Capstone Project - The Battle of Neighborhoods

Restaurants at Brussels Capital Region

Applied Data Science Capston @ Coursera

Michael Fan

I. Introduction

I am interested in restaurants in Brussels, Belgium because of personal connection to Auderghem. Auderghem is a municipality in Brussels Capital Region. Furthermore, I am interested in knowing more about Japanese restaurant in that region, because it was known that (https://en.wikipedia.org/wiki/Auderghem) as of 2004 the majority of Brussels' Japanese expatriate population lives in Auderghem. I have never been to Brussels, nor do I speak French. It will be interesting in learning about that region given the knowledge, skills, and tools I have developed from this course so far. It will be interesting to utilize the database provided by Foursquare.com API and other sources that I can access through internet.

It is generally known that number of restaurants and types of restaurants are closely related to demographics in a region. When expats feel home sick, eating some food resembled to home is best to heal it. Will the data show that?

II. Data

Since I am interested in data of restaurants at Brussels Capital Region, I will be using Foursquare API to collect venue information. The intent is to analyze the data for different types of restaurants at different locations. Foursquare venue category ID hierarchical information will be collected and analyzed. Tools will be developed to access information from the database. Demographic information from https://en.wikipedia.org/wiki/List_of_municipalities_of_the_Brussels-Capital_Region will be scraped into the dataset. Geojson information of Brussels Capital Region from http://opendatastore.brussels/dataset/region-bruxelloise for municipality boundaries will be used in the analysis. Finally, Google MAP API will be used to obtain address information from geographical coordinated obtained from Foursquare query.

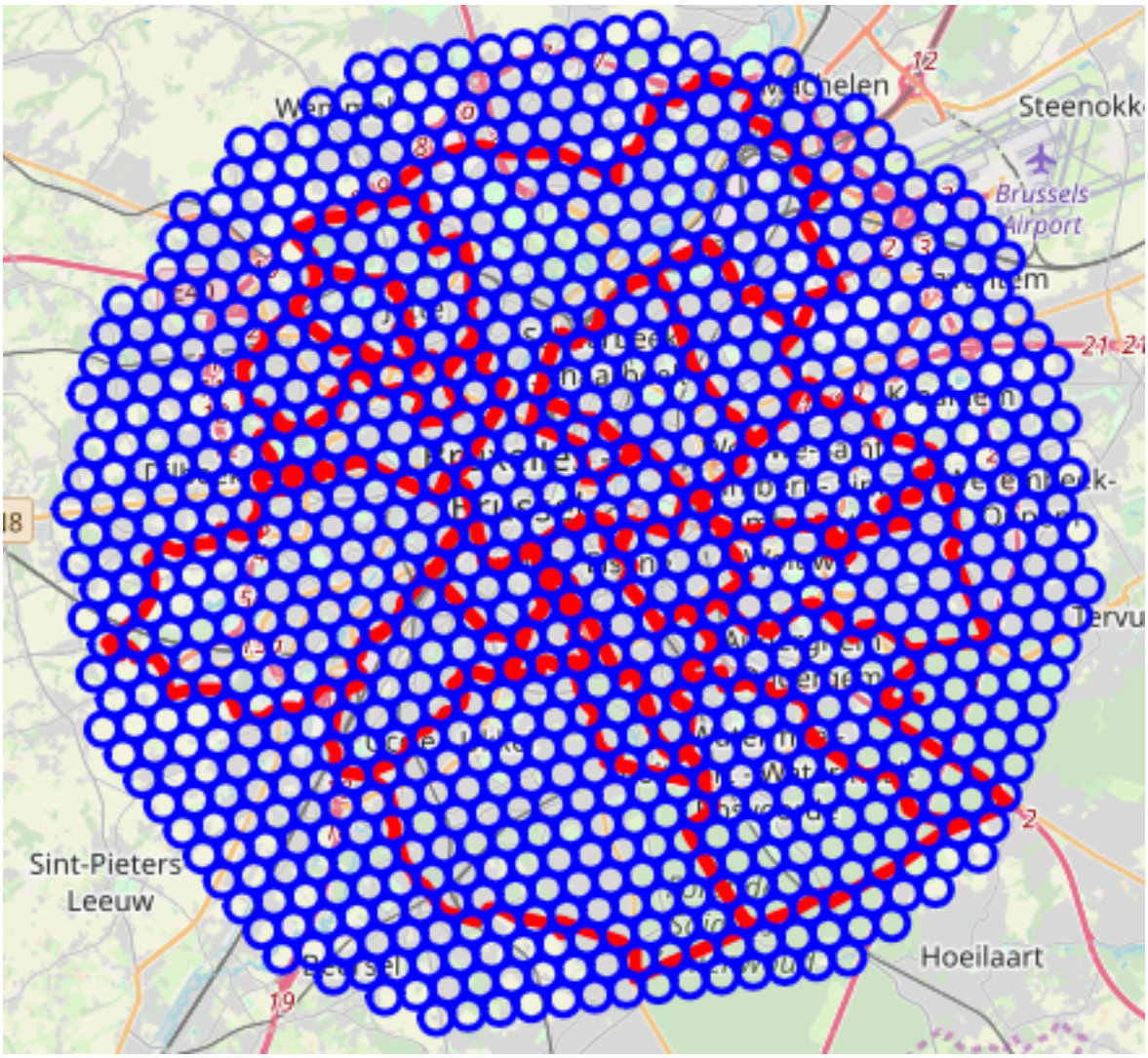
II.1 Brussels Capital Region Basic Information

