Market Segmentation for

Senses Coffee company

Hayfa Bu Hazzaa

1. Introduction

1.1 Background

Senses is a coffee company that produce three brands of coffee, Ellite, Instance, and Chill. It is based on US. The owners and shareholders decided that it is the time to expand and penetrate another market. Canada is the first option to investigate since it is near to US and the market is somehow similar in taste to US, so no need to customize the product to suite the target country culture. Canada is very big country that is why it was initially decided to dedicate three offices to manage the operations in the country. Also it was decided to focus on specific market segments which are Hotels, Bakeries and Grocery Stores. The reasons of this selection are:

1. Senses Coffee has three main coffee brand that dedicated to Hotels, Bakeries and grocery stores.

	*Ellite* is fine coffee suitable for hotels markets,
	*Instance* is average coffee that is suitable from taste and price for Bakeries markets,
	***Chill*** is an iced coffee that is very popular with good price and suitable for Grocery stores markets.

2. Canada is a very big country, and to penetrate it you need first to test the market and focus on the most venues that would generate good money.

The company investors are willing to invest in Canada market, but still need to see solid data about Canada market and specifically about the three market segments it will target. To be able to estimate the amount of the investment needed based on the initial market size. Data scientist team was hired to provide the investors

with needed data. Data scientist team will need to collect the concerned data, clean it and then analyse it, this shall provide the needed insight for the investors.

1.2 Problem

Senses Coffee company hired data scientist to collect and analyse the data based on geo-demographic market segmentation. to provide insight that will drive the decision of entering the Canadian market and segment it across the three operation offices.

Essential data which will contribute to the market segmentation process includes: Canadian municipalities, Canadian municipalities' geographical coordinates (i.e. the longitudes and latitude data of the municipality), and the number and the category of potential customers in each municipality like hotels, Grocery Store and Bakery).

This project is a data clustering project and it is aimed to segment the Canadian municipalities into three marketing segments (i.e. 3 clusters). Although the above-mentioned data will contribute in the segmentation process, the segmentation itself will be done according to the number and the category of potential customers in each municipality (i.e. the number of hotels, the number of Grocery stores and the number of Bakeries). In this approach, each operation office will operate a single of the new market.

Figure 1 below views an example of the potential market segmentation, and Figure 2 views an example of a geo-demographic market segmentation based on this descriptive analytics data clustering approach.

Figure 1 - Potential market segmentation

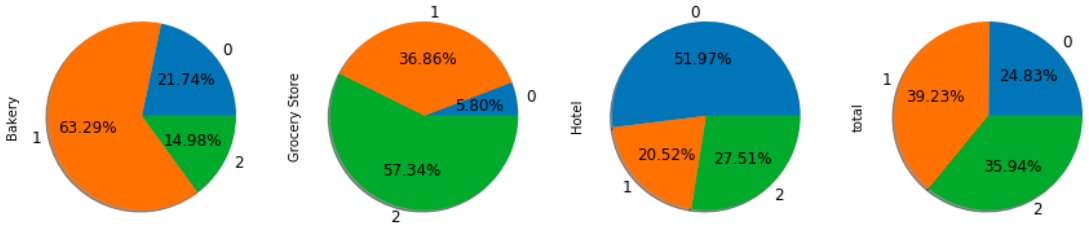
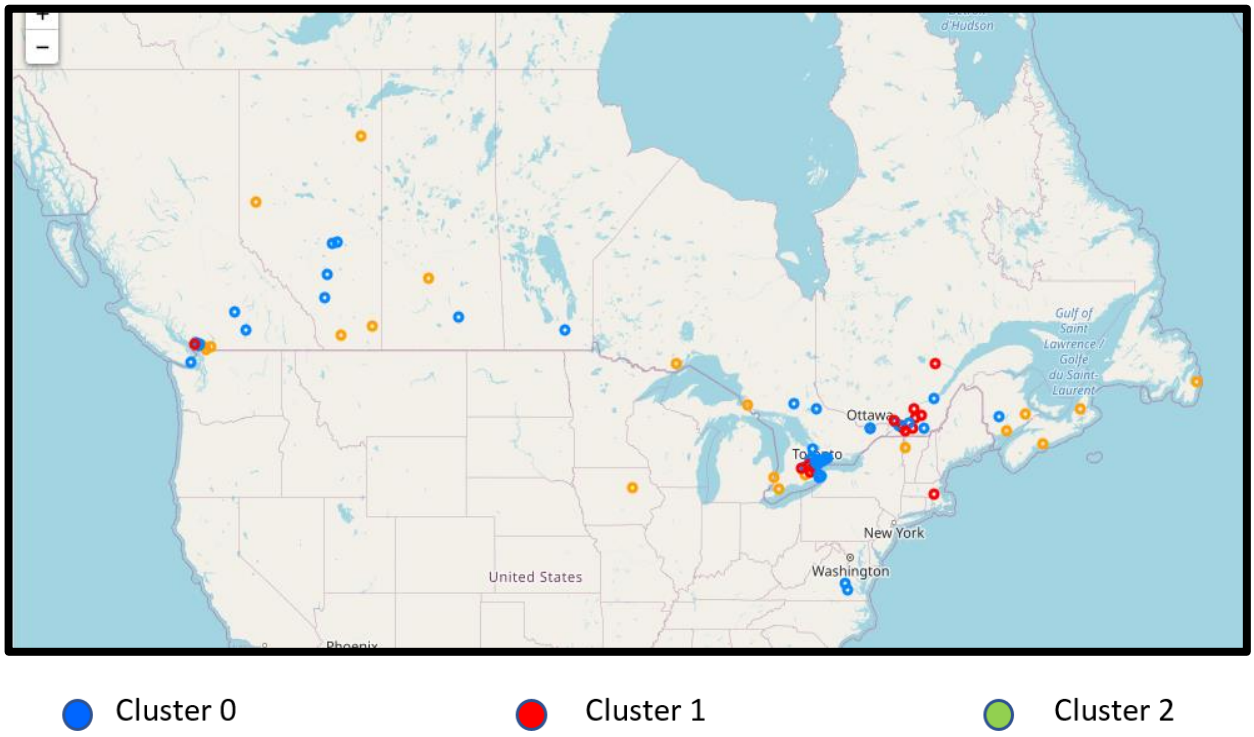


Figure 2 - Geo-demographic market segmentation



1.3 Interest

Senses Coffee company needs to select the municipalities that will be targeted by operation offices. Therefore, the company management and operation offices are interested to know the size of the new potential market in each segment. Other audiences who care about this problem include the company shareholders and Human capital department (see Table 1 below for an example).

Table 1 - An example of potential market segment size

(expressed in number of customers in each business line in each cluster)

Cluster Labels	Bakery	Grocery Store	Hotel
0	58	205	76
1	139	65	66
2	18	18	99

2. Data acquisition and cleansing

2.1 Data sources

Table 2 below describes the datasets used to build the clusters and their corresponding data sources.

Table 2 - the datasets and their data sources

No	Dataset	Description	Data Source
1	List of the largest 100 municipalities in Canada	Data fields include Municipalities, Province, Growth rate and population. See Appendix I for an example of this dataset.	I scraped the following Wikipedia site to obtain this data https://en.wikipedia.org/wiki/List_of_the_100_largest_municipalities_in_Canada_by_population
2	Geo-Location data of each municipalities in Canada	Data fields include the longitude and latitude coordinates of each municipality. See Appendix II for an example of this dataset.	I obtained this data using the Python geocoding web services API.
3	Potential customers' data	Data fields include the venue name, category, longitude and latitude, See Appendix III for an example of this dataset.	I obtained this data by exploring the municipality's venues using the Foursquare API
4	Canada map GIS data	Data of Canada with the largest municipalities. See Appendix IV for an example if this dataset.	I obtained this data using the Folium API

2.2 Data Cleansing

I followed below steps to clean the data in order to be ready for further analysis.

1. I scraped the data of 100 largest municipalities in Canada from the Wikipedia page using Python. After investigation, I discovered that the column names in the Wikipedia page are not put in standard naming convention. a column name Population(2016) use special characters, and this jeopardize the Python program code. So, I modified the column name to Population2016 to include only the standard alphabetic character set.
2. I inserted the latitude and longitude coordinate columns structure to the data frame structure of the table read from the Wikipedia page.

3. I added the coordinates data from the geocoding web services and include it in the data frame.
4. Then I checked for any missing coordinates to drop nan values cells. Fortunately, all coordinates data were successfully retrieved by the API. Then I combined the venue data with the location data and the master data acquired from the Wikipedia (Table 3).
5. I then used Folium to create Canada map with all municipalities superimposed on top and used this map to visually verify the correctness of acquired data on the map (see Appendix IV).

Table 3 - Combined Wikipedia data, location data and venue data

	Municipality	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
8251	North Vancouver	49.320713	-123.073783	Indian Fusion	49.327991	-123.072434	Indian Restaurant
8252	North Vancouver	49.320713	-123.073783	Earnest Ice Cream	49.312330	-123.079756	Ice Cream Shop
8253	North Vancouver	49.320713	-123.073783	iLoveKickboxing.com Vancouver	49.283143	-123.100976	Boxing Gym
8254	North Vancouver	49.320713	-123.073783	Waterfront Park	49.311548	-123.085778	Park
8255	North Vancouver	49.320713	-123.073783	Coal Harbour Seawall	49.291304	-123.123276	Trail
8256	North Vancouver	49.320713	-123.073783	Fairmont Pacific Rim	49.288227	-123.116932	Hotel
8257	North Vancouver	49.320713	-123.073783	Revolver	49.283187	-123.109288	Coffee Shop
8258	North Vancouver	49.320713	-123.073783	Stanley Park	49.302488	-123.141718	Park
8259	North Vancouver	49.320713	-123.073783	Loden Hotel	49.287690	-123.123574	Hotel
8260	North Vancouver	49.320713	-123.073783	Rosewood Hotel Georgia	49.283429	-123.118911	Hotel
8261	North Vancouver	49.320713	-123.073783	Lynn Headwaters Regional Park	49.355070	-123.023958	Trail

2.3 Feature Selection

After data cleansing, there were 8719 venues listed under all categories. First, I took out the duplicated ro ws it was 100 rows and I took it out. To end up with total 8619 venues. Then I listed the unique categories (310 unique categories), this is important to select the right category name that I need to keep and drop ot hers.

