MTA EXPLORATORY DATA ANALYSIS

Yingqing Qiu
Metis Data Science Bootcamp
September 2021

Introduction

- ## <u>Motivation</u>

  NYC DOT and MTA want to add/construct several new stations to satisfy the growing post-pandemic daily public transit needs.

- ## <u>Objectives and goals</u>

  We plan to send out several groups of surveyors to collect information from the passengers entering the subway stations. So that it is necessary to come out a reasonable and efficient survey plan to save time and labor. We want to find out the busiest station and time blocks for the preliminary survey plan.
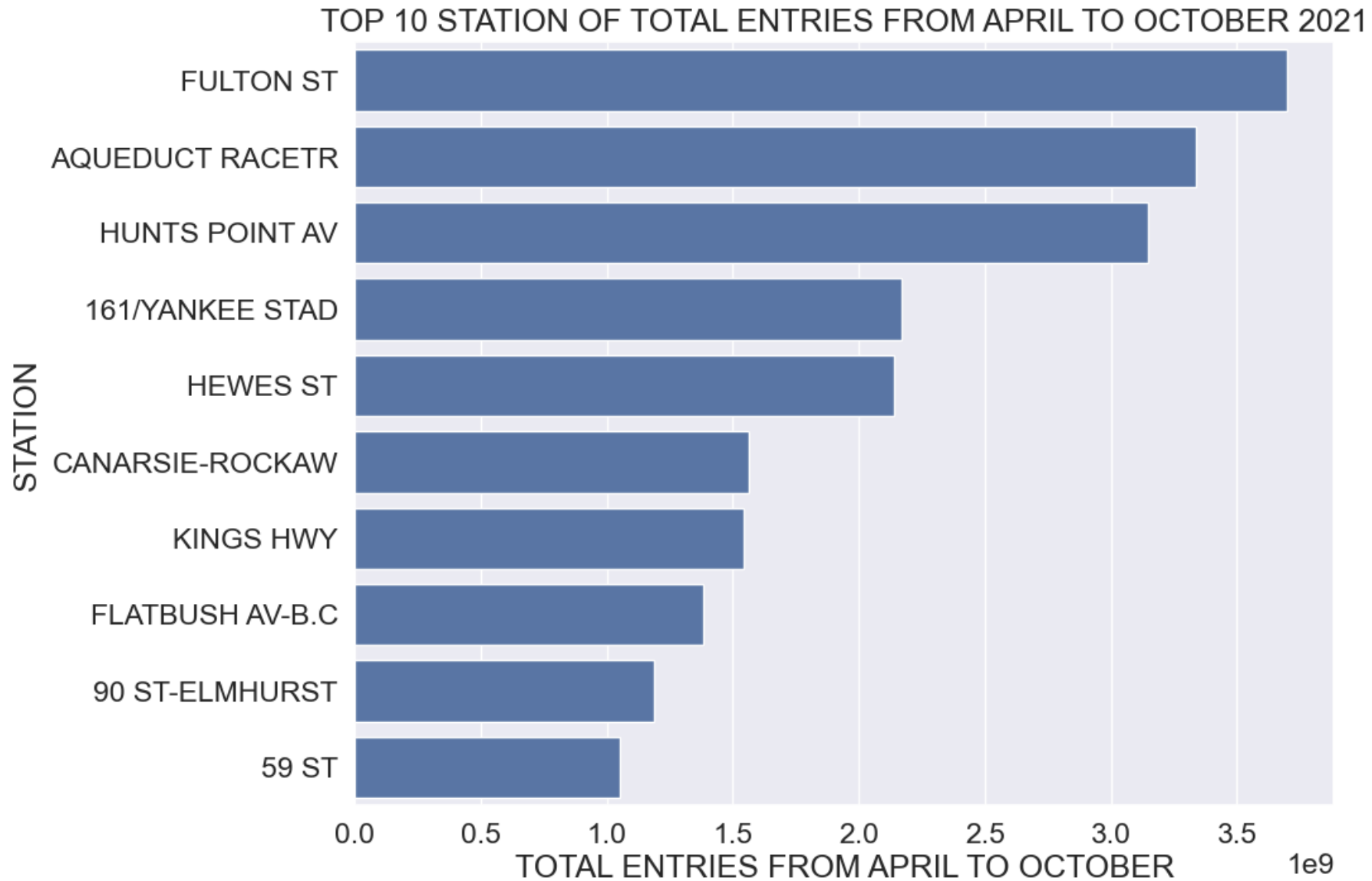
- ## <u>Final results</u>

