Digital Pioneers initiative

Project idea discussion

Team Names

Mohamed Amin

Mohamed Gado

Zainab Hedaya

Shimaa Mohamed

Kariem Elghandor

Reham Elsaid

Our Project Idea is..

UK Train Rides

and Why!

- Apply analytical skills to a real-world context by exploring UK train ride data.
- Transform raw transport information into meaningful insights.
- Highlight travel patterns, efficiency, and accessibility in the UK rail system.
- Contribute to understanding mobility challenges in the UK.
- Build transferable skills for solving transportation issues in our own communities.

Our ideas

We Decided that we can enrich our data with adding more columns that will give us more insights that can help us more

BOOKED TRIPS

We added columns that can give us information if the trains is fully or partially booked

RAIL CARD

We will add columns that will show us if the passenger used rail card or not and the amount of discounts

SEATS

We added 2 columns that show us the total seats and how many are booked

BOOKING DATE

We will add column that will shows if the ticket is booked before or in the same day

FINAL PRICE

We added column that shows us the final price of the ticket

THE ISSUES WE FACED

MISSING DATA (THAT WE SOLVED)

- Missing Data in Rason of Delay: there was no reason of delay for the on-time trips so we filled the missings with no reason of delay and we also added now columns (Minutes of delay)

- Missing Data in Rail card column: We Found that if there was no card used the cells were empty, so we filled the empty cells with not used referred to that there were no cards used.

- We found that the time and date of each of the departure and arrival time were merged and the dates were wrong, so we separated the columns and generate new columns for the dates of arrival and departure

- We found null values the columns of Duration of the trip and Delay minutes for the canceled trips so we replaced the null values with zero

MISSING DATA (THAT WE COULDN'T SOLVE)

- The missing data in the actual arriving time for the cancelled trips: we couldn't add zeros so it will turn into 12 am, and we couldn't add text as we can't change the column data type.

- We Found negative values in thousands like (-1324) for the on time trips in the duration of the trip columns

THE ISSUES WE FACED

EXAMPLE FROM THE DATASET (Not Solved)

Actual Arrival Time	Journey Status
	Cancelled
	Cancelled
	Cancelled

THE ISSUES WE FACED

EXAMPLE FROM THE DATASET (The Solved Part)

DepartureDate	Date of Journey	ArrivalDate	Departure Time	Arrival Time	Reason for Delay	Used_Railcard
1/1/2024	1/1/2024	1/2/2024	11:30:00 PM	12:50:00 AM	No Delay Reason	No
1/2/2024	1/2/2024	1/2/2024	2:15:00 AM	3:35:00 AM	Technical Issue	No
1/2/2024	1/2/2024	1/2/2024	4:00:00 PM	5:20:00 PM	Technical Issue	Yes
1/2/2024	1/2/2024	1/2/2024	2:15:00 AM	3:35:00 AM	Technical Issue	No
1/3/2024	1/3/2024	1/4/2024	11:30:00 PM	12:00:00 AM	No Delay Reason	No
1/3/2024	1/3/2024	1/4/2024	11:45:00 PM	1:35:00 AM	No Delay Reason	Yes
1/3/2024	1/3/2024	1/4/2024	11:45:00 PM	12:35:00 AM	No Delay Reason	Yes
1/3/2024	1/3/2024	1/4/2024	11:45:00 PM	12:35:00 AM	No Delay Reason	Yes
1/2/2024	1/2/2024	1/2/2024	5:00:00 AM	6:20:00 AM	Staffing	No
1/2/2024	1/2/2024	1/2/2024	8:00:00 AM	10:15:00 AM	Staff Shortage	No
1/2/2024	1/2/2024	1/2/2024	8:00:00 AM	10:15:00 AM	Staff Shortage	No
1/2/2024	1/2/2024	1/2/2024	8:00:00 AM	10:15:00 AM	Staff Shortage	No
1/2/2024	1/2/2024	1/2/2024	8:00:00 AM	10:15:00 AM	Staff Shortage	No
1/2/2024	1/2/2024	1/2/2024	8:00:00 AM	9:20:00 AM	Signal Failure	No
1/2/2024	1/2/2024	1/2/2024	8:00:00 AM	10:15:00 AM	Staff Shortage	No
1/3/2024	1/3/2024	1/3/2024	6:15:00 AM	6:45:00 AM	Signal Failure	Yes
1/2/2024	1/2/2024	1/2/2024	9:45:00 AM	11:35:00 AM	Staff Shortage	Yes
1/2/2024	1/2/2024	1/2/2024	9:45:00 AM	11:35:00 AM	Staff Shortage	Yes
1/3/2024	1/3/2024	1/3/2024	8:15:00 AM	9:35:00 AM	Signal Failure	Yes
1/2/2024	1/2/2024	1/2/2024	1:30:00 PM	1:55:00 PM	Staffing	No
1/3/2024	1/3/2024	1/3/2024	12:45:00 PM	1:45:00 PM	Technical Issue	Yes
1/2/2024	1/2/2024	1/2/2024	4:00:00 PM	5:20:00 PM	Technical Issue	No
1/2/2024	1/2/2024	1/2/2024	4:00:00 PM	5:20:00 PM	Technical Issue	Yes
1/2/2024	1/2/2024	1/2/2024	4:00:00 PM	5:20:00 PM	Technical Issue	Yes
1/2/2024	1/2/2024	1/2/2024	4:00:00 PM	5:20:00 PM	Technical Issue	Yes
1/3/2024	1/3/2024	1/3/2024	3:30:00 PM	4:50:00 PM	Weather Conditions	Yes
1/3/2024	1/3/2024	1/3/2024	3:30:00 PM	4:50:00 PM	Weather Conditions	Yes
1/3/2024	1/3/2024	1/3/2024	6:45:00 PM	7:15:00 PM	Staff Shortage	Yes
1/3/2024	1/3/2024	1/3/2024	5:45:00 PM	7:05:00 PM	Technical Issue	Yes

