

# **Coursera Capstone**

IBM Applied Data Science Capstone

Opening A Bubble Tea Shop in Los Angeles

By: Ning Chu

August 2020



# Introduction:

Los Angeles often known by its initials L.A., is the largest city in California. With nearly four million people, it is the second-most populous city in the United State. Los Angeles is known for its Mediterranean climate, ethnic diversity, Hollywood entertainment industry, and its sprawling metropolis.

Many Taiwanese immigrants settled in California, leading to a number of bubble tea shops opening around Los Angeles. Bubble tea, popularly known as boba tea, is a Taiwanese drink. It is prepared with tea or milk as the base component. This tea is glazed with boba, which are soft, chewy, and gummy tapioca pearls extracted from cassava root. Since the '90s, Los Angeles has a strong passion for milk tea studded with chewy pearls of boba. From Downtown to West Hollywood, Los Angeles has a large variety of boba houses offering both customizable teas and classic favorites. According to a new report published by Allied Market Research, the bubble tea market size was valued at \$2.4 billion in 2019 and is estimated to reach \$4.3 billion by 2027. Therefore, the analysis and results of this project would interest stakeholders who are interested in opening a bubble tea shop in Los Angeles.

## Business Problem

The objective of this Capstone project is to analyze and select the best locations in the Los Angeles to open a new Bubble Tea shop. Using Data Science methodology and instruments such as Data Analysis and Visualization, this project aims to provide solutions to answer the business question: Where in LA, should the investor open a Bubble Tea shop?

## Data

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

- Los Angeles data containing the neighborhoods and boroughs.
- Venue data, particularly data related to restaurants. This data will be used to perform further analysis of the neighborhoods
- Latitude and longitude coordinates of those neighborhoods. This is required to plot the map and get the venue data.
- Average neighborhood rent

### Data Source

- List of all neighborhoods in LA — [https://en.wikipedia.org/wiki/List\\_of\\_districts\\_and\\_neighborhoods\\_of\\_Los\\_Angeles](https://en.wikipedia.org/wiki/List_of_districts_and_neighborhoods_of_Los_Angeles)
- Coordinates of all neighborhoods and venues — GeoPy Nominatim geocoding
- Number of restaurants and their type and location in every neighborhood — Foursquare API
- LA rent data - <https://www.rentcafe.com/average-rent-market-trends/us/ca/los-angeles/>

# Problem Statement

- What is / are the best location(s) for opening a Bubble Tea shop in LA?
- In what Neighborhood and/or borough should the investor open a Bubble Tea Shop to have the best chance of being successful?
- Where would I go in LA to have the best Bubble Tea?