  After the data analysis, we selected several stations for the survey plan. Additional metrics that can be discussed in the future are considered and presented.
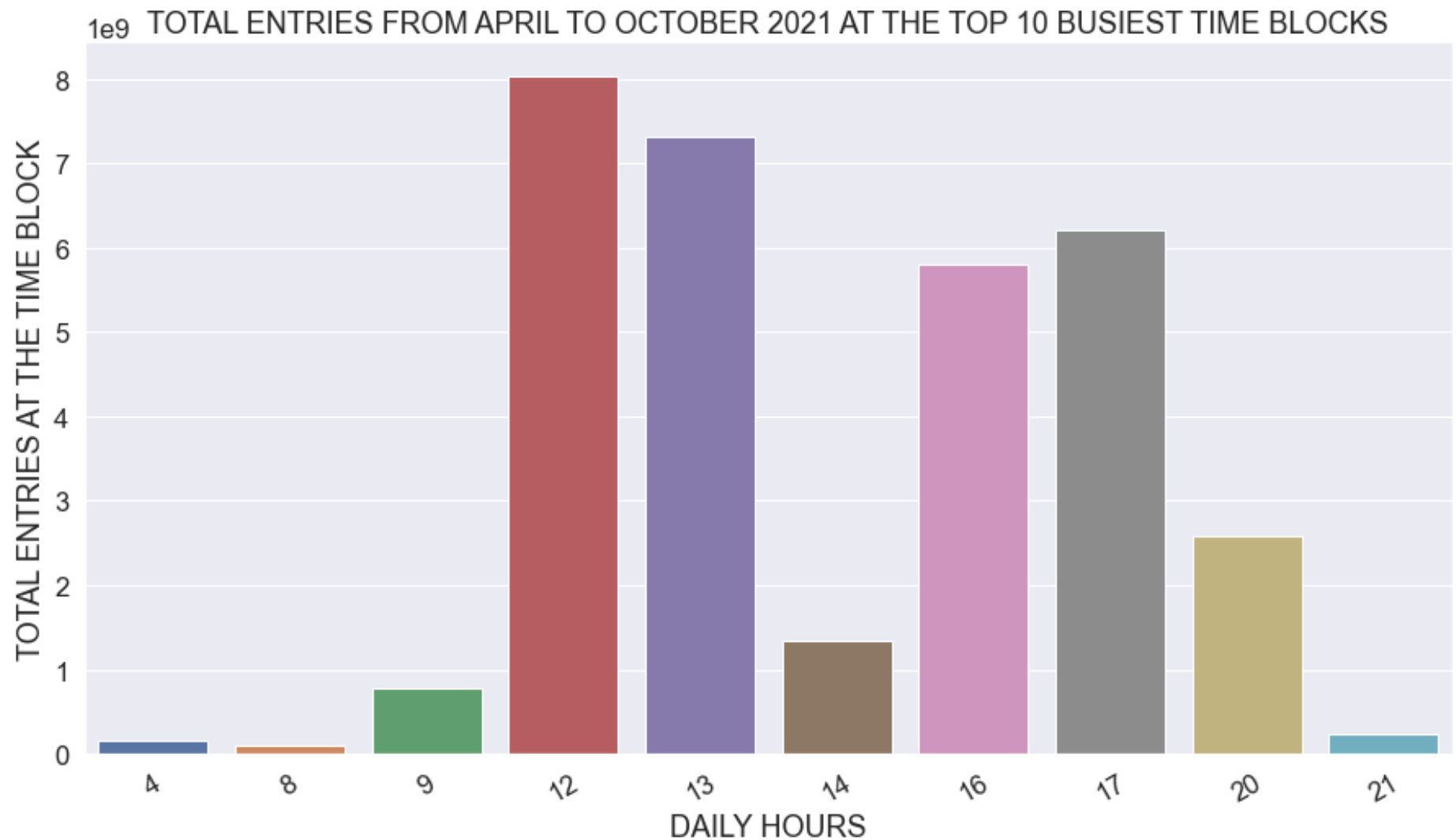
Methodology

