Predicting promising locations for opening a bakery in Toronto

Population growth = Business growth
But where should you build?

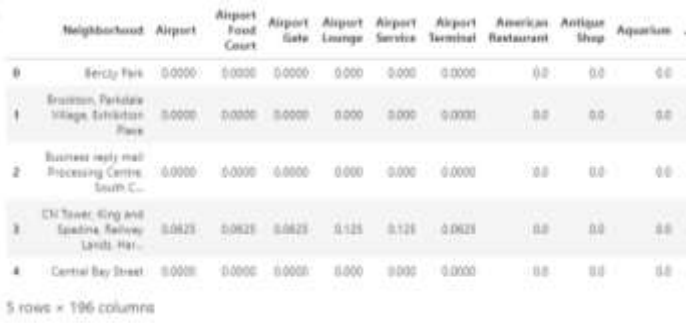
- According to a 2019 study from Ryerson University, Toronto was the second fastest growing Metropolitan Area and the top growing city in all of the United States and Canada.
- New business will be needed to support the growth in population
- Location is a key consideration when opening a business and can determine its success or failure
- K-Means clustering will help identify areas suitable for development

Data acquisition and cleaning

- List of postal codes of Canada from Wikipedia:
  - 'Not assigned' boroughs or neighborhoods were dropped
- Latitude and Longitude of Canada postal codes from Cocl.us
- Venue information including category and location were obtained from Foursquare.com
- Postal codes and Latitude and Longitude datasets were combined to form a complete 'picture' of Toronto

Foursquare Data – average occurrence of venue

- The average occurrence of each venue was obtained was calculated for each neighborhood



	Neighborhood	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium
0	Berczy Park	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.0	0.0	0.0
1	Brockton, Parkdale Village, Exhibition Place	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.0	0.0	0.0
2	Business reply mail Processing Centre, South C...	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.0	0.0	0.0
3	CN Tower, King and Spadina, Railway Lands, Har...	0.0625	0.0625	0.0625	0.125	0.125	0.0625	0.0	0.0	0.0
4	Central Bay Street	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.0	0.0	0.0

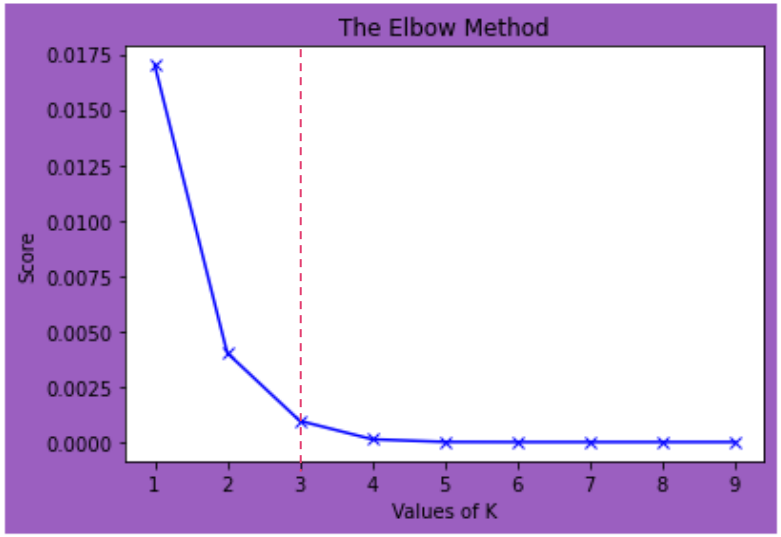
5 rows × 196 columns

- Next, all businesses except bakeries were filtered out

	Neighborhood	Bakery
0	Berczy Park	0.033333
1	Brockton, Parkdale Village, Exhibition Place	0.043478
2	Business reply mail Processing Centre, South C...	0.000000
3	CN Tower, King and Spadina, Railway Lands, Har...	0.000000
4	Central Bay Street	0.000000
5	Christie	0.000000
6	Church and Wellesley	0.000000
7	Commerce Court, Victoria Hotel	0.033333
8	Davisville	0.000000
9	Davisville North	0.000000
10	Dufferin, Dovercourt Village	0.153846

Elbow method

- In order to determine the optimal amount of clusters, the Elbow method was deployed



- Based on the 'elbow', we can see that the optimal number of clusters is 3

K-Means Clustering

- Using K-Means clustering, clusters were assigned to our bakery data and a map was created displaying its results

1 Cluster 0 = Red
- Top choice for new bakery location

3 Cluster 1 = Purple
- Not recommended for new bakery location

2 Cluster 2 = Green

Recommendation and Conclusion

K-Means was used to help predict the optimal location for a new bakery

- Cluster 0 has the least occurrences of bakeries among the clusters which makes it the best choice when considering locations for a new business
- Cluster 1 had the most occurrences of bakeries and is not recommended for a new bakery location

- This same model can be used to consider other business locations in Toronto, Canada
  - With a few modifications, it can also be used to evaluate other cities worldwide