BRAINSTATION
DATA SCIENCE FULL TIME COURSE

COLLECT AND STORE PROJECT: BIXI

STUDENT NAME: SEBASTIAN LANGER
DATE: APRIL 7, 2019

Table of Contents

Answers

1. Usage Volume Overview

PLEASE READ FIRST: Before continuing, execute the code under the section “PERFORMANCE IMPROVEMENTS 1” in the .sql file.

Part 1

1.1.1 The total number of trips for the years of 2016.

The number of trips in 2016 was 3917401

1.1.2 The total number of trips for the years of 2017.

The number of trips in 2017 was 4666765

1.1.3 The total number of trips for the years of 2016 broken-down by month.

Please, execute the query in the section 1.1.3 in the .sql file.

1.1.4 The total number of trips for the years of 2017 broken-down by month .

Please, execute the query 1.1.4 in the section in the .sql file.

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

Please, execute the queries in the section 1.1.5 in the .sql file.

Part 2

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

Please, execute the in the section query 1.2.1 in the .sql file.

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

Please, execute the query in the section 1.2.2 in the .sql file.

Part 3

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

The best 3 months are June, July and August, which is summer in Montreal

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

I would launch special promotions in the first two months and the last two months of the biking season after and before the snow.
That would be April and May and October and November.
Those are the lower months in terms of trips, and it's probably a good idea to encourage people to use the service lowing the prices.
I would like to know what happens with the bikes during the December, January and February and March.

My advice should be to explore the possibility of offering the service in warmer cities at the best price of the market which is better than leave them in storage with them during that time.

2. Trip Characteristics

2.1 Calculate the average trip time across the entire dataset.

The average trip time is 824 seconds or 13.74 minutes.

2.2 Let's dig a bit deeper and slice the average trip time across a couple of interesting dimensions. Calculate the average trip time broken-down by:

2.2.1 Membership status

The average trip for Members is 732 seconds or 12.20 minutes.
The average trip for Non-Members is 1221 seconds or 20.35 minutes.
Non-Members make trips of 37% more distance on average.

2.2.2 Month

Please, execute the in the section query 2.2.2 in the .sql file.

As an observation, we can see that the months July, August, and June (in that order) are the months with the most extended average trips.

2.2.3 Day of the week

Please, execute the in the section query 2.2.3 in the .sql file.

As an observation, we can see that Sundays and Saturdays (in that order) are the days with the most extended average trips.

2.2.4 Station name

Before continuing, execute the code under the section “PERFORMANCE IMPROVEMENTS 2” in the .sql file.

To get the information required for this question, please, execute the in the section query 2.2.4 in the .sql file.

2.2.4.1 Which station has the longest trips on average?

The **start station** with the longest trips on average is Métro Jean-Drapeau (code 6501) with 1899.16 seconds or 31.65 minutes.

The **end station** with the most extended trips on average is ALSO Métro Jean-Drapeau (code 6501) with 1941 seconds or 32.35 minutes

2.2.4.2 Which station has the shortest trips on average?

The **start station** with the shortest trips on average is Métro Georges-Vanier (St-Antoine / Canning) (code 6408) with average rides of 499 seconds or 8.31 minutes.
The **end station** with the shortest trips on average is also Métro Georges-Vanier (St-Antoine / Canning) (code 6408) with average rides of 544 seconds or 9.06 minutes.

2.2.4.3 Extremely long / short trips can skew your results. How would avoid that?

The general answer that kind of problem is to filter those values in the WHERE clause establishing some range and excluding exceptional values both the highest and lowest extremes.
In this case, if we group the trips per quantity of minutes, we find out that there are trips between 1 minute to 119 and. In all those observations the number of trips it's more or less significant, so I wouldn't eliminate any observation.

Please, execute the query in the section 2.2.4.3 in the .sql file to see that.

2.3 Let's call trips that start and end in the same station "round trips". Calculate the fraction of trips that were round trips and break it down by:

2.3.1 Membership status

Non-members seem to make roundtrips more often than Members; the fraction is 0,0488 for non-members and 0.014 from Members. Nevertheless, the numbers of trips made by the members it's significantly higher than not members in quantity with 6959342 for members and 1624824 for non-members.

