



Exploratory Data Analysis: Advice for NYC Metro Scheduling

Josh Jingtian Wang

Metis

Jan 25, 2022

Backstory

- Client
 - Metropolitan Transportation Authority (MTA)
- Question
 - What factors potentially affect NYC metro ridership?
- Purpose
 - Modify train scheduling to reduce budget
- Deliverables
 - COVID vs Metro
 - Uber/Lyft vs Metro
 - Weather vs Metro
 - Metro ridership trend in a year/week

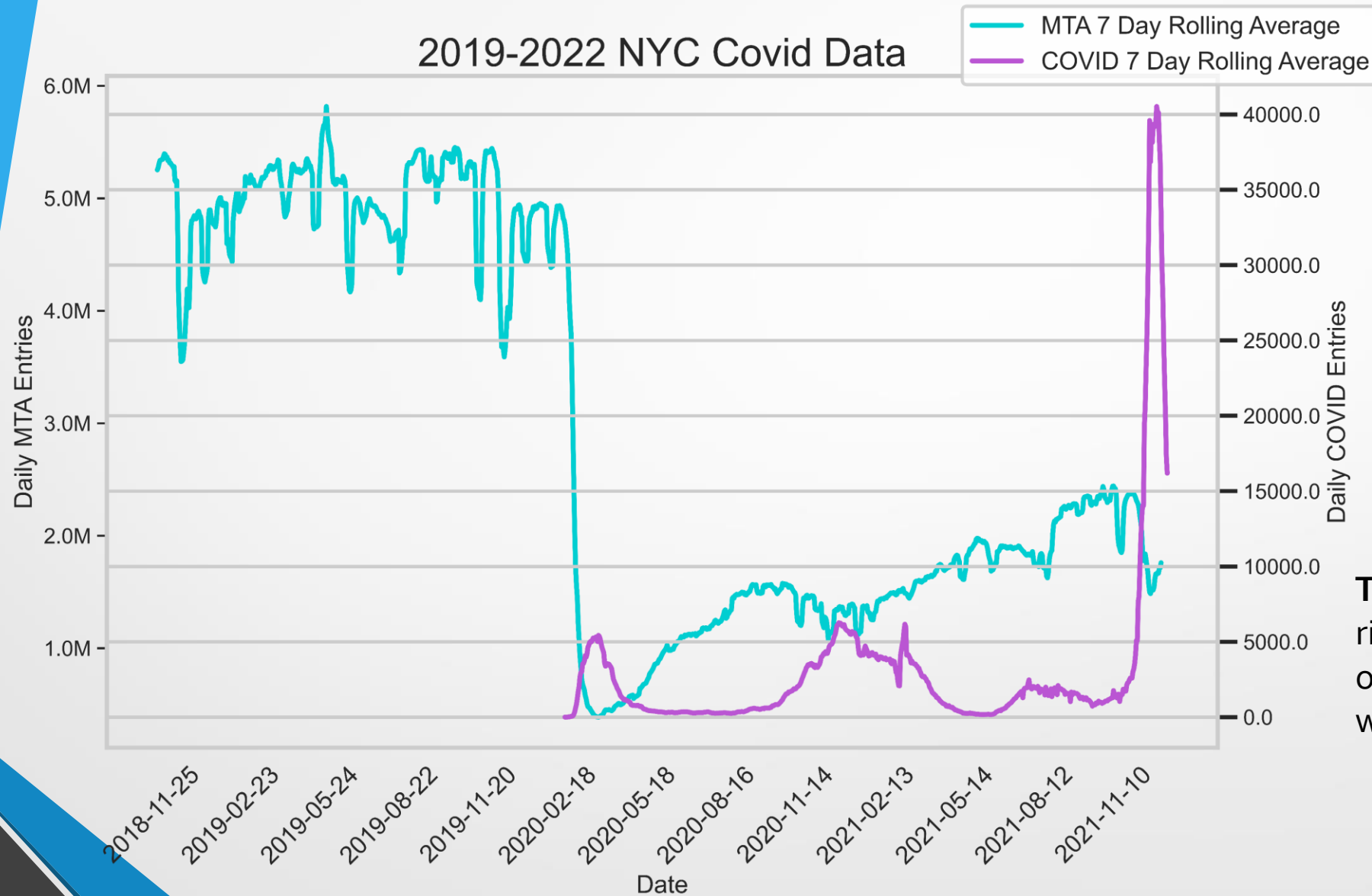
Data

- Metro data
 - <http://web.mta.info/developers/turnstile.html>Question
- COVID data
 - <https://github.com/nytimes/covid-19-data>
- Uber/Lyft data
 - <https://data.cityofnewyork.us/Transportation/FHV-Base-Aggregate-Report/2v9c-2k7f>
- Weather data
 - <https://www.ncdc.noaa.gov/cdo-web/>

