Total number of trips for 2016 and 2017 is 3917401 (Question 1.1) and 4666765 (Question 1.2), respectively. Monthly breakdown of trips for 2016 (Question 1.3) and 2017 (Question 1.4) is as follows:

|  |  |  |
| --- | --- | --- |
| **Month** | **2016** | **2017** |
| 4 | 189923 | 195662 |
| 5 | 561077 | 587447 |
| 6 | 631503 | 741835 |
| 7 | 699248 | 860732 |
| 8 | 672778 | 839938 |
| 9 | 620263 | 731851 |
| 10 | 392480 | 559506 |
| 11 | 150129 | 149794 |

(Question 1.5) average number of trips per day for each year-month combination is as follows:

|  |  |  |
| --- | --- | --- |
| **Year** | **Month** | **Trip per day** |
| 2016 | 4 | 11870.1875 |
| 2016 | 5 | 18099.2581 |
| 2016 | 6 | 21050.1000 |
| 2016 | 7 | 22556.3871 |
| 2016 | 8 | 21702.5161 |
| 2016 | 9 | 20675.4333 |
| 2016 | 10 | 12660.6452 |
| 2016 | 11 | 10008.6000 |
| 2017 | 4 | 12228.8750 |
| 2017 | 5 | 18949.9032 |
| 2017 | 6 | 24727.8333 |
| 2017 | 7 | 27765.5484 |
| 2017 | 8 | 27094.7742 |
| 2017 | 9 | 24395.0333 |
| 2017 | 10 | 18048.5806 |
| 2017 | 11 | 9986.2667 |

(Question 2.1) Total number of trips for member and non-members is 3784682 and 882083, respectively.

(Question 2.2) Fraction of total trips that were done by members for the year of 2017 broken-down by month is as follows:

|  |  |
| --- | --- |
| **Month** | **Fraction** |
| 4 | 0.8352 |
| 5 | 0.8197 |
| 6 | 0.8081 |
| 7 | 0.7643 |
| 8 | 0.7811 |
| 9 | 0.8258 |
| 10 | 0.8641 |
| 11 | 0.9246 |

Additional experiment:

Two queries was used that lead to the same result. Both queries were based on obtaining the results for total and member trips and then joining them at the end to obtain the fraction. The difference between queries were merely in the method that generates the table total and member trips. The one that used views took 24.2 seconds to run while in the other that used subqueries took 24.7 seconds.

(Question 3.1) Peak demand occurs at month 7 because people like biking in the summer.

(Question 3.2) The best time to offer membership promotion would be month 7 because both number of trips and fraction of the non-member trips is largest (i.e., the fraction of members is the lowest). Therefore, offering promotion is likely to convert more nonmembers to members.

(Question 4.1) Name of the top 5 most popular starting stations are shown below:

|  |  |
| --- | --- |
| **Station Name** | **Trips** |
| Mackay / de Maisonneuve | 97150 |
| Métro Mont-Royal (Rivard / du Mont-Royal) | 81279 |
| Métro Place-des-Arts (de Maisonneuve / de Bleury) | 78848 |
| Métro Laurier (Rivard / Laurier) | 76813 |
| Métro Peel (de Maisonneuve / Stanley) | 72298 |

(Question 4.2) The code without subqueries takes 12.2 seconds while the one with subqueries takes 5.8 seconds. This reduction in computation is because the subqueries help reduce the size of the tables that are going to be joined thus reduce computations related to join operation. The reduction of computation due to joining smaller tables is larger in magnitude than increase in computation due to the additional subquery. Therefore, the total computation is reduced.

