Indian Railway Data Analysis — Train Schedule & Operational Trends Report

# Project Title:

Indian Railway Data Analysis

# Author:

Gokulraj S

# Date:

27 May 2025

# Table of Contents

1. Project Overview

2. Dataset Description

3. Data Cleaning & Preprocessing

4. Exploratory Data Analysis (EDA)

5. Data Transformation & Aggregation

6. Advanced Analysis

7. Visualizations

8. Key Insights & Recommendations

9. Conclusion

10. Appendix

# 1. Project Overview

Problem Statement:  
Analyze Indian Railway train operation schedules to identify patterns in train frequency, operational days, and station-wise trends to support operational planning.  
  
Objectives:  
- Understand train operation patterns across days of the week  
- Identify busy source and destination stations  
- Categorize trains by operational days  
- Visualize trends using appropriate charts

# 2. Dataset Description

Source: CSV file containing train schedules  
Rows: 500  
Columns: 5  
  
Column Name | Description  
------------|------------------------------  
Train\_No | Train Number  
Train\_Name | Name of the Train  
Source\_Station\_Name | Starting station  
Destination\_Station\_Name | Destination station  
days | Day on which the train operates  
  
Missing Values: Checked and handled

# 3. Data Cleaning & Preprocessing

- Renamed columns for consistency  
- Converted station names to UPPERCASE  
- Removed or handled missing values in ‘days’ column  
- Standardized data types

# 4. Exploratory Data Analysis (EDA)

- Total number of unique trains: 500  
- Unique source stations: 200  
- Unique destination stations: 198  
- Most common source station: NEW DELHI  
- Most common destination station: HOWRAH JN.

Basic Visualizations:  
- Bar plot of number of trains per day  
- Pie chart of most frequent source stations

# 5. Data Transformation & Aggregation

- Filtered trains operating on Saturday  
- Created new dataframe for trains starting from CHENNAI CENTRAL  
- Grouped by source station to count number of trains  
- Calculated average number of trains per day per source station  
- Added new column `Day\_Type` categorizing trains as Weekday or Weekend

# 6. Advanced Analysis

- Identified distribution of train operations across the week  
- Found high traffic on Friday and Saturday  
- Pattern: Weekend trains often start from metropolitan cities  
- Correlation check between number of trains and operational days showed Friday and Saturday peak frequencies

# 7. Visualizations

Tools Used:  
- Seaborn  
- Matplotlib  
  
Charts:  
- Bar chart: Number of trains by day  
- Pie chart: Source stations share  
- Heatmap: Train frequency by source and destination station

# 8. Key Insights & Recommendations

✔️ Maximum trains operate on Friday and Saturday  
✔️ NEW DELHI is the most frequent starting point  
✔️ Increase weekend train frequency on high-demand routes  
✔️ Optimize operational schedules for under-utilized days

# 9. Conclusion

The analysis successfully identified operational trends in Indian Railways, highlighting the importance of weekends and metro cities in train schedules. Further data like passenger count and on-time performance can enhance decision-making.

# 10. Appendix

- Python code snippets  
- Extra plots  
- Raw data tables