Methods

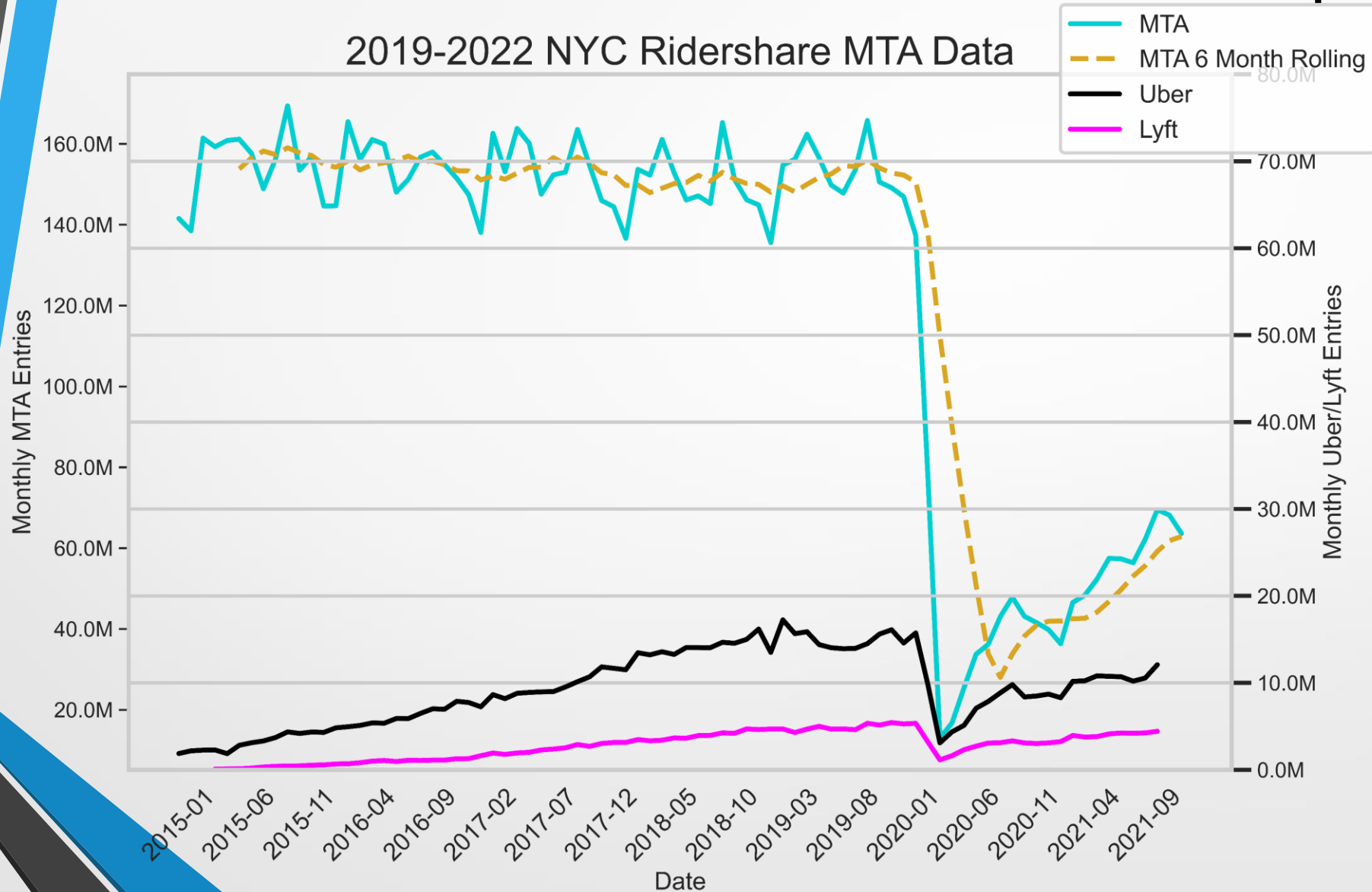
- Data acquisition
 - SQL
 - Python (SQLAlchemy)
- Data Cleaning and Manipulation
 - Python (pandas)
- Visualization
 - Python (matplotlib, seaborn, calplot (calendar heatmaps))

COVID Cases vs Metro Ridership



Takeaway: Rises in metro ridership (re-opening) are often followed by COVID waves.

