



# MTA Turnstiles

Presented by: Wafaa Alharbi

# Outline

- ▶ Introduction.
- ▶ Scope.
- ▶ Preprocessing .
- ▶ Results.
- ▶ Conclusion.



# Introduction

October is the Breast Cancer Awareness Month, Unfortunately, breast cancer touches 1 in 8 women (12.4%) over the course of her life time according to statistics from [www.breastcancer.org](http://www.breastcancer.org).

In US for women, breast cancer death rates are higher than those for any other cancer, except lung cancer

the Estimated new cases in 2021 is 284,200 for all states in us and 17,540 cases in NewYork since the beginning of the year.

Glamified brand is proud to debut 2 new eyeshadow palettes that will support 3 organizations that do outstanding work for breast cancer education, research, awareness, and direct patient support.

From launch on 09/01/2021 at 3PM CST through the end of year 2020, \$2 from each sale of the palettes (up to \$200,000) will benefit The Breast Cancer Research Foundation, Living Beyond Breast Cancer, and The National Breast Cancer Foundation.



# Scope of the Project

- Sample size:  
Four months worth of data (3.5M rows) Sep-Dec 2020
- Dataset Used:
  - MTA turnstile dataset.
  - MTA subway stations zip code dataset.
  - NYC-Census dataset.







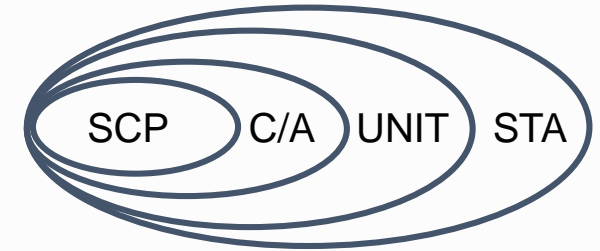
02.

