Global Consultants

Anju Antony

October 14,2019

1. Introduction

1.1 Business Problem

Global Consultants is a consulting company that helps/advise their customers who is planning to shift to a new location for settling down. People leave their current location for various reasons like job transfers or for a better living. One of the most important factors that needs to be considered while planning to shift to a new city/place is job opportunities, amenities or facilities provided by the place where you chose to live. We would be able to enjoy life if we have a steady income. We would be happier or satisfied in our residence or place, if there are variety of options to earn a living and enjoy life. Therefore, main aim of company is to predict the best options by which customers can select a best location to earn a living or enjoy life. For example, if customer is planning to moving to California , we would like to predict the nearby town or place that provides opportunities to earn a living and all amenities for them to enjoy.

Based on the area a customer would like to shift, with the help of location data of that area and the categories of interest of the customer including basic amenities, we will provide various options to the customer. This project aims to group the areas in a location as different clusters and make decisions based on it.

The Global Consultants Company would be able to provide helpful information for their customers and they would be able to pick the best place to live accordingly. The customers would be highly satisfied with the service the Company is providing and this would help to increase the reputation of the company and gain more customers. This will help make their life easy in a new city or country.

1.2 Client

Customers include the people who wish to migrate to different city or country for a better living. It would be helpful to such people to find venues that provides all facilities as well as some opportunities to start their initial earning and start a better life, nearby to the city where they are willing to migrate. Our company focus on such customers who need advise on which location to move and provide them with options or facilities available in the city. Customer willing to relocate to California will get an idea about nearby venues and opportunities provided by those places.

2. Data acquisition and cleaning:

2.1 Data:

One city will be analysed in this project. Based on the city, the latitude and longitude values will be obtained by using geopy module. The data used in this project is provided by Foursquare location data. The city's latitude and longitude values will be used to extract the venues data from Foursquare location data using explore endpoint. From the obtained results which is a JSON file, we will filter the columns and create a table/dataframe.

2.2 Data Cleaning:

Remove unnecessary columns from the dataframe created .

2.3 Feature Selection:

The features of the data should be analysed, and the unnecessary features could be removed. For our project the main feature would be the latitude and longitude values. Using latitude and longitude, the venues would be grouped into different clusters. Based on the clusters the decisions can be taken accordingly.

For this project we have chosen to explore a city in US. One of the best places to live in US is California. Below is the image of the Foursquare API data of California City, in the form of dataframe after cleaning.

	city	name	categories	lat	lng
0	Sequoia	General Sherman Tree	Tree	36.581690	-118.751456
1	Sierra	Kings Canyon National Park	National Park	36.740316	-118.963127
2	Hume	General Grant Tree	Tree	36.748025	-118.971573
3	Sequoia National Park	Kings Canyon Outlook	Scenic Lookout	36.719582	-118.897462

3. Methodology

3.1 Business Understanding

Our main goal is to suggest the best location to a customer who would like to relocate to California City .

3.2 Analytic Approach:

Around 100 venues are identified using Foursquare location data. K Means Clustering Algorithm is used to group the data in the dataframe to different clusters and these clusters are analysed to find suitable location for customer.

