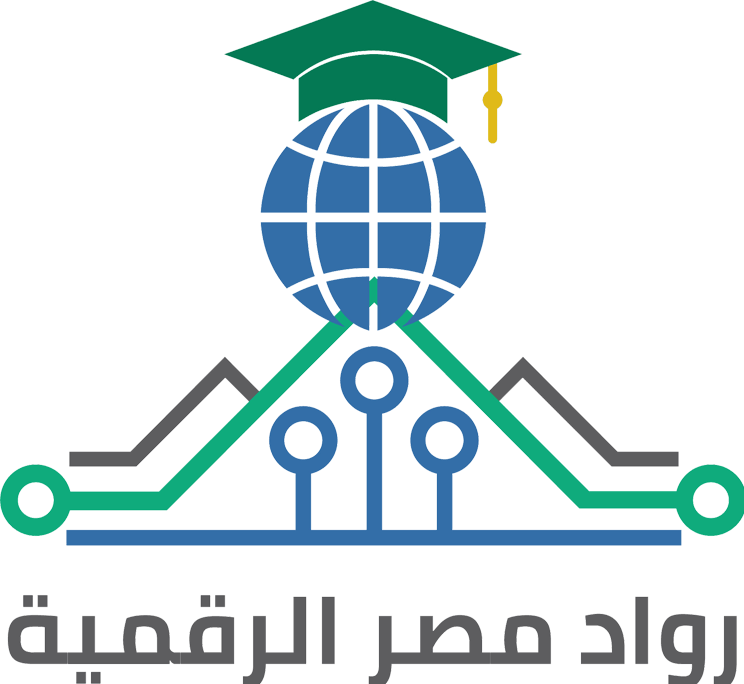
**YAT 402A Group 1 Project**

**UK Train Rides**

# **Team Members**

**Abdullah Maher Abdullah**

**Hanan Mohammed Mohammed**

**Michael Sherif Youssef**

**Mohamed Mahmoud Ahmed**

**Omnia Hussein Ali**

Contents

[**Team Members** 2](#_heading=h.66x83wphx1g2)

[Project Planning & Management 4](#_heading=h.8nqlej1l30je)

[1.](#_heading=h.ndqn27jwln7a) Project Proposal: UK Train Rides Analysis 4

[2.](#_heading=h.flbvnxx8m8w2) Project Plan 5

[3.](#_heading=h.hc7x1a54dw77) Risk Assessment & Mitigation Plan 7

[4.](#_heading=h.6og57ji1w2zg) Key Performance Indicators (KPIs) 8

[Requirements Gathering 10](#_heading=h.etm1wycs6zmf)

[1.](#_heading=h.k6s5ibm63x39) Stakeholder Analysis 10

[2.](#_heading=h.8cqq5uchptlb) User Stories & Use Cases 12

[3.](#_heading=h.4mtvhvgls8lu) Functional Requirements 13

[4.](#_heading=h.msvc9h3zhsny) Non-Functional Requirements 15

[System Analysis & Design 16](#_heading=h.aeolvawcsqcw)

[1.](#_heading=h.njvjra809bdu) Problem Statement & Objectives 16

[Key Business Questions for Dashboard Design 18](#_heading=h.4e9uxxxrkmrn)

[Technical Documentation 19](#_heading=h.9v6ylgn4wtu)

[1.](#_heading=h.nqmvuuom67y3) Data Loading and Structure 19

[2.](#_heading=h.rxkb04978kut) Measures and Calculated Fields 19

[3.](#_heading=h.deqqvmf001uy) Report Pages 22

[4.](#_heading=h.9w0i98lg27xc) Slicers 22

[5.](#_heading=h.l3p9a1tgofap) Cards 23

[6.](#_heading=h.8kl62ed10mv) Page-Level Visuals 23

[User Manual 26](#_heading=h.i6ftng5gqvus)

[Insights 28](#_heading=h.l8rkg1npa0ed)

[Actions and Recommendations 29](#_heading=h.hfagw5cmjy14)

[1.](#_heading=h.737ozcwlfui9) General Actions 29

[2.](#_heading=h.4i10f6d6ebda) RCA (Manchester - Liverpool Route) 30

[3.](#_heading=h.r3bs06ot5eq5) PDCA Plan (Manchester - Liverpool Route) 30

[4.](#_heading=h.flto8huwbo1) RCA (General Delays & Refunds) 31

[5.](#_heading=h.k71921bj3ja) DMAIC Plan (General Delays & Refunds) 31

[6.](#_heading=h.okekfps91o52) PDCA Plan (First Class Ticket Usage) 32

# Project Planning & Management

## Project Proposal: UK Train Rides Analysis

* 1. Overview

This project is all about digging into UK train travel data to uncover useful insights. Using Power BI, we’ll clean up the data, explore key trends, build a solid dashboard, and wrap things up with a final presentation. The aim? To give decision-makers clear, data-backed insights into train travel patterns and performance.

