Large Cities, Weather and Venues in Canadian Provinces

HEBA CHEHADE

MARCH 19, 2020

Introduction

Aim

To uncover the top 100 venues in each of the largest cities in the ten Canadian provinces in order to undertake a cluster analysis on the venue data.

Why?

To uncover insights that could serve entrepreneurs, marketing strategists and even those who support business development and growth in each of these provinces, who need to understand popular choices in each of the cities, how they differ and how to inform respective start-up and/or growth strategies and support programs.

Data - Cities

10 Canadian provinces with capital cities in each (while capital cities are often the largest cities, it is not always the case).

This study aims to uncover venue insights from the largest cities in each of the provinces to aid comparison. Source: Statistics Canada.

Data - Cities

Province	Largest City	Population (2019) *2017
Newfoundland and Labrador	St. John's	212,433
Prince Edward Island	Charlottetown	74,541*
Nova Scotia	Halifax	440,348
New Brunswick	Fredricton	60,879*)
Quebec	Montreal	4,319,000
Ontario	Toronto	6,472,000
Manitoba	Winnipeg	844,566
Saskatchewan	Saskatoon	330,674
Alberta	Calgary	1,515,000
British Columbia	Vancouver	2,691,000

Data - Weather

Extracted daily data between Jan 2014 until Dec 2018 to have a better picture of weather patterns over time.

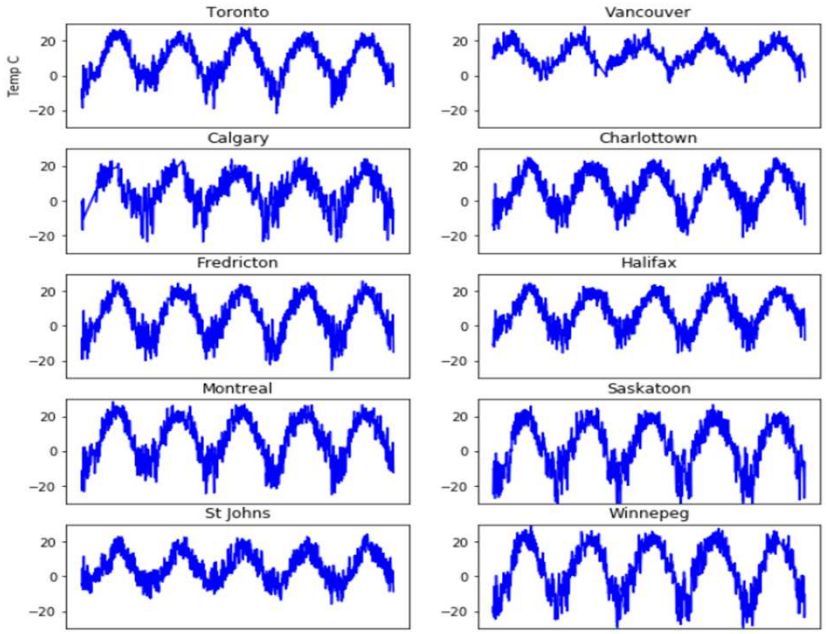
As some larger cities had more than one weather monitoring station, I used data from monitoring stations that were in close proximity to each of the largest cities above.

Data was extracted in csv format for use in the analysis. Source: Environment Canada.

