Capstone Project

Esteban Hernan Elia

Battle of the Neighborhoods Week 2

Introduction.

Toronto, the provincial capital of Ontario, with a population of over 2.7MM people, is one of Canada biggest cities.

According to the Toronto Food Sector Update Report of 2010, "The food and beverage cluster is the second largest employment sector in Toronto, the largest sector of its kind in Canada and one of the largest in Canada or the U.S., rivalling Chicago and Los Angeles as a major continental center for the industry"

Toronto is also a world leader in recognizing the role of cities in using food as a lever to promote health, well -being and sustainability. In 1991, the Toronto Food Policy Council (TFPC) was established to advise the Board of Health and the city on issues relating to food security. Currently composed of 30 members including City Councilors, community, business and urban farmers, the TFPC was the first food policy council to be embedded within the government of a major city. It is widely considered a leader for food policy councils across the world, and a voice for a healthy, just and sustainable food system.

The current work is an exploratory analysis for a stakeholder seeking to permanently move to Canada and start a business in the Food Industry. It Analyzes the commercial areas of Toronto and the density of the several business in them, with the aim of getting familiar with the commercial landscape needed for accurately pinpoint the location of this future endeavor.

Data Gathering

Using the Forsquare API, we have been gathering data about a lot of venues in the Toronto area, so we have a starting point. Since this analysis is going to be industry specific, the data that we currently have its not going to be enough of what we need and we have to drop a lot of what we have.

So instead of doing that, we decided to reshape our searches to filter for food industry only venues, and for a bigger radius, and from that, try to segment into areas according to density, so the stakeholders can have a better understanding about how commercia areas are distributed.

Once we gather all the data we need, we should rank it for the top 5 type of Food Business in Toronto and cluster them so we have the big picture about how commercial areas are distributed all around Toronto.

Methodology

First, we will need to get all the food related venues from Toronto, and after that, clean the data so we can start clustering.

We define the new search url with the following parameters.

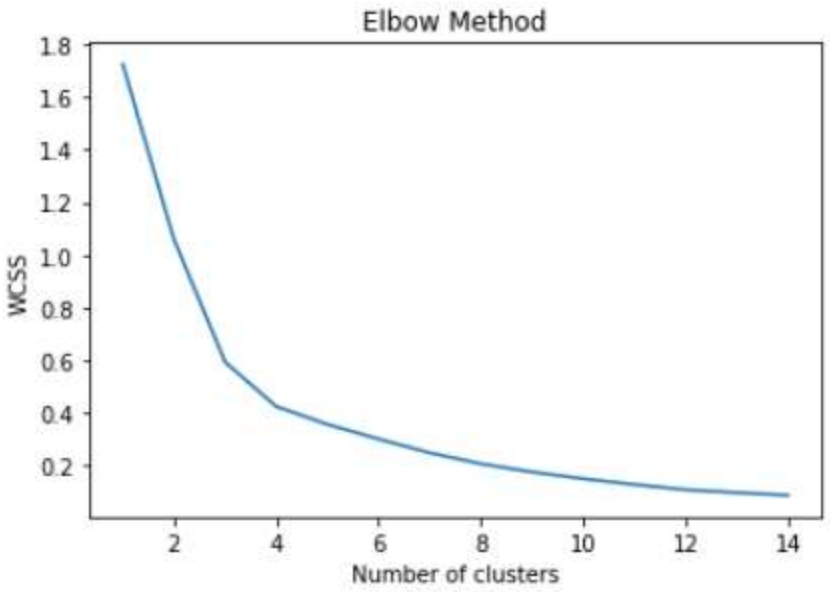
```
'https://api.foursquare.com/v2/venues/explore?&client_id=Lxxx&client_secret=A
xxx&v=20180605&ll=43.806686299999996,-97.19435340000001 &radius=1200
&intent=browse&limit=100 category=4d4b7105d754a06374d81259'
```

