WTWY: Street Team Placement

Group 2:
Adam Ariel
Summer Yuan
PK
Matias Beeck

1

Project Summary

Intro

- With the **annual gala** happening in June 2018, we set out to analyze the optimal stations for you to send your street teams to in the month leading up to the event.

Dataset

- **MTA turnstile data** from the month of May 2017

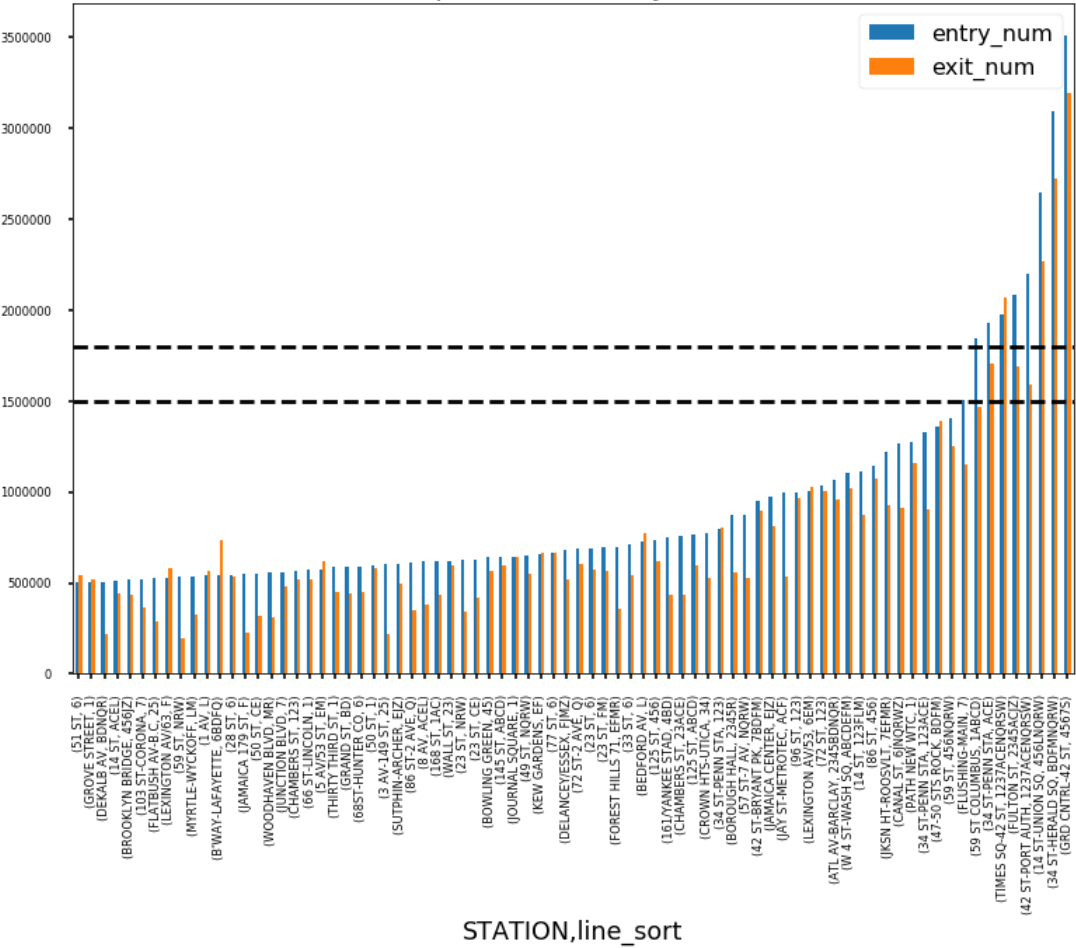- Income data from **City-Data.com**

Overview

- **Total station traffic**

- **Weekday vs weekend activity**

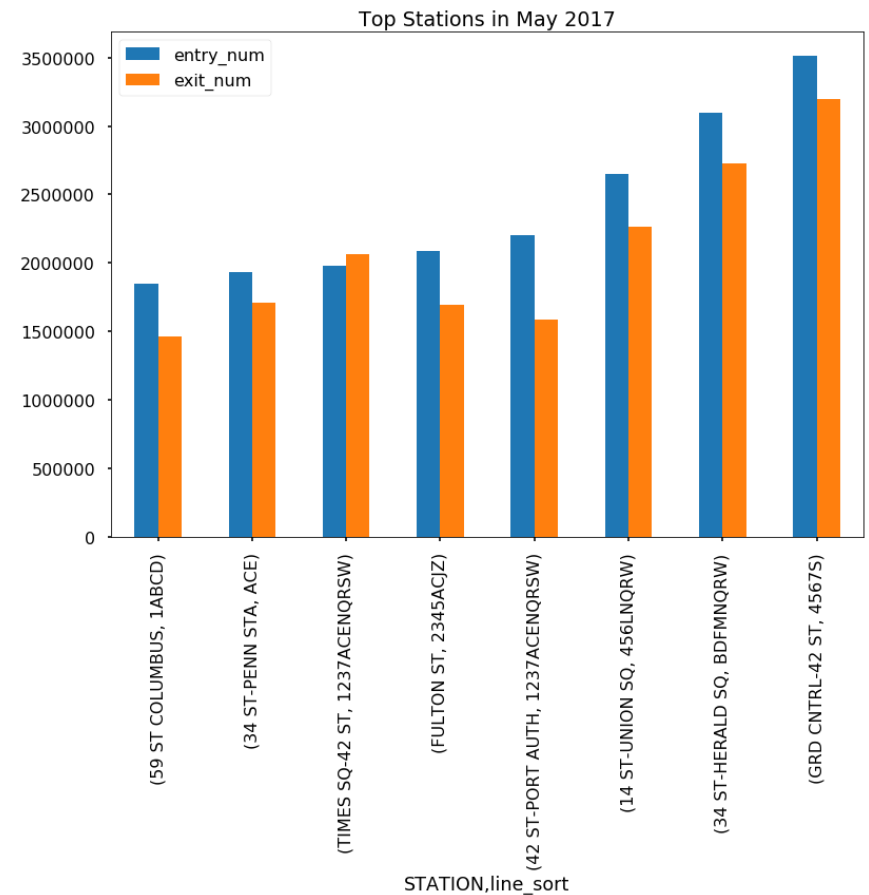- **Morning vs afternoon**

- **Income & Traffic**

2

Initial Analysis

Top Stations in May 2017

Top Stations

- In May 2017, these stations had **significantly higher traffic** than the others

- Five of these stations were located in the **midtown area**

Top Stations in May 2017

- **GRD CNTRL-42 ST**

- **34 ST-HERALD SQ**

- **14 ST-UNION SQ**

- **42 ST-PORT AUTH**

