

INFO CHALLENGE 2025

DC WMATA Metro Ridership Analysis



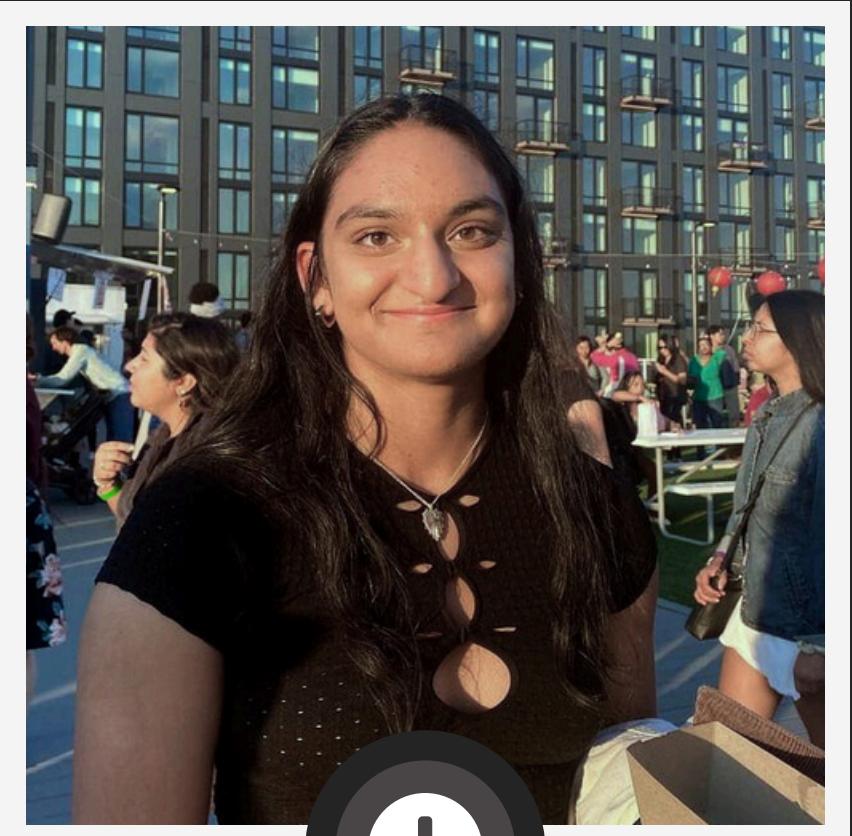
Team #:
IC25069

Team Members:
Maya Patel, Illia Polishchuk, Adrien Rozario

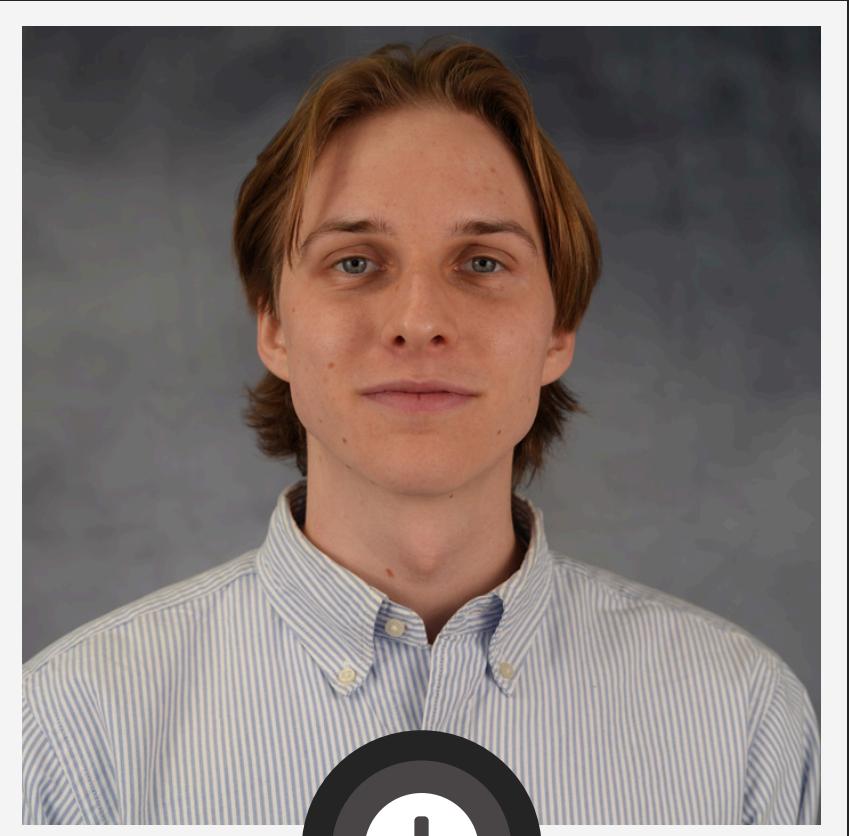




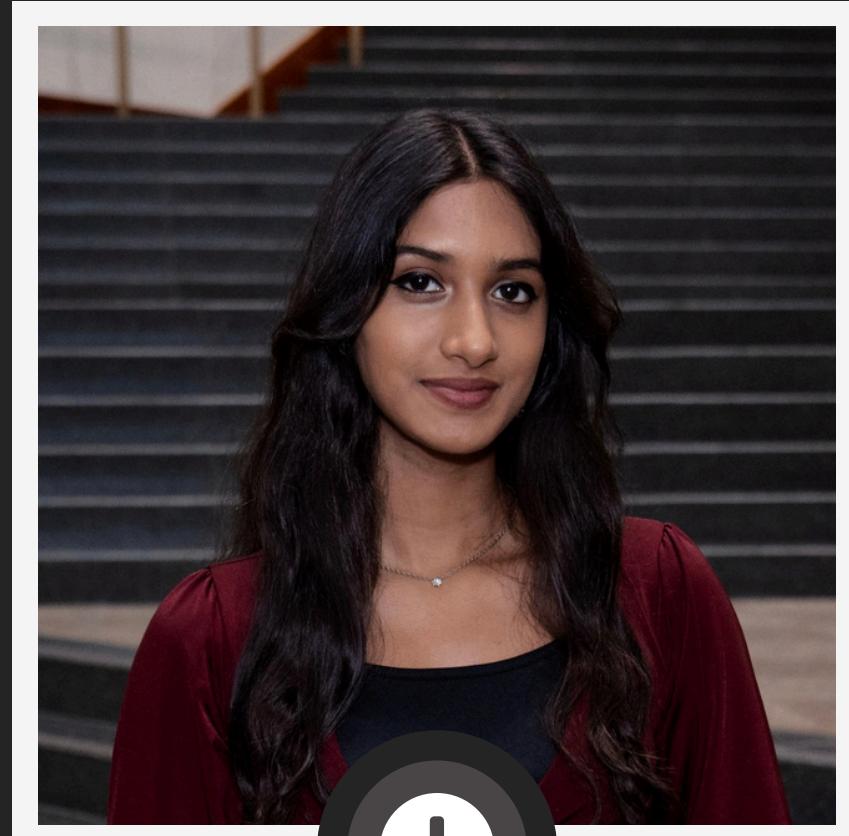
Our Team



Maya Patel



Illia Polishchuk

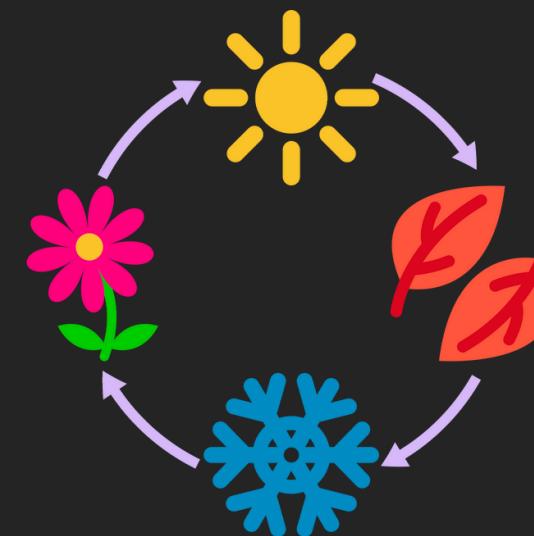


Adrien Rozario



Background

The Washington Metropolitan Transit Authority (WMATA) experiences fluctuating ridership patterns throughout the year, influenced by



seasonal trends



station locations



metro line usage



Problem Statement

However, without a data-driven approach to understanding these patterns, Metro may struggle with



resource allocation



congestion management



pricing optimization

Additionally, businesses near metro stations may lack insights into potential customer foot traffic.



