling Unstructured/Mixed Data

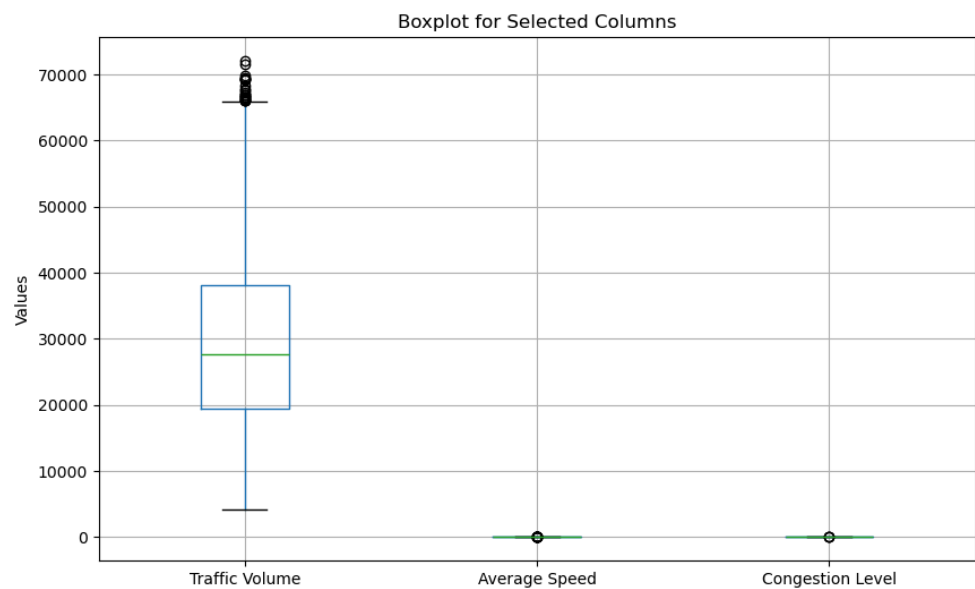
All columns were inspected for unstructured or mixed data types. The dataset was found to be consistent, with no invalid or non-standard entries. Hence, no additional data type conversion or text cleaning was required.

b) Date Column Handling

The date column in the dataset was originally stored as a string. It was successfully converted into a proper datetime format using pandas' to_datetime() function to ensure accurate time-based analysis and consistency in date operations.

c) Outlier Management

- Used IQR method to cap/remove outliers for columns like baseline_value, accelerations, uterine_contractions.
- Outliers defined as values beyond [Q1-1.5×IQR,Q3+1.5×IQR][Q1-1.5×IQR,Q3+1.5×IQR].



6.Exploratory Data Analysis (EDA)

a) Statistical Overview

- Used describe() to summarize mean, variance, skewness.

	Traffic Volume	Average Speed	Travel Time Index	Congestion Level	Road Capacity Utilization	Incident Reports	Environmental Impact	Public Transport Usage	Traffic Signal Compliance	Parking Usage	Pedestrian and Cyclist Count	
count	8936.0	8936.0	8936.0	8936.0	8936.0	8936.0	8936.0	8936.0	8936.0	8936.0	8936.0	8936.0
mean	29236.0	39.0	1.0	81.0	92.0	2.0	108.0	45.0	80.0	75.0	115.0	81.0
std	13002.0	11.0	0.0	24.0	17.0	1.0	26.0	20.0	12.0	14.0	37.0	24.0
min	4233.0	20.0	1.0	5.0	19.0	0.0	58.0	10.0	60.0	50.0	66.0	5.0
25%	19413.0	32.0	1.0	64.0	97.0	0.0	89.0	27.0	70.0	63.0	94.0	64.0
50%	27600.0	39.0	2.0	92.0	100.0	1.0	105.0	45.0	80.0	75.0	102.0	92.0
75%	38058.0	47.0	2.0	100.0	100.0	2.0	126.0	62.0	90.0	88.0	111.0	100.0
max	72039.0	90.0	2.0	100.0	100.0	10.0	194.0	80.0	100.0	100.0	243.0	100.0

b) Skewness Analysis

1.Column Name:congestion Level

The skewness value of the Congestion Level column is -0.9979, which indicates a negative (left) skewness.

