IBM COURSERA DATA SCIENCE CAPSTONE

The Battle of The Neighborhoods

In Search of the best location for an Italian restaurant in Brussels, Belgium

Juli 2019, Jolanda Dünner

Table of Contents

1 Business Case

Our client, whose dream has always been to open his own restaurant, wants to find the best location for his establishment. He already made the decision to open the restaurant in Brussels, Belgium. But he is not familiar with the areas in the city and doesn't know which location is the best choice. As he knows about the importance of choosing your location, he is asking for an analysis of the best neighbourhood to open his restaurant.
The problem question is, in other words: **What neighbourhood in Brussels has the best characteristics for our client to open up his Italian restaurant?**

2 The Data

The data for this project is coming from different sources. Our client wants to open an Italian Restaurant. Therefore, it is important to analyse the location of Italian Restaurants as well as the location of Restaurants in general.

2.1 Foursquare

Foursquare is a social networking service that allows users to discover and share information about businesses and attractions around them. The collection of different types of entertainment, drinking and dining venues make it a very valuable source for our analysis. Foursquare has an API that can be utilised to query their database and get information related to different areas, venues, reviews and tips.

2.2 zipcode-belgium

This source is used to obtain the neighbourhood location information from Brussels. It contains all Zip codes, neighbourhood names, longitudes and latitudes from Belgium. For our analysis, we will only need the information for Brussels and therefore drop all the other information.

3 Methodology

3.1 Data pre-processing

During this stage, we prepare the data to be used for our analysis. The csv for the neighbourhood information already has a very good structure, but we need to add column names and drop unnecessary columns. Also, we have to get rid of unnecessary information by removing all Zip codes outside of Brussels. The range of zip codes for Brussels goes from 1000 to 1210.

Out[105]:

	index	Unnamed: 0	1000	Bruxelles	4.351697	50.8465573
0	0	NaN	1020	Laeken	4.348713	50.883392
1	1	NaN	1030	Schaerbeek	4.373712	50.867604
2	2	NaN	1040	Etterbeek	4.389510	50.836851
3	3	NaN	1050	Ixelles	4.381571	50.822285
4	4	NaN	1060	Saint-Gilles	4.345668	50.826741

Unfortunately, the data set doesn't include column names, so we have to add them manually.

In [106]:
```
df_html.columns = ["index", "Unnamed", "Zip Code", "Neighourhood", "longitude", "latitude"]
df_html.head()
```

Out[106]:

	index	Unnamed	Zip Code	Neighourhood	longitude	latitude
0	0	NaN	1020	Laeken	4.348713	50.883392
1	1	NaN	1030	Schaerbeek	4.373712	50.867604
2	2	NaN	1040	Etterbeek	4.389510	50.836851
3	3	NaN	1050	Ixelles	4.381571	50.822285
4	4	NaN	1060	Saint-Gilles	4.345668	50.826741

After removing the unnecessary columns and filtering on all Zip codes in Brussels, we have 21 entries left to further analyse:

Out[113]:

	Zip Code	Neighourhood	longitude	latitude
0	1020	Laeken	4.348713	50.883392
1	1030	Schaerbeek	4.373712	50.867604
2	1040	Etterbeek	4.389510	50.836851
3	1050	Ixelles	4.381571	50.822285
4	1060	Saint-Gilles	4.345668	50.826741
5	1070	Anderlecht	4.312340	50.838141
6	1080	Molenbeek-Saint-Jean	4.322778	50.854355
7	1081	Koekelberg	4.325708	50.862263
8	1082	Berchem-Sainte-Agathe	4.292702	50.863984
9	1083	Ganshoren	4.317510	50.871240
10	1090	Jette	4.326090	50.877763
11	1120	Neder-Over-Heembeek	4.390489	50.897796
12	1130	Haren	4.412571	50.891966
13	1140	Evere	4.402160	50.870452
14	1150	Woluwe-Saint-Pierre	4.443297	50.829243
15	1160	Auderghem	4.433139	50.815657
16	1170	Watermael-Boitsfort	4.415818	50.799394
17	1180	Uccle	4.337235	50.801820
18	1190	Forest	4.317751	50.809143
19	1200	Woluwe-Saint-Lambert	4.428484	50.846693
20	1210	Saint-Josse-Ten-Noode	4.372336	50.853074