DATA SHOW TIME!

EXAMPLE FROM THE DATASET

Transaction ID	Departure Station	Arrival Station	Delay (mins)	Seats Booked	Occupancy Level	Final Price (£)
da8a6ba8-b3dc	London Paddington	Liverpool Lime Street	0.0	43	Partially Booked	£30.10
25a03e9d-d281	Manchester Piccadilly	London Euston	12.5	96	Fully Booked	£16.10
353d60e2-46fd	Birmingham New Street	Edinburgh Waverley	0.0	246	Fully Booked	£3.00
43b60b16-62b5	London Kings Cross	York	8.3	201	Fully Booked	£13.00
84b07ecd-57e5	London St Pancras	Reading	0.0	248	Fully Booked	£76.00
8eee62c7-5030	Liverpool Lime Street	Manchester Piccadilly	0.0	172	Partially Booked	£35.00
953b9db0-34fa	Bristol Temple Meads	London Paddington	5.7	83	Partially Booked	£1.40
aedbb16b-fbe9	Oxford	London Euston	0.0	297	Fully Booked	£37.00

Data Schema (Part 1): Transaction & Operational Details

Column Name	Data Type	Description
Transaction Details		
Transaction_ID	String	Unique identifier for each transaction or ticket
Purchase_Date	Date	Date of ticket purchase or booking
Purchase_Time	Time	Time of ticket purchase or booking
Purchase_Channel	String	Channel of purchase (Online or Station)
Ticket_Class	String	Ticket class (Standard or First)
Ticket_Type	String	Ticket type (Advance, Off-Peak, or Anytime)
Passenger_Type	String	Passenger type (Adult, Disabled, or Senior)
Railcard_Used	Boolean	Indicates whether a Railcard discount was used
Operational Metrics		
Departure_Station	String	Scheduled departure station
Arrival_Station	String	Scheduled arrival station
Scheduled_Departure_Time	Time	Scheduled departure time
Actual_Arrival_Time	Time	Actual train arrival time
Delay_Mins	Float	Actual delay upon arrival in minutes (0 if on time)

