20-Tableau data analysis

*Lingzi Xiaoli*

The dataset for this homework contains trip history data collected from first half of year 2019 and the final data visualization has been posted at Tableau public.

Data source:

<https://s3.amazonaws.com/tripdata/index.html>

Data cleaning:

csv merging was performed in python and the formatting for merged dataset was in excel.

Tableau visualization: <https://public.tableau.com/profile/daisy4454#!/vizhome/hm20_15960820702530/Citibikes?publish=yes>

Data analysis will be divided into four sections:

1. Bike demand: from Feb to Jun 2019, the bike demand at New York City kept increasing, with a significant increase from March to April with 40.03% increase. One of the potential reasons might because the weather change from winter to spring led to more outdoor activities including riding bikes. Compared two main customer sources for bike trips, subscribers tended to have more bike rides than regular customers.
2. The top five popular bike stations for starting and ending a journey are overlapped, including Grove St PATH, Hamilton Park, Sip Ave, Newport PATH and Harborside. Majority of the popular places are either park or PATH which are bike friendly. Those places gather closely at household with very high household income and with an average of trip duration time of around 10 mins.
3. The least five popular bike stations for starting and ending a journey did not overlap. Most of them are near tourist sites.
4. Bike usage: six bikes with ids of 26202, 29299, 21078, 26293 and 29437 had been used most by riders. Across the entire week, Tue and Wed are having the highest amount of bike usages. Among the riders, male is the dominant sex group. Riding a bike is most popular witg age groups of 18, 22 and 51. Younger people tend to ride bike longer than seniors.