# Potential Business Opportunities in

# Los Angeles County

**Introduction**

Los Angeles county is the most populous county in the United States with more than ten million inhabitants as of 2019. It has 88 incorporated cities and many un-corporate areas. Different culture, languages and cuisines also make Los Angeles County a international metropolis. Due to the population and broad area, there are a lot of business opportunities in this area. The city governments may think about how to improve the quality of life and convenience for the residents living in the cities. On the other hand, many companies are also thinking and looking for the business opportunities in Los Angeles area.

It could be hard for companies to make any business decisions if they do not have a big picture of what the cities actually need. In this project, we will use geographic information for each city to fetch neighborhood venues via Foursquare API. Based on the most popular venues information found in each city, we can compare the category of venues among the cities in Los Angeles county. To make the result more accurate, we will also include the data of city population, average age, household size and income. We will analyze and find out where the potential opportunities are for the cities in Los Angeles county or what kind of business can be successful in each city.

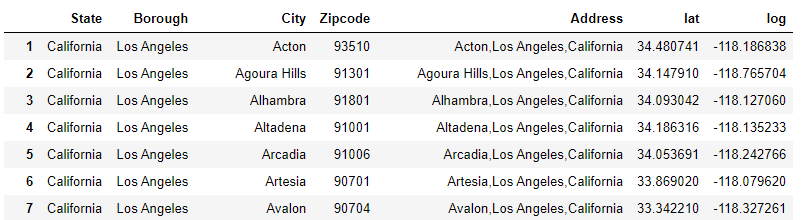
**Data Description**

The following data files are used for the analysis

1. [Los Angeles County City List](https://www.mapsofworld.com/usa/zipcodes/california/los-angeles/) - City and Zip code Mapping for the cities in Los Angeles County.
2. [Los Angeles County Population and Household File](https://data.lacity.org/Community-Economic-Development/2010-Census-Populations-by-Zip-Code/nxs9-385f) - Total Population, Total Male, Total Female, Total Household, Avg. Household for the cities of Los Angeles County.
3. [Los Angeles County Household Income](https://en.wikipedia.org/wiki/Demographics_of_Los_Angeles_County) - Medium Household Income and Medium Family Income for the cities of Los Angeles County
4. Neighborhood Venues from Foursquare based on coordinate of cities.

**Methodology**

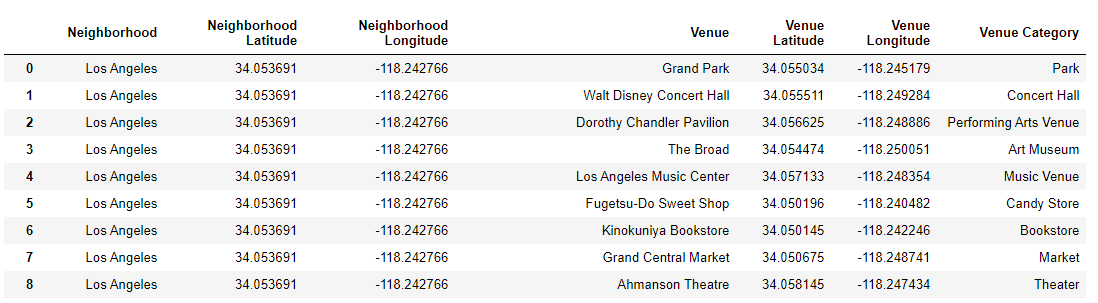
1. Since we will analyze the venues based on city locations, we need to first get the coordinates of each city. I use “Los Angeles County City List” as my data source and import geocoder model to get latitude and longitude as below.



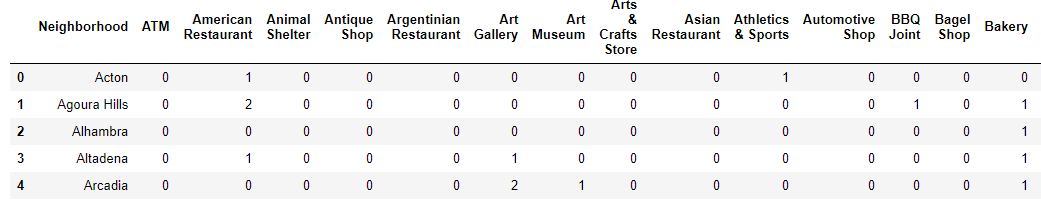
1. All 3 data sources files are joined together and become the final base dataset for further analysis as below. Final dataset has been cleaned up by removing invalid lines due to missing data from original datafile and reorganizing column order to make the dataset more readable.



