PRIYADARSHINI ENGINEERING COLLEGE

Public Transport Efficiency Analysis IBM Cognos Phase-3

Introduction:

Analysing public transport efficiency with IBM Cognos involves utilizing its robust business intelligence and data analytics capabilities to gain valuable insights from the data collected. Here's a step-by-step guide on how to perform a public transport efficiency analysis using IBM Cognos.

Objectives:

This project emphasizes the importance of data analysis and visualization to enhance the efficiency, reliability, and sustainability of public transportation services. Key takeaways from this project include:

- Data-Driven Decision Making: The project demonstrates the power of data-driven decision-making in the domain of public transportation. By collecting, processing, and visualizing data, stakeholders can make informed decisions to improve the quality of service.

- Real-Time Monitoring: The inclusion of real-time data analysis allows for on-the-fly adjustments to routes, schedules, and maintenance. This ensures that transportation services can respond dynamically to changing conditions.

- Passenger Insights: Analysis of passenger data, such as boarding and alighting patterns, preferences, and behavior, helps optimize service delivery and improve the overall passenger experience.

- Environmental Impact: The project can also include an analysis of the environmental impact of the transportation system, providing insights into sustainability and opportunities for reducing emissions.

- Multi-Modal Integration: By integrating various modes of transportation and visualizing data from multiple sources, the project can provide passengers with seamless, convenient, and interconnected mobility options.

- Data Visualization: Data visualization plays a crucial role in presenting complex transportation data in an easily understandable and actionable format. Visualizations like maps, charts, and graphs help stakeholders identify trends, bottlenecks, and areas for improvement.

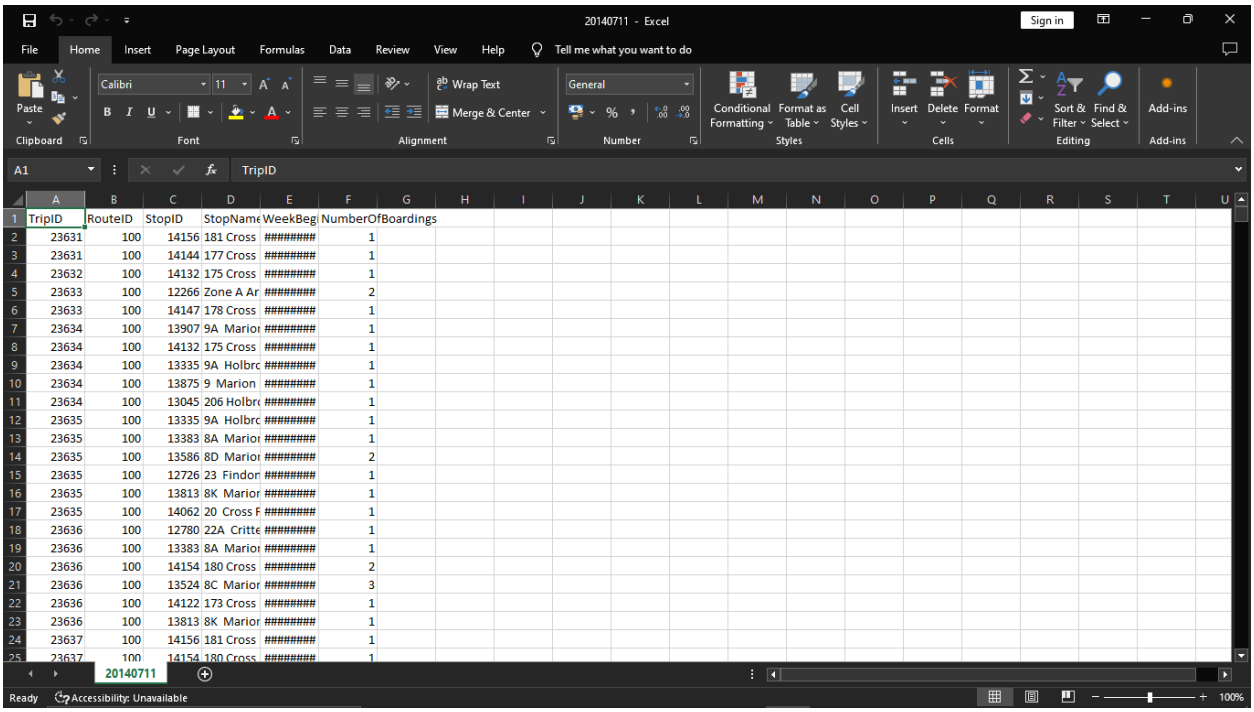
- Efficiency Optimization: The program demonstrates the use of algorithms and analysis to optimize routes, schedules, and infrastructure to minimize inefficiencies and reduce congestion.

