

Opening Upscale Italian Restaurants in Los Angeles

Introduction/Business Problem

Company A is planning to open several upscale Italian restaurants in Los Angeles. They are now deciding the areas that are suitable for their new restaurants. In so doing, the company creates a set of criteria for the areas, sorted by their importance:

First, the areas should have a high median income. The company is targeting high-income customers, and thus prefer to open its restaurants near the areas where its target customers live. In addition, locating the restaurants in such areas will also help the branding of the restaurants.

Second, the areas should not have many Italian restaurants around it. The company wants to catch business opportunities in areas where there are not many competitors. The company hopes to be a pioneer in the areas and thus can attract more customers.

Third, the areas should have a fairly large population size, since the main target customers will be the people who live in the areas. By opening restaurants in areas with a sizeable population, the company hopes to have more customers.

To solve this area selection problem, I decided to help company A by clustering Los Angeles areas based on the criteria mentioned above using the K-means method.

Data

The following is the descriptions of data used in this project and also the sources:

1. Los Angeles median income by ZIP code (2017 data):

I divide Los Angeles into areas based on their ZIP codes. I then retrieve the data on the median income of these areas (in US\$) from the website of Los Angeles Almanac (<http://www.laalmanac.com/employment/em12c.php>)

2. Los Angeles population by ZIP code:

I measure the size of the areas by using the number of people living in the area. I retrieve the data from ZIPAtlas.com

(<http://zipatlas.com/us/ca/los-angeles/zip-code-comparison/population-density.htm>)

3. Italian restaurant by ZIP code:

I use the data on the number of Italian restaurants within the radius of 1 kilometer from the centre of the ZIP code area/neighborhood to measure the competitiveness in the areas. I retrieve the data from Foursquare location data.

Methodology

I retrieve the data on median income and population size from the aforementioned sources. I then use the Foursquare API to get information about venues within the radius of 1 kilometer (1,000 meters) from the areas. After getting this data, I then group them based on the area and get the count number of each venue category. Since I am only interested in the data on Italian restaurants, I remove information related to other venue types. At this point, I have the number of Italian restaurants within a radius of 1 kilometer for each area. I then merge this information with the table of median income and population size.

I use the K-means method to cluster the areas. This method is one of the most common clustering methods and is suitable for this project. This method will cluster the areas based on the criteria that I have specified: the median income of the population, the size of the population and the number of Italian restaurants within a radius of 1 kilometer from the center of the area. Before I apply the method, I standardize all the columns in the table I have created using StandardScaler.

I set the number of clusters to five, which will allow me to group the neighborhood based on the following level of desirability: High, Mid-High, Mid, Low-Mid and Low.