Source: https://en.wikipedia.org/wiki/List_of_municipalities_of_the_Brussels-Capital_Region

	municipality	Dutch name	postcode	Population(1/1/2017)	Area	Population density(km²)	population	area	populationDensity
0	Anderlecht	Anderlecht	1070	118,241	17.717.7 km2 (6.8 sq mi)	6,680	118241	17.717	6680
1	Auderghem	Oudergem	1160	33,313	09.09.0 km2 (3.5 sq mi)	3,701	33313	9.090	3701
2	Berchem-Sainte-Agathe	Sint-Agatha-Berchem	1082	24,701	02.92.9 km2 (1.1 sq mi)	8,518	24701	2.920	8518
3	Bruxelles-Ville	Stad Brussel	1000,1020,1120,1130	176,545	32.632.6 km2 (12.6 sq mi)	5,415	176545	32.632	5415
4	Etterbeek	Etterbeek	1040	47,414	03.13.1 km2 (1.2 sq mi)	15,295	47414	3.130	15295
5	Evere	Evere	1140	40,394	05.05.0 km2 (1.9 sq mi)	8,079	40394	5.050	8079
6	Forest	Vorst	1190	55,746	06.26.2 km2 (2.4 sq mi)	8,991	55746	6.260	8991
7	Ganshoren	Ganshoren	1083	24,596	02.52.5 km2 (1.0 sq mi)	9,838	24596	2.520	9838
8	Ixelles	Elsene	1050	86,244	06.36.3 km2 (2.4 sq mi)	13,690	86244	6.360	13690
9	Jette	Jette	1090	51,933	05.05.0 km2 (1.9 sq mi)	10,387	51933	5.050	10387
10	Koekelberg	Koekelberg	1081	21,609	01.21.2 km2 (0.5 sq mi)	18,008	21609	1.210	18008
11	Molenbeek-Saint-Jean	Sint-Jans-Molenbeek	1080	96,629	05.95.9 km2 (2.3 sq mi)	16,378	96629	5.950	16378
12	Saint-Gilles	Sint-Gillis	1060	50,471	02.52.5 km2 (1.0 sq mi)	20,188	50471	2.520	20188
13	Saint-Josse-ten-Noode	Sint-Joost-ten-Node	1210	27,115	01.11.1 km2 (0.4 sq mi)	24,650	27115	1.110	24650
14	Schaerbeek	Schaarbeek	1030	133,042	08.18.1 km2 (3.1 sq mi)	16,425	133042	8.180	16425
15	Uccle	Ukkel	1180	82,307	22.922.9 km2 (8.8 sq mi)	3,594	82307	22.922	3594
16	Watermael-Boitsfort	Watermaal-Bosvoorde	1170	24,871	12.912.9 km2 (5.0 sq mi)	1,928	24871	12.912	1928
17	Woluwe-Saint-Lambert	Sint-Lambrechts-Woluwe	1200	55,216	07.27.2 km2 (2.8 sq mi)	7,669	55216	7.270	7669
18	Woluwe-Saint-Pierre	Sint-Pieters-Woluwe	1150	41,217	08.98.9 km2 (3.4 sq mi)	4,631	41217	8.980	4631

II.3 Brussels Capital Region Municipality Maps

Shown in the map is the region divided into municipality with boundary of red color, and blue circles covering the whole region. Each small circle is a block for which Foursqaure API query is performed. Because of limited results provided by each query, dividing the region of interest into small blocks is required.



II.4 Foursqaure Venue Category ID Database

Foursquare API provides abundant information about local businesses. In data analysis, usually categorial questions are being asked, instead of those about a specific business venue. Hence collecting relevant data through Foursqare API is essential. Venue category ID is used to query categorial information from Foursqaure. The venue category ID database is a hierarchical of json document. Multiple category ID's can be supplied for each query. However, single category ID may provide more consistent result. Efforts have established connection between the research question and the Foursquare venue category ID for query. The developed tools allow for asking the following questions:

- What are venue category ID's that include certain keyword?
- For a given category, what are its subcategories? What are their venue category ID's?

