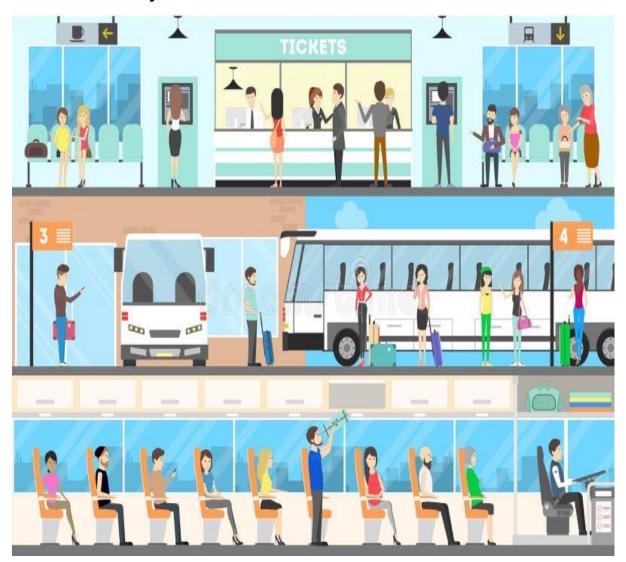
PUBLIC TRANSPORTATION ANALYSIS

Phase 5: Project Documentation and Submission



Project Objective:

The project involves analysing public transportation data to assess service efficiency, on time performance, and passenger feedback. The objective is to provide insights that support transportation improvement initiatives and enhance the overall public transportation experience. This project includes defining analysis objectives, collecting transportation data, designing relevant visualizations in IBM Cognos, and using code for data analysis.

Design Thinking:

- 1.Analysis Objectives: Define specific objectives for analyzing public transportation data, such as assessing on-time performance, passenger satisfaction, and service efficiency.
- 2.Data Collection: Identify the sources and methods for collecting transportation data, including schedules, real-time updates, and passenger feedback.
- 3. Visualization Strategy: Plan how to visualize the insights using IBM Cognos to create informative dashboards and reports.
- 4.Code Integration: Decide which aspects of the analysis can be enhanced using code, such as data cleaning, transformation, and statistical analysis.

Development phases:

1.Loading the dataset:

Loading a dataset is a fundamental step in data analytics, where you import your data into the chosen data analysis tool or software for further exploration and analysis. The specific steps for loading a dataset can vary depending on the software or programming language you are using. Here's a general overview of how to load a dataset:

- Choose Your Data Analysis Tool
- Prepare Your Data File
- Import Your Data
 - a. Python (using Pandas)
 - b. **R**

2. Preprocessing:

Some common data preprocessing tasks include:

Data cleaning:

- This involves identifying and correcting errors and inconsistencies in the data.
- For example, this may involve removing duplicate records, correcting typos, and filling in missing values.

Handling Duplicate Data:

Identify and remove duplicate records to avoid skewing analysis results.

3.Data Visualization:

• Use data visualization techniques to explore the dataset, identify trends, and gain insights before analysis begins.

Analysis Objectives:

- On-Time Performance: Measure and improve the punctuality of public transportation by analyzing data to determine the percentage of buses/trains arriving on time and identifying factors contributing to delays.
- **Passenger Satisfaction:** Assess passenger satisfaction by collecting and analyzing feedback through surveys or social media sentiment analysis, focusing on aspects like cleanliness, safety, and overall experience.
- Service Efficiency: Evaluate service efficiency by analyzing operational data, such as the cost per passenger mile, frequency of breakdowns, and utilization of resources.

Data Collection:

On-Time Performance:

Data Sources: Real-time GPS tracking systems, historical schedule data, traffic data, incident reports.

Data Collection Methods: Integration with GPS APIs, sensors on vehicles, and automated schedule tracking.

• Passenger Satisfaction:

Data Sources: Passenger surveys, social media mentions, customer service logs.

Data Collection Methods: Online surveys, sentiment analysis tools, and manual entry of feedback.