- Sustainability and Innovation: The project promotes sustainable transportation and encourages innovation, such as integrating electric or autonomous vehicles into the system.

In summary, the public transportation efficiency analysis project with data visualization using Python empowers transportation authorities, city planners, and policymakers to make evidence-based decisions that lead to more efficient, sustainable, and passenger-friendly public transportation systems. Data analysis and visualization are essential tools in the ongoing effort to enhance urban mobility and reduce the environmental impact of transportation.

DATASET:

Source:
https://www.kaggle.com/datasets/rednivrug/unisys?select=20140711.CSV



TripID	RouteID	StopID	StopName	WeekBegi	NumberOfBoardings
23631	100	14156	181 Cross	########	1
23631	100	14144	177 Cross	########	1
23632	100	14132	175 Cross	########	1
23633	100	12266	Zone A Ar	########	2
23633	100	14147	178 Cross	########	1
23634	100	13907	9A Mario	########	1
23634	100	14132	175 Cross	########	1
23634	100	13335	9A Holbro	########	1
23634	100	13875	9 Marion	########	1
23634	100	13045	206 Holbr	########	1
23635	100	13335	9A Holbro	########	1
23635	100	13383	8A Mario	########	1
23635	100	13586	8D Mario	########	2
23635	100	12726	23 Findor	########	1
23635	100	13813	8K Mario	########	1
23635	100	14062	20 Cross F	########	1
23636	100	12780	22A Critte	########	1
23636	100	13383	8A Mario	########	1
23636	100	14154	180 Cross	########	2
23636	100	13524	8C Mario	########	3
23636	100	14122	173 Cross	########	1
23636	100	13813	8K Mario	########	1
23637	100	14156	181 Cross	########	1
23637	100	14154	180 Cross	########	1

The above dataset has: 1048553 Rows.

What is data visualization?

Data visualization is the representation of data in graphical or pictorial format. It involves the use of visual elements like charts, graphs, maps, and other graphical elements to help people understand and interpret data. Data visualization is a powerful tool for conveying complex information, patterns, and insights in a more intuitive and accessible way than raw data or text alone.

The primary goals of data visualization are:

Data Exploration: It allows analysts and data scientists to explore data to identify patterns, trends, anomalies, and relationships within the dataset.

Data Communication: Data visualization makes it easier to communicate data-driven insights to a broader audience, including stakeholders and decision-makers. Visualizations can simplify complex concepts and facilitate understanding.

Data Analysis: Visualizations can assist in the analysis of data, helping to test hypotheses and derive meaningful conclusions.

Decision-Making: Visualizing data can aid in making informed decisions, as it provides a clear and concise way to understand data and its implications.

Common types of data visualizations include:

- Bar Charts: Suitable for comparing categories or groups of data.
- Line Charts: Ideal for showing trends and changes over time.
- Pie Charts: Useful for illustrating parts of a whole or proportions.
- Scatter Plots: Display relationships between two variables.
- Heatmaps: Depict data using colour intensity, suitable for matrices and correlations.
- Geospatial Maps: Show data on geographical maps.
- Histograms: Display the distribution of a single variable.
- Box Plots: Show the distribution, central tendency, and outliers of a dataset.
- Sankey Diagrams: Illustrate flow and connections in a system.

Importing libraries in Python is a fundamental step in most programming tasks. Libraries contain pre-written code and functions that you can use to perform various tasks, from data manipulation to machine learning. Here's how you import libraries in Python:

1.Using the import Statement:

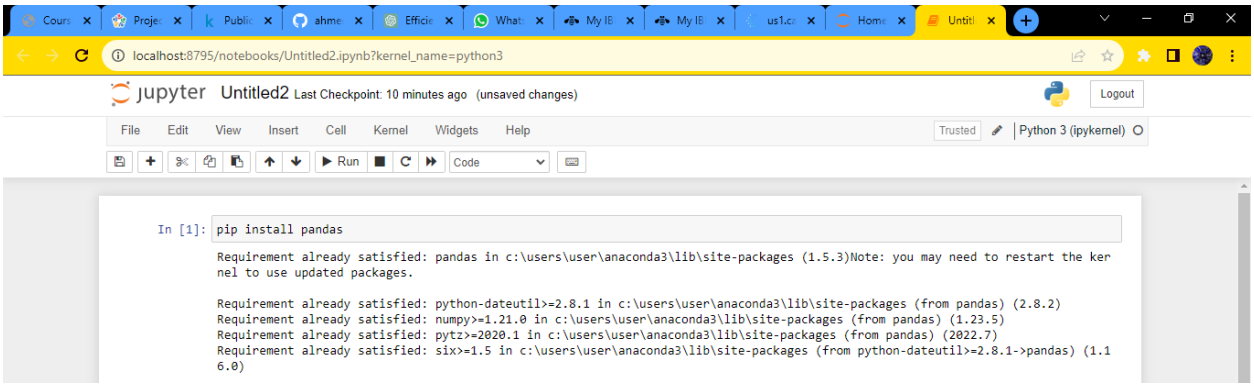
The most common way to import libraries is by using the import statement. You simply specify the library/module name after import. You can also use an alias to make it easier to refer to the library in your code.

