This capstone project ‘Battle of Neighborhoods’ is part of Applied Data

Science capstone course, Coursera, which is a sub-course of

‘IBM Data Science Professional Certificate’

---

Ankush Sharma

Table of Contents

## Problem statement

The most trending, essential and leading business in most of the western countries and other is the food industry.

In this project, a recommendation system is created analyzing various factors to suggest business ventures, entrepreneurs, investors and others to find suitable neighborhood / towns in Toronto city, Canada to start and establish their restaurant business successfully!

## Description of the background

According to the statistics provided by CNBC report, around ***60%*** of new restaurant fail within the ***first year*** and nearly ***80%*** shutter before ***fifth anniversary***. The most common and ***top reason*** is the ***location of the restaurant*** and the lack of self-awareness about the business's appropriate location.

To help them identifying the right business location among many towns in Toronto city for starting the restaurant, this research project provides a recommendation system considering various factors including but not limited to analyzing venues in and around the neighborhoods of Toronto city, comparing and providing the best and suitable place to start and run the restaurant successfully!

**Key benefiters:** Ventures, Entrepreneurs, startups companies, vendors etc.

## Data | Implementation (Solution to problem)

This research has utilized and implemented following,

### Dataset

The dataset used for the project is extracted from Wikipedia link

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M **Details of dataset**

- Postcode
- Borough
- Neighborhood
- Latitude
- Longitude

### Details of Neighborhood

- Venues
- Borough

- Latitude
- Longitude
- Venue category

Geographical details are queried from **Foursquare** platform, performing **API call traversal** and processing the resulting **JSON file to pandas dataframe**.

### Libraries used

- Pandas
- Matplotlib
- Numpy
- Folium
- Scikit-learn

### Algorithms • K-means clustering

It is one of the most common and simplest unsupervised machine learning algorithms. This is used for clustering the similar group of datapoints together which is not similar to other set of clusters of datapoints.

This is done by randomly selecting centroids, which is the beginning point of every clusters and performing iterative calculation to optimize the positions of the centroids.

The identification and optimization of clusters stops when there is no change in their values as the clustering is found to be successful.

### Purpose of K-means in project

- This clustering algorithm is used for identification of most common venues in the neighborhoods / towns of Toronto city, Canada to identify the best and most suitable location for intended business amongst them all. This will help the key benefiters to precisely choose location suitable for their restaurant business.

### End-result

- The outcome will suggest the suitable business location automatically reducing the cost of analysis, time and efforts for the same.
- Henceforth, **preventing the occurrence of 60% and 80% shut-off** level of restaurants as per the reports of CNBC (***60%*** of new restaurant fail within the ***first year*** and nearly ***80%*** shutter before ***fifth anniversary***)

This project is published in my Github repo https://github.com/eshrnks/Coursera_Capstone