PUBLIC TRANSPORT EFFICIENCY ANALYSIS

PROBLEM DEFINITION

The Public Transportation Analysis project is centered on the examination of public transportation data to evaluate service efficiency, on-time performance, and gather passenger feedback. Its core objective is to generate actionable insights that support transportation improvement initiatives and enhance the overall public transportation experience. The project involves setting clear analysis objectives, collecting and processing transportation data, creating informative visualizations in IBM Cognos, and employing data analysis techniques to inform decision-making.

This project is crucial because it addresses the pressing need for efficient, reliable, and passenger-oriented public transportation services. By leveraging data-driven approaches, it seeks to optimize service operations, improve punctuality, and cater to passenger preferences. In doing so, it not only enhances the daily lives of commuters but also contributes to sustainable urban development and resource optimization for transportation providers.

ANALYSIS OBJECTIVES

On-Time Performance:
Analyzing on-time performance involves assessing the punctuality of public transportation services. This includes evaluating how well services adhere to their schedules and identifying any patterns of delays or deviations. By understanding the on-time performance, we aim to pinpoint specific areas or routes where improvements can be made, enhancing the overall reliability and timeliness of public transportation.

Passenger Satisfaction:
To gauge passenger satisfaction, we will collect and analyze feedback from passengers. This involves understanding their experiences, preferences, and any pain points they may encounter during their journeys. By identifying areas of dissatisfaction, we can address specific concerns and enhance overall satisfaction. This could involve aspects such as cleanliness, staff behavior, and the overall comfort of the transportation services.

Service Efficiency:
Assessing service efficiency goes beyond timeliness and satisfaction, considering broader factors like route optimization and resource utilization. This objective aims to evaluate how effectively resources such as vehicles, fuel, and personnel are utilized in delivering transportation services. By optimizing routes and resource allocation, we can enhance efficiency, reduce costs, and contribute to a more sustainable and effective public transportation system.

DATA COLLECTION

Identify the sources and methods for collecting transportation data, including schedules, real - time updates, and passenger feedback.

Sources and methods:

1.Government Agencies and Open Data Portals:
Public Transit Authorities: Public transit agencies often collect data on schedules, routes, and ridership. This data can be obtained through official records and reports.
Some agencies also release their data for research and analysis for development.

2.Transportation Apps and Websites:
Ride-Hailing Apps: Companies like Uber and Lyft collect real-time data on driver locations and passenger pickups and drop-offs and they also collect passenger feedback on their services.
Transit Apps: Apps like Google Maps, Moovit, and Transit collect real-time transit data, including schedules, service disruptions, and route information.

3.GPS and Telematics:
Many vehicles, especially public transit buses and commercial trucks, are equipped with GPS and telematics systems that track their movements in real-time.

4.Traffic Cameras and CCTV Cameras:
Traffic cameras provide real-time traffic data. This includes congestion levels and accident information. In transportation hubs like airports and train stations, CCTV cameras capture real-time footage for data analysis..

5.Automatic Passenger Counters (APCs):
Public transit vehicles often use APCs to count the number of passengers getting on and off at each stop. This data helps in analyzing ridership patterns.

6.Smart Cards and Fare Collection Systems:
Smart cards like contactless transit cards or electronic toll collection systems (e.g., E-ZPass) collect data on passenger trips and toll booth usage.

7.Surveys and Questionnaires :
Conducting surveys and questionnaires among passengers can provide valuable feedback on their transportation experiences, including satisfaction, preferences, and complaints.

8.Crowdsourced Data:
Apps and platforms that allow users to report incidents, delays, or real-time updates contribute to crowdsourced transportation data. Examples include Waze for traffic updates and apps like Transit for public transit information.

9.Sensors and IoT Devices:
Sensors and Internet of Things (IoT) devices installed on transportation infrastructure, such as roads and bridges, can provide real-time data on environmental conditions, traffic flow, and infrastructure health.

Collecting transportation data is essential for improving the efficiency, safety, and user experience of transportation systems. However, it's crucial to handle and protect sensitive passenger information in compliance with privacy regulations.

VISUALISATION STRATEGY

For now, key metrics can be identified as

- On-time performance percentage
- Ridership numbers

With these key metrics, after deciding on dataset, the following steps would be incorporated iteratively

- Create Data Models
    - Define relationships between the data tables and create calculated fields within IBM Cognos Data Modules to support our metrics.
- Design Visualizations
    - Cognos Report Studio can be used to create the following visualizations:
    - For on-time performance: Line charts to show trends over time.
    - For ridership: Bar charts to compare ridership across different routes or time periods.
- Create Dashboards
    - Build a dashboard by adding visualizations:
    - Use summary tiles or KPI cards to display high-level metrics.
    - Add interactive filters for users to select specific routes or time periods.
    - Incorporate the visualizations created previously..
    - Include text boxes for comments or explanations.
- Make dashboards interactive
    - Add drill-through capabilities, interactive filters, and prompts to facilitate data exploration and analysis.

