Professional Cleaning Salesperson Assignment

Introduction

My client is a professional cleaning firm, specialized in restaurants. They are open new business and seeking to take new clients in Ottawa-ON. They requested me to build a list of potential clients to visit and offer their services. They have four salesperson and they want to assign them to different areas of the city. My job is to search for restaurants in Ottawa and group it in four areas to each salesperson visit.

Analytic approach

The fastest approach to address the problem is to load a dataset of restaurants from foursquare and cluster it using some clustering algorithms. Searching for restaurants at Foursquare is a pretty easy task. However, the client cannot operate in Quebec, so, the neighbouring city Gatineau-QC should not be visited by any salesperson. So, any restaurant in Quebec should be filtered from the list prior to clustering. The algorithm chosen to perform the clustering is Birtch, so the closest restaurants should be visited by the same salesperson. This algorithm was selected because it can cluster according to geographic shape.

Data Requirements

The dataset is going to be collected using Foursquare API, using a area between two coordinates, that includes the city of Gatineau-QC. But, the dataset should contain only venues from Ottawa. The Dataset should contain the name of the venue, type of restaurant, GPS location and address.

Data Collection

The dataset was collected from Foursquare API, which results in 179 unique rows. But some cleaning should be made, such as, filtering the venues that is not in Ottawa, and removing duplicates.
The final dataset contains 160 inputs and 9 columns. There are information about the location, that is going to be used for clustering, name, and category, that is useful for the final report.
As soon as the dataframe was concluded, it was uploaded to IBM-DB2 server on the cloud.

The sample of the dataframe:

	id	name	lat	lng	address	city	state	categories
0	582670997d7fb761d064e9e9	Sutherland Restaurant Bar and Coffee House	45.443455	-75.670020	224 Beechwood Avenue	Ottawa	ON	Breakfast Spot
1	4c0a619d7e3fc928ae70f382	McDonald's	45.441750	-75.643841	594 Montreal Rd.	Ottawa	ON	Fast Food Restaurant
2	4ba405edf964a520d07838e3	Pilo's Greek Restaurant	45.446167	-75.630297	876 Montreal Rd	Ottawa Division	ON	Greek Restaurant
3	52d71865498e489acd17525d	Le restaurant-école Les Jardins de la Cité	45.439384	-75.627370	La Cité collégiale, 801	Ottawa	ON	Restaurant
4	4b5f23d4f964a5208ea829e3	Navarra	45.429785	-75.693345	93 Murray St.	Ottawa	ON	Spanish Restaurant

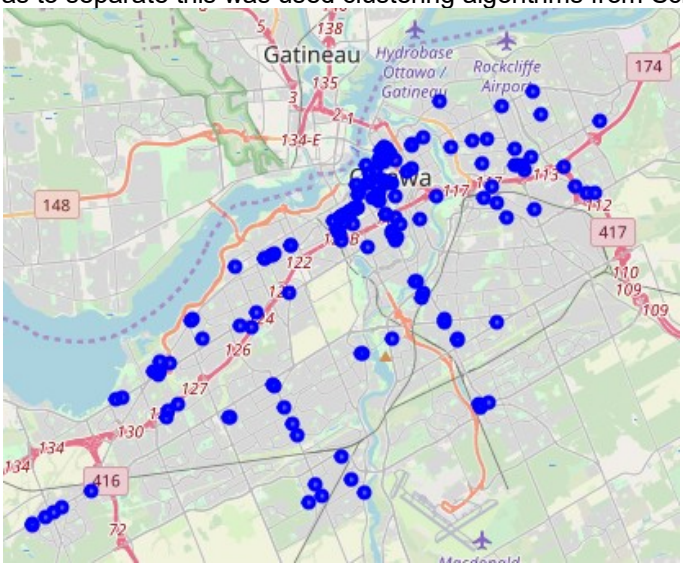
Methodology

Database

In this project the dataframe was uploaded to the cloud, because it is easier to exchange with many scripts and it is safer than keep it. IBM-DB2 database was chosen to store the data.
The script was split in two parts; the first one is focused on building the dataframe and the second one to perform the clustering algorithm. The resulted dataframe also has uploaded to the cloud, so it is accessible in any computer.