- **Pandas**
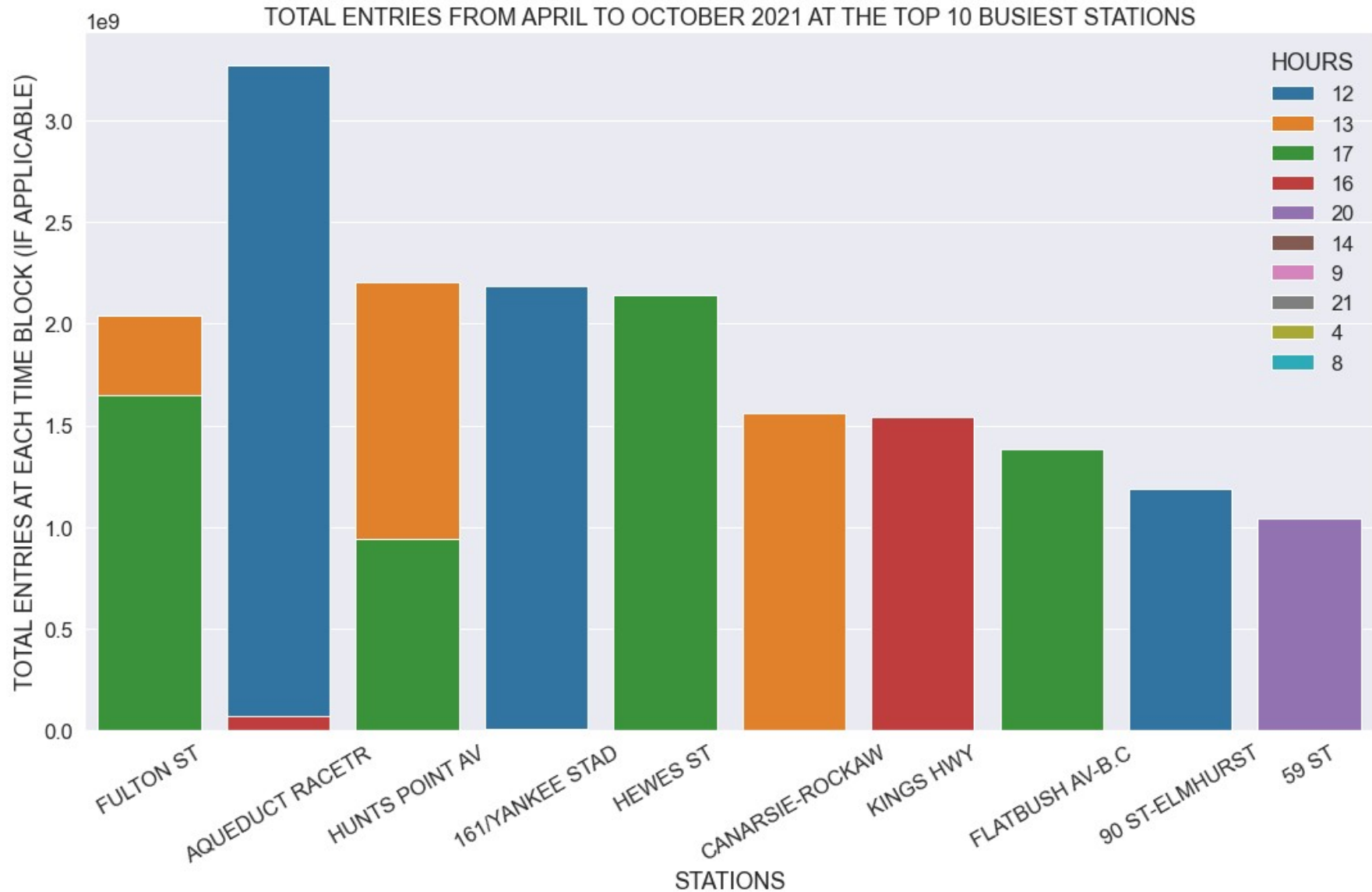