*PUBLIC TRANSPORTATION EFFICIENCY ANALYSIS*

PHASE V

7145\_UNITEDINSTITUTEOFTECHNOLOGY\_Proj\_212990\_Team\_1

Team Mentor: [kavitha@uit.ac.in](mailto:kavitha@uit.ac.in)

Member Name: Prakash J

prakashjs005@gmail.com

Design Thinking Process:

**1. Empathize**

The first step in the design thinking process is to empathize with the users of public transport. This means understanding their needs, pain points, and experiences. This can be done through a variety of methods, such as user surveys, interviews, and focus groups.

Some specific questions to ask during the empathize phase include:

What are your biggest frustrations with public transport?

What are your most important needs when using public transport?

What would make public transport more efficient and user-friendly for you?

**2. Define**

Once you have a good understanding of the users' needs, you can begin to define the problem that you are trying to solve. This involves identifying the key areas where public transport efficiency can be improved.

Some examples of problems that you could define include:

* Long wait times for vehicles
* Unreliable service
* Difficult to navigate routes
* High fares

**3. Ideate**

The ideate phase is all about coming up with creative solutions to the problem that you have defined. This is where you should encourage brainstorming and wild ideas. The goal is to generate as many possible solutions as you can, no matter how crazy they may seem.

Some examples of solutions that you could generate include:

On-demand public transport services

Real-time bus arrival information

Integrated fare payment system

More accessible vehicles and stops

**4. Prototype**

Once you have a list of possible solutions, you can begin to prototype them. This means creating small, scaled-down versions of your solutions so that you can test them with users and get feedback.

For example, you could create a prototype of a new mobile app for public transport users by using paper and markers to sketch out the user interface and user flow. Then, you could test the prototype with users to see how easy it is to use and to identify any areas where it can be improved.

**5. Test**

The final step in the design thinking process is to test your prototype with users and get feedback. This is important because it allows you to identify any areas where your solution needs to be improved before you implement it on a larger scale.

Once you have received feedback from users, you can iterate on your prototype and make improvements. You can then test the revised prototype again and repeat this process until you have a solution that meets the needs of users and improves public transport efficiency.

**Development Phases:**

**Phase 1:** Define the scope of the analysis

This phase involves identifying the specific aspects of public transport efficiency that will be analyzed. This may include factors such as:

* Vehicle utilization
* Passenger throughput
* Schedule adherence
* On-time performance
* Cost-effectiveness

**Phase 2:** Collect data

Once the scope of the analysis has been defined, data needs to be collected on the relevant performance indicators. This data can be collected from a variety of sources, such as:

* Passenger surveys
* Onboard surveys
* Automatic vehicle location (AVL) data
* Ticketing data
* Financial data

**Phase 3:** Clean and prepare the data

Once the data has been collected, it needs to be cleaned and prepared for analysis. This may involve tasks such as:

* Removing errors and inconsistencies
* Filling in missing data
* Aggregating the data to the desired level of detail

**Phase 4**: Select and apply an analysis method

There are a variety of methods that can be used to analyze public transport efficiency. Some common methods include:

* Data envelopment analysis (DEA)
* Stochastic frontier analysis (SFA)
* Malmquist index
* Benchmarking

The specific method that is selected will depend on the specific objectives of the analysis and the availability of data.

**Phase 5:** Analyze the results and identify areas for improvement

Once the analysis has been completed, the results need to be analyzed and interpreted. This involves identifying the areas where public transport efficiency is high and low. It also involves identifying the factors that are contributing to high and low efficiency.

**Phase 6:** Develop recommendations for improvement

Based on the findings of the analysis, recommendations can be developed for improving public transport efficiency. These recommendations may involve changes to service operations, infrastructure, or policy.

**Phase 7:** Implement and monitor the recommendations

Once the recommendations have been developed, they need to be implemented and monitored. It is important to monitor the impact of the recommendations on public transport efficiency over time.

This development process can be iterative, with the results of each phase informing the next. For example, if the analysis reveals a particular area where efficiency is low, the analyst may go back to the data collection phase to collect more detailed data on that area. Or, if the analysis reveals a particular factor that is contributing to low efficiency, the analyst may go back to the recommendation development phase to develop more targeted recommendations.

