Food Venus in Toronto FSAs

1 Introduction

1.1 Problem statement

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,731,571 as of 2016. There are 103 FSAs in Toronto.

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 belong to a visible minority population group, and over 200 distinct ethnic origins are represented among its inhabitants. While the majority of Torontonians speak English as their primary language, over 160 languages are spoken in the city.

With the diverse population mix in Toronto, it would be interesting to analyze the restaurants in each FSAs and understand whether different restaurants are more popular.

The purpose of this report is to analyze the restaurants (the number and mix) in the area.

1.2 Target Audience

The following groups of people may be interested in this study:

- Businessmen who wants to know where to invest or open a restaurant, with a typical cuisine.
- Home-run business who plan to open a particular restaurant in the neighborhood and want to know what type of restaurant is most welcome in the area

2 Data acquisition and cleaning

The following data are obtained and used in the analysis:

2.1 Toronto FSA codes and geolocations

Source: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
http://cocl.us/Geospatial_data

The FSA data helps segregates the Toronto city into different borough for comparison and visualization.

Data clean-up:

- The web data are downloaded using beautiful soup.
- Postal codes (FSA codes) where no borough is assigned is removed from analysis.
- The geospatial data is obtained from previous course content. Alternatively, it could be extracted by python code using the geocoders library

dataframe: fsa_df, shape: (103, 2)

	FSA	Borough
0	M1B	Scarborough
1	M1C	Scarborough
2	M1E	Scarborough

dataframe: geo_df, shape: (103, 3)

	FSA	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711

2.2   Toronto FSA populations

It helps includes the population of Toronto city by FSA in latest census (2016). This may be of interest to see if there is a relationship between population in the FSA and the number of restaurants.

Data clean-up:

- Csv file could directly be downloaded from the site and the data could be directly read by python.
- There is 1 FSA in Toronto with no population, and thus that FSA is excluded from analysis.

```
dataframe: pop_df,    shape: (523, 3)
```

	FSA	Province	Population
0	K0A	Ontario	103474
1	K0B	Ontario	20945
2	K0C	Ontario	52154

2.3   List of venues in Toronto

Source: Foursquare https://api.foursquare.com/v2/venues/explore

A list of venues in Toronto is extracted from foursquare. For the purpose of this report, 100 venues were extracted from each FSA within a radius of 500.

Data clean-up:

- A python code is used to extract data from Foursquare to result in pandas dataframe format for further analysis.

```
dataframe: venues,    shape: (2205, 7)
```

	FSA	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	M3A	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	M3A	43.753259	-79.329656	Careful & Reliable Painting	43.752622	-79.331957	Construction & Landscaping
2	M3A	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop

2.4   Categories of venues in Foursquare

A list of categories used in Foursquare with different level of categories. This helps to limit our extracted list of venues to only food venues, which are of interest in this study.

Data clean-up:

- Data may have different levels of categories and venues are assigned to lowest level category. However, not all categories have level 3 venues. Therefore, for categories without level 3 venues, the upper category (level 1 or 2) needs to be applied.

```
dataframe: foursq_cat_df,    shape: (833, 4)
                        id       Venue Category              cat1                 cat2
0   4d4b7104d754a06370d81259   Arts & Entertainment   Arts & Entertainment   Arts & Entertainment
1   56aa371be4b08b9a8d5734db   Amphitheater           Arts & Entertainment   Amphitheater
2   4fceea171983d5d06c3e9823   Aquarium               Arts & Entertainment   Aquarium
```

3 Data Analysis – Methodology

3.1 Visualizing the food venues in Toronto

As only food venues are considered, the popular venues extracted from Foursquare are linked to the venue category and filter is applied to extract only food venues from the list, resulting in a dataframe with only food venues by Toronto FSAs.

```
dataframe: foods,    shape: (1203, 8)
        FSA   Latitude    Longitude          Venue   Venue Latitude   Venue Longitude   Venue Category   cat1
62   M4A   43.725882   -79.315572    Tim Hortons       43.725517        -79.313103      Coffee Shop      Food
63   M5A   43.654260   -79.360636    Tandem Coffee     43.653559        -79.361809      Coffee Shop      Food
64   M5A   43.654260   -79.360636    Arvo              43.649963        -79.361442      Coffee Shop      Food
```

