

## Bixi Project Deliverable 1

The goal is to gain a high level understanding of how people use Bixi bikes, what factors influence the volume of usage, popular stations, and overall business growth.

Following are a number of questions asked about the performance of the company keeping specific metrics in mind.

### QUESTION 1

1. The total number of trips for the year of 2016. - **3917401**

2. The total number of trips for the year of 2017 - **4666765**

3. The total number of trips for the year of 2016 broken down by month.

APR	189923
MAY	561077
JUN	631503
JUL	699248
AUG	672778
SEP	620263
OCT	392480
NOV	150129
<b>TOTAL</b>	<b>3917401</b>

4. The total number of trips for the year of 2017 broken down by month.

APR	195662
MAY	587447
JUN	741835
JUL	860732
AUG	839938
SEP	731851
OCT	559506
NOV	149794
<b>TOTAL</b>	<b>4666765</b>

5. The average number of trips a day for each year-month combination in the dataset.

YEAR	MONTH	ACTIVE DAYS	AVG. TRIPS PER DAY
2016	APR	16	11870.19
2016	MAY	31	18099.26
2016	JUN	30	21050.10
2016	JUL	31	22556.39
2016	AUG	31	21702.52
2016	SEP	30	20675.43
2016	OCT	31	12660.65
2016	NOV	15	10008.60
2017	APR	16	12228.88
2017	MAY	31	18949.90
2017	JUN	30	24727.83
2017	JUL	31	27765.55
2017	AUG	31	27094.77
2017	SEP	30	24395.03
2017	OCT	31	18048.58
2017	NOV	15	9986.27

6. Save your query results from the previous question (Q1.5) by creating a table called `working_table1`.

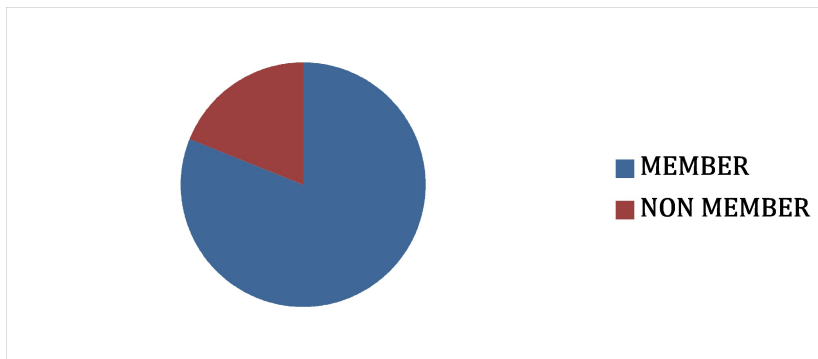
A. Please check SQL code for `working_table1`

## QUESTION 2

1. The total number of trips in the year 2017 broken down by membership status (member/non-member).

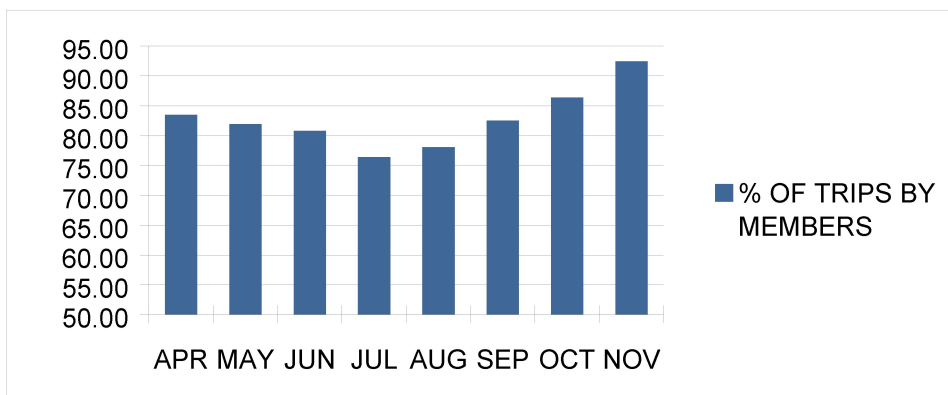
MEMBER STATUS	TRIPS
MEMBER	3784682
NON-MEMBER	882083