This means that:

- The tail of the distribution is on the left side, representing a few roads with very low congestion levels.
- The peak is on the right side, showing that most roads experience higher congestion levels.
- Therefore, the data is moderately left-skewed and not perfectly symmetrical.

This implies that while congestion is generally high across most areas, a few locations have much lower congestion, pulling the average slightly toward the left.

2. Average Speed

The skewness value of the Average Speed column is 0.2432, which indicates a slightly positive (right) skewness.

This means that:

- The tail of the distribution lies on the right side, representing a few cases with very high speeds.
- The majority of values are concentrated around the center, showing that the data is almost symmetrical.
- Since the skewness value is close to zero, the Average Speed data can be considered approximately normally distributed.

This suggests that most areas in Bengaluru have typical average speeds, with only a few roads showing unusually high speeds.

3.Traffic Volume

The skewness value of the Traffic Volume column is 0.4595, which indicates a slightly positive (right) skewness.

This means that:

- The tail of the distribution lies on the right side, representing a few areas with very high traffic volumes.
- The peak is slightly toward the left, where most roads have moderate traffic flow.
- Since the skewness value is close to zero, the data is nearly symmetrical with a slight right skew.

This suggests that while most roads experience moderate traffic levels, a few areas have exceptionally high traffic, slightly pulling the distribution toward the right.

4.Public Transport Usage

The skewness value of the Public Transport Usage column is -0.0098, which indicates a very slight negative (left) skewness.

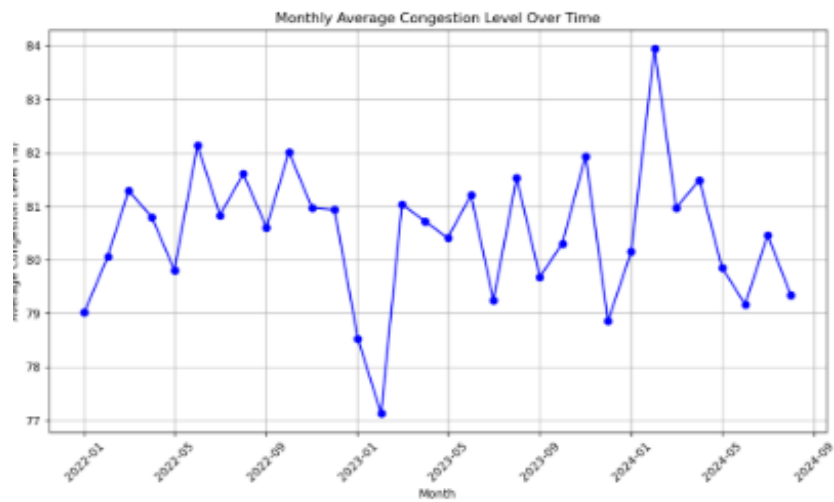
This means that:

- The distribution is nearly symmetrical, as the skewness value is extremely close to zero.
- The tail lies slightly on the left side, showing a few areas with lower public transport usage.
- The majority of data points are evenly spread, suggesting a balanced distribution of public transport usage across Bengaluru.

