IBM Applied Data Science

Coursera Capstone Project

Toronto Neighborhoods Report

Optimal Location for a Specialty Pastry Shop

Rudolph Persaud

Feb 2021

1.0 Introduction

This report describes an analysis of available data that was undertaken to establish the retail potential of the many distinct neighborhoods that exist in the city of Toronto and to determine the optimal location for a specialty bakery of pastries and other delicacies that originate from the country of Guyana in the Caribbean region.

1.1 Background

Toronto is one of the largest and the most populous cities in Canada and is situated in the southern part of the province of Ontario. The city is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world. With a population of 2.7 million, Toronto it the fourth most populous city in North America, after Mexico City, New York City, and Los Angeles.

The diverse population of Toronto reflects its current and historical role as an important destination for immigrants to Canada. More than 50 percent of residents in the city belong to a visible minority population group and over 200 distinct ethnic origins are represented among its inhabitants. This diversity is reflected in Toronto's ethnic neighbourhoods, which include Chinatown, Corso Italia, Greektown, Kensington Market, Koreatown, Little India, Little Italy, Little Jamaica, Little Portugal and Roncesvalles (Polish community).

The city of Toronto encompasses a geographical area formerly administered by many separate municipalities. These municipalities have each developed a distinct history and identity over the years and their names remain in common usage that includes East York, Etobicoke, Forest Hill, Mimico, North York, Parkdale, Scarborough, Swansea, Weston and York. Throughout the city there exist hundreds of small neighbourhoods and some larger neighbourhoods covering a few square kilometres.

There are 140 neighbourhoods officially recognized by the City of Toronto and were developed to help government and community organizations with their local planning by providing socio-economic data at a meaningful geographic area. The boundaries of these neighbourhoods do not change over time and are reflected in the many datasets published by the city.

The focus of this analysis is on the expatriate Guyanese community residing in the city of Toronto. Guyana is part of the South American mainland however it is considered a culturally Caribbean country even though it is not an island nation located in the Caribbean Sea as are most Caribbean nations. Based on the 2016 Canadian population census from Statistics Canada, there were over 87k immigrants from Guyana with the majority residing in Toronto.

Every year in the summer, the city is the host of the Caribana Festival which is one of the largest celebrations of Caribbean culture and traditions in North America and brings in over two million people each year. During this celebration, there is a significant increase in demand for all types of Caribbean foods and pastries.

1.2 Problem Description

The objective of this detailed analysis of the neighbourhoods in the city of Toronto, was to identify an ideal location for a bakery of Guyanese pastries with a storefront shop for sale to retail consumers, to local businesses including restaurants and subsequently to the general population as these delicacies become known and accepted.

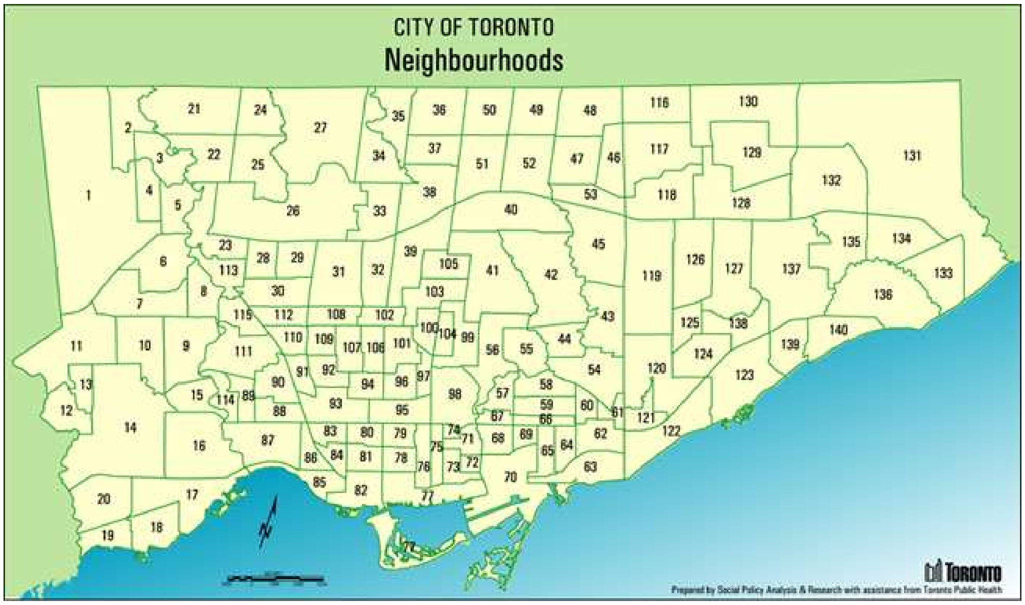


Figure 1 – Toronto Neighbourhoods

Pastries from Guyana are distinct and unique to the region and not readily available in retail stores or widely known to the larger population in Toronto. The market opportunity is for a bakery of Guyanese pastries that initially would be sold locally to the expat community in the initial growth phase and subsequently to the general population in the neighbourhood as product recognition improves over time.

The ideal neighbourhood location for the pastry shop must have several characteristics that are supportive of sustained high demand and sales. The location should be in a neighbourhood with a geographically small area as the local population would be concentrated near the store and result in greater foot traffic. The neighbourhood selected should be economically prosperous with a large number of local businesses and high local employment as these characteristics would be conducive to greater number of walk-in customers. Locations in a crime safe are with a large expat Guyanese community and large general population are preferred.

This market entry approach of anchoring on the expat community in a desired neighbourhood is the same strategy adopted by the highly successful Uncle Tetsu Cheesecake as it targeted the Toronto market. These Japanese Cheescakes were widely popular in Asia but unknown to the North America market. The initial storefront front was opened in the downtown Toronto core in an area with a large expat community and dense high foot traffic from local businesses and visitors. In rapid succession, more stores were opened in other neighbourhoods and there are usually long lines in front of its stores as consumers wait for a chance to purchase these cheesecakes with the Uncle Tetsu stamp.

