DATA ANALYTICS WITH POWER BI

PROJECT REPORT

(Project Semester September-December 2025)

Delhi Metro Analytics

Submitted by

Irfan khan

12317691

B-TECH Computer Science and Engineering

INT374

Under the Guidance of

Aashima (UID : 28968)

Discipline of CSE/IT

Lovely School of Computer Science

Lovely Professional University, Phagwara

CERTIFICATE

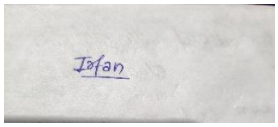
This is to certify that Irfan khan bearing Registration no. 12317691 has completed INT374 project titled, **"Delhi Metro Analytics "** under my guidance and supervision. To the best of my knowledge, the present work is the result of her original development, effort and study.

Signature and Name of the Supervisor
Aashima (UID: 28968)
School of Computer Science
Lovely Professional University
Phagwara, Punjab.

Date: 15-12-2025

DECLARATION

I, Irfan khan student of B-Tech CSE under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.



Date: 15-12-2025

Registration No. 12317691 Irfan khan

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my project guide, **Aashima mam**, for their invaluable support, motivation, and expert guidance throughout the course of this project. Their constructive feedback and continuous encouragement have been instrumental in shaping the direction and outcome of this work.

I am deeply thankful to the faculty of the **Computer Science Department, Lovely Professional University**, for providing me with a solid foundation in data analysis and for encouraging an analytical approach toward solving real-world problems. Their teachings and the academic environment they fostered have played a key role in the successful completion of this Data Science Minor Project.

I also extend my heartfelt thanks to my classmates and friends who supported me with their valuable inputs, constant motivation, and helpful discussions that enriched my understanding during the project development process.

I owe special thanks to my family, whose unwavering belief in my potential and consistent moral support gave me the strength and determination to complete this work with confidence and commitment.

Finally, I express my appreciation to all the authors, online communities, and platforms that provided valuable resources and tutorials on Power bi dashboards and data visualization techniques, which proved extremely helpful during the project.

This project would not have been possible without the collective support and contributions of everyone mentioned above.

INTRODUCTION

Urban public transportation plays a crucial role in the daily lives of millions of people, especially in metropolitan cities like Delhi. The Delhi Metro is one of the largest and most widely used rapid transit systems in India, serving lakhs of passengers every day. While the metro has significantly improved connectivity, reduced road congestion, and supported sustainable urban mobility, passengers still face several real-world challenges such as overcrowded stations, long ticket queues, delays during security checks, and discomfort during peak travel hours.

Being a resident of Delhi, these challenges have been personally experienced multiple times, especially during peak hours when excessive crowding at metro stations leads to long waiting times and, in some cases, passengers missing important trains or flights. These issues highlight the need for better crowd management, efficient resource allocation, optimized routes, and improved operational planning. Addressing such problems requires a data-driven approach rather than assumptions or manual observation.

In this context, data analytics and business intelligence tools play a vital role in understanding passenger behavior, identifying congestion patterns, and supporting informed decision-making. Microsoft Power BI is a powerful visualization and analytics platform that enables the transformation of raw transportation data into meaningful insights through interactive dashboards, charts, and key performance indicators (KPIs).

This project, titled “Delhi Metro Dashboard Analysis”, focuses on analyzing metro-related data to identify critical operational issues and propose potential solutions through visual analytics. The dashboard was designed to extract insights related to passenger crowd patterns, station usage, payment methods, revenue contribution, and route utilization. By consolidating multiple aspects of metro operations into a single interactive dashboard, the project aims to provide clarity on where problems occur and how they can be addressed effectively.

The analysis carried out in this dashboard is structured around five key objectives, including the identification of highly crowded stations, evaluation of payment methods, analysis of exit stations, assessment of revenue-generating stations, and identification of the busiest routes. Each objective addresses a specific operational challenge faced by metro passengers and authorities, offering insights that can help improve efficiency, reduce delays, and enhance passenger experience.

The dashboard was developed using a structured methodology involving data cleaning, transformation, modeling, and visualization. Power BI features such as slicers, filters, interactive charts, and calculated measures were used to ensure flexibility and usability. The visual design was carefully planned to ensure clarity, simplicity, and easy interpretation, even for non-technical users.

