

# MTA DATA ANALYSIS

Yonga Lhamo

# AGENDA

1. Introduction
2. Exploration
3. Filter
4. Visualization
5. Solution

# INTRODUCTION

## Outline of a Problem

WomenTechWomenYes (WTWY) has their annual summer gala to raise funds for their growing organization and also to spread awareness about increasing participation of women in technology. They need help in optimizing the placement of their street teams around New York City to gather maximum amount of signatures and attendees for the gala.

MTA TURNSTILE  
DATASET

Subway traffic information for the month of March, April and May 2019

ZIP CODES  
DATASET

Google API

NEIGHBORHOOD  
DATASET

Millennials account for about a quarter of the \$48 billion spent on other products in 2018

TURNSTILE  
COORDINATES

Loss of productivity costing consumers thousands of dollars

## PROBLEM

# EXPLORATION

Each Turnstile in the MTA has a unique identification

A single or a cluster of turnstiles can make up each station

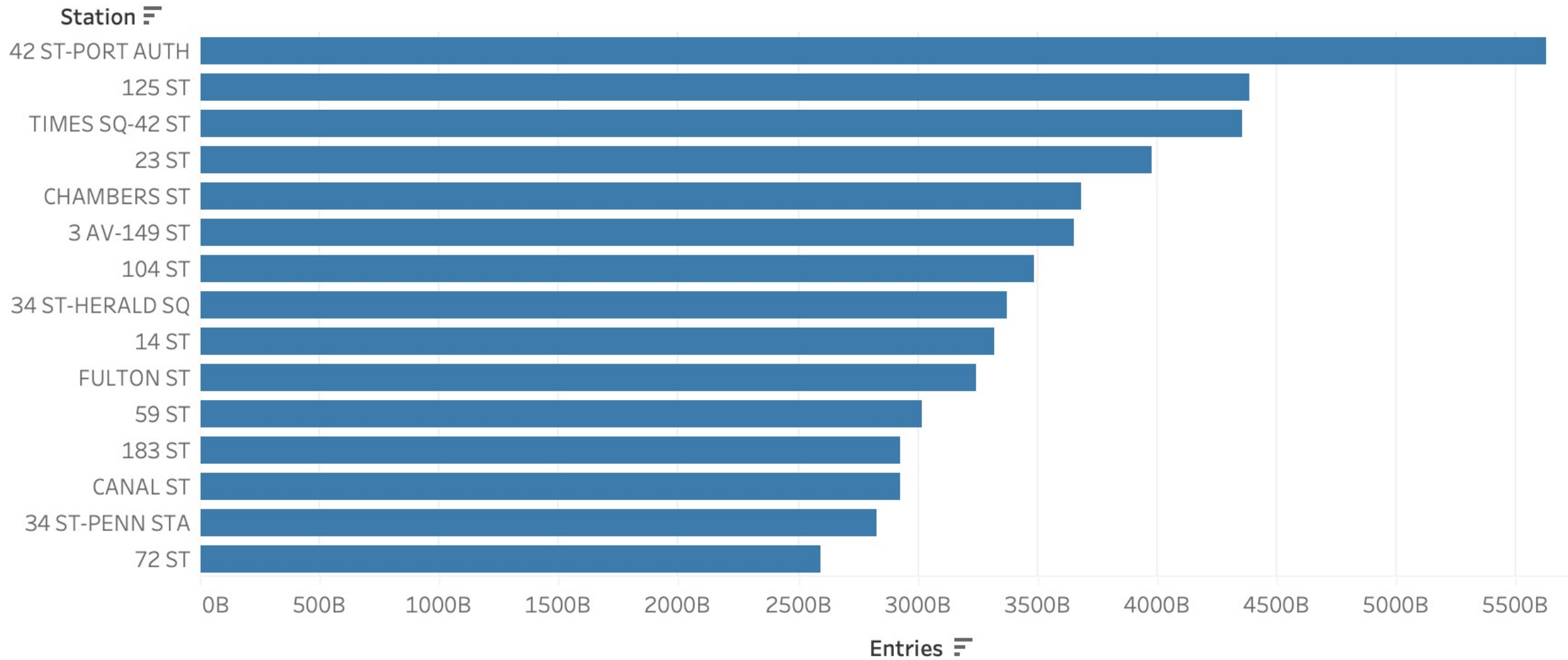
The Audits are most done with 4 hour time intervals

