**Data Engineering Task**

**Level 1: Data Exploration and Basic Operations**

Task 1.1: Load and Inspect Data

Load the dataset and display the first 10 rows.

Understand the basic structure of the data, including data types and missing values.

Task 1.2: Basic Statistics

Calculate basic statistics such as the number of trains, unique source stations, and unique destination stations.

Find the most common source and destination stations.

Task 1.3: Data Cleaning

Identify and handle missing values in the dataset.

Standardize the format of station names (e.g., make all names uppercase).

**Level 2: Data Transformation and Aggregation**

Task 2.1: Data Filtering

Filter the data to show trains that operate on specific days (e.g., only on Saturdays).

Extract and create a new dataframe for trains that start from a specific station.

Task 2.2: Grouping and Aggregation

Group the data by source station and count the number of trains originating from each station.

Find the average number of trains per day for each source station.

Task 2.3: Data Enrichment

Add a new column that categorizes trains based on their operating days (e.g., 'Weekday', 'Weekend').

**Level 3: Advanced Data Analysis**

Task 3.1: Pattern Analysis

Analyze the distribution of train journeys throughout the week. Visualize the results using bar plots or histograms.

Identify patterns or trends in the train operations based on source and destination stations.

Task 3.2: Correlation and Insights

Explore if there is any correlation between the number of trains and specific days of the week.

Provide insights and recommendations based on the analysis.

**Level 4: Data Visualization and Reporting**

Task 4.1: Visualization

Create visualizations (bar charts, line charts, heatmaps) to represent various aspects of the data (e.g., number of trains per station, day-wise distribution).

Use visualization tools like Matplotlib, Seaborn, or Plotly.

Task 4.2: Reporting

Compile a comprehensive report that includes all the findings, visualizations, and insights gained from the data.

Present the report in a clear and concise manner, suitable for stakeholders.