



TORONTO NEIGHBORHOODS REPORT

OPTIMAL LOCATION FOR A SPECIALTY PASTRY SHOP

RUDOLPH PERSAUD

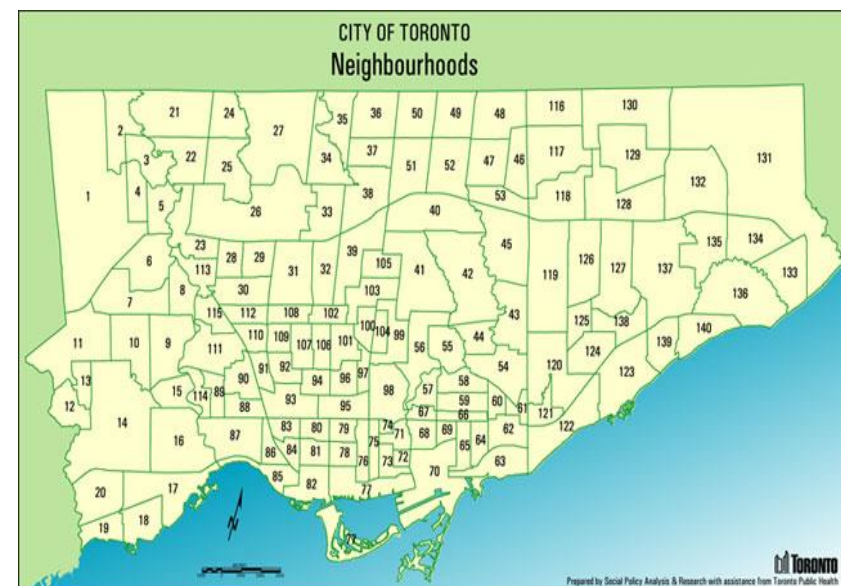
FEB 2021





BACKGROUND

- With a population of 2.7 million, Toronto is one of the largest and the most populous cities in Canada and the fourth largest city in North America.
- There are 140 neighbourhoods officially recognized by the City of Toronto that were developed to help government and community organizations with their local planning. The boundaries of these neighbourhoods do not change over time and are reflected in the many datasets published by the city.
- Toronto 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.
- More than 50 percent of residents in the city belong to a visible minority population group.
- 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.
- Based on the 2016 Canadian population census from Statistics Canada, there were over 87k immigrants from Guyana with the majority residing in Toronto.





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.
- 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 ideal neighborhood location for the pastry shop should have several characteristics to be attractive:
 - geographically small area that would support heavy foot traffic
 - economically prosperous with a large number of local businesses, high local employment and high average house prices
 - large expat Guyanese community and large general population
 - low major crime rates and vehicle thefts
- The analysis and observations in this report are of use to the following entities:
 - small business owners who are interested in opening a retail storefront for the sale of specialty goods and services to the local expat community
 - municipal governments and community organizations to identify areas where local communities are underserved and would benefit from the delivery of services





