

# 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.

## Datasets

- MTA turnstile data from the month of May 2017
- Income data from City-Data.com
- Charitable donation data from Wolters Kluwer Tax & Accounting US, 2015

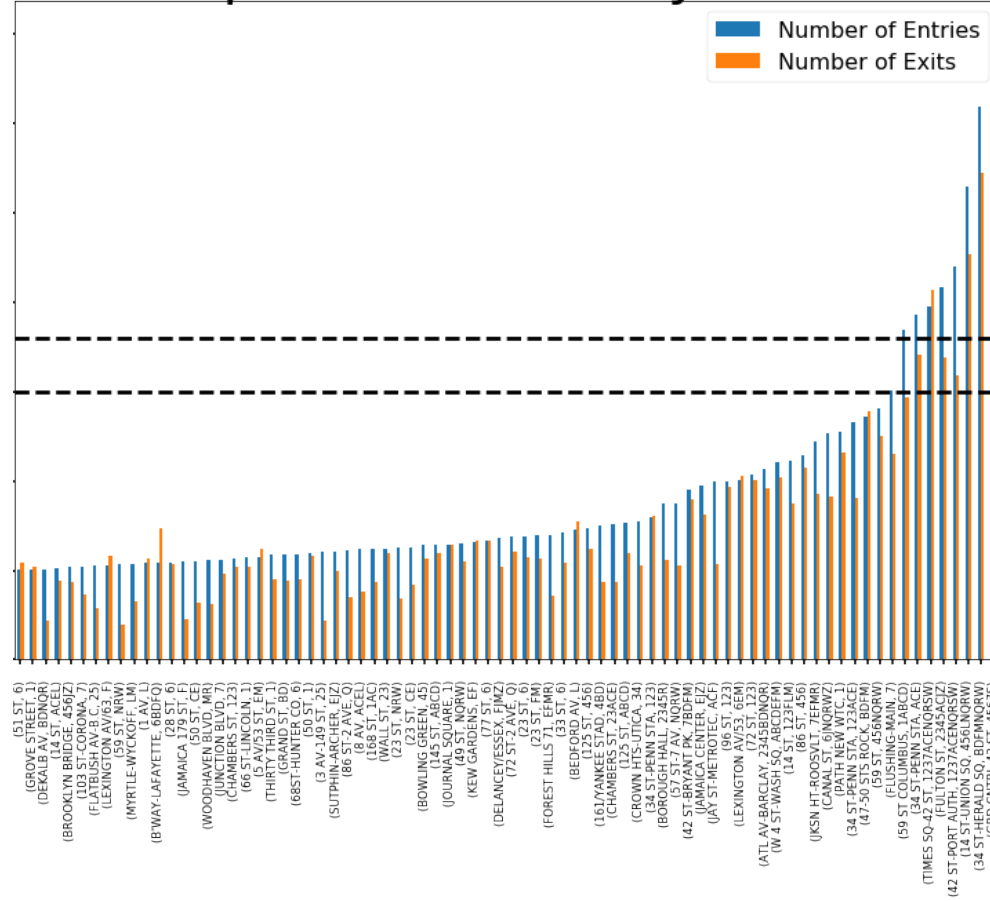
## Overview

- Total station traffic
- Weekday vs weekend activity
- Morning vs afternoon
- Income, Giving & Traffic