Overall, this project demonstrates how Power BI dashboards can be effectively used to analyze real-world transportation problems and support data-driven decision-making in urban transit systems like the Delhi Metro. The insights derived from this analysis can assist metro

authorities, planners, and stakeholders in improving crowd management, operational efficiency, revenue optimization, and overall passenger satisfaction.

SOURCE OF DATASET

The dataset used for the **Delhi Metro Dashboard Analysis** was obtained through a personal and reliable source. Instead of using publicly available or generic datasets, the data was collected with the help of a friend who works at a Delhi Metro station as a ticket collector. Accessing such operational-level data was not straightforward, as metro-related datasets are rarely available openly due to security and privacy concerns. After considerable effort and difficulty in finding suitable data from online sources, this dataset was finally acquired through personal coordination.

Although the project could have been developed using any publicly available transportation dataset, this dataset was intentionally chosen because it closely reflects **real-world metro operations and passenger behaviour**. The data represents actual scenarios faced by metro commuters, making the analysis more practical, relevant, and problem-oriented rather than theoretical.

The motivation behind selecting this dataset is deeply connected to real-life experiences. Being a resident of Delhi, the challenges of overcrowded metro stations, long queues for tickets, delays during checking, and congested routes have been personally faced multiple times. These issues are not limited to Delhi alone but are common across metro cities in India. Therefore, the dataset was selected to highlight such everyday problems and direct attention toward identifying potential solutions using data analytics.

This dataset enables the analysis of critical operational factors such as crowded stations, payment methods, exit patterns, revenue-generating stations, and busy routes. By focusing on these parameters, the dashboard aims to shift attention from raw numbers to meaningful insights that can support better planning, crowd management, and passenger experience improvement.

DATASET PREPROCESSING

Data preprocessing is a crucial step in ensuring that the dataset is clean, consistent, and ready for analysis in Power BI. Since raw data from government sources may contain inconsistencies, formatting issues, and unnecessary noise, preprocessing enhances data quality and improves the accuracy of insights generated by the dashboard. For this project, the following preprocessing steps were carried out:

1. Trimming of Whitespaces

Unwanted leading and trailing spaces were removed from all text-based columns. This step prevents issues such as mismatched categories, duplicate entries, and inaccurate filtering during analysis. Trimming helps maintain uniformity across all categorical fields.

2. Capitalization of Words

To ensure consistent formatting and improve readability, all text fields were standardized by capitalizing the first letter of each word. This step enhances the visual quality of the dashboard and eliminates inconsistency caused by mixed-case entries (e.g., “male”, “Male”, “MALE”).

3. Formatting Columns

Column formatting was applied to make the dataset clean and visually uniform. This included:

- Standardizing formats for dates, numbers, and text
- Applying appropriate titles
- Structuring values for better interpretation
- Proper formatting ensures that the visuals in Power BI appear clean and professional.

4. Assigning Correct Data Types

Each column was assigned its appropriate data type—such as **Whole Number, Decimal Number, Text, Date, or Categorical Value**—based on the nature of the information. Assigning correct data types improves data modelling, enhances filtering accuracy, and ensures correct visualization behaviour.

5. Adding New Columns ,

Additional columns were created to enhance analysis and improve data usability. These derived columns helped in categorizing stations, analysing crowd patterns, and evaluating routes more effectively. Creating these columns enabled better insights and supported accurate visualization in the Power BI dashboard.

6. Handling Duplicate Values

Duplicate rows were identified and removed to maintain data integrity. Eliminating duplicates avoids skewed insights, incorrect averages, and misleading trends. Only unique, validated records were retained for analysis.

ANALYSIS OF DATASET

To analyze the **Delhi Metro Analytics** effectively, **five interactive dashboard pages** were created using Microsoft Power BI. Each dashboard page focuses on a specific aspect of Metro data And Specify its Distribution Across metro cities and is designed to provide clear, meaningful, and actionable insights for users.

Page 1: Overview Dashboard

Purpose: Provides a high-level summary of Delhi Metro Analytics performance

This page displays key performance indicators such as **Total users Traveeling through metro on daily bases , revenue from all the stations and busy routes Crowded stations it tells a lot of activities of metro**. It offers a quick understanding of Delhi metro Performance across all time periods and acts as the entry point for deeper analysis.

