PHASE 5: PROJECT DOCUMENTATION & SUBMISSION

TOPIC: PUBLIC TRANSPORT EFFIENCY ANALYSIS

INTRODUCTION:

The project aims to analyze public transportation data to assess service efficiency, on-time performance, and passenger feedback. The analysis is crucial for transportation improvement initiatives and enhancing the overall public transportation experience.

ANALYSIS OBJECTIVES:

1. Assess On-Time Performance: Evaluate the punctuality of public transportation services, helping identify routes or stops with on-time issues.

2. Analyze Passenger Satisfaction: Examine passenger feedback and satisfaction to pinpoint areas needing improvement.

3. Evaluate Service Efficiency: Understand the efficiency of public transportation services, with a focus on the number of boardings at specific stops.

DATA COLLECTION PROCESS:

- Data Source: The data is collected from a reliable source, such as the provided dataset, which includes information about trips, routes, stops, and passenger data.

- Data Cleaning and Preprocessing: Python code is integrated to clean and preprocess the raw transportation data. This includes handling missing values, standardizing data formats, and removing outliers.

DATA VISUALIZATION USING IBM COGNOS:

- Visualization Strategy: Data is visualized using IBM Cognos, which offers powerful tools for creating informative dashboards and reports. The visualization strategy includes scatter plots, bar plots, box plots, violin plots, and line plots to convey insights effectively.

- Scatter Plot: Used to display on-time performance by visualizing the number of boardings against WeekBeginning, with the option to distinguish performance by RouteID.

- Bar Plot: Depicts the total number of boardings per StopName, enabling insights into high-traffic stops.

- Box Plot: Illustrates the distribution of the number of boardings by RouteID, providing a view of performance across different routes.

- Violin Plot: Shows the distribution of the number of boardings by StopName, helping to identify variations in performance at specific stops.

- Line Plot: Displays trends of the number of boardings by WeekBeginning, allowing for analysis of passenger trends over time.

