Location Analysis For Launch Of *"Winter-Coma"* In Canada

Applied Datascience Capstone Project

-- Manika Babbar

Introduction

Background

- *"Brick and Mortar Retail"* has been on the brink of a major apocalypse.
  - 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.
- Ecommerce vs Brick & Mortar
  - E-commerce is at 16% while in-store is at 4%
  - Revenue for in-store is 3 times that of online stores
- Imminent need to invest and be successful in the retail development industry
- Most important decision for a new retail business is *"where to open the store?"*.

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 through location data analysis before it fixates upon a site to open a store.

"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)"

Data Description

Four different datasets have been used in the analysis.

Weather Station Data	Demographic, Census Data	Postal Code Data	Foursquare API Data

- To identify colder region
- Features like Temperature, precipitation, location coordinates etc

- To select region with better business prospects
- Features like Population and income trends

- To get the Borough level data
- Features like Borough, neighbourhoods

- To analyse the Boroughs basis the neighbouring venues
- Features like Venues, Venue Categories

Methodology

DBSCAN clustering on location and temperature data from weather dataset to identify colder regions in Canada

- -

Data Analysis on the demographic, census dataset to select the region with better business prospects

- -

Scraping the postal code(Borough level) data from Wikipedia and linking the location coordinates with the help of Geospatial dataset

- -

Utilising Foursquare API data to get the neighbourhood level venue categories

- -

Analysing each area basis the type of neighbourhood

- -

Identifying Cold Regions In Canada

Fig 1: Weather Station Dataset

- *Weather Station data from "Environment Monthly, July'15"*

- *25 Data points for 1341 weather stations spread across Canada*

- Location coordinates, min. and max. temperature readings, precipitation, snowfall, sunshine etc.

- Location-based data, hence using *DBSCAN (Density Based Spatial Clustering of Applications with Noise)*.
  - Works best with spatial data with noise
  - Locates regions of high density that are separated from one another by regions of low density.

- Radius of 300 mtrs and min. sample size of 10 taken

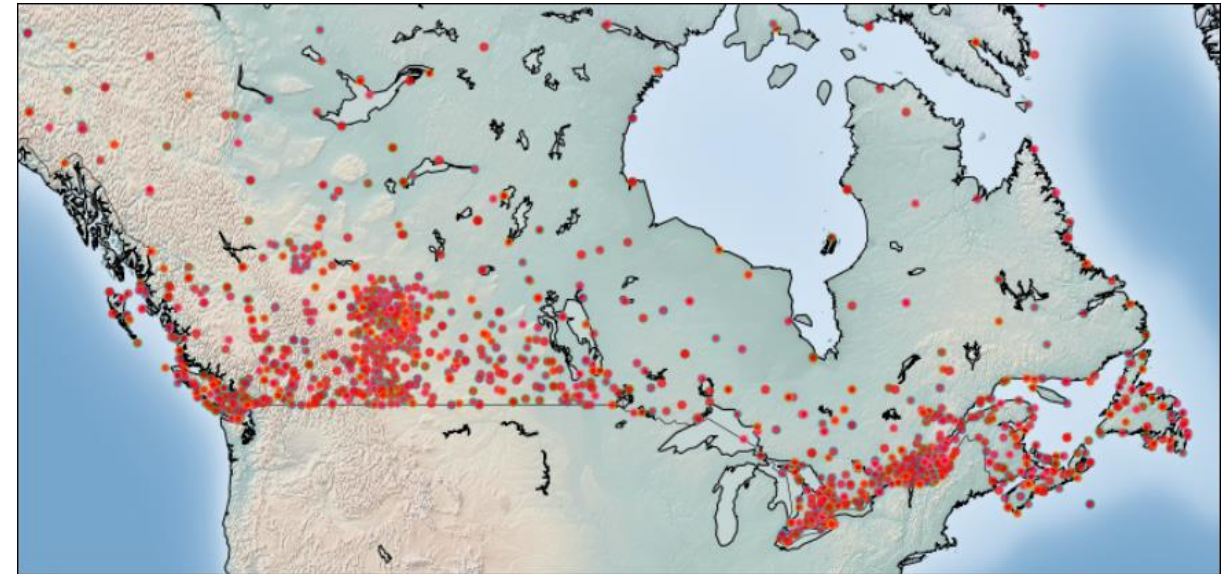- Total of 9 clusters and 428 outliers observed