Issues: lack of accuracy and consistancy

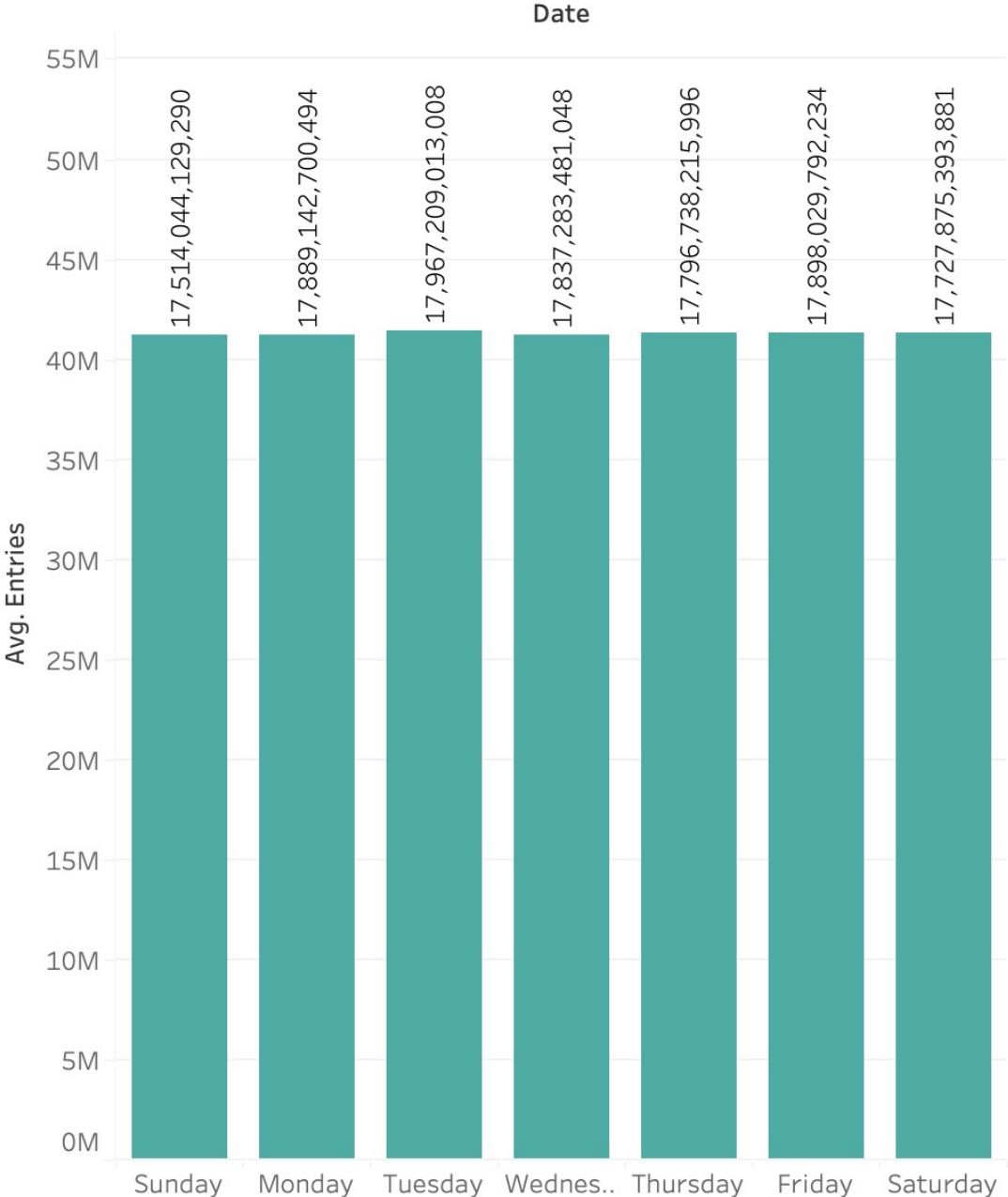
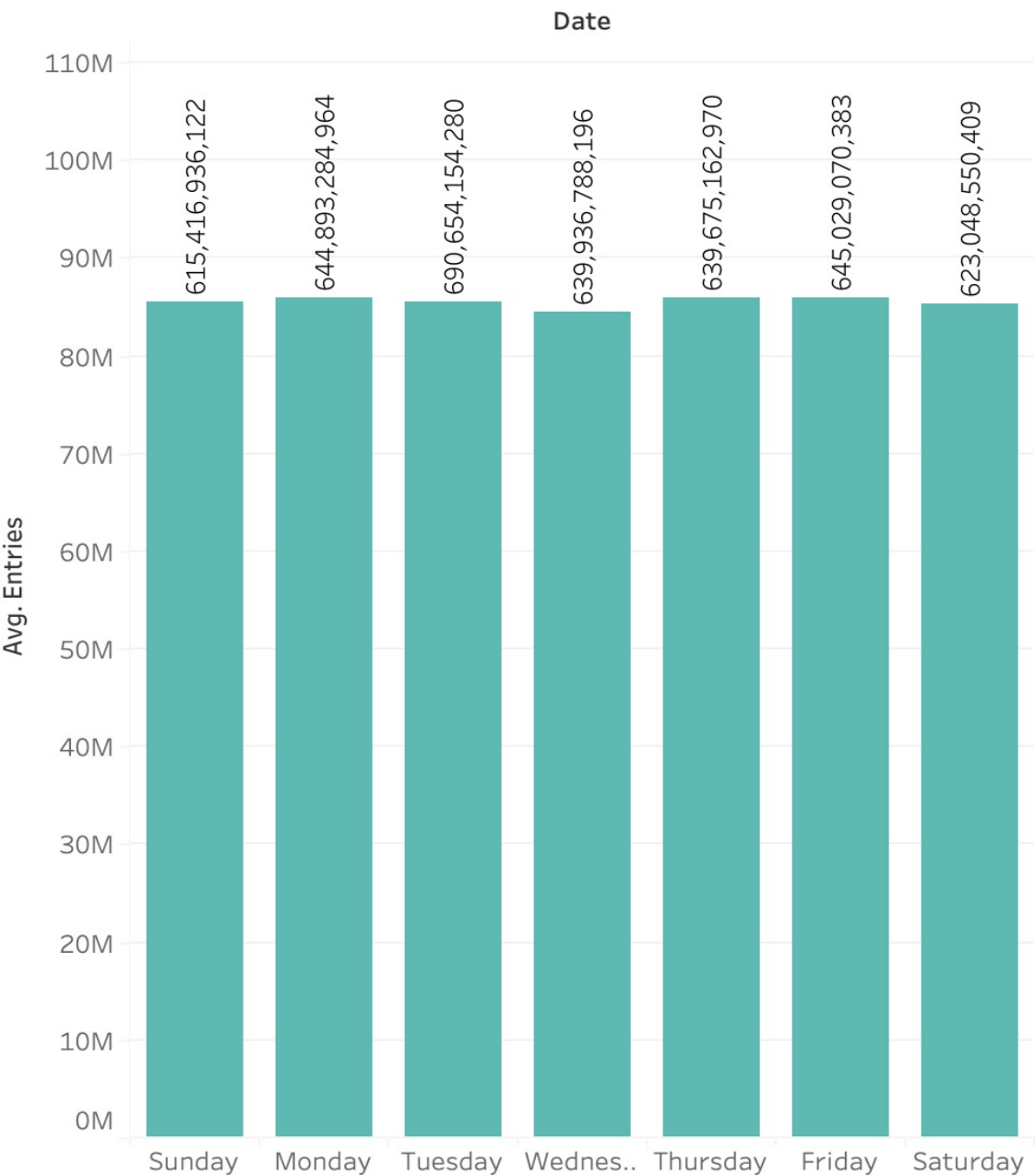
# EXPLORATION

Entries and Exits have counters that are continuous since inception for each turnstile