1.3 Target Audience

The target audience for the analysis and observations in this report is a local business entrepreneur with an interest in opening a storefront in Toronto to sell pastries and other delicacies from Guyana to the retail and business segments of the population. In order to support the eventual expansion of this concept across the city, this report details the top five most attractive neighbourhoods for consideration.

More generally, this analysis is of interest to any business owner with a desire to open a retail storefront in Toronto for the sale of goods and services to the local communities. The report highlights considerations to incorporate into the location decision in a particular neighbourhood and which of the 140 neighbourhoods will be more receptive to specialty products targeted at the expat community as the market entry to the larger community in the area.

Local municipal governments and community organizations can also use this and similar analysis of neighbourhoods in the city of Toronto to identify areas where local communities are underserved and would benefit from increased focus on the delivery of tailored services.

2.0 Data Sources

The City of Toronto launched its Open Data Portal in the fall of 2018 to meet growing demand for open data. The Open Data portal provides a variety of spatial and tabular datasets pertaining to the city's population, infrastructure, and services. Data sets are available in various standard downloadable file formats such as XLS, CSV, DGN & SHP. All datasets contain basic meta information associated with them such as who created the data, date, format, projection, attributes, contact information etc.

The City of Toronto Open Data Portal was the main source of the datasets used in this analysis to assess the attractiveness of neighbourhoods for the location of a pastry bakery shop. Details of these datasets and the neighbourhood attributes extracted from these sources are listed below.

1. **Listing of the 140 Toronto neighbourhoods and boroughs - Wikipedia**

   The initial listing of the Toronto neighbourhoods containing an identification code and a descriptive name was sourced from the Wikipedia website (see reference 1) along with the borough attribute that is needed to lookup the geospatial coordinates of the area.

2. **Neighbourhood Dataset - City of Toronto Open Data Portal**

   Geographic attributes for the neighbourhoods were extracted from this dataset (see reference 2) to obtain the neighbourhood identification code, latitude, longitude and physical area size. The latitude and longitude coordinates were not populated in this dataset and were assigned using the Nominatim service with neighbourhood addresses constructed from the descriptive name and related borough.

3. **Neighbourhood Profiles Dataset - City of Toronto Open Data Portal**

The neighbourhood profiles dataset contains a wealth of information describing the local population. Demographic attributes for the neighbourhoods were extracted from this dataset (see reference 1) to obtain the number of Guyanese expats residing in the neighbourhood based on the 2016 census from Statistics Canada.

4. **Wellbeing Toronto Economics Dataset - City of Toronto Open Data Portal**

   Economic attributes for the neighbourhoods were extracted from this dataset (see reference 1) to obtain the number of businesses in the area, local employment and average house prices.

5. **Wellbeing Toronto Population Dataset - City of Toronto Open Data Portal**

   Demographic attributes for the neighbourhoods were extracted from this dataset (see reference 1) to obtain the size of the local population.

6. **Wellbeing Toronto Safety Dataset - City of Toronto Open Data Portal**

   Safety attributes for the neighbourhoods were extracted from this dataset (see reference 1) to obtain the number of major crimes and vehicle thefts.

7. **Neighbourhood Geospatial Coordinates - Nominatim**

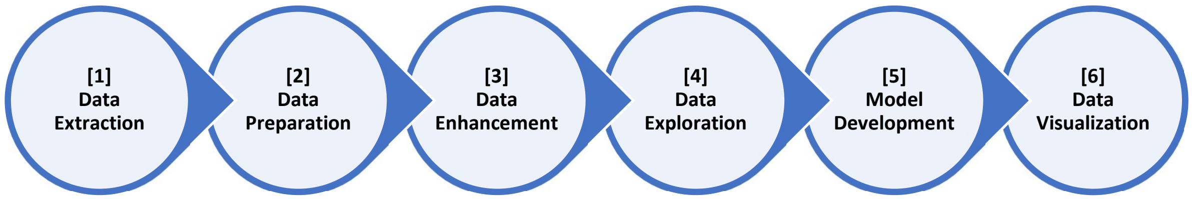
   Geospatial coordinates for the neighbourhoods defining the latitude and longitude of the center of the area were obtained using address lookups in the Nominatim service that performs a search of the OpenStreetMap dataset.

8. **Neighbourhood Venues - FourSquare**

   The FourSquare platform was used to obtain a listing of the venues located within a 1609 km (1 mile) radius of the neighbourhood center covering the venue name, latitude, longitude and venue category.

3.0 Methodology

The methodology utilized in the analysis is based a structured workflow with activities for collecting, preparing, enhancing and exploring neighbourhood data attributes followed by the development of models and presenting the results in visual charts. The analysis workflow is illustrated in the chart below.



The profile of the neighbourhoods in Toronto are first developed using a selected set of features to perform a ranking and at a second level using a listing of venues located in the neighbourhood.

3.1 Data Extraction

<u>Neighbourhood Features</u>

The input datasets that are used in the analysis of Toronto neighbourhoods are read into memory and stored without filtering or manipulation in a set of pandas dataframes.

Dataframe	Source	Features
df_gta	Wikipedia Website	Listing of neighbourhoods
df_nh	Toronto Open Data Portal	Neighnourhood dataset
df_pop	Toronto Open Data Portal	Wellbeing population dataset
df_nhp	Toronto Open Data Portal	Neighbourhood profiles dataset
df_eco	Toronto Open Data Portal	Wellbeing economics dataset
df_sft	Toronto Open Data Portal	Wellbeing safety dataset
df_geo	Nominatim	Neighbourhood latitude and longitude

<u>Venue Features</u>

The FourSquare service was used to retrieve a listing of the venues located within a radius of 1,609 km radius of the latitude and longitude coordinates for each of the neighbourhoods. A limit of 100 venues was used for each of the FourSquare lookups and the returned venue listing was stored in a dataframe df_venues without filtering or manipulation.

Dataframe	Source	Features
df_venues	FourSquare	[Venue, Venue Latitude, Venue Longitude, Venue Category]

3.2 Data Preparation

<u>Neighbourhood Features</u>

The base neighbourhood features dataframe df_gta was initialized with a listing of the neighbourdhoods extracted from the Wikipedia website and extended with the additional features of interest filtered from the dataframes representing the external datasets.

