ANALIZING FOOD BUSINESS PENETRATION IN TORONTO

Introduction

- Nowadays there is a fast increase of number of restaurants or bars on Toronto.

- Many people enjoy tasting different kind of food not only on weekends but also everyday. In

- There are some aspects to study such as location, kind of food, number of restaurants and population.

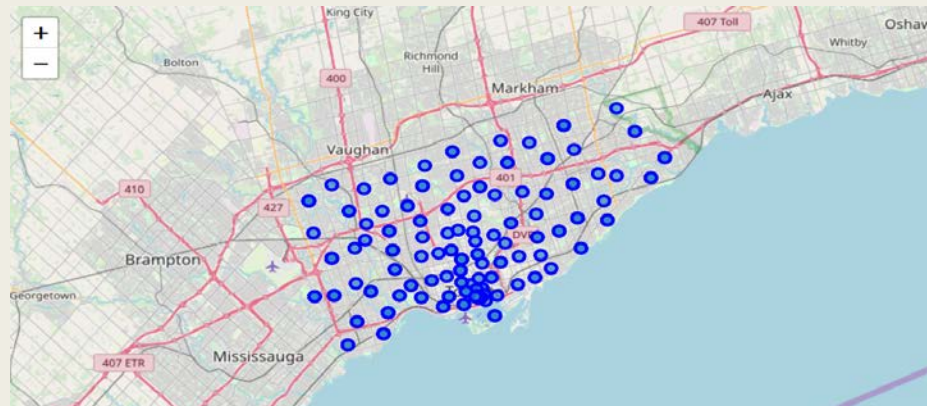# Data acquisition and cleaning

Data sources

- Datasets obtained of Internet are used to solve the problem describe above. Wikipedia provides valuable information to explore and cluster the neighborhoods in Toronto.

- Wikipedia page Here was scrape in order to filter the data.

- Also Foursquare API allows to extract the important information such as top N sites or venues in a neighborhood

Data acquisition and cleaning

Data cleaning

■ Data downloaded or scraped from sources were filter into one table. Then read it into a pandas data frame to get a structured format.

■ Use data from 2017 year and after, because this period shows the last tends.

■ BeautifulSoup was used for pulling data out of HTML and XML files

■ Folium to visualize data that's been processing by Python on an interactive map.

Data acquisition and cleaning

Feature selection

- Features were selected taking into account the objective of this project. In the first Toronto post code dataset were selected the following features:

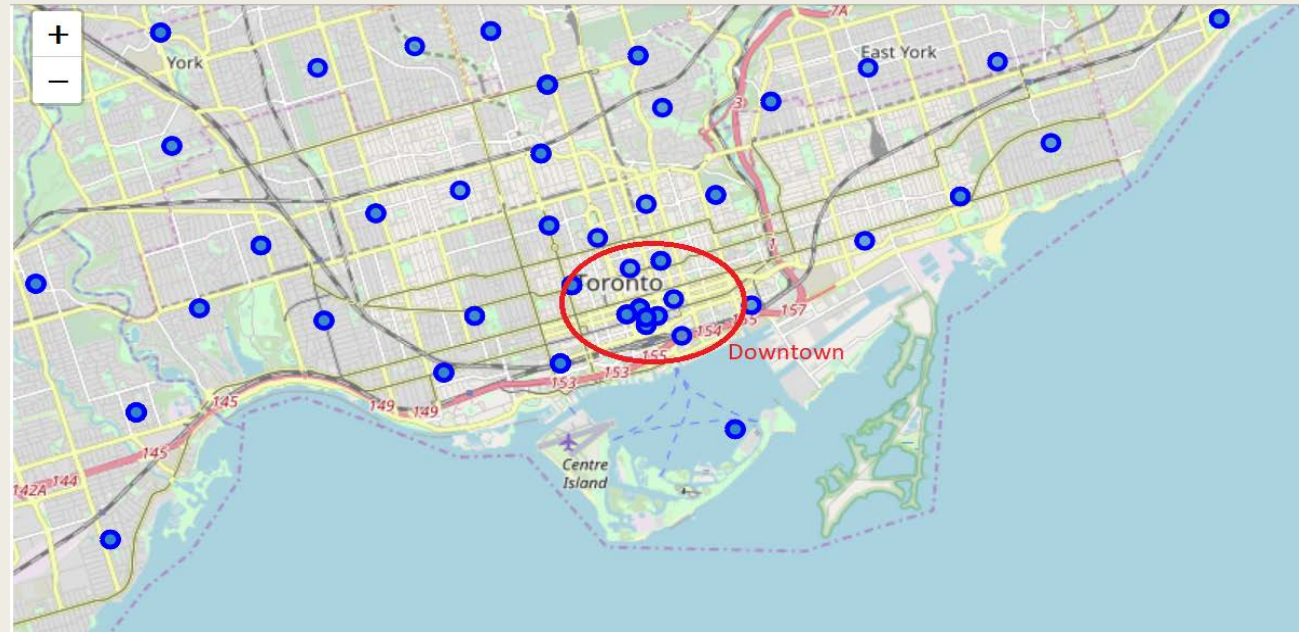| PostalCode	Borough	Neighborhood	Latitude	Longitude
M3A	North York	Parkwoods	43.752420	-79.329242
M4A	North York	Victoria Village	43.730600	-79.313265
M5A	Downtown Toronto	Harbourfront	43.650295	-79.359166
M6A	North York	Lawrence Heights	43.723270	-79.451286
M6A	North York	Lawrence Manor	43.723270	-79.451286

