INT374:DATA ANALYTICS WITH POWER BI
PROJECT REPORT

(Project Semester August-January 2026)

Railway Delay & Passenger Impact Analysis – 2025

Submitted by:

Aditya Raj

Roll No. 62

Registration No:12305995

Programme and Section: BTech CSE K23KS

Course Code: INT374

Under the Guidance of

Ms. B Monica

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that Aditya Raj, bearing Registration No. 12305995 and Roll No. 62, a student of B.Tech CSE (K23KS), has successfully completed the project titled “Railway Delay & Passenger Impact Analysis – 2025” for the course INT374: Data Analytics with Power BI, under my guidance and supervision.

To the best of my knowledge, the present work is the result of his original effort, development, and study. The project fulfills the requirements prescribed for the award of the course.

Signature and Name of the Supervisor: Ms. B. Monica
School of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab

Date: 16/12/2025

DECLARATION

I, Aditya Raj, a student of B.Tech CSE (K23KS) under the Discipline of Computer Science and Engineering at Lovely Professional University, Punjab, hereby declare that the work presented in this project report entitled "Railway Delay & Passenger Impact Analysis – 2025" is my original work and has not been submitted previously to any other university or institution for any academic purpose.

All sources of information used in this project have been duly acknowledged.

Date: 16/12/2025 Signature

Registration No: 12305995 Aditya Raj

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my project guide, Ms. B. Monica, for her continuous guidance, encouragement, and valuable suggestions throughout the course of this project. Her insights and constructive feedback played a crucial role in shaping the outcome of this work.

I am also thankful to Lovely Professional University for providing the necessary infrastructure, learning resources, and a supportive academic environment that enabled the successful completion of this project. Additionally, I acknowledge the use of open-source datasets, Microsoft Power BI, and GitHub resources, which contributed significantly to the development and analysis presented in this project.

TABLE OF CONTENTS

INTRODUCTION

Indian Railways is one of the largest railway networks in the world, transporting millions of passengers daily across thousands of stations. Due to high traffic volume, technical issues, weather conditions, and operational challenges, train delays are a frequent concern. These delays not only affect operational efficiency but also have a significant impact on passengers in terms of travel inconvenience, missed connections, and dissatisfaction.

With the availability of large-scale railway operational data, data analytics tools can be used to study delay patterns and passenger impact in a structured manner. Microsoft Power BI provides an interactive platform to analyze, visualize, and interpret such complex datasets efficiently.

This project, **"**Railway Delay & Passenger Impact Analysis – 2025**"**, focuses on analyzing train delays across Indian railway stations and understanding their impact on passengers using Power BI dashboards. The objective is to transform raw railway delay data into meaningful insights that support better monitoring and decision-making.

SOURCE OF THE DATASET

The dataset used in this project is **etrain_delays.csv**, which contains structured data related to train operations, delays, stations, and passenger impact across Indian Railways.

Dataset Overview

- Covers railway operations for the year **2025**
- Includes train-level and station-level information
- Contains delay metrics and passenger-related attributes
- Suitable for time-based, geographical, and categorical analysis

