Proposal for Clothing store Build site

Outline

+ Introduction
+ Data Description
+ Methodology
+ Results
+ Recommendations and future work

Introduction

- The Waterloo region has experienced linear growth patterns in population since 2011. According to the 2016 Census of Canada, Waterloo Region is home to 535,154 people, which represents 4.0% of the total population of Ontario (Smale, 2017).
- Its fast growing and businesses like Google and Amazon are quickly setting up shop within the Region
- My client wants to know the best locations for setting up their clothing store.
- They want to know where the most populated areas are and what is nearby

Data Description

- Consist of :
  - List of neighborhoods for the region of Waterloo
  - Latitude and Longitude for each Neighborhood
  - Venues that are in each of the neighborhoods

Methodology – Data Cleaning

- The data was not available online so I had to manually input the Neighborhoods, City it belonged to, the Province, Postal code and Country.

	Neighborhood	Borough	Province	Country	Post
0	Centreville Chicopee	Kitchener	Ontario	Canada	N2A
1	Grand River South	Kitchener	Ontario	Canada	N2A
2	Heritage Park	Kitchener	Ontario	Canada	N2A
3	Idlewood	Kitchener	Ontario	Canada	N2A
4	Stanley Park	Kitchener	Ontario	Canada	N2A

Methodology – Data Cleaning

- Next, I had to concatenate the 5 columns to make a new column named Address.

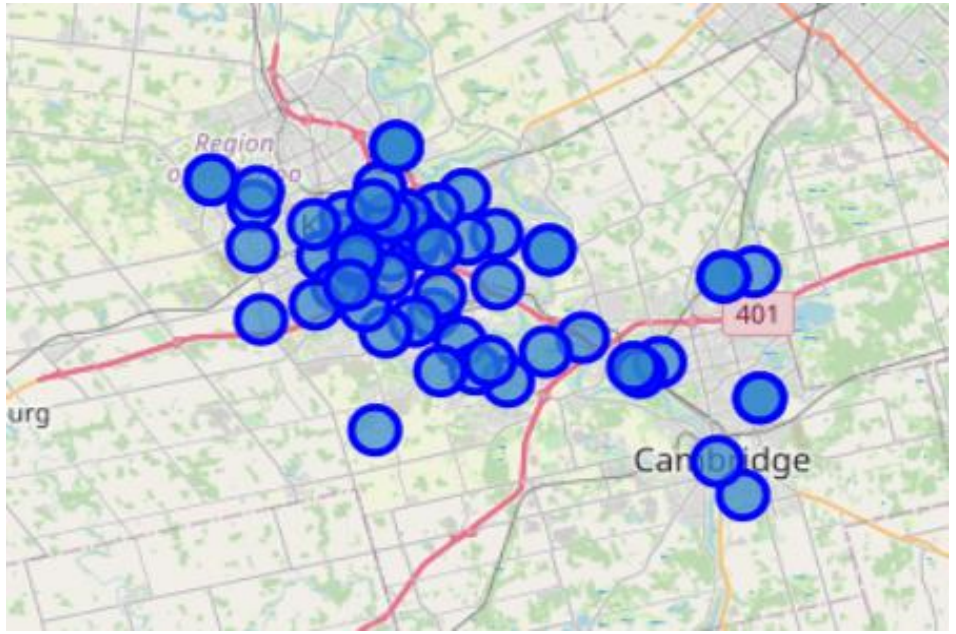
  This was due to inaccurate Lat and Longs coordinates from Foursquare when only using the Neighborhood column.

- Now I was able to acquire the Geospacial coordinates of the Neighborhoods using foursquare

	Neighborhood	Borough	Province	Country	Post	address
0	Centreville Chicopee	Kitchener	Ontario	Canada	N2A	Centreville Chicopee,Kitchener ,Ontario,Canada
1	Grand River South	Kitchener	Ontario	Canada	N2A	Grand River South,Kitchener ,Ontario,Canada
2	Heritage Park	Kitchener	Ontario	Canada	N2A	Heritage Park,Kitchener ,Ontario,Canada
3	Idlewood	Kitchener	Ontario	Canada	N2A	Idlewood,Kitchener ,Ontario,Canada
4	Stanley Park	Kitchener	Ontario	Canada	N2A	Stanley Park,Kitchener ,Ontario,Canada

Methodology – Data Exploration

- Now I was able to create a map using folium
- Note: All the neighborhoods in the Waterloo Region were set with a radius of 1000m.