1. Based on coordinates found for each city, Foursquare API is called to get nearest venues as below.

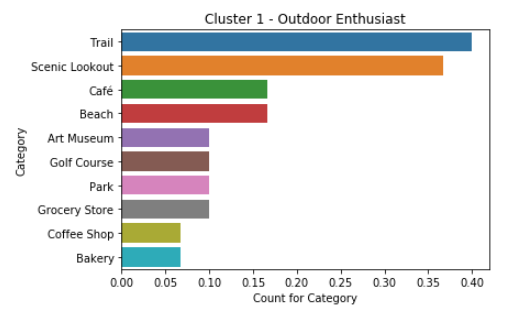


One-hot dataset is also generated for cluster preparation:

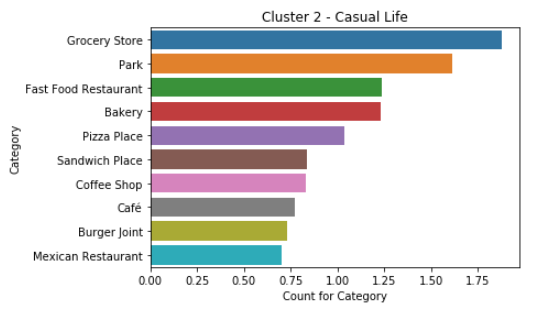


1. Using K-means model to cluster the cities based on their popular venues. In this project, we cluster the cities to 5 groups. After clustering, we get the groups as below:

Cluster 1 – “Trail” and “Scenic Lookout” are the top 2 categories in those cities. The people living there love outdoor activities.



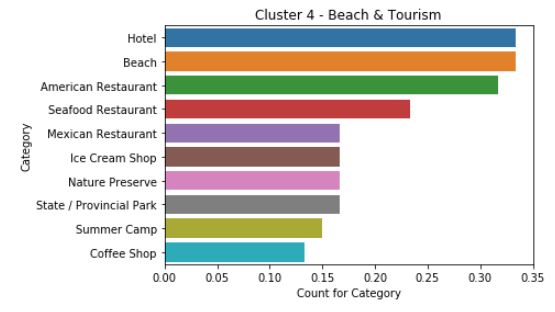
Cluster 2- “Grocery Store”, “Park” and food places are the top 10 categories. People living in those cities seems having casual lifestyles.



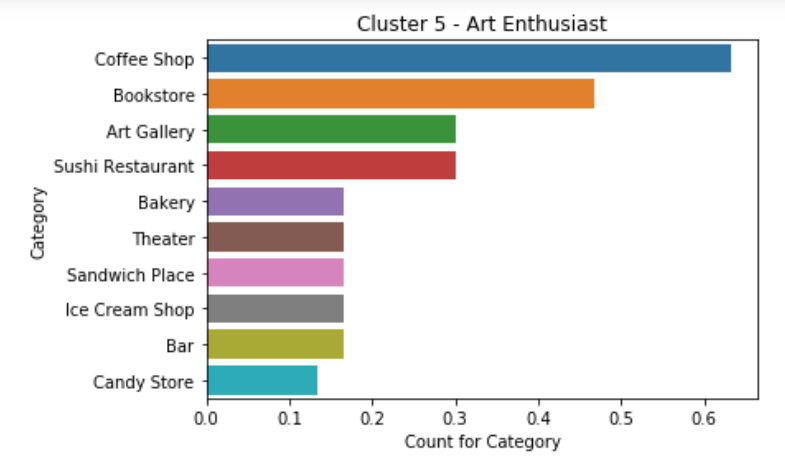
Cluster 3: All top 10 categories are food places. The people living in those cities really enjoy the different kind of cuisines.



