



# National Rail ANALYTICS PROJECT

Leveraging data to transform rail  
transportation insights

# UK NATIONAL RAIL ANALYTICS PROJECT

## Instructor

Ahmed Abd Ellatif (G1)

## Data Analysis Team

Asmaa Ibrahim

Heba Essam

Maha Muhammed

Mahmoud Abd El-Ghani

Mahmoud El-Sherif

Ahmed Sakr

Mohammad Hussain



## CONTENT

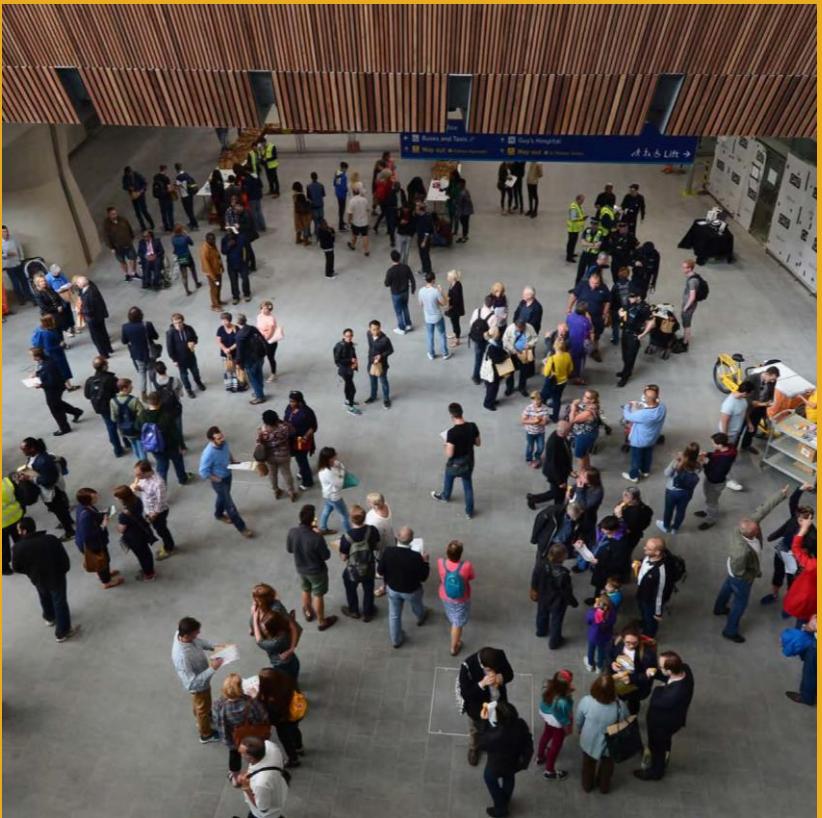
- PROJECT OVERVIEW
- DATASET DESCRIPTION
- KEY DIAGNOSTIC QUESTIONS
- DATA CLEANING & PREPARATION
- DATA ENRICHMENT & DIMENSION TABLES
- EXPLORATORY DATA ANALYSIS (EDA)
- DATA MODELING
- DAX MEASURES OVERVIEW
- DASHBOARD WALKTHROUGH
- KEY INSIGHTS & FINDINGS
- CONCLUSIONS
- RECOMMENDATIONS / ACTIONS
- FORECASTS



UK NATIONAL RAIL  
ANALYTICS PROJECT

# PROJECT OVERVIEW

# PROJECT OVERVIEW



This project focuses on analyzing UK National Rail journeys to uncover patterns in passenger behavior, ticket purchasing, travel performance, and operational efficiency. Using a cleaned dataset of train ride transactions, the analysis aims to identify key trends related to ticket types, pricing, travel timings, delays, and refund requests.

The insights generated help demonstrate practical data-analysis skills, including data preparation, dashboard design, and the extraction of meaningful conclusions from real-world transport data.

UK NATIONAL RAIL  
ANALYTICS PROJECT

# DATASET DESCRIPTION

# DATASET DESCRIPTION



- **Size:** The dataset contains **31,653 train ride records** with a wide range of journey, ticket, passenger, and performance attributes.
- **Grain:** Each row represents **a complete train-ride transaction** for a single passenger, including ticket purchase details and the journey's actual outcome.
- **Purpose:** The dataset is designed to support **operational analysis**, including ticketing behavior, travel performance, delay patterns, and customer refund activity. It provides the foundation for building dashboards that reveal key insights about passenger habits and railway efficiency.

# DATASET DESCRIPTION

## Key Components of the Dataset

- **Transaction & Purchase Details:**  
Transaction ID, Date of Purchase, Time of Purchase, Purchase Type (Online or Station), Payment Method.
- **Passenger & Ticket Information:**  
**Columns:** Railcard type, Ticket Class, Ticket Type, Ticket Price.
- **Journey Information:**  
**Columns:** Departure Station, Arrival Destination, Date of Journey, Scheduled Departure Time, Scheduled Arrival Time.
- **Performance & Delay Metrics:**  
**Columns:** Actual Arrival Time, Journey Status (On Time, Delayed, Cancelled), Reason for Delay.
- **Customer Response:**  
**Columns:** Whether a Refund Request was submitted.

# DATASET DESCRIPTION

## Dataset Highlights

- Covers a broad range of ticket types and purchase behaviors.
- Includes both scheduled and actual arrival times, enabling delay analysis.
- Provides rich categorical data suitable for segmentation and performance insights.