We added &category=4d4b7105d754a06374d81259, as the Foursquare API docs indicate is for food industry (https://developer.foursquare.com/docs/resources/categories)

After this, we clean the data reusing the functions defined before, and get a table with the top 5 Food Business

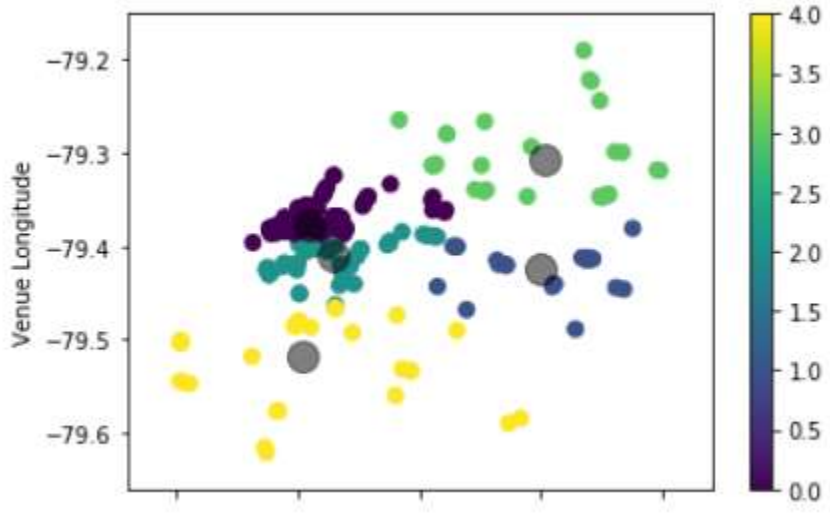
Venue Category	Venue Latitude	Venue Longitude
Bakery	50	50
Café	98	98
Coffee Shop	192	192
Pizza Place	51	51
Restaurant	61	61

We use Wcss to find the optimal K for K-Means



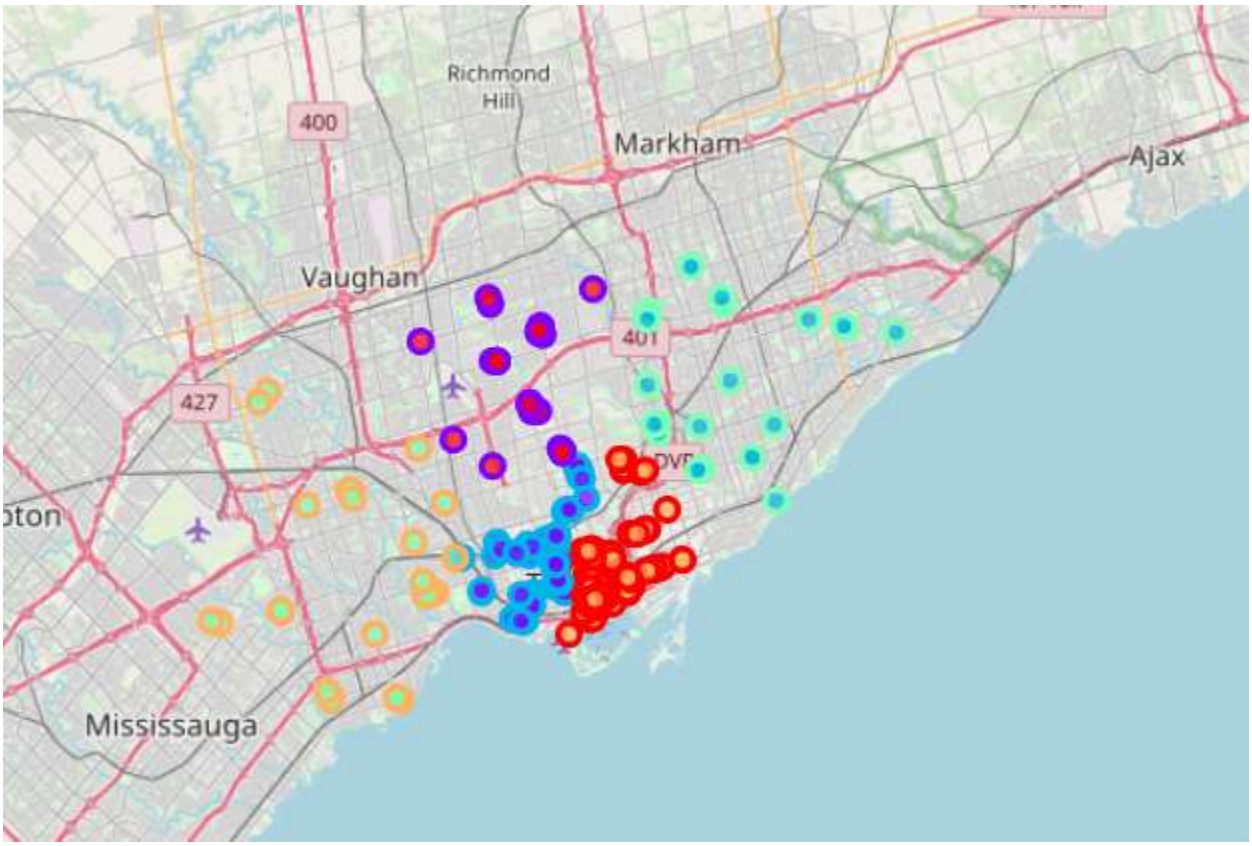
The optimal number of clusters is between 3 and 5, after that , we don't really marginally improve much.

