

Location Analysis for Launch of "Winter-Coma" in Canada

Introduction

Background

Over the past few years, "*Brick and Mortar Retail*" has been on the brink of a major apocalypse. This is one space where demand for space is far greater than the supply yet the industry is struggling to survive. As per the CNN Business data, close to 51,000 retail stores opened in U.S in last 3 years and at the same time almost 42,000 stores closed their doors. Out of these 9,300 closures were in city of New York which is touted as the world's leading financial centre and the most financially powerful city in the world.

While most believe that online retail industry has a major part to play in this, truth is far from this misconception. Analysing the e-commerce sales, e-commerce growth is at 16% while in-store is at 4% and at the same time revenue for in-store is 3 times that of online stores. Not to mention the fact that brick and mortar stores offer a lot more to the customers rather than just the product. Hence the need to invest and be successful in the retail development industry is imminent.

The most important decision for a new retail business is "*where to open the store?*". Location is one of the most strategic decision a business takes considering its long-term impact on revenue and competitive advantage.

"Maximizing sales is a primary objective for retailers, hence, finding the perfect site location that will facilitate both footfall and growth, is of key importance (O'Malley, et al, 1995)"

Problem

"Winter-Coma" is a winter wear clothing brand looking to enter Canada. However, owing to the sensitive brick and mortar retail industry status, it wants to perform a thorough location data analysis before it fixates upon a site to open a store.

Data Description

Varied datasets have been utilised for this location analysis, weather station data across Canada, province level demographic data, geospatial data and foursquare API location data.

The weather station data is analysed to identify the colder regions. This data provides details such as max. and min. temp for each station, the snowfall, precipitation etc. This data is used to do location and temperature-based density clustering. This has been sourced from "*Environment Canada Monthly Values for July – 2015*"

The demographic and census data for 2016, is utilised to select the region with better business prospects. Features like population density, avg. household income are analysed. This has been retrieved from <https://www150.statcan.gc.ca/n1/en/type/data?MM=1>, a publicly available data library by the Canadian government.

Once the region is selected, postal code data extracted from Wikipedia is utilised to capture the Borough level data. Within the Borough, the local neighbourhood data extracted from the Foursquare API, is utilised to analyse the neighbourhoods. Features like no. of eateries, no. of recreational centres, connectivity of the area are analysed. Basis this final analysis the districts to be considered for opening the stores are shortlisted.

For opening a retail store, various factors like macro location evaluation pertaining to political, socio, economic situations and micro evaluation pertaining to the local district level psychographic analysis, financial matters, site location availability etc.

In this report due to limited availability of quality data, a controlled number of factors have been analysed.

Methodology

The first step in the analysis was to find the colder regions in Canada. This was achieved by analysing the weather station data from across Canada. The *Environment Monthly for July '15* was utilised for the same. Given below is a snapshot of the features we could extract from the data:

	Stn_Name	Lat	Long	Prov	Tm	DwTm	D	Tx	DwTx	Tn	...	DwP	P%N	S_G	Pd	BS	DwBS	BS%	HDD	CDD	Stn_No
0	CHEMAINUS	48.935	-123.742	BC	8.2	0.0	NaN	13.5	0.0	1.0	...	0.0	NaN	0.0	12.0	NaN	NaN	NaN	273.3	0.0	1011500
1	COWICHAN LAKE FORESTRY	48.824	-124.133	BC	7.0	0.0	3.0	15.0	0.0	-3.0	...	0.0	104.0	0.0	12.0	NaN	NaN	NaN	307.0	0.0	1012040
2	LAKE COWICHAN	48.829	-124.052	BC	6.8	13.0	2.8	16.0	9.0	-2.5	...	9.0	NaN	NaN	11.0	NaN	NaN	NaN	168.1	0.0	1012055
3	DISCOVERY ISLAND	48.425	-123.226	BC	NaN	NaN	NaN	12.5	0.0	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1012475
4	DUNCAN KELVIN CREEK	48.735	-123.728	BC	7.7	2.0	3.4	14.5	2.0	-1.0	...	2.0	NaN	NaN	11.0	NaN	NaN	NaN	267.7	0.0	1012573

Fig 1: Weather Station Dataset

The dataset has 25 data points for each of the 1341 weather stations spread across Canada. Among the basic things like station name and id, the data points include location coordinates, min. and max. temperature readings, precipitation, snowfall, sunshine etc. data. For our analysis we will use the location coordinates along with the temperature data.

