MTA Project – AZ Real Estate

Exploratory Data Analysis

Jay Yee

Agenda

- *Overview and Strategy*

- *Building out the Model*

- *Data Results*

- *Conclusions / Recommendation*

- *Appendix*

Overview and Strategy

Overview:

AZ Real Estate (the "Client"), a local NYC real estate developer, believes that the new norm of remote work culture is here to stay. In turn, the Client wants to better understand how that has impacted subway usage and to use this data as a proxy for changes in lifestyle and housing demands that will help find undervalued opportunities that have not priced in these changes.

The Client's main questions:
1. Confirm if subway stations in traditionally white-color work-places have decreased in demand and if so by how much?
2. Is this drop in public transport consistent across all of New York? Are certain areas not as affected as others?

Building out the Model

What does the raw data look like?

	C/A	UNIT	SCP	STATION	LINENAME	DIVISION	DATE	TIME	DESC	ENTRIES	EXITS	Borough
0	A006	R079	00-00-00	5 AV/59 ST	NQRW	BMT	2019-03-02	01:00:00	REGULAR	3,758,261.0	6444610	M
1	A006	R079	00-00-00	5 AV/59 ST	NQRW	BMT	2019-03-02	05:00:00	REGULAR	3,758,275.0	6444624	M
2	A006	R079	00-00-00	5 AV/59 ST	NQRW	BMT	2019-03-02	09:00:00	REGULAR	3,758,310.0	6444811	M
3	A006	R079	00-00-00	5 AV/59 ST	NQRW	BMT	2019-03-02	13:00:00	REGULAR	3,758,378.0	6445249	M
4	A006	R079	00-00-00	5 AV/59 ST	NQRW	BMT	2019-03-02	17:00:00	REGULAR	3,758,617.0	6445578	M

What is considered a 'station'?
- Has to have its own unique entrance and exit and serviced by specific subway lines.
- There were several stations that have the same name in the data but different subway lines ran through them. For example:
    - Two stations labeled '86 ST' but serviced by the R and 1 line. The R-line station is actually located in the Upper East Side while the 1-line station is in the Upper West Side

	C/A	UNIT	SCP	STATION	LINENAME	DATE	ENTRIES	PREV_DATE	PREV_ENTRIES	DAILY_ENTRIES
123956	C025	R215	00-05-01	86 ST	R	2021-08-12	458,752.0	2021-08-11	458,752.0	0.0
731845	R165	R167	01-00-03	86 ST	1	2021-08-04	7,928.0	2021-08-03	7,529.0	399.0

Building out the Model

ENTRIES vs DAILY_ENTRIES

- ENTRIES is a datapoint provided by the MTA and shows an aggregate of all entries through a specific period
- To calculate only the passenger entries for that day, a new data set called PREV_ENTRIES was created that took the previous day's entries.

- To calculate DAILY_ENTRIES:
  - DAILY_ENTRIES = ENTRIES – PREV_ENTRIES

Data filtered by turnstile:

|	C/A	UNIT	SCP	STATION	LINENAME	DATE	ENTRIES	Borough	PREV_DATE	PREV_ENTRIES	DAILY_ENTRIES
1	A006	R079	00-00-00	5 AV/59 ST	NQRW	2019-03-03	3,759,045.0	M	2019-03-02	3,758,261.0	784.0
2	A006	R079	00-00-00	5 AV/59 ST	NQRW	2019-03-04	3,759,690.0	M	2019-03-03	3,759,045.0	645.0
3	A006	R079	00-00-00	5 AV/59 ST	NQRW	2019-03-05	3,761,124.0	M	2019-03-04	3,759,690.0	1,434.0
4	A006	R079	00-00-00	5 AV/59 ST	NQRW	2019-03-06	3,762,715.0	M	2019-03-05	3,761,124.0	1,591.0
5	A006	R079	00-00-00	5 AV/59 ST	NQRW	2019-03-07	3,764,345.0	M	2019-03-06	3,762,715.0	1,630.0
6	A006	R079	00-00-00	5 AV/59 ST	NQRW	2019-03-08	3,765,987.0	M	2019-03-07	3,764,345.0	1,642.0
7	A006	R079	00-00-00	5 AV/59 ST	NQRW	2019-03-09	3,767,541.0	M	2019-03-08	3,765,987.0	1,554.0
8	A006	R079	00-00-00	5 AV/59 ST	NQRW	2019-03-10	3,768,406.0	M	2019-03-09	3,767,541.0	865.0
9	A006	R079	00-00-00	5 AV/59 ST	NQRW	2019-03-11	3,768,954.0	M	2019-03-10	3,768,406.0	548.0
10	A006	R079	00-00-00	5 AV/59 ST	NQRW	2019-03-12	3,770,567.0	M	2019-03-11	3,768,954.0	1,613.0