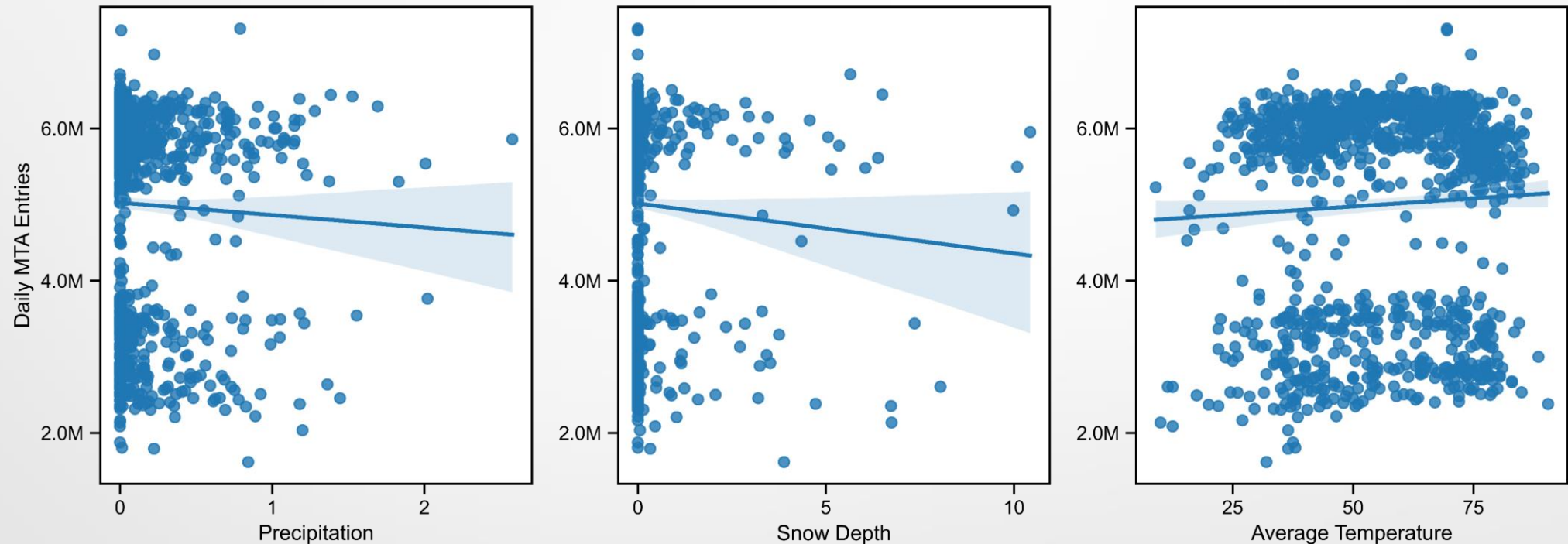
Rideshare Services vs Metro Ridership



Takeaway: The rise of Uber/Lyft may have contributed to the slight decline of pre-COVID NYC metro ridership.

Weather vs Metro Ridership

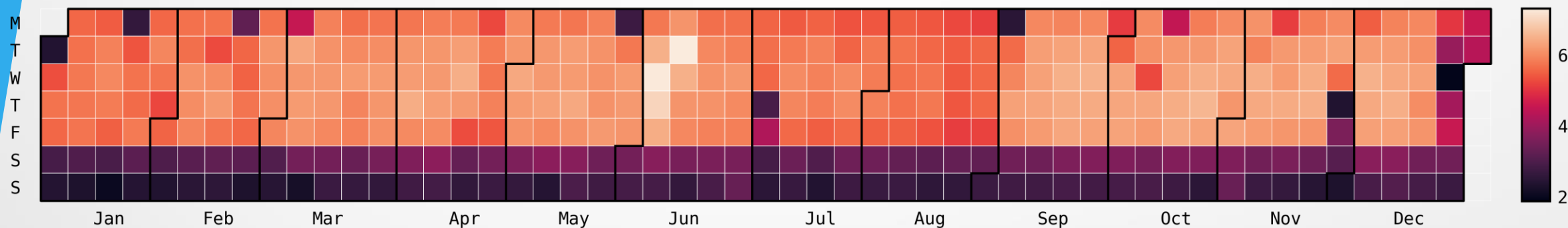
Weather vs MTA Ridership (2017-2019)



Takeaway: NYC metro ridership is negatively correlated with rainfall and snowfall, and positively correlated with ambient temperature. The relationships are not strong.