Table 4- the first 10 rows in the venue categories

	Venue Category	Count_Category
0	Park	630
1	Coffee Shop	470
2	Café	341
3	Restaurant	298
4	Grocery Store	288
5	Brewery	246
6	Hotel	236
7	Bakery	210
8	Ice Cream Shop	204
9	Pizza Place	171

The above table gives the confidence that the Canadian market is good and there are many potential hotels , Grocery stores and Bakery in Canada. Then I dropped all venues except the venues with categories (Hotels, Grocery Store and Bakery), the numbers of each venue category are shown in table 4.

Table 4 – Venues counts of Grocery Store, Hotels, Bakery

	Venue Category	Count_Category
4	Grocery Store	288
6	Hotel	236
7	Bakery	210

3. Methodology

3.1 Exploratory data analysis

3.1.1 Master dataset

The master dataset includes Canada municipalities, the province, Growth rates, Population 2016, and the municipalities coordinates (longitude and latitude). After combining the master dataset as explained in section 2, the master dataset was explored by printing the master dataset data frame, obtaining its summary information and displaying each municipality on Canada map. Table 5 shows a sample of the master dataset and its attributes. Figure 3 shows that the master dataset includes 100 municipalities, and Figure 4 depicts the location of each municipality on Canada map.

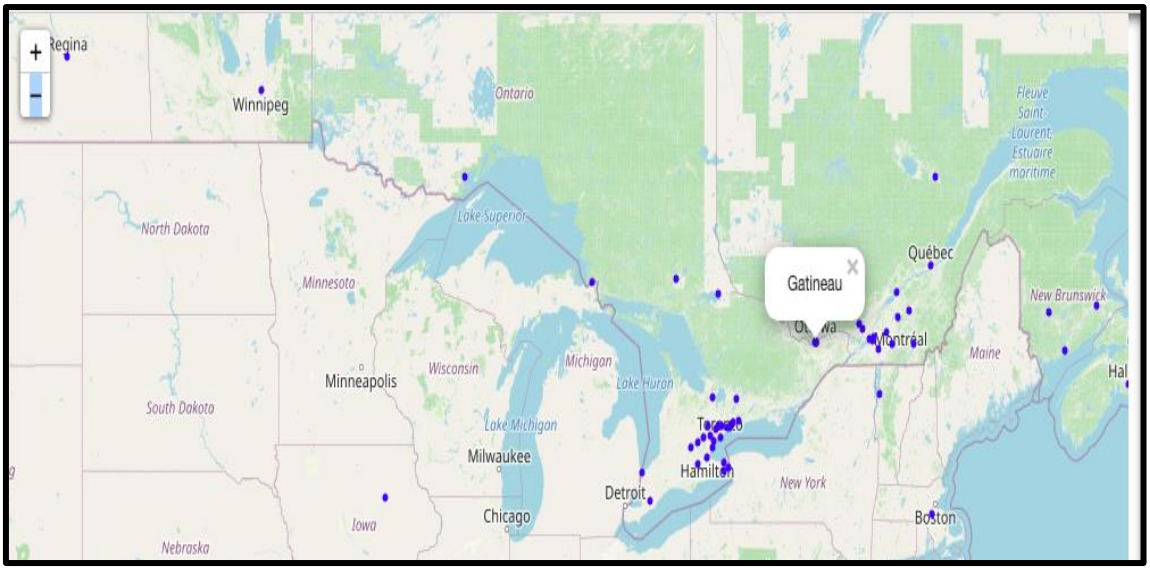
Table 5 - Sample of the master dataset and the master dataset attributes

	Municipality	Province	Land area(km2, 2011)	Growth Rate 2011–2016	Population2016	lat	lng
0	Toronto	Ontario	630.20	4.46%	2731571	43.6535	-79.3839
1	Montreal	Quebec	365.10	3.34%	1704694	45.4972	-73.6104
2	Calgary	Alberta	825.30	12.99%	1239220	51.0534	-114.063
3	Ottawa	Ontario	2790.20	5.76%	934243	45.4211	-75.6903
4	Edmonton	Alberta	684.40	14.82%	932546	53.5354	-113.508
5	Mississauga	Ontario	292.40	1.14%	721599	43.5903	-79.6457
6	Winnipeg	Manitoba	464.10	6.27%	705224	49.8955	-97.1385
7	Vancouver	British Columbia	115.00	4.64%	631486	49.2609	-123.114
8	Brampton	Ontario	266.30	13.31%	593638	43.6858	-79.7599
9	Hamilton	Ontario	1117.20	3.26%	536917	43.2561	-79.8729
10	Quebec City	Quebec	454.10	2.96%	531902	46.826	-71.2352

Figure 3 – Master Dataset Information (100 Municipalities)

```
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 7 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   Municipality           100 non-null    object
 1   Province               100 non-null    object
 2   Land area(km2, 2011)   100 non-null    float64
 3   Growth Rate 2011–2016  100 non-null    object
 4   Population2016         100 non-null    int64
 5   lat                    100 non-null    object
 6   lng                    100 non-null    object
dtypes: float64(1), int64(1), object(5)
memory usage: 5.6+ KB
```

Figure 4 –Location of each municipality on Canada map



3.1.2 Venues dataset

The venues dataset of Canada municipalities includes: the municipality name, municipality longitude, municipality latitude, the venue name, venue longitude, venue latitude and venue category. After combining the venue dataset with the master dataset as explained in section 2, the venue dataset was explored using the following descriptive statistics charts and summary information printouts:

- A printout of sample venue dataset and its attributes (Table 6).
- Venues raw dataset summary information (Figure 5). This dataset includes the venue information before dropping the duplicate rows. This figure shows that the venue raw dataset includes 8671 records.
- Venues dataset summary information printout (Figure 6). This dataset includes the venue information after dropping the duplicate rows. This figure shows that the venue raw dataset includes 8571 records.

- A bar chart that depicts the total number of venues for each Canadian municipality (Figure 7)
- Venues dataset descriptive statistics (Figure 8)
- Venues dataset logarithmic scale histogram (Figure 9)

Table 6 - Sample of the venue dataset and its attributes

	Municipality	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Toronto	43.653482	-79.383935	Downtown Toronto	43.653232	-79.385296	Neighborhood
1	Toronto	43.653482	-79.383935	Byblos Toronto	43.647615	-79.388381	Mediterranean Restaurant
2	Toronto	43.653482	-79.383935	Elgin And Winter Garden Theatres	43.653394	-79.378507	Theater
3	Toronto	43.653482	-79.383935	Art Gallery of Ontario	43.654003	-79.392922	Art Gallery
4	Toronto	43.653482	-79.383935	St. Lawrence Market (South Building)	43.648743	-79.371597	Farmers Market
5	Toronto	43.653482	-79.383935	Hailed Coffee	43.658833	-79.383684	Coffee Shop
6	Toronto	43.653482	-79.383935	Alo	43.648574	-79.396243	French Restaurant
7	Toronto	43.653482	-79.383935	Delta Hotels by Marriott Toronto	43.642882	-79.383949	Hotel
8	Toronto	43.653482	-79.383935	Yeti Nails & Spa	43.647938	-79.396330	Cosmetics Shop
9	Toronto	43.653482	-79.383935	Pai	43.647923	-79.388579	Thai Restaurant

Figure 5 – Venues raw dataset summary information

(8671 Canadian municipalities before dropping the duplicate rows)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8671 entries, 0 to 8670
Data columns (total 7 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   Municipality     8671 non-null   object
 1   Latitude         8671 non-null   float64
 2   Longitude        8671 non-null   float64
 3   Venue            8671 non-null   object
 4   Venue Latitude   8671 non-null   float64
 5   Venue Longitude  8671 non-null   float64
 6   Venue Category   8671 non-null   object
dtypes: float64(4), object(3)
memory usage: 474.3+ KB
```

Figure 6 – Venues dataset summary Information
(8571 Municipalities after dropping the duplicate rows)

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 8571 entries, 0 to 8670
Data columns (total 7 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   Municipality     8571 non-null   object
 1   Latitude         8571 non-null   float64
 2   Longitude        8571 non-null   float64
 3   Venue            8571 non-null   object
 4   Venue Latitude   8571 non-null   float64
 5   Venue Longitude  8571 non-null   float64
 6   Venue Category   8571 non-null   object
dtypes: float64(4), object(3)
memory usage: 535.7+ KB
```

Figure 7 – Sample of the total number of venues for each Canadian municipality
(after dropping the duplicate rows)