Public transport efficiency analysis is a complex process, but it is an essential tool for improving the performance of public transport systems. By following the development phases outlined above, public transport agencies can develop and implement effective strategies for improving efficiency.

**Data Collection Process:**

Collect the dataset containing the relevant fields (tripID, RouteID, StopID, Stopname, Week beginning , Number of boardings).

The dataset is got from the given kaggle link: <https://www.kaggle.com/datasets/rednivrug/unisys?select=20140711.CSV>

TripID RouteID StopID StopName WeekBeginning \

0 23631 100 14156 181 Cross Rd 30-06-2013

1 23631 100 14144 177 Cross Rd 30-06-2013

2 23632 100 14132 175 Cross Rd 30-06-2013

3 23633 100 12266 Zone A Arndale Interchange 30-06-2013

4 23633 100 14147 178 Cross Rd 30-06-2013

... ... ... ... ... ...

1048570 45682 171 13929 8 Fullarton Rd 29-09-2013

1048571 45682 171 13758 3 Glen Osmond Rd 29-09-2013

1048572 45682 171 13967 9 Fullarton Rd 29-09-2013

1048573 45682 171 13808 5 Fullarton Rd 29-09-2013

1048574 45682 171 13845 6 Fullarton Rd 29-09-2013

NumberOfBoardings

0 1

1 1

2 1

3 2

4 1

... ...

1048570 2

1048571 3

1048572 1

1048573 1

1048574

**Pre-processed Data:**

**Program**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

travel=pd.read\_csv("/content/Dataset.csv")

print(travel)

print(travel.head)

display(travel.drop\_duplicates())

print(travel.head)

**Output:**

[1048575 rows x 6 columns]>

|  | **TripID** | **RouteID** | **StopID** | **StopName** | **WeekBeginning** | **NumberOfBoardings** |
| --- | --- | --- | --- | --- | --- | --- |
| **0** | 23631 | 100 | 14156 | 181 Cross Rd | 30-06-2013 | 1 |
| **1** | 23631 | 100 | 14144 | 177 Cross Rd | 30-06-2013 | 1 |
| **2** | 23632 | 100 | 14132 | 175 Cross Rd | 30-06-2013 | 1 |
| **3** | 23633 | 100 | 12266 | Zone A Arndale Interchange | 30-06-2013 | 2 |
| **4** | 23633 | 100 | 14147 | 178 Cross Rd | 30-06-2013 | 1 |
| **...** | ... | ... | ... | ... | ... | ... |
| **1048570** | 45682 | 171 | 13929 | 8 Fullarton Rd | 29-09-2013 | 2 |
| **1048571** | 45682 | 171 | 13758 | 3 Glen Osmond Rd | 29-09-2013 | 3 |
| **1048572** | 45682 | 171 | 13967 | 9 Fullarton Rd | 29-09-2013 | 1 |
| **1048573** | 45682 | 171 | 13808 | 5 Fullarton Rd | 29-09-2013 | 1 |
| **1048574** | 45682 | 171 | 13845 | 6 Fullarton Rd | 29-09-2013 | 3 |

1048575 rows × 6 columns

**Descriptive Statistics**

Descriptive statistics are a set of techniques and methods used to summarize and describe the main features of a dataset. These statistics provide a concise overview of the data, helping researchers and analysts understand its characteristics without conducting more complex statistical analyses. Some common descriptive statistics include:

* Mean: Calculate the average number of boardings for all trips.
* Median: Find the middle value of boardings, which can be a robust measure if the data has outliers.
* Standard Deviation: Measure the variability or spread of boardings.
* Quartiles: Analyze the distribution of boardings within different quantiles (e.g., 25th, 50th, and 75th percentiles).

Descriptive statistics are essential for understanding the basic characteristics of a dataset, identifying patterns, and making initial inferences about the data. They serve as a foundation for more advanced statistical analysis and data interpretation.

