The Cuisine Project

By: Jenish Justin Palathingal

https://github.com/jenish1989/The_Cuisine_Project.git

INTRODUCTION

People are always interested to try different cuisines while having food from restaurants. Developments in food processing industry made lots of changes in Restaurant businesses. Stakeholders are always trying to take response from customers about their experience to improve quality of food and services.

Now restaurants are not only focused on preparing food. They also interested to create thematic ambiance for their customers, providing door step delivery services, some owners are going to an extent by providing party hall& other amenities to get volume orders.

Multi-cuisine restaurants are common in city areas. Restaurants are trying to leverage demand for country food by respective ethnic communities. Some investors are daring to start restaurant, serves exclusive cuisine to satisfy demands from ethnic group settlements. Countries like *Canada, Dubai, Singapore* are well known for its multi-cultural characteristics. Immigration rates are increasing on every year to make well mixed human settlement around the world. If we ignore current Covid-19 situation for sake of simplicity, above mentioned countries may face record number of immigration on recent years.

PROBLEM

The real challenge for restaurant owner, especially if restaurant located in multi-cultural city, is to understand of changes in food culture due to intermixed settlement of people. With the help of available demographic data, here trying to find ethnic settlements in neighborhoods of *Toronto, Canada*. By comparing with existing restaurant characteristics, we can find some insights for restaurant stakeholders to improve business by attracting more customers. Results also can be used to find suitable locations for new restaurants.

DATA SOURCES

To complete this project, we need two sets of data as follows.

1. Ethnic group settlement locations in *Toronto* neighborhoods, with population density

2. Existing restaurant data with location.

The real challenge here is to find immigrants data with details of ethnic origin. Migration & settlement data are not available for web scrapping. However, on following Wikipedia page shows demographic data of *Toronto* neighborhoods.

https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods

Name	FM	Census Tracts	Population	Land area (km2)	Density (people/km2)	% Change in Population since 2001	Average Income	Transit Commuting %	% Renters	Second most common language (after English) by name	Second most common language (after English) by percentage	Map
Crescent Town	EY	0190.01	8,157	0.4	20,393	-10.0	23,021	24.5	20.3	Bengali (18.1%)	18.1% Bengali	
Governor's Bridge/Bennington Heights	EY	0186.00	2,112	1.87	1129	4.0	129,904	7.1	13.3	Polish (1.4%)	01.4% Polish	
Leaside	EY	0195.00, 0196.00	13,876	2.81	4938	3.0	82,670	9.7	10.5	Bulgarian (0.4%)	00.4% Bulgarian	

In this web page details of 3 main neighborhoods of *Toronto* are given in table format. They are *East York, Etobicoke and Scarborough*. One of interesting data on these table is "<u>Second most common language-(After English)</u>". We can use this data to find ethnic minority groups scattered all over *Toronto*.

 Details about presence of local languages will lead us towards their food preferences.

Coordinate details of *Toronto* neighborhoods can be found using **Geocoder.** Using **Foursquare API** we can find restaurant locations around *Toronto* neighborhoods which are listed using **Geocoder**.

 Finally, data from *Toronto* Government's official website utilized to make sure findings are suitable for real world applications.

https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/ward-profiles/

METHODOLOGY

Scripts of this project executed using **Python programming language**. As mentioned in previous section, there are two set of data feed into Python program. The entire project runs on **IBM Cloud machine by IBM Watson studio**.

<u>Web scraping</u>

Demographics of *Toronto* neighborhoods are available in **Wikipedia**, will be collected using web scraping technique by **beautifulsoup**. Sufficient data cleaning will be performed before stored into data frame.

	Neighborhood	Density	Language	Percentage
0	Crescent Town	20393.0	Bengali	18.1
1	Governor's Bridge/Bennington Heights	1129.0	Polish	1.4
2	Leaside	4938.0	Bulgarian	0.4
3	O'Connor–Parkview	3591.0	Urdu	3.2
4	Old East York	6577.0	Greek	4.3
...	...	...	...	...
167	Woburn	3636.0	Gujarati	9.1
168	Wychwood	6150.0	Portuguese	2.7
169	York Mills	2409.0	Korean	4.0
170	York University Heights	1979.0	Italian	6.6
171	Yorkville	10795.0	French	1.9

172 rows × 4 columns

There are 172 Neighborhood locations are available with details of Second most using language (Other than English) with population density and percentage of language presence.

