



Exploratory Data Analysis (EDA) on NYC Transit Data for Optimal Advertisement

Jay Kwon

Metis Data Science and Engineering (flex program)

Module 1 - EDA

October 13, 2021



Motivation:

- Pitching to advertising agency
 - I propose that I can help them make better decisions on ad placement by analyzing NYC transit data
 - Stations, buses, subways, and commuter rails in the Greater New York area
 - more effective ad placements on interior and exterior of these vehicles, walls of the stations, human street teams

Goal:

- perform Exploratory Data Analysis (EDA) on NYC transit data to find insights on how to optimize ad placement
 - Busiest stations, busiest day of the week, seasonal trends
 - Event driven spikes in traffic

Dataset:

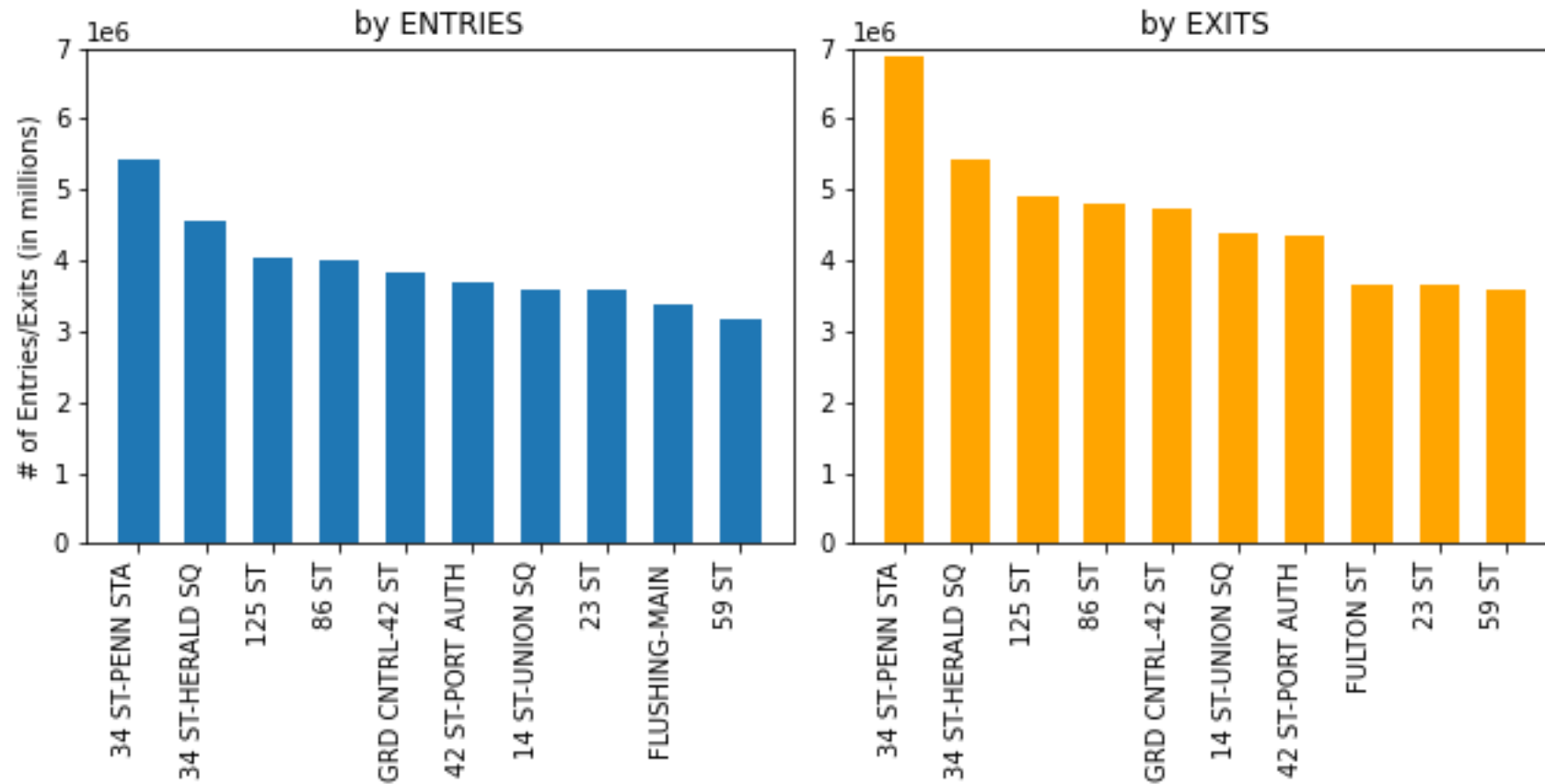
- Turnstile entries and exits data from the Metropolitan Transportation Authority (MTA): <http://web.mta.info/developers/turnstile.html>
- **July 25, 2020 to January 29, 2021 : 27 weeks (188 days)**
 - Turnstile count of entries and exits take every 4 hours
 - 379 Stations, focused on Penn Station