## Data Limitations

- The dataset covers only **four months** of the year, limiting the ability to analyze full-year seasonality.
- The dataset lacks a user demographics **or customer dimension table**, making it challenging to analyze segmentation by user characteristics.
- Journey records do not include **distance or route-length information**, preventing analysis linking **delay duration to journey distance**, or route complexity.

UK NATIONAL RAIL  
ANALYTICS PROJECT

## KEY DIAGNOSTIC QUESTIONS

# KEY DIAGNOSTIC QUESTIONS

## I. Operational Performance & Reliability

- Punctuality vs. Benchmark
- Root Cause Identification
- Service Disruption Severity

## II. Financial Health & Cost of Disruption

- Revenue Stream Segmentation
- Financial Leakage from Refunds
- Refund Policy Effectiveness

## III. Customer Behavior & Growth Momentum

- Channel Preference
- Growth Sustainability



UK NATIONAL RAIL  
ANALYTICS PROJECT

# TECHNICAL DEEP DIVE

# TECHNICAL DEEP DIVE



## Data Cleaning & Preparation

- **Source Data Ingestion:** Loaded 31,653 raw transaction records.
- **Data Type Conversion:** Ensured proper date and time formats for analysis.
- **Data Standardization:** Unified and cleaned text entries (e.g., Journey Status, Reason for Delay).
- **Handling Missing Values:** Imputed missing values in arrival times to correctly categorize journeys as Delayed or Cancelled.
- **Calculated Columns:** Created derived columns for actual ride time and delay duration.

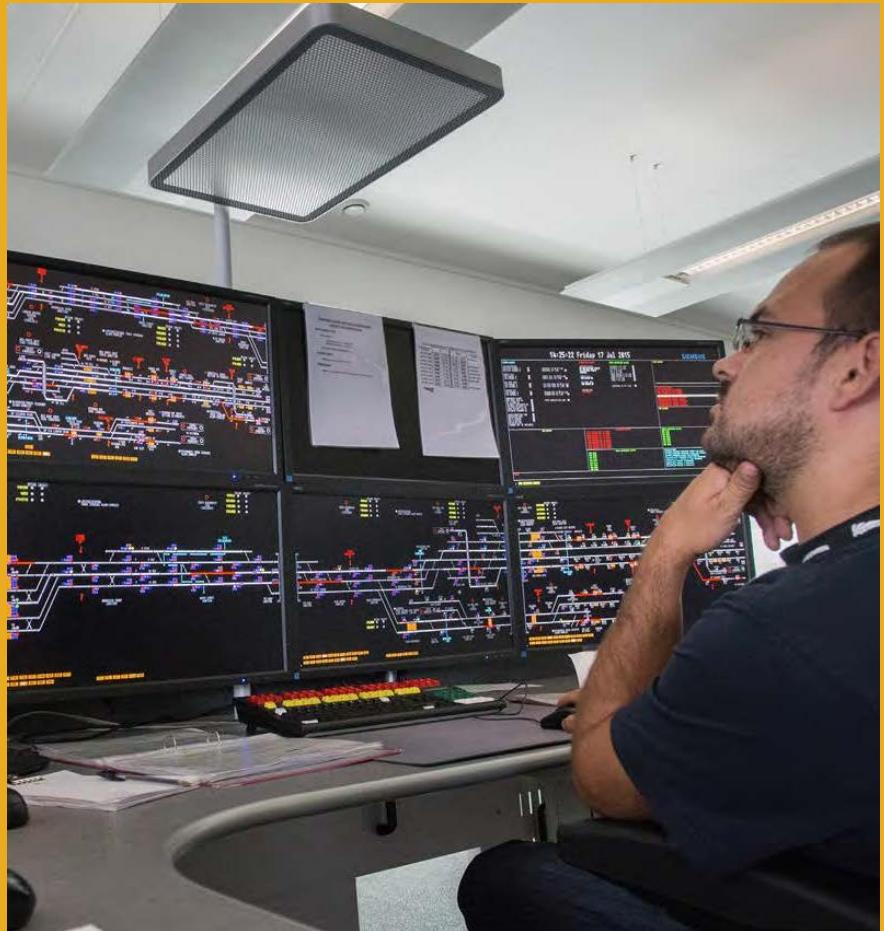
# TECHNICAL DEEP DIVE

## Data Enrichment & Dimension Tables

- **Date Dimension:** Created a dedicated Date Dimension table to enable robust Time Intelligence calculations (MoM, MTD).
- **Hierarchies:** Established hierarchies for geographical and temporal analysis (e.g., Station, Date).
- **Segmentation:** Grouped and categorized Passenger attributes (e.g., Railcard holders) and Ticket details.



# TECHNICAL DEEP DIVE



## Exploratory Data Analysis (EDA)

- Total Rides: 31,653 total transactions.
- Total Revenue: £742K generated.
- Performance Baseline: 86.82% of rides were On Time.
- Disruption Volumes: 7.24% Delayed rides and 5.94% Cancelled rides.
- Financial Leakage: £39K paid in refunds, with an Incident Refund Rate of 26.80%.
- Growth Trajectory: Strong Month-over-Month Growth of 33.89%.

# TECHNICAL DEEP DIVE

## Data Modeling

- **Model Architecture:** Implemented a **Star Schema** to optimize performance and query speed in Power BI.
- **Fact Table:** Central UK Trains Rides table containing all transactional data.
- **Dimension Tables:** Linked Date and Stations dimensions.
- **Relationships:** Established robust One-to-Many relationships to ensure accurate measure calculations across filters.



