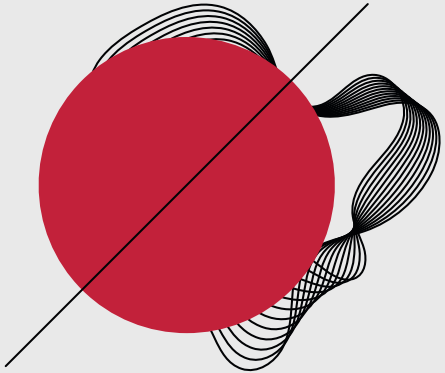
COVID-19 VACCINE CENTERS

A faster way to herd immunity.



Prepared by:
Yazeed Musallam

NYC DATA SCIENCE

01.

Introduction

NYC DATA SCIENCE

Back story of the project

- **40% percent of new yorkers uses the subway daily!**
- **Placing the vaccination centers near to subway stations is great idea!**
  - People can get can get vaccinated while going to work/home!
- **New York City Department of Health seeked to NYC Data Science company to answer:**
  - Where is the most effective stations to place the vaccination centers near by?

NYC DATA SCIENCE

Scope of the project

Sample size:

- Four months worth of data (3M rows); reason:
NYC started to distribute vaccine for all ages before 4 months.

Datasets used:

- MTA turnstile dataset.
- MTA subway stations zip code dataset.
- NYC-health COVID-19 vaccine dataset.

NYC DATA SCIENCE

02.

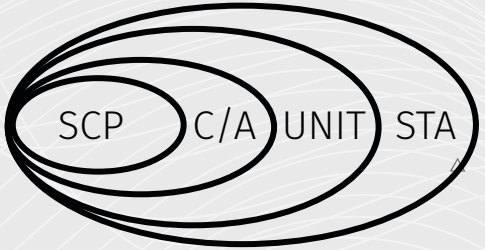
Exploratory Data Analysis

NYC DATA SCIENCE

Working with the dataset

- **Grouping each turnstile**
- **Assumation:**

Total number of turnstiles: 5032




```
Data columns (total 7 columns):
 #   Column      Dtype
---  ------      -----
 0   STATION     object
 1   UNIT        object
 2   C/A         object
 3   SCP         object
 4   DATE_TIME   datetime64[ns]
 5   ENTRIES     int64
 6   EXITS       int64
dtypes: datetime64[ns](1), int64(2), object(4)
memory usage: 201.2+ MB
```

```
Data columns (total 5 columns):
 #   Column         Dtype
---  ------         -----
 0   STATION        object
 1   TURNSTILE ID   object
 2   DATE_TIME      datetime64[ns]
 3   ENTRIES        int64
 4   EXITS          int64
dtypes: datetime64[ns](1), int64(2), object(2)
memory usage: 143.7+ MB
```