II.5 Collected and Prepared Data:

Python programing is used to scrape data from internet and perform query from API. It is followed by information combination and data preparation for analysis. The following is a snapshot of the dataset for which analysis is performed.

```
df_restaurants_brussels.head()
```

	vid	name	latitude	longitude	postcode	municipality	venueCategory2
47	5718fb6b498e930b29d77485	glacier lamy	50.775501	4.354569	1180	Uccle	Restaurant
49	56abd8c6498eaae8293f874f	À table 🍔🍝	50.774065	4.362790	1180	Uccle	Fast Food Restaurant
50	4f5fd2c3e4b00cb4dfcaf272	Burger	50.776336	4.372664	1180	Uccle	Burger Joint
61	533be9b3498e7c582a2e2d22	Pile poele	50.786713	4.417077	1170	Watermael-Boitsfort	Sandwich Place
74	4def7a001fc7e81f6da505c4	L' Entre-Pots (Uccle)	50.780449	4.325775	1180	Uccle	Italian Restaurant

III. Methodology

Number of restaurants in a region largely depends on demographics and its economic condition. More restaurants should be found in more populous regions. Thriving local businesses, especially those like restaurants, require good local economic conditions. Answering this question in general is beyond the scope of this work. The author is not an expert in such fields, nor the data collected include such information to draw upon.

Instead, the scope of this work is defined by the data collected. The focus is on:

- The research question.
- Data available from the major source, Foursqaure.com. And those data can be collected reasonably and economically.
- Other sources of data that may be available, and can be incorporated into data analysis, from which a conclusion can be drawn to answer the research question.

As a result, significant efforts were invested in understanding the data, collecting the data, and preparing the data for statistical analysis.

Foursquare API provides abundant information about local businesses. In data analysis, usually categorial questions are being asked, instead of those about a specific business venue. Hence collecting relevant data through Foursqare API is essential. Venue category ID is used to query categorial information from Foursqaure. Efforts, as shown in Data section, have established connection between the research question and the Foursquare venue category ID for query. Understanding what kind of data Foursquare API can provide is an important step to data collection.

Not included in this report, however, quality of Foursquare data needs to be investigated for its intended purpose. Like for any business venture, small scaled operation has to be tested before

significant commitment. It was such effort through which the author concluded to use one category ID for each query instead of multiple category ID's as allowed by the API.
As suggested by previous work, area of interest is divided into small blocks so that data can be queried in a more consistent manner.

One focus of this work is on comparing different municipalities in Brussels capital region. Foursquare queired results include geographic information such as latitudes and longitudes information and others. However, Google Map API provides more consistent manner to extract municipality information from latitudes and longitudes.

In addition to Foursquare API and Google Map API, relevant regional information about Brussels and geojson for geographical boundary information is obtained from internet.

Data collection through internet is cheap and even free. However, it does not mean it is easy. Collecting large amount of data impose certain difficulties. Certain difficulties have been addressed through Python programing as shown in the Jupyter Notebook submitted along with this report.

Once the data collected, it is cleaned and prepared for further analysis with the focus on the relevance to research question.

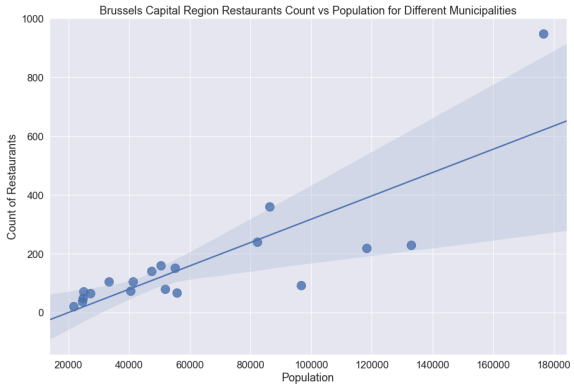
The author hopes the methodology established here allowed all relevant data to be collected. Of course, further systematic investigation is necessary to establish better confidence in this approach.

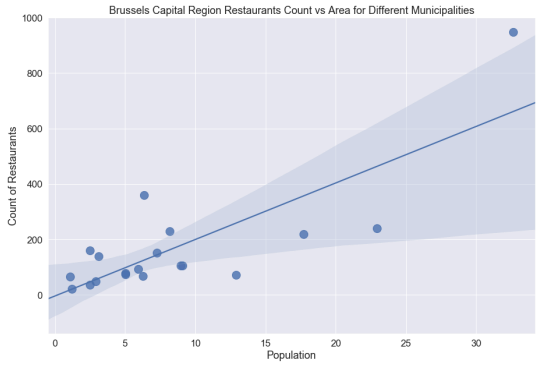
Notwithstanding, from the understood, collected, and prepared data, the analysis is ...