3.3 Exploratory Data Analysis

1. We use geopy and get the latitude and longitude values of the city to explore.

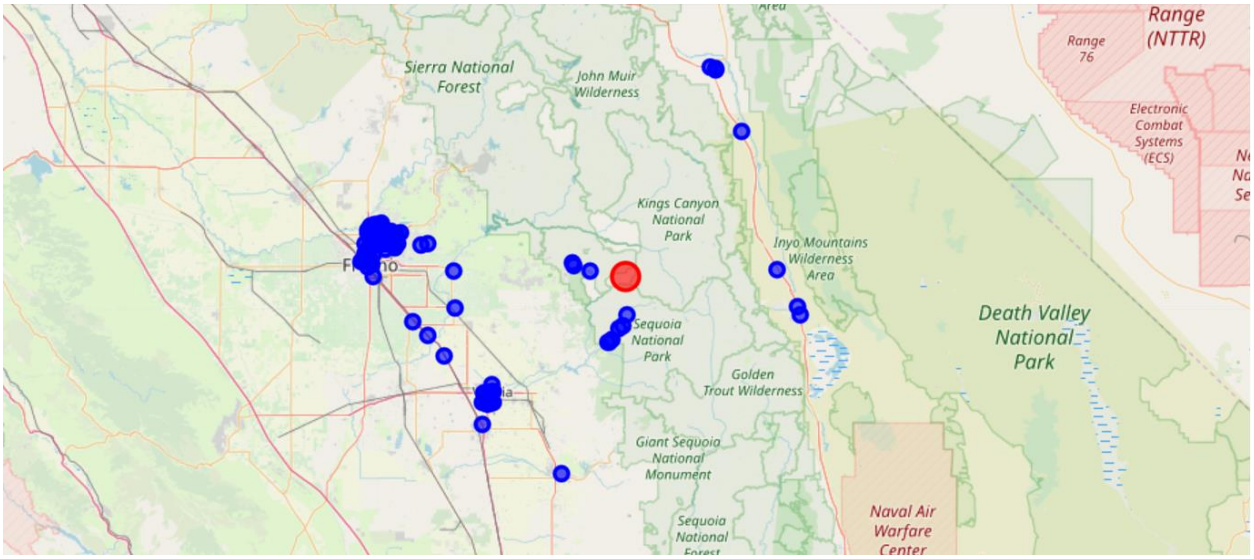
2. The list of venues around the city can be obtained from Foursquare by using the Explore endpoint.

3. The data obtained from Foursquare is transformed to pandas dataframe.

4. We will use folium libraries to create map of the city with venues superimposed on top.

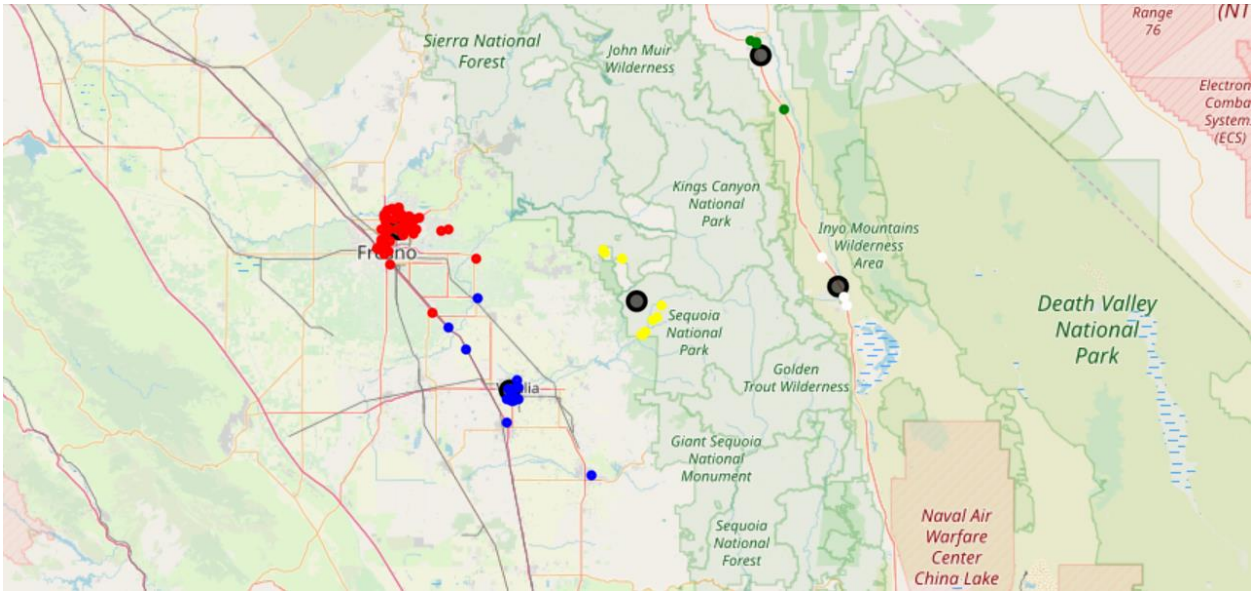
5. We will use K Means clustering algorithm to group the nearby clusters.