- Service Efficiency Overview
    - Develop visualizations, such as bar charts or line graphs, to represent service efficiency metrics like route utilization rates, service frequency, and passenger load. Use color coding to highlight areas that require attention and improvement.
- Advanced Analytics
    - Conduct advanced analytics and predictive modeling through integration with IBM Watson and other data science tools.
- Feedback and Iteration
    - Based on user feedback, existing mechanisms can be altered or improved.

CODE INTEGRATION

Data Collection Automation and Error Handling:
Utilize Python scripts with libraries like requests and pandas for automating data collection from diverse sources, including APIs and external data providers. Implement error-handling mechanisms in your code to manage network interruptions and data source changes.

Data Cleaning and Preprocessing:

- **Data Validation:** Write code to validate collected data to ensure it meets predefined quality and consistency criteria.
- **Missing Data Handling:** Use code to automatically impute missing values or flag records for further investigation.
- **Data Transformation Rules:** Implement code-based transformation rules, such as removing duplicates and normalizing data formats.
- **Data Transformation:** Automate data aggregation by specific time intervals or geographic regions. Use code to create derived features like average wait times and perform geospatial analysis tasks.

Statistical Analysis:
Leverage code libraries like NumPy, pandas, and SciPy for statistical analysis. Implement hypothesis testing (e.g., t-tests, ANOVA) and regression analysis to assess service efficiency and performance factors.

Data Visualization:
Create custom visualizations and interactive dashboards using libraries like Matplotlib, Seaborn, Plotly, or ggplot2. Automate visualization generation to provide stakeholders with updated data visualizations.

Predictive Modeling:
Develop machine learning models for predicting service efficiency or passenger demand using code. Automate model deployment for regular updates.

Feedback Analysis:
Utilize NLP libraries like NLTK or spaCy to analyze passenger feedback data for sentiment analysis, topic modeling, and keyword extraction. Generate automated reports summarizing common issues and sentiment trends from comments.

Integration with External Data:
Write code to integrate external data sources, such as weather or traffic data, into your analysis. Ensure real-time integration when necessary.

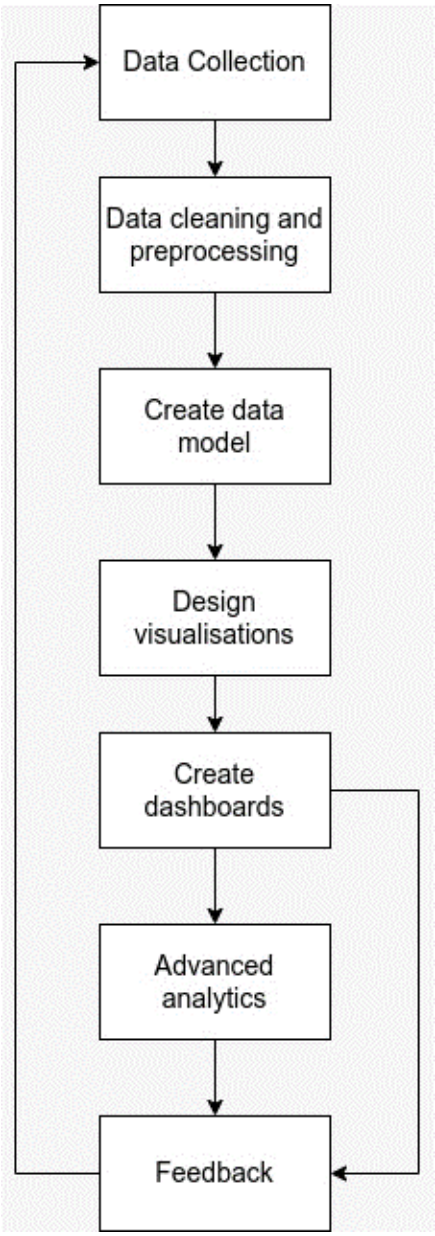
Continuous Monitoring and Alerts:
Set up scheduled code scripts to continuously monitor data sources and trigger alerts when predefined thresholds or anomalies are detected.

Decision Support System:
Develop a decision support system (e.g., web-based dashboard or desktop app) using code to provide real-time insights to transportation authorities. Automate data updates to reflect the latest analysis results.

A/B Testing:
Use code to design A/B tests, allocate test groups, and implement experiments for evaluating transportation improvement initiatives. Automate data collection and analysis for efficiency.

Feedback Loop:
Write code to automatically update analysis and recommendations based on real-time data changes, creating a feedback loop for continuous improvement.

OVERVIEW


Data Collection
Data cleaning and preprocessing
Create data model
Design visualisations
Create dashboards
Advanced analytics
Feedback