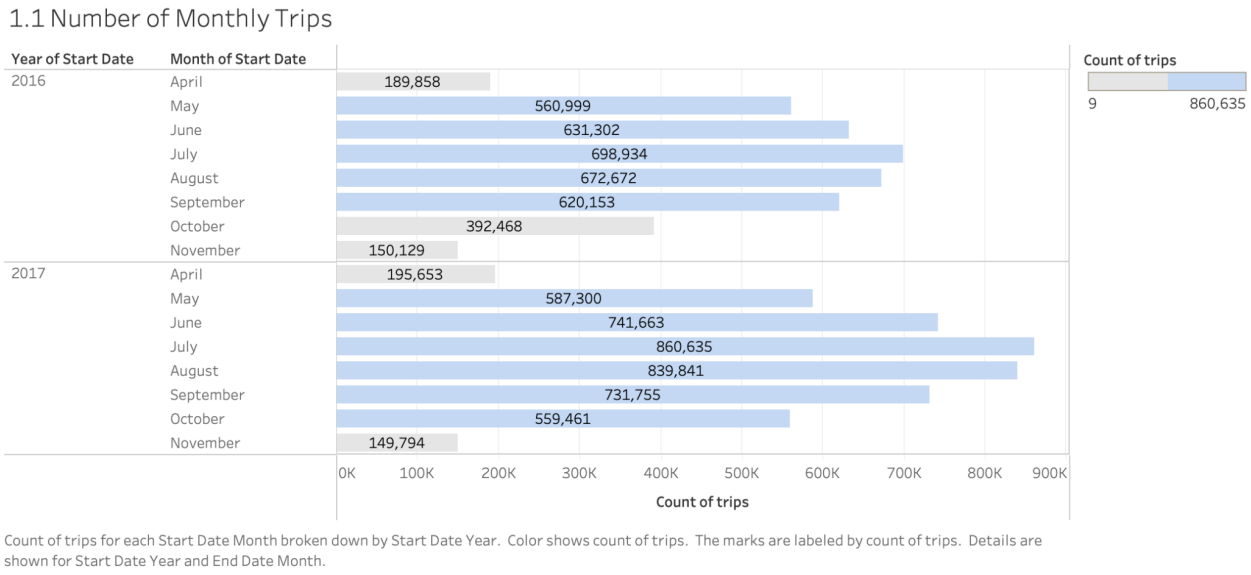
Visual Analytics - Bixi Project Part 2

By Allan Isla

Question 1

1. What differences do you notice about the usage of the Bixi service between the two years?

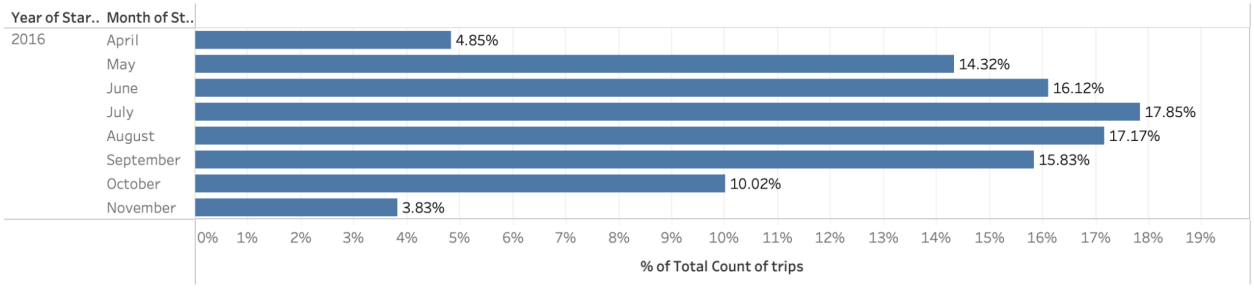


I noticed that the proportion of usage for each month is the same between both years. The summer months were the busiest and the other months were less busy in both years.

However, usage increased in every comparable month in 2017 from 2016 with the exception of November. This was the only month where 2016 surpassed usage than in 2017.