Folium Map of California



Folium Map of California after applying kmeans clustering

We can see 5 clusters**.** Centroids of cluster is displayed as a black circle mark. Spots in cluster 0 is displayed in blue. Spots in cluster 1 is displayed in red. Spots in cluster 2 is displayed in green. Spots in cluster 3 is displayed in yellow. Spots in cluster 4 is displayed in white.



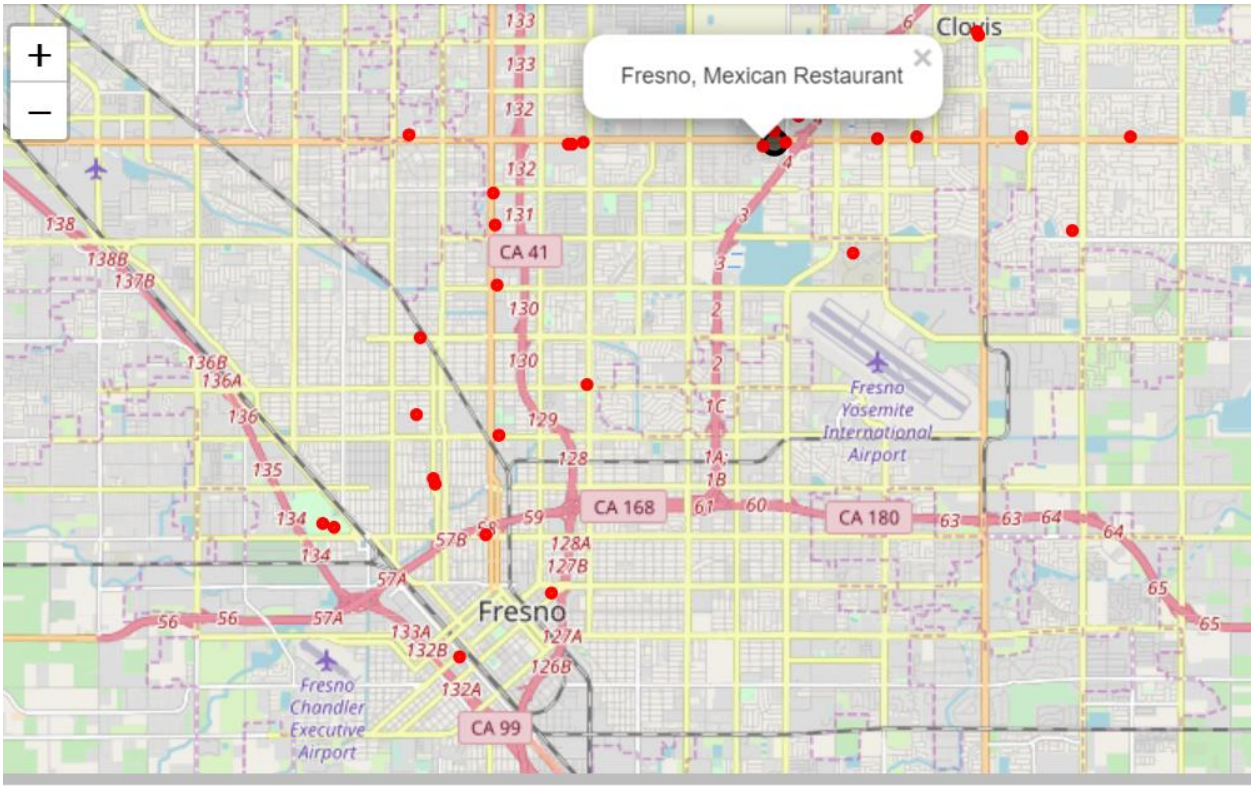
4. RESULTS

From the map and dataframe it is clear that the maximum number of restaurants, movie theatres, grocery shops, coffee shops, shopping malls etc are available in cluster1.One shifting to cluster1 can easily find a earn of living and spend life as other facilities are available. Cluster2 can be ignored as there are only three venues and they are bakery, BBQ and Coffee Shop. Cluster0 is second best option where we can find many restaurants, movie theatres and shops. Cluster 3 should be ignored because there is only national park, mountain and scenic which cannot