Finding Co-ordinates

Using **Geocoder**, Co-ordinates of 172 neighborhood locations can be found, and stored in table.

```python
# Defining a function to get coordinates
def get_latlng(neighborhood):
    # initialize your variable to None
    lat_lng_coords = None
    # loop until you get the coordinates
    while(lat_lng_coords is None):
        g = geocoder.arcgis('{}, Toronto, Canada'.format(neighborhood))
        lat_lng_coords = g.latlng
    return lat_lng_coords
# Call the function to get the coordinates, store in a new list using list comprehension
coords = [ get_latlng(neighborhood) for neighborhood in neighborhood_data["Neighborhood"].tolist()]
coords
```

```
0]: [[43.694750000000056, -79.29529999999994],
    [43.6904400301587, -79.36866491537944],
    [43.70023653076849, -79.35106505955042],
    [43.70872000000003, -79.30676999999997],
    [43.69622000000004, -79.33288999999996],
    [43.70732000000004, -79.34352999999999],
    [43.60496000000006, -79.54115999999993],
    [43.78321081505108, -79.14896358663243],
    [43.74803000000003, -79.63121999999998],
    [43.68958082516563, -79.49475148447905],
    [43.62686000000008, -79.47670999999997],
    [43.65224745356488, -79.48669681306212],
    [43.725164920865495, -79.62155460840363],
    [43.641466413806995, -79.49253715139403],
    [43.634870000000035, -79.53051999999997],
    [43.702510000000075, -79.57208999999995],
    [43.593540000000075, -79.53274999999996],
    [43.63391000000007, -79.56947999999994],
```

Data frame will be generated by combining neighborhood data and coordinate details.

	Neighborhood	Density	Language	Percentage	Latitude	Longitude
0	Crescent Town	20393.0	Bengali	18.1	43.694750	-79.295300
1	Governor's Bridge/Bennington Heights	1129.0	Polish	1.4	43.690440	-79.368665
2	Leaside	4938.0	Bulgarian	0.4	43.700237	-79.351065
3	O'Connor–Parkview	3591.0	Urdu	3.2	43.708720	-79.306770
4	Old East York	6577.0	Greek	4.3	43.696220	-79.332890

Neighborhood locations plotted using **Folium Map**
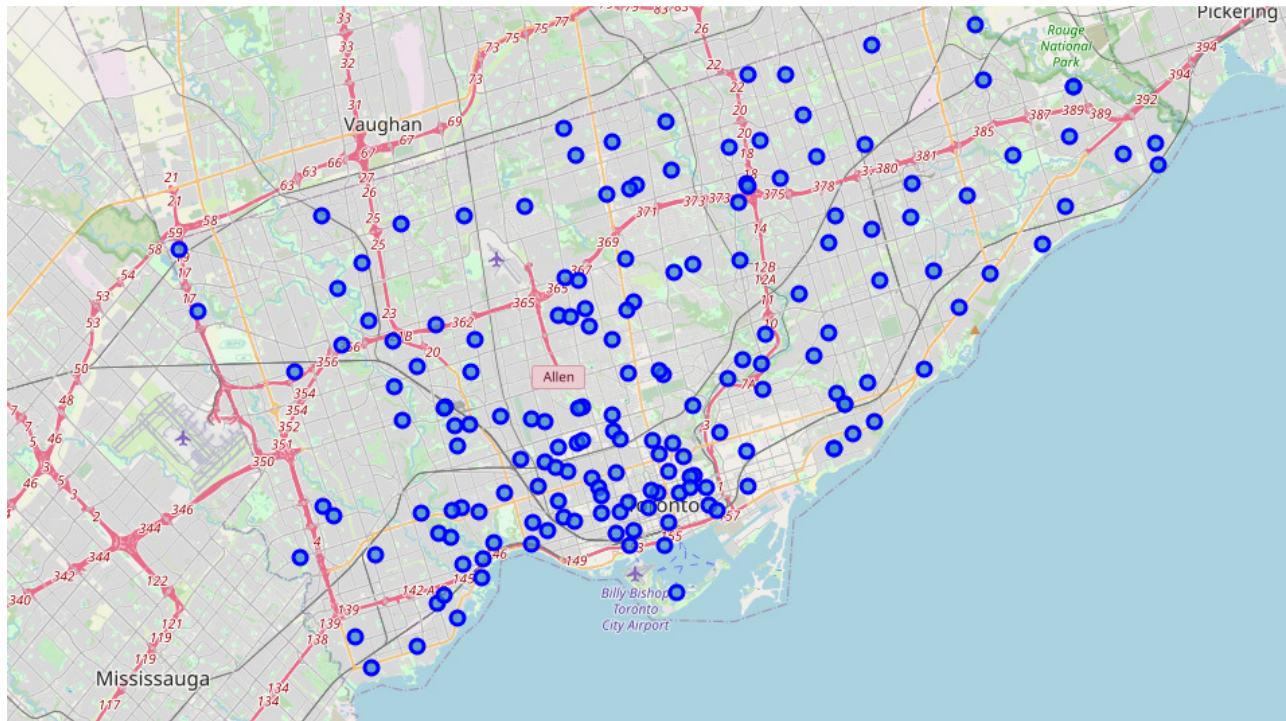
172 Locations are spread all over Toronto, sufficient to generate more realistic result.