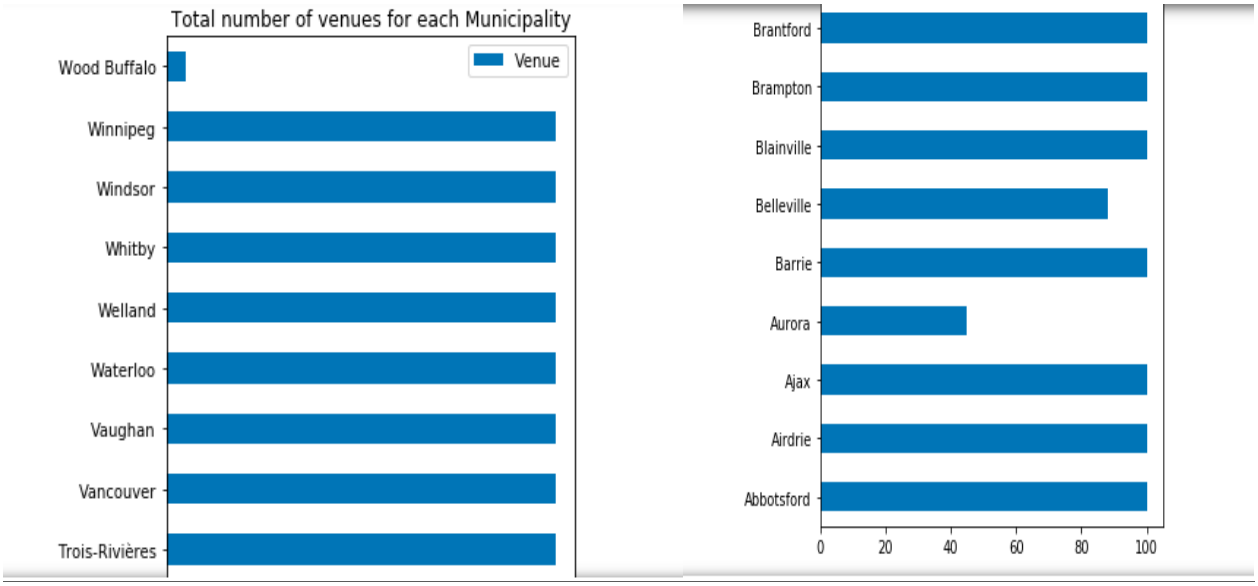
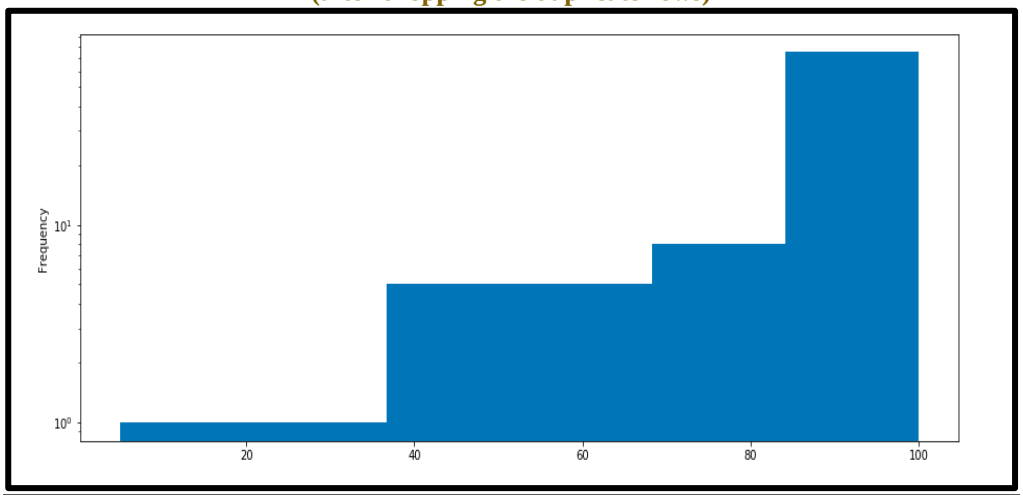


Figure 8 – Venues dataset descriptive statistics (after dropping the duplicate rows)

	Latitude	Longitude	Venue Latitude	Venue Longitude
count	8698.000000	8698.000000	8698.000000	8698.000000
mean	45.245500	-75.895224	45.227122	-75.914225
std	8.458246	41.698651	8.464220	41.709157
min	-36.598610	-123.937972	-36.925230	-124.554516
25%	43.685815	-97.138458	43.649963	-97.210771
50%	45.627484	-79.383935	45.524161	-79.385297
75%	49.243380	-72.940636	49.278501	-73.158504
max	57.652783	144.678005	57.377950	145.154135

Figure 9– Venues log-scale histogram (after dropping the duplicate rows)

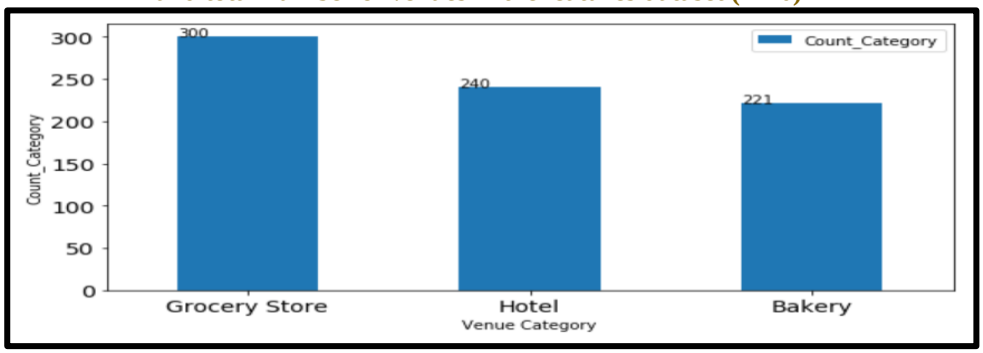


3.1.3 Exploring the features dataset

The features dataset is a subset of the venues' dataset. It includes the hotels venues, the bakeries venues and grocery stores venues. It consists of 761 cleansed data records. These records were obtained by filtering the 8698 records of venues dataset to obtain only the venues belonging to the hotels, bakeries and the grocery stores categories. The selected features dataset was explored using the following descriptive statistics charts and summary information printouts:

- A bar chart that shows the total number of each venue category in the selected features dataset (i.e. the hotels, grocery stores and bakery), and a printout of the total number of all venues in the selected features dataset (Figure 10).

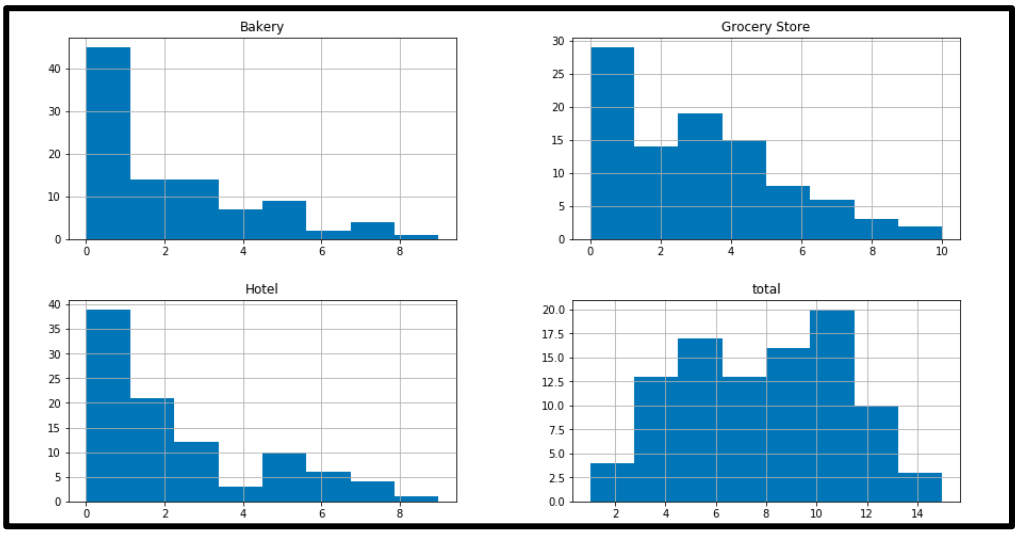
Figure 10 - Number of venues in each feature category, and total number of venues in the features dataset (2220)




```
Venue Category    Grocery StoreHotelBakery
Count_Category                         761
dtype: object
```

- Four histograms that depict characteristics of the selected features dataset, i.e. the hotels, grocery stores, bakeries and the total number of selected features (Figure 11).

Figure 11 - Selected features dataset histograms



3.1.4 Exploring the relationship between the municipalities and the features

The relationship between the municipalities and each feature in the features dataset is explored using the following descriptive statistics charts, maps and summary information printouts:

- A bar chart that shows the relationship between the municipalities and each feature in the features dataset (i.e. the number of hotels, the number of grocery stores and the number of bakeries in the municipalities). This bar chart is shown in Figure 12.
- A map that shows the density of total number of venues in each municipality (i.e. the number of hotels + the number of grocery stores + the number of bakeries). In order to show the density of venue distribution across the municipalities, the data was classified into three type of densities (low density, medium density and high density) and a colour coding was used to depict each type of density on the map (Figure 13).

