The Battle of Neighborhoods
Opening a Pizza Shop in Toronto

Capstone Project – Presentation - Week 2

Mustafa Ayub

Jan 31st, 2021

Sources:Flytap.com

Source:everydaydishes.com

Agenda

- Introduction
  - Business Problem

- Data Requirement
  - Data Source & Pre-processing
  - Data Exploration

- Methodology
  - Predictive Modeling & Clustering

- Results & Discussions
  - Cluster Analysis & Discussions

- Conclusions

Introduction

▶Business Problem

Business Problem

Problem Statement:

▶ A new investor is willing to invest in a food business that is liked by majority of Canadians. Following an extensive research, he has landed on decision to open up a pizza shop; however, he is not sure which neighborhood to pick. He is asking for advice from a Data Scientist to conduct analysis and provide a recommendation as to which neighborhood is the best one for a pizza shop

Goal Statement:

▶ Though this capstone project, I will leverage my data science knowledge gained throughout the course along with necessary data to identify where would be an ideal location for a pizza shop.

Data Requirement

- Data Source & Pre-processing
- Data Exploration

Data Source & Pre-processing

The following data were collected to use during the analysis through this project:

Data	Data Source	Data pre-processing
List of neighborhood by postal code	Wikipedia: *Link*	✓ Scraped using BeautifulSoup ✓ Remove the "\n" from the columns ✓ Exclude unassigned boroughs
Geospatial data	CSV file: *Link*	✓ Ensure no empty cells ✓ Merged data postal code data
List of Toronto venues	Foursquare API: *Link*	✓ Used one-hot-encoding to convert the categorical data into numerical data ✓ Grouped the data by neighborhood

Data Exploration – Postal Code Data

1 *Import Postal Code data*

	Postal Code	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

2 *Import Geospatial data*

	A	B	C	D
1	Postal Cod	Latitude	Longitude	
2	M1B	43.80669	-79.1944	
3	M1C	43.78454	-79.1605	
4	M1E	43.76357	-79.1887	
5	M1G	43.77099	-79.2169	
6	M1H	43.77314	-79.2395	
7	M1J	43.74473	-79.2395	
8	M1K	43.72793	-79.262	
9	M1L	43.71111	-79.2846	
10	M1M	43.71632	-79.2395	

3 *Merge 1 & 2 to obtain data frame*

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

*10 Boroughs & 103 Neighborhoods!*

Data Exploration – Foursquare API

1 *Import data from Foursquare API & group by neighborhood*

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
3	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
4	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop

271 unique venues!