DATASET DESCRIPTION

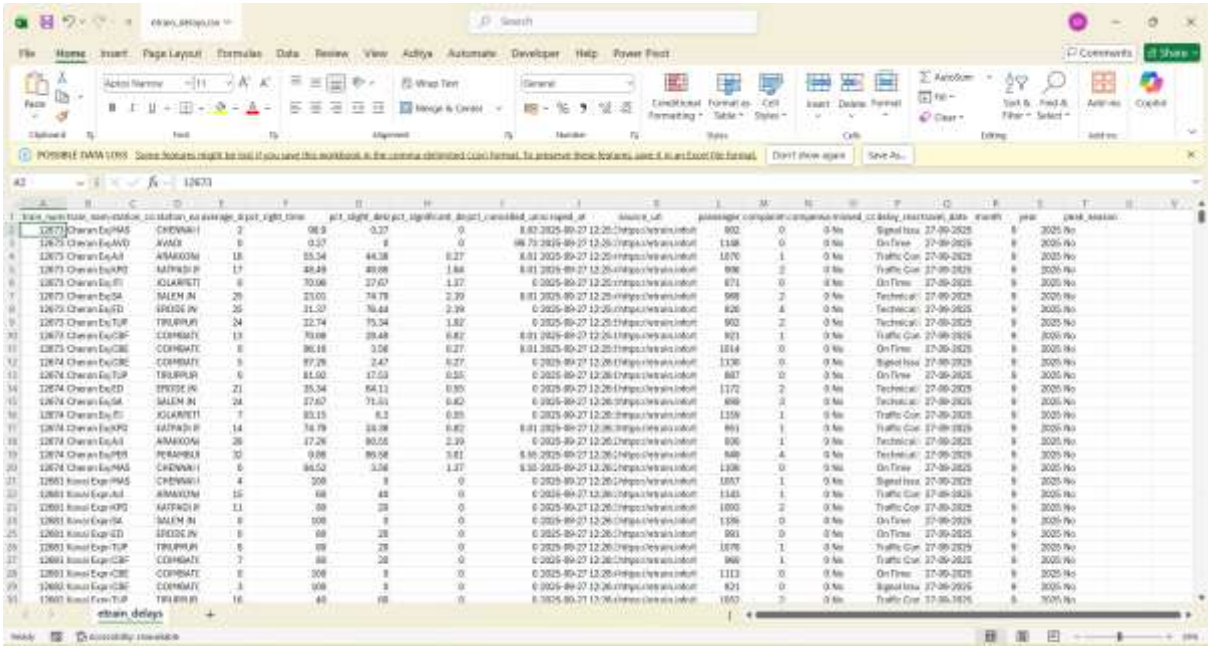
The dataset consists of multiple attributes capturing operational and passenger impact details. Important columns include:

- **Train Number & Train Name** – Identifies each train
- **Station Name** – Railway station where delay is recorded
- **Average Delay (Minutes)** – Mean delay at the station
- **Delay Reason** – Weather, Technical Issue, Traffic Congestion, Signal Issue, On Time
- **Passenger Count** – Number of passengers affected
- **Month & Year** – Time-based attributes for trend analysis

This dataset enables both operational performance analysis and passenger impact assessment.

Dataset Download Link:-

The dataset was downloaded from Kaggle using the following link:
https://www.kaggle.com/datasets/naijilaji/indian-railways-passenger-train-delays-dataset



DATASET PREPROCESSING

Dataset preprocessing is one of the most critical stages in any data analytics project, as the accuracy and reliability of the insights generated depend heavily on the quality of the data used. The raw dataset obtained from the CSV file etrain_delays.csv contained train-level, station-level, delay-related, and passenger-related information. Before performing any analysis or visualization in Power BI, the dataset was carefully cleaned, transformed, and structured using Power Query Editor.

The following preprocessing steps were carried out to prepare the dataset for dashboard development:

Steps Involved in Dataset Preprocessing

1. Data Loading and Inspection

The CSV file was imported into Power BI Desktop. An initial inspection was performed to

understand the structure of the dataset, number of rows and columns, and the nature of each attribute. Columns related to train details, station names, delay values, delay reasons, passenger counts, and time attributes (month and year) were identified as key variables for analysis.

2. Handling Missing and Null Values

The dataset contained missing and null values in delay-related and passenger-related columns. These values were handled using appropriate strategies:

- Numerical delay and passenger fields with missing values were replaced with **zero or average values**, where logically applicable.
- Records with incomplete station or train information were reviewed and removed if they could affect analysis accuracy.
- Delay reasons with missing values were categorized under **"Unknown"** to maintain consistency.

This step ensured that calculations such as averages, totals, and KPIs were not distorted by incomplete data.

3. Removal of Duplicate Records

Duplicate records were identified using key attributes such as train number, station name, date, and delay reason. These duplicates were removed to prevent overcounting in metrics such as:

- Total trains
- Passenger impact
- Average delay values

Removing duplicates improved the reliability of the dashboard insights.

