NYC Subway Analysis

Chicago Metis Cohort 1
Masa Kato, Hasan Haq, Tiffany Li

Objectives

- Raise awareness for WomenTechWomenYes (WTWY), an organization focuses on building an inclusive community for women in the technology industry
- Maximize attendance at the organization's annual gala in the beginning of summer
- Optimize attendees with people willing to assist WTWY's fundraising efforts

Questions

1. Is it enough to allocate teams to the busiest stations?
    - Focus on stations with people most likely to contribute to the cause
2. How to effectively reach this target demographic?
    - Allocate teams at stations where tech-passionate individuals are likely to be
3. What is the best day to station the teams?
    - Station teams during the busiest day of the week

Approach

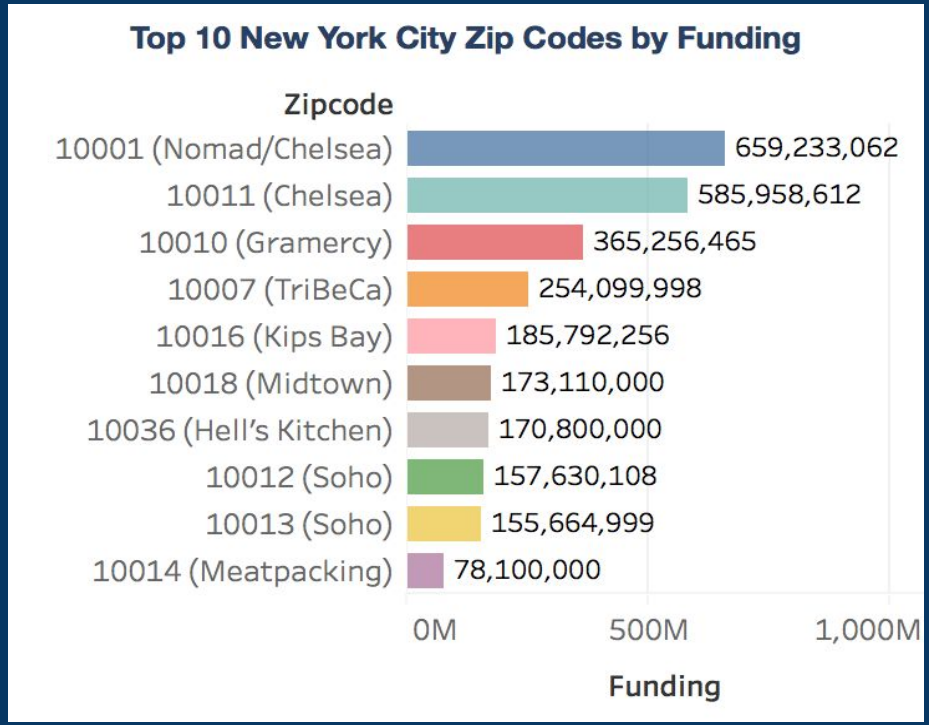
Data Collection	• Obtained from the Metropolitan Transportation Authority in NYC
Data Wrangling	• Used Python, Pandas, Numpy
Analysis	• Studied 2016 turnstile data as a proxy for 2017 • Found the optimal day to market the event
Visualization	• Utilized tools like Seaborn, Tableau • Interpreted and communicated the findings

The Data

NYC Zip Codes by Funding

Top 10 zip codes in New York City ranked by venture capital raised by tech companies between January 1st, 2016 & July 31, 2016.

Source: http://www.builtinnyc.com/2016/08/09/nyc-fundings-zipcode-2016



The Data

MTA Turnstile Data

MTA published data on entry & exit counts for each turnstile in New York City given in 4-hour intervals from 2010 onwards.