DATA SOURCES

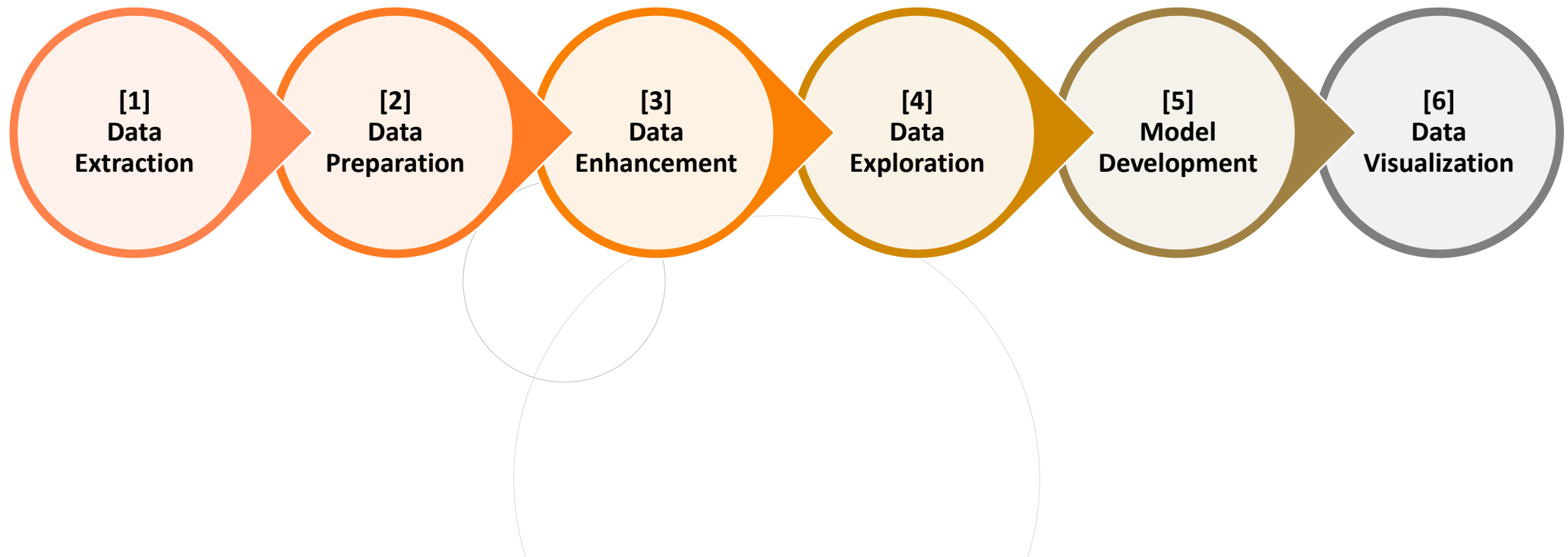
Dataset Description		Dataset Source
1	Listing of the 140 Toronto neighbourhoods <ul style="list-style-type: none">neighbourhood identification code, name, borough	Wikipedia website - List of neighbourhoods in Toronto
2	Neighbourhood Open Data Dataset <ul style="list-style-type: none">geographic attributes - Area	City of Toronto Open Data Portal
3	Neighbourhood Profiles Open Data Dataset <ul style="list-style-type: none">demographic attributes – Guyanese expats	City of Toronto Open Data Portal
4	Wellbeing Toronto Economics Open Data Dataset <ul style="list-style-type: none">economic attributes – businesses, local employment, house prices	City of Toronto Open Data Portal
5	Wellbeing Toronto Population Open Data Dataset <ul style="list-style-type: none">demographic attributes – local population size	City of Toronto Open Data Portal
6	Wellbeing Toronto Safety Open Data Dataset <ul style="list-style-type: none">Safety attributes – major crimes, vehicle thefts	City of Toronto Open Data Portal
7	Neighbourhood Geospatial Coordinates <ul style="list-style-type: none">geographic attributes – latitude and longitude	Nominatim - Tool to search of OpenStreetMap dataset
8	Neighbourhood Venues <ul style="list-style-type: none">Local Venues – name, latitude, longitude, category	FourSquare – Location platform



METHODOLOGY

The neighbourhood profile analysis is based a structured workflow

- activities for collecting, preparing, enhancing and exploring neighborhood data attributes
- development of models and presenting the results in visual charts.





METHODOLOGY

[1]
Data
Extraction

Neighbourhood Features

Dataframe	Source	Features
df_gta	Wikipedia Website	Listing of neighbourhoods
df_nh	Toronto Open Data Portal	Neighbourhood 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

Neighbourhood Venues

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



METHODOLOGY

[2]
Data
Preparation

Neighbourhood Features

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"]

Neighbourhood Venues

Dataframe	Source	Features
df_onehot	df_venues	Onehot encoding of Venue Category



METHODOLOGY

[3]
Data
Enhancement

Neighbourhood Features

- The latitude and longitude coordinates for each of the neighbourhood was not populated in the neighbourhoods dataset from the Toronto Open Data Portal
- To fill in the missing geospatial coordinates, an address lookup with the Nominatim service for the latitude and longitude was performed
- Due to unpredictable performance, the Nominatim lookup was executed several times with each new iteration focused only on the address lookups that failed in the previous iteration
- Each of the failed lookups was repeated up to 300 times using the full neighbourhood address and related addresses constructed by splitting the neighbourhood on a separator



METHODOLOGY

[4]
Data
Exploration

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



METHODOLOGY

[5]
Model
Development

Feature Ranking

- Scale the feature values to a common scale between 0 and 1 using the MinMaxScaler function from the sklearn preprocessing module

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

- Weighted score model with a weight of 1 to combine individual feature rankings to compute an overall rank for the neighbourhoods

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

Feature Clustering

- Unsupervised k-means clustering model to group neighbourhoods into similar clusters
- KElbowVisualizer function in the Yellowbrick python module to determine the optimal model for clustering using the elbow method to select the number of clusters to use

