Capstone Project- The Battle of the Neighborhoods (Week 2)

Applied Data Science Capstone by IBM/Coursera

By: Zahi Masarwa

Introduction: Best Area to open a business in Ottawa region

In this project, I will be analyzing the Ottawa region for the best area in which a business can succeed.

Since the Ottawa region is big and there are a lot of different neighborhoods, it's difficult to predict which area is the best to open up a venue.

I will use data science to analyze the region by neighborhood and its population, and I will add the number of tips that people leave in that area to see which area is the most active.

Data

Based on the definition of our problem, the factors that will influence our decision are:

* number of existing restaurants in the neighborhood (any type of restaurant)

* the population of each neighborhood

* number of tips people leave

We decided to use the regularly-spaced grid of locations to define our neighborhoods.

The following data sources will be needed to extract/generate the required information:

* neighborhood areas will be generated from CSV file and approximate addresses of centers of those areas will be obtained using **Google Maps API reverse geocoding**

* number of restaurants and their type and location in every neighborhood and tips will be obtained using **Foursquare API**

* coordinates of Ottawa will be obtained using **Google Maps API geocoding** of well-known Ottawa location

Methodology section

The main mythology that used in the analysis is analysis based on open.Ottawa web site which I take all the neighborhoods of Ottawa and I merge with longitude and latitude of each neighborhoods that I exported using geolocator.geocode from google and one of the problem that I encounter that the geo locater don't recognize every neighborhoods which means I had to skip those and from 111 areas I found 80 and on those I based the analysis.

Based on the data frame that I have I mapped every neighborhood and circle it on the map.

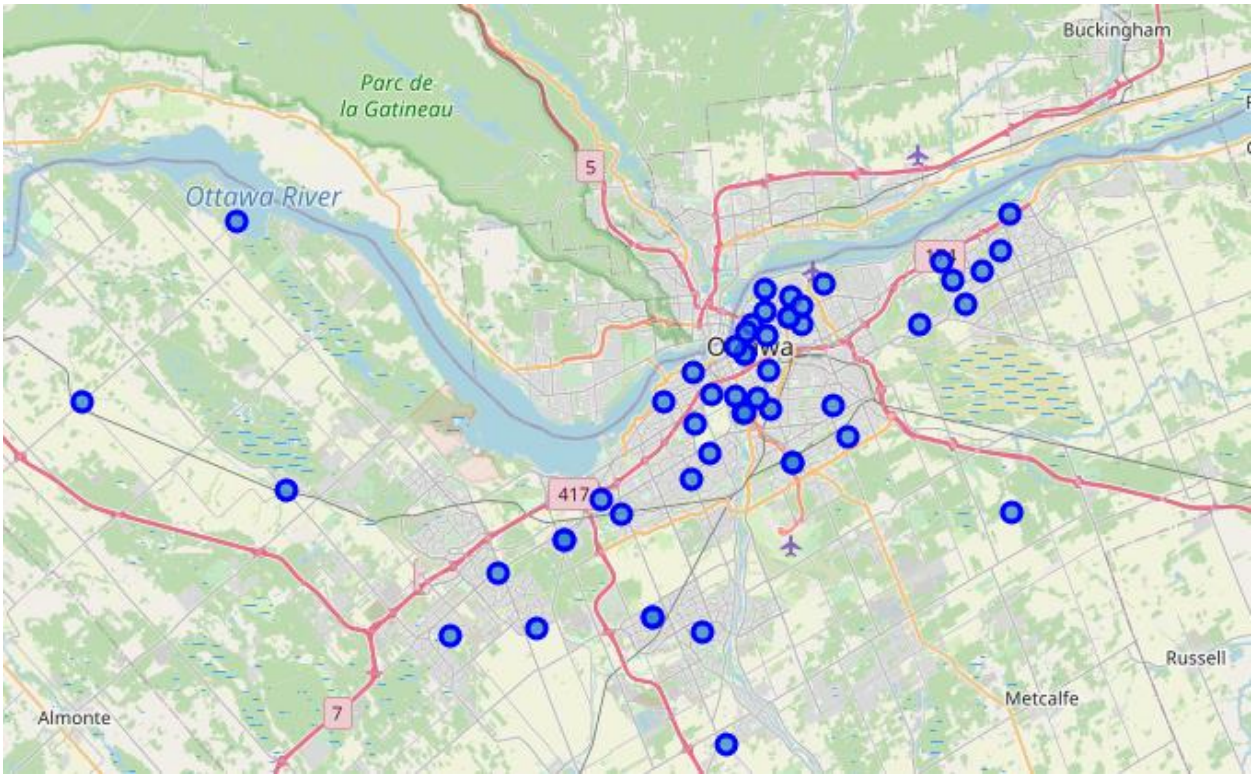
Foursquare api:

I used Foursquare api to bring all the venues in each neighborhood and bring the longitude and latitude of each one and brought the tip count of each one of those neighborhoods

Finally, I used all that info to bring the percentage of the tips by each population of the neighborhood to make sure which one is the highest's and used machine learning to cluster the areas of the highest possibility to have a good business.

Results

This is the map of the Ottawa region and blue circle to map all the neighborhoods that I was able to find



After I found all the areas and mapped them, I showed a sample of the area to make sure that the data is back that include venues name and id and longitude and latitude and the category

Out[350]:

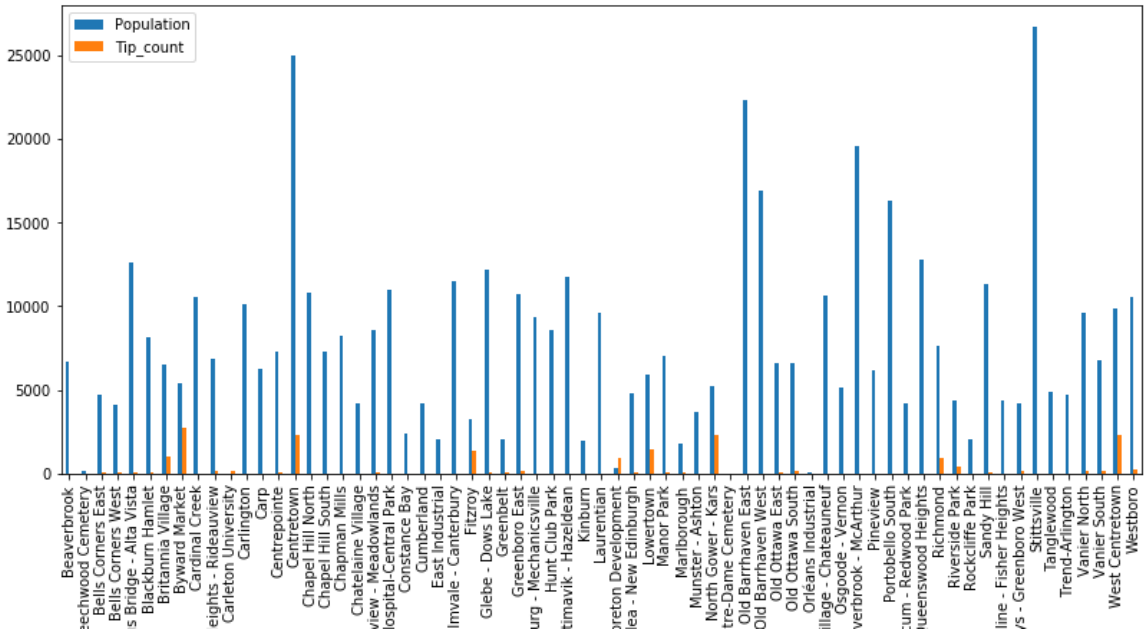
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue ID	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Old Barrhaven West	45.279020	-75.763036	4b832e3ef964a52015fc30e3	Walter Baker Sports Centre	45.280650	-75.762124	Stadium
1	Old Barrhaven West	45.279020	-75.763036	4e6cbc94d4c09a17a5b74b9a	Walter Baker Swimming Pool	45.280489	-75.762156	Pool
2	Old Barrhaven West	45.279020	-75.763036	4b0586e0f964a520b77222e3	Malvern Park	45.281032	-75.759745	Playground
3	Beaverbrook	35.278471	-81.193967	51fab68f498e1537497c43ee	Highland Athletic Center	35.280519	-81.191863	Athletics & Sports
4	Beaverbrook	35.278471	-81.193967	5690ad92498ebce4bf579c00	Universal Travel	35.274718	-81.193613	Rental Car Location