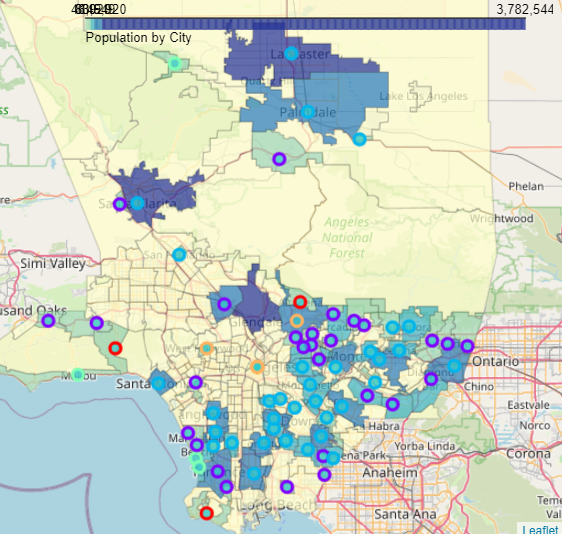
Cluster 4: “Hotel” and “Beach” are the top 2 categories for this group. Also “American Restaurant” and “Seafood Restaurant” make those cities the good places for tourists or for the people who likes nearby beaches.



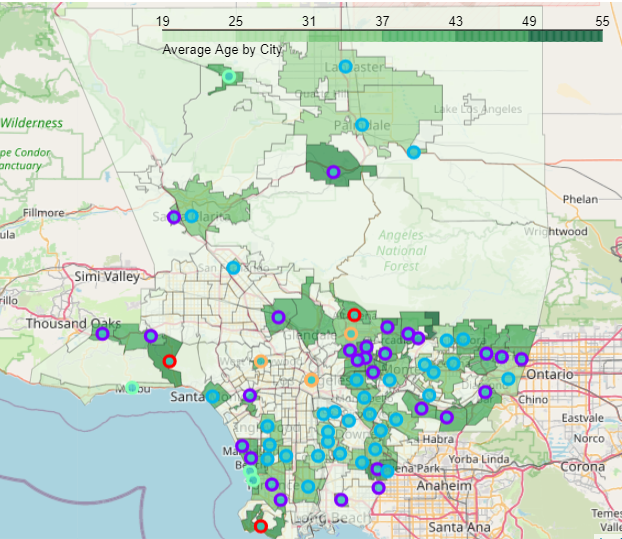
Cluster 5 – “Coffee ship”, “Bookstore”, ”Theater” and “Art Gallery” are good combinations for the art lovers.



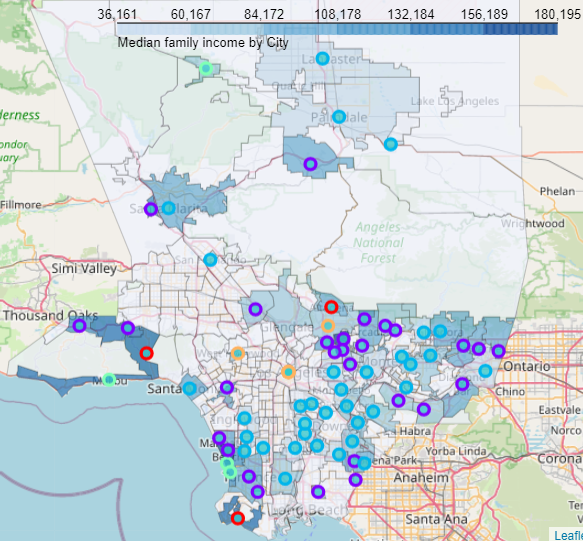
1. Visualize the clusters on the top of other relevant data layers using Folium model and choropleth map.
2. Cluster group over “City Population”



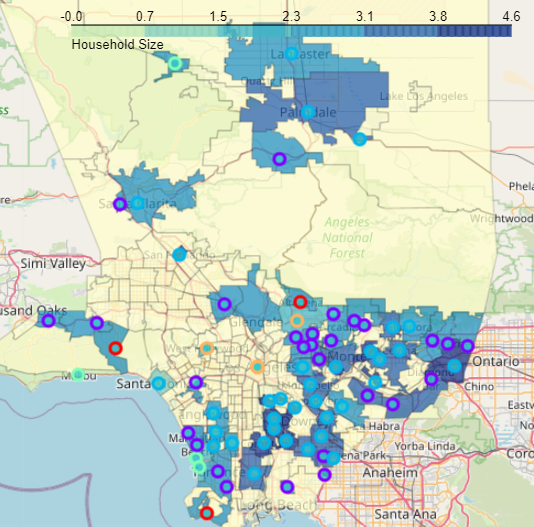
1. Cluster group over “Average Age”



