

Data Analysis - Bixi Project Part 1

by Allan Isla

Question 1

1. The total number of trips for the year of 2016.

3,917,401

2. The total number of trips for the year of 2017.

4,666,765

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

April - 189,923

May - 561,077

June - 631,503

July - 699,248

August - 672,778

September - 620,263

October - 392,480

November - 150,129

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

April - 195,662

May - 587,447

June - 741,835

July - 860,732

August - 839,938

September - 731,851

October - 559,506

November - 149,794

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

2016

April - 11,870

May - 18,099

June - 21,050

July - 22,556

August - 21,703

September - 20,675

October - 12,601

November - 10,009

2017

April - 12,229

May - 18,950

June - 24,728

August - 27,095

September - 24,395

October - 18,049

July - 27,766

November - 9,986

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

I created a table in MySql and used the same query from the previous answer.

Question 2

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

Members - 3,784,682

Non-members - 882,083

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

April - 83.5%

August - 78.1%

May - 82%

September - 82.6%

June - 81%

October - 86.4%

July - 76.4%

November - 92.5%

Question 3

1. Which time of the year the demand for Bixi bikes is at its peak?

July is the month with the most demand. The summer months of June, July and August are the 3 most popular months of the year.

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

I would offer the promotions during the least popular months of the year, which are April, May, October and November.

These months are right before summer begins and right after summer ends. People tend to want to ride bikes when the weather is warmest, so promotions during these months can incentivize people to become members during colder months.

Question 4

1. What are the names of the 5 most popular starting stations?

1. Mackay / de Maisonneuve
2. Metro Mont-Royal (Rivard / du Mont-Royal)
3. Métro Place-des-Arts (de Maisonneuve / de Bleury)
4. Metro Laurier (Rivard / Laurier)
5. Métro Peel (de Maisonneuve / Stanley)

2. Is there a difference in query run time between 4.1 and 4.2?

The run time for the query on answer 4.1 takes over 5 seconds while the query for answer 4.2 takes about 1 second.

Answering the question with the subquery speeds up the run time approximately 5x times faster.

Question 5

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

Starts	Ends
Evening - 36,781	Evening - 31,983
Afternoon - 30,718	Afternoon - 30,429
Morning - 17,384	Morning - 26,390
Night - 12,267	Night - 10,326

2. Explain the differences you see and discuss why the numbers are the way they are.

The evening hours are the most popular times for starting and ending rides throughout the day. This is usually a time people get off work and likely ride home or ride to get exercise.

I noticed that this station has more trips starting rather than ending for each time of the day with the exception being in the morning. This is the only part of the day where

trips ending outnumber trips starting at this station. This is likely due to it being close to downtown where people are often arriving here to get to work.

Night time, as expected, is the least busiest time likely due to it being too dark to ride and people are typically asleep.

Question 6

1. Count the number of starting trips per station.

I used the same query here that I used to answer question 4.1. To find the answer, the key is joining the trips tables and stations tables using the trips starting station code and the stations code.

2. Count the number of round trips for each station.

I used the same query as the previous answer. To figure out round trips I made sure my query added the WHERE clause to include trips where the start station code and end station code matched.

3. Calculate the fraction of round trips to the total number of starting trips for each station.

I used the same query to answer 6.1 as my FROM subquery, but included the stations.code in the SELECT line. I used that code to join my subquery to the trips table to help match up the starting stations to the number of roundtrips they each had.

Then from there, I was able to divide the amount of roundtrips from the total amount of trips that started at each station. I multiplied the answer to 100, so I could get a percentage to make it look more clear.

4. Filter down to only stations that had at least 500 trips starting from them and had at least 10% of their trips as roundtrips.

I used the same query that I used to answer the previous question, I just added the HAVING clause to filter and found only 14 results that met this criteria.

5. Where would you expect to find stations with a high fraction of round trips?

I did a google search of the stations that were listed in the 14 results I found in the previous question.

I noticed that many of these bike stations are next to train or bus stations located inside or nearby a park. People who often plan to go to a park, ride a train or bus that takes them there.

Then they find a bike station that is conveniently close by and rent a bike to ride around the park for leisure or exercise, not necessarily to go anywhere and typically return it to the same place they found it, so they can get back on the train or bus to venture off to their next destination.