

# Top NYC Subway Stations for Rental Scooter Deployment

Andrew Wong



## Project Goal

- Perform EDA on NYC MTA Turnstile data
- Produce results/findings that can benefit rental scooter companies (Bird, Lime, Lyft etc.)
  - Determine best stations to deploy scooters





# Design

- Focus : Station Exit Data
- Higher exit traffic → Higher potential for rental
- High traffic areas ↻ ↻ Recirculation of scooters
- ↑ Revenue
- ↑ Brand visibility



## Data

- (Pre-Covid) 4/27/2019 to 7/26/2019 (91 days)
- Stations with highest average daily exits
- Top Stations per borough (Bronx, Brooklyn, Queens, Manhattan)



## Data

- Exit counts updated every 4 hours, last count at 20:00
- Scooters collected for recharging starting at 21:00



# Algorithms (Pandas)

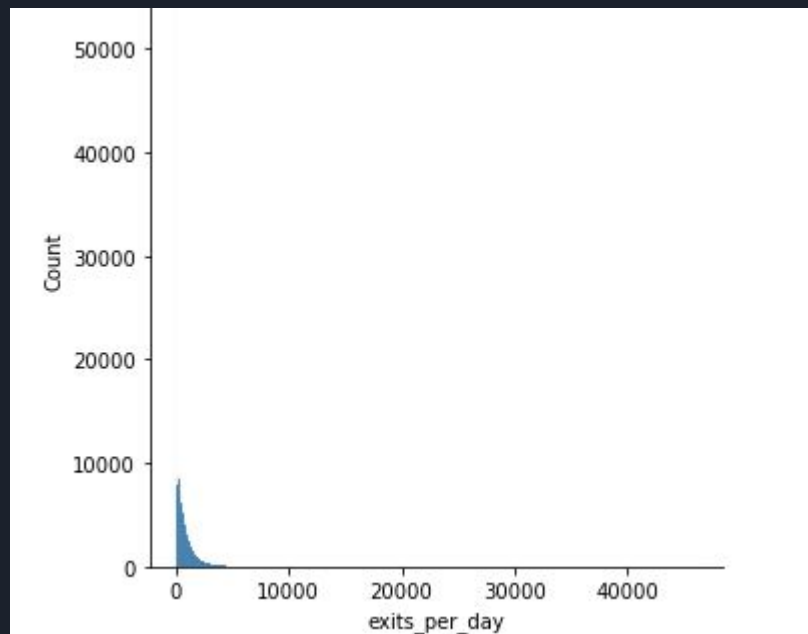
- SCP daily exit count
  - max value - min value
  - dataframe sorted by DESC
  - found inexplicable increases

C008	R099	00-00-00	DEKALB AV	BDNQR	BMT	07/26/2019	00:00:00	REGULAR	0006897017	0008702864
C008	R099	00-00-00	DEKALB AV	BDNQR	BMT	07/26/2019	04:00:00	REGULAR	0006897017	0008702864
C008	R099	00-00-00	DEKALB AV	BDNQR	BMT	07/26/2019	08:00:00	REGULAR	2000093069	1850343553
C008	R099	00-00-00	DEKALB AV	BDNQR	BMT	07/26/2019	12:00:00	REGULAR	2000092817	1850343947
C008	R099	00-00-00	DEKALB AV	BDNQR	BMT	07/26/2019	16:00:00	REGULAR	2000092406	1850344366
C008	R099	00-00-00	DEKALB AV	BDNQR	BMT	07/26/2019	20:00:00	REGULAR	2000091967	1850344989

# Algorithms (Pandas)

```
total_scp_exit.describe()
```

	EXITS_x	EXITS_y	exits_per_day
count	4.135410e+05	4.135410e+05	4.135410e+05
mean	3.615268e+07	3.610136e+07	5.132585e+04
std	2.013725e+08	2.012513e+08	7.086530e+06
min	0.000000e+00	0.000000e+00	0.000000e+00
25%	2.074170e+05	2.064270e+05	1.520000e+02
50%	1.422380e+06	1.420913e+06	4.610000e+02
75%	4.961140e+06	4.958675e+06	1.006000e+03
max	2.124127e+09	2.124125e+09	1.841642e+09





