
Data Analysis Project Idea :- UK Train Rides .

1- Project Overview :-

The primary objective of this railway data analysis project is to develop an interactive Power BI dashboard that helps analyze train journeys, operational performance, customer behavior, and route efficiency. The dashboard aims to support railway companies in making data-driven decisions for service improvements.

2- Scope of Analysis :-

- **Journey Performance:** Identify the most used routes, peak travel times, and average journey duration.
 - **Customer Insights:** Analyze ticket types, customer demographics, and repeat passenger behavior.
 - **Delay Analysis:** Evaluate delay causes, delay frequencies by station, and seasonal impacts.
 - **Regional Demand Trends:** Analyze demand across various stations and regions .
-

3-Dataset Overview :-

Column Name	Description
Transaction ID	Unique identifier for each ticket purchase.
Date of Purchase	When the ticket was bought.
Time of Purchase	Time when the ticket was bought
Purchase Type	Online or station purchase.
Payment Method	Credit Card, Contactless, etc
Railcard	Type of railcard used

Ticket Class	Standard or First-Class ticket.
Ticke Type	Advance, Off-Peak, etc
Price	Cost of the ticket
Departure Station	Journey starting station
Arrival Destination	Journey ending station.
Date of Journey	Scheduled date of the journey
Departure Time	Scheduled departure time.
Arrival Time	Scheduled arrival time.
Actual Arrival Time	Actual arrival time if available.
Journey Status	Whether the train was on time or delayed
Reason For Delay	Cause of delay if applicable.
Refund Request	If a refund was requested.

4-Methodology :-

Data collected from railway management systems, ticketing platforms, and IoT sensors installed on trains.

Data Cleaning & Transformation

- Handling missing values (e.g., unrecorded arrival times).
- Calculating delays:
 - **Departure Delay** = Actual Departure - Scheduled Departure
 - **Arrival Delay** = Actual Arrival - Scheduled Arrival
- Creating new fields: Journey Duration, Delay Category.

Data Analysis & Metrics Creation

KPIs Identified:

- Total Journeys Completed
- Average Journey Duration

- Average Delay (Departure / Arrival)
- Peak Travel Hours
- Passenger Volume by Route
- Revenue by Route and Ticket Type
- On-Time Performance %

Tools Used:

- **Power BI**
- **Python / Pandas (for initial data prep if needed)**

5-Dashboard Design Highlights:-

Visual Elements:

- Line charts for delay trends over time
- Heat maps for delay frequency by station
- Bar charts for revenue and passenger volume by route
- Slicers for filtering by date, route, delay reason, and ticket type
- Interactive map showing top-performing routes geographically

6-Risk Assessment :-

<u>Risk</u>	<u>Mitigation Strategy</u>
Data inconsistency	Apply validation checks and filters
Performance issues in Power BI	Optimize data model, use aggregation tables
Missing data in IoT streams	Use predictive modeling or flag incomplete journeys

7. Findings / Insights (Sample Expected Outcomes)

- 20% of delays caused by infrastructure issues
- Routes from **City A to City B** have the highest passenger volume during weekends

- First-Class tickets make up 15% of total revenue but only 5% of journeys
 - On-Time Performance is 85% during weekdays but drops to 72% during weekends
-

8. Conclusion & Recommendations

The interactive dashboard provides a real-time overview of railway operations, enabling:

- Proactive delay management
 - Resource allocation based on peak demand routes
 - Revenue optimization strategies
-

Next Steps:

- ✓ Conduct stakeholder review
- ✓ Refine dashboard filters and drill-downs
- ✓ Finalize for deployment on **Power BI Service** for management access