3.2 Using Foursquare API

We now have the basic information on the regions of interest. To get more information on each Neighbourhood, we are using Foursquare City Guide, commonly known as Foursquare. The local search-and-discovery app provides us with additional information about the venues in each neighbourhood.

In [33]:
```
brussels_venues.head()
```

Out[33]:

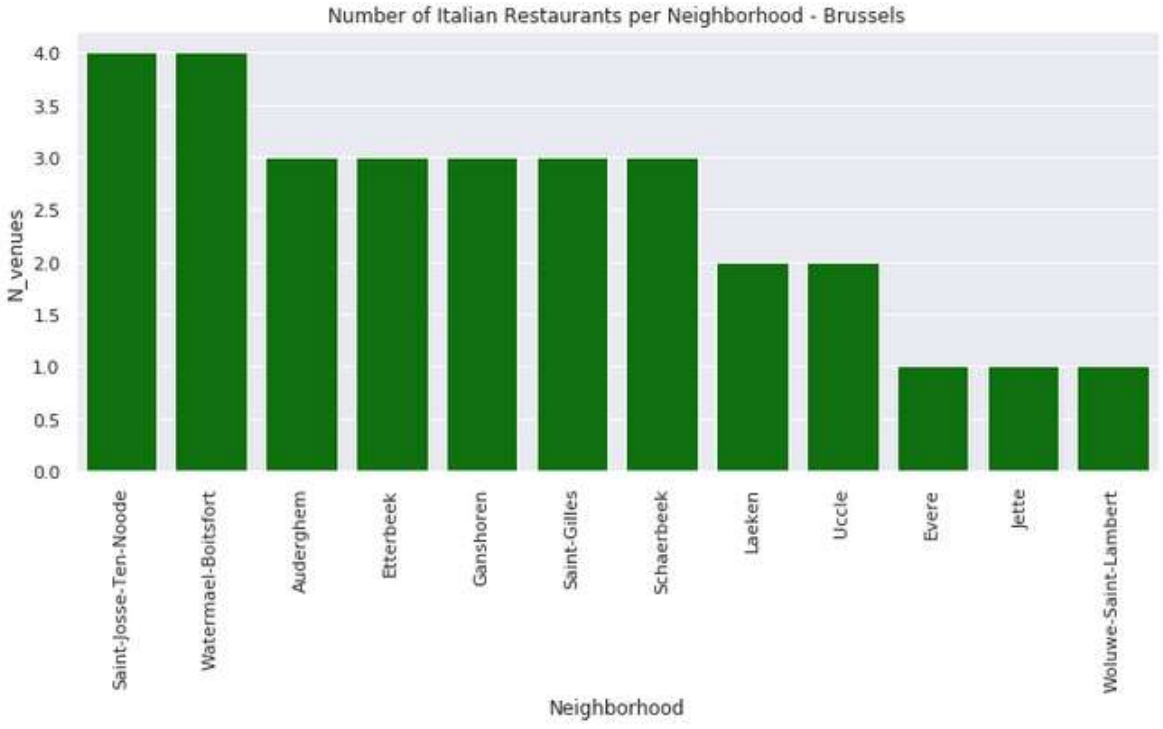
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Laeken	50.883392	4.348713	Les Jardins du Fleuriste	50.885331	4.345573	Park
1	Laeken	50.883392	4.348713	Steylemans	50.883638	4.344359	Electronics Store
2	Laeken	50.883392	4.348713	Pizzicato 2	50.884178	4.344376	Italian Restaurant
3	Laeken	50.883392	4.348713	Parc Sobieskipark (Sobieskipark)	50.884826	4.346192	Park
4	Laeken	50.883392	4.348713	Prins Leopoldsquare / Square Prince Léopold (P...	50.884269	4.342278	Plaza

