

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)			dataframe: geo_df, shape: (103, 3)		
	FSA	Borough		FSA	Latitude Longitude
0	M1B	Scarborough	0	M1B	43.806686 -79.194353
1	M1C	Scarborough	1	M1C	43.784535 -79.160497
2	M1E	Scarborough	2	M1E	43.763573 -79.188711

2.2 Toronto FSA populations

Source: <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/Table.cfm?Lang=Eng&T=1201&S=22&O=A>

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

Source: <https://developer.foursquare.com/docs/api/venues/categories>

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	4fceeaa171983d5d06c3e9823	Aquarium	Arts & Entertainment	Aquarium