- Why this time period?
- Seasonal trends. How does ridership change...
 - as we move out of summer (Aug) and into the new school year (Sept)
 - onto the holiday season in December, and into the new year

Tools used:

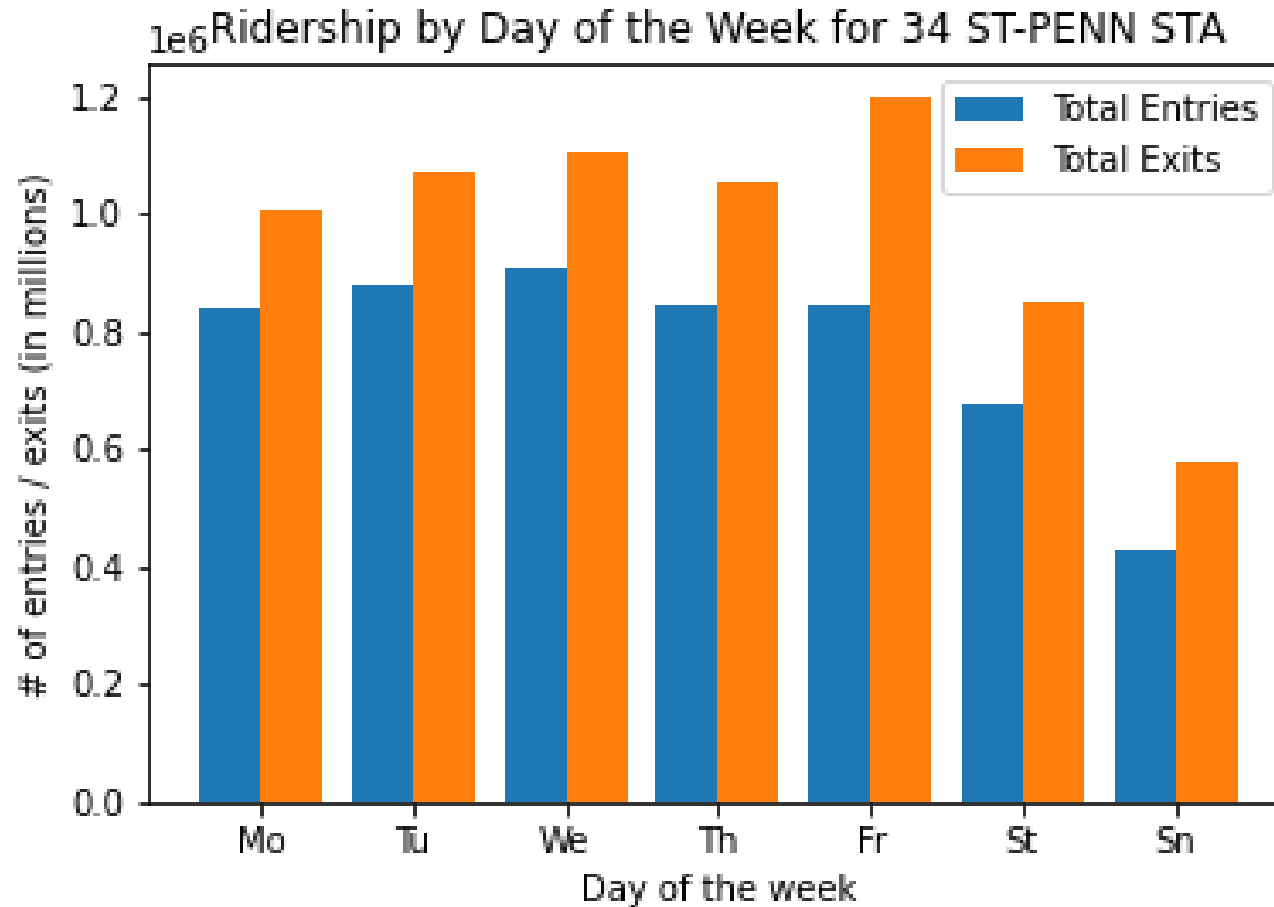
- SQL, SQLAlchemy, Pandas, Numpy, Matplotlib