Listing venues nearby neighborhoods

With the help of **Foursquare API**, Lists popular venues nearby 172 neighborhood locations tabulated from **Wikipedia** page. Later restaurants sorted and stored in data frame with co-ordinate details.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
11	Leaside	43.700237	-79.351065	Hakka Garden	43.704578	-79.349770	Indian Restaurant
15	O'Connor–Parkview	43.708720	-79.306770	Harvey's	43.710964	-79.309085	Fast Food Restaurant
18	O'Connor–Parkview	43.708720	-79.306770	Swiss Chalet	43.712871	-79.307668	Restaurant
21	Old East York	43.696220	-79.332890	Little Coxwell Restaurant	43.696180	-79.328958	Thai Restaurant
27	Old East York	43.696220	-79.332890	Kouzina	43.697407	-79.329175	Greek Restaurant
...	...	...	...	...	...	...	...
3294	Yorkville	43.672250	-79.385690	KOKO Share Bar	43.671034	-79.391457	Asian Restaurant
3298	Yorkville	43.672250	-79.385690	Okonomi House お好みハウス	43.668448	-79.386884	Japanese Restaurant
3299	Yorkville	43.672250	-79.385690	Meet Dumplings	43.667851	-79.385763	Chinese Restaurant
3300	Yorkville	43.672250	-79.385690	The One Eighty	43.668575	-79.388210	American Restaurant
3302	Yorkville	43.672250	-79.385690	K-Wok	43.667835	-79.385990	Korean Restaurant

775 rows × 7 columns

Creating clusters based on second most common language

Since many language groups can be found on neighborhoods, we can select most prominent 12 language groups, (selects languages appears more than 5 neighborhoods). Also we assumed number of clusters are 5, these assumptions sufficient to analyze & generating clusters spread over *Toronto* neighborhoods.

- Clustering technique used – **K-Means Clustering**
- Number of clusters – 5
- Number of second language present - 12

	Neighborhood	Density	Language	Percentage	Latitude	Longitude	Cluster Labels	1st Most Common Language	2nd Most Common Language	3rd Most Common Language	4th Most Common Language	5th Most Common Language	6th Most Common Language	7th Most Common Language
0	Crescent Town	20393.0	Bengali	18.1	43.694750	-79.295300	0	Bengali	Polish	Urdu	UnspecifiedChinese	Ukrainian	Tamil	Tagalog
1	Governor's Bridge/Bennington Heights	1129.0	Polish	1.4	43.690440	-79.368665	1	Polish	Bengali	Urdu	UnspecifiedChinese	Ukrainian	Tamil	Tagalog
2	Leaside	4938.0	Bulgarian	0.4	43.700237	-79.351065	0	Bulgarian	Bengali	Polish	Urdu	UnspecifiedChinese	Ukrainian	Tamil
3	O'Connor–Parkview	3591.0	Urdu	3.2	43.708720	-79.306770	0	Urdu	Bengali	Polish	UnspecifiedChinese	Ukrainian	Tamil	Tagalog
4	Old East York	6577.0	Greek	4.3	43.696220	-79.332890	0	Greek	Bengali	Polish	Urdu	UnspecifiedChinese	Ukrainian	Tamil

Visualizing clusters of language groups and restaurant locations.