Post the data cleaning (removing the null entries), the weather stations can be visualised as follows:

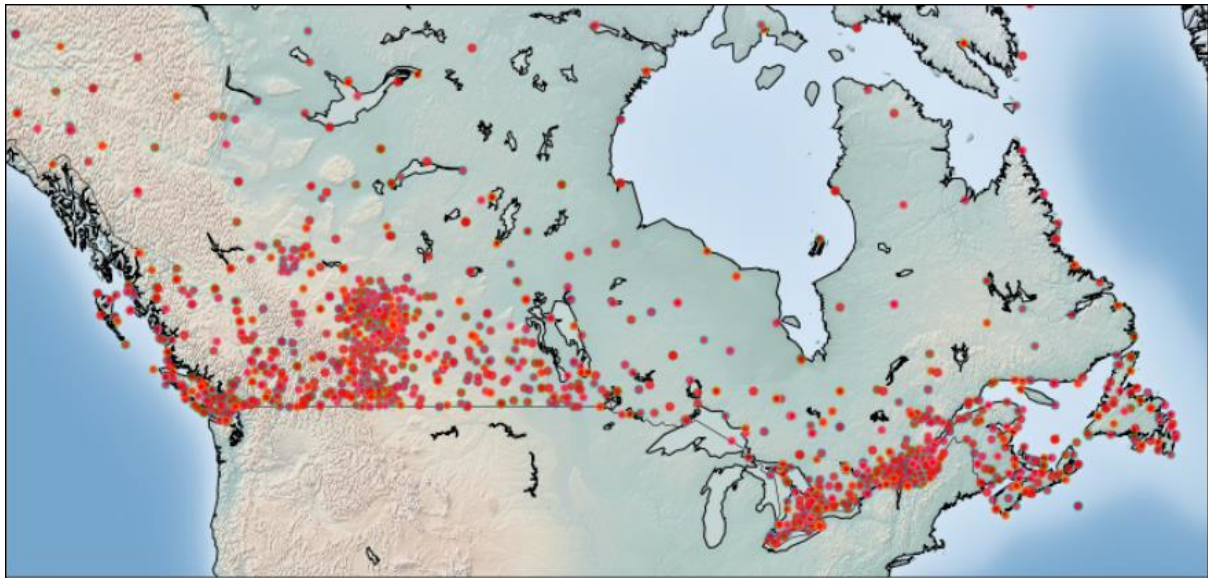


Fig 2: Weather Station Visualisation on map of Canada

This dataset is further divided into homogeneous clusters for in-depth analysis. Since this is a location-based data, the best clustering algorithm would be *DBSCAN (Density Based Spatial Clustering of Applications with Noise)*. This algorithm works best with spatial data where we are looking at arbitrary shaped clusters and noise. This algorithm locates regions of high density that are separated from one another by regions of low density.

For this analysis we will consider the location data points along with the min. and max. temperatures. Considering the wide data, we have taken a radius of 300 mtrs and min. sample size of 10. The analysis revealed a total of 9 cluster and 428 outliers. The following are the clusters as visualised on the map of Canada:

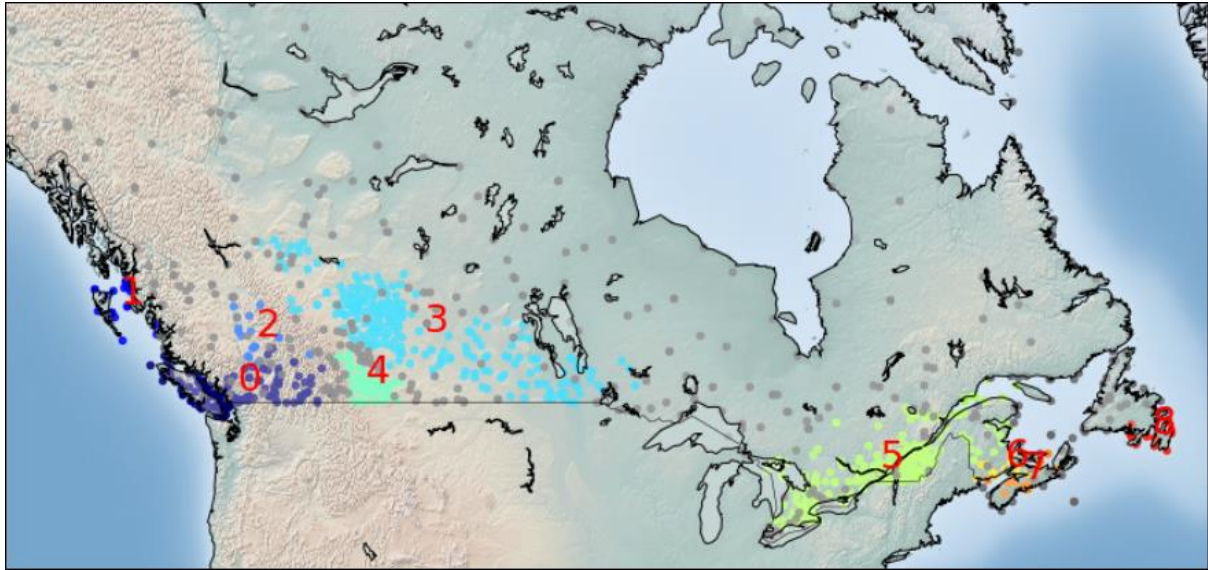


Fig 3: DBSCAN Clusters of Weather Stations

The clusters 5, 3 and 6 are the ones with lowest mean min. temperatures of around -15 °C. Other clusters in the western part are pretty warm compared to these with an avg. min. temperature above 0°C, while the remaining clusters in eastern part i.e. 7 & 8 have a min. temperature of -6°C.

With further analysis of data, eastern part seems more suitable owing to following factors:

1. More no. of weather stations, 266 compared to 215 in western part; hence the assumption of bigger area to cover
2. Lower mean of minimum temperature, -16.3°C compared to -13.8°C
3. Proximity of clusters 6, 7 and 8 which too have low temperatures of -13°C, -10°C and -4°C resp.; Unlike in western part surrounded by clusters with a mean temp of -2°C.
4. High number of clusters implies more footfalls hence more demand in the vicinity of our product.

This eastern part consists majorly of two regions QC (Quebec) and ON (Ontario). The demographic and census data is then further utilised to select the region with better business prospects. Features like population density, avg. household income are analysed. The dataset has varied features:

CENSUS_YEAR	GEO_CODE (POR)	GEO_NAME	GNR	GNR_LF	DATA_QUALITY_FLAG	ALT_GEO_CODE	DIM: Profile of Canada, Provinces and Territories (2247)	Member ID: Profile of Canada, Provinces and Territories (2247)	Notes: Profile of Canada, Provinces and Territories (2247)	Dim: Sex (3): Member ID: [1]: Total - Sex	Dim: Sex (3): Member ID: [2]: Male	Dim: Sex (3): Member ID: [3]: Female
0	2016	1	Canada	4.0	5.1	20000	1	Population, 2016	1	1.0	35151728	...
1	2016	1	Canada	4.0	5.1	20000	1	Population, 2011	2	2.0	33476688	...
2	2016	1	Canada	4.0	5.1	20000	1	Population percentage change, 2011 to 2016	3	NaN	5	...
3	2016	1	Canada	4.0	5.1	20000	1	Total private dwellings	4	3.0	15412443	...
4	2016	1	Canada	4.0	5.1	20000	1	Private dwellings occupied by usual residents	5	4.0	14072079	...

Fig 4: Demographic, Census Dataset for Canada

Post data cleaning and filtering out the required features, we analyse the population and income trends for the two regions.

