**Applied Data Science Capstone project report**

**Event locator for the data scientist meet - 2021 in Los Angeles**

**By Sathya Narayanan Muralidaran**

**Project Description:**

A company in California was planning to set up a Data Science Annual meet in 2020.But due to Covid 19,they planned for online meeting. Hence, they come up with the new proposal for the year 2021 with an idea of 5-day event and in order to engage the people differently with multiple activities during their free time, venue needs to be planned accordingly. Now the challenge is to choose the right spot where people can engage their leisure times in Restaurants, Parks and shopping malls. Hence the venue should give some idea of choosing the right places nearby.

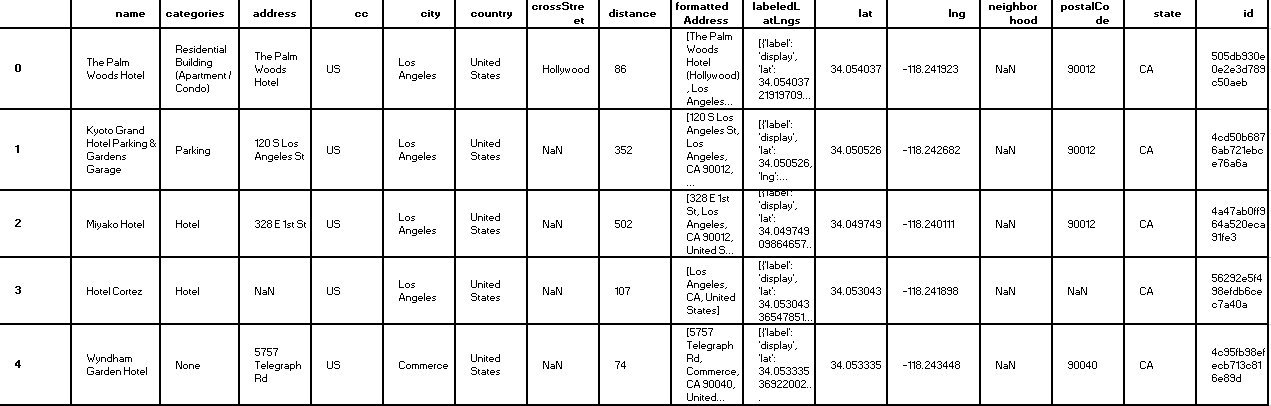
**Data Selection:**

By Using Four Square API, we need to identify the Hotel, Restaurant, Parks and Shopping malls in Los Angeles. Then after applying data cleaning activities, all the data are combined, clustered and visualized using Folium

**Activities to be performed:**

* Libraries Import
* Using Foursquare credentials publishing connections with Foursquare API by passing the necessary parameters
* Collecting data for Hotels
* Creating Data frames for Hotels, Restaurants, Parks and Shopping Malls
* Removing unwanted Columns for Hotels, Restaurants, Parks and Shopping Malls
* Deleting Blank or NAN rows and also ensure no duplicates
* Repeat the above 4 steps for Restaurants, Parks and Shopping Malls
* Consolidate the all the data and visualize the same in folium map.

Hotels: Initial sample data



**Cleaning activities :**

We need to perform the below activity to ensure the data cleaning

* Removing unwanted columns
* Removing Duplicates
* Removing data with NAN values
* Removing invalid data
* Removing data that belongs to other category

After that we are assigning it to our actual data frame

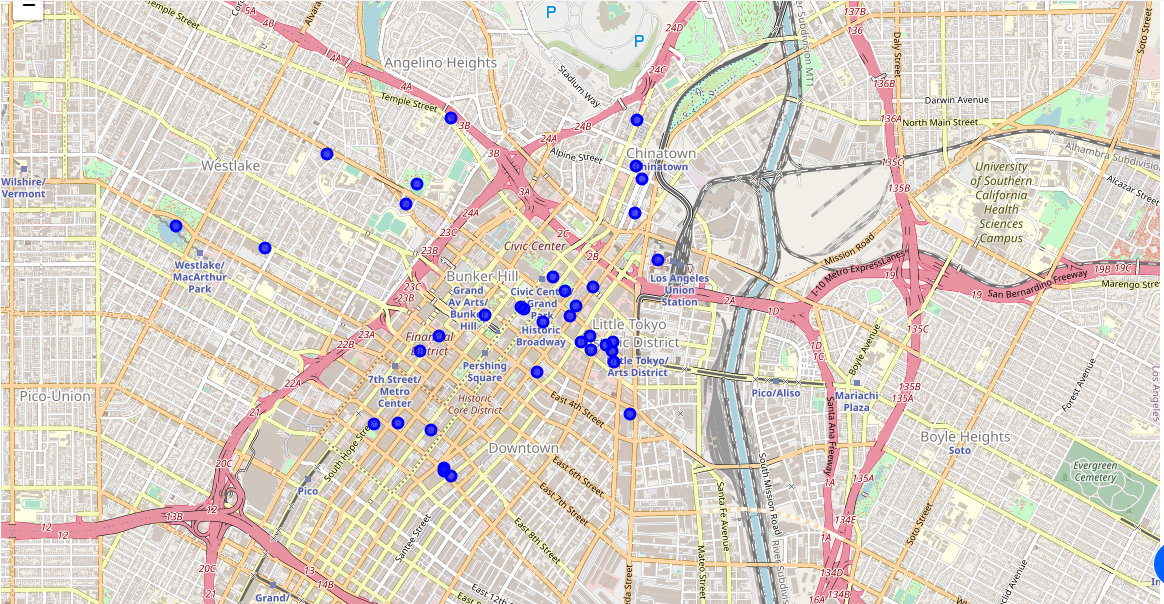
Then our data looks like



Similarly, we need to do data cleansing activity for other data frame such as Restaurants ,Parks and Shopping malls and consolidate the data to neighbourhood frame. This is the base data frame for clustering.

**Plotting and generating the folium map :**

Plot all the details and Generate the folium map to get the idea on the cluster



**Conclusion**:

Accordingly, we finalize that best place is southern California area of Los Angeles with the postal code 90012.This is the best place to conduct an event where all the Restaurants, Parks and Shopping malls are near to hotels.