EXPLORATORY DATA ANALYSIS PROJECT

Analyzing MTA Turnstile Data

By: 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. 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.

Design

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 commuters. Also, there are time series for each station. I try to analyze the raw data for traffic through the MTA subway system and compare a single month of each year for several years hoping to find a pattern and trend analysis. Then perform some visualization to present and display information in a way that encourages appropriate interpretation, selection, and association.

Data

The dataset is from the MTA website and contains 820644 instances and 11 features, including S eptember 2019, September 2020 and September 2021. A few feature highlights include station, d ate and entries.

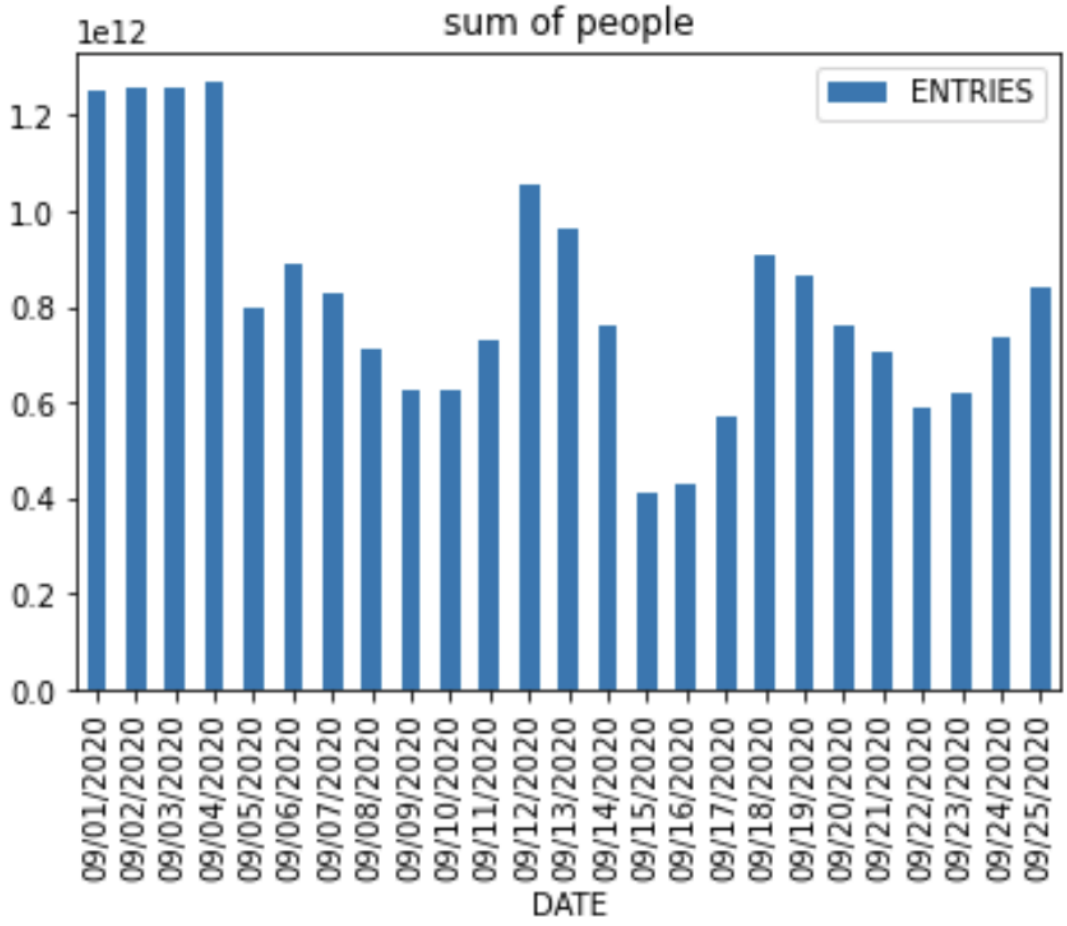
Feature Engineering

1. Cleaning up the data and removing duplicates.
2. Turning the dates and times into datetimes.
3. Making multiple subsets of data to work with.
4. Doing more exploratory data analysis.