4. Data Type Standardization

To enable accurate calculations and filtering in Power BI, data types were standardized:

- Delay values were converted to decimal or whole number formats.
- Passenger count fields were converted to whole numbers.
- Month and year columns were converted to appropriate date and numeric formats.

- Categorical fields such as station names and delay reasons were set as text data types.

Proper data typing ensured smooth functioning of slicers, charts, and DAX measures.

5. Column Renaming and Data Cleaning

Column names were renamed to make them more readable and meaningful. Unnecessary columns that were not relevant to the objectives of the project were removed, reducing complexity and improving performance.

6. Creation of Calculated Columns and Measures

Several calculated columns were created to enhance analytical depth:

- Categorization of delay severity based on delay duration
- Flag columns for peak travel periods
- Logical grouping for delay impact analysis

These calculated columns supported more meaningful visual analysis.

:- Creation of DAX Measures for KPIs

DAX (Data Analysis Expressions) was used to create dynamic measures, including:

- **Total Trains KPI**
- **Average Delay KPI**
- **Total Stations Covered**
- **Passengers Affected KPI**

These measures update automatically based on slicer selections such as year, station name, and delay reason.

7. Data Modeling

Although the dataset was primarily flat, relationships between date fields and categorical attributes were verified to ensure proper cross-filtering. This ensured that slicers and visuals interacted correctly across different dashboard elements.

ANALYSIS ON DATASET

1. Delay Reason Analysis

The donut and bar charts show that:

- Weather / Major Failures cause the highest average delays
- Technical Issues are the second most significant delay reason
- Traffic Congestion contributes to moderate delays
- Signal Issues cause smaller but recurring delays
- On-Time category reflects minimal or zero delay

This analysis highlights that infrastructure and environmental factors play a major role in delays

2. Station-wise Delay Analysis

- Certain stations such as Akola JN, Tatanagar, Bilaspur, Dankuni, and Rampur show higher average delays
- Busy junctions experience more delays due to higher traffic volume
- Smaller stations generally show lower average delay values

The bar charts clearly indicate variation in delay across stations.

3. Passenger Impact Analysis

- Stations with higher delays also show higher passenger counts affected
- Approximately 2 million passengers are impacted due to delays
- Passenger impact increases significantly during peak traffic stations

This highlights the direct relationship between operational delays and customer experience.

4. Geographical Analysis

The map visualization shows:

- Passenger impact spread across major regions of India
- Dense clusters around high-traffic railway corridors
- Metro and junction stations have higher passenger density

Geographical visualization helps in identifying regional hotspots of delay impact.

5. Train-wise and Month-wise Analysis

- Certain trains show higher frequency of delays across multiple months
- Month-wise distribution helps identify seasonal congestion patterns
- Peak months show increased delay frequency and passenger impact

This analysis supports time-based performance monitoring.

Specific Requirements:-

The primary objectives of the analysis are as follows:

- To identify trains and stations with the highest frequency of delays across the railway network
- To analyze month-wise and year-wise trends in train delays for understanding time-based patterns
- To study the distribution of delay reasons such as weather issues, technical failures, traffic congestion, and signal problems
- To evaluate the station-wise and region-wise impact of delays across Indian railway stations
- To measure the passenger impact of delays using passenger count and average delay metrics

These objectives guide the selection of KPIs, charts, maps, and analytical measures used in the Power BI dashboard, enabling effective monitoring of railway performance and passenger experience.

Visualization

- **Visualization**
- The Power BI dashboard uses a variety of interactive visualizations to effectively represent railway delay patterns and passenger impact. Each visualization is carefully selected to highlight specific insights and support user-driven analysis through filters and slicers.
- **1. KPI Cards**
- KPI cards are used to present high-level performance indicators at a glance. These include:
- **Total Trains** – Displays the total number of trains analyzed
- **Average Delay (Minutes)** – Shows the overall average delay across all stations
- **Total Stations Covered** – Represents the number of railway stations included in the analysis
- **Passengers Affected** – Indicates the total number of passengers impacted by train delays
- These KPIs provide a quick summary of railway performance and passenger impact.