Service Efficiency:

Data Sources: Financial records, maintenance logs, fleet management systems.

Data Collection Methods: Data extraction from financial systems, monitoring of maintenance logs, and integration with fleet management software.

Visualization Strategy (Using IBM Cognos):

On-Time Performance:

Create a real-time dashboard displaying the current status of buses/trains. Generate line charts showing historical on-time performance trends.

Passenger Satisfaction:

Develop a sentiment analysis dashboard summarizing positive and negative feedback.

Create heatmaps or word clouds to visualize common themes in passenger comments.

Service Efficiency:

Design a cost analysis dashboard with bar charts or pie charts showing cost breakdowns.

Create a maintenance performance dashboard to track vehicle breakdowns and repairs.

Code Integration:

Data Cleaning:

Use Python or R scripts to automate data cleaning processes, removing duplicates, handling missing values, and standardizing data formats.

• Data Transformation:

Write code to transform raw data into a format suitable for analysis, e.g., aggregating GPS data by time intervals.

Statistical Analysis:

Utilize statistical libraries in Python or R to perform in-depth analysis, such as regression analysis to identify factors affecting on-time performance.

How the insights from the analysis can support transportation improvement initiatives:

Insights from public transportation analysis play a crucial role in supporting transportation improvement initiatives. They provide valuable information and data-driven recommendations that transportation authorities, policymakers, and planners can use to enhance the efficiency, accessibility, and sustainability of public transportation systems. Here's how these insights support transportation improvement initiatives:

1.Route Optimization:

 Insights from the analysis can identify routes that are underutilized or prone to congestion. This information helps transportation authorities optimize routes by adjusting schedules, reallocating resources, or redesigning routes to reduce commute times and enhance overall service quality.

2. Service Planning:

Analysing ridership patterns and behaviour helps transportation providers plan services
that align with actual demand. Insights facilitate efficient service planning, reducing
operational costs and enhancing service quality.

3.Performance Monitoring:

• Regular analysis allows for ongoing performance monitoring. Authorities can identify and address issues as they arise, ensuring that the transportation system operates effectively and meets established performance metrics.

4.Emergency Preparedness:

• Insights from historical data can inform emergency preparedness strategies.

Understanding how the transportation system was affected during emergencies, such as natural disasters or public health crises, helps authorities plan for resilience and recovery.

Insights from public transportation analysis are a cornerstone of evidence-based decision-making and improvement initiatives. By understanding the challenges, needs, and opportunities within public transportation systems, authorities can develop and implement strategies that lead to more efficient, accessible, and sustainable transportation services, ultimately benefiting commuters and the broader community.

Instructions on how to replicate the analysis and generate visualizations using IBM Cognos and perform data analysis using code:

Replicating an analysis and generating visualizations using IBM Cognos involves several steps. IBM Cognos is a powerful business intelligence and reporting tool that allows you to create interactive reports and dashboards. Here's a simplified step-by-step guide to replicate an analysis and generate visualizations:

Prerequisites:

1. Access to IBM Cognos Analytics: You need access to the IBM Cognos Analytics platform.

Replicating the Analysis:

1.Log in to IBM Cognos:

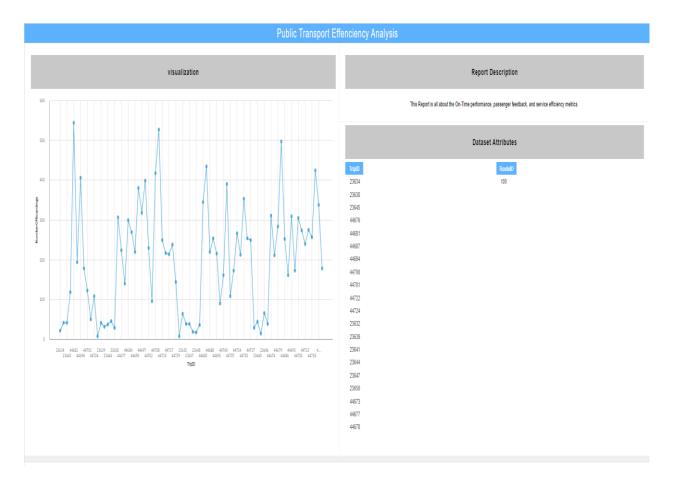
Open your web browser, go to the URL where IBM Cognos is hosted, and log in with your credentials.

