Analysis on MTA Ridership

Victor Fung, Shenghua You, Simon Ling

Big Data Management, Spring 2017

Table of Contents

Overview

An analysis the data gathered by the MTA from their turnstiles which would allow us to analyze any trends that occur throughout different parts of the year, such as by month, days of the week and weather.

Link to the repository: https://github.com/shenghuayou/MTA-Turnstile

Goals

- Examine the trends of ridership after price hikes
  - We will be examining whether the amount of people using the MTA increases or decreases after each price hike and comparing it with the ridership prior to the price hike.
- Examine the relationship between ridership and different types of weather
  - We will be examining how different type of weather affects ridership of the MTA.
- Examine the most popular station by the amount of traffic at a given station
  - We will be observing for the top ten popular stations during different seasons of the year.

Specifications

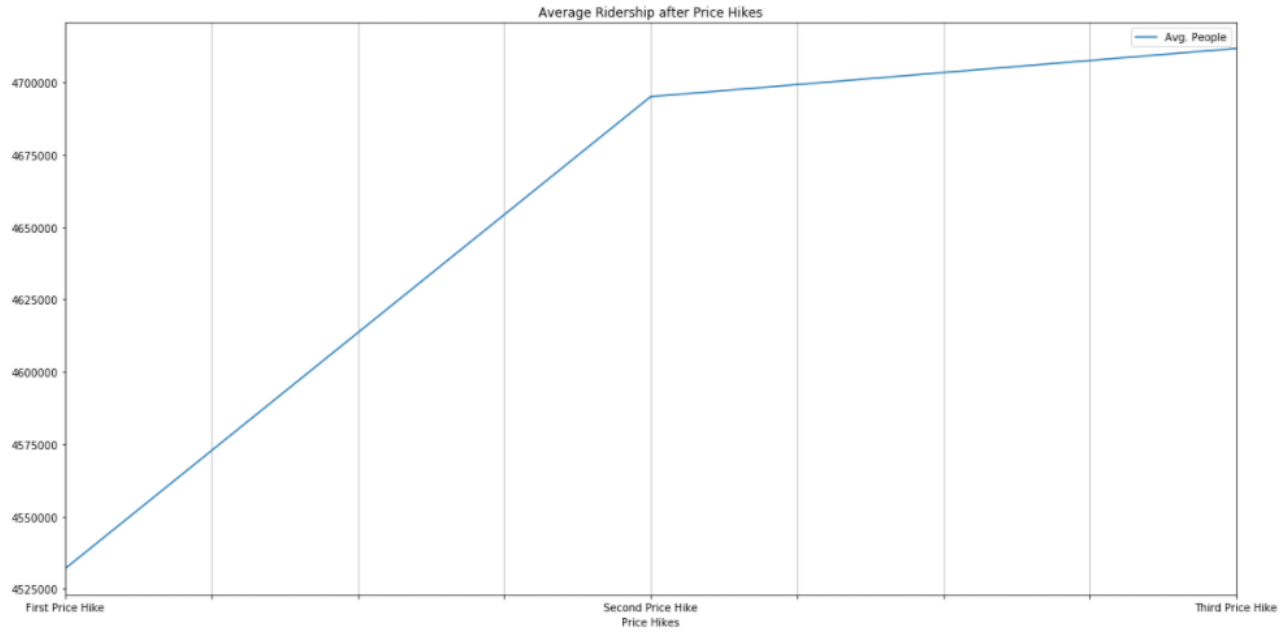
- Languages:
  - Python, Javascript
- Framework:
  - Spark
- Cluster:
  - Hadoop, HDFS (HUE)

Datasets

- MTA Turnstile Dataset (2010 – 2016)
- Wunderground Weather Data (2010 – 2016)