Proposal

This project aims to analyze ridership trends across different time periods and factors to:



- provide meaningful insights for Metro's operational planning
- assist businesses seeking to optimize marketing and service strategies





Research Question 1

1. Overall Ridership Trends (Big Picture for Metro & Businesses)

- What season has the highest overall ridership? (Metro can prepare for peak demand; businesses can plan marketing strategies.)
- What months have the highest overall ridership? (Reinforces seasonal trends.)





Research Question 2

2. Metro Line Analysis (Broad Perspective → Business Relevance)

- Which metro lines handle the most ridership overall? (Metro can prioritize improvements; high-traffic lines are valuable for ads.)
- How does ridership fluctuate across metro lines by month? (Identifies seasonal trends for each line.)





Research Question 3

3. Station-Specific Ridership (Metro Service & Business Insights Together)

- What are the top 10 busiest stations each month of the year? (Frequent usage → Metro needs more service; businesses near these stations can thrive.)
- What are the middle 10 busiest stations each month of the year? (Provides mid-range perspective on ridership.)
- What are the least 10 busiest stations each month of the year? (Metro may adjust resources; are there untapped business opportunities?)



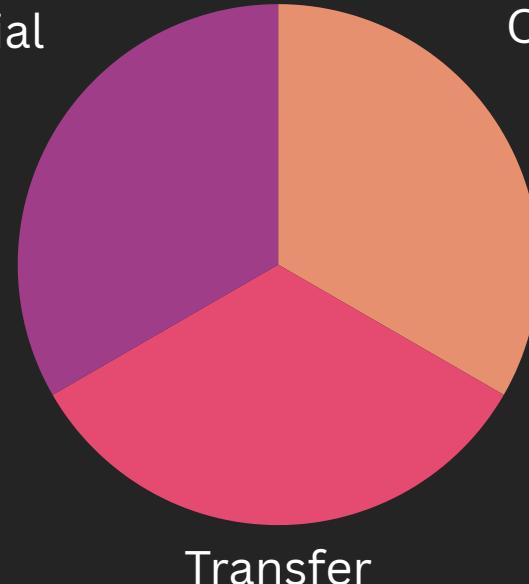


Research Question 4

4. Metro Line & Station Type Analysis (Broad Perspective → Business Relevance)

Residential

Commerical



- How is ridership distributed across metro stations and lines over time? (Helps Metro assess usage trends and service needs.)
- How many stations of each type (residential, commercial, transfer) exist on each metro line? (Reveals structural distribution of different station types.)
- Which station types (residential, commercial, transfer) handle the most traffic? (Tells businesses where customer foot traffic is strongest.)



Research Question 5

5. Business-Driven Optimization (Finally, Data-Backed Opportunities)

- Can Metro adjust pricing based on congestion? (Helps optimize revenue and ridership.)





Data Collection & Processing

We collected data from secondary sources with publicly available raw data, such as [WMATA Metrorail Ridership Summary](#) and [WMATA Corridor Data Maps](#). In order to process the data, we normalized it to have the necessary tables and relationships.

The columns with relevant information were merged into data frames to be used in visualizations.

Datasets were extracted from these sites and manipulated with Python scripts to optimize our usage.

The dataset on the geographical data of each metro station (metro_stations.csv) had station names altered to match those of the other datasets to allow merge on station name.



Data Collection & Processing

- All_Months.csv
- Annual_Station_Boarding_Compiled.csv
- entries_exits_transformed_data.csv
- everyday_2024_w_metro_stations.csv
- merged_df.csv
- metro_stations.csv





Technological Tools



Google Sheets



Visual Studio
Code



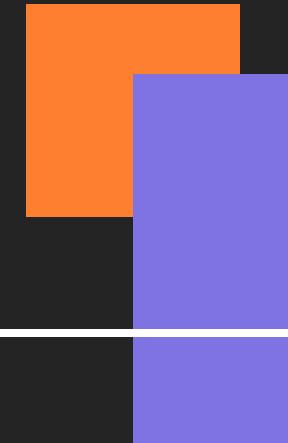
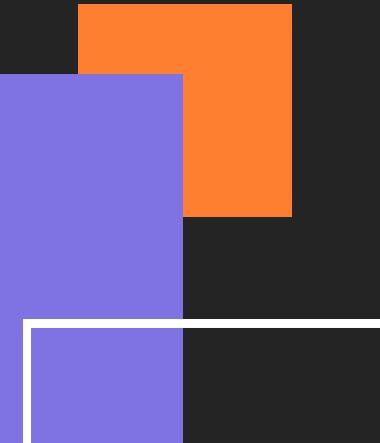
Jupyter
Notebook



Python



Github



Exploratory Data Analysis



Overall Ridership Trends

- Histogram of Average Daily Ridership by Season in 2024
- Histogram of Average Total Daily Entries by Month and Season

Metro Line Analysis

- Histogram of Lines and Ridership (Entries + Exits)
- Line Plot of Total Ridership Per Metro Line by Month