Figure 12 - Relationship between the municipality and each selected feature

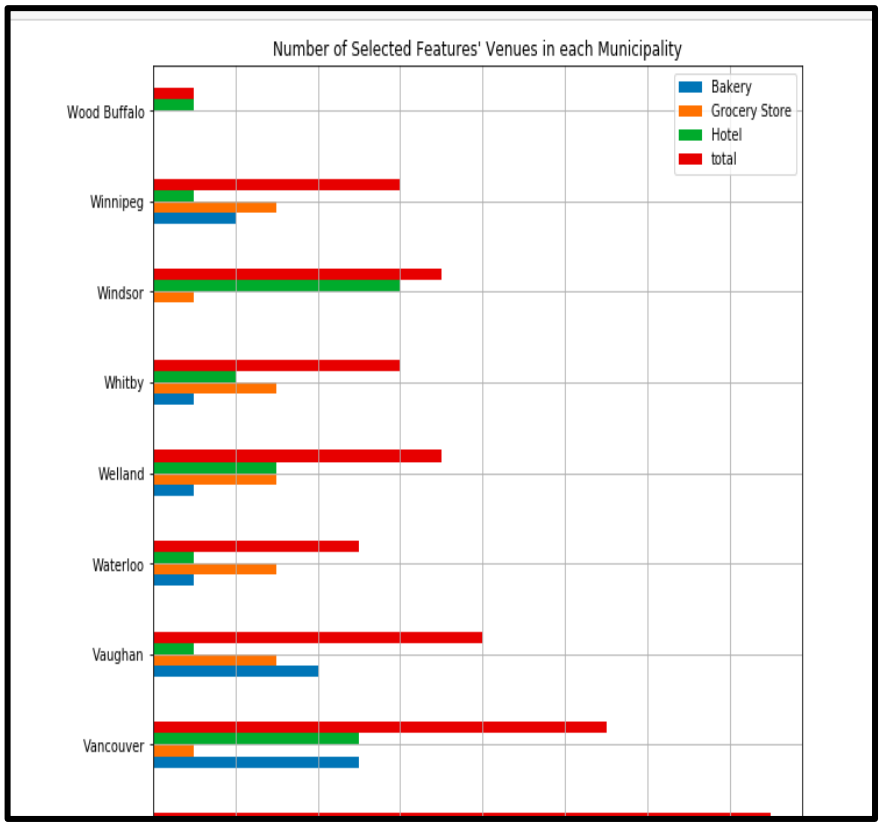
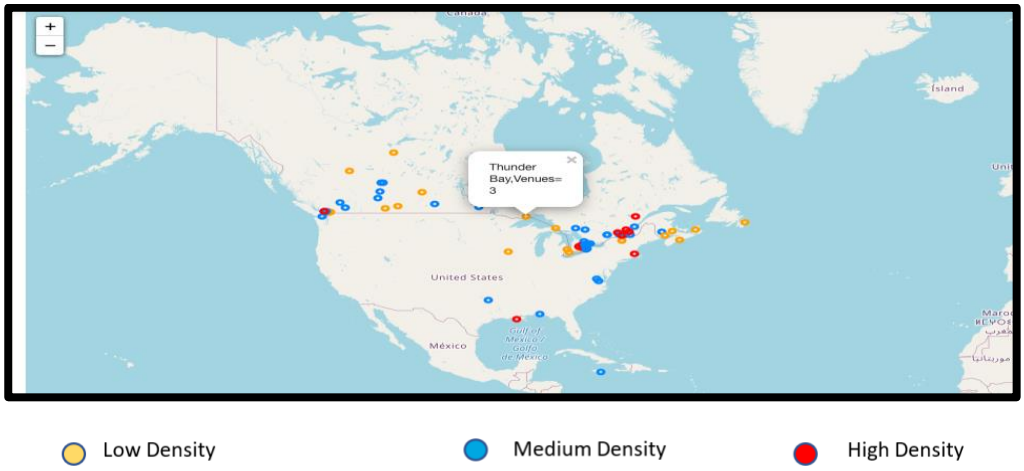
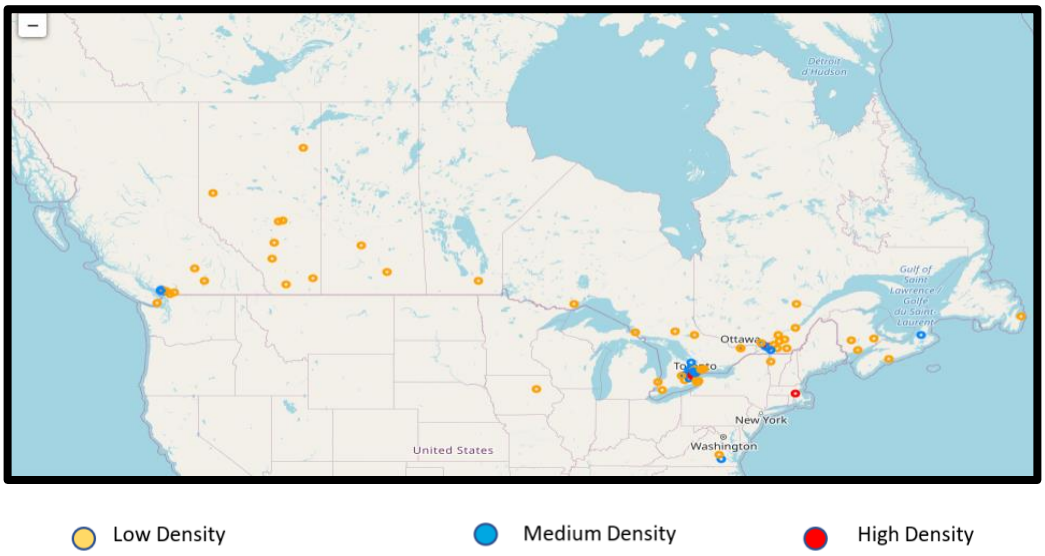


Figure 13 - Density of total number of venues in each municipality



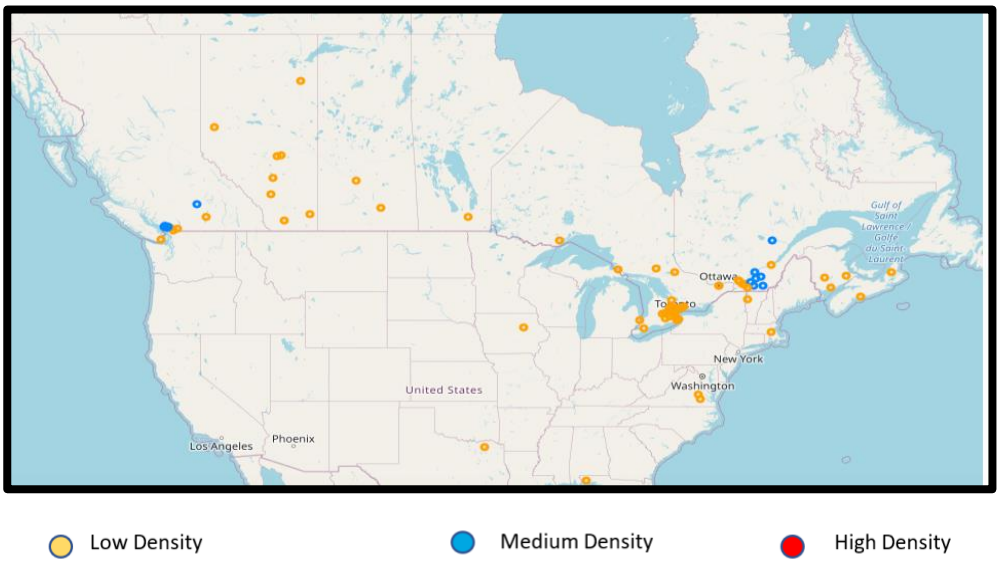
- A map that shows the density of the bakeries in each municipality. In order to show the density of the bakerys' distribution across the municipality, the data was classified into three type of densities (low density, medium density and high density) and a colour coding was used to depict each type of density on the map (Figure 14).

Figure 14 - Density of total number of bakeries in each municipality



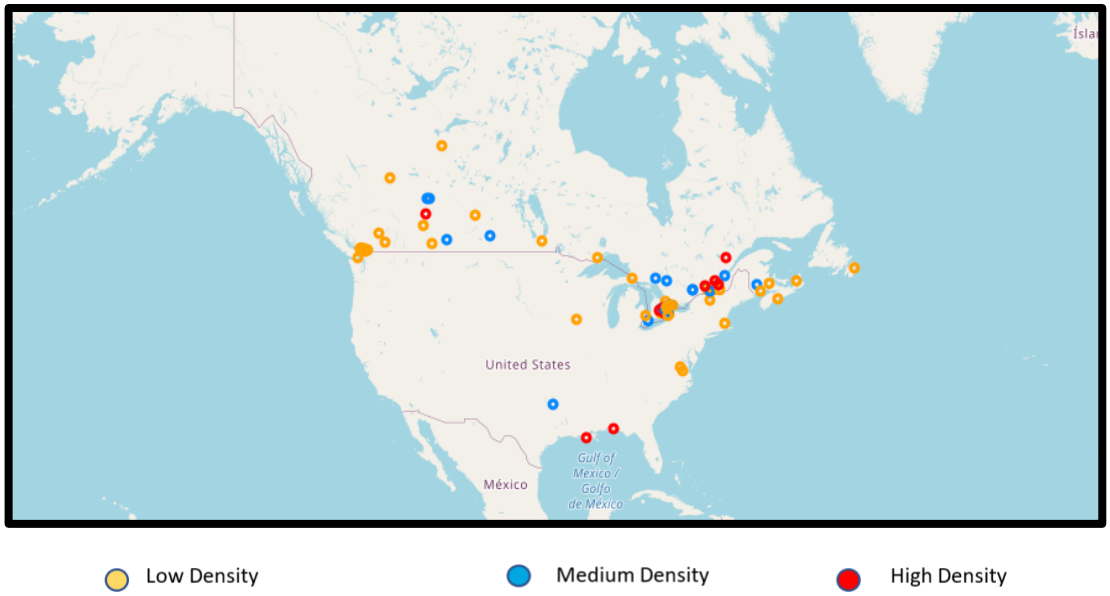
- A map that shows the density of the hotels in each municipality. In order to show the density of the Hotels distribution across the municipality, the data was classified into three type of densities (low density, medium density and high density) and a colour coding was used to depict each type of density on the map (Figure 15).

Figure 15 - Density of total number of Hotels in each municipality



- A map that shows the density of the Grocery Store in each municipality. In order to show the density of the Grocery Stores distribution across the municipalities, the data was classified into three type of densities (low density, medium density and high density) and a colour coding was used to depict each type of density on map (Figure 16).

