IBM Data Science Capstone Project (C10-W5)

Opening a new Fast-Food Restaurant in Los Angeles, California

By: Phaneendra Annamaraju

June 2021

Contents

Introduction

Los Angeles is very density populated city in west coast of USA. It is one of the major cultural hubs in the world with immigrants from many countries live in harmony. It is well known for Hollywood movie industry, night life and restaurants. Los Angeles has a very big market for restaurants and fast-food restaurants are very famous due to the tourism. Due to this, there are thousands of restaurants in the city and there is huge competition in the business. It is a very challenging task for any restaurant chain to plan their business expansion.

Problem Statement

Opening a new Fast-Food Restaurant in Los Angeles, California

A fast-food restaurant chain wants to expand their business and wants to know suitable neighborhood to open a new restaurant in Los Angeles.

Objective

The objective is to use data science methodology and machine learning algorithms to answer the following critical questions in order recommend potential location for a new fast-food restaurant in Los Angeles

- Which areas have potential Fast Food Restaurant Market?
- Which areas are lacking Fast Food Restaurants?
- What are the suitable neighborhoods in Los Angeles for opening a new Fast-Food Restaurant?

Target Audience

This is very useful for some of the major fast food restaurant chains such as Burger King, Subway, McDonald's etc… It will help them in preparing a strategy for their business expansion also understand the areas with potential market for fast food restaurants.

It will provide much more value add with further analysis on other influencing factors, measures and key external indicators. For example: impact of other similar restaurants, user ratings, market share, per capita income etc…

Approach

To answer the above questions, we need to

- Analyze number of restaurants in each neighborhood

- Population density of each neighborhood
- Select the neighborhood that has lowest restaurant to population density ratio

Data

1) Data required

- We first need list of neighborhoods
- Population density of each neighborhood
- Latitude and longitude coordinates of each neighborhood
- Venue data, specifically, list of restaurants around each neighborhood

2) Data source

- LA Times for neighborhood details and population density
    - http://maps.latimes.com/neighborhoods/neighborhood/list/
    - http://maps.latimes.com/neighborhoods/population/density/neighborhood/list/
- Geocoder library in python for latitude and longitude coordinates of neighborhoods
- Foursquare API for restaurants around each neighborhood

Methodology

1) Web scraping:

- We will first get neighborhoods and population density from LA Times portal using web scraping

2) Geocoder library in python

- Geocoder library helps get latitude and longitude coordinates

3) Foursquare API

- Get list of restaurants around each neighborhood

4) Pre-process data

- combine all data from sources by merging the data frames and create 2 new data frames
- neighborhoods : holds a unique list of neighborhoods and details of each (such as coordinates, number of restaurants, number of fast-food restaurants etc...)

- restaurants : holds list of restaurants in each neighborhood and their coordinates along with the restaurant category
- on-hot encoding
- identify 10 most common restaurant categories in each neighborhood
- prepare a new data frame to be used for clustering

5) Clustering

- perform clustering on the neighborhoods based on the restaurant categories

6) Folium map

- Visualize the clusters in a map using folium mapping library in python

7) Analysis

- Review the data to answer the questions in problem statement