Impact of Weather on New York City Subway Ridership

By Makai Charles

**Abstract**

The research for the project was inspired by the ever changing climate. I sought to find if there was a relationship between the weather in New York City and the amount of people using the subway station at several locations. I used turnstile data from the MTA and historical weather data, both for the full year of 2021. With this data I found trends in commuter patterns throughout the year during well and poor weather conditions. I also looked for relationships between well and poor weather conditions and specific subway stations to see what stations lost the most riders.

**Design**

I looked at every turnstile at every station entering and leaving New York and combined them to get how many people left or entered each station everyday of 2021. Then I found days where it either rained or snowed and categorized those days as my “bad weather” days. Then I combined these datasets to see how many people traveled on either good or bad weather days. With this combined dataset I was able to use graphs to see patterns in ridership and determine which station lost the most riders, if any at all. I was further able to see what ridership patterns were like on bad weather days and good weather days.

**Data**

For the data provided by the MTA, I was able to see how many people entered and exited each turnstile in each subway station in New York. I was able to see this at 4 hour intervals of each day of the year. The weather data I received gave me information about each hour of everyday of 2021. That information included, the current temperature, amount of rain and snow, amount of snowfall on the ground, wind speed, cloud coverage, and the icon the website used to visually describe the weather at that hour

**Algorithms**

With these datasets I was able to aggregate data in multiple ways.

1. I found days of the years where it either rained or snowed, and if it did at least one of those I considered that to be a bad weather day.
2. I found how many people entered and exited each station for everyday of the year.
3. I found how many days of the year I was considering to have good or bad weather.
4. I found which stations had the biggest change in their ridership from good weather to bad weather.
5. Then I found the change in ridership from good weather to bad weather at every station by month track stations that had the most amount of people passing through them

**Models**

1. I used line charts to visualize the amount of people passing through stations during times of the year.
2. I used bar charts to see what stations had the biggest impact from ridership because of the weather
3. I used pairplots to track ridership patterns during only good weather days or bad weather days.
4. Again, I used a line chart to track the most busy stations throughout the year to see how weather during different months impacted ridership

**Tools**

I used NumPy and Pandas to aggregate and store information in DataFrames. I also used Matplotlib and Seaborn to visualize my data

**Communication**

Information is presented in slides and visualizations