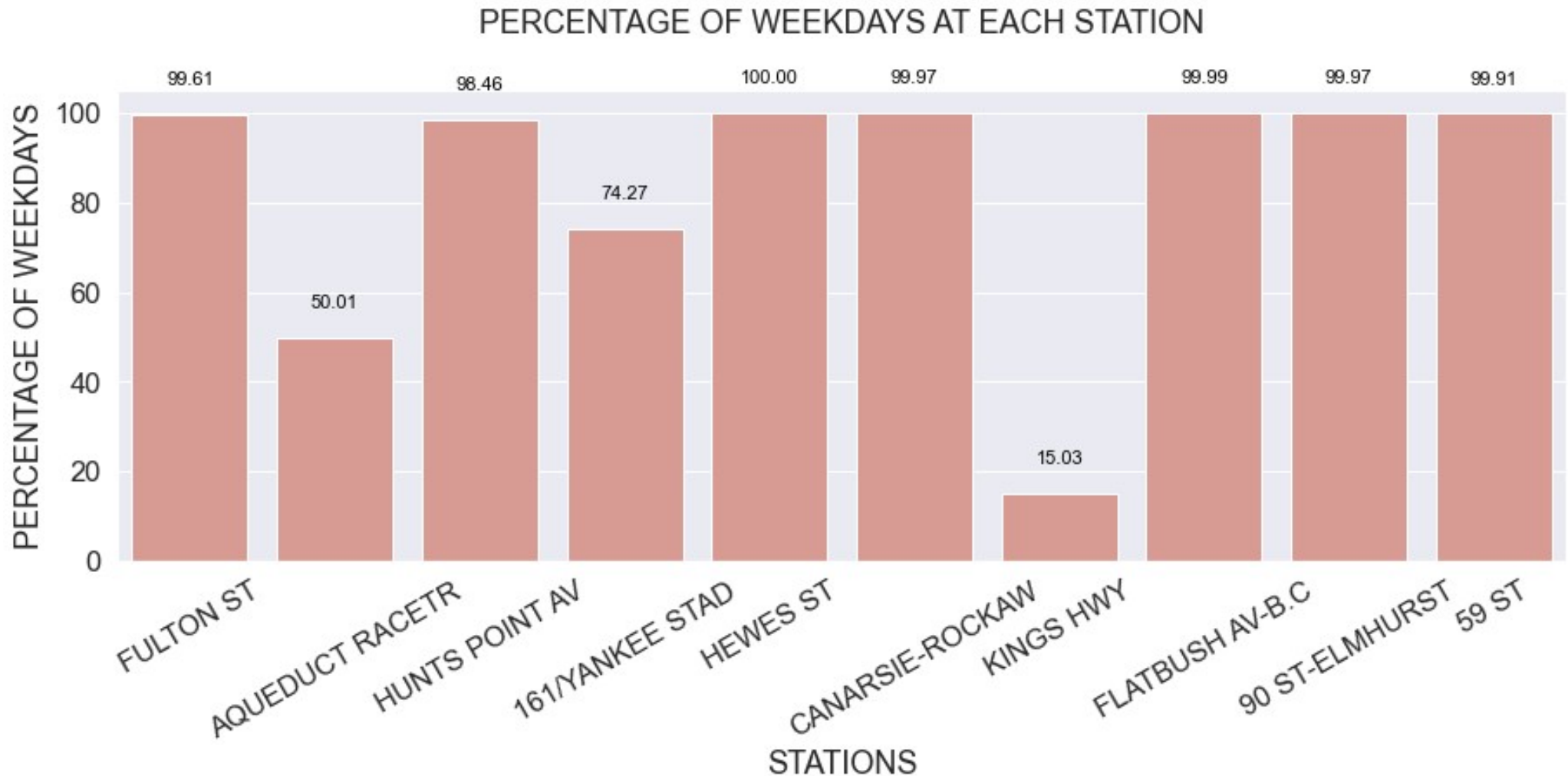- **Numpy**
- **Matplotlib**
- **Seaborn**