NYC DATA SCIENCE

Working with the dataset

MISSING VALUES

- NO MISSING VALUES ARE FOUND

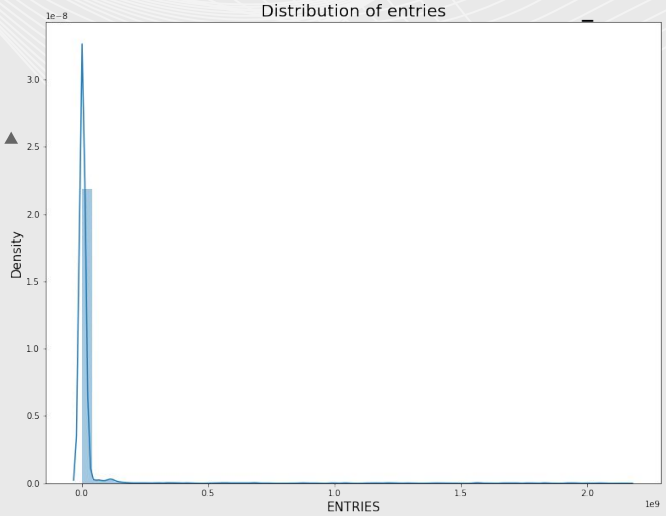
DUPLICATES

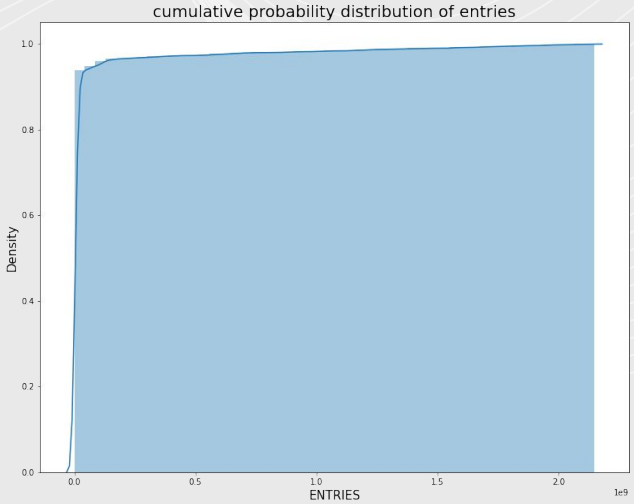
- THERE ARE DUPLICATES IN DESC COLUMN
- DUPLICATE ARE REMOVED

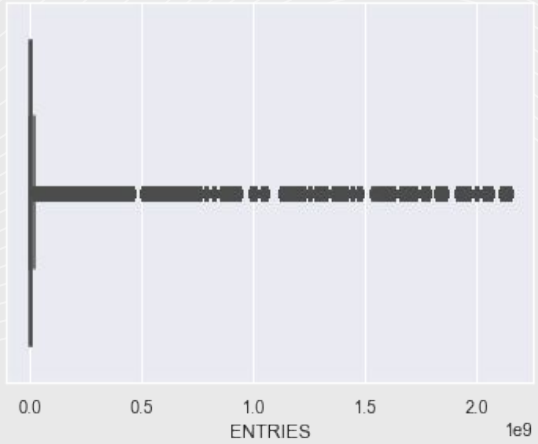
NYC DATA SCIENCE

Visualizing the data

Dataset suffers from severely outliers!





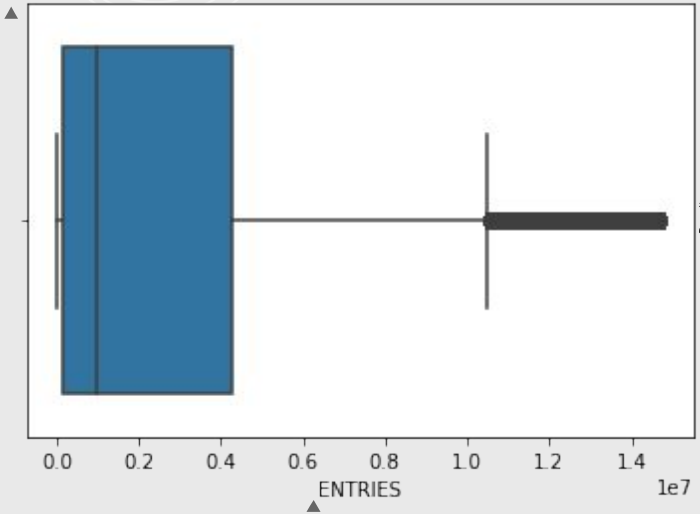


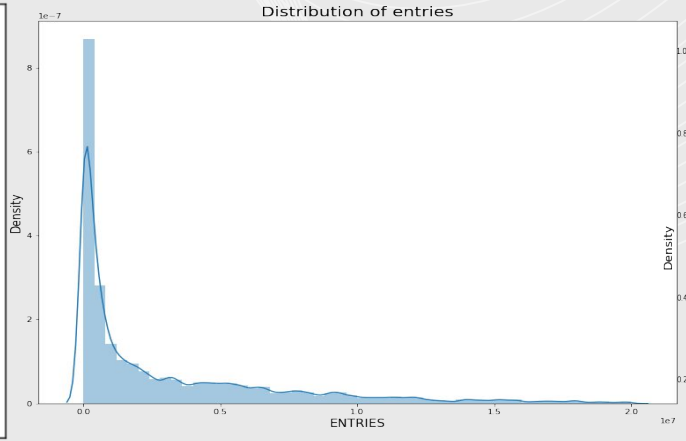
NYC DATA SCIENCE

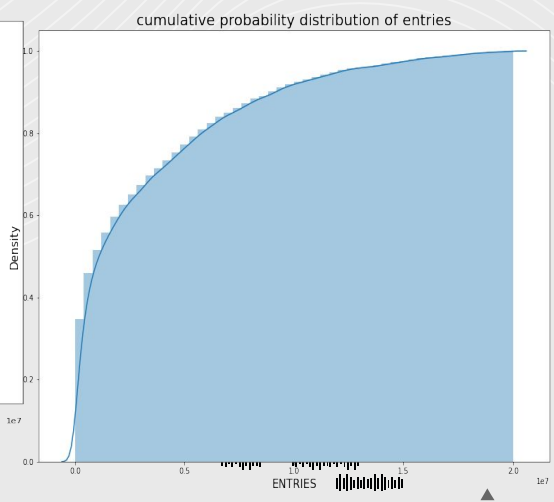
Working with outliers

Outliers threshold?