3.3 Clustering

We can now use clustering to get a better understanding of the data. First, we explore the most common venues:

```
Out[40]: Venue Category
         Italian Restaurant    30
         Bar                   28
         Plaza                 28
         Supermarket           26
         French Restaurant     24
         Bakery                20
         Restaurant            17
         Sandwich Place        17
         Park                  15
         Snack Place           14
         Name: Neighborhood, dtype: int64
```

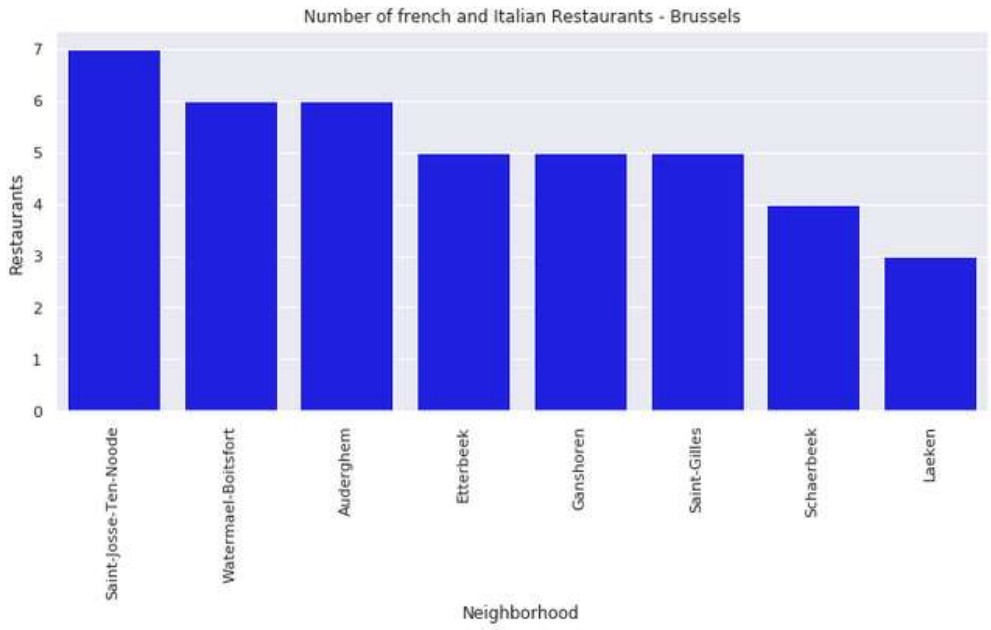
As we can see, we have a lot of Italian Restaurants in the city already. To get a better understanding of where those restaurants are located, we will filter on Italian Restaurants and cluster them by their Neighborhood.



The Neighbourhood with the most Italian Restaurants are **Saint-Josse-Ten-Noode** and **Watermael-Boitsfort.**

4 Results

From our analysis, we can say that there are already 30 Italian Restaurants in Brussels. It is difficult to get a clear conclusion from this statement, as it can mean both that there is already a high number of suppliers as well as that there is a high demand for Italian Restaurants. It is also common for similar venues to cluster in a specific region where the demand is high.



Adding up the Italian and the French Restaurants also shows a clear trend. The top 3 places are the same regions with a lot of Italian Restaurants, but Saint-Gilles is clearly dominated by the French cuisine.

5 Conclusion

The top 3 neighborhoods for restaurants are:

1. Saint-Josse-Ten-Noode
2. Watermael-Boitsfort
3. Auderghem

However, our analysis has shown that there are a lot of Italian Restaurants in the city already. Our client might consider opening a different restaurant to add a new variety or take other cities in Belgium in consideration as well. He now knows where the popular neighborhoods for Restaurants are, and also that there are already a high number of Italian Restaurants in those areas.