Opening a New Hotel in LA

Dong-Geun Kim. July 2020

1. Introduction

LA is the one of the most popular city for travelers. There are lots of places to visit such as Hollywood, Disneyland and Santa Monica beach and plenty of fine restaurants. As a result, opening a hotel allows property developers to earn significant income. Opening a new hotel is very complex problems and requires serious consideration. Especially, the location of the hotel is the one of the most important decision that would be determined.

1-1. Business Problems

The objective of this capstone project is to recommend the best location in LA to open a new hotel. Using data science and machine learning techniques, this project aims to provide solutions to the business problem.

1-2. Taget Audience

This project would be useful to property developers and investors looking for a place to open a new hotel in LA. Of course, traveling industry has had very hard time because of COVID-19. But it is needed more than 3 years to build a brand new hotel and traveling demand would be recovered after we overcome the desease. So it is very important to find a right place for a hotel and te be ready for the hay days.

2. Data

To solve the problem, I will use the following data

- List of neighborhoods in LA
- Latitude and longitude of the neighborhoods
- Venue data related to hotels

Data Source

Wikipedia contains a list of neighborhoods of LA(https://en.wikipedia.org/wiki/List_of_districts_and_neighborhoods_of_Los_Angeles). I will use web scrolling techniques to extract data from the page, with Python requests and Beautiful Soup packages. Then I will match the geographical data with the venue data from Foursquare API. This project will require lots of data analytics skills such as web scrolling, data cleansing, data visualization, and machine learning(K-mean clustering).

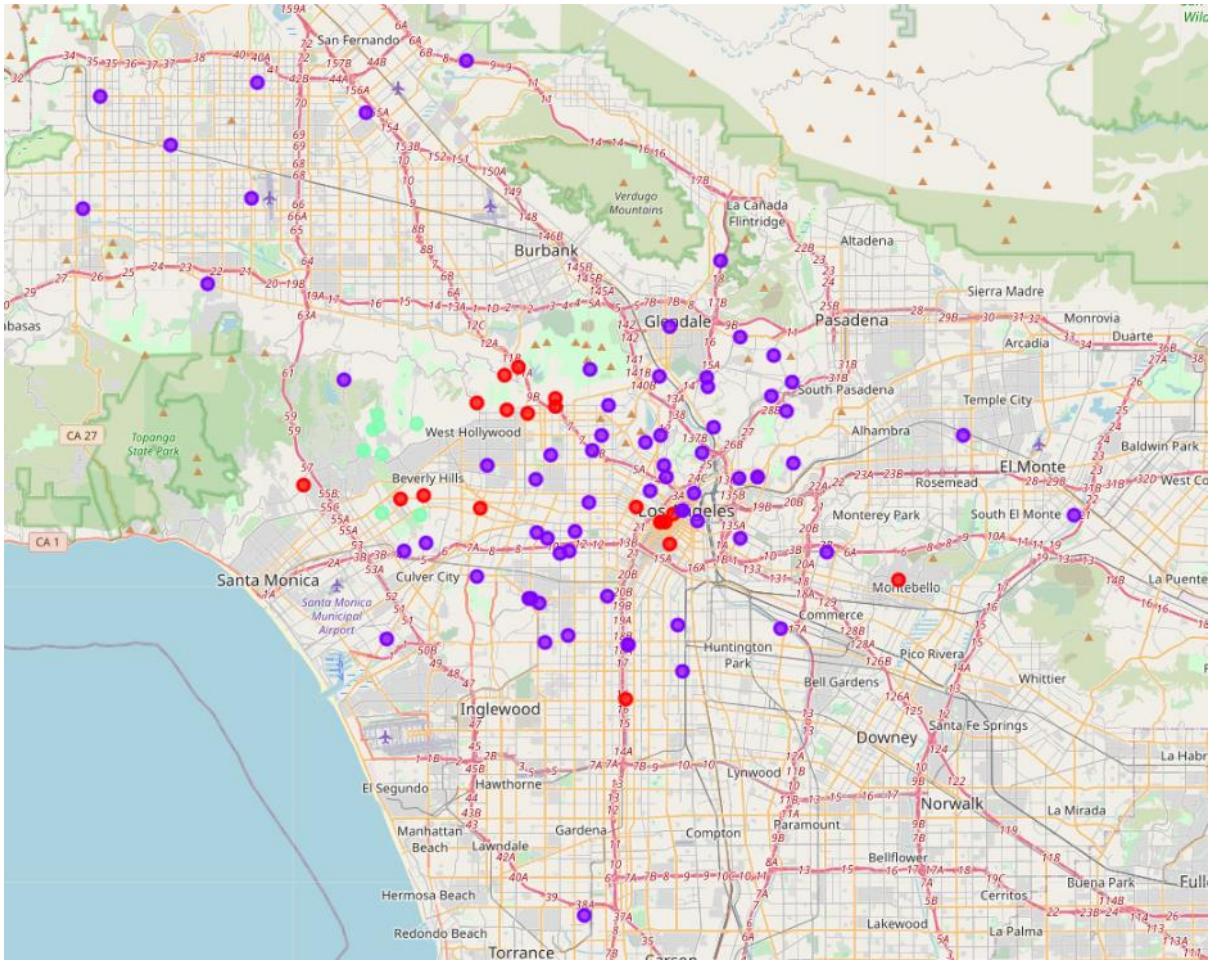
3. Methodology

First of all, I need to get a list of neighborhoods in LA from the Wikipedia(https://en.wikipedia.org/wiki/List_of_districts_and_neighborhoods_of_Los_Angeles). By using Python requests and beatifulsoup packages, I'm able to bring the list of neighborhoods. Geocoder package makes me convert address into geographical coordinates in the form of latitude and longitude. After gathering the data, I can put the data into a pandas dataframe and visualize the neighborhoods on LA map. Second, I will use Foursquare API to get the top 100 venues that are within a radius of 1000 meters. Foursquare will return the venue data in JSON format and I will extract the venue name, venue category, and geographical data. Before I conduct clustering analysis, I can show how many venues are returned for each neighborhood and analyze characteristics of each neighborhood.

At last, I will do clustering analysis by using k-mean clustering. In this case, I will cluster the neighborhoods into 3 clusters based on each neighborhood's characteristics. Then I can show which neighborhoods have higher concentration of hotels and which neighborhoods have fewer number of hotels. Based on the result, I will answer the question that which neighborhoods are more suitable to open a new hotel in LA.

4. Results

Most of hotels are concentrated in the western part of LA, with the hightest number in cluster 2. Also, downtown and sightseeing areas hold some hotels when we see cluster 0. On the other hand, cluster 1 has very low number of hotels in the area. We can guess that this area is residential area.
Hotels in cluster 2 are likely to suffer from intense competition due to over supply and high concentration of hotels. Also, Hotels in cluster 1 would face low demand of hotels. Therefore, this project recommends property developers to open new hotels in neighborhoods in cluster 1 with moderate competition.



Cluster Labels	Hotel
0	0.034436
1	0.002121
2	0.088645