# Exploratory Data Analysis

# Exploratory Data Analysis

## ○ Grouping each turnstile

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
```



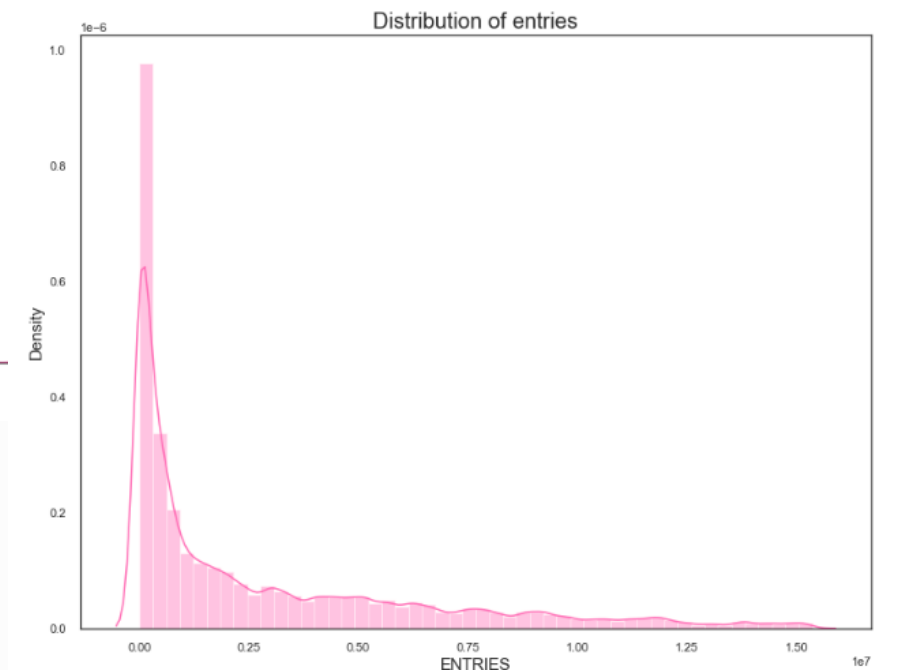
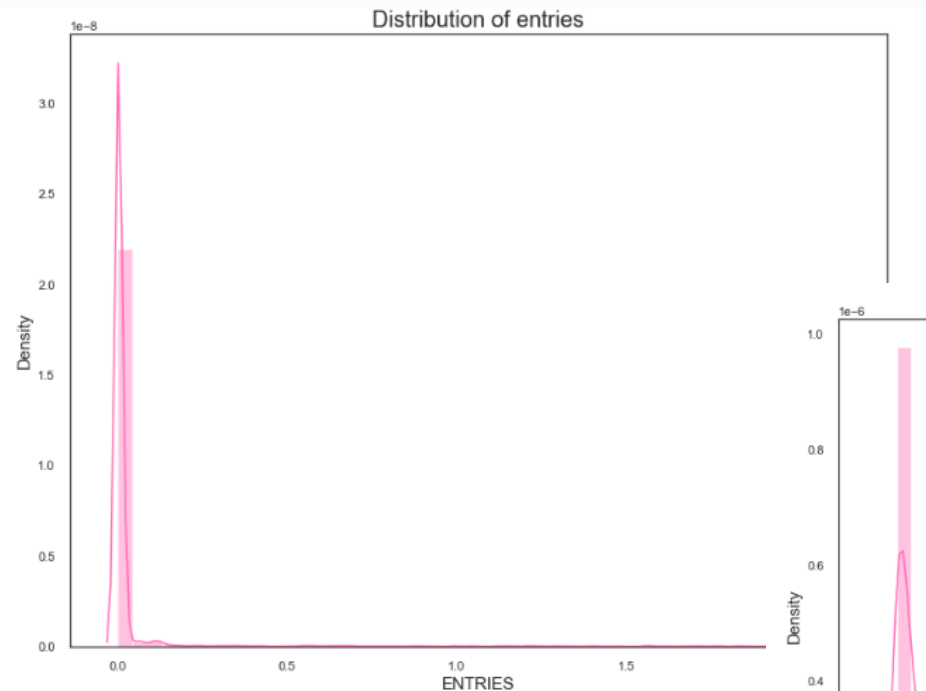
# Exploratory Data Analysis



- Removing white space from columns.
- Check any missing values.
- Check duplicates.
- Remove DESC column.

# Dealing with outliers

- Find outliers by IQR then remove it.
- Similar process was done to exits.





# NYC-Census dataset

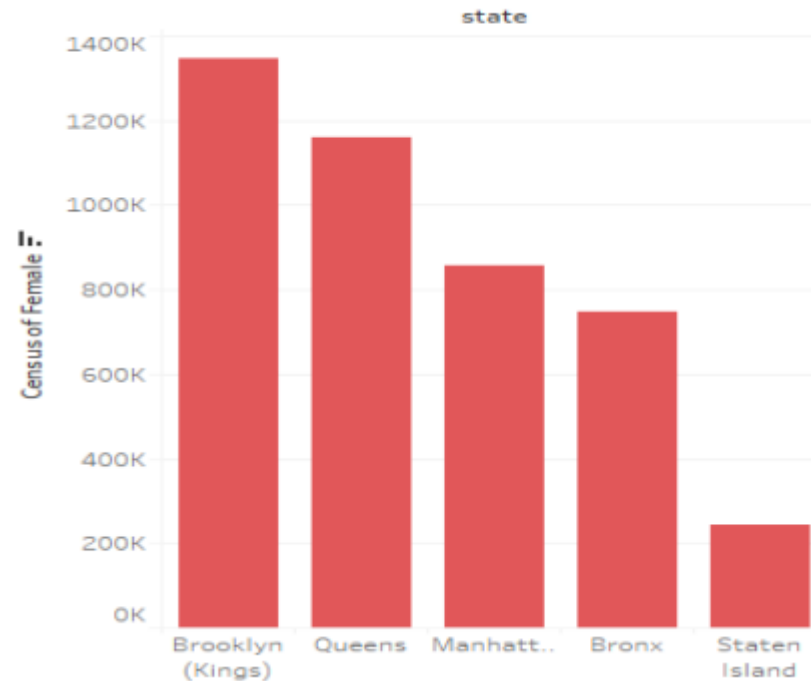
