**Assignment 1:**

**Spatial-Temporal Distribution Analysis of Bike-Sharing System**

Bike share systems are a form of public transportation that provides bicycles for short-term use to riders (Eren, 2020). These systems typically consist of a fleet of bicycles and docking stations located in public areas throughout a city. In this system, a person is able to rent a bicycle from one station ride it towards his destination and then return it to a different station close to his destination. This concept makes transport much easier and more accessible.

Citi Bike is a bike-sharing system in New York City that operates in partnership with the Department of Transportation. It was introduced in 2013 and has since grown to become the largest bike-share system in the United States, with over 12,000 bikes and 800 docking stations. Citi Bike provides riders with a convenient and eco-friendly way to travel short distances within the city, making it a popular choice for commuting and sightseeing. The system is available 24 hours a day, seven days a week, and offers various membership options to suit the needs of different users. Citi Bike has become an integral part of the transportation landscape in New York City, offering a sustainable and affordable option for short trips.

**Data**

Based on simple research, I found that the largest Bikeshare operator in the US is Citi-Bike System. So the data source for the various analyses on this assignment is going to be based on the citi-bike system’s trip data download from their website (<https://s3.amazonaws.com/tripdata>). This website does not require authentication before download, so it was easy to automate a python script to download and extract the trip data from the website.

Trip data for 2021 was considered for a number of reasons;

* It contained attributes which I found to be necessary for analysis e.g. ride-id, trip start and end date, latitude and longitude of all start and end trips.
* Also, the data for previous years showed a lot of inconsistencies in the way they were gathered. The data contained different columns almost every year. However, from February 2021, there was some consistency in the data columns. For 2022 data, one of the months data from their website was not downloadable possibly because it was corrupted. Therefore, to maintain very consistent data. 2021 data was downloaded. But even with 2021 trip data, data for January contained the inconsistent attributes, so the data for January, 2021 was separated and analyzed separately.

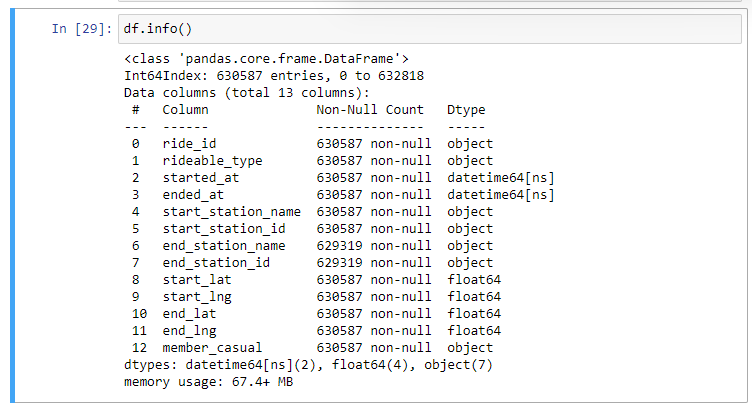
**Process of retrieving the data**

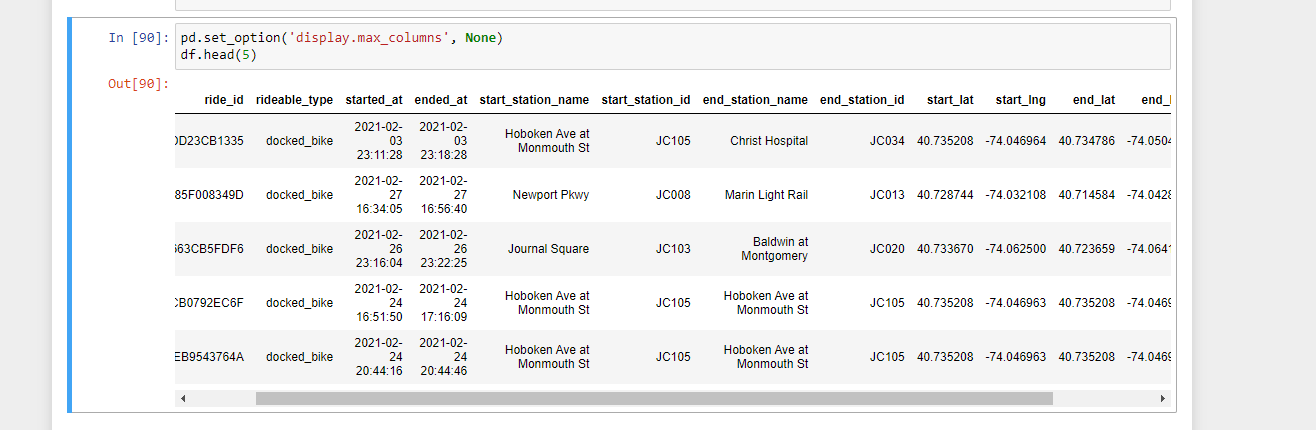
The trip data downloaded was to a large extent organized except for the inconsistent data gathered in January, 2021 which was separated from the rest of the data.

