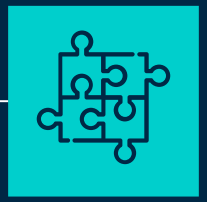
Using Data Analytics to Inform and Improve WTWY Fundraising and Outreach

Matt Ryan

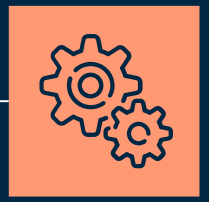
THE PROBLEM



01

MOTIVATION

WTWY is looking to expand their expand their reach, raise awareness, and attract a large gala turnout



02

OBJECTIVES

Use the data available to us to provide unique insights into our problem

03

GOALS

Provide actionable recommendations based on our newfound understandings

METHODOLOGY

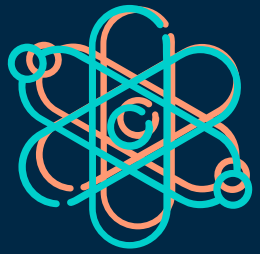
1. Start with our dataset:
   - NYC subway turnstile data from Jan 2021 to Aug 2021
2. Tools:
   - SQLAlchemy to bring the data in
   - Pandas and numpy to sanitize and manipulate our data
   - Matplotlib to visualize our data
   - SQLite for higher level data exploration

METHODOLOGY

RAW DATA

	CA	UNIT	SCP	STATION	LINENAME	DIVISION	DATE	TIME	DESC	ENTRIES	EXITS	STATION_LINE
0	A002	R051	02-00-00	59 ST	NQR456W	BMT	2020-12-26	03:00:00	REGULAR	7508509	2557569	59 ST, NQR456W
1	A002	R051	02-00-00	59 ST	NQR456W	BMT	2020-12-26	07:00:00	REGULAR	7508511	2557581	59 ST, NQR456W
2	A002	R051	02-00-00	59 ST	NQR456W	BMT	2020-12-26	11:00:00	REGULAR	7508531	2557636	59 ST, NQR456W
3	A002	R051	02-00-00	59 ST	NQR456W	BMT	2020-12-26	15:00:00	REGULAR	7508610	2557667	59 ST, NQR456W
4	A002	R051	02-00-00	59 ST	NQR456W	BMT	2020-12-26	19:00:00	REGULAR	7508767	2557689	59 ST, NQR456W
5	A002	R051	02-00-00	59 ST	NQR456W	BMT	2020-12-26	23:00:00	REGULAR	7508852	2557704	59 ST, NQR456W
6	A002	R051	02-00-00	59 ST	NQR456W	BMT	2020-12-27	03:00:00	REGULAR	7508858	2557705	59 ST, NQR456W
7	A002	R051	02-00-00	59 ST	NQR456W	BMT	2020-12-27	07:00:00	REGULAR	7508862	2557711	59 ST, NQR456W
8	A002	R051	02-00-00	59 ST	NQR456W	BMT	2020-12-27	11:00:00	REGULAR	7508885	2557743	59 ST, NQR456W
9	A002	R051	02-00-00	59 ST	NQR456W	BMT	2020-12-27	15:00:00	REGULAR	7508955	2557778	59 ST, NQR456W

