NYC MTA Subway Traffic Flow Analysis

Presented by: Satenik Safaryan
Young Suh
Sara Zong

Jan 8th, 2021

Goals of our analysis

- **WomenTechWomenYes(WTWY) has a street team that could be deployed to MTA subway stations to collect emails for fundraising and event promotion**
  - Determine WHEN and WHERE to send the street team to collect the most emails

Goals of our analysis

- WomenTechWomenYes(WTWY) has a street team that could be deployed to MTA subway stations to collect emails for fundraising and event promotion

➢ **Determine WHEN and WHERE to send the street team to collect the most emails**

EDA is NOT ONLY About Data Visualization

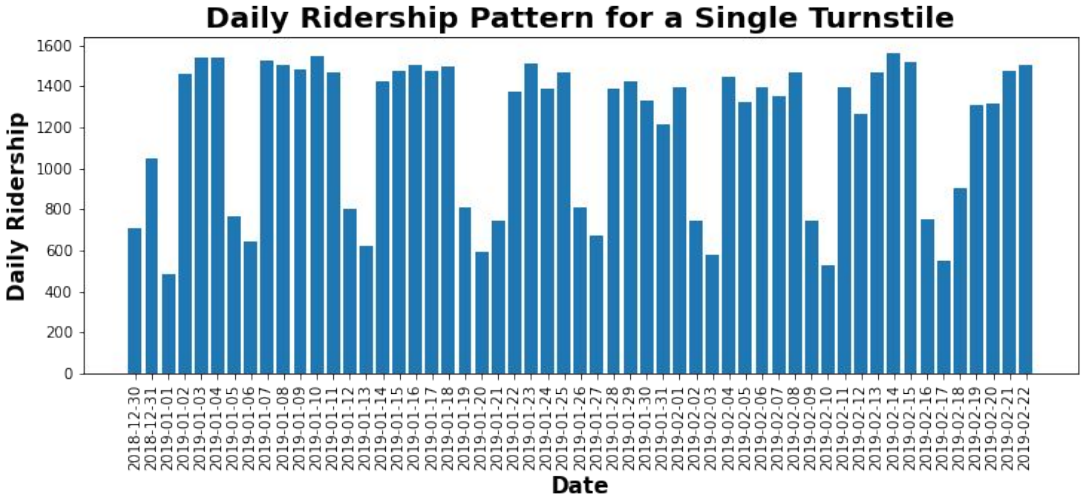
- MTA Turnstile Data
  - Jan, Feb of year 2019
- Time series regression model
  - Predict the number of people using certain station at specific time

Table View

	C/A	UNIT	SCP	STATION	LINENAME	DIVISION	DATE	TIME	DESC	ENTRIES	EXITS	DATE_TIME	WEEKDAYS
0	A002	R051	02-00-00	59 ST	NQR456W	BMT	2018-12-29	03:00:00	REGULAR	6889287	2335920	2018-12-29 03:00:00	Saturday
1	A002	R051	02-00-00	59 ST	NQR456W	BMT	2018-12-29	07:00:00	REGULAR	6889299	2335936	2018-12-29 07:00:00	Saturday
2	A002	R051	02-00-00	59 ST	NQR456W	BMT	2018-12-29	11:00:00	REGULAR	6889364	2336038	2018-12-29 11:00:00	Saturday
3	A002	R051	02-00-00	59 ST	NQR456W	BMT	2018-12-29	15:00:00	REGULAR	6889605	2336101	2018-12-29 15:00:00	Saturday
4	A002	R051	02-00-00	59 ST	NQR456W	BMT	2018-12-29	19:00:00	REGULAR	6889966	2336173	2018-12-29 19:00:00	Saturday