* 1. Objectives
* Clean and prep the dataset so it's accurate and reliable.
* Identify key questions that will lead to valuable insights.
* Build a Power BI dashboard to showcase trends and patterns.
* Summarize findings in a clear, actionable report.
  1. Expected Outcomes

By the end of the project, we’ll have a clean dataset, a solid list of analysis questions, an interactive Power BI dashboard, and a well-structured final report. The goal is to make train travel data more accessible and actionable for decision-makers.

* 1. Project Scope

The UK Train Rides Analysis project focuses on leveraging Power BI to clean, analyze, and visualize train travel data. The project will provide actionable insights for decision-makers by identifying trends, patterns, and key performance metrics within the dataset.

* In-Scope
  + Data collection and preprocessing to ensure accuracy and completeness.
  + Identification of relevant analysis questions to derive meaningful insights.
  + Development of a Power BI dashboard for interactive visualization.
  + Summarization of findings in a final report and presentation for stakeholders.
* Out-of-Scope
  + Real-time data integration or live tracking of train rides.
  + Predictive modeling or machine learning-based forecasting.
  + Changes or modifications to the dataset outside of preprocessing.
  + Implementation of operational changes based on insights.

## Project Plan

* 1. Timeline & Milestones
* Week 1: Data Cleaning & Preprocessing
  + Milestone: Clean dataset ready for analysis.
  + Deliverables:
    - A fully cleaned and structured dataset that removes inconsistencies, duplicates, and missing values.
    - Clear documentation of all preprocessing steps, including handling missing data, formatting, and transformations.
    - Initial exploratory data analysis to ensure data usability.
* Week 2: Define Analysis Questions
  + Milestone: Finalized set of analysis questions.
  + Deliverables:
    - A well-defined list of key questions that align with business objectives and decision-making needs.
    - Justification for each question based on dataset capabilities and potential insights.
    - Prioritization of questions to ensure meaningful analysis within the project timeline.
* Week 3: Dashboard Development
  + Milestone: Completed interactive Power BI dashboard.
  + Deliverables:
    - A fully functional Power BI dashboard displaying key insights, trends, and performance metrics.
    - Interactive filters and visualizations that allow users to explore different aspects of the data.
    - Optimized design for clarity, usability, and decision-making support.
* Week 4: Final Report & Presentation
  + Milestone: Final report and presentation completed.
  + Deliverables:
    - A detailed report summarizing the entire project, including data preprocessing, key findings, and dashboard insights.
    - A PowerPoint presentation highlighting key results, visualizations, and recommendations.
    - A walkthrough of the Power BI dashboard for stakeholders to ensure ease of use and understanding.
  1. Resource Allocation

| Role | Responsibilities | Tools Used |
| --- | --- | --- |
| Omnia Hussein Ali | Oversee the project, ensure milestones are met, coordinate team efforts | PowerPoint, Power BI |
| Mohamed Mahmoud Ahmed | Handle data cleaning, transformation, and ensure data integrity | Power BI, Excel |
| Hanan Mohammed Mohammed | Explore data, define key analysis questions, and create insights | Power BI |
| The Entire Group | Worked together to create the measures and columns using DAX in Power BI | Power BI |
| Abdullah Maher Abdullah | Design and build the Power BI dashboard for data visualization | Power BI |
| Michael Sherif Armanyous | Summarize findings in a structured report and presentation | PowerPoint, Word |

## Risk Assessment & Mitigation Plan

* 1. Data Quality Issues
* Risk: Missing, inconsistent, or messy data could mess up the analysis.
* Mitigation: Set up validation checks, clean the data properly, and document everything to keep it structured.
  1. Scope Creep
* Risk: Too many extra requests or changes could push the project off track.
* Mitigation: Stick to the planned analysis, set clear boundaries, and confirm scope with stakeholders upfront.
  1. Technical Challenges
* Risk: Running into Power BI limitations, slow performance, or unexpected issues.
* Mitigation: Allocate time for troubleshooting, lean on team members with experience, and keep learning as needed.
  1. Stakeholder Expectations
* Risk: Deliverables might not match what stakeholders expect.
* Mitigation: Keep communication open, share progress updates, and get feedback early to stay aligned.
  1. Team Availability & Workload
* Risk: Team members could be overloaded or unavailable, causing delays.
* Mitigation: Distribute tasks clearly, set achievable deadlines, and track progress regularly to adjust if needed.