Methodology – Nearby Venues

- Here I used my Foursquare credentials and grabbed each Venue within a 100m radius or every Neighborhood.

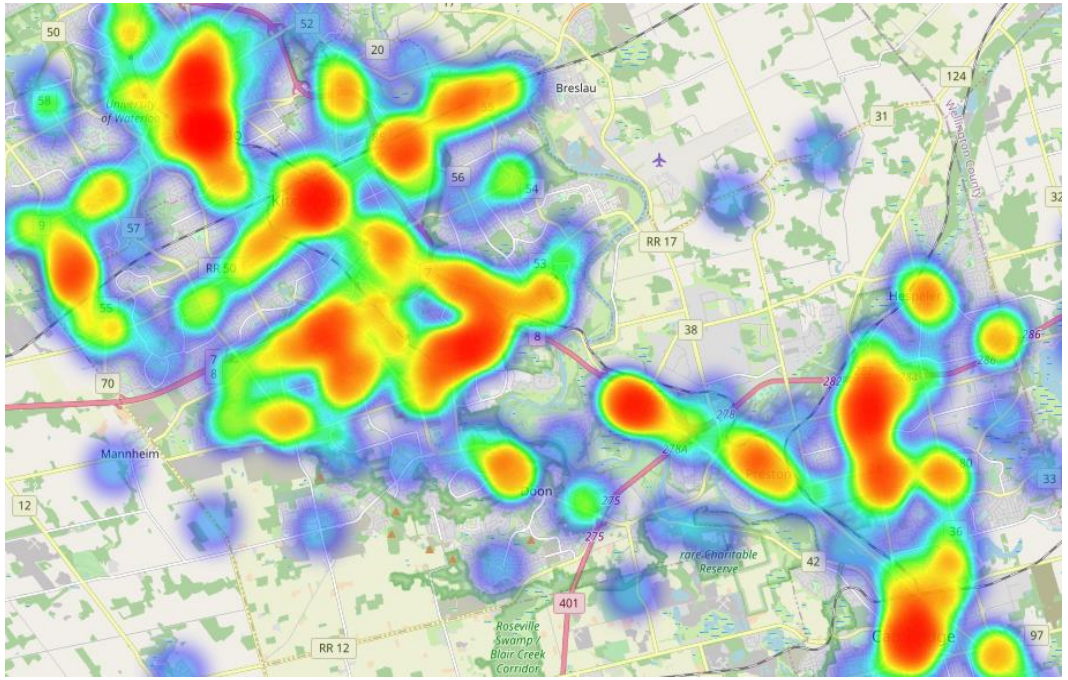
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
0	Centreville Chicopee	43.428576	-80.429814	Martini's restaurant	43.428779	-80.432767
1	Centreville Chicopee	43.428576	-80.429814	The Bent Elbow	43.430146	-80.436878
2	Centreville Chicopee	43.428576	-80.429814	Robert's Boxed Meats	43.428291	-80.421474
3	Centreville Chicopee	43.428576	-80.429814	Bangkok Cuisine	43.431065	-80.437437
4	Centreville Chicopee	43.428576	-80.429814	Chicopee Ski & Summer Resort	43.435920	-80.418340
5	Centreville Chicopee	43.428576	-80.429814	Firehouse Subs	43.421662	-80.443819
				GoodLife		

Methodology – Removing Duplicates

- Since the radius was set at 1000m there was a lot of overlap between neighborhoods.
- So to compensate for that I removed all duplicate Venue values based on its proximity to the nearest Neighborhood.

