



## **Bixi Bike Sharing data Analysis with SQL**

**October 6, 2019**

**PREPARED BY:**

MONA VAHDAT

First, we will attempt to gain an overall view of the volume of usage of Bixi Bikes and what factors influence it.

The total number of trips for the years of 2016.

It shows that the total number of trips in 2016 was 3,917,401, while the total number of trips for the years of 2017 was 4,666,765. By breaking down the dataset into months, we can see that the most two months that are popular to use Bixi bikes in 2016 and 2017 was July and followed by August.

It seems that if the company wants to have any promotions for non-members people, these two months would be great time.

The average number of trips a day based on our dataset :

Year	Month	Average_daily_trips
2017	7	27766
2017	8	27095
2017	6	24728
2017	9	24395
2016	7	22556
2016	8	21703
2016	6	21050
2016	9	20675
2017	5	18950
2016	5	18099
2017	10	18049
2016	10	12661
2017	4	12229
2016	4	11870
2016	11	10009
2017	11	9986

The table above illustrates that our conclusion about months is in the proper way

We can expand our analysis through the membership status.

membership_status	2017	total_num_trips_2017
1	2017	3,784,682
0	2017	882,083

membership_status	2016	total_num_trips_2016
1	2016	3174660
0	2016	742741

The tables above show the amount of user has a membership is significantly higher than the other group in both years.

From the table below we can see that the average trip by bike's tripper has the highest amount on November.

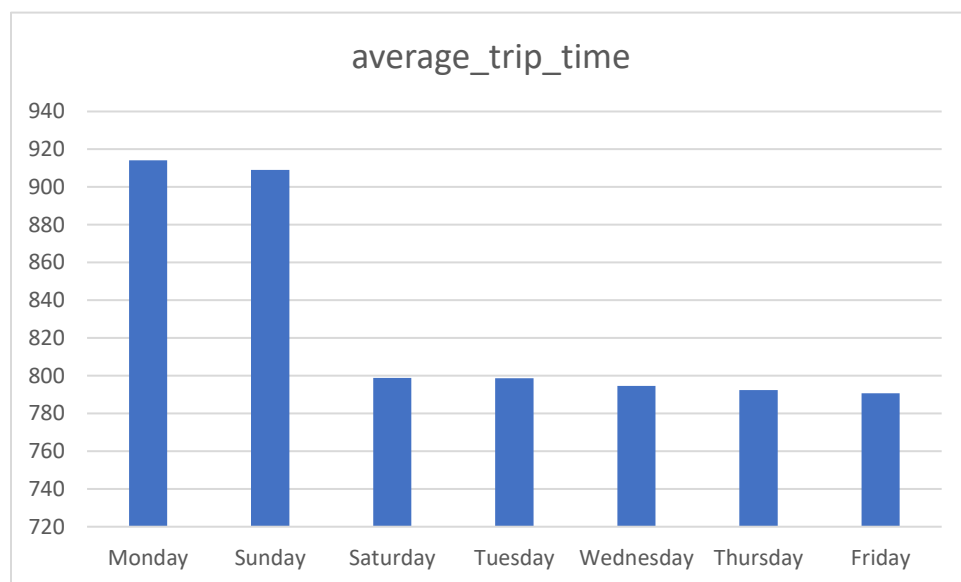
month	average_trips_2017
4	0.8352
5	0.8197
6	0.8081
7	0.7643
8	0.7811
9	0.8258
10	0.8641
11	0.9246

In this part we want to dive into the dataset to calculate the average trip time across a couple dimensions like membership status, months, day of the week and the station name.

Non-memberships have the highest average trip time than memberships.

is_member	average_trip_time
0	1221
1	732

month	average_trip_time
7	879.7803
8	855.783
6	844.4313
5	839.2297
9	803.0599
4	801.7489
10	730.4162
11	654.0278



From the above table we can see that Monday is the highest demand day for bikes followed by Sunday.

