CAC-2 Project : Case Study Presentation

Title:

Analyzing Urban Traffic in Bengaluru: Impact of Environmental and Situational Factors on Congestion and Mobility in Key Local Areas

Presented By:

Kumar Yashu (2448062)

Sreneedhi Konnor (2448055)

Aaditya Kumar Dhaka (2448001)

Sayandip Ghosh (24480048)

Neelanjan Dutta (2448040)

Project Assignment Roles:

Name	Work Done
1. Aaditya Kumar Dhaka (2448001) , 2. Neelanjan Dutta (2448040)	Data Cleaning and Transformation, Code Writing in R, Documentation Report
1. Sreenidhi Konnur (2448055) , 2. Sayandip Ghosh (24480048) 3. Kumar Yashu (2448062) 4. Aaditya Kumar Dhaka (2448001) ,	Power Point Presentation , Dataset Sourcing
1. Kumar Yashu (2448062) , 2. Sreenidhi Konnur (2448055) , 3. Aaditya Kumar Dhaka(2448001), 4. Neelanjan Dutta (2448040) and 5. Sayandip Ghosh (24480048)	QUESTION – 1 QUESTION – 2 QUESTION – 3 QUESTION – 5 QUESTION - 4

Introduction:

<u>Problem</u>: " Understanding Traffic Dynamics in Bengaluru's Key Local Areas "

<u>Statement</u>: Bengaluru's rapid urban growth has intensified traffic congestion in areas like MG Road, Indiranagar, Koramangala, Jayanagar, and Electronic City. Environmental factors like weather and situational elements such as roadwork and traffic volume significantly affect mobility. This study aims to explore these impacts to provide data-driven solutions for congestion management.

<u>Purpose of the Study</u>:

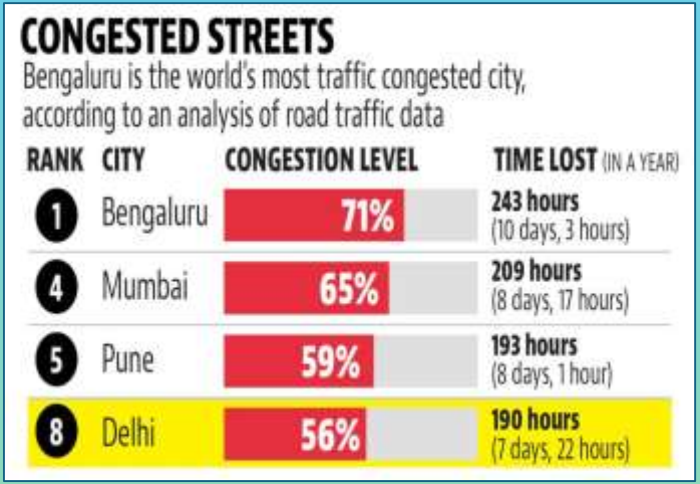
- To analyze how environmental factors (e.g., weather conditions) and situational factors (e.g., roadwork, traffic volume) affect traffic congestion and mobility in Bengaluru's key areas.
- To identify patterns and relationships between these factors and traffic dynamics through statistical analyses.
- To provide actionable insights for improving traffic management and enhancing urban mobility in MG Road, Indiranagar, Koramangala, Jayanagar, and Electronic City.

Relevance of the Study

Relevance of the Study:

- **Addressing Urban Challenges:** Bengaluru’s traffic congestion in areas like MG Road, Indiranagar, Koramangala, Jayanagar, and Electronic City impacts daily commuting, productivity, and environmental quality.

- **Improving Urban Mobility:** Understanding how environmental (e.g., weather) and situational (e.g., roadwork) factors influence traffic dynamics can provide actionable insights to mitigate congestion and enhance mobility.



Key Areas of Analysis

Areas	Description
1. Environmental Factors	Exploring the effects of weather conditions (e.g., rain, fog) on congestion and traffic volume.
2. Situational Factors	Assessing the impact of roadwork and traffic signal compliance on congestion levels
3. Public Transport Usage	Analyzing the relationship between weather conditions and public transport trends
4. Travel Time Reliability	Examining how congestion correlates with the travel time index

Main Statistical Tests Used

Tests
ANOVA (one and two way)
T-tests (two sample) and Mann-Whitney U Test
Chi-square Test (for independence & goodness of fit)
Correlation Analysis (Spearman)

Questions and Objectives

The study investigates how environmental and situational factors influence traffic congestion and mobility in Bengaluru's key areas. By formulating targeted questions and objectives, it aims to uncover actionable insights through statistical analysis, helping address traffic challenges effectively.

Question 1: Investigating the Effect of Area and Weather on Congestion Levels

- Test if there is a significant
- difference in congestion levels
- between two areas (e.g., Indiranagar
- and Koramangala).

- **2.** Investigate the impact of weather
- conditions on congestion levels
- using statistical methods and
- visualizations.

- **3.** Evaluate if the observed distribution
- of weather conditions aligns with the
- expected distribution.

Question 2: Investigating the Effect of Weather Conditions on Traffic Volume

- **1.** Compare traffic volumes across
- different weather conditions using
- the Kruskal-Wallis test.

- **2.** Analyze road capacity utilization
- under contrasting weather conditions
- using the Mann-Whitney U Test.

- **3.** Determine the relationship between
- congestion levels and traffic signal
- compliance using correlation
- analysis.

Question 3: Impact of Roadwork and Construction Activity on Traffic Volume

- Assess differences in traffic volume
- between roads with and without
- roadwork/construction activity
- using the Mann-Whitney U Test.

- **2.** Examine the relationship between
- roadwork activity and traffic
- volume categories using the Chi-
- square test for independence.

- **3.** Conduct a two-way ANOVA to
- compare traffic volumes across
- different areas under the presence
- and absence of roadwork.

Questions and Objectives

Question 4: Exploring the Effect of Weather Conditions on Public Transport Usage

- **1.** Perform one-way ANOVA to compare public transport usage
- across weather conditions.

- **2.** Determine the correlation between environmental impact and
- public transport usage using Spearman rank correlation.