Results

Overall ridership after each price hike

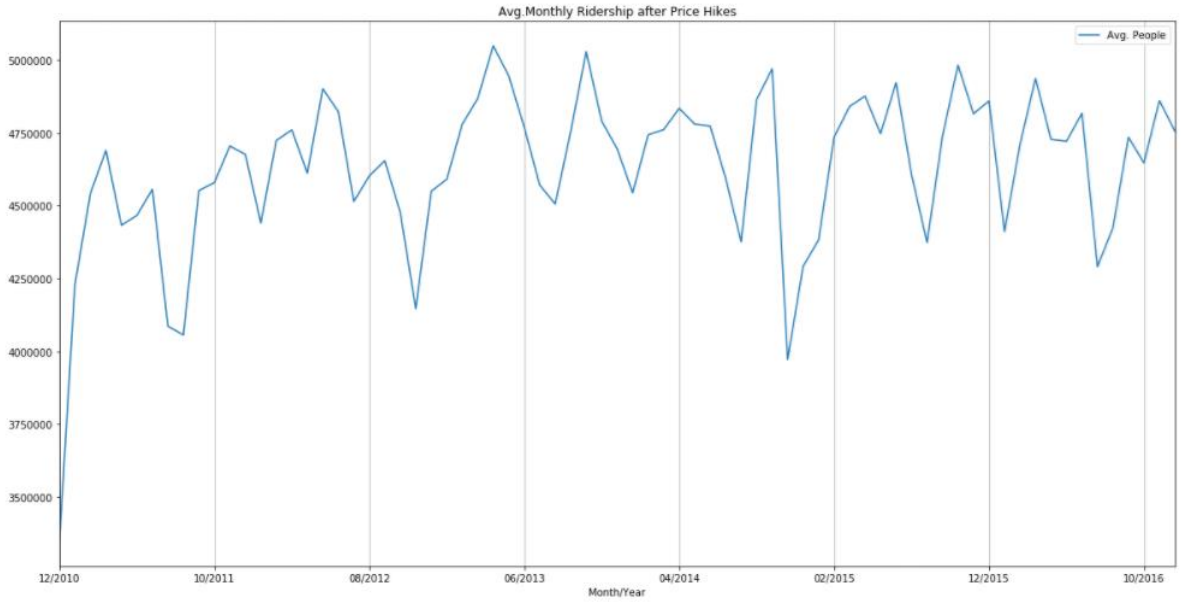


- *The base fare for the MTA during December 30th, 2010 – March 2nd, 2013 was $2.25*
- *The base fare for the MTA during March 3rd, 2010 – March 21st, 2015 was $2.50*
- *The base fare for the MTA during March 22nd, 2015 – 2016 is $2.75*
- *The average is calculated by taking the overall ridership from each period and dividing over the number of days each price hike period.*

From this chart, we can observe that the overall ridership over the years of 2010 – 2016 despite the increase in cost in the transit fare by $0.25 each time. This can be explained with the increase in the number of people living in New York City during the period of 2010 – 2015.[1]

[1] "Bronx, Queens and Brooklyn Fastest Growing Boroughs; NYC Population at All-Time High of 8,550,405." The Official Website of the City of New York. N.p., 24 Mar. 2016. Web. 11 May 2017.