To have an idea of were the centroids ended, we draw a quick scatterplot



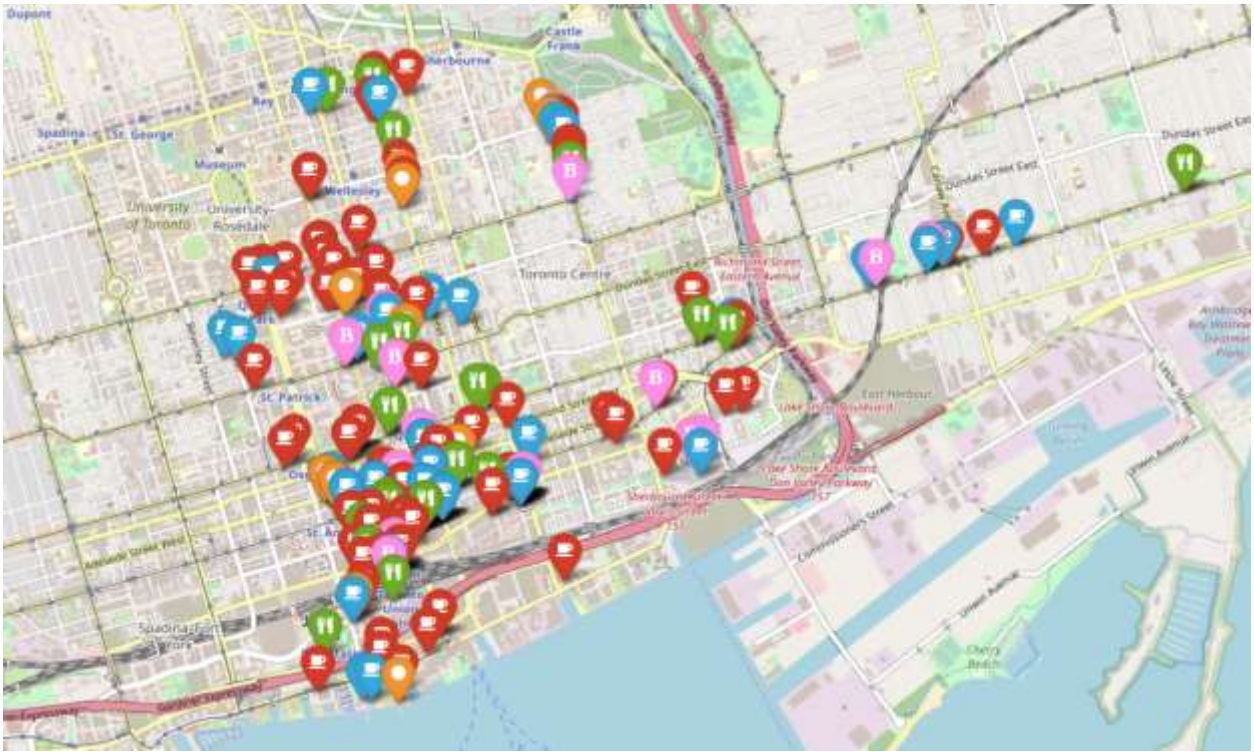
We can see here that there are mainly 2 centric areas where it has a heavy density of samples, and 3 outside areas where the food venues are scattered around the neighborhood.

