Public Transportation Analysis

Team Member

Name: Avula Devi Vara Prasad

Roll Number: 211521243027

Problem definition:

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. Some specific analysis objectives based on the provided dataset:

1. **On-Time Performance Assessment:**
    - Analyse the dataset to evaluate the on-time performance of public transportation services. This could involve calculating metrics like the percentage of trips that are on time, average delays, and distribution of delays.
2. **Passenger Satisfaction Evaluation:**
    - Use any available passenger feedback data to assess passenger satisfaction with the public transportation services. This could involve sentiment analysis on comments or surveys.
3. **Service Efficiency Analysis:**
    - Evaluate the efficiency of public transportation services. This might include examining metrics like route optimization, frequency of service, and utilization of resources.
4. **Route Analysis:**

- Analyse the dataset to determine the most popular routes, busiest times of day, and areas with high passenger demand. This can help optimize routes and schedules**.**

5. **Incident Analysis:**
    - Investigate incidents or disruptions in service. Identify the frequency, nature, and impact of incidents on the transportation system.
6. **Comparison with Benchmarks:**
    - Compare the performance metrics obtained from this dataset with industry benchmarks or standards to identify areas for improvement**.**
7. **Trend Analysis:**
    - Identify trends over time in on-time performance, passenger satisfaction, and service efficiency. This could involve looking at historical data to spot patterns.

8. **Correlation Analysis:**
    - Explore potential relationships between different aspects of public transportation (e.g., does on-time performance impact passenger satisfaction.
9. **Geospatial Analysis:**

- Utilize location data to perform geospatial analyses, such as visualizing service coverage, identifying areas with limited access, or optimizing routes.

10. **Cost-Benefit Analysis:**
    - Assess the cost-effectiveness of different routes or services in terms of both operational costs and benefits to passengers.

.

---

2. To collect transportation data for your analysis, you'll need to tap into various sources and methods. Here's a breakdown of potential sources and methods for collecting transportation data:

1. **Official Transportation Agencies:**
    - Contact local or regional transportation authorities, such as city or state transit agencies. They often provide official schedules, routes, and real-time updates for public transportation services.
2. **Open Data Portals:**
    - Many cities and municipalities provide open data portals where they share a wide range of data, including transportation data. Look for

datasets related to public transit schedules, routes, and real-time updates**.**

3. **APIs (Application Programming Interfaces):**
    - Some transit agencies provide APIs that allow developers to access real-time data for public transportation. These APIs can provide information on bus/train locations, estimated arrival times, and service disruptions.
4. **GTFS (General Transit Feed Specification) Data**:**
    - GTFS is a standardized format for public transit data that includes information about routes, stops, schedules, and more. Many transit agencies publish their data in GTFS format, making it easily accessible for analysis.
5. **Real-Time Tracking Systems:**
    - Some transportation systems have GPS tracking systems on their vehicles that provide real-time location data. This information can be used to monitor the actual movement of vehicles**.**
6. **Passenger Feedback Surveys:**
    - Conduct surveys or collect feedback from passengers to gauge their satisfaction with the

public transportation services. This can **be** done through online surveys, in-person interviews, or mobile apps.

7. **Social Media and Forums:**
   - Monitor social media platforms and online forums where passengers may discuss their experiences with public transportation. This can provide valuable qualitative insights.
8. **Third-Party Apps and Services:**
   - Utilize third-party apps or services that aggregate transportation data. Some apps provide real-time updates, user reviews, and other relevant information.
9. **Web Scraping**

   In cases where data isn't available through official channels, consider web scraping to extract relevant information from transportation websites. Be sure to check the website's terms of service and policies regarding data scraping.
10. **Historical Data Archives:**

    Some transportation agencies maintain archives of historical data, including schedules, routes, and performance metrics. This data can be valuable for trend analysis.

11. Crowdsourced Data:

Leverage platforms where users contribute data related to public transportation, such as crowd-sourced transit apps or mapping services**.**

.....

3. Visualization Strategy: Plan how to visualize the insights using IBM Cognos to create informative dashboards and reports

1. Dashboard Layout and Structure:

- **Overview Dashboard:**
  - Provide a high-level summary of key performance indicators (KPIs) related to on-time performance, passenger satisfaction, and service efficiency.
  - Include visualizations like gauges, cards, and summary charts for quick reference.
- **Detailed Performance Dashboard:**
  - Break down performance metrics by specific categories, such as routes, days of the week, and times of day.
  - Utilize interactive charts like line graphs, bar charts, and heatmaps for detailed analysis.
-
- **Feedback and Satisfaction Dashboard:**

    - Display sentiment analysis results from passenger feedback.
    - Use word clouds, sentiment score trends, and bar charts to visualize feedback sentiment.
- **Geospatial Dashboard:**
    - Utilize maps to display the coverage area of public transportation services.
    - Overlay data points like stops, routes, and passenger feedback on the map for spatial context.
- **Comparative Analysis Dashboard:**
    - Compare performance metrics across different time periods (e.g., month-to-month or year-over-year).
    - Use line charts and tables to highlight trends and variations**.**

2. Visualizations:

- Gauges and Cards:
    - Use gauges to show at-a-glance metrics like on-time percentage or passenger satisfaction scores.
    - Cards can display summary statistics for various KPIs.
- **Line Charts and Bar Charts:**

    - Represent time-series data, such as on-time performance trends over different days or months.
    - Bar charts can be used for comparisons, like route performance or passenger satisfaction ratings.
- **Heatmaps:**
    - Display patterns or density of data, such as peak travel times or popular routes, with color intensity**.**
- **Word Clouds:**
    - Visualize frequently mentioned words or sentiments from passenger feedback. This can offer a quick insight into common themes.
- **Maps and Geo-spatial Visualizations:**
    - Use maps to show the geographical distribution of stops, routes, and areas with high passenger demand**.**

3. Interactivity and Drill-Downs:

- Enable interactive elements like filters, drop-down menus, and date selectors to allow users to explore the data in more detail.