Dataframe	Source	Features
df_gta	df_nh	["Code", "Latitude", "Longitude", "Area"]
	df_geo	["Code", "Latitude", "Longitude"]
	df_pop	["Code", "Population"]
	df_nhp	["Code", “Expats”]
	df_eco	["Code", "Businesses", "LocalEmployment", "HomePrices"]
	df_sft	["Code", "MajorCrimes", "VehicleThefts"]

At the completion of the data preparation activity, the base df_gta dataframe contained all of the neighbourhood features that are required to perform the analysis and is shown in the chart below.

[22]:

Code	Neighbourhood	Borough	Latitude	Longitude	Area	Population	Expats	Businesses	LocalEmployment	HomePrices	MajorCrimes	VehicleThefts
1	West Humber-Clairville	Etobicoke	43.735990	-79.276510	5.775131e+07	33312	1895	2463	58271	317508	1119	288
2	Mount Olive-Silverstone-Jamestown	Etobicoke	43.657000	-79.428000	8.893568e+06	32954	1510	271	3244	251119	690	62
3	Thistletown-Beaumond Heights	Etobicoke	43.779280	-79.303700	6.402351e+06	10360	215	217	1311	414216	192	12
4	Rexdale-Kipling	Etobicoke	43.729200	-79.403250	4.801397e+06	10529	200	144	1178	392271	164	18
5	Elms-Old Rexdale	Etobicoke	43.756868	-79.385751	5.616463e+06	9456	145	67	903	233832	185	22
...	...	...	...	...	...	...	...	...	...	...	...	...
136	West Hill	Scarborough	43.724880	-79.253970	1.846768e+07	27392	1090	424	4500	308229	749	46
137	Woburn	Scarborough	43.787490	-79.150770	2.366499e+07	53485	1530	1073	16190	316584	808	45
138	Eglinton East	Scarborough	43.737390	-79.410930	6.179031e+06	22776	910	296	2351	274020	492	23
139	Scarborough Village	Scarborough	43.690160	-79.475000	6.040096e+06	16724	580	228	1851	356096	474	27
140	Guildwood	Scarborough	43.798130	-79.382970	7.294790e+06	9917	165	67	924	444309	113	5

<u>Venue Features</u>

The listing of venues in the neighbourhood is the second basis for profiling each of the neighbourhoods and grouping into clusters based on similarity. Each of the 343 venue categories returned by FourSquare was transformed into a binary feature using the onehot encoding method. The onehot encoding was further aggregated and scaled to produce a frequency vector for each of the neighbourhoods.

Dataframe	Source	Features
df_onehot	df_venues	Onehot encoding of Venue Category

3.3 Data Enhancement

As the geospatial coordinates of latitude and longitude for each of the neighbourhoods were not available in the input neighbourhood dataset, a lookup is done in the Nominatim service using addresses constructed from the neighbourhood name and borough.

As this service is somewhat unpredictable, the Nominatim lookup was performed several times with each new iteration focused only on the address lookups that failed in the previous iteration and repeating the lookup 300 times using the full neighbourhood address and related addresses constructed by splitting the neighbourhood on a separator. A wait time of 1 second was included to space out the lookup requests that were sent for lookup in Nominatim.

This enhancement process resulted in the retrieval of latitude and longitude coordinates for all of the neighbourhoods after several iterations with lookups in Nominatim.

3.4 Data Exploration

After all of the neighbourhood features were assembled, an exploratory data analysis activity was performed to summarize the data characteristics and to identify any anomalies and missing values.

Neighbourhood Feature Summary Statistics

	Area	Population	Expats	Businesses	Local Employment	Home Prices	Major Crimes	Vehicle Thefts
Count	140	140	140	140	140	140	140	140
Mean	8,794,110	19,511	246	536	9,409	548,193	351	30
Std	8,950,625	10,034	402	637	19,125	267,667	236	34
Min	811,304	6,577	0	47	438	204,104	81	3
25%	3,563,607	12,020	45	170	2,070	374,965	185	13
50%	6,306,846	16,750	113	346	4,053	491,210	294	21
75%	10,376,123	23,855	235	591	10,127	590,216	413	35

Neighbourhood Feature Correlation Matrix

	Area	Population	Expats	Businesses	Local Employment	Home Prices	Major Crimes	Vehicle Thefts
Area	1.0000	0.5944	0.6007	0.3919	0.2762	-0.1152	0.4520	0.7253
Population	0.5944	1.0000	0.5141	0.5311	0.4105	-0.2588	0.7227	0.4934
Expats	0.6007	0.5141	1.0000	0.1653	0.0971	-0.3944	0.4810	0.4848
Businesses	0.3919	0.5311	0.1653	1.0000	0.8803	-0.1054	0.7310	0.5477
LocalEmployment	0.2762	0.4105	0.0971	0.8803	1.0000	-0.0699	0.6199	0.3723
Home Prices	-0.1152	-0.2588	-0.3944	-0.1054	-0.0699	1.0000	-0.3477	-0.2309
Major Crimes	0.4520	0.7227	0.4810	0.7310	0.6199	-0.3477	1.0000	0.6171
Vehicle Thefts	0.7253	0.4934	0.4848	0.5477	0.3723	-0.2309	0.6171	1.0000

