Choosing the best place to open a new restaurant.

Philippe Zahlen

February 2021

1. Introduction

In a city (I am using Bruxelles in my example, just to change it a bit from always using a city in North America), where would be the best place to open a new restaurant in a specific category (French restaurant, Italian restaurant, Chinese restaurant, etc.). I will structure the city in different categories based on their postal code and gather information from a couple of restaurants in each postal code. Based on the gathered information and there rating, it would be possible to find the best place for a restaurant based on lack of some good ones. This is just one idea to search for a place, in the future this can be extended to make better guesses on different criteria. The foursquare API gives me the possibility to gather the necessary information in the different postal codes for the interpretation. I am guessing I will search for 50 establishments per postal code.

2. Data acquisition

The data will come from foursquare. I used the API to search in the different postal codes (19 are present in this city) for 50 restaurants and put the information in one single data frame. Later the ratings will be extracted and put into the data frame. The radius of 1 km is chosen in the center of every postal code. The issue will arise then of course that we will receives doubles at the borders of different postal codes because of the unevenness of those neighborhoods. Those doubles will be cleaned later.

	id	name	categories	referralId	hasPerk	location.address	location.crossStreet	location.lat	location.lng	location.labeledLatLngs	location.distance	location.pos
0	509d18bde4b07eb4d8229833	Frenchkick	[{'id': '4bf58dd8d48988d104951735', 'name': 'B...	v-1613492433	False	Rue Marche Au Charbon 11	11	50.846585	4.350772	[{'label': 'display', 'lat': 50.846585, 'lng':...	65	
1	4eb2fbb8b80328580a017302	Brasserie de l'Ommegang	[{'id': '57558b36e4b065ecebd306b0', 'name': 'B...	v-1613492433	False	Grote Markt 9 Grand Place	NaN	50.846566	4.352609	[{'label': 'display', 'lat': 50.84656550922511...	64	
2	4c0e1a5cb1b676b0d77be186	't Kelderke	[{'id': '4bf58dd8d48988d1c4941735', 'name': 'R...	v-1613492433	False	Grote Markt 15 Grand Place	NaN	50.846477	4.352807	[{'label': 'display', 'lat': 50.84647738575094...	78	
3	51d96c40498ec79de3eea97d	Restaurant Poissons Grill	[{'id': '4bf58dd8d48988d1ce941735', 'name': 'S...	v-1613492433	False	Rue de Brabant 67	NaN	50.846563	4.353374	[{'label': 'display', 'lat': 50.84656277189522...	117	
4	4d1f8ac6756e8cfafd3c6454	Thai Talks	[{'id': '4bf58dd8d48988d149941735', 'name': 'T...	v-1613492433	False	Steenstraat 51 Rue des Pierres	NaN	50.846893	4.350882	[{'label': 'display', 'lat': 50.84689343001388...	68	

Note: Due to restriction of mine foursquare developer account, I did not receive as much as data as predicted and I was restricted to 395 ids.

Note: The issue appeared here that because of the restricted developer account and that getting the ratings is a premium service. To remedy this, I chose with regrets to use random ratings and this does not take away from the problem at hand.

