PRIYADARSHINI ENGINEERING COLLEGE

Public Transport Efficiency Analysis IBM Cognos Phase-5

PROJECT DOCUMENTATION & SUBMISSION

Abstract:

- Transit authorities have been searching for the indicators to measure transit service quality and the key factors to attract citizens who do not prefer public transport.
- The aim of this study is to propose metrics to improve transit service.
- The focus is on bus transportation since it is more flexible compared to rail transportation and widely preferred by the masses in cities.

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.
- **1. Data Collection and Integration:**
- **Data Sources:** Gather data from various sources including ridership data, operational data, financial records, customer surveys, and external factors like weather and traffic conditions.
- **Data Integration:** Use IBM Cognos Data Manager to integrate data from different sources, ensuring it's clean, consistent, and ready for analysis.
- **2. Data Modeling:**
- Create Data Models: Use IBM Cognos Framework Manager to create data models that represent the integrated data. Define relationships, calculations, and business rules to prepare the data for analysis.
- **3. Report and Dashboard Creation:**

- ❖ Interactive Reports: Build interactive reports using IBM Cognos Report Studio to visualize key performance indicators (KPIs) such as on-time performance, ridership trends, and cost per passenger.
- ❖ Dashboards: Develop dynamic dashboards using IBM Cognos Workspace to provide a real-time overview of public transport metrics. Dashboards should include widgets displaying KPIs, route performance, and customer satisfaction scores.
- ❖ **4. Data Analysis:**
- ❖ Ad-Hoc Analysis: Use IBM Cognos Analysis Studio for ad-hoc analysis. Explore data, identify patterns, and generate insights on the fly.
- ❖ Predictive Analytics: Apply predictive modelling using IBM Cognos Statistics to forecast ridership trends and optimize routes and schedules.

Phase-1 Project definition & design thinking

2.Methodology:

- ❖ In this study, the focus is on bus transportation,since it is more flexible compared to rail transportation and widely preferred by the masses in cities.
- ❖ The primary data source of this study comes from the Department of Transportation for the City of Antalya.
- ❖ We load the complete boarding data of December 18,2019 which is a standard weekday.
- ❖ The data set formed consist of 305 lines and 608 routes. A route consist of a sequential list of bus stops in either forward or backward (return) directions.
- ❖ Each line has opposite two directions except two lines which are omitted in analysis.
- ❖ On December 18, 2019, a total of 7347 trips (single direction services) were made and with these trips a total of 381962 passengers were carried.
- ❖ However, in this study we slowly focus on the route efficiency.

2.1 Traversal Route Evaluation:

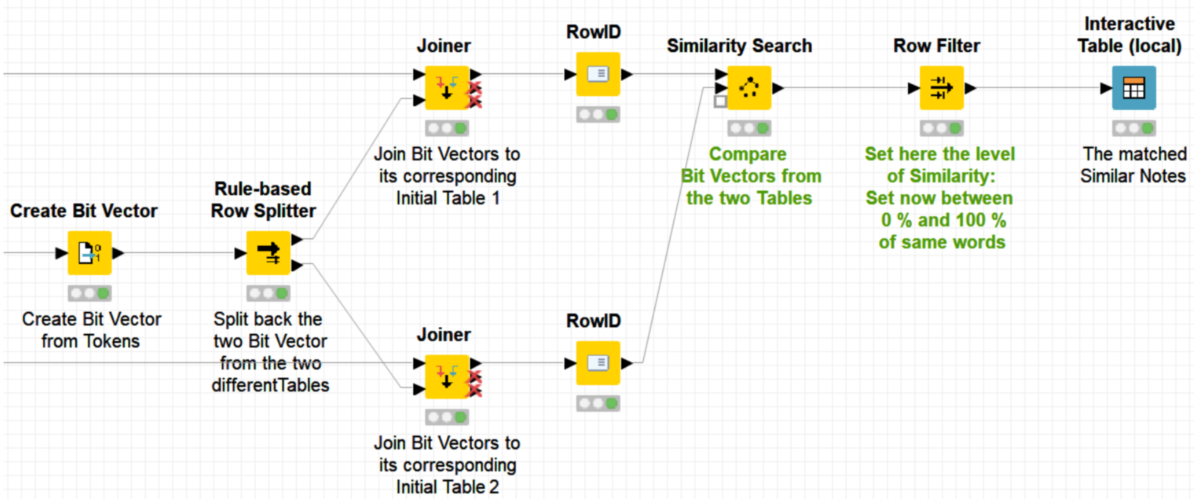
- When lines are designed they often start from an origin and return back to where it start.
- Each line may have consider to have 2 routes, a forward and backward route.
- Typically, when trips are completed at the end of the day. the number of passengers at the forward and backward routes are nearly equal.

2.2 Variance-Area Curves for Route Boarding

- Simply put, the between-area variance curves are obtained by calculating the variances of gray scale pixel values while varying rectangular unit areas within the 2-D gray scale matrix.
- This 2-D gray scale matrix is constructed from the heatmap image

2.3 Bus Stop Analysis :

- Analysis of bus stops can be comprehensive.
- Location of bus stops depends on demand nearby. For the purpose of this study, we do not suggest new bus stops but rather evaluate the boarding demand on existing routes.
- – Bus Stop Id / Name
- – Boarding Count
- – Route Count
- – Service Count
- – Average Boarding Per Route Arrival at Stop
- – Average Boarding Per Bus Service at Stop



Calculation of Route Similarity and Subsequent Hierarchical Clustering Knime Workflow.

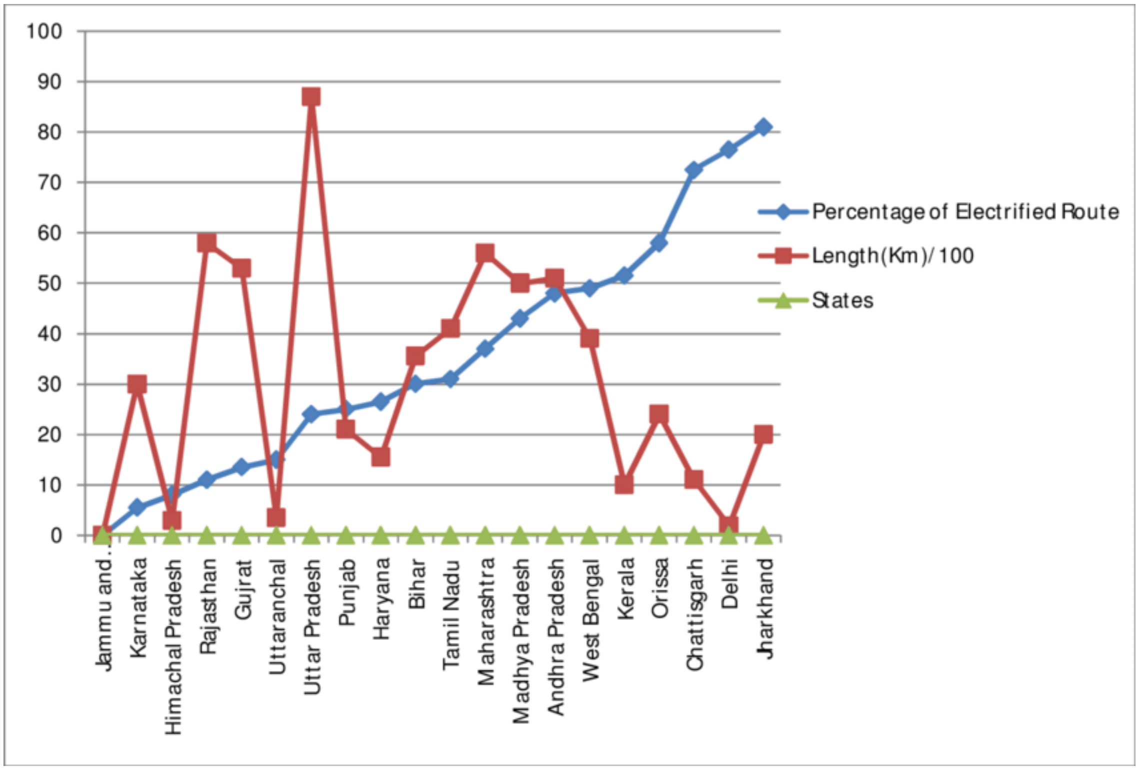
2.4 Results and Discussion

- ❖ Several metrics are proposed in Methodology section executed on the transportation data obtained from Antalya Municipality. We have discussed Route Efficiency, Traversal Route Evaluation,Variance-Area Curves for Route Board ing, Bus Stop Analysis, and Clustering of Bus Lines with respect to overlapping bus stops.

3.Route Efficiency- RE:

- ❖ RE The simplest way to evaluate route efficiency (RE) is to calculate the number of passengers per unit distance (km) travelled.
- ❖ RE for a route i is given by: RE = Pactual(i) Di (1) where Pactual(i) be the actual number of passengers for Route i and D(i) is the total travel distance in kilometers for this Route.
- ❖ Currently, there is not a single number on literature that is suggested to be the benchmark for this metric.

❖ Because, such metric depends on multiple factors including the dynamics of the population, geography and resources, we calculated RE for all routes and normalize routes with respect to total travel distance via Linear Regression.



UP C = Pactual(i) + σr

LCL = Pactual(i) – σr

3.1 Traversal Route :

❖ Evaluation When lines are designed they often start from an origin and return back to where it start.

❖ Each line may have consider to have 2 routes, a forward and backward route.

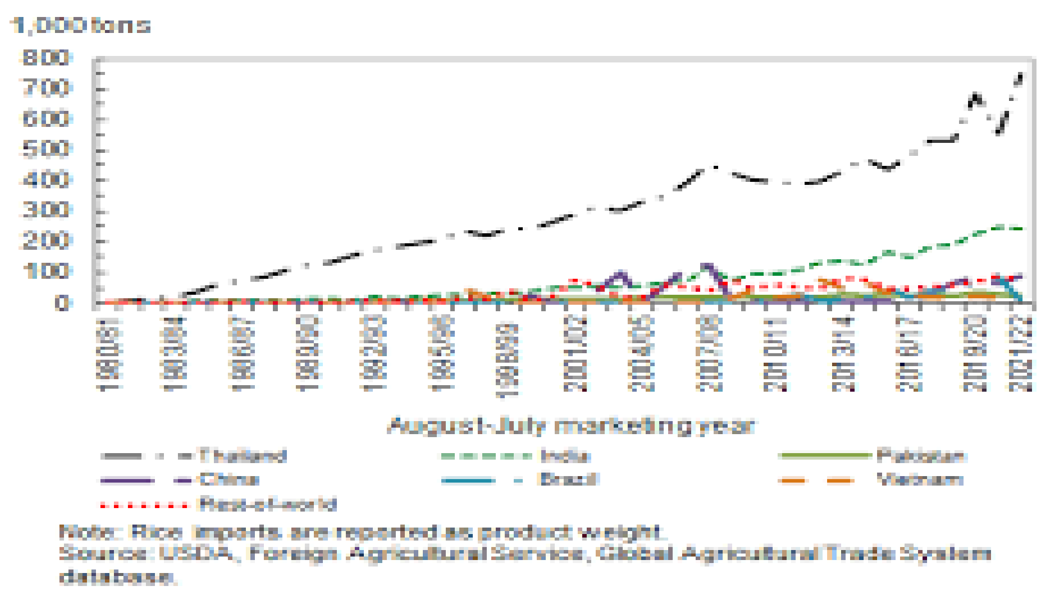
3.2 Time-Location Variance Analysis of Route Boarding :

- ❖ Time-Location mapping of route boarding provides us to analyze the variation of boarding in both time domain and as well as route bus stop sequences.
- ❖ Mapping involves counting boarding in subsequent bus stops for each trip on a route

3.3 Bus Stop Analysis:

- ❖ For the selected date, there were 3140 distinct bus stops and among which 2804 has boarding data (at least 1 boarding).
- ❖ Among them infrequent bus stops are listed in Table 3. One can conclude from the table that there are 567 (% 18).

Figure 4. U.S. rice imports by source, 1980/81-2021/22



3.4 Hierarchical Clustering of Routes for Bus Stop Similarity:

- ❖ In Antalya, a major reason for low REs, low bus utilization, high boarding variances on bus stops may be resulted from lengthy bus lines with many overlapping sections.

- ❖ This can be investigated by clustering of routes according to their common bus stops.
- ❖ The resulted dendrogram presented in Figure 8 shows the hierarchical clustering of bus routes.
- ❖ Here x- axis denotes route Ids and y-axis represent similarity distances (y=1 means no similarity whereas y=0 means one hundred percent similarity).