be chosen as a place of living. It can be better chosen as tourist spot. Cluster 4 can be ignored as there is only historic site and tourist information centre.

Cluster 1

```
[29]: dataframe_cluster1.categories.unique()
```

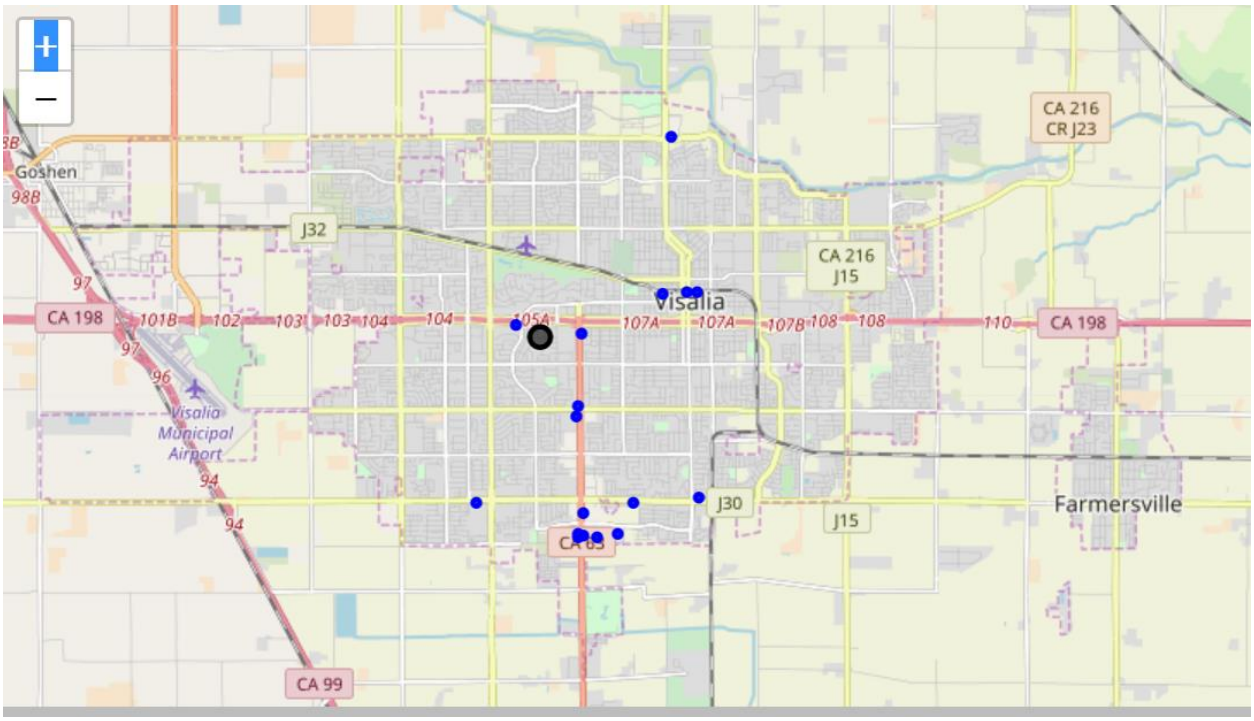
```
[29]: array(['New American Restaurant', 'Mexican Restaurant',
       'Frozen Yogurt Shop', 'American Restaurant', 'Sports Bar',
       'Grocery Store', 'Pizza Place', 'Sushi Restaurant',
       'Breakfast Spot', 'Burger Joint', 'Ice Cream Shop', 'Bakery',
       'Coffee Shop', 'Park', 'Water Park', 'Gym', 'Theater',
       'Steakhouse', 'Market', 'Brewery', 'Multiplex',
       'Mediterranean Restaurant', 'Garden Center', 'Zoo Exhibit', 'Bar',
       'Indian Restaurant', 'Fast Food Restaurant', 'Pet Store',
       'Gastropub', 'Shopping Mall', 'Arts & Crafts Store', 'Lounge',
       'Italian Restaurant', 'Zoo', 'Seafood Restaurant', 'Diner',
       'Sandwich Place', 'College Basketball Court'], dtype=object)
```



