**Public Transport Analysis Project**

**Phase 1**

**Phase 2: Innovation - Transforming the Design**

**Step 1: Data Preparation**

Objective: Prepare the existing dataset for in-depth analysis.

1. Data Quality Assessment: Before diving into analysis, it's essential to review the dataset for potential issues. This step involves identifying and documenting issues, such as missing values, outliers, or inconsistencies. Understanding the dataset's quality is crucial for the reliability of any analysis.

2. Data Cleaning: Address any issues identified during the quality assessment by cleaning the dataset. Common tasks may include handling missing values, removing duplicates, and correcting errors. A clean dataset is fundamental for meaningful analysis.

3. Data Transformation: Prepare the dataset for analysis. This step may involve aggregating data, creating derived variables, or any other necessary transformations to make the data suitable for the planned analysis. Transformation ensures that the data aligns with the objectives of your analysis.

**Step 2: Exploratory Data Analysis (EDA)**

Objective: Gain a deep understanding of the dataset through exploratory analysis.

1. Descriptive Statistics: Calculate basic statistics for key variables. This includes measures like means, medians, standard deviations, and ranges. Descriptive statistics provide an initial overview of the data's characteristics.

2. Data Visualization: Create a variety of visualizations to explore the data. Examples include histograms to examine distributions, scatter plots to understand relationships between variables, and box plots to identify outliers. Visualization aids in understanding the data's patterns.

3. Correlation Analysis: Investigate correlations between variables to uncover how different factors within the dataset relate to each other. Understanding these relationships can be crucial in identifying key drivers or trends.

**Step 3: Advanced Analytics**

Objective: Apply advanced analytical techniques to extract insights from the dataset.

1. Time-Series Analysis: If your dataset contains time-related information, conduct time-series analysis. This technique helps identify trends and patterns in on-time performance, passenger satisfaction, and service efficiency over time. For example, you can uncover seasonality or long-term trends.

2. Cluster Analysis: Group similar routes, regions, or other elements based on specific criteria. Cluster analysis can reveal patterns in service efficiency or passenger satisfaction. For instance, it might help you identify routes that share common characteristics.

3. Regression Analysis: Use regression analysis to explore the relationships between variables and understand how different factors influence on-time performance or passenger satisfaction. Regression models can help quantify the impact of various factors.

**Step 4: Data Visualization**

Objective: Visualize the findings from your analysis.

1. Create Informative Visualizations: Utilize appropriate visualization techniques to represent your analysis results. For example, line charts can help illustrate time-series trends, heatmaps can display relationships among variables, and scatter plots can highlight correlations.

2. Interactive Dashboards: If the data lends itself to this format, consider building interactive dashboards. These dashboards allow stakeholders to explore the data and insights interactively. They provide a dynamic way to present findings.

**Step 5: Reporting and Communication**

Objective: Summarize your analysis and findings for stakeholders.

1. Generate Detailed Reports: Create comprehensive reports that document the analysis process, highlight key insights, and provide actionable recommendations if deemed necessary. Reports should be clear, well-organized, and informative.

2. Presentation: Communicate the results and findings with stakeholders effectively. This may involve presenting the information through reports, presentations, or meetings. Ensure that stakeholders have a clear understanding of the insights derived from the analysis.

**Conclusion:**

The Phase 2 plan for advanced data analysis focuses on preparing and analyzing the existing dataset to gain a deeper understanding of public transportation efficiency, on-time performance, and passenger satisfaction. By addressing data quality, conducting exploratory analysis, applying advanced analytics, and visualizing the results, the project aims to provide valuable insights without the primary goal of active improvement. These insights can serve as a foundation for future decision-making and a more data-informed approach to public transportation management.

Team Members:

Joel Ananth A

Karthi M

Monish R

Neelanand R