4. Conclusion :

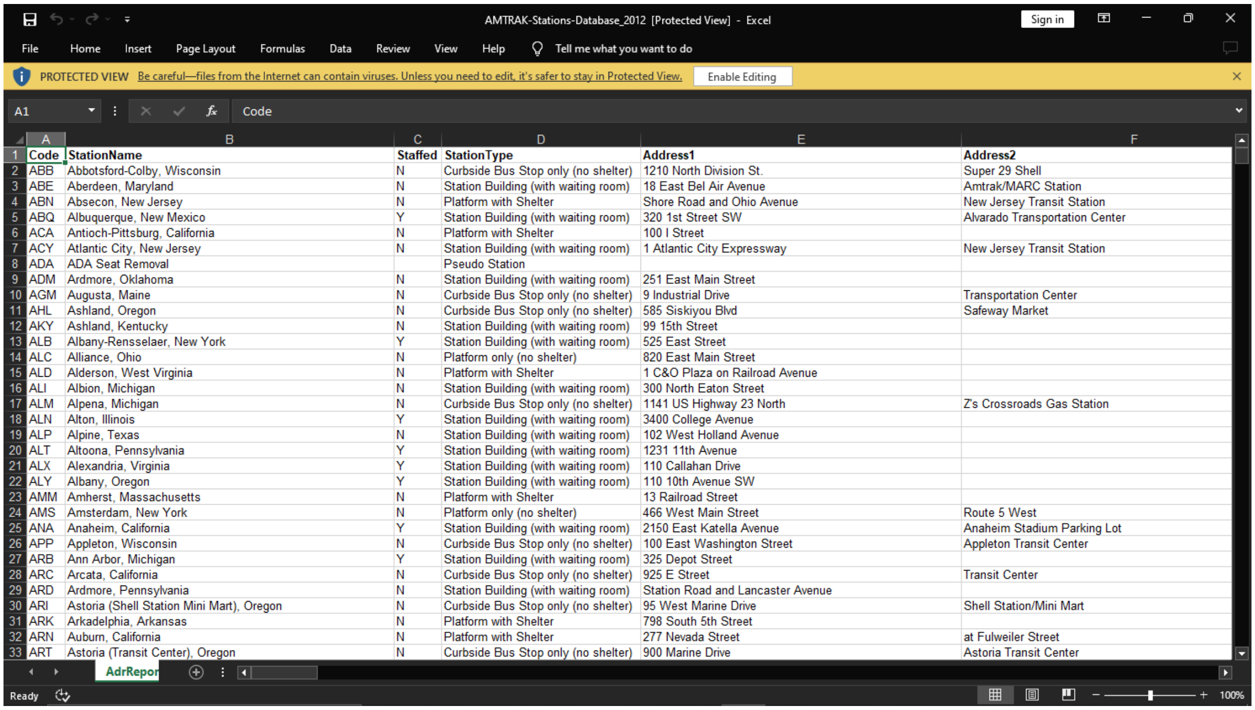
- ❖ The aim of this study is to identify inefficient routes and to propose improvements by the evaluation of route efficiencies, the analysis of bus stop boarding counts and the clustering of routes.
- ❖ Several metrics are proposed in Methodology section executed on the transportation data obtained from Antalya Municipality.

Phase-2 Innovation

DATASET:

Source: data.world.com

Sample dataset



Code	StationName	Staffed	StationType	Address1	Address2
ABB	Abbotsford-Colby, Wisconsin	N	Curbside Bus Stop only (no shelter)	1210 North Division St.	Super 29 Shell
ABE	Aberdeen, Maryland	N	Station Building (with waiting room)	18 East Bel Air Avenue	Amtrak/MARC Station
ABN	Absecon, New Jersey	N	Platform with Shelter	Shore Road and Ohio Avenue	New Jersey Transit Station
ABQ	Albuquerque, New Mexico	Y	Station Building (with waiting room)	320 1st Street SW	Alvarado Transportation Center
ACA	Antioch-Pittsburg, California	N	Platform with Shelter	100 I Street	
ACY	Atlantic City, New Jersey	N	Station Building (with waiting room)	1 Atlantic City Expressway	New Jersey Transit Station
ADA	ADA Seat Removal		Pseudo Station		
ADM	Ardmore, Oklahoma	N	Station Building (with waiting room)	251 East Main Street	
AGM	Augusta, Maine	N	Curbside Bus Stop only (no shelter)	9 Industrial Drive	Transportation Center
AHL	Ashland, Oregon	N	Curbside Bus Stop only (no shelter)	585 Siskiyou Blvd	Safeway Market
AKY	Ashland, Kentucky	N	Station Building (with waiting room)	99 15th Street	
ALB	Albany-Rensselaer, New York	Y	Station Building (with waiting room)	525 East Street	
ALC	Alliance, Ohio	N	Platform only (no shelter)	820 East Main Street	
ALD	Alderson, West Virginia	N	Platform with Shelter	1 C&O Plaza on Railroad Avenue	
ALI	Albion, Michigan	N	Station Building (with waiting room)	300 North Eaton Street	
ALM	Alpena, Michigan	N	Curbside Bus Stop only (no shelter)	1141 US Highway 23 North	Z's Crossroads Gas Station
ALN	Alton, Illinois	Y	Station Building (with waiting room)	3400 College Avenue	
ALP	Alpine, Texas	N	Station Building (with waiting room)	102 West Holland Avenue	
ALT	Altoona, Pennsylvania	Y	Station Building (with waiting room)	1231 11th Avenue	
ALX	Alexandria, Virginia	Y	Station Building (with waiting room)	110 Callahan Drive	
ALY	Albany, Oregon	Y	Station Building (with waiting room)	110 10th Avenue SW	
AMM	Amherst, Massachusetts	N	Platform with Shelter	13 Railroad Street	
AMS	Amsterdam, New York	N	Platform only (no shelter)	466 West Main Street	Route 5 West
ANA	Anaheim, California	Y	Station Building (with waiting room)	2150 East Katella Avenue	Anaheim Stadium Parking Lot
APP	Appleton, Wisconsin	N	Curbside Bus Stop only (no shelter)	100 East Washington Street	Appleton Transit Center
ARB	Ann Arbor, Michigan	Y	Station Building (with waiting room)	325 Depot Street	
ARC	Arcata, California	N	Curbside Bus Stop only (no shelter)	925 E Street	Transit Center
ARD	Ardmore, Pennsylvania	N	Station Building (with waiting room)	Station Road and Lancaster Avenue	
ARI	Astoria (Shell Station Mini Mart), Oregon	N	Curbside Bus Stop only (no shelter)	95 West Marine Drive	Shell Station/Mini Mart
ARK	Arkadelphia, Arkansas	N	Platform with Shelter	798 South 5th Street	
ARN	Auburn, California	N	Platform with Shelter	277 Nevada Street	at Fulweiler Street
ART	Astoria (Transit Center), Oregon	N	Curbside Bus Stop only (no shelter)	900 Marine Drive	Astoria Transit Center

INNOVATION OBJECTIVES:

1. **Real-time Data Analytics**:
   - Utilize real-time data from sensors, GPS devices, and passenger feedback to monitor bus and train performance. Analyze this data to make on-the-fly adjustments, such as route optimization and vehicle maintenance.
2. **Machine Learning for Predictive Maintenance**:
   - Implement predictive maintenance using machine learning algorithms. By monitoring the health of transportation assets like buses and trains, you can schedule maintenance proactively, reducing breakdowns and service disruptions.
3. **Rider Behavior Analysis**:
   - Analyze passenger data to understand ridership patterns. This includes where people board and alight, their travel times, and preferences for routes. This information can be used to optimize schedules and routes.
4. **Fare and Pricing Optimization**:
   - Analyze fare structures and pricing models to maximize revenue and ridership. Dynamic pricing based on demand and season, as well as introducing different ticket options, can be explored.
5. **Multi-Modal Integration**:

- Promote seamless integration between different modes of transportation, such as buses, trains, trams, and bicycles. Analyze data on how passengers transfer between these modes and look for opportunities to improve connectivity.

6. **Traffic Flow Modeling**:
    - Use traffic flow modeling software to optimize signal timings and road layouts to reduce congestion and improve the efficiency of public transportation routes.
7. **Energy Efficiency Assessment**:
    - Analyze the energy consumption of public transportation vehicles. Consider alternatives such as electric buses and evaluate the feasibility and benefits of transitioning to more eco-friendly options.
8. **Crowdsourcing Solutions**:
    - Engage the community in collecting data and suggesting improvements. Crowdsourced data can be valuable in identifying issues that passengers encounter daily.
9. **Digital Ticketing and Contactless Payments**:
    - Implement digital ticketing systems and contactless payment options. Analyze transaction data to understand passenger behavior, peak travel times, and fare evasion rates.
10. **Traffic Signal Priority (TSP)**:
    - Implement TSP systems that give public transportation vehicles priority at traffic signals. Analyze the impact of TSP on travel times and on-time performance.
11. **Autonomous Vehicles and Mobility as a Service (MaaS)**:
    - Investigate the feasibility of integrating autonomous vehicles into public transportation fleets. Explore MaaS platforms that allow passengers to plan and pay for multi-modal trips.
12. **Route Optimization Algorithms**:
    - Develop and implement advanced route optimization algorithms that consider real-time traffic, weather, and passenger demand data to provide efficient and dynamic route planning.
13. **Environmental Impact Assessment**:
    - Analyze the environmental impact of public transportation services and explore strategies for reducing emissions and promoting sustainability.
14. **Customer Satisfaction Surveys**:

- Regularly survey passengers to gather feedback on their experiences. Analyze the data to identify pain points and areas for improvement in service quality.

15. **Smart Infrastructure and IoT**:
    - Utilize IoT sensors and smart infrastructure to monitor traffic and transportation assets. Analyze this data to enhance efficiency, safety, and sustainability.
16. **GIS and Spatial Analysis**:
    - Use Geographic Information Systems (GIS) and spatial analysis to optimize bus stop and station locations, taking into account population density and urban development.
17. **Benchmarking with Other Cities**:
    - Compare the efficiency and performance of your public transportation system with similar systems in other cities to identify best practices and areas for improvement.
18. **Collaboration with Ride-Sharing Services**:
    - Partner with ride-sharing companies to complement public transportation services. Analyze data on shared rides and explore cost-effective integration.

CONCLUSION:

Thus by implementing these innovative ideas in the upcoming phases we can achieve the public transport efficiency.

Phase-3 Development part-1

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.

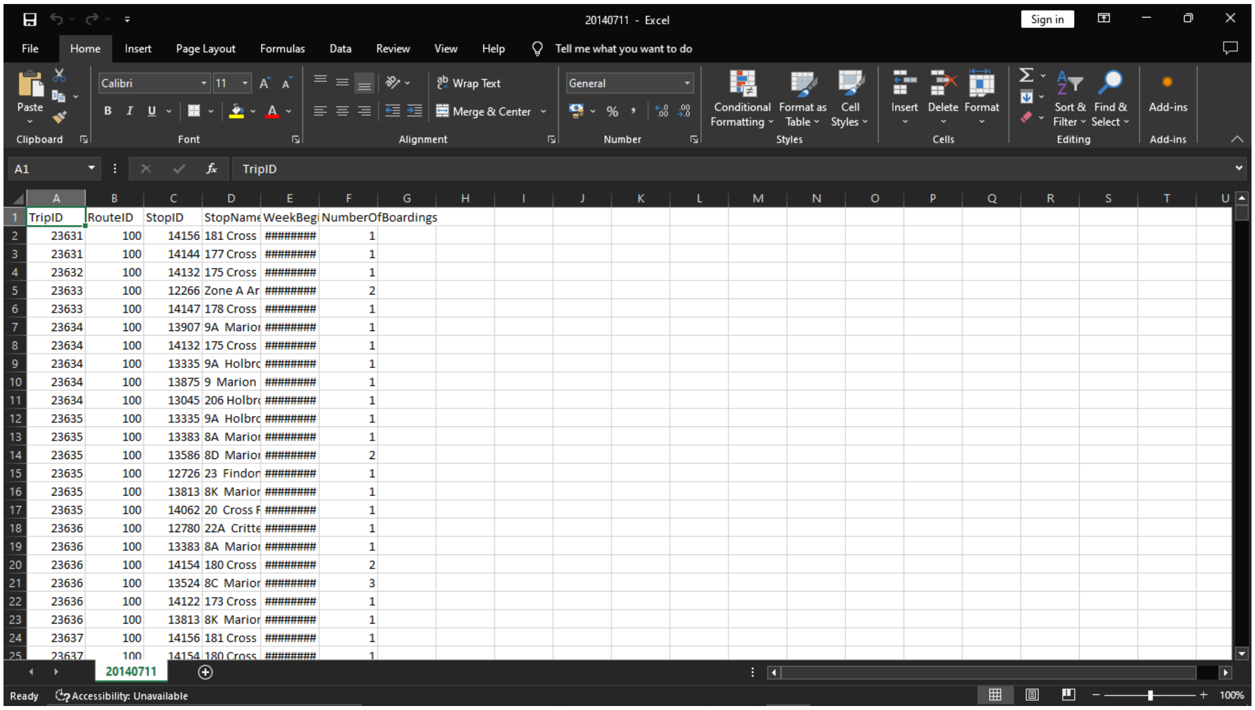
- Efficiency Optimization: The program demonstrates the use of algorithms and analysis to optimize routes, schedules, and infrastructure to minimize inefficiencies and reduce congestion.

- 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 Marioı	#########	1
23634	100	14132	175 Cross	#########	1
23634	100	13335	9A Holbrc	#########	1
23634	100	13875	9 Marion	#########	1
23634	100	13045	206 Holbrc	#########	1
23635	100	13335	9A Holbrc	#########	1
23635	100	13383	8A Marioı	#########	1
23635	100	13586	8D Marioı	#########	2
23635	100	12726	23 Findon	#########	1
23635	100	13813	8K Marioı	#########	1
23635	100	14062	20 Cross F	#########	1
23636	100	12780	22A Critte	#########	1
23636	100	13383	8A Marioı	#########	1
23636	100	14154	180 Cross	#########	2
23636	100	13524	8C Marioı	#########	3
23636	100	14122	173 Cross	#########	1
23636	100	13813	8K Marioı	#########	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:

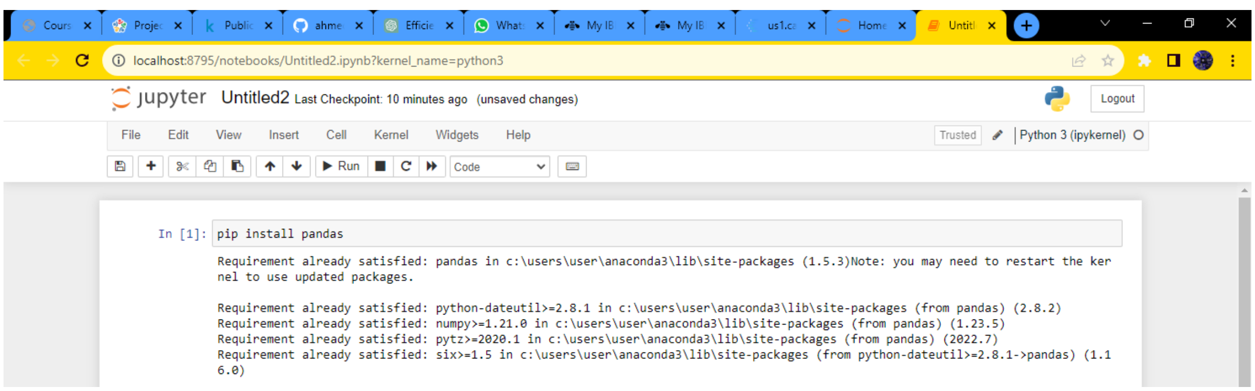
The most common way to import libraries is by using the import statement. You simply specify the library/module name after import. You can also use an alias to make it easier to refer to the library in your code.