IV. Results and Discussion

The results and discussion are combined in one section by analyzing a research question followed by its discussion, instead of separating them into two sections required by the assignment.

Research Question 1: Is there association between population and number of restaurants?





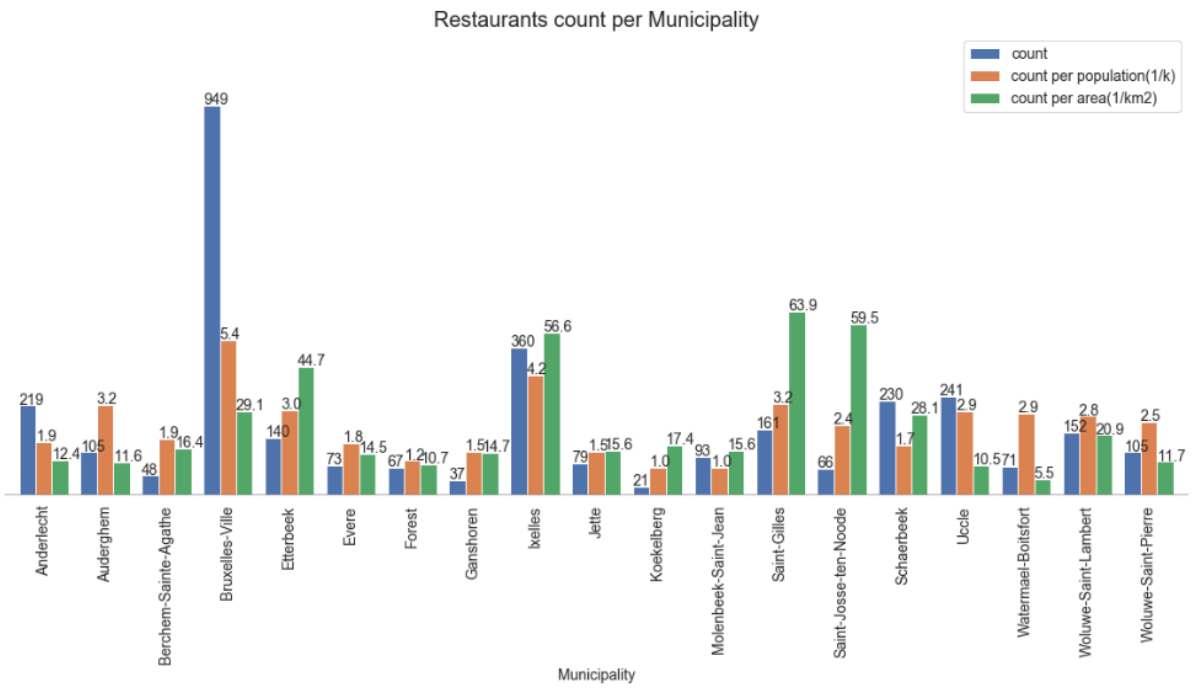
The regression plot indeed shows correlation between number of restaurants and population/area of a region.

Averaging over municipality, the following is a summary table:

	Restaurant count per municipality	avg
0	count	169.315789
1	count per population (1/k)	2.413389
2	count per area (1/km2)	24.174397

Considering the total population, area, and total count of restaurants, there are 169.3 restaurants per municipality in average; 2.4 per 1000 residents of population, and 24.1 per km^2 of area.

Showing for the data per municipality:

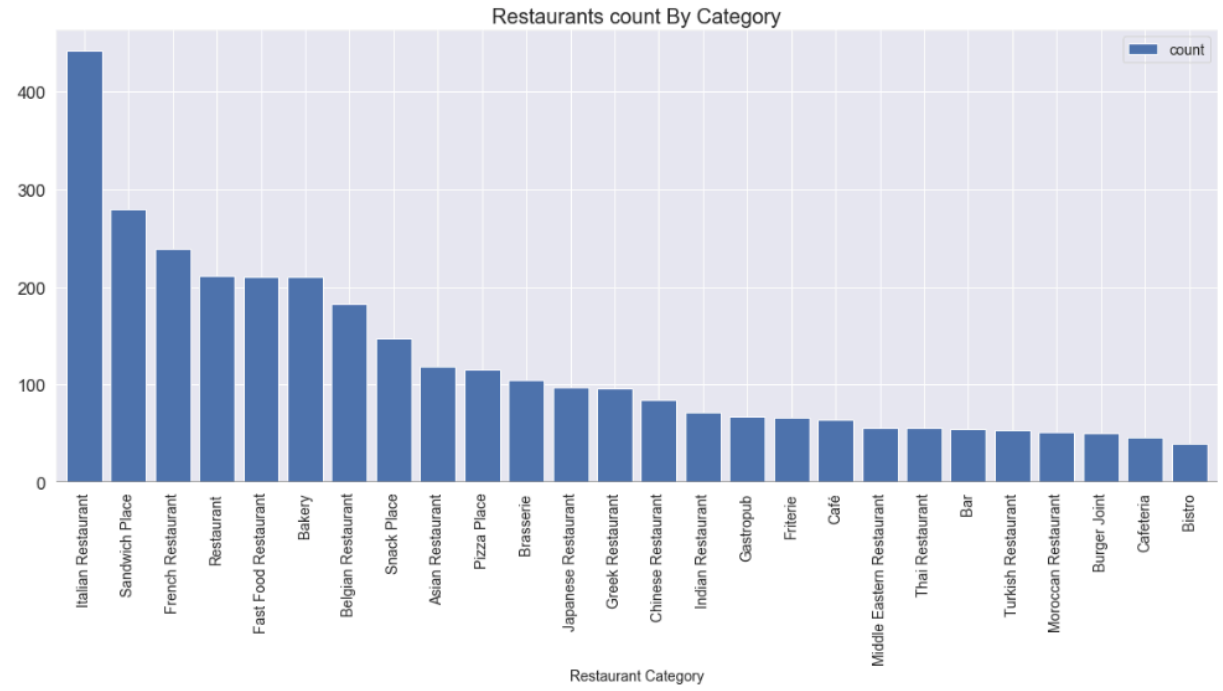


The data indeed show greater restaurant count if the poplulation and/or area is greater. In average, there are 2.4 restaurants per 1000 population for each municipality; 24.2 per area of km2. Koekelberg and Molenbeek-Saint-Jean seem to have lower than trend count. Bruxelles-Ville and Ixelles seem to have greater than trend count. This may be due to economic conditions in different municipalities.

Saint-Gilles and Saint-Josse-ten-Noode show quite dense restaurants in their area, 63.9 per km2 and 59.5 per km2. However, only in-trend number of restaurants per 1000 potential customers. This implies that even though it feels there are many restaurants when we look around in that region. But still, the competition may not be as saturated as some other municipalities. Of course, the data show these two municipalities have the greatest population density.

It will be interesting to know how economic condition, as measured by income level per capita, will affect the restaurant distribution. Combining information about density of restaurants and income level may indicate how profitable each restaurant is. It may provide insight into whether or not a particular market is saturated or under served.

Research Question 2: What are the top categories of restaurants in each municipality?



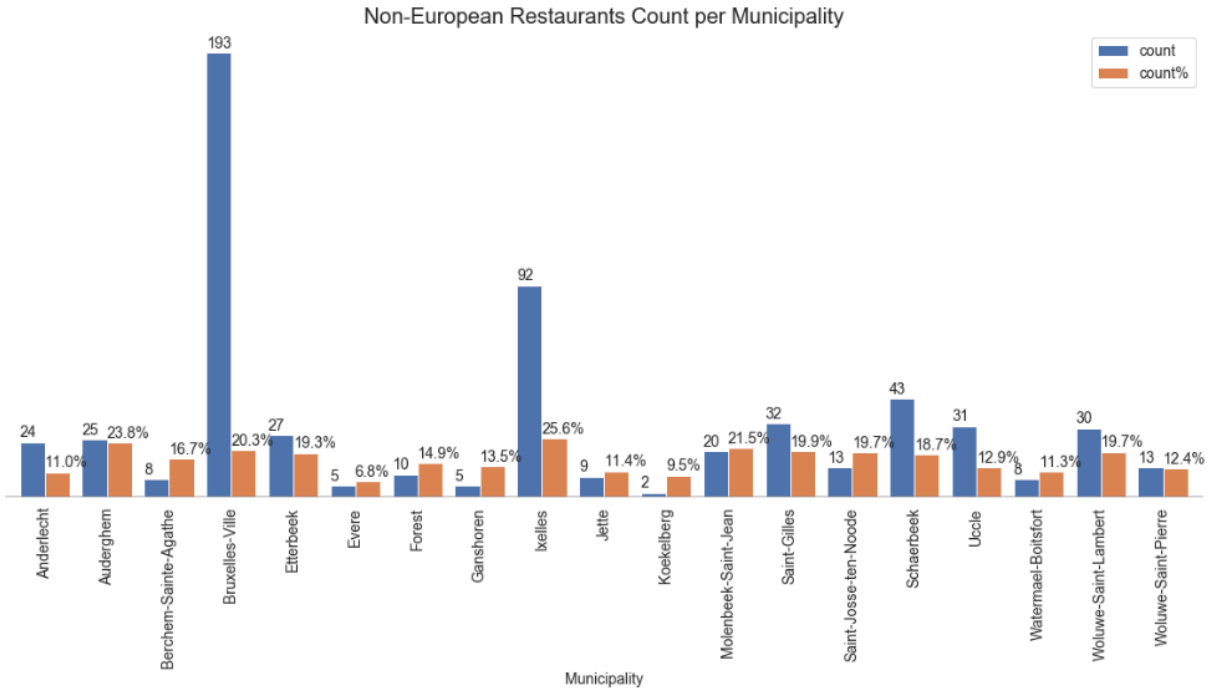
Italian restaurants are most popular categorized restaurants in Brussels, followed by Sandwidch Place and French Restaurant. Category Restaurant seems refers to restaurant that are not put into any particular restaurant category, and is ranked the fourth.

