DEPI Round 3 Graduation Project

UK Train Rides Analysis & Visualization

Team members:

1. Tarek Ahmed Ahmed
2. Mariam Helmy Ahmed
3. Mohamed Saad Zakaria
4. Mora Magdy Eshak
5. Alaa Osama Mohamed

Supervised by:
Eng. Kareem Bakli

Table of Contents:

1. Project Planning & Management

1. Project Planning & Management

Project Proposal

Objective: Analyze UK train operating data to identify performance trends, root causes of delays, and passenger journey impacts using Power BI for data integration, analysis, and visualization.

Business Objectives:

- **Improve Travel Planning: Identify the most reliable train operators, routes, and times for travel.**
- **Understand Delay Causes: Analyze root causes and patterns of delays and cancellations across the network.**
- **Performance Benchmarking: Compare Train Operating Companies (TOCs) against each other and national standards.**
- **Data-Driven Decisions: Provide comprehensive, interactive dashboards for commuters, leisure travelers, and transport analysts.**

Project Plan

Phase	Tasks	Duration	Milestone
1. Data Preparation & Cleaning	Find/download train data files, clean data, fix errors, handle missing values	Week 1-3	Clean, ready-to-use dataset in Power BI
2. Analysis Planning	Define key questions, create initial charts, sketch dashboard mockups	Week 4	Agreed-upon questions and dashboard sketch
3. Dashboard Development	Build charts/graphs, add filters/interactivity, team testing	Week 4-5	Finished, interactive Power BI dashboard

4. Reporting & Presentation	**Write summary report, create presentations, practice delivery**	**Week 6**	**Project completed with report and presentation**

Resources

Power BI, Excel, Data Analysis Tools

Task Assignment & Roles

- **Data Understanding & Structure: Tarek**
- **Data Quality Assessment: Mariam**
- **Data Model Design: Mora**
- **Data Cleaning & Transformation: Mohamed**
- **KPI Calculations: Alaa**
- **Dashboard Development: Team collaboration**
- **Reporting & Presentation: All team members**