Page 2: Passengers by Payment Mode

Purpose: Analyze the users payment mode which type of method they are using

These dashboards examine the relationship between **Metro and people who are living in metro cities and facing daily life problems**

Page 4: Top 10 most crowded Stations

Purpose: For easy Travel experience and not to miss their Trains And Flights

This page focuses on **how the people miss their trains and flights and even Job meetings , interviews and some families leaves their loved ones on another stations**

Page 5 Routes it focus on passengers are travelling on which type of routes on daily bases

Purpose: Measure the most busy routes

Purpose: Enable flexible and interactive data exploration.

Objective1 Top 10 Most Exited Stations

i. General Description

This objective focuses on identifying metro stations with the highest passenger exits. The analysis helps understand exit congestion and passenger flow at major stations.

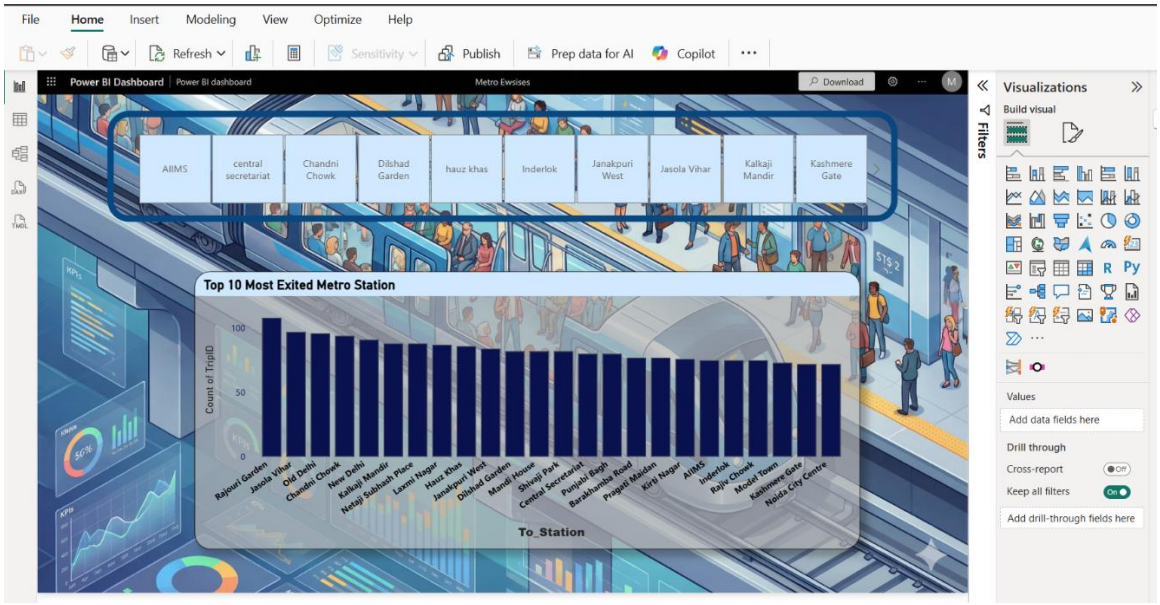
ii. Specific Requirements

- Visual Type: Bar Chart / Column Chart
- Values: Exit Count
- Axis: Station Name
- Filters: Station slicer

iii. Analysis Results

The analysis highlights the **Top 10 most exited stations**, where maximum passengers leave the metro system. These stations usually experience heavy crowd movement during peak hours.

iv. Visualization



Objective 2 Passengers by Payment Mode

i. General Description

This objective analyzes passenger distribution based on different payment methods used for metro entry**.**