The dataframe is then grouped by venue category to show the most popular food venues. As there are many categories, only the top 15 were visualized in the bar chart below for further analysis.
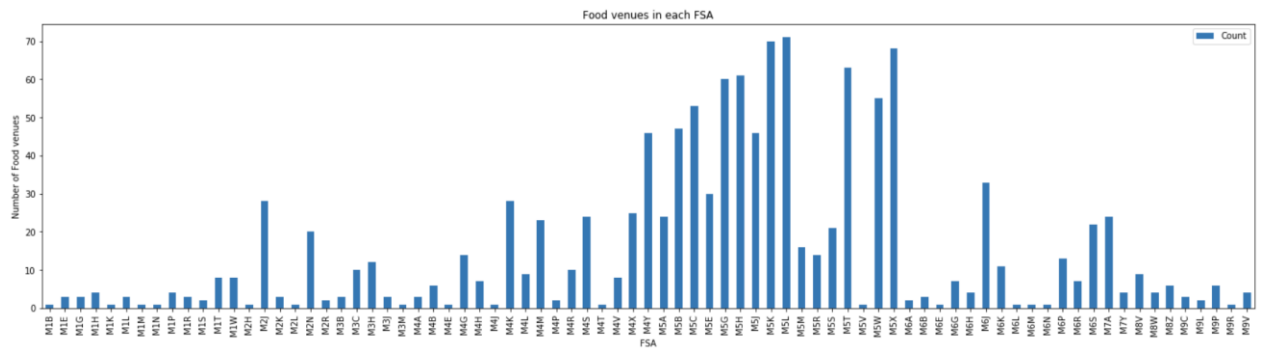
The most popular food venue in Toronto is coffee shop and café.
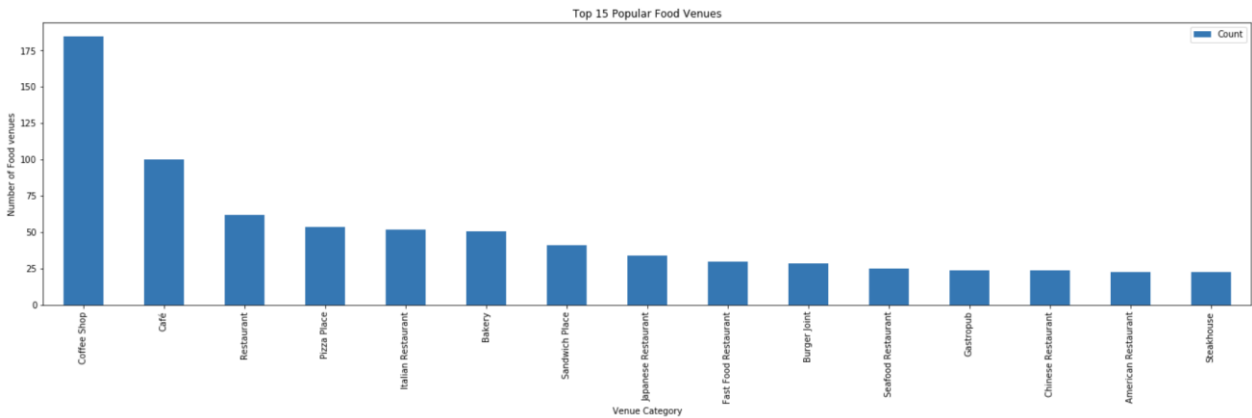
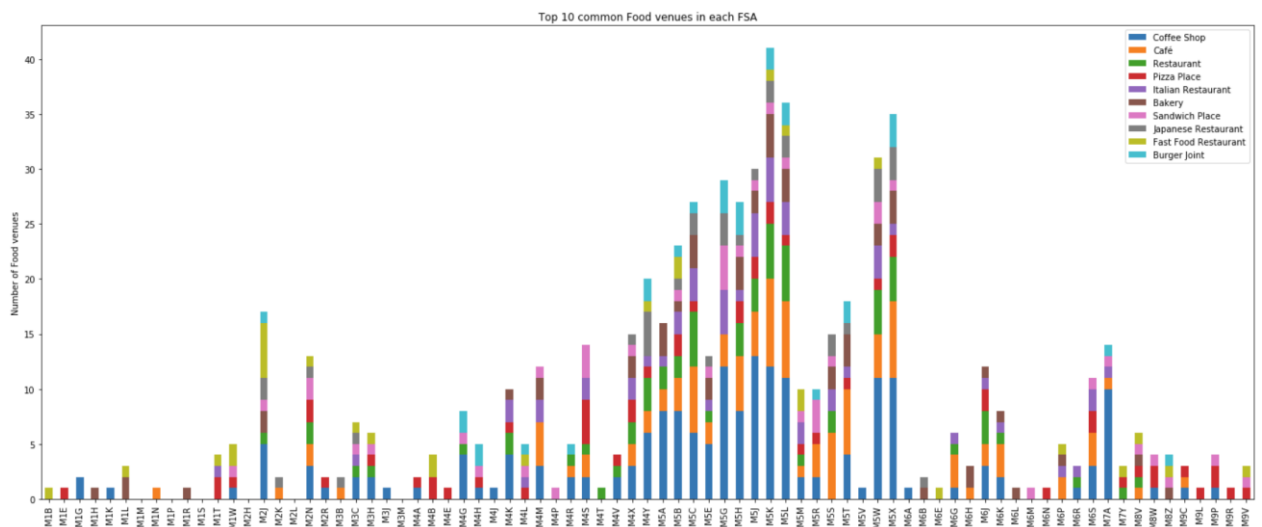## 3.2 Visualizing the food venues in Toronto FSAs