Data acquisition and cleaning

Feature selection

■ After use Foursquare the features that were selected are:

Postal Code	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Summary	Venue Category	Distance
M5A	Harbourfront	43.650295	-79.359166	The Distillery Historic District	This spot is popular	Historic Site	13
M5A	Harbourfront	43.650295	-79.359166	Distillery Sunday Market	This spot is popular	Farmers Market	216
M5A	Harbourfront	43.650295	-79.359166	Arvo	This spot is popular	Coffee Shop	187
M5A	Harbourfront	43.650295	-79.359166	Cacao 70	This spot is popular	Dessert Shop	127
M5A	Harbourfront	43.650295	-79.359166	SOMA chocolatemaker	This spot is popular	Chocolate Shop	91

Methodology

- **Exploratory Data Analysis**
  - *The first task of analysis of data was to examine population density on Toronto by create a map*
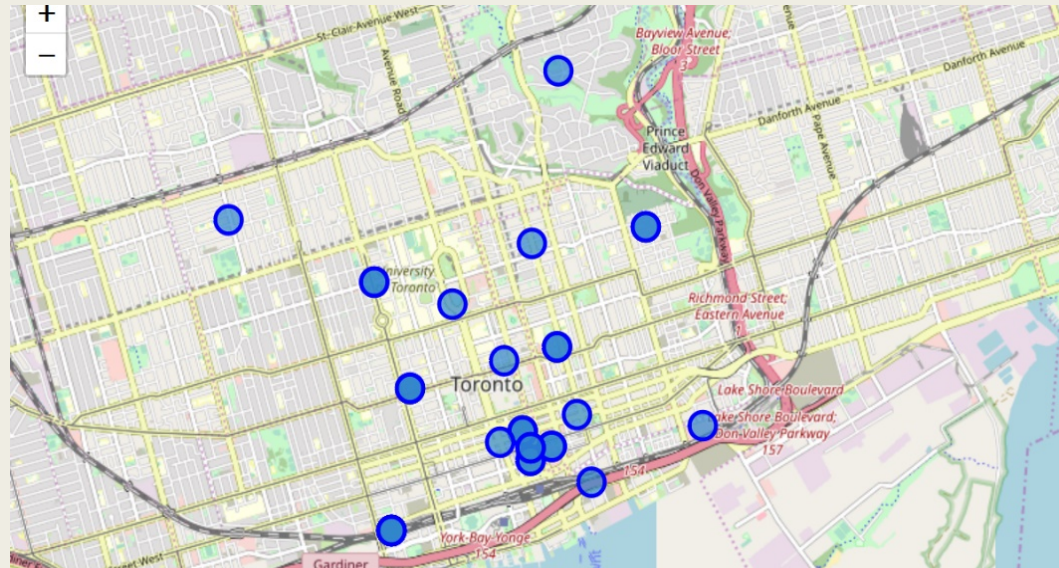
Methodology

- **Exploratory Data Analysis**
  - *The study focused in Downtown Toronto based population density.*