5. Performing some visual analytics on data.
6. Selecting subsets of the total unique values

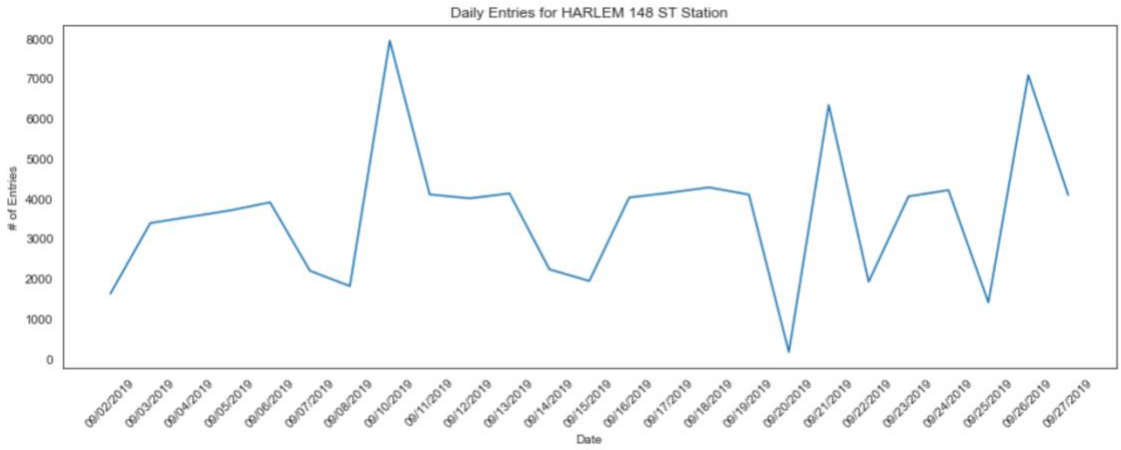
Visual Analytics

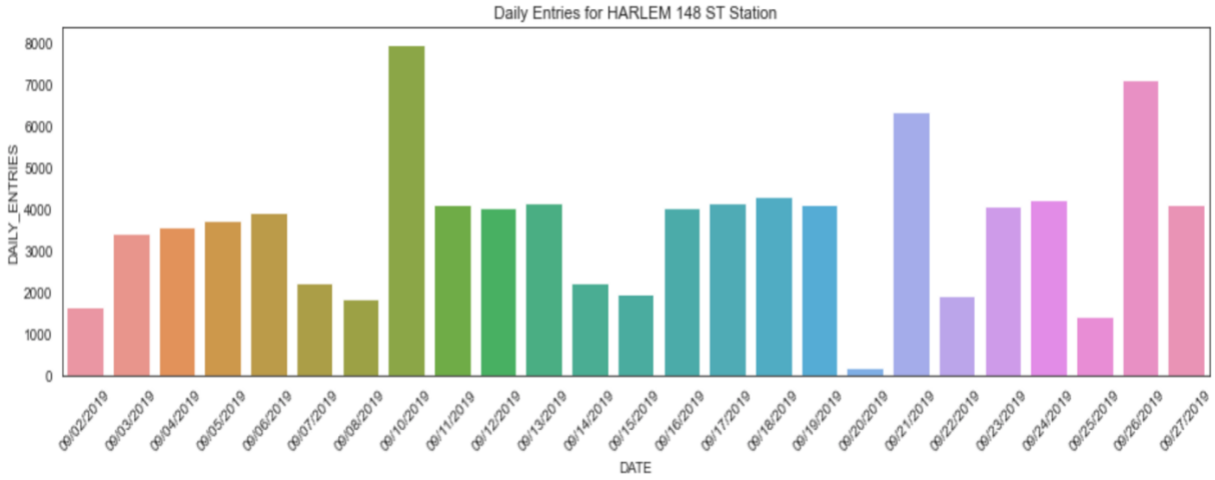
In the bottom graph we can see the sum of number of people who commute every day.