Another view is presented by grouping the dataframe of popular food venues in Toronto by FSAs.

There are most food venues in M5L, M5K and M5X.

Food venue in each FSAs are separated into different food venue categories to identify any differences in preferences of food venues in each FSAs.

We can see that there are some slight differences in each for the FSAs in their preferred food.

Clustering into 4 clusters is performed with k-means to facilitate further analysis if needed:

|	FSA	Borough	Population	Latitude	Longitude	Cluster Labels	Top 1 Food Venue	Top 2 Food Venue	Top 3 Food Venue	Top 4 Food Venue	Top 5 Food Venue	Top 6 Food Venue	Top 7 Food Venue	Top 8 Food Venue
0	M1B	Scarborough	66108	43.806686	-79.194353	0	Fast Food Restaurant	Wings Joint	Eastern European Restaurant	Cupcake Shop	Deli / Bodega	Dessert Shop	Diner	Doner Restaurant
2	M1E	Scarborough	46943	43.763573	-79.188711	3	Mexican Restaurant	Pizza Place	Breakfast Spot	Eastern European Restaurant	Cupcake Shop	Deli / Bodega	Dessert Shop	Diner
3	M1G	Scarborough	29690	43.770992	-79.216917	2	Coffee Shop	Korean Restaurant	Wings Joint	Eastern European Restaurant	Deli / Bodega	Dessert Shop	Diner	Doner Restaurant
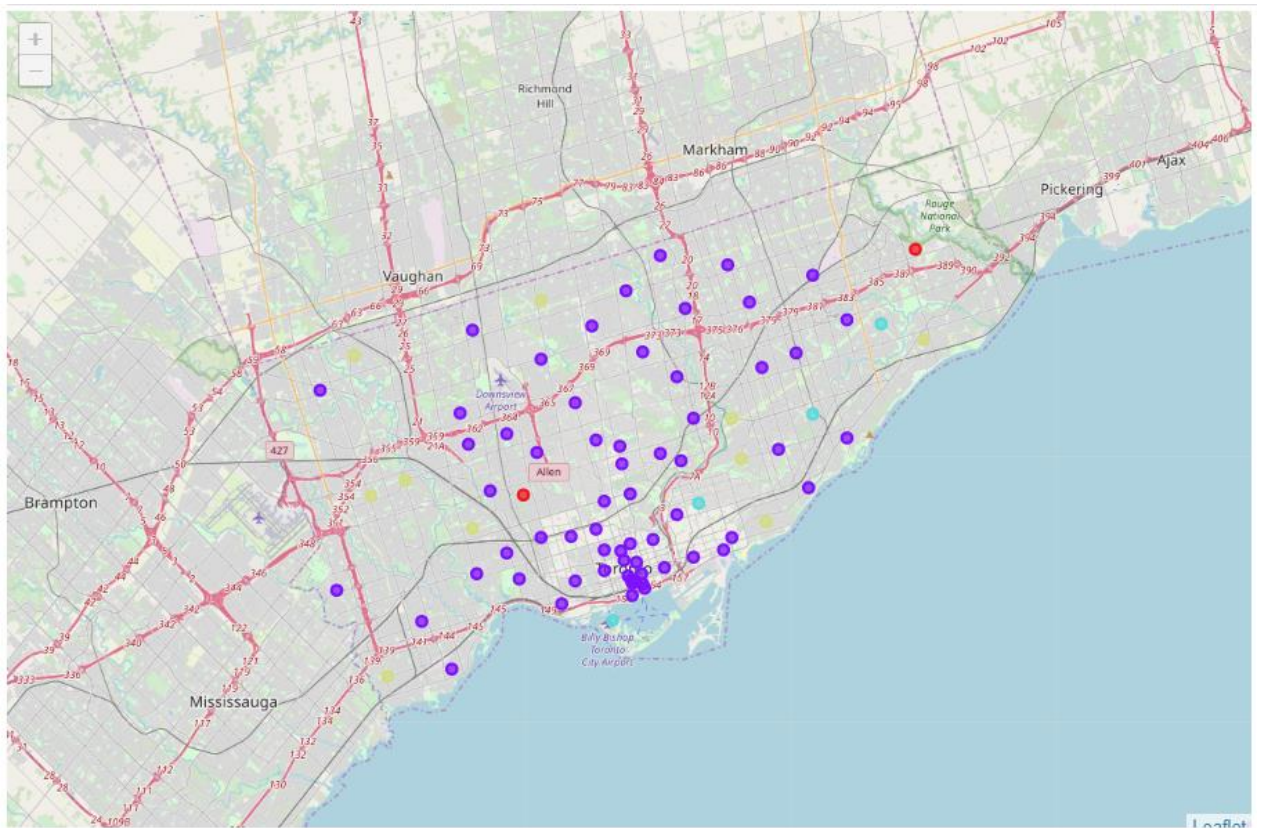
3.3 Visualizing the number of food venues on Toronto FSAs