2.Using Specific Functions/Classes:

You can also import specific functions or classes from a library if you don't need the whole library. This can reduce memory usage and improve code readability.

Python Program For Data Visualization For Public Transportation Efficiency Analysis:

1. Installing pandas:

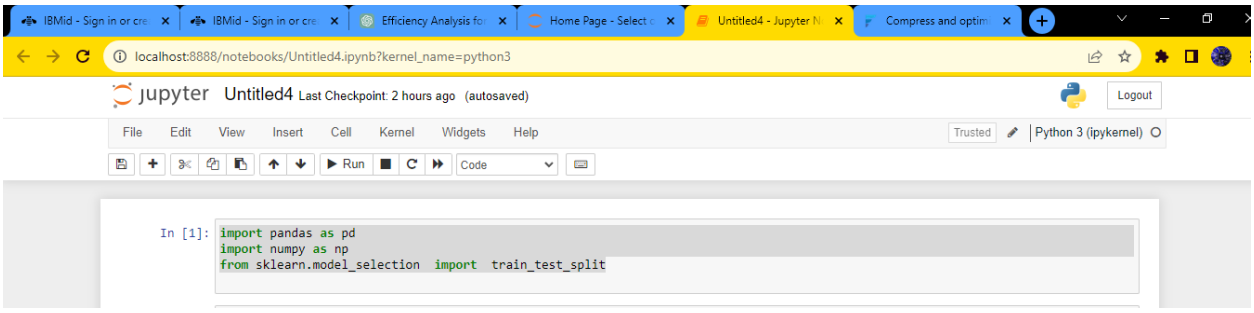


2. Importing Libraries:

import pandas as pd

import numpy as np

from sklearn.model_selection import train_test_split



3. Reading the dataset:

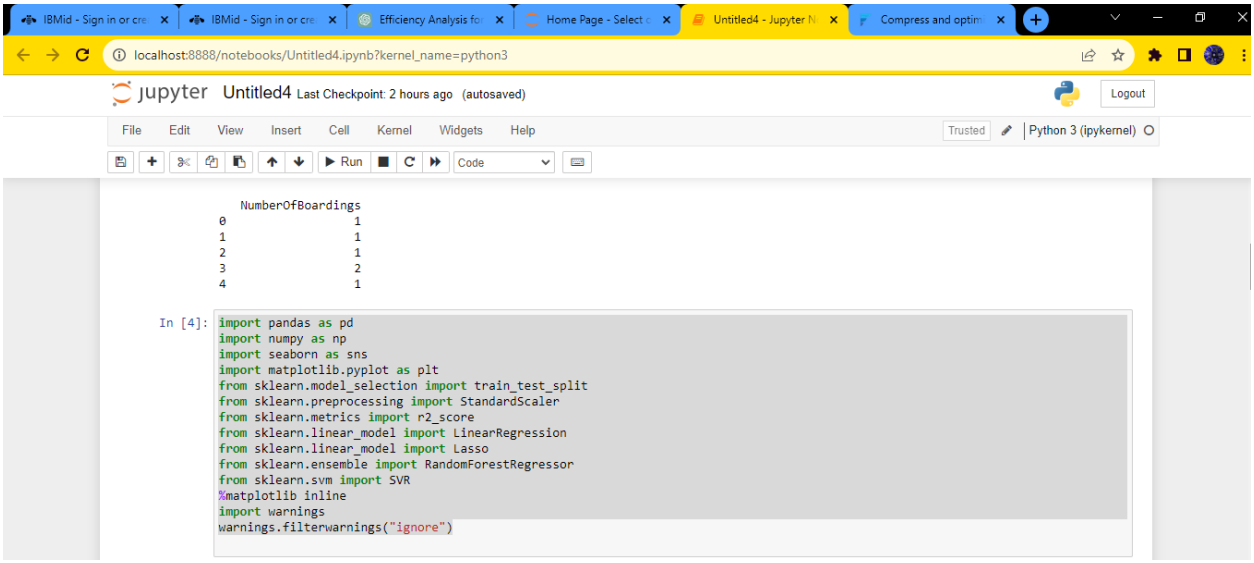
```
import pandas as pd


# Replace 'your_file.csv' with the actual path to your CSV file.
file_path = 'C:/Project_dataset.csv'


# Read the CSV file using pandas
try:
    df = pd.read_csv(file_path)


    # Display the first few rows of the dataframe
    print("First few rows of the CSV file:")
    print(df.head())
except FileNotFoundError:
    print(f"File not found at {file_path}. Please provide a valid file path.")
except Exception as e:
    print(f"An error occurred: {e}")
```