## Key Performance Indicators (KPIs)

* 1. Data Cleaning and Preprocessing (Week 1) – 25%

| KPI | Metric | Target | Evaluation Method |
| --- | --- | --- | --- |
| Task Completion Rate | % of preprocessing tasks completed on time | ≥ 90% | Checked against project timeline |
| Data Quality Improvement | % reduction in missing values, duplicates, and errors | ≥ 80% | Comparison of raw vs. cleaned dataset |
| Team Contribution Score | Peer evaluation (average score out of 10) | ≥ 7/10 | Anonymous team member ratings |
| Documentation Clarity | Lecturer’s evaluation (score out of 10) | ≥ 8/10 | Assessed on completeness, clarity, and readability |

* 1. Analysis Questions Phase (Week 2) – 20%

| KPI | Metric | Target | Evaluation Method |
| --- | --- | --- | --- |
| Relevance of Questions | % of questions aligned with dataset purpose | ≥ 80% | Lecturer review and approval |
| Question Diversity | % of dataset attributes covered by analysis questions | ≥ 70% | Number of unique dataset fields used |
| Team Collaboration Score | Peer evaluation (average score out of 10) | ≥ 7/10 | Team members rate cooperation and idea sharing |
| Lecturer Feedback Score | Lecturer’s rating (score out of 10) | ≥ 7/10 | Direct feedback on question quality |

* 1. Dashboard Phase (Week 3) – 30%

| KPI | Metric | Target | Evaluation Method |
| --- | --- | --- | --- |
| Dashboard Completeness | % of analysis questions answered through visualizations | ≥ 90% | Cross-check dashboard vs. proposed questions |
| Functionality Score | Lecturer’s rating (score out of 10) | ≥ 8/10 | Evaluates interactivity, usability, and filters |
| Team Efficiency Score | Avg. time per dashboard task (hrs) | ≤ 3 hrs | Tracked per member’s contributions |
| Peer Support Score | Peer evaluation (average score out of 10) | ≥ 7/10 | Team members rate each other’s supportiveness |
| Lecturer’s Overall Dashboard Score | Lecturer’s rating (score out of 10) | ≥ 7/10 | Overall assessment of dashboard design |

* 1. Final Presentation & Report (Week 4) – 25%

| KPI | Metric | Target | Evaluation Method |
| --- | --- | --- | --- |
| Report Completeness | % of required sections covered | 100% | Lecturer review based on report structure |
| Presentation Engagement | Lecturer & peer rating (score out of 10) | ≥ 7/10 | Based on clarity, audience engagement, and delivery |
| Teamwork Balance Score | Peer evaluation (average score out of 10) | ≥ 7/10 | Team members rate fairness of workload distribution |
| Learning Outcome Score | Self-assessment (score out of 10) | ≥ 7/10 | Team members rate their own learning experience |
| Lecturer’s Final Evaluation | Lecturer’s rating (score out of 10) | ≥ 7/10 | Overall assessment of project effort and quality |

# Requirements Gathering

## Stakeholder Analysis

1. Project Sponsor (The One Who Needs the Analysis)

* What They Need:
  + Clean, structured data that makes sense.
  + Insights that help with decision-making (not just numbers on a screen).
  + A dashboard that’s easy to navigate, not a mess of charts.
  + A final report that’s straight to the point with useful takeaways.
* How We Deliver:
  + Proper data cleaning to remove errors and inconsistencies.
  + Well-thought-out analysis questions that lead to real insights.
  + A dashboard that’s clear, interactive, and doesn’t require a manual.
  + A report that highlights key findings without unnecessary fluff.

1. Data Analysts & Project Team (Us, the Ones Doing the Work)

* What We Need:
  + A clear project scope—what are we solving?
  + Good-quality data that doesn’t require hours of fixing.
  + The right tools (Power BI) to get the job done efficiently.
  + A structured approach so the work is organized, not chaotic.
* How We Make It Happen:
  + Define analysis goals from the start—no wasted effort.
  + Keep track of cleaning and transformation steps for transparency.
  + Use Power BI to automate and visualize insights instead of manual work.
  + Split tasks logically so everything runs smoothly.

1. End Users (The Ones Who Use the Dashboard & Insights)

* What They Need:
  + A dashboard that’s actually useful, not overwhelming.
  + Quick access to key trends and important stats.
  + Filters and drill-down options to explore different angles.
  + Clear recommendations, not just data dumps.
* How We Make It Work for Them:
  + Keep the design clean, intuitive, and easy to navigate.
  + Highlight key metrics upfront—no need to dig for answers.
  + Add filters so users can get the exact data they need.
  + Summarize key takeaways so insights are actionable.

1. Business Decision Makers (Railway Operators, Transport Planners, etc.)

* What They Need:
  + Insights on passenger trends, peak times, and demand patterns.
  + A way to track train service performance (delays, efficiency, etc.).
  + Data-driven recommendations to improve operations.
  + A summary that’s easy to digest without deep analysis.
* How We Help Them:
  + Structure the analysis around real-world decision-making needs.
  + Include trend analysis and performance metrics in the dashboard.
  + Present findings in a way that helps them make better calls.
  + Offer scenario-based insights for potential improvements.