Each clusters are represent by colors
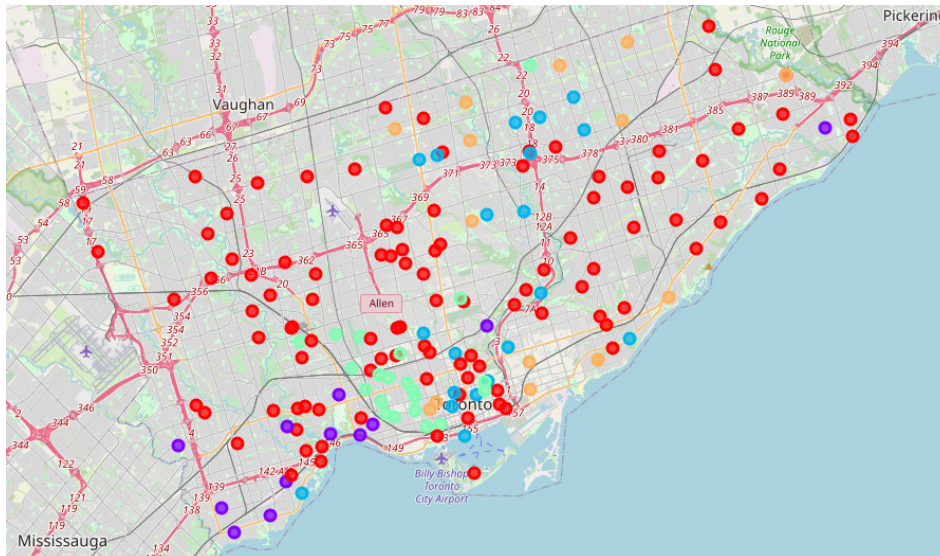
- Cluster 0 – Red
- Cluster 1 – Violet
- Cluster 2 – Blue
- Cluster 3 – Light green

- Cluster 4 – Orange

Clusters and Restaurant distribution map.

Based on population size, diameter of cluster marker increased proportionally to compare with other clusters. This will help to compare strength of ethnic community around respective neighborhoods. Green colored dots are representing existing restaurants.

Radius of Marker Proportioned to - (Population density x Percentage of second language presence)/1000
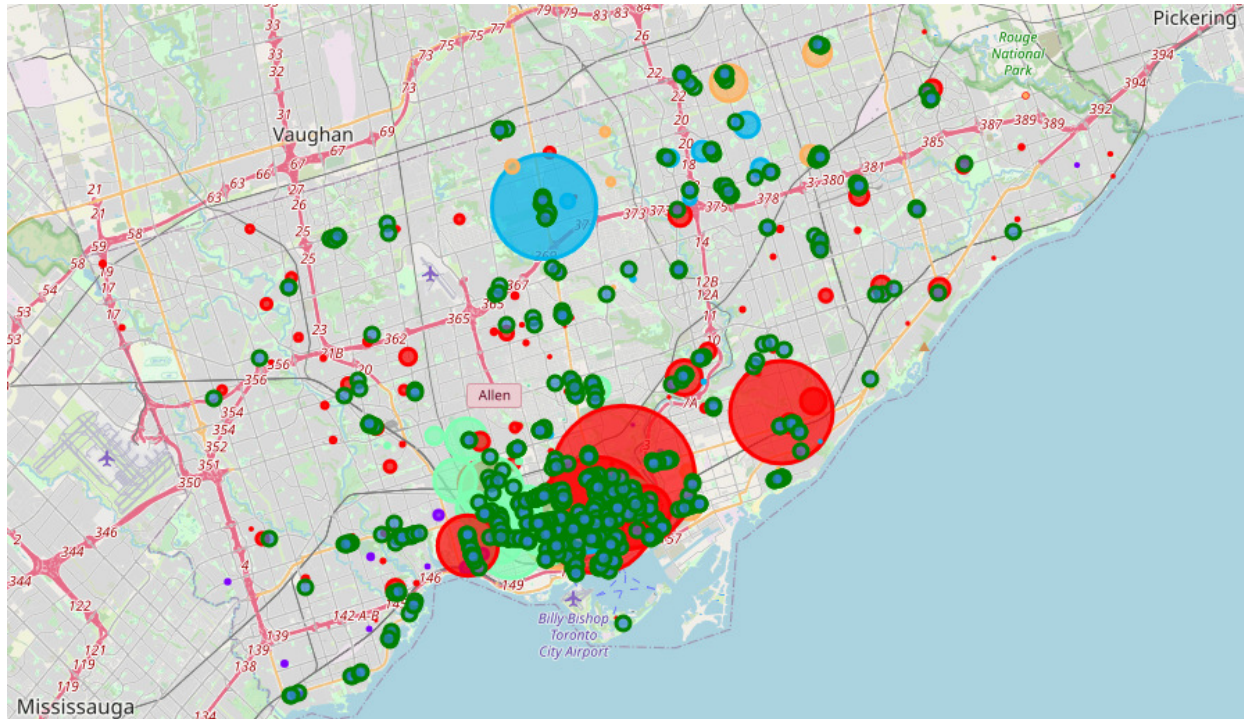
CLUSTER ANALYSIS