## By Gender

### Observation from data:

- 1) The total percentage of Female in NYC: 4,357,727
- 2) More than 52% of Brooklyn's census is female with Total census : 1,346,912



Census of Female in Boroughs of NYC



# Mapping MTA dataset to stations zip code dataset

MTA dataset

index	STATION	TUR	index	STATION	TURNSTILE	DATE_TIME	DATE	ENTRIES	EXITS	PREV_DATETIME	PREV_ENTRIES	PREV_EXITS	DAILY_ENTRIES	Daily_EXITS	ZIPCODE
2263453	FULTON ST	N095A-01-FUL	2266633	FULTON ST	N095A-01-FUL	2020-09-01 17:00:00	09/01/2020	3834089	6589803	09/01/2020	984108.0	352635.0	984108.0	352635.0	11217
1466936	AVENUE U	D015-R03-00-A	1468996	AVENUE U	N095-R014-00-03-08-FULTON ST	2020-12-19 12:00:00	12/19/2020	11	32	12/19/2020	28491.0	111953.0	28480.0	111921.0	11217
1111982	71 ST	E009-R06-01	1113545	71 ST	R208-R014-03-03-00-FULTON ST	2020-11-26 16:00:00	11/26/2020	5	6	11/25/2020	2355.0	49913.0	2350.0	49907.0	11217
3315597	SUTTER AV-RUTLD	R627-R00-00-S AV-	3320263	SUTTER AV-RUTLD	N422-R318-00-00-01-FULTON ST	2020-11-05 19:00:00	11/05/2020	5634714	5079658	11/05/2020	5634298.0	5079516.0	416.0	142.0	11217
