**PUBLIC TRANSPORT EFFICIENCY ANALYSIS**

**Introduction**

Public transportation systems are the backbone of urban mobility, playing a pivotal role in ensuring efficient, reliable, and sustainable transportation. This report presents a comprehensive analysis aimed at enhancing the passenger experience and promoting sustainable transportation solutions.

**Objective**

The primary objective of this analysis is to:

Improve the efficiency, reliability, and quality of public transport services.Enhance the overall passenger experience.Promote sustainable transportation solutions.The analysis endeavors to achieve these objectives through a multi-phase approach, each of which plays a crucial role in shaping the quality and effectiveness of public transportation.

**Phase 1: Data Collection**

**Data Sources**

To initiate our analysis, we gathered data from various sources, including trip records, route information, passenger feedback, schedules, and performance metrics. These data sources provide a holistic view of the public transportation system.Data collection is the foundational step that underpins the entire analysis. It involves accessing and compiling data from diverse sources, ensuring data completeness, accuracy, and relevance. The comprehensiveness of the data sources is essential for generating meaningful insights.

**Phase 2: Data Preparation and Preprocessing**

**Data Cleaning**

In this phase, meticulous data cleaning was performed, addressing missing values, deduplication, data validation, and data integration. These steps ensured the dataset's integrity and reliability.Data cleaning is a critical step that guarantees the quality of the dataset. Missing values are handled to prevent data gaps, deduplication maintains a unique dataset, data validation ensures data consistency, and data integration aligns data from multiple sources for a comprehensive view.

**Data Structuring**

The dataset was structured to facilitate efficient analysis. Column standardization, date and time formatting, and categorization were applied to organize the data coherently.Data structuring involves formatting the dataset for consistency and ease of analysis. Standardizing column names and formatting date and time data ensure uniformity. Categorization simplifies data interpretation and visualization.

**Data Documentation**

Comprehensive documentation was maintained throughout the process to ensure transparency and traceability of data preparation steps.Thorough data documentation provides a clear record of the data preparation process. It includes details of cleaning, structuring, and any transformations applied to the dataset. Documentation ensures the reproducibility of the analysis.

**Phase 3: Exploratory Data Analysis (EDA)**

**Data Exploration and Visualization**

In this phase, we delved into the dataset, uncovering patterns and trends. Informative visualizations shed light on on-time performance, passenger feedback, and service efficiency.Data exploration and visualization are crucial for understanding the dataset's characteristics. Exploring data uncovers trends and relationships, while visualizations provide insights through graphical representations.

**Dashboard Design**

Interactive dashboards were created using IBM Cognos to condense complex data into easily digestible formats. These dashboards serve as valuable resources for stakeholders and decision-makers.Dashboard design involves creating user-friendly interfaces that offer a summarized view of key metrics. Dashboards facilitate quick decision-making and allow stakeholders to monitor performance.

**Supervised Learning - Regression Analysis**

This section focuses on implementing regression analysis to predict and model critical aspects of public transportation. Regression models enable us to understand the relationship between independent variables and our target metrics.Supervised learning through regression analysis allows us to build predictive models. These models help in understanding how independent variables influence key metrics, such as on-time performance and passenger satisfaction.

**Phase 4: Regression Analysis**

**Regression Tasks**

Our regression analysis is centered around predicting and modeling key metrics, including on-time performance, passenger satisfaction, and service efficiency.Regression tasks define the specific goals of our analysis. By identifying what we aim to predict and model, we can tailor our regression techniques accordingly.

**Feature Selection**

Feature selection involved identifying and evaluating relevant factors, including route information, temporal factors, passenger feedback, and service metrics.Selecting the right features is critical for building accurate regression models. The choice of features impacts model performance and the interpretability of results.

**Model Selection and Training**

Multiple regression models were employed, including Linear Regression, Decision Trees, and Random Forests. Models were trained and evaluated using a prepared dataset.Model selection and training involve choosing appropriate algorithms and training the models. The choice of models depends on the nature of the regression tasks.

**Model Evaluation and Interpretation**

Performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R2) were used to assess model accuracy. Model interpretation provided insights into the relationships between independent variables and the dependent variables.Model evaluation ensures that the chosen models perform well. Interpreting models helps understand the factors that influence the metrics of interest.