Related to the name of stations the results get from dataset:

Station *Métro Jean-Drapeau* has the longest trip on average. Average trip time 1899 seconds.

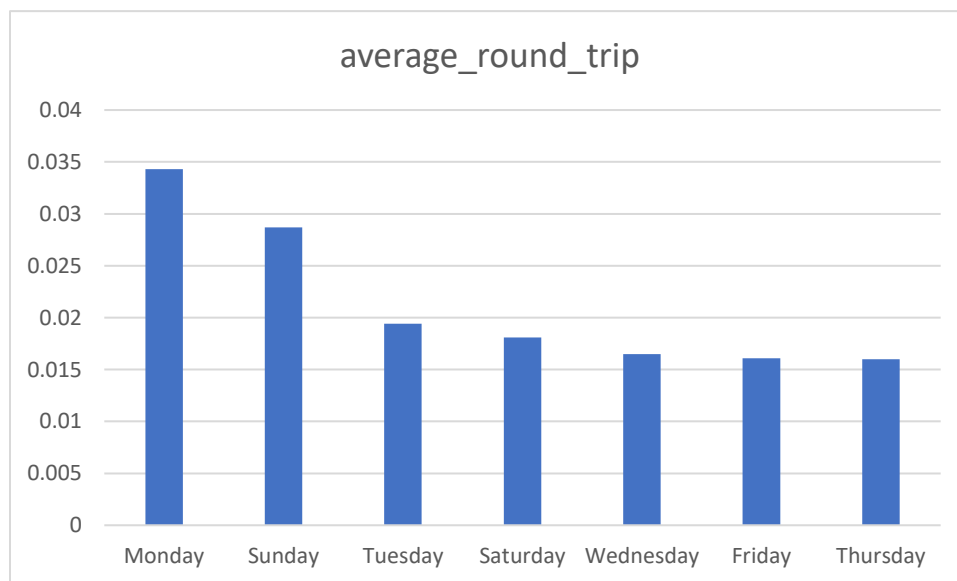
Station *Métro Georges-Vanier (St-Antoine / Canning)* has the shortest trip on average. Average trip time 499 seconds.

**Q:** Short trips can skew your results. How would avoid that?

**A:** use LIMIT to only show the first few rows

At this point we want to know more about round trips which the start point and the end point are the same.

average_round_trip	membership_status
0.0142	1
0.0488	0



From the analysis that we have, it seems that non-members have higher fraction of round trips. We can see non-members generally have longer trip time and also have higher fraction of round trips than members. Also, weekend has the highest trip time as the same as the month of

July. This may be due to the free time and leisure activities on weekend and summer done by non-members.

The five most popular stations at the starting point are listed below:

station_name	total_num_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

Also we have the 5 most popular station at the end point of trips:

Column1	Column2
station_name	total_num_trips
Berri / de Maisonneuve	103720
Mackay / de Maisonneuve	99128
Métro Place-des-Arts (de Maisonneuve / de Bleury)	95343
Métro St-Laurent (de Maisonneuve / St-Laurent)	86886
Métro Peel (de Maisonneuve / Stanley)	76551

How is the number of starts and ends distributed for the station Mackay / de Maisonneuve throughout the day?

station_name	time_of_day	num_trips_as_start_station
Mackay / de Maisonneuve	evening	36781
Mackay / de Maisonneuve	afternoon	30718
Mackay / de Maisonneuve	morning	17384
Mackay / de Maisonneuve	night	12267

The number of trips by this station as start and end station increase from morning until evening and then drops at night. In the morning, more trips are done as ending station, while in the evening most user use it as a start station.

One of the reasons that why this station is most popular may due to the location of this station which is Downtown.

There is another factor that is worth to mention is about the proportionally of the stations.

The *Ste-Catherine / Ste-Élisabeth* station has the least number of member trips as the starting and ending points. While the *Fullum / Gilford* has the most number of members trips.

All in all, the highest average of round trips is from Jean-Drapeau, about 30%, and previously we also found out this station has the longest trip time in average about 32 minutes.