3315599	SUTTER AV-RUTLD	R627-R00-00-S AV-	3320265	SUTTER AV-RUTLD	N422-R318-00-00-02-FULTON ST	2020-12-18 19:00:00	12/18/2020	35945	14042	12/18/2020	35553.0	13912.0	392.0	130.0	11217

stations zip code of  
Brooklyn

STATION	ZIPCODE
ST-PROSPECT	11215
4 AV-9 ST	11215
4AV-9 ST	11215
7 AV	11215
PROSPECT AV	11215
...	...
NASSAU AV	11222
JUNIUS ST	11212
LOCKAWAY AV	11212
SARATOGA AV	11212
ER AV-RUTLD	11212



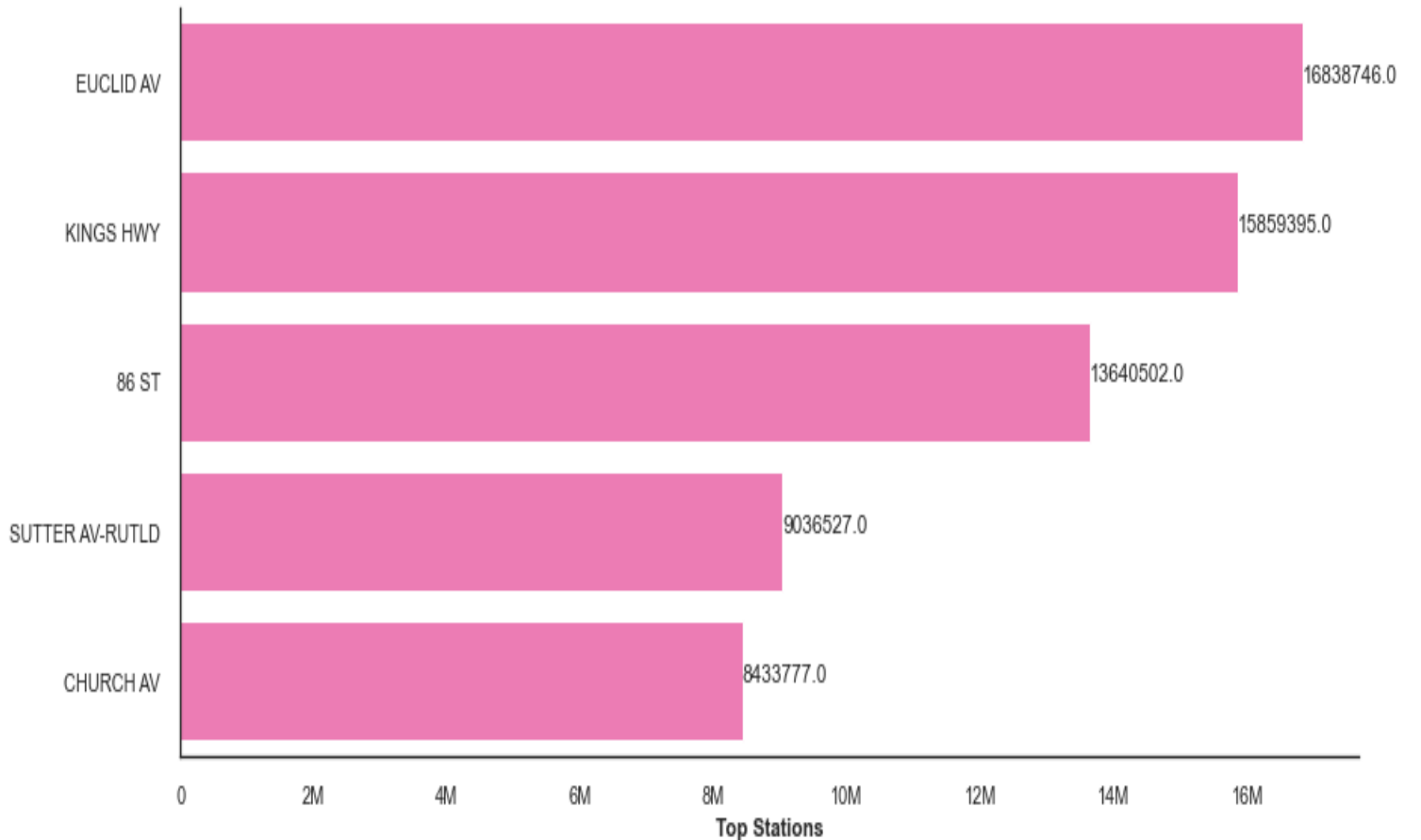
## 03. Insights

# Most Crowded Stations in Brooklyn



Total Traffic PER 17-WEEKS