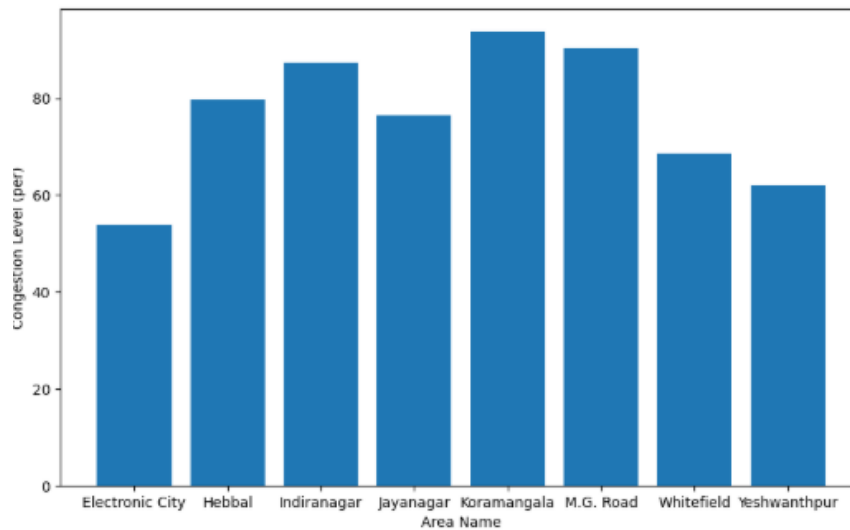
Overall, the Public Transport Usage data can be considered approximately normal, with no significant skewness observed.

7)Visualization

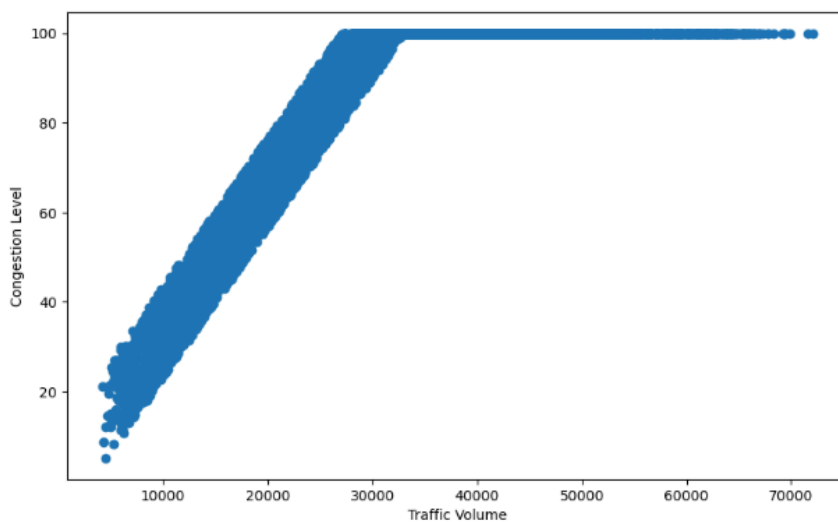
1)How congestion changes over time (line plot)