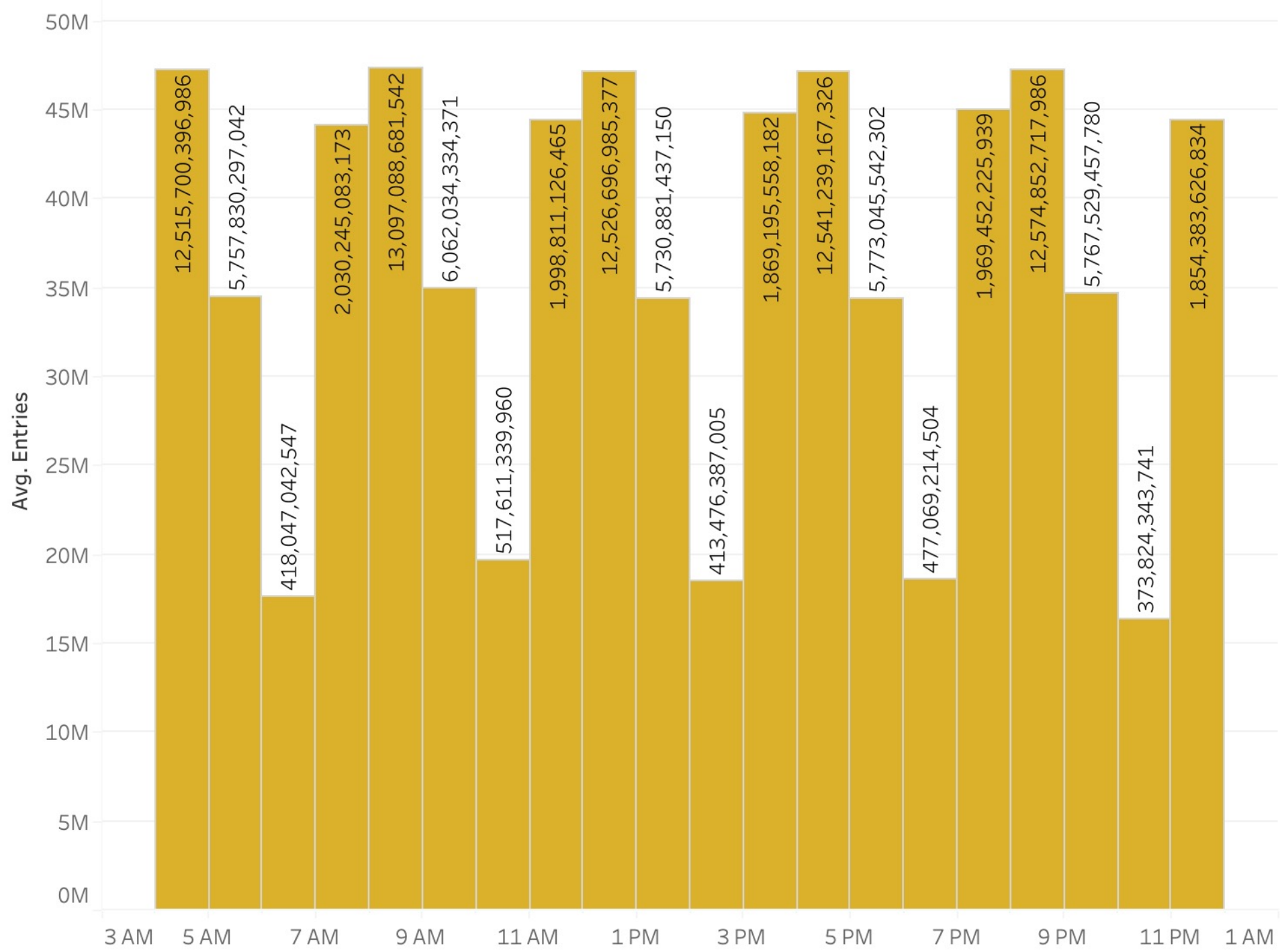
Stations with highest volume of commuters can be found by adding up the traffic for each of its turnstiles over the given timeframe