Data Schema (Part 2): Occupancy & Capacity Metrics

Column Name	Data Type	Description
Occupancy & Capacity Data (continued)		
Train_Capacity	Integer	Total capacity of the train/service
Actual_Passengers	Integer	Actual number of passengers on the train
Occupancy_Level	String	Train occupancy level (Partially Booked or Fully Booked)
Seats_Booked_Total	Integer	Total number of seats booked in this transaction
Seats_Booked_Discount	Float	Number of seats booked using a discount (e.g., Railcard)
Seats_Booked_Full_Fare	Float	Number of seats booked at full fare (without discount)
Booking_Category	String	Booking category (e.g., Advance Booking, Same-Day Booking)
Is_Cancelled	String	Indicates whether the journey was cancelled (Yes/No)

Detailed Project Plan

Phase	Detailed Description	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 clean and ready-for-analysis dataset.	clean 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 (railwayypbsh.pbix)	Shimaa Mohammed, Mohamed Amin
5.Documentation	Compiling all project artifacts, analysis results, KPI definitions, and the project plan into a comprehensive documentation file.	Final Documentation File, Presentation	Zainab Elsaid

Key Performance Indicators: Measuring Rail Network Health

Operational Performance

Focused on service reliabilities and on time delivery

- **Punctuality Rate**
- **Average Delay Duration**
- **Main Delay Causes**

Capacity & Efficiency

Focused on resource utilization and train occupancy

- **Occupancy Rate**
- **Fully Booked Trains Ratio**
- **Seat Utilization**

Sales & Booking

Focused on commercial performance and revenue

- **Total Seats Booked**
- **Railcard Usage Rate**
- **Revenue Distribution**

Operational Insights: High Punctuality, Focused Delays

92.00%

Punctuality Rate

3%

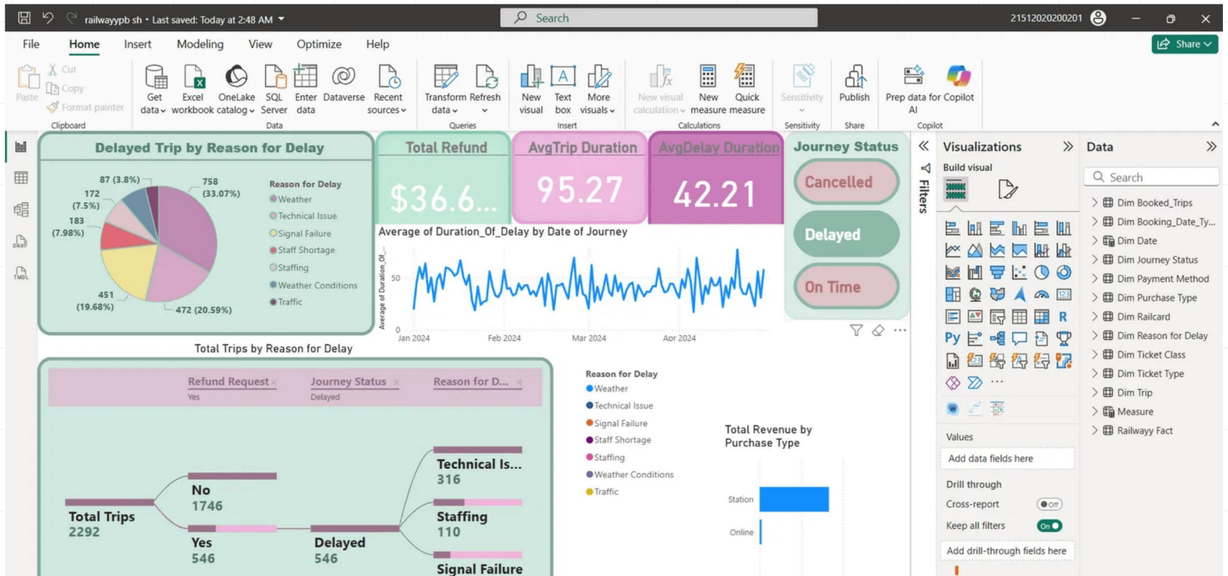
Delayed Trips

42.21

Avg Delay Duration (mins)