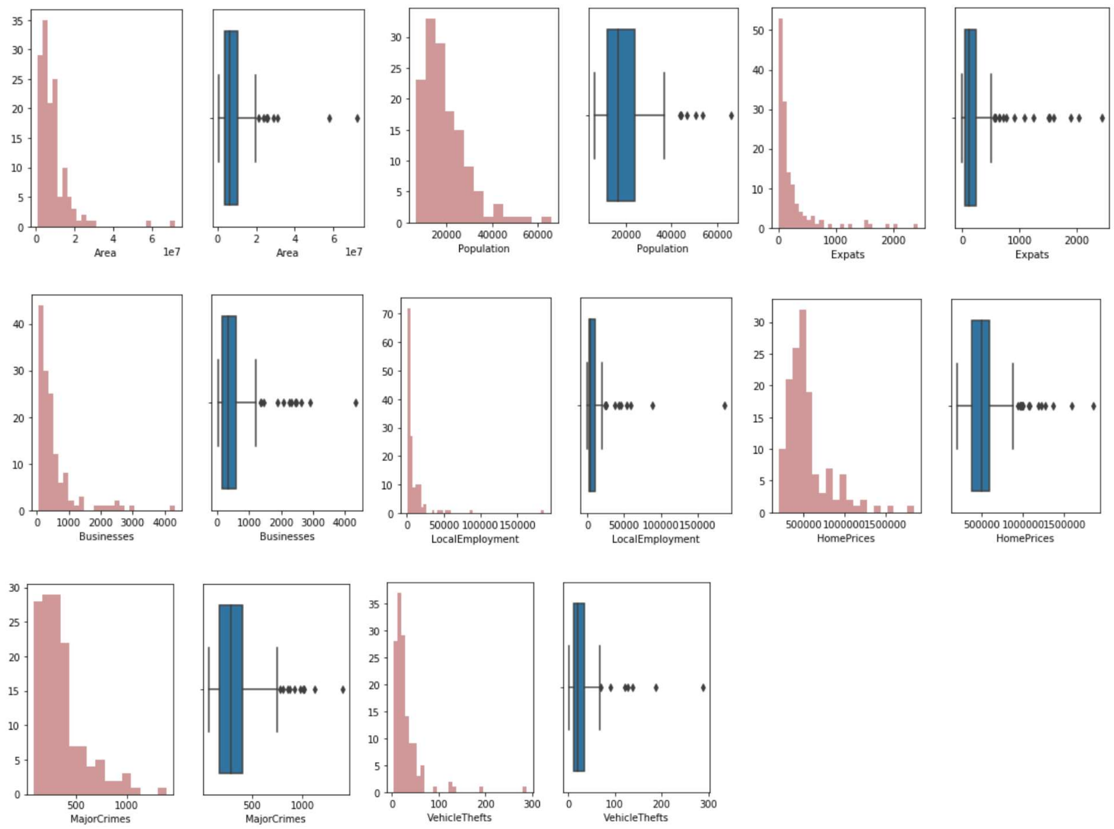
Observations:

1. There is a strong correlation between the number of neighbourhood businesses and local employment and major crimes.
2. There is a strong correlation between the neighbourhood physical area and vehicle thefts.
3. There is a strong correlation between population and vehicle thefts.

The review of summary statistics and the correlation matrix for the selected neighbourhood features do not identify any basis to change and all features are retained for the analysis.

Neighbourhood Feature Frequency Distributions

The charts below were generated using the Seaborn charting module in order to display the frequency distributions and box plots for each of the data features.



Observations:

1. These frequency distributions illustrate positive skewness with a long right tail representing feature value concentration in the lower ranges with some outliers in the outer ranges.

3.5 Model Development

Feature Ranking

A weighted score model was developed to compute an overall ranking for all of the neighbourhoods in Toronto based for the location of the new pastry shop. Each of the features was first individually ranked based on the attractiveness of the feature and a weighted score computed and ranked to produce an overall ranking for the neighbourhood.

Feature	Feature Rank Order	
Area	Ascending	Smallest to Largest
Population	Descending	Largest to Smallest
Expats	Descending	Largest to Smallest
Businesses	Descending	Largest to Smallest
LocalEmployment	Descending	Largest to Smallest
Home Prices	Descending	Largest to Smallest
Major Crimes	Ascending	Smallest to Largest
Vehicle Thefts	Ascending	Smallest to Largest

In order to compute the weighted score across all features, a feature weight of 1 was applied as all of the features were expected to contribute equally to the attractiveness of a neighbourhood for the proposed pastry shop.

$$Weighted\ Score = \sum_{i=1}^{n} w_i\ .r_i$$

Feature Clustering

The neighbourhoods in Toronto are grouped into a set of clusters based on the similarity of the associated features using the k-means model.

The unsupervised k-means clustering model partitions the input n observations into k clusters in which each observation belongs to the cluster with the nearest cluster mean which serves as a representative for each cluster. The initial assignment of the k cluster means is done randomly and means are updated after each iteration of the algorithm and repeated until there is no change in the computed means. A measure for model fit is the Sum of Squared Error statistic or distortion which is computed as the sum of the squared differences between the cluster centroid and the observations in the cluster.

The optimal model specifying the number of clusters k is determined using the elbow method which runs clustering for a range of values for k and computes an average score for all clusters using the distortion statistic. When the average distortion is plotted on a line chart against the number of clusters, the inflection point or elbow on the chart identifies the model that best fits the data.

The KElbowVisualizer function in the yellowbrick python module combines the steps to determine the optimal model, visually presents a chart of distortion against k and the best value for k.

As the neighbourhood features are stated on different scales, the data must first be standardized onto a common scale. The MinMaxScaler function from the sklearn preprocessing module scales the input data values to range between 0 and 1 using the minimum and maximum statistics.

$$f_{scaled} = \frac{f - f_{min}}{f_{max} - f_{min}}$$

The neighbourhood features are standardized using the MinMaxScaler function which results in the dataframe below.

[30]:

	Area	Population	Expats	Businesses	LocalEmployment	HomePrices	MajorCrimes	VehicleThefts
0	0.798222	0.450570	0.779835	0.564882	0.311847	0.068939	0.791159	1.000000
1	0.113303	0.444536	0.621399	0.052373	0.015131	0.028581	0.464177	0.207018
2	0.078379	0.063756	0.088477	0.039747	0.004707	0.127729	0.084604	0.031579
3	0.055936	0.066604	0.082305	0.022679	0.003990	0.114389	0.063262	0.052632
4	0.067362	0.048520	0.059671	0.004676	0.002507	0.018072	0.079268	0.066667
...	...	...	...	...	...	...	...	...
135	0.247519	0.350799	0.448560	0.088146	0.021903	0.063299	0.509146	0.150877
136	0.320378	0.790549	0.629630	0.239888	0.084938	0.068378	0.554116	0.147368
137	0.075248	0.273005	0.374486	0.058218	0.010315	0.042503	0.313262	0.070175
138	0.073301	0.171009	0.238683	0.042319	0.007619	0.092397	0.299543	0.084211
139	0.090890	0.056290	0.067901	0.004676	0.002621	0.146023	0.024390	0.007018

140 rows × 8 columns

Venue Ranking

A ranking of the top 10 venue categories located in each of the neighbourhoods was developed using the standardized onehot encoding of the venue categories returned by the FourSquare lookups. This ranking provides an alternative profile of the neighbourhoods from the distribution of venues such as coffee shops and fast-food restaurants located in the neighbourhood.