Data Analysis

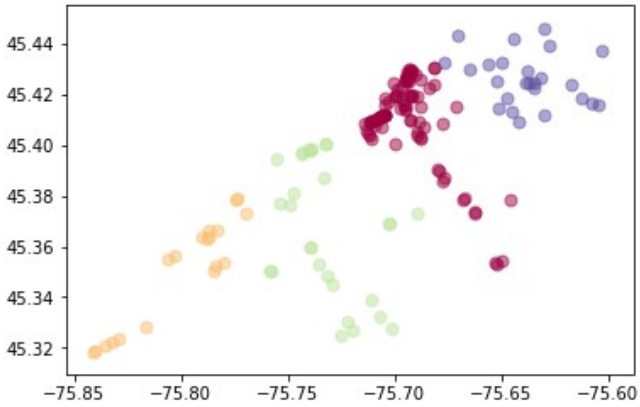
Exploring the locations of restaurants in Ottawa is clear that the majority of them are close to Downtown. But also has regions outside from city centre that has some agglomeration. To identify the better was to separate this was used clustering algorithms from Scikit-Learn toolbox.



The main information from dataframe for clustering the venues is the GPS location, shown as two columns: latitude and longitude.

Clustering Models

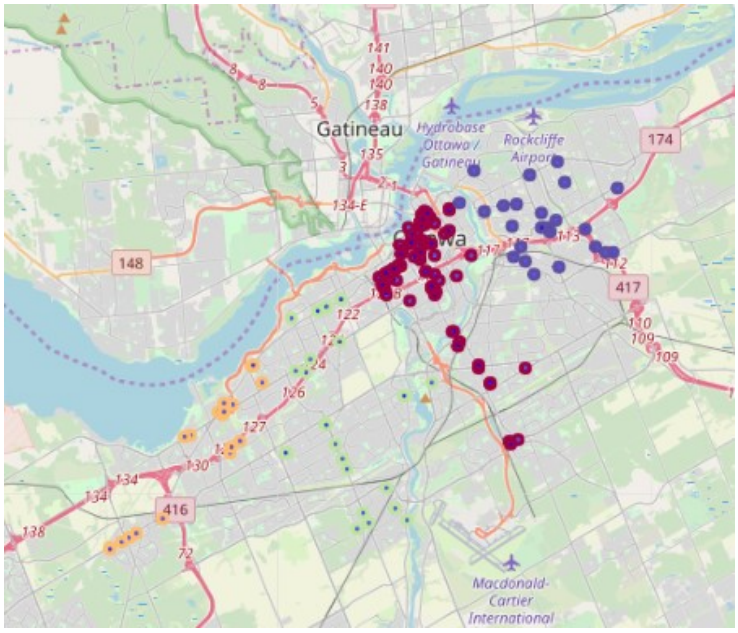
Were applied clustering models such as BDSCAN, K-Means and Birtch, all those have the same problem of aggregating too many venues over one cluster in Downtown. The result from Birtch was chosen because it has less inequality between clusters.



The scatter plot was useful to make a preview where the clusters have been assigned.

Results

The final result:



Were built four dataframes with address, location, category and name of the restaurants and this is to deliver to client. Some samples of those are below:

	index	ID	NAME	LAT	LNG	ADDRESS	CITY	STATE	CATEGORIES	Cluster
4	4	4b5f23d4f964a5208ea829e3	Navarra	45.429785	-75.693345	93 Murray St.	Ottawa	ON	Spanish Restaurant	0
5	5	58c098aa01f43341813efdd0	Sur-Lie	45.429970	-75.692940	111, Murray Street	Ottawa	ON	French Restaurant	0
6	6	4b5347adf964a520949527e3	Saffron Restaurant	45.430578	-75.681453	426 Rideau St.	Ottawa	ON	Middle Eastern Restaurant	0
7	7	4b9c3786f964a520e95536e3	Sitar Indian Restaurant	45.430496	-75.681453	417 Rideau St.	Ottawa	ON	Indian Restaurant	0
13	13	4c1e0804b4e62d7f0c9bdd93	73 North Restaurant	45.423101	-75.696786	73 Sparks St	Ottawa	ON	Middle Eastern Restaurant	0