## User Stories & Use Cases

* 1. Project Sponsor (The One Who Needs the Analysis)
* User Story:

As a project sponsor, I want a cleaned dataset and an easy-to-use dashboard so I can quickly understand trends in UK train rides and make informed decisions.

* Use Case:
  + Opens the dashboard to review key metrics on train performance and passenger trends.
  + Uses insights to assess service efficiency and identify areas for improvement.
  + Prepares reports for stakeholders using key findings from the dashboard.
  1. End Users (The People Using the Dashboard & Insights)
* User Story:

As an end user, I want an intuitive dashboard with key train ride metrics so I can quickly find the data I need without technical expertise.

* Use Case:
  + Accesses the dashboard to check real-time or historical train data.
  + Uses filters to analyze trends for specific routes, stations, or time periods.
  + Downloads or shares relevant reports for internal use.
  1. Business Decision Makers (Railway Operators, Transport Planners, etc.)
* User Story:

As a transport planner, I want to analyze passenger trends and train performance so I can optimize schedules and improve service reliability.

* Use Case:
  + Reviews peak travel times and demand patterns to adjust scheduling.
  + Identifies routes with frequent delays and investigates causes.
  + Uses predictive trends to plan future infrastructure improvements.

## Functional Requirements

* 1. Data Cleaning & Preprocessing
* Ensure all Transaction IDs are unique and valid.
* Standardize dates and times (e.g., Date of Purchase, Departure Time, Arrival Time) to avoid inconsistencies.
* Handle missing or incorrect values in fields like Reason for Delay and Refund Request.
* Ensure Ticket Class, Ticket Type, and Payment Method categories are correctly labeled.
* Remove duplicates and fix any formatting issues in station names (Departure Station, Arrival Destination).
  1. Data Analysis & Metrics

The system must calculate key insights based on the available data:

* Sales & Revenue Trends:
  + Total ticket sales over time (based on Date of Purchase).
  + Average ticket price by Purchase Type, Ticket Class, and Railcard.
  + Most common Payment Methods used.
* Passenger Travel Patterns:
  + Most popular Departure Stations and Arrival Destinations.
  + Peak travel hours based on Time of Purchase and Departure Time.
  + Preferred Ticket Types (e.g., single, return).
* Punctuality & Service Performance:
  + Average Journey Status breakdown (on-time vs. delayed).
  + Delay analysis using Actual Arrival Time vs. Scheduled Arrival Time.
  + Most common Reasons for Delay.
* Refund & Customer Satisfaction:
  + Percentage of trips with Refund Requests.
  + Correlation between delays and refund claims.
  1. Interactive Dashboard
* The dashboard must allow users to:
  + View key stats at a glance (sales trends, travel patterns, performance metrics).
  + Filter data by date range, departure station, ticket type, and more.
  + Compare time periods (e.g., this month vs. last month’s ticket sales).
  + Drill down into specific transactions (e.g., check purchase details for a certain day).
* Must include:
  + Clear charts & trendlines (sales growth, travel demand spikes, etc.).
  + Heatmaps for peak travel times based on Departure Time.
  + Delay analysis breakdown by route and Reason for Delay.
  1. Reporting & Exporting
* Users must be able to download reports with summarized insights.
* Exportable charts and tables for use in presentations.
* Ability to extract refund request summaries for further analysis.
  1. User Experience
* Simple and intuitive design—no unnecessary complexity.
* Fast filtering and loading speeds even with large datasets.
* Clear tooltips & labels explaining what each metric means.

## Non-Functional Requirements

* 1. Performance (Speed & Efficiency)
* The system should load quickly without noticeable delays.
* It should handle large amounts of data without slowing down or freezing.
* Interactions like clicking, filtering, and navigating should feel responsive and smooth.
  1. Security (Keeping Data Safe)
* Only the right people should have access to the project files and dashboard.
* Data should be protected from accidental loss or unwanted changes.
* Any stored information should be kept private and not shared without permission.
  1. Usability (Easy to Use)
* The system should be clear and simple so users can understand it without training.
* Labels and instructions should help users navigate the dashboard without confusion.
* Everything should be organized logically so users can find what they need quickly.
  1. Reliability (Always Works as Expected)
* The system should work consistently without frequent errors or crashes.
* If something goes wrong, there should be a way to fix issues without losing important work.
* The system should be able to handle different types of user inputs without breaking.

# System Analysis & Design

## Problem Statement & Objectives

Train operators and transport planners need data-driven insights to improve service efficiency, track delays, and optimize schedules. This project provides a cleaned dataset and an interactive Power BI dashboard to analyze sales trends, passenger behavior, and service performance.

1. Use Case DiagramA diagram of a project

   AI-generated content may be incorrect.

The system has three key actors:

* Project Sponsor reviews high-level metrics and delay reports to assess service efficiency.
* End Users use the dashboard to check train data, filter trends, and generate reports for internal use.
* Business Decision Makers analyze trends, identify frequent delays, and adjust schedules to improve operations.