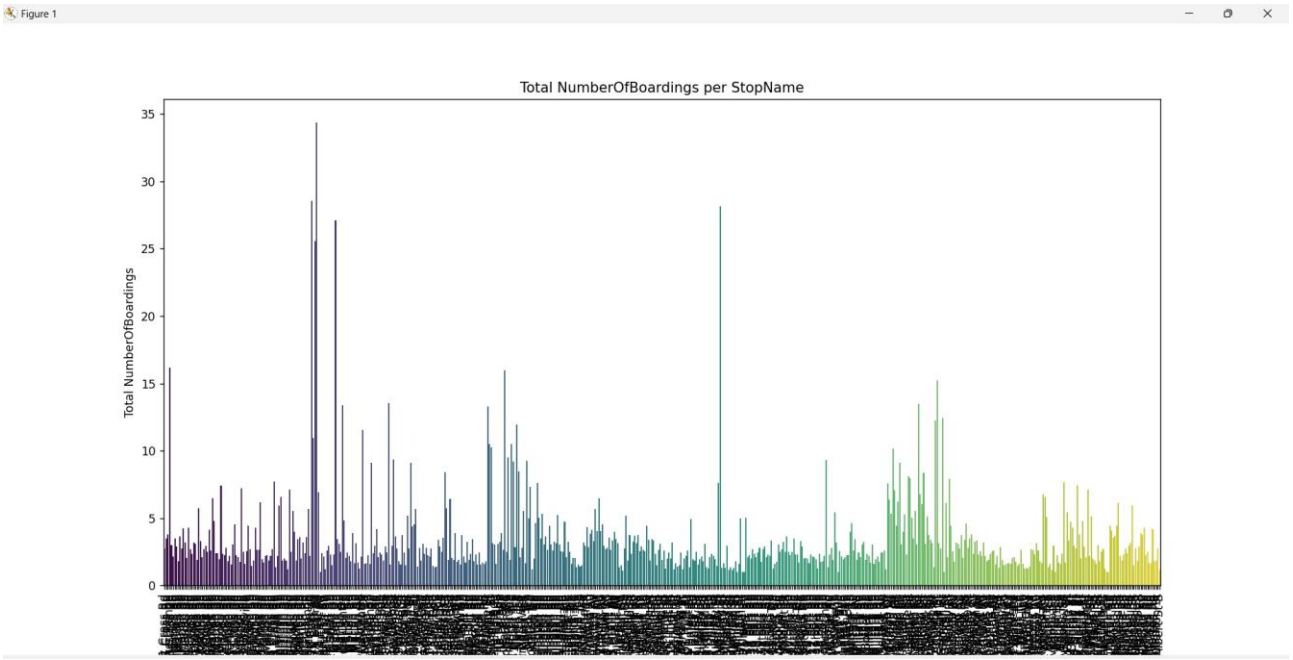
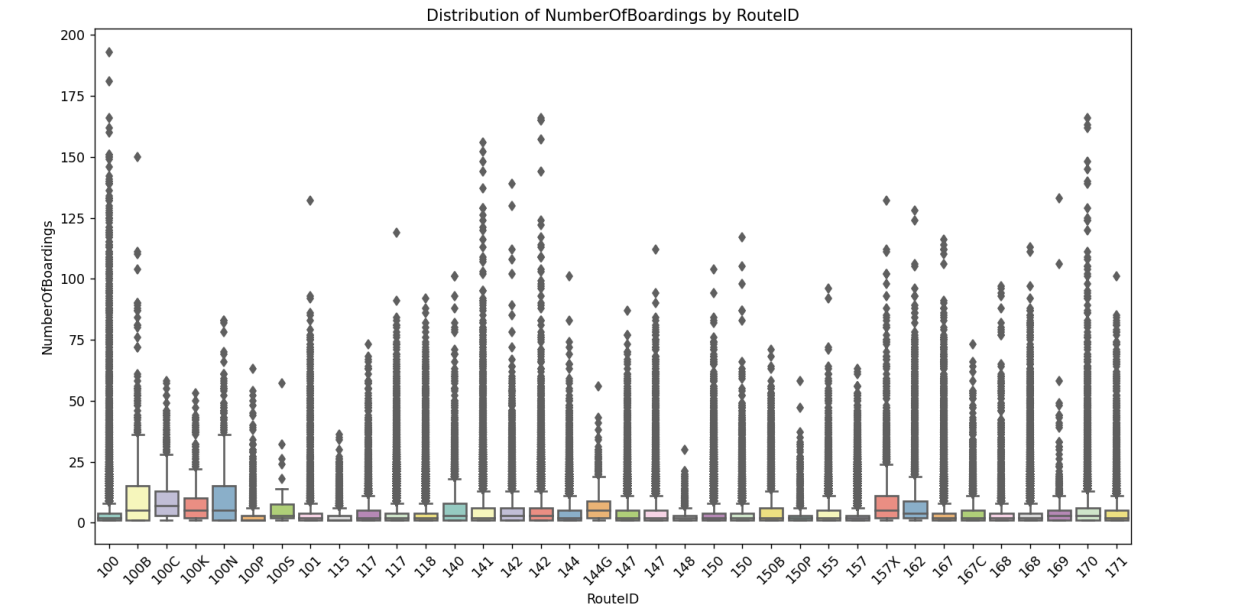