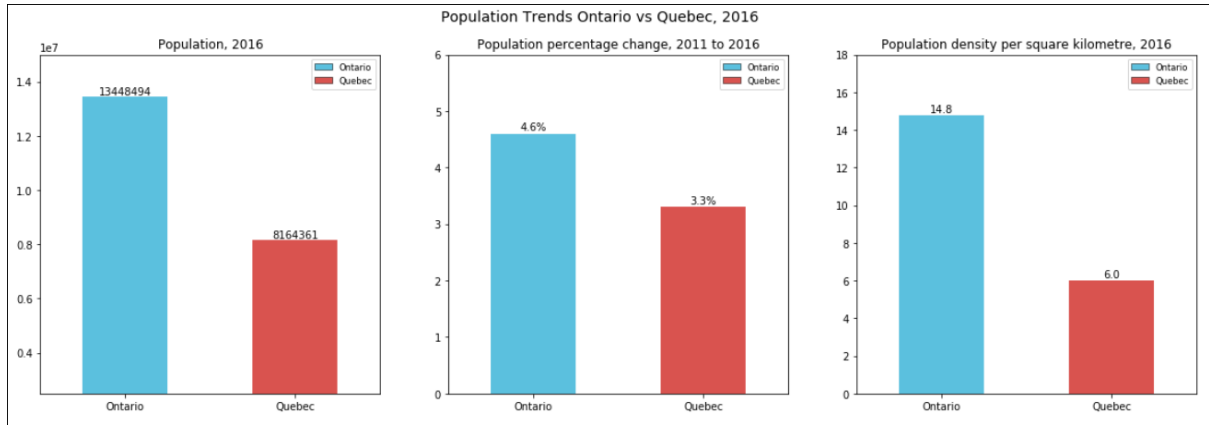


Fig 5: Population Trends - Ontario vs Quebec, 2016

The Population trends show that Ontario has a higher population than Quebec. High percentage increase in population indicates high economic as well as social activities in Ontario. Population density of Ontario is more than double of that of Quebec, indicating high demand generation possibilities in that region.

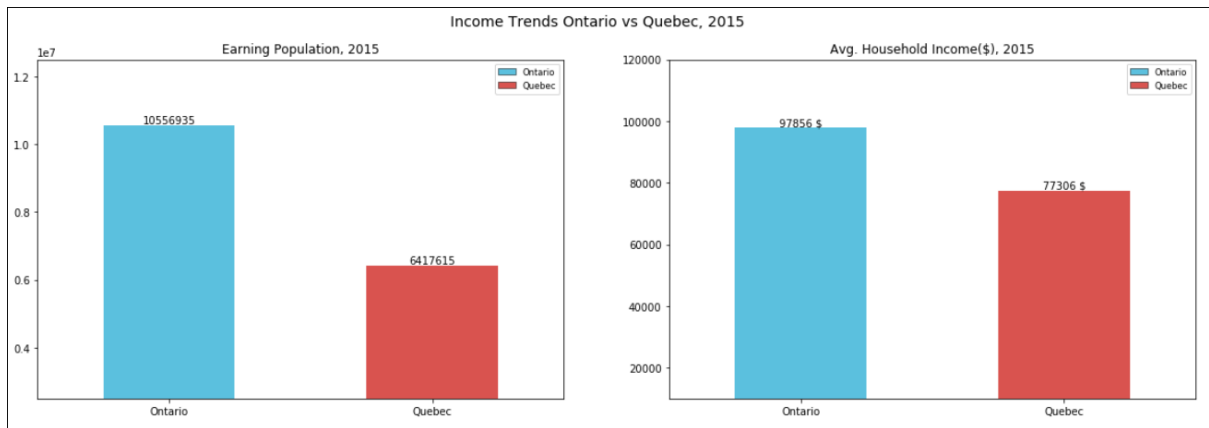


Fig 6: Income Trends - Ontario vs Quebec, 2016

Income trends show high earning potential of Ontario with high number of earning individuals along with high average household income, indicating high consumer spending.

Following inferences were drawn from the data analysis:

1. Ontario has higher population, higher population density hence more coverage for our brand
2. Ontario has had a higher percentage increase in population over the years indicating a stable and rising economy
3. Ontario has higher number of earning individuals, indicating more consumers
4. Ontario has higher average household income than Quebec, indicating better spending potential
5. Both the regions have 66% of population in the age group of 15-64 years and average age of the population is 41 years

Hence Ontario comes out as the region with far better business prospects than Quebec. Basis this analysis, we can safely assume that Ontario will give the brand better demand generation opportunities along with high consumer spending.

Within Ontario, we will analyse Toronto city. This is due to paucity of good quality district level demographic and economic dataset.

The postal code data from Wikipedia is scraped to extract the Borough level data for Toronto City. The data has information regarding the Borough of each of the 103 postal codes in 10 Boroughs around Toronto city along with the neighbourhood places. Further the Geospatial dataset is utilised to access the location coordinates of these postal codes.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

Fig 7: Postal Code Data with Location Coordinates

Since these are spread all over Toronto with a very wide coverage, we will use *DBSCAN clustering* with a radius of 400 mtrs and minimum sample size of 5 for identifying the outliers and shortening our area of focus.

Below are the shortlisted 56 postal codes/areas in 8 Boroughs, which have been chosen for further analysis.