- **3.** Fit regression models (Poisson and Negative Binomial) to
- predict public transport usage based on weather conditions.

Question 5: Analyzing the Relationship Between Congestion Level and Travel Time Index

- **1.** Test whether the median travel time index significantly
- deviates from a hypothesized value using the Wilcoxon
- Signed-Rank Test.

- **2.** Analyze the monotonic relationship between congestion levels
- and travel time indices using Spearman's rank correlation.

- **3.** Compare congestion levels across different ranges of travel
- time indices using the Kruskal-Wallis test.

Dataset Description

- **Source:** The dataset was obtained from Kaggle. (**Link:** https://www.kaggle.com/datasets/preethamgouda/banglore-city-traffic-dataset)
- **Overview:** This dataset contains traffic-related data collected across various areas and roads. It includes both numeric and categorical features, providing insights into traffic volume, speed, congestion, and other transportation metrics.

Column Name	Description
Date	The date of data collection
Area Name	Name of the area
Road Name	Name of the road or junction
Traffic Volume ***(in k)***	Total number of vehicles recorded on the road
Travel Time Index ***(ratio)***	A ratio comparing actual travel time to free-flow travel time.
Congestion Level ***(in %)***	Percentage-based measure of road congestion.
Road Capacity Utilization ***(in %)***	Percentage of road capacity being utilized.
Environmental Impact ***(in gm)***	A calculated metric quantifying environmental effects, such as CO_2 emissions.
Public Transport Usage ***(in %)***	Percentage of people using public transport
Traffic Signal Compliance ***(in %)***	Percentage compliance with traffic signals.
Weather Conditions ***(categorical)***	Weather conditions during data collection (e.g., "Clear", "Rain").
Roadwork and Construction Activity ***(categorical)***	Indicates if roadwork or construction activity was present ("Yes" or "No").

Dataset Cleaning and Transformation

Initial Data

	Date	Area Nam	Road/Intersection Nam	Traffic Volum	Average Spee	Travel Time Inde	Congestion Lev	Road Capacity Utilizatio	Incident Report	Environmental Impac	Public Transport Usag	Traffic Signal Complianc	Parking Usag	Pedestrian and Cyclist Cou	Weather Condition	Roadwork and Construction Activi
2	01-01-2022	Indiranagar	100 Feet Road	50590	50.23	1.5	100	100	0	151.18	70.63	84.04	85.4	111	Clear	No
3	01-01-2022	Indiranagar	CMH Road	30825	29.38	1.5	100	100	1	111.65	41.92	91.41	59.98	100	Clear	No
4	01-01-2022	Koramangala	Sony World Junction	60874	43.82	1.5	100	100	1	171.75	32.77	75.55	63.57	111	Clear	No
5	01-01-2022	Koramangala	Sarjapur Road	57292	41.12	1.5	100	100	3	164.58	35.09	64.63	93.16	104	Clear	No
6	01-01-2022	M.G. Road	Trinity Circle	47848	34.24	1.5	100	100	3	145.7	39.93	61.02	55.39	94	Overcast	No
7	01-01-2022	M.G. Road	Anil Kumble Circle	36574	29.98	1.5	100	100	3	123.15	72.48	81.57	90.37	115	Clear	No
8	01-01-2022	Jayanagar	Jayanagar 4th Block	25379	38.46	1.5	79.04	100	2	100.76	46.32	88.12	68.18	92	Clear	No
9	01-01-2022	Jayanagar	South End Circle	25022	35.04	1.5	78.98	100	1	100.04	44.26	99.43	62.1	105	Clear	No
10	02-01-2022	Indiranagar	100 Feet Road	22050	52.87	1.13	78.43	100	3	94.1	35.74	70.61	62.2	115	Fog	No
11	02-01-2022	Indiranagar	CMH Road	37877	26.43	1.5	100	100	1	125.75	75.83	98.16	95.74	96	Overcast	No
12	02-01-2022	Koramangala	Sarjapur Road	29106	43.56	1.5	90.41	100	3	108.21	53.05	99.6	57	100	Fog	No

- Dataset comprised over **5,000 rows** spanning **January 2023 to 2024**; analysis was focused on the years **2024** for greater relevance and depth.
- Refined the dataset from **16 columns** to **12 key columns** by removing four less relevant features.
- Removed **null values** and **duplicate entries** to ensure accuracy and reliability in the analysis.

- Utilized **log transformation** and **Box-Cox method** to normalize certain non-normal columns for more robust statistical analysis.