A majority of trips in 2017 were taken by members



2. The percentage of total trips by members for the year 2017 broken down by month.

MONTH	MEMBER TRIPS	TOTAL	% OF TRIPS BY MEMBERS
APR	163417	195662	83.52
MAY	481540	587447	81.97
JUN	599509	741835	80.81
JUL	657865	860732	76.43
AUG	656049	839938	78.11
SEP	604358	731851	82.58
OCT	483445	559506	86.41
NOV	138499	149794	92.46

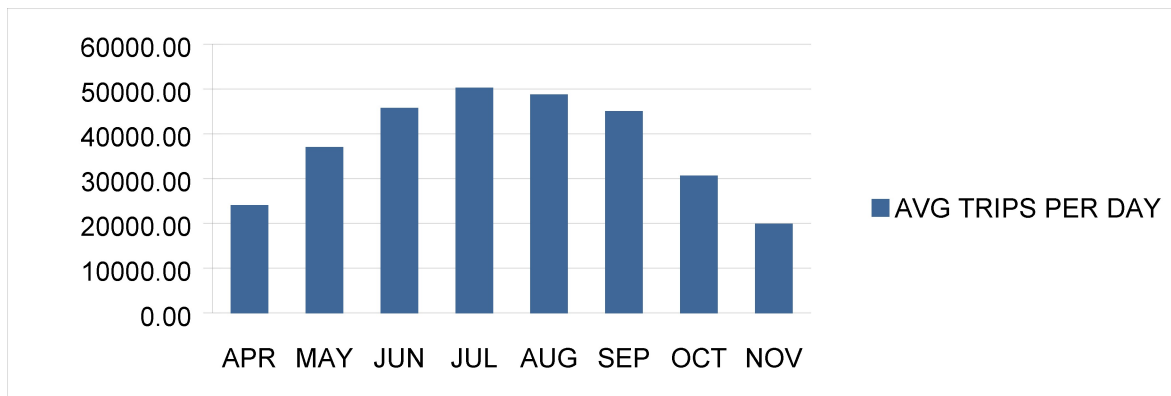


The data indicates that as the temperature starts to drop, non-members are less likely to use the service.

### QUESTION 3

1. At which time(s) of the year is the demand for Bixi bikes at its peak?

MONTH	AVG TRIPS PER DAY
APR	19994.87
MAY	24099.06
JUN	30709.23
JUL	37049.16
AUG	45070.47
SEP	45777.93
OCT	48797.29
NOV	50321.94



Peak demand occurs in the warmer months of June, July and August.

2. If you were to offer non-members a special promotion in an attempt to convert them to members, when would you do it? Describe the promotion and explain the motivation and your reasoning behind it.

**A.** The average ridership drastically declines in the colder months, but also during the warmer times there is peak activity for non members riding with bixi.

This signifies that temperature is a major factor in bixi usage.

Members continue to use the service even in the colder months.

When a non-member is regularly using the service (warmer months), it makes sense to push the season membership at a reduced cost (probably 25%) as it might push the non member to ride more if he has the service.

Also monetarily speaking, It will secure subscription fees for more people (irrespective of whether they ride in the colder months or not)

## QUESTION 4

1. What are the names of the 5 most popular starting stations? Determine the answer without using a subquery.

Without using a subquery the query ran in **11.485 s**

STATION CODE	STATION NAME	TRIPS
6100	Mackay / de Maisonneuve	97150
6184	Métro Mont-Royal (Rivard / du Mont-Royal)	81279
6078	Métro Place-des-Arts (de Maisonneuve / de Bleury)	78848
6136	Métro Laurier (Rivard / Laurier)	76813
6064	Métro Peel (de Maisonneuve / Stanley)	72298

2. Solve the same question as Q4.1, but now use a subquery. Is there a difference in query run time between 4.1 and 4.2? Why or why not?