- Outliers after 15M entries for each turnstile (Using IQR)
- Empirically 20M was chosen to avoid loss of important information.
- Similar process was done to exits







1.1) Working with data - Hourly entries/exits

Challenges and solution

Observation:

1) If turnstile counting are reversed.
2) If counter increasing to more than limit of 900 per turnstile for each four hours.
3) If turnstile is suddenly overshooting in unusual behavior.

Hourly_Entries	
	430
DAY A	313

Same time different day

Hourly_Entries	
	899
DAY B	379

NYC DATA SCIENCE

1.2) Working with data - Hourly entries/exits

Challenges and solution

1) Solution to overshooting ?
   1.1) Get the median of same day at same time in different weeks.
   2.2) Removing outliers for each station.

STATION	Fixed_Hourly_Entries
34 ST-PENN STA	4352214
34 ST-HERALD SQ	3256149
125 ST	3132274
86 ST	2907853
23 ST	2849891
42 ST-PORT AUTH	2721105
FLUSHING-MAIN	2686299
14 ST-UNION SQ	2644451
JKSN HT-ROOSVLT	2592408
TIMES SQ-42 ST	2445775
96 ST	2295068

NYC DATA SCIENCE

1.2) Working with data - Hourly entries/exits

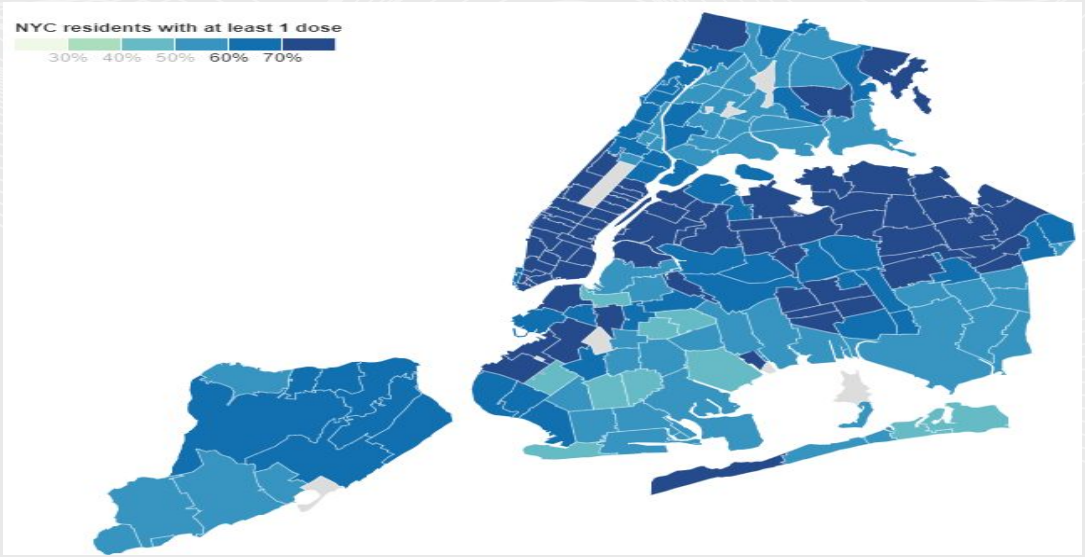
Results by removing only outliers for the whole dataset.