Clean Data

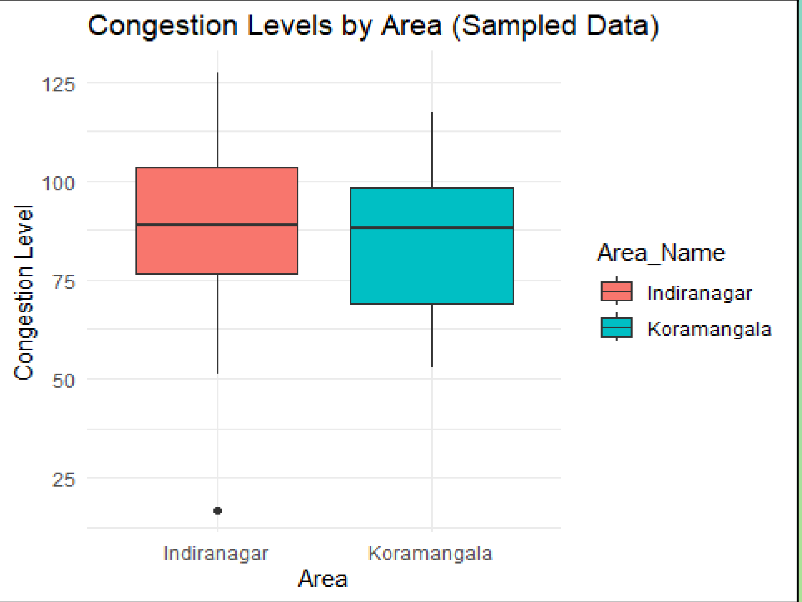
	Date	Area_Name	Road_Name	Traffic_Volume	Travel_Time_Index	Congestion_Level	Road_Capacity_Utilization	Environmental_Impact	Public_Transport_Usage	Traffic_Signal_Compliance	Weather_Conditions	Roadwork_and_Construction_Activity
2	01-01-2024	Indiranagar	CMH Road	25253.08892	1.25	112.6641073	60.1	150.6128032	28.44057627	64.72	Clear	No
3	01-01-2024	Koramangala	Sony World Junction	29605.27961	1.5	91.77776706	100	113.8716307	25.24314237	93.27	Clear	No
4	01-01-2024	Koramangala	Sarjapur Road	53176.62561	1.5	92.40328977	100	121.783839	32.60246474	65.42	Rain	Yes
5	01-01-2024	M.G. Road	Trinity Circle	33567.28262	1.5	94.31854418	100	167.3341302	24.93947257	66.56	Clear	No
6	01-01-2024	M.G. Road	Anil Kumble Circle	34341.79167	1.5	104.7179294	100	101.3962943	48.98936905	89.25	Clear	No
7	01-01-2024	Jayanagar	South End Circle	55236.86737	1.09	93.54273136	98.45	149.4329552	51.73017981	80.67	Clear	No
8	02-01-2024	Indiranagar	100 Feet Road	38711.51131	1.5	98.34388712	100	99.36409079	37.7318273	66.37	Clear	No
9	02-01-2024	Indiranagar	CMH Road	15969.0806	1.5	99.03564483	100	160.4890322	61.59282521	98.34	Clear	No
10	02-01-2024	Koramangala	Sony World Junction	23587.87316	1.5	90.09427399	100	97.80329112	54.00097698	64.13	Rain	No
11	02-01-2024	Koramangala	Sarjapur Road	26765.93724	1.5	100.8377527	100	93.89596512	19.22183614	63.07	Clear	No

Methodology and Inferences

Question 1: Investigating the Effect of Area and Weather on Congestion Levels

Objective 1

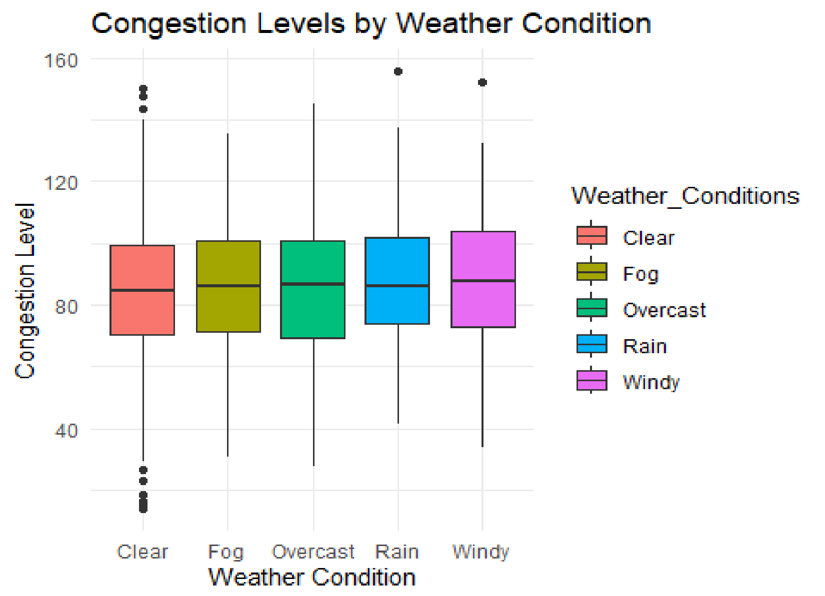
```r
# Step 5: Independent t-test (assuming normality and equal variances)
t_test_result <- t.test(Congestion_Level ~ Area_Name, data = sampled_d
ata, var.equal = TRUE)
cat("T-test result: t =", t_test_result$statistic, "p-value =", t_test
_result$p.value, "\n")

## T-test result: t = -0.1378291 p-value = 0.890952
```



The T-test and boxplot reveal no significant difference in congestion levels between Indiranagar and Koramangala, suggesting that observed variations are due to random chance.

Objective 2

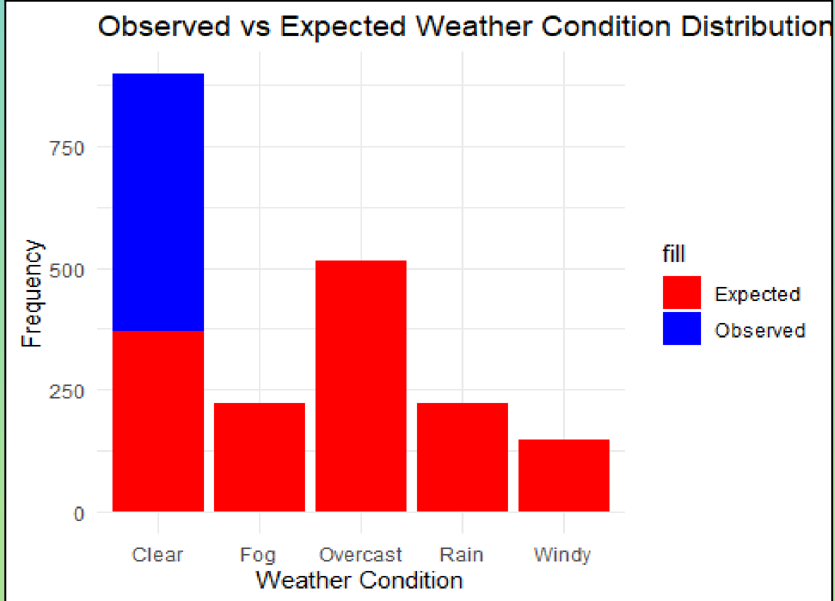
```r
# Step 4: Extract F-value and p-value for comparison with critical F-v
alue
f_value <- anova_summary[[1]]["Weather_Conditions", "F value"]
p_value <- anova_summary[[1]]["Weather_Conditions", "Pr(>F)"]
cat("F-value:", f_value, "p-value:", p_value, "\n")

## F-value: 0.4077076 p-value: 0.8032037
```



The one-way ANOVA and boxplot confirm no significant impact of weather conditions on congestion levels, as medians and IQRs are similar across categories, with differences likely due to random variation.