- **FULTON ST**

- **TIMES SQ-42 ST**

- **34 ST-PENN STA**

- **59 ST COLUMBUS**

Stations By Traffic

3

Time Analysis

Weekday vs Weekend

- Created subsets of data for weekday/ weekends

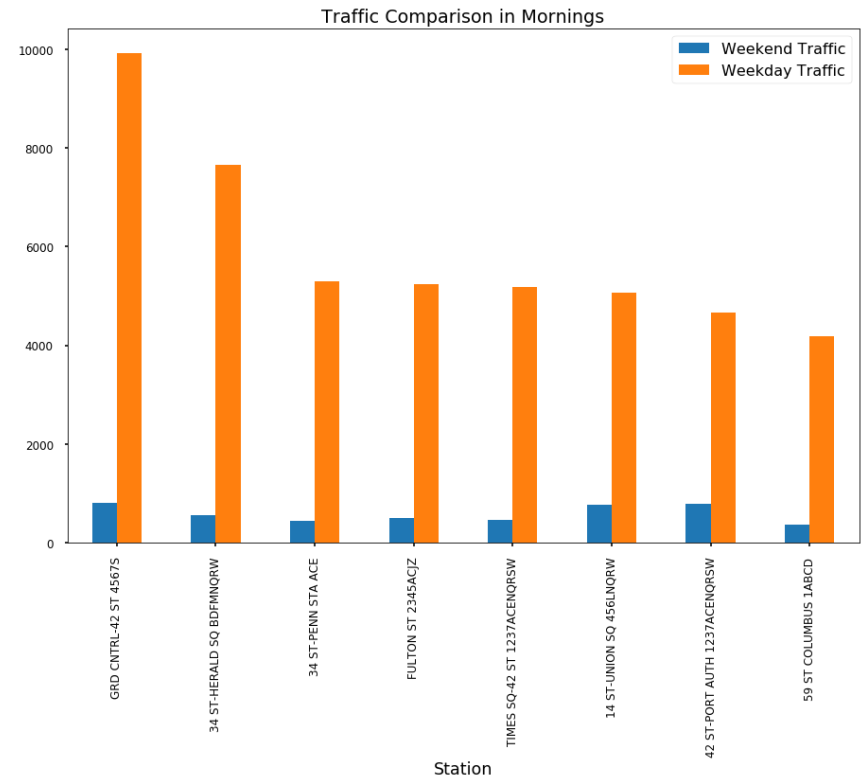| STATION	LINENAME	parsed_datetime	Weekend	station_id
59 ST	NQR456W	2017-04-29 04:00:00	True	0
59 ST	NQR456W	2017-04-29 08:00:00	True	0
59 ST	NQR456W	2017-04-29 12:00:00	True	0
59 ST	NQR456W	2017-04-29 16:00:00	True	0
59 ST	NQR456W	2017-04-29 20:00:00	True	0
59 ST	NQR456W	2017-04-30 00:00:00	True	0
59 ST	NQR456W	2017-04-30 04:00:00	True	0
59 ST	NQR456W	2017-04-30 08:00:00	True	0
59 ST	NQR456W	2017-04-30 12:00:00	True	0
59 ST	NQR456W	2017-04-30 16:00:00	True	0
59 ST	NQR456W	2017-04-30 20:00:00	True	0
59 ST	NQR456W	2017-05-01 00:00:00	False	0
59 ST	NQR456W	2017-05-01 04:00:00	False	0
59 ST	NQR456W	2017-05-01 08:00:00	False	0
59 ST	NQR456W	2017-05-01 12:00:00	False	0

