**Introduction**

In this notebook, I want to examine the neighborhoods in Los Angeles and cluster them based on similarities in terms of location and nearby services. Finally, with this information, I want to determine the following questions:

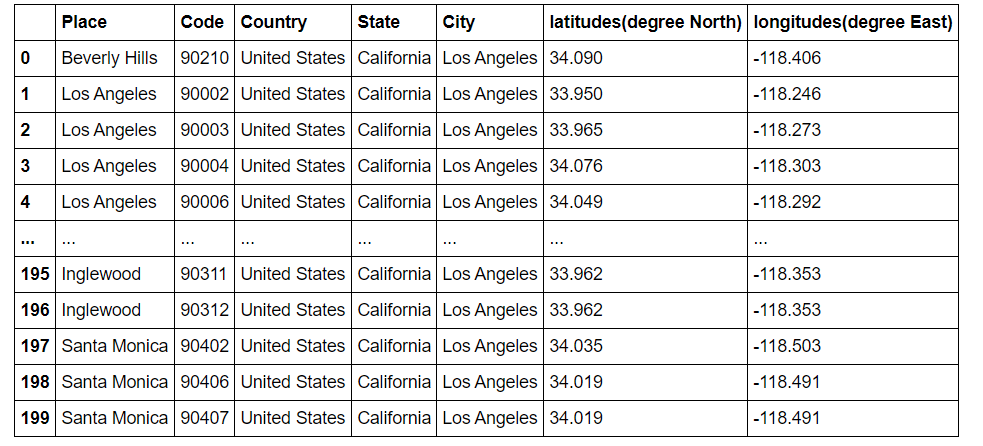
1. Which part of Los Angeles has the most venues
2. What category of venue is most common in LA
3. If I cluster LA venues into 5 groups, what does the clusters look like

# **Data**

The data I will use include:

* FourSquare location data
* A list of Postal Codes in Los Angeles: https://www.geonames.org/postal-codes/US/CA/california.html

The postal code dataset contains a list of locations in Los Angeles, their postal codes, and their longitudes and latitudes. From these information, I composed a Pandas dataframe, which, after applying data wrangling, contains 200 locations in Los Angeles and looks as the following:

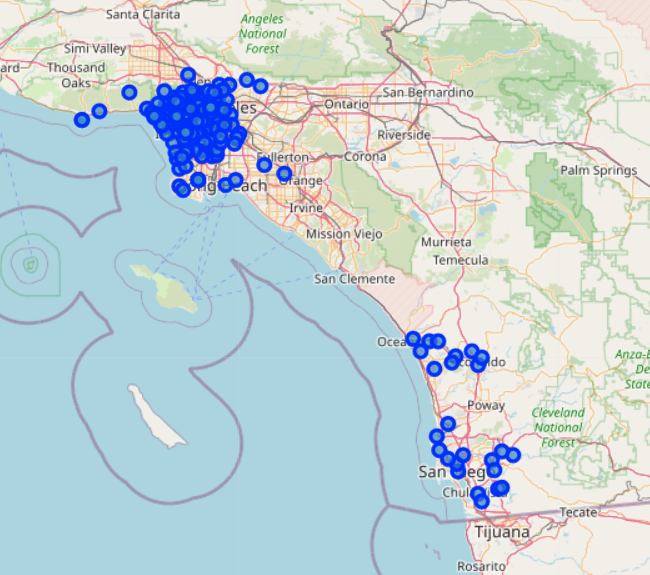


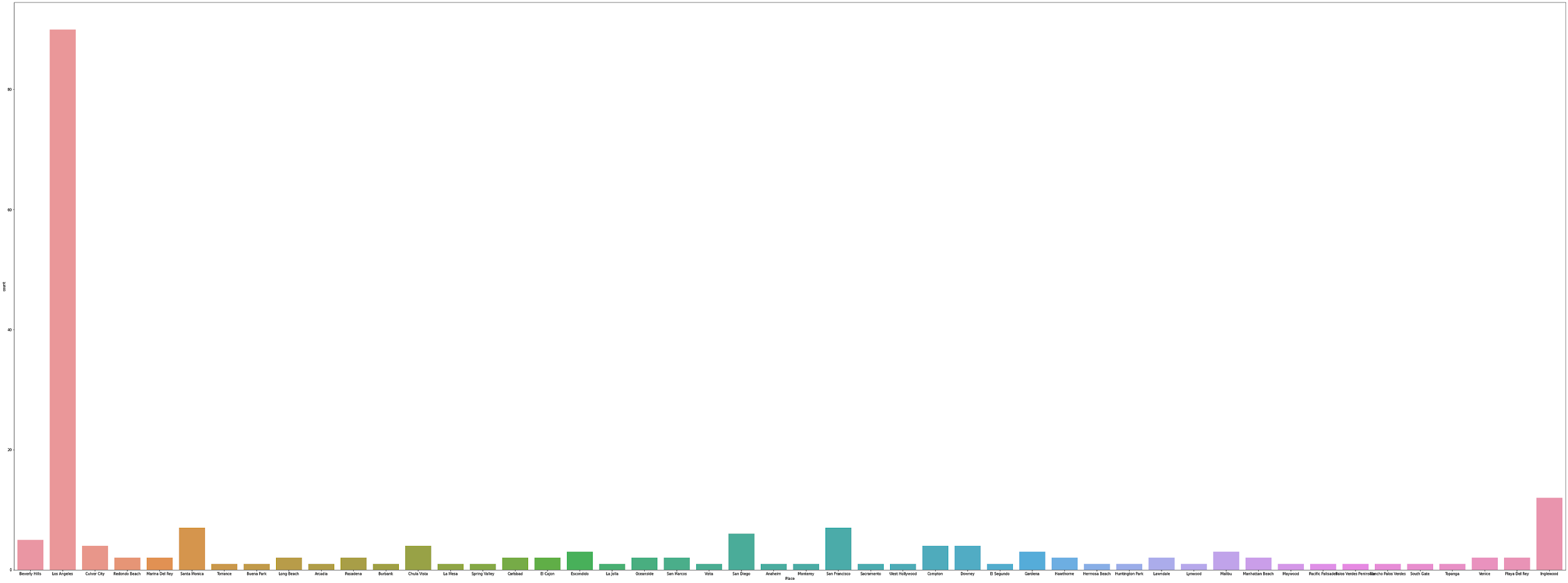
**Methodology**

To better understand the data, I first applied exploratory data analysis to it and then did a clustering analysis using K-means algorithm to determine how the venues will be grouped.

**EDA:**

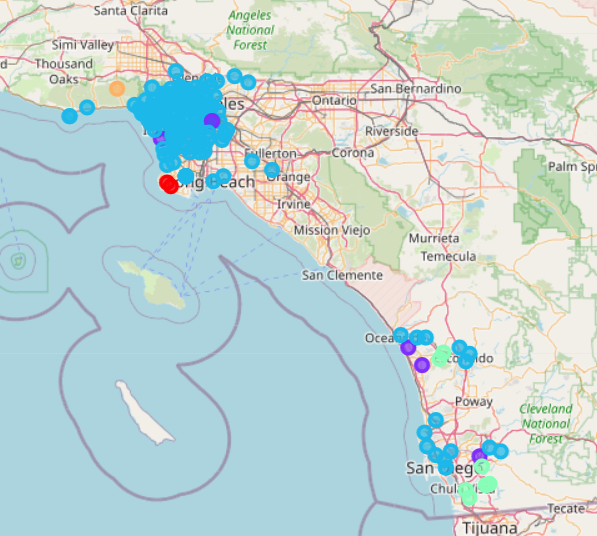
* Locations of each data point





Then, I used FourSquare API to collect the venues around these locations and performed a clustering to the venues.

**K-Means Clustering:**



**Result**

We see that most venues are located in LA and then in Inglewood and then other places.And in fact LA surpasses other places by a lot. Also, the above map illustrates how the cluster looks like

**Discussion**

We see that most venues are located in LA and then in Inglewood and then other places.And in fact LA surpasses other places by a lot. Also, the above map illustrates how the cluster looks like

# **Conclusion**

From this analysis, we see that

1. Most venues are located in Los Angeles, then Inglewood, then other places
2. Although the category for each venue isn't general enough, it seems like restaurant is the most common venue in LA
3. The above map illustrates how the cluster looks like