2 *Create a visual of the data to see the top 10 venues in Toronto*

List of Top 10 Venues
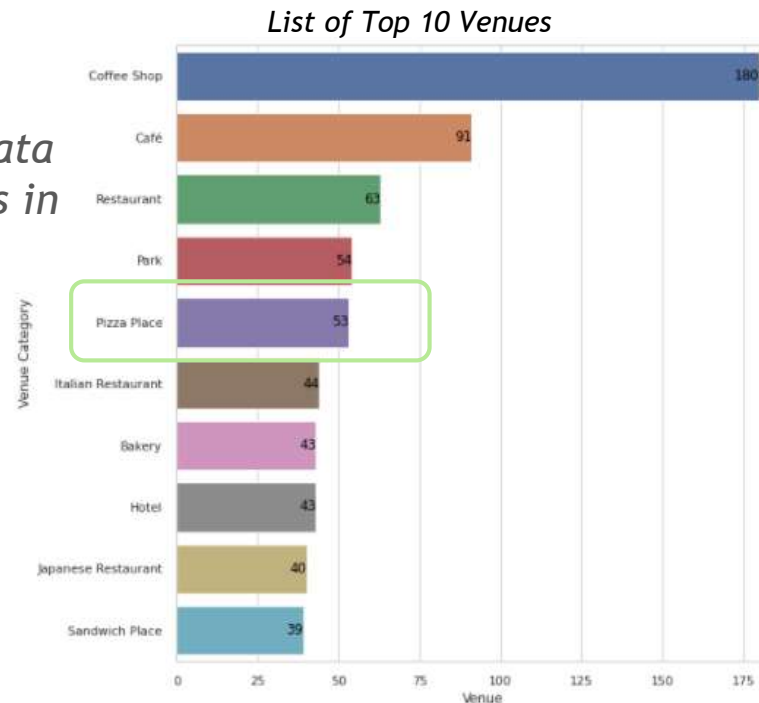

Pizza Place is top 5th venue in the dataset!

Methodology

- Predictive Modeling & Clustering

Predictive Modeling & Clustering

▶ Data preparation

    ▶ Used one-hot-encoding to convert the categorical data into numerical data

    ▶ Grouped the data by neighborhood to obtain the average frequency of the venue occurrence