	C/A	UNIT	SCP	STATION	LINENAME	DATE	TIME	ENTRIES	EXITS	DATE_TIME	DAY_OF_WEEK	ZIPCODE
37	A002	R051	02-00-00	59 ST	NQR456	2016-04-01	00:00:00	5599639	1896005	2016-04-01 00:00:00	Friday	11209
38	A002	R051	02-00-00	59 ST	NQR456	2016-04-01	04:00:00	5599650	1896015	2016-04-01 04:00:00	Friday	11209
39	A002	R051	02-00-00	59 ST	NQR456	2016-04-01	08:00:00	5599690	1896110	2016-04-01 08:00:00	Friday	11209
40	A002	R051	02-00-00	59 ST	NQR456	2016-04-01	12:00:00	5599694	1896111	2016-04-01 12:00:00	Friday	11209
41	A002	R051	02-00-00	59 ST	NQR456	2016-04-01	16:00:00	5599946	1896170	2016-04-01 16:00:00	Friday	11209

Source: http://web.mta.info/developers/turnstile.html

Data Wrangling

Put Pandas to work

- Performed operations to process the data into a format preferable to the analysis and grouped the data by various categories

  (i.e. Counts at stations per day)

Design Decisions / Assumptions

1. Used MTA Turnstile Data from April 1st, 2016 to June 1st, 2016
    - Since the annual Gala is at the beginning of summer, we decided to use the data for the three months preceding summer.
    - Used data from 2016 because it would provide the most recent trends for station & turnstile traffic
2. Focused on stations with the most traffic that are in the top 10 zip codes for venture capital funding
    - Matched stations with zip codes by incorporating latitude & longitude data and reverse geocoding them into zip codes

Design Decisions / Assumptions

3. Some turnstiles were replaced & as a result the running counts of the entry/exits would change
   - Replaced the outlier value with the median for the month

	DATE	ENTRIES
0	2016-04-01	140543
1	2016-04-02	49230
2	2016-04-03	35043
3	2016-04-04	156321
4	2016-04-05	155212
5	2016-04-06	43919788
6	2016-04-07	150811

Example: Grd Cntrl - 42 St Station

	C/A	UNIT	SCP	STATION	LINENAME	DATE	TIME	ENTRIES	EXITS	DATE_TIME	DAY_OF_WEEK	ZIPCODE
148882	R238	R046	00-00-08	GRD CNTRL-42 ST	4567S	2016-04-06	04:00:00	43773605	9870803	2016-04-06 04:00:00	Wednesday	10011
148883	R238	R046	00-00-08	GRD CNTRL-42 ST	4567S	2016-04-06	08:00:00	43773605	9871234	2016-04-06 08:00:00	Wednesday	10011
148884	R238	R046	00-00-08	GRD CNTRL-42 ST	4567S	2016-04-06	12:00:00	3070	622	2016-04-06 12:00:00	Wednesday	10011

Analysis

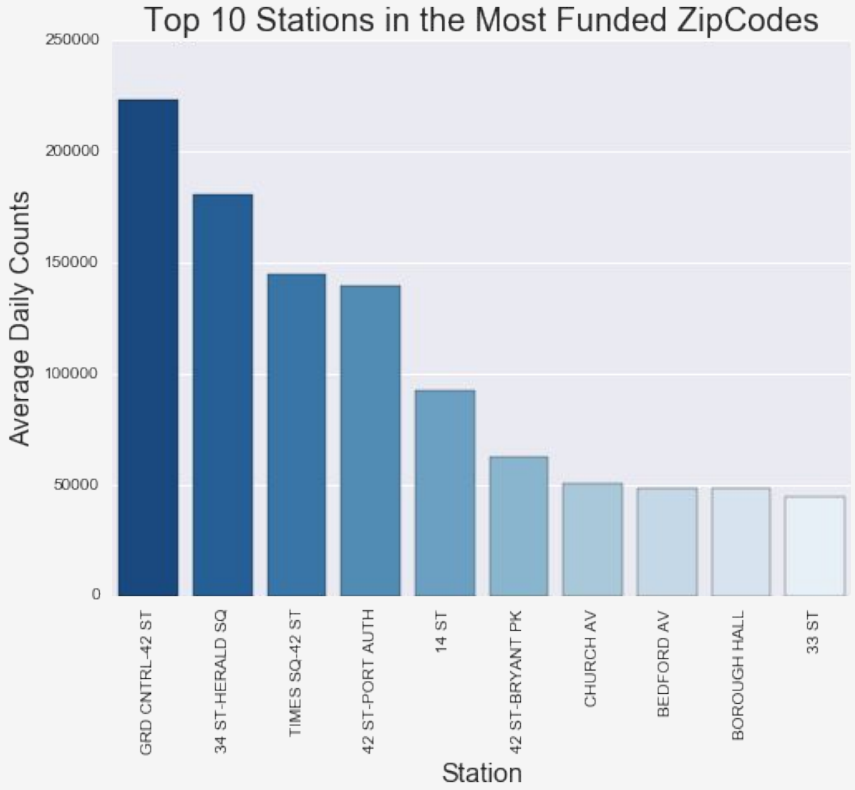

Stations in the Top 10 Most Funded Zip Codes for Technology Companies
Target
Month & Day with the Most Foot Traffic
Stations with Heavy Foot Traffic