Data filtered by station / line name:

|	STATION	LINENAME	DAILY_ENTRIES	ENTRIES	Borough
0	1 AV	L	3,109,217.0	14,471,608.0	M
1	103 ST	1	2,091,736.0	1,798,471.0	M
2	103 ST	6	2,316,813.0	36,655,681.0	M
3	103 ST	BC	899,196.0	13,967,624.0	M
4	103 ST-CORONA	7	3,677,694.0	15,470,589.0	Q
5	104 ST	A	316,260.0	182,364.0	Q
6	104 ST	JZ	458,263.0	6,026,022.0	Q
7	110 ST	6	1,894,045.0	4,587,421.0	M
8	111 ST	7	2,128,613.0	16,646,021.0	Q
9	111 ST	A	451,492.0	50,331,790.0	Q

Results

Top 25 Trafficked Stations in NYC - Pre Covid
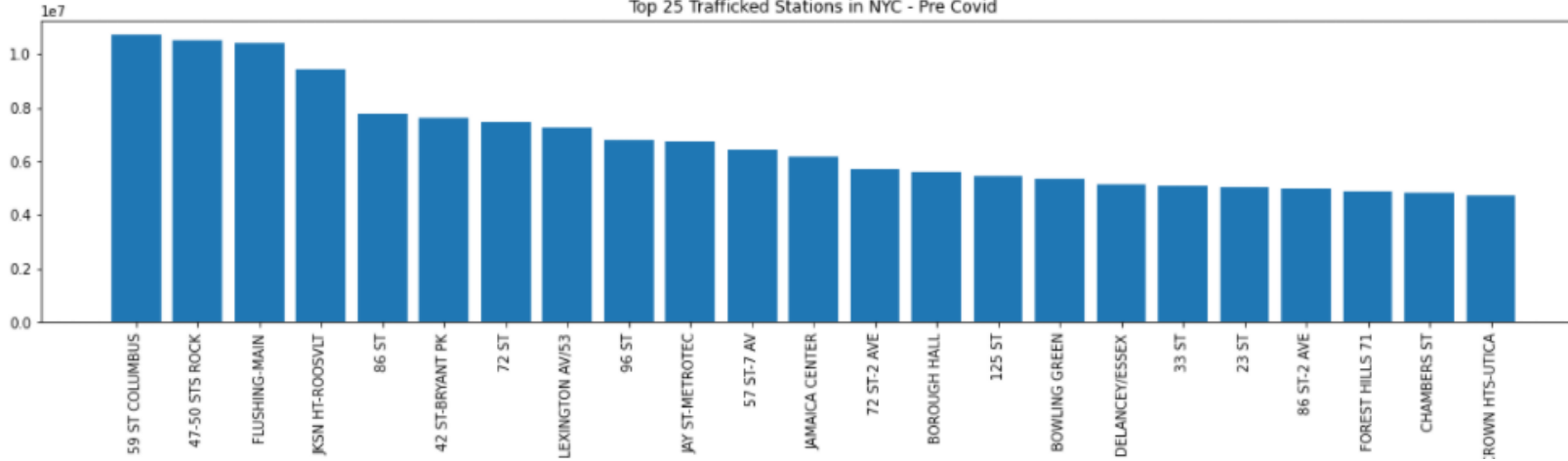