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:

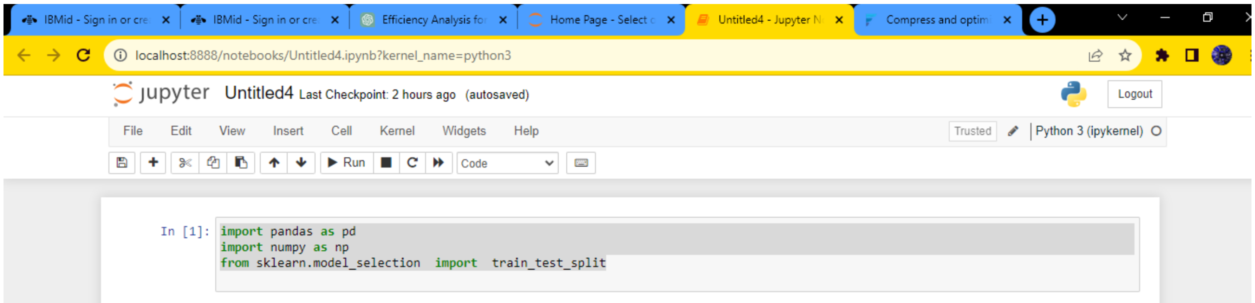


2. Importing Libraries:

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}")
```

```
In [3]: 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:

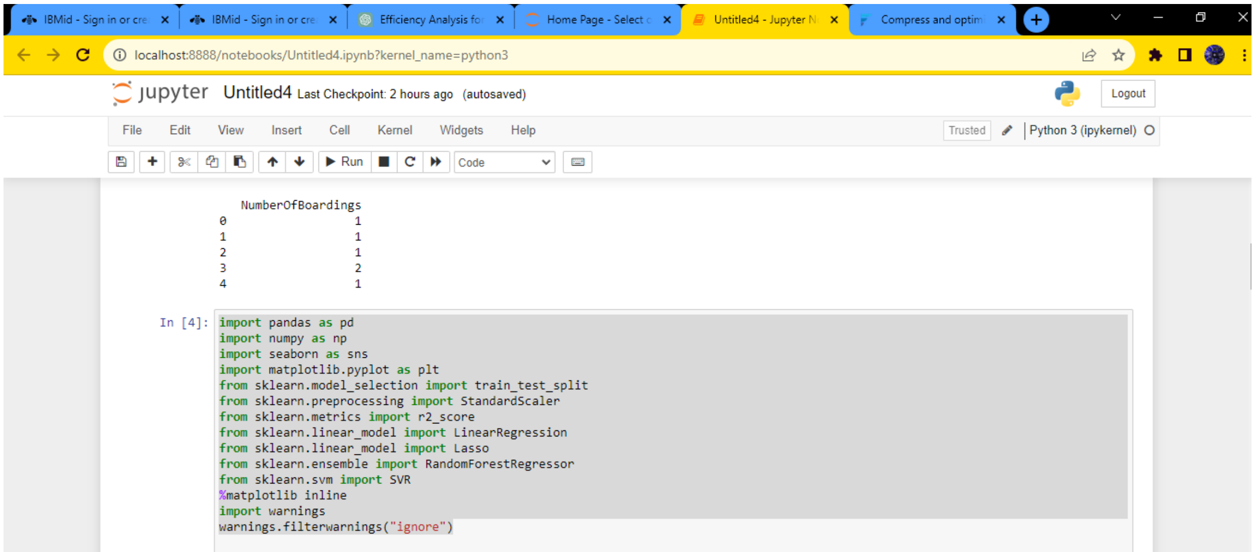
```
C:\Users\User\AppData\Local\Temp\ipykernel_3484\2929465056.py:8: DtypeWarning: Columns (1) have mixed types. Specify dtype opti
on on import or set low_memory=False.
  df = pd.read_csv(file_path)

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
```

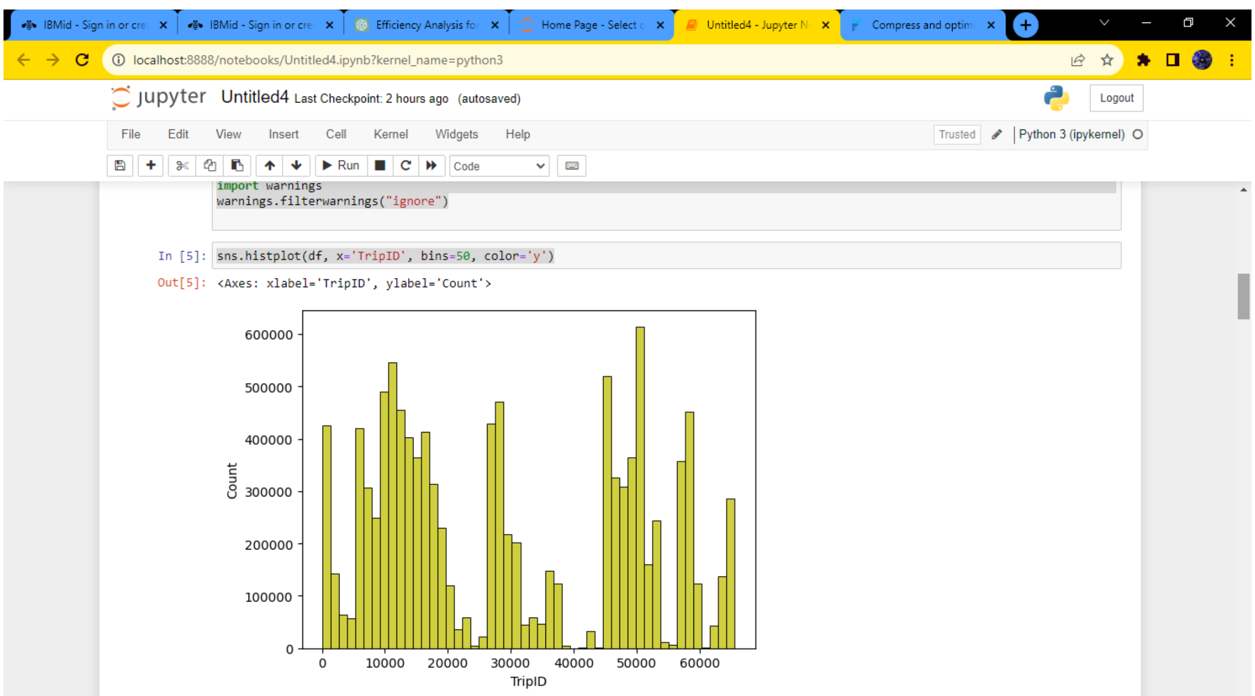
import warnings

warnings.filterwarnings("ignore")



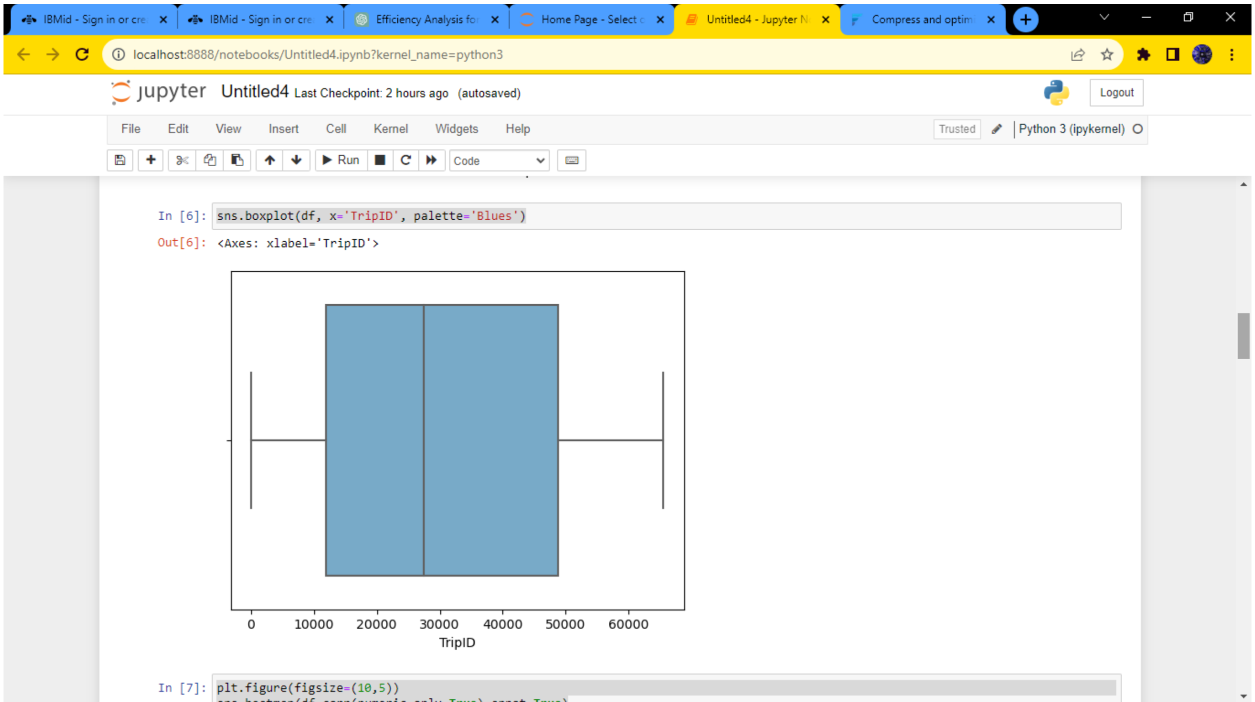
5. Bar chart:

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



6. Boxplot:

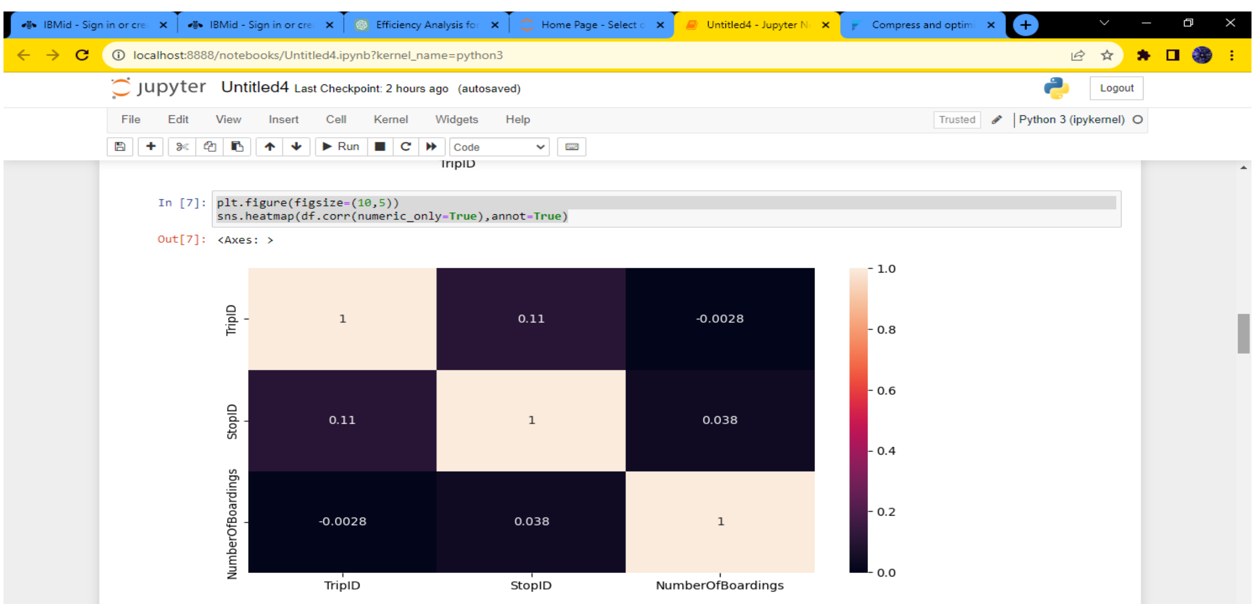
sns.boxplot(df, x='TripID', palette='Blues')



7. Heatmap:

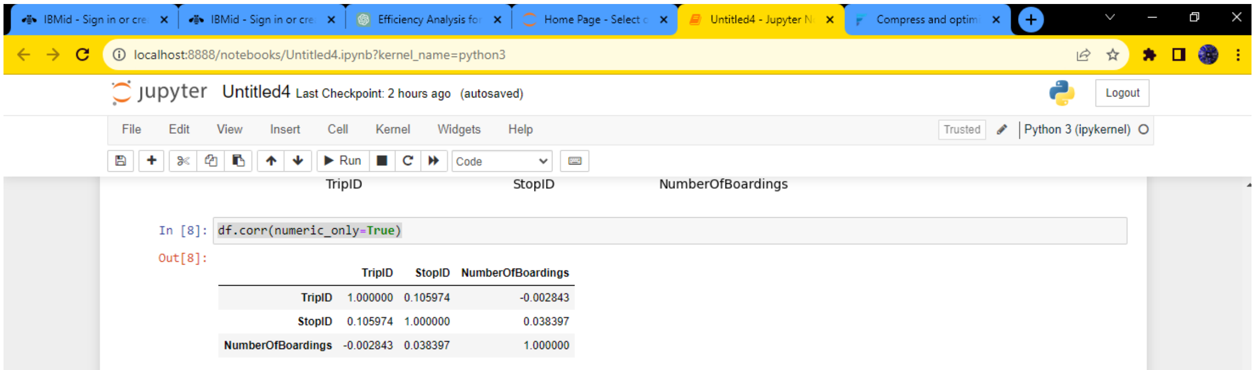
plt.figure(figsize=(10,5))