Fig 2: Weather Station Visualisation on map of Canada
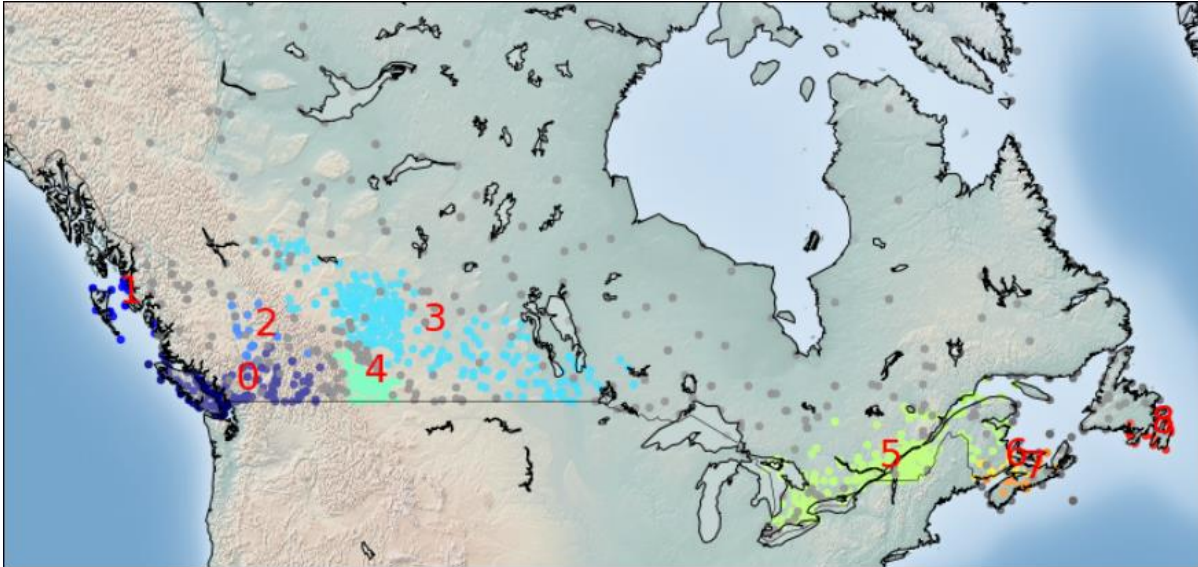
Identifying Cold Regions In Canada

Fig 3: DBSCAN Clusters of Weather Stations

- Clusters 5, 3 and 6 are the ones with lowest mean min. temperatures of around -15 ℃.

- Clusters in the western part are warmer compared to eastern

- Avg. min. temperature of western part above 0°C

- 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:

- More no. of weather stations, 266 compared to 215 in western part;

- Lower mean of minimum temperature, -16.3°C compared to -13.8°C

- 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.

- High number of clusters implies more footfalls hence more demand in the vicinity of our product.

Selecting the Region with Better Business Prospects

Eastern part consists majorly of two regions *QC (Quebec) and ON (Ontario).* The demographic and census data is further analysed
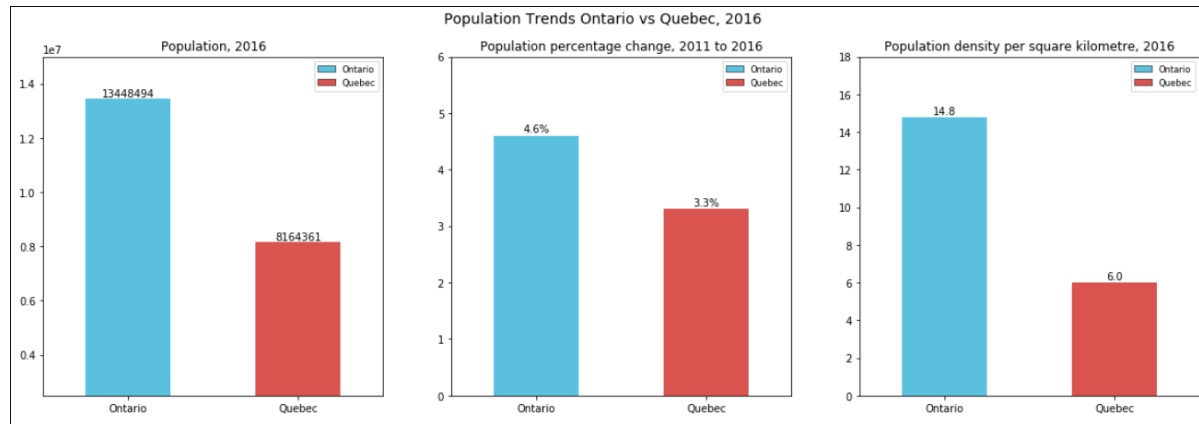
Fig 4: Population Trends - Ontario vs Quebec, 2016