2.Data Source Connection:

Ensure that you have a connection to your data source. You can connect to various data sources, including databases, Excel files, or other data repositories.

3. Create a New Report or Dashboard:

- In the Cognos interface, navigate to the area where you want to create a new report or dashboard (e.g., Reports, Dashboards, or Workspaces).
- Create a new report or dashboard and select your data source.



4. Select Data:

Drag and drop the data elements (dimensions, measures) you want to analyse onto your report canvas.

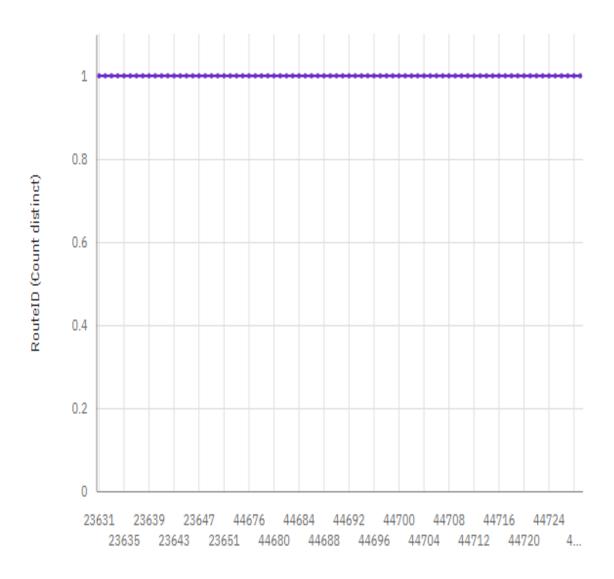
5. Data Transformation and Filtering:

Apply data transformations and filters as needed to focus on the specific aspects of the data you want to analyse.

6. Create Visualizations:

- IBM Cognos offers a range of visualization options, including charts, tables, maps, and more.
- Drag and drop the appropriate visualization type onto your canvas.
- Configure the visualization by assigning data elements to the required fields (e.g., x-axis, y-axis, colour).