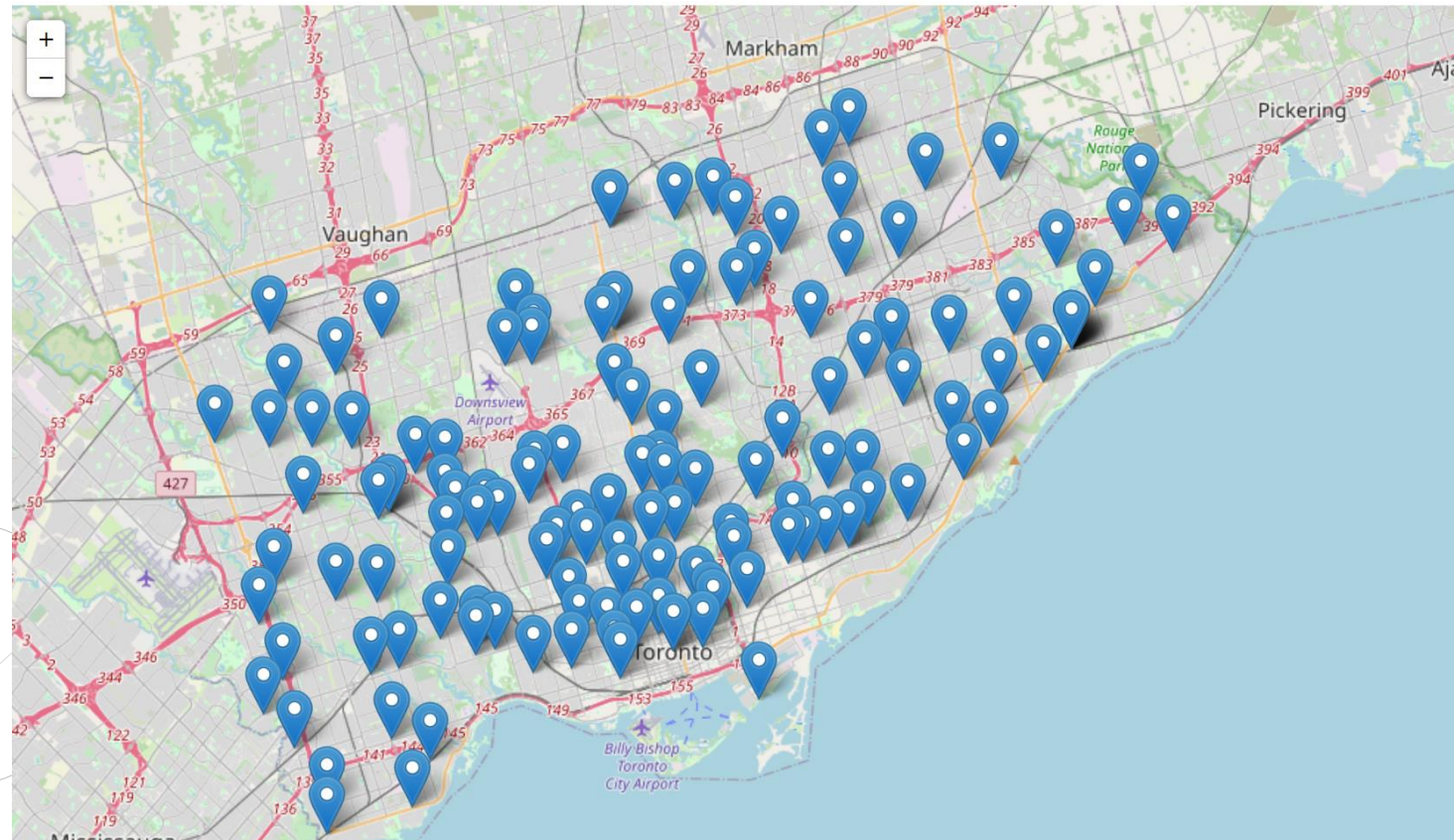


METHODOLOGY

[6]
Data
Visualization

Neighbourhoods of Toronto

- The Folium service was used to display the 140 neighbourhoods on a map of Toronto using the latitude and longitude coordinates for each area