Top 25 Trafficked Stations in NYC - Post Covid
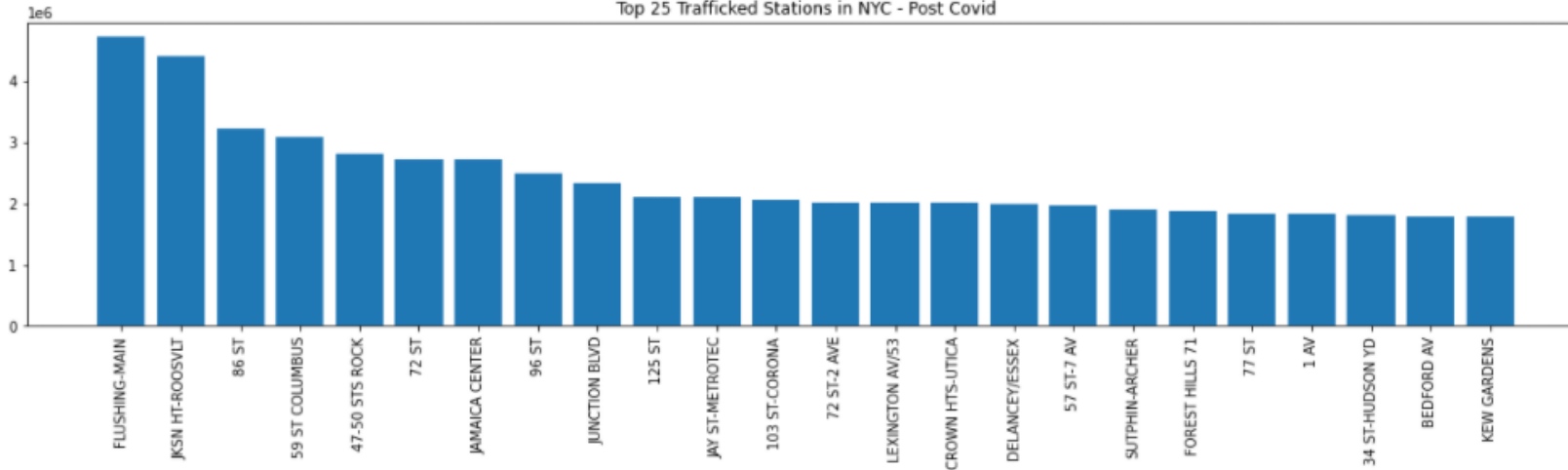
Major Takeaways

- Pre-covid, the top 25 busiest stations ranged from 10.7MM passengers to 4.7MM passengers over the 6-month span.
  - The average station traffic was 6.6MM

- During the same time span, post-covid passenger traffic ranged from 4.7MM to 1.8MM passengers.
  - The average station traffic was 3.4MM

- Although there has been a drop in traffic, the top stations remain mostly static with only 8 new stations for both pre- and post-covid

