Project Title : Public Transportation Efficiency 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.

23647	100	13331	9A Holbro	########	3
23647	100	14083	165 Cross	########	1
23647	100	12747	23 Findon	########	1
23647	100	13499	8C Marior	########	2
23647	100	14135	175 Cross	########	1
23647	100	14152	179 Cross	########	1
23647	100	14129	174 Cross	########	1
23648	100	13331	9A Holbro	########	1
23648	100	13931	9B Marior	########	1
23648	100	13269	10 Holbro	########	1
23648	100	13817	8K Marior	########	1
23648	100	14155	180 Cross	########	1
23649	100	12939	18 Critten	########	1
23649	100	14152	179 Cross	########	1
23649	100	13874	9 Marion	########	1
23649	100	14142	176 Cross	########	1
23649	100	14131	11A Maric	########	2
23649	100	14000	10 Marior	########	1
23649	100	14047	10A Maric	########	2
23649	100	13730	8G Marior	########	1
23650	100	13499	8C Marior	########	1
23650	100	13905	9A Marior	########	1
23650	100	14047	10A Maric	########	3
23650	100	14131	11A Maric	########	1
23650	100	13931	9B Marior	########	1
23651	100	12806	21 Critten	########	1
23651	100	13269	10 Holbro	########	1

Design Thinking:

1.Project Objectives:

Objective 1: Assessing On-time Performance

- Develop a machine learning model to predict the estimated arrival time of government buses.
- Achieve a high level of prediction accuracy and reliability to improve on-time performance.

Objective 2: Customer Experience and Satisfaction Improvement

- Passenger Feedback Analysis: Collect and analyze customer feedback to identify pain points and areas for improvement in service quality.
- Technology Integration: Assess the integration of modern technology, such as mobile apps and real-time tracking, to enhance the overall passenger experience and convenience.

Objective 3 : Cost Reduction and Financial Sustainability

- Fare Structure Assessment: Analyze fare pricing and payment systems to ensure they cover operational costs while remaining affordable for passengers.
- Energy Efficiency: Evaluate the environmental impact and energy consumption of public transportation modes, aiming for cost-effective and sustainable alternatives.

Analysis Approach:

Step 1: Data Collection and Integration

Data Source Identification:

Begin by identifying and collecting data from various sources. These sources may include GPS tracking systems on buses and trains, fare collection systems, passenger surveys, traffic cameras, weather reports, and more. It's crucial to pinpoint the relevant data sources that provide information about routes, schedules, ridership, and operational metrics.

Real-Time and Historical Data:

Public transportation analysis benefits from both real-time and historical data. Real-time data provides immediate insights into the current state of the system, such as bus locations or delays. Historical data, on the other hand, helps identify trends and patterns over time, allowing for long-term planning and performance evaluation. Both types of data should be integrated to provide a comprehensive view.

Geospatial Integration:

Geographic information plays a crucial role in public transportation analysis. Geospatial data, such as GIS (Geographic Information System) data and mapping information, allows you to analyze routes, stops, and spatial relationships. Integrating geospatial data with other datasets enables the visualization of transportation networks and helps in optimizing routes and schedules.

Collecting User Feedback:

Collecting passenger feedback for public transportation analysis is essential for improving the quality of services and making data-driven decisions. Here are some steps and methods you can use to collect passenger feedback:

a. **Online Surveys:**

- We can Create online surveys using platforms like SurveyMonkey, Google Forms, or custom-built survey tools.
- Share the survey links through social media, the transportation agency's website, or email newsletters.
- Ask passengers about their overall satisfaction, specific issues, suggestions for improvements, and demographic information.

b. Mobile Apps:

- We Can Develop a dedicated mobile app for passengers to provide feedback easily.
- In that app we can Include features like rating rides, reporting problems, and leaving comments.
- Use push notifications to encourage users to provide feedback after using the service.

c. In-App Feedback:

- If the transportation system has a mobile ticketing or tracking app, integrate a feedback feature directly into the app.
- We can Allow passengers to report issues, submit complaints, or offer suggestions through the app.

d. Social Media Monitoring:

- We can Monitor social media platforms for mentions, comments, and messages related to public transportation.
- Engage with passengers and address their concerns promptly.

e. Customer Service Centers:

- We can Set up customer service centers at key transportation hubs where passengers can provide feedback in person or via forms.
- Ensure staff are trained to handle feedback professionally.

f. Onboard Surveys:

- We can Conduct onboard surveys using paper forms or electronic devices.
- We can Ask passengers to rate their experience and provide comments during their journey.

g. **Focus Groups:**

- We can Organize focus group discussions with passengers to dive deeper into specific issues and gather qualitative insights.
- We can Use these sessions to brainstorm solutions and improvements.

h. Comment Cards:

- We can Place comment cards or suggestion boxes at transportation stops, stations, and vehicles.
- Encourage passengers to drop their comments or suggestions anonymously.

i. Email and Text Surveys:

- We can Send periodic email or text surveys to passengers who have registered with the transportation agency.
- Keep the surveys short and focused on key issues.

j. 12. Data from Complaints and Incident Reports:

- We can Analyze data from formal complaints and incident reports to identify recurring issues.
- We can Use this information to address systemic problems.

k. 14. Regular Feedback Cycles:

- We can Implement a regular feedback cycle and communicate to passengers how their input has led to improvements.
- This can incentivize ongoing participation.

