Project Title: Public Transport Analysis

Problem Definition:

The project involves analyzing 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.

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 Holbrc	########	1
23634	100	13875	9 Marion	########	1
23634	100	13045	206 Holbr	########	1
23635	100	13335	9A Holbrc	########	1
23635	100	13383	8A Mario	########	1
23635	100	13586	8D Mario	########	2
23635	100	12726	23 Findon	########	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
23637	100	13335	9A Holbrc	########	3
23637	100	12266	Zone A Ar	########	5
23637	100	13196	13 Holbrc	########	1
23638	100	12562	218 Findo	########	1
23638	100	12266	Zone A Ar	########	3
23638	100	13875	9 Marion	########	1
23638	100	14133	11A Mari	########	1
23638	100	12472	220 Wood	########	2
23638	100	12257	25 Torren	########	4
23638	100	13618	8E Mario	########	1
23638	100	12216	224 Wood	########	2
23639	100	12266	Zone A Ar	########	6
23639	100	14170	183 Cross	########	1
23639	100	13813	8K Mario	########	1

Design Thinking:

Based on the provided information, here's a structured approach to address the project objectives:

1. Project Objectives:

On-Time Performance Assessment:

Measure and evaluate the punctuality of public transportation services, including buses, trains, and trams.
Identify patterns of delays and their causes, such as traffic congestion, maintenance issues, or operational inefficiencies.
Set benchmarks for on-time performance and develop strategies to improve reliability.

Passenger Satisfaction Analysis:

Gather and analyze passenger feedback through surveys, focus groups, or online reviews.
Assess the factors that contribute to passenger satisfaction, such as cleanliness, safety, courtesy of staff, and comfort.
Use insights to implement improvements, enhance the overall passenger experience, and retain or attract riders.

Service Efficiency Optimization:

Evaluate the efficiency of public transport services, considering factors like route optimization, vehicle utilization, and energy consumption.
Identify areas where resources can be allocated more effectively to reduce costs and environmental impact.
Implement changes to schedules, routes, and maintenance procedures to enhance service efficiency and sustainability.

2. Data collection :

Schedule Data:

Official Timetables: Obtain official schedules from transportation agencies or providers. These timetables are typically available online or in printed format and provide planned departure and arrival times for various routes.

GTFS (General Transit Feed Specification): Many transit agencies publish GTFS data, a standardized format that includes schedule information. This data can be accessed through data feeds and APIs, making it suitable for analysis and real-time updates.

Archived Data: Historical schedule data can be collected from previous years to analyze trends and performance over time. These archives can help identify long-term patterns and areas for improvement.

Real-Time Updates:

GPS and AVL Systems: Many modern public transportation vehicles are equipped with GPS (Global Positioning System) and AVL (Automatic Vehicle Location) systems. These systems provide real-time location and movement data, which can be accessed through APIs or data feeds to track vehicle positions and predict arrival times accurately.

Mobile Apps and Websites: Public transportation providers often have mobile apps and websites that offer real-time updates to passengers. Data from these sources can be scraped or accessed through official APIs to provide real-time information on vehicle locations and delays.

Crowdsourced Data: Some mobile apps and platforms collect real-time data from passengers' smartphones. Passengers willingly or unknowingly contribute data on vehicle locations and travel times, which can be aggregated for real-time updates.

Passenger Feedback:

Surveys and Questionnaires: Transportation agencies can conduct surveys and distribute questionnaires to collect feedback from passengers. These surveys may be paper-based or conducted online through the agency's website or mobile apps.

Social Media Monitoring: Monitor social media platforms like Twitter, Facebook, and Instagram for passenger feedback and complaints. Passengers often share their experiences and concerns, providing valuable insights into service quality.

Customer Service and Complaint Logs: Data can be collected from customer service centers and complaint logs. Analyzing these logs can help identify recurring issues and areas for improvement based on direct feedback from passengers.

3. Visualization Strategy:

Step 1: Data Preparation

Before creating dashboards and reports, ensure that your data is properly prepared. This includes data cleaning, transformation, and integration from various sources. Use IBM Cognos to connect to your data sources and create a data model for easy visualization.

Step 2: Define Dashboard Objectives

Clearly define the objectives of your dashboard. What insights are you trying to convey? Common objectives for public transportation analysis might include on-time performance, passenger satisfaction, and service efficiency. Consider the key performance indicators (KPIs) that matter most.

Step 3: Identify Key Metrics

Determine the key metrics and data points that will be included in your dashboard. For example, you might include metrics like on-time performance percentages, customer satisfaction scores, route efficiency data, and more.

Step 4: Design the Dashboard

Design the layout and components of your dashboard. IBM Cognos offers a range of visualization options, including charts, tables, maps, and more. Consider the following components:
KPI Cards: Display important metrics in a summarized format.
Charts and Graphs: Use bar charts, line graphs, and pie charts to visualize trends and comparisons.
Maps: Show route performance and passenger density using geographic data.
Tables: Present detailed data in tabular format.
Filters and Prompts: Allow users to interact with the data by applying filters and prompts to focus on specific aspects.

Step 5: Create Reports

Reports can provide detailed insights and are often used in conjunction with dashboards. Use IBM Cognos to generate reports that support the insights presented in your dashboards. These reports can be in the form of PDFs, Excel sheets, or other formats.

4. Code Integration:

Data Cleaning:

Automated Data Validation: Code can be used to implement data validation checks, flagging or correcting inconsistencies, missing values, and outliers in the dataset.
Standardization and Cleaning Rules: Develop code to apply standardization and cleaning rules consistently across large datasets, ensuring data uniformity.
Data Deduplication: Code can automatically identify and remove duplicate records, reducing errors and redundancy in the data.

Data Transformation:

Feature Engineering: Code can create new features or variables based on existing data, providing richer insights into passenger behavior, route efficiency, and performance metrics.
Aggregation and Summarization: Code allows for efficient aggregation and summarization of data, enabling analyses at different levels (e.g., daily, weekly, or by geographic region).
Geospatial Analysis: Use code to process and analyze geospatial data, such as calculating distances between stops, mapping routes, or identifying service gaps.

Statistical Analysis:

Hypothesis Testing: Code can automate hypothesis tests to assess the significance of factors affecting public transportation, such as weather conditions, service interruptions, or demographics.
Regression Analysis: Implement code for regression analyses to understand the relationships between variables, such as how changes in service frequency impact ridership.
Time Series Analysis: Code can facilitate time series analysis to uncover patterns, seasonality, and trends in transportation data, helping with forecasting and decision-making.

Conclusion Questions:

1.What are the most significant findings or insights from the analysis, and how do they impact the objectives set at the beginning?

2.Have the objectives of the analysis been met, and to what extent? Are there any unexpected or secondary insights that emerged during the analysis?

3.What are the practical implications of the insights for improving public transportation services or making informed decisions?

4.How might the insights obtained from this analysis inform future strategies, policies, or investments in the public transportation system?

5.Were there any limitations or challenges encountered during the analysis that should be addressed in future research or data collection efforts?

6.How can the analysis results be effectively communicated to stakeholders, decision-makers, or the general public?

7.What recommendations or action steps can be derived from the analysis, and how can they be prioritized and implemented?