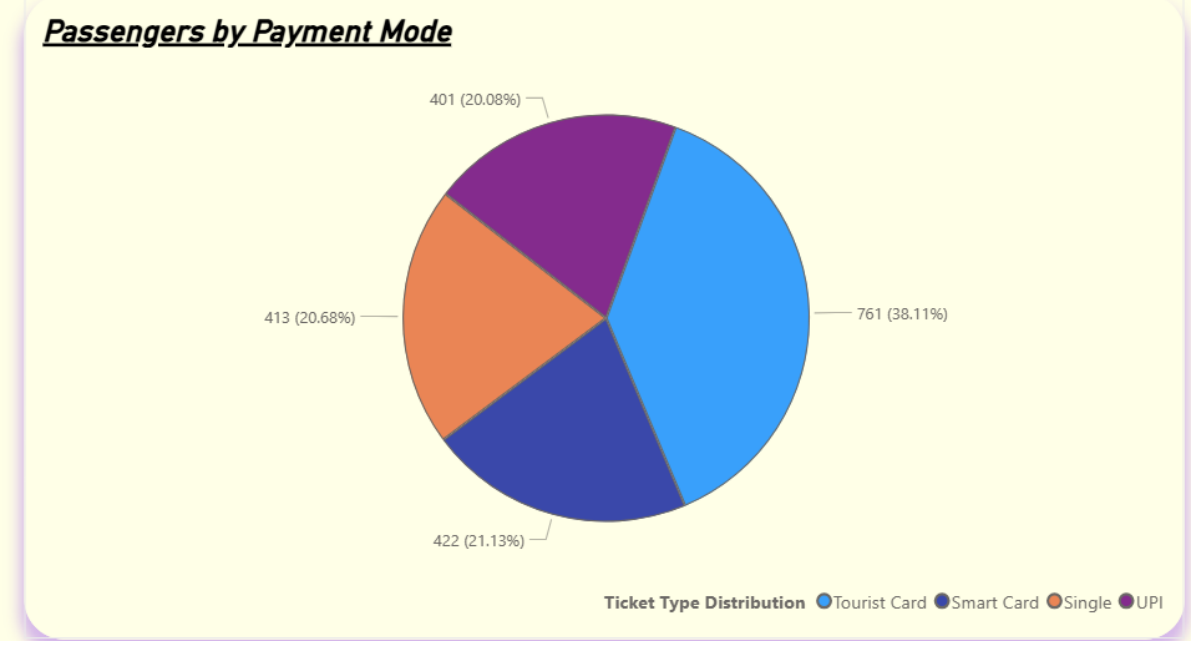
ii. Specific Requirements

- Visual Type: Donut Chart / Pie Chart
- Values: Passenger Count
- Legend: Payment Mode (UPI, Smart Card, Token, Digital Card)
- Filters: Payment Mode slicer

iii. Analysis Results

The visualization clearly shows which payment methods are most frequently used. Token-based entries contribute more to queue formation, while digital payments support faster entry.

iv. Visualization



Objective 3 Top 10 Crowded Stations

i. General Description

This objective identifies metro stations with the highest passenger crowd density.

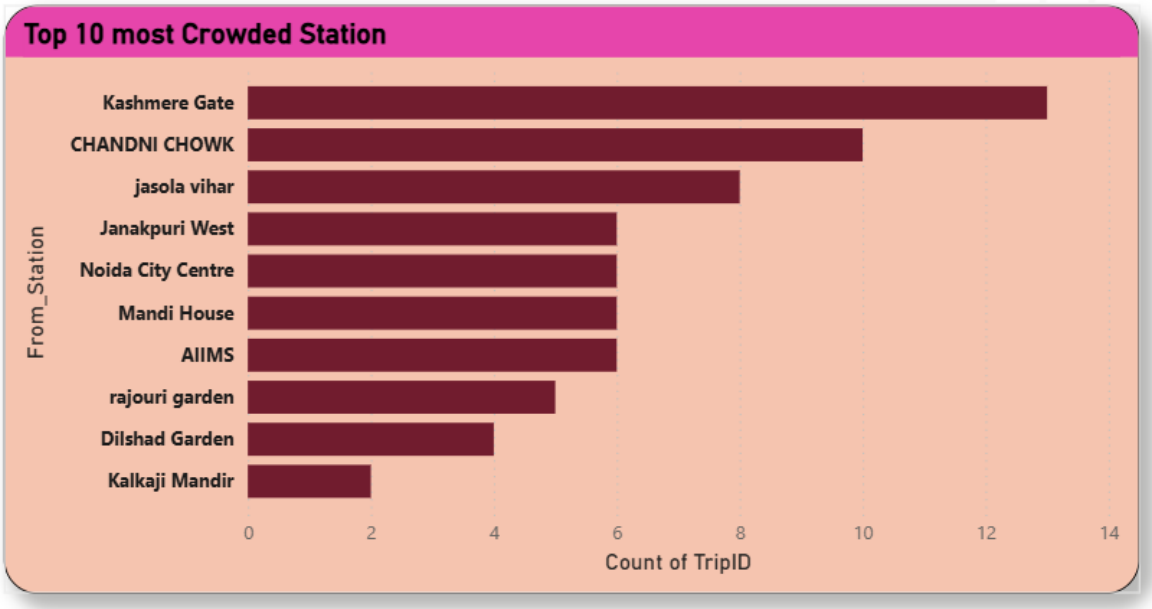
ii. Specific Requirements

- Visual Type: Bar Chart
- Values: Passenger Footfall
- Axis: Station Name
- Filters: Station slicer

iii. Analysis Results

The dashboard highlights the Top 10 most crowded stations, especially during peak travel hours. These stations require better crowd control and staff management.

iv. Visualization



Objective 4 Top 10 Revenue-Generating Stations

i. General Description

This objective focuses on analyzing revenue contribution from different metro stations.