34 ST PENN STATION



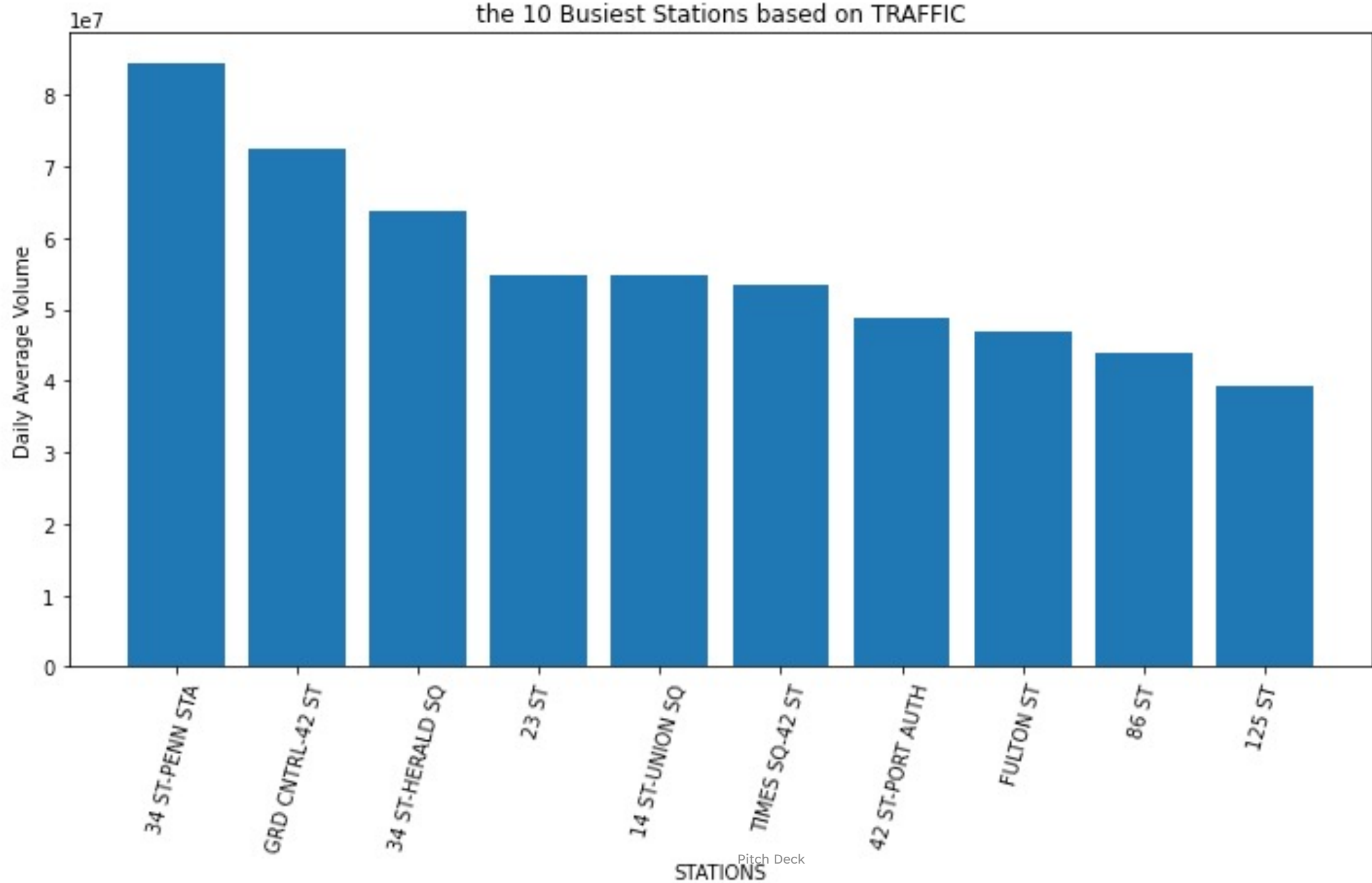




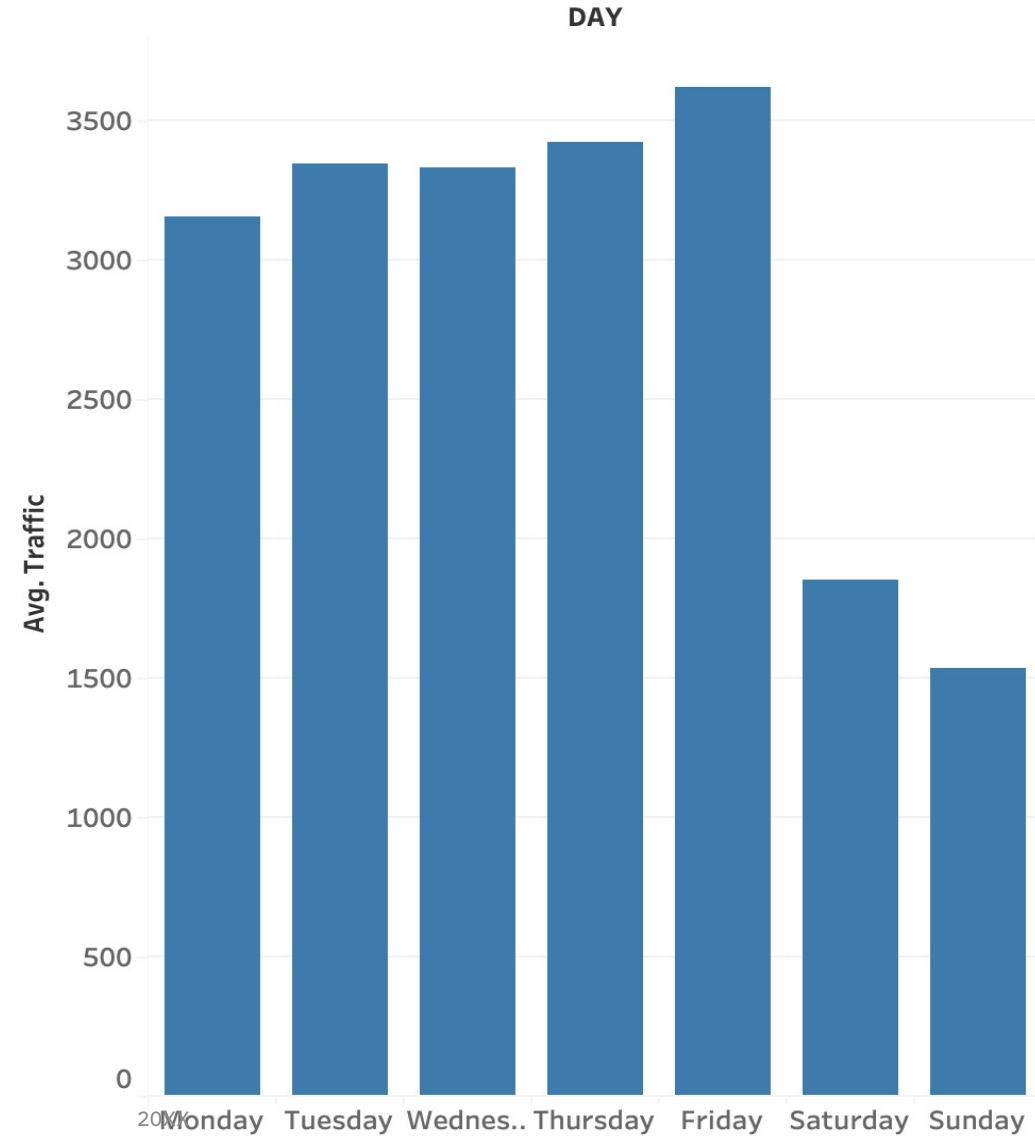
## AFTER SOME REMEDIES

- For each turnstile, take the last audit value and negate from previous
- Set an upperbound and reset count to best save and estimate count
- Set the negative values positive

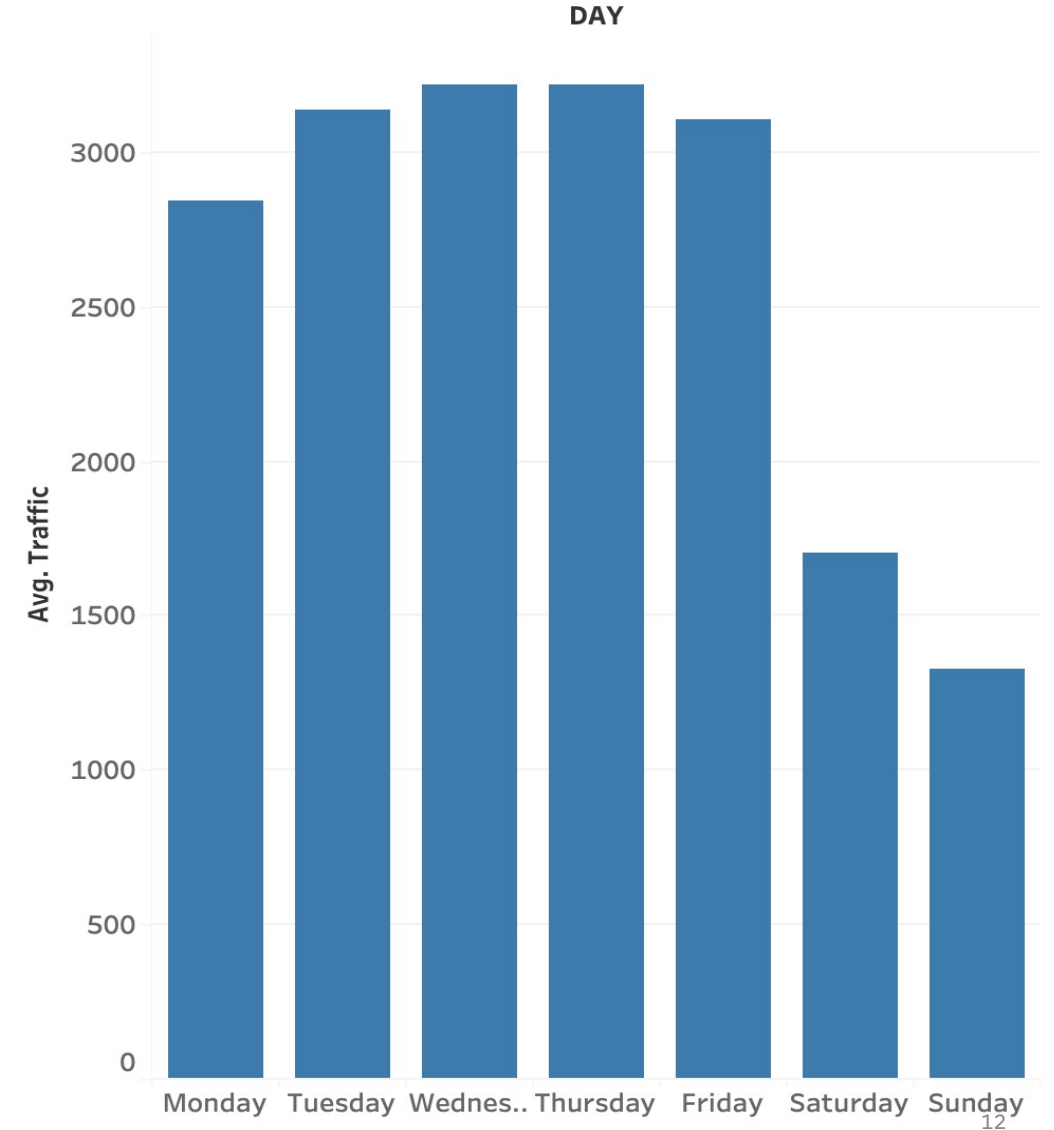
the 10 Busiest Stations based on TRAFFIC