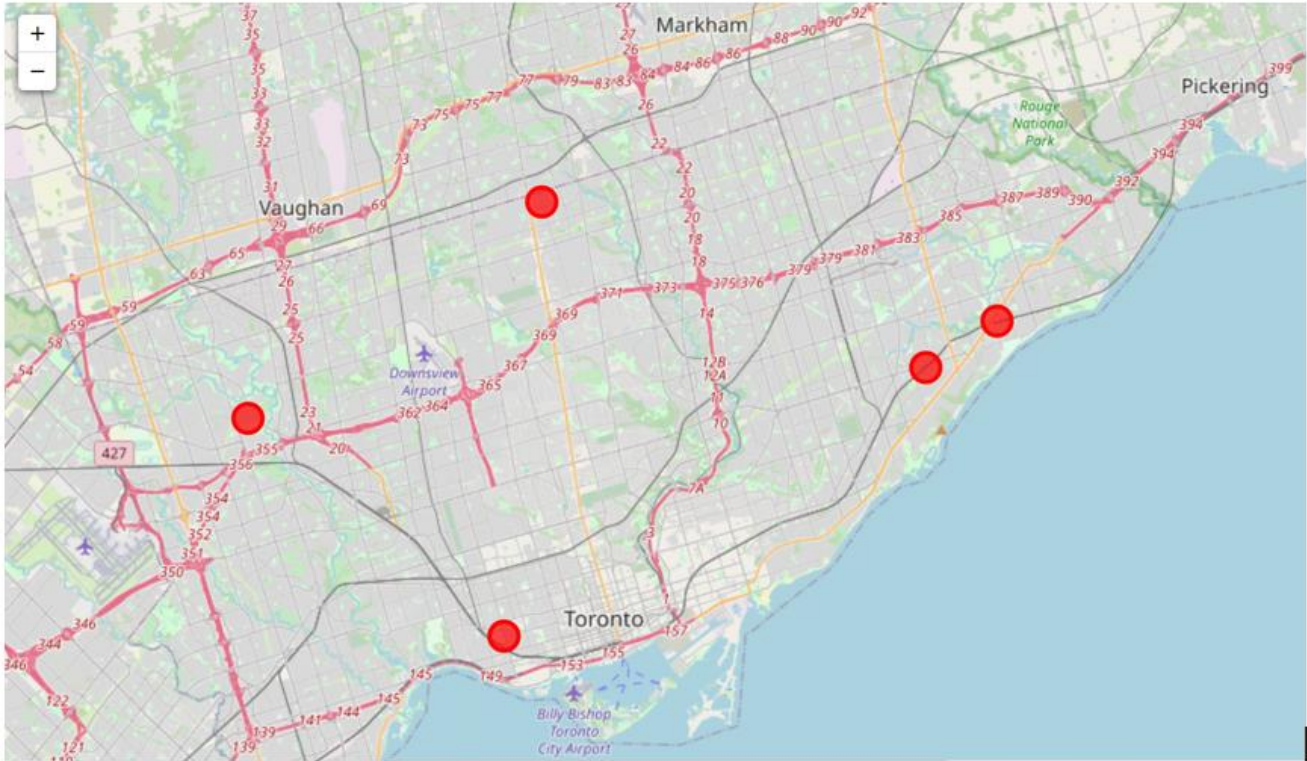




RESULTS

Feature Ranking

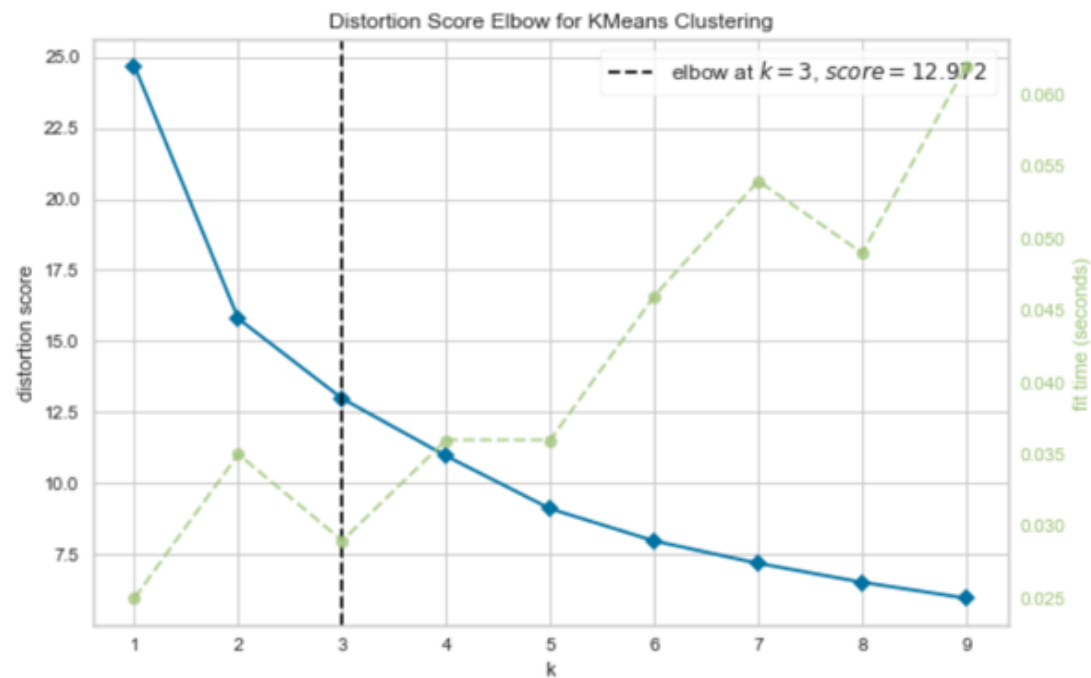
Code	Neighbourhood	Feature Score	Feature Rank
104	Mount Pleasant West	331.5	1
97	Yonge-St.Clair	369.5	2
53	Henry Farm	387.5	3
99	Mount Pleasant East	406.0	4
42	Banbury-Don Mills	414.5	5