The other kinds of processing that was done included;

* Data type conversion: I had to convert the date columns from string formats to date-time formats.

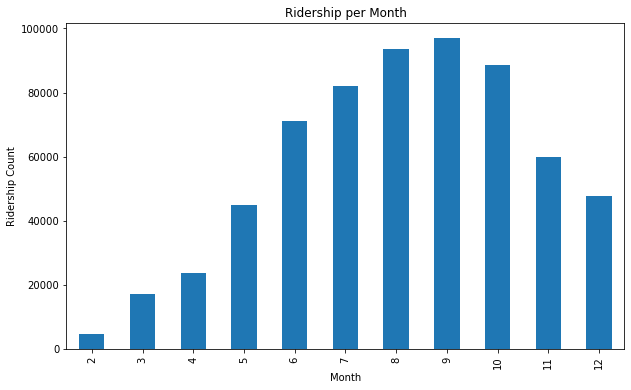
Removal of empty records: some of the months contained null records or null cells. Especially location data like latitude and longitude values which were missing were removed in order to enable the functions to receive consistent data. Therefore, they were removed using the pandas’ library **dropna** function.

With all the 630587 records, and 13 columns, the data was imported into python as a data frame using pandas’ library.

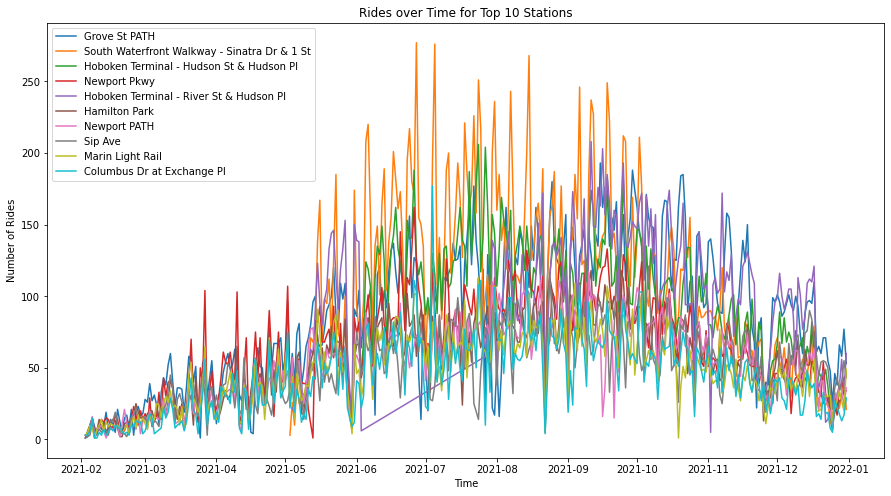


**Temporal Analyses**

1. **Ridership per month for 2021**

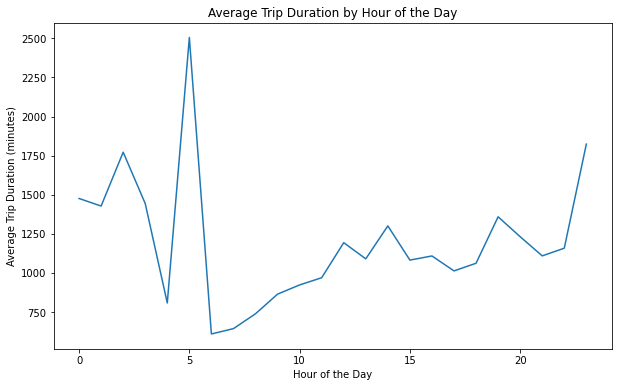


This shows that ridership started low at the beginning of the year, peaked in September at a ridership of about 9,9000 rides that month then started dropping as the year was ending. It is likely that the weather played a role in the fluctuation of ridership per month. In New York, September is usually considered to have warm and pleasant weather, which could have contributed to the high ridership in that month. On the other hand, the drop in ridership towards the end of the year could be due to the cold and harsh winter weather, leading to fewer people opting for bike rides.



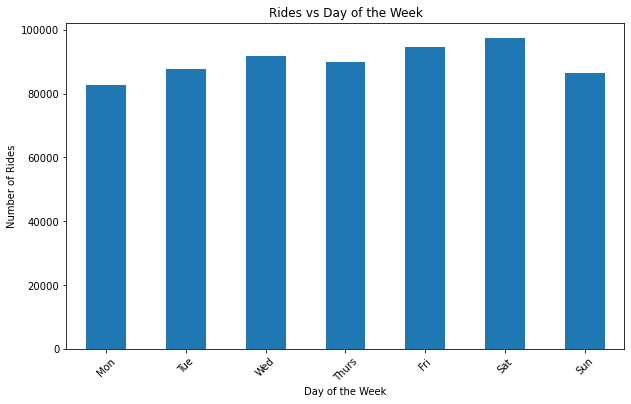