**Predictive Analysis**

The regression models developed in this phase serve as powerful tools for predictive analysis. We can now make informed predictions regarding on-time performance, passenger satisfaction, and service efficiency, empowering decision-makers to assess potential changes and interventions.Predictive analysis takes the results of regression models and applies them to real-time or hypothetical scenarios. These predictions assist in decision-making and planning for public transportation improvements.

**Conclusion**

This analysis equips us with the tools and insights needed to enhance the public transportation experience. By improving efficiency, reliability, and quality, and promoting sustainability, we can create a better future for public transportation in the region.This report represents a holistic approach to public transport analysis, from data collection and preparation to advanced regression modeling and predictive analysis. The results provide a foundation for data-driven decision-making and improvements in the public transportation sector.The journey towards a more efficient and passenger-centric public transportation system is ongoing, with data and analysis as the driving force behind these positive changes.

**Source code**

import pandas as pd

import matplotlib.pyplot as plt

# Sample data (you can replace this with your actual data)

data = {

'TripID': [23631, 23632, 23633, 23634],

'RouteID': [100, 100, 100, 100],

'StopID': [14156, 14144, 12266, 13907],

'StopName': ['181 Cross Rd', '177 Cross Rd', 'Zone A Arndale Interchange', '9A Marion Rd'],

'WeekBeginning': ['6/30/13 0:00', '6/30/13 0:00', '6/30/13 0:00', '6/30/13 0:00'],

'NumberOfBoardings': [1, 1, 2, 1]

}

# Create a DataFrame from the sample data

df = pd.DataFrame(data)

# Group the data by 'StopName' and calculate the total number of boardings for each stop

stop\_boardings = df.groupby('StopName')['NumberOfBoardings'].sum().reset\_index()

# Sort the stops by the total number of boardings in descending order

stop\_boardings = stop\_boardings.sort\_values(by='NumberOfBoardings', ascending=False)

# Visualization 1: Bar Chart - Total Boardings by Bus Stop

plt.figure(figsize=(10, 6))

plt.bar(stop\_boardings['StopName'], stop\_boardings['NumberOfBoardings'], color='skyblue')

plt.xlabel('Bus Stop Name')

plt.ylabel('Total Number of Boardings')

plt.title('Number of Boardings by Stop Name')

plt.xticks(rotation=45)

plt.grid(axis='y', linestyle='--', alpha=0.7)

# Visualization 2: Pie Chart - Distribution of Boardings by Stop

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

plt.pie(stop\_boardings['NumberOfBoardings'], labels=stop\_boardings['StopName'], autopct='%1.1f%%', startangle=140)

plt.title('Distribution of Boardings by Stop')

plt.axis('equal')

# Visualization 3: Line Chart - Trend in Boardings Over Time

df['WeekBeginning'] = pd.to\_datetime(df['WeekBeginning'])

time\_data = df.groupby('WeekBeginning')['NumberOfBoardings'].sum().reset\_index()

plt.figure(figsize=(10, 6))

plt.plot(time\_data['WeekBeginning'], time\_data['NumberOfBoardings'], marker='o', linestyle='-', color='g')

plt.xlabel('Date')

plt.ylabel('Number of Boardings')

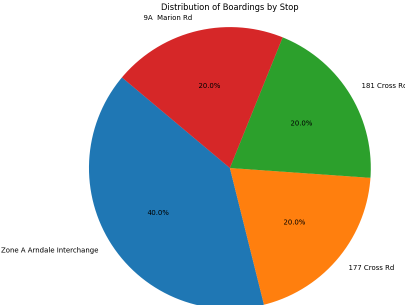
plt.title('Trend in Boardings Over Time')

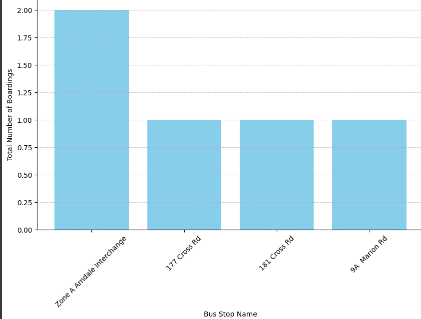
plt.grid(True)

