

UK Train Rides

Project Team :

1. Aya Ahmed Abdelfattah
2. George Gerges Georgy
3. Omnya mamdouh Ibrahim
4. Ahmed Emadeddin Zaghloul
5. Esraa Ahmed

Project Proposal (UK Train Rides)

Overview

This project aims to analyze train journey data in the United Kingdom (National Rail) for the period from January to April 2024. The dataset includes information on departure and arrival stations, ticket prices, journey times, and ticket categories. The goal is to derive important insights that help understand passenger behavior and improve services and operations.

Objectives

1. Identify the most popular routes: Determine which stations and lines have the highest passenger density.
2. Analyze revenue by ticket type: Understand how income is distributed across various ticket types and classes.
3. Measure on-time performance: Evaluate the extent to which journeys adhere to their scheduled times.
4. Forecast demand trends: Use predictive models to estimate future demand for train journeys.

Scope

- Data Cleaning & Preprocessing
- Exploratory Analysis & Analysis Questions
- Predictive Modeling
- Interactive Dashboard Development
- Recommendations & Final Presentation

Project Plan

Timeline

The project can be organized over 4 weeks (adjustable as needed):

1. Week 1: Data Cleaning & Preprocessing
2. Week 2: Developing Analytical Questions & Exploratory Analysis
3. Week 3: Building Predictive Models & Validating Their Accuracy
4. Week 4: Creating the Interactive Dashboard & Final Presentation

Milestones

- End of Week 1: Have a clean, analysis-ready dataset.
- End of Week 2: Identify main analytical questions and prepare initial reports.
- End of Week 3: Complete predictive models and test their accuracy.
- End of Week 4: Finalize the interactive dashboard and deliver the final presentation.

Deliverables

1. Clean Dataset
2. Analysis Reports
3. Forecasting Models
4. Interactive Dashboard
5. Final Presentation

Timeline

Week 1: Data Cleaning & Preprocessing

Tasks:

- Build and structure the data model.
- Clean and preprocess the data using SQL and Python (pandas, Matplotlib).

Proposed Distribution:

- Esraa Ahmed: Clean the data using SQL.
- George Gerges Georgy: Clean the data using Python (pandas).
- Ahmed Emadeddin Zaghoul: Implement data processing code and ensure data quality.

Deliverables:

- A clean, well-prepared dataset ready for analysis.
- Data processing scripts (e.g., Jupyter Notebook or SQL script).

Week 2: Data Analysis (Analysis Questions)

Tasks:

- Identify key analysis questions based on the dataset.
- Extract patterns and trends using SQL and Python.

Proposed Distribution:

- Omnya Mamdouh Ibrahim: Define the analytical questions and required reports.
- Aya Ahmed Abdelfattah: Perform data analysis with SQL to answer specific questions.
- Ahmed Emadeddin Zaghoul: Perform data analysis with Python (pandas, Matplotlib) and create initial visualizations.

Deliverables:

- A list of analytical questions that can be answered with the data.
- Preliminary reports and findings.

Week 3: Forecasting

Tasks:

- Determine forecasting questions (e.g., future ridership, revenue).
- Build a predictive model using Python (scikit-learn).

- Generate forecasts (e.g., number of future rides, projected revenue).

Proposed Distribution:

- George Gerges Georgy: Develop the predictive model using scikit-learn.
- Ahmed Emadeddin Zaghloul: Create visualizations to present the forecasting results.

Deliverables:

- Reports on future trends and forecasts.
- Charts and graphs illustrating the predictions.

Week 4: Final Visualization & Presentation

Tasks:

- Develop an interactive dashboard using Tableau.
- Prepare the final report and presentation.

Proposed Distribution:

- Omnya Mamdouh Ibrahim: Draft the final report and presentation.
- Esraa Ahmed: Design the Tableau dashboard.
- Aya Ahmed Abdelfattah: Test the dashboard and suggest improvements.

Deliverables:

- An interactive dashboard showcasing the analysis and forecasts.
- A final report and presentation ready for submission and discussion.

Risk Assessment & Mitigation Plan

Potential Risks

1. Data inconsistencies or missing values
 - Impact: May lead to inaccurate analyses.
 - Mitigation: Use data validation tools and techniques to handle missing values.
2. Time constraints versus data volume
 - Impact: Could delay project completion.
 - Mitigation: Organize tasks clearly and hold weekly progress meetings.
3. Inaccurate predictive models
 - Impact: Weak forecasting results that undermine the project's credibility.
 - Mitigation: Apply cross-validation and hyper parameter tuning to improve model accuracy.
4. Communication or data-sharing issues
 - Impact: Work disruptions and delays in exchanging files.
 - Mitigation: Use reliable collaboration platforms (e.g., GitHub, Google Drive) and set clear communication protocols.

Key Performance Indicators (KPIs)

1. Forecast Accuracy
 - Measured via metrics such as MAE, RMSE, or MAPE to assess how closely predictions match actual values.
2. Query Response Time
 - The time it takes to execute queries or analyses in the dashboard or database.
3. User Adoption Rate
 - The percentage of users or decision-makers who regularly rely on the reports and dashboard.
4. On-time Performance
 - Measured as the percentage of journeys that arrive on schedule out of the total journeys; improvements can be tracked after data-driven decisions are implemented.
5. Daily/Weekly Revenue
 - Tracking changes in ticket revenue over time and after applying new strategies.

Conclusion

This proposal provides a clear framework for managing and executing the UK Train Rides project, starting with data collection and cleaning, moving through analysis and forecasting, and culminating in the development of an interactive dashboard. The project will enable stakeholders to better understand train usage patterns, improve services, increase customer satisfaction, and enhance operational efficiency.