

UK Train Rides Project

- Project Description: This project focuses on analyzing train ticket sales data in the United Kingdom to understand passenger behavior and travel patterns. By exploring ticket types, classes, prices, and delays, the project aims to uncover insights that can support better pricing decisions, improve service quality, and increase overall revenue.
- Group Members & Roles:
 - Marina Youssef:
 - Cleaning: Python
 - Analysis Questions: SQL
 - Forecasting Questions: Python
 - Visualization: Power BI
 - Menna-Allah Mahmoud:
 - Cleaning: Power Query
 - Analysis Questions: Python
 - Forecasting Questions: Python
 - Visualization: Power BI
 - Helana Hany:
 - Cleaning: Python
 - Analysis Questions: Power Query
 - Forecasting Questions: Python
 - Visualization: Power BI
 - Salma Hazem:
 - Cleaning: Python
 - Analysis Questions: SQL
 - Forecasting Questions: Python
 - Visualization: Power BI
 - Mariam Ahmed:
 - Cleaning: Power Query
 - Analysis Questions: Python
 - Forecasting Questions: Python
 - Visualization: Power BI
- Team Leader: Marina Youssef
- Objectives: The main purpose of this project is to **analyze UK train ticket sales data** in order to:
 1. Understand passenger behavior and usage patterns across different categories (including accessibility for disabled travelers).
 2. Identify which classes and ticket types (standard, first class) are most popular and most profitable.
 3. Observe temporal trends and seasonality in ticket sales, including shifts between 2023 and 2024.
 4. Evaluate the booking channels (online vs station) and how their usage evolves over time.
 5. Assess service performance (delays) and understand root causes.
 6. Benchmark stations by performance, highlighting strengths and weaknesses to support resource allocation and operational planning.

7. Provide actionable insights for pricing policy, service improvement, and inclusive access strategies.

- Tools & Technologies:
Cleaning & Preprocessing: Python, Power Query, Power BI.
Analysis and questioning: SQL, Python, Power Query
Forecasting questions: Python
Visualization Dashboard and Final Presentation: Power BI
- Milestones & Deadlines:
 - Week 1 → Data Understanding: Download dataset, explore columns, and define project scope.
 - Week 2 → Data Cleaning & Preparation: Handle missing values, fix data types, and clean the dataset using Power Query/Python.
 - Week 3 → Data Analysis (EDA): Perform exploratory analysis to identify patterns and insights.
 - Week 4 → Dashboard Design: Build interactive visuals and KPIs in Power BI.
 - Week 5 → Final Report & Presentation: Finalize documentation, insights, and present findings.
- KPIs (Key Performance Indicators):

1. Passenger Demographics & Access

- Percentage of ticket users by category (Adult / Senior / Disabled)
- Number & percentage of disability-friendly ticket options available
- Ticket prices for disabled passengers

2. Ticket Sales & Revenue

- Sales volume by class (First class vs Standard)

3. Booking Channel Usage

- Percentage of tickets booked online vs at station
- Monthly or quarterly change in online vs station bookings (2023 → 2024)

4. Service Performance & Reliability

- Percentage of trips delayed
- Delay reasons distribution (e.g. infrastructure, weather, operational)

5. Temporal & Seasonal Trends

- Peak seasons or months with highest ticket sales
- Year-over-year growth or decline in bookings

6. Station-level Performance

- Highest-performing stations (by ticket volume, revenue)
- Lowest-performing stations