Assignment 4: Exploratory Analysis Using Tableau Data Science for Business Decisions

Student ID: 261032796
Name: Ahmed Ibrahim
Email: ahmed.ibrahim6@mail.mcgill.ca

Development

Question 1

a) At first glance, these two visualizations look similar, both being bar charts, using the same color to fill the bars and both have start station codes as their columns. However, the length of the bar charts and the gradient of the x-axis are not the same for both visualizations. Thus, both bar charts are essentially not the same.

b) Presumably, the right chart refers to the number of records associated with each start station. In contrast, the left chart illustrates only the unique instances of durations (seconds) for each start station. Using the script found in Appendix A, it was possible to confirm that all the number of records had a value greater or equal to the number of unique duration instances.

c) The merged plot uses the start station code data as rows and the counts of trip durations (seconds) as columns. The percentage of distinct values are visualized via a color gradient, and labels are assigned to each bar (script in Appendix B). The merged bar chart is found in appendix B being figure 1.

d) The merged bar chart conveys the information found in both bar charts by illustrating the number of records per station via each bar's length, and the sequential color gradient illustrates what percentage of those records are the distinct duration of trips. The labels are added to aid in the viewers' process of comparing both of those fields. Thus, the use of the labels and color gradient illustrates the comparison of the distinct and total records within a singular bar chart.

Question 2

a) The filtered bar chart is generated by first conducting a right join between station data and the trip data from august 2019. A right join is used to ensure all stations are including within the visualization. Any rows from the trip data that do not have any matches with the station rows are excluded. The key value used to conduct the join operation is the station codes found in both datasets respectfully. The bar chart columns are the station names, and the rows are the total duration of trips (seconds). The sum of trip durations is filtered at a breaking point of 50,000 seconds. Only four stations had such sums within 0 to 50,000 seconds, while four other stations were found to have no trips (see figure 2 in Appendix C).

b) The four stations with no trips recorded are in Laval (see Figure 3 in Appendix C), such that these stations were made available starting in October 2019 [1]. Six BIXI stations are in Montréal-Est, where four of these stations have total trip durations below 50,000

seconds, and the other two have higher sums of 62,566 and 81,624 seconds, respectively (see figure 4 in Appendix C). The low usage rate in Montréal-Est can be potentially attributed to a low bike-station density [2], being further away from metro stations, and having fewer commercial enterprises within the region [3].

Appendix

Appendix A

The following is a conditional script to verify that whether the number of records is greater or equal to the distinct counts of durations (seconds) for each start station (see the "Test Number of Records vs. Distinct Values" sheet in the *Tableau* file named *YCBS-256.assignment4_q1.twb* to find the calculated field named *CheckRecordsvsDistinctDurations*):

```
IF (COUNTD([Duration Sec]) <= SUM([Number of Records])) THEN
TRUE ELSE FALSE  END
```