Output:

```
C:\Users\User\AppData\Local\Temp\ipykernel_3484\2929465056.py:8: DtypeWarning: Columns (1) have mixed types. Specify dtype opti
on on import or set low_memory=False.
  df = pd.read_csv(file_path)

First few rows of the CSV file:
   TripID RouteID  StopID                    StopName        WeekBeginning  \
0   23631     100   14156               181 Cross Rd  2013-06-30 00:00:00
1   23631     100   14144               177 Cross Rd  2013-06-30 00:00:00
2   23632     100   14132               175 Cross Rd  2013-06-30 00:00:00
3   23633     100   12266  Zone A Arndale Interchange  2013-06-30 00:00:00
4   23633     100   14147               178 Cross Rd  2013-06-30 00:00:00

   NumberOfBoardings
0                  1
1                  1
2                  1
3                  2
4                  1
```

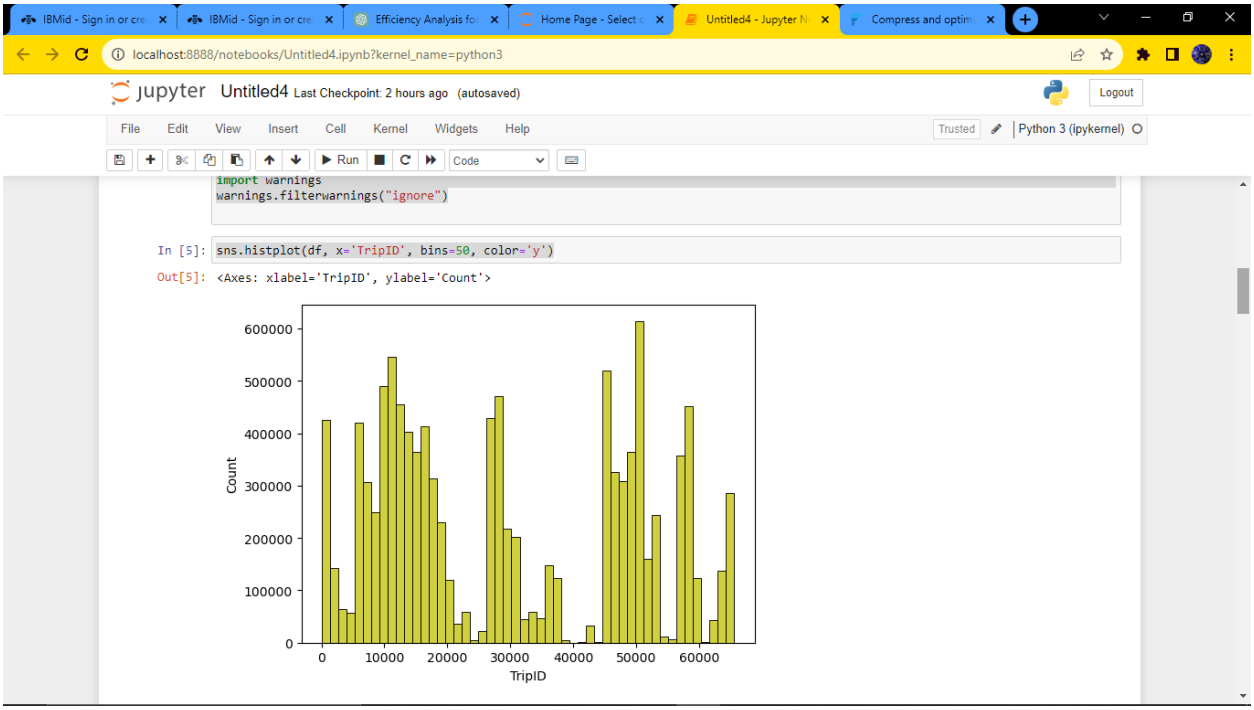
4. Loading the dataset:

```
import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split

from sklearn.preprocessing import StandardScaler

from sklearn.metrics import r2_score

from sklearn.linear_model import LinearRegression

from sklearn.linear_model import Lasso

from sklearn.ensemble import RandomForestRegressor

from sklearn.svm import SVR

%matplotlib inline

import warnings

warnings.filterwarnings("ignore")
```



