Data Processing

* 2018 December and July csv’s downloaded from website
* Added a column to each csv referencing the year and month
* Combined 2 csv’s via Pandas.
* December data will be classified as winter, while July data will be classified as summer in the Analysis.
* Generation Data – exclude “Other” category holds individuals with a birth year prior to 1946 b/c the possibility is low individuals in this age bracket participated in the program.

Phenomena 1

* Citi Bike Program is more productive in the winter than summer (opposite of expectation)
* Visualization 1: Biker Trip Duration Ave
  + By taking the top 50 repeated Bikers, we notice trip duration average decreases at a slightly higher rate (differences between point B and A is larger) in the summer than the winter.
  + Out of the top 50, we notice majority participated in the winter.
* Visualization 2: Biker Cycles
  + By assessing the top 50 most bikers’ participation (participation is the number of completed cycles by a biker), we notice majority of bikers fall under the winter event and number of completed cycles per biker is greater than those in the summer.

Phenomena 2

* Generations’ performance is consistent in summer and winter (change in weather has no impact on activity).
* Visualization 1: Generation Participation
  + Number of completed cycles per generation is relatively consistent between winter and summer.
* Visualization 2: Generation Trip Duration Ave
  + The trip duration average within a generation in both seasons is relatively consistent

Maps

* 2 Maps display top 100 Start and 100 End Stations used by bikers
* Color change based on increase of popularity (popularity is determined by the count of stations)
* **Note**: Filter for top 100 stations is applied b/c displaying all stations isn’t appealing.

Resources

* Tableau Public Link

<https://public.tableau.com/profile/nader6919#!/vizhome/HW_Final_15719763823600/CitiBikeProgram?publish=yes>