Station-Specific Ridership

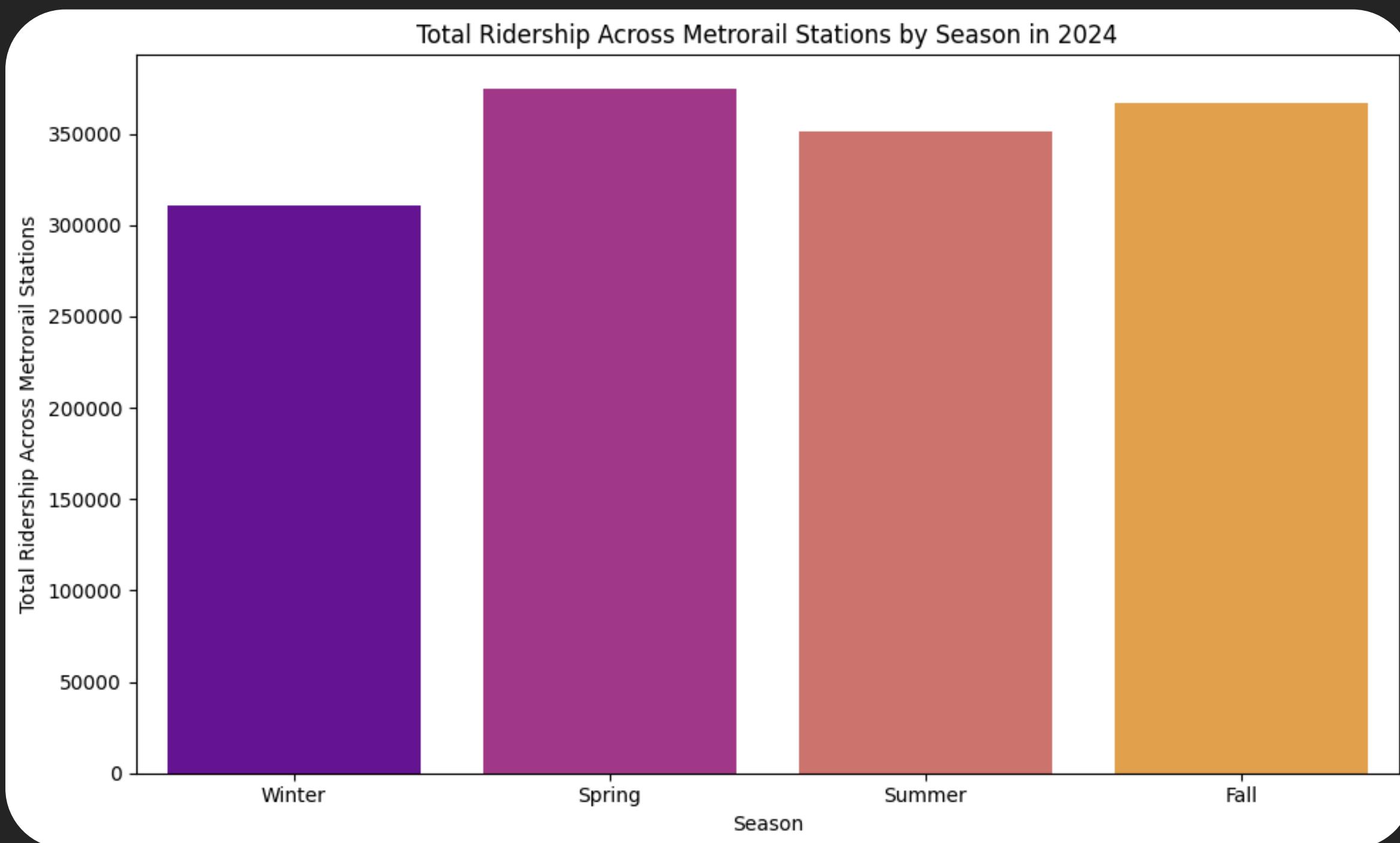
- Line Graph of Average Daily Entries for the Top 10 Busiest Metro Stations by Month in 2024
- Line Graph of Average Daily Entries for the Middle 10 Most Used Metro Stations by Month in 2024
- Line Graph of Average Daily Entries for the 10 Least Busy Metro Stations by Month in 2024

Metro Line & Station Type Analysis

- Heatmap of Monthly Total Daily Entries by Station by Line
- Grouped Bar Chart of Number of Each Station Type by Line
- Grouped Bar Chart and Pie Chart of Ridership by Station Type

Overall Ridership Trends:

What season has the highest overall ridership?

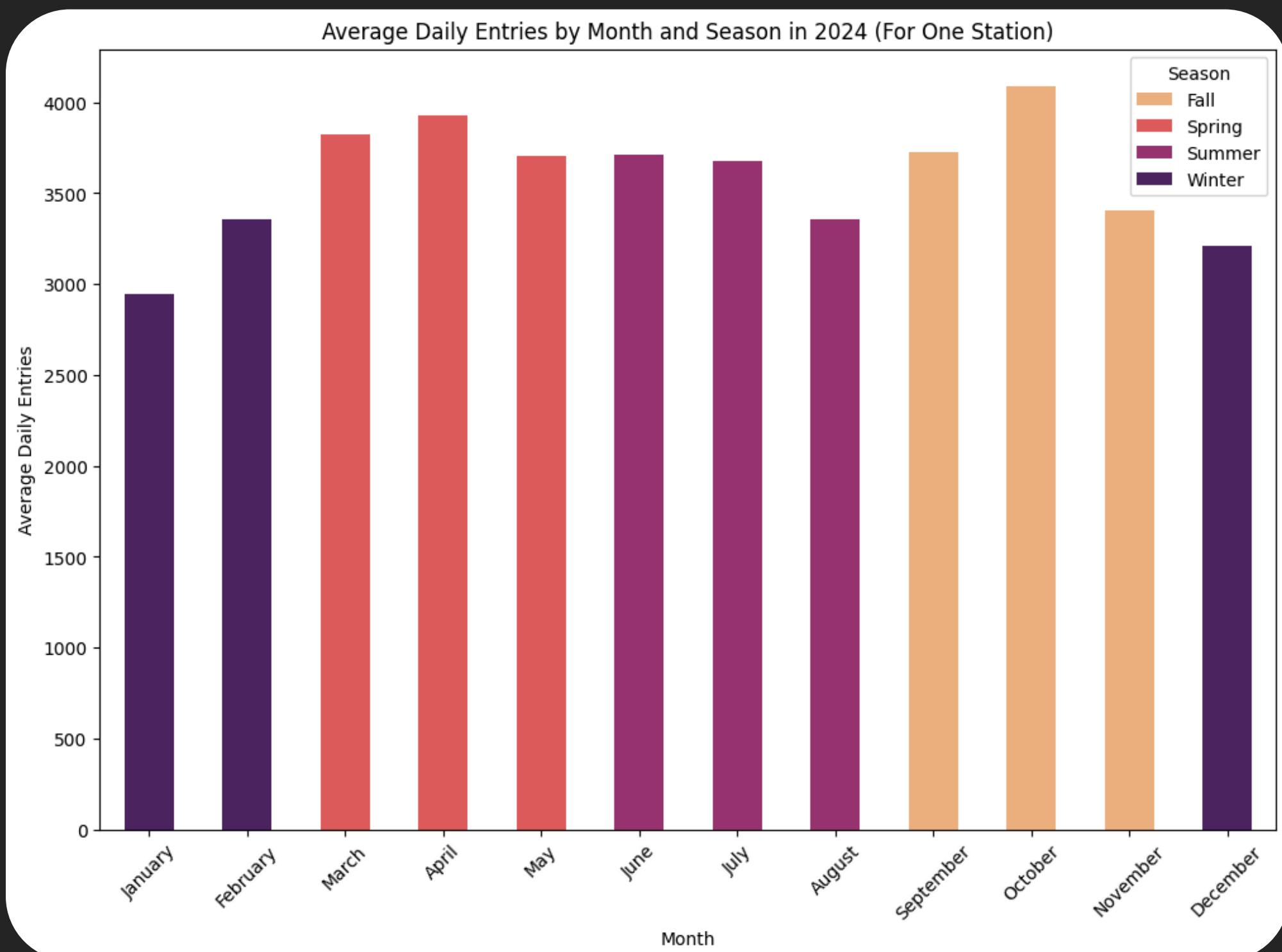


Although ridership fluctuates, peaking in spring and dropping in winter.

This graph does not provide enough data to draw concrete conclusions about ridership patterns,

So we decided to look into total ridership every month of the year.

Overall Ridership Trends: What months have the highest overall ridership?