(Question 5.1) distribution of number of starts and ends throughout the day is as table below:

|  |  |  |
| --- | --- | --- |
| **Day time** | **# starts** | **# ends** |
| afternoon | 30718 | 30429 |
| evening | 36781 | 31983 |
| morning | 17384 | 26390 |
| night | 12267 | 10326 |

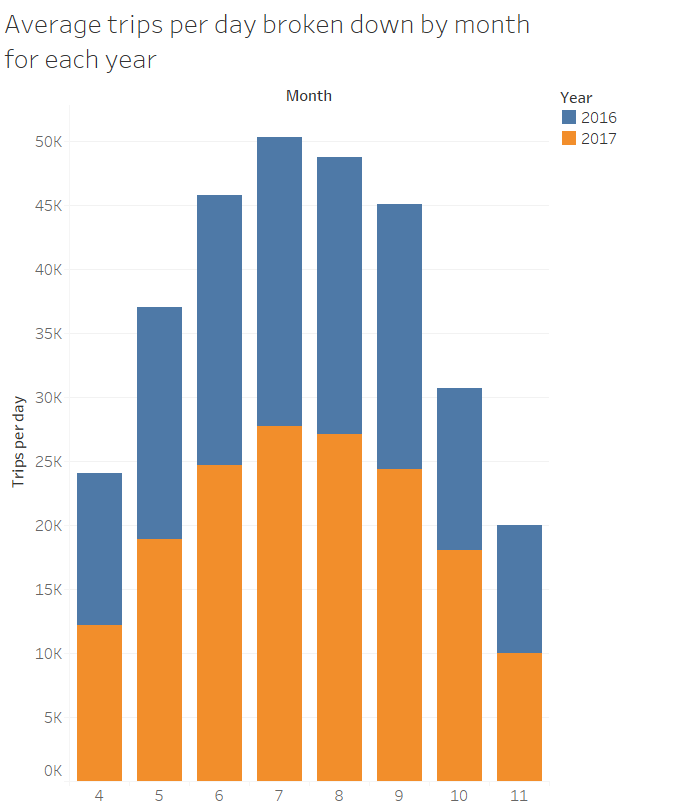
(Question 5.2) results show least trips at night, which is due to cold weather and low visibility. In the morning, there are more trips ending at this station. In the evening, there are more trips starting there than ending. These are because the station is located at downtown where people commute to in the morning for work and from in the evening. Furthermore, as data includes weekends trips, trips in the afternoon and evening are higher than in the morning when people usually sleep in over the weekends.

(Question 6.1 – 6.4) please see the attached SQL queries file for step by step queries develpment

(Question 6.5) Below are top 10 stations with highest fraction of roundtrip to total trip:

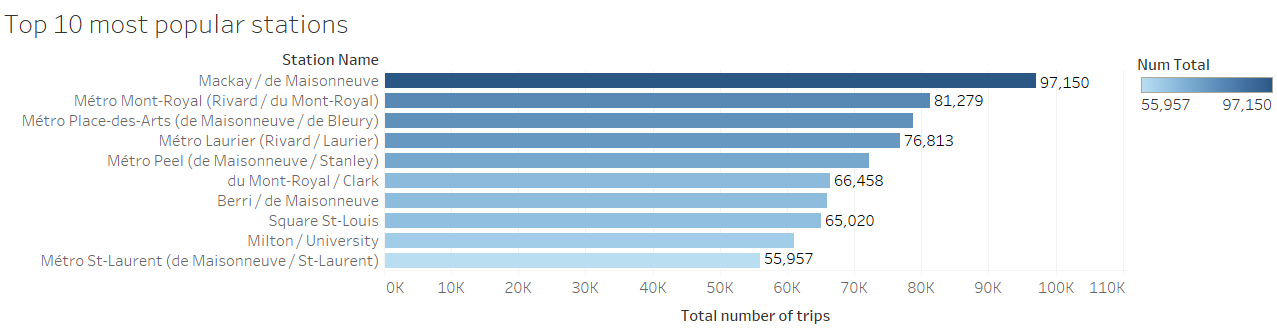
|  |  |
| --- | --- |
| **Station name** | **Fraction** |
| Métro Jean-Drapeau | 0.3020 |
| Métro Angrignon | 0.2331 |
| Berlioz / de l'Île des Soeurs | 0.2043 |
| LaSalle / 4e avenue | 0.2006 |
| Basile-Routhier / Gouin | 0.1932 |
| Parc Plage | 0.1846 |
| Gare Canora | 0.1792 |
| LaSalle / Sénécal | 0.1473 |
| Casino de Montréal | 0.1437 |
| Quai de la navette fluviale | 0.1376 |

