**PHASE 3:**

**PUBLIC TRANSPORTATION ANALYSIS**

**INTRODUCTION:**

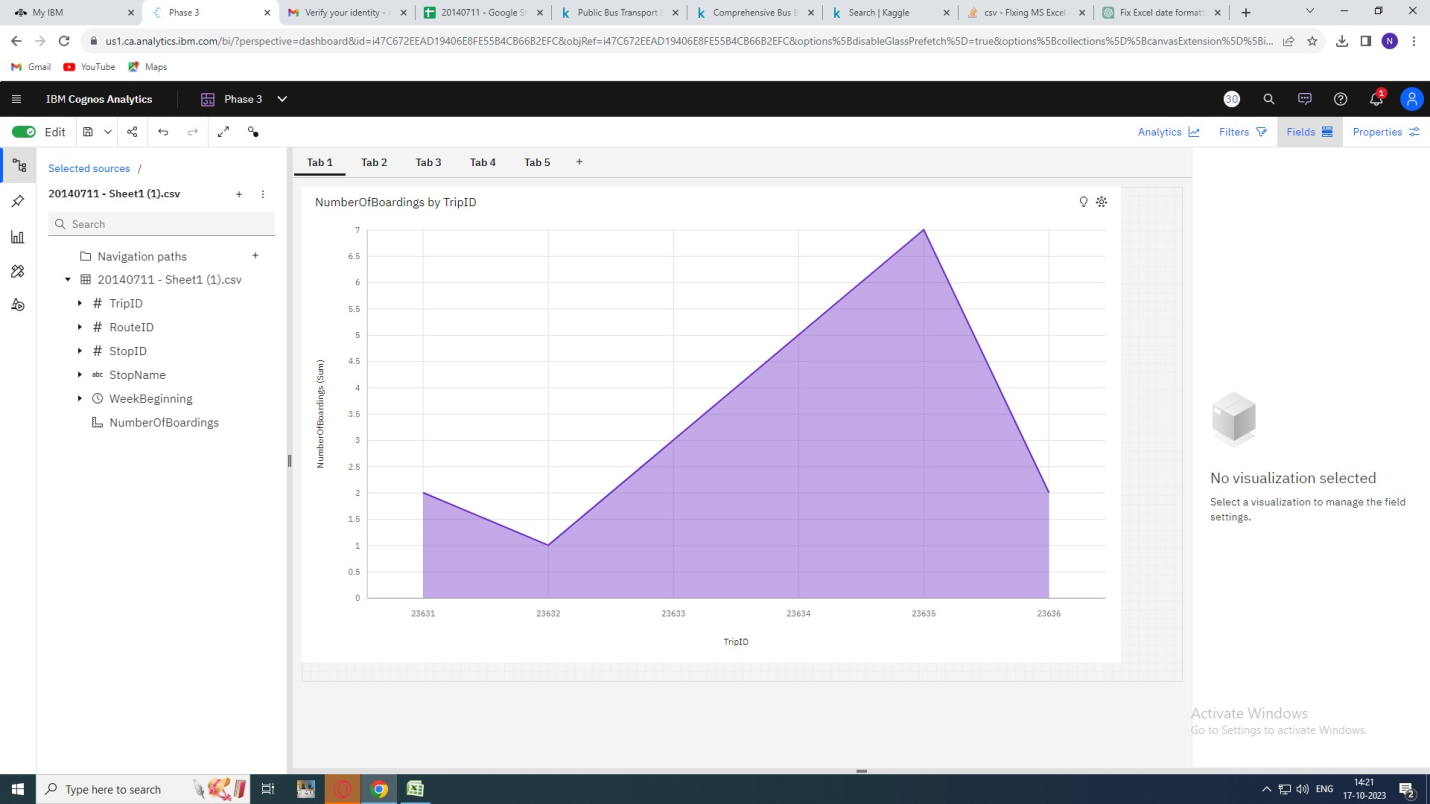
The "Public Transportation Passenger Boarding Data" dataset records information regarding passenger boardings at different bus stops within a public transportation network. This data is vital for understanding the utilization of public transportation services and analyzing the popularity of specific routes and stops. The dataset encompasses a variety of fields that provide insights into individual trips, routes, stops, and the number of boardings that occurred during a specific period. Using the given data set visualization is done using cognos tools.