Results

Average % Decrease in Station Traffic by Borough
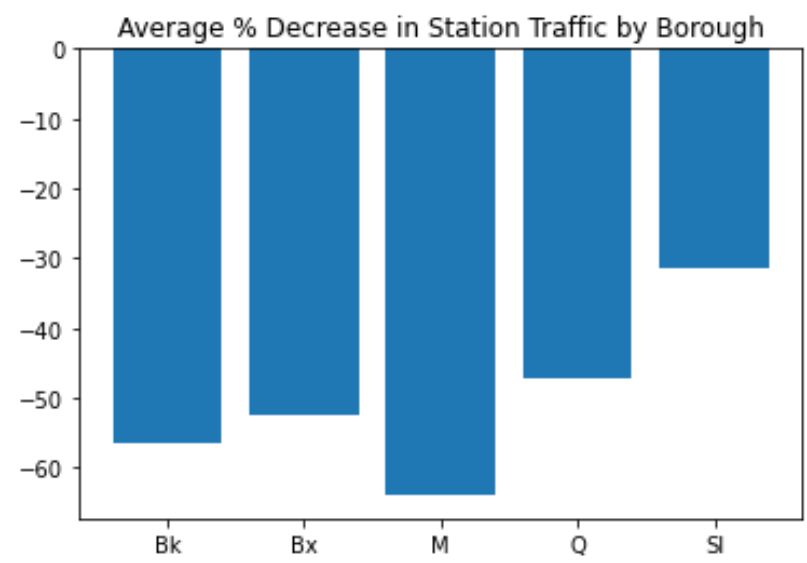
Borough
Bk	-56.314230769230775
Bx	-52.51942028985508
M	-64.00974358974364
Q	-47.23099999999998
SI	-31.48

Takeaways

- Average decrease in station traffic is 56%.
- Only three stations saw an *increase* in traffic compared to the same period, 2019.
    - *ASTORIA BLVD was under construction from March 2019 – Dec 2019.

- Manhattan stations had an average decrease of 64%, significantly above the other boroughs.
    - Reflects the impact of the new remote work culture.
    - Essential workers (service industry, healthcare, construction, etc.) make up a larger population of the other boroughs.

	STATION	LINENAME	DAILY_ENTRIES_x	ENTRIES_x	Borough	DAILY_ENTRIES_y	ENTRIES_y	CHANGE
156	ASTORIA BLVD	NQW	130,171.0	6,478,005.0	Q	845,325.0	6,572,220.0	549.4
10	111 ST	J	162,828.0	3,744,683.0	Q	190,427.0	3,938,689.0	16.95
398	TOMPKINSVILLE	1	256,051.0	576,891.0	SI	278,123.0	637,944.0	8.62

Conclusions / Recommendation

- Confirm if subway stations in traditionally white-color work-places have decreased in demand and if so by how much?
Yes, there has been a material decrease in subway usage specifically within Manhattan. The average decrease in subway traffic for Manhattan stations is 64%.

- Is this drop in public transport consistent across all of New York? Are certain areas not as affected as others?
Yes, 3 out of the 5 boroughs in NYC saw at least a 50% decrease in subway ridership. However, the most affected were Manhattan subways. This could be a combination of factors, most likely because of the new remote work culture and also the fact that there are more essential workers living in the outer boroughs than in Manhattan.

Recommendations

Based on the data, there has been a material decrease in subway usage across all boroughs in NYC. More specifically, the data confirms Manhattan has been more adversely affected by the pandemic than other boroughs, likely from the new remote work culture.

It is recommended that new development opportunities in other boroughs be explored.

