What is the safe places to open a restaurant?

What type of dish should you sell?

Ayoub Wahbi

Avril ,17,2020

1-**Introduction:**

There are many investors which investigates in restaurants but there investment do not success, that is not by chance, because off the place chosen, the behaviors of people in this place, the type of product that the buy do not pay attention of client

1.2-Problem:

Now, I have an investor who want to open a restaurant in Canada, Toronto, his fist problem that he don't know a better place, and he looking for a safety place, and his second problem is which type of dishs that the can make to pay attention of all categories of people.

1-3Target Audience:

To solve this problems, as data scientist my objective is to locate safety place in Toronto, and search in this place for what are the preferable dish for people in this place, by looking as the restaurants in this place what type of dish buy.

Please note that we gonna take into consideration population.

2-Data acquisition and cleaning 1 :

2.1 Data sources:

Source of our data : Crimes in Canada between 2014 and 2019 from: http://data.torontopolice.on.ca/datasets/neighbourhood-crime-rates-boundary-file-/data.

This first data contain the average of crimes for each type of crime for 5 years between 2014 and 2019 in all neighbors in Toronto and contain population too ,

Witch can make as to see the safe places with good population .

Neighbourhood	Population	Assault_AVG	AutoTheft_AVG	BreakandEnter_AVG	Homicide_AVG	Homicide_CHG	Robbery_AVG	TheftOver_AVG
Yonge-St.Clair	12528	31.0	4.3	23.3	0.0	0.0	5.7	4.3
York University Heights	27593	333.2	106.3	113.2	0.8	-1.0	75.8	36.3
Lansing-Westgate	16164	70.7	23.7	38.8	1.7	-1.0	14.7	7.0
Yorkdale-Glen Park	14804	160.2	55.5	63.3	1.2	-0.5	31.5	22.5
Stonegate-Queensway	25051	83.2	28.7	52.8	0.0	0.0	20.7	6.0

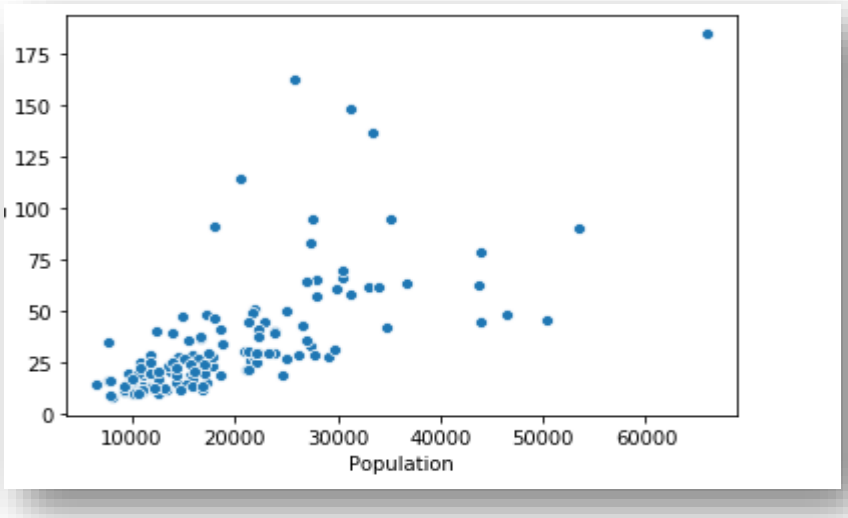
cleaning data : In the first data we gonna do a lot ,because the are many column that we gonna drop and add a column .First we gonna calculate the average of crime between 2014 and 2019,and drop all columns just leaving neighbors and population and the new column .

	Neighbourhood	Population	Total_avg
0	Yonge-St.Clair	12528	9.800000
1	York University Heights	27593	94.942857
2	Lansing-Westgate	16164	22.228571
3	Yorkdale-Glen Park	14804	47.671429
4	Stonegate-Queensway	25051	27.342857

3-Exploratory Data Analysis and modeling 1 :

Exploring:

First we gonna scatter the population and the average of crime.



As we see in this figure that number of population increase the avg of crime increase .