Folium map is used to create interactive leaflet maps in python to show the locations of popular food venues in Toronto.

The following plot shows the popular food venues in Toronto.
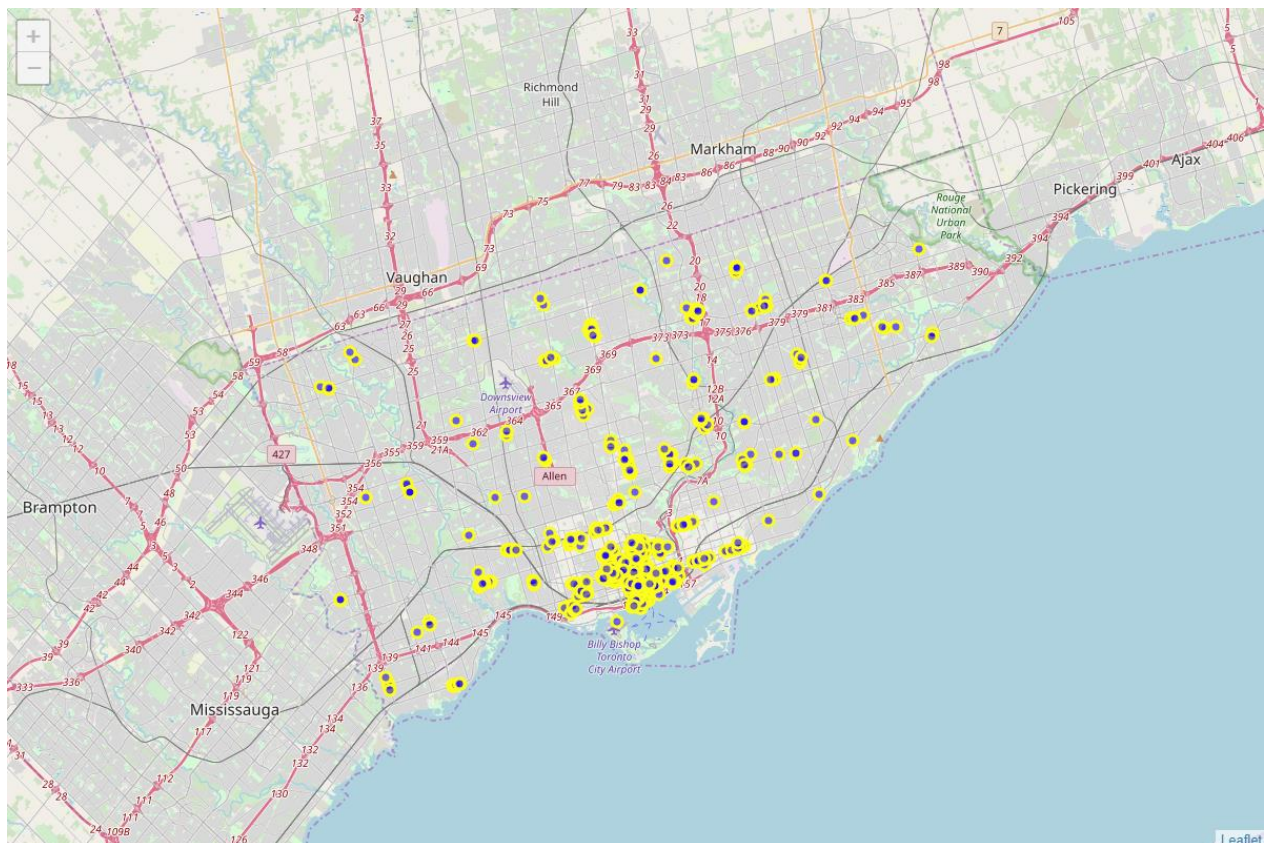
Food venue locations are group into clusters using MarkerCluster object. Each cluster is represented by the number of popular food venues in each location.

