**Homework 10 Tableau**

**Mark Ludwick**

1. Data Background:
   1. Used five percent random sample of the 35,757,984 records from June 2017 through May 2019.
   2. Records with unidentified gender also had unreliable birth years
   3. Trips taken by anyone above life expectancy for New York for their gender (females 85, males 79) at the time of the ride were removed from all analyses. This removed a very small percentage of the sample but filtered out some very extreme values.
   4. Trips that were longer than 30 miles were filtered, again they were very few.
   5. A few hundred rides had null values for the Station IDs and Station Names, these were filtered out as well.

|  |  |  |
| --- | --- | --- |
| **Question / Question Area** | **Tableau Worksheet Containing Analysis & Answers** | **Comments in addition to the Tableau Workbook** |
| How many trips have been recorded total during the chosen period? | Gender Integrated |  |
| By what percentage has total ridership grown? | Ridership Growth |  |
| How has the proportion of short-term customers and annual subscribers changed? | Proportion of Short-term vs Subscribers |  |
| What are the peak hours in which bikes are used during summer months? | Peak Hours by Season |  |
| What are the peak hours in which bikes are used during winter months? | Peak Hours by Season |  |
| Today, what are the top 10 stations in the city for starting a journey? (Based on data, why do you hypothesize these are the top locations?) | Top 10 Start Stations | Most or all of the top 10 start stations appear to be in Manhattan, volume is heaviest during the morning and evening commutes so commuting to and from home/subway/train/bus to and from work drives. |
| Today, what are the top 10 stations in the city for ending a journey? (Based on data, why?) | Top 10 End Stations | The list of the top 10 End Stations is very similar to the top 10 start stations. Likely that riders are using the bikes to get both to and from work. |
| Today, what are the bottom 10 stations in the city for starting a journey? (Based on data, why?) | Bottom 10 Start Stations | These locations seem to be off the beaten path, probably more residential and with more dispersed stations to access and drop off the bikes. |
| Today, what are the bottom 10 stations in the city for ending a journey (Based on data, why?) | Bottom 10 End Stations | These are essentially the same stations as the bottom 10 start stations. |
| Today, what is the gender breakdown of active participants (Male v. Female)? | Gender Breakdown by Quarter | There is no user IDs in the dataset and not enough person ID information to create a unique user ID (just birth year and gender), so this question cannot be answered with the available data. Setting that aside, females make up 25.8% of bike rides in the most recent quarter, up from 22.7% in the previous quarter. That would seem to indicate that female bike rides are on the rise, but these percentages are nearly identical to those in the same quarters of the previous year |
| How effective has gender outreach been in increasing female ridership over the timespan? | Gender Breakdown by Quarter |  Gender % Increase Over Months] | From the beginning of the study data timeframe, the overall increase in monthly female bike rides has increased by 16.5%, while the increase is lower for males (14.2%) |
| How does the average trip duration change by age? | Trip Duration by Age |  Trips Taken by Age] | Trip Duration generally declines as age increases, but much more pronounced is that the number of rides decreases after a high around 30, dip and plateau ending at 48, then declining to near zero by age 85, which was the maximum age included. There is an anomaly that occurs at age 83 with the average trip duration going up to 35.64 minutes, I did not track down the source for this, but would ordinarily do that. I'll just say that the number of trips were relatively few and someone keeping a bike overnight by accident a time or two could drive this number up. |
| What is the average distance in miles that a bike is ridden? | Mean Miles and Total Trips Per Bike | My analysis for this and the next two questions does not take into account that some bikes came on-line after others. This is important and could be calculated from the available data, but I didn't go there for this homework. Mean miles per Bike ID is 107; and the mean number of trips for a Bike ID is 97. |
| Which bikes (by ID) are most likely due for repair or inspection in the timespan? | Bike IDs Ordered by Maintenance Needs | This sheet essentially operationalizes the previous sheet [Mean Miles and Total Trips Per Bike] to rank order each bike first by the trip distance and then by the number of trips. Trip Distance is the primary driver so it is the primary sort variable, then when percentiles of Trip Distance are the same, then Number of Trips breaks the tie. |
| How variable is the utilization by bike ID? | Mean Miles and Total Trips Per Bike | Sixty-eight of the bikes have been ridden between 58.3 and 155.6 miles; given that the bottom of the range is just over one third of the top of this range, I'd say the variable in miles traveled is high. Ninety-five percent of the bikes have been ridden between 9.7 and 204.3 miles. Similarly, 68% of the bikes have gone on between 54.2 and 139.4 trips, a wide range for similar reasons. |
| 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. | Start Station Total Volume |  End Station Total Volume |  |
| If you're working with a merged dataset: a dynamic map that shows how each station's popularity changes over time (by month and year) -- with commentary pointing to any interesting events that may be behind these phenomena. | Start Station Volume Over Time |  End Station Volume Over Time | Over the two years, station locations have been added and removed. With more being added to northern Manhattan, the few in New Jersey either shut down or are no longer have their data captured in this database. Seasonal fluctuations are evident as volumes drop in the winter months. Governor’s Island has no bike usage at all between October and April each year. |
| Find at least two unexpected phenomena in the data and provide a visualization and analysis to document their presence. | Peak Hours by Season | I was a little surprised to see that during the past two years, while summer has the highest volume, Fall and not Spring is the season with the second highest volume. |