Results

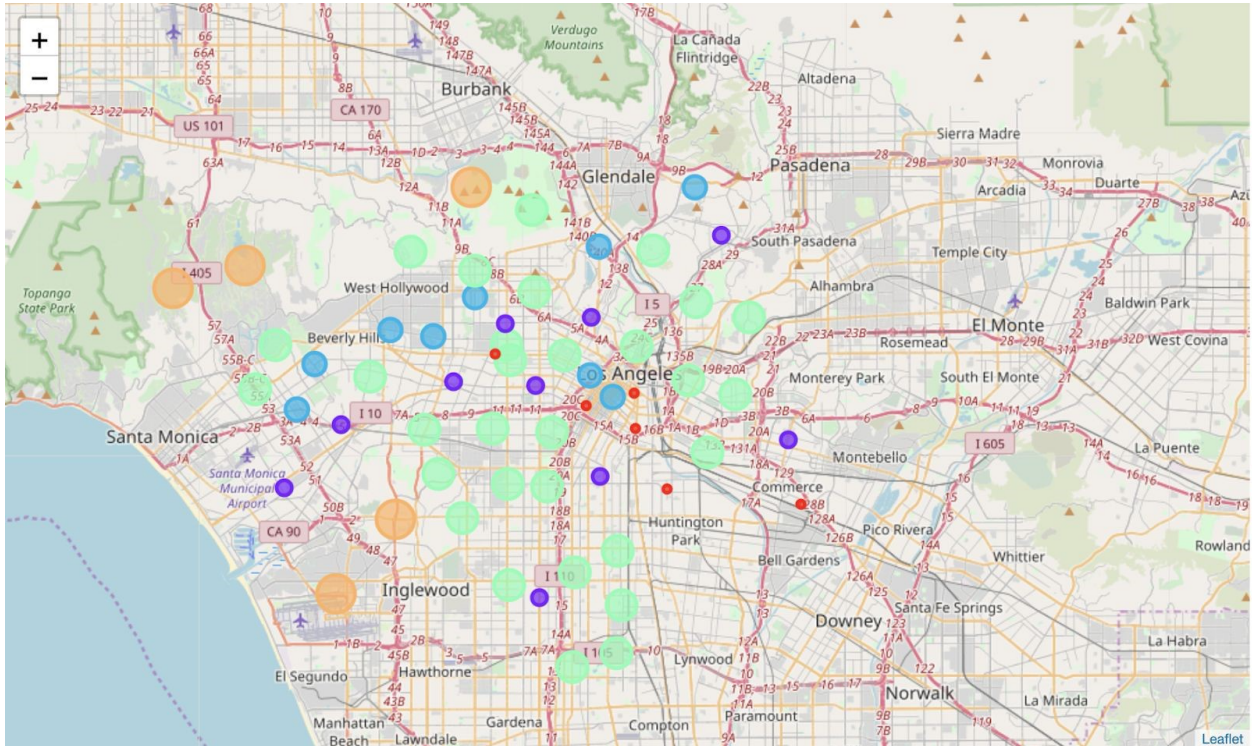
The K-Means clustering gives the following results for the clusters of the areas in Los Angeles:

	Cluster Labels	Population	Median Income	Italian Restaurant	Desirability	Number of Neighborhood
0	0	7204.833333	29296.000000	0.666667	Low	6
1	1	70711.700000	48913.900000	0.700000	Mid-Low	10
2	2	21728.444444	70406.111111	3.000000	Mid	9
3	3	43232.166667	43308.900000	0.300000	Mid-High	30
4	4	22624.200000	112789.600000	0.000000	High	5

The cluster with the highest desirability (i.e. the cluster with a high median income and population size but a low number of median income) is the cluster number 4. The table shows that this cluster consists of 5 areas. The average number of population is 22,624, which is much lower than that of clusters of 1 and 3. However, the median income of this cluster stands out at \$112,789, which is significantly higher than the rest of the clusters. In addition, there is no Italian restaurant around the areas in this cluster. The combination of these characteristics leads to the high desirability of this cluster.

On the other hand, cluster 0 is the least desirable cluster among the 5 clusters. This cluster consists of 6 neighborhoods with an average size of population of just 7,205 people. Even though the average number of Italian restaurants is less than 1 for this cluster, the median income is relatively lower at \$29,296.

The clusters are visualized by the following map:



The size of the circles represents the desirability of the areas. The areas with an orange circle are the most desired whereas those with a red circle are the least desired. The map shows that the desired areas are located in the north-west and south-west part of Los Angeles.