Generally, the late spring and summer seasons are considered the best for riding in New York which occur from May to September. During these seasons, temperatures are warm and pleasant, with little or no rain, which makes it ideal for outdoor activities such as cycling. Spring, in particular, has mild temperatures and less precipitation compared to other seasons, making it an ideal time for cycling. The graph above clearly follows the general pattern of the seasons.

1. **Average Trip Duration by Hour of the Day**



The figure above suggests that, the longest trips are taken in the morning between 4 and 6am. Although there are relatively little activities happening in this period because most users might be asleep by then, the longest activities are occurring in this period as well as parts of the late night. This suggest that the trips are taken for physical training purposes.

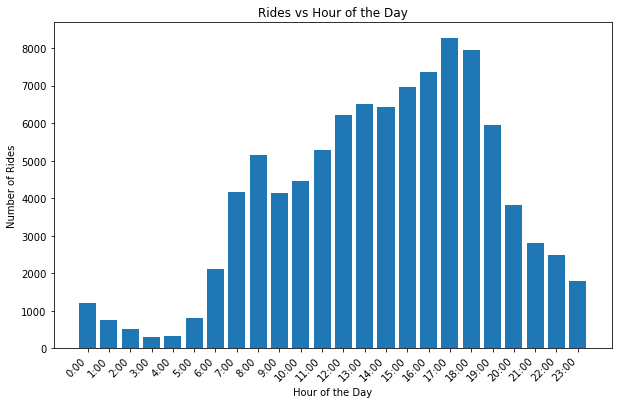
1. **Number of Rides Per Day of Week**



The bar graph above suggests that most rides occur towards the weekends peaking on Saturdays. This could be because most people are generally free from work and can therefore take casual rides during the weekends.

1. **Rides vs. Hour of the Day**

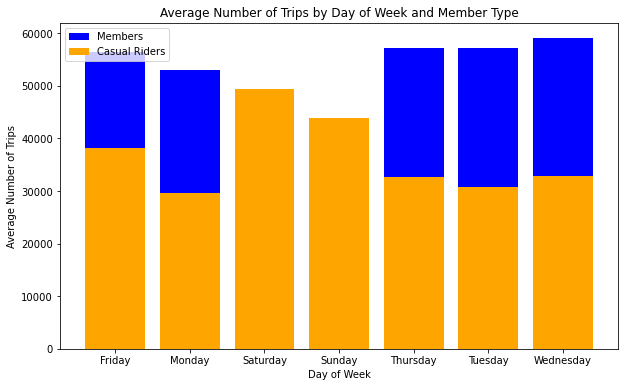
To understand the frequency of rides for all the 24hours in one day, the graph below was plotted. It shows that there are relatively little activities between 12am and 5am. However, after that period (which is mostly when people get out of their homes) there is a growing ridership peaking at around 6pm when people are likely to be commuting to and from work or engaging in other activities outside their homes. After this peak, the ridership starts to decline and reaches its lowest point at 11pm, likely due to the fact that people have completed their activities and are returning home.