Objective 3

```r
# Step 4: Perform Chi-squared test
chi_sq_result <- chisq.test(x = observed, p = expected_proportions)
cat("Chi-squared Test Result:\n")
print(chi_sq_result)

## Chi-squared Test Result:
##
##  Chi-squared test for given probabilities
##
## data:  observed
## X-squared = 1045.4, df = 4, p-value < 2.2e-16
```

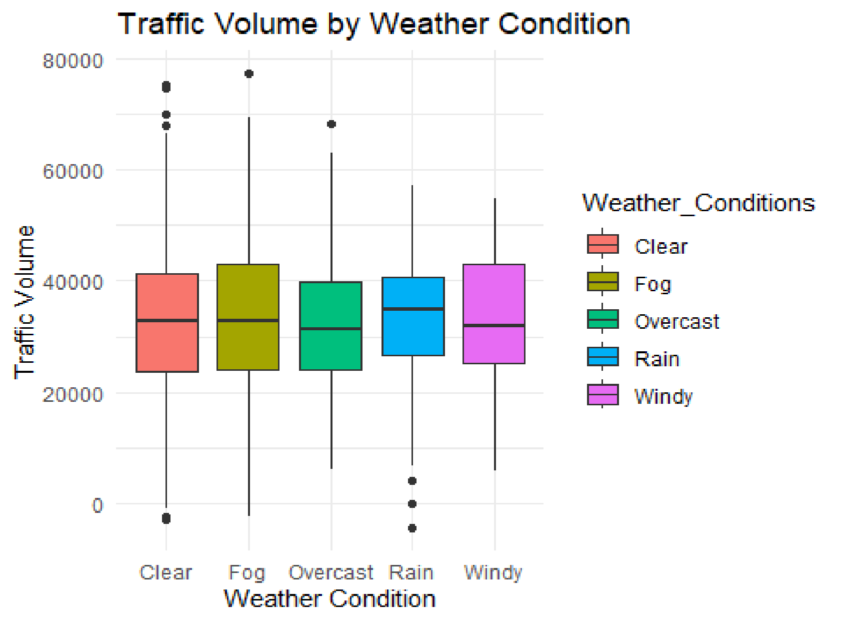


The Chi-squared test shows a significant mismatch between observed and expected weather condition frequencies ($p < 2.2e\text{-}16$), highlighting substantial deviations from the expected distribution.

Question 2 :Investigating the Effect of Weather Conditions on Traffic Volume

Objective 1

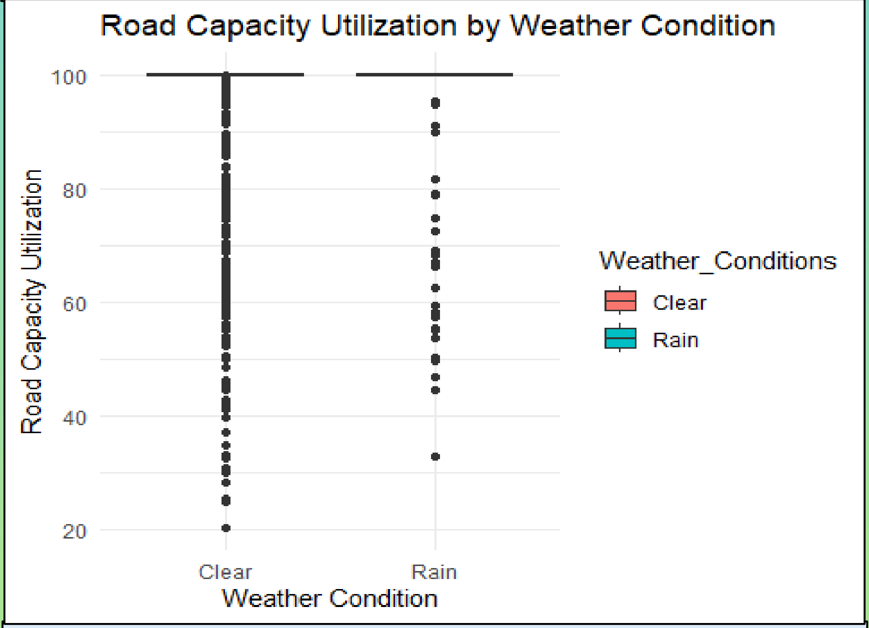
```r
# Main Test: Kruskal-Wallis Test
kruskal_result <- kruskal.test(Traffic_Volume ~ Weather_Conditions, da
ta = data)

# Display results
cat("Kruskal-Wallis Test Statistic:", kruskal_result$statistic, "\nP-v
alue:", kruskal_result$p.value, "\n")

## Kruskal-Wallis Test Statistic: 1.842464
## P-value: 0.7647049
```



The Kruskal-Wallis test (p = 0.7647) and boxplot indicate no significant difference in traffic volumes across weather conditions, suggesting minimal influence of weather on traffic volume.

Objective 2

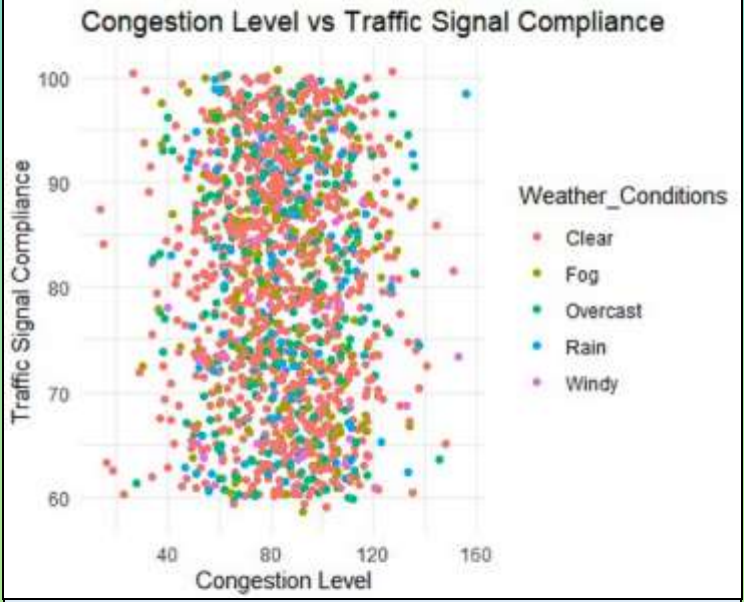
```r
Mann-Whitney U Test (Wilcoxon rank-sum test)
mann_whitney_result <- wilcox.test(clear_weather_data, rain_weather_data,
alternative = "two.sided", conf.level = 0.95)
print(mann_whitney_result)

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  clear_weather_data and rain_weather_data
## W = 62751, p-value = 0.8368
```



