*Ricia Daniels-Data Analysis <Module 20>*

Citibike NYC Rider and Station Analysis: 2022 vs. 2023.

In Tableau, I imported my data sets and joined them on common fields such as ‘station name’ and ‘longitude’ and ‘latitude’.

Citibike Ridership: May 2022, June 2022, May 2023, and June 2023 \*Has the type of rider changed? \*Have the top station locations changed? \*Have the trip totals gone up or down?

The following visualizations will illustrate citibike ridership data from May and June of 2022 and 2023.

***Data Cleaning***

I collected the data from Citi Bike Data. I used Citi Bike trip history csv files from January, February and March of 2014. The files are very large and include trip and rider data from every station trip for the entire month. I used pandas in a jupyter notebook to clean the data. I used the concat function to combine all the csv files into one dataframe.

***Analysis***

After reviewing the visualizations, I concluded that the ridership had increased.

***Tableau Story***

Below is the final Tableau Story. You can also view it on the Tableau Public site- citibike Analysis: intro reader type, start stations, top 10 start, top 10 start map, end stations, top 10 end, top 10 end map, conclusion

***Resources***

CitiBike Data Sources:

Citi Bike Data may\_2022\_1 = pd.read\_csv("Data/202205-citibike-tripdata\_1.csv", dtype={'start\_station\_id': str, 'end\_station\_id': str}) may\_2022\_2 = pd.read\_csv("Data/202205-citibike-tripdata\_2.csv", dtype={'start\_station\_id': str, 'end\_station\_id': str}) may\_2022\_3 = pd.read\_csv("Data/202205-citibike-tripdata\_3.csv", dtype={'start\_station\_id': str, 'end\_station\_id': str}) jun\_2022\_1 = pd.read\_csv("Data/202206-citibike-tripdata\_1.csv", dtype={'start\_station\_id': str, 'end\_station\_id': str}) jun\_2022\_2 = pd.read\_csv("Data/202206-citibike-tripdata\_2.csv", dtype={'start\_station\_id': str, 'end\_station\_id': str}) jun\_2022\_3 = pd.read\_csv("Data/202206-citibike-tripdata\_3.csv", dtype={'start\_station\_id': str, 'end\_station\_id': str}) jun\_2022\_4 = pd.read\_csv("Data/202206-citibike-tripdata\_4.csv", dtype={'start\_station\_id': str, 'end\_station\_id': str}) may\_2023\_1 = pd.read\_csv("Data/202305-citibike-tripdata\_1.csv", dtype={'start\_station\_id': str, 'end\_station\_id': str}) may\_2023\_2 = pd.read\_csv("Data/202305-citibike-tripdata\_2.csv", dtype={'start\_station\_id': str, 'end\_station\_id': str}) may\_2023\_3 = pd.read\_csv("Data/202305-citibike-tripdata\_3.csv", dtype={'start\_station\_id': str, 'end\_station\_id': str}) may\_2023\_4 = pd.read\_csv("Data/202305-citibike-tripdata\_4.csv", dtype={'start\_station\_id': str, 'end\_station\_id': str}) jun\_2023\_1 = pd.read\_csv("Data/202306-citibike-tripdata\_1.csv", dtype={'start\_station\_id': str, 'end\_station\_id': str}) jun\_2023\_2 = pd.read\_csv("Data/202306-citibike-tripdata\_2.csv", dtype={'start\_station\_id': str, 'end\_station\_id': str}) jun\_2023\_3 = pd.read\_csv("Data/202306-citibike-tripdata\_3.csv", dtype={'start\_station\_id': str, 'end\_station\_id': str}) jun\_2023\_4 = pd.read\_csv("Data/202306-citibike-tripdata\_4.csv", dtype={'start\_station\_id': str, 'end\_station\_id': str})