	Cluster Labels	Neighborhood	Zip	Name	Latitude	Longitude	Population	Median Income	Italian Restaurant
51	4	90056 - Ladera Heights	90056	Ladera Heights	33.987945	-118.370442	8108	84438	0
53	4	90045 - Los Angeles(Los Angeles International ...	90045	Los Angeles(Los Angeles International Airport,...	33.954017	-118.402447	39315	90399	0
54	4	90068 - Los Angeles(Hollywood)	90068	Los Angeles(Hollywood)	34.137411	-118.328915	21713	82718	0
55	4	90049 - Los Angeles(Bel Air Estates, Brentwood)	90049	Los Angeles(Bel Air Estates, Brentwood)	34.091829	-118.491244	33520	121671	0
58	4	90077 - Los Angeles(Bel Air Estates, Beverly G...	90077	Los Angeles(Bel Air Estates, Beverly Glen)	34.102084	-118.451629	10465	184722	0

The table above shows the areas in cluster 4, the most desirable cluster. There is no Italian restaurant within a radius of 1 kilometer from all the five neighborhoods in the cluster. The population sizes of these neighborhood range from 8,108 people in Ladera Heights to 39,315 people in the LA International Airport area. Lastly, the median income also varies, from \$84,438 for Ladera Heights to 184,722 for Bel Air Estates - Beverly Glen.

Overall, all the areas in this cluster are characterized by the non-existence of nearby Italian restaurants, a modest population size and a high median income.

	Cluster Labels	Neighborhood	Zip	Name	Latitude	Longitude	Population	Median Income	Italian Restaurant
30	0	90013 - Los Angeles(Downtown Central, Downtown...	90013	Los Angeles(Downtown Central, Downtown Fashion...	34.044639	-118.240413	9727	22808	1
41	0	90015 - Los Angeles(Downtown Fashion District, ...	90015	Los Angeles(Downtown Fashion District, South Pa...	34.039224	-118.266293	15134	32979	1
52	0	90010 - Los Angeles(Hancock Park, Wilshire Cen...	90010	Los Angeles(Hancock Park, Wilshire Center, Win...	34.062125	-118.315709	1943	47115	1
56	0	90040 - Commerce, City of	90040	Commerce, City of	33.994524	-118.149953	9798	43585	0
57	0	90021 - Los Angeles(Downtown Fashion District, ...	90021	Los Angeles(Downtown Fashion District, Downtow...	34.029043	-118.239504	3003	12864	1
59	0	90058 - Los Angeles(Southeast Los Angeles), Ve...	90058	Los Angeles(Southeast Los Angeles), Vernon	34.001617	-118.222274	3624	16425	0

The table above shows the areas in cluster 0, the least desirable cluster. In four out of six areas in the cluster, there is already an Italian restaurant within a radius of 1 kilometer from the center. The population sizes of these areas are also relatively small, with the largest area having only 15,134 people in it. The median incomes are also significantly smaller compared to those in cluster 4. The highest median income in the cluster is \$47,115, which is only slightly higher than half of the lowest median income in cluster 4.

Discussion

Based on the results presented above, it is clear that the company should open their restaurants in the areas that make up cluster 4. The reason being these areas fulfill the criteria set by the company: high median income, decent size of population and less or no competition. These areas are Ladera Heights, the LA Airport Area, Hollywood, Bel Air Estates - Brentwood and Bel Air Estates - Beverly Glen. They are located in either the north-west or south-west part of Los Angeles and known for having affluent residents.

Areas in cluster 2 also are also similar to those in cluster 4 in terms of population size. In addition, the median income of these neighborhoods are on average higher than that of other clusters except for cluster 4. However, the competition is high in these areas, with an average of three Italian restaurants around each of them. On the other hand, clusters 3 and 1 are not as desirable as cluster 4 due to their low median income, despite the high

population and low competition. Lastly, areas in cluster 0 are the least desirable due to the combination of low population and median income.

Conclusion

In this project, I apply the K-Means clustering method to find the best areas in Los Angeles for opening new upscale Italian restaurants. I apply three criteria to determine the grouping and desirability of the areas: high median income, modest size of population and low competition (as measured by the number of Italian restaurants within a radius of 1 kilometer).

The results show that the areas with the highest level of desirability are Ladera Heights, the LA Airport Area, Hollywood, Bel Air Estates - Brentwood and Bel Air Estates - Beverly Glen. These areas are characterized by the three criteria mentioned above.