Maximizing Advertising Reach Using MTA Data

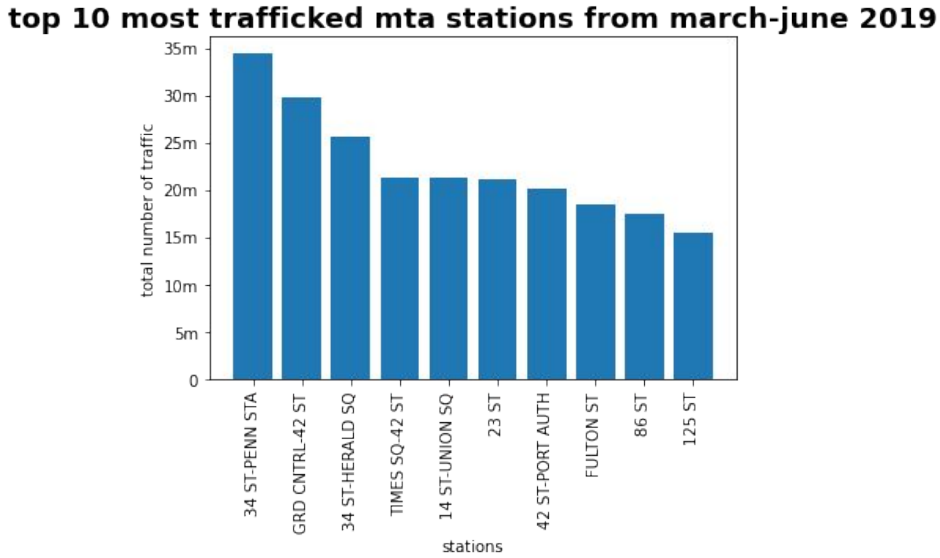
Objectives & Goals

- NY-based matcha company is coming out with a new bottled beverage.
- They want to roll out a series of digital advertisements through subway stations in NY.
- When and where should they advertise to maximize product exposure?

Methodology

- Using data from March-June 2019
  - Looking at top 10 most trafficked stations
  - Weekday vs weekend traffic
  - Peak rush hours
- Importing, cleaning, analyzing data
  - SQLite/DB Browser
  - SQLAlchemy
  - Pandas DataFrame
    - Looking for duplicates, missing, outliers
    - Biggest issues were the entry/exit counter and figuring out how to calculate time intervals
  - Visualization - MPL & Seaborn

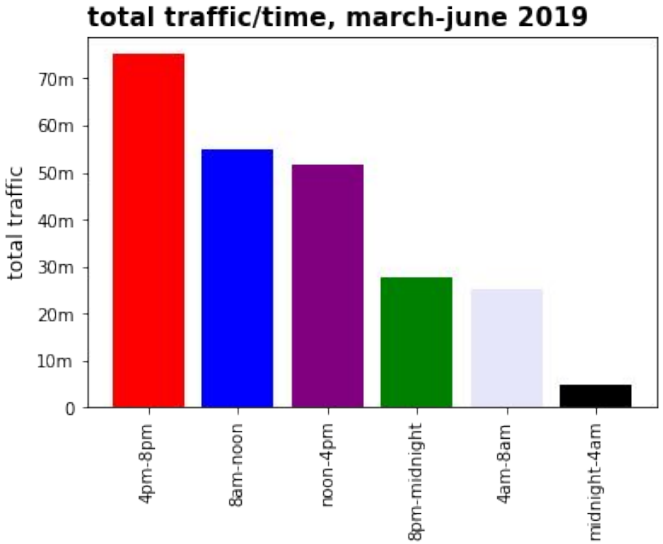
```python
def hourly_entries(max_counter):
    num_entries = mta_df["ENTRIES"] - mta_df["PREV_ENTRIES"]
    num_entries = num_entries.abs()
    num_entries = num_entries.apply(lambda x: 0 if x>max_counter else x)
    return num_entries
```

