2018 Citi Bike Data Analysis

You can view my Tableau dashboard from the file 2018 Citi Bike Data.twbx from this repo.

* How many trips have been recorded total during the chosen period? Since Tableau Public edition only allows for 1M rows, and my laptop was having trouble with large files, I decided to take a subset of data from the CSV files published on the [Citi Bike Data](https://www.citibikenyc.com/system-data) webpage. For each monthly CSV file for 2018, I was able to head -10000 and then cat 2018\*.csv > 2018-citi-120k.csv. As you will see from my chart NYC Citi Bike | Trips by Month there are a total of 120,000 trips because of how I split the monthly data to 10,000 each.
* By what percentage has total ridership grown? I was not able to determine this from the data that I used.
* How has the proportion of short-term customers and annual subscribers changed? I was not able to determine this from the data that I used.
* What are the peak hours in which bikes are used during summer months? As you can see from my bar graph in NYC Citi Bike | Peak Usage by Hour, during June and July the hour 18:00 (or 6pm) is the busiest. Unfortunately, you cannot see all available data for the month of August with the subset of data that I extracted.
* What are the peak hours in which bikes are used during winter months? As you can see from my bar graph in NYC Citi Bike | Peak Usage by Hour, during December the hour 10:00 (or 10am) is the busiest. Unfortunately, you cannot see all available data for the month of December with the subset of data that I extracted.
* Today, what are the top 10 stations in the city for starting a journey? (Based on See the chart NYC Citi Bike | Top 10 Start Stations for the list of top 10.
* Today, what are the top 10 stations in the city for ending a journey? (Based on data, why?) See the chart NYC Citi Bike | Top 10 End Stations for the list of top 10.
* Today, what are the bottom 10 stations in the city for starting a journey? (Based on data, why?) See the chart NYC Citi Bike | Bottom 10 Start Stations which shows the stations with the fewest starting trips.
* Today, what are the bottom 10 stations in the city for ending a journey (Based on data, why?) See the chart NYC Citi Bike | Bottom 10 End Stations which shows the stations with the fewest ending trips.
* Today, what is the gender breakdown of active participants (Male v. Female)? See the chart NYC Citi Bike | Trips by Age & Gender which shows Male v. Female by age breakdown.
* How effective has gender outreach been in increasing female ridership over the timespan? I was not able to determine this from the subset of data collected.
* How does the average trip duration change by age? See the chart NYC Citi Bike | Trips by Age & Gender which shows Male v. Female by age breakdown.
* What is the average distance in miles that a bike is ridden?
* Which bikes (by ID) are most likely due for repair or inspection in the timespan?
* How variable is the utilization by bike ID?

**Additionally, city officials would like to see the following visualizations:**

* A static map that plots all bike stations with a visual indication of the most popular locations to start and end a journey with zip code data overlaid on top. See the maps on sheet NYC Citi Bike | Trip Frequency by Start Station and NYC Citi Bike | Trip Frequency by End Station.