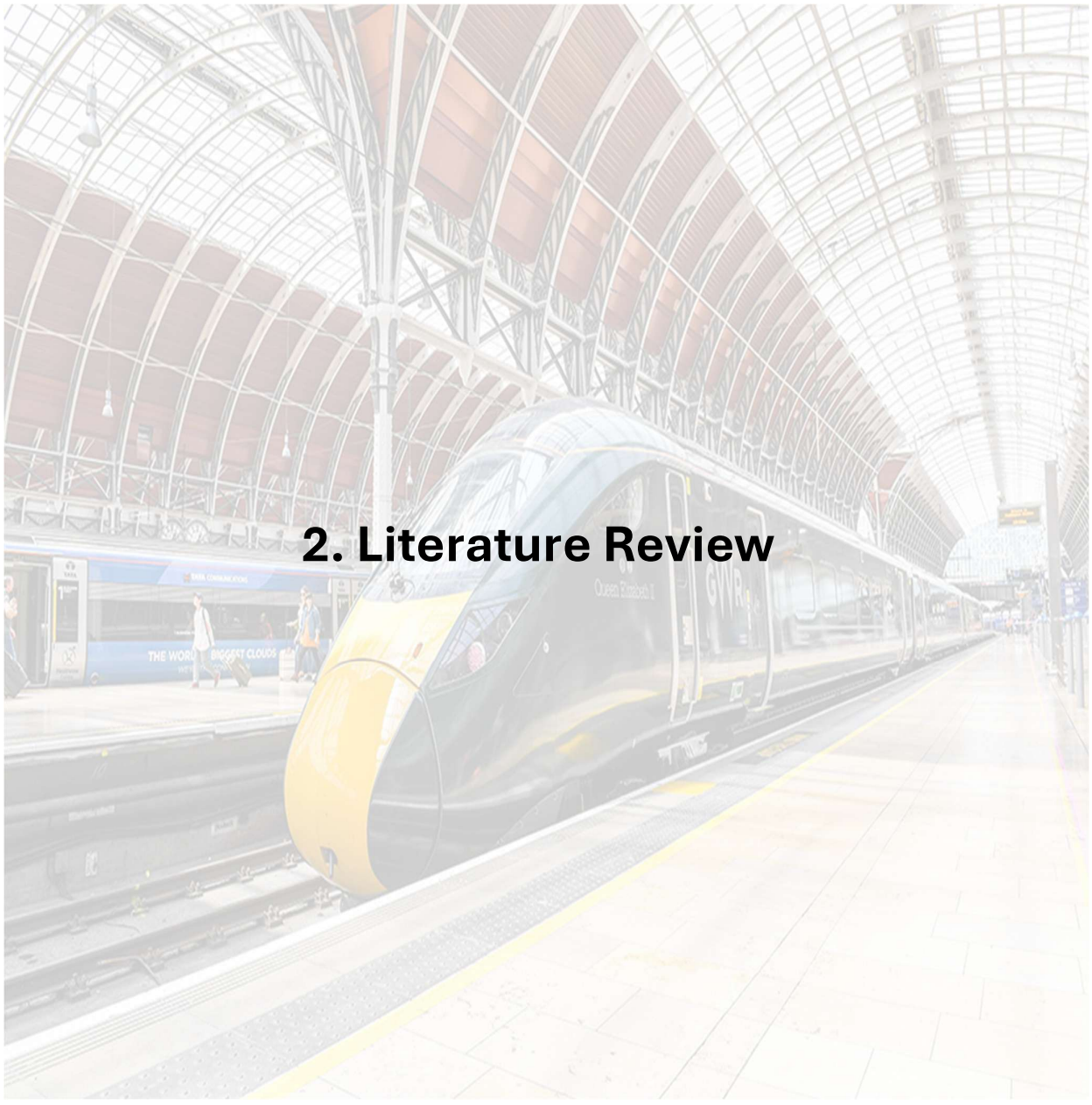
Risk Assessment & Mitigation Plan

- **Risk: Data inconsistency and missing values → Solution: Data validation techniques and conditional calculations**
- **Risk: Performance issues with large datasets → Solution: Efficient DAX queries and data model optimization**
- **Risk: Inaccurate time calculations → Solution: Adjusted formulas for cross-midnight journeys**

KPIs (Key Performance Indicators)

- **Delay frequency and duration**
- **Cancellation rates**
- **Journey times**
- **Performance by TOC (Train Operating Company)**

- **Route reliability metrics**



2. Literature Review

2. Literature Review

Feedback & Evaluation

Project focuses on practical analysis of transportation data with clear business applications. Strong emphasis on data cleaning and visualization.

Suggested Improvements

- **Expand analysis to include seasonal trends**
- **Add predictive analytics for delay forecasting**
- **Incorporate passenger volume data for impact analysis**

Final Grading Criteria

Based on data accuracy, visualization effectiveness, user interaction, analytical insights, and presentation quality.

`

3. Requirements Gathering

3. Requirements Gathering

Stakeholder Analysis

- **Travelers: Commuters and leisure travelers seeking reliable journey planning**
- **Transport Analysts: Professionals needing performance insights**
- **Train Operating Companies: Management requiring operational benchmarking**
- **Government Agencies: Regulatory bodies monitoring network performance**

User Stories

- **"As a daily commuter, I want to see which train operators have the best punctuality records to choose reliable services."**
- **"As a transport analyst, I need to understand root causes of delays to recommend operational improvements."**
- **"As a TOC manager, I want to compare my company's performance against competitors to identify areas for improvement."**

Functional Requirements

- **Performance analysis by operator, route, and time period**
- **Delay cause analysis and trending**
- **Interactive filtering by date, operator, route**
- **Geographical visualization of performance**
- **Drill-down capabilities to detailed journey data**

Non-functional Requirements

- **Dashboard load time under 5 seconds**
- **Clear, intuitive data visualization**
- **Responsive design for different devices**
- **Accurate and up-to-date information**

4. System Analysis & Design

4. System Analysis & Design

Problem Statement & Objectives

Manual analysis of UK train performance data is time-consuming and lacks interactivity. This project automates the process through Power BI dashboards to provide immediate, actionable insights for various stakeholders.

Use Case Diagram

- **Travelers: View performance dashboards for journey planning**
- **Analysts: Perform detailed analysis and generate reports**
- **TOC Managers: Monitor company performance metrics**
- **Administrators: Manage data updates and system maintenance**

Software Architecture

Power BI dashboard for visualization, file-based data storage (CSV/Excel), with data flowing from source files through Power Query transformation into interactive visuals.

Database Design & Data Modeling

- **ER Diagram: Relationships between Operators, Routes, Journeys, Delays**
- **Logical Schema: Tables for Journey facts with dimension tables for Time, Operators, Routes**
- **Physical Schema: Optimized for Power BI with proper keys and relationships**

Data Flow & System Behavior

- **DFD: Data extraction → Power Query transformation → Power BI visualization**
- **Activity Diagram: Load data → Clean/transform → Calculate metrics → Visualize → Generate insights**

UI/UX Design & Prototyping

- **Wireframes: Dashboard layout with performance overview, operator comparison, delay analysis, and geographical views**
- **UI/UX Guidelines: Professional color scheme, intuitive navigation, clear typography, interactive filtering**