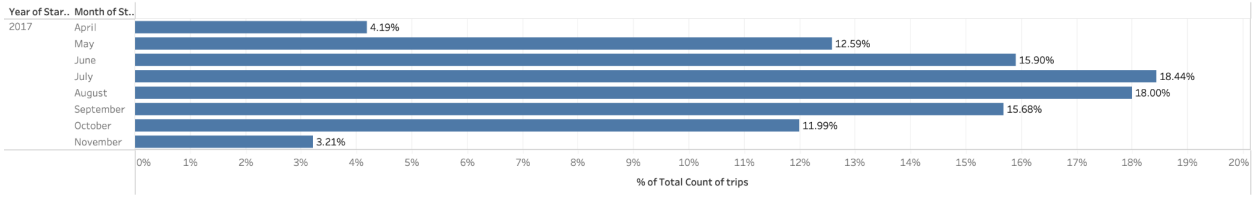
2. How does the proportional monthly usage differ between 2016 and 2017?

1.2a Percentage of Monthly Trips 2016



% of Total Count of trips for each Start Date Month broken down by Start Date Year. The marks are labeled by % of Total Count of trips. The view is filtered on Start Date Year, which keeps 2016.

1.2b Percentage of Monthly Trips 2017



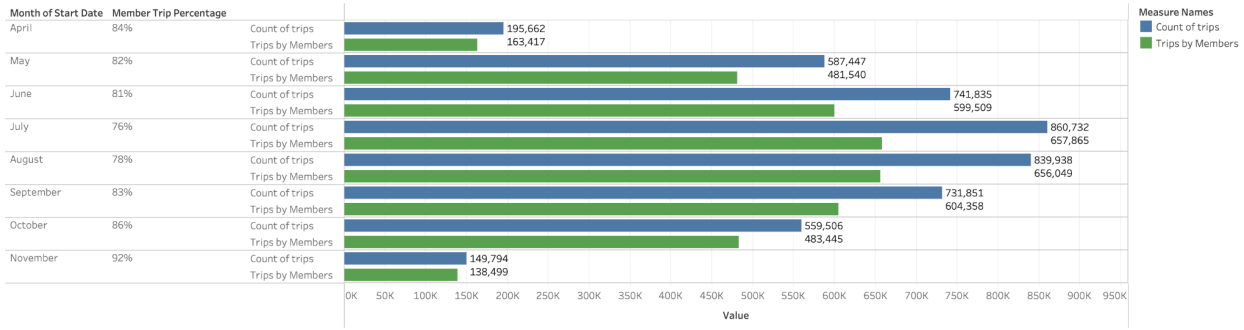
% of Total Count of trips for each Start Date Month broken down by Start Date Year. The marks are labeled by % of Total Count of trips. The view is filtered on Start Date Year, which keeps 2017.

Similar to question 1, the proportion of usage for each month is the same between both years. However, the most popular months of July and August received a higher percentage of users in 2017 than in 2016. All the other months except October had a higher percentage of users in 2016 than 2017.

July and August in 2017 had significantly more users compared to the other months throughout the year. In 2016, while those 2 months were also the most popular, they were not as considerably busier than the other months like they were in 2017.

3. Visualization of the percentage of trips per month by members in 2017.

1.3 Percentage of Trips by Members 2017



Count of trips and Trips by Members for each Member Trip Percentage broken down by Start Date Month. Color shows details about count of trips and Trips by Members. The marks are labeled by count of trips and Trips by Members. The data is filtered on Start Date Year, which keeps 2017.

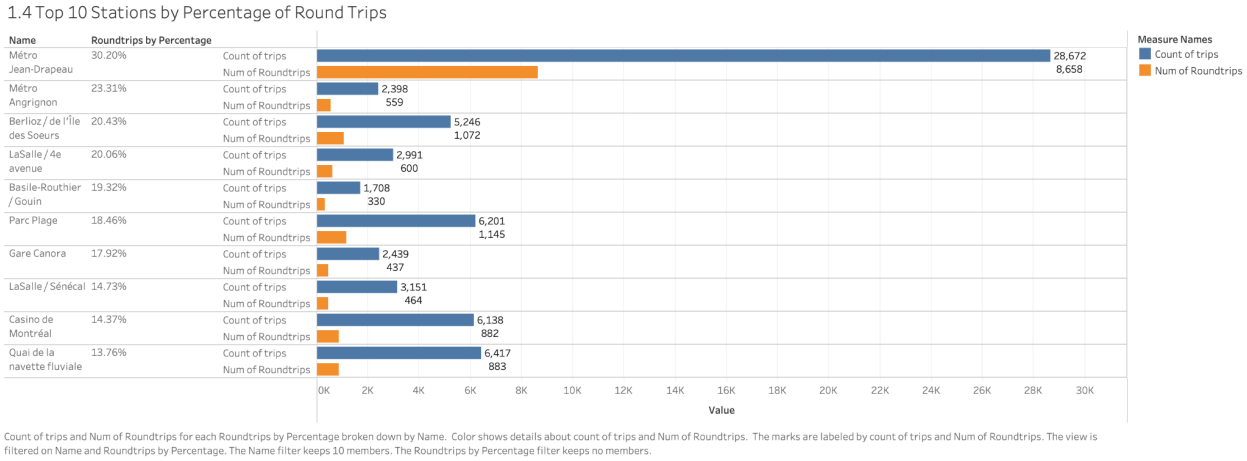
I created a calculated field called **Member Trip Percentage**. I wrote the statement ROUND(SUM([Is Member]) / COUNT([Id]), 2) to calculate the total amount of trips by members compared to the overall amount of trips.