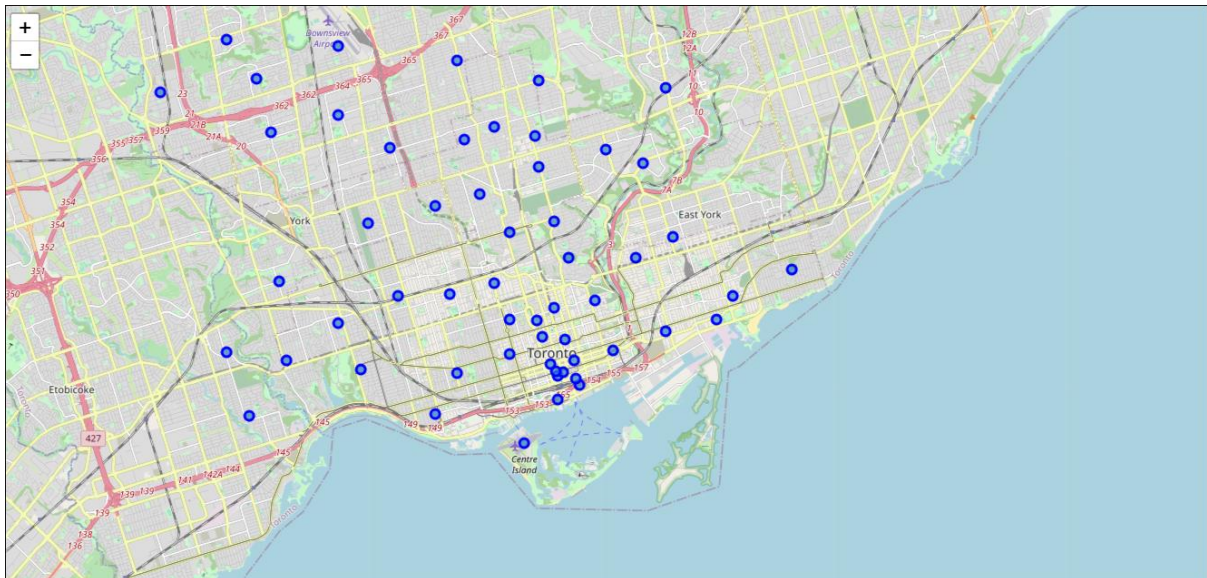


Fig 8: Shortlisted postal codes via DBSCAN

Foursquare API data is used on these shortlisted areas to explore their neighbourhoods. For each area, we can explore the number of venues, their categories among other things. With the help of such data we can further understanding the area, whether it's a residential area or an open field or a market place.

Snapshot of the venue category data for these neighbourhoods:

- There are total of 1025 venues in these neighbourhoods in 213 unique categories

- Highest number of venues belong to eateries like Coffee shop, café, restaurants, pubs and parks
- Boroughs with highest no. of venues belong to mostly M4 and M5 postal codes which belong to Downtown and Central Toronto boroughs

Coffee Shop	75
Café	64
Park	38
Restaurant	35
Italian Restaurant	24
Sandwich Place	22
Bakery	22
Pizza Place	19
Pub	16
Bar	16

Fig 9: Most Common Venue Categories

Using One hot coding methodology, we will analyse the venue categories in each neighbourhood. Number of venues along with the most common venue categories across the areas are some of the important factors analysed with this data.

Postal Code	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0 M3C	Restaurant	Coffee Shop	Asian Restaurant	Beer Store	Gym	Art Gallery	Dim Sum Restaurant	Bubble Tea Shop	Sporting Goods Shop	Italian Restaurant
1 M3K	Park	Snack Place	Business Service	Airport	College Gym	College Rec Center	Distribution Center	Discount Store	Diner	Dim Sum Restaurant
2 M3L	Grocery Store	Park	Shopping Mall	Bank	Hotel	Cupcake Shop	Distribution Center	Discount Store	Diner	Dim Sum Restaurant
3 M3M	Business Service	Baseball Field	Food Truck	Cuban Restaurant	Dog Run	Distribution Center	Discount Store	Diner	Dim Sum Restaurant	Dessert Shop
4 M4E	Neighborhood	Pub	Trail	Health Food Store	Yoga Studio	Creperie	Discount Store	Diner	Dim Sum Restaurant	Dessert Shop

Fig 10: Common Venue Categories for each Postal Code

K-Means cluster analysis is used at this point to cluster together similar areas basis the venues in the neighbourhood. *Elbow method* of reducing the within cluster inertia is used to find the best value of K. The most optimum value of “K” for cluster formation comes out to be “4”.

