

---

# Coursera Capstone

## IBM Applied Data Science with Capstone

**Predicting the best location for a business start ups in a  
particular neighborhood of Toronto**

By:Saranesh ManiRatna.K  
November 2019

---

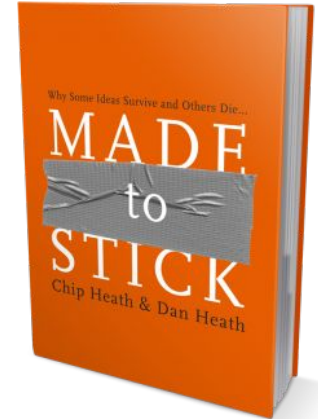
# Business Problem

A way to increase the productivity and reach of a startup is to set it up in a location which is popular and there is a better chance for business. So in order to find this place, we have to research the data of various venues and places in the region. There are various datasets and information regarding different places and venues available on the Internet.

We use this data to find the best places which would enable the business startups to succeed. Even if you have a well functioning business startup, it is important that it is set up at the right place.

So, the above mentioned process will help in maximizing the efficiency of the startup.

Instead of selecting with an basic assumption it would be nice to use the data relating to the nearby venues in particular neighborhood (needed by the customer) and providing him the best areas to open his/her business.





# 1. Introduction

- Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,731,571 in 2016. Toronto is an international center of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.
- The diverse population of Toronto reflects its current and historical role as an important destination for immigrants to Canada. The city is home to the Toronto Stock Exchange, the headquarters of Canada's five largest banks, and the headquarters of many large Canadian and multinational corporations. Its economy is highly diversified with strengths in technology, design, financial services, life sciences, education, arts, fashion, business services, environmental innovation etc.



---

## Data Section

### 1. Data Sources

1. Data of neighborhoods and respective postal codes are gathered by scraping the web page (Wikipedia).
2. Using the postal codes the latitude and longitude locations of those respective neighborhoods are gathered from a ".csv" file provided on the Internet.
3. Data relating to the venues in a certain neighborhood according to the postal codes are gathered using the FOURSQUARE API.

### 2. Data Cleaning

1.The data obtained from the Wikipedia page is taken as HTML text at first so it taken into a pandas data frame.

- Then the rows with 'Borough' as 'Not assigned' are dropped.
- All the rows with 'Neighborhood' feature as 'Not assigned' are replaced with it 'Borough' feature.
- Then the locations are grouped according to the postal codes.

---

## Methodology:

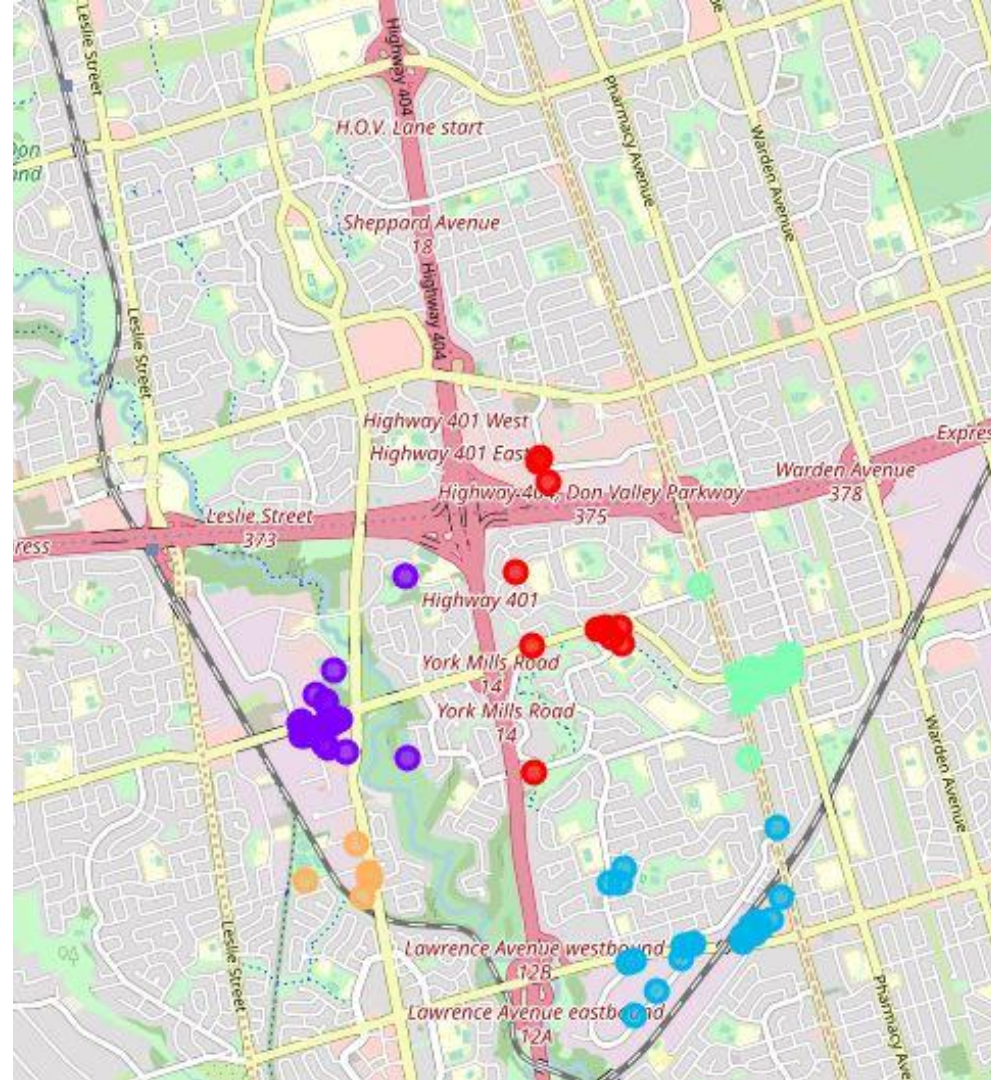
# Methodology section

Once the customer has provided you with

- The details of the business he wanted to start.
- The region he wanted to start that business.
- . The venues nearby that region are taken from API of Foursquare. After the venues are gathered the locations are plotted on the map of Toronto and are clustered as per the need which indicates the crowdedness of shops around it. For clustering the venues, the K-Means clustering is used.

## Results

A list of best possible cluster centers (latitude, longitude) are give back along with the a visualization map by folium which also indicates the best locations for the business customer wanted to start in that area. With minimum risk involved, there will be less possible competition from other businesses of that area. It is taken care to provide the location with minimum competition for that business. And also making sure that the area nearby is populated or having popular venues.





## Discussion Section

- If someone wants to start a Coffee Shop in Parkwoods, the best way to do so, is by analyzing the number of Coffee Shops and the total number of venues present in the various clusters present in the map. We do this to ensure that the startup faces as little competition from other similar venues as possible. Here, it is better to start a Coffee Shop in the 4<sup>th</sup> cluster as there are no other similar types of venues. This will help in the startup getting off to a good start.

---

## Conclusion.

People are turning to big cities to start a business or work. For this reason, people can achieve better outcomes through their access to the platforms where such information is provided.

Not only for investors but also city managers can manage the city more regularly by using similar data analysis types or platforms.