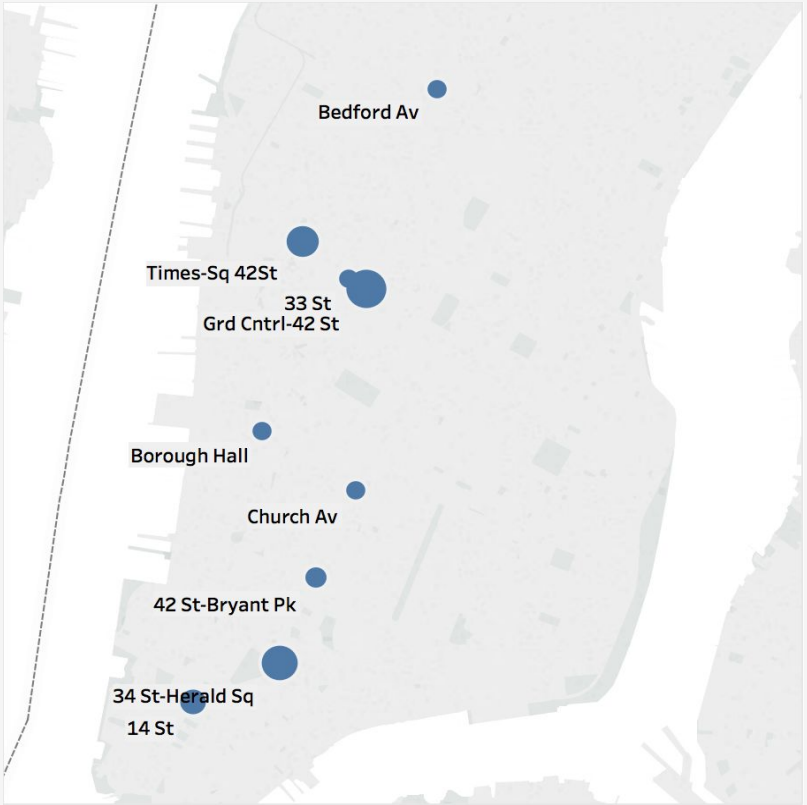
Analysis

Target stations ranked by foot traffic within target zip codes:

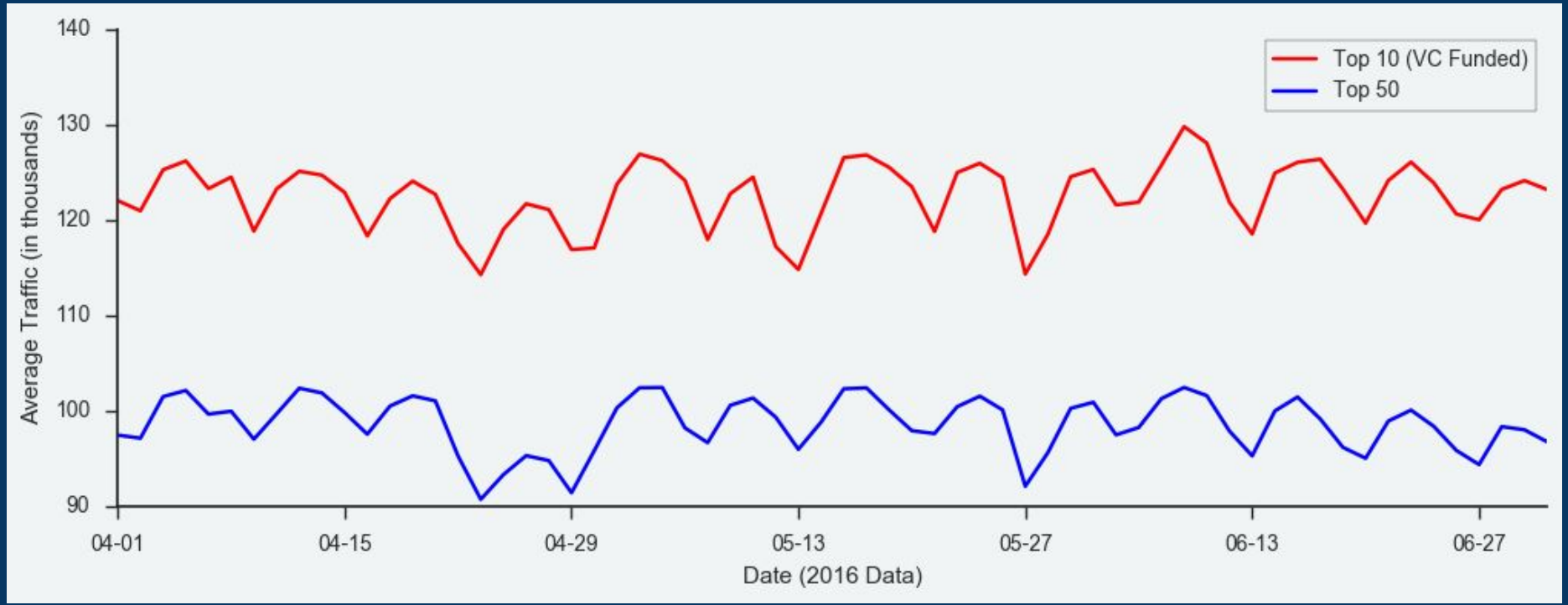
1. GRD CNTRL-42 ST (2)
2. 34 ST-HERALD SQ (3)
3. TIMES SQ-42 ST (7)
4. 42 ST-PORT AUTH (8)
5. 14 ST (17)
6. 42 ST-BRYANT PK (27)
7. CHURCH AV (34)
8. BEDFORD AV (35)
9. BOROUGH HALL (36)
10. 33 ST (40)