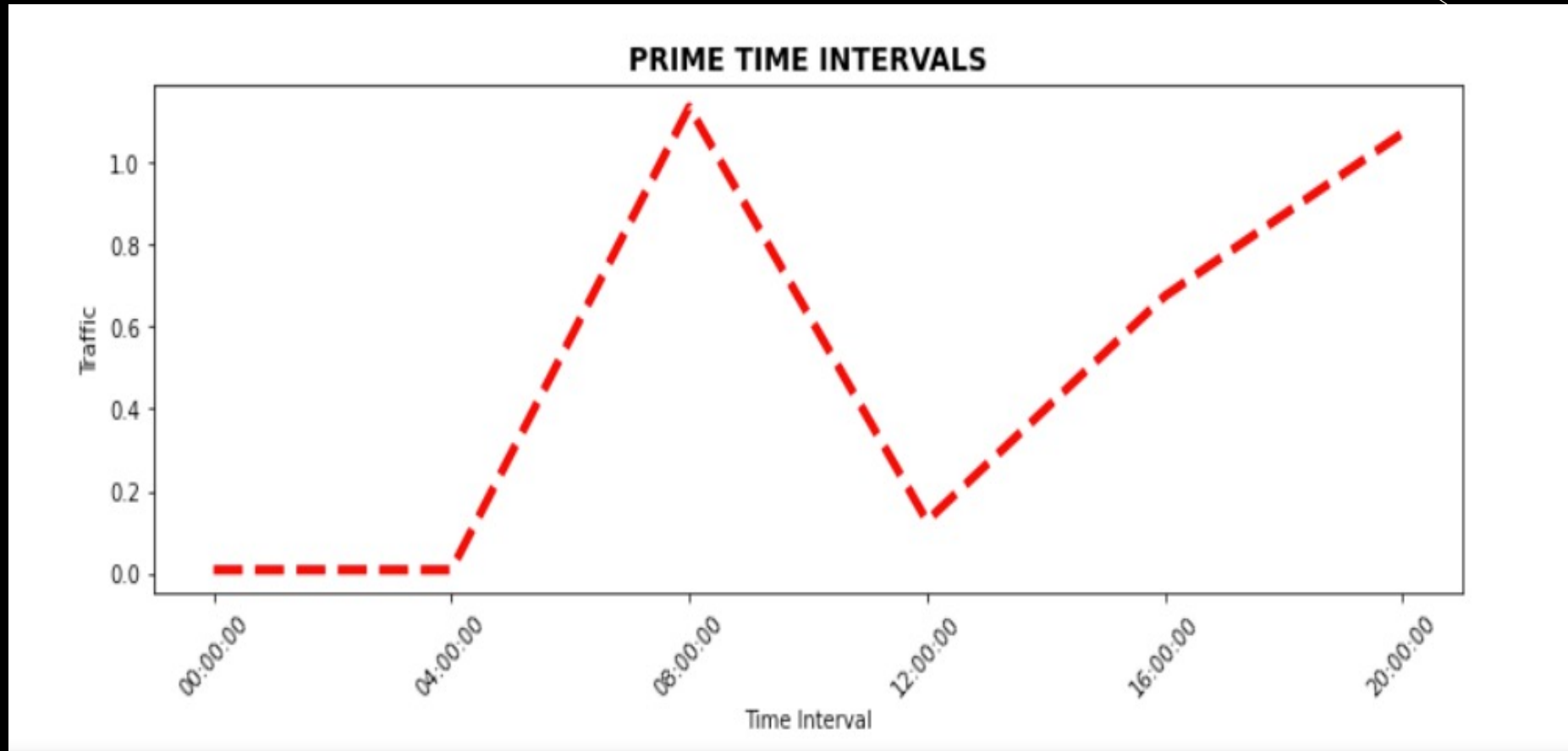


## AVERAGE TRAFFIC 34-ST PENN STATION

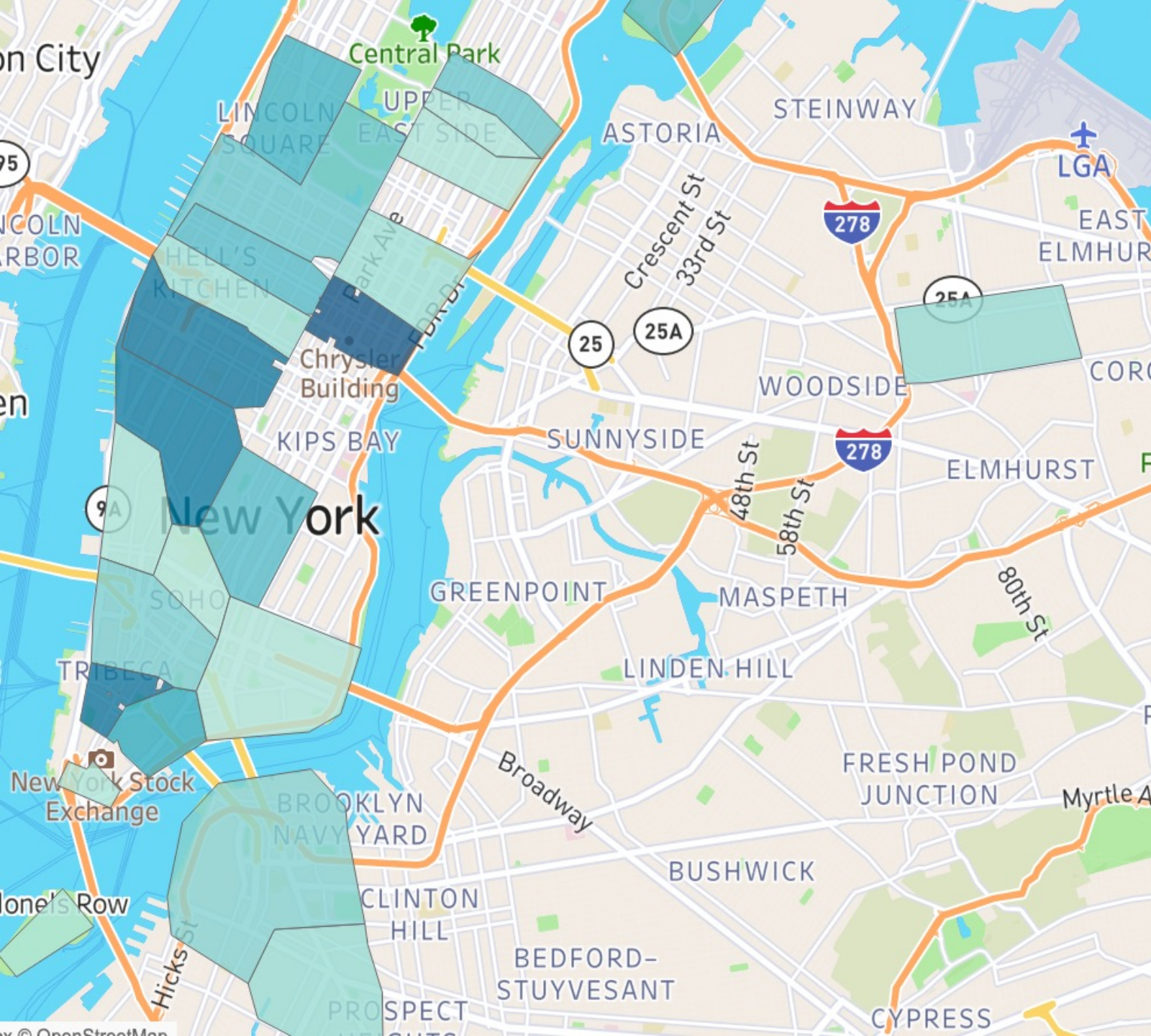


## AVERAGE TRAFFIC FOR ALL STATION





The prime time intervals are between 7 am to 9am and 4pm to 8pm



## FILTER

Filter data to the top 30 stations