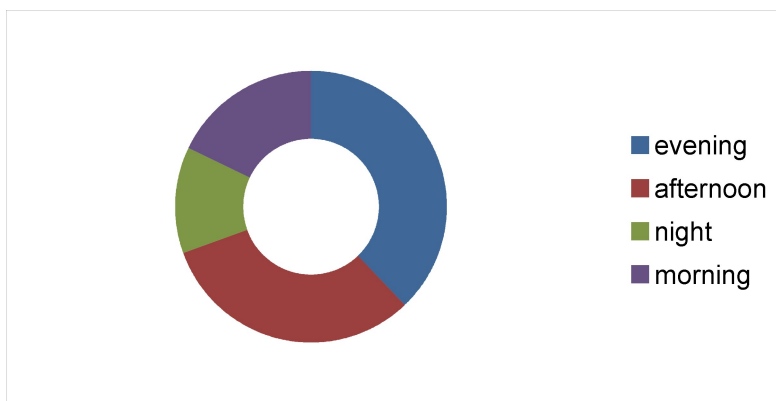
Using a subquery the run time for the query goes down to **2.546 s**  
(down by 78 %)

This is probably because after using the subquery, SQL has to parse through a much smaller set of data - hence speeding operations.

## QUESTION 5

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

TIME OF DAY	TRIPS	%
evening	36781	37.86
afternoon	30718	31.62
night	12267	12.63
morning	17384	17.89



2. Explain and interpret your results from above. Why do you think these patterns in Bixi usage occur for this station? Put forth a hypothesis and justify your rationale.

**A. 68% of all trips happen in the afternoon and in the evenings.**

This suggests that Mackay might be a station close to an entertainment hub or more of a place where **people might go after work to explore on bike** and relax.

## QUESTION 6

List all stations for which at least 10% of trips are round trips. Round trips are those that start and end in the same station. This time we will only consider stations with at least 500 starting trips. (Please include answers for all steps outlined here)

1. First, write a query that counts the number of starting trips per station.

A. Only the first 5 (of 540) entries (based on no of trips taken in descending order) are shown here in the interest of saving space.

STATION CODE	TRIPS
6100	97150
6184	81279
6078	78848
6136	76813
6064	72298

2. Second, write a query that counts, for each station, the number of round trips.

A. Only the first 5 (of 540) entries (based on no of round trips taken per station in descending order) are shown here in the interest of saving space.

STATION CODE	ROUND TRIPS
6501	8658
6026	5622
6036	4123
6023	2591
6050	2182

3. Combine the above queries and calculate the fraction of round trips to the total number of starting trips for each station.

A. Only the first 5 (of 540) entries (based on the % of round trips taken per station in descending order) are shown here in the interest of saving space.

STATION CODE	STATION NAME	ROUND TRIPS	TOTAL TRIPS	% ROUND TRIPS
6501	Métro Jean-Drapeau	8658	28672	30.20
7048	Métro Angrignon	559	2398	23.31
	Berlioz / de l'Île des			
6428	Soeurs	1072	5246	20.43
7015	LaSalle / 4e avenue	600	2991	20.06
	Basile-Routhier /			
6736	Gouin	330	1708	19.32

4. Filter down to stations with at least 500 trips originating from them and having at least 10% of their trips as round trips.

A. Below are the 14 results of the query that satisfy the asked conditions,

STATION CODE	STATION NAME	ROUND TRIPS	TOTAL TRIPS	% ROUND TRIPS
6026	de la Commune / Place Jacques-Cartier	5622	50822	11.06
6502	Casino de Montréal	882	6138	14.37
6736	Basile-Routhier / Gouin	330	1708	19.32
6501	Métro Jean-Drapeau	8658	28672	30.20
6429	Place du Commerce	927	8569	10.82
5006	Collège Édouard-Montpetit	144	1439	10.01
6109	Quai de la navette fluviale	883	6417	13.76
6428	Berlioz / de l'Île des Soeurs	1072	5246	20.43
6714	LaSalle / Sénécal	464	3151	14.73
6016	Jacques-Le Ber / de la Pointe Nord	300	2719	11.03
6359	Parc Plage	1145	6201	18.46
7048	Métro Angrignon	559	2398	23.31
7015	LaSalle / 4e avenue	600	2991	20.06
7007	Gare Canora	437	2439	17.92



5. Where would you expect to find stations with a high fraction of round trips? Describe why and justify your reasoning.

A. Based on the highlighted items from the above table and a few google searches for the station names, It can be inferred that the above stations with a high number of round trips starting from them are **almost all places of commerce.**

This means that users **used other means to get to this station, used a bike from this station, finished their specific work and returned the bike to this location.**

They **could also be metro stations** as people might be using the metro to get to the station and then using a bike to get to a specific location and returning the bike to the metro location in the same trip.