# TECHNICAL DEEP DIVE

## DAX Measures Overview

- **Operational Measures:** Calculated core metrics like PPM%, Delay Time Averages, and Service Failure Rates.
- **Financial Measures:** Defined revenue totals, Average Ticket Price, and the financial impact of Refunds.
- **Time Intelligence:** Utilized DAX functions (DATEADD, TOTALMTD) for Month-over-Month growth and trend analysis.
- **Segmentation Measures:** Quantified passenger groups (Railcard holders vs. Non-Railcard holders).

UK NATIONAL RAIL  
ANALYTICS PROJECT

# DASHBOARD WALKTHROUGH

# DASHBOARD WALKTHROUGH

Excitement builds...

200<sup>th</sup>

...for Railway's 200th anniversary in 2025

Executive Summary

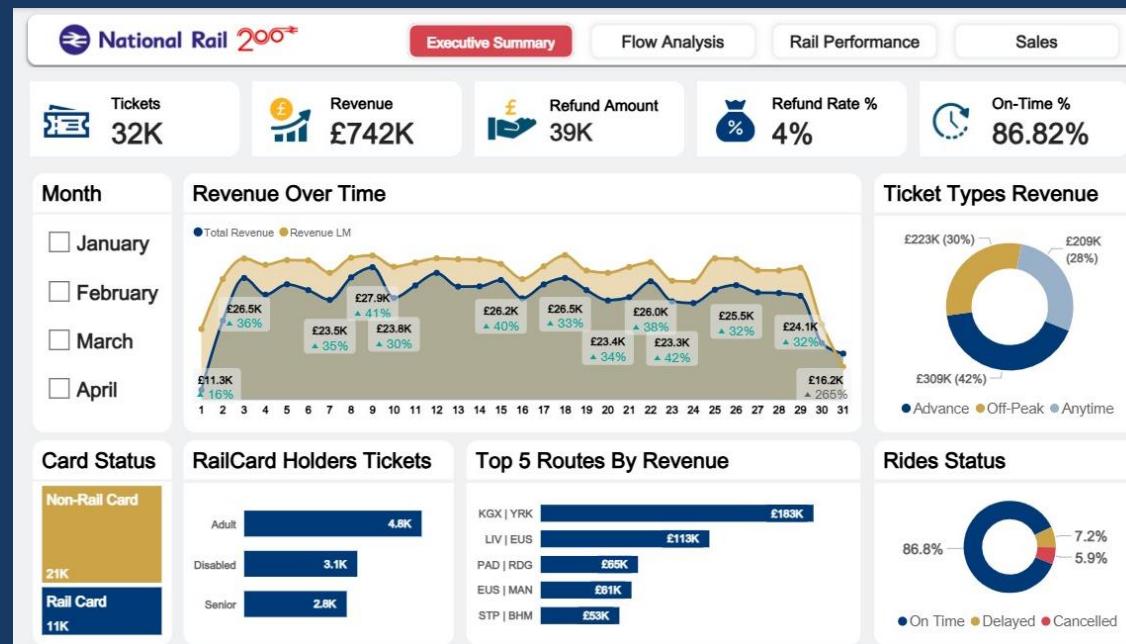
Flow Analysis

Rail Performance

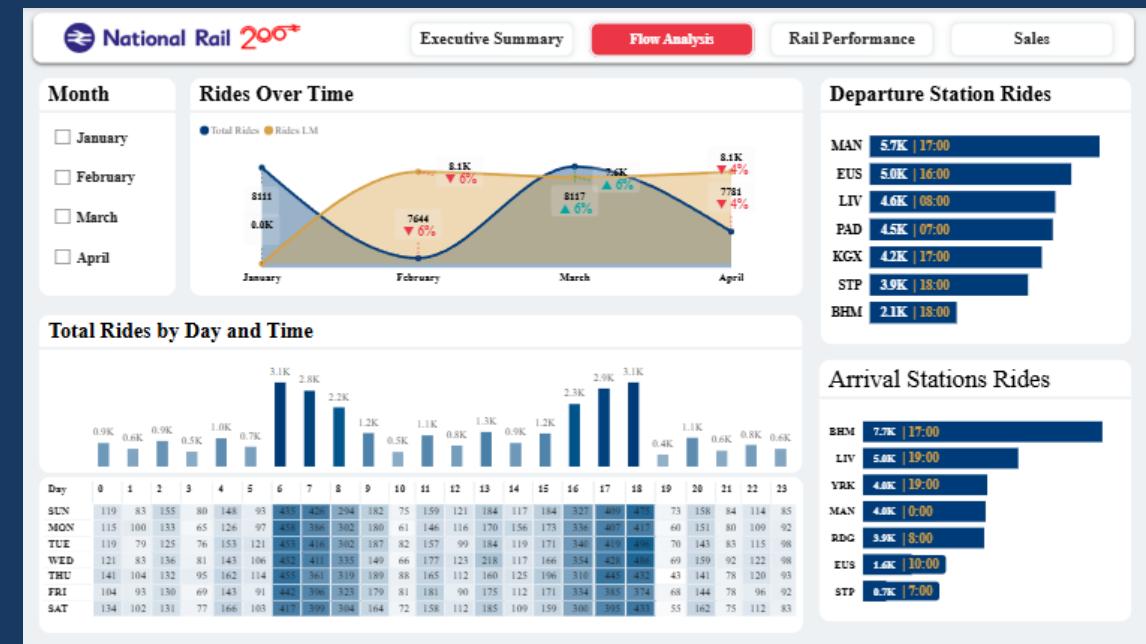
Sales

Landing Page

# ANALYTICAL REVIEW



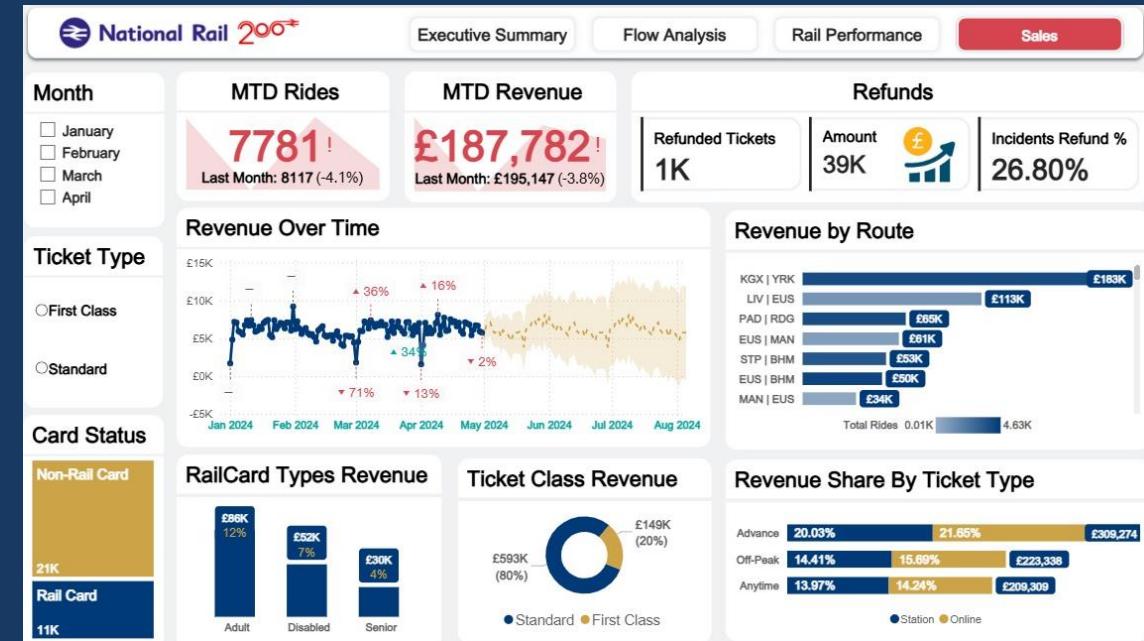
Executive Summary



# ANALYTICAL REVIEW

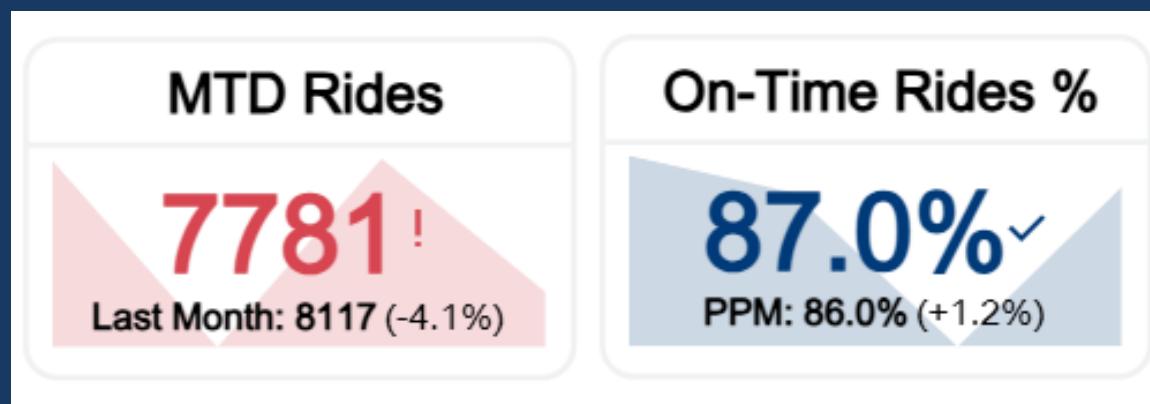
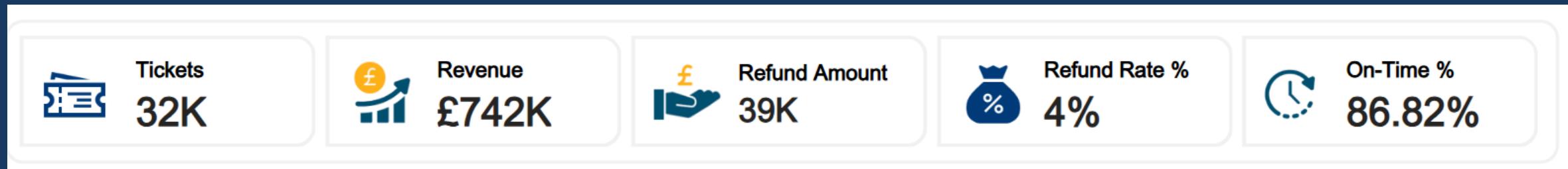