The Mann-Whitney U test (p = 0.8368) and boxplot show no significant difference in road capacity utilization between "Clear" and "Rain" conditions, indicating weather has minimal impact.

Objective 3

```r
# Main Test: Spearman Correlation
correlation_test <- cor.test(data$Congestion_Level,
data$Traffic_Signal_Compliance, method = "spearman")
}

print(correlation test)

##
##  Spearman's rank correlation rho
##
## data:  data$Congestion_Level and data$Traffic_Signal_Compliance
## S = 548492723, p-value = 0.2894
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.02761366
```

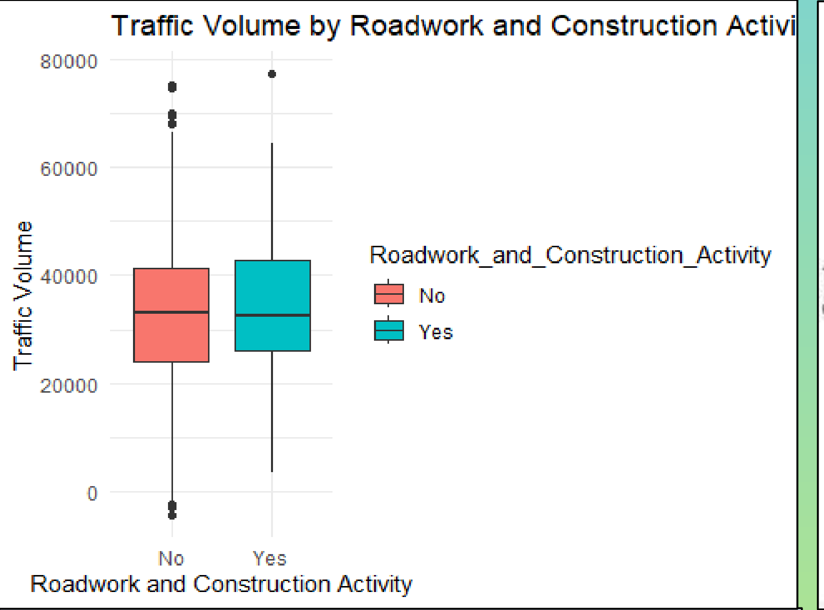


The Spearman's corr (ρ = -0.0276, p = 0.2894) and scatterplot indicate no significant monotonic relationship btw Congestion Level and Traffic Signal Compliance, suggesting no strong correlation between the two variables.

Question 3: Impact of Roadwork and Construction Activity on Traffic Volume

Objective 1

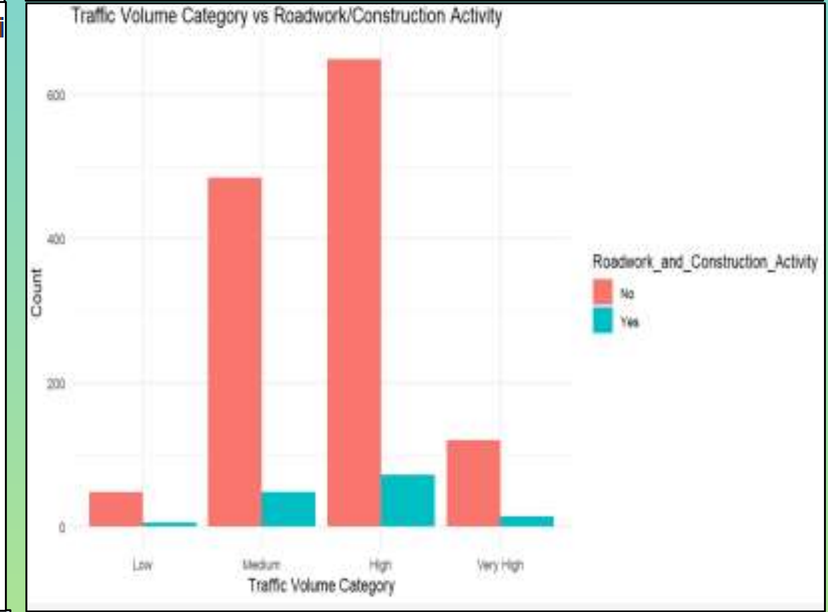
```
# Main Test: Mann-Whitney U Test (Wilcoxon rank-sum test)
mann_whitney_result <- wilcox.test(Traffic_Volume ~
Roadwork_and_Construction_Activity, data = data_subset)
print(mann_whitney_result)

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  Traffic_Volume by Roadwork_and_Construction_Activity
## W = 92315, p-value = 0.4777
```



The Mann-Whitney U test (p = 0.4777) and boxplot show no significant difference in traffic volume between groups with and without roadwork, suggesting that roadwork has no substantial impact on traffic volume.