- Ontario has higher population, higher population density hence more coverage for our brand
- Ontario has had a higher percentage increase in population over the years indicating a stable and rising economy

- Ontario has higher number of earning individuals, indicating more consumers
- Ontario has higher average household income than Quebec, indicating better spending potential
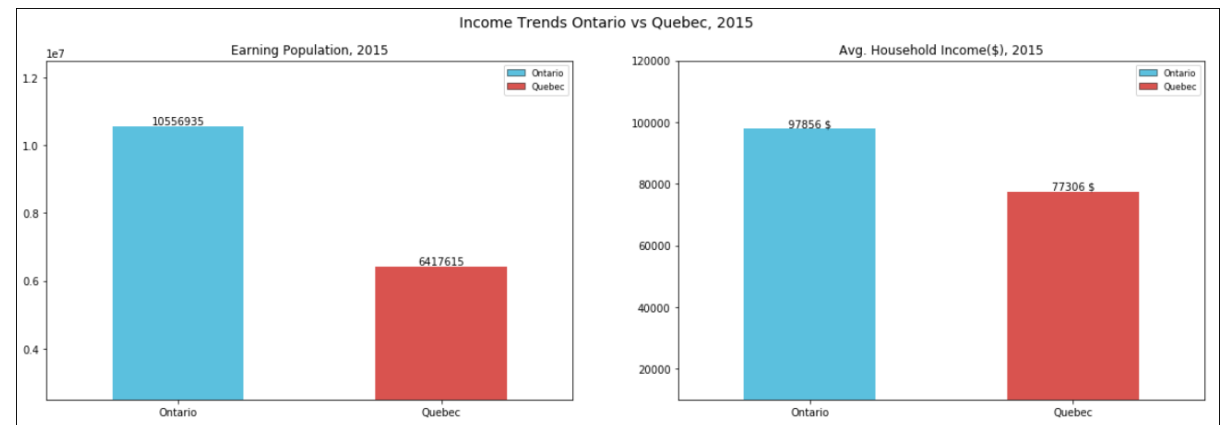
Quebec ✗ *Ontario* ✓

Fig 5: Income Trends - Ontario vs Quebec, 2016

Postal Code Data Analysis for Toronto City

In Ontario, due to limited availability of good quality district level demographic data, we will analyse Toronto City

- Postal code data scraped from Wikipedia to extract the Borough level data for Toronto City

- 103 postal codes in 10 Boroughs around Toronto city

- Geospatial dataset used 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 6: Postal Code Data with Location Coordinates

Fig 7: Shortlisted postal codes via DBSCAN

- DBSCAN clustering used to restrict area of focus

- Radius of 400 mtrs and minimum sample size of 5

- 56 postal codes/areas in 8 Boroughs shortlisted for neighbourhood analysis

Neighbourhood Data Analysis

Foursquare API data has been used to explore the neighbourhoods. For each area, we can explore the number of venues, their categories. This data helps in 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:

- 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 8: Most Common Venue Categories

	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 9: Common Venue Categories for each Postal Code

