As a city heavily troubled by congestion, lots of New Yorkers choose to ride a bike to commute or do a in-city travel. Although riding through a city is a great way to feel the city, more likely to be shocked by the awful ground transportation in this metropolitan, visitors accept riding bikes as an awesome option to travel from one location to another in New York. As a major bike sharing provider, Citi Bike is likely to be a perfect representative to depict the bike sharing environment in New York City. As a cycling lover, I would like to investigate in the topic of bike sharing in New York, though the subject itself is interesting.

Thanks to Citi Bike, it is easy to access Citi's bike sharing data on its website and thus allowing me to start such a project more easily.

There are downloadable monthly data since 2013 on Citi Bike official website. What makes things even greater is Citi offers the public access to real-time system data. For the sake of carrying out an as comprehensive analysis as possible, I would like to divide the whole project into 4 stages. In the first stage, I would like to investigate a recent month (2024 July) to get a basic understanding of the recent bike sharing environment. For the second stage, I would attempt to analyze it on the scale of a year by using the data in the year of 2023. When it comes to third stage, I would try my best to conduct the analysis with "all" the data dating back to 2013 when is the earliest available data to summarize the bike sharing developing trend yearly. In the same stage, a future development forecast should be presented. In the fourth stage, I am going to estimate the revenue of Citi Bike. As one kind of the output of this project, some Tableau dashboards are to be created for the readers to explore.

Considering the huge size of the datasets, it is far from plausible for me to use all these data to carry out the analysis. Statistically, it is good to sample these data with logically correct methods

and use sampled data which is of a much smaller size to do analysis. Such an analysis is expected to be appliable to the whole population including these in the near future. Bike sharing environment could change at any time. However, with those accessible real-time data, we are likely to have our already obtained result automatically adjusted if we have automated analyzing process. Thus, developing an automated analyzing process for the data of the same format is also an objective of the project, in addition to analysis itself.