We scatter the areas on the map.



We will start with the area coded in red, we will name it RED AREA (we will refer to the border color for the circles, the filling is for contrast so the area can be seen on the map), for clarification here is the map legend for the following 5 maps as well.

RED AREA (Downtown)

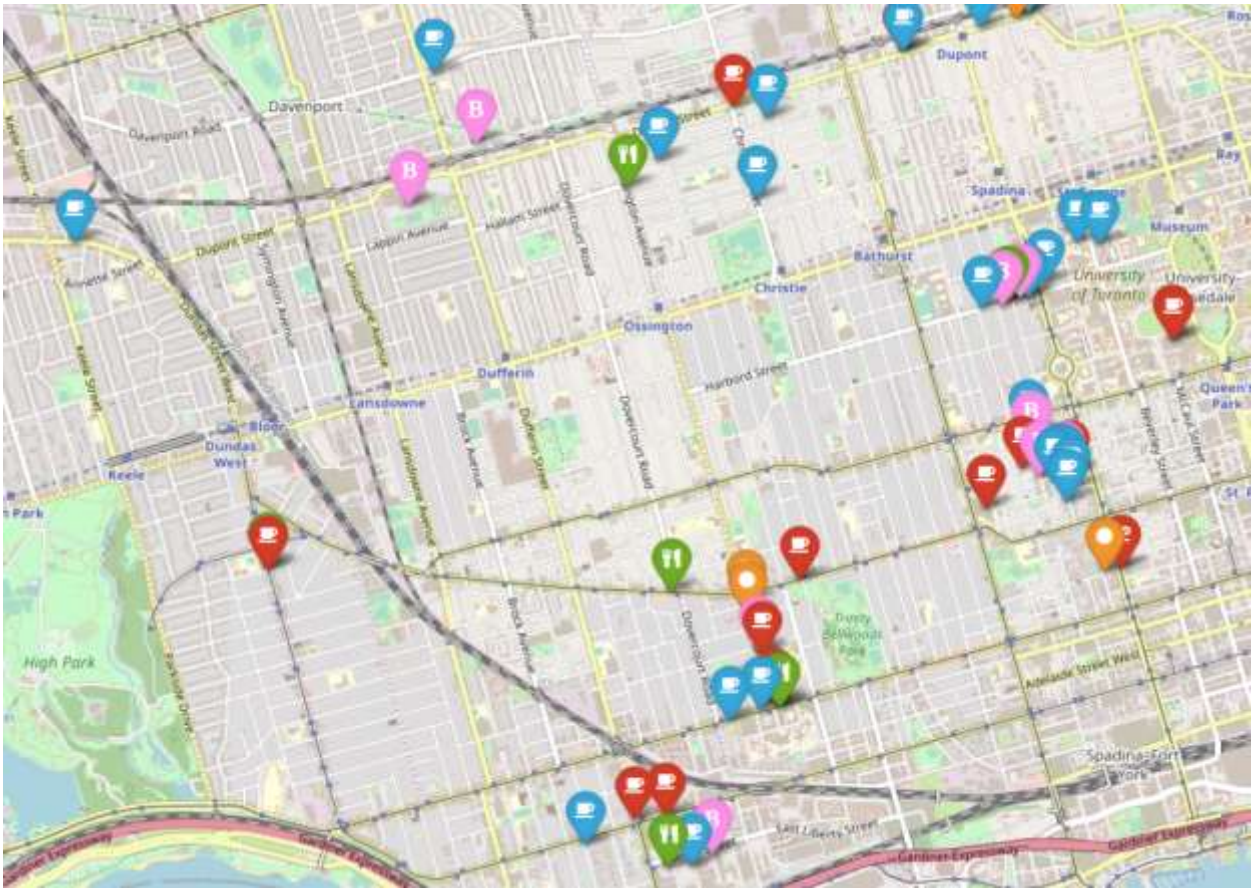


Filtering by amount of total venues in the area we have the following.

Venue Category	Amount
Coffee Shop	141
Café	58
Restaurant	40
Bakery	30
Pizza Place	19

We can see that is area has a LOT of Coffee related business and clearly is the denser of the 5 areas. We should explore a little further before jumping into conclusions.

BLUE AREA



The total counts for the blue area are as follow:

Venue Category	Amount
Café	25
Coffee Shop	19
Restaurant	11
Bakery	9
Pizza Place	8

We can see that we are near the downtown area but concentration is lower by a significant amount.

SUBURB AREAS

I decided to group this 3 areas together since they show similar characteristics, they all have very low density and are on the outside of downtown.

Venue Category	Amount
Coffee Shop	11
Café	5
Pizza Place	5
Restaurant	5
Bakery	1

Venue Category	Amount
Coffee Shop	13
Bakery	6
Pizza Place	6
Café	3
Restaurant	1

Venue Category	Amount
Pizza Place	13
Coffee Shop	8
Café	7
Bakery	4
Restaurant	4

Analysis:

Red Area (Downtown)

Venue Category	Amount	Percentage
Coffee Shop	141	48.96
Café	58	20.14
Restaurant	40	13.89
Bakery	30	10.42
Pizza Place	19	6.60

If we focus on the Red Area, we can see that the vast majority of venues are Cafes and Coffee Stores. And there's very few pizza places in comparison (only 6% of the total venues). This could be an opportunity since the area is as dense as it gets for food where you have some restaurants but very few Pizza Places in comparison.