# Algorithms (Pandas)

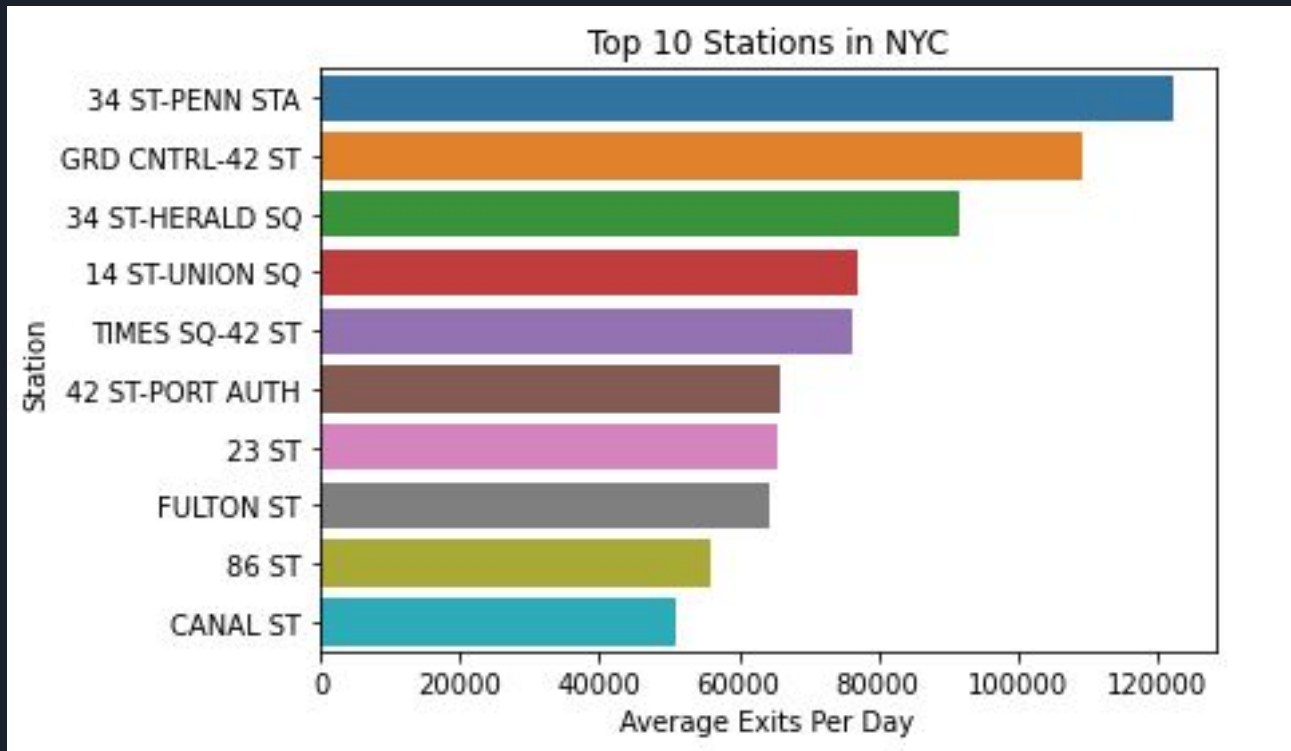
```
#Remove SCP exits_per_day values > 51325  
total_scp_exit2 = total_scp_exit[(total_scp_exit.exits_per_day < 51325)]
```

```
#Find AVG EXITS per day per STATION.  
avg_daily_exit = total_scp_exit2.groupby(['STATION', 'DATE', 'AREA'])[['exits_per_day']].sum()
```

```
#Group by STATION and return the mean of daily exits  
station_daily = avg_daily_exit.groupby(['STATION', 'AREA'])[['exits_per_day']].mean()
```



# Results - All Manhattan







## Results Considerations

- Manhattan is dense with stations
  - Less need for scooter rides
- Busiest areas  $\neq$  best areas for scooters
  - Sidewalk/street conditions not ideal for placement, riding, safety

