

# EXPLORATORY DATA ANALYSIS

Final Project

## Analyzing MTA Turnstile Data

Marjan Rezvani

# Abstract

- In this project, I first get familiar with MTA data set and its related concepts, explore the features and instances, do the exploratory data analysis, then I look through some algorithms to do the data visualization. Getting information from graphs is far more time-saving than looking through text and numbers – that's why many decision makers would rather have information presented to them in graphical form, as opposed to a written or textual form.

# Introduction

- People in New York City use the subway as their primary transportation every day. There are several stations in the dataset which has a high number of commuter. Also, there are time series for each station. I try to analyze the raw data for traffic through the MTA subway system.

# motivation

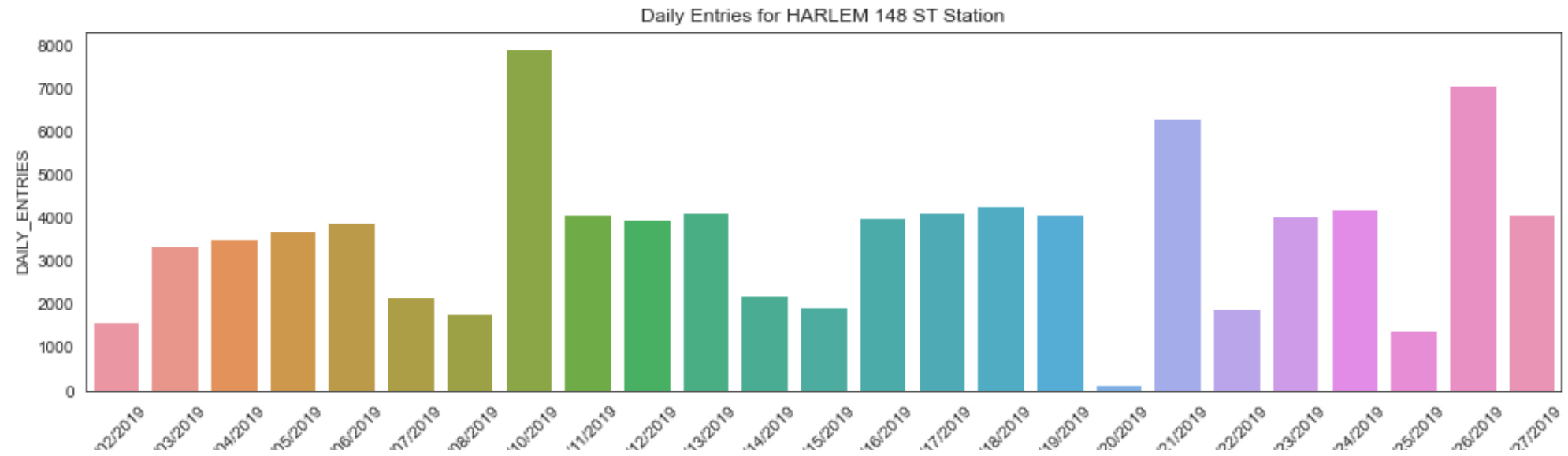
- I have been working to compare one month of each year for several years hoping to find a pattern and trend analysis. And perform some visualization to present and display information in a way that encourages appropriate interpretation, selection, and association. My goal is comparing the same month in several years to identify differences between them such as population of people who commute from a specific station.