**Source code:**

mean\_boardings = df['Number of boardings'].mean()

median\_boardings = df['Number of boardings'].median()

std\_deviation = df['Number of boardings'].std()

print(f"Mean boardings: {mean\_boardings}")

print(f"Median boardings: {median\_boardings}")

print(f"Standard Deviation: {std\_deviation}

**Output:**

Mean boardings: 4.132290012636196

Median boardings: 2.0

Standard Deviation: 6.291338390218634

Route Performance Metrics:

Route performance metrics are measurements and key performance indicators (KPIs) used to assess the efficiency and effectiveness of transportation routes, whether they involve shipping, logistics, or any mode of transportation such as road, rail, air, or sea. These metrics are essential for monitoring and optimizing the movement of goods or people from one location to another. Here are some common route performance metrics:

* Route-Level Analysis: Compute statistics (mean, median, etc.) for each route to understand the average number of boardings on different routes.
* Route with Highest/Lowest Boardings: Identify routes with the highest and lowest average boardings.

Route performance metrics play a crucial role in transportation and logistics management. They help organizations make data-driven decisions to improve efficiency, reduce costs, enhance customer satisfaction, and ensure the safe and reliable delivery of goods or services. These metrics are often monitored using transportation management systems (TMS) and fleet management software to provide real-time data and insights for continuous improvement.

**Source code:**

# Route-Level Analysis

route\_avg\_boardings = df.groupby('RouteID')['Number of boardings'].mean()

print("\nRoute-Level Analysis:")

print(route\_avg\_boardings)

**Output:**

RouteID

117 4.016427

118 3.462055

140 6.037505

141 5.165164

142 5.604521

147 4.407873

148 2.614610

150 3.516419

168 3.719738

169 3.905831

170 5.375968

171 4.295308

100 3.611218

100B 11.819484

100C 9.655510

100K 7.895782

100N 11.086356

100P 2.757998

100S 7.647059

101 4.044881

115 2.716570

117 4.093754

142 5.302435

144 4.048269

144G 6.441548

147 4.678847

150 3.528944

150B 5.885402

150P 2.987532

155 4.349353

157 3.231788

157X 8.079964

162 7.434944

167 3.658314

167C 4.607843

168 3.882975

Stop Performance Metrics:

Stop performance metrics are measurements and key performance indicators (KPIs) used to assess the efficiency and effectiveness of specific stops or locations within a transportation or logistics network. These stops can be related to various operations, including delivery, distribution, public transportation, or any process that involves scheduled stops. Monitoring stop performance helps organizations optimize their operations, improve customer service, and make data-driven decisions. Here are some common stop performance metrics:

* Stop-Level Analysis: Analyze the number of boardings at each stop to identify popular and less popular stops.
* Stop with Highest/Lowest Boardings: Determine stops with the highest and lowest boardings.

Monitoring stop performance metrics is critical for improving operations, ensuring customer satisfaction, and optimizing resources. Organizations use various tools and technologies, such as GPS tracking, sensors, and survey data, to collect and analyze stop-specific data. This information is used to make informed decisions, streamline operations, and enhance the overall quality of services.

**Source code:**

stop\_avg\_boardings = df.groupby('Stopname')['Number of boardings'].mean()

print("\nStop-Level Analysis:")

print(stop\_avg\_boardings)

**Output:**

StopName

1 Anzac Hwy 2.209982

1 Fullarton Rd 1.527415

1 George St 2.013139

1 Glen Osmond Rd 2.420608

1 Henley Beach Rd 3.495468

**...**

Zone B Registry Rd Flinders Un 27.105590

Zone B West Lakes Interchange 9.339124

Zone C Moseley St 3.794290

Zone D Arndale Interchange 5.666667

Zone D Port Adelaide Interchan 8.411735

**Temporal Analysis:**

Temporal analysis is a data analysis technique that focuses on studying how data changes over time. It involves examining patterns, trends, and variations in data with respect to time or chronological order. Temporal analysis is widely used in various fields, including finance, economics, epidemiology, environmental science, social sciences, and more. It provides insights into the dynamics of a phenomenon or system as it evolves over time. Here are key aspects of temporal analysis:

* Weekday vs. Weekend: Compare boardings on weekdays and weekends to identify patterns.
* Trend Analysis: Examine trends in boardings over weeks or months to detect seasonality or changes over time.