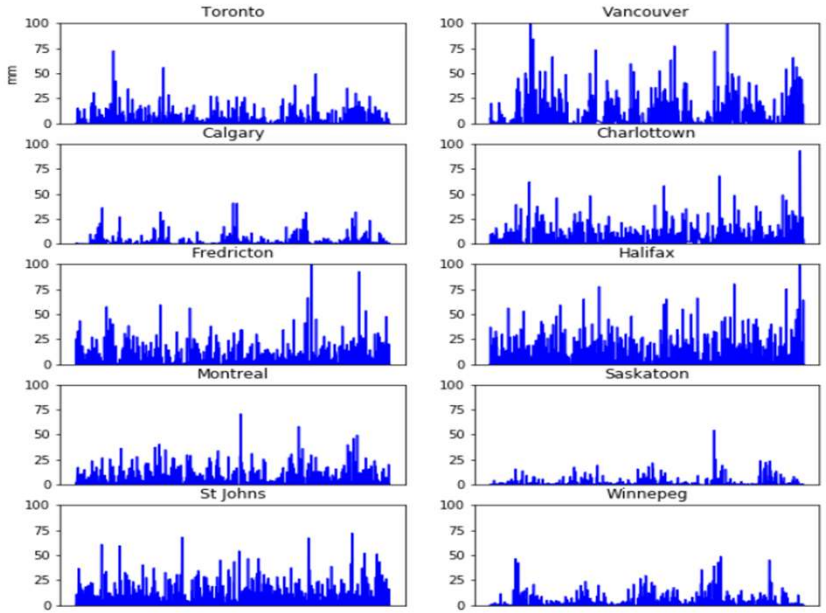
Data - Weather

City	Mean Temperature(C)	Min Temperature (C)	Max Temperature (C)	Mean Precipitation (mm)	Max* Precipitation (mm)
Toronto	9	-22	28	2	72
Vancouver	12	-4	28	5	100
Calgary	6	-24	25	1	41
Charlottetown	6	-21	25	3	93
Fredricton	6	-26	27	3	137
Halifax	8	-16	28	4	115
Montreal	7	-25	28	3	71
Saskatoon	3	-32	27	1	54
St John's	5	-16	24	4	72
Winnipeg	5	-29	29	1	49

Data – Temperature (2014-2018)



Data – Precipitation (2014 – 2018)



Data - Venues

Using a developer account and respective API credentials, Python code was written within Jupyter notebook to:

- extract the Top 100 venues
- within a 10 km radius from the latitude and longitude of the cities' centers
- 10 km was chosen because it is most likely the furthest distance that can be travelled by foot
- Data extracted March 14, 2020

Charlottetown and St. John's only had 91 and 90 venues extracted respectively.

Data - Venues

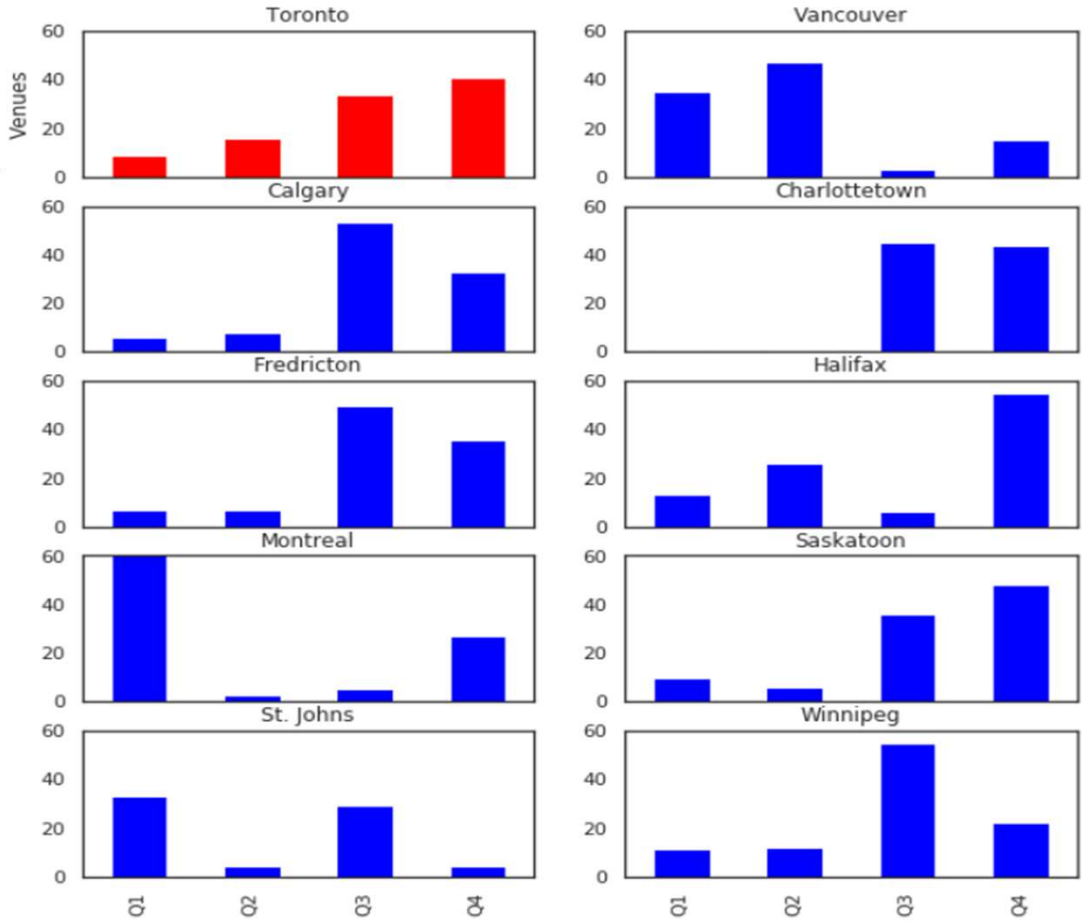
City	Unique Catgories	Categories that account for 50%+ of venues	Multiple check-ins?	Furthest Venue from City Center
Toronto	63	20	2	2.1 km
Vancouver	60	18	2	2.7 km
Calgary	53	12	1	2.8 km
Charlottetown	49	8	4	9.9 km
Fredricton	52	9	8	9.6 km
Halifax	57	11	1	3.6 km
Montreal	58	20	2	3.7 km
Saskatoon	61	15	1	6.6 km
St. John’s	40	14	1	9.9 km
Winnipeg	56	13	0	8.2 km

Data – Dispersion of Venues

Using venue coordinates as extracted from foursquare, the dispersion – in quarters - of the top 100 (exception of Charlottetown and St. John's) venues in each of the cities was examined with quarters defined clockwise starting from the midnight position and going all the way around.