5. Implementation & Deployment

5. Implementation & Deployment

Technology Stack

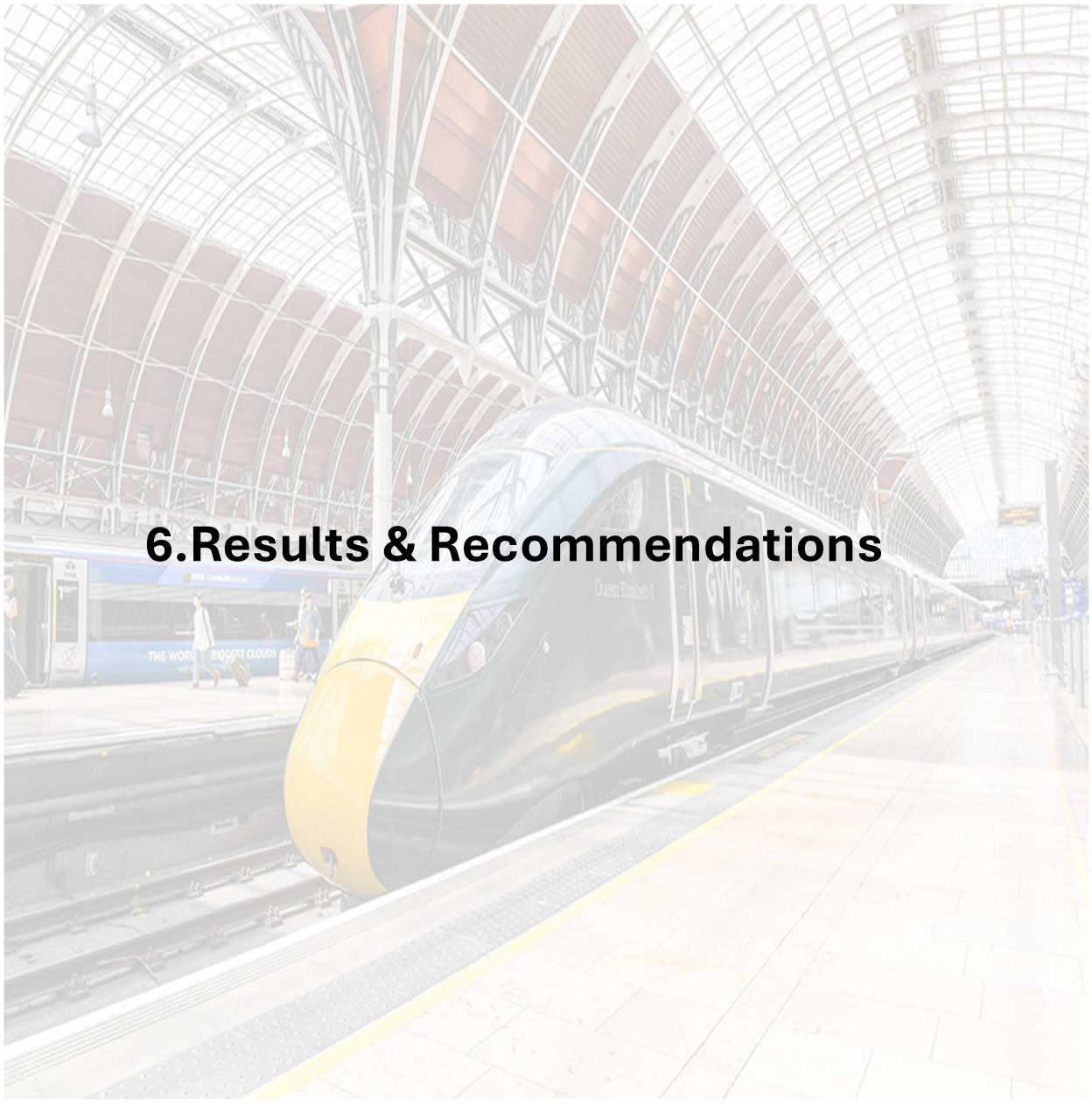
- **Dashboard: Power BI**
- **Data Storage: CSV files**
- **Analysis: Power Query, DAX**
- **Sharing: Power BI Service**

Deployment Strategy

- **Data refresh schedules configured in Power BI Service**
- **Dashboard published to relevant stakeholder groups**
- **Access controls implemented based on user roles**

Key Implementation Challenges & Solutions

1. **Negative Duration Values: Fixed by adjusting time calculation formulas to handle cross-midnight journeys**
2. **Missing Delay Reasons: Addressed by replacing nulls with "No Delay" for on-time journeys**
3. **Inconsistent Data Entries: Standardized delay reason categories (e.g., unified "Signal Failure" variations)**



6.Results & Recommendations

6.Results & Recommendations