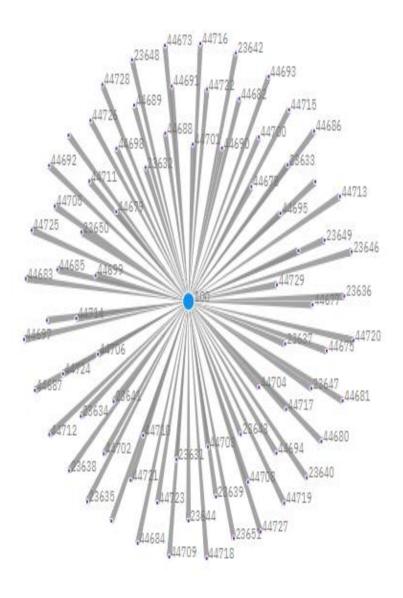


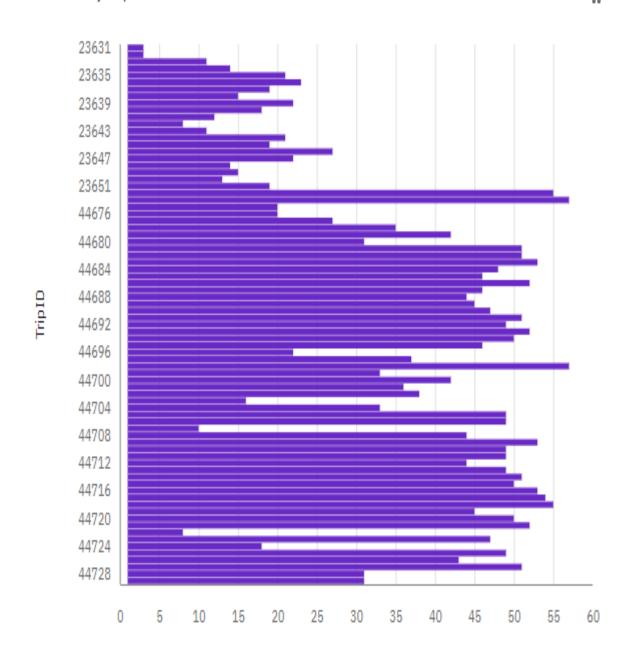
TripID

TripID to RouteID

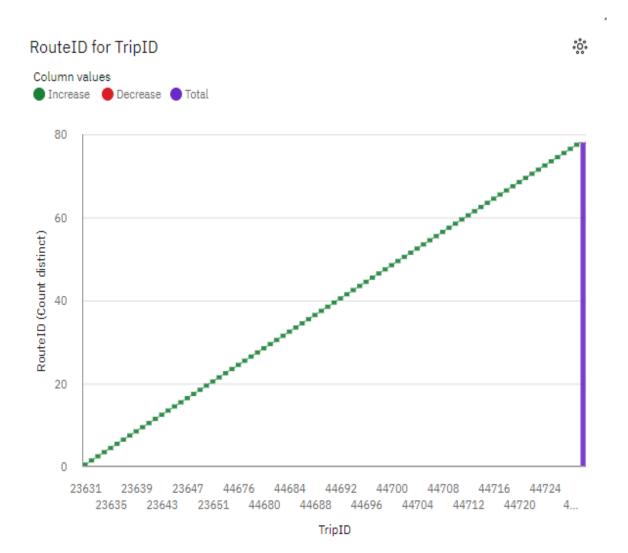






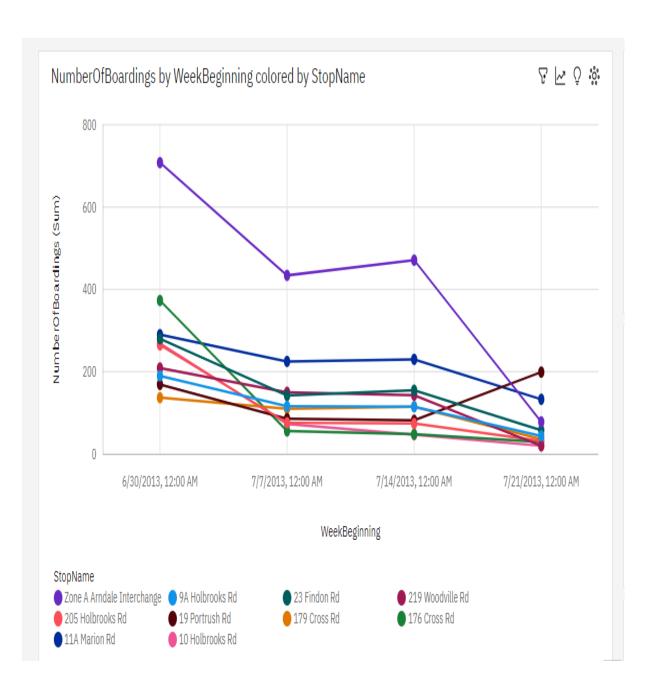


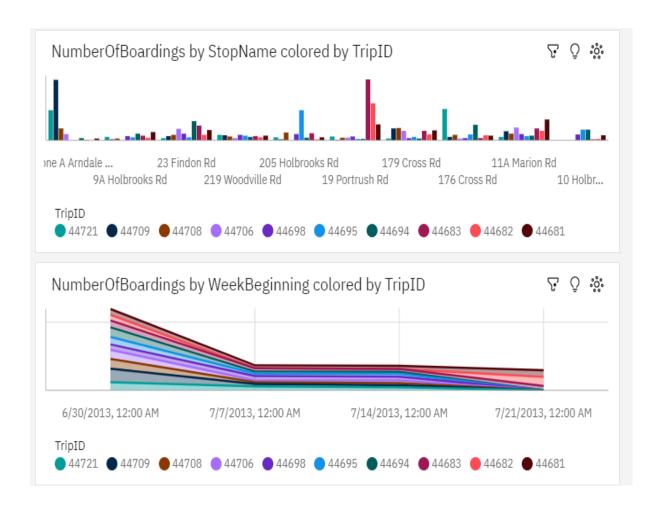
RouteID (Count distinct)