▶ Clustering

    ▶ *Leveraging KElbow method, identified that the data can be divided into 3 clusters; k = 3*

    ▶ *Generated clusters using KMean Clustering technique*
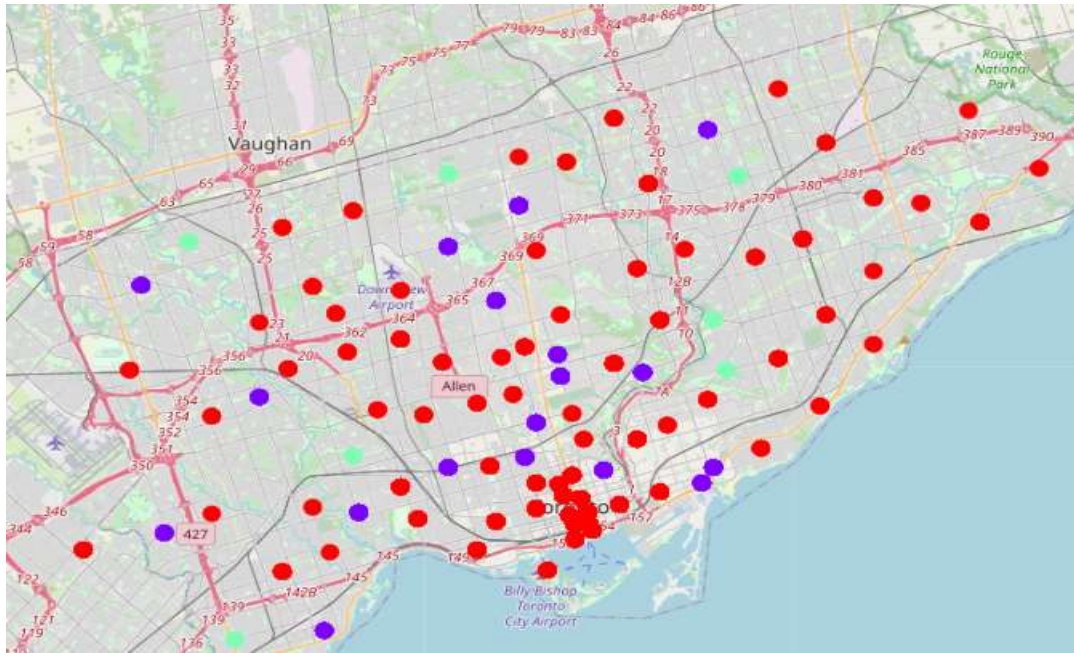
Legend:
- *Cluster 0 (red)*
- *Cluster 1 (dark blue)*
- *Cluster 2 (light blue)*
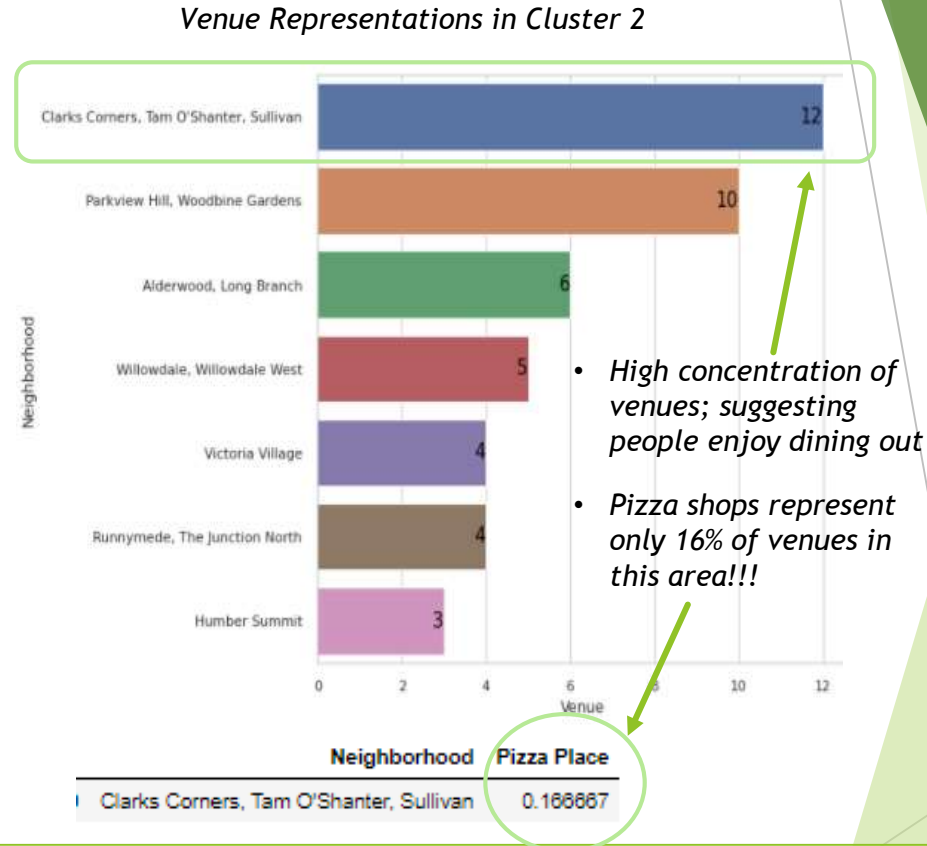
Please refer to appendix B for details

