Project 2: BIXI

QUESTION 1:

1.1) The total number of trips for 2016 was: **3,916,401**
1.2) The total number of trips for 2017 was: **4,666,765**

1.3) The total number of trips broken down by month for 2016 and 2017

Months 2016	**Total number of trips 2016**	**Total number of trips 2017**
April	195,653	189,858
May	587,309	561,064
June	741,810	631,380
July	860,807	699,135
August	839,938	672,986
September	731,852	620,259
October	559,557	392,578
November	149,839	150,141

1.4) The average number of trips per day for each year month combination in the database.

YEAR	**MONTH**	**Average number of trips per day**
2016	April	11,866
2016	May	18,099
2016	June	21,046
2016	July	22,553
2016	August	21,709
2016	September	20,675
2016	October	12,664
2016	November	9,384
2017	April	12,228
2017	May	18,945
2017	June	24,727
2017	July	27,768
2017	August	27,095
2017	September	24,395
2017	October	18,050
2017	November	9,365

Question 2:

2.1) The total number of in the year of 2017 broken down my members and non-members

	Total number of members & non-members	fraction of members vs non members
Member	3,784,682	81.10%
Not a member	882,083	18.90%

2.2) The fraction of members in 2017 by month.

2017	Number of members	Fraction of members by month
April	195,653	83.52%
May	587,309	81.98%
June	741,810	80.81%
July	860,807	76.43%
August	839,938	78.10%
September	731,852	82.58%
October	559,557	86.40%
November	149,839	92.45%

Question 3:

3.1) Which time of the year, the demand for Bixit bikes is at it's peak?

Summer Is by far the peak season with demand peaking in the month of July, as can be seen from the data above.

3.2) If you were to offer non-members a special promotion in an attempt to convert them to members, when would you do it?

If our goal was to convert the maximum number of non-members to become members then we would do it in the peak of summer (July & August). July and August is the time of year where you have the most people cycling and where the percentage of non members cycling is at its highest. Nevertheless, I suspect that to attract more customers to use their service in the quieter months, Bixi would be offering a promotion during those times. Given the data, I would not be surprised if in the past they gave special promotions in the months of October and November, hence maybe why there are so many more members relative to non-members in those months.

Question 4:

There are many journeys with null values for the duration they take, thus all journeys with null values within them have been removed from the data.
Any trip that last 1 minute or less has also been removed as it's very unlikely that a bike was actually used within such a short period of time.
Finally all journeys that last longer than 30 minutes were also removed. This is because, according to the Bixi website, non members are charged a fee for using the bike for more than 30 minutes which may skew any future analysis.
After removing this data from the original data set (into a view), there are still 4,824,088 journeys in our data.
All journey times will be dealt with in minutes to simplify our analysis. This shouldn't change our results very much.

4.1) Calculate the entire trip time over the entire dataset.
The average trip time is **7.97 minutes** with our cleaned data.

NOTE: If we were to use the TIMESTAMPDIFF() to calculate the time difference between the start and the end of the trip we would get very different results. The average time in this case would be 13.39 minutes. To understand why there is such a big difference one would have to investigate further with Bixi. For now we will keep on using the data provided in the trip duration column.

Question 5:
5.1) The Average trip time broken-down by membership status is:

Is member	Average trip time
Yes	7.34
No	10.68

5.2) The Average trip time broken-down by Month is:

Month	Average trip time (min)
April	7.74
May	8.08
June	8.15
July	8.41
August	8.23
September	7.83
October	7.21
November	6.54

5.3) The Average trip time broken-down by day of the week is:

Day of the Week	Average trip time
Sunday	8.65
Monday	7.78
Tuesday	7.78
Wednesday	7.75
Thursday	7.74
Friday	7.77
Saturday	8.56

5.4) The Average trip time broken-down by station name (5 longest trip and 5 shortest tirps):
Extremely long and short trips could skew our data. To deal with this we should remove any extreme cases. In the previous step by only taking trip times larger than 1min and shorter than 60 min we have already removed such data that could skew our data.

Locations	Average trip time
Métro Jean-Drapeau	15.42
LaSalle / 4e avenue	15.22
Casino de Montréal	14.77
Basile-Routhier / Gouin	14.77
Parc Plage	14.53
Marquette / du Mont-Royal	6.05
Pontiac / Gilford	5.71
Métro Laurier (Rivard / Laurier)	5.66
Métro Lucien-L\Allier (Lucien l\'Allier / Argyle)'	5.41
Métro Georges-Vanier (St-Antoine / Canning)	5.19

Question 6:

6.1) Fraction of round trips broken down by membership status:

Member	Percentage of round trips
Yes	1.13%
No	3.53%

6.2) Fraction of round trips broken down by day of the week:

Day of the week	Percentage of round trips
Sunday	2.71%
Monday	1.47%
Tuesday	1.28%
Wednesday	1.23%
Thursday	1.22%
Friday	1.37%
Saturday	2.17%

Question 7:

These finding are not so surprising. Commuters, and other regular riders that use the service to get between point a and point b are more likely to be members as they use the bikes on a regular basis. A regular rider that often uses the service will have a much larger economic incentive to be a member than someone who doesn't often use the service. Secondly, we can assume that occasional leisure riders are much less likely to be members and are also more likely just to use the service for a round trip visiting the city.

With regards to the day of the week, the reasons are similar to the first part of this question. On weekends there is a clear increase in round trips. This is most probably due to there being less people using the service for commuting and people have more time to go for a bike ride around the park. Additionally, there are more people riding bikes of leisure on weekends than on week days.