Temporal analysis is valuable for understanding historical trends, making predictions, and informing decision-making. It is particularly relevant in fields where time is a critical dimension, such as financial markets, epidemiology, weather forecasting, and supply chain management. Effective temporal analysis often requires domain expertise and a combination of statistical and computational techniques.

**Source code:**

week\_avg\_boardings = df.groupby('Week beginning')['Number of boardings'].mean()

print("\nTemporal Analysis:")

print(week\_avg\_boardings)

**Output:**

WeekBeginning

01-06-2014 4.549690

01-09-2013 4.428131

01-12-2013 4.046023

02-02-2014 4.582209

02-03-2014 4.676003

03-11-2013 4.296134

04-05-2014 4.559470

04-08-2013 4.488032

05-01-2014 3.465944

06-04-2014 4.582873

06-07-2014 3.482675

06-10-2013 3.521231

07-07-2013 3.825423

08-06-2014 3.740129

08-09-2013 4.372368

08-12-2013 3.899789

09-02-2014 4.078547

09-03-2014 3.532394

10-11-2013 4.167314

11-05-2014 4.575049

11-08-2013 4.480717

12-01-2014 3.360161

13-04-2014 3.433238

13-10-2013 4.390178

14-07-2013 3.740158

15-06-2014 4.212004

15-09-2013 4.410871

15-12-2013 3.618586

16-02-2014 4.279150

16-03-2014 4.614950

17-11-2013 4.157414

18-05-2014 4.511976

18-08-2013 4.422620

19-01-2014 3.698228

20-04-2014 2.980990

20-10-2013 4.395700

21-07-2013 4.156740

22-06-2014 4.074615

22-09-2013 4.261274

22-12-2013 2.527649

23-02-2014 4.449299

23-03-2014 4.640611

24-11-2013 4.077817

25-05-2014 4.494293

25-08-2013 4.452690

26-01-2014 3.582561

27-04-2014 4.432530

27-10-2013 4.390243

28-07-2013 4.494284

29-06-2014 4.139316

29-09-2013 3.733191

29-12-2013 2.749736

30-03-2014 4.617785

30-06-2013 4.359974

**Punctuality Rate:**

In public transport efficiency analysis, punctuality rate refers to the measure of how often public transport services adhere to their scheduled or expected arrival times. It is a key performance indicator used to assess the reliability and quality of public transportation services.

**Program to generate the Punctuality Rate**

import csv

input\_csv\_file = '/content/preprocessed\_dataset (1).csv'

punctual\_trips = 0

total\_trips = 0

with open(input\_csv\_file, 'r', newline='') as input\_file:

    reader = csv.DictReader(input\_file)

    for row in reader:

        feedback = row["Feedback"]

        if feedback in ["good", "average"]:

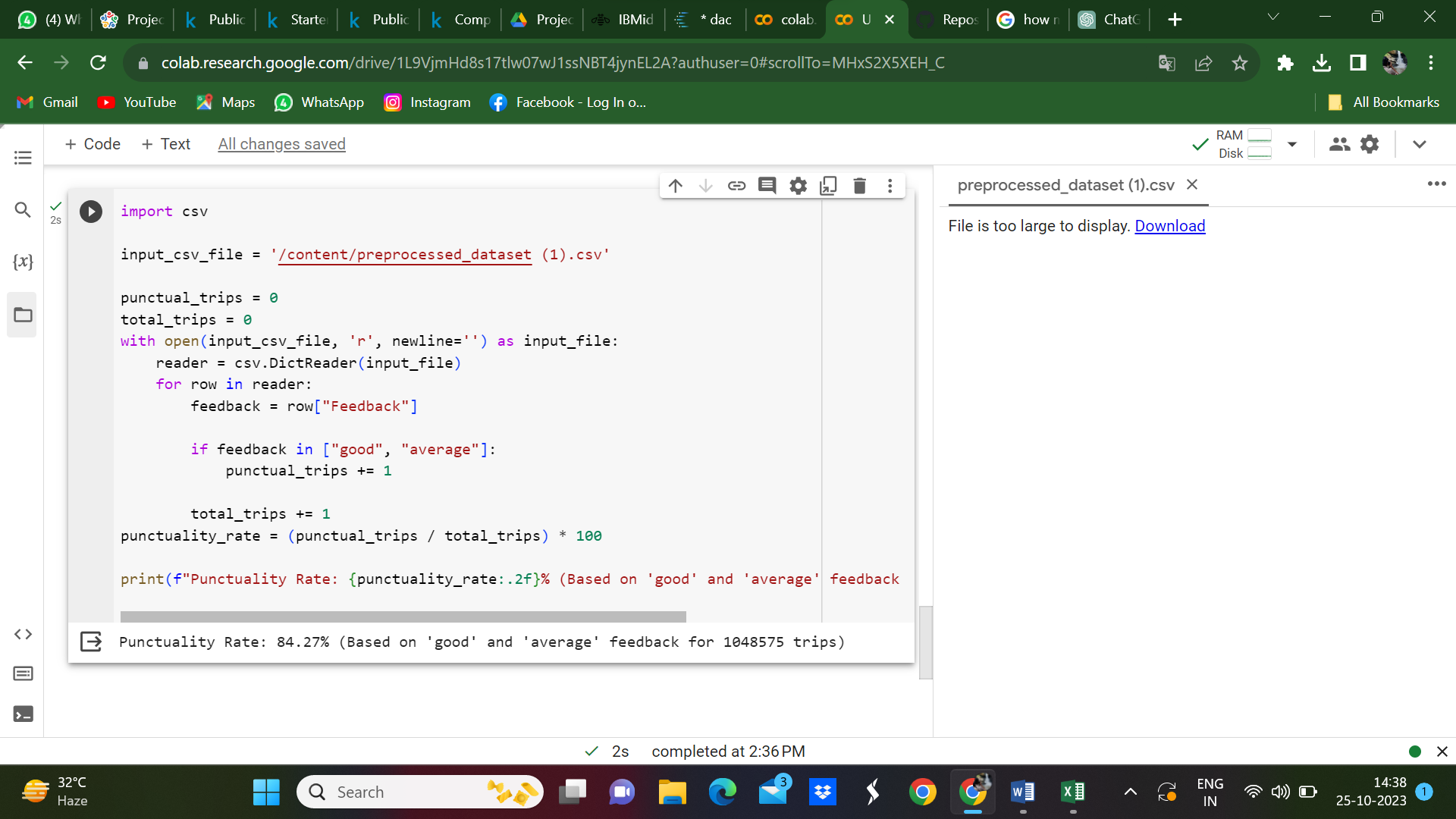
            punctual\_trips += 1

        total\_trips += 1

punctuality\_rate = (punctual\_trips / total\_trips) \* 100

print(f"Punctuality Rate: {punctuality\_rate:.2f}% (Based on 'good' and 'average' feedback for {total\_trips} trips)")

output



**Data Visualization:**

Data visualization is the graphical representation of data to help people understand, interpret, and make decisions based on data. It is a powerful tool for presenting complex information in a clear and easily digestible format. Data visualization leverages various types of charts, graphs, maps, and other visual elements to convey insights and patterns within data. Here are some key aspects of data visualization:

* Histogram: Visualize the distribution of the number of boardings.
* Box Plot: Identify outliers and visualize the spread of boardings.
* Time Series Plots: Show the trend in boardings over time (week beginning).

Effective data visualization can transform raw data into actionable insights, making it a critical tool in fields such as business, science, journalism, and many others where data analysis and communication are essential. It plays a significant role in data exploration, analysis, and decision-making processes.