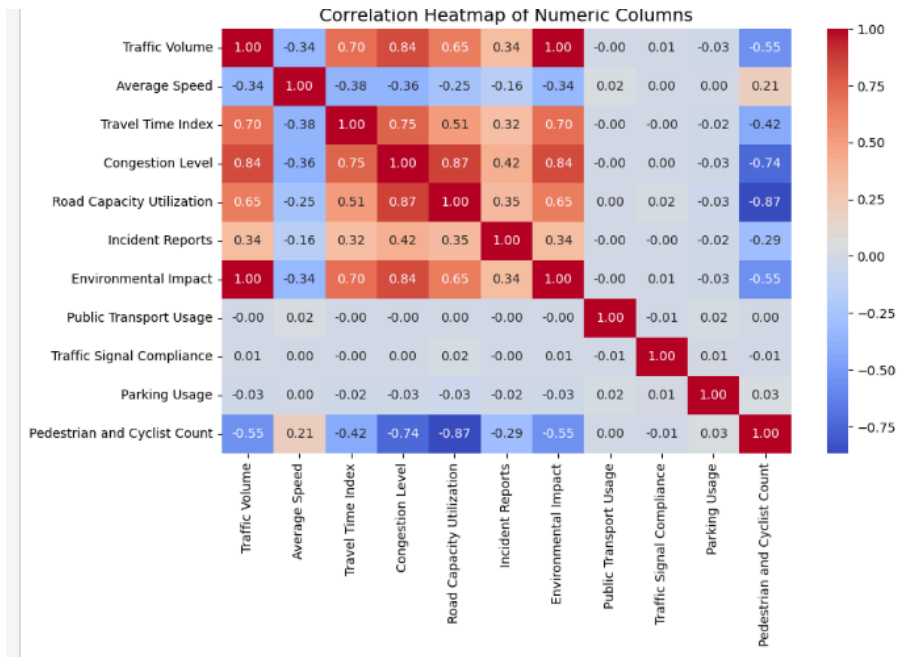
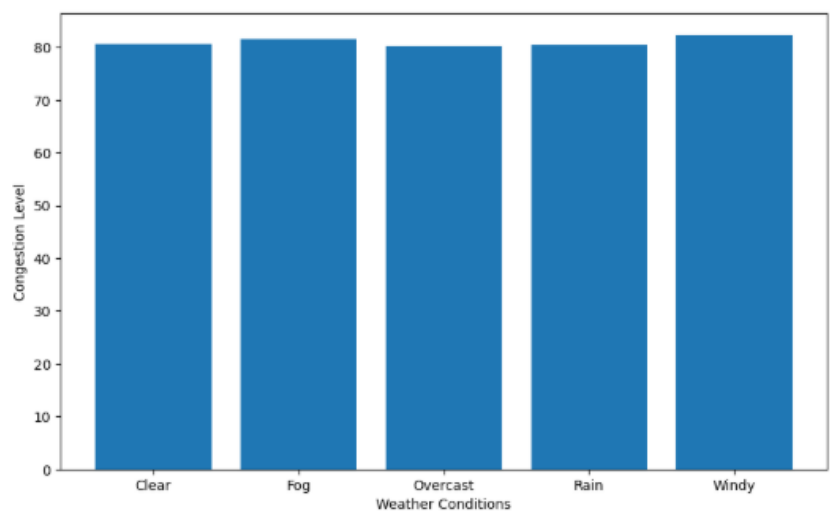
2)Which area has highest congestion



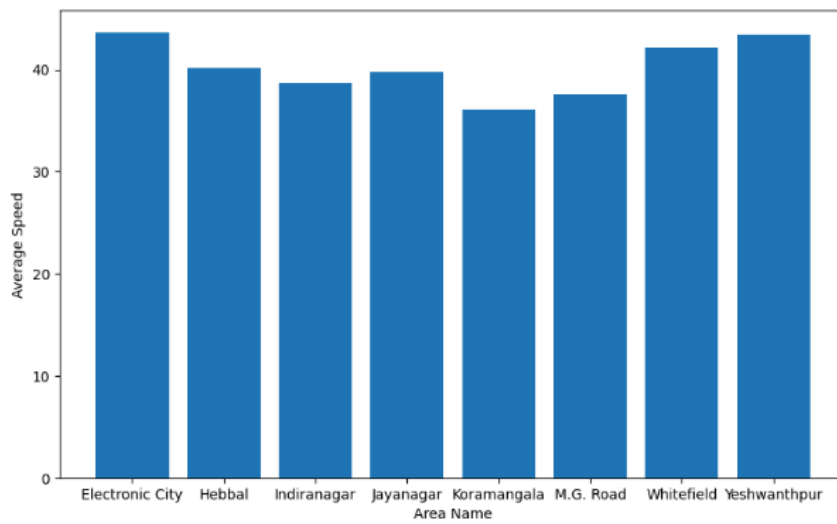
3)Relation between traffic volume & congestion