Figure 16 - Density of total number of Grocery Stores in each municipality

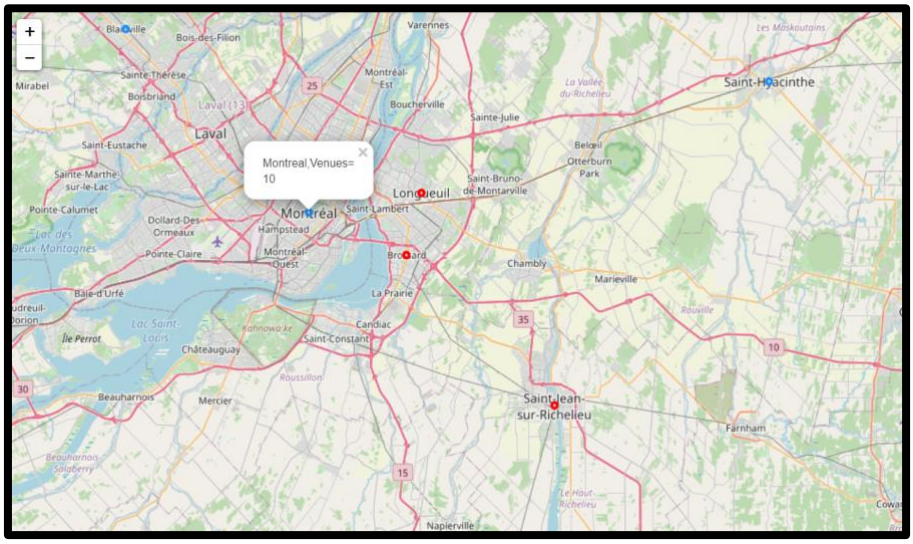


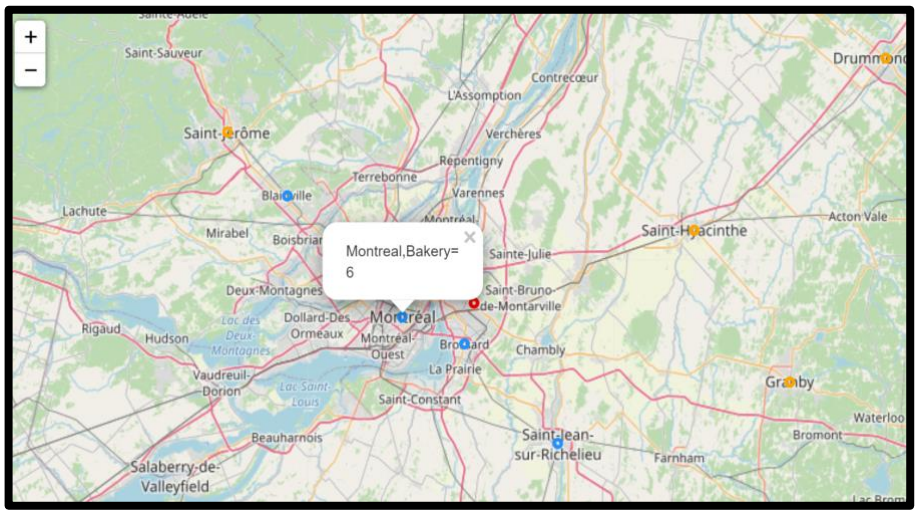
To verify the correctness on the data displayed on the above mentioned four maps, the features data of Montreal municipality was extracted from the features dataset and displayed in the form of a data frame and compared with the corresponding data displayed on the maps. Figure 17 depicts the features data of Montreal municipality in a data frame and the corresponding number of features in four separate maps.

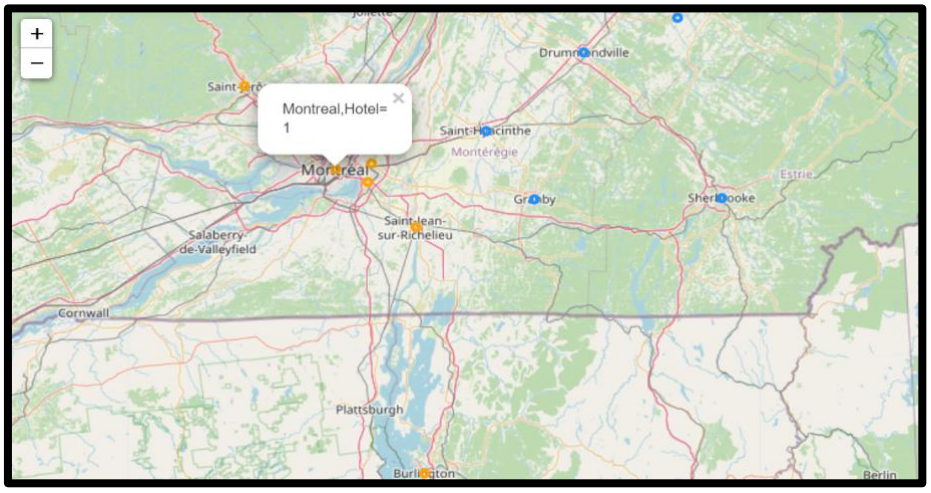
Figure 17.A – The features data of Montreal municipality

	Municipality	Bakery	Grocery Store	Hotel	total	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Province	Lar
48	Montreal	6	3	1	10	100	100	100	100	100	100	Quebec	

Figure 17.B – The corresponding maps of features data of Montreal municipality







3.2 Inferential statistical testing

No inferential statistical testing was needed since the required datasets were fully acquired through the Internet. The datasets were then cleansed, filtered and prepared to generate the full master dataset and the full features dataset required for building the clustering models.

3.3 Model selection

3.3.1 Selecting the Machine Learning Technique

Table 7 summarizes the most important predictive and descriptive machine learning techniques.
It also shows why I selected the clustering technique to solve this business problem.

Table 7 – Justification of selecting the clustering technique

No	ML Technique	Analytical Approach	Description	Selected?	Reason
1	Regression	Predictive	Supervised learning technique used for predicting a continuous value	No	• The business problem solution is descriptive in nature. • The data is unlabelled, and the process must be unsupervised
2	Classification	Predictive	Supervised learning technique used for predicting the class or category of a case	No	• The business problem solution is descriptive in nature. • The data is unlabelled, and the process must be unsupervised
3	Recommender systems	Predictive	Supervised or unsupervised learning technique used to offer relevant suggestions to users. Categorized as either a collaborative filtering or a content-based system	No	• The business problem solution is descriptive in nature. • The data is unlabelled, and the process must be unsupervised • No data for similar companies' performance is available
4	Clustering	Descriptive	Unsupervised learning technique used for finding groups of similar cases, for example, can be used for customer segmentation	**Yes**	• The business problem needs to find similar municipalities (i.e. market segments) in order to target them by specific operations teams in the company • The data is unlabelled, and the process must be unsupervised
5	Association	Descriptive	A learning technique (commonly unsupervised) used for finding items or events that often co-occur	No	• The business problem does not need finding items that often co-occur
6	Anomaly detection	Descriptive	Supervised, unverified or semi-supervised learning technique used for discovering abnormal and unusual cases	No	• The business problem does not need to detect anomalies
7	Sequence mining	Descriptive	A learning technique (commonly unsupervised) used for determining sequential patterns in data	No	• The business problem does not need to detect sequential patterns in data
8	Dimension reduction	Descriptive	Unsupervised learning technique used for reducing the size of data	No	• No need to reduce data size. Data size is not too large (100 municipalities and 761 venues).

3.3.2 Selecting the Machine Learning Model

Having selected the clustering technique, the second step in my model selection approach was to decide which clustering model is suitable for solving the business problem.

Table 8 summarizes the most important clustering models and shows why I selected the K-Means clustering model and the agglomerative clustering model to solve the business problem.

Table 8 – Justification of clustering models selection

No	Clustering Model	Description	Selected?	Reasons
1	K-Means clustering	It divides the data into K non-overlapping subsets or clusters without any cluster internal structure or labels (unsupervised algorithm)	Yes	• It is an easy and simple model, with fewer hyperparameters than the other clustering algorithms (we mainly need to specify the K value). • In our case, there is no difficulty in determining the value of K since a specific number of market segments (3) is already predefined by the management of company • We have a medium size dataset (100 municipalities and 761 venues), and this algorithm is optimal for dealing with medium or large size datasets • In our case, there is no large computation cost during runtime especially because the sample size is not large (100 municipalities and 761 venues). • In our case, we do not need to care about the complexity associated with providing a proper scaling (for fair treatment among features) since all the features are homogeneous and hence no scaling is needed (each feature represents the number of specific category of venues in the municipality). • In our case, we are not interested in the notion of outliers (since every single municipality must be classified in a distinct market segment). Therefore, using this algorithm which has no notion of outliers, does not represent a problem in our case. • The variance of the features dataset variables has almost the same value across the three features dataset values
2	Hierarchical clustering	It builds a hierarchy of clusters where each node is a cluster consisting of the clusters of its daughter nodes.	Yes	• Hierarchical clustering is normally used for small size datasets. For the size of our dataset (100 municipalities), it is not so difficult to use the dendrogram. • For the size of our dataset, K-means is more efficient. Hierarchical clustering takes longer computation times in comparison with K-means. However, in our case, the algorithm would not take long computation times because the size of the dataset is not too big.
3	Density-based spatial clustering of applications with noise (DBSCAN)	It groups together points that are closely packed together (points with many nearby neighbors)	No	• DBSCAN does not fit our business problem (in which every single municipality must be classified in a distinct market segment) since DNSCAN has the notion of noise (outliers) and it ignores less dense areas or noises based on the two parameters: radius and minimum points • DBSCAN needs a careful selection of its parameters. The radius and the minimum points parameters are indeterministic in our case, since no constraints are imposed by the company management on them. • DBSCAN is much slower than K-Means • DBSCAN doesn't work well over clusters with different densities

3.3.3 Applying the K-Means clustering model to the business problem

Based on the justification described in the previous section, I applied the K-Means clustering model to segment the company potential market into three distinct segments. The model was applied using the parameters shown in Table 9 and the datasets described in Table 10. The obtained results are described in section 4.

Table 9 – Parameters used in the K-Means Clustering model

No	Parameter	Value
1	Number of clusters (K)	K=3 i.e. we have 3 clusters (since three market segments are specified as business requirement in the problem definition section (Section 1)
2	Distance calculation method	Euclidean Distance (default)

Table 10 – Description of the used datasets

No	Dataset	Dataset Type	Description	Number of records
1	Municipalities (raw data)	Master dataset	Raw data of the municipalities (extracted from the Wikipedia website)	100
2	Municipalities (cleansed data)	Master dataset	Data of municipalities that have venue information in the Four-Square database.	100 (all municipalities already have data in the Four-Square database)
3	Municipalities venues	Features dataset	Data of the venues belonging to the 100 municipalities	761

3.3.4 Applying the hierarchical agglomerative clustering model to the business problem

Based on the justification described in the previous section, I have also applied the hierarchical agglomerative clustering model to segment the company potential market into three distinct segments. The model was applied using the parameters shown in Table 11 and the same datasets described before in Table 10. The obtained results are described in section 4.

Table 11 – Parameters used in the hierarchical agglomerative clustering model

No	Parameter	Value
1	Number of clusters (K)	K=3
2	Distance calculation method	• Single • Complete • Average

4. Results

4.1 Results of the K-Means clustering model

Table 12 shows the number of municipalities in each of the three clusters (i.e. in each market segment), and Table 13 shows the centroid value of each of them.

Table 12 – Number of municipalities in each market segment

Clus_km	Municipality	Bakery	Grocery Store	Hotel
0	43	43	43	43
1	31	31	31	31
2	21	21	21	21

Total = 100 municipalities

Table 13 – Centroid value of market segments

Clus_km	Bakery	Grocery Store	Hotel
0	0.171411	0.622044	0.206544
1	0.517191	0.236480	0.246328
2	0.131883	0.113590	0.754527

Figures 18,19, 20 and 21 show the distribution of municipalities based on the frequency of occurrence of Bakeries, Hotels and Grocery Stores venues in each municipality.