	STATION	entries of station (4months)	Monthly exits of station (4months)
0	34 ST-PENN STA	4091308	5042959
1	GRD CNTRL-42 ST	3277735	4428445
2	34 ST-HERALD SQ	3064163	3495343
3	23 ST	3016865	3464967
4	FULTON ST	2721699	3346058
5	59 ST	2536612	2705119
6	14 ST-UNION SQ	2448403	3207576
7	JKSN HT-ROOSVLT	2443236	2251527
8	86 ST	2419090	2886234
9	TIMES SQ-42 ST	2284576	3402594
10	125 ST	1913196	2458149

Rank	Station/Complex
1	Times Sq-42 St
2	Grand Central-42 St
3	34 St-Herald Sq
4	14 St-Union Sq
5	Fulton St
6	34 St-Penn Station
7	34 St-Penn Station
8	59 St-Columbus Circle
9	74-Broadway
10	Flushing-Main St

Source: MTA website - 2020

NYC-health vaccine dataset



Observation from data:

1) The total percentage of fully vaccinated: 61%

2) Total station in areas that has more than 70% vaccinated: 52

NYC DATA SCIENCE

Mapping MTA dataset to NYC COVID dataset

MTA DATA SET

STATION	ENTRIES (Daily)	EXI
34 ST-PENN STA	32471	400
GRD CNTRL-42 ST	26014	351
34 ST-HERALD SQ	24319	277
23 ST	23943	275
FULTON ST	21601	265

	STATION	ENTRIES (Daily)	PERC_FULLY
207	GRD CNTRL-42 ST	26014	93
54	34 ST-HERALD SQ	24319	99
11	14 ST-UNION SQ	19432	69
228	JKSN HT-ROOSVLT	19391	84
310	TIMES SQ-42 ST	18132	100
...	...	...	...
292	ROCKAWAY PARK B	274	53
131	BEACH 98 ST	180	53
124	BEACH 105 ST	121	53
145	BROAD CHANNEL	112	46
272	ORCHARD BEACH	5	66

COVID-19 vaccine set

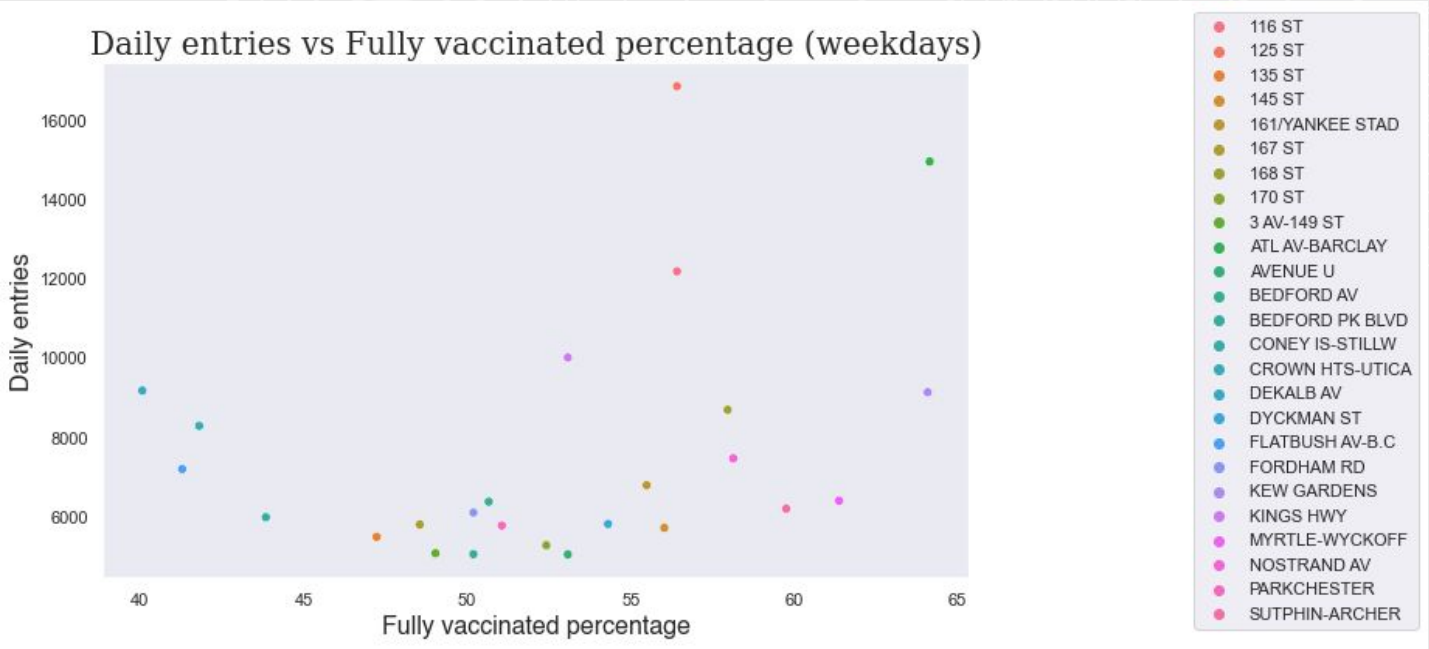
RHOOD_NAME	BOROUGH	ZIPCODE	Label	PERC_FULLY
loMad/West	Manhattan	10001	10001, 10118	99
/Lower East Side	Manhattan	10002	10002	76
amercy/Greenwich	Manhattan	10003	10003	69
District	Manhattan	10004	10004	100
District	Manhattan	10005	10005	84