DATA MAGIC

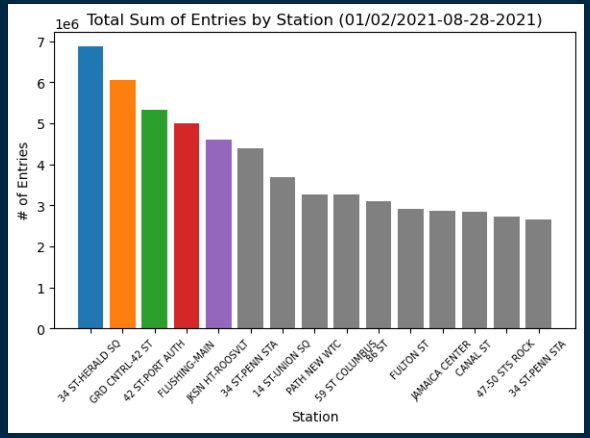


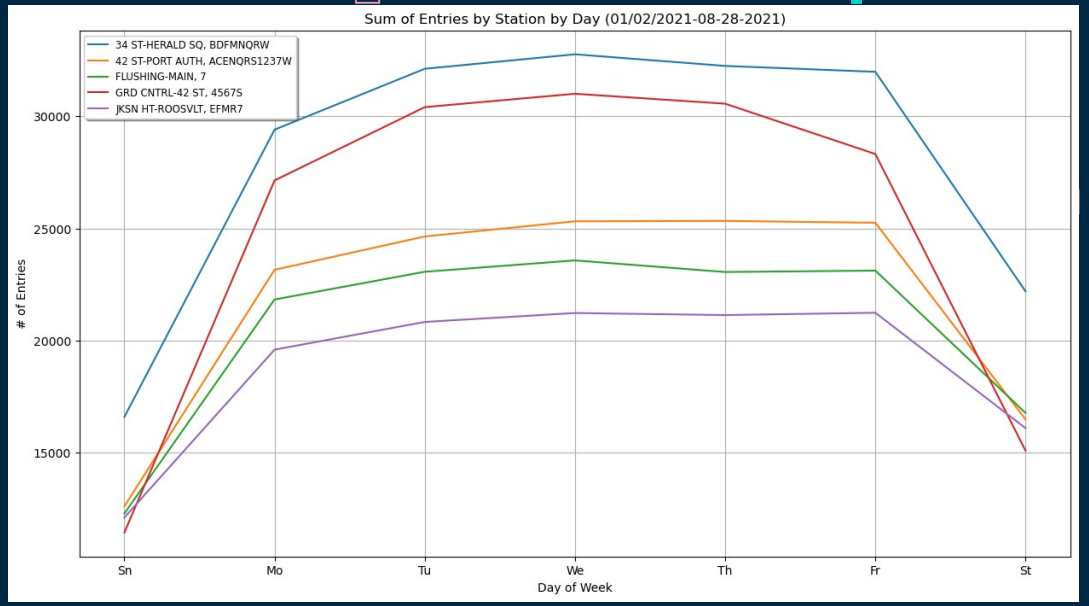
(EDA CLEANING AND MANIPULATION)

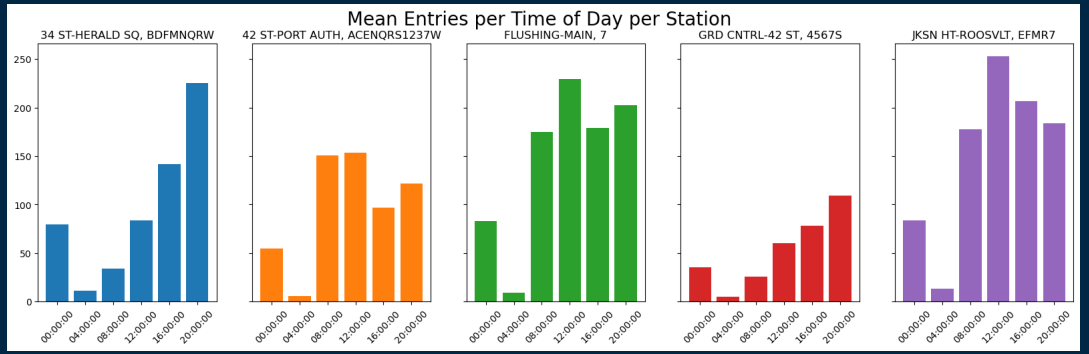
NEW DATA

	CA	UNIT	SCP	STATION_LINE	DATE	ENTRIES	DAILY_TOT	DAILY_TOT_ABS
1	A002	R051	02-00-00	59 ST, NQR456W	2020-12-27	7508858	349.0	349.0
2	A002	R051	02-00-00	59 ST, NQR456W	2020-12-28	7509173	315.0	315.0
3	A002	R051	02-00-00	59 ST, NQR456W	2020-12-29	7509828	655.0	655.0
4	A002	R051	02-00-00	59 ST, NQR456W	2020-12-30	7510464	636.0	636.0
5	A002	R051	02-00-00	59 ST, NQR456W	2020-12-31	7511071	607.0	607.0
6	A002	R051	02-00-00	59 ST, NQR456W	2021-01-01	7511448	377.0	377.0

RESULTS


Total Sum of Entries by Station (01/02/2021-08-28-2021)
1e6
# of Entries
Station
34 ST-HERALD SQ
GRD CNTRL-42 ST
42 ST-PORT AUTH
FLUSHING-MAIN
JKSN HT-ROOSVLT
34 ST-PENN STA
14 ST-UNION SQ
PATH NEW WTC
59 ST COLUMBUS
86 ST
FULTON ST
JAMAICA CENTER
CANAL ST
47-50 STS ROCK
34 ST-PENN STA


Sum of Entries by Station by Day (01/02/2021-08-28-2021)
34 ST-HERALD SQ, BDFMNQRW
42 ST-PORT AUTH, ACENQRS1237W
FLUSHING-MAIN, 7
GRD CNTRL-42 ST, 4567S
JKSN HT-ROOSVLT, EFMR7
# of Entries
Day of Week
Sn
Mo
Tu
We
Th
Fr
St


Mean Entries per Time of Day per Station
34 ST-HERALD SQ, BDFMNQRW
42 ST-PORT AUTH, ACENQRS1237W
FLUSHING-MAIN, 7
GRD CNTRL-42 ST, 4567S
JKSN HT-ROOSVLT, EFMR7

CONCLUSIONS

- Target the 5 high volume stations with street teams
- The middle/end of the work week is consistently high volume across all 5 stations
- Focus night teams on 34 St – Herald; Flushing – Main and Jksn Ht – Roosevelt see big influxes around noon

CONCLUSIONS

“This is a quote. Words full of wisdom that someone important said and can make the reader get inspired.”

—SOMEONE FAMOUS

Future Work

1. Further sanitization of data
2. Deeper investigation into trends or relation to events
3. Potential to tie in US Census Data
    - Look to areas with higher women/men ratios
    - Using median income by zipcode to target deeper pockets or people who might need a leg up

THANK YOU

APPENDIX

RESOURCES

- Github Repo
- MTA Data