I displayed the amount of trips per month in 2017 and the amount of trips by members per month that year. I also provided the percentage of trips completed by members in each month.

This chart shows that at least 76% of trips each month are completed by members. The highest percentage of trips completed by members are in the colder months, which also has the fewest bike users. The lowest percentage of trips by members are in the summer months, which is the busiest time of the year.

These trends are likely attributed to seeing a higher percentage of non-members riding during the summer months due to ideal weather and a lower percentage during the colder months since fewer people overall have the desire to ride then.

4. Visualization of the top 10 stations by percentage of round trips.



I created a calculated field called **Num of Roundtrips**. I used the statement SUM(If [Start Station Code] = [End Station Code] then 1 ELSE 0 END)) to calculate the number of roundtrips for each station.

I also created a calculated field called **Roundtrips by Percentage**. I used the statement SUM(If [Start Station Code] = [End Station Code] then 1 ELSE 0 END) / COUNT([trips]) to calculate the percentage of roundtrips for each station.

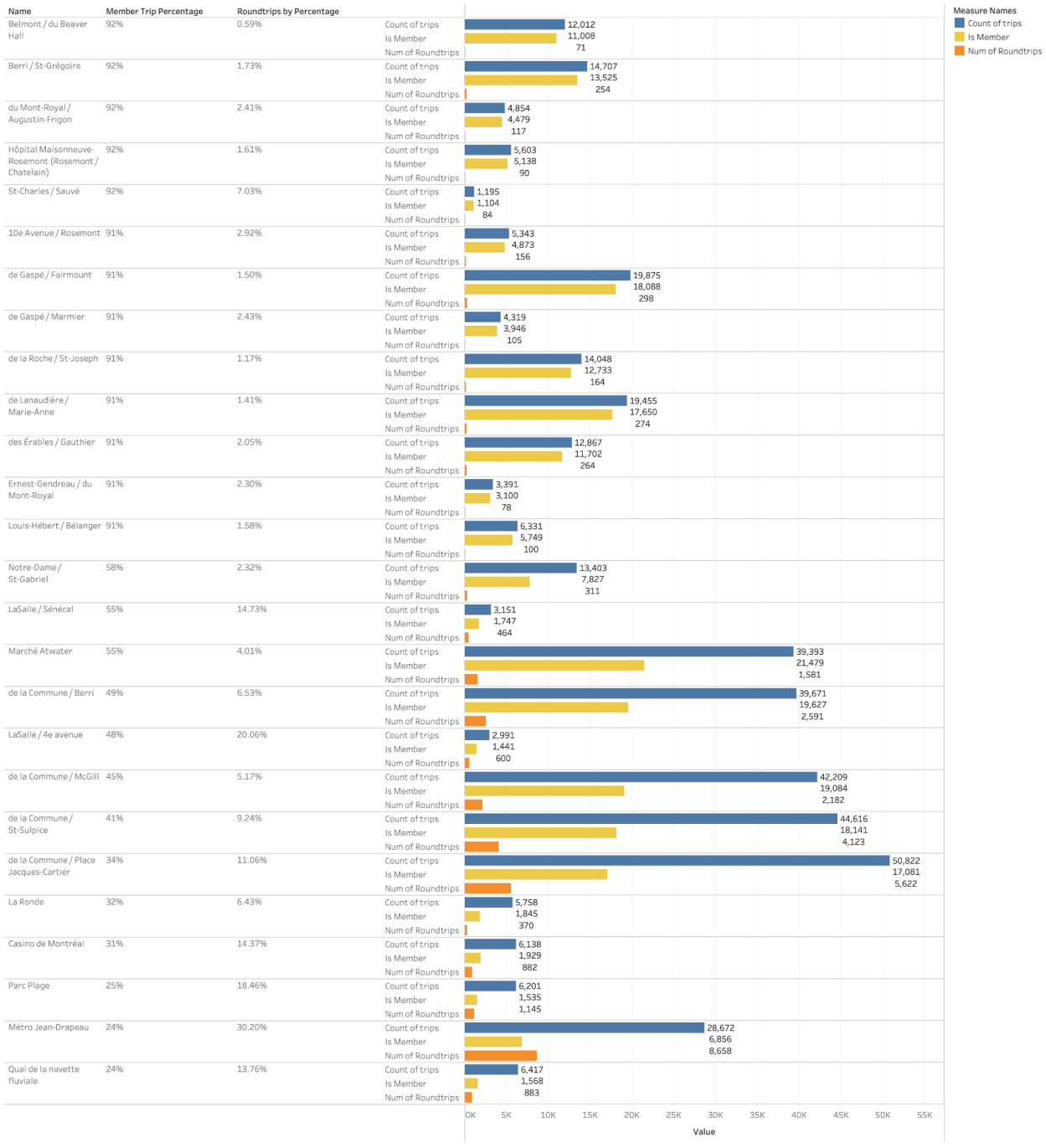
I displayed the amount of overall trips and number of roundtrips per station. I sorted the stations that had the highest percentage of roundtrips and narrowed the list to the top 10 stations.