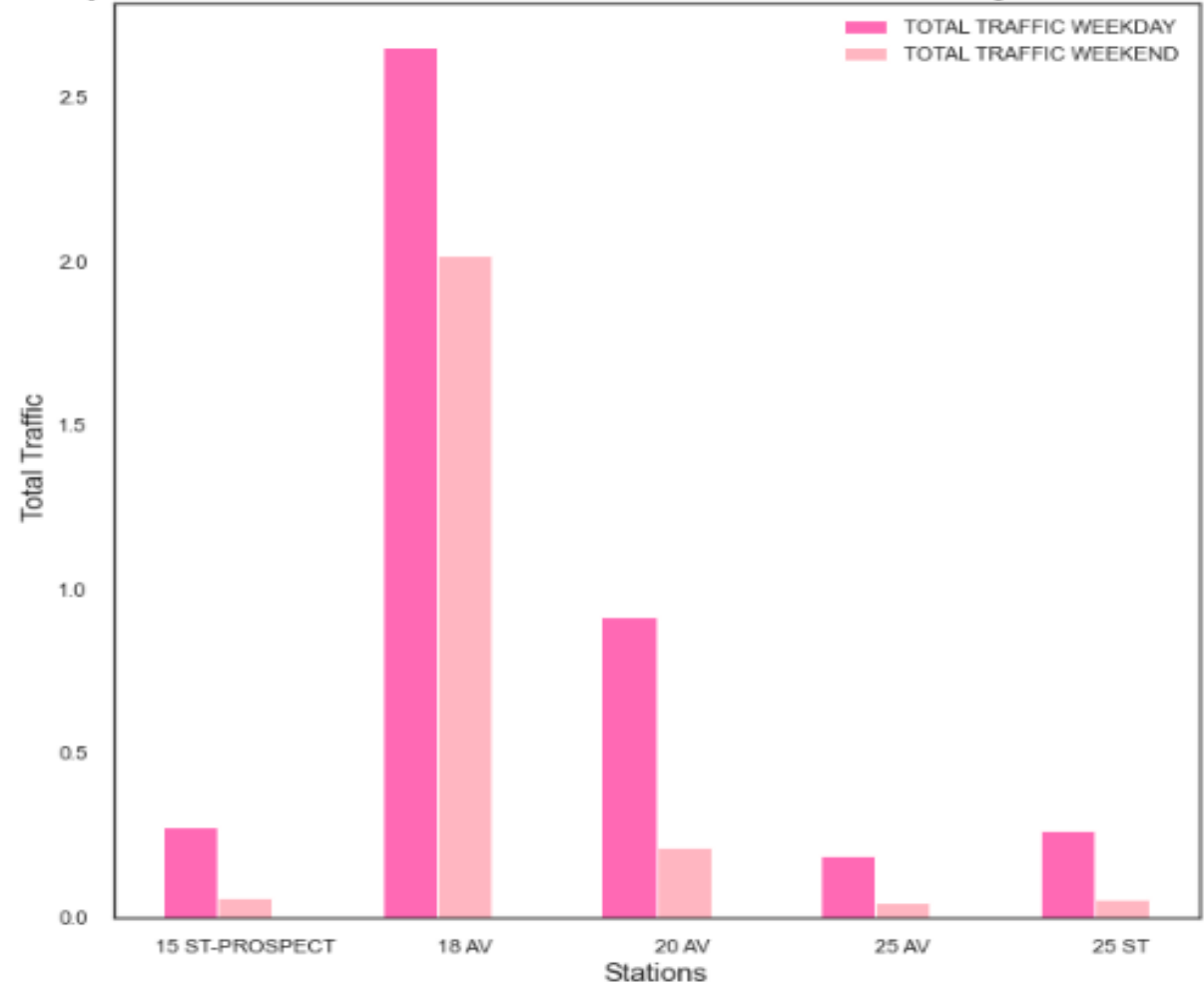
## MOST CROWDED STATIONS



# Top 5 Stations in Weekday and Weekend



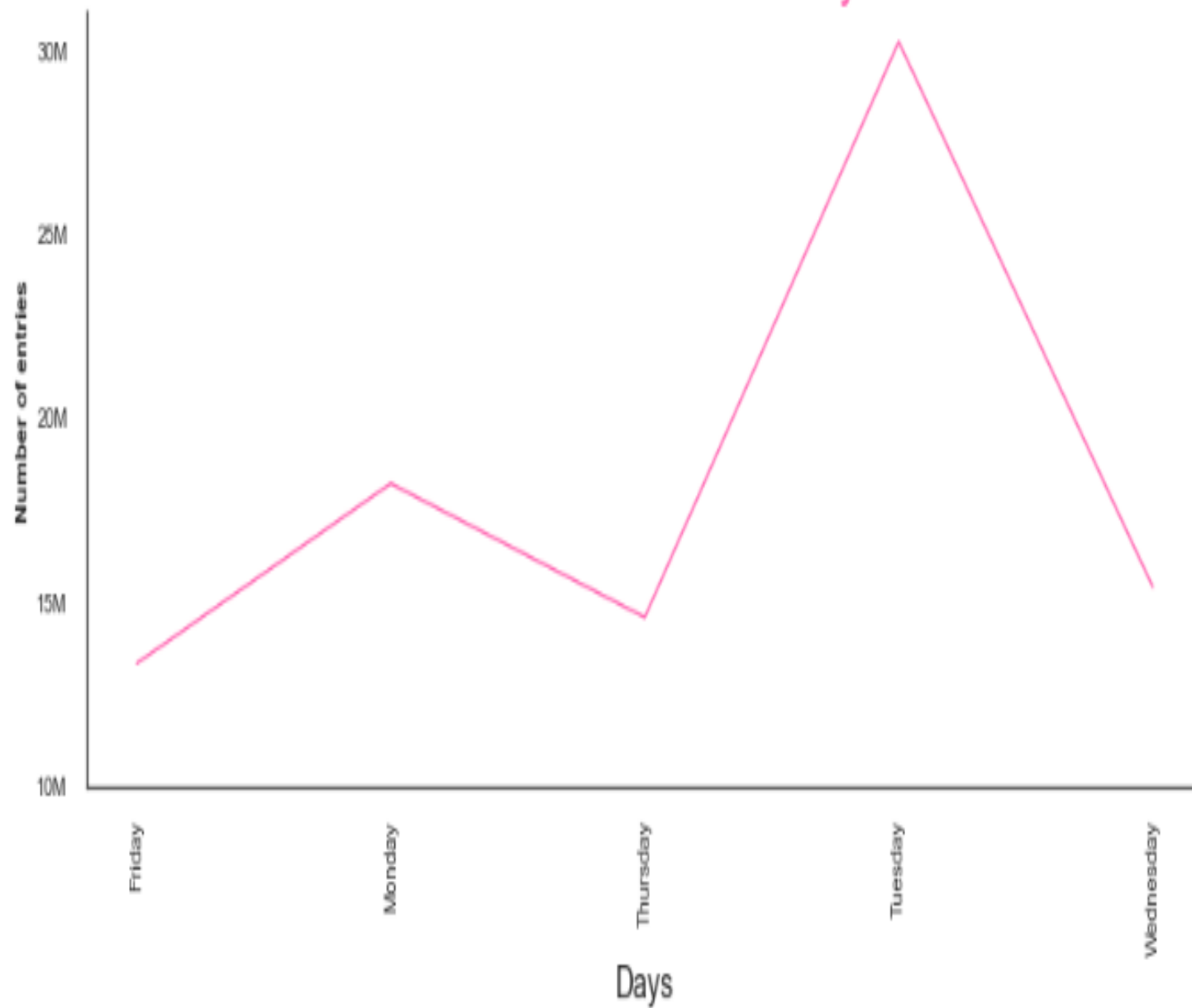
Top 5 stations are more crowded in each of weekday and weekend



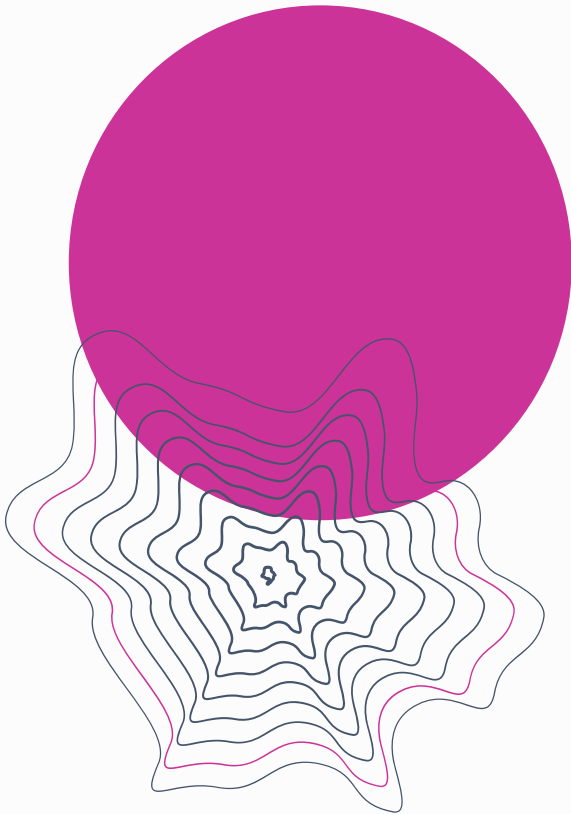




Number of entries for each day



# THANKS!



## Conclusion:

The street team will be placed at the entrances to subway stations in weekday especially in Tuesday.

## Future works:

Automated reports every year for best day and date .