Blue Area (Next to Downtown)

Venue Category	Amount	Percentage
Café	25	34.72
Coffee Shop	19	26.39
Restaurant	11	15.28
Bakery	9	12.50
Pizza Place	8	11.11

Here, things even out a little bit, we still have majority of Coffee related venues but he rest seems to have even out, making the other half.

This area could be an opportunity depending how close to downtown you could get a location, and if we check the map, we can see that the Pizza places for this area are also pretty far away from the downtown area, except for one. So this could be also an opportunity if the stakeholders can find a good location near downtown (bordering the left side of the blue area map)

Suburb area

Given the low density of venues, we should disregard this areas for this analysis, since we cant conclude much other than the venues located are randomly spaced over the area.

Conclusions

We can conclude that there are some opportunities to invest in a venue around the Red Downtown area, specifically in the Pizza Place category, since there's seems to be way fewer offers than a dense area like the one we analyze should offer, only 6% of the total venues, where Cafes and Coffee Stores sum up to 70%.

There could be similar opportunities on the Blue Area if the stakeholders could find a location closer to the downtown area, because the vast majority of the pizza places in this area are far away, even though the percentages are closer together for the remaining categories excluding Coffee Store and Café.

References:

https://www.toronto.ca/legdocs/mmis/2018/hl/bgrd/backgroundfile-118079.pdf

https://www.toronto.ca/wp-content/uploads/2017/08/9601-2010_Toronto_Food_Sector.pdf