2. Bar Charts

Bar charts are used to compare values across different categories:

- **Station-wise Average Delay** – Highlights stations with higher delay durations
- **Train-wise Delay Distribution** – Identifies trains that experience frequent delays
- **Delay Reason Distribution** – Compares delay causes such as weather, technical issues, traffic congestion, and signal failures

Bar charts make it easy to identify patterns and outliers.

These visualizations enhance user interaction and support intuitive data exploration, enabling users to derive insights efficiently.

3. Line Charts

Line charts are used for **time-based analysis**:

- **Month-wise Delay Trends** – Shows how delays vary across different months
- **Passenger Impact Over Time** – Illustrates changes in the number of affected passengers

These charts help in identifying seasonal trends and peak congestion periods.

4. Donut / Pie Charts

Donut charts visually represent proportional data:

- **Delay Reason Contribution** – Shows the percentage contribution of each delay reason
- **On-Time vs Delayed Trains** – Provides a clear comparison between punctual and delayed services

This visualization helps in understanding relative impact.

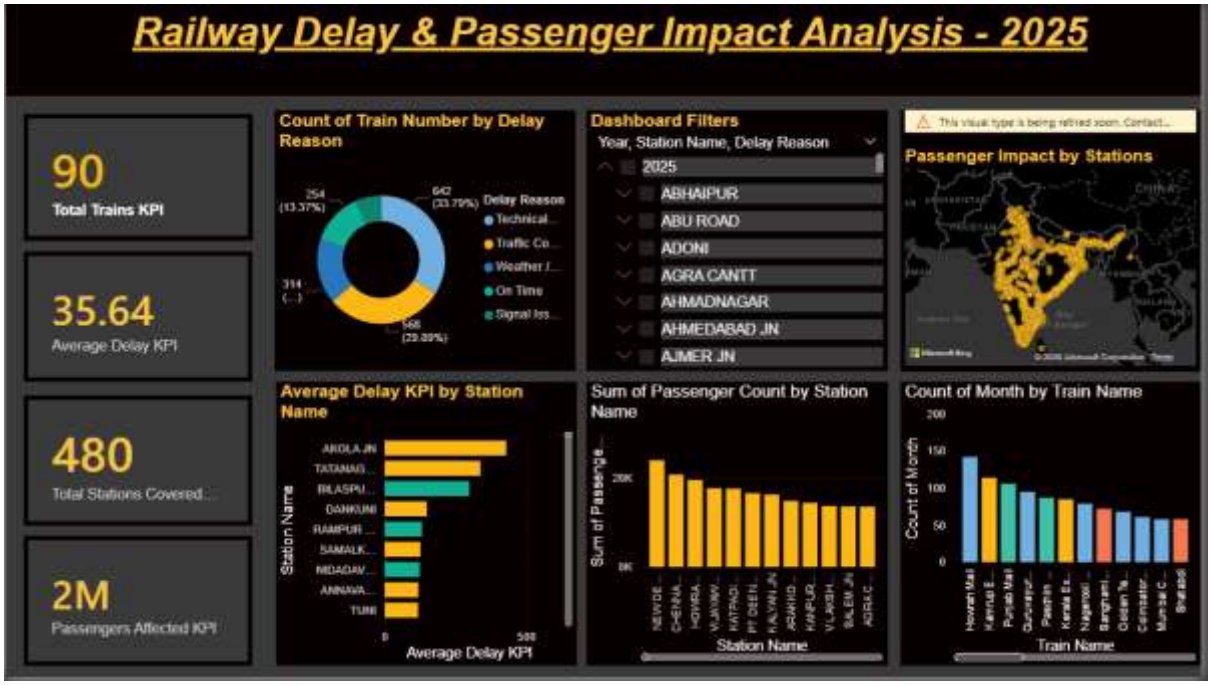
Slicers:-

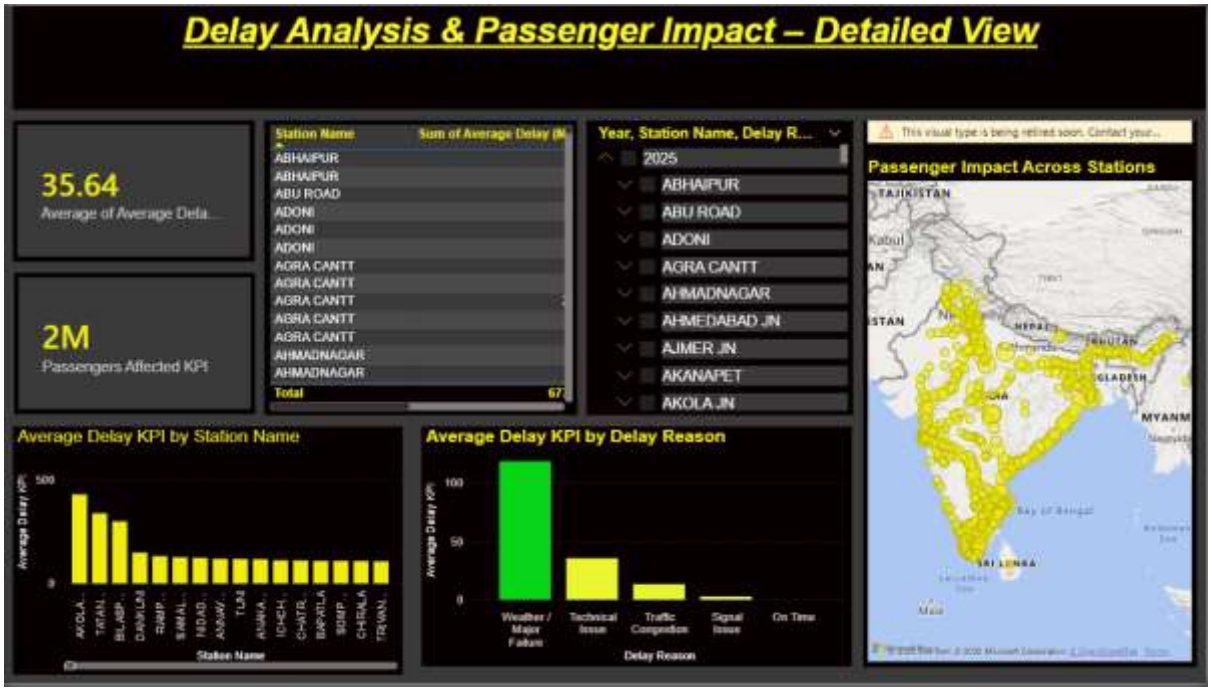
Slicers enable interactive filtering of the dashboard:

- **Month**
- **Year**
- **Station Name**
- **Train Name**
- **Delay Reason**

Slicers allow users to customize analysis dynamic

Dashboard:-





Linkedln Post :-

Linkedlnpost/Adityaraj/powerbidashboard

CONCLUSION:-

The project **"Railway Delay & Passenger Impact Analysis – 2025"** successfully demonstrates how Power BI can be used to analyze large-scale railway operational data. By transforming raw CSV data into interactive dashboards, the project provides meaningful insights into delay reasons, station performance, passenger impact, and geographical distribution.

The dashboard enables quick decision-making by highlighting critical delay factors and high-impact stations. This project showcases practical skills in data preprocessing, KPI creation, visualization, and analytical interpretation.

FUTURE SCOPE:-

The project can be enhanced further by:

- Integrating real-time railway data
- Applying predictive analytics for delay forecasting
- Adding weather and infrastructure data
- Performing route-level performance analysis
- Developing alert systems for high-delay zones

REFERENCES:-

1. Indian Railway Open Data Sources
2. Microsoft Power BI Documentation
3. Power Query Editor Documentation
4. Data Analytics Course Material (INT374)