Rail Performance

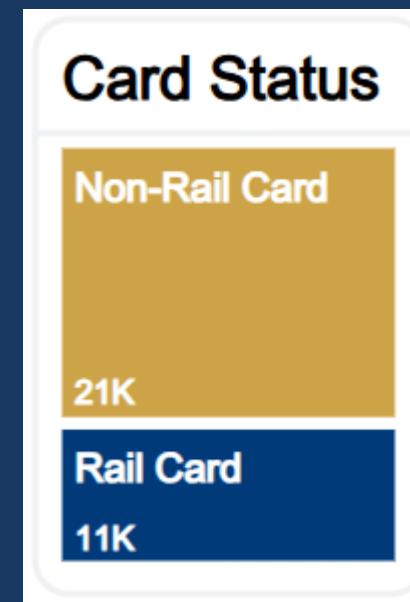


Sales Analysis

# ANALYTICAL REVIEW

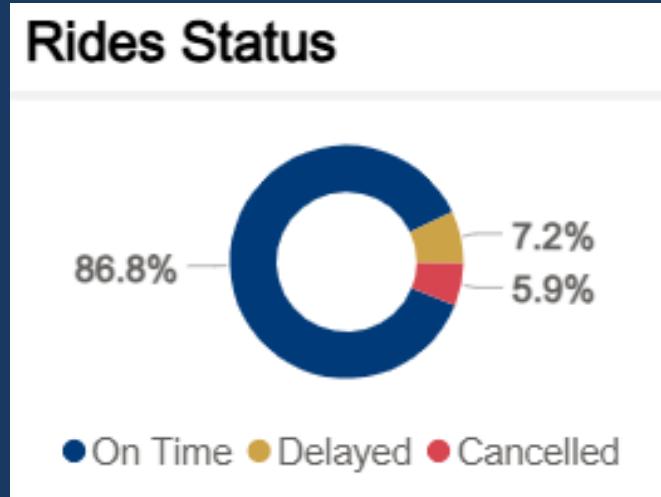
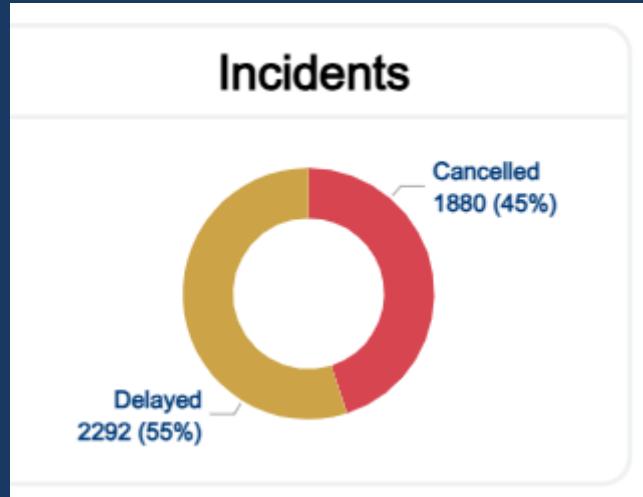


KPIs and Cards

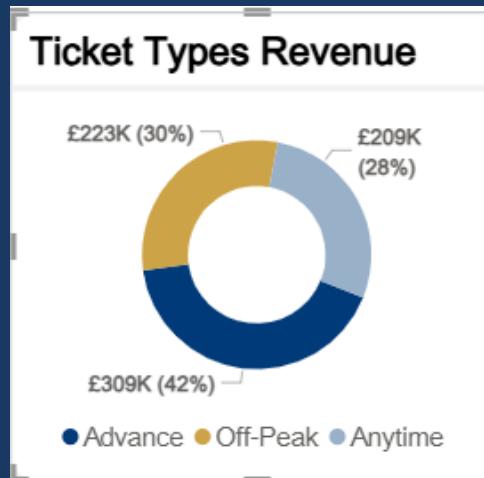
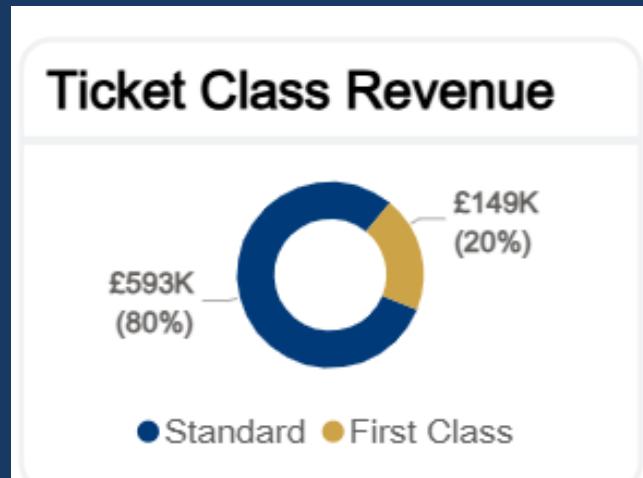