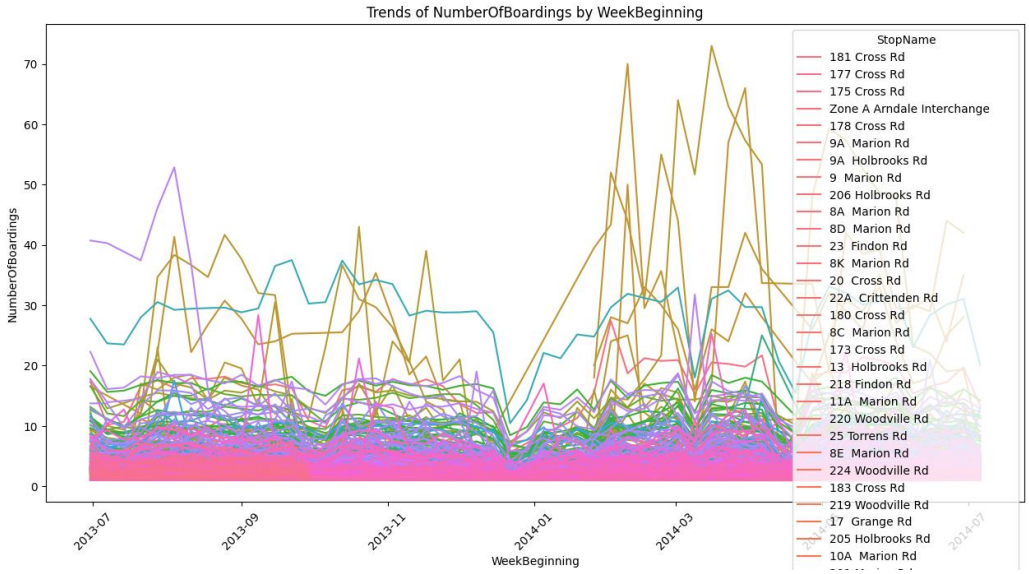
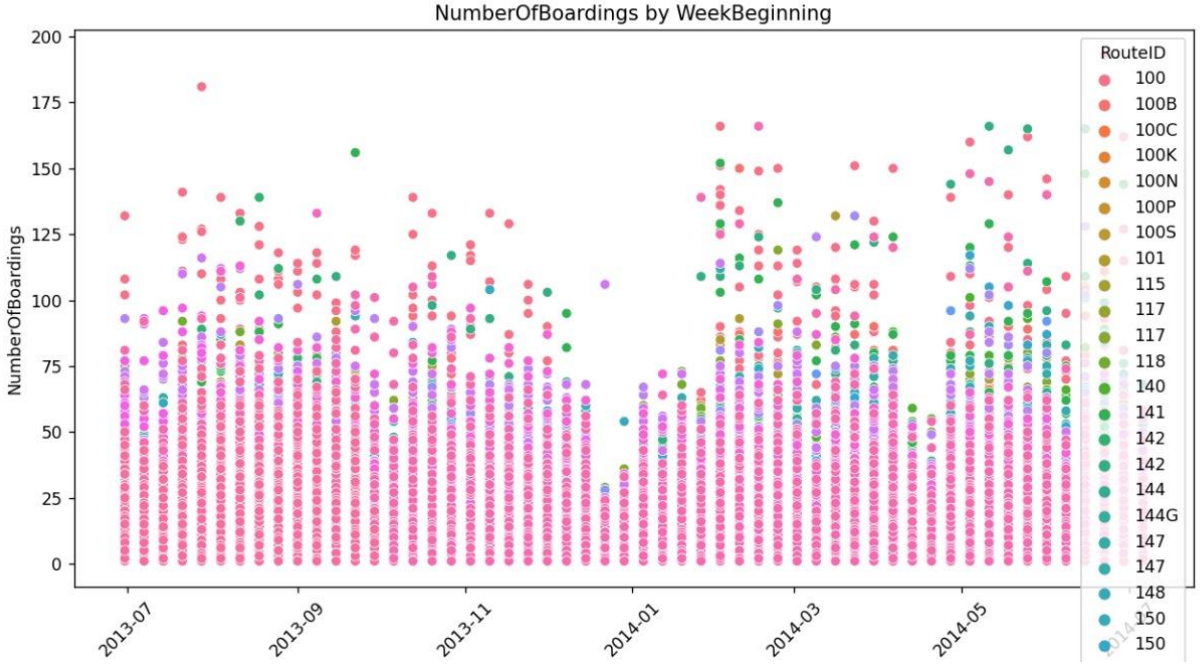