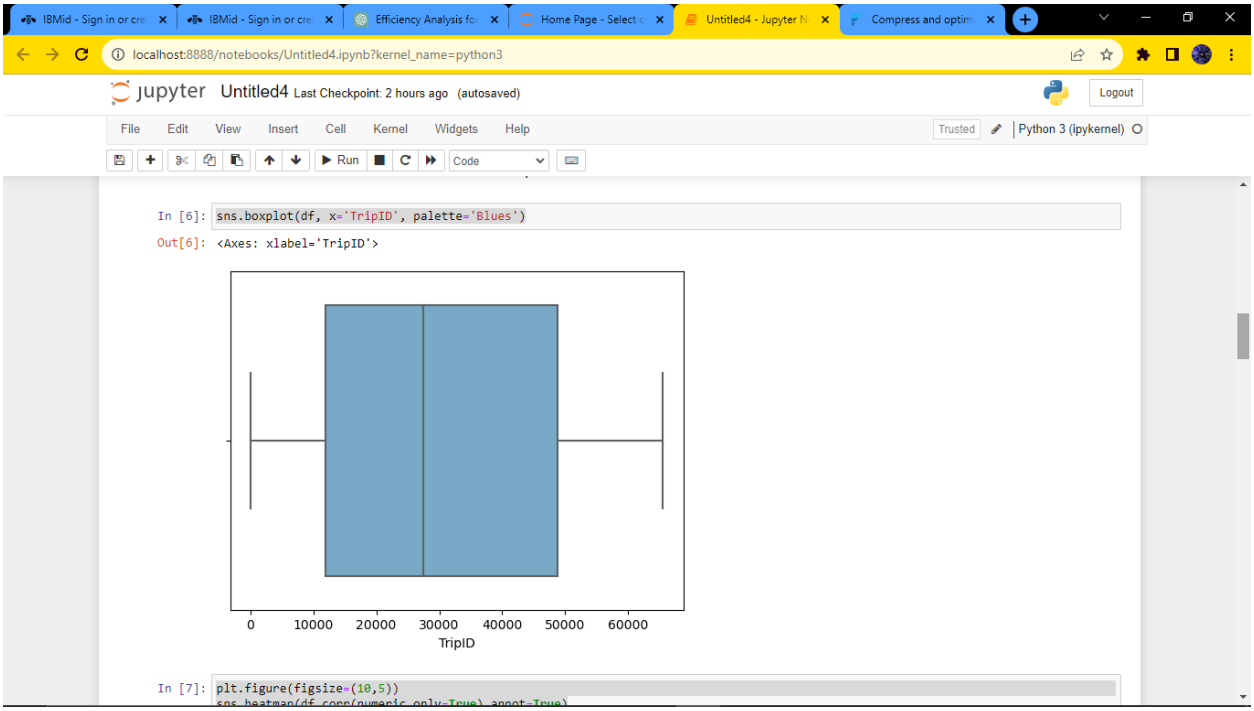
5. Bar chart:

sns.histplot(df, x='TripID', bins=50, color='y')



6. Boxplot:

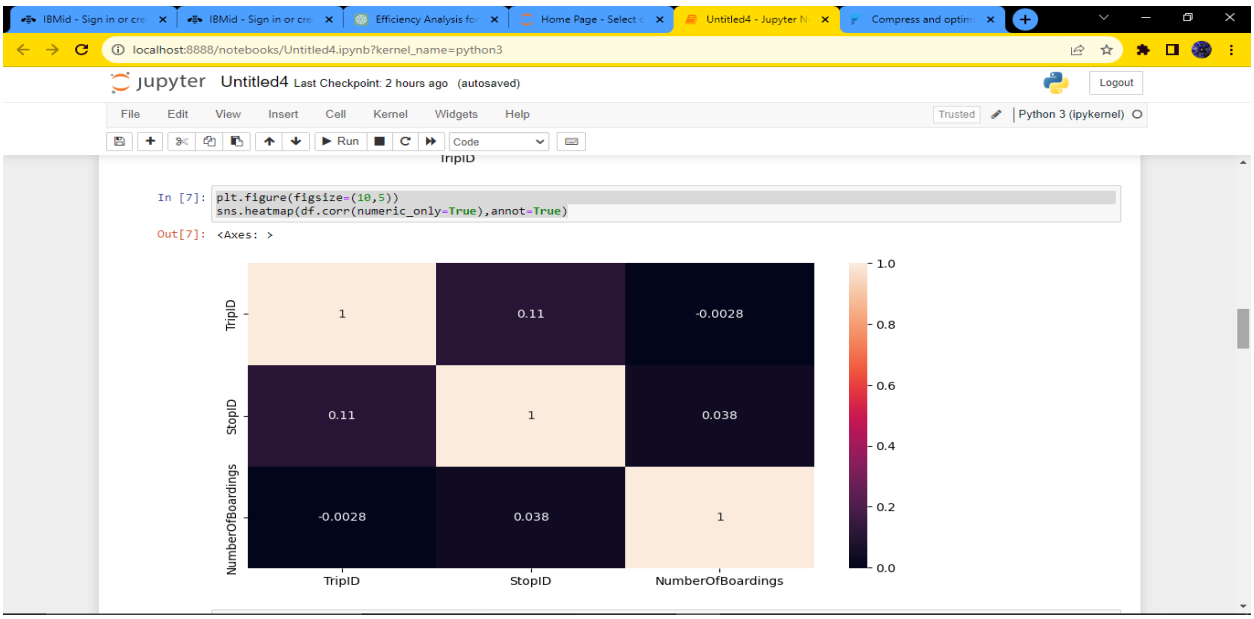
sns.boxplot(df, x='TripID', palette='Blues')