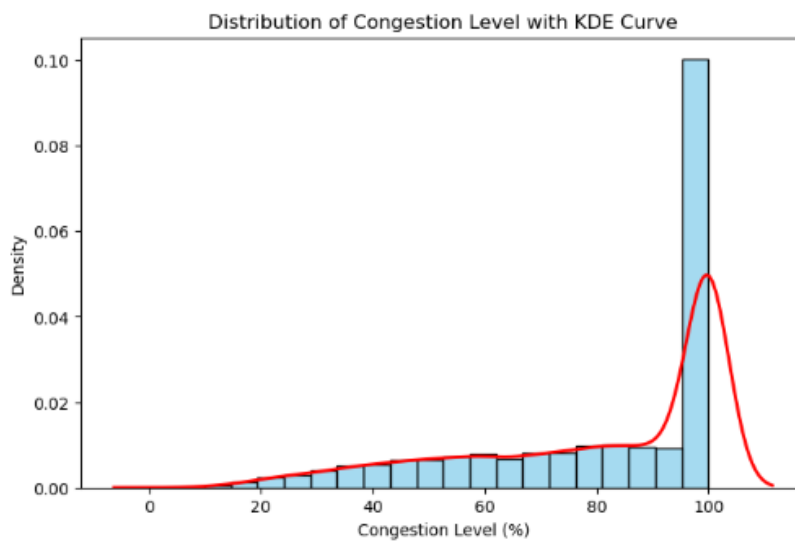
4)Which weather causes more congestion



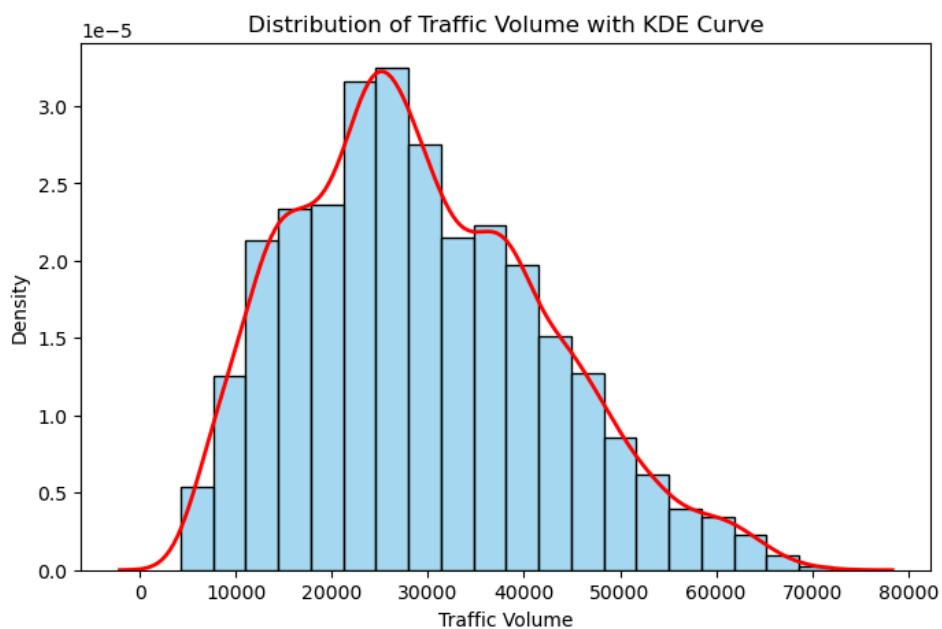
6)How does average speed vary across different areas?



7)What is the overall distribution of congestion levels across Bengaluru?



8)What is the overall distribution of Traffic Volume levels across Bengaluru?



8)Insight Generation

1)How congestion changes over time

Observation:

- The congestion level fluctuated between 77% and 84% from 2022–2024, with a slight dip during early 2023 (Jan–Feb).

Insight:

- This dip indicates a temporary improvement in traffic flow, possibly due to holidays or reduced travel activity. However, the overall congestion trend remains consistently high, showing that traffic is a persistent problem.