These bike stations tend to be close to train or bus stations located nearby parks or tourist attractions for bikers to ride around leisurely and enjoy the scenery then return to the same bike station before getting back on the train or bus they originally arrived on.

Question 2

1. Relationship between percentage of round trips and percentage of member trips by station. Are there any interesting patterns?

2.1 Member Trip % & Roundtrip % per Station



Count of trips, Is Member and Num of Roundtrips for each Member Trip Percentage and Roundtrips by Percentage broken down by Name. Color shows details about count of trips, Is Member and Num of Roundtrips. The marks are labeled by count of trips, Is Member and Num of Roundtrips. The view is filtered on Name, which keeps 26 members.

I listed each station and I then provided the member trip percentage and the roundtrips by percentage for each station. I sorted the list in order by member trip percentage and filtered the stations to display those that had the highest and lowest percentage.

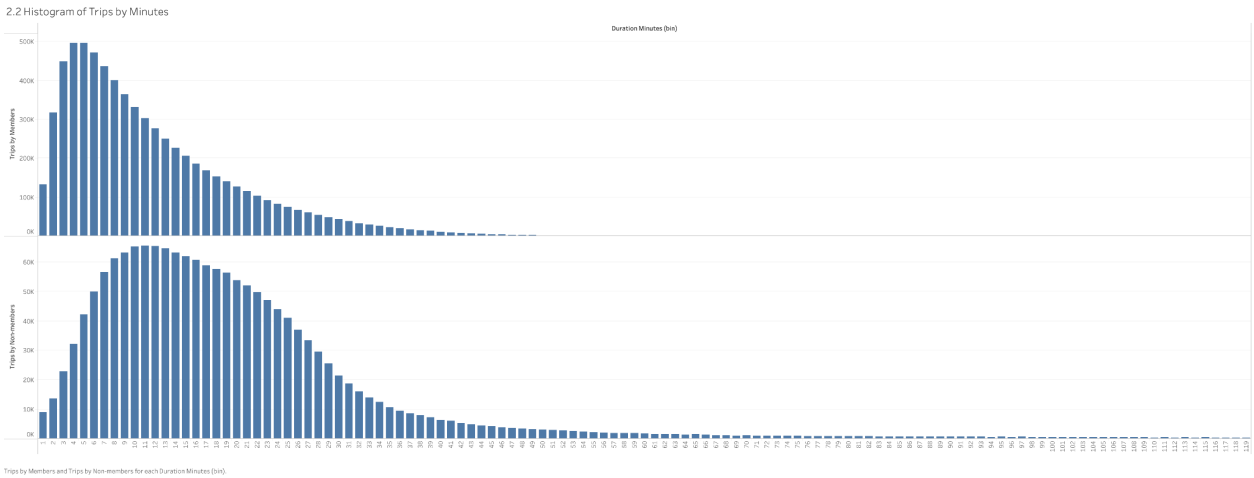
The stations that had the highest percentages had 91-92% of trips taken by members while the stations with the lowest percentage of trips taken by members ranged from as low as 24% to 58%.