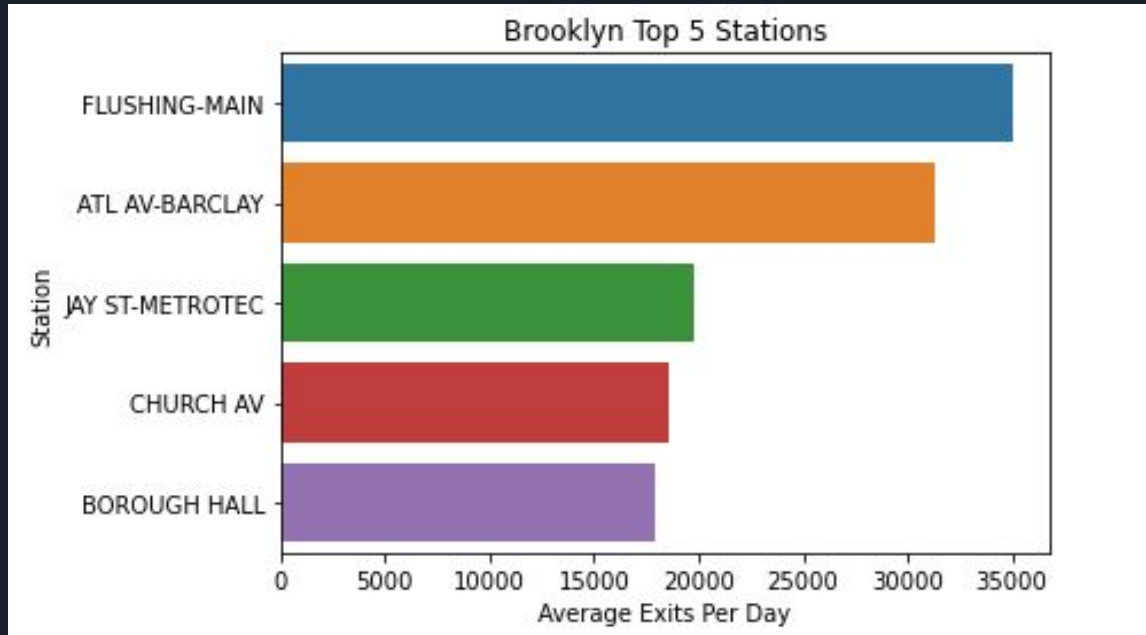


## Algorithms cont.

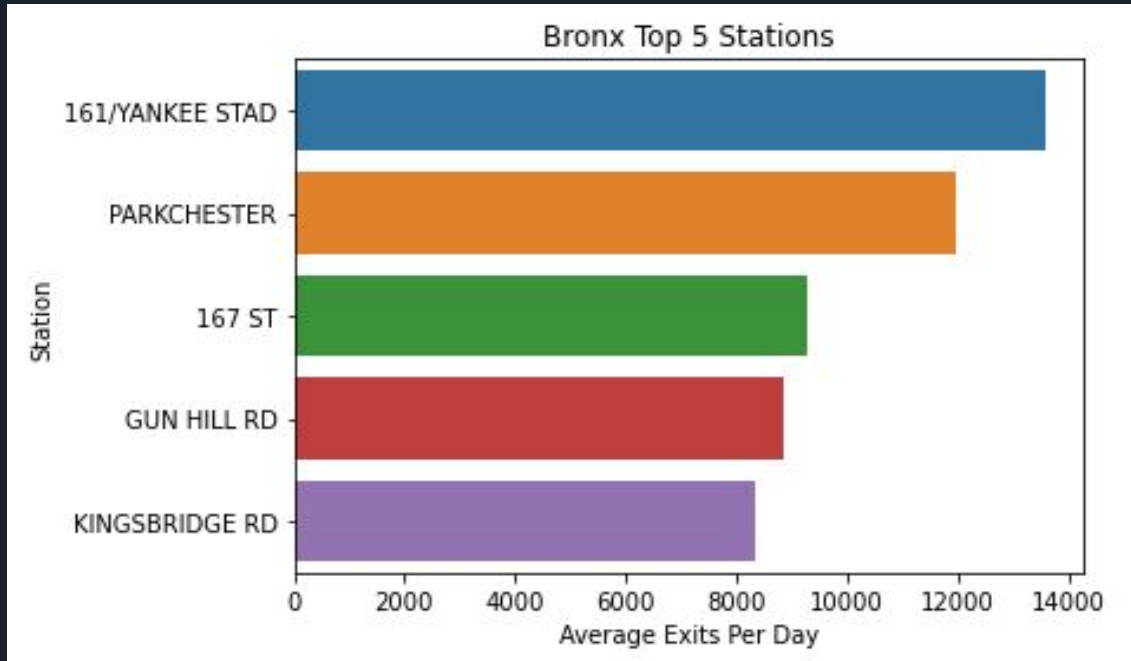
### Finding TOP 5 Stations Per Borough

```
#Masks for each borough of interest  
brooklyn = (station_daily.AREA == 'BROOKLYN')  
bronx = (station_daily.AREA == 'BRONX')  
queens = (station_daily.AREA == 'QUEENS')  
manhattan = (station_daily.AREA == 'MANHATTAN')
```