2

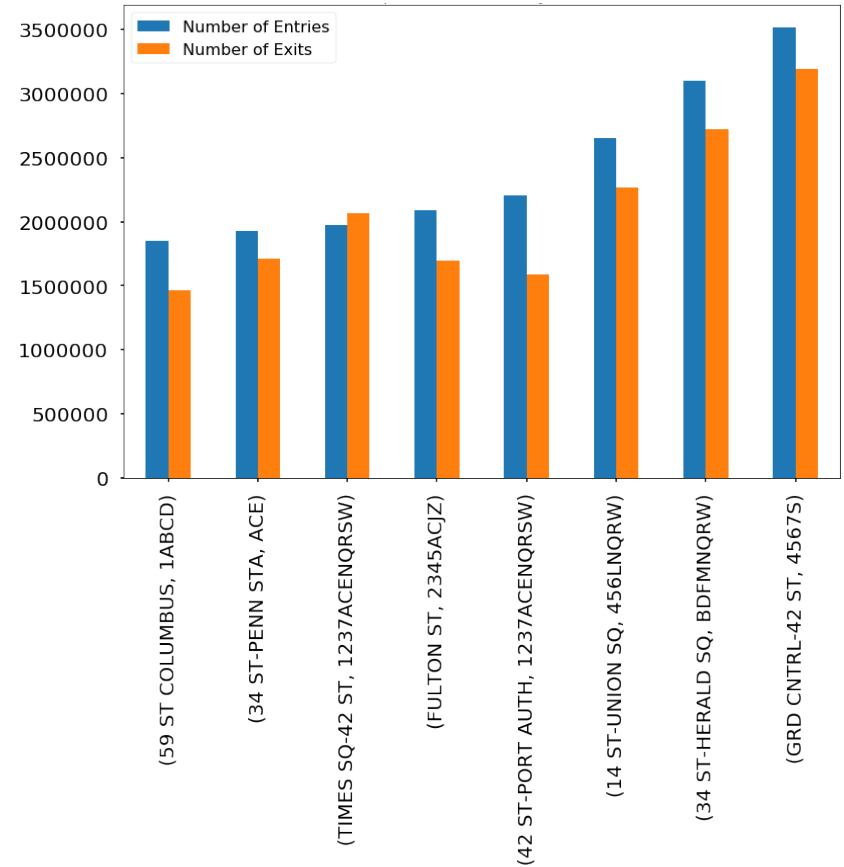
# **Initial Analysis**

# 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**







- GRD CNTRL-42 ST (10017)
- 34 ST-HERALD SQ (10001)
- 14 ST-UNION SQ (10003)
- 42 ST-PORT AUTH (10018)
- FULTON ST (10038)
- TIMES SQ-42 ST (10036)
- 34 ST-PENN STA (10001)
- 59 ST COLUMBUS (10023)

## Stations By Traffic

3

# **Time Period Analysis**

## Data Prep

- Created subsets of the top stations' data filtered by weekday and weekend
- Bucketed data by time of day
- Morning: 6am – 1pm
- Afternoon: 2pm – 10pm

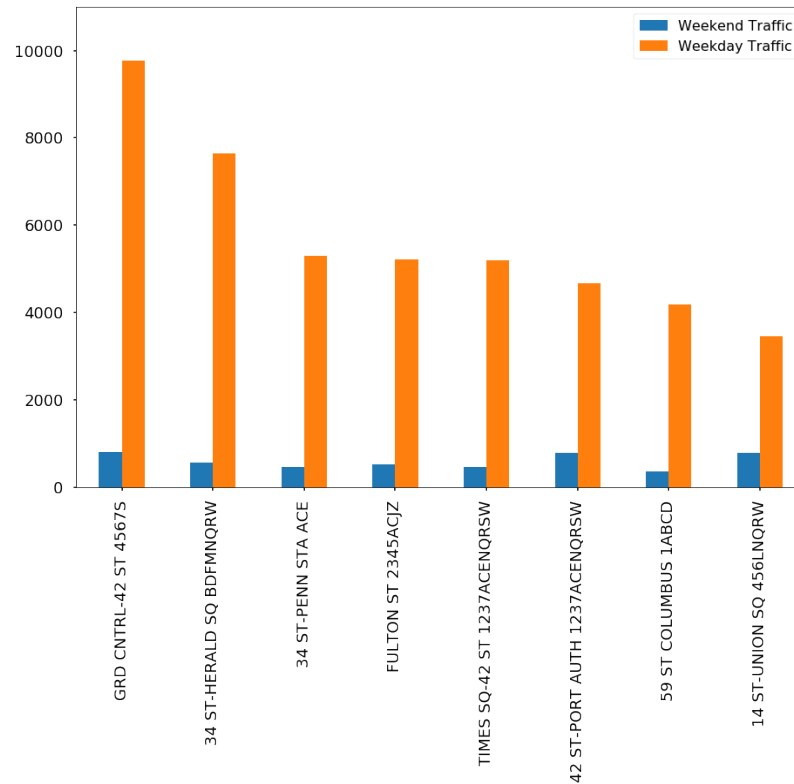
STATION	LINENAME	parsed_datetime	Weekend	station_id
59 ST	NQR456W	2017-04-29 04:00:00	True	0
59 ST	NQR456W	2017-04-29 08:00:00	True	0
59 ST	NQR456W	2017-04-29 12:00:00	True	0
59 ST	NQR456W	2017-04-29 16:00:00	True	0
59 ST	NQR456W	2017-04-29 20:00:00	True	0
59 ST	NQR456W	2017-04-30 00:00:00	True	0
59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

	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

- Traffic numbers represent average station traffic per hour
- For mornings, weekday traffic volume was found to be much higher than weekends, as expected
- Street teams should not be sent out on weekend mornings