Appendix B

The following is a conditional script to compute the percentage of distinct counts for durations of trips (seconds) in ratio to the sum of the number of records (see the "Merged Bar Chart for Distinct Trip Durations per Start Station" sheet in the *Tableau* file named *YCBS-256.assignment4_q1.twb* to find the calculated field named *Ratio of Distinct Values to Total Values*):

```
COUNTD([Duration Sec])/SUM([Number of Records])
```
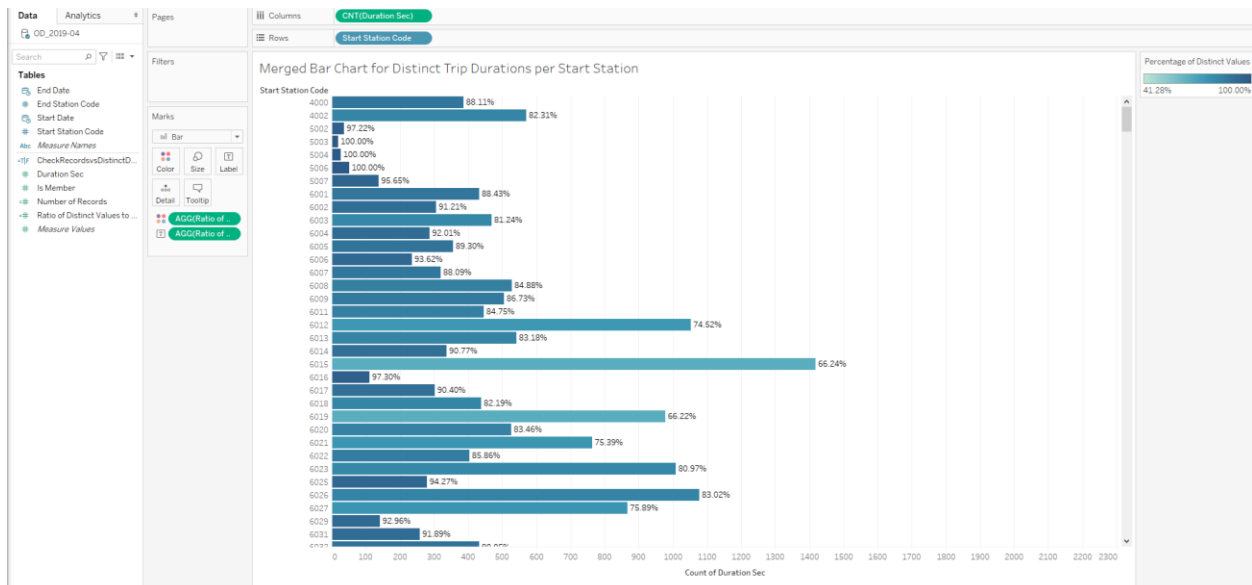

Figure 1. Screenshot of a bar chart for distinct duration of BIXI trips per start station code. The blue teal color was used as sequential color. This screenshot was taken from the "Merged Bar Chart for Distinct Trip Durations per Start Station" sheet in the *Tableau* file named *YCBS-256.assignment4_q1.twb*.
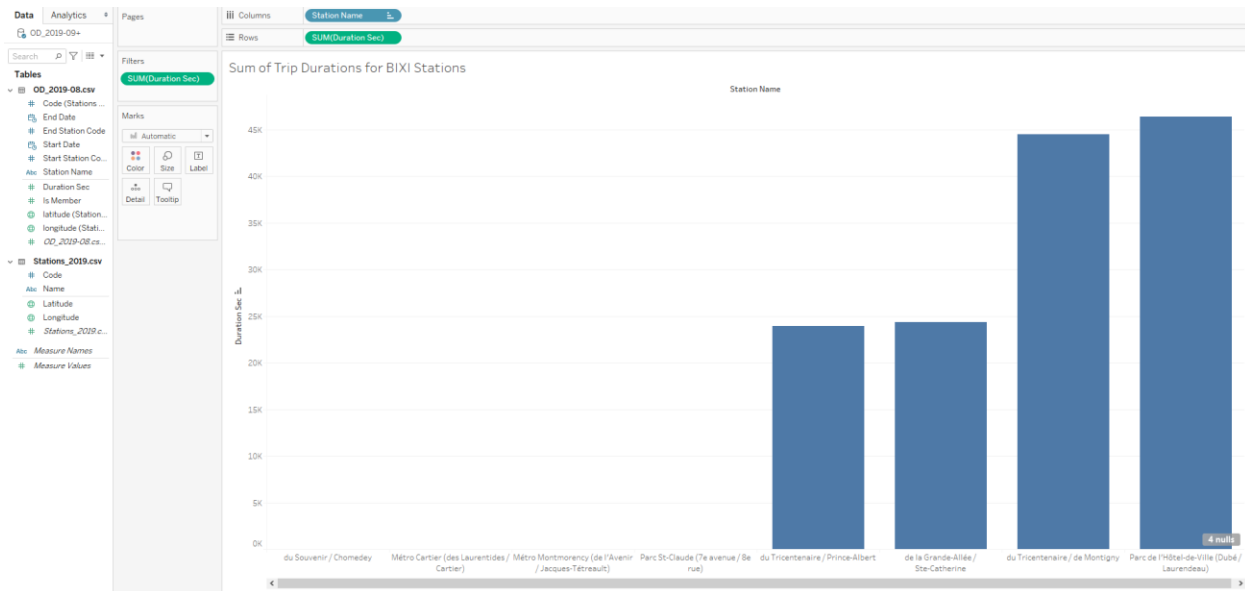
Figure 2. Screenshot of a bar chart for the total duration of BIXI trips (seconds) per station. The stations shown have less than 50,000 seconds of total trip duration. This screenshot was taken from the "Bar Chart for BIXI Stations with less than 50,000 seconds of Total Trip Duration or no Trips" sheet in the *Tableau* file named *YCBS-256.assignment4_q2.twb*.