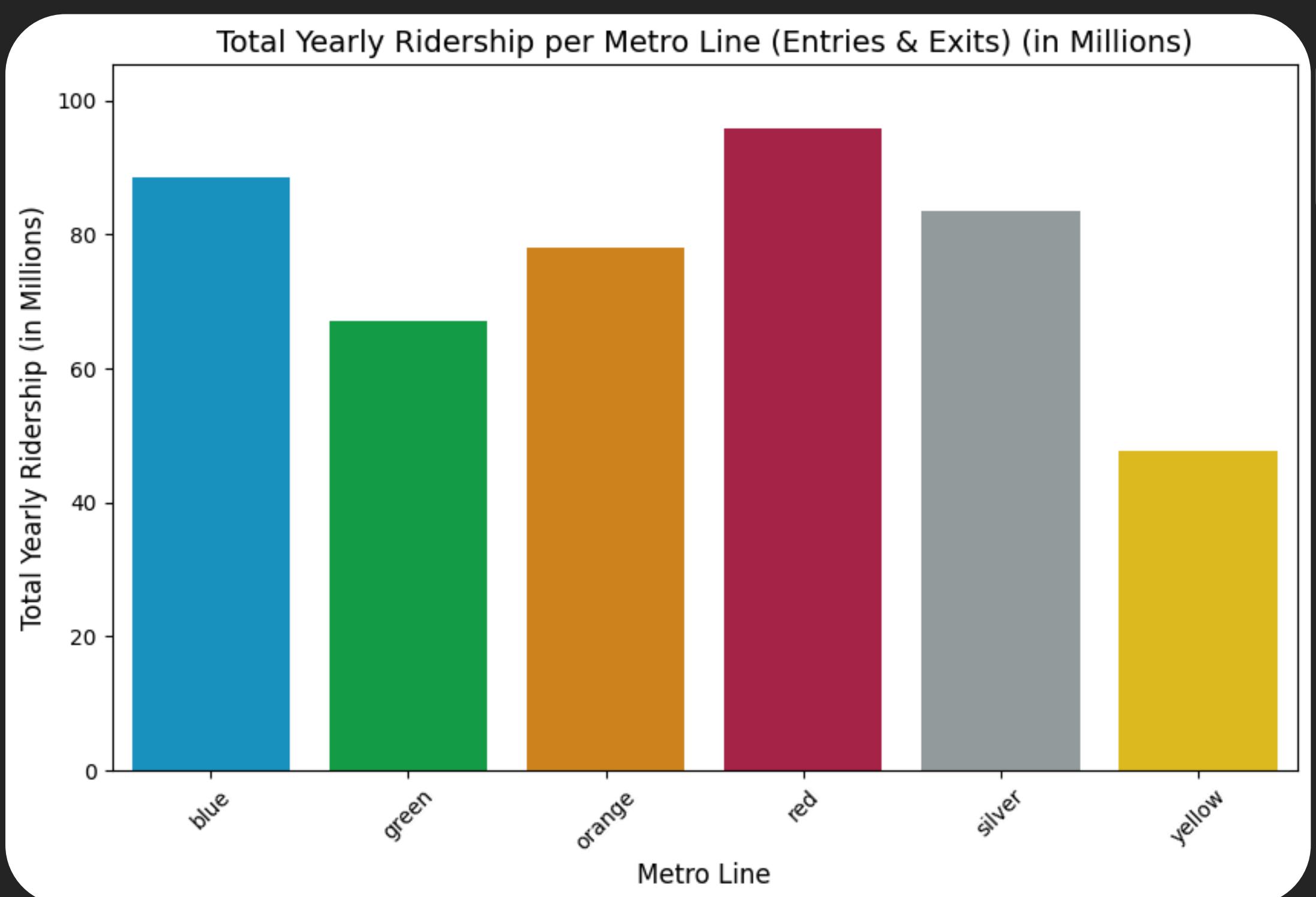
Overall trends for months correspond to average ridership to their respective seasons.

The spikes in April and October show that some months have peaks, but that the distribution of any season is not significantly impacted.

This prompted us to explore if the metro lines followed the same pattern, and if any particular metro line skews any month or season.

Metro Line Analysis:

Which metro lines handle the most ridership overall?



Metro Line	Total Yearly Ridership (in Millions)
blue	88.558285
green	67.031008
orange	78.045014
red	95.759772
silver	83.468273
yellow	47.767306

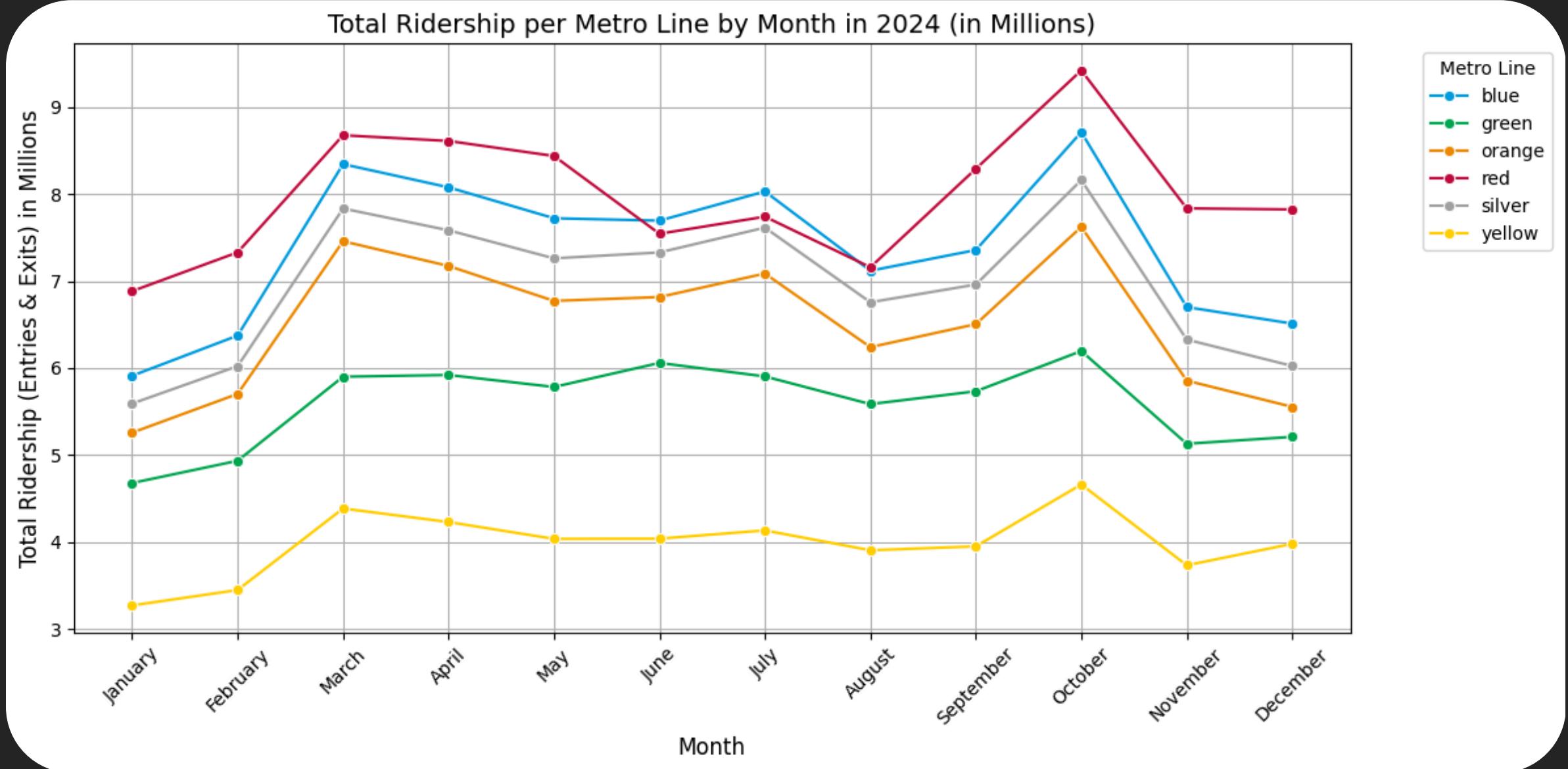
Based on the graph of total ridership in millions, the red line comes in first, and the blue line second.

This helped understand that some lines are more frequented than others,

which set ground for the upcoming line graphs analyzing ridership per line per month.

Metro Line Analysis:

How does ridership fluctuate across metro lines by month?

