Best Neighborhood for a Szechuan Restaurant
Rui Ma
Feb 1 2021

Introduction
In this hypothetical scenario, a successful Chinese restaurant is looking to expand their business into a new city in the United States.

This Szechuan restaurant was opened up in 2013 at a college town and sooner become one of most popular Chinese restaurants in town. Recently, the owners would like to expand the family business to West Coast. Los Angeles, California is the first choice.

The owners had conducted market research and is confident that the taste of their restaurant would be the dominator in town. However, not knowing the city well, the owners have difficulties to identify the best neighborhood to open their business.

Based on the family's experiences with their restaurant, seeking out an area that already attracts diners who enjoy Chinese cuisine would be a good indicator for the location of the new restaurant.

Approach
For the purposes of this project this problem will be investigated using the Foursquare API to identify which neighborhoods meet the preferences outlined by the client.

Data Collection

Part 1 - Data Preparation: Neighborhoods
In order to utilize the Foursquare API the following information was needed:

- A list of Los Angeles neighborhoods
- The post codes of each neighborhood
- Latitude and longitude co-ordinates for each post code.

This data was found at the following website: https://www.geonames.org/postalcode-search.html?q=Pittsburgh%2C+Pennsylvania&country=US

The data was scrapped and then cleaned through a series of steps into a data frame of 4 features, Neighborhood, Postcode, Latitude, and Longitude.

The Folium library was then used as a visual aid while examining the data, rendering maps of Los Angeles's neighborhoods. Later in the analysis folium was be used to show the different venue clusters that were identified using machine learning methods. Once the neighborhood data was arranged into a workable data frame, The Foursquare API software identified:

- The different types of venues (e.g., restaurants, parks) in each neighborhood.
- The popularity of each venue in each of these neighborhoods.
- The focus of our search will be on the " Food " category, specifically on Chinese cuisine.

The following steps show the data cleaning steps needed to present a data frame of the relevant information needed. This data frame should have 4 features "PostCode", "Neighborhood", "Latitude", "Longitude".

To do this the API to build a list of venues of up to 100 candidates, within a 500-metre radius of each of Los Angeles's neighborhood centroids. The number of candidates was limited to 100 as that is the maximum number allowed as a guest user of this website.

The venue search include only data from the "Food" and "Chinese restaurant" categories (Table: 1). The category hierarchy ID codes were taken from:
https://developer.foursquare.com/docs/buildwith-foursquare/categories/

	Subcategory	ID
1	Anhui Restaurant	52af3a5e3cf9994f4e043bea
2	Beijing Restaurant	52af3a723cf9994f4e043bec
3	Cantonese Restaurant	52af3a7c3cf9994f4e043bed
4	Cha Chaan Teng	58daa1558bbb0b01f18ec1d3
5	Chinese Aristocrat Restaurant	52af3a673cf9994f4e043beb
6	Chinese Breakfast Place	52af3a903cf9994f4e043bee
7	Dim Sum Restaurant	4bf58dd8d48988d1f5931735
8	Dongbei Restaurant	52af3a9f3cf9994f4e043bef
9	Fujian Restaurant	52af3aaa3cf9994f4e043bf0
10	Guizhou Restaurant	52af3ab53cf9994f4e043bf1
11	Hainan Restaurant	52af3abe3cf9994f4e043bf2
12	Hakka Restaurant	52af3ac83cf9994f4e043bf3
13	Henan Restaurant	52af3ad23cf9994f4e043bf4
14	Hong Kong Restaurant	52af3add3cf9994f4e043bf5
15	Huaiyang Restaurant	52af3af23cf9994f4e043bf7
16	Hubei Restaurant	52af3ae63cf9994f4e043bf6
17	Hunan Restaurant	52af3afc3cf9994f4e043bf8
18	Imperial Restaurant	52af3b053cf9994f4e043bf9
19	Jiangsu Restaurant	52af3b213cf9994f4e043bfa

	Subcategory	ID
20	Jiangxi Restaurant	52af3b293cf9994f4e043bfb
21	Macanese Restaurant	52af3b343cf9994f4e043bfc
22	Manchu Restaurant	52af3b3b3cf9994f4e043bfd
23	Peking Duck Restaurant	52af3b463cf9994f4e043bfe
24	Shaanxi Restaurant	52af3b633cf9994f4e043c01
25	Shandong Restaurant	52af3b513cf9994f4e043bff
26	Shanghai Restaurant	52af3b593cf9994f4e043c00
27	Shanxi Restaurant	52af3b6e3cf9994f4e043c02
28	Szechuan Restaurant	52af3b773cf9994f4e043c03
29	Taiwanese Restaurant	52af3b813cf9994f4e043c04
30	Tianjin Restaurant	52af3b893cf9994f4e043c05
31	Xinjiang Restaurant	52af3b913cf9994f4e043c06
32	Yunnan Restaurant	52af3b9a3cf9994f4e043c07
33	Zhejiang Restaurant	52af3ba23cf9994f4e043c08

At the end of this process, we get a data frame of 7 features and 985 rows, comprising of 29 unique categories.