The map could be zoom in/out to change the level of grouping of these clusters.

Most restaurants are clustered in King Street and Queen Street near the Lake Ontario.
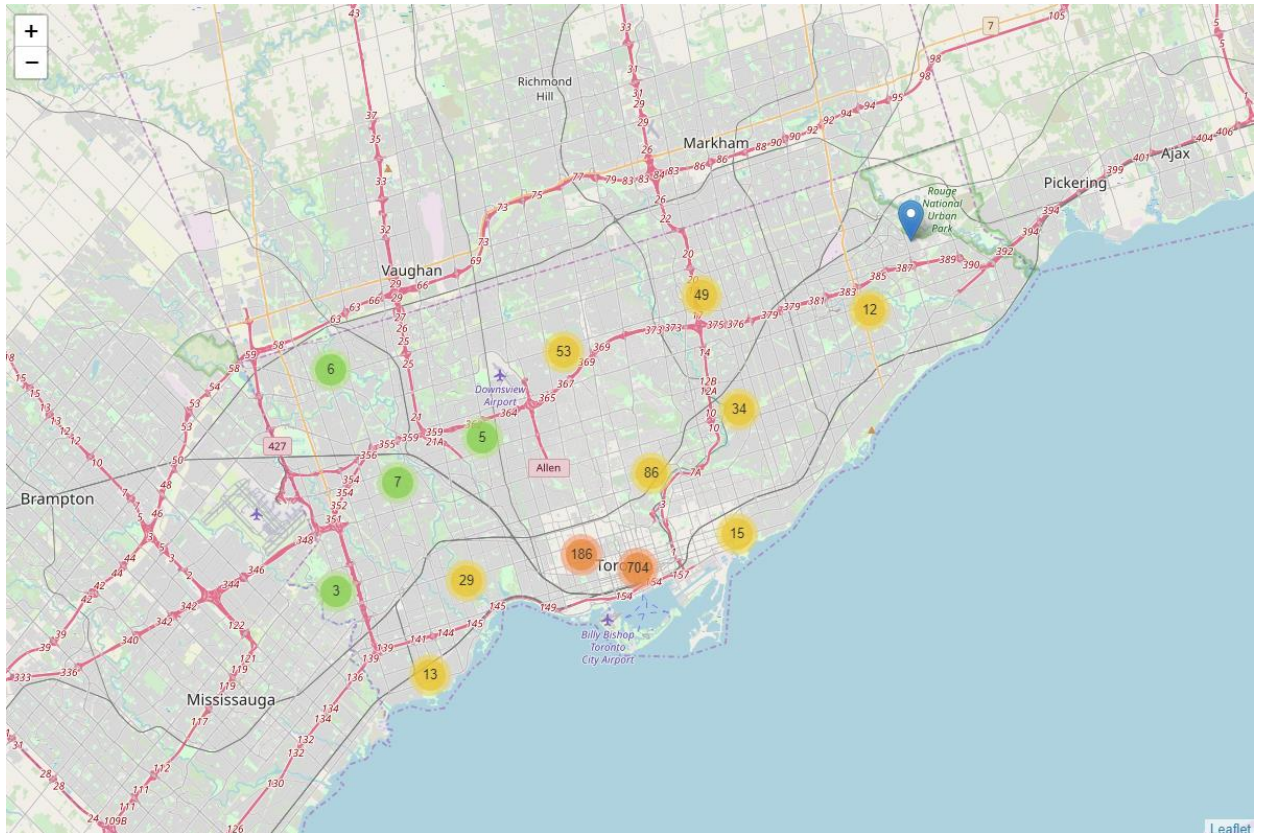

3.4   Visualizing the population on Toronto FSAs

By plotting the population by FSA on map using folium, a geojson file separating different FSA is used. To obtain the geojson file, the Canada FSA ArcGIS shapefile was downloaded from https://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-2016-eng.cfm and QGIS is used to convert the ArcGIS shapefile. (Reference: Converting shapefiles to geojson).

