Final Project Summary

Intro

1. **Problem & background:** Someone is looking to open a coffee shop in Toronto and needs a recommendation on where they can open, based on competition, location to foot traffic, and population.

   Audience: An entrepreneur looking to open a coffee shop in Toronto

2. **Data & usage:** We will use a number of different data sources to explore neighborhoods and postal codes.

   -- as this is a good indicator of local foot traffic

   a) 2016 population census Canada for population: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/Tables/File.cfm?T=1201&SR=1&RPP=9999&PR=0&CMA=0&CSD=0&S=22&O=A&Lang=Eng&OFT=CSV

   b) postal codes and their neighborhoods: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

   -- as people may be more familiar with Nieghborhoods then with postal codes

   c) foursqure data API (foursquare developer credentials required)

   -- We will be pulling in latitudes and longitudes of all venues within 500 metres (as this is an indication of foot traffic)

   -- We will also pull in categories of these venues and find out how many of these venues are coffee shops and cafes to understand the competition in the area.

3. **Assumptions:**

   a) Venues that are not coffee shops (e.g. parks, businesses, restaurants), create good foot traffic for coffee shops

   b) Coffee shops within the same Postcode are bad for traffic as it is competition

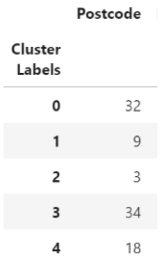
   c) A % of the Population of that Postcode will be considered foot traffic for that coffee shop

Methodology

1. We collected Data of Toronto Neighborhoods by postal code, and population by postal code.
2. We then enriched the data with all venues within 500 metre radius of that postal code.
3. We also enriched our data by categorizing all Coffee shopes within that zip code.
4. We attributed an estimated foot traffic to our coffee shope by: a) total population of Toronto/Total venue/ Total coffeeshops of the postal code in Toronto *1% (this is an estimate of how much foot traffic we would get as a result of nearby venues(e.g. parks, schools, restaurants) b) Total population of the postal code/ Total cofeeshops of the postal code *1%
5. We used a cluster analysis to group each postal code by lowest to highest total foot traffic

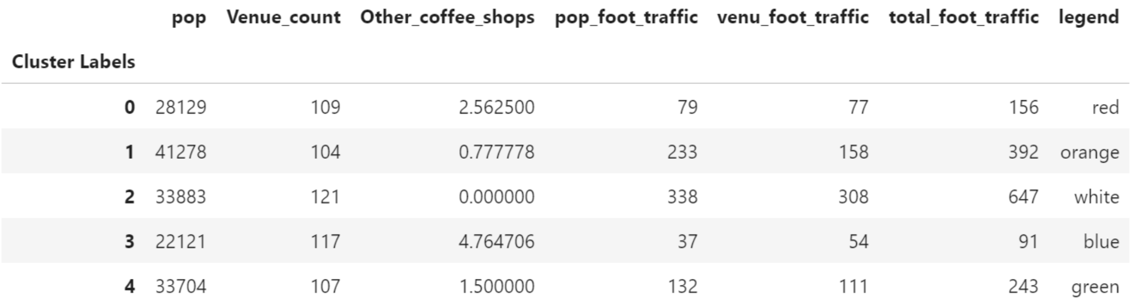
Results

Postal code within each cluster:

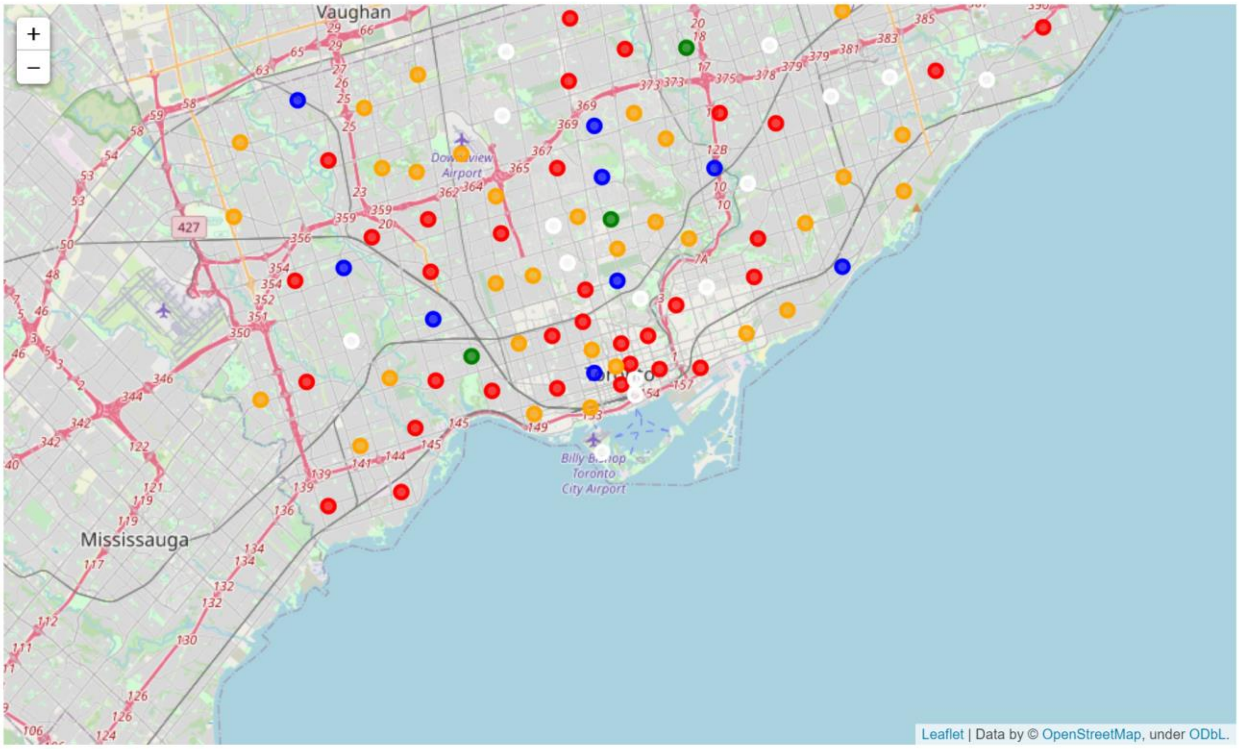


	Postcode
Cluster Labels	
0	32
1	9
2	3
3	34
4	18

Mean of each cluster:



	pop	Venue_count	Other_coffee_shops	pop_foot_traffic	venu_foot_traffic	total_foot_traffic	legend
Cluster Labels							
0	28129	109	2.562500	79	77	156	red
1	41278	104	0.777778	233	158	392	orange
2	33883	121	0.000000	338	308	647	white
3	22121	117	4.764706	37	54	91	blue
4	33704	107	1.500000	132	111	243	green


+
−
Vaughan
Downsview Airport
427
Mississauga
Billy Bishop Toronto City Airport
Leaflet | Data by © OpenStreetMap, under ODbL.

Discussion

It is important to note a few gaps and limitations within this data analysis.

1. As postal codes can contain different geographical sizes, population density is better measure then just population. Due to limited data, we had to use population
2. Cost of land, or business is not factored in. It would have been great to look into businesses for sale and do a comparison with the recommendation.
3. Venues can be deeper classified and weighted. For instance, parks, museums as a category should be weighted higher than a convenient store as it relates to bringing in foot traffic.

Recommendation

Based on the analysis, we recommend looking at locations within **cluster 4** as it is characterized by many venues and few competition, and a high population which should bring a high number of foot traffic to the area.