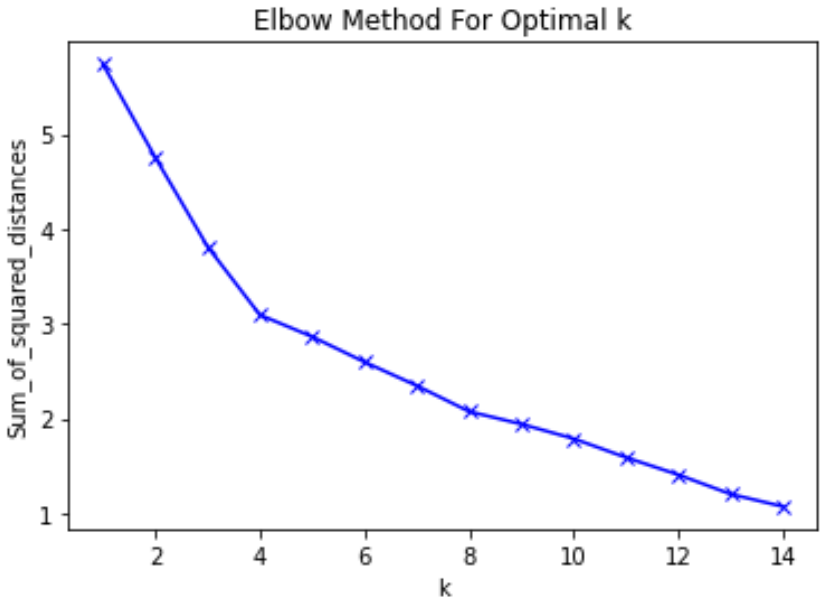
Methodology – Heat Map

- To see the spread of the data I created a heatmap to make sure that the venues all had the correct latitude and latitudes assigned by the Foursquare API.
- This also allowed me to see the distribution of the Venues.