Venue Clustering

The neighbourhoods in Toronto are grouped into a set of clusters based on the similarity of the venue categories in the standardized onehot encoding using the k-means model and the elbow method to determine the optimal number of clusters.

3.6 Data Visualization

The Folium service was used to display the 140 neighbourhoods on a map of Toronto with color coded circles to highlighted the top ranked neighbourhoods and the related clusters of similar neighbourhoods.

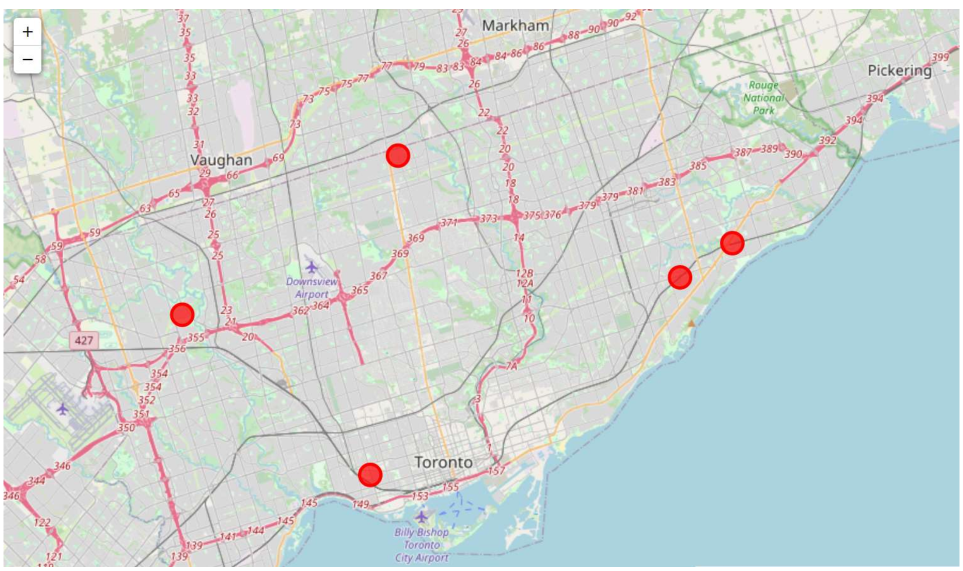
4.0 Results

Feature Ranking

Using the weighted score model and ranking, the top five most attractive neighbourhoods for the new pastry shop were identified and shown in the table below.

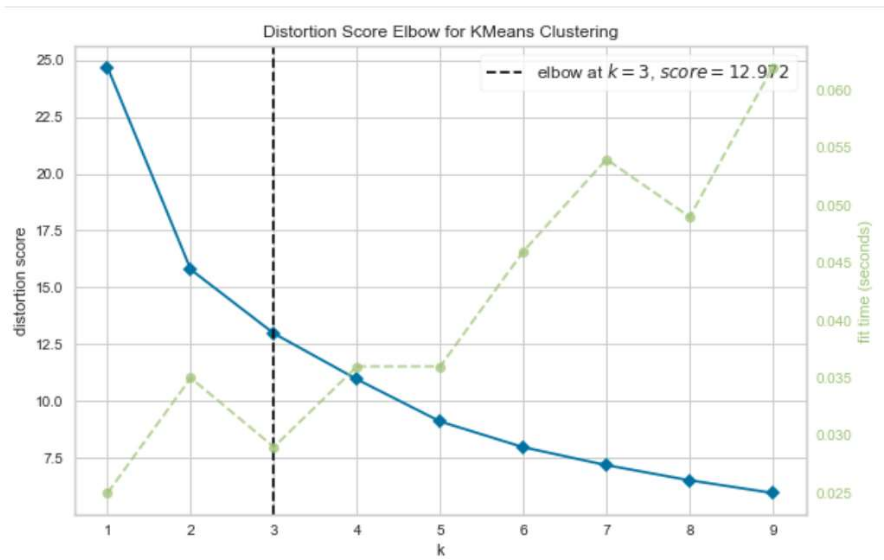
Code	Neighbourhood	Feature Score	Feature Rank
104	Mount Pleasant West	331.5	1
97	Yonge-St.Clair	369.5	2
53	Henry Farm	387.5	3
99	Mount Pleasant East	406.0	4
42	Banbury-Don Mills	414.5	5

The top ranked neighbourhoods are visually depicted on a map of the city of Toronto as red circles on the Folium chart below.