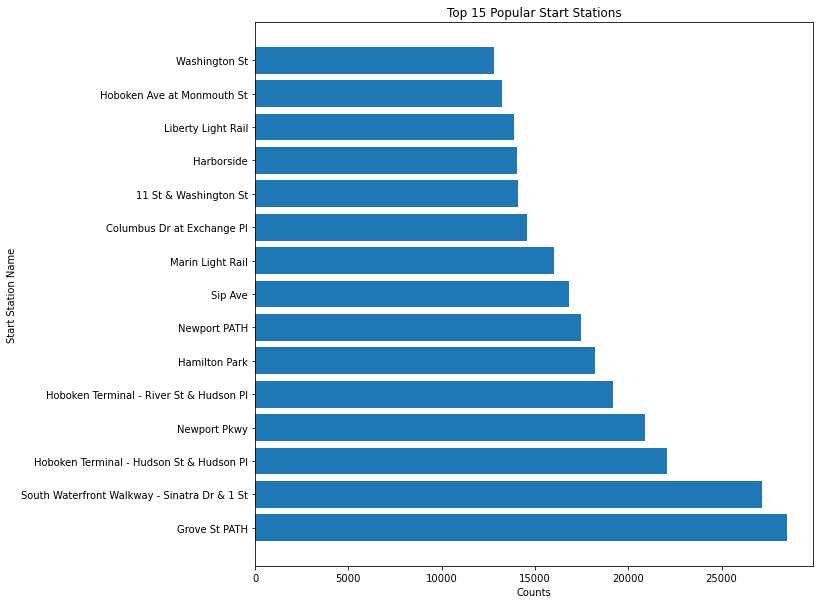
1. **Average Number of Trips by Day of Week and Member Types**

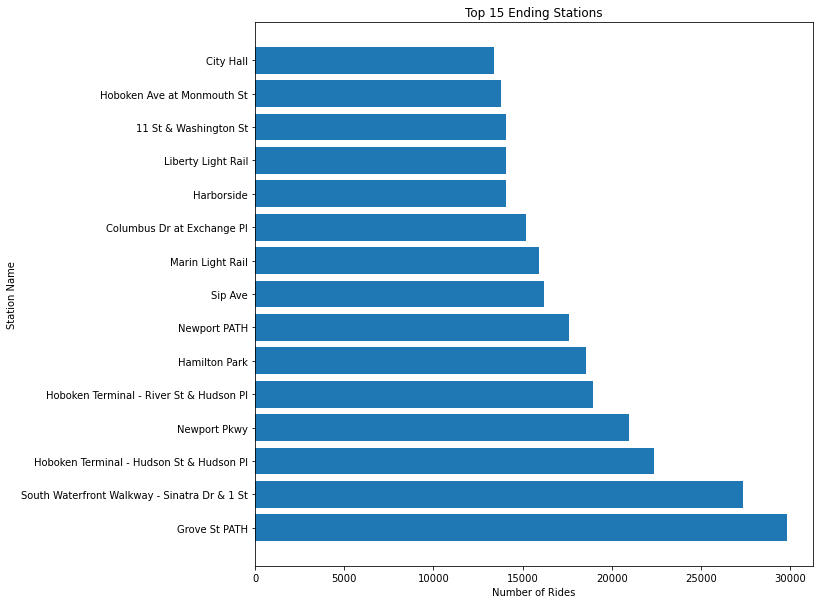
From the previous graph, it was noted that the bikeshare system was mostly patronized during the weekends. Therefore, it was hypothesized that, if most of the rides occur during the Saturday possibly because it’s a free day for most people, then most of the riders must be casual riders. Therefore, the plot below – Number of Trips by Day of week and Member Type was graphed.