Then I show a sample of the result of tip count included after I extracted the it using the foursquare api and also include the population of each neighborhood

]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue_ID	Venue	Venue Latitude	Venue Longitude	Venue Category	Tip count	Population
0	Old Barrhaven West	45.27902	-75.763036	4b832e3ef964a52015fc30e3	Walter Baker Sports Centre	45.280650	-75.762124	Stadium	4	16880
1	Old Barrhaven West	45.27902	-75.763036	4e6cbc94d4c09a17a5b74b9a	Walter Baker Swimming Pool	45.280489	-75.762156	Pool	2	16880
2	Old Barrhaven West	45.27902	-75.763036	4b0586e0f964a520b77222e3	Malvern Park	45.281032	-75.759745	Playground	1	16880
3	Old Barrhaven East	45.27902	-75.763036	4b832e3ef964a52015fc30e3	Walter Baker Sports Centre	45.280650	-75.762124	Stadium	4	22286
4	Old Barrhaven East	45.27902	-75.763036	4e6cbc94d4c09a17a5b74b9a	Walter Baker Swimming Pool	45.280489	-75.762156	Pool	2	22286

This is a chart of that shows the count of tips, population by neighborhood



After I compere that I created a new data frame to show the top area by the percentage

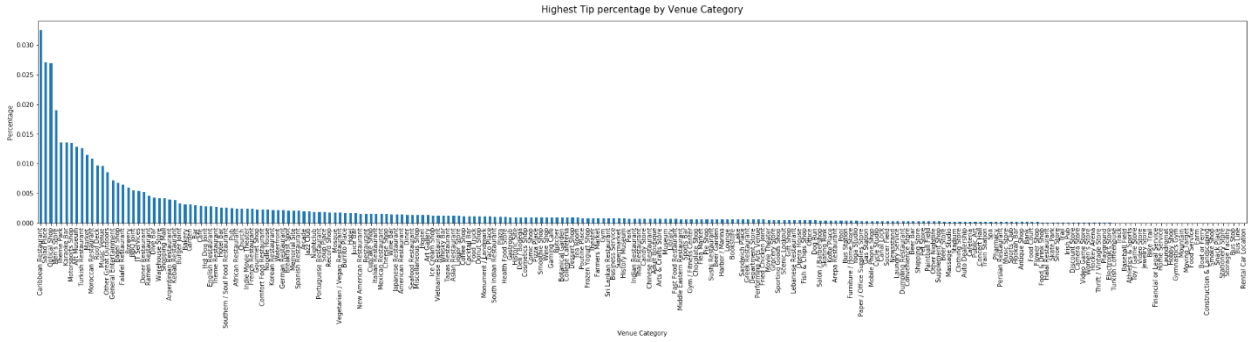
Neighborhood	Population	Tip_count	Percentage	Cluster Labels
Carleton University	19	153	8.052632	3
Lebreton Development	369	912	2.471545	3
Byward Market	5382	2717	0.504831	4
North Gower - Kars	5213	2279	0.437176	4
Fitzroy	3299	1391	0.421643	4

