**Benjy Manning’ CitiBike Analysis for City Leaders**

Prepared by: Benjy Manning

**Overview of Data and Approach:**

The analysis below and visualized is for the year of 2020 and includes all rides that were recorded during the year.

Initial data was obtained from the CitiBike Public Website link (<https://s3.amazonaws.com/tripdata/index.html>) and was extracted and transformed using python programming and then imported to Tableau to analyze and visualize for this presentation.

**Unexplained Phenomena:**

There were numerous phenomena observed during the analysis. The following were most the most unexplainable and may warrant further review to enhance program tracking and customer experience:

# 1 - Demographics:

* When reviewing different age data, the following items were observed:
  + There is a significant amount of Unknown gender/aged 52 participants.
  + This is observed in the ages by decade when removing non suspect ages as well as in the trip analysis by age visualizations.
  + There are also 21 participants aged 100 or older.
* Some users may not feel comfortable sharing gender data, particularly non subscribed users.
* Some non-subscribing users may accept a default birth year when initiating rides which may explain the high number of participants in the age of 52.

# 2 – Ridership

* + Customers tended to ride more during leisure hours where as subscribers tended to ride more during commuting hours during the winter months. This may indicate that subscribers tend to use the program for commuting.
  + Subscribers tend to ride the most during evening leisure hours with some concentration during commuting hours where as customers tend to increase ridership from afternoon to evening.
  + Ridership increases most during the period of April to June which also corresponds with the largest increase in subscribers.

**Notable Questions Answered from the Data:**

1. How many trips have been recorded total during the chosen period?
   1. 336,802
2. What are the peak hours in which bikes are used during summer months?
   1. Peak Hours in the summer months are 5:00 PM to 8:00 PM
3. What are the peak hours in which bikes are used during winter months?
   1. 8:00 AM and 5:00 – 7:00 PM (subscriber) Commuter
   2. 12:00 – 5:00 (customer) Leisure
4. 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?)
   1. All appear generally located near transit stations or parks which would support both daily commuter and leisure users.
5. Today, what are the top 10 stations in the city for ending a journey? (Based on data, why?)
   1. Nearly the same as the beginning journeys
6. Today, what are the bottom 10 stations in the city for starting a journey? (Based on data, why?)
   1. These stations appear to be located in lower income neighborhoods and further away from Transit stations as well as more urban zip codes.
7. Today, what are the bottom 10 stations in the city for ending a journey (Based on data, why?)
   1. The Bottom 10 stations are appearing in Manhattan and other NYC boroughs.
8. Today, what is the gender breakdown of active participants (Male v. Female)?
   1. There are 84,798 female participants, 199,807 male participants and 52,197 participants of unknown gender.
9. How effective has gender outreach been in increasing female ridership over the timespan?
   1. Male Riders have grown faster and by more than females.
   2. Female ridership grew by ~80,000 particpants where as male participant ridership grew by ~200,000 particpants.
   3. Marketing programs targeted to female riders may need to be re-assesed.
10. How does the average trip duration change by age?
    1. After adjusting for some outlier age data discussed in the phenomena, section, average duration is generally higher for younger and older participants. This would make sense as the younger participants generally rode further while older particpants may ride shorter distances but at slower speeds.
11. Which bikes (by ID) are most likely due for repair or inspection in the timespan?
    1. Bikes 44683 and 40831 may likely require repair as they appear to have over 100X the distance as the nearest other 10 ten bikes.
    2. Bike ID 42308 may also require service as its recorded durtation in service is nearly 2x the the bike with the next highest duration in service and also appears in the top 10 bikes by number of rides.

**Visualizations Created:**

* Total Number of Rides by Month
* Percent Change in Rides Month over Month
* Ridership Change Month over Month
* Total Rides by Rider Type and Month
* Summer Peak Hours
* Winter Peak Hours
* Top 10 Starting Stations
* Top 10 Ending Stations
* Bottom 10 Starting Statoins
* Bottom 10 Ending stations
* Ages by Decade 10 Year Band and Gender
* Riders by Gender
* Annual Rider Count by month and Gender
* Trip Analysis by Age – Average Trip Duration in Minutes by Age
* Trip Analysis by Age – Average Trip Distance by Age
* Trip Analysis by Age – Number of Rides by Age
* Top 12 Bikes by Distance Riden
* Top 10 Bikes by Duration Riden
* Top 10 Bikes by Number of Rides taken
* Total Minutes Ridden in 2020 by month
* Starting Station Map by Volume
* Bottom Starting Station Map with Median Income and Zip Codes
* Ending Station Map with Zip with Zip Codes
* Bottom 10 Ending Station Map
* Most Popular Starting and Ending Stations Combinations

**Dashboards**

* Top 10 Stations and Stations Combinations
* Annual Rides and Monthly Overview
* Peak Hours Dashboard
* Annual and Monthly Ride Data

**Storys:**

* Operation Rider, A Citibike Data Story
  + Top 10 Stations and Combinations
  + Peak Hours of Ridership by Summer and Winder