 

Submitted by Rick Gentile on 10/3/2020

Analysis of Citi Bike Data.

* The analysis of the Citi Bike Data I undertook involved the following data selection, data addition and data cleaning:
  + Citi Bike daily ridership data from August 2018, August 2019 and August 2020.
    - Additional data elements added to the file were as follows:
      * Month Year – August\_2020, August\_2019 and August\_2018.
      * Age of the rider – Example 47. Which is Birth Year minus data point year i.e. August\_2020.
  + Weather Data from WeatherUnderground.com was scraped for each day of August 2018, August 2019 and August 2020. The data points utilized were the temperature average and precipitation average for each day.
    - Additional data elements added to the file were as follows:
      * Month-Year - August\_2020, August\_2019 and August\_2018.
      * Data points removed were High and Low for each category of Temperature, Dew Point, Humidity, Wind Speed, Pressure and Precipitation.
* The data was selected to determine the impact of COVID-19 on ridership by the quantity and quality trips taken and to also determine if there was an adverse effect to overall ridership or to riders by gender and age group and to use weather elements of temperature and precipitation to account for lowered ridership.
* Based on the above the following are the two phenomenae observed. Please note, when I wrote about quantity and quality, quanity represented the number of riders and quality represented trip duration and where those trips began and ended.
  + Phenomena One – Ridership totals rose 9.6% in 2019 from 2018. Although ridership fell 12% during 2020 it was only 2.6% lower than in 2018 which of course was not a pandemic year. Other ridership data also showed that most age groups were consistent from each year with the exception of the male age group of 30-44. As for total by gender, the male gender was the biggest dropff in 2020 of which in part may be due to the increase in the ‘Unknown” gender group.
    - Conclusion – Upon review of the data detailed above the ridership numbers although a decrease from 2019, did not fall as much as I would expect during a pandemic wherein New York and New Jersey were still in virtual shutdown. Based on the the data people were still willing to ride bikes during this crisis which should be a predictor for increased ridership once the pandemic lockdowns have been lessened or eliminated.
  + Phenomena Two – Weather Data was utilized to determine if the conditions or quality of a rider’s experience was effected by extremely warm weather or heavy precipitation prohibiting rider’s from engaging in the activity. In reviewing the data across the three years neither the temperature or rain had an impact on ridership as compared from year to year. In short, the weather from 2018-2020 did not give any year an ‘advantage’ over other years in terms of better weather. Also reviewed was the trip duration based on Trips of 10 days or less (14,440 minutes). This was done to eliminate any outlier data found in August 2020 data which had data points of trip durations that were well outside the maximum from other years. Based on that data the Trip Duration for August 2020 doubled that of August\_2019. The number of trips may have been less but they were longer. In terms of stations where trips began and ended, August 2020 did see less starting points for trips but did see nearly the same trip ending points for previous years.
    - Conclusion – Upon review of the detailed data, the quality of a trip i.e. duration increased in 2020. Although less starting points are reported in August 2020, the ending trip points were similar to those of the other years. Once again, it was my expectation during a pandemic ridership quality of a Bike Ride would be less not more in terms of average trip duration. Although the starting points were less, the ending destination points were similar to other years.