and merge zip code dataset. These are the some of the locations.

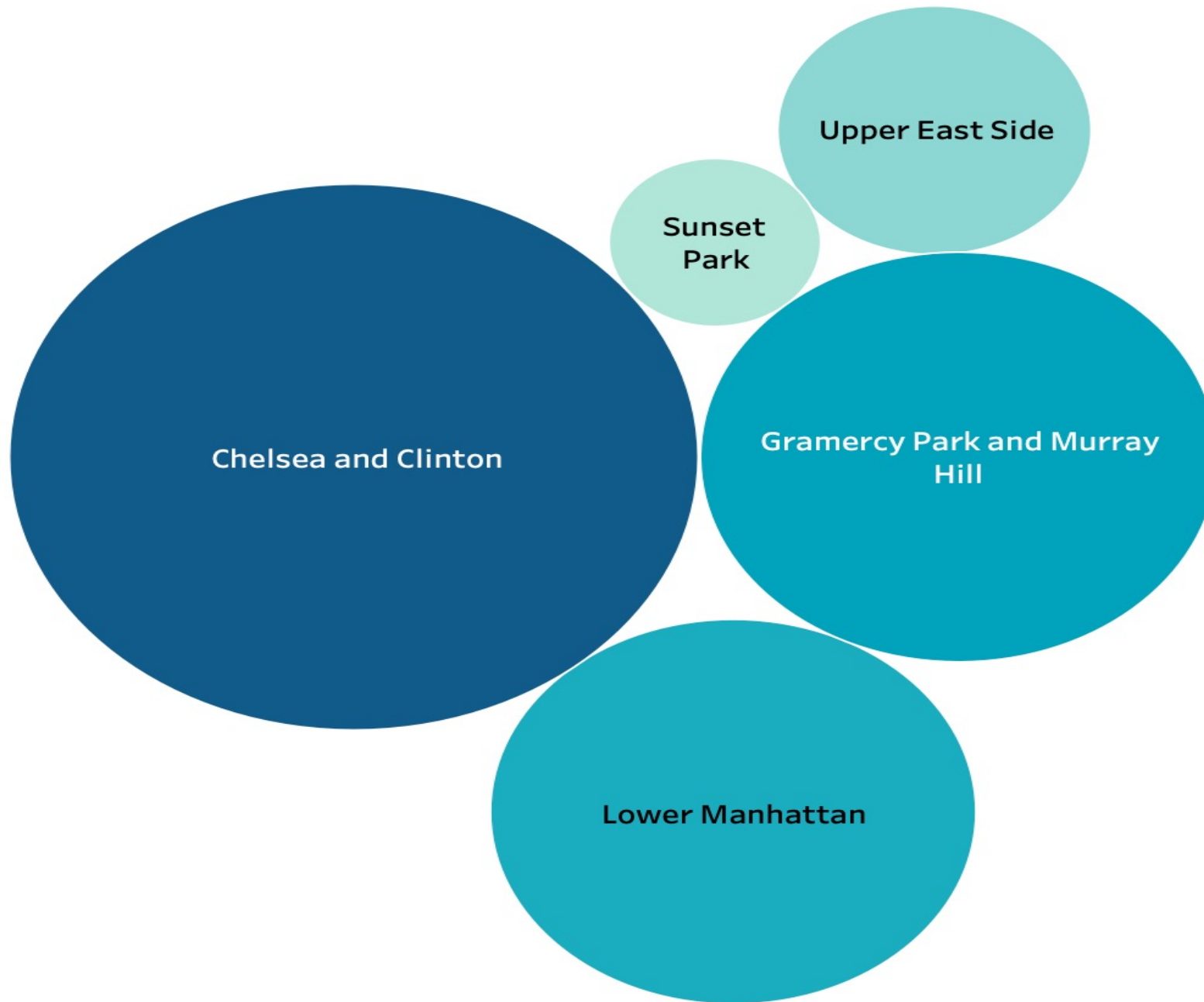




Merge with neighborhood data .  
Visualized here.  
Further filter the dataset by  
selecting five neighborhoods that  
pertain to the objective of the  
analysis.