The core use cases include:

* Viewing key metrics is relevant for all actors, as they need a broad overview of performance.
* Analyzing trends helps both end users and decision-makers track patterns in passenger behavior and delays.
* Generating reports is a shared function across all actors, as each needs insights for their specific roles.
* Filtering data by routes and dates is mainly for end users and business decision-makers who require specific analysis.
* Checking real-time data is only relevant for end users managing daily operations.
* Identifying frequent delays is a key function for business decision-makers and project sponsors monitoring service performance.
* Optimizing schedules is exclusive to business decision-makers who use the data to improve timetables.
* Investigating reasons for delays is also specific to business decision-makers who need to determine causes and implement solutions.

1. Functional & Non-Functional Requirements

* Functional Requirements:
  + The system must clean, preprocess, and validate train ticket data to ensure accuracy.
  + Users must be able to filter, view trends, and generate reports using the Power BI dashboard.
  + Performance tracking features must include analysis of delays, refund requests, and ticket sales.
* Non-Functional Requirements:
  + The dashboard must load insights quickly to support real-time decision-making.
  + The user interface should be simple and intuitive, ensuring accessibility for all users.
  + Data reliability must be maintained, ensuring consistency across reports.
  + Security should be implemented with restricted access based on user roles.

1. Software

* The system is structured for efficiency using:
* Power BI for data visualization and interactive dashboards.
* Excel/CSV for raw data input and preprocessing before analysis.
* Cloud Storage (if needed) for scalability and data backup.

# Key Business Questions for Dashboard Design

**General Questions**

* What is the total number of trips made?
* What is the total revenue generated from ticket sales?
* What is the average ticket revenue?
* How many routes are captured in the dataset?

**Revenue Related Questions**

* Which railcard types generate the most revenue?
* Which stations contribute the most to refund revenue?
* Which departure stations have the highest total and refund revenue?

**Stations & Time Related Questions**

* What are the peak travel hours based on trip volume?
* What is the on-time rate compared to the ideal target (100%)?
* Which are the busiest departure stations?
* Which routes have the highest demand?
* How is trip distribution spread across AM/PM time blocks and days of the week?

**Passengers Related Questions**

* How does travel volume differ between weekdays and weekends by hour?
* What are the trip patterns by day of the week?
* Which railcards are most frequently used?
* What is the most popular ticket class among passengers?
* How do payment methods vary by purchase type?

**Refund & Delay Related Questions**

* What are the most common reasons for delays and cancellations?
* What percentage of passengers requested refunds?
* What proportion of trips were on time, delayed, or cancelled?

# Technical Documentation

## Data Loading and Structure

The original dataset consists of the following columns:

* Transaction ID
* Date of Purchase
* Time of Purchase
* Purchase Type: Online, Station
* Payment Method: Contactless, Credit Card, Debit Card
* Railcard: Adult, None, Disabled, Senior
* Ticket Class: Standard, First Class
* Ticket Type: Advance, Off-Peak, Anytime
* Price
* Departure Station
* Arrival Destination
* Date of Journey
* Departure Time
* Arrival Time
* Actual Arrival Time
* Journey Status: On Time, Delayed, Cancelled
* Reason for Delay: Signal Failure, Weather Conditions, Traffic, Technical Issue, Staff Shortage
* Refund Request: Yes, No

Additional calculated column in Power Query:

* Route: Combination of Departure Station and Arrival Destination using Text.Combine({[Departure Station], " \_ ", [Arrival Destination]})

## Measures and Calculated Fields

Key calculated fields and measures used in the model:

* Total Trips
  + DAX: COUNT(railway[Transaction ID])
* 0
  + DAX: 0 = 0
* 1
  + DAX: 1 = 1
* Average Revenue
  + DAX: AVERAGE(railway[Price])
* Total Revenue
  + DAX: SUM(railway[Price])
* Cancelled
  + DAX: CALCULATE([Total Trips], railway[Journey Status] = "Cancelled")
* Delayed
  + DAX: CALCULATE([Total Trips], railway[Journey Status] = "Delayed")
* On\_Time
  + DAX: CALCULATE([Total Trips], railway[Journey Status] = "On Time")
* Refund\_yes
  + DAX:

CALCULATE(

COUNTROWS(railway),

FILTER(railway, railway[Refund Request] = "Yes")

)

* Total Refund Request
  + DAX:

CALCULATE(

[Total Trips],

railway[Refund Request] = "Yes"

)

* Refund Revenue
  + DAX:

CALCULATE(

[Total Revenue],

FILTER(railway, [Total Refund Request])

)

* Month
  + DAX: FORMAT([Date of Purchase], "mmmm")
* weekDay
  + DAX: FORMAT(railway[Date of Journey], "dddd")
* Weekday/Weekend
  + DAX:

IF(

railway[weekDay] IN {"Saturday", "Sunday"},

"Weekend",

"Weekday"

)

* Time\_zone\_AM/PM
  + DAX: FORMAT(railway[Departure Time], "hh AM/PM")
* Day\_Hours
  + DAX:

VAR HourText = LEFT([Time\_zone\_AM/PM], 2)

VAR Period = RIGHT([Time\_zone\_AM/PM], 2)

VAR HourNum =

SWITCH(

TRUE(),

Period = "AM" && HourText = "12", 0,

Period = "AM", VALUE(HourText),

Period = "PM" && HourText = "12", 12,

Period = "PM", VALUE(HourText) + 12

)

RETURN HourNum

* delay\_time
  + DAX:

DATEDIFF(

railway[Arrival Time],

railway[Actual Arrival Time],

MINUTE

)

* cancel\_rate
  + DAX:

VAR cancelcount = COUNTROWS(

FILTER(railway, railway[Journey Status] = "Cancelled")

)

VAR totalcount = COUNTROWS(railway)

RETURN DIVIDE(cancelcount, totalcount, "0.00%")

* delay\_rate
  + DAX:

VAR delaycount = COUNTROWS(

FILTER(railway, railway[Journey Status] = "Delayed")

)

VAR totalcount = COUNTROWS(railway)

RETURN DIVIDE(delaycount, totalcount, "0.00%")

* on\_time\_rate
  + DAX:

VAR ontimecount = COUNTROWS(

FILTER(railway, railway[Journey Status] = "On Time")

)

VAR totalcount = COUNTROWS(railway)

RETURN DIVIDE(ontimecount, totalcount, "0.00%")

## Report Pages

The dashboard includes the following pages:

* Homepage
* Revenue
* Stations & Time
* Passengers
* Refund & Delay

## Slicers

All pages except the homepage contain the following slicers:

* Weekday/Weekend
  + Field: Weekday/Weekend (measure)
  + Hierarchy:
    - Weekday: Monday, Tuesday, Wednesday, Thursday, Friday
    - Weekend: Saturday, Sunday
* Payment Method
  + Field: Payment Method (original column)
  + Values:
    - Contactless
    - Credit Card
    - Debit Card
* Ticket Type
  + Field: Ticket Type (original column)
  + Values:
    - Advance
    - Anytime
    - Off-Peak
* Ticket Class
  + Field: Ticket Class (original column)
  + Values:
    - First Class
    - Standard
* Month
  + Field: Month (measure)
  + Values:
    - December
    - January
    - February
    - March
    - April

## Cards

All pages except the homepage contain the following cards:

* Passengers
  + Fields: Total Trips (measure)
* Routes
  + Fields: Count of Route (original column)
* Refund
  + Fields: Refund Revenue (measure)
* AVG-Revenue
  + Fields: Average Revenue (measure)
* Revenue
  + Fields: Total Revenue (measure)
* Cancelled
  + Fields: Cancelled (measure)
* Delay
  + Fields: Delayed (measure)
* On Time
  + Fields: On\_Time (measure)

## Page-Level Visuals

**Visuals in Revenue Page**

* Line Chart: Total Revenue Over Time
  + X-Axis: Date of Journey (original column)
  + Y-Axis: Total Revenue (measure)
* Tree Map: Total Revenue by Railcard
  + Category: Railcard (original column)
  + Values: Total Revenue (measure)
* Decomposition Tree: Refund Revenue Analysis
  + Analyze: Refund Revenue (measure)
  + Explain by:
    - Ticket Class (original column)
    - Ticket Type (original column)
* Pie Chart: Total Revenue by Purchase Type
  + Legend: Purchase Type (original column)
  + Values: Total Revenue (measure)
* Table: Revenue by Departure Station
  + Columns:
    - Departure Station (original column)
    - Total Revenue (measure)
    - Refund Revenue (measure)

**Visuals in Stations & Time Page**

* Area Chart: Peak and Idle Hours
  + X-Axis: Day\_Hours (measure)
  + Y-Axis: Total Trips (measure)
* Gauge: On Time Gauge
  + Value: on\_time\_rate (measure)
  + Minimum Value: 0 (measure)
  + Maximum Value: 1 (measure)
  + Target Value: 1 (measure)
* Stacked Column Chart: Busiest Stations
  + X-Axis: Departure Station (original column)
  + Y-Axis: Total Trips (measure)
  + Filters:
    - Departure Station
    - Top N: Top 5
    - Count of Departure Station
* Stacked Column Chart: Busiest Routes
  + X-Axis: Route (original column)
  + Y-Axis: Total Trips (measure)
  + Filters:
    - Route
    - Top N: Top 5
    - Total Trips
* Matrix: Trips Heat Distribution
  + Rows: Day\_Hours (measure)
  + Columns: weekDay (measure)
  + Values: Total Trips (measure)
* Map: Departure Station Distribution
  + Location: Departure Station (original column)
  + Bubble Size: Count of Departure Station (original column)