Recommendation:

- Authorities can analyze traffic management practices from early 2023 to identify factors that reduced congestion. Implementing similar strategies—like optimized signal timings or encouraging public transport—may help maintain smoother traffic year-round.

2)Which area has highest congestion

Observation:

- The congestion levels vary significantly across different areas of Bengaluru. The data shows that Koramangala (94%), M.G. Road (91%), and Indiranagar (88%) have the highest congestion, while Electronic City (54%), Yeshwanthpur (62%), and Whitefield (69%) experience comparatively lower congestion levels.

Insight:

- High congestion in Koramangala, M.G. Road, and Indiranagar can be attributed to dense commercial zones, shopping centers, and major traffic intersections, which lead to slower vehicle movement. Areas with lower congestion, such as Electronic City and Yeshwanthpur, likely benefit from wider roads, industrial layouts, or better traffic distribution.

Recommendation:

- To reduce congestion in high-traffic areas, authorities can implement intelligent traffic signal systems, dedicated bus lanes, and alternate route planning. Promoting public transportation and restricting peak-hour parking can further improve traffic flow in heavily congested regions.

3)Relation between traffic volume & congestion

Observation:

- The data shows a clear positive relationship between Traffic Volume and Congestion Level. As the number of vehicles increases, the congestion level also rises. In the dataset, lower traffic volumes (around 4,000–5,000 vehicles) correspond to congestion levels below 20%, while higher volumes (above 70,000 vehicles) show congestion levels close to 100%.

Insight:

- This indicates that traffic congestion in Bengaluru is strongly influenced by vehicle density. As more vehicles enter the roads, the existing infrastructure struggles to accommodate the flow, leading to slower speeds, longer travel times, and higher congestion levels. This pattern suggests that road capacity is nearing or exceeding its limit in several areas during peak hours.

Recommendation:

- To manage congestion effectively, vehicle density control and demand-based traffic management are essential. Authorities can encourage carpooling, public transport adoption, and staggered working hours to distribute traffic more evenly throughout the day. Expanding road capacity and improving traffic signal coordination can further reduce congestion during peak traffic hours.

4) Which weather causes more congestion

Observation:

- The average congestion levels across different weather conditions—Clear (80.7%), Fog (81.6%), Overcast (80.3%), Rain (80.5%), and Windy (82.3%)—show only slight variations. The differences are minimal, remaining within a narrow range of approximately 2%.

Insight:

- This indicates that weather conditions are not a major factor influencing congestion in Bengaluru. Traffic levels remain consistently high regardless of weather, suggesting that road infrastructure, signal timing, and vehicle volume play a much larger role in determining congestion levels than external factors like rain or fog.

Recommendation:

- Instead of focusing on weather-related traffic management, authorities should prioritize infrastructure improvements, real-time traffic monitoring, and optimized signal coordination. However, during foggy or rainy conditions, enhanced visibility aids and speed control measures can help maintain road safety without significantly impacting traffic flow.

5) Relationship among numeric columns

Observation:

- The heatmap shows that Traffic Volume, Congestion Level, Travel Time Index, and Road Capacity Utilization are strongly positively correlated ($r > 0.80$). This indicates that as vehicle volume and travel time increase, congestion and road usage also rise.
- On the other hand, Pedestrian and Cyclist Count shows a strong negative correlation with Congestion Level ($r = -0.74$) and Road Capacity Utilization ($r = -0.87$), suggesting fewer pedestrians and cyclists in areas with heavy traffic.
- Average Speed is negatively correlated with most congestion-related factors, meaning traffic speed decreases as congestion rises.

Insight:

- The correlations highlight that vehicle density and travel demand are the key drivers of congestion in Bengaluru. High road capacity utilization leads to slower speeds and higher travel times, directly increasing congestion. The drop in pedestrian and cyclist count in high-congestion areas implies that unsafe or uncomfortable conditions discourage non-motorized transport.