Objective 2

```
# Perform Chi-squared test
chi_sq_test <- chisq.test(contingency_table)
cat("\nChi-squared Test for Independence:\n")

##
## Chi-squared Test for Independence:

print(chi_sq_test)

##
##  Pearson's Chi-squared test
##
## data:  contingency_table
## X-squared = 0.68579, df = 3, p-value = 0.8765
```

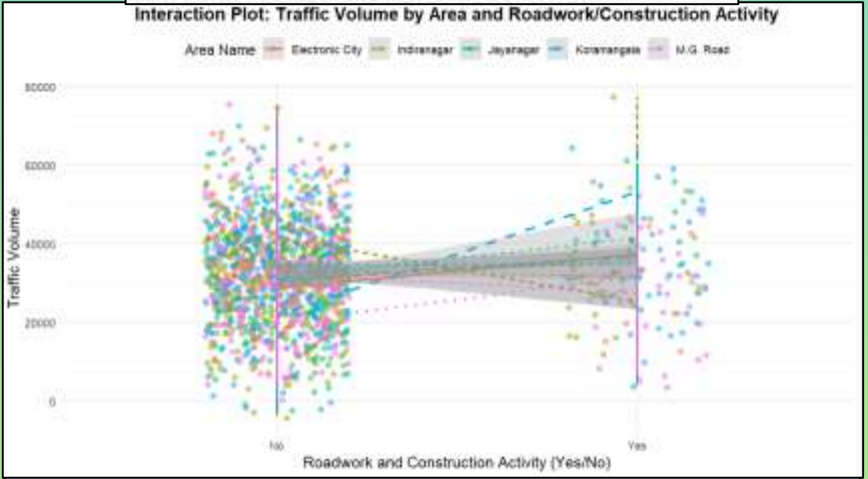


The Chi-squared test (p = 0.8765) and bar chart show no significant relationship between roadwork and traffic volume category, suggesting that roadwork does not affect traffic volume distribution.

Objective 3

```
## Conducting Two-Way ANOVA:

anova_result <- aov(Traffic_Volume ~ Area_Name *
Roadwork_and_Construction_Activity, data = data)
anova_summary <- summary(anova_result)
print(anova_summary)

##                                                Df    Sum Sq   Mean Sq F
value
## Area_Name                                        4 1.159e+09 289707987
1.723
## Roadwork_and_Construction_Activity               1 6.988e+07  69881220
0.416
## Area_Name:Roadwork_and_Construction_Activity     4 2.644e+09 661098518
3.931
## Residuals                                     1464 2.462e+11 168181159
##                                               Pr(>F)
## Area_Name                                    0.14243
## Roadwork_and_Construction_Activity           0.51929
## Area_Name:Roadwork_and_Construction_Activity 0.00353 **
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## F-value (Interaction Term): 3.930871
## F-tabulated (Critical): 2.378007
```

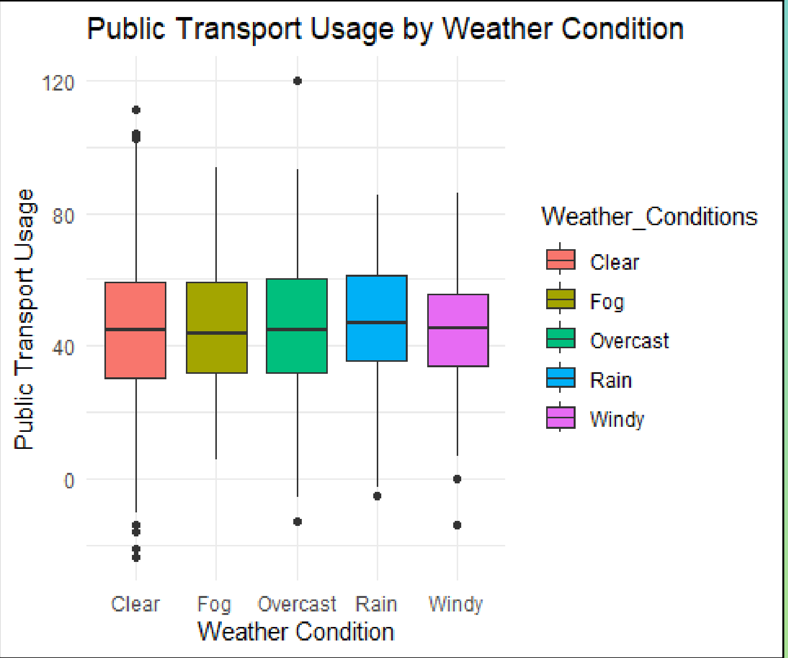


The Two-Way ANOVA reveals a significant interaction between area and roadwork, indicating that roadwork's impact on traffic volume varies by area, while the main effects of area and roadwork alone are not significant.

Question 4: Exploring the Effect of Weather Conditions on Public Transport Usage

Objective 1

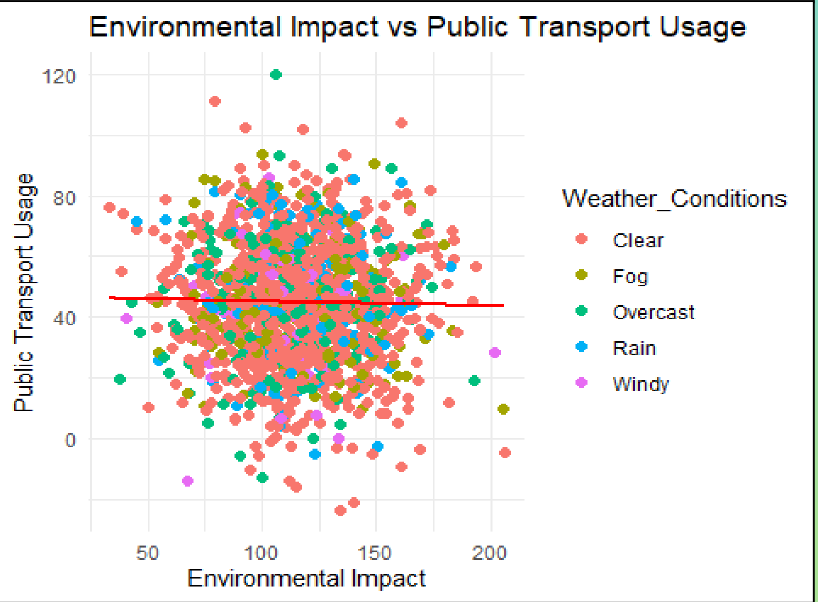
```
# Main Test: One-Way ANOVA
anova_result <- aov(Public_Transport_Usage ~ Weather_Conditions, data = data)
anova_summary <- summary(anova_result)
print(anova_summary)

##                      Df Sum Sq Mean Sq F value Pr(>F)
## Weather_Conditions    4   1249   312.3   0.787  0.534
## Residuals          1469 583018   396.9
```



The One-Way ANOVA (p = 0.534) and graph show no significant difference in public transport usage across weather conditions, suggesting weather does not strongly influence transport usage.