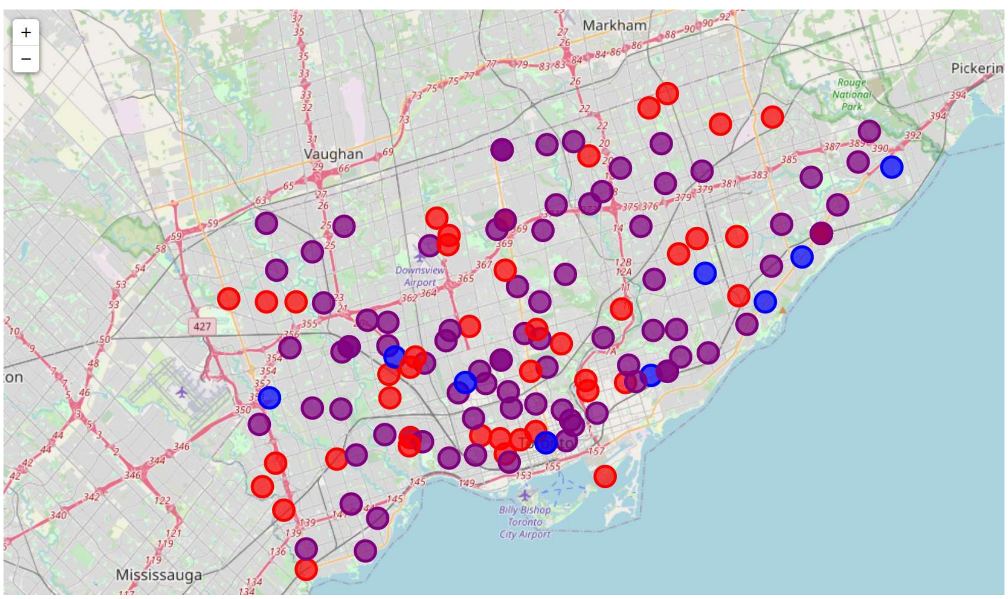
Feature Clustering

The neighbourhood features were grouped into a set of clusters using the KMeans model and the KElbowVisualizer function used to identify the optimal number of clusters. The chart below shows the relationship between the number of clusters in the tested model and the distortion statistic. The elbow was identified at 3 clusters where there was a steep increase in the processing time.




```
Optimal number of clusters: 3
```

Feature clusters of neighbourhood groups generated by the optimal KMeans model using the KElbowVisualizer function with 3 clusters is color coded and depicted in the Folium chart below.



	Cluster 0	Cluster 1	Cluster 2
Color	blue	red	Purple
Count	10	45	85

Feature cluster 2 contains the largest number of the neighbourhoods with the majority of the 5 top ranked neighbourhoods and is shown in the table below.

Code	Neighbourhood	Feature Rank	Feature Cluster
104	Mount Pleasant West	1	1
97	Yonge-St.Clair	2	2
53	Henry Farm	3	2
99	Mount Pleasant East	4	2
42	Banbury-Don Mills	5	1

The KMeans clustering results indicate that neighbourhoods 104 and 42 are more similar to each other in profile based on the distribution of feature values and neighbourhoods 97, 53 and 99 are alike in profile. This information would be of use in selecting a second and third location for the pastry shop as expansion proceeds across the city.

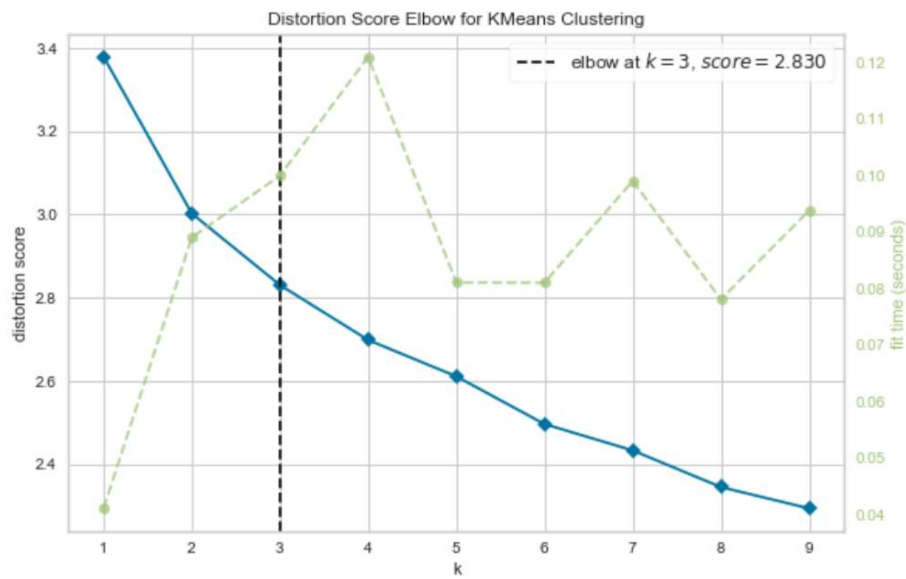
Venue Ranking

The ranking of venue categories located in each of the neighbourhoods is the basis for profiling the neigbourhood and identifying neighbourhoods that are similar in profile. The table below shows the top 10 ranked venue categories located in the feature ranked neigbourhoods. Each of these neighourhoods is characterized by the presence of a large number of restaurants and is an attractive feature for the location of the pastry shop with products that can complement a restaurant meal.

	42 - Victoria Village	53 - O'Connor-Parkview	97 - Rosedale-Moore Park	99 - Yonge and Eglinton	104 - Lawrence Park North
Top 1	Coffee Shop	Pizza Place	Italian Restaurant	Italian Restaurant	Coffee Shop
Top 2	Middle Eastern Restaurant	Park	Park	Park	Italian Restaurant
Top 3	Fast Food Restaurant	Fast Food Restaurant	CafÃ©	Coffee Shop	Sushi Restaurant
Top 4	Pizza Place	Playground	Sushi Restaurant	Restaurant	Bakery
Top 5	Sandwich Place	Pet Store	Coffee Shop	CafÃ©	Pub
Top 6	Grocery Store	Skating Rink	Pizza Place	Pizza Place	Fast Food Restaurant
Top 7	Bookstore	Bus Stop	Grocery Store	Japanese Restaurant	CafÃ©
Top 8	Shopping Mall	Breakfast Spot	Restaurant	Deli / Bodega	Bubble Tea Shop
Top 9	Gas Station	Brewery	Liquor Store	Sushi Restaurant	Golf Course
Top 10	Clothing Store	Burger Joint	Convenience Store	Bakery	Japanese Restaurant