Results & Discussions

▶ Cluster Analysis & Discussions

Cluster Analysis & Discussions

- Cluster 0 (red):
  - 18 existing Pizza Shop and high level competition
  - Close to downtown; hence high rent & other cost

- Cluster 1 (Dark Blue):
  - 25 existing Pizza Shop and highly competitive

- Cluster 2 (Light Blue):
  - 10 existing Pizza Shop
  - Easier to enter the market with an authentic pizza shop

Venue Representations in Cluster 2

- High concentration of venues; suggesting people enjoy dining out

- Pizza shops represent only 16% of venues in this area!!!

Neighborhood	Pizza Place
Clarks Corners, Tam O'Shanter, Sullivan	0.166667

Key Takeaways:
This analysis suggests that areas surrounding "Clarks Corners, Tam O'Shanter, Sullivan" neighborhood would be an ideal place to open an authentic pizza shop; people enjoy dining out & very low concentration of existing pizza shops!

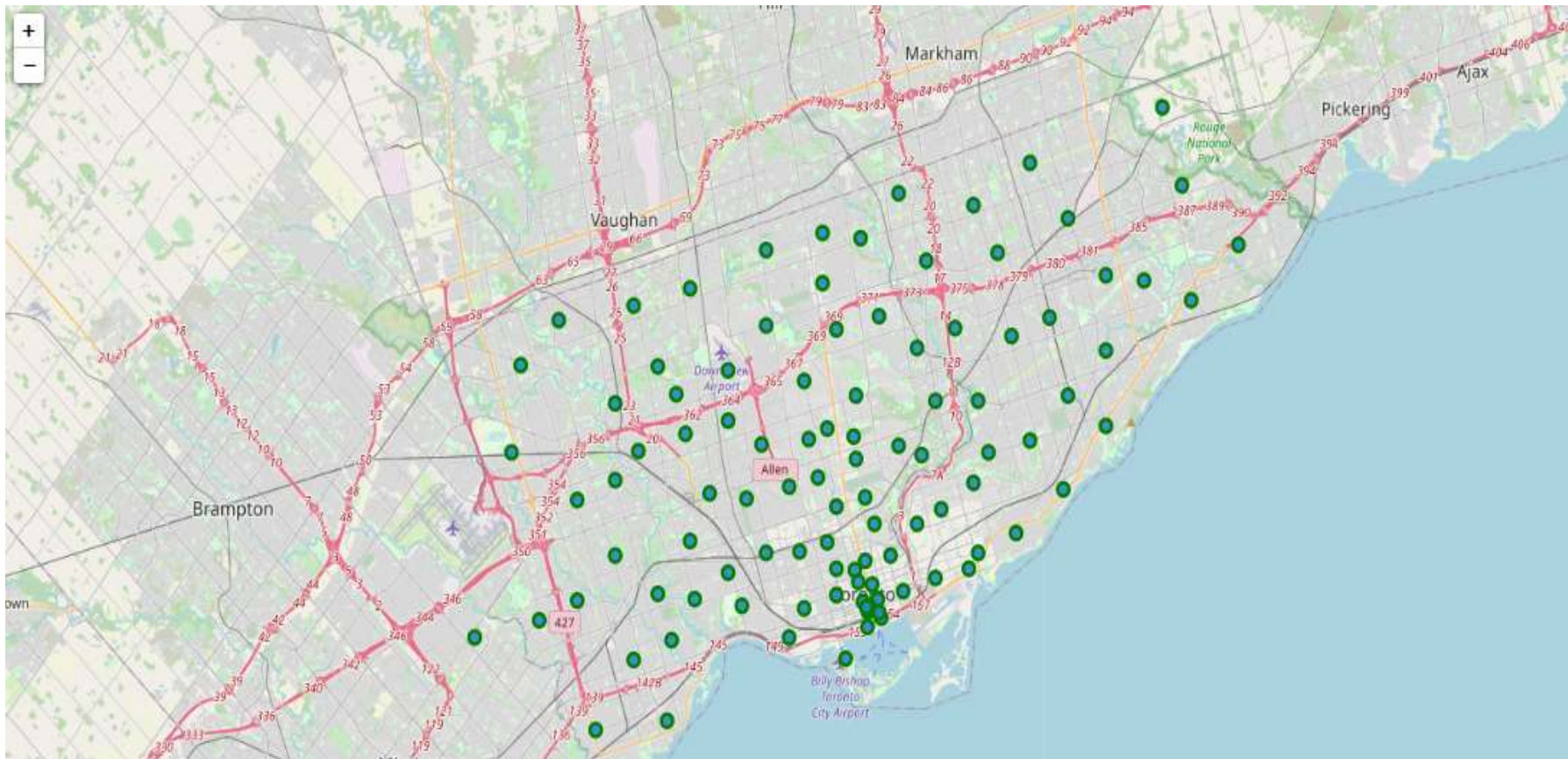*Please refer to appendix C for details*

Conclusion

▶ Conclusion

Conclusions

▶ This Capstone project's goal was to define a hypothetical problem and solve it using the knowledge gained during the course

▶ The hypothetical problem chosen was to look at the neighborhoods of Toronto and find an ideal place to open up a pizza shop

▶ The following process was used to complete the project:

  ▶ Identified sources of various data (postal code, geospatial data, list of venues)

  ▶ Pre-processed data to build the required data frame for analysis

  ▶ Identified the number of optimal clusters using KElbow

  ▶ Leveraged KMean clustering technique and divided the neighborhoods into 3 clusters