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



8. Correlation:

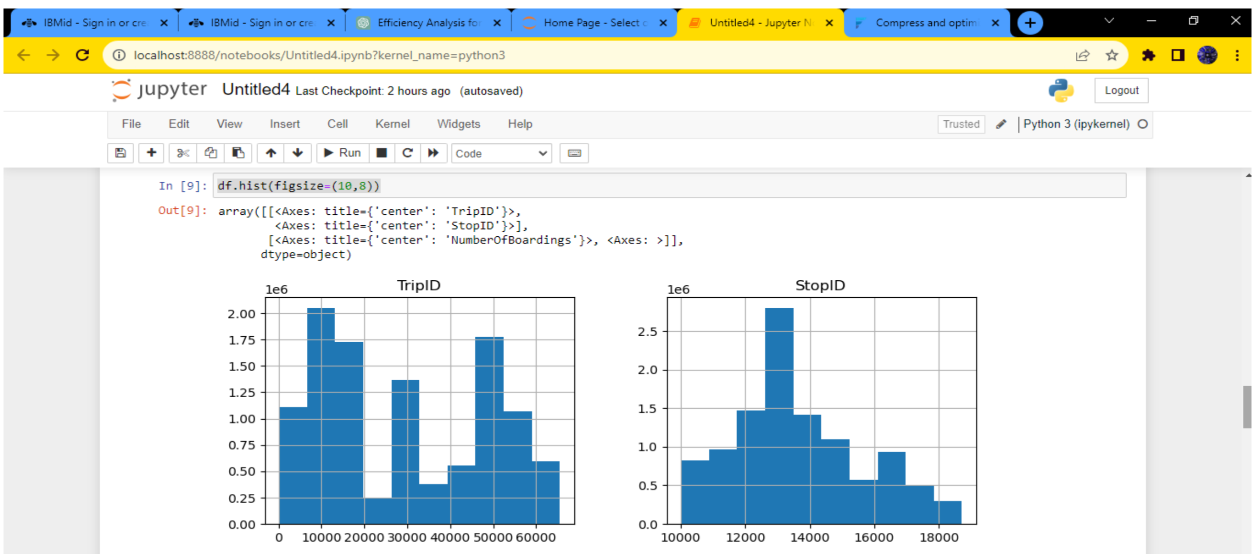
df.corr(numeric_only=True)

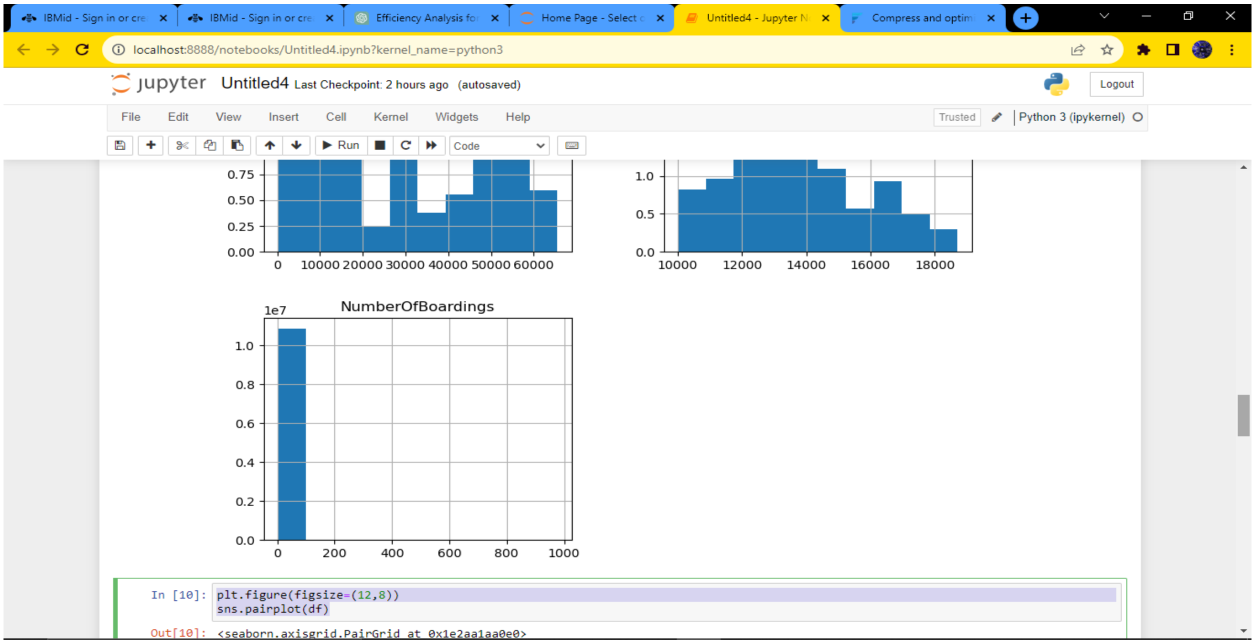


	TripID	StopID	NumberOfBoardings
TripID	1.000000	0.105974	-0.002843
StopID	0.105974	1.000000	0.038397
NumberOfBoardings	-0.002843	0.038397	1.000000

9. Histogram:

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

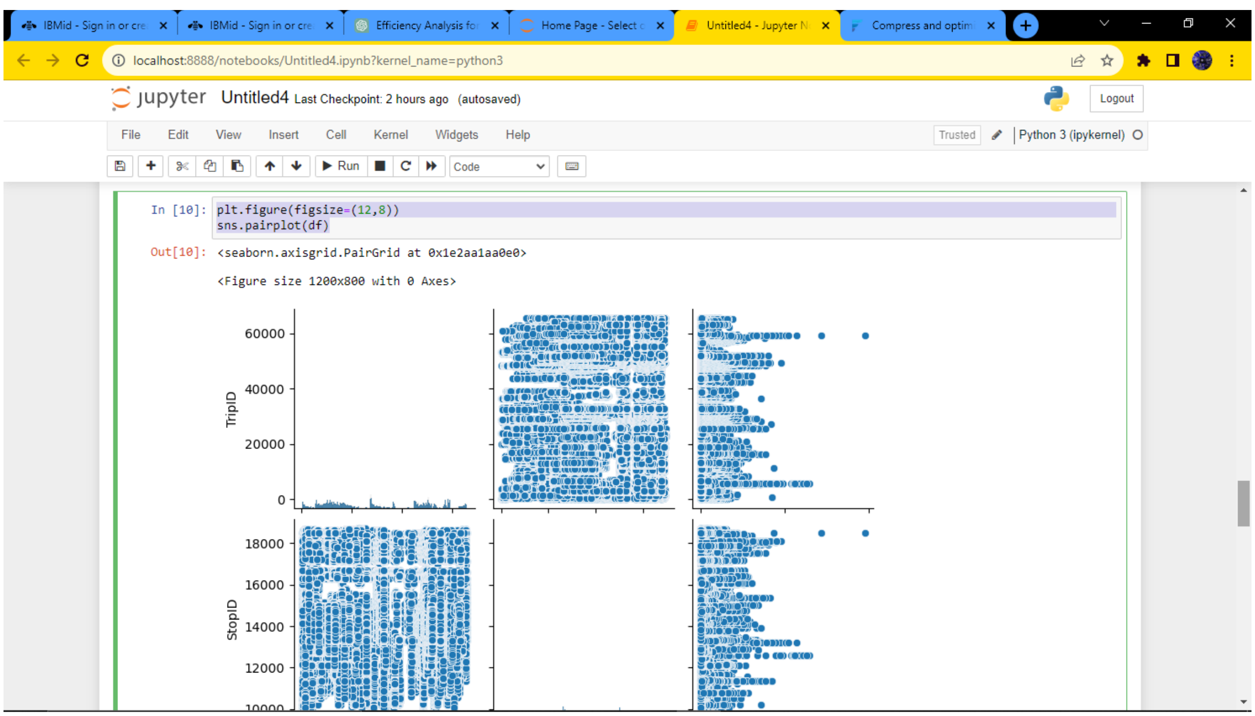


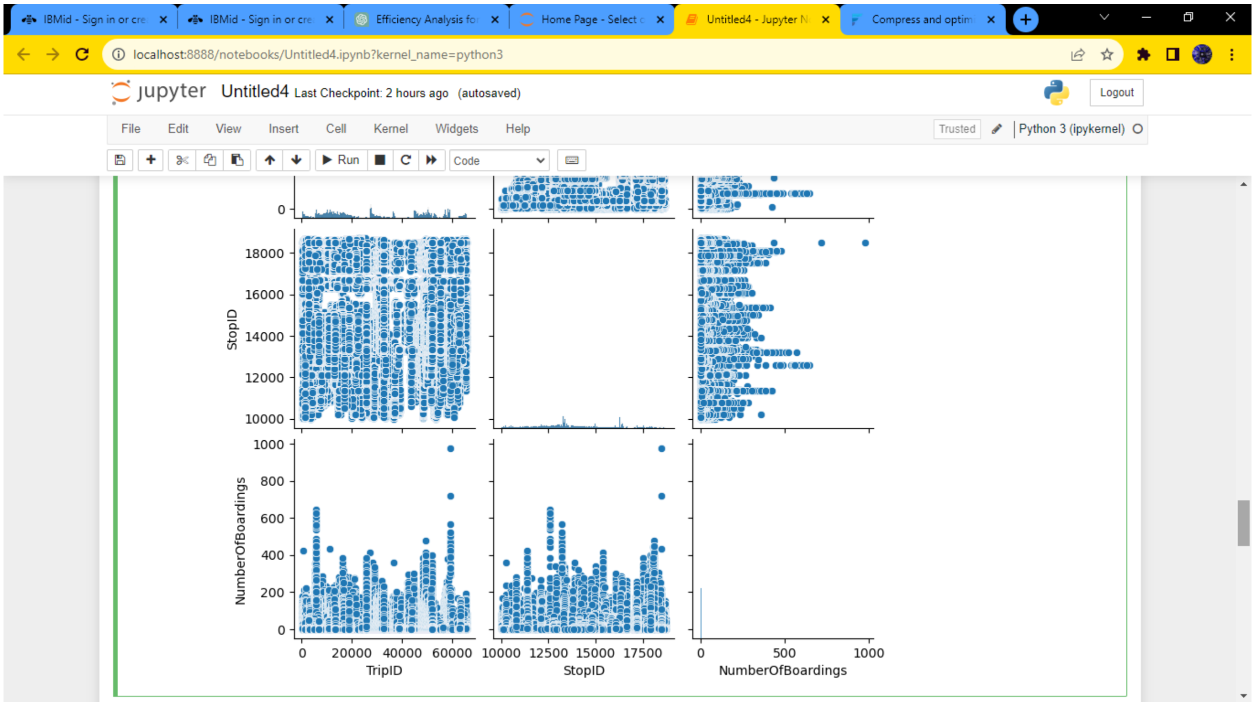


10. Pair plot:

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

sns.pairplot(df)





Conclusion:

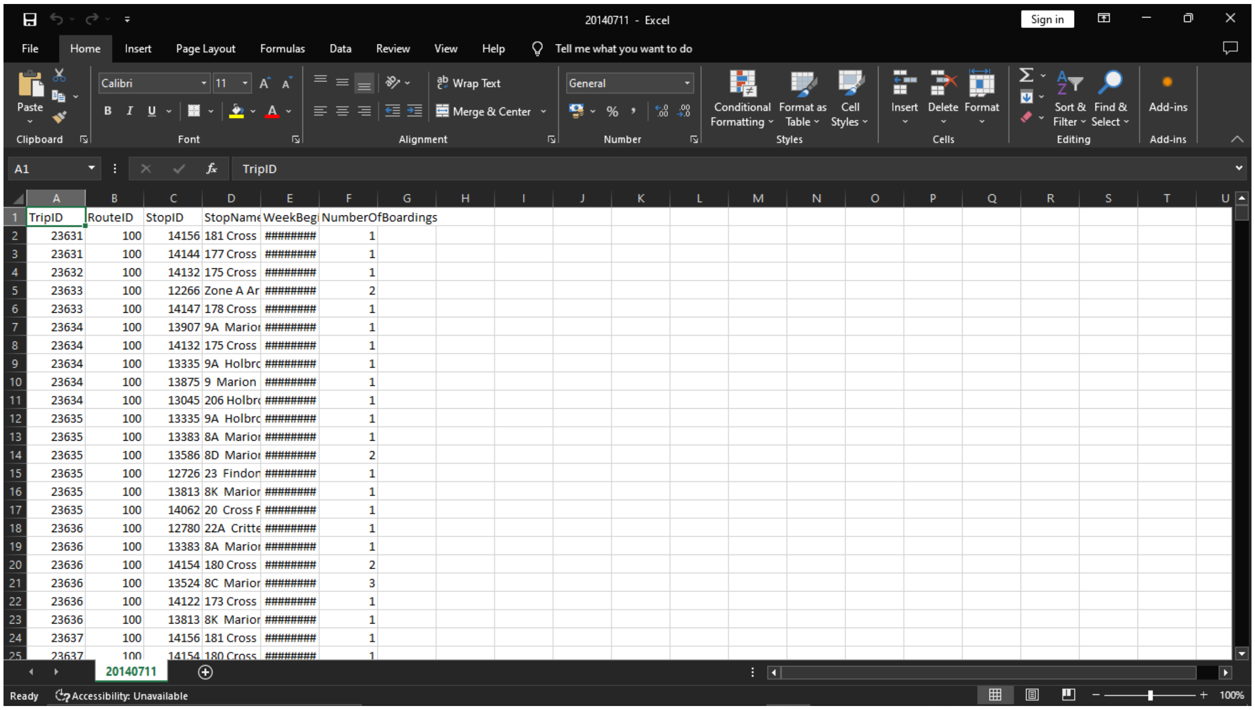
In conclusion, the public transportation efficiency analysis project with data visualization using a Python program serves as a valuable tool for understanding, optimizing, and improving public transportation systems.