3. **Visualization Selection:**

a. Exploratory Data Analysis:

- In this section, focus on visualizations that help you understand the basic structure of your dataset. Use histograms and box plots to explore the distribution of numerical variables, and bar charts or pie charts for categorical variables. Scatter plots can reveal relationships between pairs of variables. This initial analysis is crucial for gaining insights into the dataset's characteristics and potential patterns.

b. Temporal Analysis:

- If your dataset contains temporal data, consider visualizations like time series plots or calendar heatmaps. Time series plots are excellent for displaying trends and seasonality, while calendar heatmaps can highlight patterns over days, months, or years. These visualizations are essential for understanding how your data evolves over time and identifying long-term trends or periodic fluctuations.

c. Spatial Analysis:

- For datasets with geographic components, choropleth maps or heatmaps can be powerful tools. Choropleth maps use color-coding to represent data values in different geographic regions, offering insights into regional variations. Heatmaps can display concentrations of data points in specific geographic areas, helping you identify hotspots or trends related to location.

d. Correlation and Relationships:

- When exploring relationships between multiple variables, scatter plots matrix, correlation matrices, or network graphs can be valuable. Scatter plots matrix displays scatter plots for pairs of variables, allowing you to quickly identify correlations. Correlation matrices provide a numerical overview of relationships, while network graphs visually represent connections and dependencies between variables or entities in your dataset.

e. Outlier Detection:

- Detecting outliers is crucial in large datasets. Box plots, scatter plots, and violin plots are effective for identifying outliers in numerical data. Heatmaps can be used to visualize anomalies in multivariate datasets. Robust visualization techniques are essential for understanding the impact of outliers on your analyses and ensuring the integrity of your results.

f. Interactive Visualizations and Dashboards:

- Consider incorporating interactive visualizations and dashboards into your project. Tools like Tableau, Power BI, or D3.js allow users to interact with the data dynamically. Interactive charts, filters, and drill-down options enable users to explore specific aspects of the dataset, making the findings more accessible and engaging. Interactive visualizations can enhance user experience and facilitate a deeper understanding of complex datasets.

g. Comparative Analysis:

- When your project involves comparing different groups or categories, stacked bar charts, grouped bar charts, or parallel coordinates plots can be helpful. These visualizations allow you to compare multiple variables across different categories simultaneously, providing insights into how different factors interact and influence one another. Comparative analysis visualizations are vital for drawing meaningful conclusions about group differences and similarities within the dataset.

4.**Code Integration**

1. Data Cleaning:

Data cleaning involves handling missing values, removing duplicates, correcting inconsistencies, and ensuring data integrity. Python offers powerful libraries like Pandas for data manipulation and cleaning.

Example: Handling Missing Values with Pandas

```
import pandas as pd
# Load your dataset into a Pandas DataFrame
data = pd.read_csv('your_dataset.csv')
# Handling missing values by filling NaN values with mean of the column
data.fillna(data.mean(), inplace=True)
# Removing duplicates
data.drop_duplicates(inplace=True)
# Correcting inconsistencies - Example: converting text to lowercase
data['column_name'] = data['column_name'].str.lower()
# Save the cleaned data back to a CSV file
data.to_csv('cleaned_data.csv', index=False)
```

2. Data Transformation:

Data transformation involves converting data into a suitable format for analysis. This could include normalizing data, encoding categorical variables, or creating new features from existing ones.

Example: Encoding Categorical Variables with Pandas

```
# Assuming 'category_column' is a categorical column in your DataFrame
encoded_data = pd.get_dummies(data, columns=['category_column'])
# Normalizing numerical data (scaling to a range)
```

```
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
data['normalized_column'] = scaler.fit_transform(data[['numerical_column']])
```

3. Statistical Analysis:

Statistical analysis involves deriving meaningful insights from the data, such as calculating summary statistics, correlation, hypothesis testing, etc.

Example: Calculating Summary Statistics and Correlation with Pandas

```
# Summary statistics
summary_stats = data.describe()
# Correlation matrix
correlation_matrix = data.corr()
# Hypothesis testing (Example: t-test for two independent samples)
from scipy.stats import ttest_ind
group1 = data[data['group_column'] == 'group1']['value_column']
group2 = data[data['group_column'] == 'group2']['value_column']
t_stat, p_value = ttest_ind(group1, group2)
```

5.**Conclusion Questions:**

1. What is the average time taken to complete a route, and how does it vary during different times of the day?

2. What percentage of buses/trains operate on time, and how does this performance vary across different routes and days of the week?

3. What are the most common positive aspects mentioned in passenger feedback, and how can these be reinforced or expanded upon?

4. Which routes experience the highest passenger demand, and are they adequately served in terms of frequency and capacity?

5. Are there specific locations or routes with higher reported incidents, requiring increased security measures?

6. Are there trends in fuel consumption or maintenance costs that can be addressed to improve cost-efficiency?

7. How accessible is public transportation for people with disabilities, and what improvements can be made to ensure inclusivity?

8. What is the carbon footprint of public transportation, and how can initiatives like electric or hybrid vehicles be integrated to reduce environmental impact?

9. What is the impact of maintenance work or road closures on transportation schedules, and how can alternative routes be optimized during such events?

10. Based on current trends and feedback, what are the long-term recommendations for infrastructure development, route expansion, or technological integration to enhance the overall public transportation experience?