The graph confirms that all the members who use the system during the weekends are mostly casual members. Most likely for fun or for physical training purposes.



**Spatial Analyses**

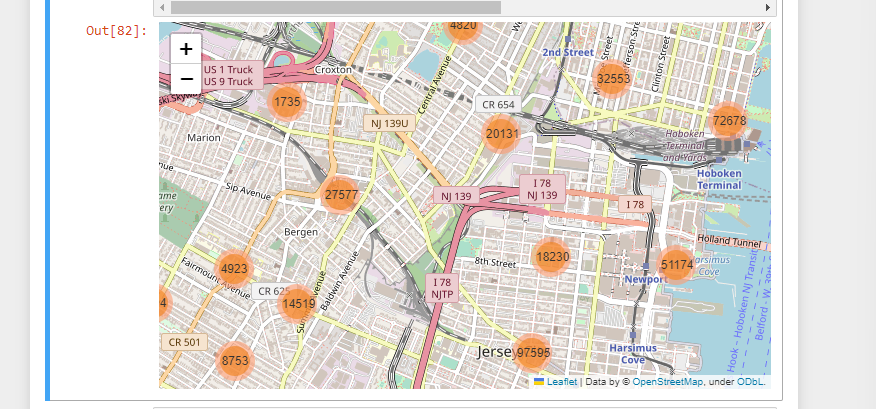
1. ****



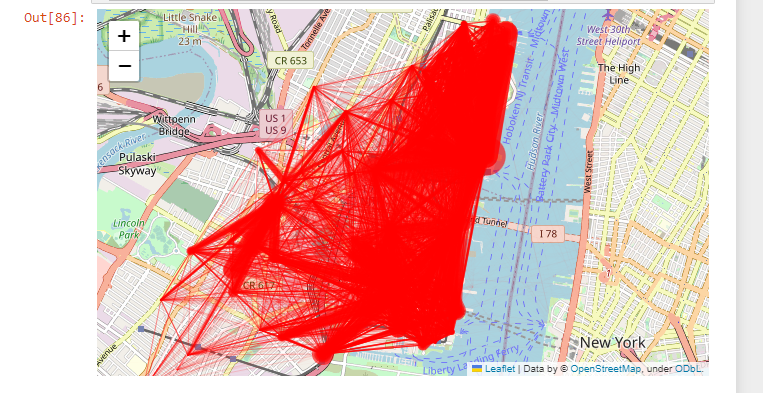
The spatial analysis included trying to answer the question; This analysis was focused on understanding the most popular start and end stations/routes among riders in New York City. Both graphs show the top 15 most frequently used start and end stations. The graph for popular start stations shows that Grove St. Path station was the most frequently used, with a high number of rides originating from that station. The graph for popular end stations shows that the second most popular station was South Waterfront Walkway to Dr & 1 St. These two stations appear to be the most popular among riders in New York City, which highlights the significance of these routes for cyclists. This information can be used to optimize resources and infrastructure in these areas, to ensure a safe and efficient riding experience for users. It can also help to identify areas for improvement, by identifying underutilized stations/routes such as Washington St. and City Hall St. that may benefit from increased investment or promotion.

1. **Areas with ride activities**

The use of a cluster map in this analysis allowed for the visualization of ride activity patterns in New York City. By utilizing the latitude and longitude values within the data, the location of each ride could be plotted on a map and grouped into clusters based on their proximity to one another. The clusters are represented by a color, and the size of the cluster reflects the number of unique rides that occurred in that area. This information provides valuable insight into where riders are frequently starting and ending their trips and can inform decisions related to bike station placement, infrastructure development, and policy making. For example, areas with high ride activity may warrant the expansion of bike lanes or the creation of bike-friendly streets to accommodate the growing demand for bike riding in the city. Additionally, a higher density of bike stations in these areas can be strategically placed to increase accessibility and convenience for riders.



The following figure gives a general indication of the origin and destinations of rides.