From this data, I gathered that the stations with the highest percentage of trips taken by members had very low percentages of roundtrips, all less than 10%. The stations that had the lowest percentage of trips by members tended to have some of the highest percentage of roundtrips, several of them had 10% or greater.

I concluded that members do not take roundtrips frequently, these are people who tend to use bikes routinely and are usually using them to go from one place to another such as to work or home.

Roundtrips tend to be for riders using them to ride around parks, islands and tourist attractions to leisurely enjoy the bike ride as opposed to going from one station to another. These are people who might be just visiting the city or people who don't plan on using the bike more than once in a day.

2. Histogram visualizing the distribution of all trips by duration in minutes between member and non-member trips. What can be said about the behavior of members vs. non-members in terms of trip length?



I created a calculated field called **Trips by Members** and **Trips by Non-members**.

I used the statement SUM([Is Member]) to create **Trips by Members** and COUNT([trips]) - SUM([Is Member]) to create **Trips by Non-members**.

I also created a calculated field called **Duration Minutes**. I simply divided the data Duration Sec by 60 to create this new data field. Then I filtered these minutes into bins to display all the trips completed by members and non-members by each incremental minute.

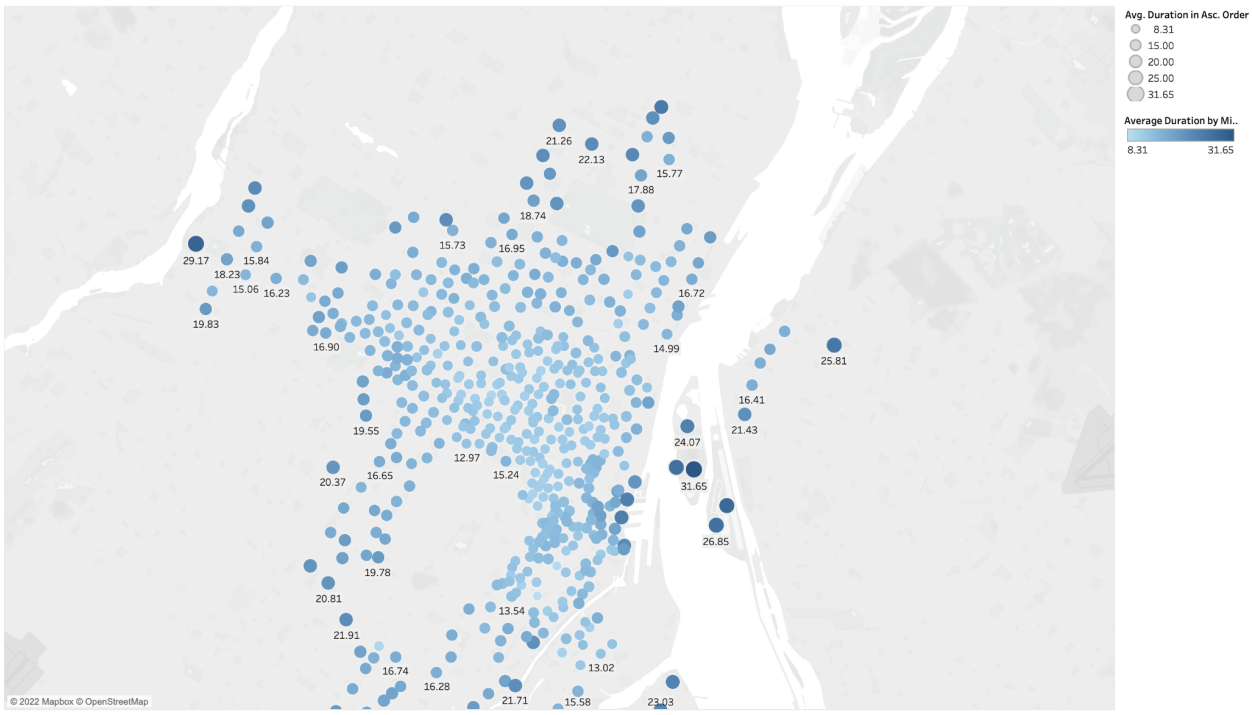
The most common duration of trips by members lasts around 4-5 minutes, while for non-members the most common duration is about 10-11 minutes. Since members prepay upfront for unlimited rides, they can ride for as long or as little as they want on each trip without any additional charge.

Even if a location is only a few minutes away, knowing that there is no additional payment required, members will often take advantage of this benefit to help them get to their destination faster than simply by walking.