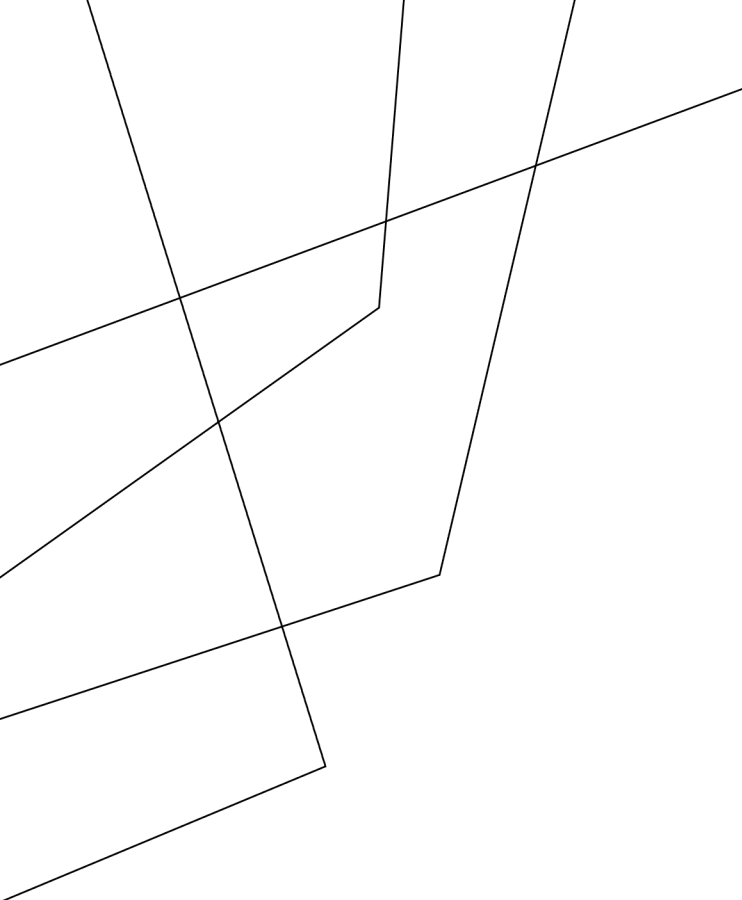
- ☐ Chelsea and Clinton
- ☐ Gramercy Park and Murray Hill
- ☐ Lower Manhattan
- ☐ Sunset Park
- ☐ Upper East side

# Neighborhoods



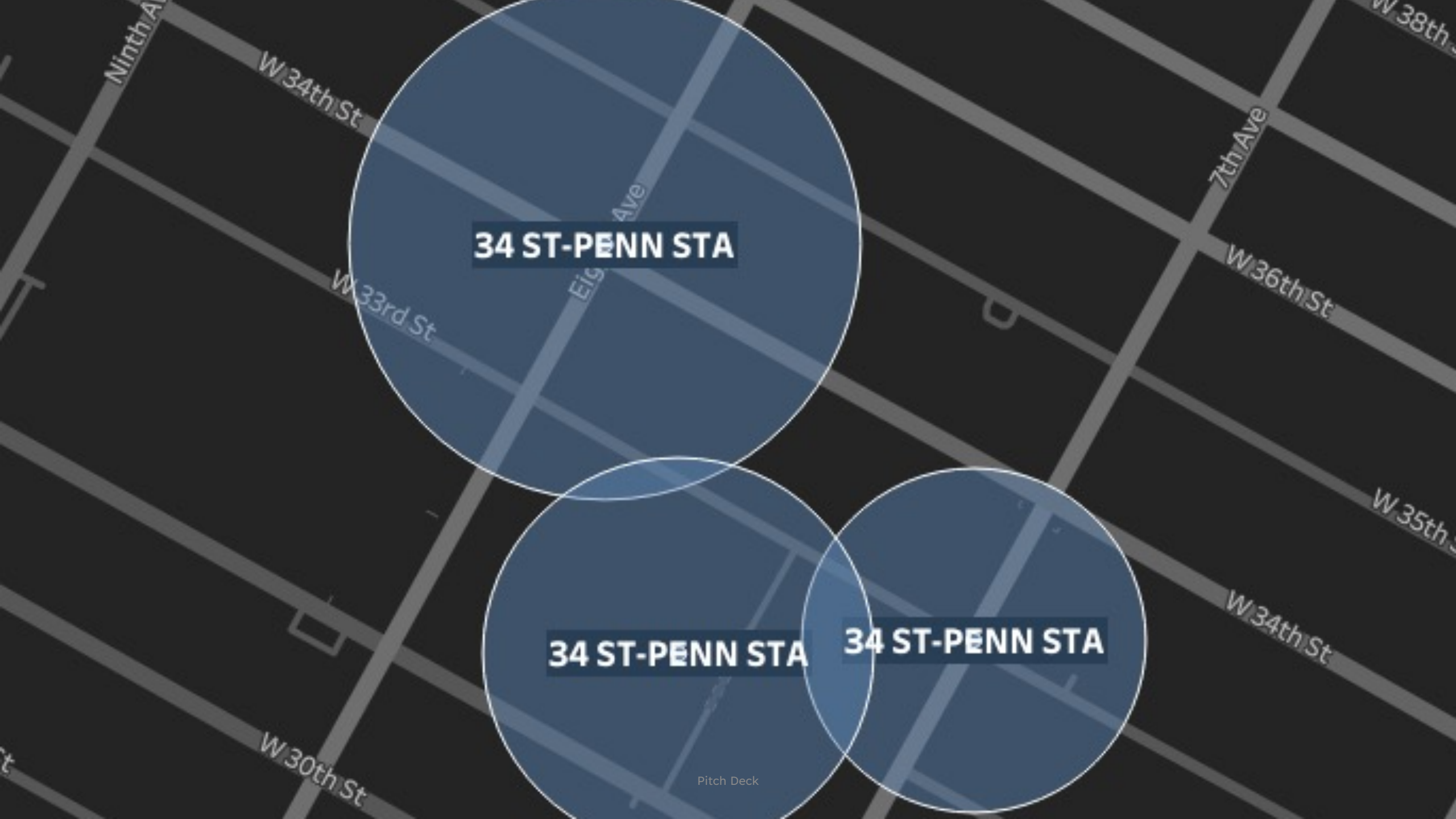
- ❑ Chelsea and Clinton  
High Traffic and tech hub
- ❑ Gramercy Park and Murray Hill  
High Traffic and tech hub
- ❑ Lower Manhattan  
High Traffic and tech hub
- ❑ Sunset Park  
Tech hub
- ❑ Upper East side  
Finance