Cluster 0

	index	ID	NAME	LAT	LNG	ADDRESS	CITY	STATE	CATEGORIES	Cluster
115	115	4b0da39af964a520e74c23e3	Mugena African Caribbean Foods & Restaurant	45.378573	-75.774348	911 Richmond Rd.	Ottawa	ON	Caribbean Restaurant	1
116	116	4db8c8284df0ded98bc535c7	Ho-Ho Chinese Restaurant	45.379299	-75.773672	875 Richmond Road	Ottawa	ON	Chinese Restaurant	1
123	123	50f1a1e4e4b0d37546f3e928	Carlingwood Family Restaurant	45.366349	-75.783176	Lincoln Heights, Britannia Bay	Ottawa	ON	Diner	1
124	124	4b62f8c1f964a5208a5b2ae3	Cora's Breakfast & Lunch	45.366697	-75.787195	1355 Richmond Rd	Ottawa	ON	Breakfast Spot	1
125	125	4b65c745f964a52026ff2ae3	Carlingwood Family Restaurant	45.373222	-75.769341	2121 Carling Avenue	Ottawa	ON	Diner	1

Cluster 1

	index	ID	NAME	LAT	LNG	ADDRESS	CITY	STATE	CATEGORIES	Cluster
101	101	522bc6c9498ee49645ff8664	moe's world famous Newport restaurant	45.394624	-75.755414	322, Churchill Avenue North	Ottawa	ON	Pizza Place	2
102	102	56b65d17498ef040e44a6998	African Village Restaurant	45.398153	-75.739573	1356 Wellington	Ottawa	ON	Ethiopian Restaurant	2
103	103	5455886f498eefcc8498012c	Viet Taste Restaurant	45.398171	-75.740157	1395 Wellington West	Ottawa	ON	Vietnamese Restaurant	2
104	104	5a49aac5b6eedb5d70d32cce	Napoli's Restaurant Pizza & Pasta	45.396812	-75.743123	81 Richmond Rd	Ottawa	ON	Restaurant	2
105	105	4bf3122c94af2d7fa9133972	Napolis Italian Restaurant	45.396936	-75.743153	81 richmond rd	Ottawa	ON	Pizza Place	2

Cluster 2

	index	ID	NAME	LAT	LNG	ADDRESS	CITY	STATE	CATEGORIES	Cluster
0	0	582670997d7fb761d064e9e9	Sutherland Restaurant Bar and Coffee House	45.443455	-75.670020	224 Beechwood Avenue	Ottawa	ON	Breakfast Spot	3
1	1	4c0a619d7e3fc928ae70f382	McDonald's	45.441750	-75.643841	594 Montreal Rd.	Ottawa	ON	Fast Food Restaurant	3
2	2	4ba405edf964a520d07838e3	Pilo's Greek Restaurant	45.446167	-75.630297	876 Montreal Rd	Ottawa Division	ON	Greek Restaurant	3
3	3	52d71865498e489acd17525d	Le restaurant-école Les Jardins de la Cité	45.439384	-75.627370	La Cité collégiale, 801	Ottawa	ON	Restaurant	3
8	8	4b6f374bf964a520bde42ce3	Habesha	45.432539	-75.676591	574 Rideau St.	Ottawa	ON	Ethiopian Restaurant	3

Cluster 3

Discussion

Analysing the results, mainly the map where the clusters were allocated, is possible to see some problems with the model. One problem is because the algorithm does not take into consideration physical barriers, such as, rivers and places that are not connected by roads. So it can assign two venues in the same cluster that there is no direct path between them. Another problem is the excessive amount of venues assigned in one cluster. For the purpose proposed by the client, it is the problem of overload one salesperson.

Conclusion

The clustering was useful to perform the task. However, itself is not enough to solve the issue presented by the client, some other features, mainly the road map, should be used to truly connect venues together. And improvements on the algorithm used should be done to avoid overload.