**Final Project – Proposal Submission**

**Project Title**: UK Train Rides Analysis

**Project Description**:

This project aims to analyze UK train ride data to extract valuable insights on trip performance, peak times, train delays, and key factors affecting service quality. We used data analysis and Power BI to create interactive dashboards that reflect critical performance indicators.

**Group Members & Roles**:

**Omar Ali**– Operation Operational Performance & Efficiency dashboard, preprocessing & cleaning, data modelling, business questions that was answered through the dashboard.

**Ahmed Ali**– Implementing interactive buttons within the dashboard and make sure that buttons perform the right actions and integrate smoothly with the dashboard's features and contribute to business questions that was answered through the dashboard.

**Basma Saber**– Customer satisfaction dashboard, data modelling, business questions that was answered through the dashboard.

**Fatma Abu Elfadl**–Responsible for writing and structuring the report and contribute to data modelling, business questions that was answered through the dashboard.

**Areej Moro**– Overview dashboard, data modelling, business questions that was answered through the dashboard.

**Shahd Saad**– Customer behavior dashboard, data modelling, business questions that was answered through the dashboard.

**Team Leader:** Omar Ali

**Objectives:**

1.Analyze train ride frequencies and identify peak travel times.

2.Examine delay patterns and potential causes.

3.Assess ticket pricing trends based on distance, time, and demand.

4.Identify the busiest routes and stations.

5.Provide recommendations for improving train service efficiency.

**Tools & Technologies:**

Power BI for data modeling and analysis and interactive reporting.

**Milestones & Deadlines:**

Data Collection & Cleaning – Week1

Data Analysis & KPI Extraction – Week2

Dashboard Development – Week3

Testing & Feedback Collection – Week4

**KPIs (Key Performance Indicators):**

**1. Data Quality Score (Accuracy & Completeness)**

After cleaning the dataset, it was found that 7% of records had missing values, especially in reason for delay record.

We used interpolation to estimate missing values to reduce missing or incorrect data to below 3% to ensure high data accuracy.

**2. Dashboard Performance (Load Time)**

Initial dashboard load time was 10 seconds due to large datasets and complex calculations.

We used aggregations, optimized DAX queries, and filtered unnecessary data to improve performance and keep load time under 6 seconds for a smoother user experience.

**3. Number of Key Insights Generated**

Identified key insights, including:

1.What is the average delay time per journey?

2.Which time slots (morning, afternoon, night) have the most delays?

3.Which stations experience the highest number of delays?

4.What percentage of delays exceed 30 minutes? 1 hour?

5.Are delays more common during certain seasons or months?

6.What percentage of delayed journeys result in missed connections (if data available)?

7.How does the actual arrival time compare to the scheduled arrival time for

different routes?

8.How often do delayed journeys lead to refund requests?

9.Are refunds more common for certain ticket types or payment methods?

10.What is the percentage of refund requests that get approved?

11.Do customers who experience delays tend to avoid booking certain routes later?

12.Are there specific routes where customers request refunds more frequently?

13.How much revenue is lost due to refunds?

14.Is there a correlation between refund requests and ticket class?

15.How has buy tickets changed over time (daily, weekly, monthly trends)?

16. What is the distribution of ticket prices? Are there peak price periods?

17. What is the average spending per passenger?

18. Which payment method is used the most, and does it impact ticket price or frequency of purchase?

19. Is there a correlation between ticket price and journey duration?

20.What is the most common journey length (short vs. long-distance trips)?

**4. Distribution & Access Report**

We implemented scheduled data refreshes to keep information up to date.

Goal: Maintain 100% accessibility and ensure reports update automatically without manual intervention.

**5. Visualization Effectiveness (Readability Score)**

We adjusted the color scheme and added data labels to enhance readability to ensure that all charts and KPIs are easy to understand at a glance, improving decision-making efficiency.