Cluster 0

- Largest among 5 resulting cluster.
- Spreads over – *Toronto Centre, Toronto-Danforth, University-Rosedale, Parkdale-High Park, Scarborough Southwest*.
- Second most common language is different. Almost all the 12 languages consider during evaluation can find here. Most of these languages are from *Asia region*. So <u>Chinese and Indian restaurants will be dominating this area.</u>
- This is an overcrowded area need further studies to check scope for new restaurant.

Cluster 1

- Spread over - *Parkdale-High Park, Etobicoke-Lakeshore*
- This is the smallest cluster among 5.
- <u>Second most language is "Polish",</u>
- Some traditional restaurants can be found in *Parkdale-High park and Royal York*. <u>We can suggest them to add Polish dishes on their menu.</u>
- New restaurants can be consider locations in South most areas of *York*.

Cluster 2

- Include ward mainly – *Willow dale. (North York center)*
- <u>Second most language is "Chinese"</u>
- Comparing with cluster 0 & 1, this location popular immigrant location. About 40% of immigrants are from *China, Philippines & Korea* combined.
- *Chinese* restaurants are already occupied in these areas, need to examine scope for new restaurants.

Cluster 3

- Include *York South-Weston, Parkdale-High Park, Davenport* wards.
- <u>Second most language " Portuguese" (People from Brazil and south Africa can be found in those areas)</u>
- This cluster concentrated towards *Toronto city*. Currently restaurants are spread all over this location. Probably no scope for new restaurants.

- Very less number of restaurants are in area *Spadina-Fort York*

<u>Cluster 4</u>

- Includes *University-Rosedale, Scarborough North* wards.
- <u>Second most language used is "Cantonese". (Origin from China and Korea)</u>
- Lots of *Chinese* restaurants are already in place
- Lots of places are still vacant without restaurant presence.
- Immigrants are more lives in these regions (more than 67%) so Multi-cuisine restaurants are always having scope for improvements.

