Harold Young

USC Data Analytics

10/15/18

Tableau Project

**Preface:** The citi bike trip duration data is in seconds. I came to this conclusion by locating the largest rental during the month of which this data was taken (September,2018). Seeing the number stated as 2.44 million, and knowing that the only timeframe which would correlate with that number and the amount of time yielded by this data (one month) would indicate seconds being used which was buttressed by calculating this for further verification.

1. **Total Bike Rentals Per Day:**

The first day of September in 2018 there were 54,323 bike rentals, on the 30th there were 66,233. This yields a monthly gain percent of 21.9%.

1. **Running Total of Bike Rentals per Day:**

This slide shows the total amount of bike rentals in the month of September 2018 which is 1877,844.

1. **How has the proportion of short-term customers and monthly subscribers changed in September 2018?**

Of the two user types, customers started and ended at 8030 and 10765 respectively whilst the subscribers started and ended at 48,999 and 64,217. The customer base yielded a 34% growth during the month whilst subscribers yielded a 31% growth. The percentage of users between the two types in relation to one another start at 6.1 times in the beginning of the month and 5.97 times at the end of the month.

1. **What are the peak hours in which bikes are used during the month?**

Whilst peak hours vary between the day timeframe of September 2018, the 17th and 19th hour, 5-7pm, are on average the peak hours of each day.

1. **10 most popular bike nodes based on trip duration:**

The 10 nodes shown on this map are the most popular based on the sum of the user’s trip duration, respectively. The range of values of these nodes are from 1.74million to 2.92 million. This indicates that these nodes are popular regarding having users near those areas that rent the bikes for weeks at a time.

1. **10 least popular bike nodes based on trip duration:**

The 10 nodes shown on this map are the least popular regarding the sum of the customers trip duration. The range of these durations varies from 65 to 114. Whilst this can not be verifiable with the data yielded from the citi bike dataset, but it seems that these rentals are either accidental due to the range of time for these nodes being between 1 to 2 minutes or the nodes are extremely close to each other in which it would allow the user to go from one node to the other for extremely quick trips.

1. **Today, what are the top 10 stations in the city for ending a journey (Based on data, why?)**

For rental completion, these nodes are the most popular regarding trip duration. The values of these nodes range from 9.3 million to 19.667 million. Most of these nodes are within the 11-12 million range and the highest node (the bright red) being the max of the range.

1. **Today, what are the bottom 10 stations in the city for ending a journey (Based on data, why?)**

These are stations that have a very low trip duration. The range of these values are between 387 and 1000.

1. **Today, what is the gender breakdown of active participants (Male v. Female)?**

Of the total amount of rentals during the month male participants (customers and subscribers accounted for a total of 1,240,964 rentals. This is 33.89% more than the female participants which account for 926,866 rentals.

1. **How effective has gender outreach been in increasing female ridership over the timespan?**

At the beginning of the month the female ridership was 26,750. The month ended with a total of 35,340. This is a 32.11% increase in female ridership for the month.

1. **How effective has gender outreach been in increasing male ridership over the timespan?**

The male ridership started out at 30,279 and ending with 39,642 members. This is a 30.92% increase for the male participants.

1. **How does the average trip duration change by birth year:**

The bike data seems to be unclean regarding the birth year? Due to the citi bike program starting in 2013 and rider birth year starting at 1880, leads one to believe (at first glance) that there were riders ranged from 133 to despite the oldest person to have ever lived(Jeanne Calment) died at the age of 122. This could be due to riders yielding false information when participating in the citi bike program or person mis-filed the data into their database.

The only consistency in the data is from the range of birth years from 1950 to 1987 with a spike in trip duration of riders born in 1967.The trip duration average is between 789 to 860.

1. **What is the average trip duration per citi bike node?**

Whilst the data is too large to explain concisely. The view on this sheet allows the user to understand the average of every trip duration per node.

1. **Which bikes (by ID) are most likely due for repair or inspection in the timespan?:** During the month, it is evident that some bikes are used much more than others but this isn’t indicative of which bikes are most likely due for repairs as the rental program allows users to rent the bikes for however long they like. Due to this, and the data not being based on what distance the bikes actually traveled during that time in which they were rented causes inconclusion regarding the repair and or inspection correlating to trip duration.