Neighbourhood Data Analysis

Using One hot coding methodology, we will analyse the venue categories in each neighbourhood.
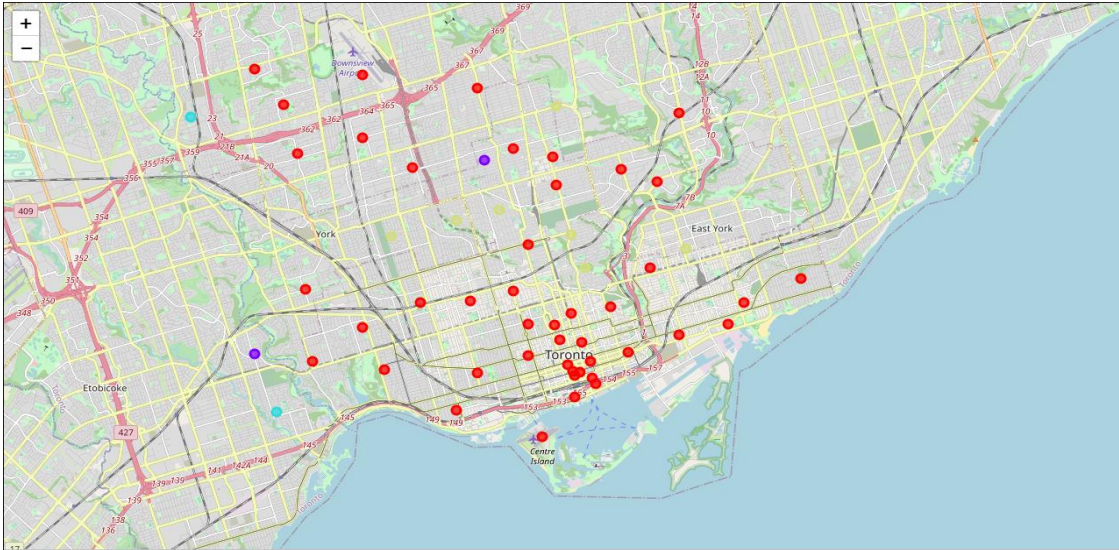
Fig 10: Clusters via K-Means

- K-Means cluster analysis used to cluster together similar areas basis the venues in the neighbourhood.

- Elbow method of reducing the within cluster inertia used to find the best value of K.

- Optimum value of "K" for cluster formation comes out to be "4".

- Red ones are the 45 areas belonging to one cluster while the remaining 11 belong to the other three clusters.

- 11 areas in three clusters have parks, fields, yoga studios, playgrounds along with one off stores, hence filtering them out

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 11: Clusters with parks, playgrounds and one-off stores

Neighbourhood Data Analysis

- 23 out of the red areas fall in Toronto Borough

- Shortlisted Boroughs:

  - *Central Toronto*

  - *Downtown Toronto*

- Areas spread over a radius of 5 km

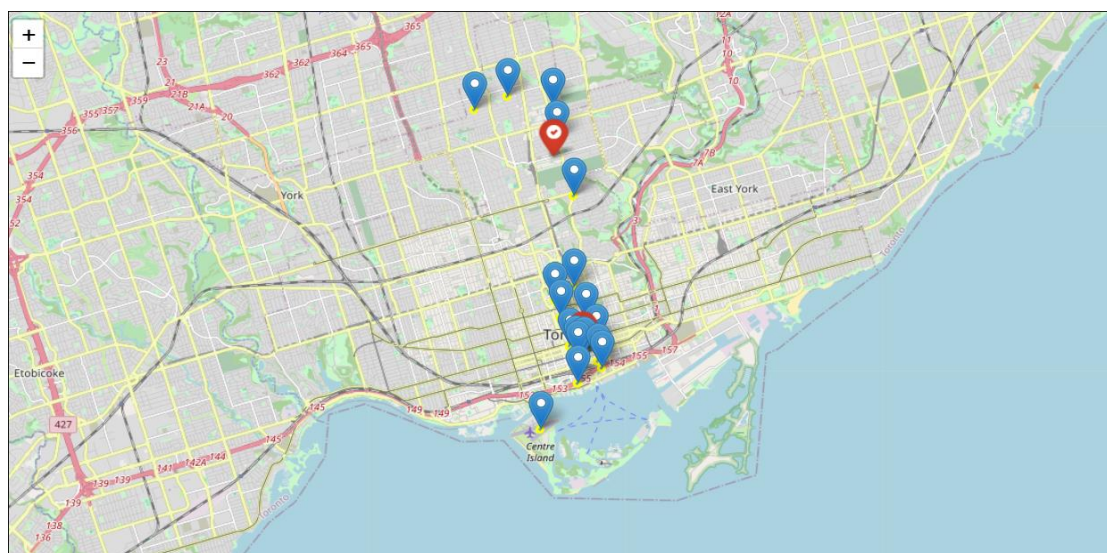- DBSCAN clustering used to cluster together high density areas

Fig 12: Areas with relevant venue categories in Central and Downtown Toronto

- 5 areas of Central Toronto and 13 areas in the high-density cluster of *Downtown Toronto.* within a radius of 1.5 km.

- 2 clothing stores and 1 shopping mall in each of these Boroughs

- No. of eateries, departmental stores, recreational centres are far more in Downtown.

- Presence of Union train station in Downtown , the busiest in Ontario.