RESULTS & FINDINGS

- Unoccupied areas, <u>suitable for new restaurants</u> are plotted in **Folium map**
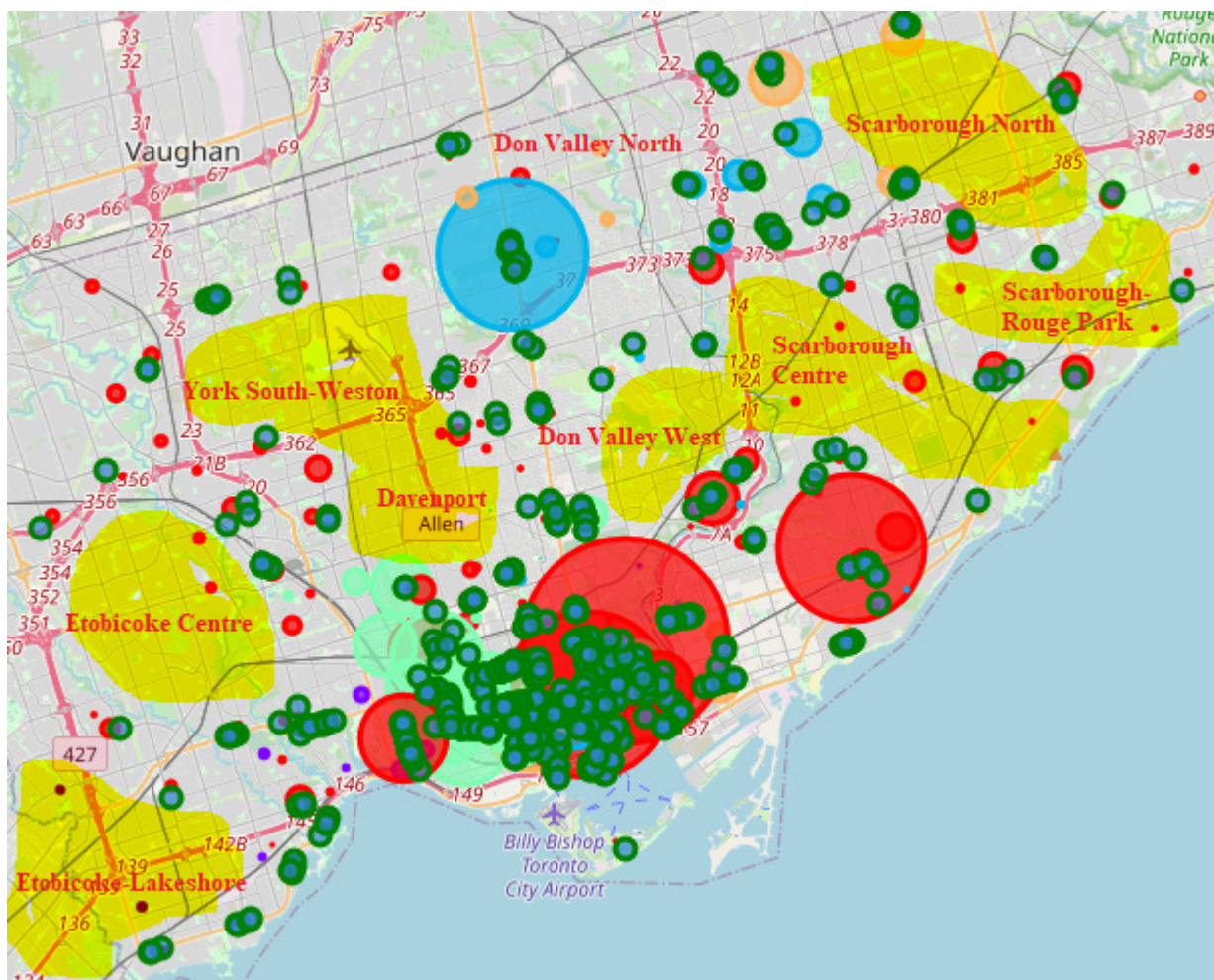
- *Toronto center, Toronto Danforth* (Cluster 0), are not suitable for new restaurants, since it's packed with many restaurants already. Existing restaurants can try to add Chinese and Indian Cuisines to their menu.

- *Etobicoke-Lakeshore* (Cluster 1), suitable for new restaurants since restaurants are less in this area. Study shows *Polish* cuisines will be successful in these areas. But not recommend for exclusive restaurant because *Polish* population not significantly large compared to other communities

- *Don Valley North, Scarborough-Agincourt* (Cluster 2) mostly occupied by *Chinese* ethnic group, but *Chinese* restaurants are already present. Even though some areas are vacant for new restaurants.

- *Davenport* (Cluster 3) Occupied by *Portuguese* speaking community. But this location is near to *Toronto* , where large number of restaurants available. Existing restaurants can be consider to add *Portugal* cousins to existing menu.

- *Scarborough North* (Cluster 4) Occupied by *Cantonese* speaking communities, which is related to *Chinese* culture. Because of prominent presence of migrant population especially from Asian countries, new restaurant with *Indian-Chinese* cuisines will be having scope.

SCOPE FOR FURTHER STUDIES

- Above studies are based on Language groups presented in each neighborhoods. Some communities may present in 3-4 generations, their food habits may change due to prolonged presence in *Canada*.

- Communities speaking *Cantonese and Chinese* are much related. More studies required to find difference between food cultures.

- This study based on Censes details of *Toronto* carried out in 2006, which is available on Wikipedia. Need updated dataset to arrive more accurate search results.

- Current **Covid-19 restrictions** may affect our studies due to disrupted working conditions of restaurants. So **Foursquare API** venue search may show partial result, that will affects our analysis.

- Presence of *Bengali* language we can find all the clusters, their cuisines related to *Indian* food style.

- **K-means clustering** technique can also performed by increasing number of clusters for in-depth study.

- Other demographic factors also need to consider while looking for restaurant locations such as transportation facility, rental expenses, municipal permissions etc.

REFERENCE

- Details of polish cuisines can be found on Wikipedia (**https://en.wikipedia.org/wiki/Polish_cuisine**)
- *Portuguese* cuisines can be found on **https://en.wikipedia.org/wiki/Portuguese_cuisine**
- Toronto ward profiles : **https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/ward-profiles/**
- Git hub link :

  https://github.com/jenish1989/The_Cuisine_Project.git