Part 1 - Data Preparation: Income

The client wishes to attract returning patrons who are seeking a high-end dining experience. An initial way to explore this question would be to look at household incomes in the potential neighborhoods, as household incomes (or disposable incomes) may provide an indication of the frequency potential customers might seek to dine out. Higher earners may be more inclined to seek out the more exclusive dining experiences the client wishes to provide in her restaurant.

Median household income data was taken from U.S. Census Bureau. The table, "MEDIAN INCOME IN THE PAST 12 MONTHS (IN 2019 INFLATION-ADJUSTED DOLLARS)" were downloaded from http://data.census.gov.

The final data frame consisted of 5 features: income, postcode, latitude, longitude and neighborhood

	PostCode	Income	Neighbourhoods	Latitude	Longitude
72	90250	31905.0	Hawthorne	33.914	-118.349
41	90044	29029.0	Los Angeles	33.955	-118.290
43	90046	29025.0	Los Angeles	34.107	-118.365
24	90026	26045.0	Los Angeles	34.077	-118.265
56	90066	25785.0	Los Angeles	34.003	-118.430
31	90034	25248.0	Los Angeles	34.029	-118.400
114	91342	25121.0	Sylmar	34.305	-118.432
17	90019	24464.0	Los Angeles	34.048	-118.334
9	90011	24433.0	Los Angeles	34.008	-118.258
113	91335	24391.0	Reseda	34.201	-118.539
112	91331	23447.0	Pacoima	34.256	-118.421

Method

Part 1:

The analysis began with determining the density of restaurants in each neighborhood, followed by grouping them into which venues were the most popular. In this case the top 10 venues for each neighborhood were presented.

The next step was to employ K-means clustering to organize the top venues into clusters of neighborhoods based on the most common venues. The clusters were displayed using Folium, and tables generated for each cluster.

Part 2:

The neighborhoods from the selected cluster were further investigated by analyzing the household income and Neighborhood:

The data frame of median household incomes was investigated with basic statistics to determine the mean incomes for each of the selected neighborhoods found in the "Part 1 - Analysis".

Results and Discussion

Foursquare revealed a total of 62 unique FOOD categories were identified in the 200 LA neighborhoods that make up the greater LA City region. This data set was reduced to just the top 10 venues food venues and the frequency in which they appeared in each neighborhood.

K-means clustering reduced the neighborhoods into 6 clusters with the 4th cluster grouping eight neighborhoods: 'Beverly Hills', 'Los Angeles', 'Chatsworth', 'Sherman Oaks', 'Studio City', 'Gardena', 'Cerritos', 'Lomita' (Figure 1).