Objective 2

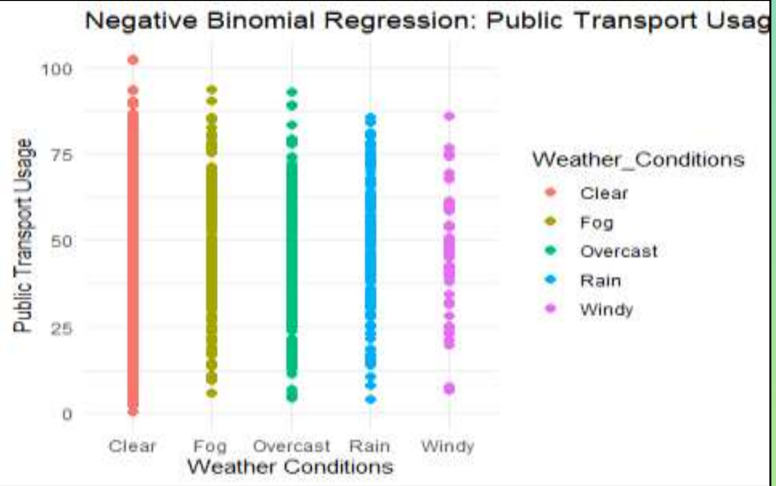
```
# Calculate Spearman rank correlation
spearman_corr <- cor.test(data$Environmental_Impact,
data$Public_Transport_Usage, method = "spearman")
print(spearman_corr)

##
##  Spearman's rank correlation rho
##
## data:  data$Environmental_Impact and data$Public_Transport_Usage
## S = 539285728, p-value = 0.6909
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##         rho
## -0.01036415
```



The Spearman rank correlation (ρ = -0.0104, p = 0.6909) and scatterplot show no significant monotonic relationship between Environmental Impact and Public Transport Usage, indicating no meaningful association between the variables.

Objective 3

```
## Negative Binomial Model Summary:
summary(negbinom_model)
##
## Call:
## glm.nb(formula = Public_Transport_Usage ~ Weather_Conditions,
##     data = data, init.theta = 5.443490081, link = log)
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                3.809159   0.015295 249.041   <2e-16 ***
## Weather_ConditionsFog      0.004649   0.037109   0.125    0.900
## Weather_ConditionsOvercast 0.012253   0.035239   0.348    0.728
## Weather_ConditionsRain     0.065160   0.041291   1.578    0.115
## Weather_ConditionsWindy    0.025212   0.068538   0.368    0.713
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(5.4435) family taken to
be 1)
##
##     Null deviance: 1531.3  on 1448  degrees of freedom
## Residual deviance: 1528.7  on 1444  degrees of freedom
## AIC: 12737
##
## Number of Fisher Scoring iterations: 1
##
##
##              Theta:  5.443
##          Std. Err.:  0.228
##
##  2 x log-likelihood:  -12725.254
```

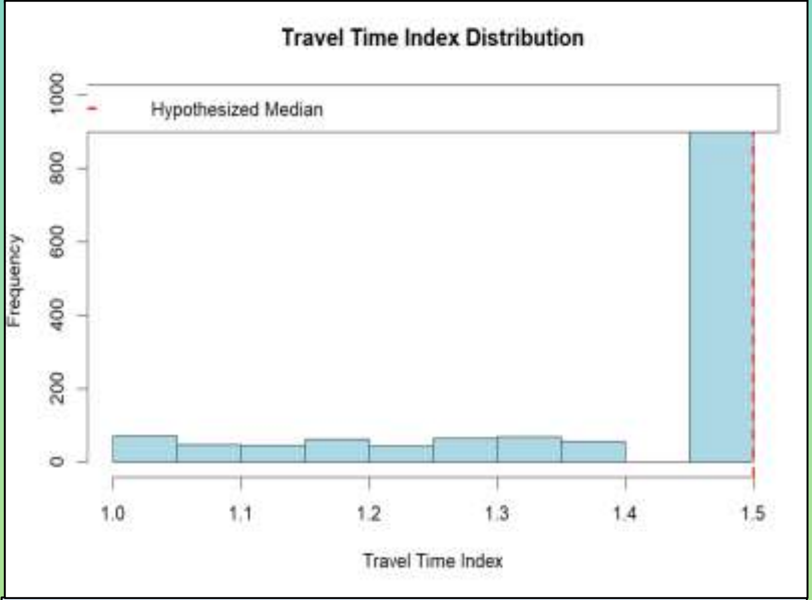


The Poisson regression shows that rainy weather significantly increases public transport usage (p < 0.0001), but overdispersion suggests that the Negative Binomial model is more appropriate for accurately capturing the data.

Question 5: Analyzing the Relationship Between Congestion Level and Travel Time Index

