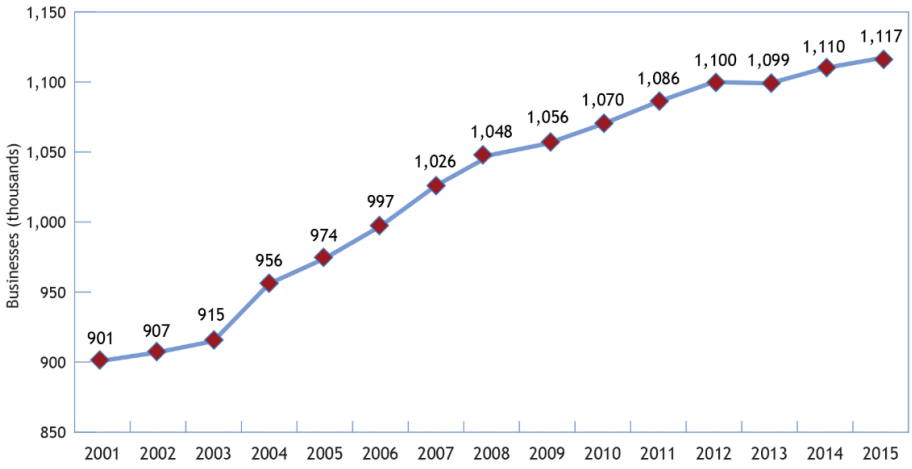
The Battle of The Neighborhoods

Background:

The number of businesses in Canada is growing every year. As you can see in the chart below, Canada has consistently had a steady increase in number of businesses. With every new business, business owners need to setup a shop or headquarters. And based on the data shown below, this need will only continue to grow. The given project can help such business owners setup their new shops/businesses in the most strategic and beneficial locations.

Figure 1: Number of businesses with at least one employee, Canada, 2001–2015



Source: Statistics Canada, Canadian Centre for Data Development and Economic Research (CDER), National Accounts Longitudinal Microdata File.

Introduction/Business Understanding

This project is to give us a glimpse into some of the major neighbourhoods of one of the biggest cities in Canada, namely Toronto. The goal of this project is to help someone who wants to open a shop/business in Toronto determine a suitable location to open there business. By leveraging data science and the Foursquare

API they can be provided with details about existing and competing businesses and position their new shop/business accordingly.

Some of the business requirements of this project include:

- Determining the top most popular venues in a given neighbourhood
- Categorizing and providing data on particular venues such as Restaurants, Parks etc.
- Producing the locations (latitude and longitude), names, ratings, categories, addresses etc. of the produced venues as required
- Producing information specific to the category of shop/business that the stakeholder is trying to create
- and more

Example Scenario

For example lets say an aspiring Restaurant owner would like to open a restaurant somewhere in Toronto. This project should help him by providing data regarding similar restaurants in each neighbourhood, their locations (latitude and longitude), and additional details about the restaurant that can be produced by the Foursquare API such as rating, menus, reviews etc.

Stakeholders

So for this project the stakeholders are business owners looking for locations for their new business/shop. The information generated and produced by this project could also be of interest to various industries.

Analytical Approach

Given the nature of this project and business requirements, we will be choosing a descriptive analytic approach

Data

Gathering Data

We will be gathering data using data provided by Wikipedia on the city of Toronto stored in csv format as well as leveraging the data provided by Foursquare API. Some of the parameters of data we will be using includes the following

- latitude and longitude coordinates of neighbourhoods and venues
- top venues in each neighbourhood
- category of each venue (Restaurant, Park, Museum etc.)
- Venue names
- Neighbourhood names
- Venue details etc.