Top 10 busiest stations (27 week period)



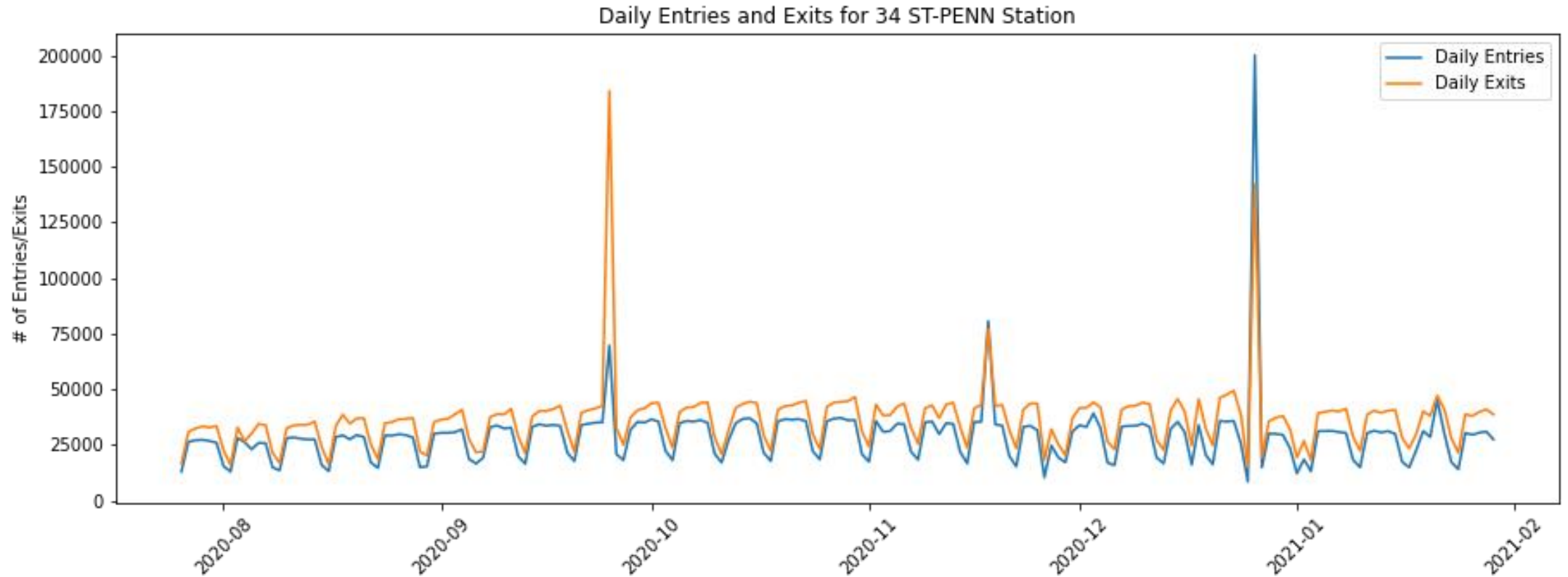
- 9 out of the top 10 stations by Entries is in the top 10 by Exits
- more people exit the stations than enter; NOTE: also true for ALL stations combined, 3.7% that enter, exit at stations not covered by the dataset (e.g. commuter rails in other states)
- Penn Station is #1 for both Entries and Exits

Penn Station



- Weekdays are the busiest: Mo, Tu, We, Th, Fr
- Weekends are less busy: St, Sn
- Spike in Exits on Friday: time to party in the city after work, ride cab back home after subway closes

Penn Station



- Weekly cyclical trend (peak during weekdays, drop on weekends)
- higher exits than entries
- 3 unusual spikes in traffic after capping count at 200,000

Penn Station

Day after Christmas

Announcement of public school shutdown due to 3% Covid test positivity rate

Error in Data Entry?
(further investigation needed)