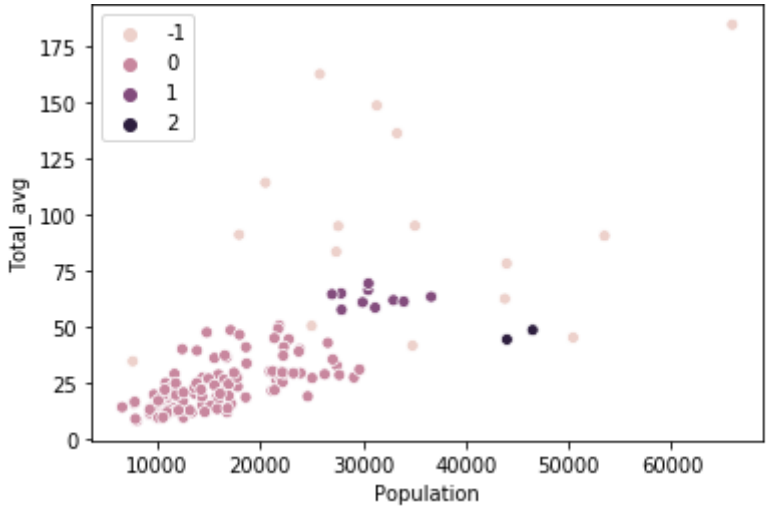
we see that there are some place with great numbers of population and the avg is low or under 50 crimes in 5 years witch good for us . we see some point that are clustered and like they have some similarly.

For that we need to cluster this data and choose the best cluster for as.

Modeling :

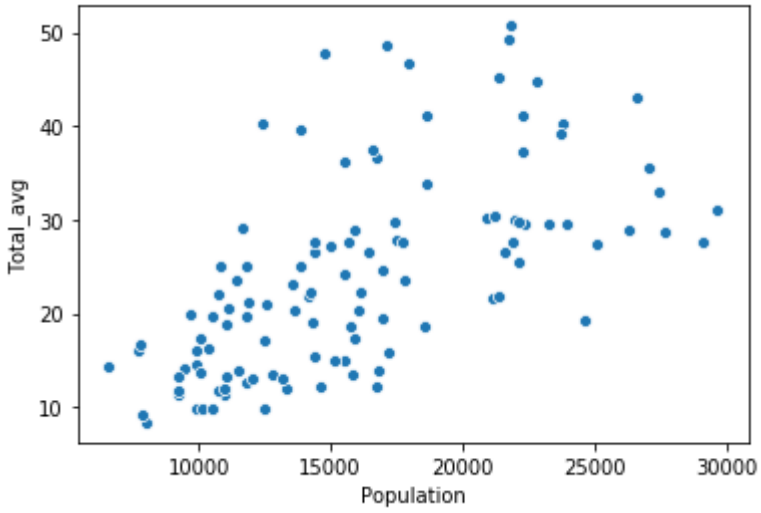
for this problem we gonna use DBSCAN to cluster this data , why ?

because DCSCAN we don't need to oblige the data to have a number of clusters and the number of cluster depend on data .

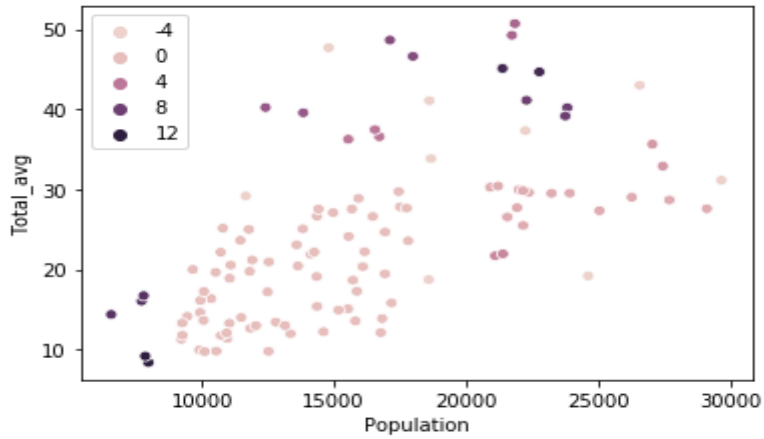


the result put as to choose cluster with label 0 because he is the one who give us place with great population with avg of crime under 50 crime in 5 years.