**Data preprocessing:**

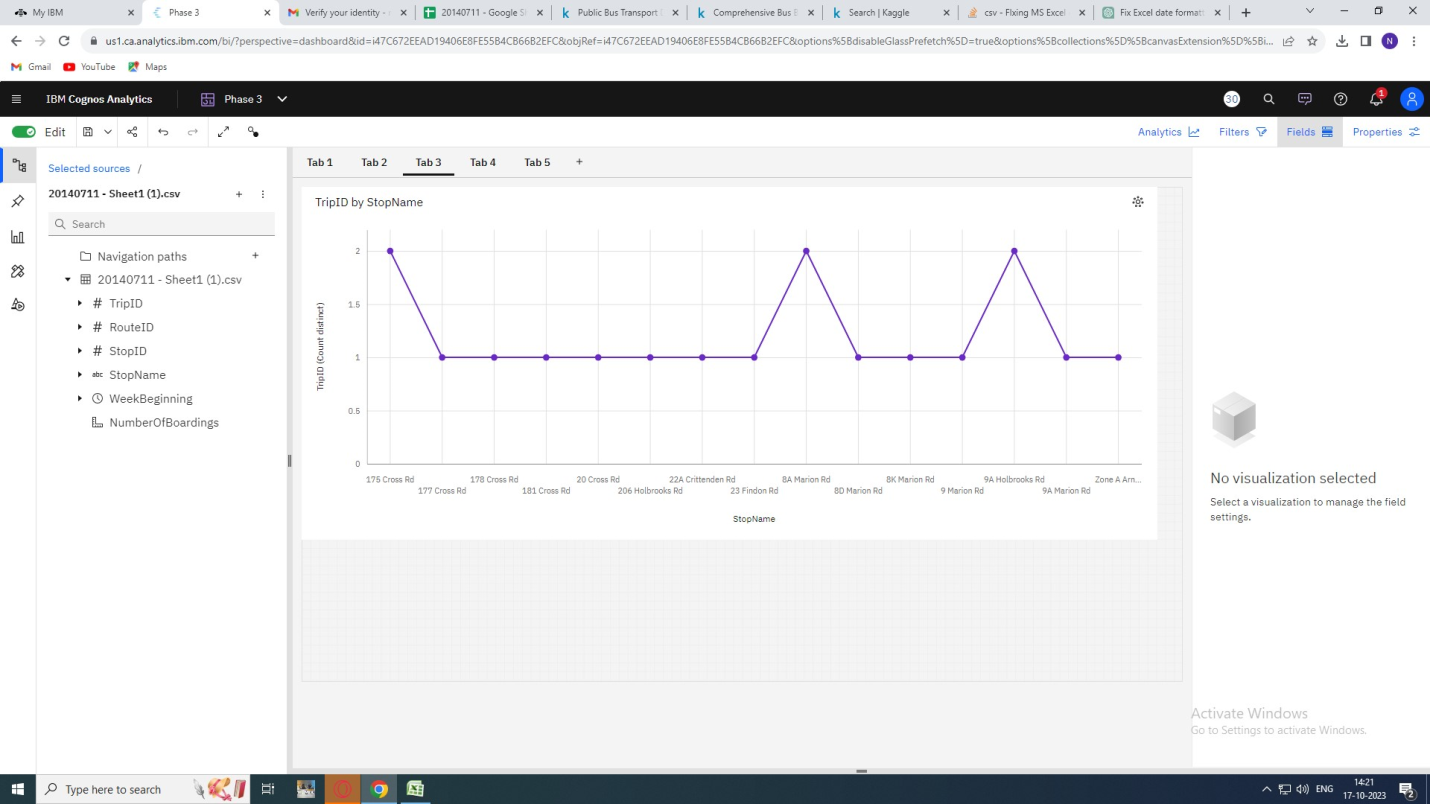
* Data Visualization:
* **Create visualizations to explore the data and understand patterns. For instance, you could plot the number of boardings over time or compare boardings across different stops or routes**.
* Data analysis:
* **Conduct various analyses based on your specific goals. For instance, you might want to identify the most popular routes, stops, or times of day for boardings.**