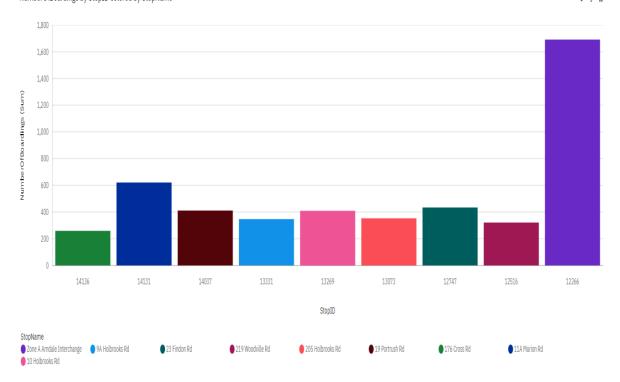
7. Customize Visualizations:

- Customize your visualizations with labels, titles, formatting, and other options.
- Apply aggregation, grouping, sorting, and filtering to the data in your visualization.









8. Interactivity:

- Add interactivity to your dashboard or report, such as drill-through actions, filters, and prompts.

9. Save and Publish:

- Save your report or dashboard to the IBM Cognos platform.
- Optionally, publish it to a location accessible to others.

Performing Data Analysis Using Code:

If you want to complement your IBM Cognos analysis with additional data analysis using code (e.g., Python or R), you can follow these steps:

1. Data Extraction:

- Export the data from IBM Cognos in a format that can be used by your code. Common formats include CSV, Excel, or connecting directly to the database from your code.

2. Data Analysis Using Code:

- Use your preferred data analysis tool (e.g., Jupyter Notebook for Python or RStudio for R) to load the data and perform further analysis. You can apply statistical analysis, machine learning, or any other data science techniques as needed.

Code:

import pandas as pd

Load data from a CSV file (assuming the data is in a file named 'data.csv')

```
data = pd.read csv('project dataset.csv')
```

Calculate punctuality rate (assuming 'NumberOfBoardings' represents successful boardings)

```
total_boardings = data['NumberOfBoardings'].sum()
total_trips = len(data['TripID'].unique())
```

```
punctuality_rate = (total_boardings / total_trips) * 100
print(f"Punctuality Rate: {punctuality_rate:.2f}%")
```

- 3. Data Visualization Using Code:
- Generate additional visualizations using code libraries like Matplotlib or Seaborn for Python, or ggplot2 for R.
- Customize and tailor the visualizations to your specific requirements.

4. Combine Results:

- Integrate the results from your code-based analysis with the insights obtained from IBM Cognos, creating a comprehensive analysis.

5. Documentation:

- Document your analysis, data sources, methodologies, and findings for future reference and sharing with others.

By following these steps, you can replicate the analysis in IBM Cognos, generate visualizations, and enhance your analysis with code-based data analysis and visualization techniques, creating a more comprehensive and data-driven understanding of your data.

Conclusion:

In conclusion, the analysis of public transportation reveals both strengths and areas for improvement. The system demonstrates efficiency and accessibility, but safety and environmental concerns persist. Affordability remains a critical factor, and maintenance and infrastructure upgrades are needed. Embracing innovation and technology could enhance the passenger experience. It is crucial for stakeholders to engage in collaborative efforts to address these challenges and shape the future of public transportation. Prioritizing safety, sustainability, and customer satisfaction while considering the evolving needs of the community is paramount for a robust and successful public transportation system.