Fig 13: High Density areas mapped via DBSCAN
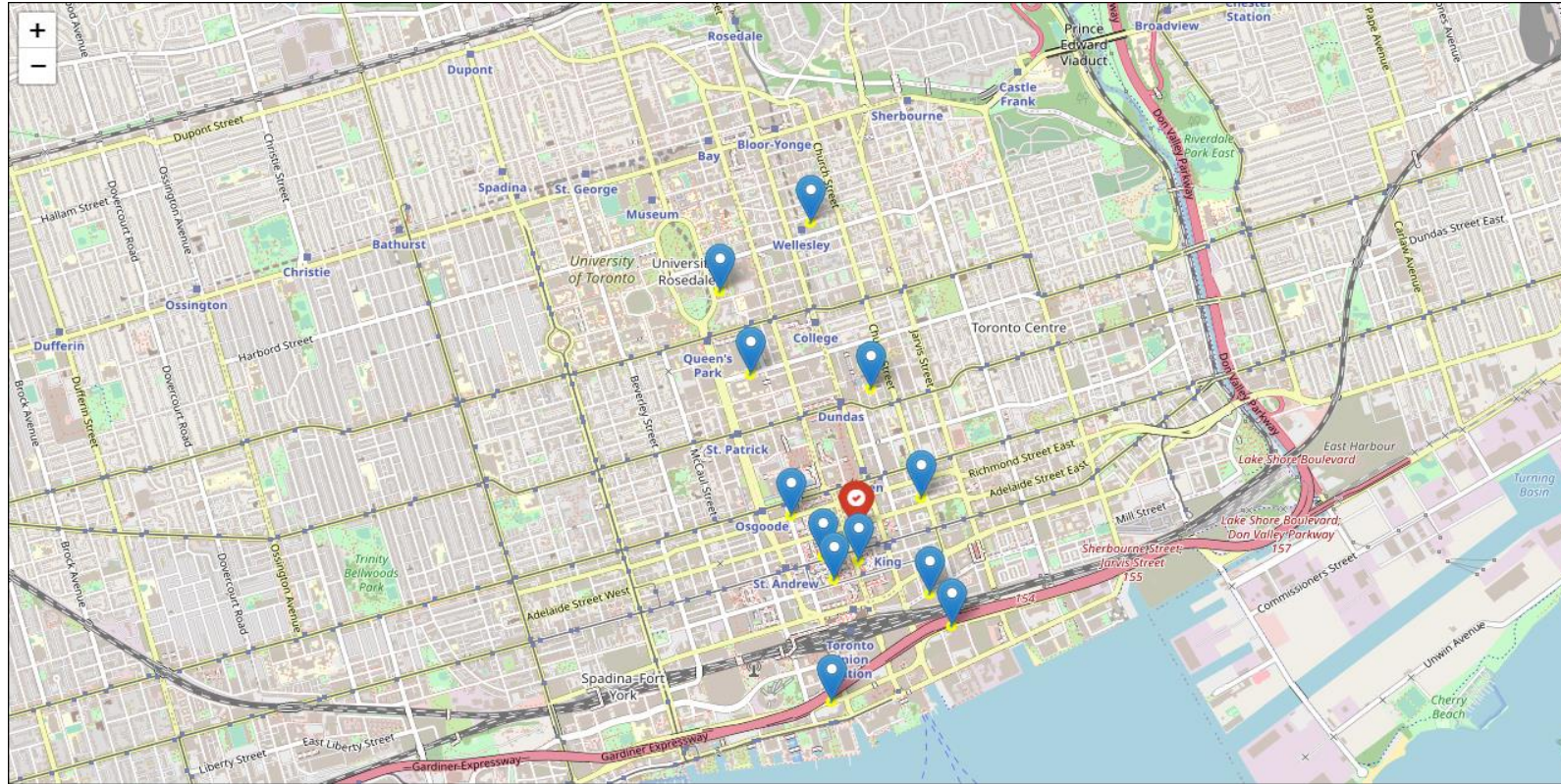
Results

Fig 14: Downtown Toronto

- Downtown is the centre of economic activity
  - High paying capacity individuals
  - High market potential.
- Presence of numerous eating outlets, recreational centres, centre of all activities
  - High visibility of the brand
  - High pedestrian traffic
  - Constant customer pull
- Union railway station, a central location
  - Easy accessibility

The 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.

"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."

- Economic Survey'18

Thankyou

Manika Babbar