

Coursera Capstone
IBM Applied Data Science Capstone Project

Opening New Restaurants in Downtown Los Angeles

By: SAEDEH FARIVAR

Introduction

For most of this decade, some grungy neighborhoods in downtown Los Angeles like the Arts District region have been irresistible magnet for builders. Construction works are heating up in these regions to keep pace with the high demand for office space. Some of the biggest projects already complete and a lot of projects planned for the area. If ultimately built, they would bring more than 5,000 new apartments and condos, along with hundreds of hotel rooms and more offices, shops and restaurants. Building restaurants of any type in an accessible walking distance to the new developed office buildings would provide the opportunity for office staff to find a convenient place to eat at their lunch time and avoid driving in downtown traffic.

Business Problem

The objective of this capstone project is to analyze and select the best locations in downtown Los Angeles to open new restaurants. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to business developers if they are looking to open new restaurants, where would you recommend that they open it?

Data

To solve the problem, we will need the following data:

- List of all districts or neighborhoods in downtown Los Angeles. This is required to define the scope of project, which is confined to downtown area.
- Latitude and longitude coordinates of those districts. This is required to plot the map and also get the venue data.
- Venue data, particularly data related to restaurants. This data will be used to perform clustering on neighborhoods.

Sources of Data

This Wikipedia page (https://en.wikipedia.org/wiki/Downtown_Los_Angeles#Districts) contains a list of neighborhoods or districts in downtown Los Angeles, with the total of 17 districts. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and BeautifulSoup packages. Then we will get the geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.

After that, we will use Foursquare API to get the venue data for those neighborhoods.

Foursquare has one of the largest databases of places and is used by lots of developers. Foursquare API (<https://foursquare.com/>) will provide many categories of the venue data, we are particularly interested in the Restaurant category in order to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.

Methodology

Firstly, we need to get the list of neighborhoods in downtown Los Angeles. The list is available in the Wikipedia page (https://en.wikipedia.org/wiki/Downtown_Los_Angeles#Districts). We will do web scraping using Python requests and beautifulsoup packages to extract the list of neighborhoods. However, this is just a list of names. We need to get the geographical coordinates in the form of latitude and longitude in order to be able to use Foursquare API. To do so, we will use the Geocoder package that will allow us to convert address into geographical coordinates in the form of latitude and longitude. Then we will transfer this data into a Pandas DataFrame and visualize the neighborhoods in a map using Folium package. This allows us to make sure that the geographical coordinates data returned by Geocoder are correctly plotted in downtown Los Angeles. Next, we will use Foursquare API to explore the neighborhoods and get the top 100 venues that are within a radius of 500 meters¹. We need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical coordinates of the neighborhoods in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. With the data, we can check how many venues were returned for each neighborhood and examine how many unique categories can be curated from all the returned venues. Since, we are analyzing the “Restaurant” data, we will filter the data by extracting the venue categories which contains ‘*Restaurant*’ in their venue names. Then, we will analyze each neighborhood by grouping the rows by neighborhood and taking the total frequency of occurrence of restaurants. By doing so, we are also preparing the data for use in clustering. Lastly, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. It identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. We will cluster the neighborhoods into 3 clusters based on their frequency of occurrence for “Restaurant”. The results will allow us to identify which neighborhoods have fewer number of restaurants. The occurrence of restaurants in different

¹ The 500-meter radius may seem a short distance, but since neighborhoods are closely located in downtown area, a shorter distance was picked to easily distinguish between neighborhoods with their venues.

