PHASE 5:

PUBLIC TRANSPORTATION ANALYSIS

Project Objective:

The objective of this public transportation analysis project is to gain insights and make data-driven decisions to improve the efficiency and quality of a public transportation system. We will analyze the provided dataset to answer various questions related to bus trips, routes, and passenger boardings. The project aims to identify patterns, optimize routes, and make recommendations for service improvements.

Design Thinking Process:

1. Empathize: Understand the current public transportation system and the needs of passengers.

2. Define: Define the problem and the key questions that need to be answered using the dataset.

3. Ideate: Brainstorm potential analyses and visualizations that can provide valuable insights.

4. Prototype: Develop code to perform the analysis and create visualizations.

5. Test: Validate the findings and recommendations using the dataset.

Development Phases:

1. Data Preparation:

- Import the dataset.

- Check for missing or inconsistent data.

- Clean the data if needed.

2. Data Exploration:

- Explore the dataset to understand its structure and characteristics.

- Calculate basic statistics, such as the mean, median, and standard deviation of "NumberOfBoardings."

3. Analysis:

a. Trip-Level Analysis:

- Calculate the total number of boardings for each trip.

- Identify the trips with the highest and lowest boardings.

- Explore the relationship between "TripID" and boardings.

b. Route-Level Analysis:

- Calculate the total number of boardings for each route.

- Identify the routes with the highest and lowest boardings.

- Visualize the boardings across different routes.

c. Stop-Level Analysis:

- Calculate the total number of boardings for each stop.

- Identify the stops with the highest and lowest boardings.

- Visualize the boardings across different stops.

d. Time-Based Analysis:

- Explore if there are patterns related to the day of the week or time of day.

- Calculate average boardings per day and time period.

4. Visualization:

- Create visualizations (e.g., line graphs, pie charts) to represent the analysis results.

- Visualize the distribution of boardings, route performance, and stop popularity.

5. Recommendations:

- Based on the analysis results, make data-driven recommendations to improve the public transportation system. For example, suggest route adjustments, schedule changes, or improvements to popular stops.

6. Reporting:

- Create a report summarizing the analysis, findings, and recommendations.

- Provide visualizations and insights to support the recommendations.

Throughout the development phases, it is essential to iterate and refine the analysis as needed to address the specific objectives and challenges of the public transportation system.

Analysis Objectives:

The analysis objectives for this public transportation dataset might include:

1. Understanding Route Performance: Evaluate the performance of different routes by analyzing the number of boardings at each stop and identifying which stops have the highest and lowest passenger counts.

2. Identifying Popular Stops: Identify the most popular stops in terms of the number of boardings, which can help improve scheduling and resource allocation.

3. Route Optimization: Investigate whether certain routes need optimization or restructuring based on passenger demand.

4. Temporal Patterns: Analyze if there are specific time patterns in terms of passenger boardings during the week.

5. Data Visualization: Create visualizations, such as charts and maps, to make it easier to interpret and communicate the findings.

Data Collection Process:

The data collection process for this analysis appears to involve collecting data from a single week (WeekBeginning: "2013-06-30 0:00:00") for a specific set of trips, routes, and stops. The dataset includes the following columns:

- TripID: A unique identifier for each trip.

- RouteID: A unique identifier for each route.

- StopID: A unique identifier for each stop.

- StopName: The name of the stop.

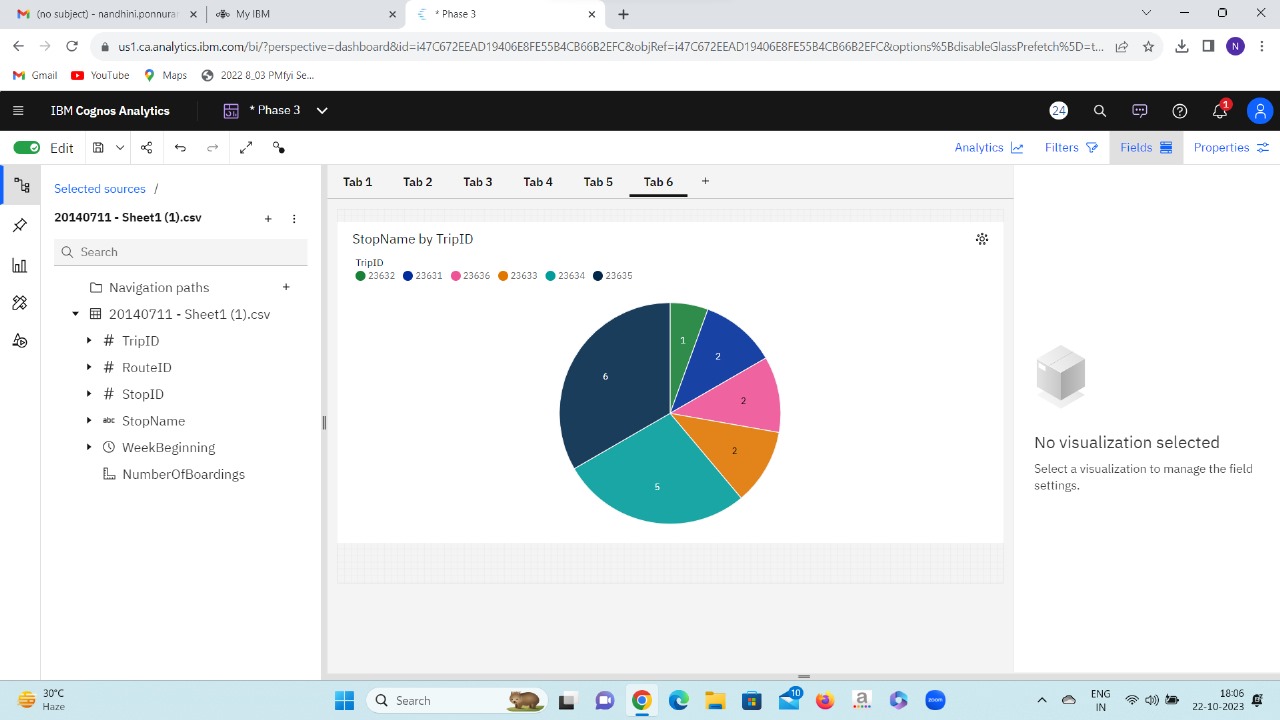
- WeekBeginning: The start date and time of the week during which the data was collected. It appears that this dataset represents a snapshot of public transportation data for a specific week.