For each municipality, the top restaurant categories are:

	1	2	3	4	5
Anderlecht	Bakery	Italian Restaurant	Snack Place	Restaurant	Sandwich Place
Auderghem	Italian Restaurant	Bakery	Sandwich Place	Asian Restaurant	Fast Food Restaurant
Berchem-Sainte-Agathe	Snack Place	Bakery	Greek Restaurant	Italian Restaurant	Japanese Restaurant
Bruxelles-Ville	Italian Restaurant	Sandwich Place	Belgian Restaurant	French Restaurant	Restaurant
Etterbeek	Italian Restaurant	Sandwich Place	Bakery	Fast Food Restaurant	Chinese Restaurant
Evere	Friterie	Snack Place	French Restaurant	Sandwich Place	Bakery
Forest	Sandwich Place	Pizza Place	Italian Restaurant	Belgian Restaurant	Snack Place
Ganshoren	Italian Restaurant	Bakery	Belgian Restaurant	Friterie	Restaurant
Ixelles	Italian Restaurant	French Restaurant	Sandwich Place	Fast Food Restaurant	Japanese Restaurant
Jette	Bakery	Sandwich Place	Snack Place	Italian Restaurant	Belgian Restaurant
Koekelberg	French Restaurant	Fast Food Restaurant	Restaurant	Bakery	Italian Restaurant
Molenbeek-Saint-Jean	Italian Restaurant	Fast Food Restaurant	Bakery	Sandwich Place	Turkish Restaurant
Saint-Gilles	Fast Food Restaurant	Italian Restaurant	French Restaurant	Sandwich Place	Greek Restaurant
Saint-Josse-ten-Noode	Italian Restaurant	Fast Food Restaurant	Sandwich Place	Middle Eastern Restaurant	Snack Place
Schaerbeek	Fast Food Restaurant	Italian Restaurant	Bakery	Sandwich Place	Snack Place
Uccle	Bakery	Italian Restaurant	French Restaurant	Restaurant	Sandwich Place
Watermael-Boitsfort	French Restaurant	Restaurant	Italian Restaurant	Cafeteria	Sandwich Place
Woluwe-Saint-Lambert	Italian Restaurant	Sandwich Place	Bakery	Fast Food Restaurant	French Restaurant
Woluwe-Saint-Pierre	Bakery	French Restaurant	Italian Restaurant	Restaurant	Sandwich Place

Research Question 3: How are non-European restaurants distributed in each municipality?

Since I am mainly interested in data association between restaurants and expats. Let's look at the non-European restaurants:



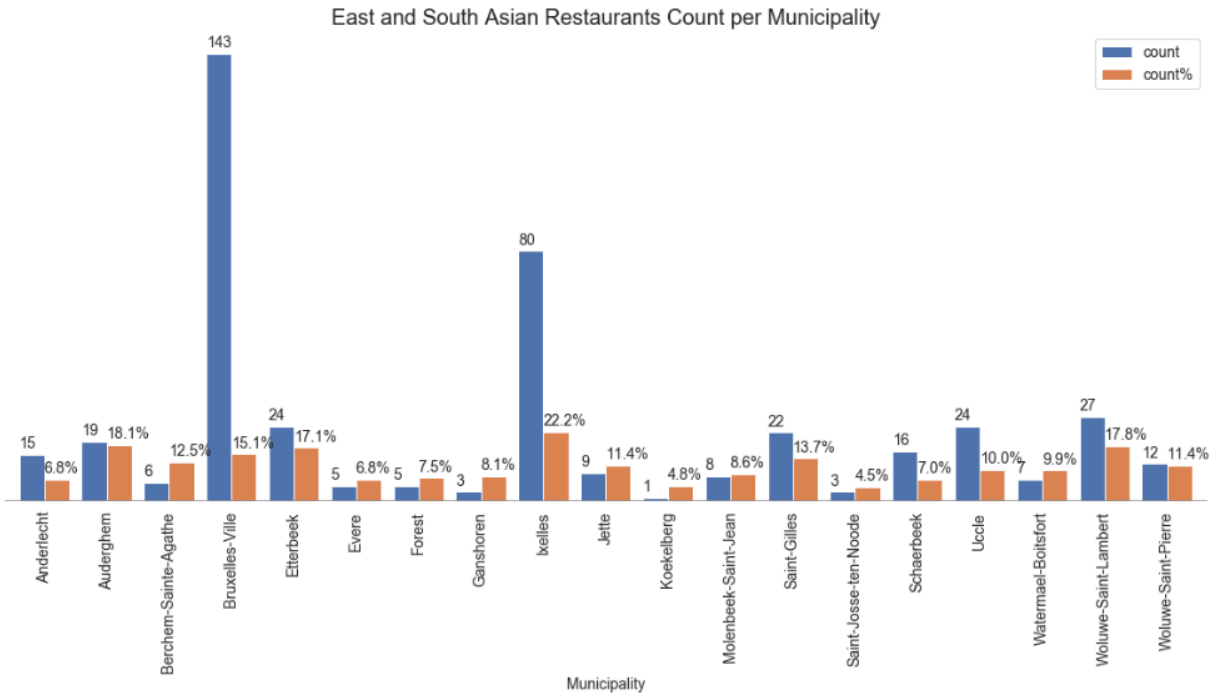
Auderghem, Bruxelles-Ville, Ixelles and Molenbeek-Saint-Jean have more (>20%) non-European restaurants than other municipalities.

Since Bruxelles-Ville has largest population, its percentage of non-European restaurants may indicate Belgium's general population's interest in such food, which is 20.3%. This may imply Auderghem, Ixelles and Molenbeek-Saint-Jean have more expats living in Brussels than other municipalities.

Research Question 4: Which municipalities have more Asian restaurants?

As mentioned in introduction, I am interested in knowing statistics of Japanese Restaurant. Japanese Restaurant belong to the category of Asian Restaurant. In the data, some venues are specified with Asian Restaurant subcategories, exclusive from other categories. It will be better we look at them as a whole.

Particularly, East and Southern Asian Restaurants for each municipality are



If we only consider East and South Asian Restaurants, Molenbeek-Saint-Jean is no longer the top municipality. Auderghem and Ixelles in addition to Bruxelles-Ville are still the top municipalities. Etterbeek and Woluwe-Saint-Lambert are other municipalities that have more East Southern Asian Restaurant count percentage than the benchmark, Bruxelles-Ville.

This may imply East Southern Asian expats may have larger portion of population in Auderghem, Ixelles, Etterbeek and Woluwe-Saint-Lambert than the benchmark. Molenbeek-Saint-Jean may have more non-East-Southern expats.