Objective 1

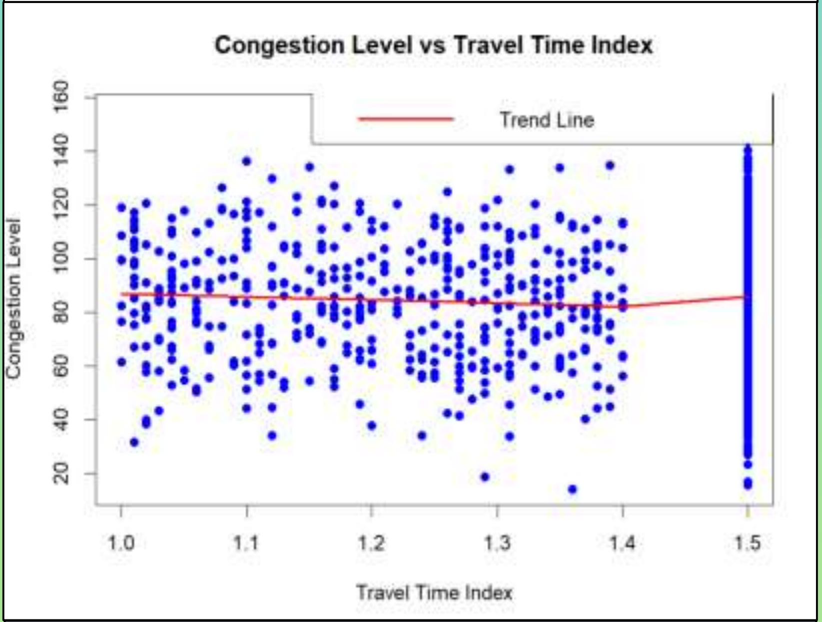
```
# Perform Wilcoxon Signed-Rank Test
wilcox_test <- wilcox.test(data$Travel_Time_Index, mu = mu, alternativ
e = "two.sided")
print(wilcox_test)

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  data$Travel_Time_Index
## V = 0, p-value < 2.2e-16
## alternative hypothesis: true location is not equal to 1.5
```



The Shapiro-Wilk and Wilcoxon Signed-Rank Test results, along with the histogram, show significant deviation from the hypothesized median of 1.5, indicating that the true median Travel Time Index differs from 1.5.

Objective 2

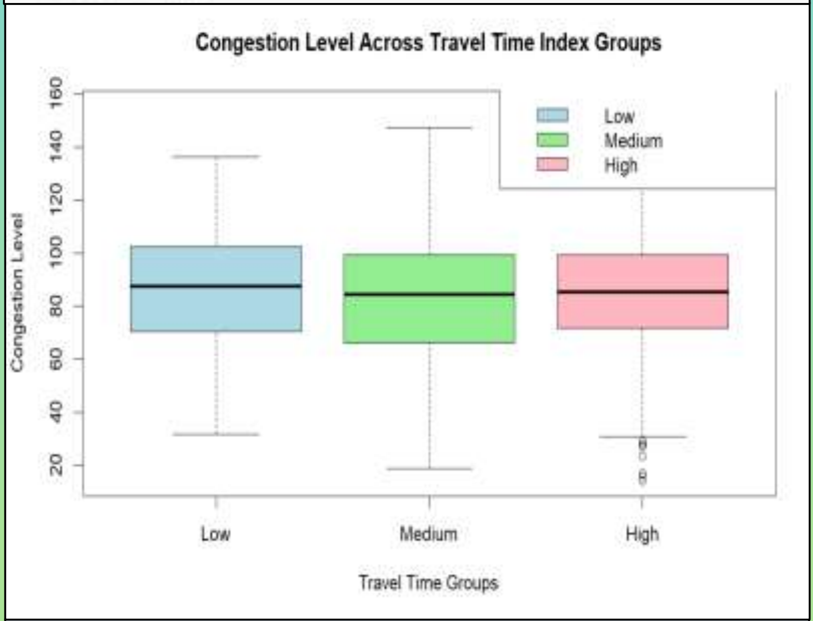
```
# Perform Spearman's correlation test
correlation_result <- cor.test(data$Congestion_Level, data$Travel_Time_Index,
                               method = "spearman")

## Spearman's Rank Correlation Coefficient: 0.01747416

cat("P-value:", correlation_result$p.value, "\n")

## P-value: 0.5062786
```



Spearman's Rank Correlation test and the scatterplot show no significant monotonic relationship between Congestion Level and Travel Time Index (ρ = 0.0175, p = 0.5063).

Objective 3

```
# Perform the Kruskal-Wallis test
kruskal_result <- kruskal.test(Congestion_Level ~ Travel_Time_Groups, data =
data)

# Display the results of the Kruskal-Wallis test
cat("Kruskal-Wallis Test Statistic:", kruskal_result$statistic, "\n")

## Kruskal-Wallis Test Statistic: 1.465099

cat("P-value:", kruskal_result$p.value, "\n")

## P-value: 0.480682
```



The Kruskal-Wallis test and boxplot show no significant difference in the medians of Congestion Level across the three Travel Time Index groups (Low, Medium, High) (p = 0.4807)

Conclusion

The study analyzed the impact of environmental and situational factors on urban traffic in Bengaluru. Employing rigorous statistical methods, the findings revealed:

1. **<u>Congestion Dynamics:</u>** No significant difference in congestion levels was observed between major areas or weather conditions, suggesting uniform traffic behavior across scenarios.
2. **<u>Weather and Traffic Volume:</u>** Weather conditions had minimal impact on traffic volume and road capacity utilization, indicating their limited influence on traffic flow.
3. **<u>Roadwork Effects:</u>** Roadwork and construction activities did not significantly affect traffic volumes, highlighting potential inefficiencies in mitigation strategies.
4. **<u>Public Transport Usage:</u>** Weather conditions did not significantly alter public transport usage, suggesting consistent usage patterns irrespective of environmental changes.
5. **<u>Interdependencies:</u>** Limited correlations between variables like congestion and compliance levels, or environmental impacts and transport usage, underscore the need for holistic policy interventions.

Future Work :

This study can be extended to other metropolitan areas to compare traffic patterns and develop generalized solutions. Incorporating real-time traffic and weather data could enhance predictive modeling for dynamic congestion management. Additionally, exploring the impact of emerging technologies like smart traffic systems and electric vehicles, along with analyzing commuter behavior, could provide deeper insights for user-centric urban mobility strategies.

References

- **Kaggle Datasets** (Link: https://www.kaggle.com/datasets/preethamgouda/banglore-city-traffic-dataset)

- **R Documentation** (Link: https://www.rdocumentation.org/)

- **Open Government Data Portal Karnataka** (Link: https://karnataka.data.gov.in/)

- **Open Government Data (OGD) Platform India** (Link: https://www.data.gov.in/)

- **Fundamentals of Mathematical Statistics Book By S.C. Gupta, VK Kapoor**

THANK YOU