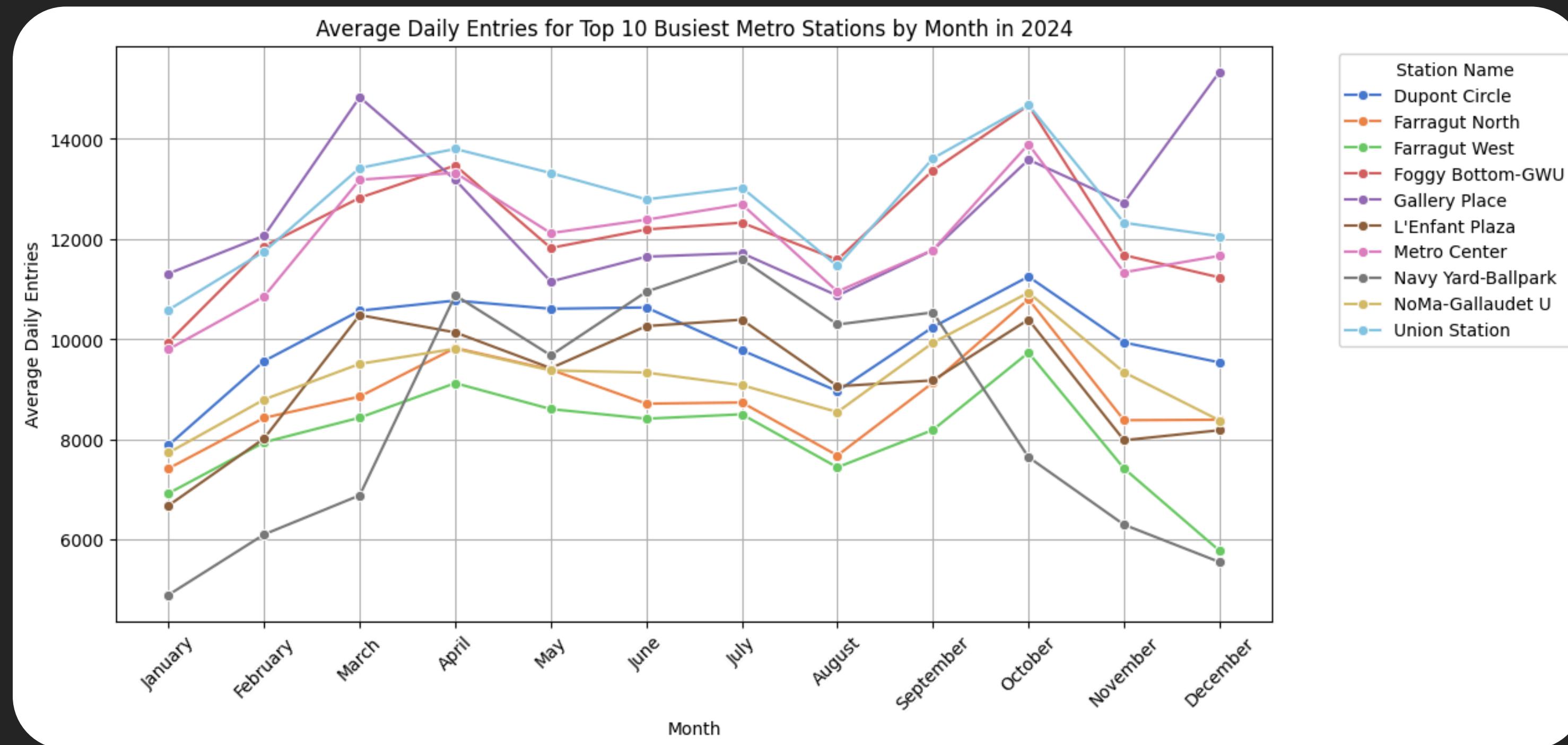


By looking at the total ridership per line for each month, we can see the patterns over time.

This may be beneficial for metro to optimize resources, and for riders to avoid congestion or seek popular attractions, perhaps near particular stations.

Station-Specific Ridership:

What are the top 10 busiest stations each month of the year?



January

Cold, no major holidays

February

Cold, no major holidays

March

Gallery Place reaches peak

Nearby attractions: National Art Gallery, Smithsonian, Arena, cherry blossom sights

April

Union Station reaches peak

Nearby attractions: Memorials, National Museums, cherry blossom sights

May

Hot, traveling out of state for vacation

June

Hot, traveling out of state for vacation

July

Hot, traveling out of state for vacation

August

Hot, traveling out of state for vacation

September

Most stations' ridership inclines

Popular celebrations: State Fair, Jazz Fest Fall-themed festivals & activities, beginning of fall semester & work events

October

Most stations' ridership peaks

Popular celebrations: Oktoberfest, Fall Wine Festival, Halloween festivities, middle of fall semester & work events

November

Most stations' ridership declines

Cold, traveling out of state for Thanksgiving

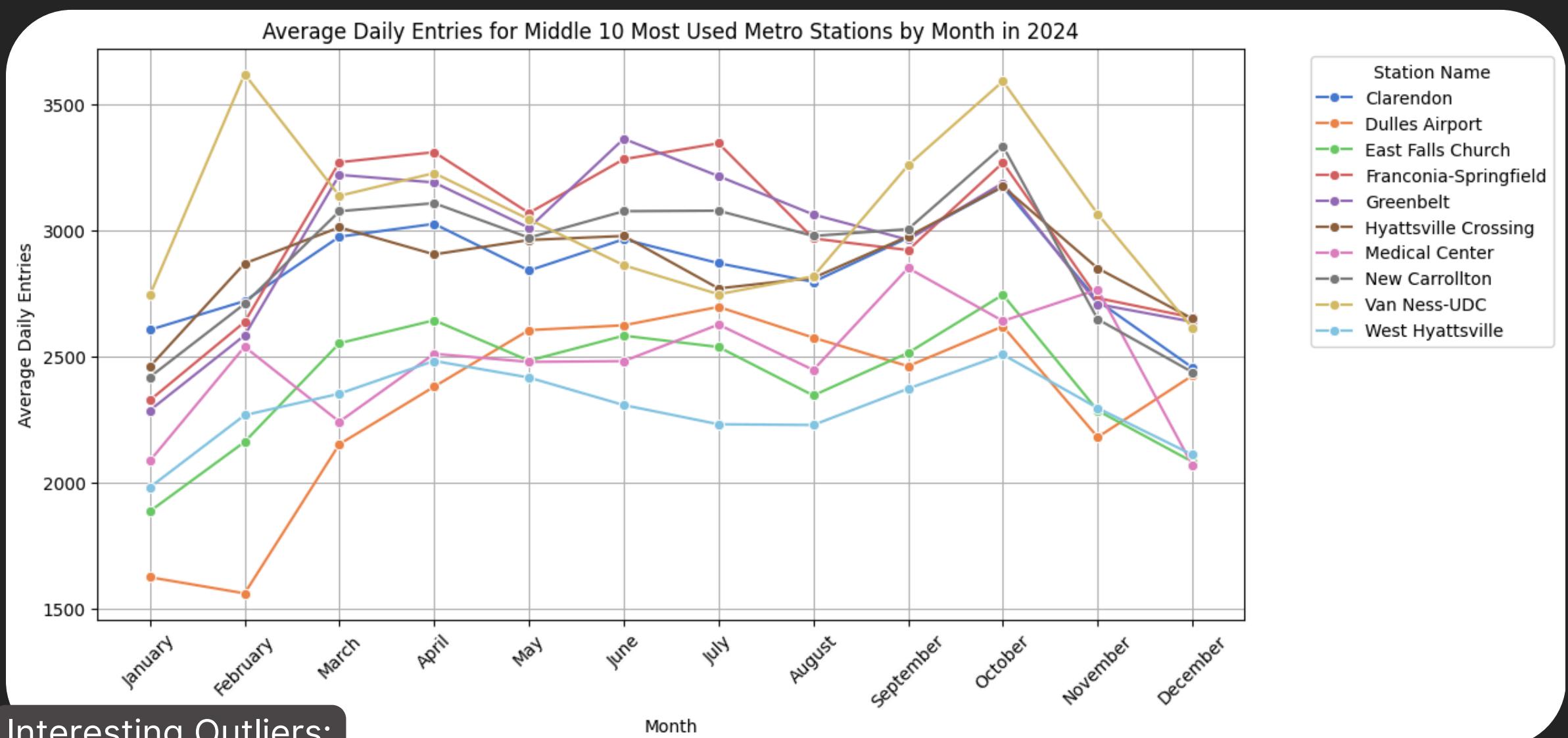
December

Most stations' ridership declines except for Gallery Place

Nearby attractions: Holiday Markets, Christmas festivities, light shows, ice shows

Station-Specific Ridership:

What are the middle 10 busiest stations each month of the year?