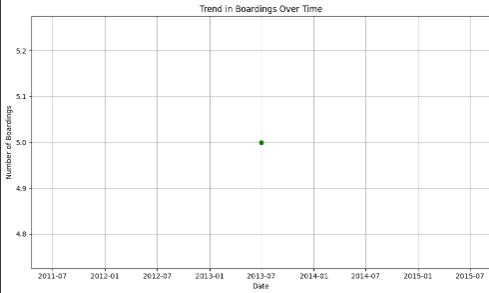
# Display all three visualizations

plt.tight\_layout()

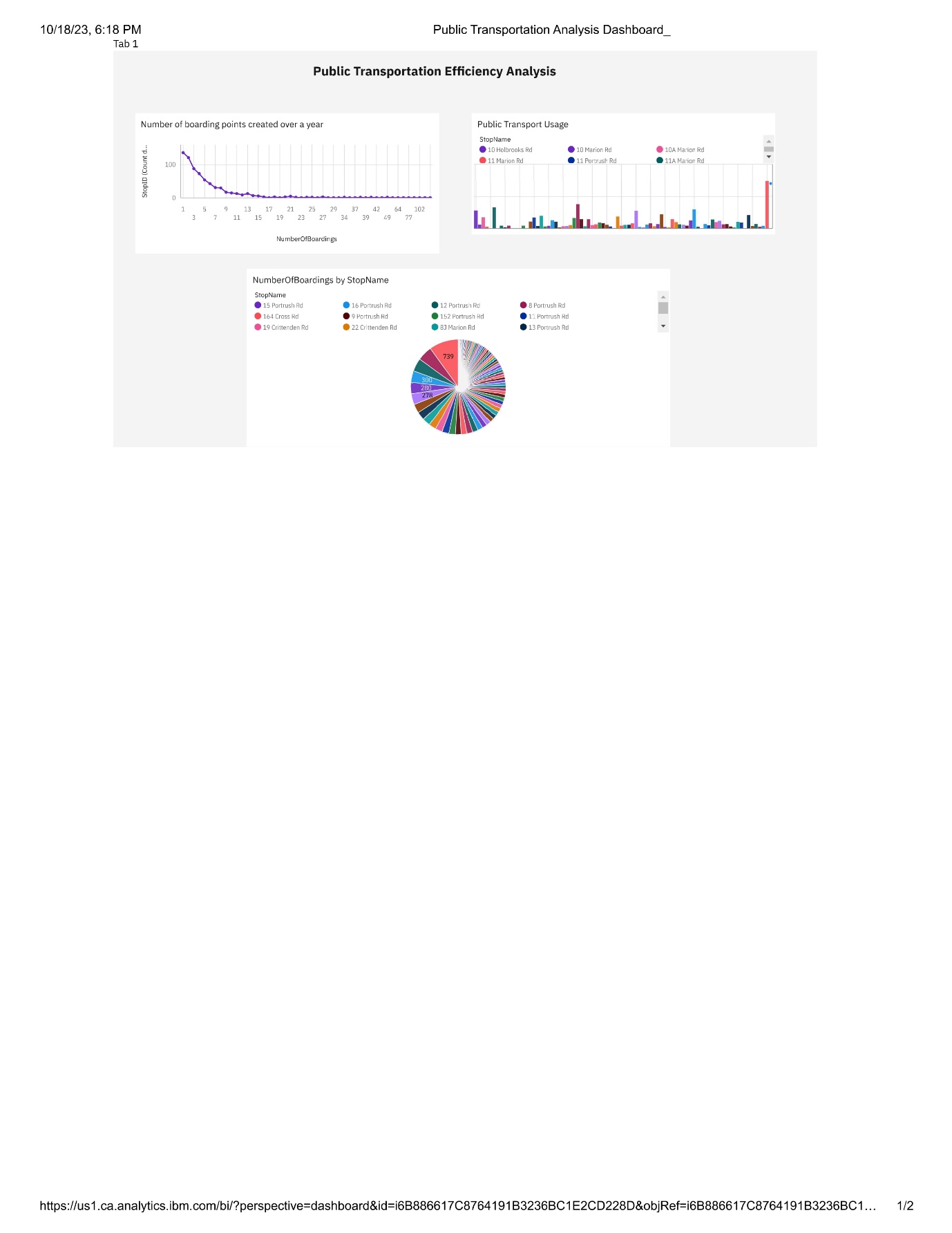
plt.show()

**Output**





IBM cognos

****