Phase-4 Development part-2

Development part-2 using IBM Cognos for Visualization

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 Marion	#########	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 Marion	#########	1
23635	100	13586	8D Marion	#########	2
23635	100	12726	23 Findon	#########	1
23635	100	13813	8K Marion	#########	1
23635	100	14062	20 Cross F	#########	1
23636	100	12780	22A Critte	#########	1
23636	100	13383	8A Marion	#########	1
23636	100	14154	180 Cross	#########	2
23636	100	13524	8C Marion	#########	3
23636	100	14122	173 Cross	#########	1
23636	100	13813	8K Marion	#########	1
23637	100	14156	181 Cross	#########	1
23637	100	14154	180 Cross	#########	1

The above dataset has: 1048553 Rows.

Development part-2 using IBM Cognos for visualization

Building a public transportation efficiency analysis using IBM Cognos for visualization involves several steps. In this process, you'll collect data, prepare it, create data models, and generate visualizations to analyse the efficiency of public transportation services. Here's a step-by-step guide:

Step 1: Define Your Objectives Clearly define the objectives of your analysis. Determine what aspects of public transportation efficiency you want to measure and improve, such as on-time performance, ridership, cost-effectiveness, or route optimization.

Step 2: Data Collection and Integration Collect relevant data from various sources, including historical transportation records, scheduling information, geographic data, and real-time tracking systems. Ensure that the data is accurate, complete, and up-to-date. IBM Cognos can connect to various data sources, including databases, spreadsheets, and APIs, to import and integrate data.

Step 3: Data Preparation Clean, transform, and shape your data to make it suitable for analysis. This involves handling missing data, removing duplicates, and converting data types. IBM Cognos provides data preparation tools to assist with this.

Step 4: Create Data Models Design data models that represent the relationships between different data elements. You can use IBM Cognos Framework Manager or the Data Modules feature to create data models that are optimized for performance and user-friendliness.

Step 5: Report and Dashboard Creation Now, it's time to build reports and dashboards in IBM Cognos. Here's how:

a. Create a New Report or Dashboard: Start by creating a new report or dashboard in IBM Cognos.

b. Select Data Sources: Connect your report or dashboard to the data models you've created.

c. Choose Visualization Types: Decide on the type of visualizations you want to use. This could include bar charts, line graphs, maps, and tables. IBM Cognos offers a variety of visualization options.

d. Build Visualizations: Design and customize your visualizations. For example, you can create a bar chart to display on-time performance by route or a map showing stop locations and their average wait times.

e. Create Interactivity: Add interactive elements to your dashboard, such as filters, drill-through capabilities, and parameters that allow users to interact with the data.

Step 6: Data Analysis and Insights Use the visualizations in your report or dashboard to analyze the efficiency of public transportation. Identify trends, outliers, and areas for improvement. For example, you can track the most and least efficient routes, analyze peak hours, or evaluate the impact of schedule changes.

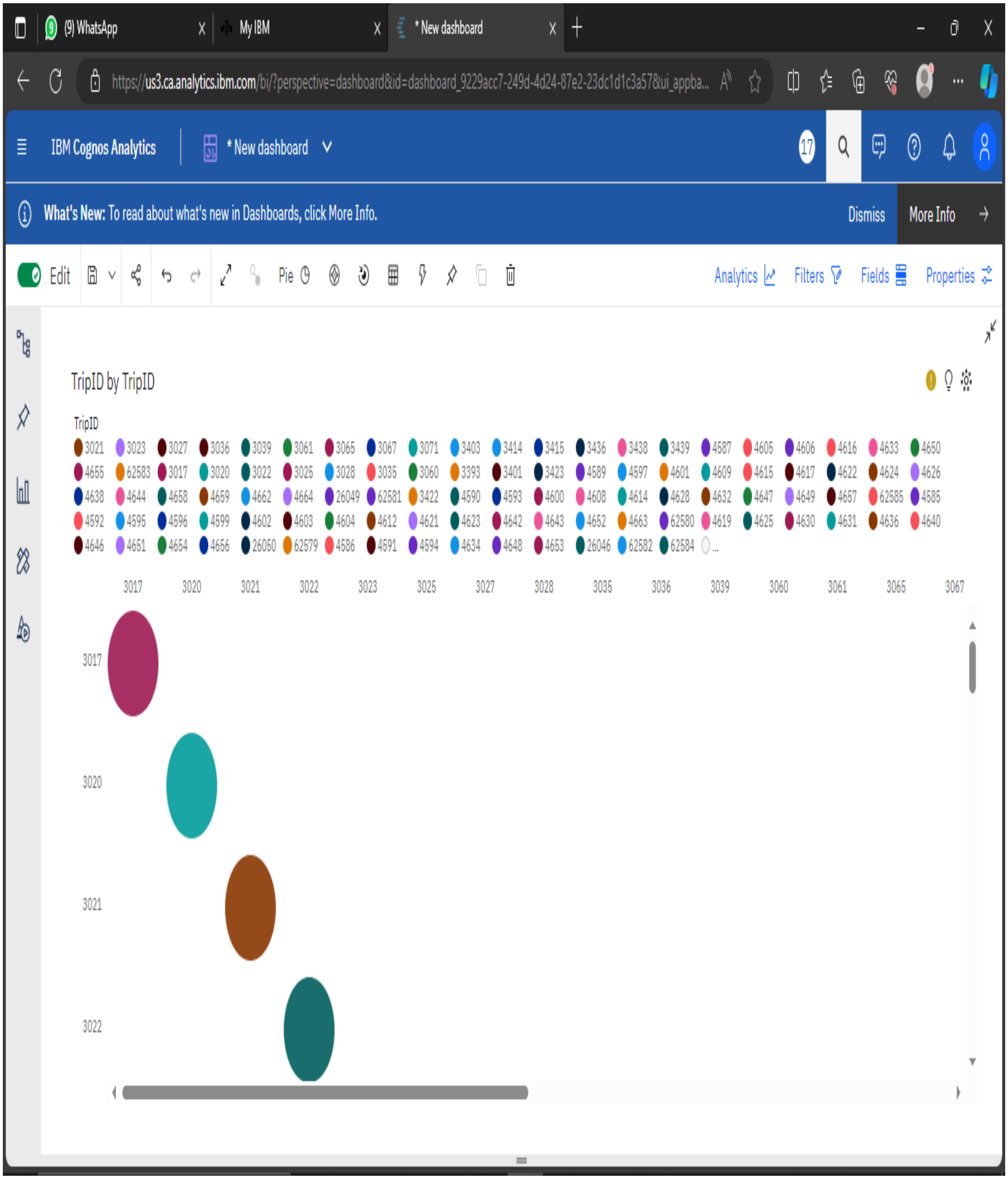
Step 7: Sharing and Collaboration IBM Cognos allows you to share your reports and dashboards with relevant stakeholders. Collaborate with decision-makers, transportation authorities, and analysts to ensure that the insights are used for making informed decisions.

Step 8: Schedule and Automation Set up schedules or triggers for your reports and dashboards to be automatically updated and shared at regular intervals. This ensures that your analysis remains up-to-date.

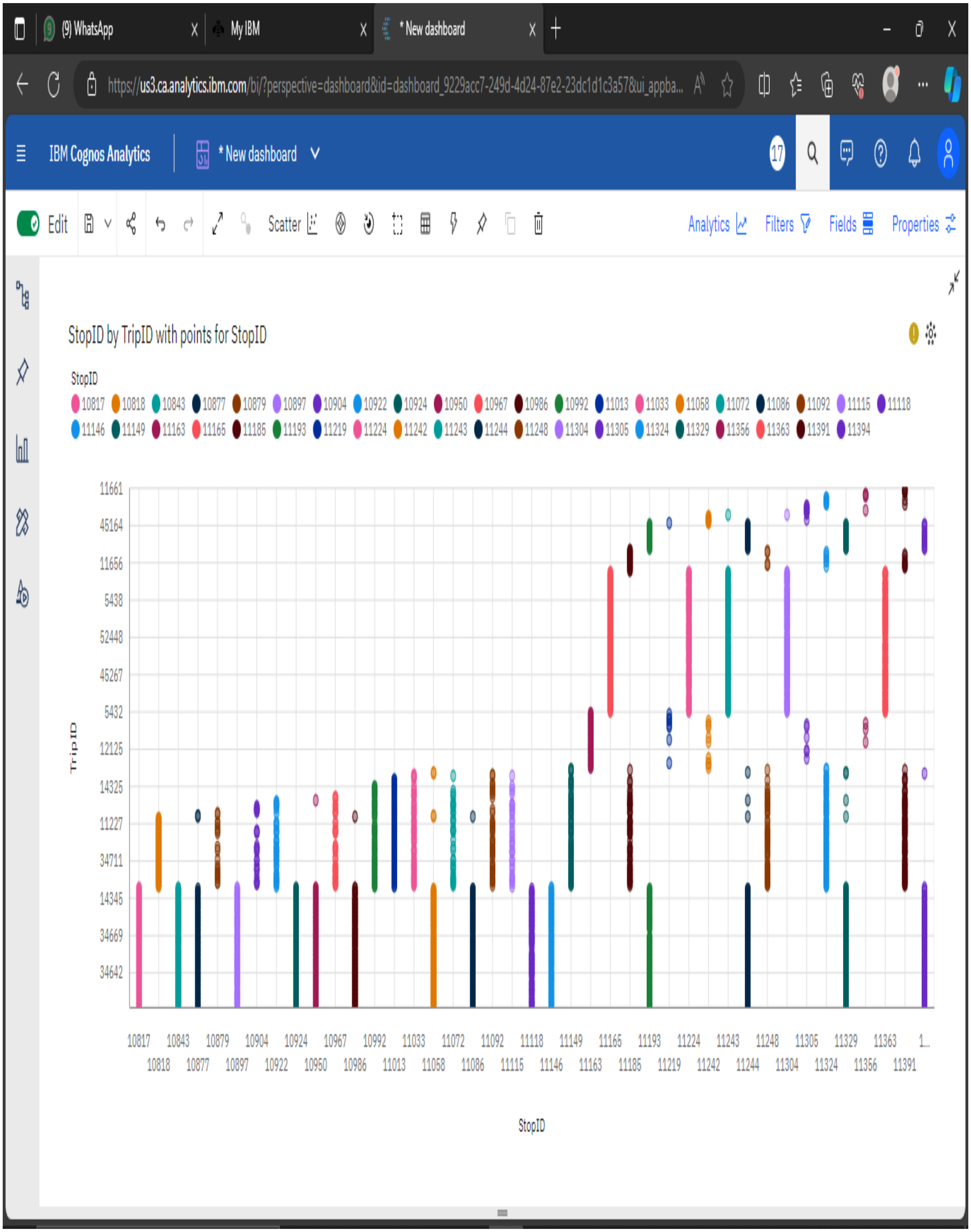
Step 9: Monitor and Iterate Continuously monitor the performance of public transportation and gather user feedback. Use this information to refine your reports and dashboards, making them more insightful and valuable.

By following these steps, you can use IBM Cognos to create an efficient and insightful analysis of public transportation. It helps decision-makers optimize transportation services and improve the overall efficiency of the system.