Two most populated areas are M2N in New York Borough and M1B in Scarborough Borough in Toronto.
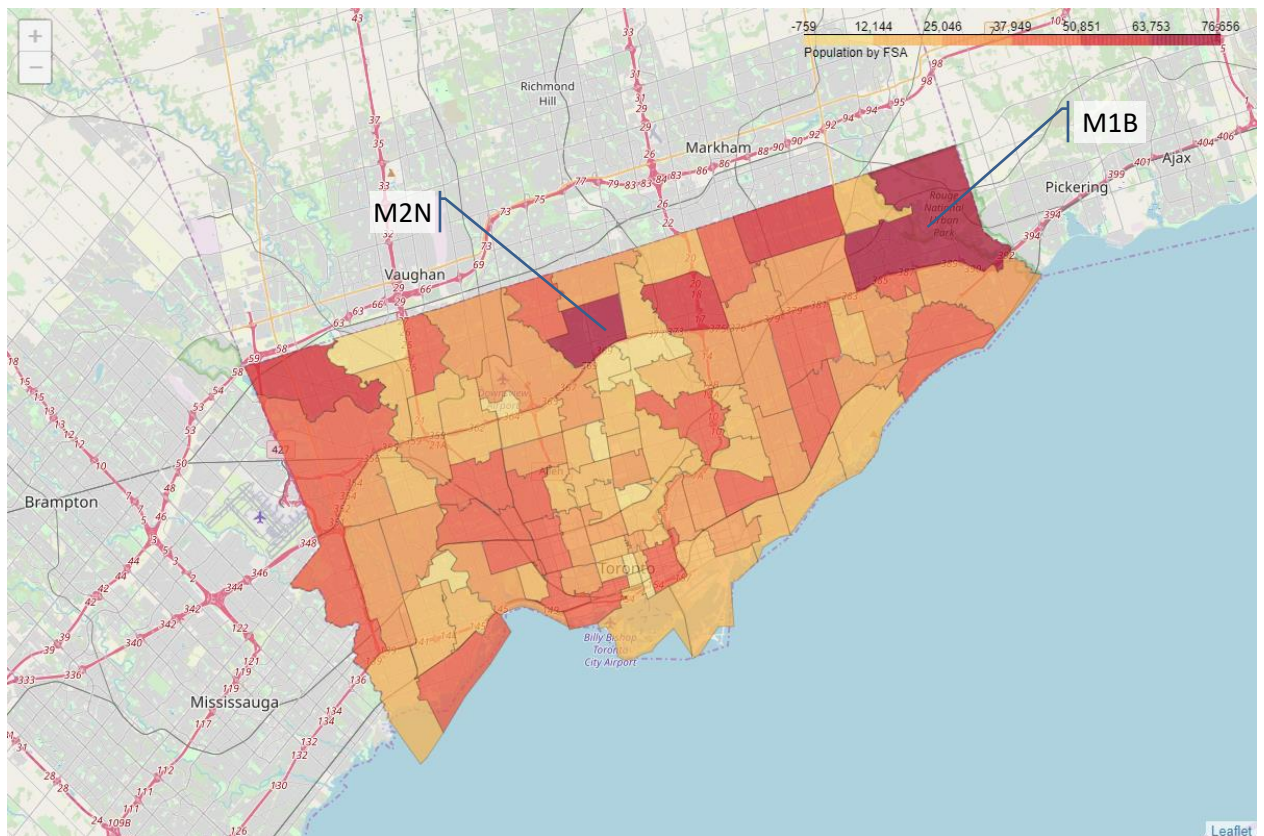
We map over the food venue cluster view on the choropleth map that shows the population to visually compare the number of population and the number of food venues.