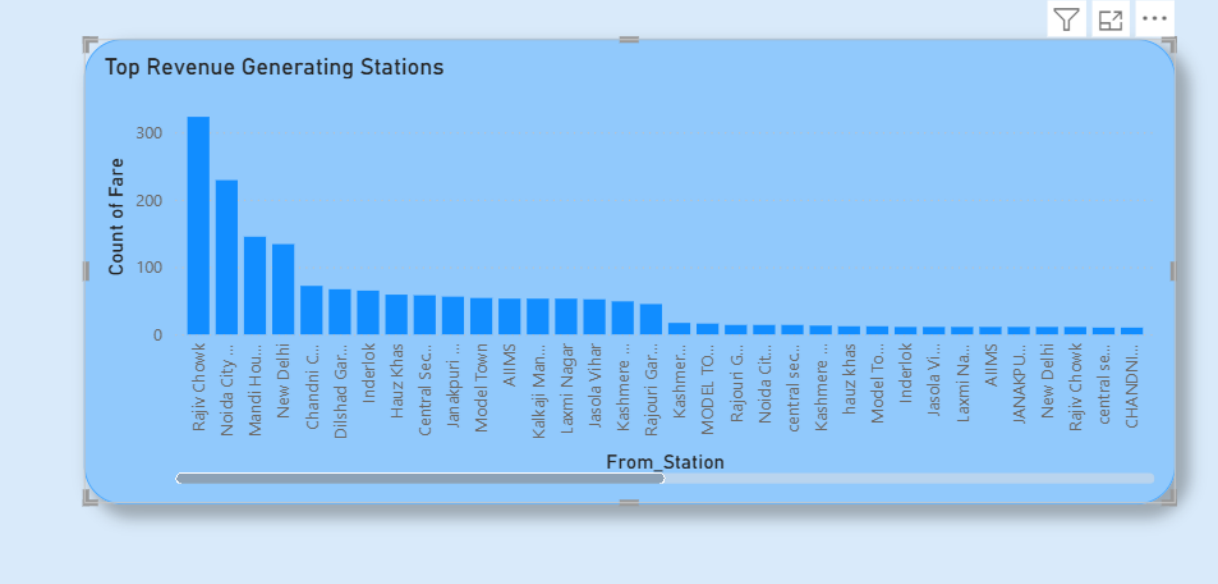
ii. Specific Requirements

- Visual Type: Column Chart
- Values: Revenue
- Axis: Station Name
- Filters: Station slicer

iii. Analysis Results

The analysis identifies the Top 10 revenue-generating stations, which are generally high-traffic and commercially important locations.

iv. Visualization



Objective 5: Most Busy Passenger Routes

i. General Description

This objective analyzes passenger movement between source and destination stations to identify busy routes.

ii. Specific Requirements

- Visual Type: Table / Bar Chart
- Values: Passenger Count
- Axis: Route (Source → Destination)
- Filters: Route slicer