Trip ID:

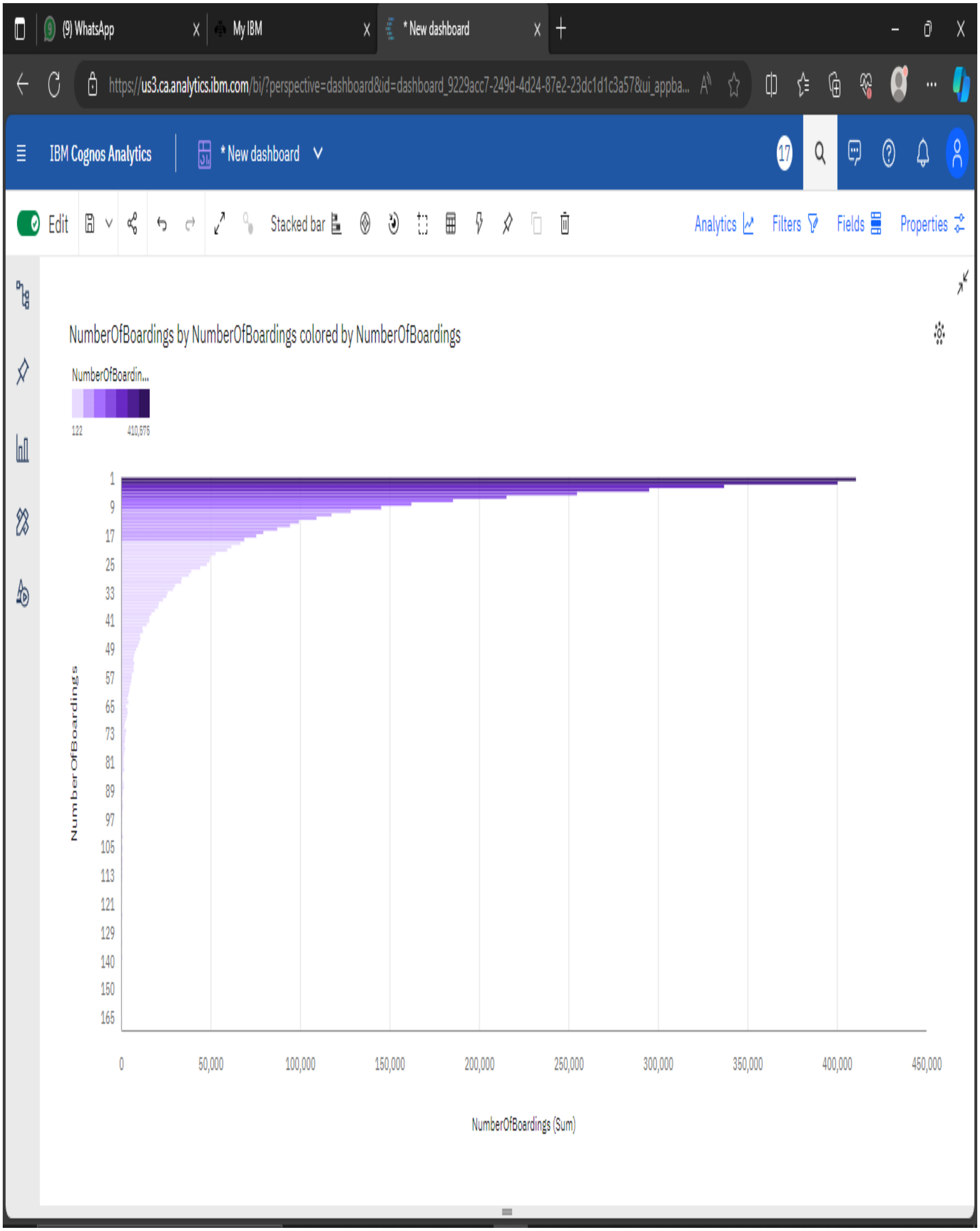

IBM Cognos Analytics
* New dashboard
What's New: To read about what's new in Dashboards, click More Info.
Dismiss
More Info
Edit
Analytics
Filters
Fields
Properties
TripID by TripID
TripID

Stop ID:

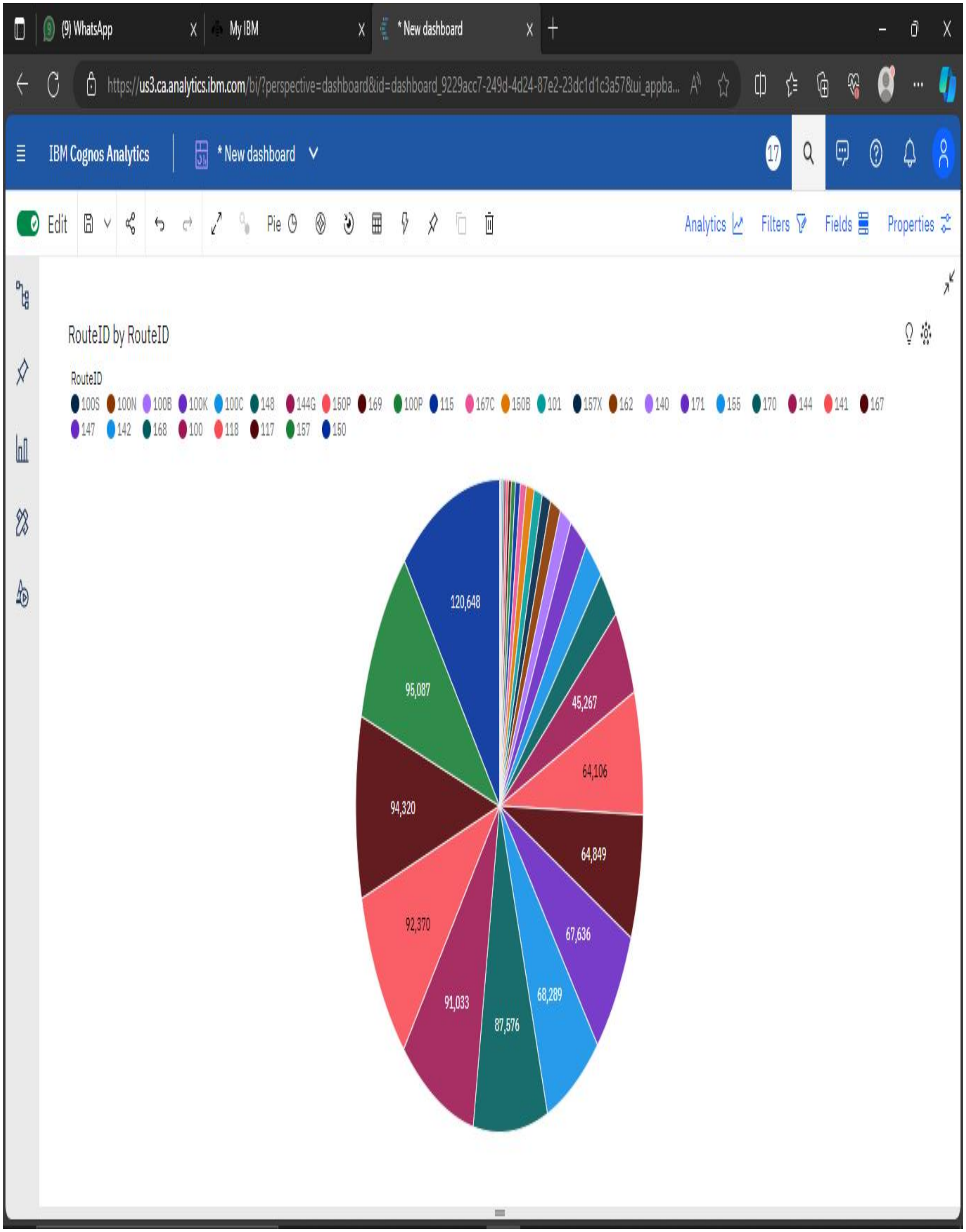


(9) WhatsApp
My IBM
* New dashboard
https://us3.ca.analytics.ibm.com/bi/?perspective=dashboard&id=dashboard_9229acc7-249d-4d24-87e2-23dc1d1c3a57&ui_appba...
IBM Cognos Analytics
* New dashboard
Edit
Scatter
Analytics
Filters
Fields
Properties
StopID by TripID with points for StopID
StopID
10817 10818 10843 10877 10879 10897 10904 10922 10924 10950 10967 10986 10992 11013 11033 11058 11072 11086 11092 11115 11118
11146 11149 11163 11165 11185 11193 11219 11224 11242 11243 11244 11248 11304 11305 11324 11329 11356 11363 11391 11394
11661
45164
11656
5438
52448
45267
5432
12125
14325
11227
34711
14345
34669
34642
TripID
10817 10843 10879 10904 10924 10967 10992 11033 11072 11092 11118 11149 11165 11193 11224 11243 11248 11305 11329 11363 1...
10818 10877 10897 10922 10950 10986 11013 11058 11086 11115 11146 11163 11185 11219 11242 11244 11304 11324 11356 11391
StopID

Number of boardings:

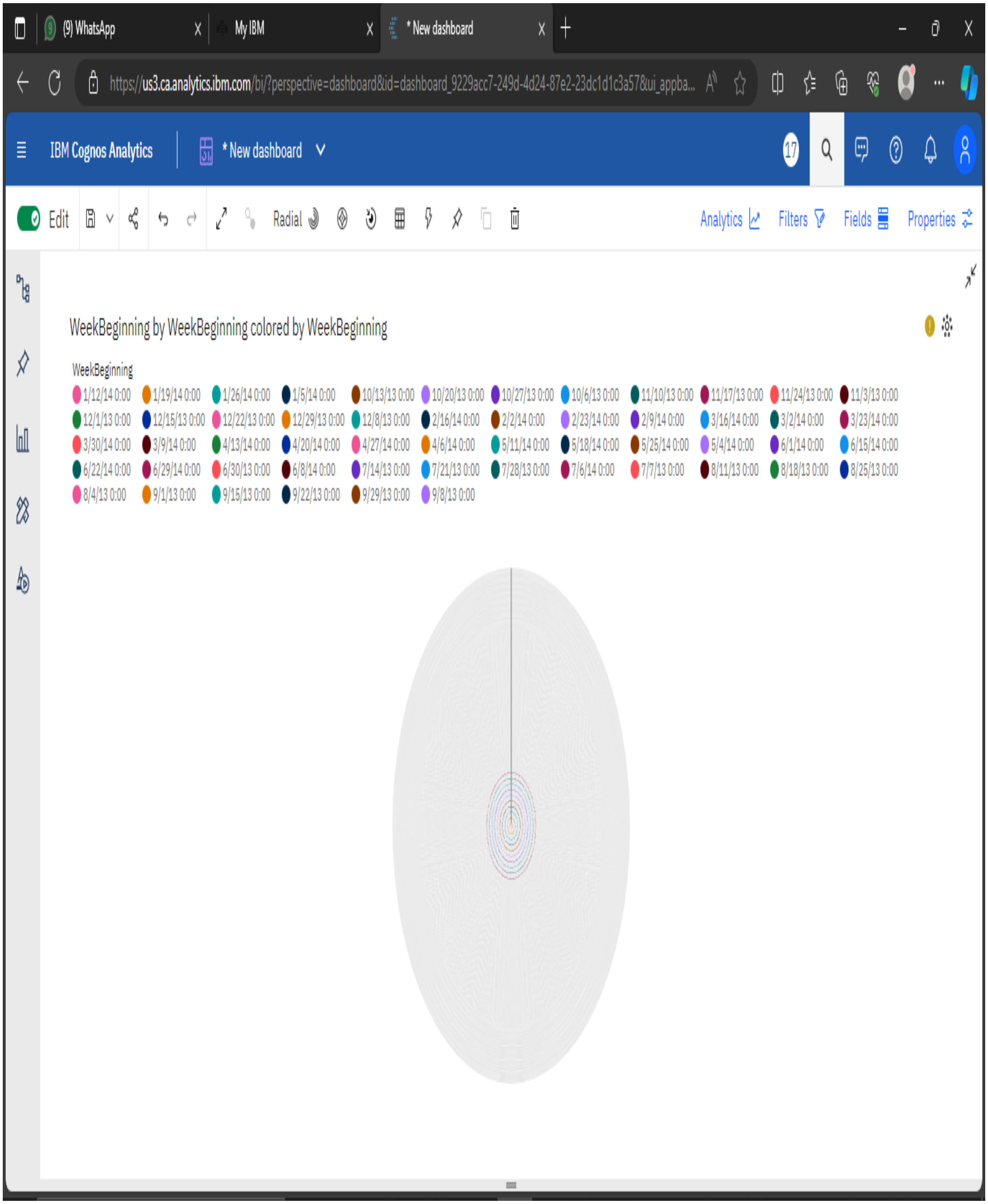


Route ID:

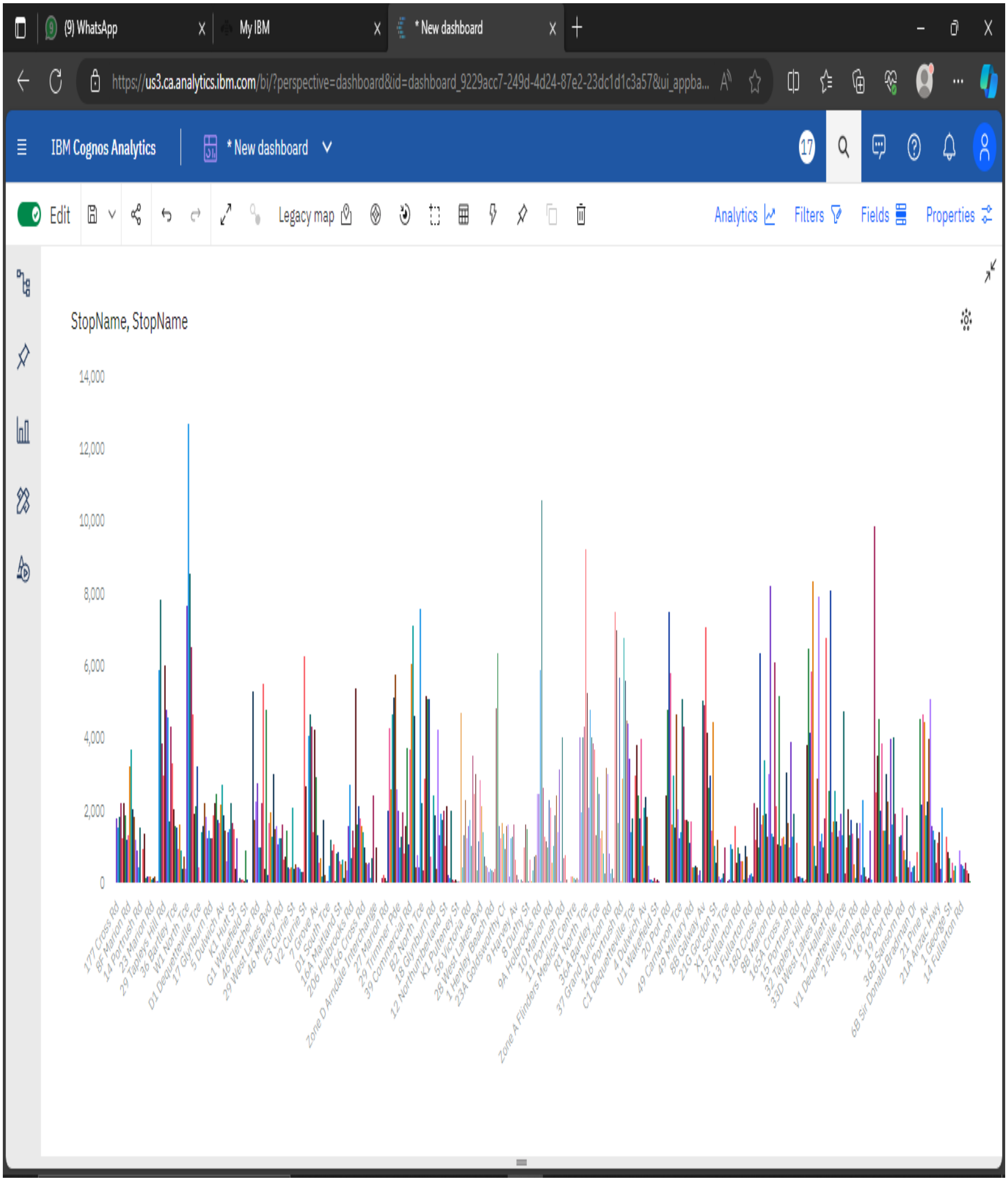


RouteID by RouteID
RouteID
100S 100N 100B 100K 100C 148 144G 150P 169 100P 115 167C 150B 101 157X 162 140 171 155 170 144 141 167
147 142 168 100 118 117 157 150
120,648
95,087
94,320
92,370
91,033
87,576
68,289
67,636
64,849
64,106
45,267