02.

Insights

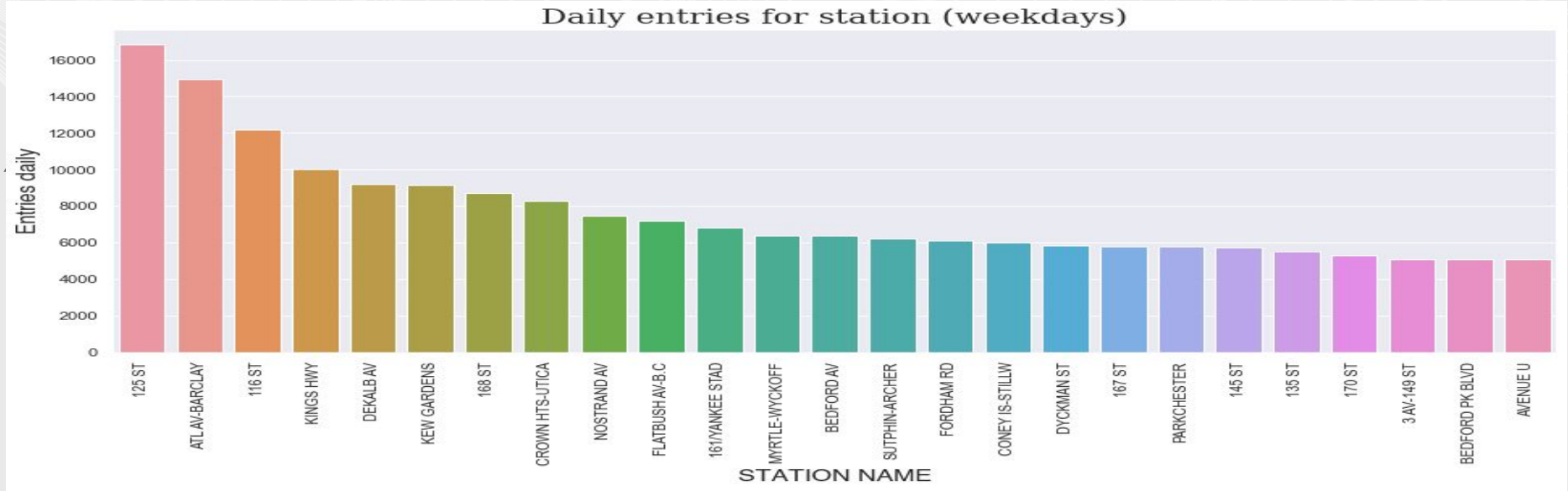
NYC DATA SCIENCE

Daily entries vs. vaccine percent by station

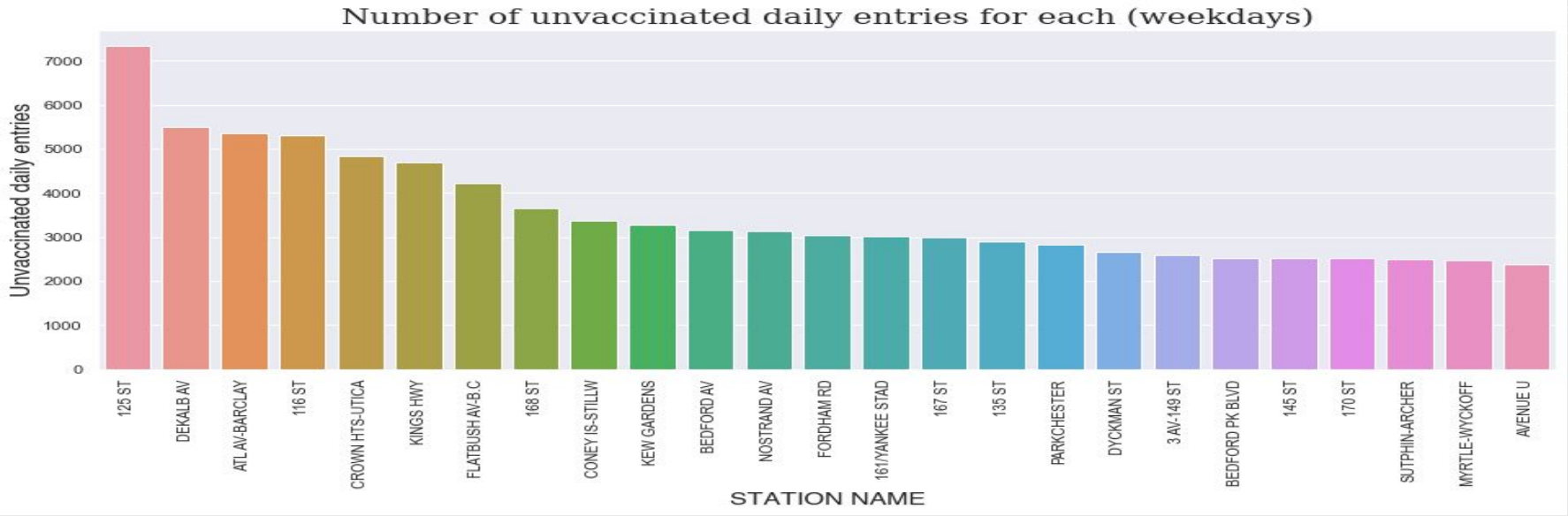


NYC DATA SCIENCE

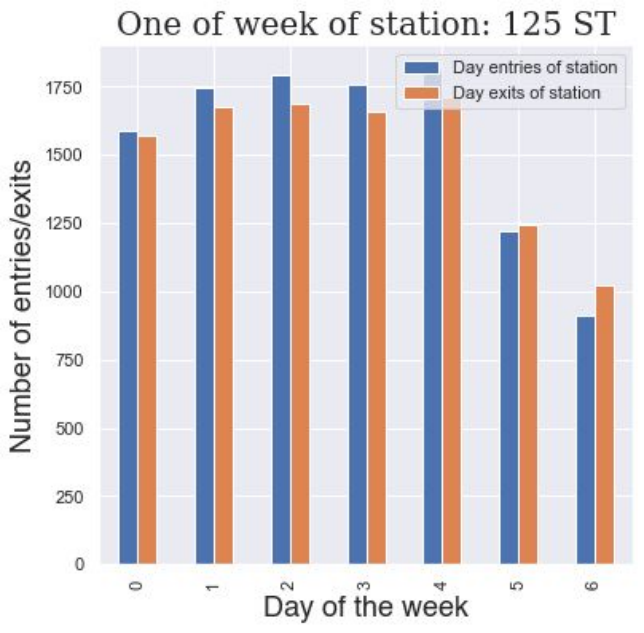
Daily entries for stations - weekdays



