

Coursera Applied Data Science Capstone Project

The Battle of Neighborhood

Author: Anoop Kohale, IBM

Introduction/Business Problem

Market analysis is an important part of any business start-up. The success or failure depends on the location where the business is opened. In big cities like New York, there is lot of competition to attract customers with your offerings. However, selection a perfect neighbourhood is often difficult and require lot of work.

Opening any business in any neighbourhood requires carefully analysing number of aspects of the business for it to become successful. There are number of factors influencing the decision, viz. will there be enough customers to buy my offerings, is there any competition around that can take away by revenue, etc. To success the businessman must carefully study these factors and come up with the strategy or plan of operating the business in the neighbourhood.

Fortunately, Advanced data analysis and machine learning will help taking this decision with the information available in abundance around the internet. Foursquare is such an information provider. Foursquare provides data about the interesting venues around any neighbourhood. We can utilize the machine learning algorithms and find out the clustering of specific business in the neighbourhood. This will empower us with the understanding of demographics and we can then take better decision that will result in making the business a success.

The main beneficiary of this project will be any entrepreneur who wishes to open a business in big city like New York. The project will try to find a suitable neighbourhood to open a business. For example, someone wants to open a bakery and is looking for a suitable neighbourhood, this project will give an insight on to the venues in a neighbourhood and then can decide whether

opening a bakery there will be a suitable option. E.g. if there are other bakeries in the neighbourhood it may not be a suitable option.

Data to be used

Any location you open a business has number of other similar businesses operating. One needs to analyse the data available at your hands to come up with a decision. During this project we will gather data from various data sources. Below is the list of the data sources used for this project.

1. Foursquare Venues data

- a. Type: API Call to Foursquare. JSON data about the venue.
- b. Description: The data has various venues around a location within specific radius. Venues are categorized and reviewed by users of Foursquare
- c. Source: <https://www.foursquare.com>

2. Geocoder data

- a. Type: Latitude and Longitude data for given location
- b. Description: The latitude and longitude data of a given location can be extracted using GEOCODER library
- c. Source: Geocoder library

3. Neighbourhood data

- a. Type: Neighbourhoods around New York city
- b. Description: Neighbourhoods of New York city
- c. Format: GeoJSON data
- d. Source: <https://geo.nyu.edu/catalog/nyu-2451-34572>

The main features of the data will be neighbourhood and their latitude and longitude. Foursquare API will provide data about the venues near by the latitude and longitude of the neighbourhood. This data will include category of the venue, its popularity in terms of user ratings and other related data.

Using this information, we can cluster the venues using clustering techniques. The clusters then will be visualized to take decision.