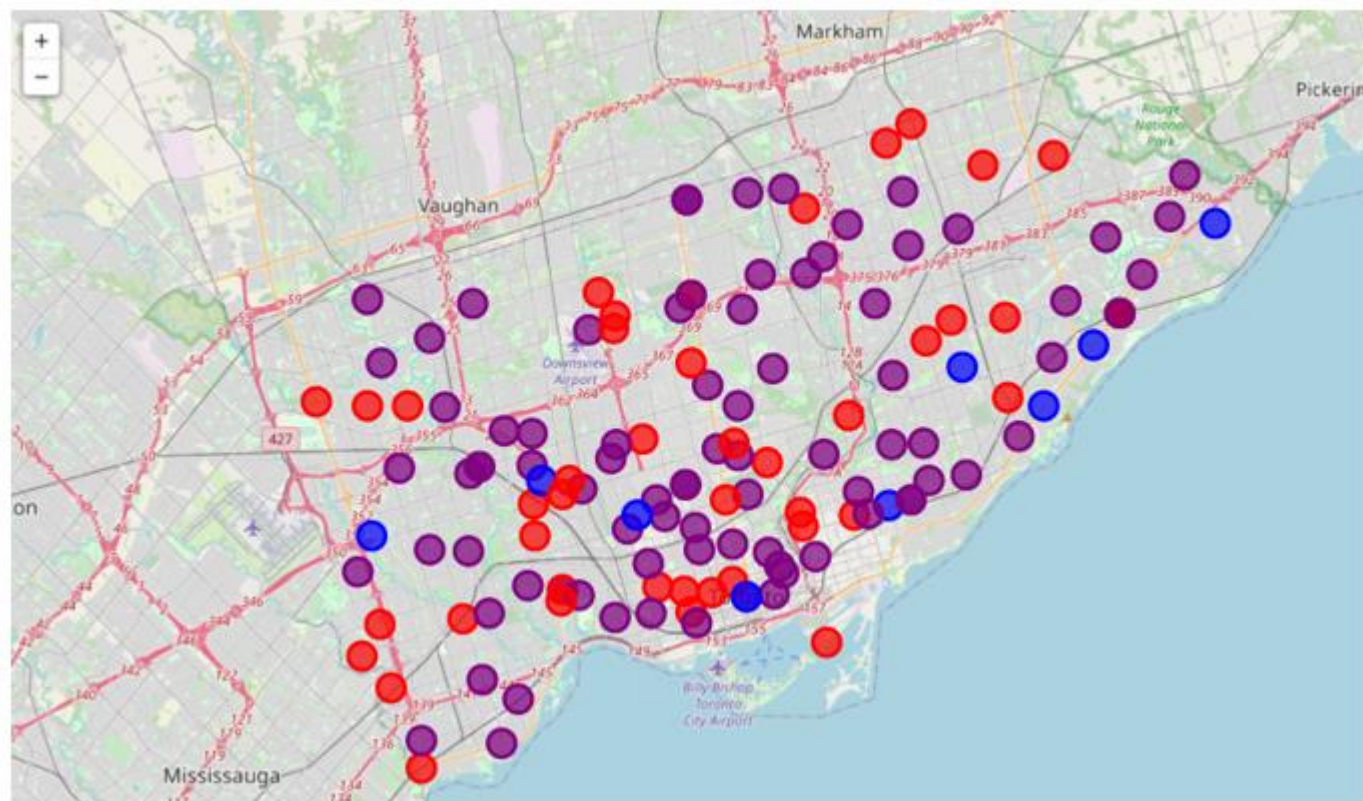


RESULTS

Feature Clustering



Optimal number of clusters: 3



CONCLUSION

- 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.