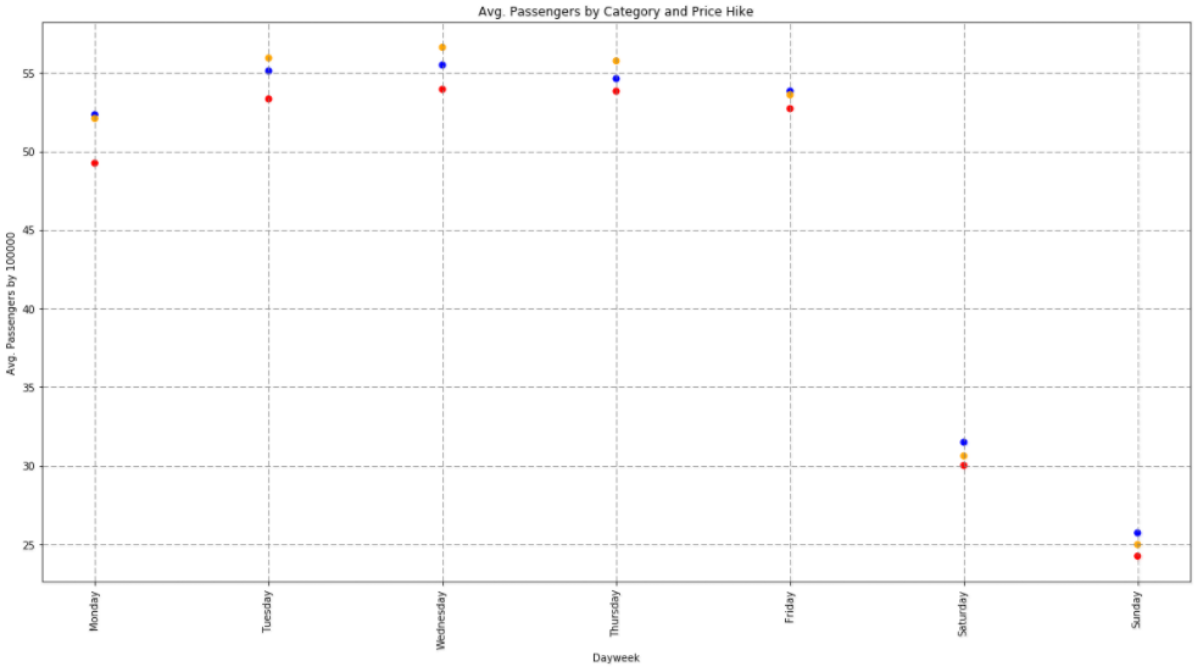
Monthly ridership after each price hike



- *The average monthly ridership is calculated by taking the overall ridership for each month and dividing it by the number of days in each month.*

From this chart, we can observe that the monthly average ridership over each price hike is stable, as the trends of when the amount of ridership increases and decreases on the same months of the years, specifically July and August of each year, which can be due factors such as students mainly K-12 are often on their summer break during these months of the year in New York City.

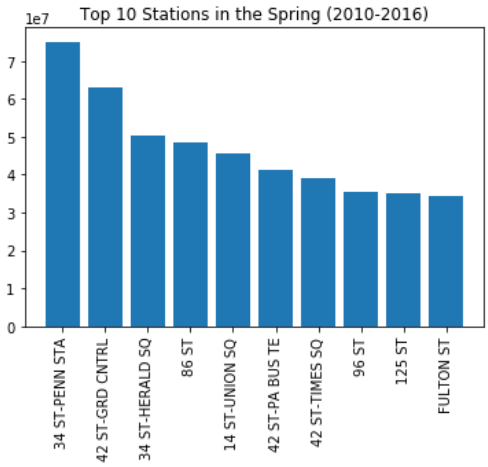
Ridership by each day of the week after each price hike

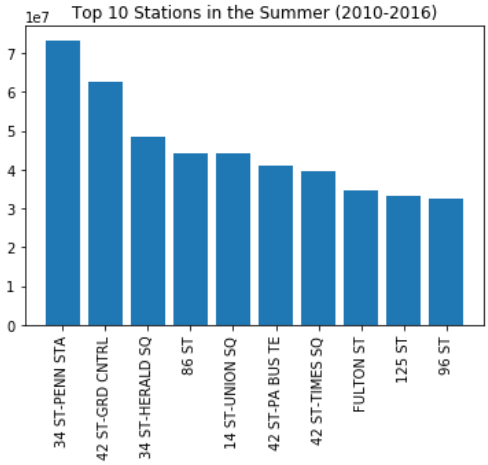


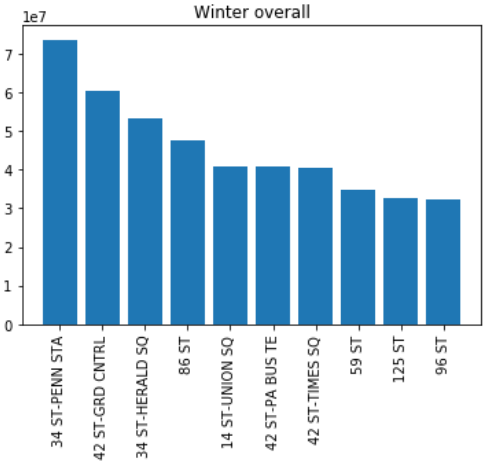
- *The first price hike is represented by the red dots.*
- *The second price hike is represented by the blue dots.*
- *The third price hike is represented by the orange dots.*

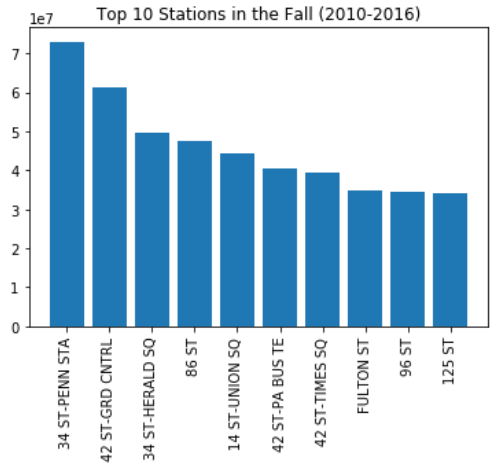
Observing the scatter plot above, it is clear that after the first increase by the MTA, the average ridership increased on every day of the week, shown by the blue and red dots. This trend changes after the next price hike, by observing the orange and blue dots, where the average ridership increased on the weekdays Monday through Thursday but began to decrease as the weekend approached, namely Friday, Saturday and Sunday.