7. Heatmap:

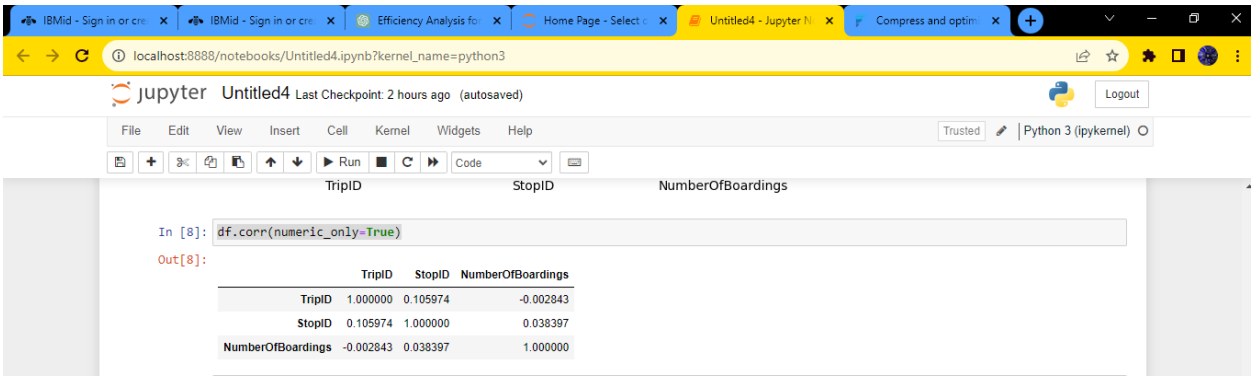
plt.figure(figsize=(10,5))

sns.heatmap(df.corr(numeric_only=True),annot=True)



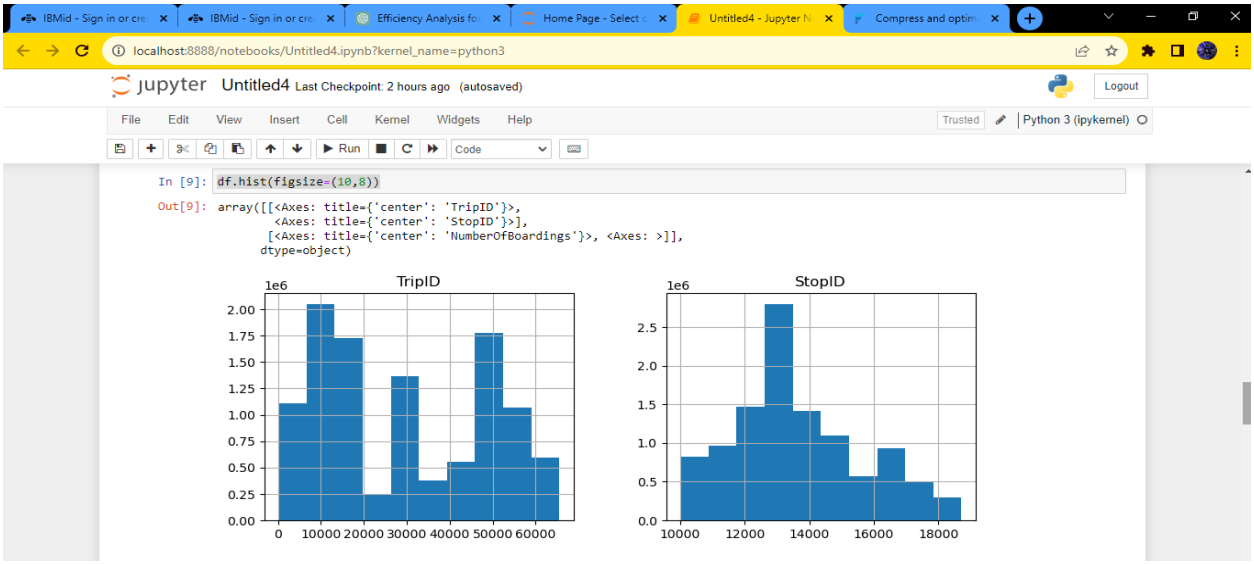
8. Correlation:

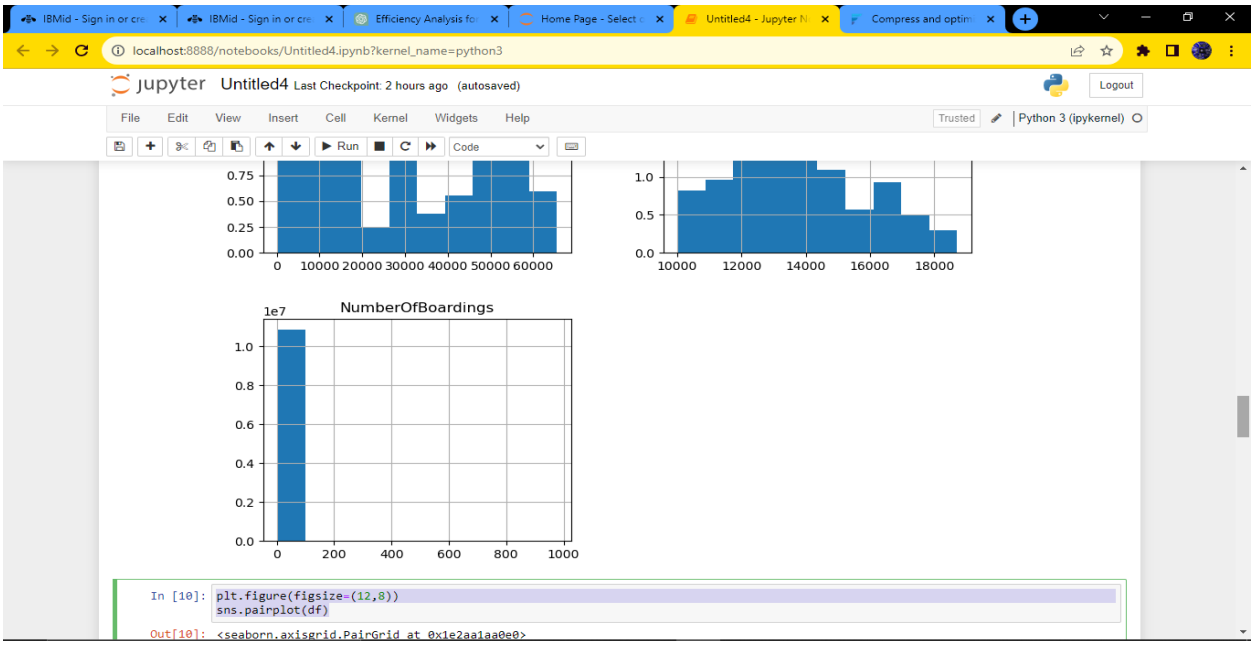
df.corr(numeric_only=True)



9. Histogram:

df.hist(figsize=(10,8))