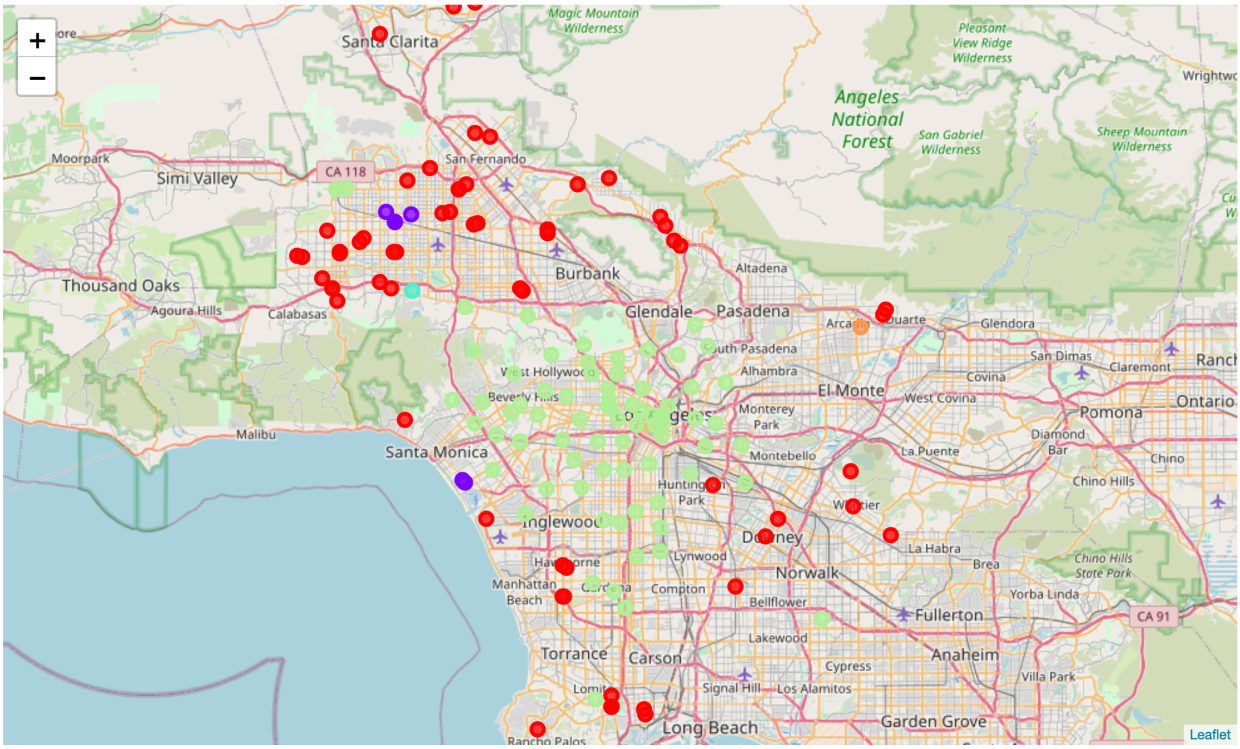

Figure 1

Cluster 2 revealed the category “Chinese Restaurant” as the most popular venue for people to visit (Table 2). A refined search of these six neighborhoods showed that the density of Szechuan Restaurant is at only 9th most common venue in all six neighborhoods. This indicated that there would be considerably low direct competition from Szechuan cuisine with high density of Chinese restaurant in these neighborhoods

	Neighbourhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Beverly Hills	Chinese Restaurant	Deli / Bodega	Thai Restaurant	Food Truck	Breakfast Spot	Bubble Tea Shop	Buffet	Café	Cantonese Restaurant	Coffee Shop
1	Los Angeles	Chinese Restaurant	Cantonese Restaurant	Taiwanese Restaurant	Market	Japanese Restaurant	Thai Restaurant	Dim Sum Restaurant	Asian Restaurant	Szechuan Restaurant	Breakfast Spot
2	Los Angeles	Chinese Restaurant	Cantonese Restaurant	Taiwanese Restaurant	Market	Japanese Restaurant	Thai Restaurant	Dim Sum Restaurant	Asian Restaurant	Szechuan Restaurant	Breakfast Spot
3	Los Angeles	Chinese Restaurant	Cantonese Restaurant	Taiwanese Restaurant	Market	Japanese Restaurant	Thai Restaurant	Dim Sum Restaurant	Asian Restaurant	Szechuan Restaurant	Breakfast Spot
4	Los Angeles	Chinese Restaurant	Cantonese Restaurant	Taiwanese Restaurant	Market	Japanese Restaurant	Thai Restaurant	Dim Sum Restaurant	Asian Restaurant	Szechuan Restaurant	Breakfast Spot
5	Los Angeles	Chinese Restaurant	Cantonese Restaurant	Taiwanese Restaurant	Market	Japanese Restaurant	Thai Restaurant	Dim Sum Restaurant	Asian Restaurant	Szechuan Restaurant	Breakfast Spot
6	Los Angeles	Chinese	Cantonese	Taiwanese	Market	Japanese	Thai	Dim Sum	Asian	Szechuan	Breakfast

Table 2

Conclusion

The purpose of this project was to find neighborhoods within Los Angeles where Szechuan cuisine is rated as relatively less common venues sort out by all other cuisines. Using Foursquare data, the search was reduced to a cluster of six areas, Beverly Hills, Los Angeles, Chatsworth, Sherman Oaks, Studio City, Gardena, Cerritos and Lomita. Foursquare also revealed a relatively light density of Szechuan cuisine restaurants in all these neighborhoods.

Based on the findings of this investigation I would recommend the client consider either the neighborhoods of Chatsworth, Sherman Oaks, Studio City or Cerritos as possible locations for her new restaurant. All these locations are the most centrally located and are on the upper end of household incomes. Further information beyond the scope of this report would assist the client decide on the best location to open her first restaurant.