Station Popularity during 2010 – 2016





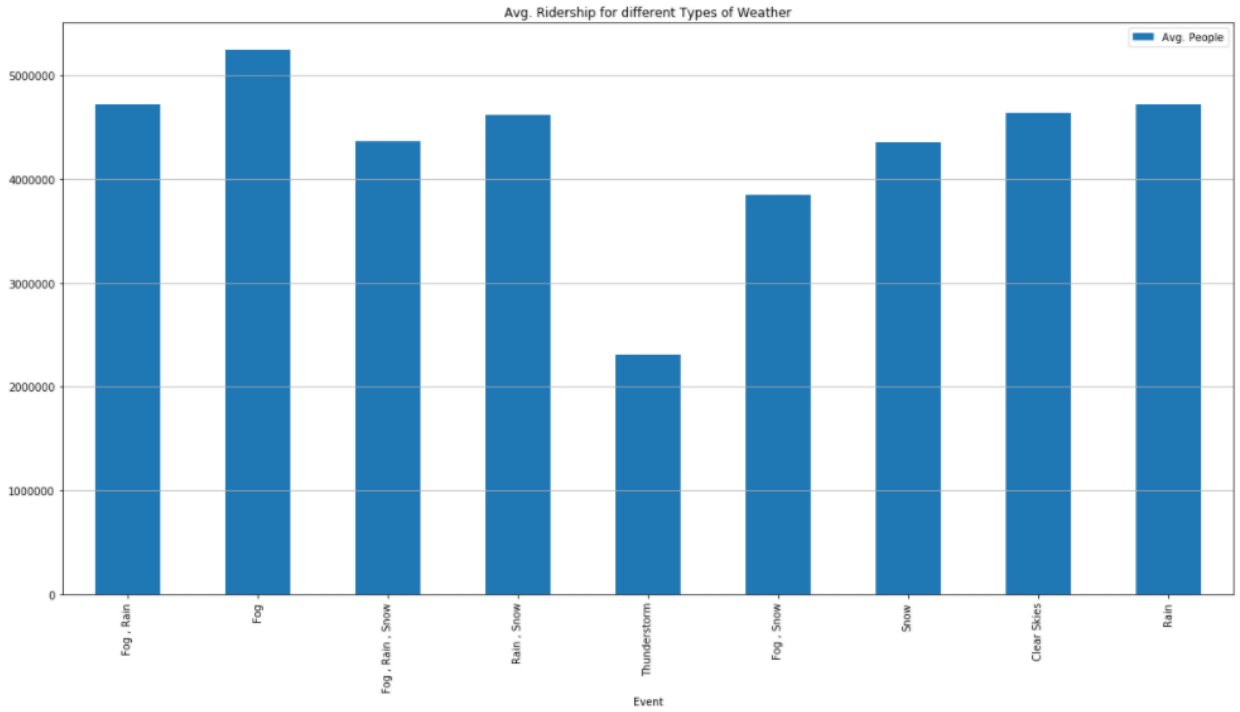




- *Station popularity ranking for each season with the leftmost column being the most popular to the rightmost column being the tenth most popular station.*

Observing the charts above, it is clear that the most popular stations tend to stay consistent, which can be explained as stations such as 34th Street – Penn Station and 42nd Street Grand Central are intersection areas where many subway systems, buses and other train systems intersect.