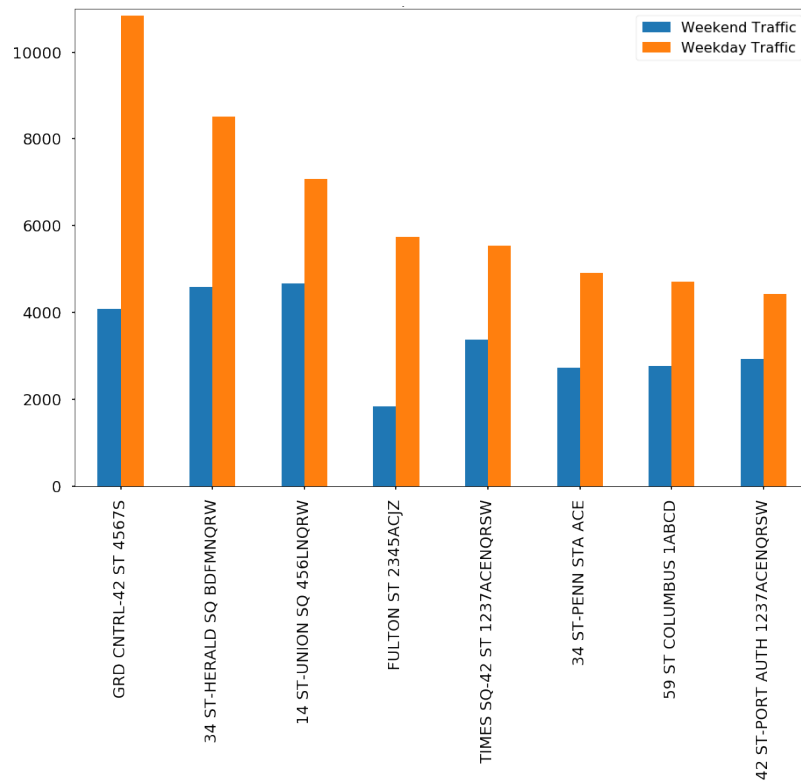
### Morning Traffic By Station



## Wkday v Wknd: Afternoons

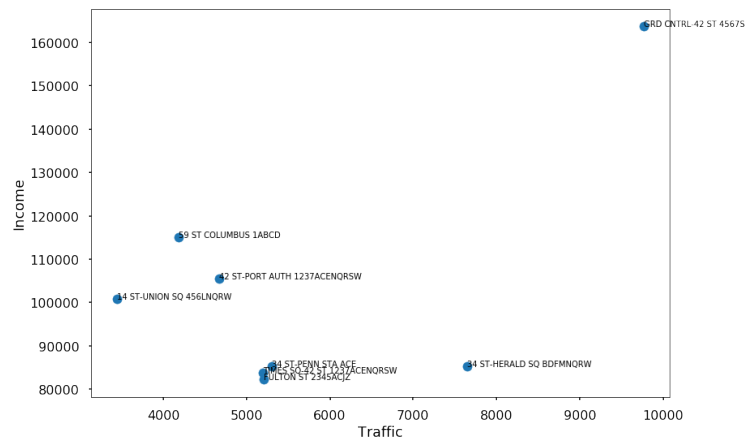
- For afternoons, the difference between weekday and weekend was not as significant
- Street teams should be sent out on weekend afternoons, but priority should be on weekdays