(Question 1.1) illustrates how the daily number of trips changed by month in 2016 vs 2017



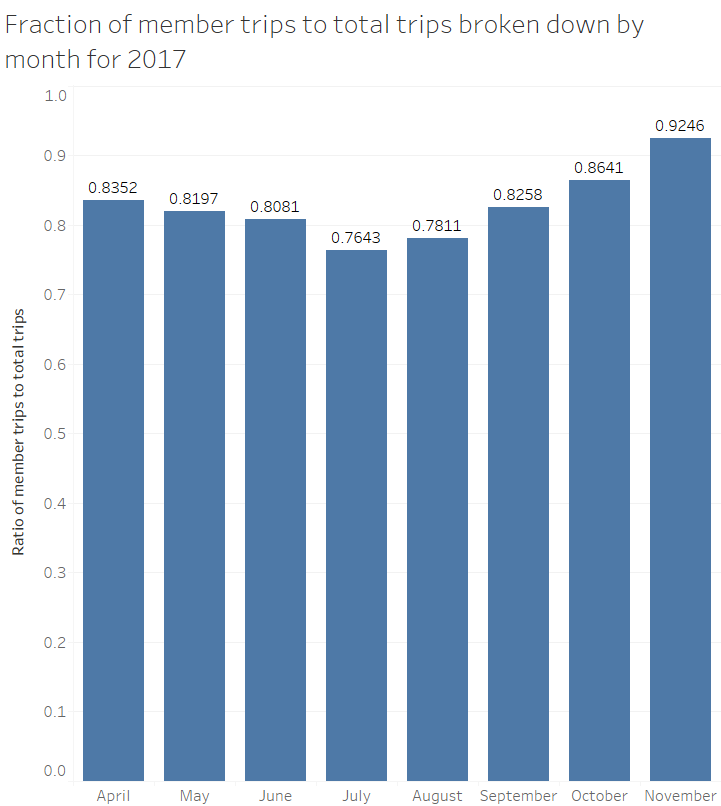
As the emphasis of the question is on the distribution over month, its better to reserve the most effective channel (i.e., the position) for that attribute and move the illustrate the less important years with stack length and color hue.

(Question 1.2) figure below illustrates the top 10 most popular stations.



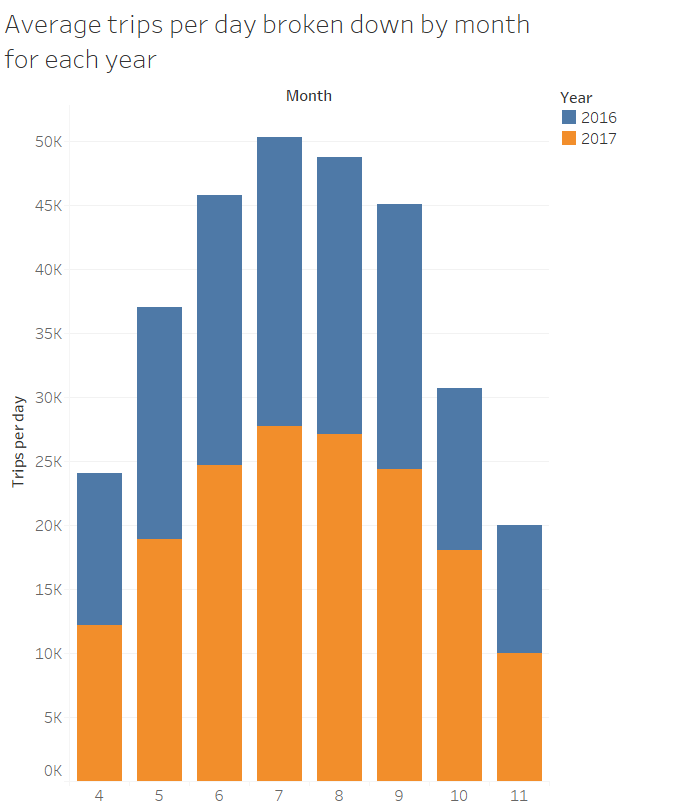
As the difference between numbers of trips for different stations is not that large, it cannot be distinguished in the pie chart. Furthermore, pie chart is based on angles, areas, and colors all of which stand way below length in terms of effectiveness. Color saturation and order is also used in the proposed bar chart to highlight the difference in number of trips. As station names are mostly quite long, horizontal bar chart is preferred to the vertical one to ensure readability of the vertical tick labels.