Interesting Outliers:

- Peak in Van Ness-UDC station in February

This could be attributed to UDC's Annual Holiday Concert.

- Dip in Dulles Airport in February

Travel around DC tends to be at its lowest during the winter season,

and tourists likely hold off on traveling in February as the following month has the Cherry Blossoms as a major tourism event.

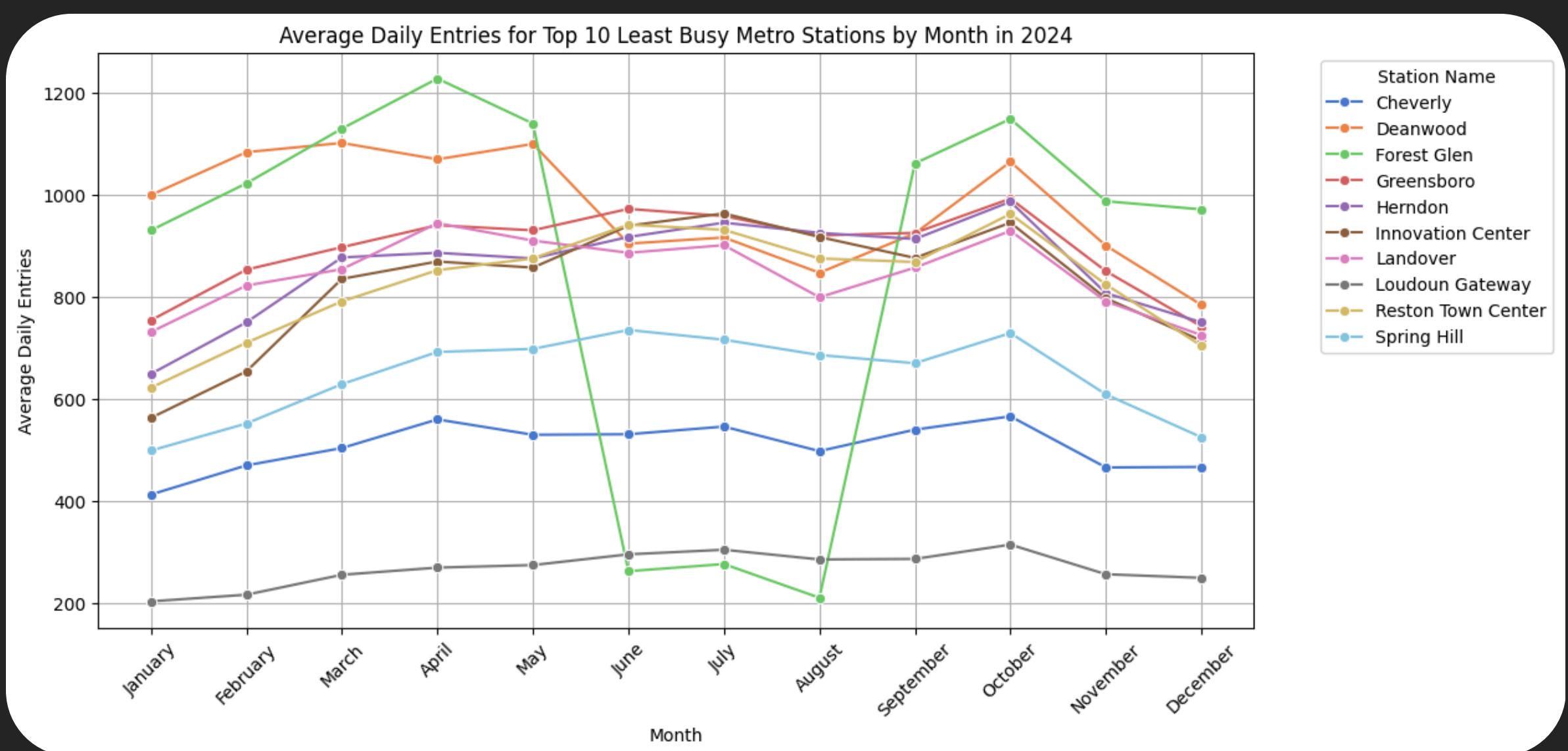
The ten middle-most frequented Metro stations, the peak is less apparent in March, and the summer plateau is less noticeable.

The October spike is still very visible, followed by the steep decrease in ridership into December.

Like the previous graph, the station trends tend to move with each other, but there are some discrepancies in February.

Station-Specific Ridership:

What are the least 10 busiest stations each month of the year?



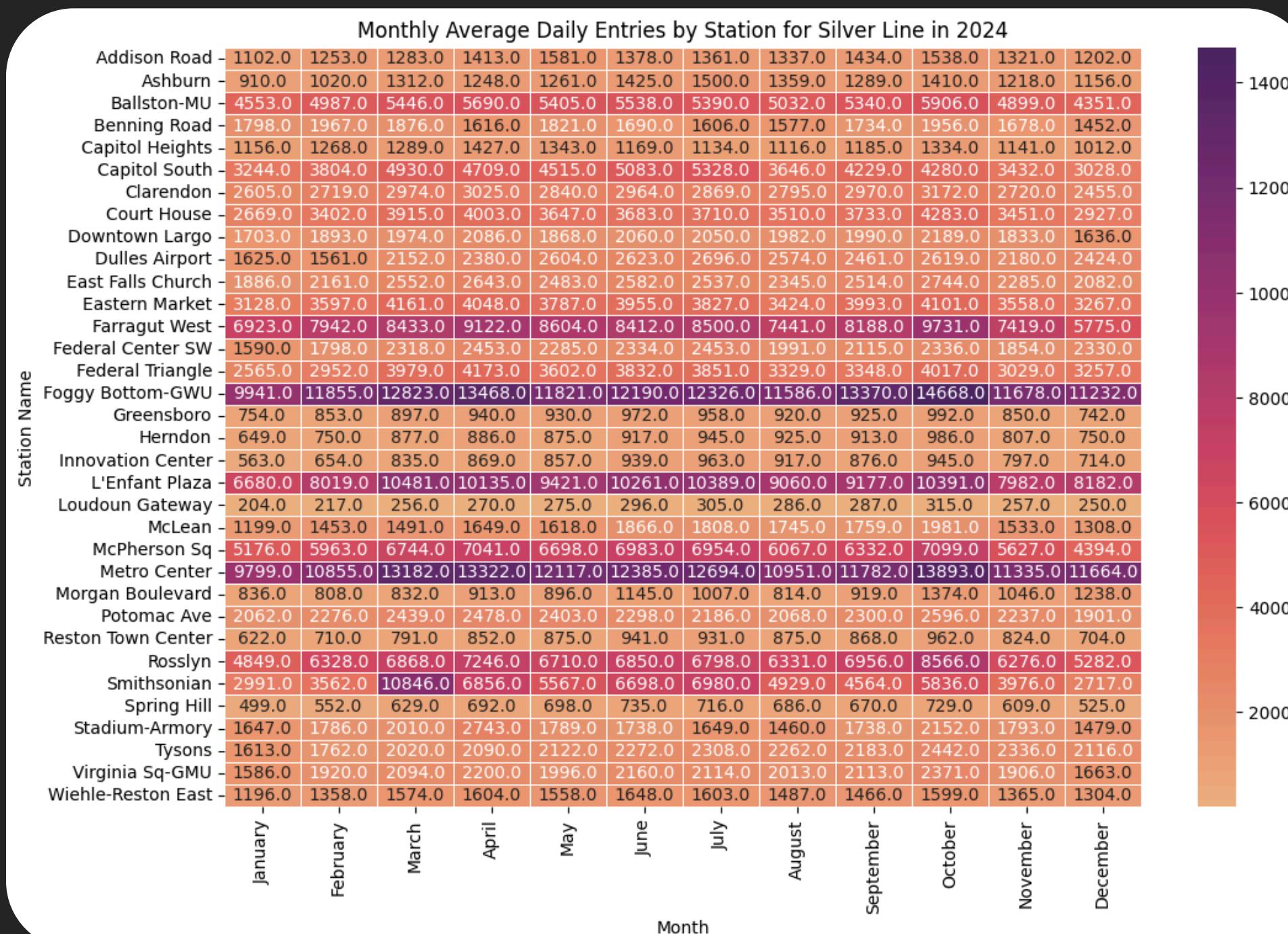
The average daily entries appear to constantly plateau with small peaks in October and overall lower ridership in Winter months.