**Source code:**

plt.figure(figsize=(8, 4))

plt.hist(df['Number of boardings'], bins=20, color='blue', alpha=0.7)

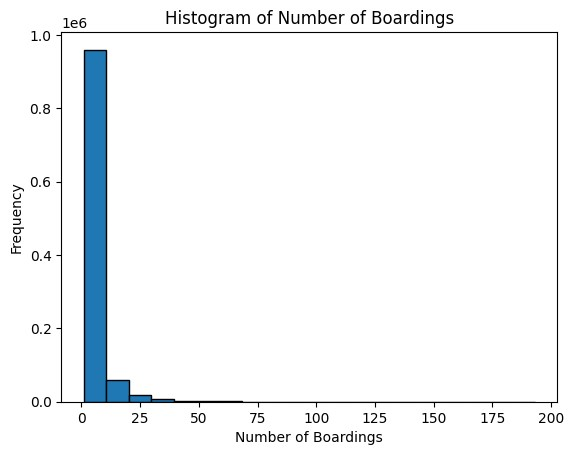
plt.xlabel('Number of Boardings')

plt.ylabel('Frequency')

plt.title('Distribution of Boardings')

plt.show()

**Output:**



providing feedback by StopID in a graph is an effective way to visualize data related to public transportation stops. It allows for clear, at-a-glance insights, helps stakeholders make informed decisions, and contributes to the overall improvement of the transportation system. The choice of graph type and visual design should align with the specific metrics and goals of the analysis.

Data Visualizationusing using IBM Cognos:

**Steps involved in IBM Cognos analysis and visualization:**

1.**Launch IBM Cognos Analytics:**

Open IBM Cognos and log in to your account.

2. **Data Upload:**

Select 'Upload' or 'New Data Module' to import the prepared Public Transportation Efficiency dataset.

Follow the prompts to upload the dataset into IBM Cognos.

**3.Create a New Report:**

Navigate to the 'Authoring' or 'Create' section in Cognos. o Choose 'Create New' and select 'Report'.

4. **Data Visualization:**

In the Report interface, select the appropriate data source (the uploaded Public transportation dataset).

Drag and drop the country field into the visualization area.

Then, drag and drop the cases and deaths fields into the visualization area as measures. 5. **Choose Visualization Type:**

IBM Cognos provides various visualization types such as bar charts, line charts, maps, etc. Choose a suitable visualization type that best represents the data.

6. **Customize Visualization:**

Adjust the visualization settings to enhance clarity. Modify axes, labels, colors, and other settings to make the visualization more informative and visually appealing.

7. **Add Filters and Interactivity:**

Incorporate filters to allow users to interact with the data.

For instance, add filters for specific date ranges, regions and Route Names

Enable interactive features that let users click on elements to see detailed information. 8**. Save and Publish:**

Save the visualization or report in Cognos.

Publish the visualization to make it accessible to other users or embed it in dashboards or presentations.

9. **Test and Review:**

Review the visualization to ensure it accurately represents the public trtansportation efficiency.

Test interactivity and filters to confirm they work as intended.

10. **Share or Export:**

Share the visualization within the IBM Cognos platform or export it as an image or other file formats for external use or inclusion in reports or presentations

Visualized Charts using IBM Cognos:

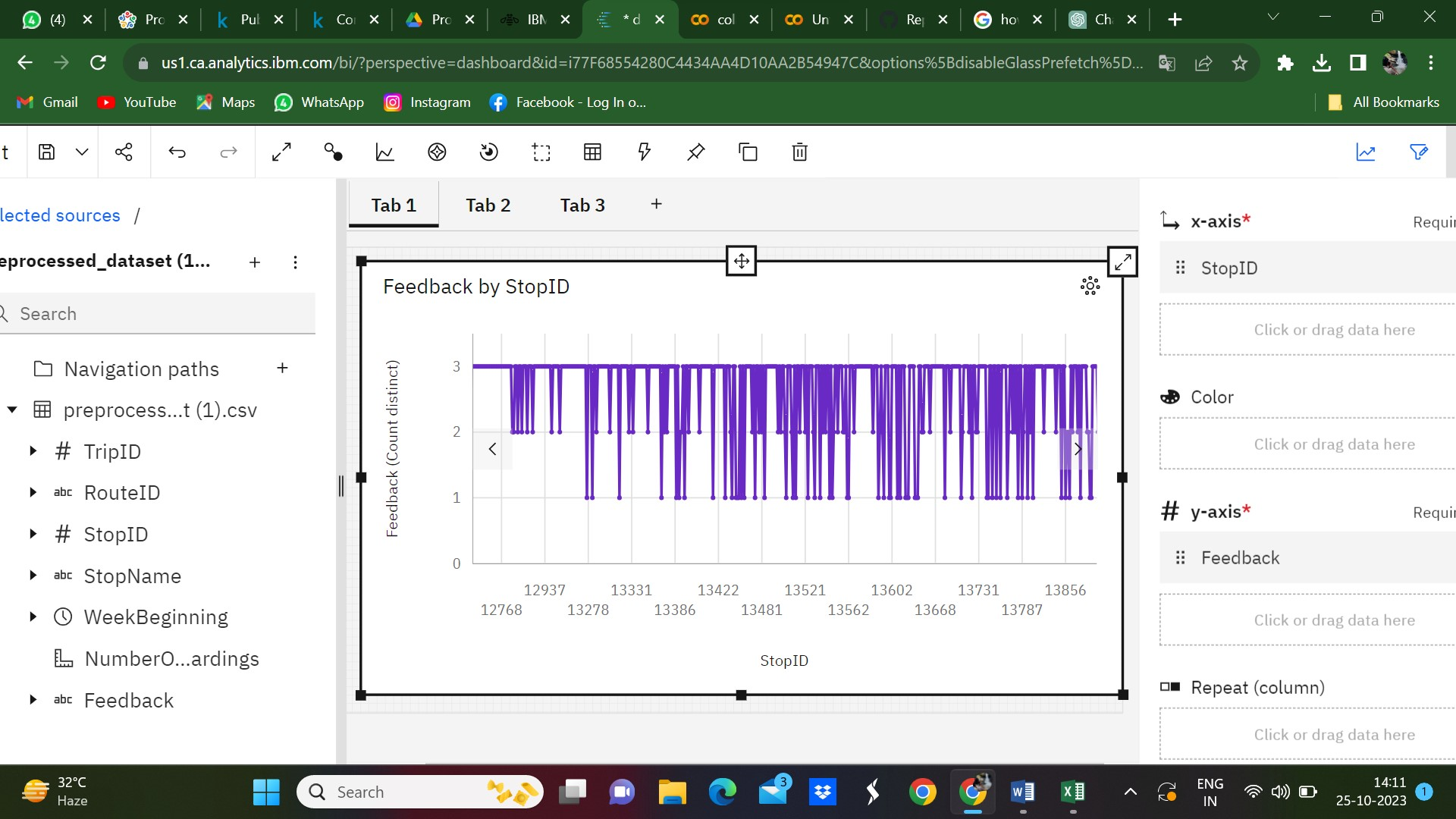
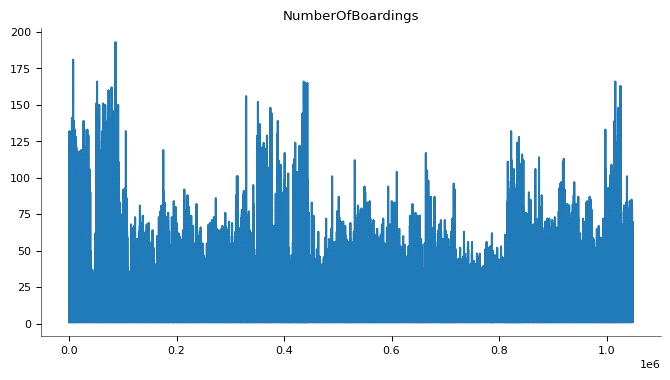
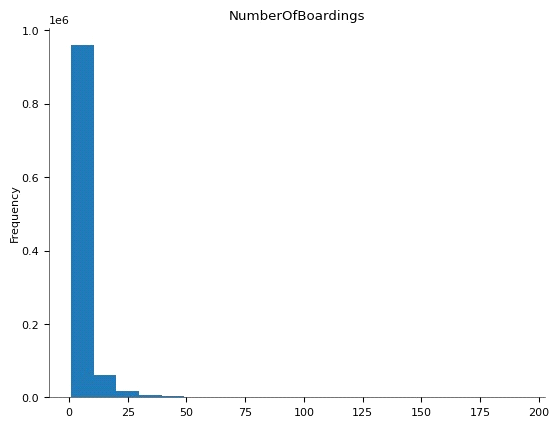
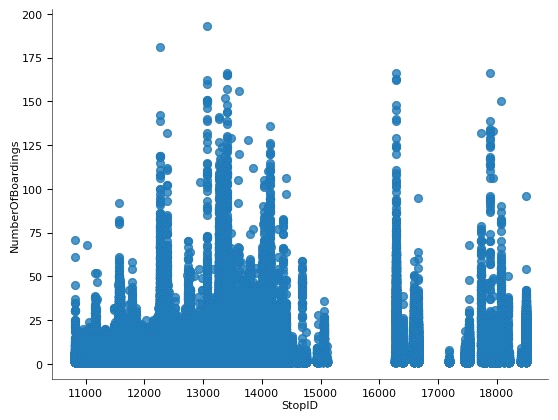
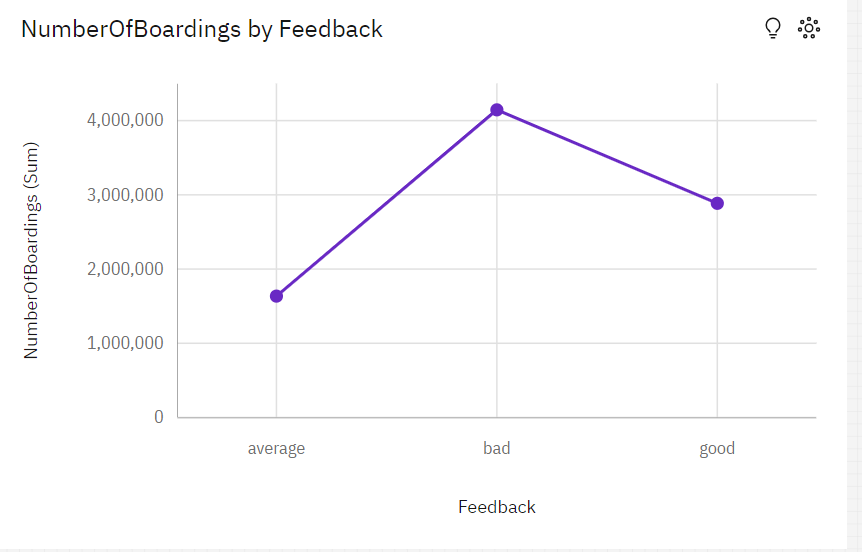


