Team members: Maya Patel,
Adrien Rozario, Illia Polishchuk
Team number: IC25069

Team Abstract

Project Proposal:

In this project, we analyzed the public DC WMATA Metro ridership data to determine:

Ridership Trends & Patterns

- What season has the highest overall ridership? (Should Metro adjust service schedules during certain seasons?)
- What times of the day are busiest across all stations? (Are there stations that need more frequent service due to peak-hour congestion?)

Station-Specific Analysis

- What is the busiest station each month of the year? (Could this create potential opportunities for business ads?)
- What are the least busiest stations? (Could this create potential opportunities for Metro to optimize resources?)
- Which station types (residential, commercial, transfer) handle the most traffic?
- Which metro lines handle the most ridership overall?

Data Collection:

We collected data from secondary sources with publicly available raw data, such as WMATA Metrorail Ridership Summary and WMATA Corridor Data Maps.

Data Processing:

In order to process the data, we normalized it to have the necessary tables and relationships. The columns with relevant information were merged into dataframes so they could be used in visualizations.

Exploratory Data Analysis:

1. Seasonal average ridership bar chart + monthly average ridership (bar plot)
2. Total daily time entries for the top 15 busiest stations in 2024 (stacked bar plot)
    a. Rank stations by highest peak-hour ridership.
3. Monthly average ridership for the top 10 busiest stations in 2024 (line plot)
    a. Rank stations by highest & lowest ridership per month.
4. Ridership by station type (stacked bar plot)
5. Total ridership by metro line (line plot)