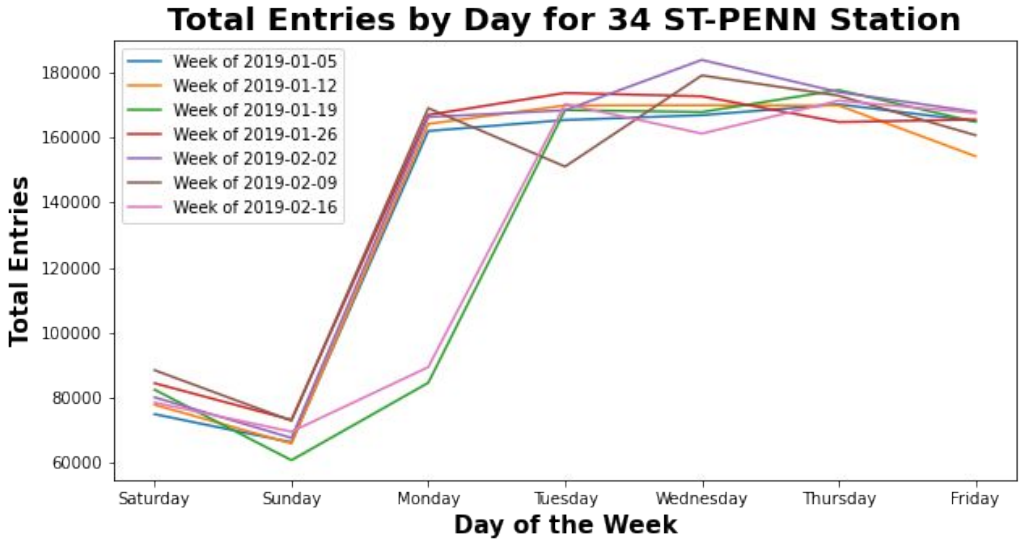
Showing 1 to 9 of 150 entries

Distribution of subway riders over dates

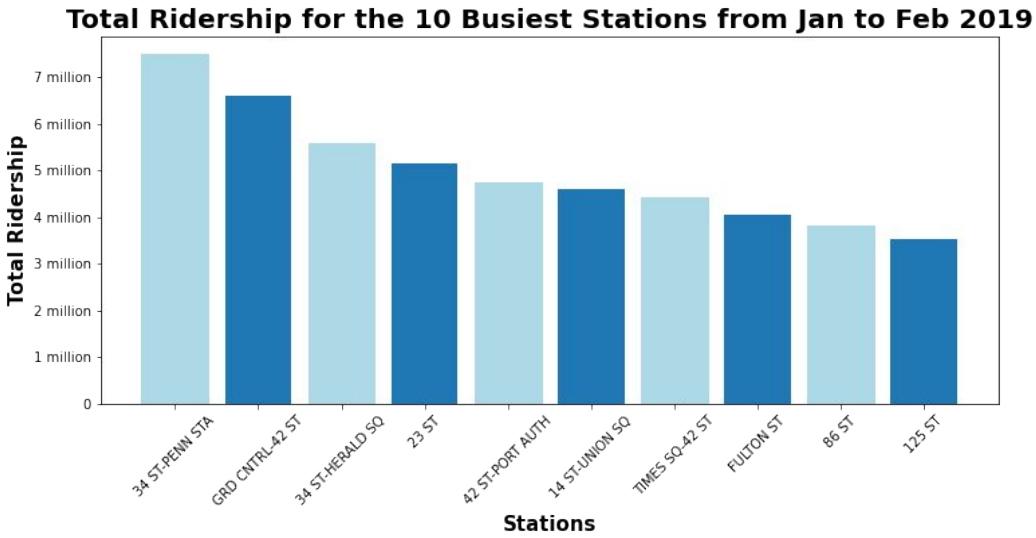


- Data from a single turnstile over time shows a weekend vs. weekday trend for passenger traffic

Distribution of subway riders over Day Of the Week (DOW)



The 10 Busiest Subway Stations



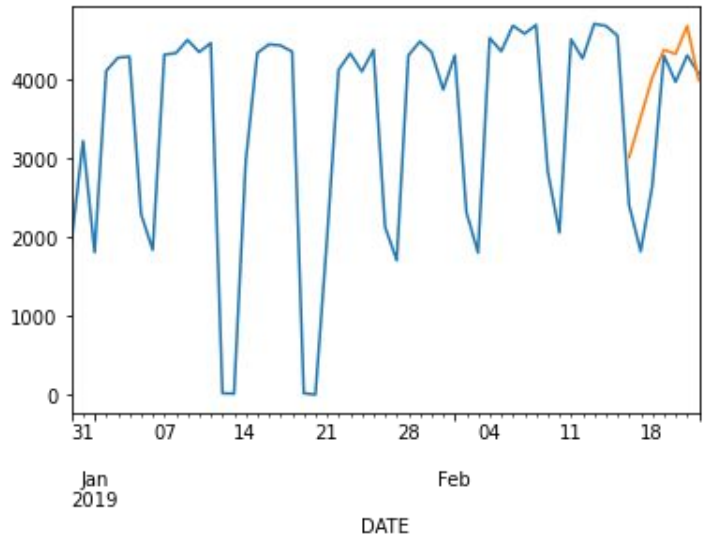
Top 10 stations:

1. 34 ST-PENN STA
2. GRD CNTRL-42 ST
3. 34 ST-HERALD SQ
4. 23 ST
5. 42 ST-PORT AUTH
6. 14 ST-UNION SQ
7. TIMES SQ-42 ST
8. FULTON ST
9. 86 ST
10. 125 ST


WE HAVE TO GO DEEPER

ARIMA

AutoRegressive Integrated Moving Average is a time series forecasting model.



This is an example for one of the stations: R641 R210 BEVERLY RD

Mock data!

Suggestions

According to the prediction we suggest to place the staff members near the following stations:

	DATE	STATION	PREDICTED_ENTRIES
0	2019-02-23	PTH22 R540 PATH NEW WTC	36442.0
0	2019-02-23	R238 R046 GRD CNTRL-42 ST	35505.0
0	2019-02-23	R533 R055 FLUSHING-MAIN	34240.0
0	2019-02-23	R138 R293 34 ST-PENN STA	33647.0
0	2019-02-23	N324 R018 JKSN HT-ROOSVLT	31682.0
0	2019-02-23	R221 R170 14 ST-UNION SQ	30198.0
0	2019-02-23	N051 R084 59 ST COLUMBUS	30135.0
0	2019-02-23	N506 R022 34 ST-HERALD SQ	28975.0
0	2019-02-23	N063A R011 42 ST-PORT AUTH	27279.0
0	2019-02-23	A025 R023 34 ST-HERALD SQ	26469.0

Mock data!

Future steps

Use Station Geolocation data as a future step.

- Ensure wide coverage of different neighborhoods to have a more diverse pool of people
- Exclude the stations with the most tourists and focus on the local population

Mock data!

Thank you for your attention!