Daily Metro Ridership Trend (2019)

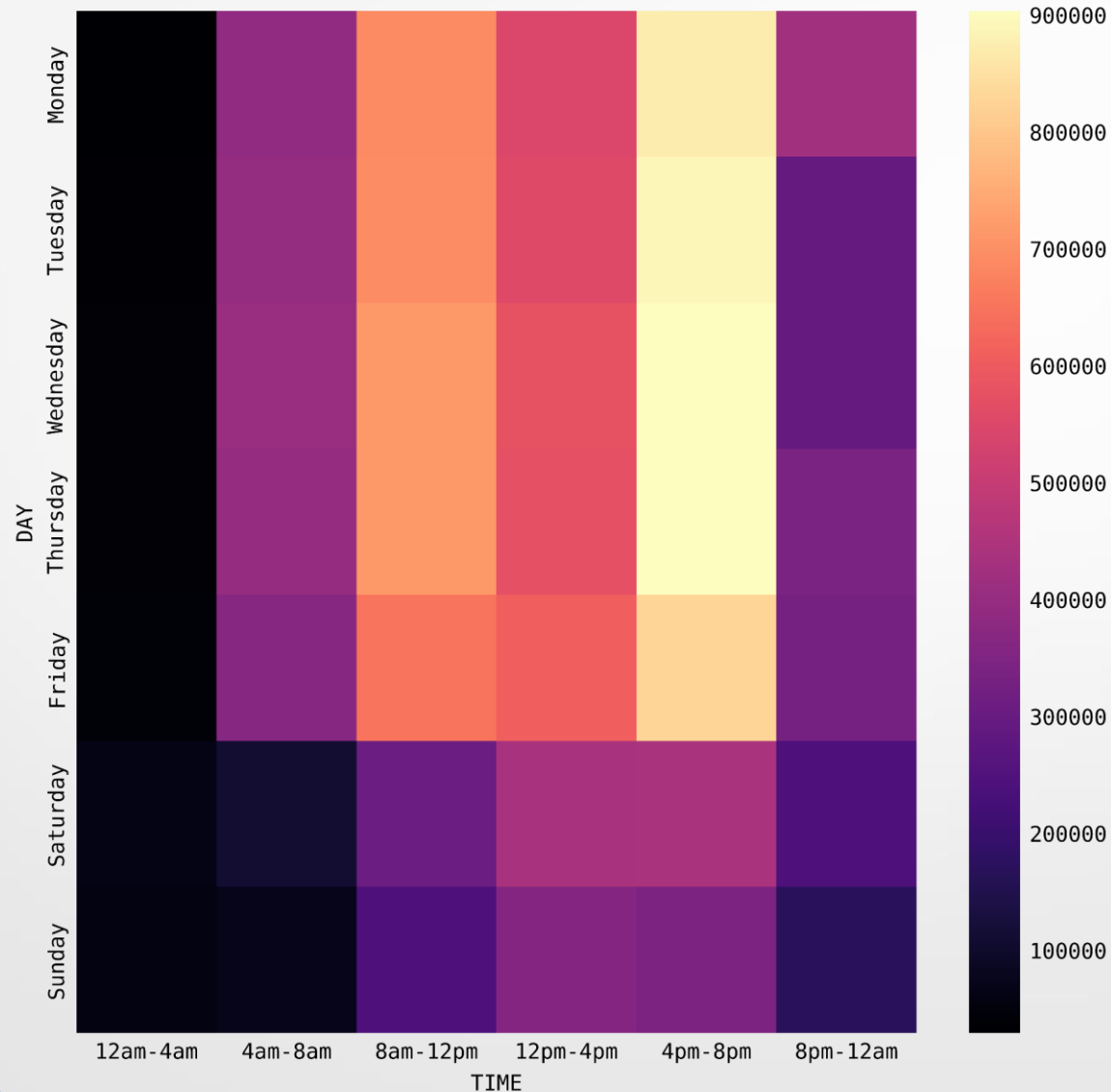
NYC Metro Daily Entries (in millions)
2019



Takeaway:

- Low ridership on holidays, weekends.
- Higher ridership in Spring and Fall than in Winter and Summer.

Hourly Metro Ridership Trend (2019)



Takeaway:

- Low ridership on weekends.
- High ridership from 8am to 8pm on weekdays.

Note:

Removed holidays
Data skewed for March to November

Advice

- COVID waves and rideshare services potentially affect NYC metro ridership
- Weather may not be a strong enough factor to warrant train scheduling change
- Fewer trains can be scheduled on holidays, weekends and at nighttime.