Figure below shows the fraction member trips to total trips broken down by month for 2017.



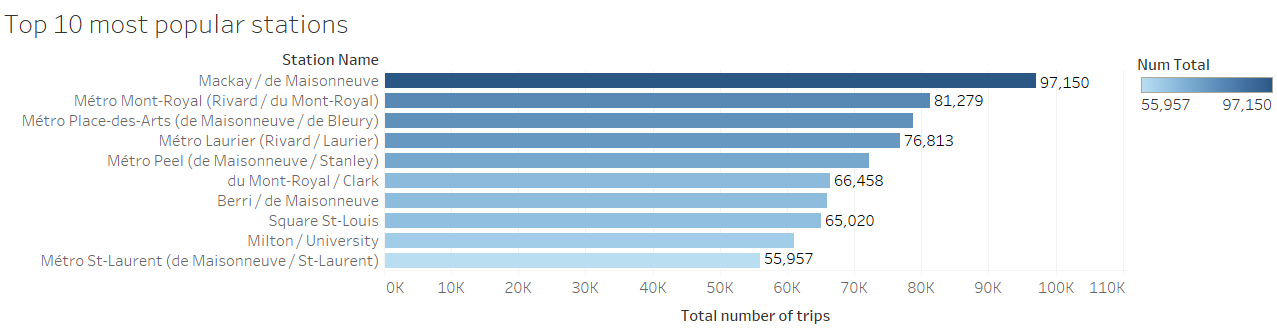
To achieve this result, a mathematical concept is exploited in that the fraction of number of ones to the total number of a rows in a field that only consist of zeros and ones is equal to the average across the entire rows. This concept is very useful in this exercise in that average is an aggregate function in Tableau and it’s only through these aggregate functions that a calculated field can be broken down by groups (e.g. months). The value of each bar is denoted on its top which shows complete consistency with results obtained in part 1.

The following are slides of the story created for this dataset. The first slide illustrates how the daily number of trips changed by month in 2016 vs 2017.