**Visuals in Passengers Page**

* Area Chart: Weekdays vs Weekends
  + X-Axis: Day\_Hours (measure)
  + Y-Axis: Total Trips (measure)
  + Legend: Weekday/Weekend (measure)
* Line Chart: Trips by Weekday
  + X-Axis: weekDay (measure)
  + Y-Axis: Total Trips (measure)
* Pie Chart: Railcard Usage
  + Legend: Railcard (original column)
  + Values: Total Trips (measure)
* Pie Chart: Ticket Class Distribution
  + Legend: Ticket Class (original column)
  + Values: Total Trips (measure)
* Stacked Bar Chart: Payment Method by Purchase Type
  + Y-Axis: Payment Method (original column)
  + X-Axis: Total Trips (measure)
  + Legend: Purchase Type (original column)

**Visuals in Refund & Delay Page**

* Stacked Area Chart: Arrival vs Actual Arrival Time
  + X-Axis: Day\_Hours (measure)
  + Y-Axis:
    - Count of Arrival Time (original column)
    - Count of Actual Arrival Time (original column)
* Clustered Column Chart: Reason for Delay and Cancellation
  + X-Axis: Reason for Delay (original column)
  + Y-Axis:
    - Delayed (measure)
    - Cancelled (original column)
* Pie Chart: Refund Request
  + Legend: Refund Request (original column)
  + Values: Count of Refund Request (original column)
* Pie Chart: Journey Status
  + Legend: Journey Status (original column)
  + Values: Total Trips (measure)
* Line and Clustered Column Chart: Refund Status Over Time
  + X-Axis: Date of Journey (original column)
  + Column Y-Axis:
    - Delayed (measure)
    - Cancelled (measure)
  + Line Y-Axis: Refund\_yes (measure)

# User Manual

**Project Sponsor**

* What they need
  + The project sponsor wants to quickly understand trends in train performance and make decisions based on key metrics like revenue, passenger activity, delays, and refunds.
* How they use the dashboard
  + Start with the Homepage
    - Use the bookmarks to navigate to other pages like Revenue, Stations and Time, Passengers, or Refund and Delay.
  + Go to the Revenue page
    - Check the line chart to see how revenue changes over time.
    - Look at the tree map to find out which Railcard types bring in the most revenue.
    - Use the pie chart to compare online and station purchases.
    - Use the decomposition tree to break down what’s causing high refund amounts.
  + Use filters
    - Select a specific month to focus on.
    - Filter by ticket type, payment method, or class to narrow down results.
  + Export visuals for reports
    - Right-click any visual and choose export or show as table to use the data in reports.

**End User**

* What they need
  + End users need quick and simple access to train data, with clear visuals and filters they can use without any technical knowledge.
* How they use the dashboard
  + Start at the Homepage
    - Use the bookmarks to open the section they need, such as Passengers or Stations and Time.
  + Go to the Passengers page
    - Look at the railcard pie chart to understand rider types.
    - Use the stacked bar chart to compare payment methods with purchase types.
    - View weekday versus weekend travel patterns using the area and line charts.
  + Use filters
    - Choose filters like payment method, ticket type and month to focus the view.
    - Try switching between weekday and weekend to see different usage patterns.
  + Check the Stations and Time page
    - Use the heatmap to see when travel peaks by day and hour.
    - View the map to understand where people are departing from the most.
    - Use the busiest routes chart to find the most used travel paths.
  + Export data
    - Right-click charts and export data if needed for internal updates.

**Business Decision Makers**

* What they need
  + These users need deeper insights into performance issues, delays, passenger demand, and refund trends to plan and improve train services.
* How they use the dashboard
  + Use the Homepage page to access Refund and Delay
    - The stacked area chart compares scheduled and actual arrival times.
    - The column chart shows top reasons for delays and cancellations.
    - Pie charts show how many trips were delayed or refunded.
  + Go to the Stations and Time page
    - Check the gauge chart to see the overall on-time performance.
    - Use the charts for busiest stations and routes to see where demand is highest.
    - The heatmap helps find peak hours for planning.
  + Use the Refund and Delay page
    - View the combined chart showing delays, cancellations, and refund requests over time.
    - This helps connect delays to customer complaints and costs.
  + Use filters to explore scenarios
    - Select certain ticket classes or travel days to see how different groups are affected.
    - Look at specific months to understand seasonal changes in demand or performance.

# Insights

**Performance Overview**

* 32,000 passengers took trips across 65 different routes.
* 86.82% of trips were on time, 7.24% delayed, and 5.94% cancelled.
  + Overall good, but about 1 in 7 trips had some kind of issue.