Average Daily Traffic for Target Stations

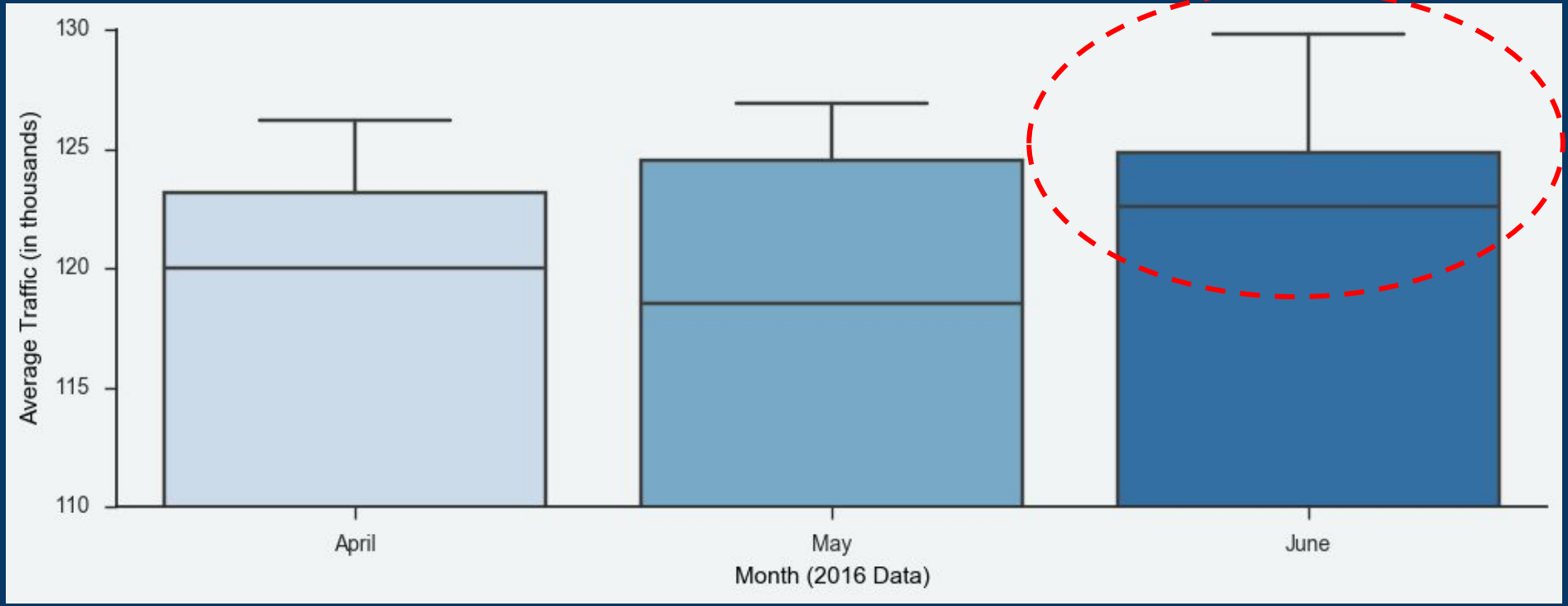




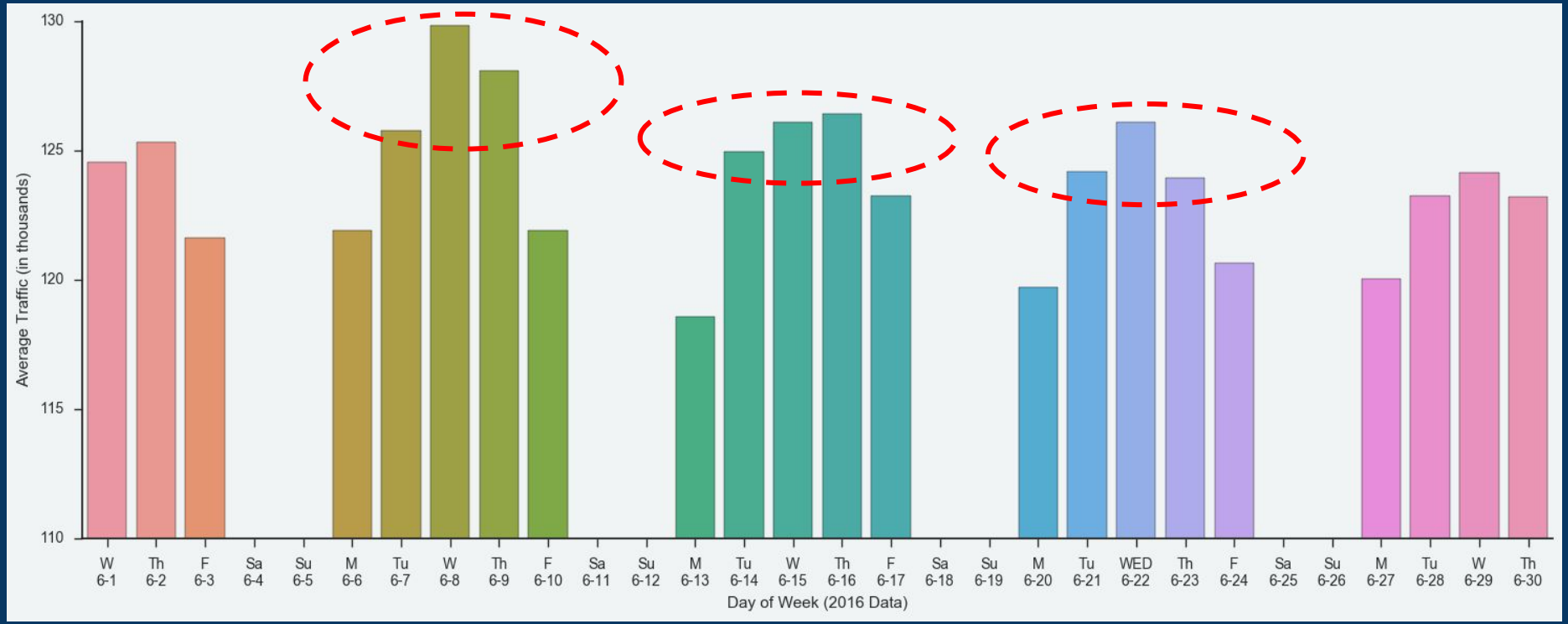
Our Recommended Stations Have High Traffic