Figure 18 – Distribution of municipalities based on the frequency of occurrence of Bakery and Grocery Store venues

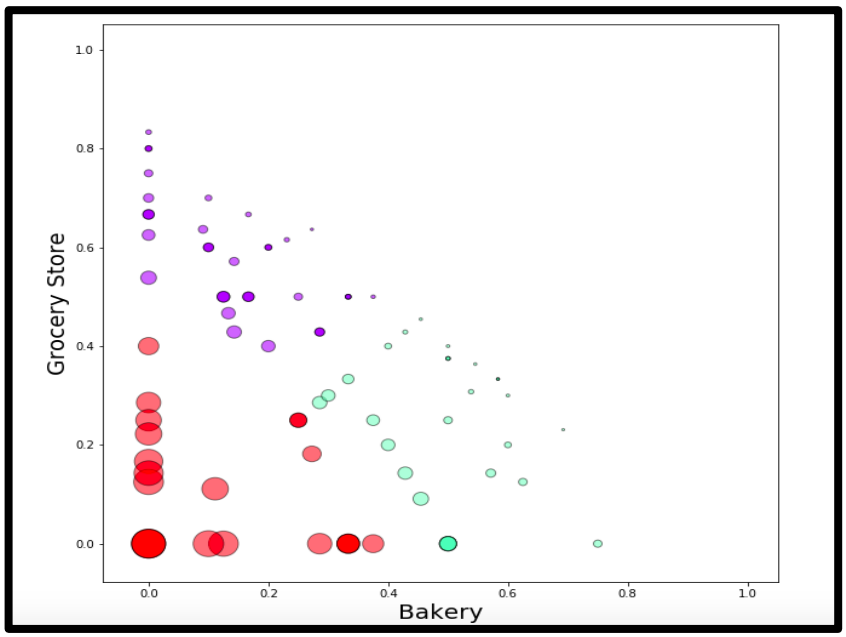


Figure 19 – Distribution of municipalities based on the frequency of occurrence of hotel and Grocery Store venues

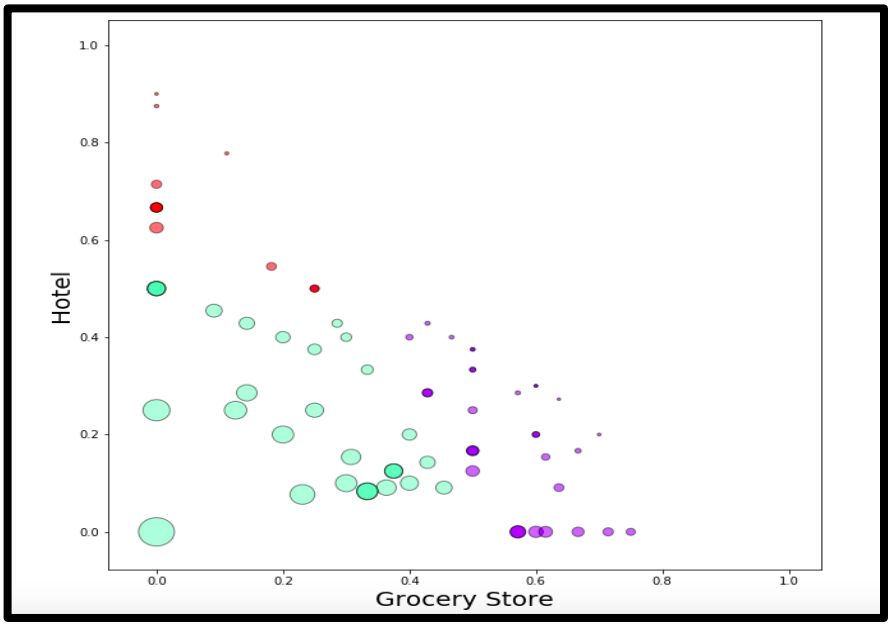


Figure 20 – Distribution of municipalities based on the frequency of occurrence of Bakery and Hotel venues

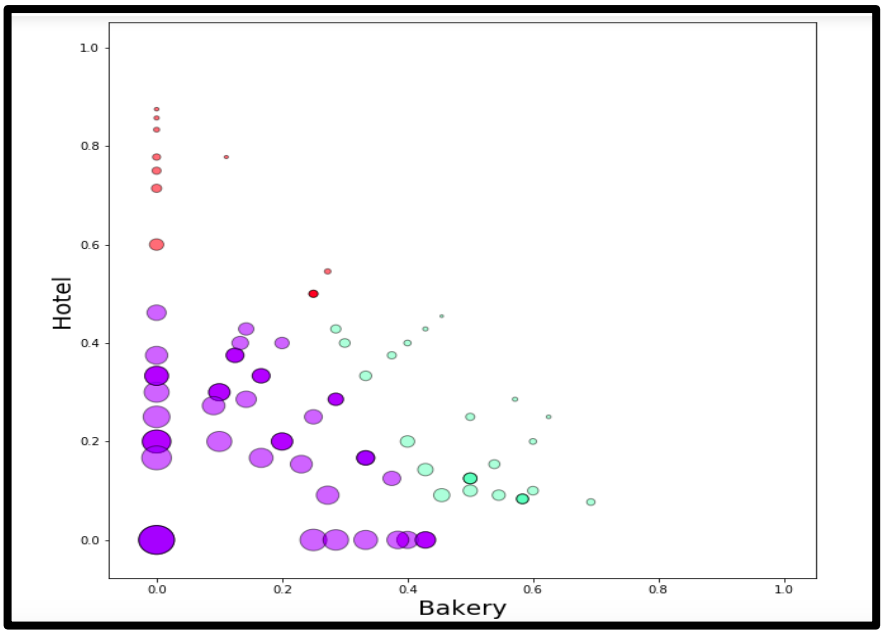


Figure 21 – Distribution of municipalities based on the frequency of occurrence of Bakery, hotel and Grocery Store venues

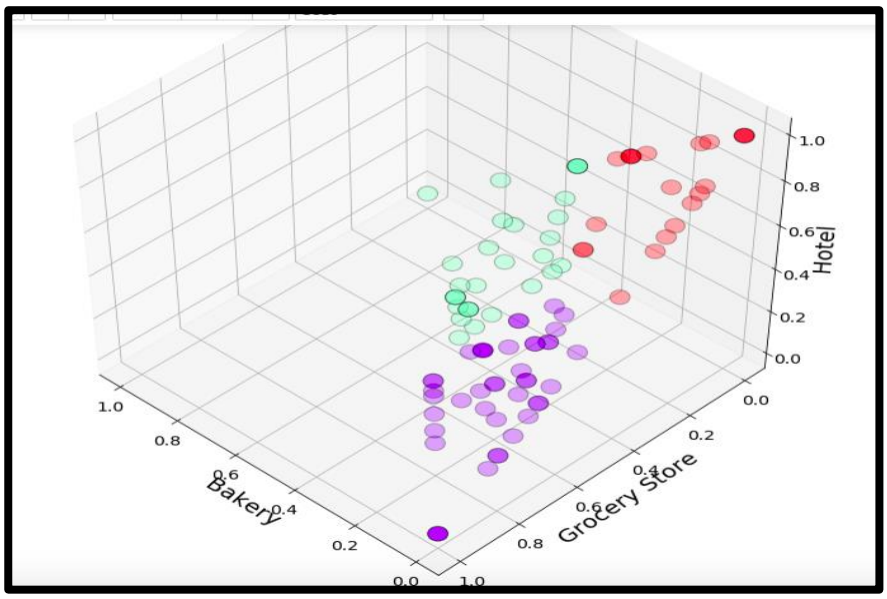


Table 14 shows the number of Grocery Stores, Hotels and Bakeries in each market segment. The same results are illustrated by the bar chart in Figure 22. Figure 23 depicts the density of Grocery Stores, Hotels and Bakeries in each market segment.

Table 14 – Number of Bakery, Grocery Stores and Hotels in each market segment

Cluster Labels	Bakery	Grocery Store	Hotel	total
0	58	205	76	339.0
1	139	65	66	270.0
2	18	18	99	135.0