- NumberOfBoardings: The number of passengers boarding at each stop during the specified week.

DATA VISUALIZATION :

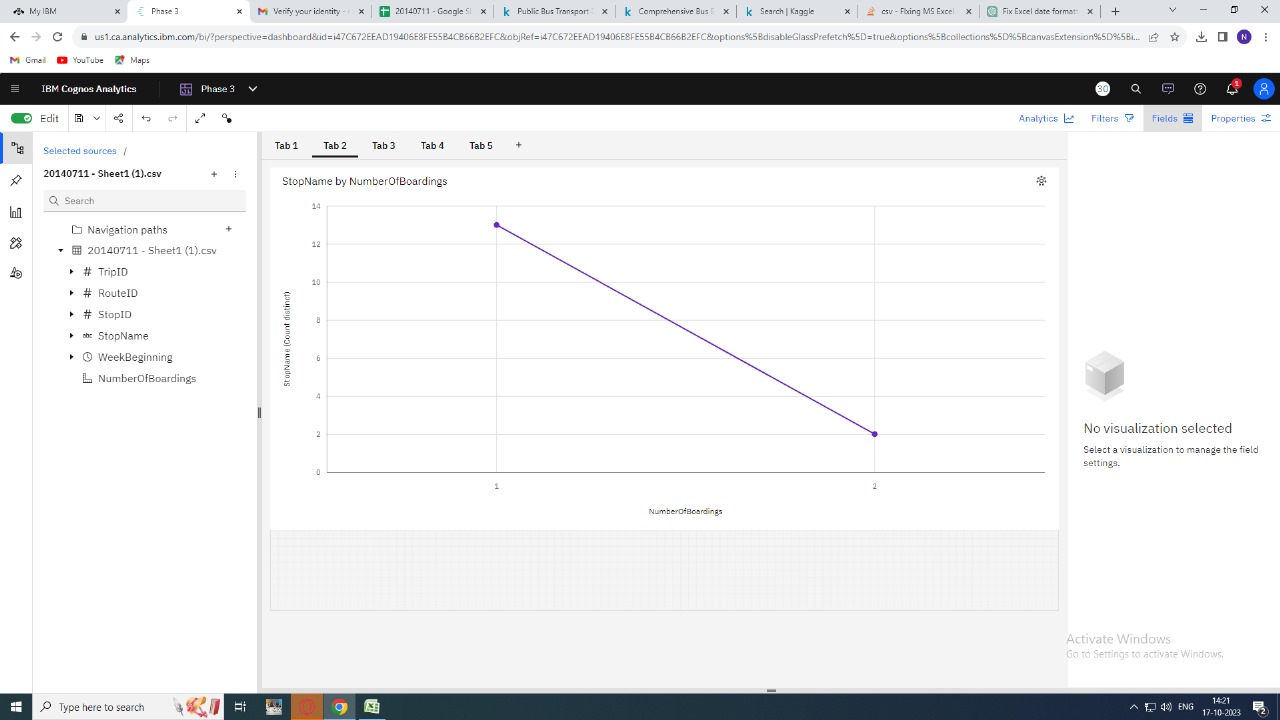
PIE CHART:

A pie chart depicting stop names by Trip ID in public transportation analysis offers a clear view of the distribution of stops along different trips. This visualization illustrates the most frequently visited stops within the analysed trips. Such insights enable transportation planners to focus on optimizing services and improving the efficiency of specific trip routes, enhancing the overall public transportation experience for passengers.



