

Opening of new restaurant in the neighborhood.

Sneha Videkar

May 13, 2020

1 Introduction

Yelp [1][2] is a company which publishes crowd-sourced reviews about businesses using the company website.

Yelp dataset captures business data from all around the world. Mainly this site captures reviews related to business of various categories like Dentist, Restaurant, etc. Along with reviews, it lists location details of every business in Yelp database.

1.1 Business Problem

Success of new business depends on many factors. Location is the key factor which should be considered while opening new business. For this purpose, location attributes present in the Yelp dataset[4] will help find a place to start a new business.

This project aim to help investors/entrepreneur to find an appropriate location for their prospective restaurant business. With the help of Yelp Business dataset and machine learning techniques like clustering, this project will locate top 10 most common restaurant businesses in a neighborhood. This will guide investors while opening a new restaurant business at particular location.

1.2 Target Audience

This project is mainly helpful for the investors/entrepreneur who are looking for an appropriate location for their prospective restaurant business. Also, with the help of exploratory analysis from this project can help restaurant business owners to identify potential competitors.

2 Data Description

2.1 Data acquisition

For this problem below dataset will be used:

1. Business.csv will be used from Yelp Dataset. This dataset is taken from Kaggel website[3] and Yelp dataset website[4]. Figure 1 gives an idea about few records from the Business.csv dataset.

Table.1 lists all features associated with every business entity.

business_id	name	neighborhood	address	city	state	postal_code	latitude	longitude	stars	review_count	is_open	categories
kCoE3yEtg8UVz5OD3GVw	"BDJ Realty"	Summerlin	"2620 Regatta Dr, Ste 102"	Las Vegas	NV	89128	36.207430	-115.268460	4.0	5	1	Real Estate Services; Real Estate; Home Services...
xcgFnd-MwkZeO5G2HQ0gAQ	"T & T Bakery and Cafe"	Markham Village	"35 Main Street N"	Markham	ON	L3P 1X3	43.875177	-79.260153	4.0	38	1	Bakeries; Bagels; Food
INMVV_ZX7CJSDWQGD0M8Nw	"Showmans Government Center"	Uptown	"600 E 4th St"	Charlotte	NC	28202	35.221647	-80.839345	3.5	7	1	Restaurants; American (Traditional)
I09JMeQ6ynYs5MCJtrcmQ	"Ailze Catering"	Yonge and Eglinton	"2459 Yonge St"	Toronto	ON	M4P 2H6	43.711399	-79.399339	3.0	12	0	Italian; French; Restaurants
IHYICS-y8AFJultv6MGpvg	"Starbucks"	Liberty Village	"85 Hanna Avenue"	Toronto	ON	M6K 3S3	43.639863	-79.419533	4.0	21	1	Food; Coffee & Tea

Figure 1: Overview of Business.csv

Table 1: Business data column list

Data columns	Type
business_id	object
name	object
neighborhood	object
address	object
city	object
state	object
postal_code	object
latitude	float64
longitude	float64
stars	float64
review_count	int64
is_open	int64
categories	object

2. Forsquare API will be used to get the most common venues for above mentioned business and in Toronto area. Only near by venues of type restaurants will be considered for further analysis to achieve the project goal.

2.2 Data Preprocessing

Yelp Business data was loaded in the IBM Watson Studio and accordingly csv file was read in dataframe called businessDF.

Yelp business data frame has captured various categories of the business data. In this project, business with "Restaurant" as a category are considered for modelling purpose.

There are in total 67741 rows in Yelp Business dataset after pre-processing.

References

- [1] <https://en.wikipedia.org/wiki/Yelp>
- [2] <https://www.yelp.com/about>

[3] <https://www.kaggle.com>

[4] <https://www.yelp.com/dataset/documentation/main>