Appendix

Top 25 Results

		Pre-Covid			Post-Covid	
	STATION	LINENAME	DAILY_ENTRIES	STATION	LINENAME	DAILY_ENTRIES
1	59 ST COLUMBUS	ABCD1	10,704,229	FLUSHING-MAIN	7	4,716,338
2	47-50 STS ROCK	BDFM	10,498,341	JKSN HT-ROOSVLT	EFMR7	4,408,243
3	FLUSHING-MAIN	7	10,410,452	86 ST	456	3,226,781
4	JKSN HT-ROOSVLT	EFMR7	9,424,987	59 ST COLUMBUS	ABCD1	3,092,007
5	86 ST	456	7,796,712	47-50 STS ROCK	BDFM	2,811,104
6	42 ST-BRYANT PK	BDFM7	7,608,385	72 ST	123	2,729,092
7	72 ST	123	7,463,459	JAMAICA CENTER	EJZ	2,728,077
8	LEXINGTON AV/53	EM6	7,251,724	96 ST	123	2,484,268
9	96 ST	123	6,778,498	JUNCTION BLVD	7	2,326,298
10	JAY ST-METROTEC	ACF	6,755,289	125 ST	ACBD	2,103,238
11	57 ST-7 AV	NQRW	6,449,788	JAY ST-METROTEC	ACF	2,098,411
12	JAMAICA CENTER	EJZ	6,182,378	103 ST-CORONA	7	2,066,141
13	72 ST-2 AVE	Q	5,686,175	72 ST-2 AVE	Q	2,015,865
14	BOROUGH HALL	2345R	5,589,295	LEXINGTON AV/53	EM6	2,007,313
15	125 ST	ACBD	5,462,211	CROWN HTS-UTICA	34	2,005,299
16	BOWLING GREEN	45	5,334,958	DELANCEY/ESSEX	FJMZ	1,980,276
17	DELANCEY/ESSEX	FJMZ	5,128,890	57 ST-7 AV	NQRW	1,965,504
18	33 ST	6	5,075,698	125 ST	456	1,909,715
19	125 ST	456	5,064,545	SUTPHIN-ARCHER	EJZ	1,895,714
20	23 ST	6	5,056,085	FOREST HILLS 71	EFMR	1,887,476
21	23 ST	FM	4,961,471	77 ST	6	1,835,543
22	86 ST-2 AVE	Q	4,958,864	1 AV	L	1,831,366
23	FOREST HILLS 71	EFMR	4,893,004	34 ST-HUDSON YD	7	1,810,447
24	CHAMBERS ST	ACE23	4,839,089	BEDFORD AV	L	1,783,904
25	CROWN HTS-UTICA	34	4,746,537	KEW GARDENS	EF	1,782,594
			164,121,064			59,501,014

Filtering Code
(16 stations in total were filtered out)

```python
#Filtering PATH stations in Manhattan that weren't caught by the 'tourist hotspot' filter
post_daily_station.loc[(post_daily_station['STATION'] == '34 ST-PENN STA') , 'LINENAME'] = 'PATH'
post_daily_station.loc[(post_daily_station['STATION'] == 'PATH NEW WTC') , 'LINENAME'] = 'PATH'
post_daily_station.loc[(post_daily_station['STATION'] == 'THIRTY THIRD ST') , 'LINENAME'] = 'PATH'
post_daily_station.loc[(post_daily_station['STATION'] == 'PAVONIA/NEWPORT') , 'LINENAME'] = 'PATH'

#Grand Central Station does not have a PATH stop but is still a major train station
post_daily_station.loc[(post_daily_station['STATION'] == 'GRD CNTRL-42 ST') , 'LINENAME'] = 'PATH'

#Filtering PATH stations in New Jersey
post_daily_station.loc[(post_daily_station['STATION'] == 'NEWARK BM BW') , 'LINENAME'] = 'PATH'
post_daily_station.loc[(post_daily_station['STATION'] == 'NEWARK C') , 'LINENAME'] = 'PATH'
post_daily_station.loc[(post_daily_station['STATION'] == 'NEWARK HM HE') , 'LINENAME'] = 'PATH'
post_daily_station.loc[(post_daily_station['STATION'] == 'PAVONIA/NEWPORT—*') , 'LINENAME'] = 'PATH'
post_daily_station.loc[(post_daily_station['STATION'] == 'NEWARK HW BMEBE') , 'LINENAME'] = 'PATH'
post_daily_station.loc[(post_daily_station['STATION'] == 'JOURNAL SQUARE') , 'LINENAME'] = 'PATH'
post_daily_station.loc[(post_daily_station['STATION'] == 'GROVE STREET') , 'LINENAME'] = 'PATH'
post_daily_station.loc[(post_daily_station['STATION'] == 'HARRISON') , 'LINENAME'] = 'PATH'
post_daily_station.loc[(post_daily_station['STATION'] == 'EXCHANGE PLACE') , 'LINENAME'] = 'PATH'

#Data set also includes aggregate bus data that which will be filtered out
post_daily_station.loc[(post_daily_station['STATION'] == 'CITY / BUS') , 'LINENAME'] = 'PATH'
```