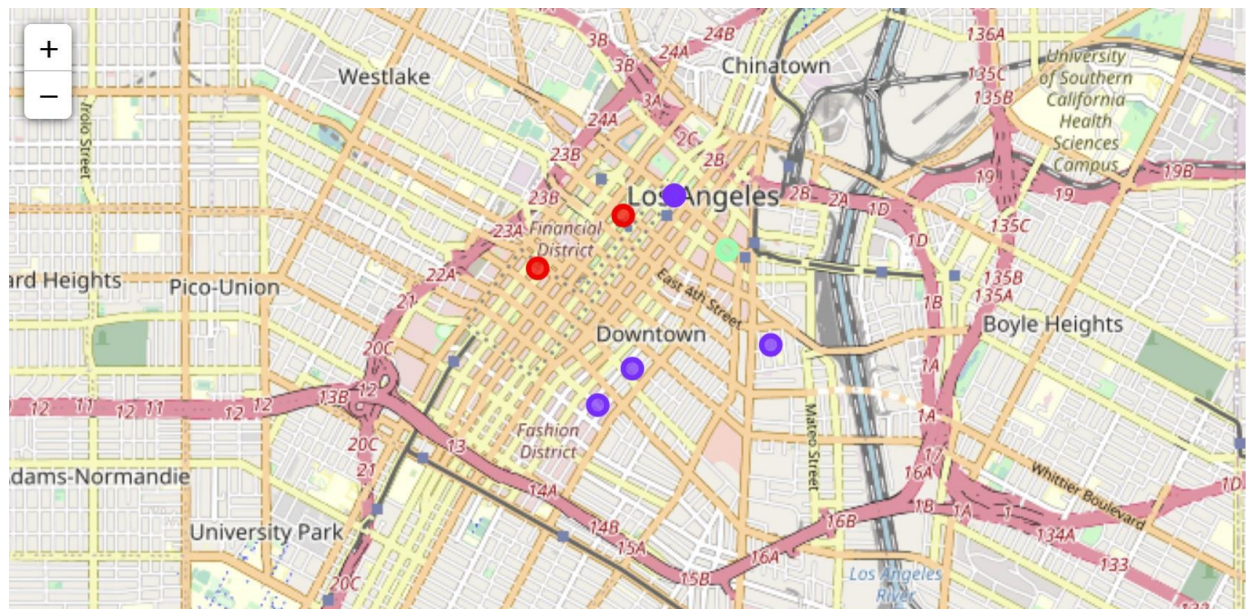
neighborhoods will help us to answer the question as which neighborhoods are most suitable to open new restaurant.

Results

The results from the k-means clustering show that we can categorize the neighborhoods into 3 clusters based on the frequency of occurrence for “Restaurant”:

- Cluster 0: Neighborhoods with moderate number of restaurants
- Cluster 1: Neighborhoods with low number of restaurants
- Cluster 2: Neighborhoods with high concentration of restaurants

The results of the clustering are visualized in the map below with cluster 0 in red color, cluster 1 in purple color, and cluster 2 in mint green color.



Discussion

As the results show, most of the restaurants are concentrated in Financial District with the total number of 35 restaurants and in Little Tokyo with the total number of 33 restaurants (in cluster 2). Cluster 0 has moderate number of restaurants between 16 to 21 in Bunker Hill, South Park, and Toy District neighborhoods. On the other hand, cluster 1 has very low number of restaurants between 1 to 12 in nine other neighborhoods (Arts District, Civic Center, Fashion District, Flower District, Historic Core, Industrial District, Jewelry District, Skid Row, and Wholesale District). This represents a great opportunity and high potential areas to open new restaurants. Therefore, this project recommends property developers to capitalize on these findings to open

new restaurants in neighborhoods in cluster 1. Property developers can also open new restaurants in neighborhoods in cluster 0 with moderate competition, but they are advised to avoid neighborhoods in cluster 2 which already have high concentration of restaurants.

Conclusion

In this project, we have gone through the process of identifying a business problem and the data required, extracting and preparing the data, performing machine learning by clustering the data into 3 clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders i.e. property developers regarding the best locations to open a new restaurant. The answer to the business question is: The neighborhoods in cluster 1 are the most preferred locations to open new restaurants. The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new restaurant.

References

Downtown Los Angeles, Geography, Districts in Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Downtown_Los_Angeles#Districts.