Project: UK-Railway

Phase 1: Data Infrastructure and Preprocessing

- **Tasks:**
  - Build a data warehouse, data model, clean and preprocess the data.
  - **Tools:** SQL, Python (pandas, numpy).
- **Deliverables:**
  - Data warehouse.
  - Cleaned dataset ready for analysis.
  - Data preprocessing notebook.

Phase 2: Analysis Questions

- **Tasks:**
  - **Determine Data Analysis Questions:** Determine all possible analysis questions that can be deducted from the dataset and would be of interest to the organization's decision makers.
  - **Tools:** SQL, Python (pandas, numpy).
- **Deliverables:**
  - A set of analysis questions that can be answered via the dataset.

Phase 3: Forecasting Questions

- **Tasks:**
  - Determine a set of forecasting questions and answer them using the trends found in the given dataset.
  - Such forecasting questions must include the prediction of the number of rides for the next month. Then, accordingly, it highlights the forecasted revenue during each day of the next month. Also, you need to specify the demand for different ticket classes.
  - **Tools:** Python (scikit-learn, pandas, Matplotlib).
- **Deliverables:**
  - Visualization plots answering forecasting questions.

Phase 4: Visualization Dashboard and Final Presentation

- **Tasks:**
  - **Build a Visualization Dashboard:** Build a PowerBI visualization dashboard that visualizes the answers to all answered questions.
  - **Final Presentation:** Prepare a report and presentation summarizing the project work, including data analysis, model development, and deployment.
  - **Tools:** SQL, Python (pandas, Matplotlib), PowerBI.
- **Deliverables:**
  - Visualization dashboard
  - Final report and presentation