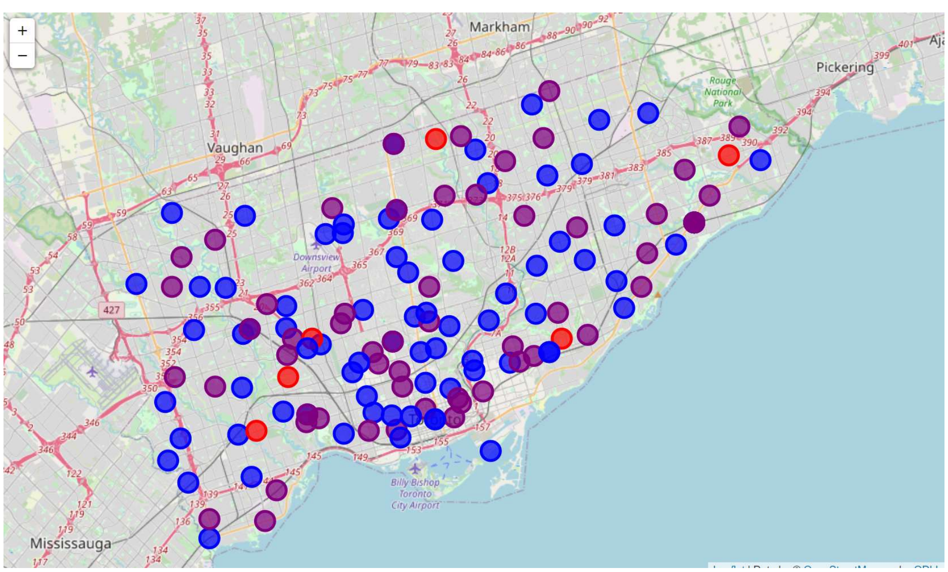
<u>Venue Clustering</u>

The neighbourhood venue categories were grouped into a set of clusters using the KMeans model and the KElbowVisualizer function used to identify the optimal number of clusters. The chart below shows the relationship between the number of clusters in the tested model and the distortion statistic. The inflection point was identified at 3 clusters where there was a steep increase in the processing time.




```
Optimal number of clusters: 3
```

The venue clusters of neighbourhood groups generated by the optimal KMeans model using the KElbowVisualizer function with 3 clusters is color coded and depicted in the Folium chart below.



	Cluster 0	Cluster 1	Cluster 2
Color	red	blue	Purple
Count	6	76	58

Venue cluster 1 contains the largest number of the neighbourhoods with the majority of the 5 top featured ranked neighbourhoods in cluster 2 and is shown in the table below.

The KMeans clustering results indicate that neighbourhoods 104, 97 and 99 are more similar to each other based on the location of venues and neighbourhoods 53 and 42 are alike in profile. This information would be of use in selecting a second and third location for the pastry shop as expansion proceeds across the city.

Code	Neighbourhood	Feature Rank	Venue Cluster
104	Mount Pleasant West	1	2
97	Yonge-St.Clair	2	2
53	Henry Farm	3	1
99	Mount Pleasant East	4	2
42	Banbury-Don Mills	5	1

5.0 Discussion

The objective of this analysis is to determine the best location for a Guyanese pastry shop given its market strategy of entry sales to the local expat community followed by growth sales to the general population and local businesses.

Profiling the neighbourhoods in Toronto based on 343 different venue categories returned by FourSquare provides an indirect means of making the location decision as the existence of ATM machines, bus stops and Parks in the area have little predictive power on the sales potential of a pastry shop.

The neighbourhood profiles developed using selected features are the most useful in predicting sales potential of an area is the basis for making the recommendation. The feature ranking and feature clustering results provide enough insights into the neighbourhood profiles across the city for decision making.

Based on the results produced by the feature ranking weighted score model, the most attractive location in the city of Toronto for the new pastry shop is the neighbourhood of **Mount Pleasant West** with an identification code of 104.

A second location to consider for the pastry shop is the neighbourhood of **Banbury-Don Mills** with an identification code of 42 as this neighbourhood is most similar to the Mount Pleasant West neighbourhood as highlighted in the feature cluster analysis of all of the neighbourhoods.

6.0 Conclusion

This detailed analysis of the neighbourhood profiles in Toronto has demonstrated the usage of data science techniques to extract, prepare, enhance and assemble data attributes from external sources for examination and the application of statistical techniques to highlight the relationships between these attributes. Unsupervised machine learning models using the KMeans algorithm were used to identify similarities between the neighbourhoods based on selected features and venues located in the area. The elbow method using the distortion statistic was used to statistically determine the optimal model to cluster the data attributes and to identify the ideal neighbourhood for the location of a pastry shop with a targeted focus for product acceptance and a strategy for sales growth.

7.0 References

1. Wikipedia Contributors. (2019, December 5). List of neighbourhoods in Toronto. Retrieved December 13, 2019, from Wikipedia website: https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Toronto
2. City of Toronto Open Data Portal. Retrieved February 7, 2021, from https://open.toronto.ca/
3. Neighbourhoods Open Data Dataset. (n.d.). Retrieved February 7, 2021, from https://open.toronto.ca/dataset/neighbourhoods/
4. Neighbourhood Profiles Open Data Dataset. Retrieved from https://open.toronto.ca/dataset/neighbourhood-profiles/
5. Wellbeing Safety Open Data Dataset. Retrieved February 7, 2021, from https://open.toronto.ca/dataset/wellbeing-toronto-safety/
6. Guyanese people. (2021, January 24). Retrieved February 8, 2021, from Wikipedia website: https://en.wikipedia.org/wiki/Guyanese_people
7. Government of Canada, Statistics Canada. (2016). Census Profile, 2016 Census - Canada [Country] and Canada [Country]. Retrieved February 8, 2021, from Statcan.gc.ca website: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=PR&Code1=01&Geo2=PR&Code2=01&Data=Count&SearchText=Canada&SearchType=Begins&SearchPR=01&B1=Ethnic%
8. Toronto. (2021, February 5). Retrieved February 7, 2021, from Wikipedia website: https://en.wikipedia.org/wiki/Toronto#Demographics
9. Demographics of Toronto - Wikipedia. (n.d.). Retrieved February 7, 2021, from en.m.wikipedia.org website: https://en.m.wikipedia.org/wiki/Demographics_of_Toronto#Ethnic_diversity