	PostalCode	Borough	Neighborhood	Latitude	Longitude
2	M5A	Downtown Toronto	Harbourfront	43.650295	-79.359166
5	M7A	Downtown Toronto	Queen's Park	43.661150	-79.391715
12	M5B	Downtown Toronto	Ryerson\n	43.657363	-79.378180
13	M5B	Downtown Toronto	Garden District\n	43.657363	-79.378180
26	M5C	Downtown Toronto	St. James Town	43.651210	-79.375481

Methodology

■ **Exploratory Data Analysis**

– *Showing Downtown Toronto and its neighborhoods*

Methodology

- **Exploratory Data Analysis**
  - *Using foursquare to locate venues in the neighborhoods inside "Downtown". In this process 3217 registers were analyzed*

Postal Code	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Summary	Venue Category	Distance
M5A	Harbourfront	43.650295	-79.359166	The Distillery Historic District	This spot is popular	Historic Site	13
M5A	Harbourfront	43.650295	-79.359166	Distillery Sunday Market	This spot is popular	Farmers Market	216
M5A	Harbourfront	43.650295	-79.359166	Arvo	This spot is popular	Coffee Shop	187
M5A	Harbourfront	43.650295	-79.359166	Cacao 70	This spot is popular	Dessert Shop	127
M5A	Harbourfront	43.650295	-79.359166	SOMA chocolatemaker	This spot is popular	Chocolate Shop	91

Methodology

■ Exploratory Data Analysis

Postal Code	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Summary	Venue Category	Distance
M4Y	Church and Wellesley	43.666585	-79.381302	Starbucks	This spot is popular	Coffee Shop	699
M4Y	Church and Wellesley	43.666585	-79.381302	Okonomi House お好みハウス	This spot is popular	Japanese Restaurant	495
M4Y	Church and Wellesley	43.666585	-79.381302	GoodLife Fitness	This spot is popular	Gym	591
M4Y	Church and Wellesley	43.666585	-79.381302	The Captain's Boil	This spot is popular	Seafood Restaurant	501
M4Y	Church and Wellesley	43.666585	-79.381302	Bulk Barn	This spot is popular	Grocery Store	594

Methodology

- **Exploratory Data Analysis**
  - Filtering *restaurant feature*

```
['Neighborhood',
 'Neighborhood Latitude',
 'Neighborhood Longitude',
 'American Restaurant',
 'Asian Restaurant',
 'Caribbean Restaurant',
 'Chinese Restaurant',
 'Fast Food Restaurant',
 'Filipino Restaurant',
 'Greek Restaurant',
 …………………………..']
```

Methodology

■ Exploratory Data Analysis

– *K-means was used to group restaurants and evaluate the best place to open a new restaurant. I decide to create 4 clusters considering the distances obtained in the previous results.*

Neighborhood Latitude	Group
43.630210	1.0
43.640815	2.0
43.645160	1.0
43.647100	4.0
43.648280	4.0
43.648395	4.0
43.648690	1.0
43.649700	4.0
43.650295	1.0
43.651210	1.0
43.653530	3.0
43.656091	1.0
43.657363	1.0
43.661150	1.0
43.663110	1.0
43.666585	1.0
43.668160	1.0
43.668781	1.0
43.682205	1.0

Results

■ The purpose of this project was to analyze the best place to open a new restaurant in Toronto. According to the results there are two ideal neighborhoods: CN Tower and Chinatown. In these neighborhoods the restaurants offer is not saturated.

```
{'Postal Code': 'M5V',
'Neighborhood': 'CN Tower',
'Neighborhood Latitude': 43.64081500000003,
 'Neighborhood Longitude': -79.39953781899999}
```

```
{'Postal Code': 'M5T',
'Neighborhood': 'Chinatown',
 'Neighborhood Latitude': 43.653530000000046,
'Neighborhood Longitude': -79.39723268299997}
```

Discussion

■ Despite of the goals reached of this project, I think that there are further improvement and development that can be done. For instance, kind of food that offer each restaurants or hours with more clients. These would deep ore the study and increase the probability of success to star the business.

Conclusion

- This study could very useful to determine the probability of success before to start a business. The initial investment to open a bar or restaurant is high, so many people need to know all information that guarantee preserve their money. The study done is good marketing tool.

- Foursquare is a good API to get venues and user preferences in data analyzing.

- There is a lot of information published on Internet that can be processed for many studies.

- Phyton and its libraries are very useful to filter data, it is a simple language with many advantages.