Please, execute the query in the section 2.3.1 in the .sql file to see that.

2.3.2 Day of the week

Sundays and Saturdays seem to be the days with more roundtrips

Please, execute the query in the section 2.3.2 in the .sql file to see that.

2.4 Discuss the differences you observed and come up with possible explanations.

Non-members use the bicycle for much longer on average, although there are many more trips made by members. The most popular months are those that are far from the cold, in the summer and the most popular days are Saturdays and Sundays.
That is, rest is synonymous with going for a ride on a bicycle.
Longer trips are made much more often from and to the Métro Jean-Drapeau station and the shorter trips to and from Métro Georges-Vanier (St-Antoine / Canning).
You might think that non-members are tourists who are passing by and the members are local people who use the service to move around the city comfortably. With this in mind, we

might think Métro Jean-Drapeau is popular among tourists because there may be exciting attractions in the area and the Métro Georges-Vanier station (St-Antoine / Canning) is more popular amongst local people, and It may be a favorite place to go to do daily tasks.

3. Popular Stations

3.1 What are the names of the 5 most popular starting stations?

- Mackay / de Maisonneuve
- Métro Mont-Royal (Rivard / du Mont-Royal)
- Métro Place-des-Arts (de Maisonneuve / de Bleury)
- Métro Laurier (Rivard / Laurier)
- Métro Peel (de Maisonneuve / Stanley)

Please, execute the query in the section 3.1 in the .sql file to see that.

3.2 What are the names of the 5 most popular ending stations?

- Berri / de Maisonneuve
- Mackay / de Maisonneuve
- Métro Place-des-Arts (de Maisonneuve / de Bleury)
- Métro St-Laurent (de Maisonneuve / St-Laurent)
- Métro Peel (de Maisonneuve / Stanley)

Please, execute the query in the section 3.2 in the .sql file to see that.

3.3 If we break-up the hours of the day as follows: morning, afternoon, evening, night

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

Please, execute the query in the section 3.3.1 in the .sql file to see that.

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

This station is a place where the people arrive mostly in the morning, maybe to go to work.
In the evening and at night this station is more popular as a place of departure, maybe when people are coming back home from their activities.
In the afternoon this place has a similar fraction of people using it both as a place of departure and arrival.

3.4 (A) Which station has proportionally the least number of member trips? (B)How about the most? To damper variance, consider only stations for which there were at least 10 trips starting and ending from it.

- A
    - The **start station** with fewer trips of members is CHSLD Éloria-Lepage (de la Pépinière / de Marseille) with 379
    - The **end station** with fewer trips of members CHSLD Benjamin-Victor-Rousselot (Dickson / Sherbrooke) with 558

- B
    - The **start station** with more trips of members is Mackay / de Maisonneuve with 80538
    - The **end station** with more trips of members is Berri / de Maisonneuve with 83453

3.5 List all stations for which at least 10% of trips are round trips. Recall round trips are those that start and end in the same station. This time we will only consider stations with at least 50 starting trips.

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

Please, execute the query in the section 3.4.1 in the .sql file.

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

Please, execute the query in the section 3.5.2 in the .sql file.

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

Please, execute the query in the section 3.5.3 in the .sql file.

3.5.4 Filter down to stations with at least 50 trips originating from them.

There are 129 stations with more than at least 0.1 round-trip trips.

Please, execute the query in the section 3.5.4 in the .sql file to see that.

3.5.5 Given what we learned above about the relation between round trips, membership status, and day of the week, where would you expect to find stations with a high fraction of round trips?

- For non-members If we take the 4 observations most relevant
    - The station de la Commune / Place Jacques-Cartier on Saturday with 8185 and Sunday with 7745
    - The station de la Commune / St-Sulpice on Saturday with 6429 and Sunday with 6176
- If we take the 4 most relevant observation for Members always in Métro Jean-Drapeau in this order of relevance:
    - A - on Sunday with 1849
    - B -on Saturday 1531
    - C- on Friday 848
    - D- on Wednesday 717

Please, execute the query in the section 3.5.5 in the .sql file to see that.