Main Delay Causes

Technical Issues	33.07%
Signal Failure	19.68%
Weather	7.50%
Staff Shortage	7.50%



Sales Insights: Revenue Distribution and Railcard Impact

$21K
Total Revenue

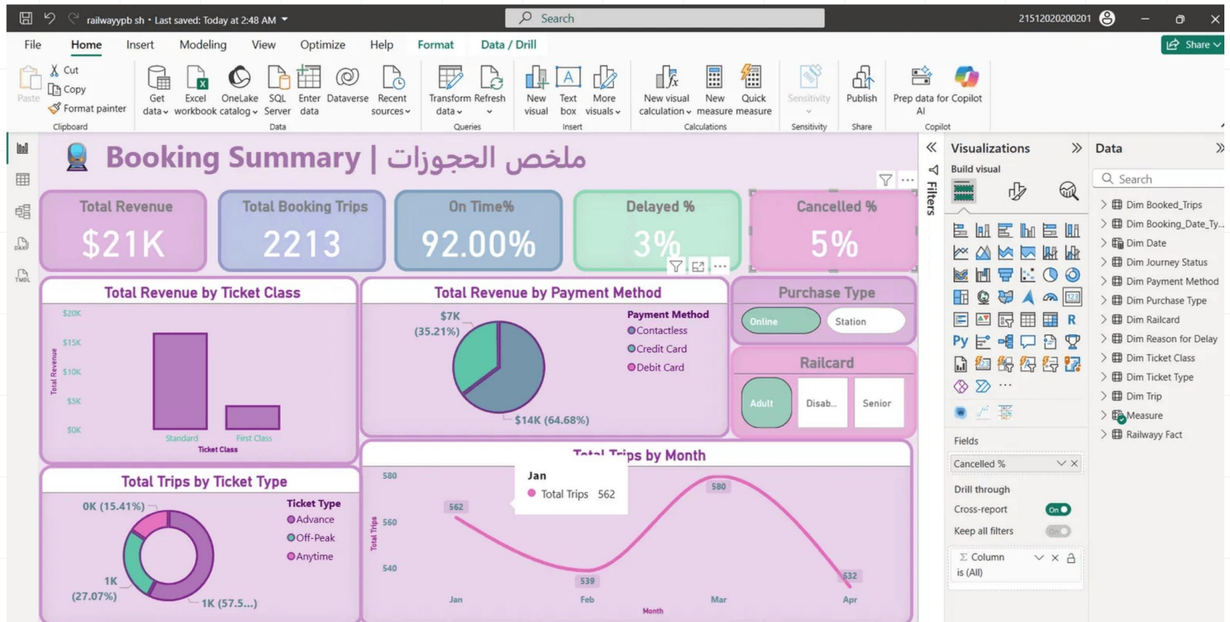
2,213
Total Booking Trips

Payment Methods

Credit Card	64.68%
Contactless	35.21%

Railcard Usage

Adult	~15K trips
Disabled	~20K trips
Senior	~5K trips



Project Methodology (Part 1): Data Foundation

1

Data Collection & Understanding

Identified required data sources (booking data, operational data, capacity data). Received and analyzed the initial dataset to understand its content, structure, and quality characteristics.

Deliverable: Initial Data File & Data Understanding Document

2

Data Cleaning & Preprocessing

Handled missing values, standardized date and time formats, removed duplicates, and transformed data types to ensure quality. Produced the clean and analysis-ready Cleaned_Data_Final.csv file.

Deliverable: Cleaned_Data_Final.csv

3

Data Analysis & KPI Definition

Conducted Exploratory Data Analysis (EDA) to identify trends and anomalies. Defined Key Performance Indicators (KPIs) across three categories: Operational Performance, Capacity & Efficiency, and Sales & Booking.

Deliverable: KPI List & Exploratory Analysis Results

Project Methodology (Part 2): Dashboard & Documentation

4

Power BI Dashboard Development

Imported clean data into Power BI, created data relationships, developed DAX measures and calculations for KPI computation, and designed an interactive, multi-page dashboard for visualization and exploration.

Deliverable: railwayypbsh.pbix Dashboard File

5

Documentation

Compiled all project artifacts, analysis results, KPI definitions, data schema documentation, and project plan into a comprehensive documentation file for reference and knowledge transfer.

Deliverable: Final Documentation File

✓ Project Methodology Complete

THANK YOU