Appendix

Change in ridership results

	STATION	LINENAME	DAILY_ENTRIES_x	ENTRIES_x	Borough	DAILY_ENTRIES_y	ENTRIES_y	CHANGE
156	ASTORIA BLVD	NQW	130,171.0	6,478,005.0	Q	845,325.0	6,572,220.0	549.4
10	111 ST	J	162,828.0	3,744,683.0	Q	190,427.0	3,938,689.0	16.95
398	TOMPKINSVILLE	1	256,051.0	576,891.0	SI	278,123.0	637,944.0	8.62
38	168 ST	1AC	997,474.0	294,241.0	M	938,437.0	401,067.0	-5.92
260	FLUSHING AV	G	395,944.0	528,501.0	Bk	330,647.0	632,866.0	-16.49
62	215 ST	1	305,041.0	181,902.0	M	238,886.0	420,257.0	-21.69
167	AVENUE U	N	608,353.0	3,828,820.0	Bk	436,836.0	4,144,718.0	-28.19
180	BEACH 60 ST	A	370,128.0	3,036,315.0	Q	263,355.0	140,496.0	-28.85
159	AVENUE H	BQ	899,568.0	663,519.0	Bk	632,928.0	38,266.0	-29.64
310	KINGS HWY	N	891,522.0	4,225,967.0	Bk	617,586.0	4,494,421.0	-30.73
15	121 ST	JZ	400,198.0	3,345,311.0	Q	262,546.0	3,558,149.0	-34.4
254	EASTN PKWY-MUSM	23	793,683.0	2,799,934.0	Bk	518,857.0	3,146,780.0	-34.63
417	WOODHAVEN BLVD	JZ	760,313.0	4,749,601.0	Q	488,881.0	5,161,237.0	-35.7
177	BEACH 25 ST	A	306,358.0	33,489.0	Q	192,796.0	346,185.0	-37.07
49	18 AV	N	719,438.0	7,829,574.0	Bk	447,823.0	8,049,676.0	-37.75
273	FT HAMILTON PKY	N	763,125.0	2,822,226.0	Bk	471,797.0	2,909,307.0	-38.18
107	59 ST	NRW	2,083,262.0	8,781,021.0	Bk	1,278,848.0	9,453,754.0	-38.61
161	AVENUE J	BQ	1,215,290.0	9,268,178.0	Bk	741,315.0	9,508,030.0	-39.0
44	174-175 STS	BD	1,041,980.0	85,391.0	Bx	635,285.0	651,090.0	-39.03
30	155 ST	BD	657,722.0	8,881,358.0	M	391,876.0	9,372,756.0	-40.42
136	82 ST-JACKSON H	7	2,856,287.0	129,765.0	Q	1,698,233.0	461,101.0	-40.54
0	1 AV	L	3,109,217.0	14,471,608.0	M	1,831,366.0	15,532,793.0	-41.1
25	145 ST	1	1,632,704.0	135,134,675.0	M	954,216.0	135,403,636.0	-41.56
8	111 ST	7	2,128,613.0	16,646,021.0	Q	1,236,845.0	18,004,051.0	-41.89
305	JUNCTION BLVD	7	4,036,166.0	13,074,599.0	Q	2,326,298.0	13,861,663.0	-42.36
264	FORDHAM RD	BD	2,151,729.0	4,030,546.0	Bx	1,234,858.0	4,918,611.0	-42.61
85	36 AV	NQW	1,075,460.0	17,598,778.0	Q	608,296.0	17,836,415.0	-43.44
416	WINTHROP ST	25	992,543.0	6,486,503.0	Bk	560,569.0	7,084,470.0	-43.52
4	103 ST-CORONA	7	3,677,694.0	15,470,589.0	Q	2,066,141.0	16,354,845.0	-43.82
256	ELMHURST AV	MR	2,405,702.0	958,514.0	Q	1,350,073.0	1,034,636.0	-43.88