# Data

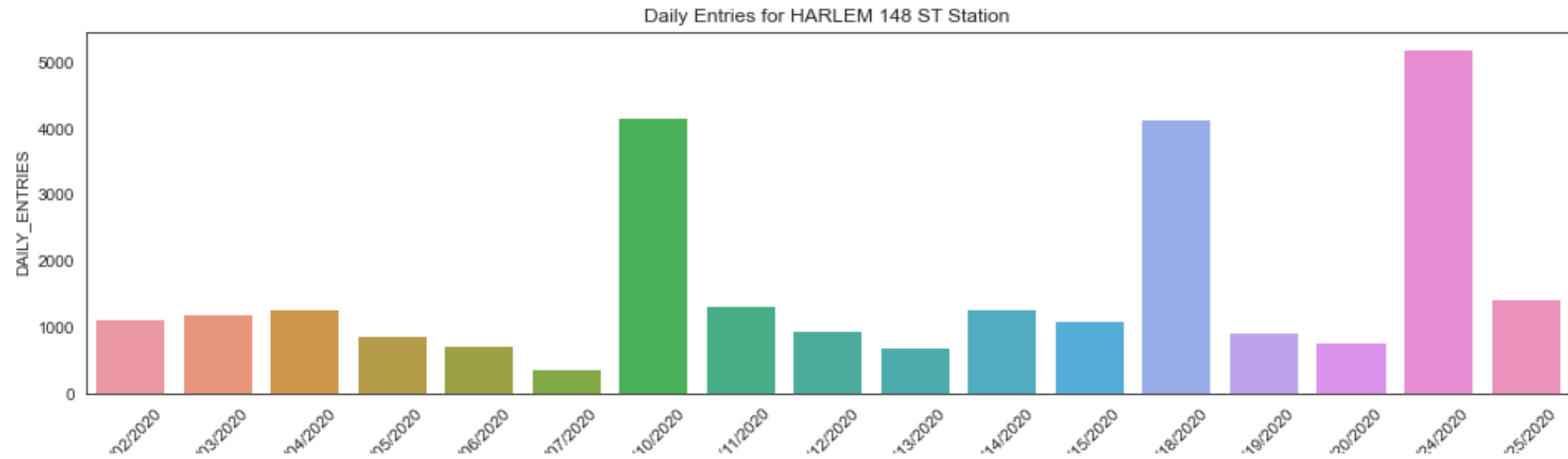
- The data is from the MTA website. I grabbed the data for whole month of September for three different years.
- It includes September 2019, 2020, and 2021.
- I worked on some highlight features including:
- STATION : Name assigned to the subway station
- DATE : Represents the date of the audit data.
- ENTRIES : the number of commuters

- Firstly, I clean up data then convert some data to datetime.
- Then I make some subsets of data to work with.
- And in continue I do more exploratory data analysis.
- Finally, I do some visual analytics on data.

# Daily Entries for HARLEM 148 ST Station in September 2019

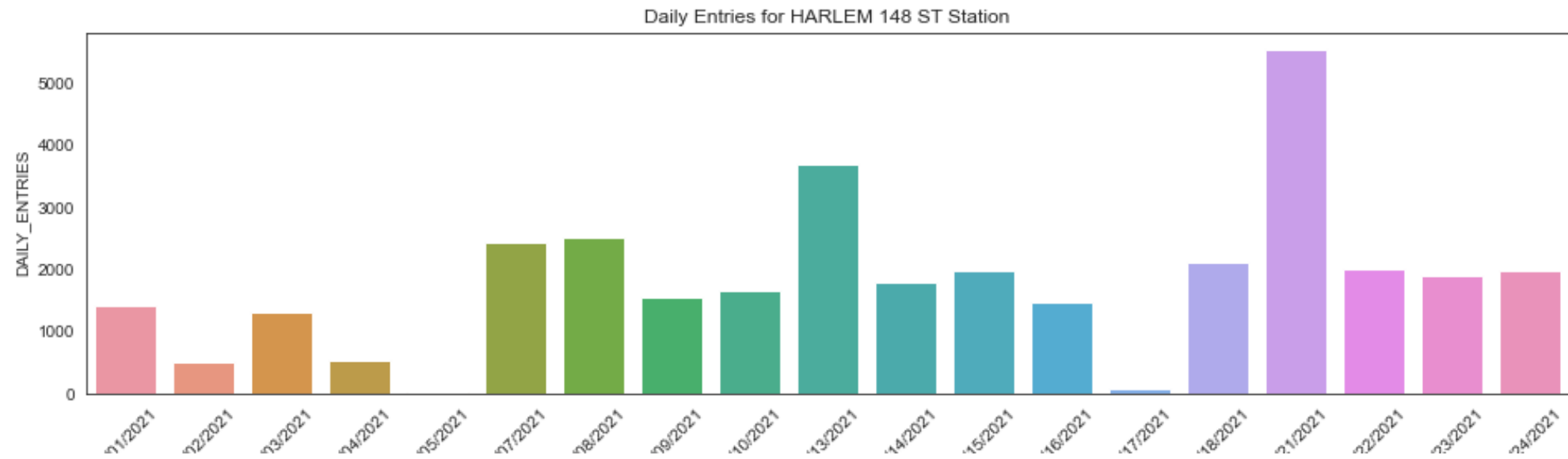


# Daily Entries for HARLEM 148 ST Station in September 2020





# Daily Entries for HARLEM 148 ST Station in September 2021



# result

- As it is obvious, in September 2019, the number of people who use the subway was much more than the other two years. The reason is quarantine period which caused decreasing of commuter.

## To do:

- If I had more time, I would work on different times of subway traffic to find the highest hour of traffic for each day and each station.

Thank you!