Data Analysis Project Technical and Financial Proposal

1. Project Title

UK Train Rides Data Analysis

2. Introduction

This project aims to analyze the United Kingdom train rides dataset to uncover insights about travel patterns, station performance, and seasonal trends.

It is conducted by a team of DEPI students, led by Mohamed Mayhoub, specializing in data analysis and visualization.

Through a combination of SQL queries, Python libraries (Pandas, Matplotlib), and Power BI. The project transforms raw transportation data into valuable, interpretable insights that can support decision-making in transport management.

3. Project Objective

i. **Clean and prepare the raw dataset for analysis.**
ii. **Perform exploratory data analysis (EDA) in hopes to discover trends, outliers, and meaningful metrics**.
iii. **Apply SQL for querying and aggregating passenger and route data.**
iv. **Use Python for visualization and statistical analysis**
v. **Present insights through Power BI interactive charts and structured reports.**

4 .Work in Scope

- **Data Preparation & Cleaning: Removing duplicates, handling missing data, and normalizing columns**.
- **Exploratory Data Analysis: Generate summary statistics, find the sales patterns, and partition data**.
- **Visualization: Make bar charts, line graphs, and heat maps showing significant results**.
- **Reporting: Present results in a report and deliver charts in Power BI formats**.

- **Review Meeting: Present results to the instructor.**

5. Deliverables

- ✓ **Cleaned dataset in Excel or CSV format**
- ✓ **Interactive Data Visualization (Power BI)**
- ✓ **Detailed Presentation report.**

6. Timeline (The real time of actionable work)

Task	Duration
Data Cleaning & Preparation	3 days
SQL Query Development	2 days
Exploratory Data Analysis (Python)	3 days
Visualization & Reporting	2 days
Review & Final Submission	2 days
Total Project Time	**12 days**

7. Tools and Technologies

Category	Tool/Library
Programming Language	Python
Data Processing	Panadas & NumPy
Visualization & Reporting	Matplotlib & Power Bi
Version Control	GitHub

8. Expected Outcome.

- ✓ **Identify the busiest stations and routes in the UK train system**

- ✓ **Reveal time-based patterns (e.g., monthly or seasonal peaks)**
- ✓ **Understand the distribution of ride frequencies across regions.**
- ✓ **Create a visual analytical summary supporting future planning decisions.**