**📘 Technical Documentation**

**UK Train RIDS – Q1 2024 Operational Review**

**📍 Project Title:**

**UK Train RIDS – Operational Review (Q1 2024)**

**1. 📌 Project Overview**

This project analyzes the operational performance of the UK train network during the first quarter of 2024 using the RIDS (Railway Information and Data System) platform. The focus is on data-driven insights to assess punctuality, frequency, delays, and service demand using Python-based data analysis and visualization techniques.

**2. 👥 Team Members**

* **Ziad Mohamed**
* **Asem Khaled**
* **Reem Khaled**
* **Ahmed Taha**
* **Ahmed Sherif**
* **YOUSEF ATTIA**

**3. 🧪 Objectives**

* Collect and clean real-time and historical train operational data.
* Analyze service patterns and delay occurrences.
* Build a model to predict operational efficiency.
* Visualize the performance metrics in an interpretable format.

**4. 🔧 Tools & Technologies Used**

| **Category** | **Tools/Libraries** |
| --- | --- |
| Programming Language | Python |
| Data Processing | Pandas, NumPy |
| Visualization | Matplotlib, Seaborn |
| Modeling | Scikit-learn |
| Development Platform | Jupyter Notebook |
| Documentation | MS Word, PowerPoint |

**5. 📥 Data Collection**

* **Source:** UK RIDS public API or CSV logs .
* **Period:** January 1st, 2024 – March 31st, 2024
* **Format:** CSV / JSON
* **Key Fields:** Train ID, Station, Arrival/Departure Time, Delay Duration, Route, Date

**6. 🧼 Data Cleaning & Preprocessing**

* Removed null/missing values.
* Filtered outliers (e.g., unrealistically long delays).
* Standardized timestamp formats.
* Renamed columns for clarity.
* Encoded categorical features where needed.

**7. 📊 Exploratory Data Analysis (EDA)**

* Assessed distribution of delays across days of the week and hours of the day.
* Identified peak usage periods and delay hotspots.
* Found correlations between delay duration and train type or route.

**8. 🔍 Data Modeling**

* **Goal:** Predict likelihood or severity of train delays.
* **Model:** Random Forest Classifier / Regressor (or another algorithm used).
* **Features:** Time, Day, Train Type, Route, Previous Delay.
* **Metrics:** Accuracy, Precision, Recall, RMSE (based on model type).
* **Outcome:** Identified significant features influencing delays.

**9. 📈 Data Visualization**

Key visualizations included:

* Line chart showing number of daily trips and average delay.
* Bar chart comparing delays across different train lines.
* Heatmap for delays by hour and day.
* Pie chart for train types in service.