Results

TOP 10 STATION OF TOTAL ENTRIES FROM APRIL TO OCTOBER 2021

4

Results

Results

TOTAL ENTRIES FROM APRIL TO OCTOBER 2021 AT THE TOP 10 BUSIEST STATIONS

1e9

y-axis: TOTAL ENTRIES AT EACH TIME BLOCK (IF APPLICABLE)

x-axis: STATIONS — FULTON ST, AQUEDUCT RACETR, HUNTS POINT AV, 161/YANKEE STAD, HEWES ST, CANARSIE-ROCKAW, KINGS HWY, FLATBUSH AV-B.C, 90 ST-ELMHURST, 59 ST

HOURS
- 12
- 13
- 17
- 16
- 20
- 14
- 9
- 21
- 4
- 8

6

Results

PERCENTAGE OF WEEKDAYS AT EACH STATION

Results
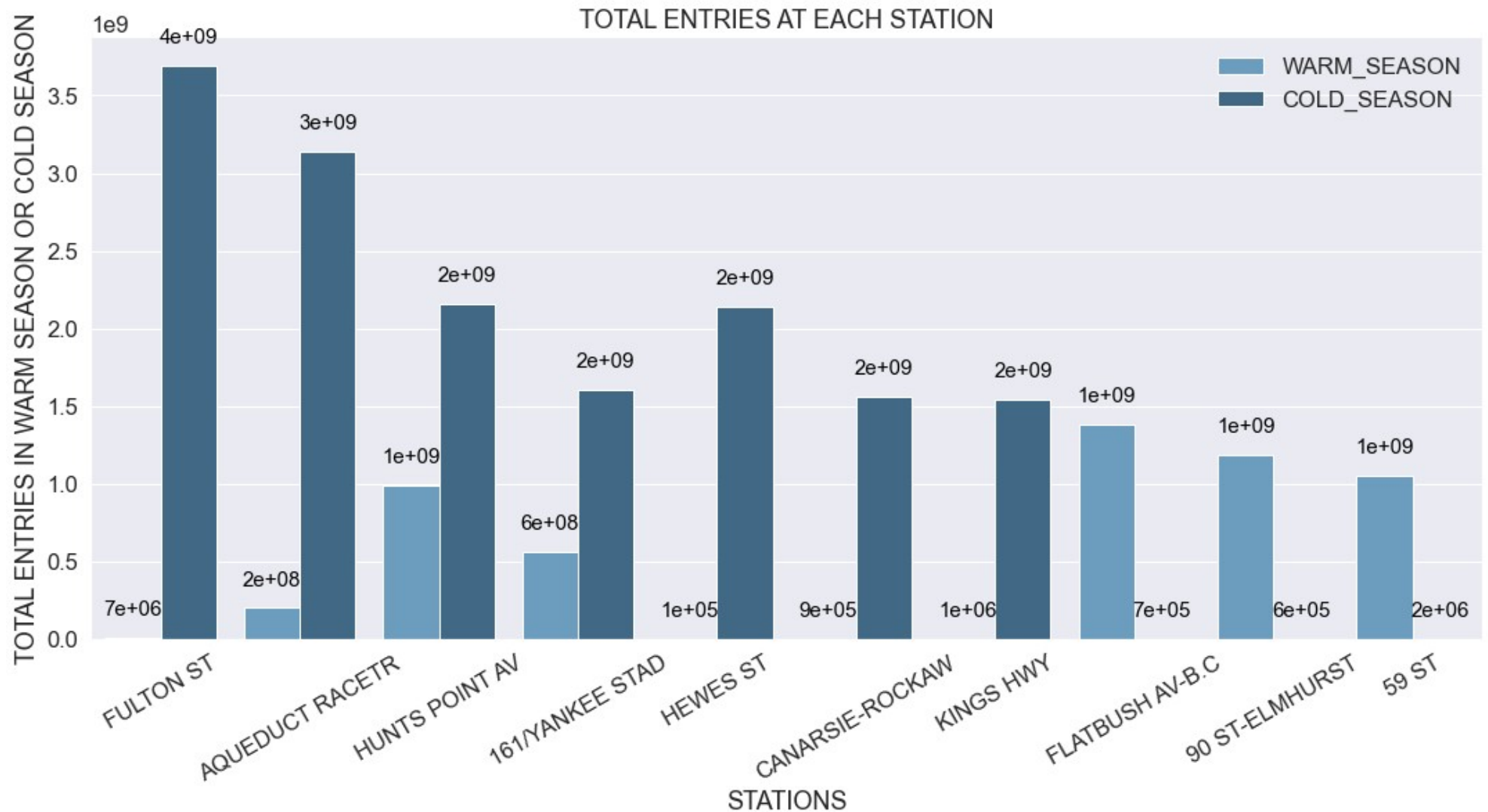
Conclusions

PRELIMINARY SURVEY PLAN – TIME AND LOCATION

Station	Time	Weekday/ Weekends	Season
FULTON ST	17:00:00	Weekday	Cold
AQUEDUCT RACETR	12:00:00	Both Weekday/Weekends	Cold
HUNTS POIN AV	13:00:00	Weekday	Cold/Warm
… …			

Future Work

- **Distance between stations**

- **Correlate the census characteristics near each station with the daily entries**

- **... ...**

Thank you!