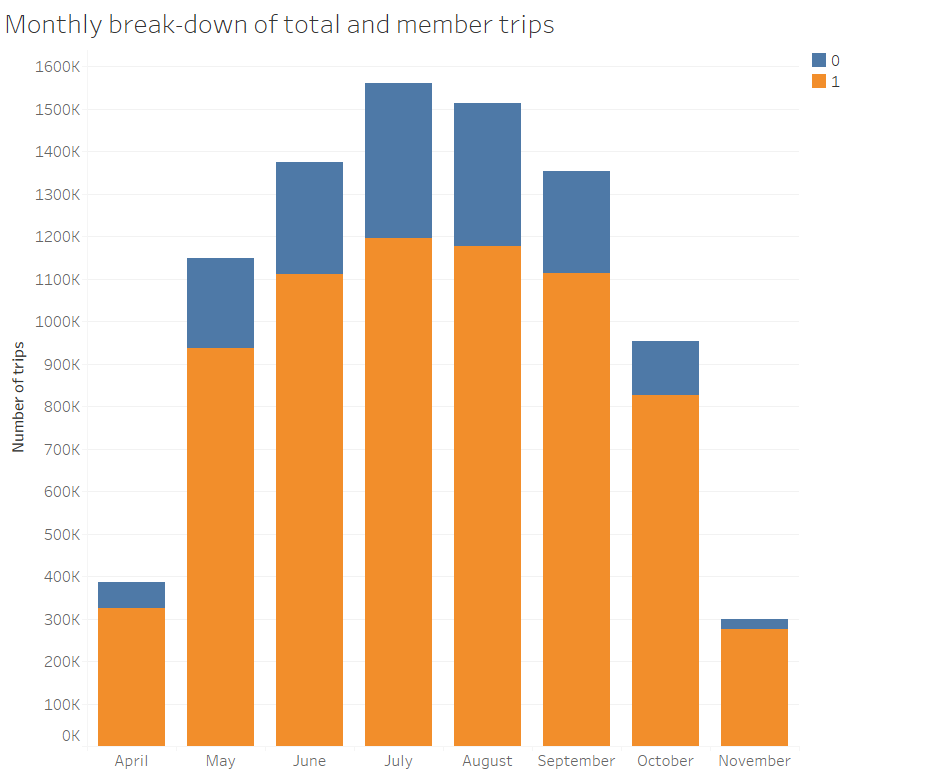


Peak demand is in summer because biking is more fun in those times. The changes in the number of trips for both 2016 and 2017 are similar as total number of trips; however, number of trips in 2017 is slightly larger than 2016.

The second slide illustrates the top 10 most popular stations in order of total trips starting from them:

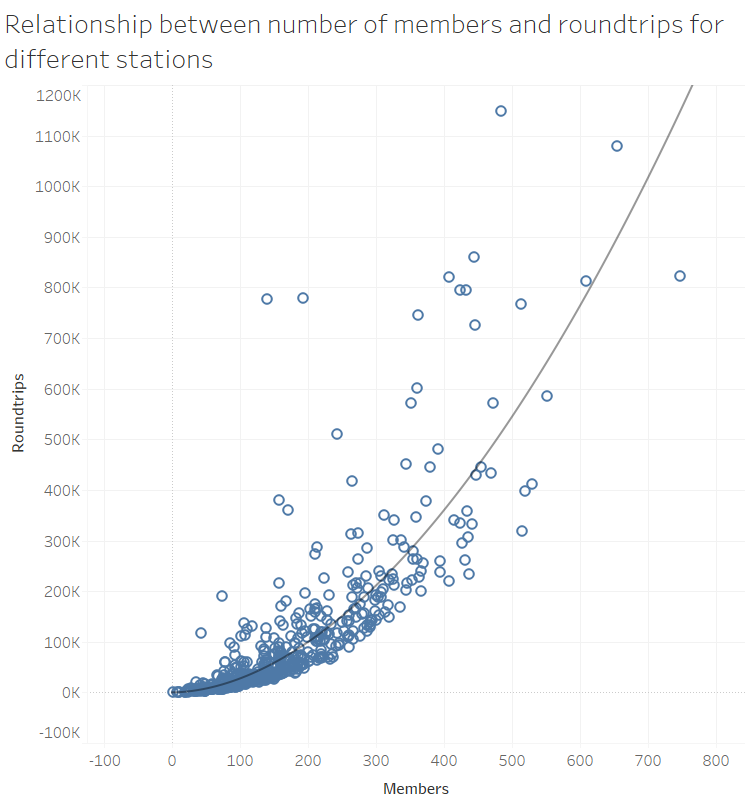


The following slide illustrates the total and member trips broken down

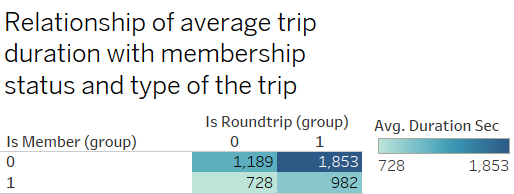


Member trips

The slide below shows the relationship between number of members and roundtrips for different stations. It can be seen that number of roundtrips grows steeper as the number of members become larger. This is due to the fact that members are more likely to pick up a bike and return to the same spot (e.g., near where they live), than a tourist for example. A tourist would use it to get from one attraction to the other, so the bikes ends up at a new station every time. The best model to fit this trend was found to be a power relationship with exponent of 1.86 and coefficient of determination equal to 0.82.



Finally, the last analysis correspond to relationship of average trip duration with membership status and type of the trip as shown below.



To make a fair comparison average duration is chosen as the measure. Results show that in general nonmembers tend to use the bikes for longer durations. Furthermore, the duration of trips is longer in roundtrips for both members and nonmembers.