Weekday vs Weekend

- Created subsets of data for morning and afternoon

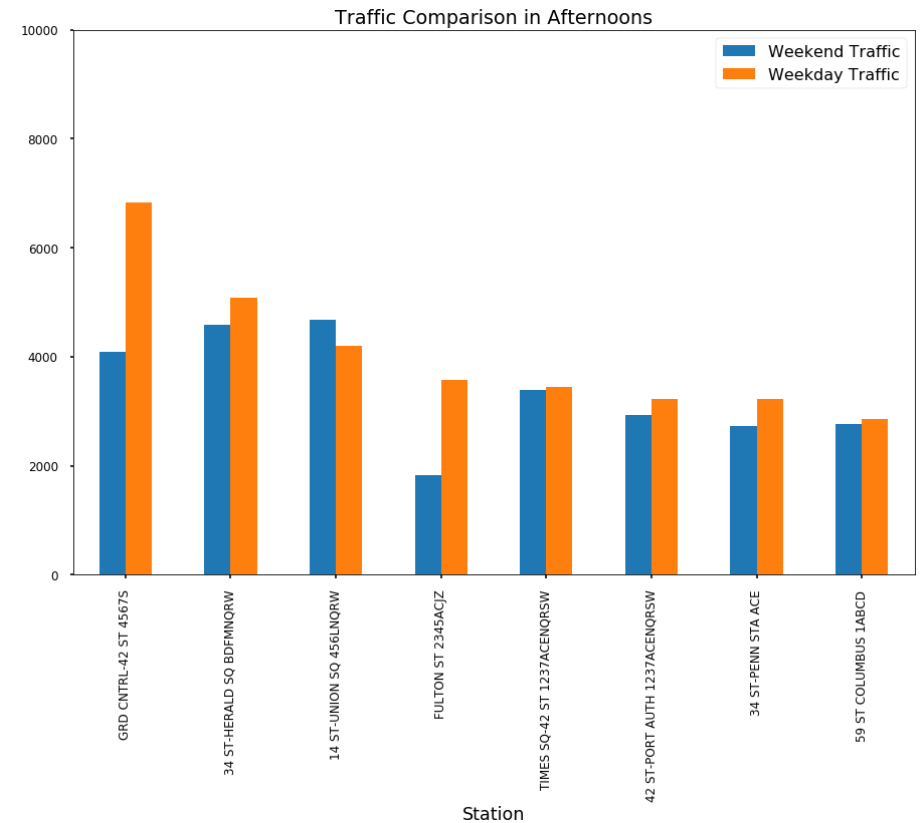|	STATION	LINENAME	ENTRIES	EXITS	parsed_datetime	station_id
147235	GRD CNTRL-42 ST	4567S	813019	2422100	2017-04-29 05:00:00	3603
147236	GRD CNTRL-42 ST	4567S	813032	2422285	2017-04-29 09:00:00	3603
147237	GRD CNTRL-42 ST	4567S	813054	2422792	2017-04-29 13:00:00	3603
147238	GRD CNTRL-42 ST	4567S	813179	2423350	2017-04-29 17:00:00	3603

Wkday v Wknd: Morning

- Weekday traffic volume was found to be much higher than Weekends in the mornings