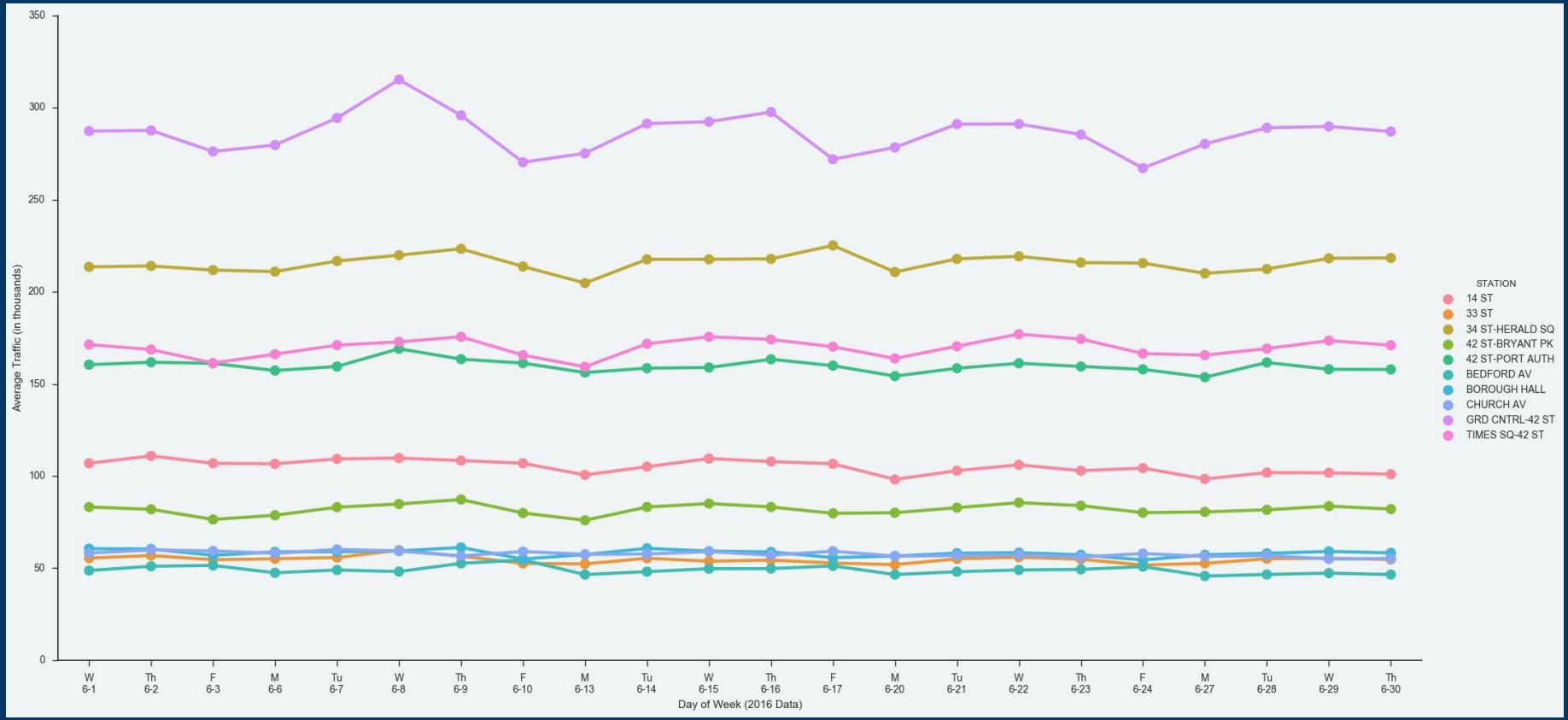
We Recommend Marketing in June



We Recommend Marketing on Wednesday



(Almost) All Stations Follow the Wednesday Trend



Final Recommendations

- Month: June
- Day of the week: Wednesday

1. GRD CNTRL-42 ST (2)
2. 34 ST-HERALD SQ (3)
3. TIMES SQ-42 ST (7)
4. 42 ST-PORT AUTH (8)
5. 14 ST (17)
6. 42 ST-BRYANT PK (27)
7. CHURCH AV (34)
8. BEDFORD AV (35)
9. BOROUGH HALL (36)
10. 33 ST (40)

Next Steps

- Include an analysis to determine what is the best time of the day
- Incorporate other demographic factors, such as income
- Rank the target stations using weighted scores
- Support our design decisions with statistical testing

We Need This.