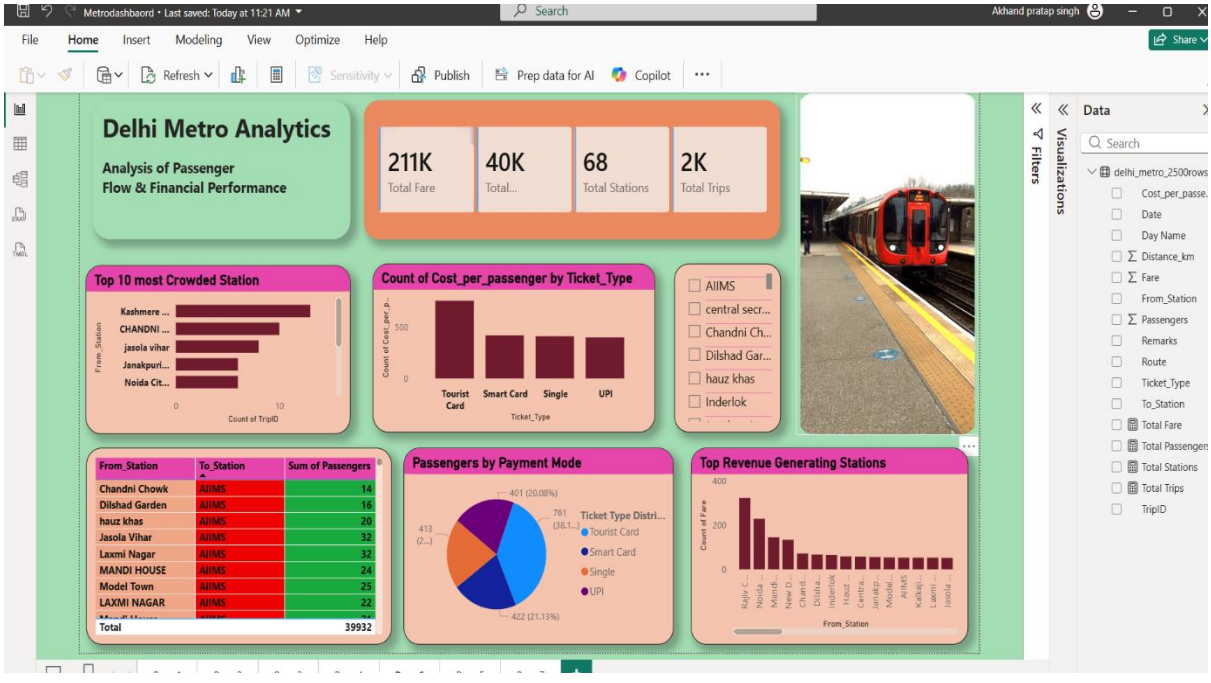
iii. Analysis Results

The analysis highlights the most frequently used routes, indicating high passenger demand and overcrowding on specific metro lines.

iv. Visualization

From_Station	To_Station	Sum of Passengers
Chandni Chowk	AIIMS	14
Dilshad Garden	AIIMS	16
hauz khas	AIIMS	20
Jasola Vihar	AIIMS	32
Laxmi Nagar	AIIMS	32
MANDI HOUSE	AIIMS	24
Model Town	AIIMS	25
LAXMI NAGAR	AIIMS	22
Mandi House	AIIMS	21
model town	AIIMS	20
central secretariat	AIIMS	17
Kashmere Gate	AIIMS	17
NOIDA CITY CENTRE	AIIMS	19
Central Secretariat	AIIMS	38
Chandni Chowk	AIIMS	92
Dilshad Garden	AIIMS	15
Hauz Khas	AIIMS	35
Inderlok	AIIMS	32
Total		39932

FINAL DASHBOARD



CONCLUSION

The **Delhi Metro Dashboard Analysis** provides a comprehensive, interactive, and data-driven view of passenger movement, crowd patterns, revenue distribution, and route utilization across the metro network. By integrating key metrics such as exited stations, crowded stations, payment modes, revenue-generating stations, and busy routes, the dashboard transforms complex metro data into clear and actionable insights. The use of multiple visualization techniques along with dynamic slicers enables users to explore operational challenges effectively and understand patterns that influence crowd management, operational efficiency, and overall passenger experience.

Page 1 – Exit, Crowd & Payment Overview

This page provides a high-level understanding of passenger movement and congestion patterns in the Delhi Metro system:

- **Top 10 Most Exited Stations** are displayed to identify locations where maximum passengers exit the metro, helping understand exit congestion and crowd flow.
- **Passengers by Payment Mode** (UPI, Smart Card, Token, Digital Card) are visualized using charts to analyze ticketing behavior and identify causes of long entry queues.
- **Top 10 Crowded Stations** highlight stations with the highest passenger density, especially during peak hours.
- **Interactive slicers** allow users to filter data by station and payment mode for focused analysis.

Page 2 – Revenue-Generating Stations Analysis

This page focuses on analyzing station-wise revenue contribution:

- **Top 10 Revenue-Generating Stations** are visualized to identify stations contributing the highest revenue to the metro system.
- The analysis helps understand the relationship between passenger volume and revenue generation.