Question 8:
The 5 most popular starting stations:

Starting trip Locations	Number of trips
Mackay / de Maisonneuve	55,118
Métro Mont-Royal (Rivard / du Mont-Royal)	46,249
Métro Place-des-Arts (de Maisonneuve / de Bleury)	45,695
Métro Laurier (Rivard / Laurier)	44,030
Métro Peel (de Maisonneuve / Stanley)	40,784

Question 9:
The 5 most popular ending stations:

Ending trip Locations	Number of trips
Berri / de Maisonneuve	60,721
Mackay / de Maisonneuve	56,539
Métro Place-des-Arts (de Maisonneuve / de Bleury)	55,152
Métro St-Laurent (de Maisonneuve / St-Laurent)	49,337
Métro Peel (de Maisonneuve / Stanley)	43,320

The 2 tables are quite similar. 5 out of the top starting stations are also 5 of the most popular ending stations.

Question 10:

10.1) The number of starts stations distributed throughout the day for all stations. This is nearly exactly the same for end stations, not surprisingly.
Morning: 7am – 11am
Afternoon: 12pm – 4pm
Evening: 4pm – 9pm

	Fraction of trips	Total number of trips
morning	24.23%	1,168,881
afternoon	30.48%	1,470,320
evening	33.49%	1,615,782
night	11.80%	569,105

10.2) The number of starts and ends distributed throughout the day for Mackay / de Maisonneuve

	Fraction starting trips	Fraction ending trips
morning	17.56%	26.84%
afternoon	31.69%	30.58%
evening	38.16%	32.30%
night	12.59%	10.28%

10.3) Explain the differences you see and discuss why the numbers are the way they are?

First of all when looking at the averages over the whole database it's interesting to see that the bikes are much more popular in the afternoon and evenings than in the morning. This must be due to the mornings often being a little colder and maybe due to the fact people have more time in the evening than in the morning.

In the second table we see that Mackay / de Maisonneuve is most probably in the city centre with all the offices. This is because it has a large number of cyclists ending at that station in the morning whiles a low level of starting trips. In the evenings it has a high number of starting trips from there whiles relatively less ending trips to that station. Checking it on google maps confirms our assumptions with RBC Royal Bank just next door.

Its Average trip time for Mackay / de Maisonneuve is at 7.54 minutes which is at the lower end of the spectrum.

Question 11:

Which station has proportionally the least number of member trips? Remove all stations with less than 10 starting or ending stations.

1.1) The minimum number of starting trips from any given station is 283 and the minimum number of ending trips from any given station is 406 trips. Given this is well above 10 we don't need to remove any station from our database.

NOTE: In mySQL I have created a view called stations_10_trips which only stores stations with more than 10 trips. If in our data there were any stations with less than 10 trips we would use this view instead of using the stations table.

1.2) The stations with the least and most member trips proportionally as starting trips are the following:

Starting trip Locations	Fraction of members
du Mont-Royal / Augustin-Frigon	92.23%
Hôpital Maisonneuve-Rosemont (Rosemont / Chatelain)	91.89%
Belmont / du Beaver Hall	91.80%
Berri / St-Grégoire	91.77%
10e Avenue / Rosemont	91.76%
Casino de Montréal	33.77%
La Ronde	32.65%
Parc Plage	26.74%
Métro Jean-Drapeau	26.28%
Quai de la navette fluviale	25.15%

1.3) The stations with the least and most member trips proportionally as ending trips are the following:

Ending trip Locations	Fraction of members
du Mont-Royal / Augustin-Frigon	91.99%
Square Victoria	91.50%
de Gaspé / Fairmount	91.37%
Marmier / St-Denis	91.33%
Hôpital Maisonneuve-Rosemont (Rosemont / Chatelain)	91.31%
Casino de Montréal	38.77%
La Ronde	36.92%
Parc Plage	28.94%
Quai de la navette fluviale	25.45%
Métro Jean-Drapeau	25.34%

NOTE: The average fraction of members, by location is 81.9%.

Question 12:

12.1, 12.2, 12.3) The fraction of round trips by starting station with at least 10% of trips starting from them are:

Starting Trip Locations	% of Round Trips	Nbr of Round Trips	Nbr of Trips
Métro Jean-Drapeau	26.35%	3,930	14,916
Métro Angrignon	22.48%	294	1,308
Parc Plage	18.68%	659	3,527
Gare Canora	16.68%	238	1,427
LaSalle / 4e avenue	16.39%	258	1,574
Berlioz / de l\Île des Soeurs'	16.32%	488	2,990
Basile-Routhier / Gouin	14.88%	142	954
Casino de Montréal	12.64%	418	3,308
Quai de la navette fluviale	12.56%	444	3,535
LaSalle / Sénécal	11.14%	183	1,643
CHSLD Éloria-Lepage	10.60%	30	283

12.4) Once again, all stations have well above 50 starting trips so we will not need to filter out any stations in our analysis.

12.5) From the table above there seems to be a clear correlation between the % of round trips and places which are associated with leisure. The top 3 starting trip locations in terms of % of round trips are all parks. Metro Jean-Drapeau, has several tourist attractions within the park and it tops off the list in first place. Additionally, stations with high round trips seem to also have lower level of members as can be seen with Metro Jean-Drapeau having one of the lowest number of members using it's station. The opposite seems to hold too where business office areas, who thus attract commuters seem to have a lower level of round trips and a much higher level of members such as Mackay / de Maisonneuve with only 1.3% of round trips.

Once again we can expect people who use the bikes for pleasure rather than as a means to travel/commute, are more likely to be using the bikes on weekend, spend more time per journey and be more affected by weather (colder and warmer time of year). The above analysis we have done seems to corelate with our assumption, with a few examples that really stand out such as Métro Jean-Drapeau, Parc Plage, and Mackay / de Maisonneuve.