  ▶ Conducted analysis to determine which cluster and subsequently which neighborhood would be ideal for a new authentic pizza shop
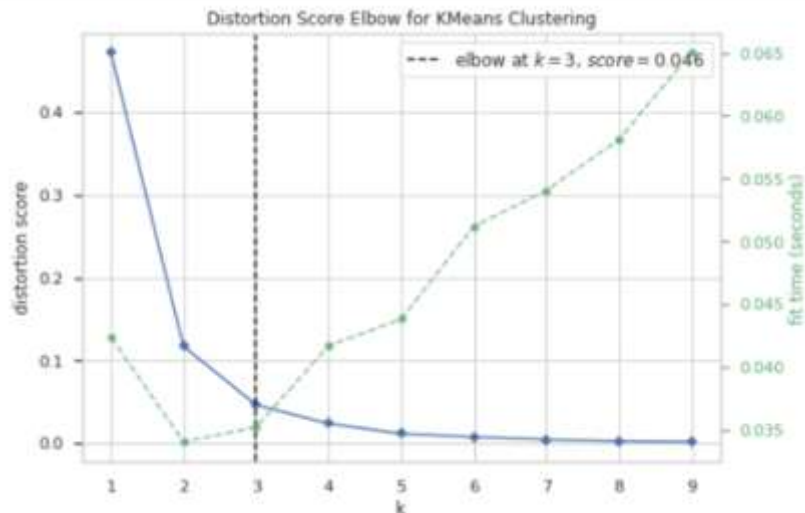
Appendix

Appendix A – Visual representation of Neighborhoods

Appendix B - Clustering

1 *Leveraging KElbow method; k = 3*

2 *Generate clusters using KMean Clustering technique*

```
: kclusters = 3

  to_grouped_clusters = pizza_df.drop('Neighborhood', 1)

  # run k-means clustering
  kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(to_grouped_clusters)

  # check cluster labels generated for each row in the dataframe
  kmeans.labels_[0:10]

63]: array([0, 2, 1, 0, 1, 0, 0, 0, 1, 0], dtype=int32)
```

4 *Visualize how the neighborhoods were clustered*

3 *Merge data frame to add cluster label for each neighborhood*

	Neighborhood	Pizza Place	Cluster Labels
0	Agincourt	0.000000	0
1	Alderwood, Long Branch	0.333333	2
2	Bathurst Manor, Wilson Heights, Downsview North	0.045455	1
3	Bayview Village	0.000000	0
4	Bedford Park, Lawrence Manor East	0.043478	1
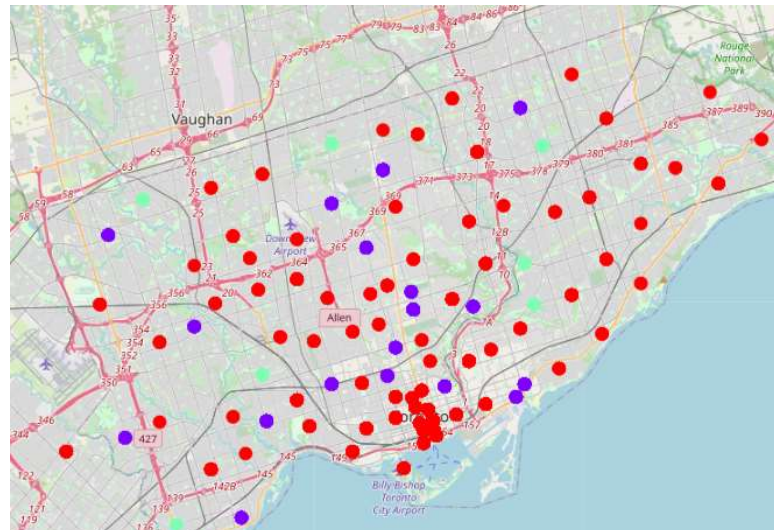
Appendix C - Cluster Analysis & Discussions

