Final Project– Proposal

Project Description:

This project focuses on analyzing multiple datasets from the Metropolitan Transportation Authority (MTA) to evaluate the performance, reliability, and usage trends of public transportation services in New York City. The analysis covers both bus and subway systems and integrates data from various operational aspects such as daily ridership, mechanical failures, route colors, performance indicators, and station information.

By exploring these datasets together, the goal is to uncover meaningful insights about system efficiency, passenger demand patterns, service reliability, and potential areas for improvement. The project will apply data cleaning, exploratory data analysis, statistical methods, and visualization techniques to transform raw transportation data into actionable knowledge.

Group Members:

Amr Mohamed: Team Leader & Data Analyst

Bishoy Mina: Data Analyst

Paula Sameh: Data Analyst

Abanoub Shawkat: Data Cleaning

Mohamed Mamdouh: Data Visualization

Mohamed Ahmed: Documentation & Presentation

Objectives

1. Analyze ridership trends across different time periods and transportation modes.
2. Evaluate bus reliability by examining mechanical failures and mean distance between failures.
3. Identify performance patterns across boroughs, routes, and agencies.
4. Integrate multiple MTA datasets to build a comprehensive view of system operations.
5. Generate data-driven insights and visualizations to support decision-making and service improvement.
6. Explore predictive or optimization techniques to forecast ridership or enhance transportation efficiency and passenger experience.

Tools & Technologies

- Excel – Initial data exploration and data cleaning
- SQL – Querying, filtering, and joining datasets efficiently
- Python – Advanced data analysis and processing
- Pandas, NumPy – Data manipulation and statistical calculations
- Matplotlib – Data visualization and trend analysis
- Power BI / Tableau – Interactive dashboards and reporting
- Jupyter, colab Notebook – Development and documentation of analysis workflow
- Git/GitHub – Version control and project collaboration

Milestones & Deadlines

1. Dataset Understanding & Project Planning
Deadline: Mid-October

- Review all datasets and sheet structures
- Define project scope and questions
- Finalize proposal structure

2. Data Cleaning & Preparation (Excel + Python)
Deadline: End of October

- Handle missing values and duplicates
- Standardize formats (dates, text, numeric types)
- Merge relevant datasets if needed

3. Exploratory Data Analysis (EDA)
Deadline: Mid November

- Analyze ridership trends, reliability, and performance
- Identify correlations and patterns
- Extract initial insights

4. Visualization & Dashboard Design
Deadline: End of November

- Build charts and graphs (Python / Excel / Power BI)
- Design interactive dashboard layout
- Highlight key metrics and KPIs

5. Advanced Analysis / Predictive or Optimization Model (optional but strong)
Deadline: Mid December

- Build forecasting or optimization model
- Evaluate performance and insights

6. Final Report & Presentation Preparation
Deadline: End of December

- Write final analysis report
- Prepare presentation / portfolio version
- Final review and polish

KPIs (Key Performance Indicators)

1. Data Cleaning & Processing

- 100% of missing/duplicate data handled
- 100% of datasets standardized and merged successfully
- 95% Data validation accuracy

---

2. Analysis & Insights

- % of agreed business questions answered (Target: ≥ 90%)
- Number of meaningful insights discovered (Target: ≥ 5)
- Accuracy of calculations and metrics (e.g., average miles before failure) (Target: ≥ 95%)

---

3. Visualization & Reporting

- Dashboard load time (Target: < 3 seconds)
- Dashboard usability (≥ 80% of users navigate without assistance)
- Number of visualizations that clearly explain key trends (Target: ≥ 5)

---

4. Final Documentation & Presentation

- Final report completeness (Target: 100%)
- Number of actionable recommendations (Target: ≥ 3)
- Presentation clarity and engagement score (Target: ≥ 85%)