Week Beginning:



Stop Names:



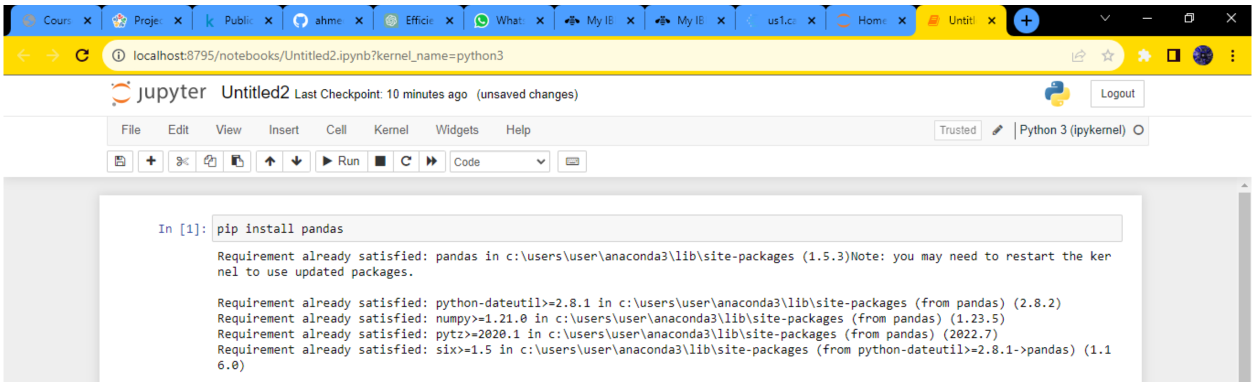
(9) WhatsApp
My IBM
* New dashboard
https://us3.ca.analytics.ibm.com/bi/?perspective=dashboard&id=dashboard_9229acc7-249d-4d24-87e2-23dc1d1c3a57&ui_appba...
IBM Cognos Analytics
* New dashboard
Edit
Legacy map
Analytics
Filters
Fields
Properties
StopName, StopName
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0

Python Integration

Objective:

In this notebook, We have explored how people are travelling from different stops in Adelaide Metropolitan area and managing the buses on each route according to the no of passenger commuting through the buses.

1. Installing pandas:

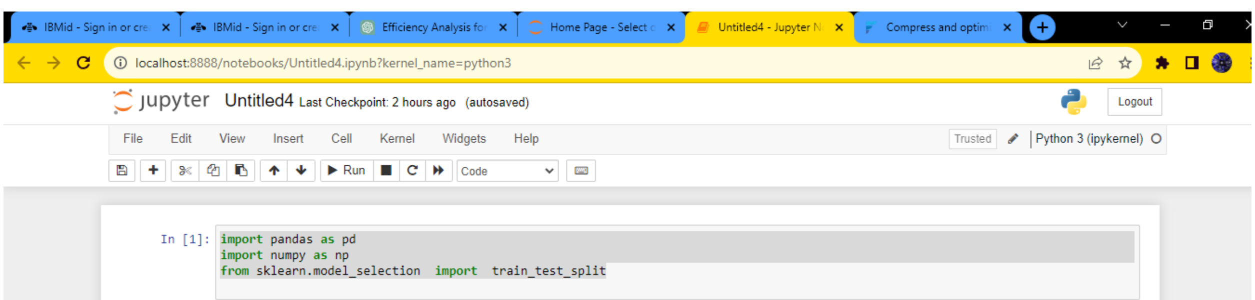


2. Importing Libraries:

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}")
```

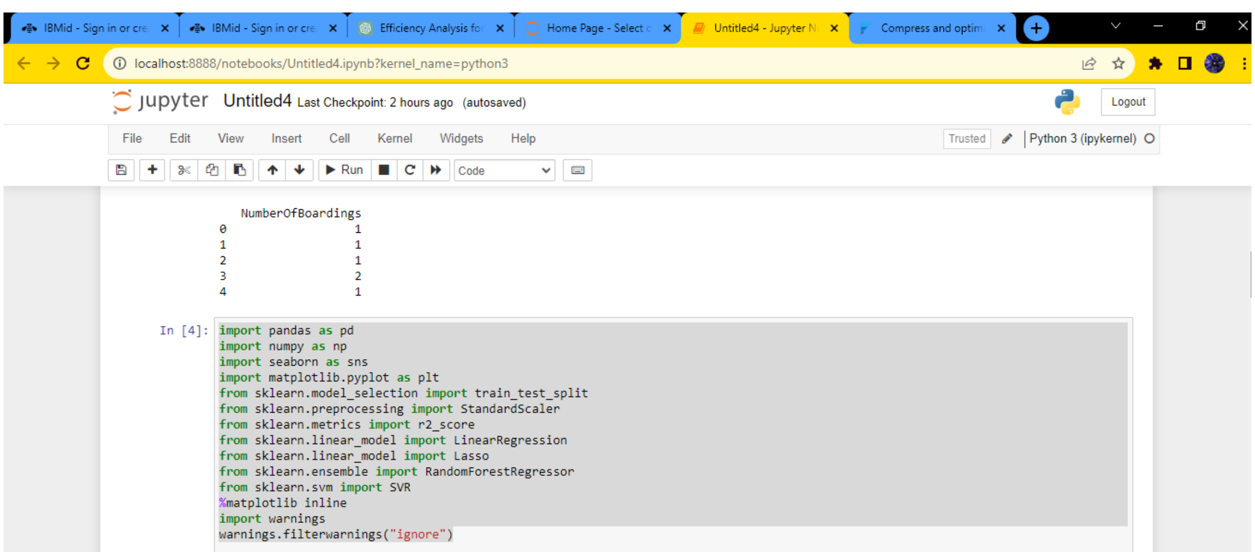
```
In [3]: import pandas as pd

        # 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
```

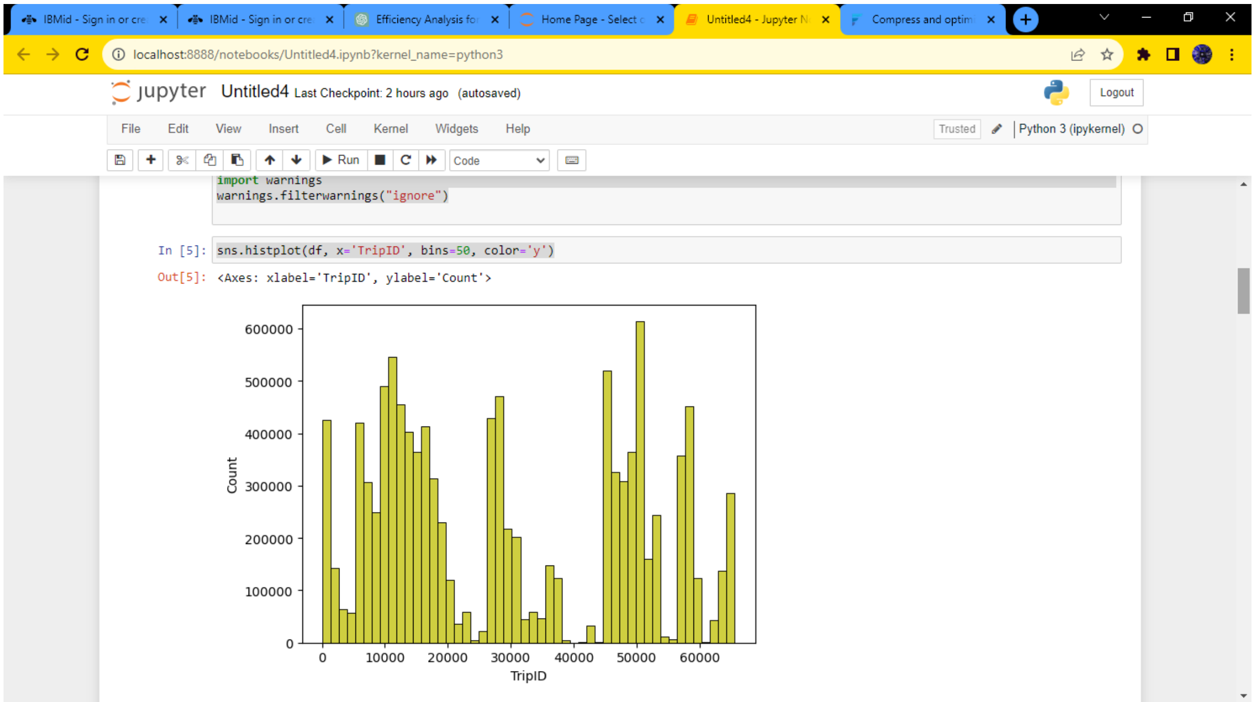
4. Loading the dataset:

```
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")
```



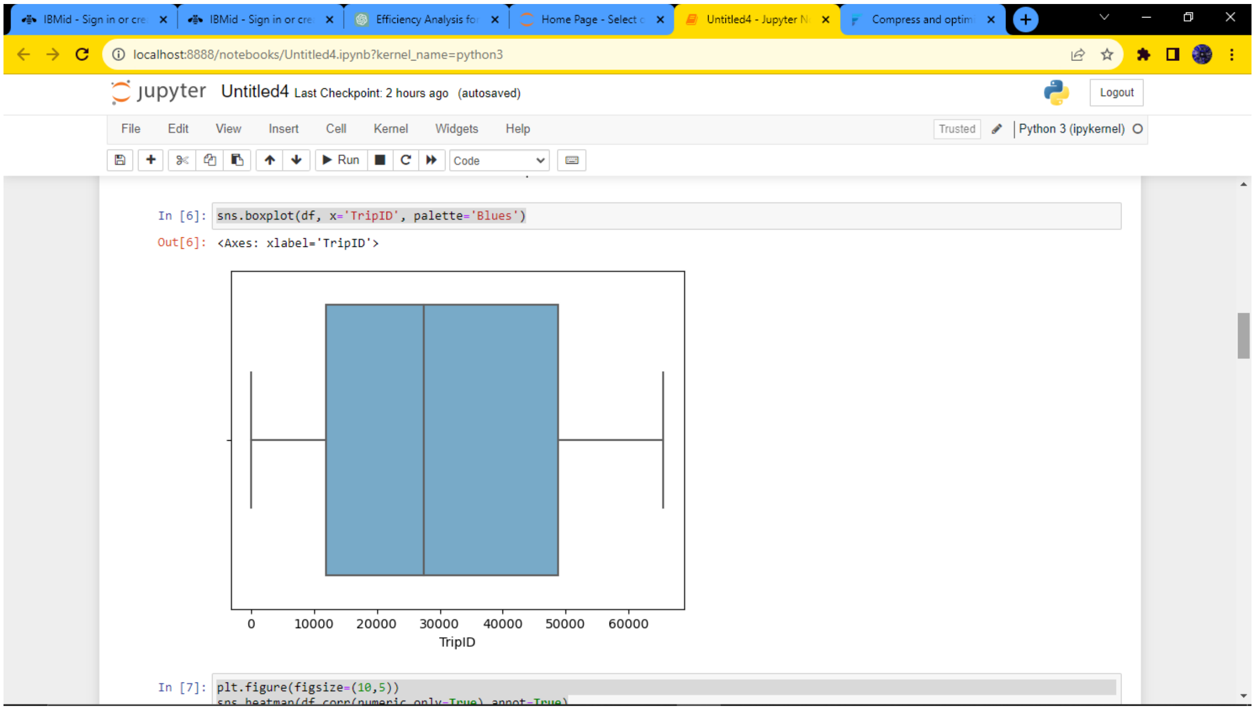
5. Bar chart:

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



6. Boxplot:

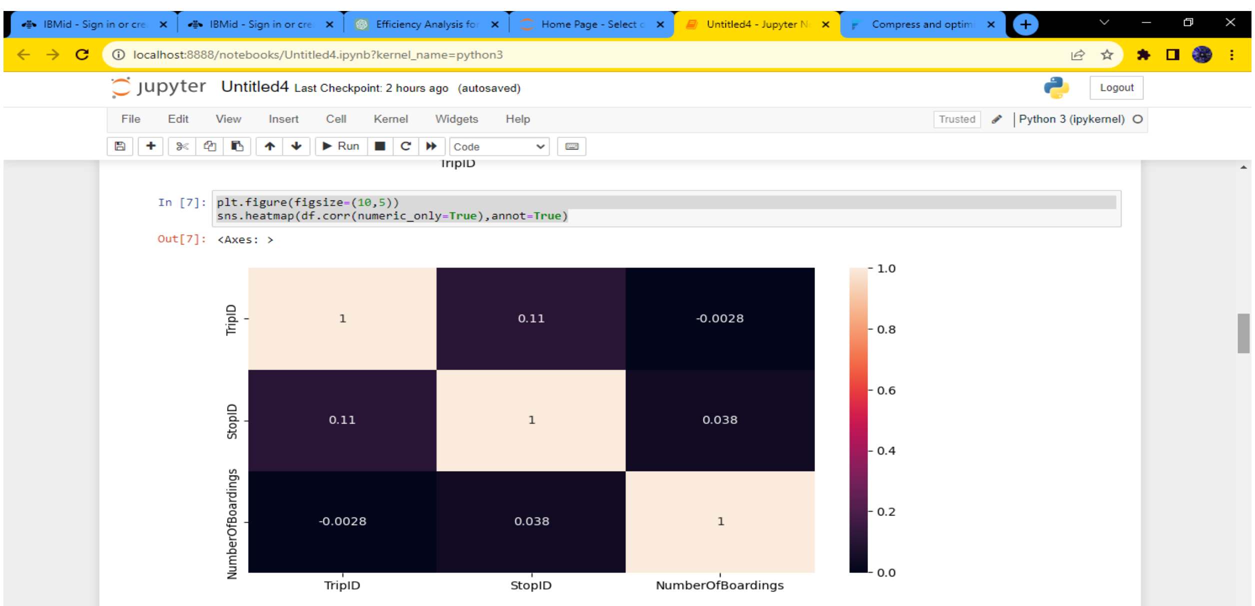
sns.boxplot(df, x='TripID', palette='Blues')



7. Heatmap:

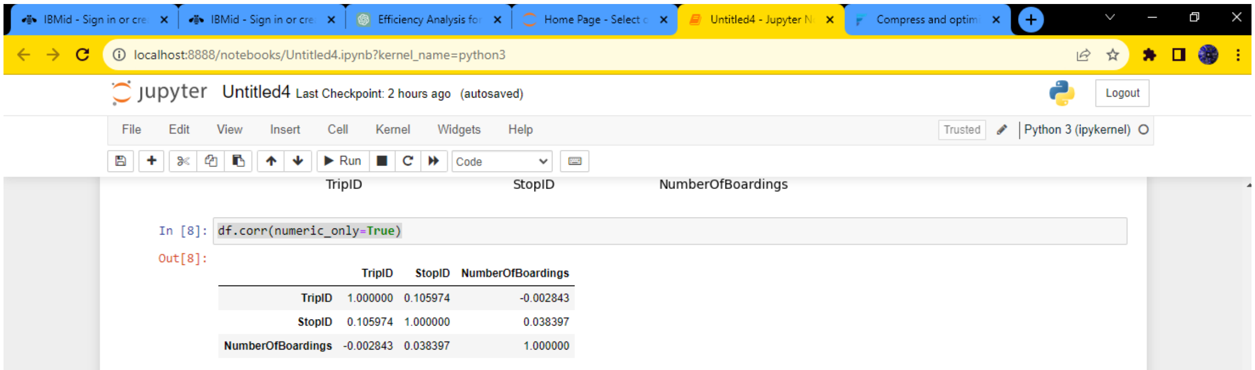
plt.figure(figsize=(10,5))

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



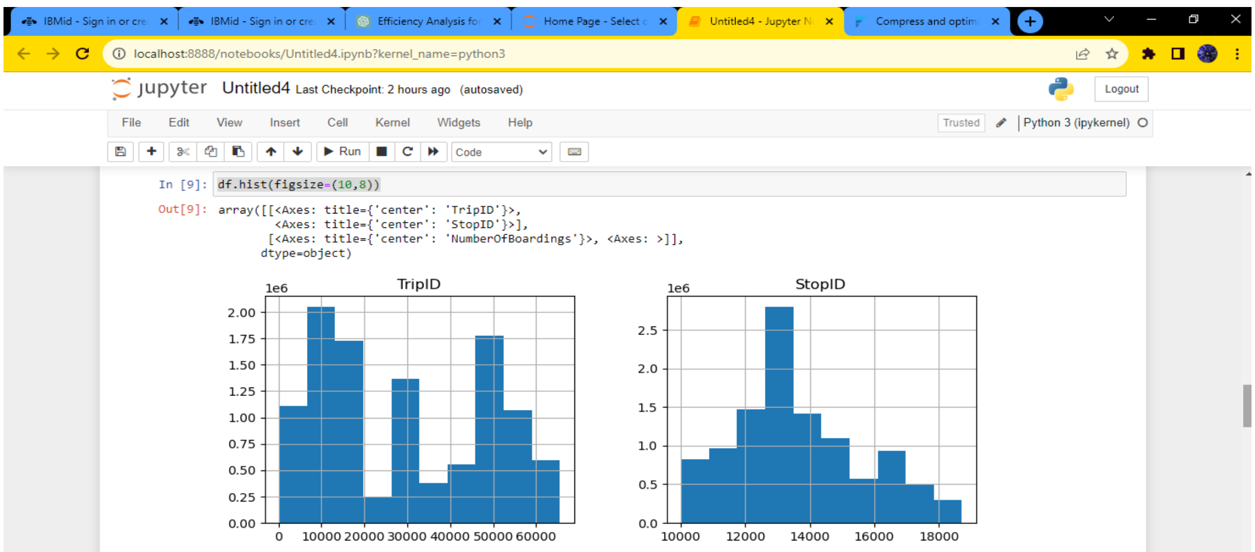
8. Correlation:

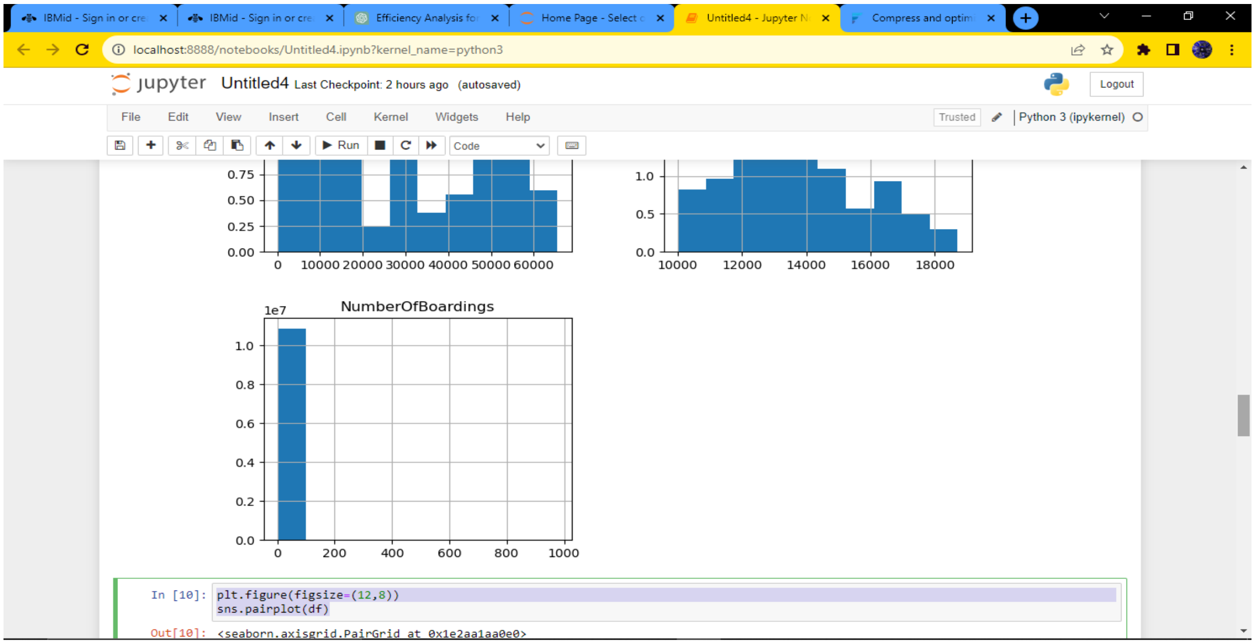
df.corr(numeric_only=True)



9. Histogram:

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

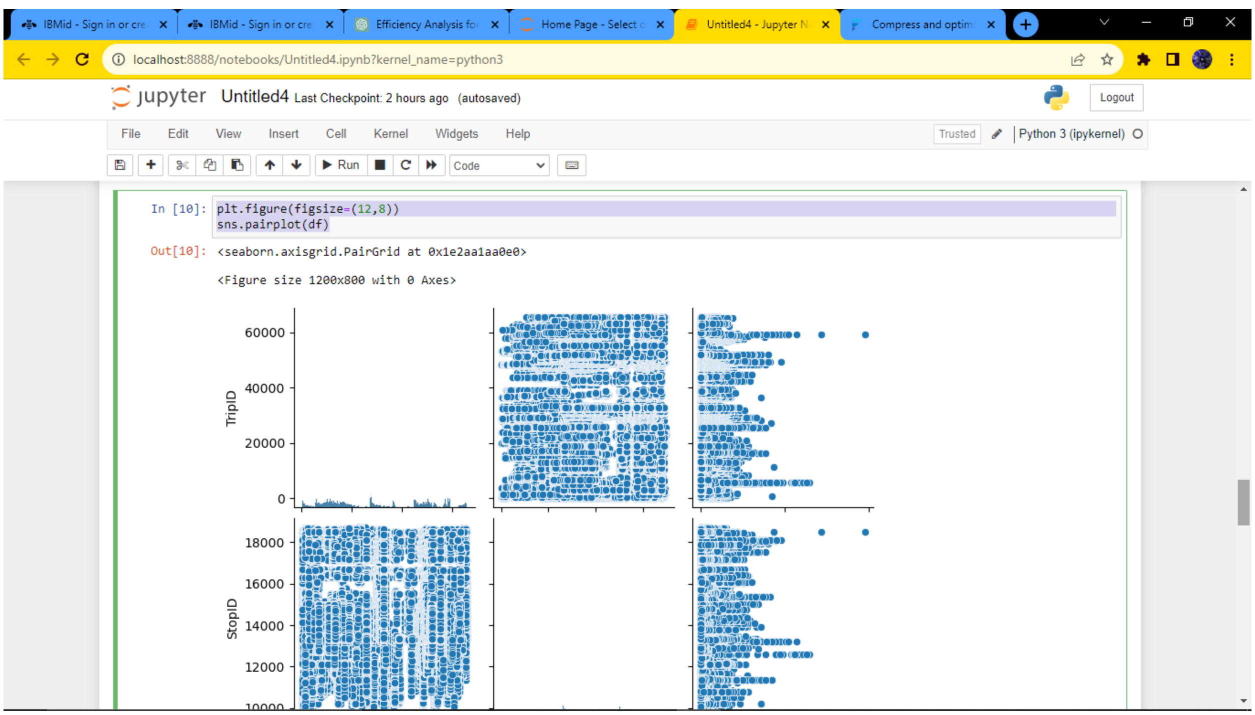


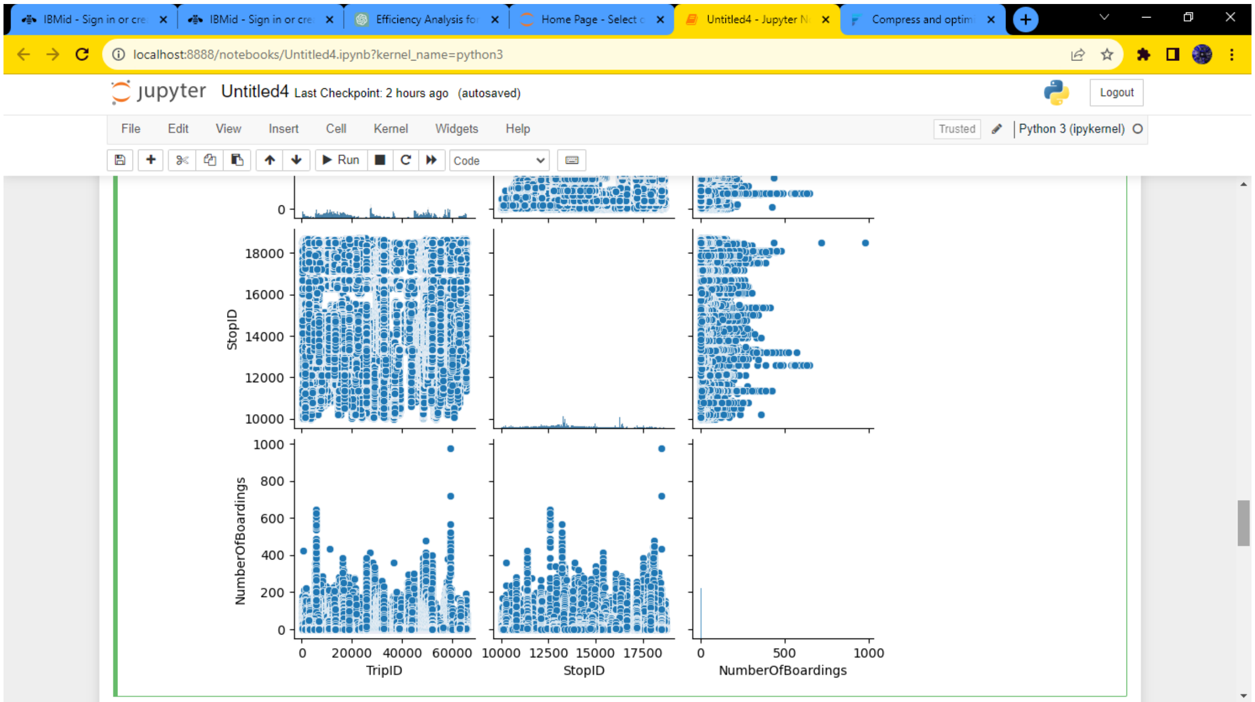


10. Pair plot:

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

sns.pairplot(df)





Conclusion:

In conclusion, the use of IBM Cognos for visualization in the public transportation efficiency analysis project has brought about positive changes, leading to more efficient and user-friendly services, better decision-making, and improved sustainability.