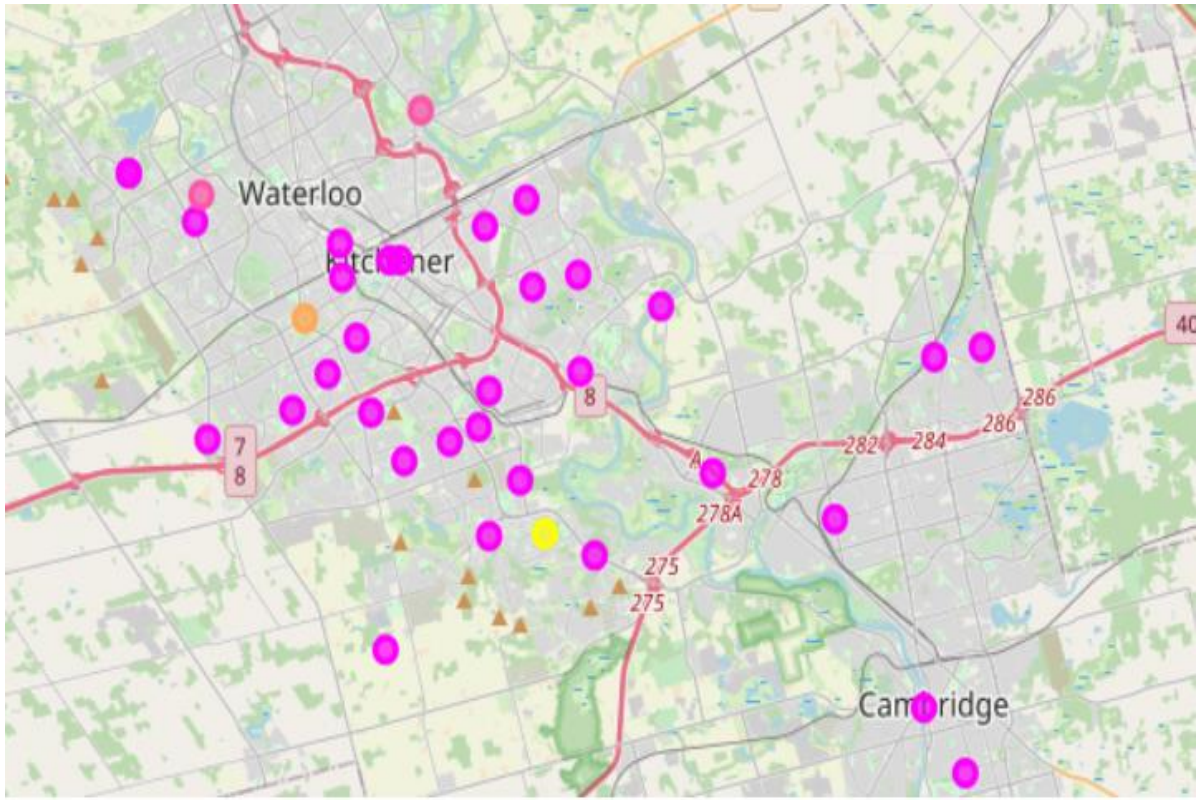
Methodology – Elbow Method

- To find the value for k I had to run an elbow method on the One encoded dataset.
- From the image I selected a value of 4 because of the slight bend see at k = 4



Methodology - Clustering

- After selecting k as being 4 I ran the KNN on the Neighborhoods to determine which venues in each cluster.
- After running KNN I produced the map seen on the left.



Results – Cluster 0

- The first cluster, cluster 0 (Orange) was considered to be on the city's perimeter. The only venue is had was a bank.

Bank	1.0
Yoga Studio	0.0
Fish Market	0.0
Fish & Chips Shop	0.0
Fast Food Restaurant	0.0
Farmers Market	0.0
Farm	0.0
Falafel Restaurant	0.0
Electronics Store	0.0
Eastern European Restaurant	0.0

Results – Cluster 1

- Cluster 1 (Purple) contained all of the venues with high foot traffic. This included Coffee shops, grocery stores, and restaurants

Coffee Shop	0.071817
Grocery Store	0.049662
Sandwich Place	0.040241
Restaurant	0.036553
Pizza Place	0.032019
Pharmacy	0.027452
Bank	0.026135
Fast Food Restaurant	0.023931
Café	0.023856
Hotel	0.020698

Results - Cluster 2

- Cluster 2 (Red) Here we have another cluster located on the perimeter of the city,. Between Waterloo and Kitchener. This cluster consisted of Skating rinks, Breakfasts spots and Breweries.

Skating Rink	0.50
Breakfast Spot	0.25
Brewery	0.25
Cupcake Shop	0.00
Department Store	0.00
Design Studio	0.00
Dessert Shop	0.00
Diner	0.00
Discount Store	0.00
Convenience Store	0.00

Results - Cluster 3

- Cluster 3 (Yellow) Was on the outskirts of the city, between Kitchener and Cambridge. Considered to be an industrial park it only had a paper supply store in its vicinity.

Paper / Office Supplies Store	1.0
Yoga Studio	0.0
Fish Market	0.0
Fast Food Restaurant	0.0
Farmers Market	0.0
Farm	0.0
Falafel Restaurant	0.0
Electronics Store	0.0
Eastern European Restaurant	0.0
Discount Store	0.0

Limitations

- The foursquare dataset was quite limited and only returned venue location for a fraction of the neighborhoods in scope.
- The lack of information available online created issues when acquiring the data.

Recommendations

- Based on the results we seen four different clusters, 1 in the city center, 2 on the perimeter, and 1 on the outskirts
- From this information we can see the venues is the city center and how many venues are in each neighborhood.
- However it is difficult to conclude the best location to build we have a good place to start. For example areas with a lot of restaurants generate foot traffic so looking at a neighborhood that has a lot of restaurants might be a good place to start.

Future Work

- In the future more metrics will be included. Such as cost of operating a business in certain neighborhoods and the rate in which neighborhoods are growing.
- Additionally being able to pinpoint an exact location given a set of conditionals would be ideal moving forward.