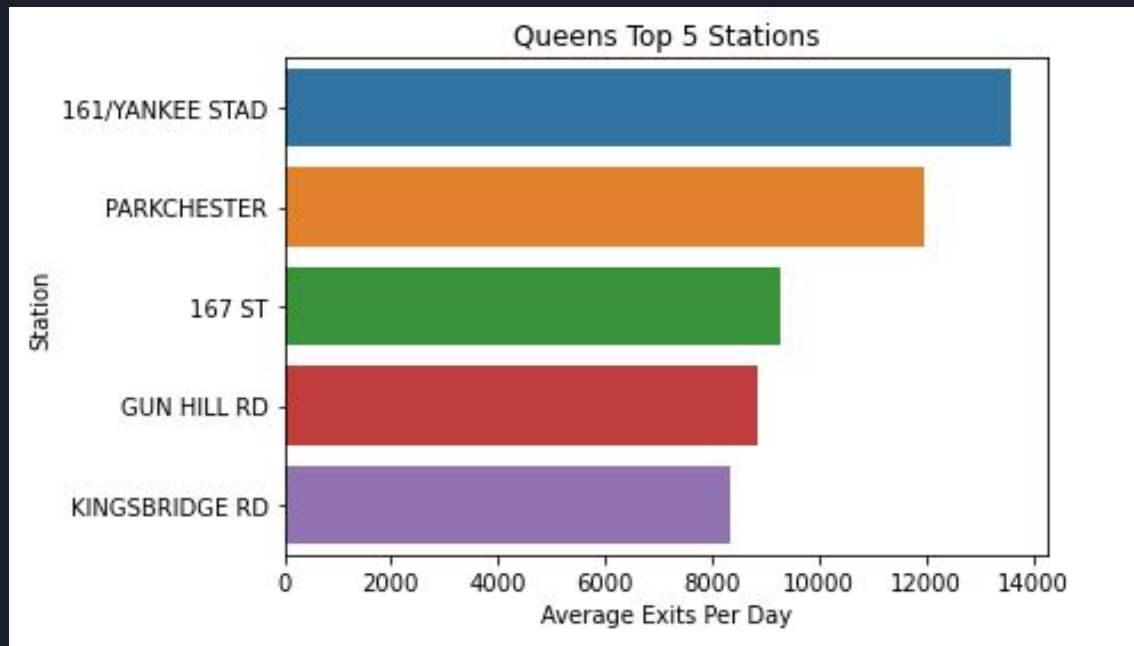
# Results - Top 5 Stations Per Borough



# Results - Top 5 Stations Per Borough



# Results - Top 5 Stations Per Borough





## Tools

- SQL for data concatenation and filtering
- Pandas for data manipulation, analysis, cleaning
- Seaborn for plotting and visualization
- Python to download data, reading data using sqlalchemy engine





## Future Work/Modeling

- Automatically pull new data
- Create predictive models incorporating scooter sales data by station
  - Factoring in holidays, special events