Note that the Y-axis scale is significantly lower in comparison to the previous two graphs, so discrepancies in ridership patterns are more sensitive in these lines.



Metro Line & Station Type Analysis:

How is ridership distributed across metro stations and lines over time?



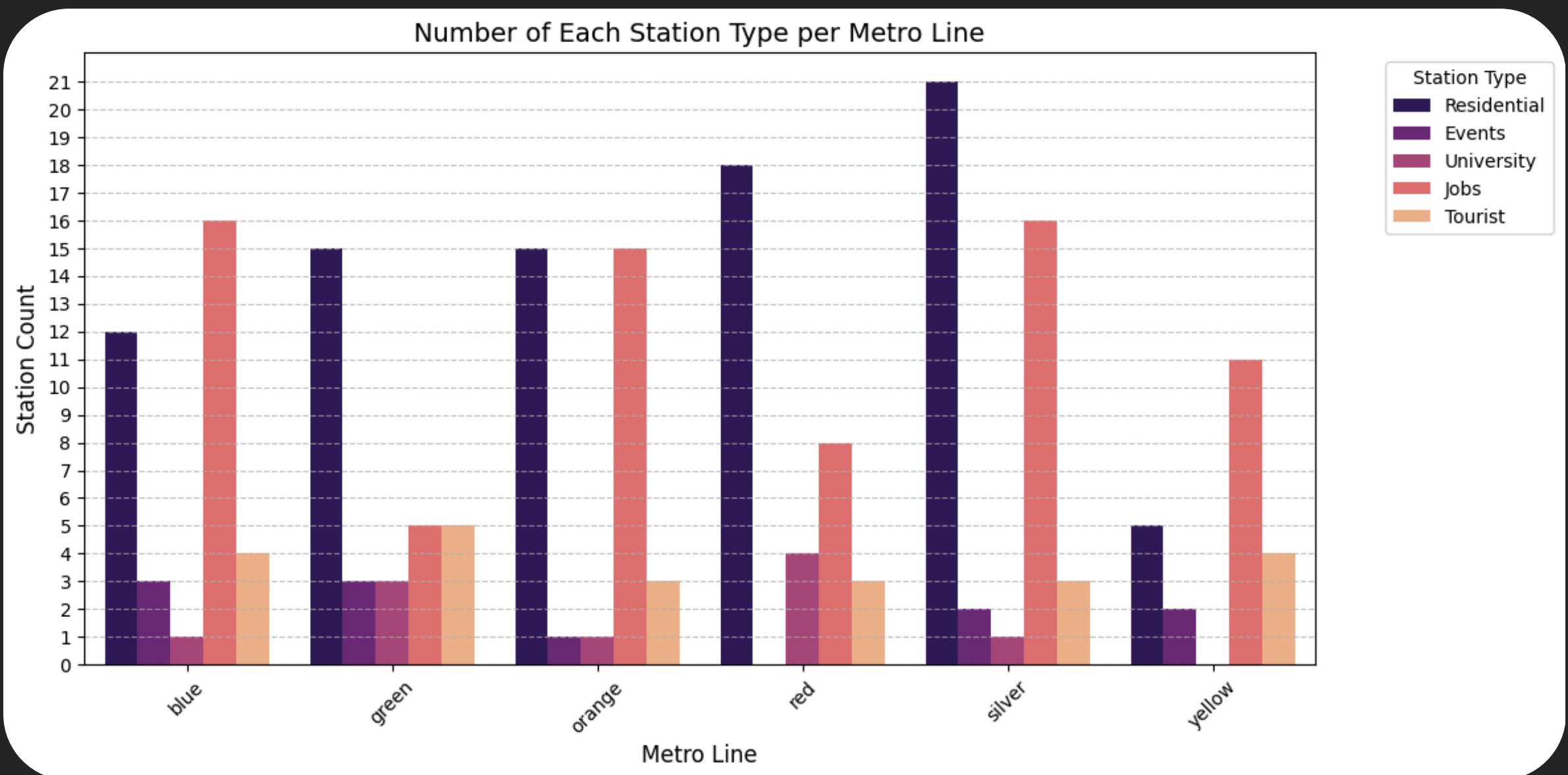
This view allows for easy identification of outliers in each station in which certain events or line closures might impact the overall ridership of a certain line during a particular month.

Outliers are significant, as the previous graphs show how overall metro trends are not unique to certain stations.

This prompted our idea to explore dynamic pricing of advertising in different stations in relation to these spikes by comparing the trends of specific stations to themselves.

Metro Line & Station Type Analysis:

How many stations of each type exist on each metro line?



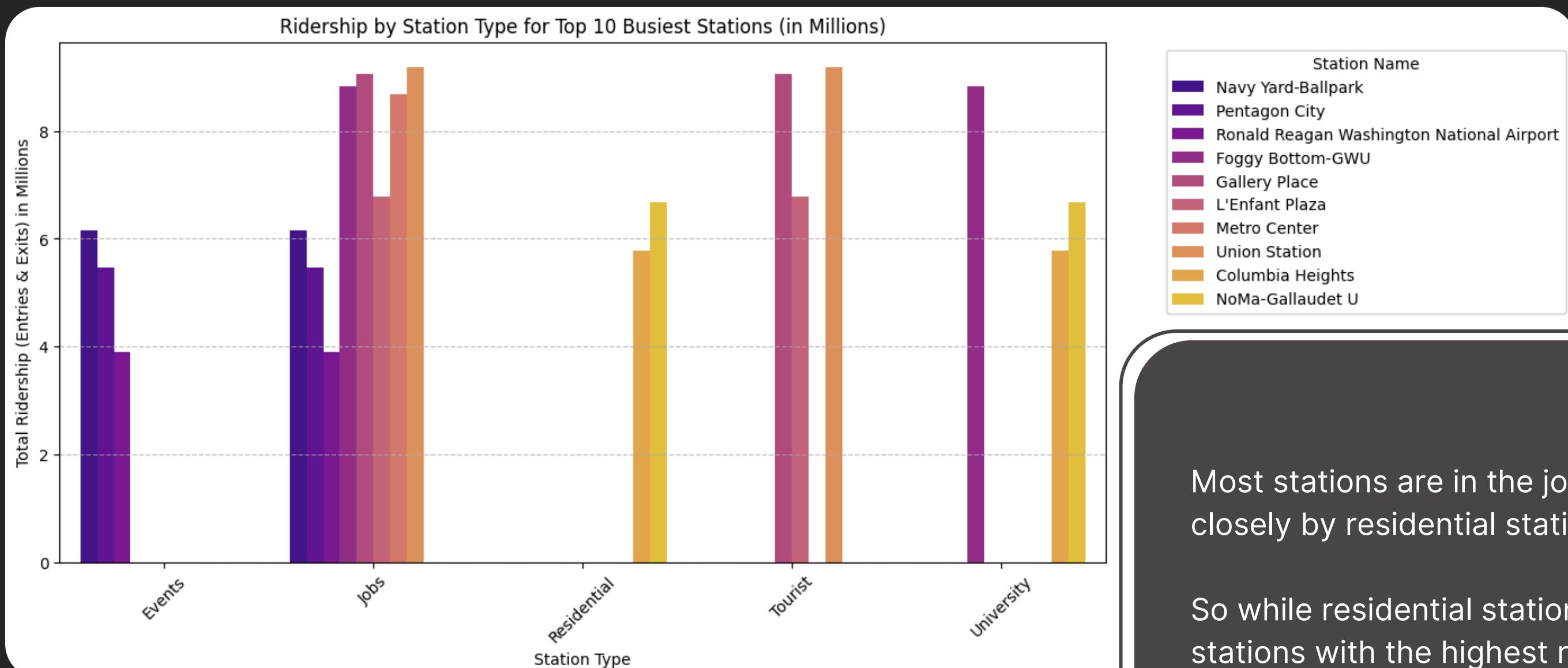
Some lines fit into two categories so we included them in both to provide accurate representation.