## Methodology

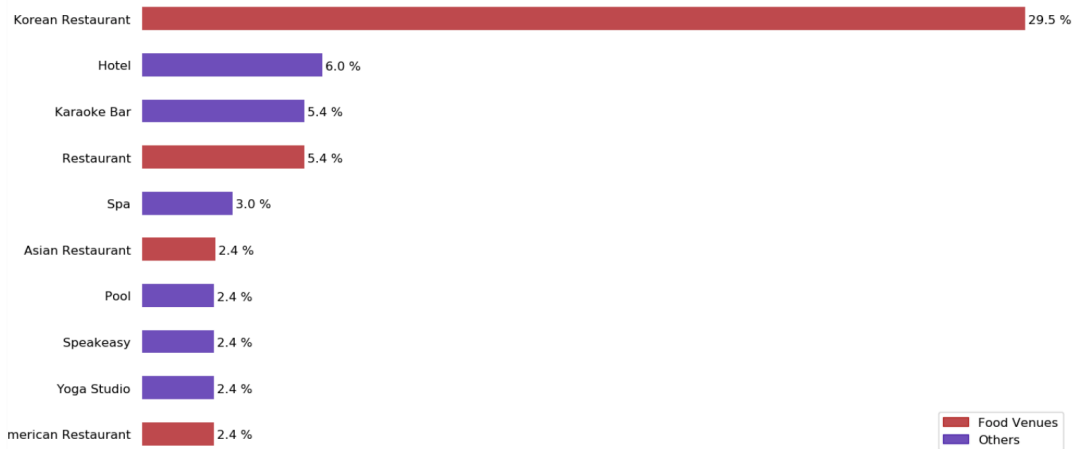
1. Collect data on the neighbourhoods of Los Angeles from the internet. There are no relevant datasets available for this and therefore, so we will need to be scraped from a webpage and processed into a dataframe. The location coordinates of each neighbourhood will be used in GeoPy Nominatim geolocator and appended to the neighbourhood data.
2. Foursquare be used to explore each of neighbourhoods and their venues using Foursquare location data. The venues of the neighbourhoods will be analyzed in detail and patterns will be discovered. This discovery of patterns will be carried out by grouping the neighbourhoods using k-means clustering.
3. Each cluster will be examined and a decision will be made regarding which cluster fits the shareholder's requirements.
4. Finally, if there are multiple neighbourhoods that fit these conditions, Los Angeles rent data can be used to influence the shareholder's decision. The results of the analysis will highlight potential neighbourhoods where a Bubble Tea Shop may be opened based on geographical location and proximity to competitors.
5. Data of Bubble Tea Shop will be sorted based on rankings in order to find where people go in can have the best Bubble Tea Shop.

## Analysis

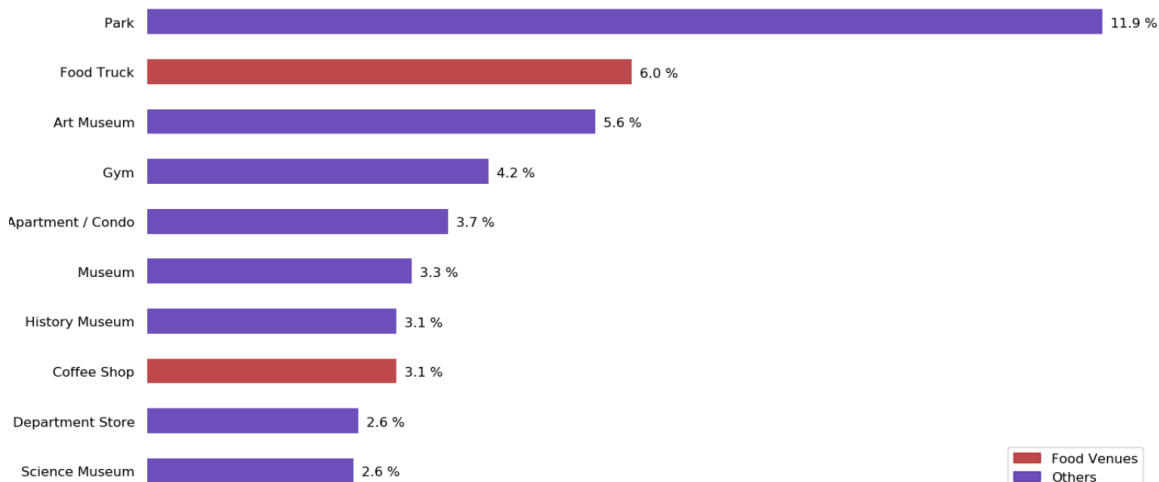
- Importing Libraries
- Web Scraping Neighbourhood Data
- Loading and Cleaning Neighbourhood
- Obtaining Neighbourhood Coordinates
- LA Neighbourhood Map
- Defining Foursquare Credentials and Version
- Exploring the first Neighbourhood
- Exploring all Neighbourhoods
- Analyzing each Neighbourhood
- Clustering Neighbourhoods
- Examining the Clusters
- Visualizing Top 10 Venues for each Cluster
- Investigating the chosen Cluster
- Web Scraping Rent Data