Number of unvaccinated for each stations - weekdays

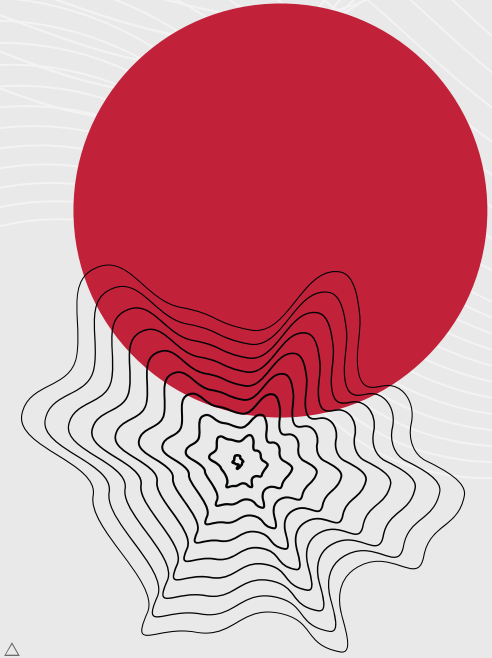


Analysis of entries/exits



NYC DATA SCIENCE

THANKS!



Future works:

1) Automated weekly reports of best vaccine centers.
2) Use dashboards i.e. using Tableau.

NYC DATA SCIENCE