Merge data with Turnstile geo locations



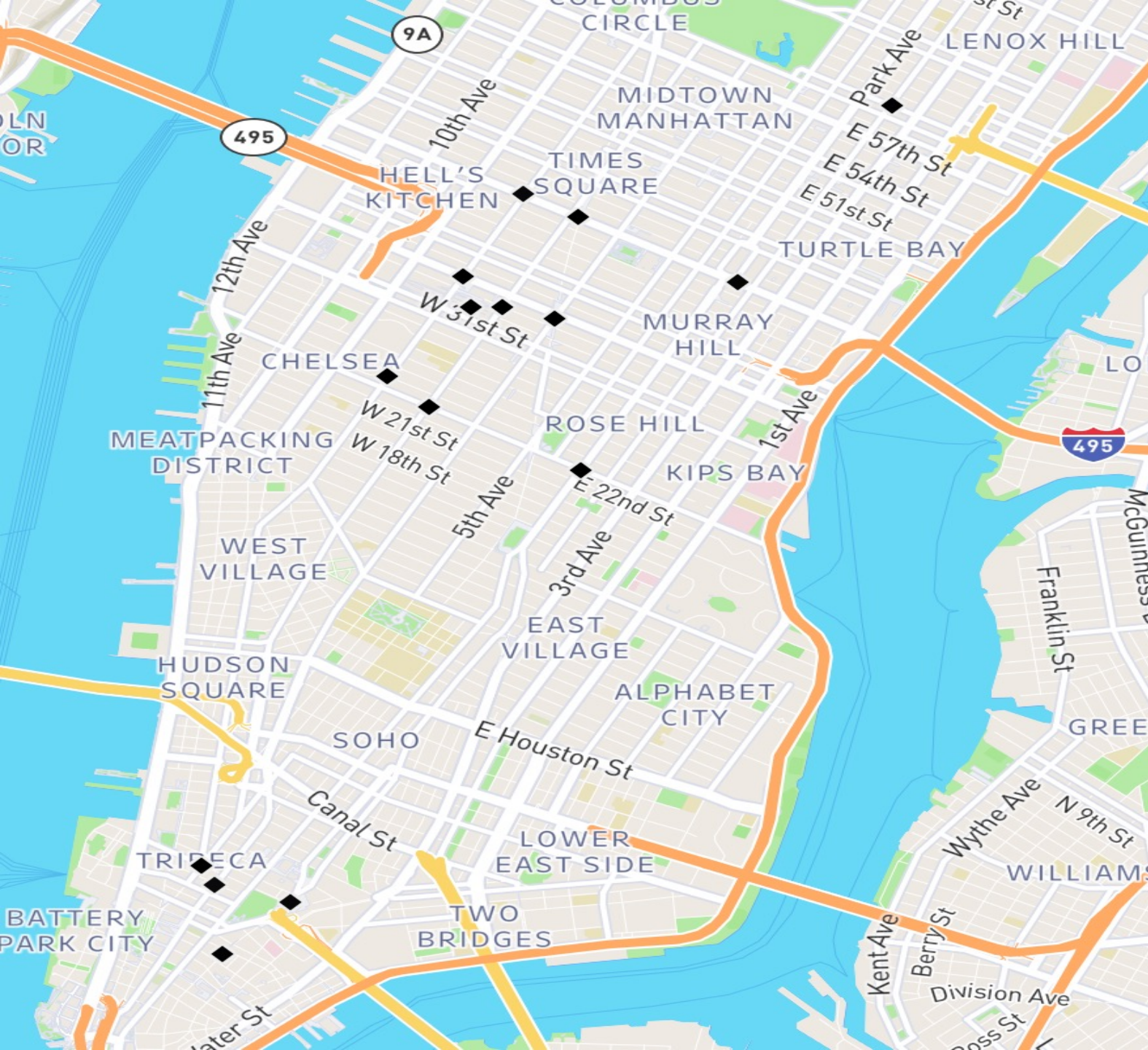


**34 ST-PENN STA**

**34 ST-PENN STA**

**34 ST-PENN STA**



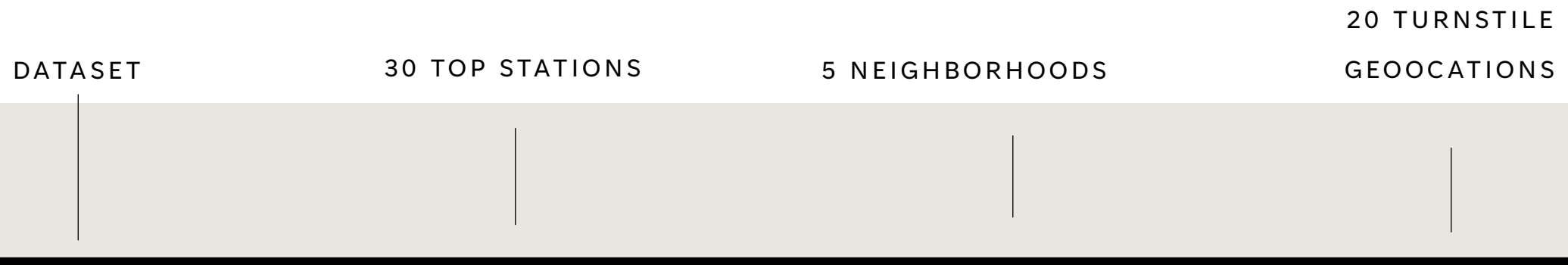


## SOLUTION

Further filter by sorting turnstile on traffic.



# SUMMARY





# FURTHER

- Bigger sample data
- donation history data based on neighborhoods
- Education dataset
- Data relating to interest in social causes

A series of white, thin, overlapping geometric lines on a black background, forming a complex, abstract shape on the left side of the slide.

THANK YOU!