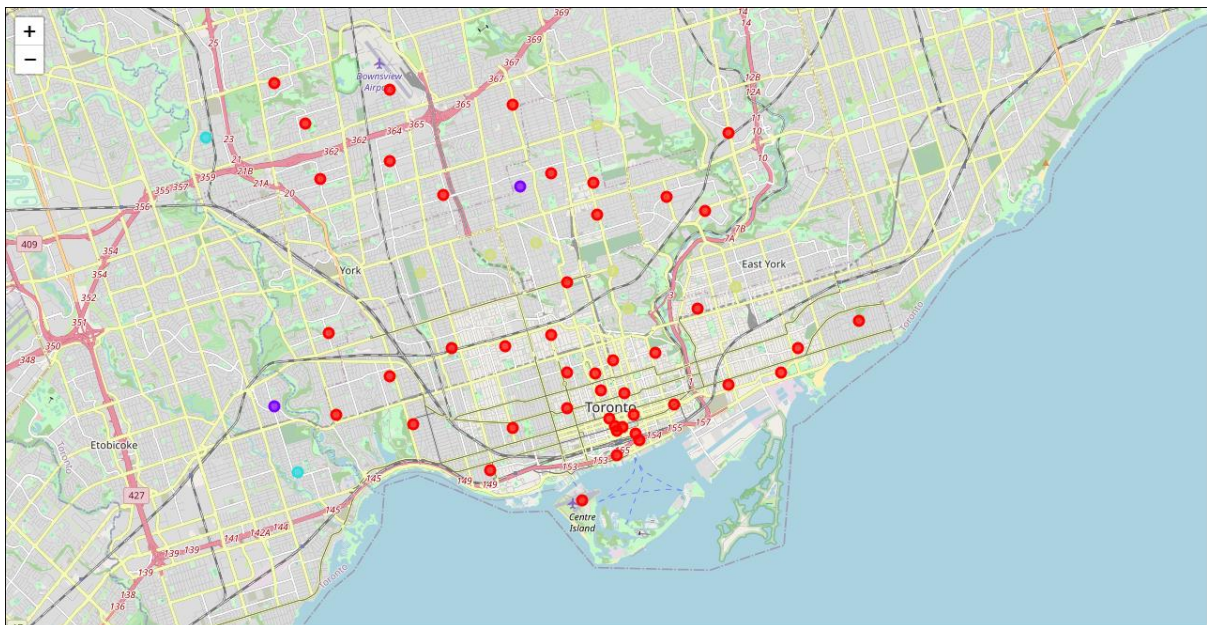


Fig 11: Clusters via K-Means

In the figure above we can see the 4 clusters formed post the k-means clustering. The red ones are the 44 areas belonging to one cluster while the remaining 12 belong to the other three clusters.

Further analysis of the said three clusters reveals that all the 12 areas in those clusters have parks, fields, yoga studios, playgrounds along with one off stores.

Toronto_merged[(Toronto_merged['Cluster Labels'] == 3) (Toronto_merged['Cluster Labels'] == 1) (Toronto_merged['Cluster Labels'] == 2)]																
	Postal Code	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
7	M6C	York	Humewood-Cedarvale	43.693781	-79.428191	3	Park	Trail	Field	Hockey Arena	Yoga Studio	Cuban Restaurant	Distribution Center	Discount Store	Diner	Dim Sum Restaurant
10	M6E	York	Caledonia-Fairbanks	43.689026	-79.453512	3	Park	Women's Store	Cuban Restaurant	Dog Run	Distribution Center	Discount Store	Diner	Dim Sum Restaurant	Dessert Shop	Department Store
17	M4J	East York	East Toronto, Broadview North (Old East York)	43.685347	-79.338106	3	Park	Pizza Place	Convenience Store	Coffee Shop	Cuban Restaurant	Distribution Center	Discount Store	Diner	Dim Sum Restaurant	Dessert Shop
31	M9M	North York	Humberlea, Emery	43.724766	-79.532242	2	Food Service	Baseball Field	Cuban Restaurant	Dog Run	Distribution Center	Discount Store	Diner	Dim Sum Restaurant	Dessert Shop	Department Store
32	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	3	Bus Line	Park	Swim School	Creperie	Distribution Center	Discount Store	Diner	Dim Sum Restaurant	Dessert Shop	Department Store
33	M5N	Central Toronto	Roselawn	43.711695	-79.416936	1	Garden	Pool	Yoga Studio	Creperie	Distribution Center	Discount Store	Diner	Dim Sum Restaurant	Dessert Shop	Department Store
36	M5P	Central Toronto	Forest Hill North & West, Forest Hill Road Park	43.696948	-79.411307	3	Park	Trail	Sushi Restaurant	Jewelry Store	College Rec Center	Colombian Restaurant	Dog Run	Distribution Center	Discount Store	Diner
44	M4T	Central Toronto	Moore Park, Summerhill East	43.689574	-79.383160	3	Park	Summer Camp	Restaurant	Creperie	Distribution Center	Discount Store	Diner	Dim Sum Restaurant	Dessert Shop	Department Store
48	M4W	Downtown Toronto	Rosedale	43.679563	-79.377529	3	Park	Playground	Trail	Creperie	Distribution Center	Discount Store	Diner	Dim Sum Restaurant	Dessert Shop	Department Store
52	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944	1	Pool	River	Yoga Studio	Coworking Space	Discount Store	Diner	Dim Sum Restaurant	Dessert Shop	Department Store	Deli / Bodega
55	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...	43.636258	-79.498509	2	Baseball Field	Yoga Studio	Eastern European Restaurant	Dog Run	Distribution Center	Discount Store	Diner	Dim Sum Restaurant	Dessert Shop	Department Store