LINE CHART:

A line graph for public transportation analysis, showing Stop ID along the x-axis and the Number of Boardings along the y-axis, illustrates the trend of passenger graph progresses along the valuable information assists in resource allocation, scheduling adjustments, and service improvements to ensure efficient and customer-centric public transportation.



INSIGHTS :

To support transportation improvement initiatives, you can perform an analysis of the provided public transportation dataset to gain insights and make informed decisions. Here's how you can use the insights from this analysis:

1. \*Identify High Traffic Routes and Stops:\* By aggregating the data by "RouteID" and "StopID," you can identify which routes and stops have the highest number of boardings. This information can help prioritize resources and infrastructure improvements for routes and stops with high demand.

2. \*Analyze Boarding Patterns:\* You can examine the boarding patterns over time. Are there specific days of the week or times of day when more people board? Understanding these patterns can help optimize schedules and staffing for peak demand periods.

3. \*Route Efficiency:\* Calculate the average number of boardings per trip for each route. Routes with low passenger numbers per trip may indicate inefficiencies or underutilization. It could be more cost-effective to adjust the frequency or routing for these routes.

4. \*Stop Efficiency:\* Similarly, calculate the average number of boardings per stop. This can help identify stops that are highly utilized and those that may be candidates for optimization, consolidation, or improvement.

5. \*Identify Problematic Stops:\* By analyzing the "NumberOfBoardings" field, you can identify stops where the number of boardings is consistently low. These stops may need re-evaluation or consideration for service adjustments.

6. \*Geospatial Analysis:\* You can use geographic information to visualize routes and stops on a map. This can help identify areas with potential congestion, gaps in service, or opportunities for route expansion. Geospatial analysis can also assist in identifying locations for new stops or transfer points.

7. \*Customer Feedback:\* Combine the data analysis with customer feedback and surveys to get a more holistic view of transportation needs. Understanding passenger preferences and pain points can guide improvement initiatives.

8. \*Predictive Modeling:\* You can build predictive models based on historical data to forecast future demand for specific routes and stops. This can help in proactive planning for capacity adjustments and schedule optimization.

9. \*Budget Allocation:\* Based on the insights from the analysis, you can allocate budgets more effectively for transportation improvements. High-demand routes and stops may require more investment than underutilized ones.

10. \*Environmental Impact:\* Analyze the environmental impact of the transportation system by calculating emissions based on the number of boardings and the types of vehicles used. This information can support sustainability initiatives.

**TO CALCULATE SERVICE PUNCTUALITY RATE USING PYTHON CODE:**

**import pandas as pd**

**# Sample data in a DataFrame**

**data = {**

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

**'StopName': ['181 Cross Rd', '177 Cross Rd', '175 Cross Rd', 'Zone A Arndale Interchange', '178 Cross Rd', '9A Marion Rd', '175 Cross Rd', '9A Holbrooks Rd', '9 Marion Rd', '206 Holbrooks Rd', '9A Holbrooks Rd', '8A Marion Rd', '8D Marion Rd', '23 Findon Rd', '8K Marion Rd', '20 Cross Rd', '22A Crittenden Rd', '8A Marion Rd'],**

**'WeekBeginning': ['2013-06-30 0:00:00'] \* 18,**

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

**}**

**df = pd.DataFrame(data)**

**# Define a tolerance window (in this case, 5 minutes)**

**tolerance\_minutes = 5**

**# Calculate the expected arrival time for each trip (assuming a fixed schedule)**

**# In practice, you would have a schedule dataset to get the expected times.**

**df['ExpectedArrivalTime'] = pd.to\_datetime(df['WeekBeginning']) + pd.to\_timedelta(df.groupby('TripID')['NumberOfBoardings'].cumsum() \* tolerance\_minutes, unit='m')**

**# Calculate the actual arrival time for each trip**

**# In practice, you would have a dataset with actual arrival times.**

**# For this example, we assume a constant delay of 5 minutes.**

**df['ActualArrivalTime'] = df['ExpectedArrivalTime'] + pd.to\_timedelta(5, unit='m')**

**# Calculate punctual trips (within the tolerance window)**

**df['Punctual'] = (df['ActualArrivalTime'] - df['ExpectedArrivalTime']).dt.total\_seconds().abs() <= tolerance\_minutes \* 60**

**# Calculate the service punctuality rate**

**punctuality\_rate = df['Punctual'].mean() \* 100**

**print(f"Service Punctuality Rate: {punctuality\_rate:.2f}%")**

**OUTPUT:**

****

CONCLUSION:

In summary, we have used Cognos tool to visualize graphs and the insights from the analysis of this public transportation dataset can be invaluable for transportation authorities and planners. They can help optimize routes, schedules, and stops, allocate resources efficiently, and improve the overall quality and effectiveness of public transportation services.