The following script is used in figure 3 to segregate the stations into three groups being the ones with no trips, total durations over and under 50000 seconds, respectively (see the "Bar Chart for BIXI Stations with less than 50,000 seconds of Total Trip Duration or no Trips" sheet in the *Tableau* file named *YCBS-256.assignment4_q2.twb* to find the calculated field named *Sort Stations by Trip Duration*):

```
IF SUM([Duration Sec]) > 50000 THEN "More than 50000 sec."
ELSEIF SUM([Duration Sec]) < 50000 THEN "Less than 50000 sec
ElSE "NULL"
END
```
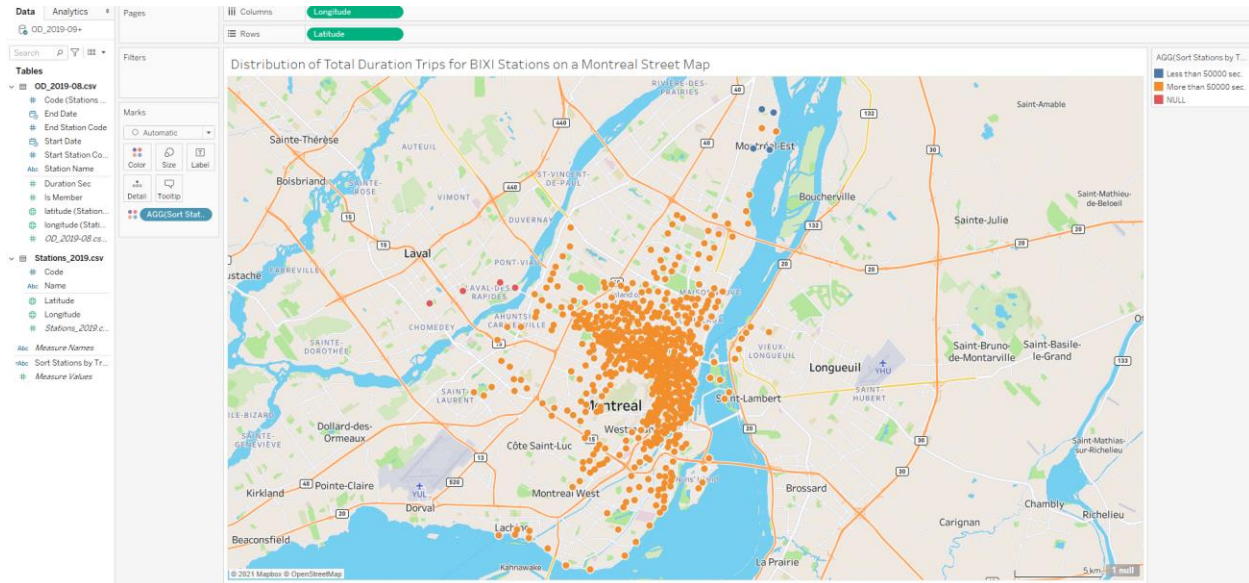
Figure 3. Screenshot of a map illustrating the distribution of total trip durations per BIXI station. All stations in blue represent a total trip duration below 50000 seconds, the ones over 50000 seconds are in orange, and the ones that have no trips recorded are in red. This screenshot was taken from the "Total Duration of Trips per BIXI Stations on a Montreal Street Map" sheet in the *Tableau* file named *YCBS-256.assignment4_q2.twb*.
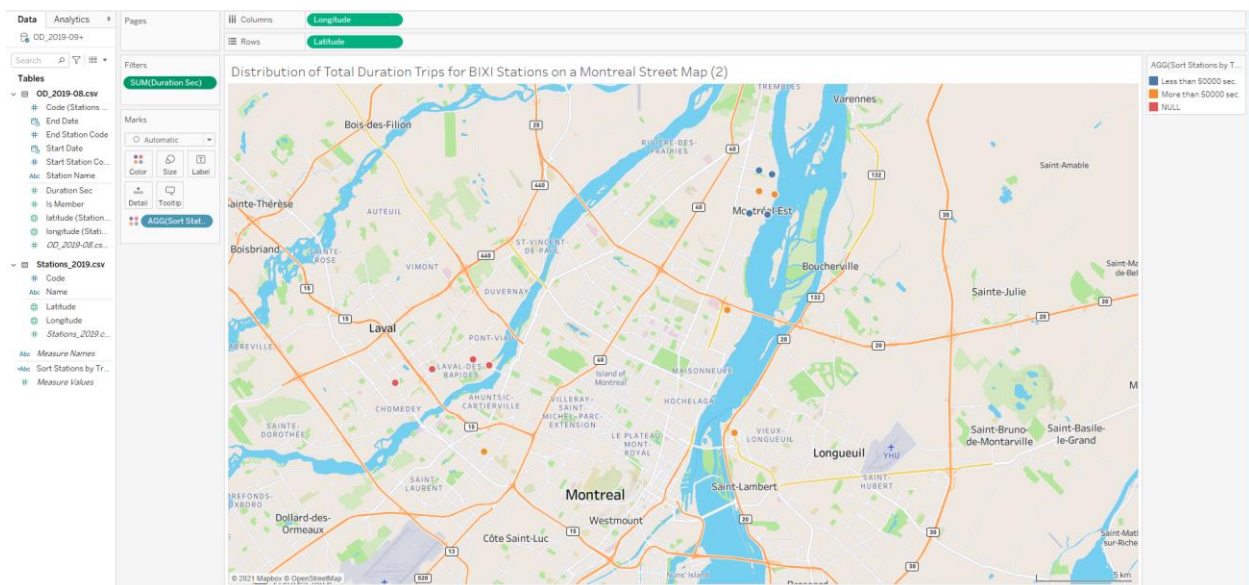