Figure 22 – Number of Grocery Stores, hotels and Bakeries in each market segment

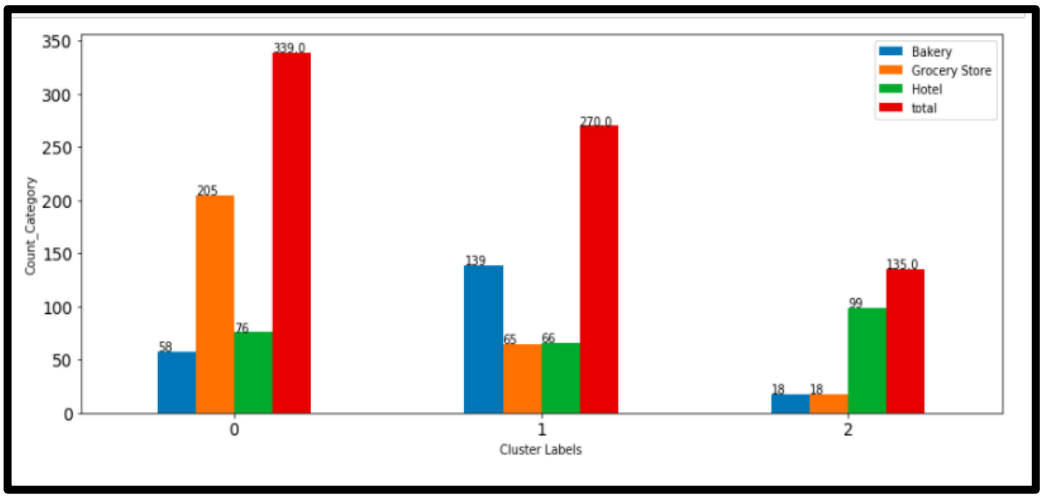


Figure 23 – Density of Bakery, Grocery Store and Hotel in each market segment

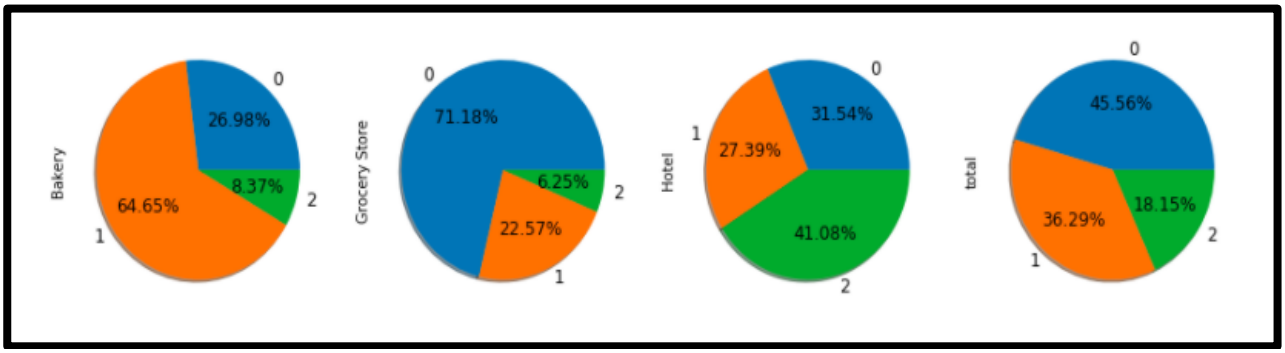
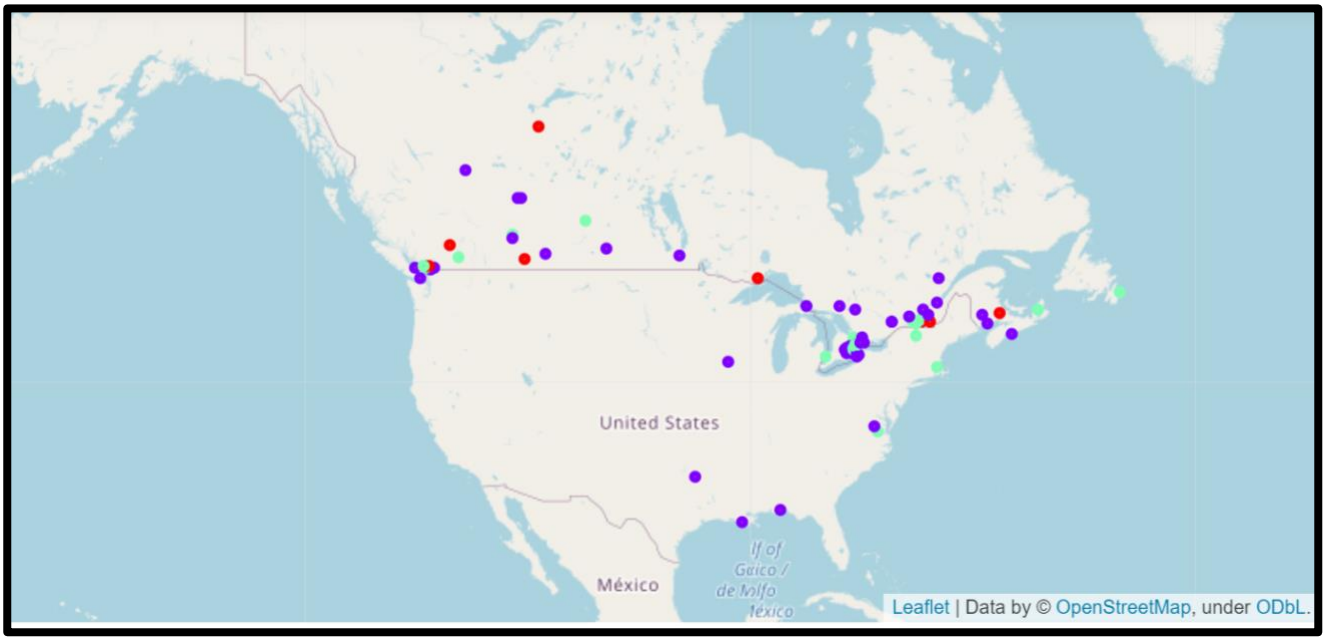


Figure 24 shows the distribution of Canadian municipalities on the three market segments, and Tables 15, 16 and 17 show sample reports of the municipalities of each market segment.

Figure 24 – Distribution of Canadian municipalities on the three market segments



Cluster 0 (Market Segment 1) Cluster 1 (Market Segment 2) Cluster 2 (Market Segment 3)

Table 15 – Report sample: Canadian municipalities of market segments 1

	Municipality	Province	Bakery	Grocery Store	Hotel	total	Venue	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	St. Albert	Alberta	0	10	5	15	90	0	Grocery Store	Hotel	Bakery
1	Trois-Rivières	Quebec	2	7	6	15	100	0	Grocery Store	Hotel	Bakery
2	Hamilton	Ontario	4	10	0	14	100	0	Grocery Store	Bakery	Hotel
3	Halton Hills	Ontario	5	8	0	13	100	0	Grocery Store	Bakery	Hotel
4	Drummondville	Quebec	0	7	6	13	99	0	Grocery Store	Hotel	Bakery
5	Saint-Jérôme	Quebec	3	8	2	13	100	0	Grocery Store	Bakery	Hotel
6	Saguenay	Quebec	0	8	4	12	72	0	Grocery Store	Hotel	Bakery
7	Terrebonne	Quebec	1	7	3	11	72	0	Grocery Store	Hotel	Bakery
8	Kitchener	Ontario	3	7	1	11	100	0	Grocery Store	Bakery	Hotel
9	Ottawa	Ontario	1	6	3	10	100	0	Grocery Store	Hotel	Bakery
10	Gatineau	Quebec	1	6	3	10	100	0	Grocery Store	Hotel	Bakery
11	Guelph	Ontario	4	6	0	10	100	0	Grocery Store	Bakery	Hotel
12	Milton	Ontario	1	7	2	10	100	0	Grocery Store	Hotel	Bakery
13	Red Deer	Alberta	0	7	3	10	85	0	Grocery Store	Hotel	Bakery
14	Quebec City	Quebec	2	6	2	10	100	0	Grocery Store	Hotel	Bakery
15	Greater Sudbury	Ontario	1	4	3	8	100	0	Grocery Store	Hotel	Bakery
16	Strathcona County	Alberta	3	4	1	8	100	0	Grocery Store	Bakery	Hotel

Table 16 – Report sample: Canadian municipalities of market segments 2

Province	Land area(km2, 2011)	Growth Rate 2011–2016	Population2016	lat	lng	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	marker_c
Ontario	266.30	13.31%	593638	43.6858	-79.7599	1	Bakery	Grocery Store	Hotel	#1E9
Ontario	1607.60	1.38%	64044	42.1539	-71.1828	1	Bakery	Grocery Store	Hotel	#FFA
Ontario	138.90	6.20%	193832	43.4474	-79.6667	1	Bakery	Grocery Store	Hotel	#1E9
Quebec	115.60	3.58%	239700	45.5172	-73.4467	1	Bakery	Grocery Store	Hotel	#1E9
Ontario	292.40	1.14%	721599	43.5903	-79.6457	1	Bakery	Grocery Store	Hotel	#1E9
Quebec	45.20	8.13%	85721	45.4555	-73.4679	1	Bakery	Grocery Store	Hotel	#1E9
British Columbia	115.00	4.64%	631486	49.2609	-123.114	1	Hotel	Bakery	Grocery Store	#FFA
Quebec	225.80	2.94%	95114	45.3057	-73.2533	1	Grocery Store	Bakery	Hotel	#1E9
Quebec	188.70	4.53%	55648	45.6275	-72.9406	1	Hotel	Grocery Store	Bakery	#FFA
Quebec	55.10	6.27%	56863	45.6793	-73.8762	1	Bakery	Grocery Store	Hotel	#1E9
British Columbia	90.60	4.27%	232755	49.2434	-122.973	1	Hotel	Bakery	Grocery Store	#FFA
British Columbia	11.80	9.76%	52898	49.3207	-123.074	1	Hotel	Bakery	Grocery Store	#FFA

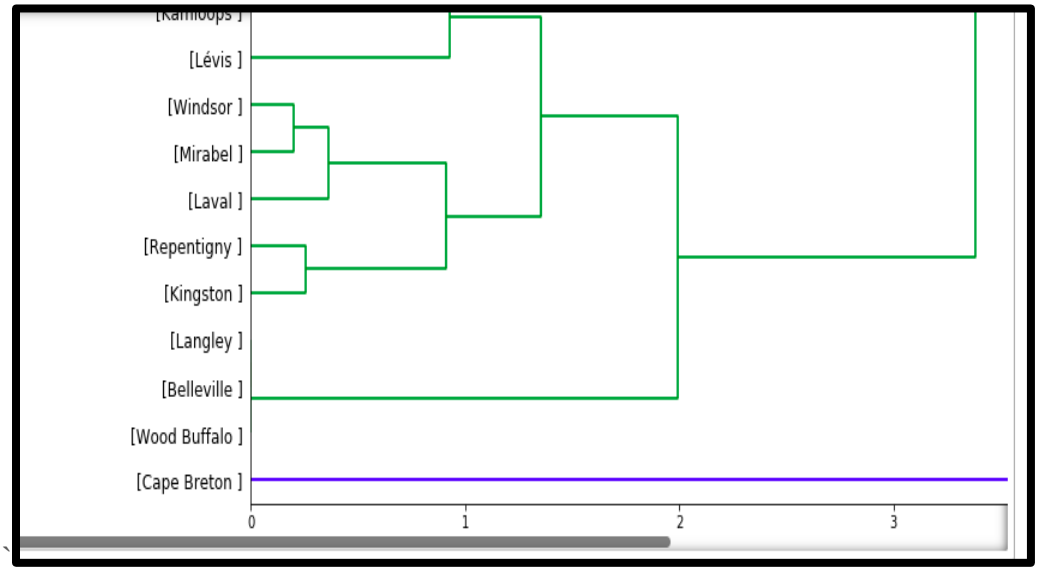
Table 17 – Report sample: Canadian municipalities of market segment 3