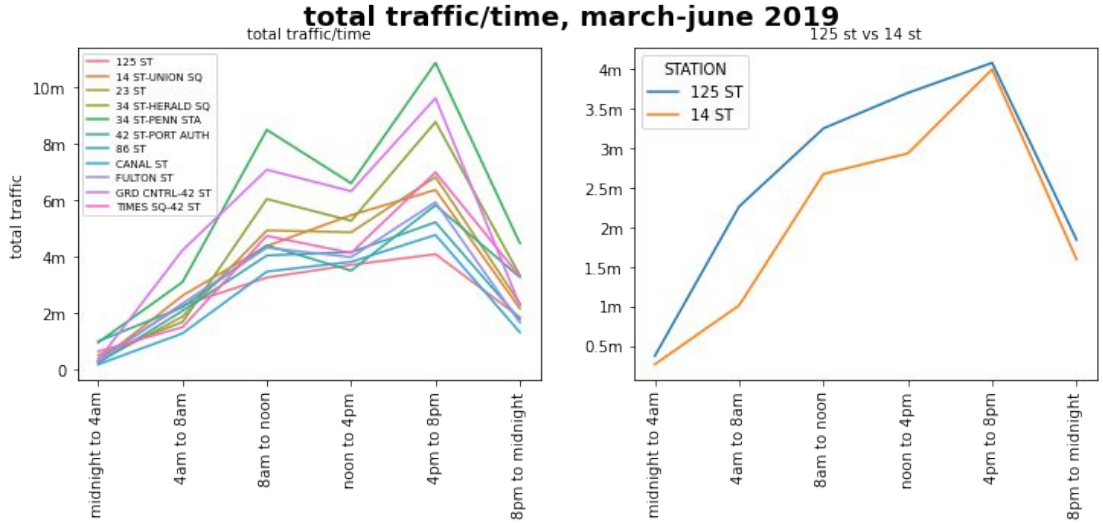


Traffic/Time Results

- Turning timestamp into a hour integer
- Grouping total traffic by hour integer

	STATION	TIME	HOUR	TIME_BUCKET
131193	FRANKLIN ST	00:00:00	0	8pm to midnight
131194	FRANKLIN ST	04:00:00	4	midnight to 4am
131195	FRANKLIN ST	08:00:00	8	4am to 8am
131196	FRANKLIN ST	12:00:00	12	8am to noon

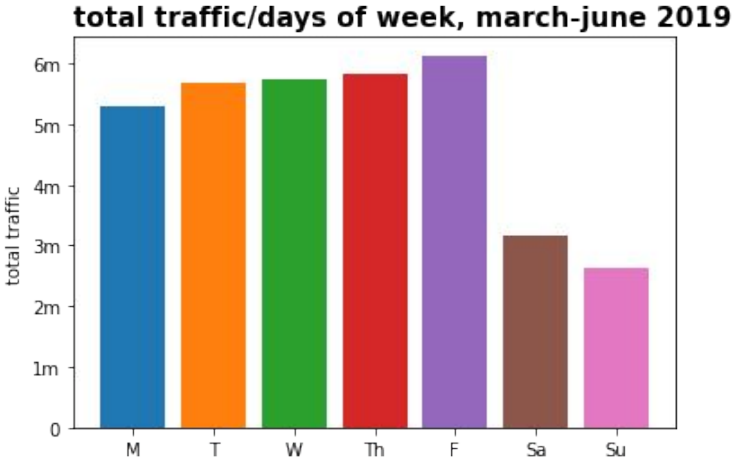


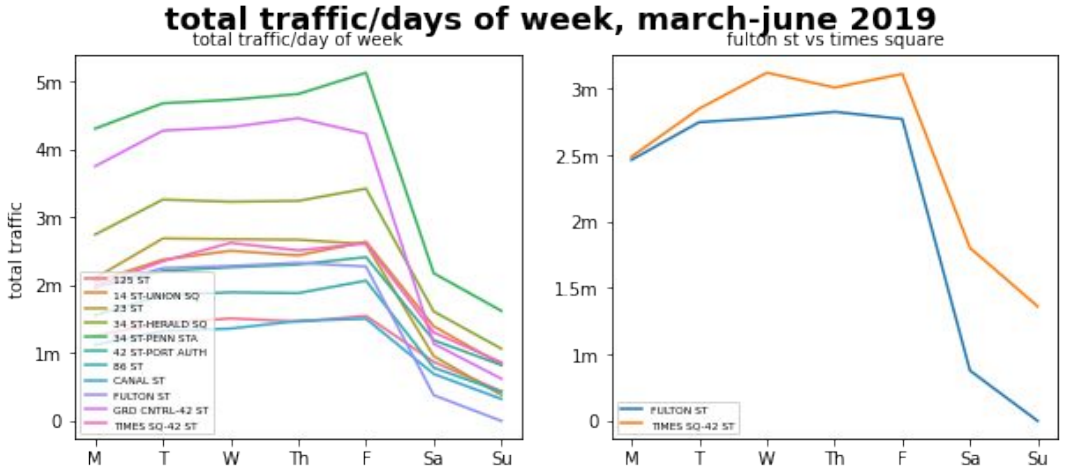


Traffic/Day of Week Results

- pd.to_datetime(date column).dt.day_name()
- pd.DatetimeIndex(date column).month