Cluster 0 (Red)

	Neighborhood	Pizza Place	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Agincourt	0.0	0	43.79420	-79.262029	Panagio's Breakfast & Lunch	43.792370	-79.260203	Breakfast Spot
59	Parkdale, Roncesvalles	0.0	0	43.64896	-79.456325	Cider House	43.650688	-79.450685	Restaurant
59	Parkdale, Roncesvalles	0.0	0	43.64896	-79.456325	Domani Restaurant & Wine Bar	43.649235	-79.450229	Italian Restaurant
59	Parkdale, Roncesvalles	0.0	0	43.64896	-79.456325	Inter Steer	43.649796	-79.450310	Eastern European Restaurant
59	Parkdale, Roncesvalles	0.0	0	43.64896	-79.456325	Revue Cinema	43.651112	-79.450961	Movie Theater

- *18 existing Pizza Shop*
- *Close to downtown*
- *High rent & cost*
- *Highly competitive*

Cluster 1 (Dark Blue)

	Neighborhood	Pizza Place	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
19	Davisville	0.117647	1	43.704324	-79.388790	Provocative Pizza Series	43.708293	-79.389546	Pizza Place
19	Davisville	0.117647	1	43.704324	-79.388790	Pizzaiolo	43.707301	-79.389830	Pizza Place
19	Davisville	0.117647	1	43.704324	-79.388790	Shoppers Drug Mart	43.707806	-79.389893	Pharmacy
8	Business reply mail Processing Centre, South C...	0.058824	1	43.662744	-79.321558	East End Garden Centre & Hardware	43.664564	-79.324471	Garden Center
19	Davisville	0.117647	1	43.704324	-79.388790	Bolan Thai Cuisine	43.706833	-79.389200	Thai Restaurant

- *25 Pizza Shop*
- *Competition*

Cluster 2 (Light Blue)

	Neighborhood	Pizza Place	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
38	Humber Summit	0.333333	2	43.756303	-79.565963	Pizza Monza	43.755043	-79.567195	Pizza Place
1	Alderwood, Long Branch	0.333333	2	43.602414	-79.543484	Il Paesano Pizzeria & Restaurant	43.601280	-79.545028	Pizza Place
1	Alderwood, Long Branch	0.333333	2	43.602414	-79.543484	Timothy's Pub	43.600165	-79.544699	Pub
1	Alderwood, Long Branch	0.333333	2	43.602414	-79.543484	Toronto Gymnastics International	43.599832	-79.542924	Gym
1	Alderwood, Long Branch	0.333333	2	43.602414	-79.543484	Tim Hortons	43.602396	-79.545048	Coffee Shop

- *10 Pizza Shop*
- *Not very competitive*

It seems any neighborhood in cluster 2 would be an ideal location to open up a pizza shop!

Appendix D - References

- Neighborhood by postal code
  - Wikipedia: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Geospatial Coordinates (Latitude & Longitude)
  - CSV file: https://cocl.us/Geospatial_data
- Available Venues per Neighborhood
  - Foursquare API: https://developer.foursquare.com/docs
- Toronto Populations
  - Population Stat : https://populationstat.com/canada/toronto
- Statistics about Pizza
  - Technomics: https://www.technomic.com/reports/consumer/consumer-trend-reports/pizza
- Animation for presentation
  - Google: www.google.ca

Neighborhood by postal code — Wikipedia: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
Geospatial Coordinates (Latitude & Longitude) — CSV file: https://cocl.us/Geospatial_data
Available Venues per Neighborhood — Foursquare API: https://developer.foursquare.com/docs
Toronto Populations — Population Stat : https://populationstat.com/canada/toronto
Statistics about Pizza — Technomics: https://www.technomic.com/reports/consumer/consumer-trend-reports/pizza
Animation for presentation — Google: www.google.ca