



# Opening Upscale Italian Restaurants in Los Angeles

IBM Data Science Capstone Project  
Andika Mulya



## Introduction/Business Problem

- Company A is planning to open several upscale Italian restaurants in Los Angeles.
- They are deciding on the areas, the criteria are:
  - the areas should have a high median income
  - the areas should not have many Italian restaurants around it
  - the areas should have a fairly large population size
- The solution is provided using the KMeans clustering method



## Data

The data is retrieved from the following sources:

- Los Angeles median income by ZIP code (2017 data):  
Los Angeles Almanac (<http://www.laalmanac.com/employment/em12c.php>)
- Los Angeles population by ZIP code:  
ZIPAtlas.com  
(<http://zipatlas.com/us/ca/los-angeles/zip-code-comparison/population-density.htm>)
- Italian restaurant by ZIP code:  
Foursquare API



## Methodology

- K-means method to cluster the neighborhoods
- Cluster the neighborhoods based on:
  - Median income
  - Population size
  - Number of Italian restaurants
- All the columns are standardized using StandardScaler



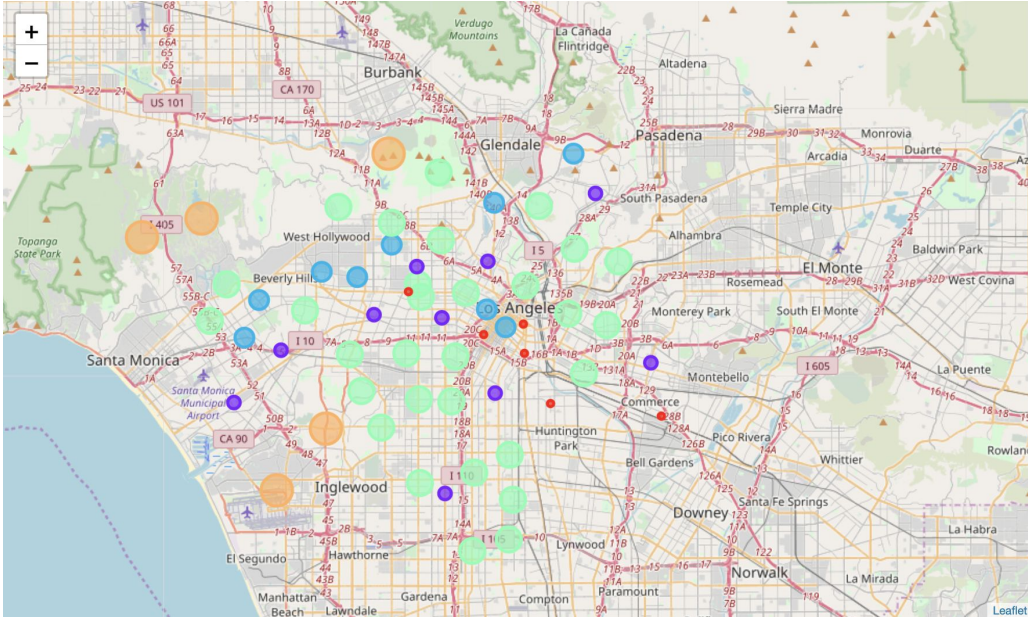
## Results

- The clusters and their characteristics are:

	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

- Cluster with the highest desirability is cluster 4

- Los Angeles map with the neighborhoods (bigger circles imply higher desirability)





## Results

- The neighborhoods in cluster 4:

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



## Results

- The neighborhoods in cluster 0:

	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 neighborhoods are characterized by a low number of population and low median income.





## Discussion

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



## Conclusion

- I apply the K-Means clustering method to find the best group of neighborhoods in Los Angeles for opening new upscale Italian restaurants.
- I apply three criteria to determine grouping and desirability of the neighborhoods: 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 neighborhoods with the highest level of desirability are Ladera Heights, the LA Airport Area, Hollywood, Bel Air Estates - Brentwood and Bel Air Estates - Beverly Glen.