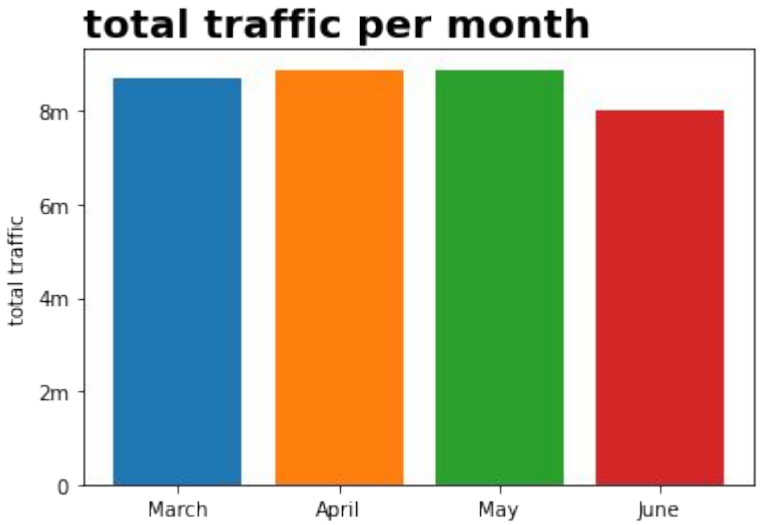
	STATION	DATE	TOTAL_TRAFFIC	DAY_OF_WEEK	WEEK_OF_YEAR	MONTH
1079	125 ST	03/01/2019	146309.0	Friday	9	3
1080	125 ST	03/02/2019	92920.0	Saturday	9	3
1081	125 ST	03/03/2019	79066.0	Sunday	9	3
1082	125 ST	03/04/2019	105194.0	Monday	10	3
1083	125 ST	03/05/2019	144338.0	Tuesday	10	3





Insights

- Weekdays during morning and late afternoon/night commuting hours
- Looking at 4 months over course of 2 seasons, not much change in total traffic -- seasons don't impact timing as much
- Some stations indicate that there are exceptions
    - Fulton St -- despite weekend drop off, maybe working professionals most interested in product.
    - Times Square -- more consistent traffic throughout the week but influx of tourists
    - 124 St & 14 St -- all day



Future Work

- Would be interesting to look at specific subway lines -- advertise inside the actual subway cars
- Compare data from 2019 with corresponding months in 2021 or 2022 to see how Covid-19 has impacted these numbers
- Research more about demographics of riders per station

Appendix 1

	MONTH	DAILY_ENTRIES			DAILY_EXITS		
		mean	std	median	mean	std	median
0	3	1021.581304	994.953216	770.0	814.430353	1326.124117	489.0
1	4	1065.386727	1021.618945	817.0	852.654585	1467.071802	525.0
2	5	1071.473182	1034.577359	816.0	861.496904	1545.358054	528.0
3	6	1059.216590	1028.202368	810.0	893.028645	1977.713077	526.0

```python
def hourly_entries(max_counter):
    num_entries = mta_df["ENTRIES"] - mta_df["PREV_ENTRIES"]
    num_entries = num_entries.abs()
    num_entries = num_entries.apply(lambda x: 0 if x>max_counter else x)
    return num_entries
```

```python
def time_convert(input): #convert timestamp
    return int(input.replace(':', "")[0:2])
```

```python
mta_df["HOUR"] = mta_df["TIME"].apply(time_convert)
```

```python
def time_buckets(num):
    if num >= 2 and num < 6:
        return 'midnight to 4am'
    elif num >= 6 and num < 10:
        return '4am to 8am'
    elif num >= 10 and num < 14:
        return '8am to noon'
    elif num >= 14 and num < 18:
        return 'noon to 4pm'
    elif num >= 18 and num < 22:
        return '4pm to 8pm'
    elif num >= 22 or num < 2:
        return '8pm to midnight'
    else:
        return "xyz"
```

```python
mta_df["TIME_BUCKET"] = mta_df["HOUR"].apply(time_buckets)
```

Appendix 2

ENTRIES	...	DATETIME	MONTH	PREV_DATE	PREV_ENTRIES	PREV_EXITS	DAILY_ENTRIES	DAILY_EXITS	HOUR	TIME_BUCKET	INTERVAL_TOTAL_TRAFFIC
6962094	...	2019-03-01 03:00:00	3	02/28/2019	6962053	2361664	41.0	8.0	3	midnight to 4am	49.0
6962111	...	2019-03-01 07:00:00	3	03/01/2019	6962094	2361672	17.0	42.0	7	4am to 8am	59.0
6962243	...	2019-03-01 11:00:00	3	03/01/2019	6962111	2361714	132.0	303.0	11	8am to noon	435.0
6962502	...	2019-03-01 15:00:00	3	03/01/2019	6962243	2362017	259.0	65.0	15	noon to 4pm	324.0
6963303	...	2019-03-01 19:00:00	3	03/01/2019	6962502	2362082	801.0	67.0	19	4pm to 8pm	868.0
...	...	...	...	...	...	...	...	...	...	...	...
5554	...	2019-06-28 05:00:00	6	06/28/2019	5554	379	0.0	0.0	5	midnight to 4am	0.0
5554	...	2019-06-28 09:00:00	6	06/28/2019	5554	379	0.0	0.0	9	4am to 8am	0.0
5554	...	2019-06-28 13:00:00	6	06/28/2019	5554	379	0.0	0.0	13	8am to noon	0.0
5554	...	2019-06-28 17:00:00	6	06/28/2019	5554	379	0.0	0.0	17	noon to 4pm	0.0
5554	...	2019-06-28 21:00:00	6	06/28/2019	5554	379	0.0	0.0	21	4pm to 8pm	0.0