Figure 1
Total NumberOfBoardings per StopName
Total NumberOfBoardings
0
5
10
15
20
25
30
35


Distribution of NumberOfBoardings by RouteID
NumberOfBoardings
0
25
50
75
100
125
150
175
200
RouteID
100
100B
100C
100K
100N
100P
100S
101
115
117
117
118
140
141
142
142
144
144G
147
147
148
150
150
150B
150P
155
157
157X
162
167
167C
168
168
169
170
171





PYTHON CODE INTEGRATION:

Python is integrated for data cleaning and preprocessing, as well as generating the visualizations. It ensures that the data is prepared for effective analysis in IBM Cognos.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset and specify data types for columns with mixed data types
data = pd.read_csv('nm_dataset.csv', dtype={'StopName': 'object'},
parse_dates=['WeekBeginning'])

# Function to create a scatter plot
def create_scatter_plot():
    plt.figure(figsize=(12, 6))  # Increase the plot size
    sns.scatterplot(data=data, x='WeekBeginning', y='NumberOfBoardings',
hue='RouteID')
    plt.title('NumberOfBoardings by WeekBeginning')
    plt.xlabel('WeekBeginning')
    plt.ylabel('NumberOfBoardings')
    plt.xticks(rotation=45)
    plt.legend(title='RouteID')
    plt.show()

# Function to create a bar plot
def create_bar_plot():
    plt.figure(figsize=(18, 10))  # Increase the plot size
    sns.barplot(data=data, x='StopName', y='NumberOfBoardings', ci=None,
palette='viridis')
    plt.title('Total NumberOfBoardings per StopName')
    plt.xlabel('StopName')
    plt.ylabel('Total NumberOfBoardings')
    plt.xticks(rotation=90)
    plt.show()

# Function to create a box plot
def create_box_plot():
    plt.figure(figsize=(14, 7))  # Increase the plot size
    sns.boxplot(data=data, x='RouteID', y='NumberOfBoardings', palette='Set3')
    plt.title('Distribution of NumberOfBoardings by RouteID')
    plt.xlabel('RouteID')
    plt.ylabel('NumberOfBoardings')
    plt.xticks(rotation=45)
    plt.show()

# Function to create a violin plot
def create_violin_plot():
    plt.figure(figsize=(16, 8))  # Increase the plot size
    sns.violinplot(data=data, x='StopName', y='NumberOfBoardings',
palette='Set2')
    plt.title('Distribution of NumberOfBoardings by StopName')
    plt.xlabel('StopName')
```

```
    plt.ylabel('NumberOfBoardings')
    plt.xticks(rotation=90)
    plt.show()

# Function to create a line plot
def create_line_plot():
    plt.figure(figsize=(16, 8))  # Increase the plot size
    sns.lineplot(data=data, x='WeekBeginning', y='NumberOfBoardings',
hue='StopName', ci=None)
    plt.title('Trends of NumberOfBoardings by WeekBeginning')
    plt.xlabel('WeekBeginning')
    plt.ylabel('NumberOfBoardings')
    plt.xticks(rotation=45)
    plt.legend(title='StopName')
    plt.show()

# Call the functions to generate the plots
create_scatter_plot()
create_bar_plot()
create_box_plot()
create_violin_plot()
create_line_plot()
```

HOW INSIGHTS IMPROVE USER EXPERIENCE:

- On-Time Performance: Identifying routes with on-time issues allows transportation authorities to take corrective actions, resulting in more reliable services and improved user experience.

- Passenger Satisfaction: Analyzing passenger feedback helps pinpoint areas of concern and allows for proactive improvements, enhancing passenger satisfaction.

- Service Efficiency: By assessing the number of boardings at different stops, transportation providers can allocate resources more effectively, reduce overcrowding, and optimize the user experience.

CONCLUSION:

In conclusion, the analysis of public transportation data provides valuable insights for transportation improvement and enhancing the overall public transportation experience. By assessing on-time performance, passenger satisfaction, and service efficiency, transportation authorities can make data-driven decisions that lead to a better user experience. The integration of Python code for data cleaning and preprocessing ensures that the data is ready for analysis using IBM Cognos, which offers powerful visualization tools for conveying these insights effectively.

TEAM MEMBERS:

1. SRIVISHWA P -2021115109
2. SURYA NARAYANAAN K C -2021115113
3. SWASTHIKA LAKSHMI PRIYA R -2021115115
4. VISHNU PRIYA K -2021115121
5. SIVA L -2021115330