When replot the result we see:

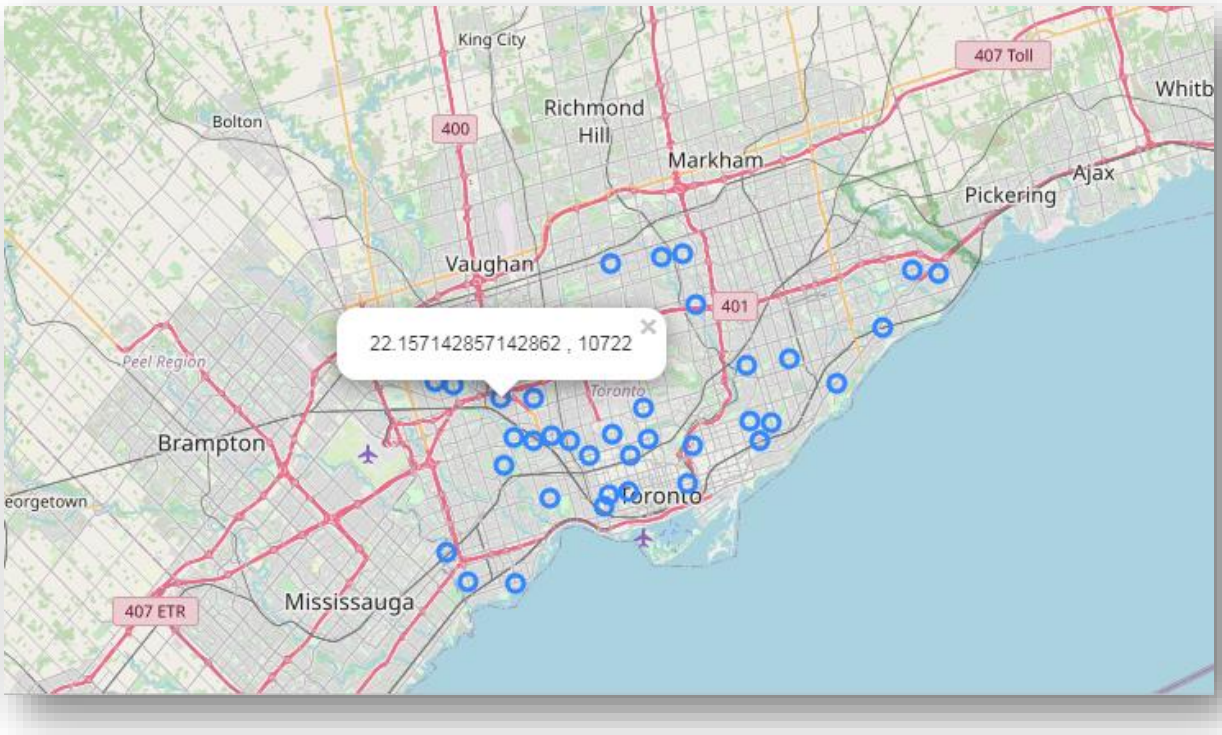


We ask a question can we modify this new data or not, why do not remove points which seems to be like noise .let's do it with DBSCAN again .



This result looks better then the others and this what we want , place with great numbers of population and descending avg of crime.

we gonna use geocoder to found latitude and longitude for every neighbor in this data which the safe places in Toronto ,giving by our model in this maps.