*Note that the red color is only for illustrative purposes and does not signify anything unique about Toronto.

Data – Dispersion of Venues



Cluster Analysis

Converted venue data into a vector to enable the ability to run a K-means cluster analysis:

- Number of clusters is recommended to generally be between 3 and 5
- 3 was more appropriate given a sample of 10 cities

Cluster A:

- Small cities , venues dispersed around the 10 km radius, multiple venue check-ins
- Charlottetown, Fredricton, St. John's

Cluster B:

- Large cities, venues within 3.5 km from the center of each of the cities, diversified venues and limited multiple check-ins
- Toronto, Vancouver, Montreal

Cluster C:

- Cities between the above two clusters, venues between 3.5 km and 8.5 km from the center of each of the cities (except for Calgary), limited multiple check-ins but common top venues
- Calgary, Saskatoon, Halifax, Winnipeg

Cluster A: Small Cities

Cluster Labels	City	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	Charlottetown	Coffee Shop	Fast Food Restaurant	Gas Station	Seafood Restaurant	Restaurant	Ice Cream Shop	Sandwich Place	Pub	Grocery Store	Brewery
0	Fredricton	Coffee Shop	Restaurant	Grocery Store	Pub	Bar	Pizza Place	Gym	Hotel	Café	Liquor Store
0	St Johns	Coffee Shop	Restaurant	Bakery	Pub	Fast Food Restaurant	Gastropub	Scenic Lookout	Café	Ice Cream Shop	Park

Cluster B: Large Cities

Cluster Labels	City	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
1	Toronto	Coffee Shop	Café	Hotel	Italian Restaurant	Japanese Restaurant	Gym	Bakery	Farmers Market	Park	Concert Hall
1	Vancouver	Coffee Shop	Bakery	Park	Hotel	Seafood Restaurant	Ice Cream Shop	Café	Trail	Dessert Shop	Garden
1	Montreal	Café	Bakery	French Restaurant	Restaurant	Park	Vegetarian / Vegan Restaurant	Hotel	Japanese Restaurant	Yoga Studio	Pizza Place

Cluster C: Medium Cities

Cluster Labels	City	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
2	Calgary	Restaurant	Coffee Shop	Hotel	Deli / Bodega	Diner	Steakhouse	Café	Bakery	Italian Restaurant	Exhibit
2	Halifax	Coffee Shop	Restaurant	Hotel	Pizza Place	Pub	Italian Restaurant	Brewery	Seafood Restaurant	Burger Joint	Breakfast Spot
2	Saskatoon	Restaurant	Hotel	Café	Coffee Shop	Bakery	Pub	Breakfast Spot	Pizza Place	American Restaurant	Steakhouse
2	Winnipeg	Hotel	Coffee Shop	Ice Cream Shop	Restaurant	Brewery	Breakfast Spot	Diner	Pub	Pizza Place	Park

Challenges

Location accuracy

As already mentioned, St. John's coordinates had to be manually entered in order to extract the top 100 venues. Having said that, what is defined as 'city center' in Foursquare is something that any researcher conducting a similar study needs to double check prior to accepting the data generated to ensure that data is representative of what a researcher is seeking.

Data extracted from Foursquare by date / time

The other is related to the fact that one could not extract venues using the foursquare API specifying date / time. This limits a researcher's ability to run the code multiple times on the same data unless one ensures that the data is saved. This particularly became an issue when I would have liked to run the cluster analysis again on the same data setting the clusters to 4 and 5 to see if the results. I did not save the data because I did not expect the global pandemic to hit Canada this quickly which, as interesting as it is, is a key environmental factor that would shift data and make this study obsolete.

Conclusion

In conclusion, Foursquare data is very beneficial and can tell stakeholders a lot of information as long as one uses objective standards to set coordinates of cities and define the study well in terms of time as it might require a longer study than originally planned given limitations in the Foursquare API.

Canada's largest cities can be clustered according to their different venue characteristics, dispersion around the center of each of the cluster cities and customer behavior which provides greater insight on how entrepreneurs and those supporting the businesses, growth and development in the largest cities within those provinces.