**DATA VISUALIZATION USING COGNOS ANALYTICS TOOLS:**

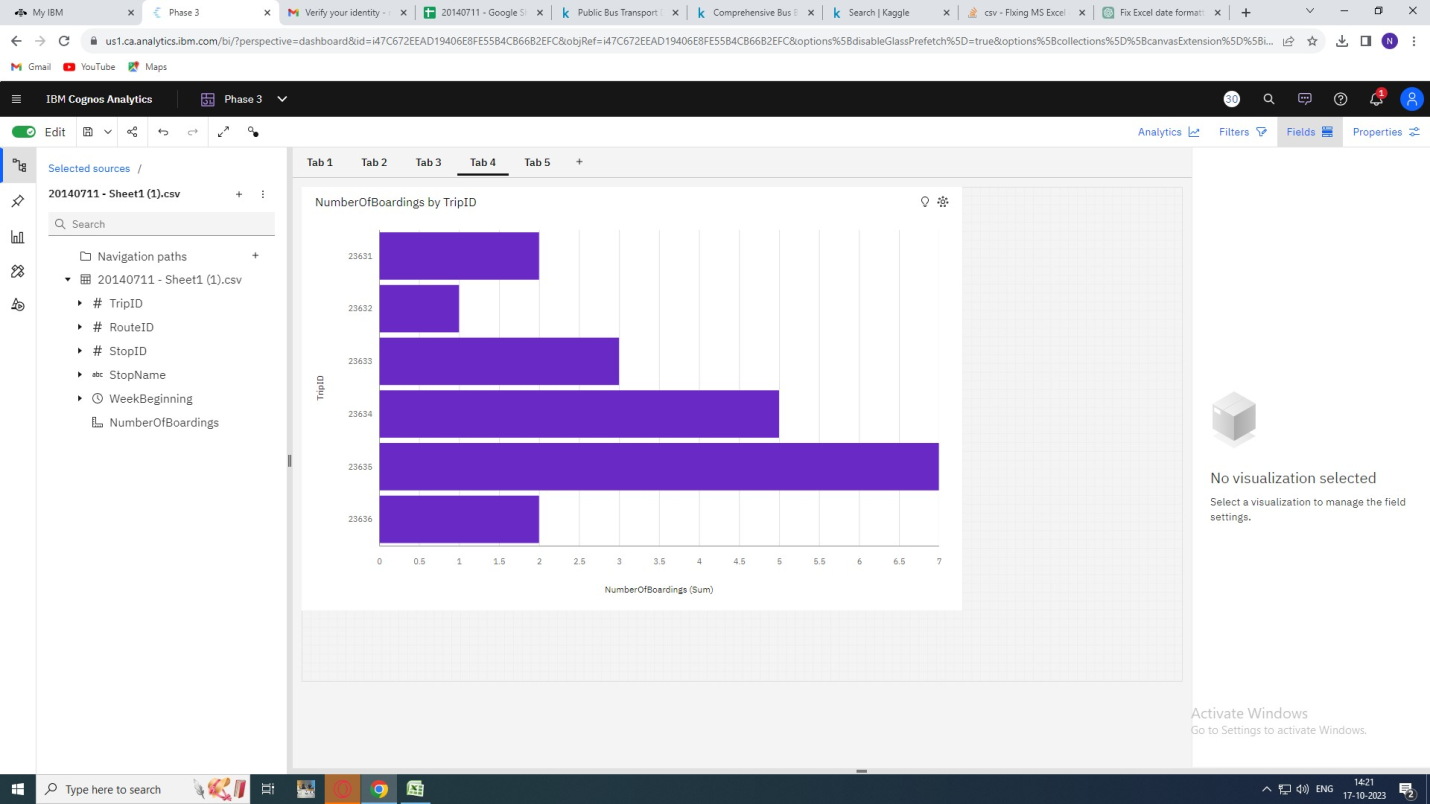
1. **The below visualization is the line chart that shows “Number of Boardings” by “TripID” which is indicated in x-axis and y-axis.**



**2. The below visualization is the line graph that shows “TripID” by “Stop name” which is indicated in x-axis and y-axis.**



**3. The below visualization is the bar graph that shows “Number of Boardings” by “TripID” which is indicated in x-axis and y-axis.**

****

**TO ENSURE QUALITY AND ACCURACY:**

**PROGRAM:**

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error, r2\_score

# Create a DataFrame from your data

data = {

'TripID': [23631, 23631, 23632, 23633, 23633, 23634, 23634, 23634, 23634, 23634, 23635, 23635, 23635, 23635, 23635, 23635, 23636, 23636, 23636],

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

'StopID': [14156, 14144, 14132, 12266, 14147, 13907, 14132, 13335, 13875, 13045, 13335, 13383, 13586, 12726, 13813, 14062, 12780, 13383, 14154],

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

}

df = pd.DataFrame(data)

# Split the data into training and testing sets

X = df[['TripID', 'RouteID', 'StopID']]

y = df['NumberOfBoardings']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Create and fit a linear regression model

model = LinearRegression()

model.fit(X\_train, y\_train)

# Make predictions on the test set

y\_pred = model.predict(X\_test)

# Evaluate the model

mse = mean\_squared\_error(y\_test, y\_pred)

r2 = r2\_score(y\_test, y\_pred)

print(f'Mean Squared Error: {mse}')

print(f'R-squared: {r2}')

# Now, you can use the model to make predictions for new data points

# For example, to predict the Number of Boardings for a new trip:

new\_trip = [[23637, 100, 14160]] # Replace with your own data

predicted\_boardings = model.predict(new\_trip)

print(f'Predicted Number of Boardings: {predicted\_boardings[0]}')

Output:

Mean Squared Error: 0.3333333333333333

R-squared: 0.0

Predicted Number of Boardings: 1.0

CONCLUSION

In this phase we have build the public transportation efficiency analysis using IBM Cognos for

Visualization.The analysis objectives and data is collected from the source shared,they are being processed and cleaned to ensure its quality and accuracy.