- Street teams should not be sent out on weekend mornings

Traffic Comparison in Mornings

Wkday v Wknd: Afternoons

- For the afternoons, the difference between weekday and weekend was not as significant

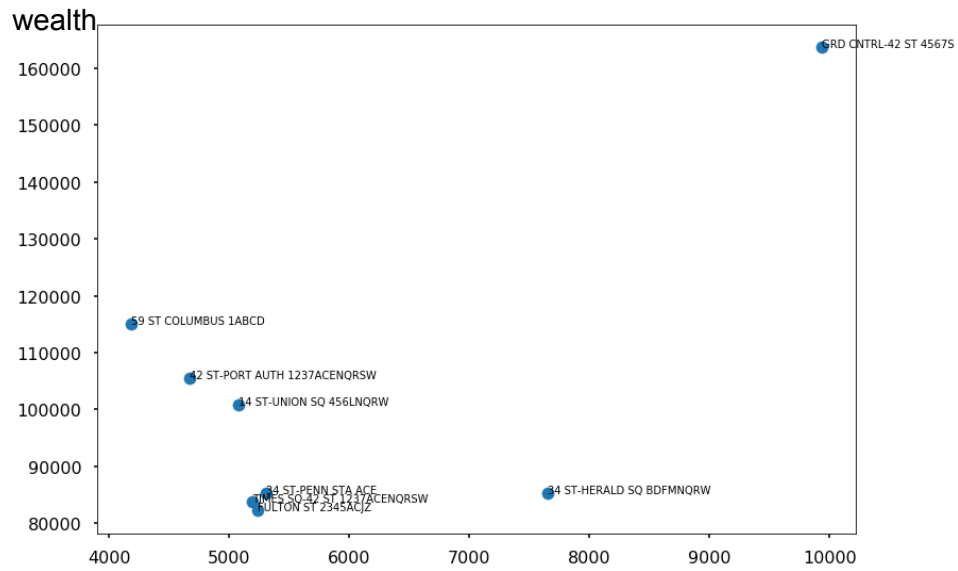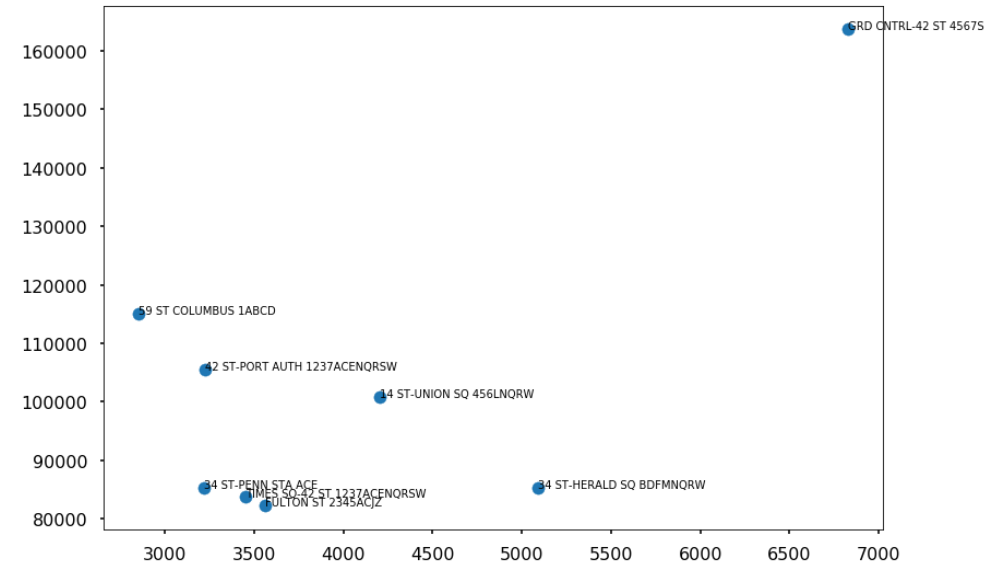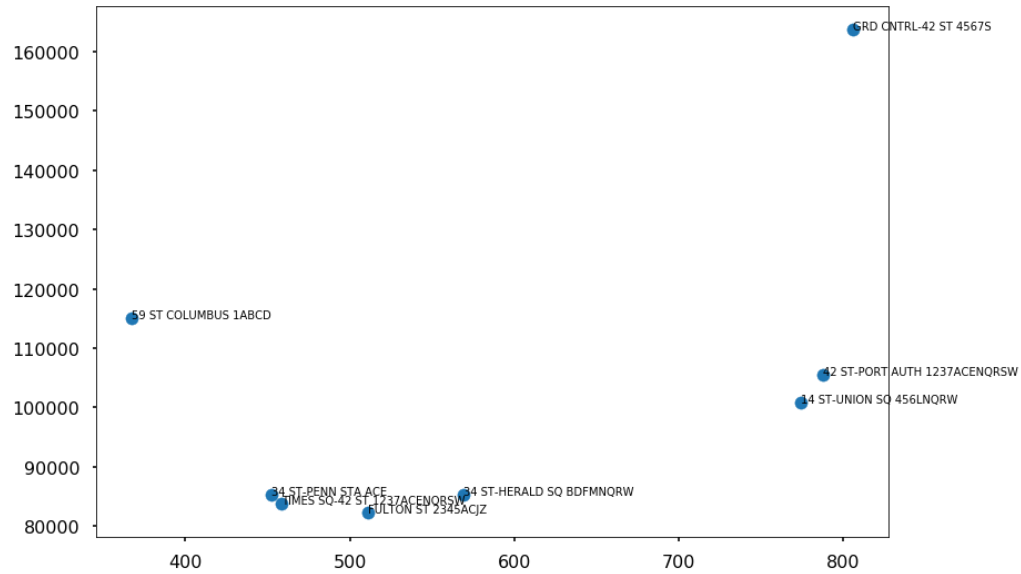- Afternoon traffic is still heavy on weekends