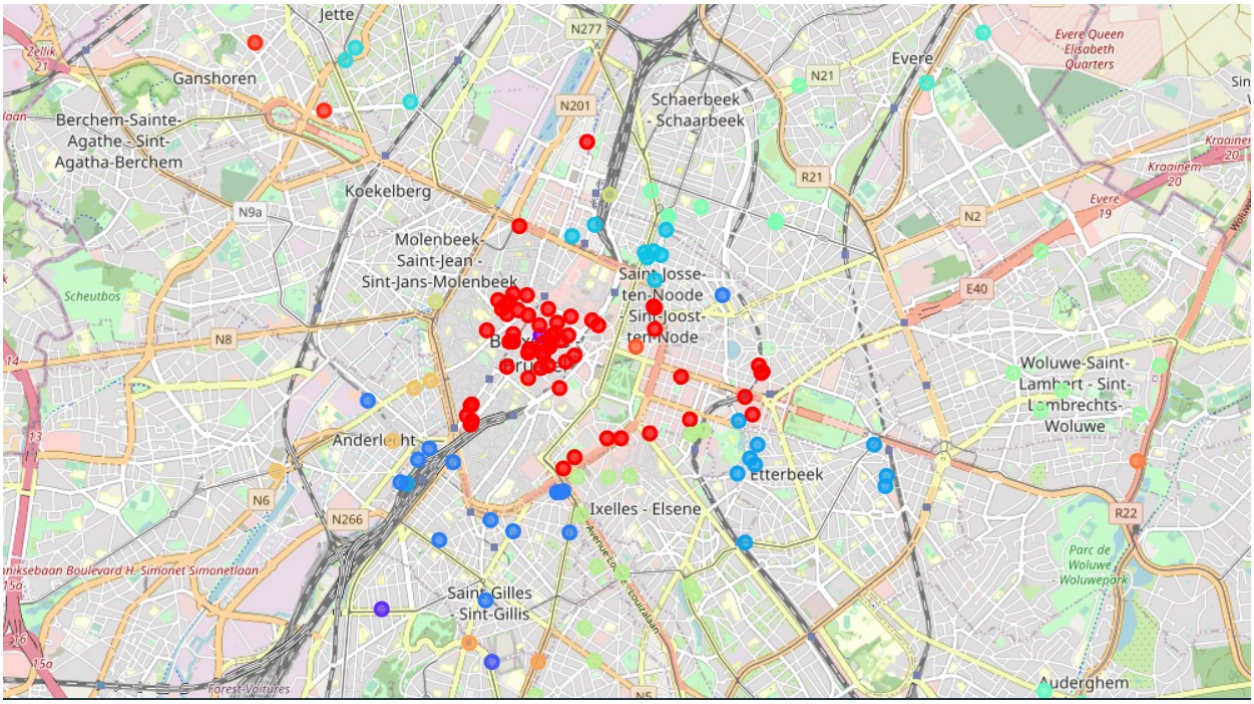
3. Data cleaning

Just using the necessary columns and changing the name to something more meaningful.
Dropping the duplicates and the rows with no values.
Lastly adding a new column for the ratings.

	id	name	address	latitude	longitude	postal code	rating
0	509d18bde4b07eb4d8229833	Frenchkick	Rue Marche Au Charbon 11	50.846585	4.350772	1000	
1	4eb2fbb8b80328580a017302	Brasserie de l'Ommegang	Grote Markt 9 Grand Place	50.846566	4.352609	1000	
2	50b5f09be4b0f5cb08770674	Restaurant NMBS / SNCB	Europakruispunt 2 Carrefour de l'Europe	50.846345	4.357042	1000	
5	52e2df4c498ed7af43f58713	Rustic - Restaurant romanesc	Rue Gretry	50.849873	4.350758	1000	
6	4c7cd7b4744d8cfadfa2bb37	French Touch	Marnixlaan 17 avenue Marnix	50.839198	4.363390	1000	

4. Exploratory Data Analysis

The data is not complicated to understand, so I already started to do a Kmeans clustering based on the coordinates and postal code.
The different clustering is then visualized on a map.



Now we could continue to tweak even more and to solve the problem graphically but there is much easier solution for our problem, and I chose a much simpler method by grouping and taking the means of every ratings.

5. Predictive Modeling

The method I chose was finding the best place by simply grouping by postal code and taking the mean. Here we use the power of pandas to do that very easily. There are of course more methods to solve this, but I took the simplest as it is easily understood.
By using the simple 'groupby' method in pandas, I received the following result.

	rating
	mean
postal code	
1000	4.617647
1030	3.750000
1040	6.000000
1050	1.666667
1080	0.000000
7000	8.000000

Based on this the best place to open a restaurant is the postal code with the lowest value and in our case then the postal code 1080.
The issue with this method is of course the lack of data and the best thing is of course to collect as much as possible.

5. Conclusions

The conclusion is that my lack of data, gives me a falsely prediction and in the future the best thing is to collect much more data. Another thing is that this was just for one category of restaurants, different categories will give different predictions, but the method still stays valid.