Tree Map Chart

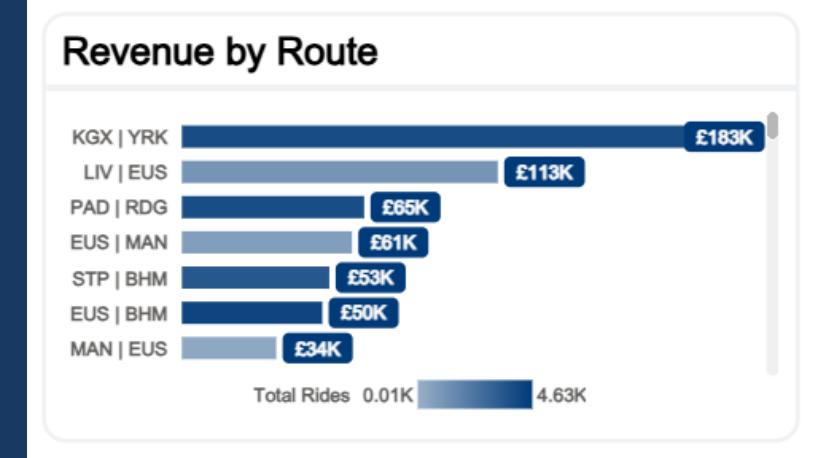
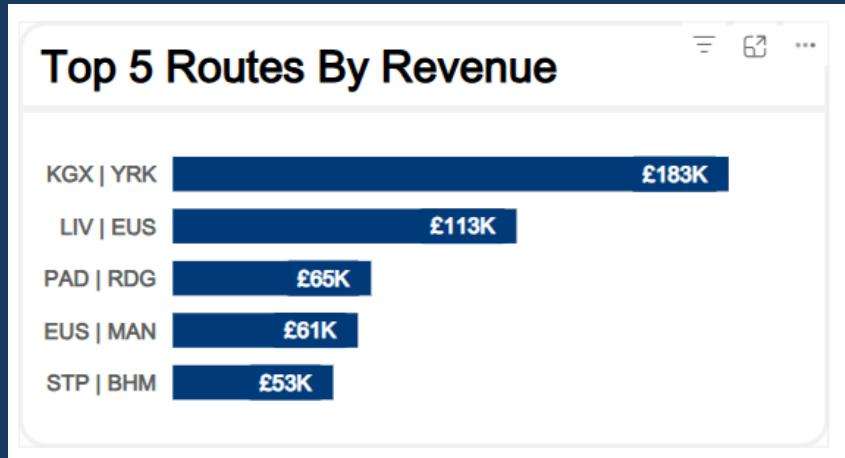
# ANALYTICAL REVIEW