Figure 4. Screenshot of a map illustrating the distribution of total trip durations per BIXI station. A filter field was added only to include stations with total durations below 100,000 seconds. All stations in blue represent a total trip duration below 50,000 seconds, the ones over 50,000 seconds are in orange, and the ones that have no trips recorded are in red. Five stations were found to have durations between 50,000 to 100,000 seconds. This screenshot was taken from the "Total Duration of Trips with less than 100,000 seconds per BIXI Stations on a Montreal Street Map" sheet in the *Tableau* file named *YCBS-256.assignment4_q2.twb*.

References

1. St-Pierre, C. (October 2019). *Bixis arrive in Laval*. https://www.thesuburban.com/news/laval_news/bixis-arrive-in-laval/article_a4ae4d53-bf40-5174-bf84-3d3c8a7c9c49.html
2. National Association of City Transportation Officials (2015). *Walkable station spacing is key to successful, equitable bike share.* NACTO. https://nacto.org/walkable-station-spacing-is-key-to-successful-equitable-bike-share/
3. Imani, A., Eluru, N., El-Geneidy, A., Rabbat, M. & Haq, U. (2014). *How does land-use and urban form impact bicycle flow: Evidence from the bicycle-sharing system (BIXI) in Montreal*. Transport Geography. http://dx.doi.org/10.1016/j.jtrangeo.2014.01.013