Most lines can be categorized with residential or jobs, but jobs has more consistency.

Hence, we should assess station type more closely per station.

Metro Line & Station Type Analysis:

Which station types handle the most traffic?

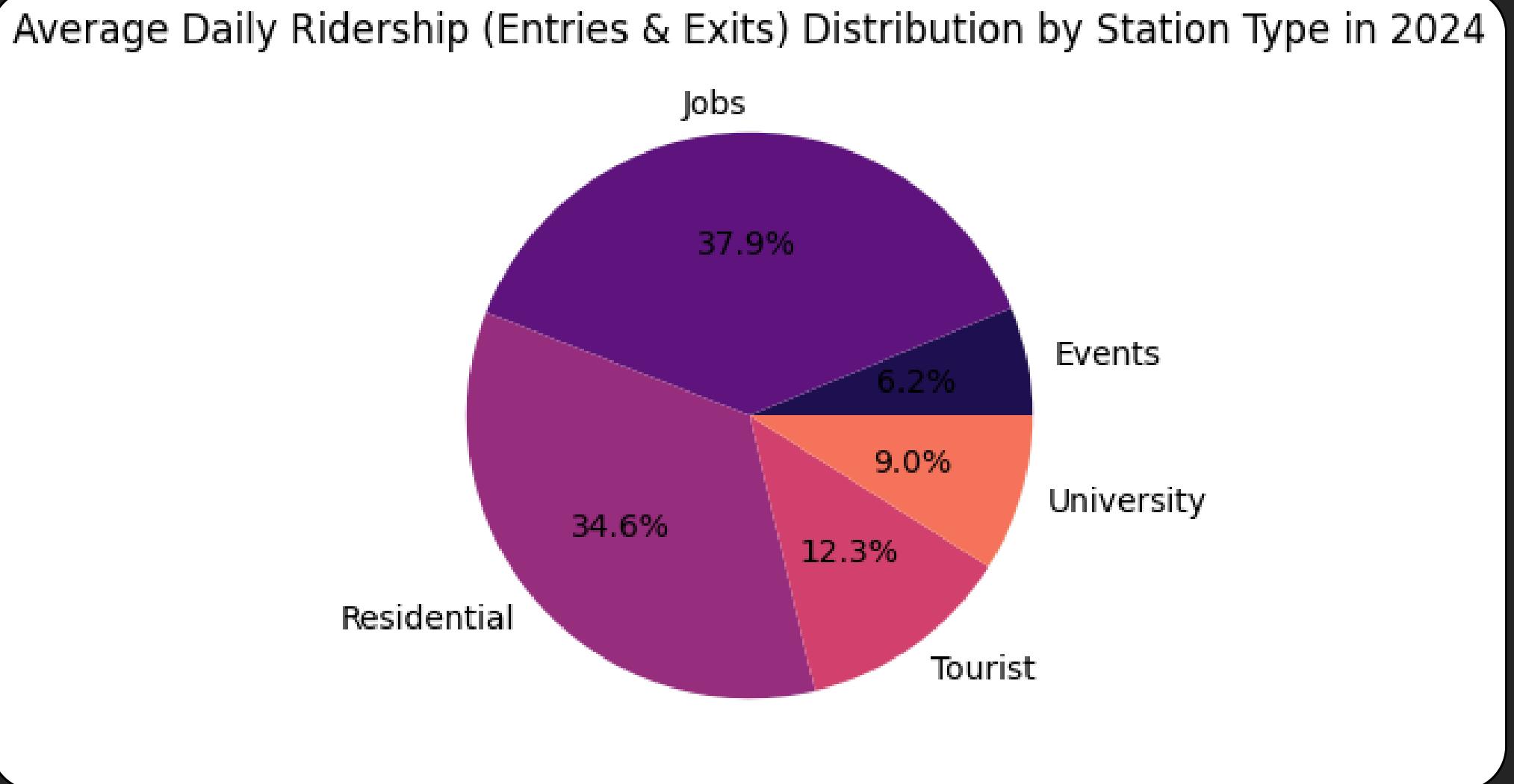


Most stations are in the jobs category, followed closely by residential stations.

So while residential stations are more common, stations with the highest ridership are those used primarily for work-related commutes.

Metro Line & Station Type Analysis:

Which station types handle the most traffic?



Alternative visual of the distributions of stations types across top active stations highlights jobs and residential stations.

People who use residential stations are later use job stations, making the distribution almost equal.

This accounts for people who work in the residential area or live in the “jobs” area.



Interpretation

This is the conclusion to our exploratory data analysis, which we have left it open to interpretation

- for Metro to thoughtfully allocate their resources,
- for businesses to come up with thoughtful marketing strategies around the Metro,
- and even for average Metro users to thoughtfully plan their rides.



**make the most of your
metro experience**

>>> One Step Further

With that said, you can check out the advertising smart pricing tool that we built on our Github, titled [Advertising_Smart_Pricing_Final.ipynb](#)



[maya-mp/
WMATA_Analysis](#)



This project analyzes WMATA Metro ridership trends to optimize transit planning and business strategies through data-driven insights.

3 Contributors 0 Issues 0 Stars 0 Forks

[WMATA_Analysis/Info_Challenge/Initial_Viz/Advertising_Smart_Pricing_Final.ipynb at main · maya-mp/WMATA_Analysis](#)

This project analyzes WMATA Metro ridership trends to optimize transit planning and business strategies through data-driven insights. - maya-mp/WMATA_Analysis

[GitHub](#)

Future Implementations



Data Accuracy

- Research the smaller lines in more depth to ensure proper assignment of station type.
- Standardize the dataset columns (with data types, and pivoting vs unpivoting) across csv's
- Utilize geographical APIs on the events and frequented areas around the stations would allow us to understand the trends with more data-based backing and would allow us to make better predictions on the trends of 2025 if future DC event data is available.

Normalization of everyday_2024_w_metro_stations.csv

- Create a SQL database and build queries and views to broaden the scope
- Connect to Tableau to make dynamic data stories
- Utilize the X, Y, and Address data we have on stations with mapping sheets

Dynamic Pricing Script

- This script provides good ground for future research and implementation of dynamic advertisement pricing within metro lines.
- Since we had limited data for this project, this script can be expanded by adding the data from the past years, not limited to 2024, to further improve the accuracy of the script.

INFO CHALLENGE 2025

Thank you

Team #: IC25069

Team Members:
Maya Patel, Illia Polishchuk, Adrien Rozario