A.0 Appendix

A.1 Neighbourhoods

Code	Neighbourhood	Borough
1	West Humber-Clairville	Etobicoke
2	Mount Olive-Silverstone-Jamestown	Etobicoke
3	Thistletown-Beaumond Heights	Etobicoke
4	Rexdale-Kipling	Etobicoke
5	Elms-Old Rexdale	Etobicoke
6	Kingsview Village-The Westway	Etobicoke
7	Willowridge-Martingrove-Richview	Etobicoke
8	Humber Heights-Westmount	Etobicoke
9	Edenbridge-Humber Valley	Etobicoke
10	Princess-Rosethorn	Etobicoke
11	Eringate-Centennial-West Deane	Etobicoke
12	Markland Wood	Etobicoke
13	Etobicoke West Mall	Etobicoke
14	Islington-City Centre West	Etobicoke
15	Kingsway South	Etobicoke
16	Stonegate-Queensway	Etobicoke
17	Mimico	Etobicoke
18	New Toronto	Etobicoke
19	Long Branch	Etobicoke
20	Alderwood	Etobicoke
21	Humber Summit	North York
22	Humbermede	North York
23	Pelmo Park-Humberlea	North York
24	Black Creek	North York
25	Glenfield-Jane Heights	North York
26	Downsview-Roding-CFB	North York
27	York University Heights	North York
28	Rustic	North York
29	Maple Leaf	North York
30	Brookhaven-Amesbury	North York
31	Yorkdale-Glen Park	North York
32	Englemount-Lawrence	North York
33	Clanton Park	North York
34	Bathurst Manor	North York
35	Westminster-Branson	North York
36	Newtonbrook West	North York
37	Willowdale West	North York
38	Lansing-Westgate	North York
39	Bedford Park-Nortown	North York
40	St. Andrew-Windfields	North York
41	Bridle Path-Sunnybrook-York Mills	North York

Code	Neighbourhood	Borough
42	Banbury-Don Mills	North York
43	Victoria Village	North York
44	Flemingdon Park	North York
45	Parkwoods-Donalda	North York
46	Pleasant View	North York
47	Don Valley Village	North York
48	Hillcrest Village	North York
49	Bayview Woods-Steeles	North York
50	Newtonbrook East	North York
51	Willowdale East	North York
52	Bayview Village	North York
53	Henry Farm	North York
54	O'Connor-Parkview	East York
55	Thorncliffe Park	East York
56	Leaside-Bennington	East York
57	Broadview North	East York
58	Old East York	East York
59	Danforth - East York	East York
60	Woodbine-Lumsden	East York
61	Crescent Town	East York
62	East End-Danforth	Old City of Toronto
63	The Beaches	Old City of Toronto
64	Woodbine Corridor	Old City of Toronto
65	Greenwood-Coxwell	Old City of Toronto
66	Danforth Village - Toronto	Old City of Toronto
67	Playter Estates-Danforth	Old City of Toronto
68	North Riverdale	Old City of Toronto
69	Blake-Jones	Old City of Toronto
70	South Riverdale	Old City of Toronto
71	Cabbagetown-South St. James Town	Old City of Toronto
72	Regent Park	Old City of Toronto
73	Moss Park	Old City of Toronto
74	North St. James Town	Old City of Toronto
75	Church-Yonge Corridor	Old City of Toronto
76	Bay Street Corridor	Old City of Toronto
77	Waterfront Communities-The Island	Old City of Toronto
78	Kensington-Chinatown	Old City of Toronto
79	University	Old City of Toronto
80	Palmerston-Little Italy	Old City of Toronto
81	Trinity-Bellwoods	Old City of Toronto
82	Niagara	Old City of Toronto
83	Dufferin Grove	Old City of Toronto
84	Little Portugal	Old City of Toronto
85	South Parkdale	Old City of Toronto
86	Roncesvalles	Old City of Toronto
87	High Park-Swansea	Old City of Toronto
88	High Park North	Old City of Toronto

Code	Neighbourhood	Borough
89	Runnymede-Bloor West Village	Old City of Toronto
90	Junction Area	Old City of Toronto
91	Weston-Pellam Park	Old City of Toronto
92	Corso Italia-Davenport	Old City of Toronto
93	Dovercourt-Wallace Emerson-Junction	Old City of Toronto
94	Wychwood	Old City of Toronto
95	Annex	Old City of Toronto
96	Casa Loma	Old City of Toronto
97	Yonge-St.Clair	Old City of Toronto
98	Rosedale-Moore Park	Old City of Toronto
99	Mount Pleasant East	Old City of Toronto
100	Yonge and Eglinton	Old City of Toronto
101	Forest Hill South	Old City of Toronto
102	Forest Hill North	Old City of Toronto
103	Lawrence Park South	Old City of Toronto
104	Mount Pleasant West	Old City of Toronto
105	Lawrence Park North	Old City of Toronto
106	Humewood-Cedarvale	York
107	Oakwood Village	York
108	Briar Hill-Belgravia	York
109	Caledonia-Fairbank	York
110	Keelesdale-Eglinton West	York
111	Rockcliffe-Smythe	York
112	Beechborough-Greenbrook	York
113	Weston	York
114	Lambton Baby Point	York
115	Mount Dennis	York
116	Steeles	Scarborough
117	L'Amoreaux	Scarborough
118	Tam O'Shanter-Sullivan	Scarborough
119	Wexford-Maryvale	Scarborough
120	Clairlea-Birchmount	Scarborough
121	Oakridge	Scarborough
122	Birchcliffe-Cliffside	Scarborough
123	Cliffcrest	Scarborough
124	Kennedy Park	Scarborough
125	Ionview	Scarborough
126	Dorset Park	Scarborough
127	Bendale	Scarborough
128	Agincourt South-Malvern West	Scarborough
129	Agincourt North	Scarborough
130	Milliken	Scarborough
131	Rouge	Scarborough
132	Malvern	Scarborough
133	Centennial Scarborough	Scarborough
134	Highland Creek	Scarborough
135	Morningside	Scarborough

Code	Neighbourhood	Borough
136	West Hill	Scarborough
137	Woburn	Scarborough
138	Eglinton East	Scarborough
139	Scarborough Village	Scarborough
140	Guildwood	Scarborough