It seems that the FSAs with high population (e.g. M1B) have only low number of food venues. The area with most restaurants (M5L, M5K and M5X) are only moderately populated.
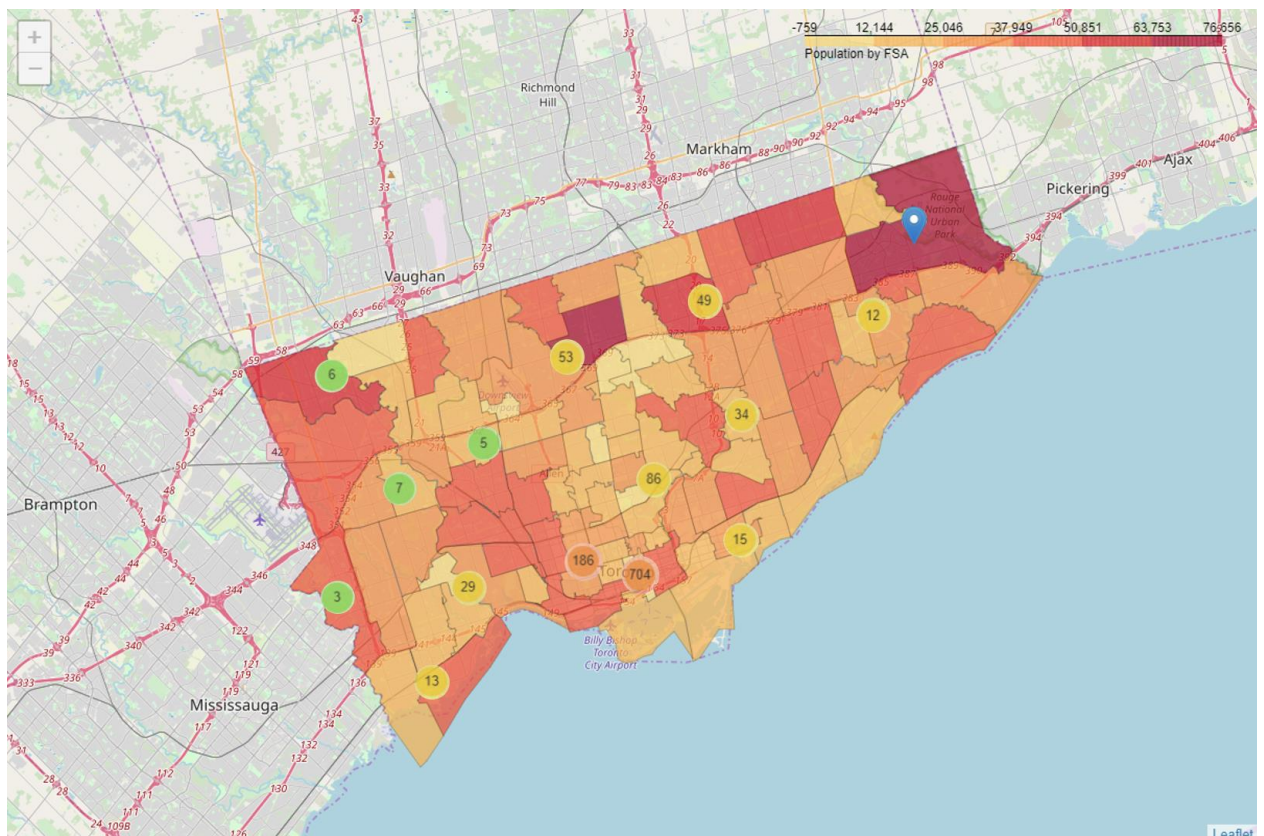
3.5   Regression Analysis

A dataframe is created to include the number of venues and the population by each FSA for analysis:

```
dataframe: regression_df,   shape: (79, 3)

    FSA  Population  Venue_no

0   M1B       66108         1

1   M1E       46943         3

2   M1G       29690         3
```