4-Data acquisition and cleaning 2:

Data sources:

After analyzing of this data we ganna achieve our first task ,then we gonna use the final data (contain only safe place in Toronto).Then we ganna use foursquare to search four restaurant in each neighbors.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	id	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Yonge-St.Clair	43.688019	-79.394571	4adc5a0bf964a520d12b21e3	The Bagel House	43.687374	-79.393696	Bagel Shop
1	Yonge-St.Clair	43.688019	-79.394571	4ae22cf6f964a520778b21e3	Cava Restaurant	43.689809	-79.394932	Tapas Restaurant
2	Yonge-St.Clair	43.688019	-79.394571	52b1e6a8498e43e8f94b4b25	9bars	43.688660	-79.391940	Café
3	Yonge-St.Clair	43.688019	-79.394571	4b15383bf964a52079a923e3	Capocaccia Café	43.685915	-79.393305	Italian Restaurant
4	Yonge-St.Clair	43.688019	-79.394571	526ecc2b498edb771548d405	DAVIDsTEA	43.688421	-79.394385	Tea Room

cleaning data: we need to clean our data from venue category that is not a restaurant .

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	id	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Yonge-St.Clair	43.688019	-79.394571	4ae22cf6f964a520778b21e3	Cava Restaurant	43.689809	-79.394932	Tapas Restaurant
1	Yonge-St.Clair	43.688019	-79.394571	4b15383bf964a52079a923e3	Capocaccia Café	43.685915	-79.393305	Italian Restaurant
2	Yonge-St.Clair	43.688019	-79.394571	5a119e738a6f175e860cac17	Mary Be Kitchen	43.687708	-79.395062	Restaurant
3	Yonge-St.Clair	43.688019	-79.394571	52acedc011d2bf5586251c3c	Union Social Eatery	43.687895	-79.394916	American Restaurant
4	Yonge-St.Clair	43.688019	-79.394571	4be349d763609c7439e11bff	Daeco Sushi	43.687838	-79.395652	Sushi Restaurant

4-Exploratory Data Analysis and modeling 2 :

Exploring:

When we analyzing we see:

Italian Restaurant	27
Restaurant	20
Sushi Restaurant	13
Fast Food Restaurant	12
Thai Restaurant	12
Korean Restaurant	11
Japanese Restaurant	10
Mexican Restaurant	10
Middle Eastern Restaurant	9
Indian Restaurant	7
American Restaurant	6
Vegetarian / Vegan Restaurant	5
Vietnamese Restaurant	5
French Restaurant	5
Seafood Restaurant	5
Caribbean Restaurant	4
Tapas Restaurant	3
Asian Restaurant	3

And that mean the fist nine types are dominating .

And we see Asiatique types dominating like Shushi Restaurant, we gonna regroup it in Asiatique type.

Finely our data will contain 5 type of restaurant.

Modeling:

Now we gonna use kmeans to cluster neighbors in our data and we gonna use 5 as number of cluster ,why when count number of type of restaurant in our data we see that 5 type are highly dominate .

And we get for each neighbors his labels with means the dominant type or types in this neighbors.

before that we gonna regroup data by the mean of each type for each neighbors

Neighborhood	Asiatique	Fast Food Restaurant	Italian Restaurant	Mexican Restaurant	Restaurant
Bathurst Manor	0.000000	1.000000	0.000000	0.000000	0.000000
Broadview North	0.000000	0.000000	0.000000	0.000000	1.000000
Corso Italia-Davenport	0.333333	0.000000	0.555556	0.111111	0.000000
Danforth East York	0.500000	0.000000	0.000000	0.500000	0.000000
Dufferin Grove	0.230769	0.076923	0.230769	0.153846	0.307692