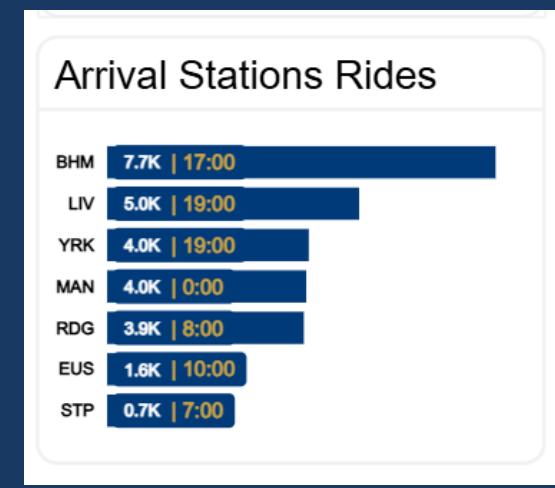
Pie Charts



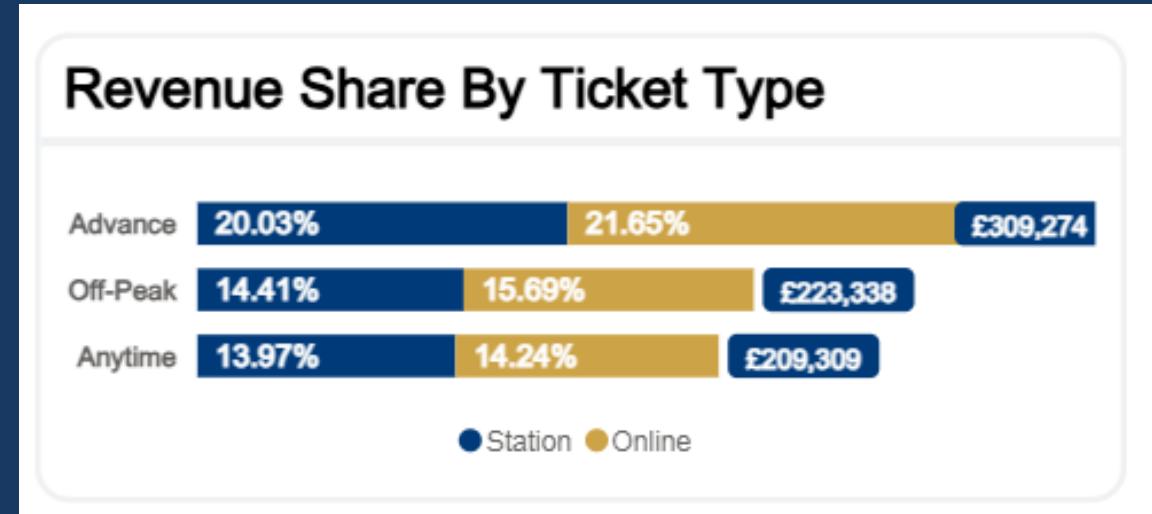
# ANALYTICAL REVIEW



## Bar Charts

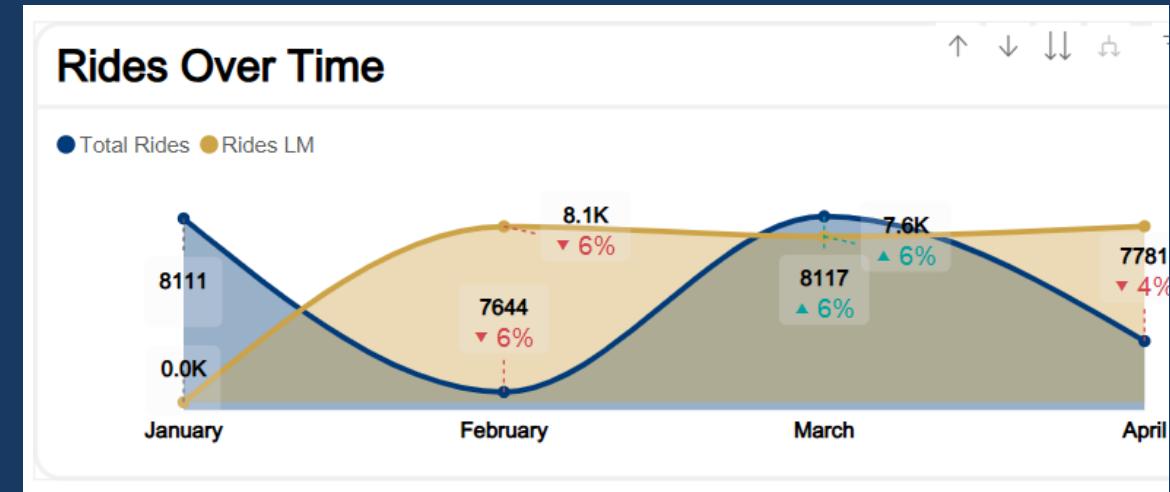


# ANALYTICAL REVIEW

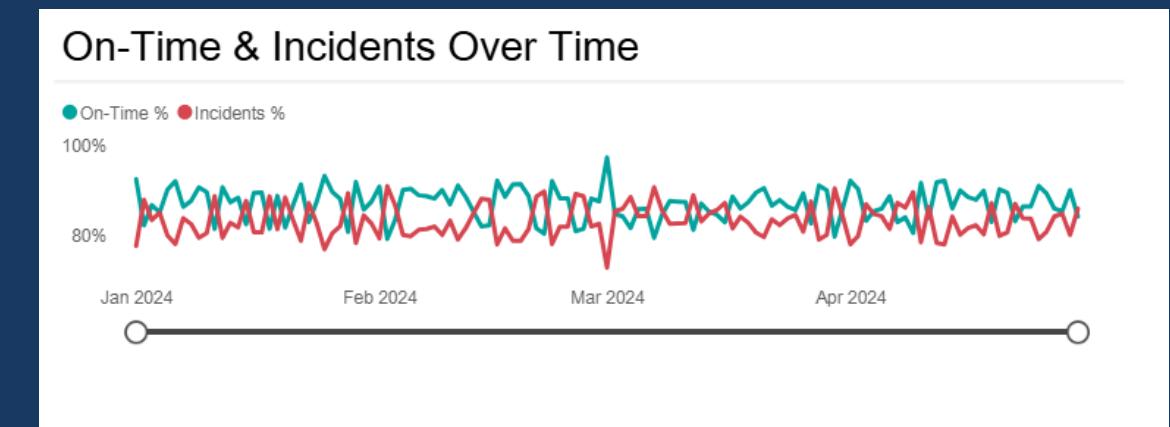


Stacked Bar Charts

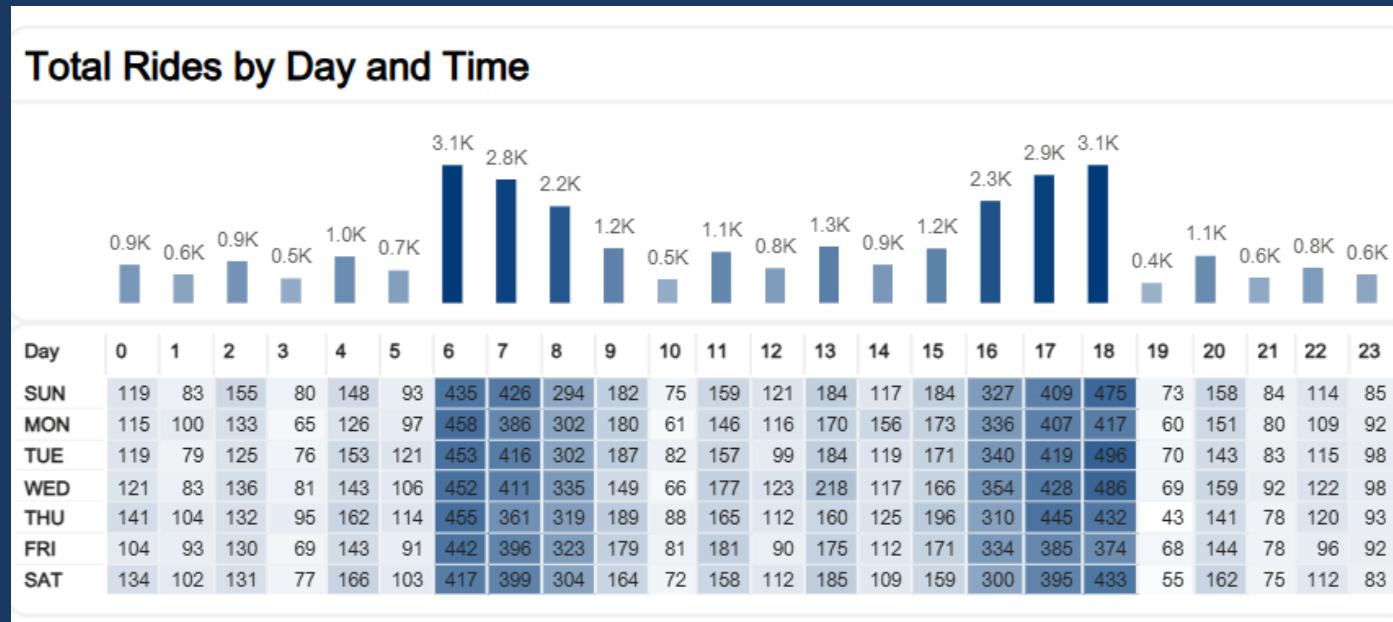
# ANALYTICAL REVIEW



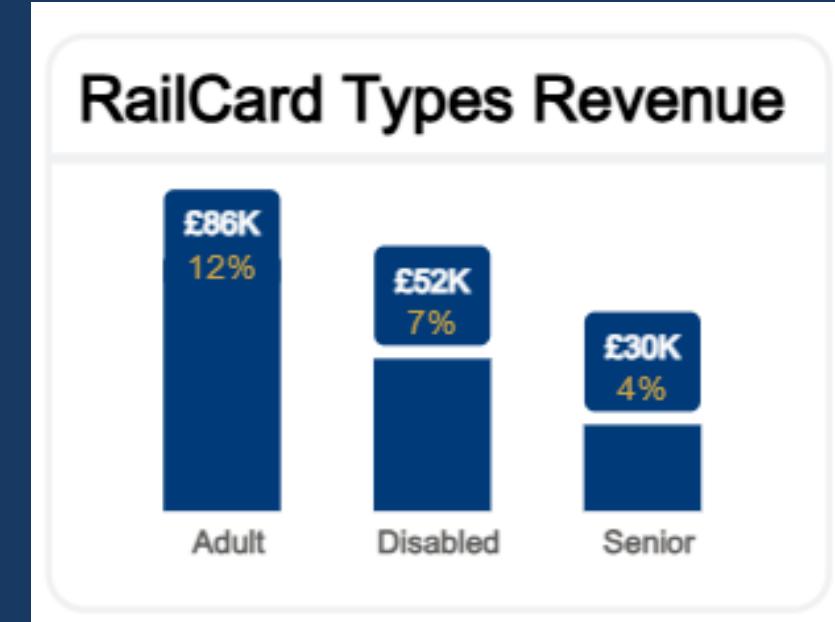
## Line Charts



# ANALYTICAL REVIEW



Combined Matrix and  
Column Charts



Column Chart with KPI  
Indicator

# UK NATIONAL RAIL ANALYTICS PROJECT

## KEY INSIGHTS & FINDINGS

# KEY INSIGHTS & FINDINGS

- **Cancellation Crisis:** The 5.94% cancellation rate **exceeds the 5.0% threshold** for severe disruption, highlighting critical service instability despite meeting the general PPM target.
- **Infrastructure Failure Dominance:** **Signal Failure** is confirmed as the single largest root cause of delays, underscoring systemic reliability issues in the railway infrastructure.
- **Direct Cost of Disruption:** £39K in refunds and a high **26.80% Incident Refund Rate** quantify the direct financial penalty of operational failures, proving that poor performance is expensive.
- **Growth Vulnerability:** The impressive **33.89% MoM Growth** is at risk unless immediate action is taken to stabilize service reliability and convert new passengers into loyal customers.

UK NATIONAL RAIL  
ANALYTICS PROJECT

# CONCLUSIONS

# CONCLUSIONS

- **Performance Gap:** The network is **punctual** (86.82% PPM compliant) but severely **unreliable** (high cancellation rate).