Also added a table of most business has high tip percentage according to population of each neighborhood

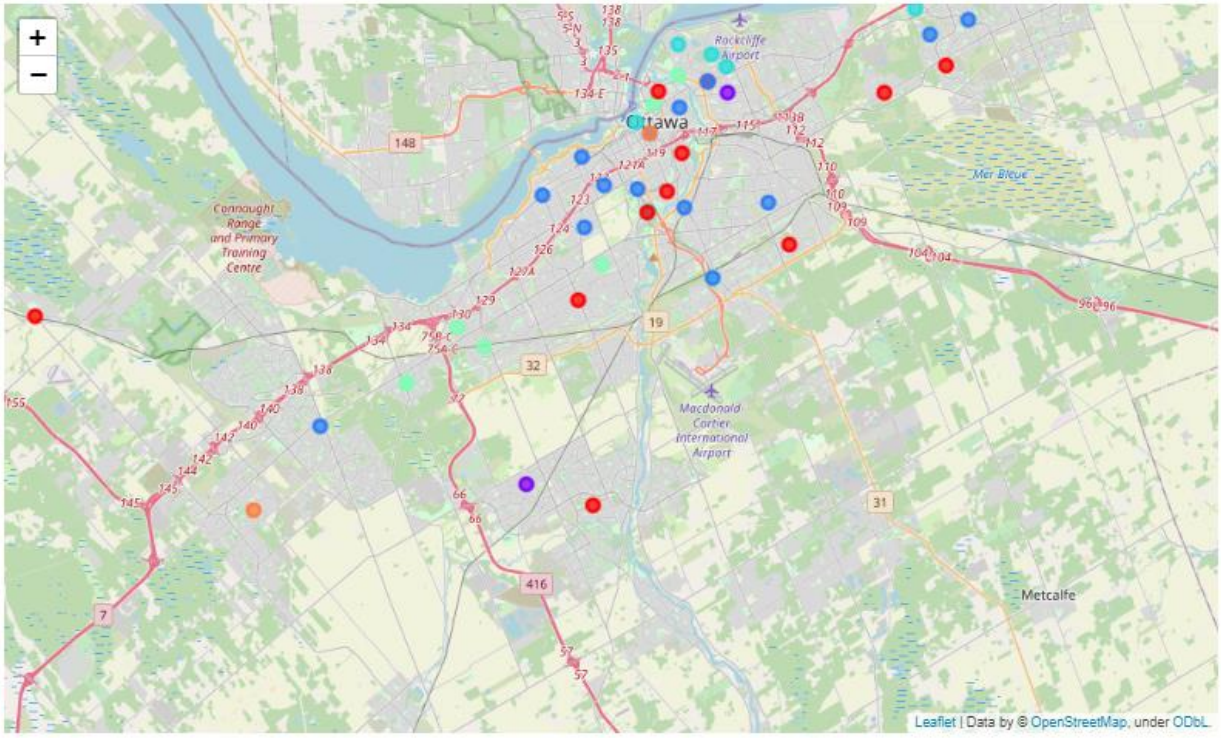
:

Venue Category	Population	Tip count	Percentage
Caribbean Restaurant	369	12	0.032520
Salad Place	738	20	0.027100
Optical Shop	74	2	0.027027
Men's Store	369	7	0.018970
Theme Park	369	5	0.013550

This is the business with a high percentage category



this all the areas clustered to show which area is the best to open a business



as you can see in this area the map the best areas are teal and baby blue are the areas that best to open a business.

Discussion section

After I finished gathering all the information about the areas and the tips if the area is very populated it doesn't mean it will be full of customers based on the tips and the info that I have of curse the data is not finished

since there still more info that need to be gather on each business and business in order to see more info about the customers that they get and are they come from the same area to their venue.

Conclusion

In conclusion we fount the best area to open a business is in cluster 3 or 4 which they are have the highest percentage of tip count and it's more likely to have your business to succeed.