Since non-members are charged a flat rate for each use, they are motivated to get their monies worth with each trip. This likely influences them to use these bikes for lengthier durations and avoid using the bikes if the ride will only take a few minutes to complete.

3. Map to visualize the average trip duration per station across the city. Are there any interesting geographic patterns? Why might this be?

2.3 AVG Duration per Station



Map based on average of Longitude and average of Latitude. Color shows AVG Duration. Size shows AVG Duration. The marks are labeled by AVG Duration. Details are shown for Name.

I created a calculated field called **AVG Duration** and wrote AVG(Duration Minutes) to find out the average trip by minutes at each station.

I noticed that the stations furthest away from the center of the city tend to have higher duration averages than the stations closest to the center.

Since the stations closest to the center are nearby to other bike stations, train stations and popular attractions, people tend to arrive at their destination in a few minutes. The stations furthest from the center require people to travel longer distances just to get to the next closest station or closer to the center of the city where it's more bustling.

Also, many of the stations that have higher average durations tend to be closer to islands and parks where people like to ride more leisurely as well to enjoy the ride as opposed to riding to get from one location to another.

Question 3

While the revenue Bixi generates from their members, and 1-day and 3-days passes is fairly well understood by the business, they would like to get a better understanding of the revenue

generated by infrequent users who make single, shorter trips an hour long or less. The pricing model for single trips is as follows:

- $2.99 flat rate for each trip that is 30 minutes or less
- $4.79 ($2.99 + $1.80) for trips greater than 30 minutes, up to 45 minutes in length
- $7.79 ($2.99 + $1.80 + $3) for trips greater than 45 minutes, up to 60 minutes

Note: You may assume every non-member trip is a single trip, subject to the pricing model above.

1. Create a calculated field to calculate the revenue generated by the above pricing model.

   I created the calculated field, **Revenue for Single Trips**. I used the statement

   IF [Duration Minutes] <= 30 then 2.99
   ELSEIF [Duration Minutes] <= 45 then 4.79
   ELSEIF [Duration Minutes] <= 60 then 7.79
   END

   This calculates the amount of times that a trip lasted 30 minutes or less and then adds 2.99 for each of these occurrences.

   Then this calculation takes trips that are less than or equal to 45 minutes that haven't been previously accounted for yet and adds up 4.79 for each of those occurrences.

   Finally, this statement adds up the trips that were 60 minutes or less that haven't been previously accounted for yet and then adds 7.79 to each of those trips.

   This calculates to **$26,640,889.24** as total revenue for single trips.

2. What are the total dollar amounts and relative percentage of revenue from single trips up to an hour in length for each of the three different pricing buckets above?

   I created the calculated field, **Revenue 30 min. & Under**, I used the **Duration Minutes** calculated field and IF statement that I used to answer the previous question, but only looking for information on the trips that lasted 30 minutes or less.

   IF [Duration Minutes] <= 30 then 2.99

END

This calculates a total revenue for these trips to be $**23,947,930**, which makes up **89.89%** of revenue from single trips.

I created another calculated field, **Revenue 45 min. & Under**. My IF statement for this is:

IF [Duration Minutes] > 30 and [Duration Minutes] <=45
then 4.79 END

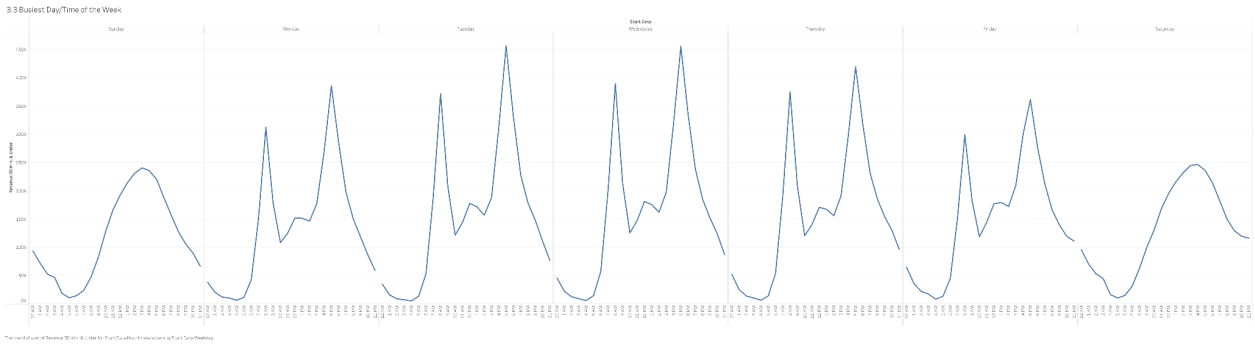
This calculates total revenue trips greater than 30 minutes, up to 45 minutes as **$2,151,041**, which makes up **8.07%** of revenue from single trips.