1. **Hotspot areas/routes for riding;**

A hotspot map is a visual representation of spatial data that highlights areas of high activity. In this case, the hotspot map was generated using the Folium library and the latitude and longitude information of the ride data. The map displays clusters of points, each representing the number of unique rides that took place in that specific area. The size of the cluster and its color intensity indicate the density of ride activity in that area. The hotspot map provides an effective way of identifying areas with high levels of bike ride activity and can be used to help understand patterns of bike use in a particular region. This information can be valuable for policymakers, bike rental companies, and other stakeholders who are interested in improving the bike-riding experience for users.



**Other Data That May Help to Give a Better Interpretation and Prediction of Trips**

To predict bikeshare trip activities, the following data and their sources could prove very relevant:

1. Weather data such as temperature, precipitation, and wind speed: This can be obtained from weather agencies such as National Weather Service (NWS) or National Oceanic and Atmospheric Administration (NOAA).
2. Demographic data such as population density and income levels: This can be obtained from the US Census Bureau or other demographic research organizations, also, data can be downloaded from worldpop.com
3. Economic data: This can be obtained from the Bureau of Labor Statistics or other economic research organizations.
4. Land use and Land Cover Information: Which can be either be downloaded from the NASA website or determined using GIS and Remote Sensing Technologies.
5. Transportation data: This can be obtained from transportation agencies such as the Federal Highway Administration or the Department of Transportation

With ample time and research, these data sources could be gathered, converted into a useable format e.g. dataframes and then used as features to help train a machine learning model capable of predicting the ridership in New York.

I think GIS software such as QGIS (open source) or ArcGIS (proprietary) could also be used to perform a lot of analysis and also to map out the result of these analyses. I use QGIS often when dealing with spatial analysis since it’s an easy tool to use.

**Supporting Research Papers**

I have found a number of research papers that have explored the relationship between various factors and bikeshare usage using google scholar as listed below:

In a research Article titled; *Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto* authored by El-Assi et al. (2017); The paper explores the relationship between the built environment and weather conditions and their effect on bike-sharing demand in Toronto, Canada. At the end, the paper proposes that the demand for bike-sharing is significantly influenced by characteristics including land use, population density, and accessibility to public transportation, as well as by meteorological conditions, particularly temperature and precipitation. Generally, the study emphasized on the need of taking the built environment and weather into account when designing bike-sharing systems.

The paper "*Investigation on the effects of weather and calendar events on bike-sharing according to the trip patterns of bike rentals of stations*" by Kim (2018) aimed to investigate the impact of weather and calendar events on bike-sharing demand. The study analyzed the trip patterns of bike rentals of stations in South Korea to determine the effects of weather and calendar events on bike-sharing. The results showed that weather conditions, such as temperature and precipitation, have a significant impact on bike-sharing demand. Additionally, the study found that calendar events, such as holidays and weekends, also have a significant effect on bike-sharing demand. Therefore, findings of this study support the conclusion that weather and calendar events play a crucial role in determining bike-sharing demand.

Considering a paper authored by Wang & Lindsey, (2019) titled; “*Neighborhood socio-demographic characteristics and bike share member patterns of use*”. After adjusting for station accessibility, adjacent bike infrastructure, the built environment, gender, and age, the results show that participants use bike share more frequently in neighborhoods with a higher concentration of minorities and lower socioeconomic status (SES) neighborhoods. These passengers travel at more varying hours of the day and on more days of the week, and they use more origin-destination pairs of stations regularly.

**References**

El-Assi, W., Salah Mahmoud, M., & Nurul Habib, K. (2017). Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. *Transportation*, *44*(3), 589–613. https://doi.org/10.1007/S11116-015-9669-Z

Eren, E. (2020). A review on bike-sharing: The factors affecting bike-sharing demand. *Elsevier*. https://www.sciencedirect.com/science/article/pii/S2210670719312387

Kim, K. (2018). Investigation on the effects of weather and calendar events on bike-sharing according to the trip patterns of bike rentals of stations. *Journal of Transport Geography*, *66*, 309–320. https://doi.org/10.1016/J.JTRANGEO.2018.01.001

Wang, J., & Lindsey, G. (2019). Neighborhood socio-demographic characteristics and bike share member patterns of use. *Journal of Transport Geography*, *79*, 102475. https://doi.org/10.1016/J.JTRANGEO.2019.102475