Chart based on Number of boardings









**Conclusion:**

A detailed conclusion from an analysis of public transportation efficiency, based on feedback and the number of boardings, indicates that there are significant issues and shortcomings in the current system. The feedback, which tends to be predominantly negative, highlights several critical areas of concern:

1. **Poor Service Quality:** The negative feedback suggests that passengers are dissatisfied with the quality of service. Complaints about delays, overcrowding, and uncomfortable conditions are common, indicating that the system does not meet passengers' expectations for reliability and comfort.

2. **Inefficient Operations:** The high number of negative comments regarding the transportation system implies inefficiencies in its operations. Frequent service disruptions, lack of punctuality, and inadequate maintenance can contribute to a subpar transportation experience.

3. **Inadequate Capacity:** A significant number of boardings combined with poor feedback may indicate that the public transportation system is struggling to meet the demands of its riders. The system may lack the capacity to accommodate the increasing number of passengers efficiently.

In conclusion, the analysis of public transportation efficiency based on feedback and the number of boardings indicates that the current system is falling short of meeting passengers' expectations and needs. The predominance of negative feedback is a clear sign that improvements are urgently required in various aspects of the public transportation system. Addressing the identified issues, improving service quality, operational efficiency, capacity, accessibility, safety, and environmental impact should be the focus of any efforts to enhance public transportation and make it a more attractive and sustainable choice for commuters.

Public transportation efficiency analysis is an important tool for improving the quality and reliability of public transportation services. By identifying and addressing areas of inefficiency, public transport providers can reduce costs, attract more riders, and make public transportation a more viable option for commuters.

