UK Railway Analysis Documentation

Project Overview

This document provides a full technical and analytical overview of the UK Railway dataset analysis and Power BI dashboard. It includes the project plan, data schema, KPIs, and workflow used to build the final analytical solution.

1. Detailed Project Plan

This section outlines the phases the project went through from its inception to the final delivery of the Power BI dashboard.

Phase	Detailed Description of the Phase	Key Deliverables	Responsible
1. Data Collection and Understanding	Identifying the required data sources (booking data, operational data, capacity data). Receiving the initial data file and understanding its content and structure.	Initial Data File. Data Understanding Document.	Mohamed Amin, Mohamed Gado, Zainab Elsaid, Shimaa Mohammed
2. Data Cleaning and Preprocessing	Handling missing values, standardizing date and time formats, removing duplicates, and transforming data types to ensure quality. The	lean and ready-for-analysis.	Mohamed Amin, Mohamed Gado, Zainab Elsaid, Shimaa Mohammed
3. Data Analysis and KPI Definition	Analyzing the data to identify key patterns and defining the main KPIs related to performance, occupancy, and passenger activity.	List of Key Performance Indicators (KPIs). Exploratory Analysis Results.	Mohamed Amin, Zainab Elsaid, Shimaa Mohammed
4. Power BI Dashboard Development	Importing the clean data into Power BI. Creating relationships between tables, Developing the necessary measures, Designing and developing the interactive dashboard.	Final dashboards with Power Bi file.	Shimaa Mohammed, Mohamed Amin
	Compiling all project artifacts, analysis results, KPI definitions,		

5.Documentation	and the project plan into a comprehensive documentation file.	Final Documentation File, and Presentation.	Zainab Elsaid

3- Key Performance Indicators (KPIs)

The Key Performance Indicators (KPIs) are organized into main categories to ensure comprehensive coverage of the project's various aspects.

Category	Key Performance Indicator (KPI)	Description	Measurement (Based on Data)
Operational Performance	**Punctuality Rate**	The percentage of journeys that arrive on time (0 minutes delay).	Number of journeys with `Overall_Status` = 'On Time' divided by the total number of journeys.
	Average Delay	The average delay in minutes for all delayed journeys.	Average value of the `Delay_Mins` column for delayed journeys.
	Main Delay Causes	Identifying the most common reasons for delays.	Analysis of the frequency of values in the `Status_Reason` column.
Capacity and Occupancy Efficiency	**Occupancy Rate**	The average percentage of seats occupied compared to the total train capacity.	Average of (`Actual_Passengers` / `Train_Capacity`).
	Fully Booked Trains Ratio	The percentage of journeys that were fully booked.	Number of journeys with `Occupancy_Level` = 'Fully Booked' divided by the total number of journeys.
Sales and Booking	**Total Seats Booked**	The total number of seats booked during the data's time period.	Sum of the `Seats_Booked_Total` column.
	Channel Booking Ratio	Distribution of bookings by purchase channel (Online, Station).	Analysis of the distribution of values in the `Purchase_Channel` column.
	Railcard Usage	The percentage of bookings that utilized a Railcard discount.	Number of journeys with `Railcard_Used` = 'Yes' divided by the total number of journeys.

4- Data Schema:

This table describes the columns in the data file, clarifying the data type, a sample value, and a detailed description for each column.

Column Name	Data Type	Sample Value	Detailed Description
Transaction_ID	Object	da8a6ba8-b3dc-4677-b176	Unique identifier for each transaction or ticket.
Purchase_Date	Object	12/8/2023	Date of ticket purchase or booking.
Purchase_Time	Object	12:41:11 PM	Time of ticket purchase or booking.
Purchase_Channel	Object	Online	Channel of purchase (e.g., Online, or Station).
Payment_Method	Object	Contactless	Payment method used (e.g., Contactless, Credit Card).
Ticket_Class	Object	Standard	Ticket class (e.g., Standard, or First).
Ticket_Type	Object	Advance	Ticket type (e.g., Advance, or Off-Peak).
Passenger_Type	Object	Adult	Passenger type (e.g., Adult, Disabled, or with Railcard).
Railcard_Used	Object	Yes	Indicates whether a Railcard discount was used (Yes/No).
Booking_Category	Object	Advance Booking	Booking category (e.g., Advance Booking).
Departure_Station	Object	London Paddington	Scheduled departure station.
Arrival_Station	Object	Liverpool Lime Street	Scheduled arrival station.
Scheduled_Travel_Date	Object	1/1/2024	Scheduled travel date.
Scheduled_Departure_Time	Object	11:00:00 AM	Scheduled departure time.
Scheduled_Arrival_Time	Object	1:30:00 PM	Scheduled arrival time.
Actual_Arrival_Time	Object	1:30 PM	Actual train arrival time.
Scheduled_Duration_Mins	int64	150	Scheduled journey duration in minutes.
Delay_Mins	float64	0.0	Actual delay upon arrival in minutes (zero if on time).

Overall_Status	Object	`On Time`	Overall journey status (e.g., `On Time`, or `Delayed`).
Status_Reason	Object	`Arrived On Time`	Detailed reason for the status (e.g., Arrived On Time, Signal Failure).
Seats_Booked_Total	int64	`43`	Total number of seats booked in this transaction.
Seats_Booked_Discount	float64	`8.6`	Number of seats booked using a discount (e.g., Railcard).
Seats_Booked_Full_Fare	float64	`34.4`	Number of seats booked at full fare.
Occupancy_Level	Object	`Partially Booked`	Train occupancy level (e.g., `Partially Booked`, `Fully Booked`).
Train_Capacity	int64	`200`	Total capacity of the train/service.
Actual_Passengers	int64	`188`	Actual number of passengers on the train.
Is_Cancelled	Object	`No`	Indicates whether the journey was cancelled (Yes/No).