**Revenue**

* Total revenue: £741.92K | Avg. per ticket: £23.44
* Refunds cost around £38.7K (about 5.2% of total revenue).
* Standard class made up over 90% of sales. First class is a small share, but higher value per ticket.
* 66% of passengers didn’t use any railcard.
* Most used railcard: Adult.
  + There’s room to promote railcards better.
* Sales split: 51.59% online, 48.41% at stations.
  + It shows good adoption of online channels.
* Stations with the most refund revenue: Liverpool Lime Street and Manchester Piccadilly.
  + It might be worth looking into what’s causing issues there.

**Passenger Trends**

* Peak travel hours are 6 AM and 6 PM, typical commuter times.
* Midday travel is steady but not as high.
* Weekdays are busier than weekends, especially Wednesdays.
  + Still decent weekend activity, not a sharp drop.
* Standard class is by far the most popular.
* Payment methods: mostly credit card, then contactless, then debit.

**Stations & Time**

* Top stations: Manchester Piccadilly, London Euston, Liverpool Lime Street.
* Busiest route: Manchester Piccadilly - Liverpool Lime Street. London routes also show high demand.
* Heatmap shows steady demand across weekdays.
* Travel drops slightly on weekends but stays active.

**Delays & Refunds**

* 3.53% of trips ended with a refund request.
* Most common delay reasons: technical issues, signal failure, and bad weather.
* Delays spike between 8 AM and 11 AM.

# Actions and Recommendations

## General Actions

**Promote Railcards:**

* Push Adult, Disabled, and Senior railcards more.
* Target full-fare passengers (66% don’t use any railcard).
* Add new railcards to cover other groups not served by current options.

**Fix Refund Issues at Key Stations:**

* Liverpool Lime Street and Manchester Piccadilly have the most refund revenue.
* Look into what’s causing problems there, service, delays, or refund process.

**Cut Down Morning Delays:**

* Most delays happen between 8 AM and 11 AM.
* Focus on fixing technical issues, staff shortages, and any process slowdowns during this time.
* Add more trains during peak hours to manage demand and reduce delays, if supported by cost-benefit analysis.

**Get More from First-Class:**

* First-class tickets are low in volume but likely higher value.
* Explore ways to make first-class more appealing or available.

**Improve Communication Delay:**

* Make real-time info during delays more accessible.
* Especially important during morning rush hours.

## RCA (Manchester - Liverpool Route)

**Problem:**

* Delays on the busiest route ,especially 8 AM to 11 AM.

**Root Causes:**

* There is too much pressure during peak hours.
* Not enough staff or poor scheduling.
* Technical issues cause slow turnarounds.

## PDCA Plan (Manchester - Liverpool Route)

**Plan:**

* Audit the causes of morning delays on this route.
* Optimize schedules and assign more staff in the early shift.
* Improve communication during delays.

**Do:**

* Roll out small operational fixes: better crew rotation, maintenance before peak time.
* Update signage and announcements to keep passengers informed.

**Check:**

* Track delay times and number of complaints during 8 AM–11 AM.
* Collect rider feedback specifically for this route and time window.

**Act:**

* If on-time rates improve, keep the changes and apply similar fixes to other high-volume routes.
* If not, adjust based on real-time data and feedback.

## RCA (General Delays & Refunds)

**Problem:**

* Too many delays and refunds system-wide, especially in the morning.

**Root Causes:**

* Delays mostly between 8 AM–11 AM (technical faults, low staff, weather).
* Refund spikes tied to problem stations and delayed services.
* System not built to handle volume well during peaks.

## DMAIC Plan (General Delays & Refunds)

**Define:**

* Goal: Reduce morning delays and lower the refund rate overall.
* Focus on 8 AM–11 AM across all routes.

**Measure:**

* Track number of delays and refunds by hour, route, and station.
* Monitor on-time rates and refund request volumes.

**Analyze:**

* Spot patterns, which stations, which times, what issues (tech, staffing).
* Look into the link between delays and refund surges.

**Improve:**

* Boost early staffing — ensure adequate crew during morning peak times.
* Better maintenance checks before morning shifts to avoid technical failures.
* Streamline ticket handling and communication during delays to keep customers informed and reduce frustration.

**Control:**

* Set a weekly delay/refund check-in routine.
* Use dashboard tracking for delays by hour and station.
* Keep collecting feedback to fine-tune what’s working.

## PDCA Plan (First Class Ticket Usage)

**Plan:**

* Check what First Class offers now, price, seats, perks.
* See what’s missing or why people avoid it.
* Plan a small push to show off the benefits (comfort, space, quiet).

**Do:**

* Run a promo or discount on busy routes.
* Update posters, emails, and app banners to promote it.
* Let staff offer upgrades at stations if seats are free.

**Check:**

* Track how many more people book First Class during the promo.
* Ask those who tried it if they’d pay full price next time.

**Act:**

* If it works, keep doing it and build on it.
* If not, look at changing the offer, price, perks, or how it’s marketed.