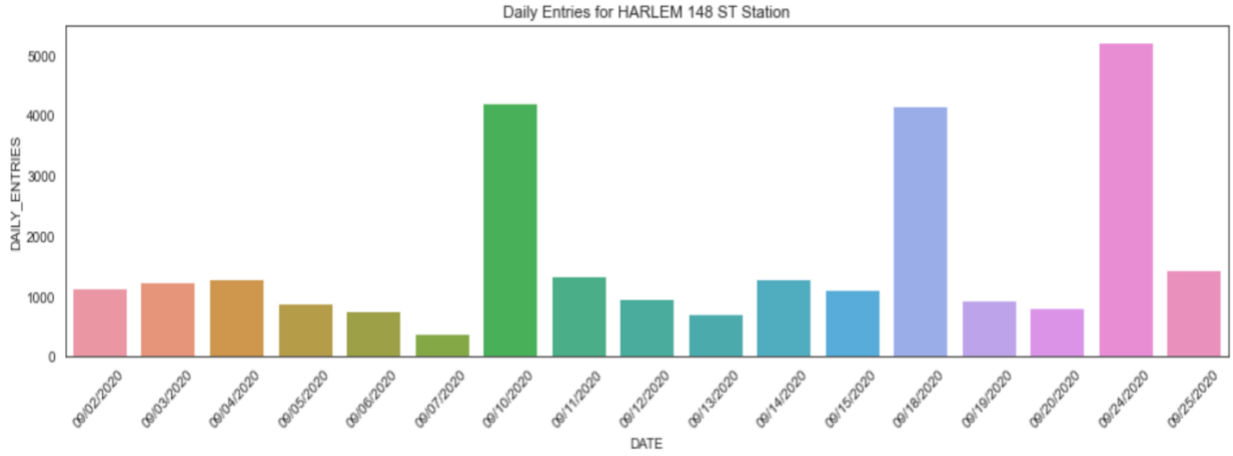


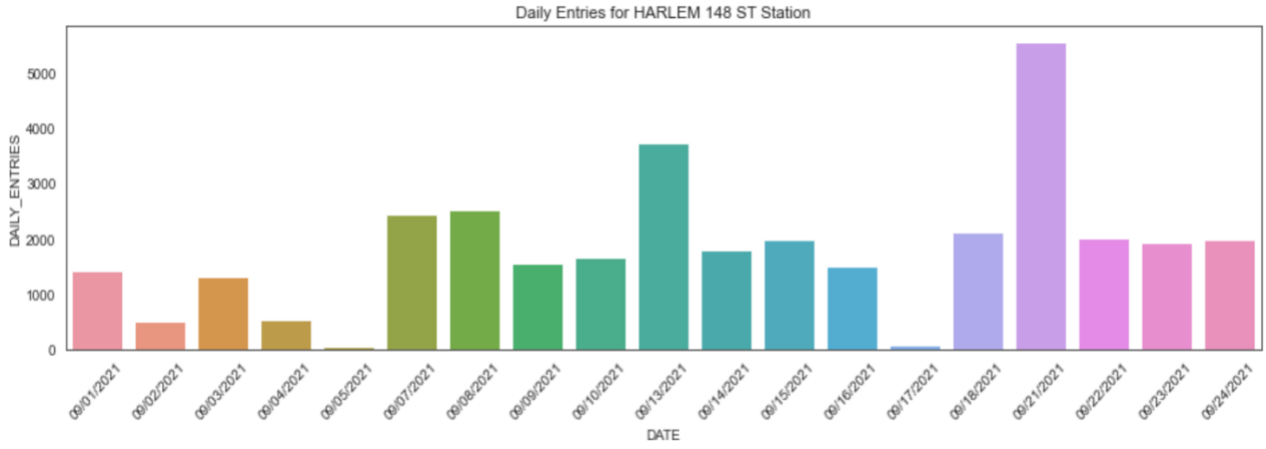
Graph's for people traffic for the Harlem Station over a single month period of September for 3 separate years

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 commuters.




Daily Entries for HARLEM 148 ST Station
DAILY_ENTRIES
0
1000
2000
3000
4000
5000
6000
7000
8000
09/02/2019
09/03/2019
09/04/2019
09/05/2019
09/06/2019
09/07/2019
09/08/2019
09/10/2019
09/11/2019
09/12/2019
09/13/2019
09/14/2019
09/15/2019
09/16/2019
09/17/2019
09/18/2019
09/19/2019
09/20/2019
09/21/2019
09/22/2019
09/23/2019
09/24/2019
09/25/2019
09/26/2019
09/27/2019
DATE


Daily Entries for HARLEM 148 ST Station
DAILY_ENTRIES
0
1000
2000
3000
4000
5000
09/02/2020
09/03/2020
09/04/2020
09/05/2020
09/06/2020
09/07/2020
09/10/2020
09/11/2020
09/12/2020
09/13/2020
09/14/2020
09/15/2020
09/18/2020
09/19/2020
09/20/2020
09/24/2020
09/25/2020
DATE


Daily Entries for HARLEM 148 ST Station
DAILY_ENTRIES
0
1000
2000
3000
4000
5000
09/01/2021
09/02/2021
09/03/2021
09/04/2021
09/05/2021
09/07/2021
09/08/2021
09/09/2021
09/10/2021
09/13/2021
09/14/2021
09/15/2021
09/16/2021
09/17/2021
09/18/2021
09/21/2021
09/22/2021
09/23/2021
09/24/2021
DATE