NYC Citi Bikes: Determining the Most Popular Docking Stations

Project Outline: Citi Bikes are a popular way to get around New York City, allowing people to explore the city while remaining outdoors. The purpose of the project will be to examine the Citi Bikes database and determine the top 10 docking stations with respect to usage. Transactions track where the customer picked up the bike and where it was dropped off (Docking Station ID & Coordinates are included); other details of the transaction such as date/time, customer type (single trip or subscription), and gender are included. Using Google Maps, a map of New York City will be created to illustrate the popularity of the docking stations.

Research Questions:

1. What are the Top 10 docking stations (in 2019) where bikes were taken from and to?
2. Looking into March 2018, 2019, & March 2020, has the usage (in # of trips) of Citi Bikes increased or decreased?
3. Which type of consumer uses Citi Bikes more, Customer (24 Hour Pass/3 Day Pass) or Subscriber (Annual Member), in 2019?
4. What is the average trip duration (in 2019) per rider?

Datasets to be used: The data set that will be analyzed is publicly available and provided by Citi Bikes. The data is provided by monthly reports (csv files) which includes all the transactions made. Google maps library will also be imported and used to illustrate a heat map of the most popular docking stations.

Breakdown of Tasks:

1. Import and merge the datasets (csv files) for the year of 2019; begin cleaning the data.
   1. Group the transactions by Docking Station.
   2. Use Matplotlib to create a bar graph; comparison of station popularity.
2. Count the number of trips for March 2018, 2019, and 2020.
   1. Datasets for March will be brought in 2018, 2019, and 2020 datasets will be analyzed
   2. Use Numpy to count number of transactions per station.
   3. Use Matplotlib to create a bar graph and scatter plot.
3. Create a new DataFrame and group by Customer type to determine which type of customer generates most transactions. Do a separate group by for age groups and bin these particular age groups.
   1. Group the transactions by Customer type.
   2. Use Numpy to count number of transactions.
   3. Use Matplotlib to create pie chart.
   4. Use pandas to create bins for each age group (magnitude of age group TBD).
   5. Use Matplotlib to create a histogram of these bins; illustrates distribution of customers by age group bins.
4. Use Google Maps & Coordinates from Data Set to create a map of NYC illustrating the most popular docking stations.