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

	name	categories	address	cc	city	country	crossStreet	distance	formatted Address	labeledLatLngs	lat	lng	neighborhood	postalCode	state	id
0	The Palm Woods Hotel	Residential Building (Apartment / Condo)	The Palm Woods Hotel	US	Los Angeles	United States	Hollywood	86	[The Palm Woods Hotel (Hollywood) , Los Angeles...	[{'label': 'display', 'lat': 34.054037 21919709...	34.054037	-118.241923	NaN	90012	CA	505db930e 0e2e3d789 c50aeb
1	Kyoto Grand Hotel Parking & Gardens Garage	Parking	120 S Los Angeles St	US	Los Angeles	United States	NaN	352	[120 S Los Angeles St, Los Angeles, CA 90012, ...	[{'label': 'display', 'lat': 34.050526, 'lng':...	34.050526	-118.242682	NaN	90012	CA	4cd50b687 6ab721ebc e76a6a
2	Miyako Hotel	Hotel	328 E 1st St	US	Los Angeles	United States	NaN	502	[328 E 1st St, Los Angeles, CA 90012, United S...	[{'label': 'display', 'lat': 34.049749 03864657...	34.049749	-118.240111	NaN	90012	CA	4a47ab0ff9 64a520eca 91fe3
3	Hotel Cortez	Hotel	NaN	US	Los Angeles	United States	NaN	107	[Los Angeles, CA, United States]	[{'label': 'display', 'lat': 34.053043 36547851...	34.053043	-118.241898	NaN	NaN	CA	56292e5f4 98efdb6ce c7a40a
4	Wyndham Garden Hotel	None	5757 Telegraph Rd	US	Commerce	United States	NaN	74	[5757 Telegraph Rd, Commerce, CA 90040, United...	[{'label': 'display', 'lat': 34.053335 36922002...	34.053335	-118.243448	NaN	90040	CA	4c95fb98ef ecb713c81 6e83d

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

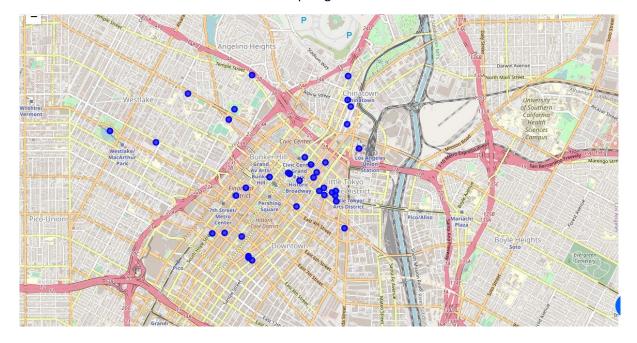After that we are assigning it to our actual data frame

Then our data looks like

	name	categories	address	lat	lng	postalCode	state
2	Miyako Hotel	Hotel	328 E 1st St	34.049749	-118.240111	90012	CA
5	Roes hotel	Hotel	2683 Euclid Avenue	34.052979	-118.243178	90001	CA
8	Kawada Hotel	Hotel	200 S Hill St	34.05287	-118.247909	90012	CA
11	Little Toyko Hotel	Hotel	327 1/2 E 1st St	34.050409	-118.240011	90012	CA
15	DoubleTree by Hilton	Hotel	120 South Los Angeles Street	34.050352	-118.242768	90012	CA

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.