Finally, I created another calculated field, **Revenue 60 min. & Under**. My IF statement for this is:

IF [Duration Minutes] >45 and
[Duration Minutes] <= 60 then 7.79
END

The total revenue for trips greater than 45 minutes, up to 60 minutes is **$541,919** which makes up **2.03%** of revenue from single trips.

3. Visualization to show the total amount of flat rate revenue for each hour and each day of the week. Which days/times is Bixi generating the most revenue from their flat rate charge?



The most revenue is being generated mid-week with Tuesday and Wednesday being the most popular. Even more specifically, the late afternoon around 5pm tends to be the busiest and 8am is typically the second busiest time of the day.

This is a time when people are either going to work or getting off work and likely using the bikes to ride to and from their jobs or perhaps to a train or bus station.

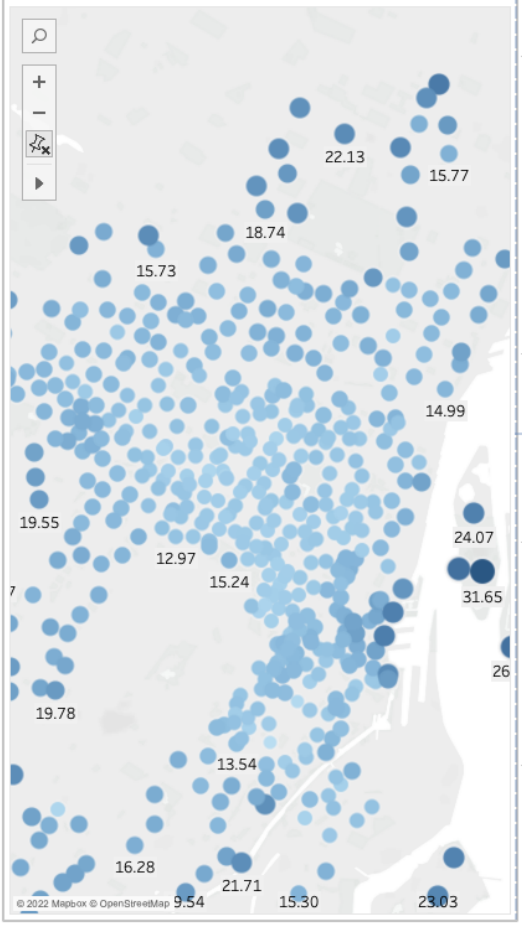
Mid-week are the days that tend to have the most people going into their jobs, whereas Monday, Thursday or Friday are closer to the weekend and those are times where people more frequently take the day off or holidays tend to land on. This can contribute to generating lesser revenue on those days in comparison to Tuesdays and Wednesdays.

Weekends are the lowest revenue producing for single trips further strengthening the belief that single trips are related to people using these bikes to go to and home from work.

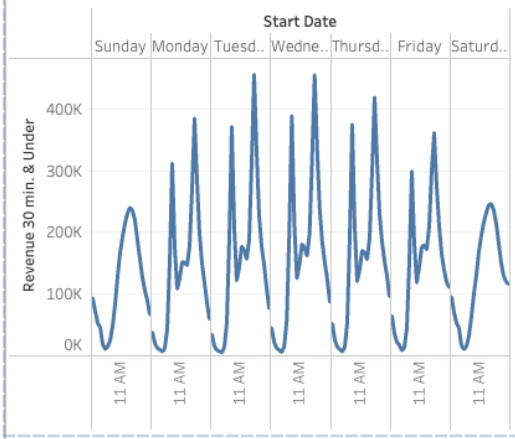
Since the majority of employees have weekends off, they aren't as likely to take single trips or more likely to sign up for memberships when they have the whole day to take a trip on multiple occasions.