Fig 12: Clusters with parks, playgrounds and one-off stores

We can assume that these neighbourhoods are mostly open areas which are of no significance to us in our analysis hence we will filter out the remaining 45 areas. Since most of the Boroughs belong to Toronto, hence filtering out Central and Downtown Toronto.

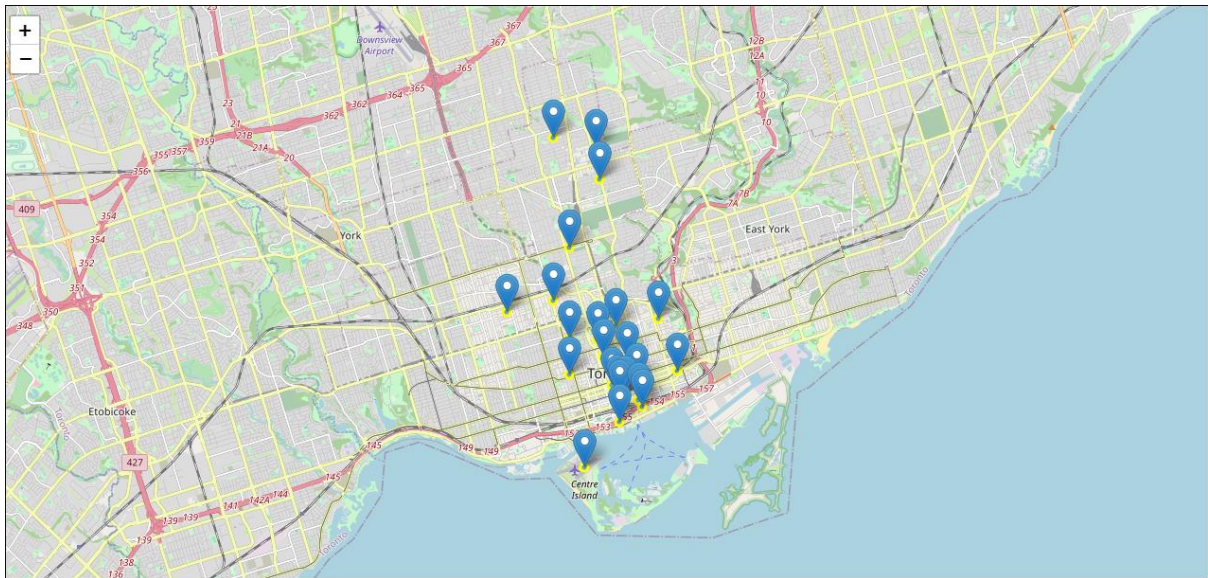


Fig 13: Areas with relevant venue categories in Central and Downtown Toronto

These areas are spread over a radius of 5 km and some seem to be one off locations. Since we are more interested in high density areas (basic assumption that high density areas will have higher footfalls), we will further reduce our area and use *DBSCAN* with a radius of 700 mtrs and minimum sample size of 5.