Recommendation:

- To reduce congestion, city planners should focus on reducing vehicle load through public transport incentives, traffic rerouting, and infrastructure upgrades. Improving pedestrian and cycling facilities can also encourage eco-friendly mobility, reducing overall traffic pressure. Additionally, smart traffic signal systems can help maintain average speeds and optimize road capacity usage.

6) How does average speed vary across different areas?

Observation:

- The average vehicle speed varies across different areas of Bengaluru. The highest average speed is observed in Electronic City (43.66 km/h) and Yeshwanthpur (43.44 km/h), while the lowest is in Koramangala (36.06 km/h) and M.G. Road (37.53 km/h).

Insight:

- Areas with commercial and densely populated zones like Koramangala, M.G. Road, and Indiranagar experience lower average speeds, indicating higher congestion and slower traffic movement. On the other hand, Electronic City and Yeshwanthpur, being on the outskirts or with wider roads, have smoother traffic flow.

Recommendation:

- Improve traffic signal coordination and road infrastructure in Koramangala and M.G. Road to reduce congestion. Encourage use of public transport and carpooling in high-density zones. Implement real-time traffic monitoring and adaptive signal systems to optimize speed and reduce delays.

7)What is the overall distribution of congestion levels across Bengaluru?

Observation:

- The distribution shows that most congestion levels are concentrated around 90–100%, while fewer data points exist for lower congestion levels. The KDE curve confirms a right-skewed distribution, indicating frequent high congestion.

Insight:

- This suggests that traffic congestion in Bengaluru is consistently high, with only a few instances of low congestion. The spike near 100% shows severe congestion events are common, possibly due to peak-hour traffic or poor road capacity in high-density areas.

Recommendation:

- Introduce intelligent traffic management systems to distribute traffic more evenly.
- Implement staggered office timings to reduce peak-hour load.
- Promote public transport and carpooling to decrease vehicle density.
- Conduct area-wise congestion audits to identify and improve critical choke points.

8)What is the overall distribution of Traffic Volume levels across Bengaluru?

Observation:

- The traffic volume distribution is nearly normal with a slight right skew (skewness ≈ 0.46). Most values are centered between 20,000 and 35,000, indicating that moderate traffic volumes are most common, with fewer instances of extremely high traffic.

Insight:

- The near-normal distribution implies consistent road usage patterns across the city. Occasional spikes in traffic may be linked to specific peak hours or key commercial zones rather than random fluctuations.

Recommendation:

- Maintain current traffic flow strategies while improving congestion handling during high-volume periods. Introducing predictive traffic management using historical patterns could help anticipate and reduce temporary surges in volume.

8.Tools & Technologies Used

- **Python:** Main language for data analysis and modeling.
- **Pandas, NumPy:** Data handling, transformation, statistics.
- **Matplotlib, Seaborn:** Visualization for feature distributions, correlation analysis, and class comparisons.
- **Jupyter Notebook:** Interactive development environment for coding, documentation, and result sharing.
- **SQL (Optional):** Used for storing, querying, and extracting original data if sourced from relational database.

10.Conclusion

- The analysis of Bengaluru's traffic data reveals that congestion levels are strongly influenced by traffic volume, while weather conditions have minimal impact. Areas like Koramangala, M.G. Road, and Indiranagar experience the highest congestion levels, mainly due to high vehicle density and commercial activity. Average speed is inversely related to congestion, with areas such as Electronic City and Yeshwanthpur maintaining higher speeds and smoother flow.
- The distribution analysis shows that traffic volume follows a near-normal pattern with a slight right skew, indicating occasional peaks in vehicle count. Overall, traffic congestion in Bengaluru is primarily driven by urban density and road usage patterns rather than external environmental factors.
- To improve traffic conditions, strategic measures such as better traffic signal coordination, promoting public transport, and implementing smart mobility systems can help reduce congestion and enhance travel efficiency.