4. Create a dashboard containing at least 3 visualizations, using two already created and one additional new one.

2.3 AVG Duration per Station



3.3 Busiest Day/Time of the Week



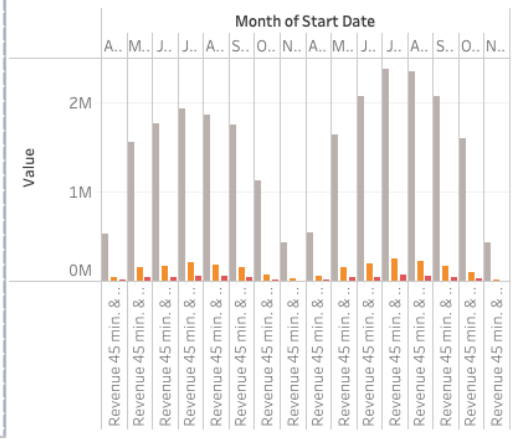
Average Duration by Min.

8.31 — 31.65

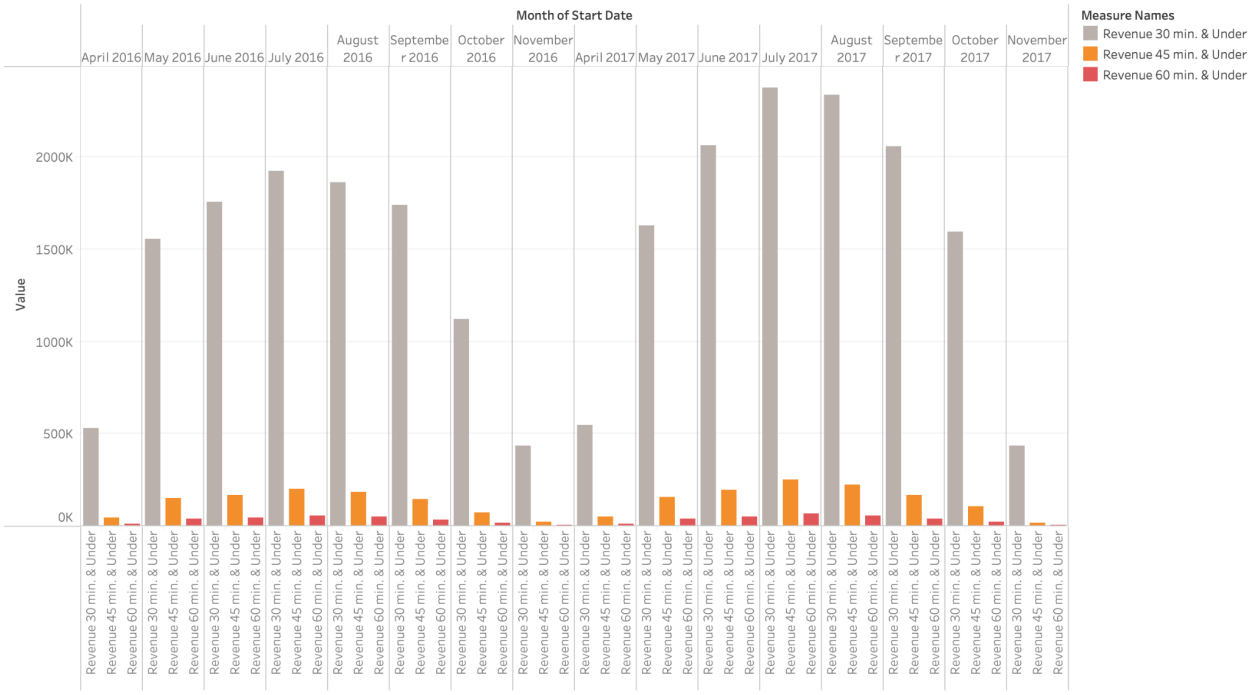
Measure Names

- Revenue 30 min. & U..
- Revenue 45 min. & U..
- Revenue 60 min. & U..

4. Revenue Breakdown by Mo./Yr.



4. Revenue Breakdown by Mo./Yr.



Revenue 30 min. & Under, Revenue 45 min. & Under and Revenue 60 min. & Under for each Start Date Month. Color shows details about Revenue 30 min. & Under, Revenue 45 min. & Under and Revenue 60 min. & Under.

I wanted to create a dashboard that paid attention to helping determine trends that show where and when bixi trips are the most popular.

I used the visualization that displays the map of Montreal with the average duration of trips per station, the graph that shows the revenue for trips lasting 30 minutes or less per each day of the week and a new graph I created that breaks down the revenue by each month of the two different years and how they compare between trips that are under each pricing bucket.

This shows that stations closest to the city center, mid-week and rides under 30 minutes especially during the summer months will generate the most revenue. This information can help the company strategize how they could be charging their rates to maximize value and how to boost frequency during less popular times and in areas around the city that have the most potential to grow.