3 Days After MLK
(work/school starts to pick up)

Top 5 and Bottom 5 daily Entries

DATE	DAILY_ENTRIES	DAILY_EXITS
2020-12-26	200026.0	142202.0
2020-11-18	80617.0	77423.0
2020-09-25	69646.0	184019.0
2021-01-21	45379.0	47194.0
2020-12-03	39225.0	44139.0
...
2020-07-26	13054.0	16710.0
2020-08-02	13005.0	16089.0
2021-01-01	12293.0	19432.0
2020-11-26	10493.0	18319.0
2020-12-25	8543.0	14923.0

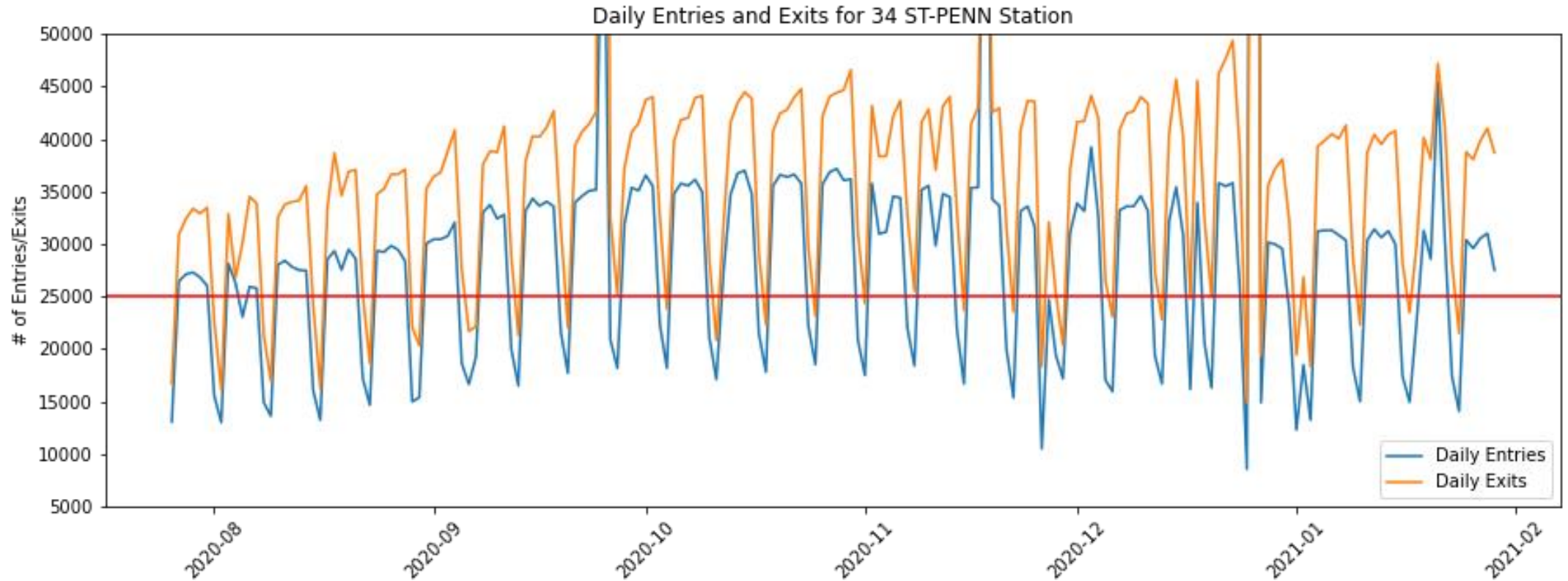
Public holidays celebrated in NY include:

Labor Day Mon, Sep 7, 2020
Columbus Day Mon, Oct 12, 2020
Veterans Day Wed, Nov 11, 2020
Thanksgiving Thu, Nov 26, 2020
Christmas Day Fri, Dec 25, 2020
New Year's Day Fri, Jan 1, 2021
MLK Jr. Day Mon, Jan 18, 2021

Lowest Traffic:

- Summer
- New Years Day
- Thanksgiving
- Christmas

Penn Station: General Trend



- Low Traffic in July and August (end of summer)
- High Traffic in October and November (fall)
- Medium Traffic in January



Recommendations:

- Target the top 10 busiest station by Entries and Exits
- Ad space near Exits are more valuable than those near Entries
- Advertise more during Weekdays especially near Exits for Penn Station
- Day after Christmas is the best day to advertise
- Keep on eye on NYC Covid policies and deploy human street teams when shutdowns are announced (e.g. 3% covid test positivity rate)
- Do not advertise on Thanksgiving, Christmas, and New Years day.