

# UK Train Rides Technical Report.

---

## Prepared by:

1- Youssef Ahmed

2- Yomna Saad Abdelazeem

## Tools Used:

Python (Pandas, NumPy, Seaborn, Matplotlib), SQL, Tableau.

## Dataset:

railway.csv

## Abstract

This report presents an in-depth analysis of a UK railway dataset with the aim of identifying delay causes, understanding travel behavior, and forecasting demand. A structured data science pipeline was implemented, including data cleaning, exploratory analysis, modeling, and visualization. Forecasts of ride volume and revenue by ticket class were also provided.

## 1. Introduction

The performance of train services in the UK is a key factor influencing commuter satisfaction and operational efficiency. This project analyzes historical ride data to uncover patterns in delay incidents and forecast future demand, which can support strategic decision-making.

## 2. Methodology

The project followed a four-week pipeline:

- Week 1: Data cleaning and model structuring.
- Week 2: Formulation of key analytical questions.
- Week 3: Forecasting future demand and revenue.
- Week 4: Dashboard creation and report generation.

The tools used include Python for data processing and modeling, and Tableau for interactive dashboard visualization.

### 3. Data Preprocessing

#### 3.1 Dataset Overview

- The dataset includes information on ticket purchase time, journey dates, departure/arrival times, delays, and ticket types.

#### 3.2 Cleaning Steps

- Removed or imputed missing values in columns like Railcard and Reason\_For\_Delay.
- Standardized and renamed columns.
- Converted string-based times and dates into datetime objects.
- Derived key features: journey\_duration, actual\_journey\_duration, and delay\_time.

### 4. Exploratory Data Analysis (EDA)

#### 4.1 Delay Analysis

- Most common delay reasons were identified (e.g., signal failures, weather).
- Delays were more frequent during peak hours and specific weekdays.

#### 4.2 Customer Behavior

- Different ticket types exhibited varying delay exposures.
- Time-of-day analysis revealed trends in ticket purchases.

#### 4.3 Visualizations

- Bar charts, heatmaps, and boxplots illustrated delay patterns and demand fluctuations.

### 5. Results and Discussion

The models predicted a steady increase in demand, particularly on weekdays.

- Revenue forecasts showed higher income from first-class tickets, especially during morning hours.
- Identified actionable insights to reduce delays and optimize scheduling.

## 6. Conclusion

This project successfully processed and analyzed the UK railway dataset, providing insights into delay trends and forecasting future ride volume. These findings can support more efficient train scheduling, targeted marketing, and operational planning.

## 7. References

- Python Documentation: pandas, matplotlib, seaborn
- Tableau Public: <https://public.tableau.com>