Understanding the data

We can understand various parameters of the data using visual analysis and descriptive statistics. We must also do this to check the data quality (missing, invalid or misleading values) to ensure a comprehensive dataset

Data Preparation

This includes aggregating various records of data to create comprehensive single records in a data frame to represent one neighbourhood, with all required parameters as columns

Methodology

Now that we have loaded our data set into a data frame we perform the following exploratory analysis and inferential statistical testing

- Adding postal code as a parameter/variable to the data frame
- Checking whether the data set is correct by displaying records where postal code is 'M5G' and where postal code is 'M9V'

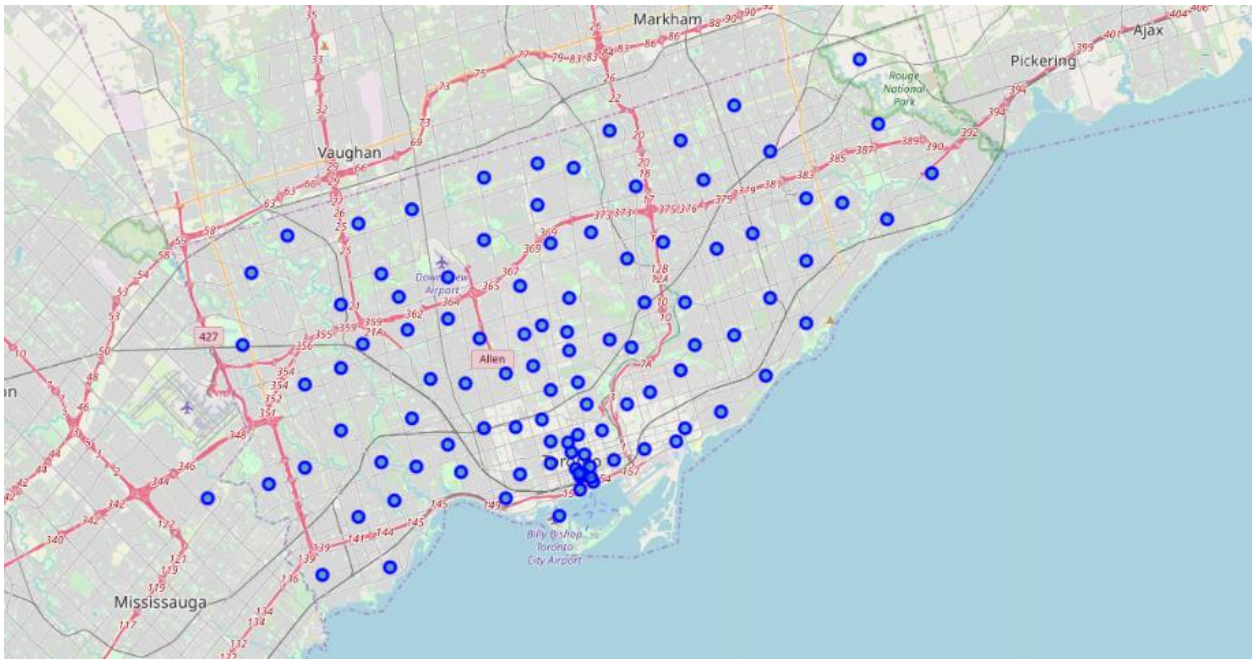
Out[18]:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
57	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383

Out[19]:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
101	M9V	Etobicoke	Albion Gardens, Beaumond Heights, Humbergate, ...	43.739416	-79.588437

- We then map the neighborhoods in Toronto on a map. (Map works on jupyter notebook but does not render on github)



- Create app on Foursquare developers tools
- Leverage data from Foursquare API by making calls with appropriate urls, Client ID and Client secret

- Display venue data near a given neighbourhood leveraged from Foursquare including Venue name, Venue locality, Venue distance from neighbourhood, Venue address, Venue category etc.

```
In [27]: nearby_venues = json_normalize(venues) # flatten JSON
         nearby_venues
```

Out[27]:

venue.id	venue.location.address	venue.location.cc	venue.location.city	venue.location.country	venue.location.crossStreet	venue.location.distance	venue.location.formattedAddress	venue.location.labeledLatLngs	venue.location.lat	venue.location.lng	venue.location.post
21b8d48842553	6210 Finch Ave. W Suite 103	CA	Etobicoke	Canada	at Albion Rd.	344	[6210 Finch Ave. W Suite 103 (at Albion Rd.), ...	[{'label': 'display', 'lat': 43.742421428838322...	43.742421	-79.589471	M
200f479154133c	1530 Albion Rd	CA	Etobicoke	Canada	Albion Mall	405	[1530 Albion Rd (Albion Mall), Etobicoke ON M9...	[{'label': 'display', 'lat': 43.741685, 'lng':...	43.741685	-79.584487	M
200f47e334153c	1530 Albion Rd.	CA	Etobicoke	Canada	at Kipling Ave. (Albion Centre)	370	[1530 Albion Rd. (at Kipling Ave. (Albion Cent...	[{'label': 'display', 'lat': 43.741201655509377...	43.741202	-79.584545	M
94a520c45522e3	1530 Albion Rd	CA	Etobicoke	Canada	at Finch Ave. W.	413	[1530 Albion Rd (at Finch Ave. W.), Etobicoke ...	[{'label': 'display', 'lat': 43.741696, 'lng':...	43.741696	-79.584379	M
b4a1cd4337535c	1530 Albion Rd	CA	Etobicoke	Canada	Near Finch Ave. W.	413	[1530 Albion Rd (Near Finch Ave. W.), Etobicok...	[{'label': 'display', 'lat': 43.7416936, 'lng':...	43.741694	-79.584373	M
c4e41bdaaf7667f	1530 Albion Road, Unit T25	CA	Etobicoke	Canada	NaN	397	[1530 Albion Road, Unit T25, Etobicoke ON M9V ...	[{'label': 'display', 'lat': 43.741568968801908...	43.741569	-79.584489	M
b0d0a78d7bcd27	1620 Albion road	CA	Toronto	Canada	Albion Road and Finch Ave	319	[1620 Albion road (Albion Road and Finch Ave),...	[{'label': 'display', 'lat': 43.74184023292111...	43.741840	-79.590561	
4e2d7f2d3a2a80	1530 Albion Road, Unit F-1	CA	Toronto	Canada	NaN	427	[1530 Albion Road, Unit F-1, Toronto ON M9V 1B...	[{'label': 'display', 'lat': 43.7417571, 'lng':...	43.741757	-79.584230	M

- Filter this data so that we only include desired parameters such as Venue name, Venue category, latitude and longitude

	name	categories	lat	lng
0	Subway	Sandwich Place	43.742421	-79.589471
1	Shoppers Drug Mart	Pharmacy	43.741685	-79.584487
2	Popeyes Louisiana Kitchen	Fried Chicken Joint	43.741202	-79.584545
3	Sheriff's No Frills	Grocery Store	43.741696	-79.584379
4	The Beer Store	Beer Store	43.741694	-79.584373

- Return the number of these venues returned by Foursquare

```
In [29]: print('{} venues were returned by Foursquare.'.format(nearby_venues.shape[0]))

         8 venues were returned by Foursquare.
```

- etc.

Results

We have successfully provided all the data required by business owners to allow them to open their new businesses in the most strategic and beneficial areas. We have met all our business requirements. We have provided our stakeholders with the data required in a filtered, structured and easy to understand format.

We have provided all sorts of relevant data statistics such as most popular venues in each respective neighborhood, most popular venues in a given neighborhood, additional details of each venue leveraged by Foursquare including addresses,

postal codes, categories, names etc. We have provided venue category specific data as well.

Observation:

We were able to make significant observations such as the surplus of restaurants and cafes in neighborhoods like Scarborough and West Toronto. We also note that neighborhoods like Central Toronto, Downtown Toronto, East York and North York have stiff competition for coffee shops. Whereas neighbourhoods like Etobicoke and York are good neighborhoods to open a coffee shops.

Discussion Section:

The future scope of this project includes making it available to business holders so that they can make use of the information it produces. This project has successfully leveraged the modern technologies of data science while making use of one of the worlds leading location data providers (Foursquare) to solve the given problem.

Conclusion

We have successfully provided all the data required by business owners to allow them to open their new businesses in the most strategic and beneficial areas. We have met all our business requirements. We have provided our stakeholders with the data required in a filtered, structured and easy to understand format.

We have provided all sorts of relevant data statistics such as most popular venues in each respective neighborhood, most popular venues in a given neighborhood, additional details of each venue leveraged by Foursquare including addresses, postal codes, categories, names etc. We have provided venue category specific data as well.