Regression analysis is performed and it shows that the number of food venues is slightly negatively correlated to population in each FSA.

```
[Text(0, 0.5, 'Number of Venues'), Text(0.5, 0, 'Population')]
```
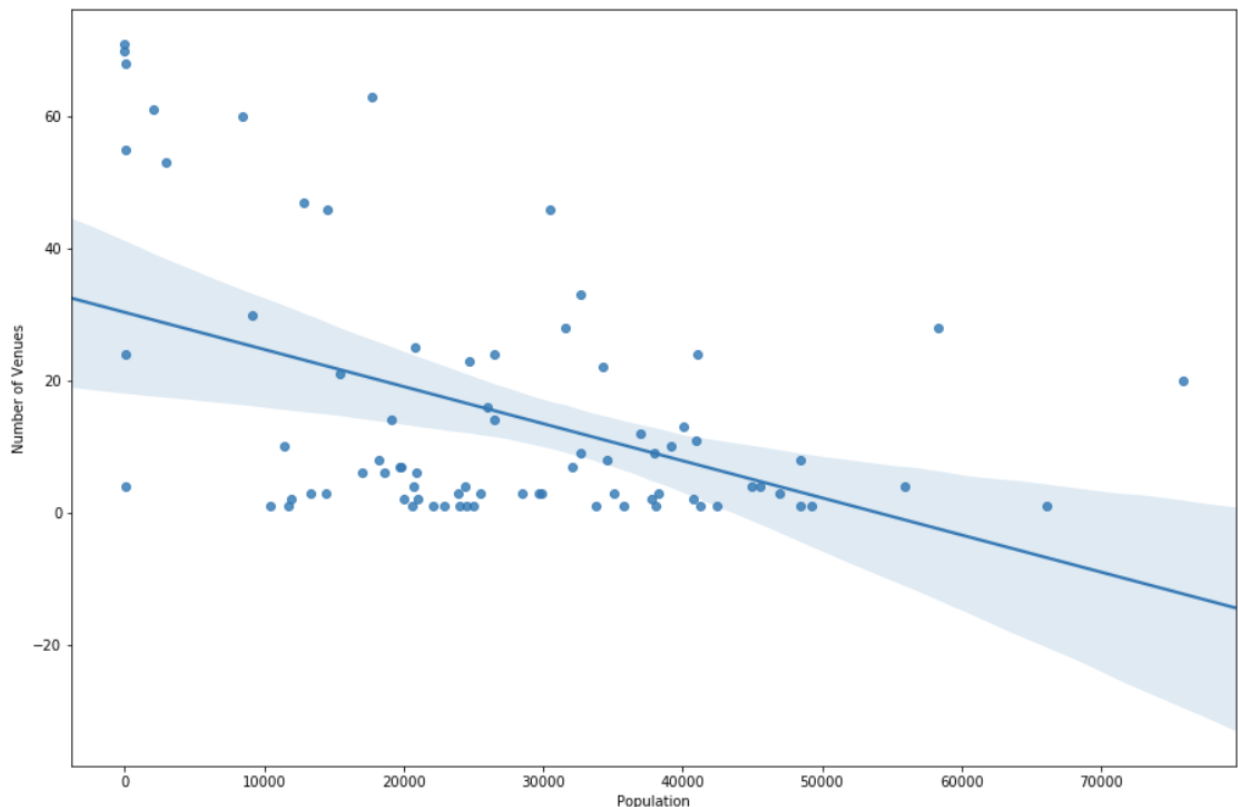

4   Results and Discussions

From the above analysis, we noted that:

- The most popular food venue in Toronto is coffee shop and café.
  This could be because of its popularity among people in Toronto, and also that coffee shops and café are usually much smaller and the number of customers there is limited. With data on size of food venues or number of tables/customers it could support, the resulting analysis might be more meaningful.
- Even though the most popular food venue is coffee shop is most of the districts, there are some slight differences in each for the FSAs in their preferred food. For example, M2J seems to have more fast food restaurant than other FSAs. Further clustering may help to further analysis the differences between FSAs.
- There are most food venues in M5L, M5K and M5X, which are the smallest FSAs in Toronto near King Street. This could be explained by the fact that it is a tourist district and attracts people there for taking a coffee or having a meal while visiting there.

- The number of restaurants in an FSA does not correlates to the population of the FSA. This is probably because Toronto is a tourist and commercial center. Other data might be more appropriate, e.g. number of tourists/visitors in the area, or the number of business buildings/offices. However, as the data is not available in this study, the analysis is not performed.

5 Conclusion

Given the limited data and time available, the analysis of food venues in Toronto is mainly limited to the analysis of categories of food venues in each FSA and the correlation with population in each FSA.

With these limitations, we have the following key observations:

- The most popular food venue in Toronto is coffee shop and café. But different FSAs may have slight differences in preference of food venues.
- The number of restaurants in a FSA does not correlates to the population of the FSA. Other factors, e.g. number of tourists/visitors or number of business building/offices may be more appropriate given it is a tourist and commercial center.

Further data may be required for more fruitful analysis, e.g. the size of the food venues, the number of tourist/visitors in each FSA, the number of business building/offices.