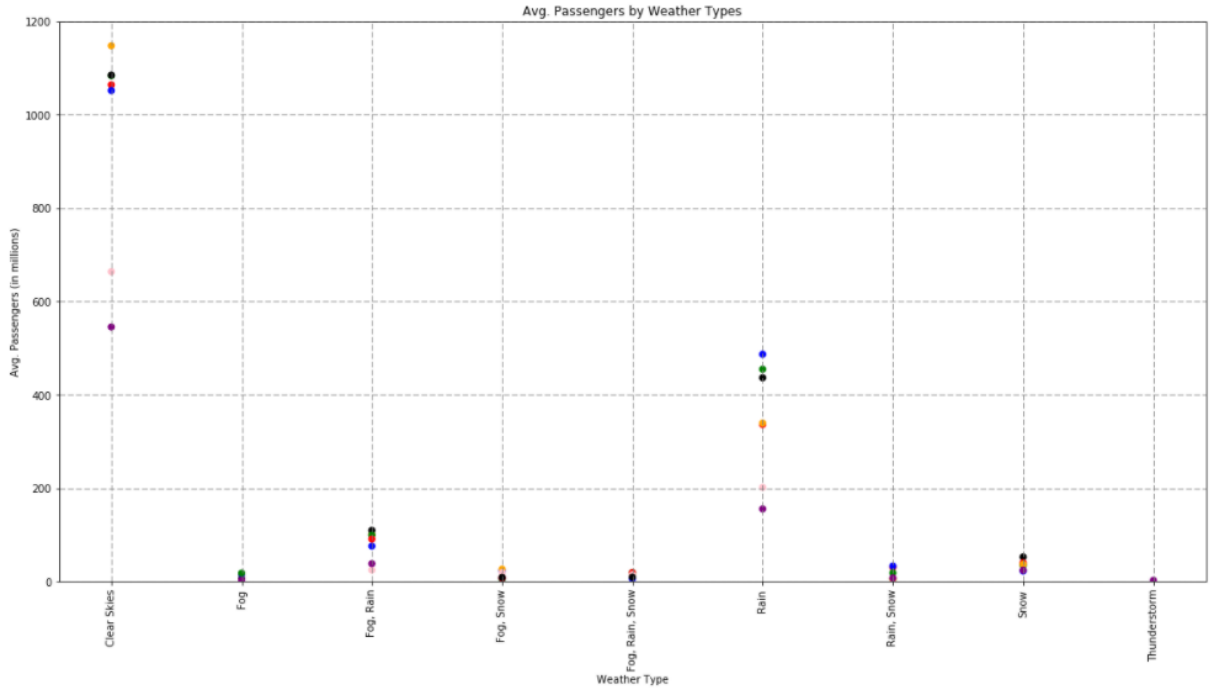
Overall ridership in different weather events



- *The average ridership is calculated by performing a summation of passengers on the type of weather on all days from 2010 – 2016, and dividing it over the amount of days for each type of weather.*

Observing the chart, we can see that the overall ridership remains stable, varying between four to five million on average for most types of weather. Outliers such as thunderstorms show a significant drop, but this can be due to the lack of the amount of days with thunderstorms.

Ridership by days of the week for different weather events



- *Monday is represented with the color red.*
- *Tuesday is represented with the color blue.*
- *Wednesday is represented with the color green.*
- *Thursday is represented with the color black.*
- *Friday is represented with the color orange.*
- *Saturday is represented with the color pink.*
- *Sunday is represented with the color purple.*

Observing different types of weather and days of the week, it is clear that on days with clear skies there is an increase in the amount of ridership. Examining closely on each weekday for different types of weather the average ridership is higher on weekdays compared to weekends.

Conclusion

The results from the popular stations indicate that the most popular stations are those that are located in hotspots of Manhattan, which attract tourists and also have the most intersections between subways systems along with other systems of transportations. The top seven spots never switch places regardless of season, whereas the last three spots for the top ten shift spots depending on the season, which can be can due to certain activities or events happening during those specific seasons of the year.

From the results of different weather events and the relationship to the average ridership, there is a consistency of ridership regardless of the weather event, but when examining the relationship by each day of the week, it is clear that on days other than clear skies there is a drop in the average ridership, especially on weekends.

Examining the results from the price hikes and ridership relationship, it is fair to say despite the increase of fare cost on the MTA subway system; there will be a trend for an increase in ridership as the subway system is one of the primary ways to get to work during the weekdays. There is also a new trend shown that during the weekends where there is a decrease in the ridership of the MTA, which allows us to conclude that with the increase in the amount of people living in New York City, the MTA can increase the base fare cost and still get an overall average increase of ridership but the trend will only reflect primarily on weekdays where residents are most likely using the subway system to get to their destination, such as school or work.