	Municipality	Province	Bakery	Grocery Store	Hotel	total	Venue	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Brampton	Ontario	7	4	2	13	100	1	Bakery	Grocery Store	Hotel
1	Norfolk County	Ontario	9	3	1	13	100	1	Bakery	Grocery Store	Hotel
2	Oakville	Ontario	7	4	1	12	100	1	Bakery	Grocery Store	Hotel
3	Longueuil	Quebec	7	4	1	12	100	1	Bakery	Grocery Store	Hotel
4	Mississauga	Ontario	7	4	1	12	100	1	Bakery	Grocery Store	Hotel
5	Brossard	Quebec	6	4	1	11	100	1	Bakery	Grocery Store	Hotel
6	Vancouver	British Columbia	5	1	5	11	100	1	Hotel	Bakery	Grocery Store
7	Saint-Jean-sur-Richelieu	Quebec	5	5	1	11	100	1	Grocery Store	Bakery	Hotel
8	Saint-Hyacinthe	Quebec	3	3	4	10	100	1	Hotel	Grocery Store	Bakery
9	Blainville	Quebec	5	4	1	10	100	1	Bakery	Grocery Store	Hotel
10	Burnaby	British Columbia	5	0	5	10	100	1	Hotel	Bakery	Grocery Store
11	North Vancouver	British Columbia	5	0	5	10	100	1	Hotel	Bakery	Grocery Store
12	North Vancouver	British Columbia	5	0	5	10	100	1	Hotel	Bakery	Grocery Store
13	Montreal	Quebec	6	3	1	10	100	1	Bakery	Grocery Store	Hotel
14	New Westminster	British Columbia	5	0	5	10	100	1	Hotel	Bakery	Grocery Store
15	Vaughan	Ontario	4	3	1	8	100	1	Bakery	Grocery Store	Hotel

4.2 Results of the hierarchical agglomerative clustering model

When I applied the hierarchical agglomerative clustering model on this business problem, I obtained the same results generated from the K – Means clustering model. Also, the results of the hierarchical agglomerative clustering model did not change when I used different options for the distance calculation method (i.e. the single, complete and average distance calculation methods). Figure 25 shows the results of applying the hierarchical agglomerative clustering model to the business problem (in the form of a dendrogram.)

Figure 25 – Result of the hierarchical agglomerative clustering model (Dendrogram)



5. Discussion

The results obtained from the K-Means clustering algorithm and the agglomerative clustering algorithm are identical, and both algorithms have provided a good solution to the business problem. As shown in Figures 22 and 23 we can notice that:

- Market segment 1 is a “Grocery Stores oriented” market segment: About 72% of the Canadian grocery stores’ target market belongs to the municipalities of this market segment. The remaining 28% of grocery stores market is shared by the bakery and hotel market segments with a percentage of 21% and 6% respectively.

- Market segment 2 is a “Bakery oriented” market segment: About 62% of the bakery target market belongs to the municipalities of this market segment. The remaining bakery target market is shared between the grocery store and hotel 3 with a share of 31% and 8% respectively.
- Market segment 3 is a “Hotel oriented” market segment About 42% of the hotels’ target market belongs to the municipality of this market segment. The remaining 58% of the hotels’ target market is shared between the grocery stores and bakery with a percentage of almost 31% and 27% respectively.

It is quite interesting to note that the total number of potential customers in each market segment is somewhat different (339, 270 and 135 for market segments 1,2, and 3 respectively). Based on this data, the management of the company is advised to consider adjusting its organization structure and uplifting its human capital capabilities in order to be able to successfully implement the new marketing strategy and cope with the requirements of the new Canadian market.

6. Conclusion

In this study, I used the K-Means and the Agglomerative clustering machine learning techniques to segment the new Canadian market of Senses Coffee company I identified the frequency of occurrence of Hotel, Grocery Store and Bakery as the most important features that affect the segmentation of this potential market. I built both K-Means clustering model and Agglomerative clustering model to build the market segments. These models are very useful in helping the company management in several ways. For example, it could help develop a new organization chart and plan the human capital and competencies necessary to implement the company’s new marketing strategy.

Appendix I –Example of the Canadian municipalities in the Wikipedia page

Rank (2016)	Municipality	Province	Municipal status	Land area (km², 2011)	Growth Rate 2011–2016	Population (2016)	Population (2011)	Population (2006)	Population (2001)	Population (1996)
1	***Toronto***	Ontario	City	630.2	4.46%	2,731,571	2,615,060	2,503,281	2,481,494	2,385,421
2	*Montreal*	Quebec	Ville	365.1	3.34%	1,704,694	1,649,519	1,620,693	1,583,590	1,547,030
3	*Calgary*	Alberta	City	825.3	12.99%	1,239,220	1,096,833	988,193	879,003	768,082
4	**Ottawa**	Ontario	City	2,790.2	5.76%	934,243	883,391	812,129	774,072	721,136
5	**Edmonton**	Alberta	City	684.4	14.82%	932,546	812,201	730,372	666,104	616,306
6	Mississauga	Ontario	City	292.4	1.14%	721,599	713,443	668,549	612,925	544,382
7	***Winnipeg***	Manitoba	City	464.1	6.27%	705,224	663,617	633,451	619,544	618,477
8	*Vancouver*	British Columbia	City	115.0	4.64%	631,486	603,502	578,041	545,671	514,008
9	Brampton	Ontario	City	266.3	13.31%	593,638	523,911	433,806	325,428	268,251
10	Hamilton	Ontario	City	1,117.2	3.26%	536,917	519,949	504,559	490,268	467,799
11	**Quebec City**	Quebec	Ville	454.1	2.96%	531,902	516,622	491,142	476,330	473,569
12	Surrey	British Columbia	City	316.4	10.60%	517,887	468,251	394,976	347,820	304,477
13	Laval	Quebec	Ville	247.1	5.34%	422,993	401,553	368,709	343,005	330,393
14	***Halifax***	Nova Scotia	Regional municipality	5,490.3	3.34%	403,131	390,096	372,679	359,111	342,851
15	London	Ontario	City	420.6	4.83%	383,822	366,151	352,395	336,539	325,669
16	Markham	Ontario	City	212.6	9.03%	328,966	301,709	261,573	208,615	173,383
17	Vaughan	Ontario	City	273.5	6.22%	306,233	288,301	238,866	182,022	132,549
18	Gatineau	Quebec	Ville	343.0	4.11%	276,245	265,349	242,124	226,696	217,591
19	*Saskatoon*	Saskatchewan	City	209.6	10.89%	246,376	222,189	202,340	196,861	193,653
20	Longueuil	Quebec	Ville	115.6	3.58%	239,700	231,409	229,330	225,761	227,408
21	Kitchener	Ontario	City	136.8	6.42%	233,222	219,153	204,668	190,399	178,420
22	Burnaby	British Columbia	City	90.6	4.27%	232,755	223,218	202,799	193,954	179,209

Appendix II – Example of Canadian municipalities' Geo-Location data

	Municipality	Province	Land area(km2, 2011)	Growth Rate 2011–2016	Population2016	lat	lng
0	Toronto	Ontario	630.20	4.46%	2731571	43.6535	-79.3839
1	Montreal	Quebec	365.10	3.34%	1704694	45.4972	-73.6104
2	Calgary	Alberta	825.30	12.99%	1239220	51.0534	-114.063
3	Ottawa	Ontario	2790.20	5.76%	934243	45.4211	-75.6903
4	Edmonton	Alberta	684.40	14.82%	932546	53.5354	-113.508
5	Mississauga	Ontario	292.40	1.14%	721599	43.5903	-79.6457
6	Winnipeg	Manitoba	464.10	6.27%	705224	49.8955	-97.1385
7	Vancouver	British Columbia	115.00	4.64%	631486	49.2609	-123.114
8	Brampton	Ontario	266.30	13.31%	593638	43.6858	-79.7599
9	Hamilton	Ontario	1117.20	3.26%	536917	43.2561	-79.8729
10	Quebec City	Quebec	454.10	2.96%	531902	46.826	-71.2352
11	Surrey	British Columbia	316.40	10.60%	517887	51.2715	-0.341452
12	Laval	Quebec	247.10	5.34%	422993	48.071	-0.77235
13	Halifax	Nova Scotia	5490.30	3.34%	403131	44.6486	-63.5859
14	London	Ontario	420.60	4.83%	383822	51.5073	-0.127647
15	Markham	Ontario	212.60	9.03%	328966	43.8543	-79.3268
16	Vaughan	Ontario	273.50	6.22%	306233	43.7942	-79.5268
17	Gatineau	Quebec	343.00	4.11%	276245	45.4284	-75.7106
18	Saskatoon	Saskatchewan	209.60	10.89%	246376	52.1318	-106.661
19	Longueuil	Quebec	115.60	3.58%	239700	45.5172	-73.4467

Appendix III – Example of potential customers data

	Municipality	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Toronto	43.653482	-79.383935	Downtown Toronto	43.653232	-79.385296	Neighborhood
1	Toronto	43.653482	-79.383935	Byblos Toronto	43.647615	-79.388381	Mediterranean Restaurant
2	Toronto	43.653482	-79.383935	Elgin And Winter Garden Theatres	43.653394	-79.378507	Theater
3	Toronto	43.653482	-79.383935	Art Gallery of Ontario	43.654003	-79.392922	Art Gallery
4	Toronto	43.653482	-79.383935	St. Lawrence Market (South Building)	43.648743	-79.371597	Farmers Market
5	Toronto	43.653482	-79.383935	Hailed Coffee	43.658833	-79.383684	Coffee Shop
6	Toronto	43.653482	-79.383935	Alo	43.648574	-79.396243	French Restaurant
7	Toronto	43.653482	-79.383935	Delta Hotels by Marriott Toronto	43.642882	-79.383949	Hotel
8	Toronto	43.653482	-79.383935	Yeti Nails & Spa	43.647938	-79.396330	Cosmetics Shop
9	Toronto	43.653482	-79.383935	Pai	43.647923	-79.388579	Thai Restaurant

Appendix IV – Example of Canadian GIS data