	STATION	TIME_BUCKET	INTERVAL_TOTAL_TRAFFIC
0	1 AV	4am to 8am	402351.0
1	1 AV	4pm to 8pm	1229892.0
2	1 AV	8am to noon	896375.0
3	1 AV	8pm to midnight	576962.0
4	1 AV	midnight to 4am	76499.0
...	...	...	...
2269	ZEREGA AV	4pm to 8pm	114077.0
2270	ZEREGA AV	8am to noon	89565.0
2271	ZEREGA AV	8pm to midnight	39795.0
2272	ZEREGA AV	midnight to 4am	15038.0
2273	ZEREGA AV	noon to 4pm	119982.0

	TIME_BUCKET	INTERVAL_TOTAL_TRAFFIC
1	4pm to 8pm	75215949.0
2	8am to noon	55074402.0
5	noon to 4pm	51740650.0
3	8pm to midnight	27651015.0
0	4am to 8am	25063801.0
4	midnight to 4am	4938946.0

Appendix 3

	CA	UNIT	SCP	STATION	DATE	DAILY_ENTRIES	DAILY_EXITS	TOTAL_TRAFFIC	DAY_OF_WEEK	DAY_INDEXED	WEEK_OF_YEAR	MONTH
0	A002	R051	02-00-00	59 ST	03/01/2019	1525.0	532.0	2057.0	Friday	4	9	3
1	A002	R051	02-00-00	59 ST	03/02/2019	742.0	207.0	949.0	Saturday	5	9	3
2	A002	R051	02-00-00	59 ST	03/03/2019	568.0	179.0	747.0	Sunday	6	9	3
3	A002	R051	02-00-00	59 ST	03/04/2019	1235.0	395.0	1630.0	Monday	0	10	3
4	A002	R051	02-00-00	59 ST	03/05/2019	1479.0	475.0	1954.0	Tuesday	1	10	3
...	...	...	...	...	...	...	...	...	...	...	...	...
580722	TRAM2	R469	00-05-01	RIT-ROOSEVELT	06/24/2019	0.0	0.0	0.0	Monday	0	26	6
580723	TRAM2	R469	00-05-01	RIT-ROOSEVELT	06/25/2019	0.0	0.0	0.0	Tuesday	1	26	6
580724	TRAM2	R469	00-05-01	RIT-ROOSEVELT	06/26/2019	0.0	0.0	0.0	Wednesday	2	26	6
580725	TRAM2	R469	00-05-01	RIT-ROOSEVELT	06/27/2019	0.0	0.0	0.0	Thursday	3	26	6
580726	TRAM2	R469	00-05-01	RIT-ROOSEVELT	06/28/2019	0.0	0.0	0.0	Friday	4	26	6

	STATION	DATE	TOTAL_TRAFFIC	DAY_OF_WEEK	WEEK_OF_YEAR	MONTH
1079	125 ST	03/01/2019	146309.0	Friday	9	3
1080	125 ST	03/02/2019	92920.0	Saturday	9	3
1081	125 ST	03/03/2019	79066.0	Sunday	9	3
1082	125 ST	03/04/2019	105194.0	Monday	10	3
1083	125 ST	03/05/2019	144338.0	Tuesday	10	3
...	...	...	...	...	...	...
42256	TIMES SQ-42 ST	06/24/2019	192700.0	Monday	26	6
42257	TIMES SQ-42 ST	06/25/2019	202468.0	Tuesday	26	6
42258	TIMES SQ-42 ST	06/26/2019	216549.0	Wednesday	26	6
42259	TIMES SQ-42 ST	06/27/2019	217181.0	Thursday	26	6
42260	TIMES SQ-42 ST	06/28/2019	208061.0	Friday	26	6

2	1 AV	Saturday	331767.0
3	1 AV	Sunday	250350.0
4	1 AV	Thursday	702865.0
...	...	...	...
2648	ZEREGA AV	Saturday	45561.0
2649	ZEREGA AV	Sunday	34457.0
2650	ZEREGA AV	Thursday	90860.0
2651	ZEREGA AV	Tuesday	89988.0
2652	ZEREGA AV	Wednesday	92842.0