# Result

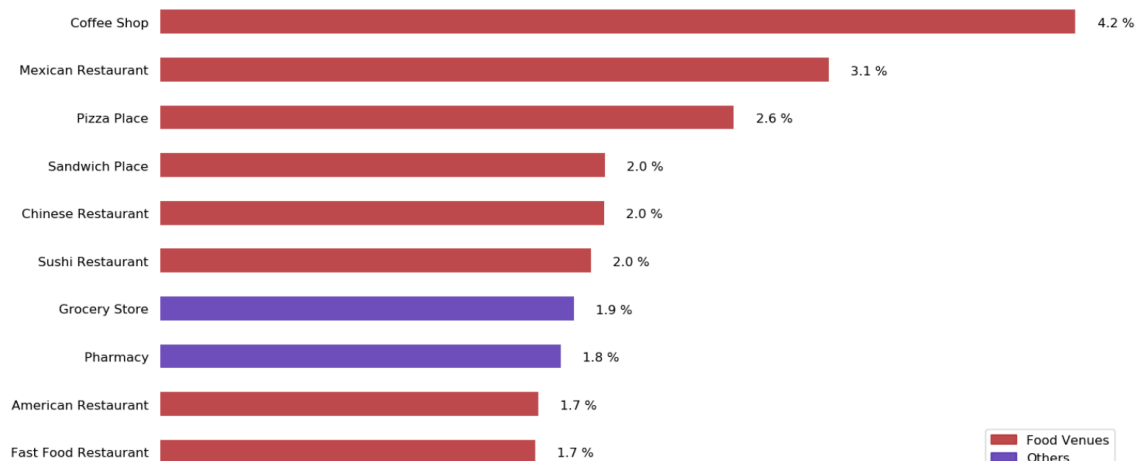
Ten Most Prevalent Venues of Cluster 1  
(in % of all venues)



Ten Most Prevalent Venues of Cluster 2  
(in % of all venues)



Ten Most Prevalent Venues of Cluster 3  
(in % of all venues)



1.It's noticeable that coffee shop is the most prevalent venue in Cluster 3, so we have to exclude those location spot in Cluster 3.Moreover, In Cluster 1, the ratio of Korean Restaurant is nearly 1/3, also 4 food venues are most prevalent in this group.As for Cluster 2, it is our best option, this is because it looks like the place might full of tourist ex: any kind of museums, department stores. Furthermore, the Food Truck is the second most prevalent venues in Cluster2 plus Coffee Shop will not bring out our customers because the ration of it only 3 %.Therefore, we choose Cluster 2 to be our best bubble tea shop location.

2.The best Bubble Tea shop is Pearl's Finest Teas in Mid City which has 8.8 score but it is in the Cluster 1, except for knowing which bubble tea shop is the hottest, it also remind us that we should open in other place because its rent is super expensive that we need to consider about our budget.

## Discussion

All of the above analysis is depended on the accuracy of Four Square data.

To get better results, future research work and more comprehensive analysis could consider using a paid account to bypass these limitations as well as incorporating data from other external databases. In addition, I got a problem of over quota in Four Square API because I forgot to download csv, so remind you guys to remember downloading the data from API

## Conclusion

The objective of this project was to identify the best potential neighbourhoods in Los Angeles where a bubble milk tea shop can be set up. All the required neighbourhood data was either scraped of the internet or obtained using a geolocator. After the neighbourhoods were visualized on a folium map, their venues were explored using Foursquare location data. Based on the frequency of occurrences of different venue types, the neighbourhoods were divided into four groups with the help of k-means clustering. The clusters were examined and the best one in which a restaurant could be set up was chosen.Average neighbourhood rent data was called upon and while it provided interesting insights, it could not influence the decision only with the information at hand.

As touched upon earlier, the results of the analysis highlight potential neighbourhoods where a bubble tea shop may be opened solely based on geographical location with many entertainment venues. This will only serve as a starting point in the overall investigation since there are a lot of other factors - availability of commercial spaces, appeal of each location, proximity to major roads, access through public transport, etc. - that influence such a decision.