- **Financial Impact:** Unstable operational performance is directly translating into quantifiable financial losses through high refund rates.
- **Focus Area:** The immediate priority must shift from simply tracking on-time performance to aggressively mitigating the **root causes** of delays and cancellations.

UK NATIONAL RAIL  
ANALYTICS PROJECT

# RECOMMENDATIONS / ACTIONS

# RECOMMENDATIONS / ACTIONS



- **Prioritize Signal Investment:** Allocate urgent capital expenditure to upgrade and maintain signalling infrastructure to directly address the primary cause of delays.
- **Operational Review of Cancellations:** Launch an immediate process review to identify and eliminate the systemic issues driving the high 5.94% cancellation rate.
- **Proactive Refund Mitigation:** Implement strategies to reduce average delay times, thereby minimizing the volume of refund-eligible disruptions and capping financial leakage.

UK NATIONAL RAIL  
ANALYTICS PROJECT

# FORECASTS (SUGGESTED FUTURE PROJECTIONS)

# FORECASTS

- **Revenue Projection:** Forecast future revenue growth based on scenario analysis (e.g., impact of reducing cancellation rates by 1.0%).
- **Service Level Projection:** Project the expected reduction in delay minutes if Signal Failure is mitigated by X%.
- **Financial Impact Forecasting:** Model the potential savings in refund expenditure over the next fiscal period by improving punctuality.

Q & A