- Implement drill-down capabilities to allow users to navigate from high-level summaries to more granular data.

4. Storytelling Elements:

- Include text boxes, annotations, and narratives to provide context and guide the viewer through the insights.

5. Accessibility and Responsiveness:

- Ensure that the dashboards are accessible to a wide audience, including those with disabilities. Use alt text for images, provide clear labels, and use color schemes that are easily distinguishable**.**

6. Data Refresh Frequency:

- Set up automatic data refreshes to ensure that the dashboards display the most up-to-date information.

7. User Training and Documentation:

- Provide training and documentation for users who will interact with the dashboards to ensure they understand how to navigate and interpret the visualizations.

.....

4. Code integration can greatly enhance various aspects of your analysis. Here are specific areas where code can be beneficial:

1. **Data Cleaning and Preparation:**
    - Python or R: Use code to automate tasks like handling missing values, removing duplicates, standardizing data formats, and applying transformations.
2. **Data Integration and Aggregation:**
    - SQL: Leverage SQL queries for joining multiple datasets, performing complex aggregations, and extracting subsets of data for analysis.
3. **Statistical Analysis:**
    - Python (with libraries like pandas, NumPy, and SciPy) or R: Perform advanced statistical analyses, such as hypothesis testing, regression analysis, and clustering.
4. **Machine Learning and Predictive Modeling:**
    - Python (with libraries like scikit-learn, TensorFlow, or PyTorch) or R (with libraries like caret or xgboost): Utilize machine learning algorithms for tasks like predictive odelling, anomaly detection, or clustering.

5. **Sentiment Analysis:**
    - Python (with libraries like NLTK, spaCy, or TextBlob): Use natural language processing libraries to perform sentiment analysis on passenger feedback.
6. **Time Series Analysis:**
    - Python (with libraries like pandas, statsmodels, or Prophet) or R (with libraries like forecast or prophet): Analyze time series data for trends, seasonality, and forecasting.
7. **Geospatial Analysis:**
    - Python (with libraries like geopandas, folium, or Shapely) or R (with libraries like sf or leaflet): Perform geospatial operations like mapping, spatial joins, and distance calculations.
8. **Automated Reporting and Documentation:**
    - Python (with libraries like matplotlib, seaborn, or ReportLab) or R (with libraries like ggplot2 or RMarkdown): Automate the generation of visualizations and reports based on the latest data.
9. **Data Validation and Quality Checks:**

Python (with assert statements or custom validation functions): Use code to implement data validation checks to ensure data integrity.

10.Web Scraping (if additional data sources are needed):

Python (with libraries like BeautifulSoup or Scrapy): Extract data from websites if additional information is required for the analysis.

11.Custom Data Transformations:

If your data requires specific transformations unique to your project, writing custom code allows you to tailor the process to your needs.

12 .**Automating Data Retrieval (if applicable):**If you're working with APIs or regularly updated datasets, use code to automate the retrieval of new data.

Conclusion :

Based on the outlined project involving the analysis of public transportation data to assess service efficiency, on-time performance, and passenger feedback, the following conclusions can be drawn:

1. On-Time Performance: The analysis of the transportation data indicates that on-time performance is a critical aspect of public transportation services. Understanding and improving punctuality is crucial for enhancing the overall public transportation experience.

2. Passenger Satisfaction: The assessment of passenger feedback provides valuable insights into the areas that passengers find satisfactory and those that require improvement. Addressing passenger concerns and preferences is essential for providing a positive transportation experience.
3. Service Efficiency: The data analysis reveals opportunities for optimizing service efficiency. This includes aspects such as route optimization, frequency of service, and resource allocation. Streamlining operations can lead to cost savings and improved service delivery.
4. Key Performance Indicators (KPIs): Defined KPIs, including on-time performance metrics and passenger satisfaction scores, serve as important benchmarks for evaluating the effectiveness of transportation services. Regular monitoring of these KPIs is crucial for continuous improvement.
5. Geospatial Insights: Geospatial analysis provides a clear understanding of service coverage and areas with high passenger demand. This information is valuable for making informed decisions regarding route planning and resource allocation.
6. Trends and Patterns: The analysis of historical data reveals trends and patterns in public transportation usage. Recognizing these trends allows for proactive adjustments to schedules and resources to better meet passenger needs.
7. Incident Management: The examination of incident data highlights areas of potential disruption in service. Developing effective incident management strategies is essential for minimizing service disruptions and ensuring passenger satisfaction.
8. Cost-Benefit Analysis: Evaluating the cost-effectiveness of different routes and services is crucial for optimizing resource allocation and achieving a balance between operational costs and passenger benefits.

9. Recommendations for Improvement: Based on the analysis, several recommendations can be made to enhance the public transportation experience. These may include adjustments to schedules, route optimization, and strategies for addressing passenger concerns.

10. Continuous Monitoring and Iteration: It is imperative to establish a system for ongoing monitoring of performance metrics and passenger feedback. This allows for timely adjustments and ensures that improvements are sustained over time.