1. Cluster group over “Median Family Income”



1. Cluster group over “Median Family Income”



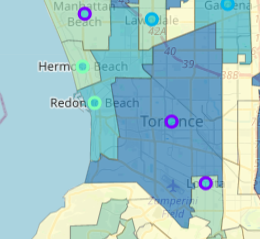
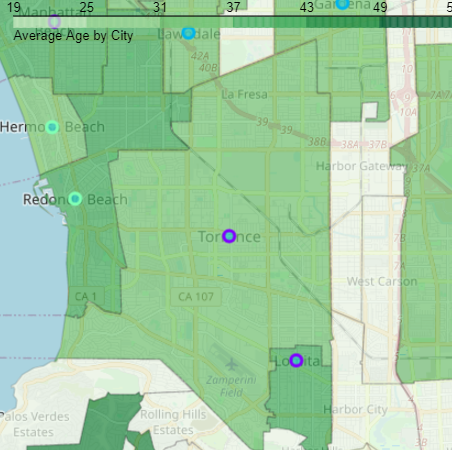
**Discussion and Observations**

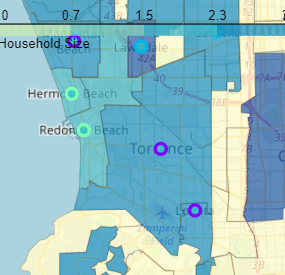
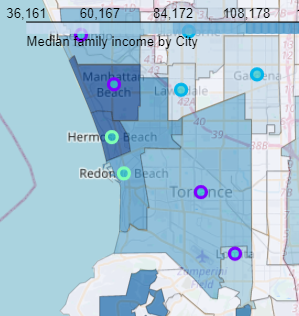
By comparing the cluster group and corresponding city data, we can easily find out where the differences are among cities. I will deep dive the data of 2 cities and provide some insights about potential business opportunities.

**Case #1 –City Torrance**

Torrance is currently labelled as “Casual Life” cluster. It has a large population (144K). The average households size is 2.94 and Median Age is 35. which indicate young families with kids.

*Recommendation:* There would be the good opportunity to open Seafood restaurant or Hotels since it is close to Redondo Beach which is popular for tourists. It would be very convenient for tourists to come to Torrance and enjoy their stays there.

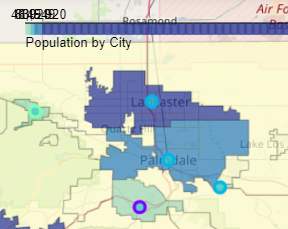
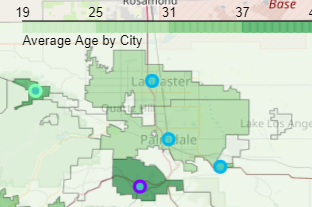


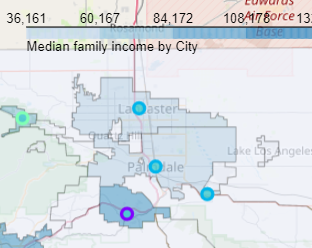
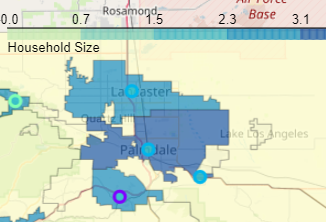
 

**Case#2 – City Lancaster and Palmdale**

Lancaster and Palmdale are labelled as “Food Lover” cluster. There are a lot of restaurants in those cities. The median incomes are low (55K-57K) and household size are a little high (2.3-3.1). The population is a little high, especially for Lancaster.

*Recommendation:* By considering all data for the 2 cities, it would be a good opportunity to open a couple of more grocery stores as it would be very convenient for the family with kids. The goods price in the grocery store should be reasonable and affordable for most of families there.

**Conclusion**

By comparing the most popular venues around the cities in Los Angeles County, business owners will have much better ideas about how and where they should expand or set their business.

With the additional data, such as population, family income, average household size and average age, we can identify those business opportunities even more accurately.

When business owners review those opportunities, it would be also valuable to consider other information, such as criminal rate and race.