Afternoon Traffic By Station

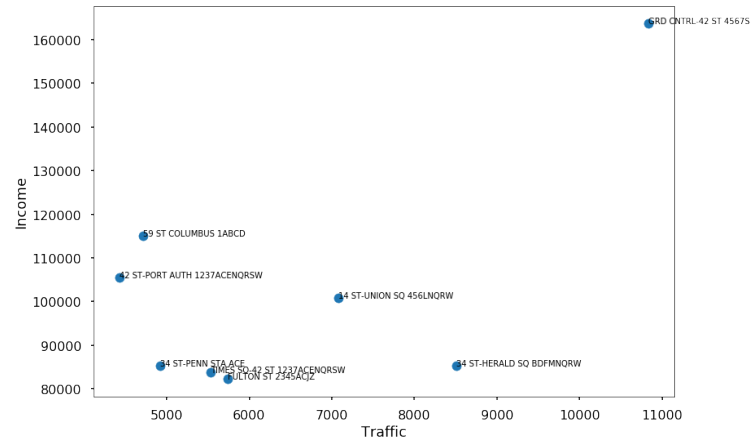


4

## **Income Data**

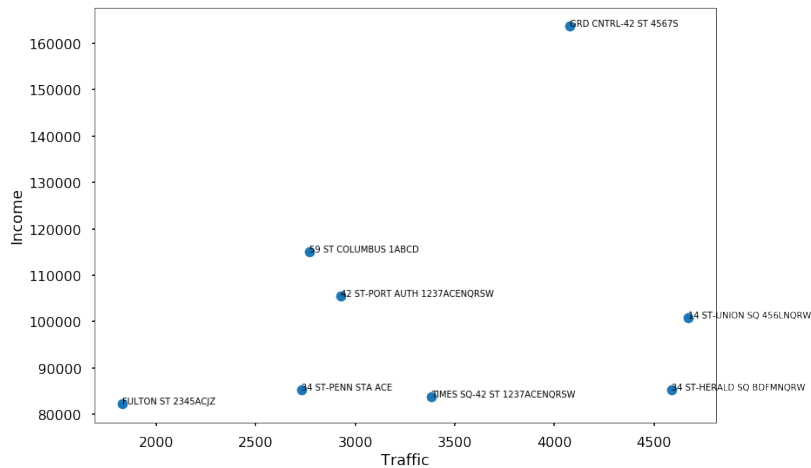


Weekday Morning



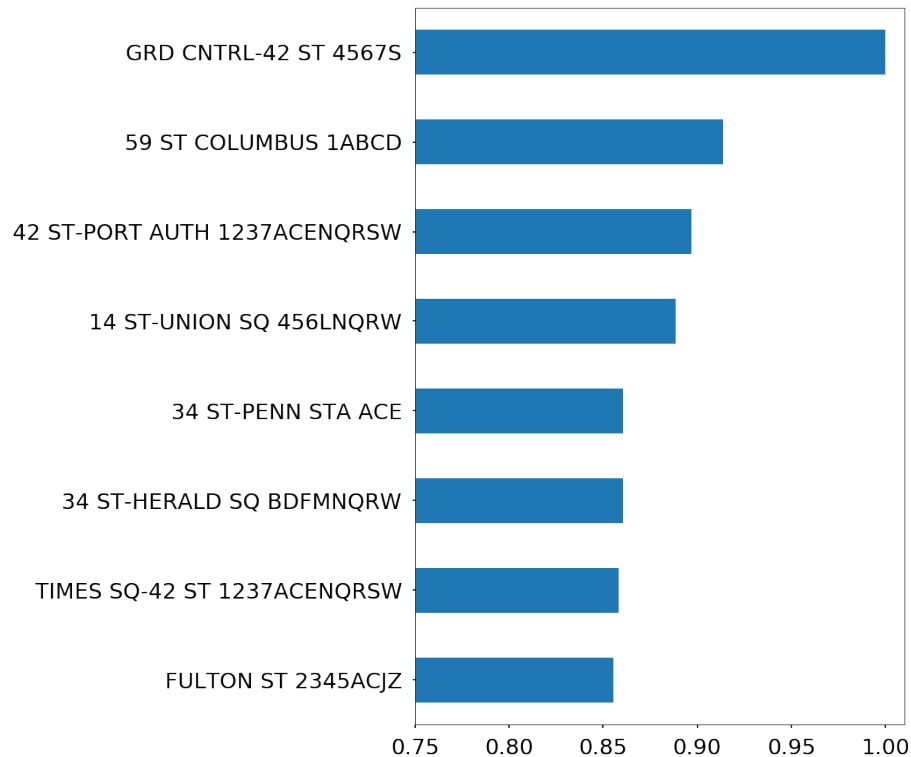
Weekday Afternoon

Weekend  
Afternoon



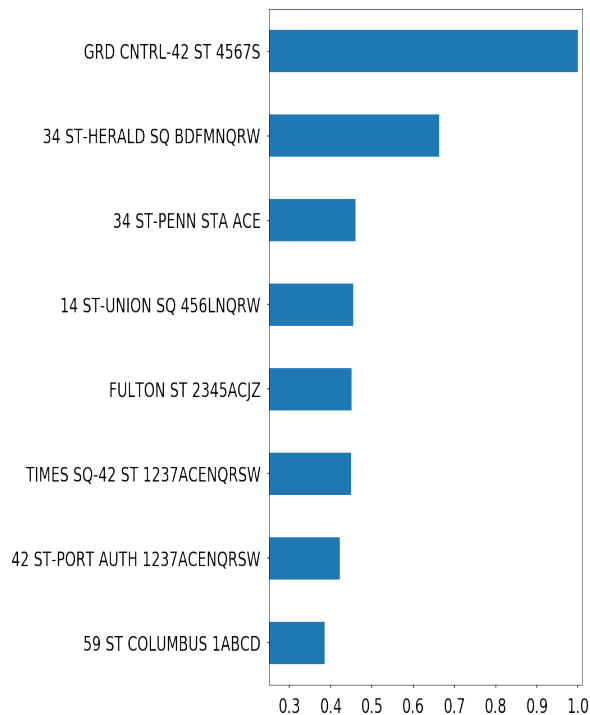
## Charitable Donations

- Our research showed that people making 50k were expected to give  $\frac{3}{4}$  as much as people making 200k
- Relative expected gift value for top stations

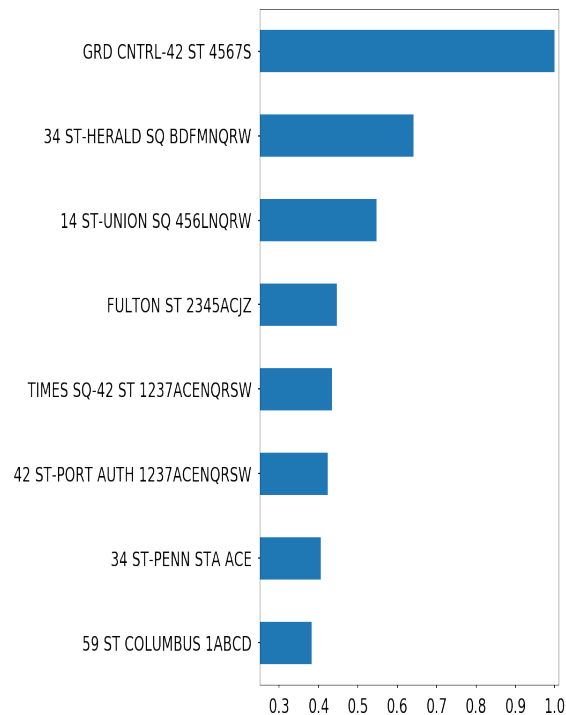




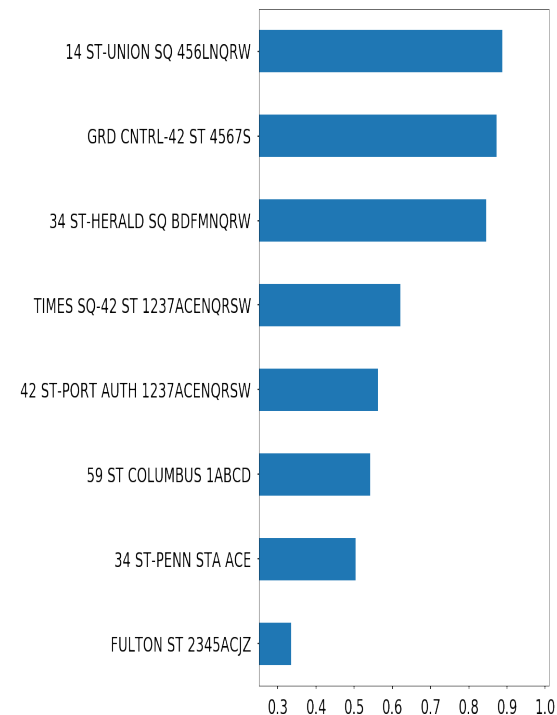
## Weekday Morning



## Weekday Afternoon



## Weekend Afternoon



Relative Expected Gift Value

## Recommendation

Weekday Mornings (6am – 1pm)

1. Grand Central
2. Herald Square
3. Penn Station

Weekday Afternoons (2pm – 10pm)

1. Grand Central
2. Herald Square
3. Union Square

Weekend Mornings (6am – 1pm)

Rest 😊

Weekend Afternoons (2pm – 10pm)

1. Union Square
2. Grand Central
3. Herald Square

**Questions?**

6

# Appendix

## Data Cleaning

- Parsed date and time into a datetime object 'parsed\_datetime'
- Assigned each station entrance a unique id by combining the 'C/A', 'UNIT', 'SCP' and 'STATION' variables
- Calculated entries and exits between each audit by taking the absolute value of the differences from row to row
- Created a boolean for if each entry was a weekday or weekend
- Sorted 'LINENAME' so that we didn't have any duplicate 'LINENAME' entries
- Dropped audits that contained above 10,000 entries/exits since those were errant rows or where the entry/exit count was reset for the station or entrance