Traffic Comparison in Afternoons

4

Income Data
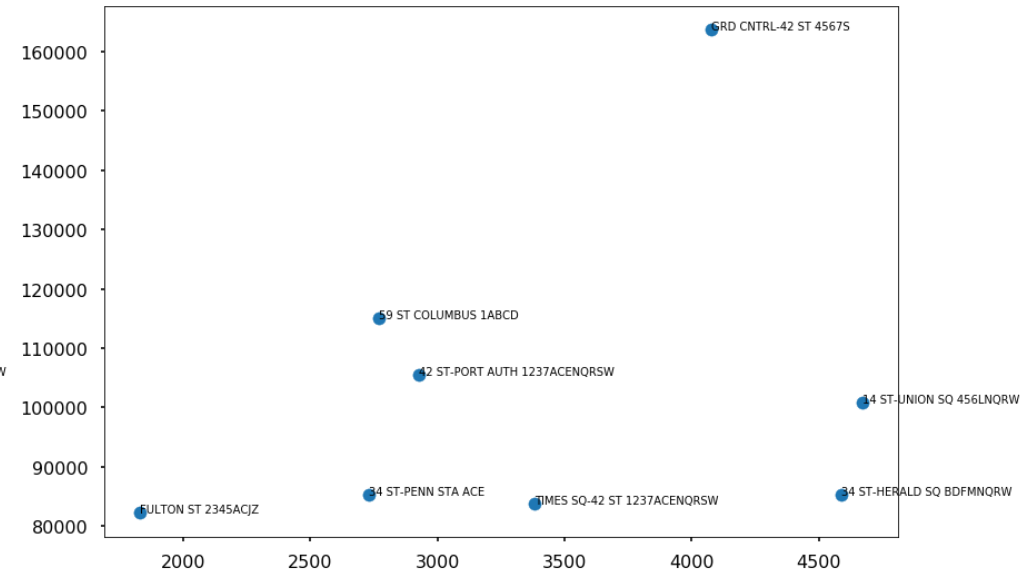
Weekday

wealth

Weekday Morning

Weekday Afternoon

Weekend

Weekend Morning

Weekend Afternoon

5

Proposal

Conclusions/Next Steps:
Max traffic

Weekday Mornings (6am – 1pm):	Weekday Afternoons (2pm – 10pm):
Grand Central Herald Square Penn Station	Grand Central Herald Square Union Square
Weekend Mornings (6am – 1pm): Rest :)	Weekend Afternoons (2pm – 10pm): Union Square Herald Square Grand Central

Conclusions/Next Steps:
Max Donation

- In all of our time scenarios, Grand Central is the most optimal one for income and traffic.

- The rest of popular stations varies in traffic and income, we recommend WTWY to determine based on its preference.

THANKS!

6

Appendix

Data Cleaning

- **Movie ratings are more difficult to predict than I initially had thought**

- **Logistic regression performed significantly better than linear regression**

- **Run a random forest regression on outcome variable**

- **Focus on one genre of movies from a more specified date range**

- **Attempt to find more features that might effect movie ratings more**
  - Analyze plot keywords using NLP
  - Analyze social data to see how it might affect movie ratings