Power BI Development Report

1. ETL – Extract, Transform and Load

The ETL process was carried out entirely in Power Query, with the goal of preparing raw data for subsequent analysis and modelling. CSV files were imported using the **"Get Data > Folder"** option, ensuring that the model automatically updates whenever new files are added to the defined directory.

Several transformations were applied during this phase:

- **Data cleaning:** removal of unnecessary columns and duplicate records to ensure dataset consistency.
- **Handling missing values:**
    - In the *Reason for Delay* column, around 87% of values were missing, which were replaced with **"Unknown"**.
    - In the *Railcard* column, the value **"None"** was replaced with **"No Railcard"**, making interpretation clearer.
- **Standardization of inconsistent values:**
    - Delay reasons such as *"Signal Failure"* and *"Signal failure"* were unified.
    - The typo *"Adul"* was corrected to **"Adult"**.
- **Data type definitions:** for example, *Price* was set as a fixed decimal number, suitable for financial calculations.
- **Calculated columns created in Power Query:**
    - **Refund Amount** – amount refunded according to delay.
    - **Refund Granted** – binary flag to identify if a refund was issued.
    - **Route** – combination of Departure Station and Arrival Destination.
    - **Delay (minutes)** – calculated delay in minutes based on scheduled vs. actual arrival time.

This process ensured that the data was clean, consistent, and ready for analysis.

2. Data Model

The data model was designed using a **star schema**, with a central fact table (**Railway_Facts**) connected to multiple dimension tables. This approach improves Power BI performance and simplifies the creation of measures and visualizations.

2.1. Fact Table – Railway_Facts

The **Railway_Facts** table stores all transactional records and key metrics. Main columns include:

- Transaction ID
- Date of Journey and Purchase Date
- Departure Station / Arrival Destination
- Ticket Type and Class
- Payment Method
- Railcard usage
- Delay in minutes
- Refund Amount and Refund Granted
- Journey Status
- Reason for Delay

2.2. Dimension Tables

Dimension tables were created as references from the fact table, keeping only relevant attributes and removing duplicates:

- **Dim_Ticket_Types:** unique combinations of *Ticket Type* and *Ticket Class*, with surrogate key *Ticket_Type&Class_ID*.
- **Dim_Railcards:** railcard categories (*Adult, No Railcard, Disabled, Senior*).
- **Dim_Purchase:** purchase channels (*Online* or *Station*).
- **Dim_Payment_Methods:** payment methods (*Contactless, Credit Card, Debit Card*).

- **Dim_Journey_Dates:** derived from *Date of Journey*, with additional fields such as Year, Month, Day of Week, Weekend flag, and Week Number.
- **Dim_Delays:** normalized reasons for delays (e.g., *Signal Failure, Traffic, Staffing*).
- **DatasUnicas:** a calendar table generated from both *Date of Journey* and *Purchase Date*, supporting time-based analysis and slicers.

3. Relationships

Clear relationships were established between **Railway_Facts** and the dimensions to ensure data integrity:

- **Railway_Facts[Date_of_Journey] → Dim_Journey_Dates[Date]** (active).
- **Railway_Facts[Purchase_Date] → Dim_Journey_Dates[Date]** (inactive, activated via DAX when required).
- **Railway_Facts[Reason_for_Delay] → Dim_Delays[ID_Reasons for Delay]**.
- **Railway_Facts[Ticket_Type] / [Ticket_Class] → Dim_Ticket_Types[Ticket_Type&Class_ID]**.
- **Railway_Facts[Railcard] → Dim_Railcards[Railcard ID]**.
- **Railway_Facts[Purchase_Type] → Dim_Purchase[Purchase Type ID]**.
- **Railway_Facts[Payment_Method] → Dim_Payment_Methods[Payment Method ID]**.

Inactive relationships (such as *Purchase_Date*) allow alternative perspectives without compromising the main model's performance.

4. DAX Measures

Several measures were created in Power BI to support analytical needs and answer key business questions:

- **Delay_Minutes:** calculates train delays in minutes.
- **Refund_Amount:** applies the refund policy based on the level of delay.
- **Refund_Granted:** binary indicator ("Yes"/"No").
- **Days_Before_Journey:** number of days between ticket purchase and journey date.
- **Transaction_Count:** counts the total number of journeys (transactions).

All measures were implemented with DAX, ensuring flexibility and efficiency, as they dynamically adapt to slicers and filters applied in reports.

5. Visualizations and Reports

The Power BI dashboard was structured into several thematic pages, each addressing specific analysis objectives:

1. **Executive Summary:** overall view with main KPIs (total sales, number of journeys, refunds, and delays).
2. **Routes:** identification of the most popular routes, share of delayed, cancelled and on-time journeys, and an interactive map for geographic analysis.
3. **Tickets:** analysis of revenue by ticket type and class.
4. **Purchases:** customer behaviour analysis, including average purchase lead time, most common purchase times, preferred channels, and impact on average ticket price.
5. **Delays:** breakdown of main reasons for delays, percentage of refunds issued, journey punctuality, and routes with the highest accumulated delay in minutes.

Additional features included:

- **Slicers** for filtering by time, ticket type/class, railcard usage, and payment method.

- **Custom tooltips** providing additional details (e.g., revenue, average delays, number of journeys).
- **Custom and AI visuals**, such as decomposition trees and heatmaps, to uncover deeper insights.

6. Conclusion

The implemented model, centred on **Railway_Facts** and supported by the dimension tables **Dim_Ticket_Types, Dim_Railcards, Dim_Purchase, Dim_Payment_Methods, Dim_Journey_Dates, Dim_Delays, and DatasUnicas**, enabled the creation of a scalable, efficient, and user-friendly dashboard.

This solution successfully answered all project requirements, including:

- What are the most popular routes?
- Is demand higher during weekends or specific periods?
- How does revenue vary by ticket type and class?
- What is the punctuality performance of journeys?
- What are the main causes of delays?
- What is the financial impact of refunds?

The project therefore covered all critical phases of the Power BI analytics cycle: **ETL → Data Modelling → DAX Measures → Visualizations and Storytelling**, resulting in a complete and decision-oriented reporting tool.