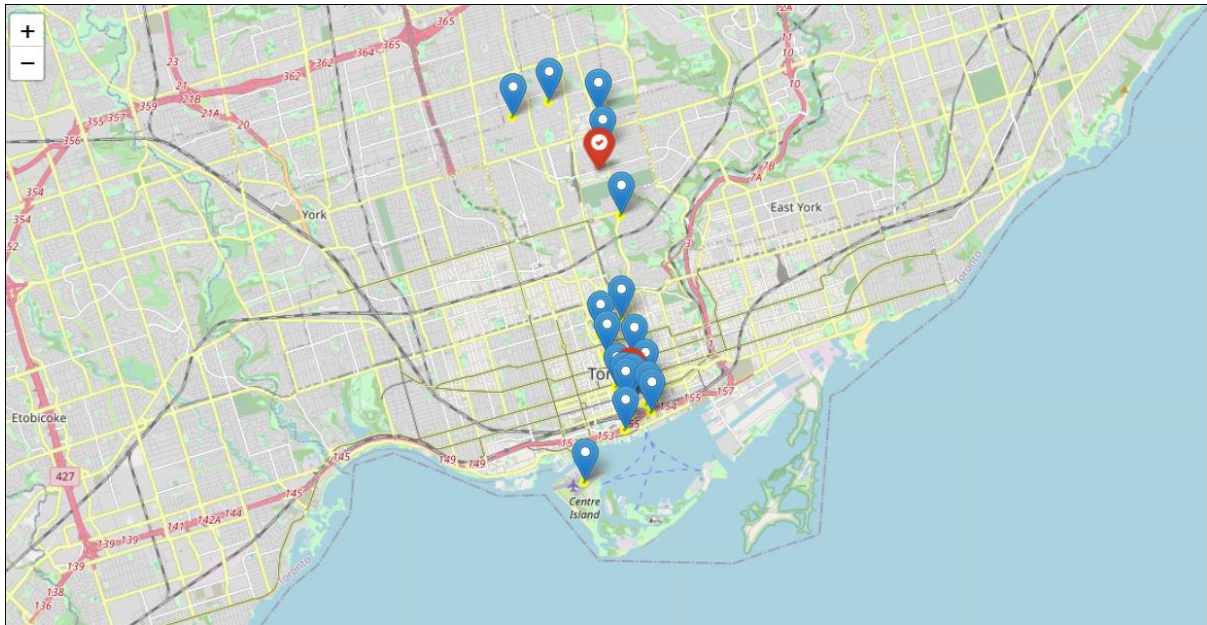


Fig 14: High Density areas mapped via DBSCAN

The analysis reveals 5 areas in Central Toronto and 13 areas in the high-density cluster of “*Downtown Toronto*”.

Central Toronto has 5 postal codes/areas within a radius of 2 km. There are 2 clothing stores and 1 shopping mall in these areas as per the data. Number of eateries, recreational centres are far lesser than Downtown Toronto.

The highly dense cluster of Downtown has 13 postal codes/areas within a radius of 1.5 km. There are 2 clothing stores and 1 shopping mall in this area as per the given dataset. Overall no. of eateries, departmental stores, recreational centres are one of the highest across Toronto. The area has Union train station which is the central station and one of the busiest train stations in Ontario.

Results

While deciding upon the location of a retail store various factors come into play. Competitive advantage, customer proximity, geographically strategic with easy accessibility are some of the most important ones.

Through the location analysis done for “*Winter-Coma*”, we have touched upon a few of these as per the availability of data. And basis that, “*Downtown Toronto*” has come out as the best choice.

Downtown is the centre of economic activity with umpteen number of company offices in the vicinity. This indicates high paying capacity individuals, hence high market potential. The presence of numerous eating outlets, recreational centres, Downtown becomes the centre of all activities. This indicates high visibility of the brand along with high pedestrian traffic which ensures a constant customer pull. Downtown has the Union railway station, which will have regular trains in and out hence this becomes a central location easily accessible by the nearby areas.

Discussion

As derived upon by the analysis, “*Downtown Toronto*” seems to be the best area for opening a retail store. This can be confirmed from the information by the economic survey which states,

“*Downtown is the main business district of Toronto and it has the retail core presence with over 600 retail stores within 1.5 km radius area.*”

Our analysis here focussed on the areas of Customer proximity, easy accessibility and high market potential. This can be further detailed by analysing the market in terms of the nature and presence of competition in the area and availability of a good location with good return on investment.

With availability of good quality demographic data, we can analyse the age group, profession, lifestyle and income of the individuals residing in these localities. This will give us a better idea on the profiling of our target audience hence our product portfolios.

Conclusion

To find the best location to open a retail store of “*Winter-Coma*” in Canada, a thorough location data analysis was done. From analysing the weather patterns across Canada, retrieving the high business prospects region, determining the high-density areas, classifying the neighbourhood’s basis the venues present within a radius of 500 mtrs, performing postal area code level neighbourhood analysis, we have been able to shortlist the best location, “*Downtown Toronto*”.

**Various data analysis and machine learning techniques have been used in this process. DBSCAN and K-Means clustering have been used for the spatial data analysis at different points. Exploratory data analysis has been used to study the demographic and census data. Folium library has been used for visualisation of the areas.*