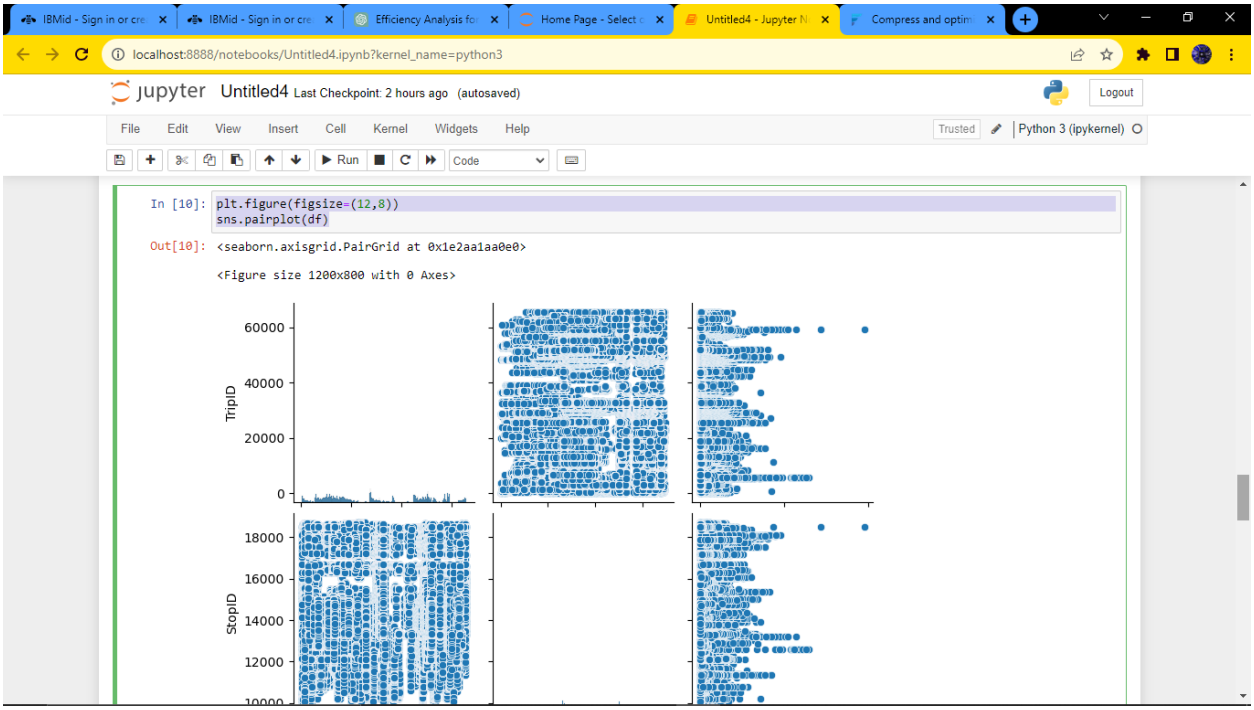


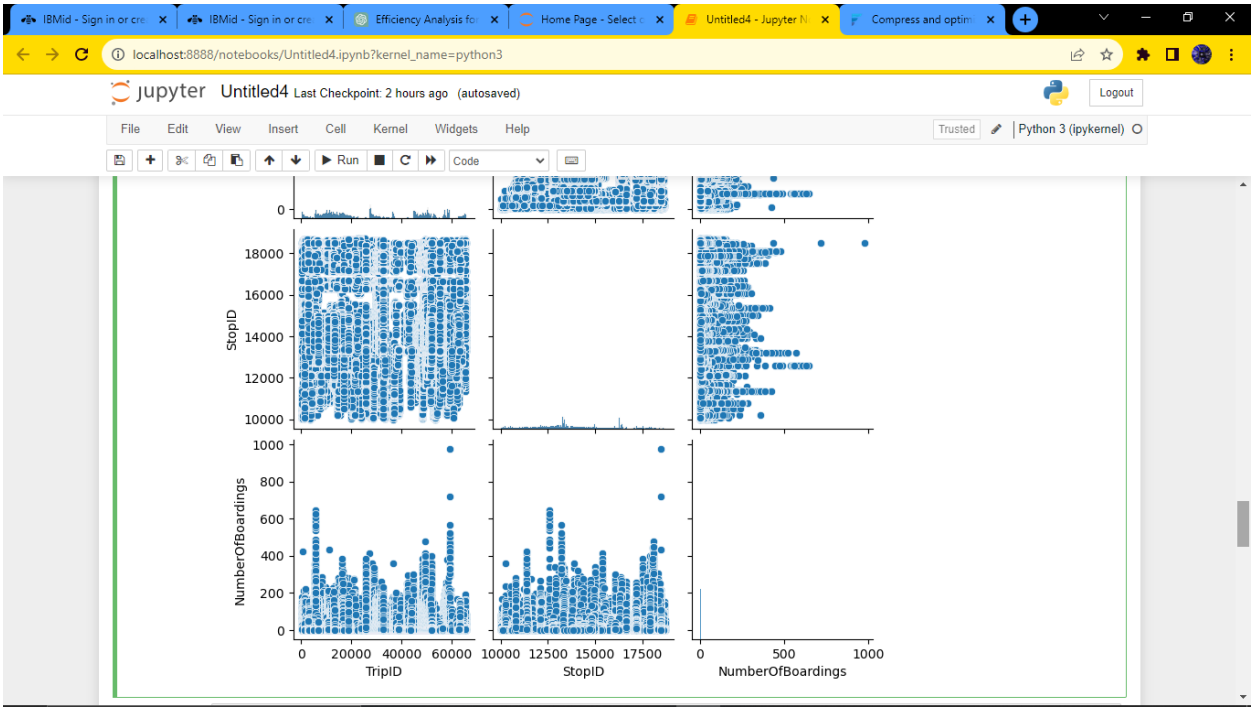


10. Pair plot:

plt.figure(figsize=(12,8))

sns.pairplot(df)





Conclusion:

In conclusion, the public transportation efficiency analysis project with data visualization using a Python program serves as a valuable tool for understanding, optimizing, and improving public transportation systems.