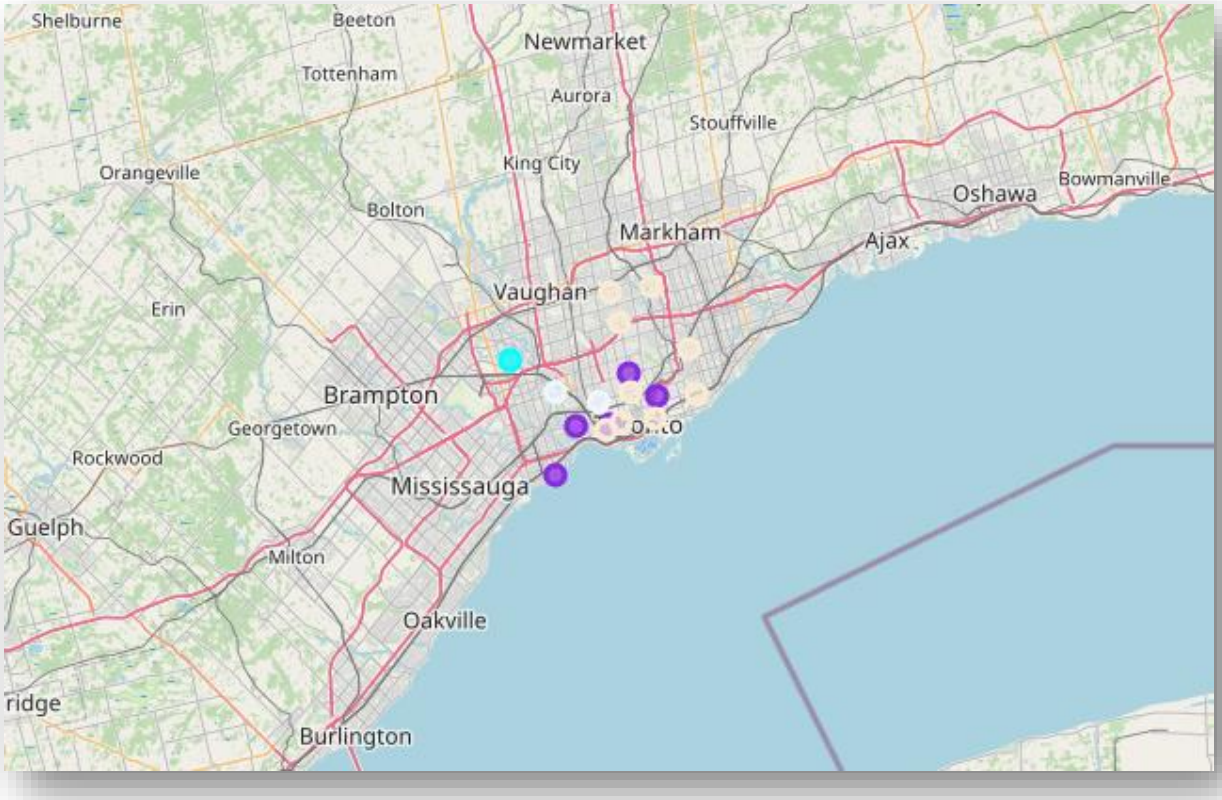
After clustering:

Neighborhood	Asiatique	Fast Food Restaurant	Italian Restaurant	Mexican Restaurant	Restaurant	labels
Bathurst Manor	0.000000	1.000000	0.000000	0.000000	0.000000	1
Broadview North	0.000000	0.000000	0.000000	0.000000	1.000000	4
Corso Italia-Davenport	0.333333	0.000000	0.555556	0.111111	0.000000	0
Danforth East York	0.500000	0.000000	0.000000	0.500000	0.000000	2
Dufferin Grove	0.230769	0.076923	0.230769	0.153846	0.307692	4
Hillcrest Village	0.500000	0.000000	0.000000	0.000000	0.500000	2
Humber Heights-Westmount	0.000000	0.000000	1.000000	0.000000	0.000000	0
Kingsway South	0.000000	1.000000	0.000000	0.000000	0.000000	1
Lansing-Westgate	1.000000	0.000000	0.000000	0.000000	0.000000	2
Lawrence Park North	0.000000	0.000000	0.000000	1.000000	0.000000	3
Little Portugal	0.571429	0.000000	0.142857	0.000000	0.285714	2
Long Branch	0.000000	0.000000	0.000000	1.000000	0.000000	3
Morningside	0.000000	0.000000	1.000000	0.000000	0.000000	0

Then we arriver to give each neighbor the dominating type or a types of restaurants .

0- Represent fast food and Italian restaurants
1- Represent fast food restaurants
2- Represent Italique and Asiatique restaurants
3- Represent Mexican restaurants
4- Represent all type restaurants

Then we put all this neighbors in map and each color define cluster.



5-conclusion:

Finally we arrived to solve our problems to found places in Toronto that are safe and we arrived to determine type of restaurants are dominating for each of this places