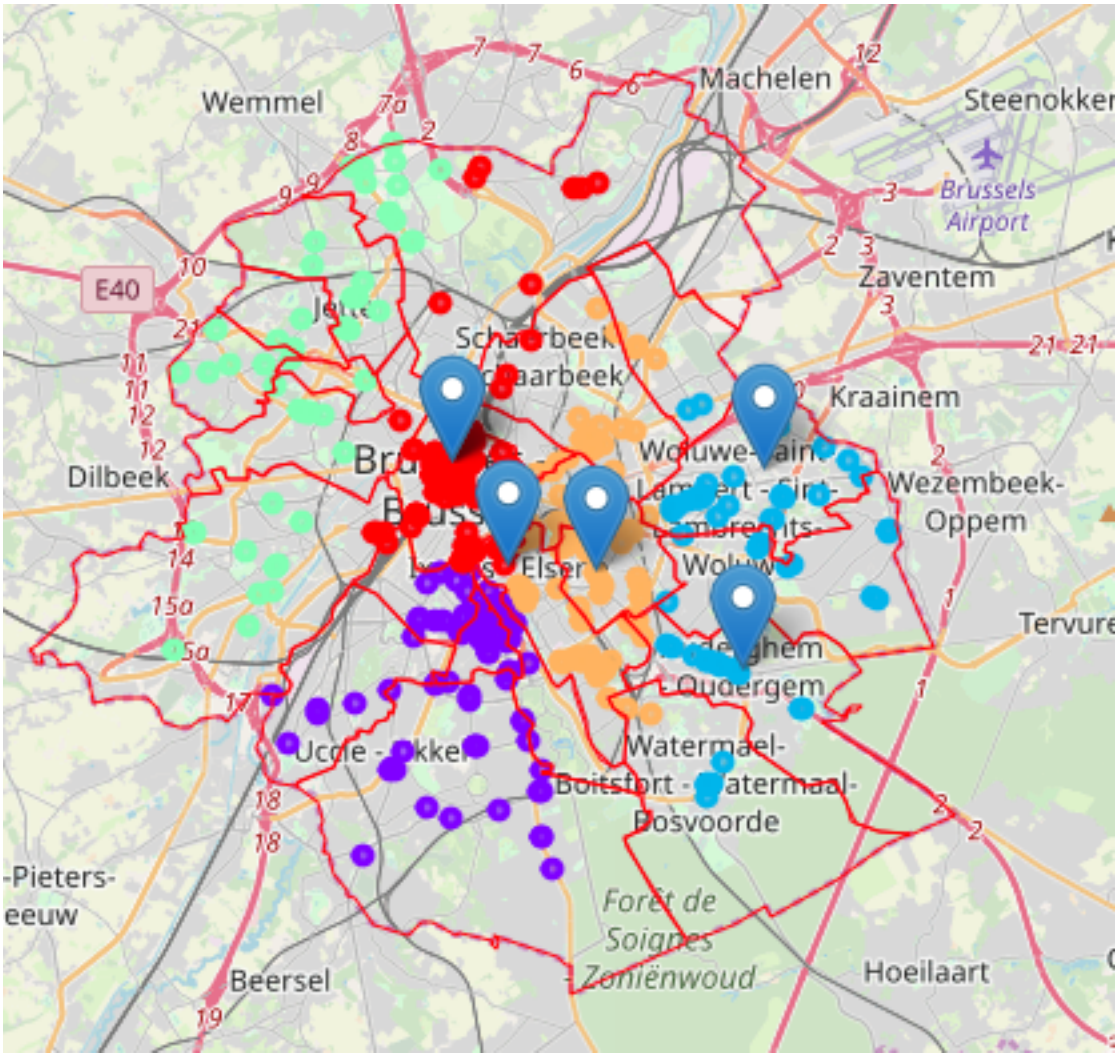
Research Question 5: What about Japanese restaurants?

My motivation to conduct this analysis was to know more about Japanese restaurants in Auderghem because it was reported that large population of Japanese expats live in Auderghem. However, from the analysis performed, I realized there are flaws in my analysis. Several are:

1. Some restaurant venues are categorized as Asian Restaurant as primary category, which means it is not specified as one of many subcategories of Asian Restaurants. Foursquare may have more venue details to provide such information. But it is beyond the scope of this work.
2. After glancing the data when performing analysis, I found that restaurants with Japanese-related name are categorized as Seafood Restaurant. Foursquare may have more venue details to provide such information. But it is beyond the scope of this work.
3. As it can be seen in the analysis performed, Auderghem does not have many restaurnt venue information provided by Foursquare: 105 in total restaurant venues excluding those categories with fewer than 50 venues in Brussels Capital Region as whole, 25 for non-european restaurants, 19 for East Southern Asian Restaurants. The statistics

population is not very large to draw much meaningful information, especially considering data are obscured by the previous two factors.

As a result, the final analysis of this work is to perform clustering analysis of Asian Restaurants in Brussels Capital Region.



5 clusters of East Southern Asian restaurants are shown in the map. Also shown are labels for Auderghem, Ixelles, Etterbeek, Wolue-Saint-Lambert, and Brussels. Indeed, in those regions, greater density of restaurants of interest is seen.

V. Conclusion

Categorial descriptive statistical analysis has been performed to understand the association between population and restaurants, using data collected from Foursqaure.com API, Google MAP API, and Wikipedia for Brussels Capital Region. In general, there is obvious association between number of restaurants and population. For municipality with larger population, there

tends to be more restaurants as data show. Some municipalities have above or below trend restaurant count per population. The reason for that may be the economic condition, i.e. income level. Another reason may be the population is either under-served or restaurants are over-crowded. Answering such questions require more information about economic condition of the municipalities. It is beyond what the collected data can be drawn conclusion upon.

The fact I am interested in Auderghem than any other municipalities is a hint that Auderghem may be a municipality with larger expat population, in general, than other municipalities in Brussels. The data indeed imply such. However, the collected data is obscure to provide more affirmative information to conclude that more Japanese restaurants in Auderghem as a result of large Japanese expat population. Some potential reasons were discussed in the Results and Discussion section. Querying more details of venue information from Foursquare API may provide further clarification.

Data collected for this work is through internet. Such approach is cheap and even free. However, it does not mean it is easy. An attempt has been made to establish a methodology to understand, collect, and prepare the data for better analysis. To achieve better confidence in such an approach, further investigation is required.