- **Station-level filters** allow users to analyze individual stations and compare revenue performance easily.

Page 3 – Busy Routes & Passenger Movement Analysis

This page emphasizes passenger travel patterns across metro routes:

- **Source-to-Destination route analysis** is used to identify the most busy and frequently used metro routes.
- The dashboard highlights routes with high passenger demand and overcrowding.
- These insights support better planning of train frequency and route-level capacity management.

Key Features and Benefits

- **Interactivity:** Slicers and filters allow users to explore station-wise, route-wise, and payment-wise data dynamically.
- **Visual Diversity:** Use of bar charts, column charts, donut charts, tables, and slicers provides multi-dimensional insights.
- **Problem Identification:** Helps identify overcrowding, exit congestion, ticketing delays, and high-demand routes.
- **User-Friendly Design:** Clean visuals and simple layouts make the dashboard easy to understand for both technical and non-technical users.
- **Decision Support:** Assists metro authorities and analysts in improving crowd management, revenue optimization, and passenger experience.

Overall Insight

The complete **Delhi Metro Dashboard Analysis** integrates passenger exits, crowd density, payment behavior, revenue contribution, and route utilization into a unified analytical solution. By combining interactive visuals, filters, and concise insights, the dashboard transforms raw metro data into meaningful information that supports data-driven decision-making for better urban transport management.

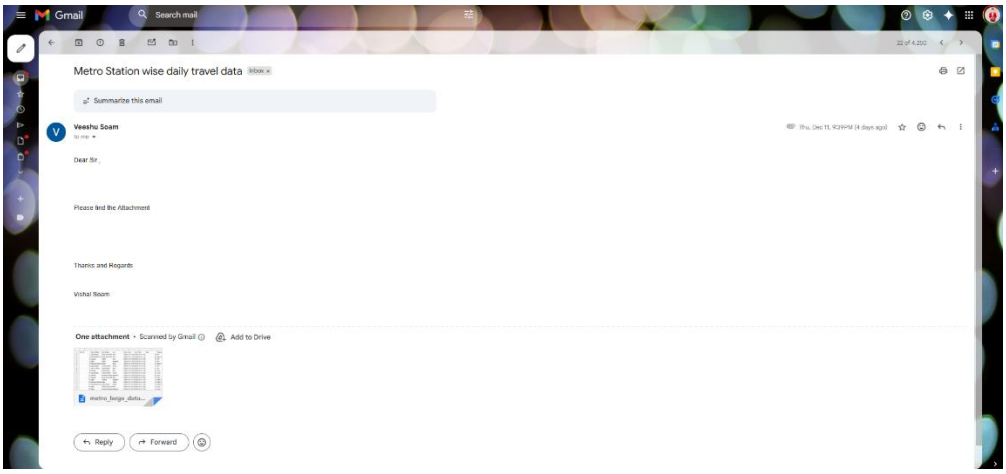
FUTURE SCOPE

The Delhi Metro Dashboard can be enhanced further to increase its analytical depth and real-world applicability:

1. **Real-Time Data Integration:** Integration of live metro data such as real-time footfall and ticketing information for continuous monitoring.
2. **Predictive Analytics:** Use of forecasting models to predict crowd levels and peak-hour congestion.
3. **Advanced Visualizations:** Addition of heatmaps for station crowd intensity and route congestion.
4. **Expanded Dataset:** Inclusion of peak/off-peak timings, special events, and seasonal trends.
5. **User Personalization:** Role-based dashboards for metro staff, planners, and management teams.

REFERENCES

https://mail.google.com/mail/u/0/#inbox/FMfcgzQdzmSjztLzzkZCCWKzJGwjgkFC(from



- Microsoft Power BI Documentation. *Visualizations, Slicers & Dashboard Guides*. https://docs.microsoft.com/power-bi
- Microsoft Excel Documentation. *Data Analysis and Dashboarding Features*. https://support.microsoft.com/excel
- TutorialsPoint. *Power BI Data Analysis Guides*.
- Project Mentor / Faculty Guidance
- Internal Training Materials and Simulated Dataset

LINKS

LINKEDIN:
https://www.linkedin.com/feed/update/urn:li:activity:7405657946589736960/

GOOGLE DRIVE:

https://drive.google.com/drive/folders/1h7jI-P-GIOY2knrqCCmbBSfYbIl265vu?usp=drive_link