Cluster 0

```
dataframe_cluster0.categories.unique()
```

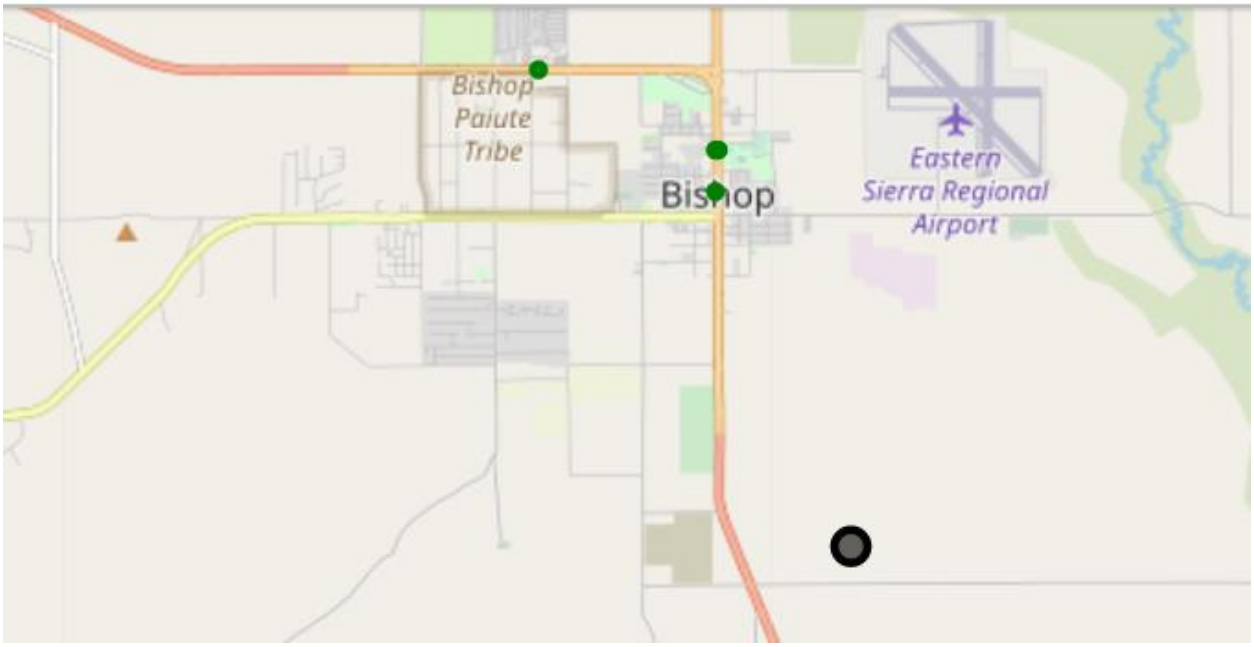
```
array(['Burger Joint', 'Gastropub', 'American Restaurant', 'Liquor Store',
       'Sushi Restaurant', 'Warehouse Store', 'Factory',
       'Japanese Restaurant', 'Donut Shop', 'Coffee Shop', 'Pizza Place',
       'Mexican Restaurant', 'Movie Theater', 'Pet Store',
       'Italian Restaurant'], dtype=object)
```



Cluster 2

```
dataframe_cluster2.categories.unique()
```

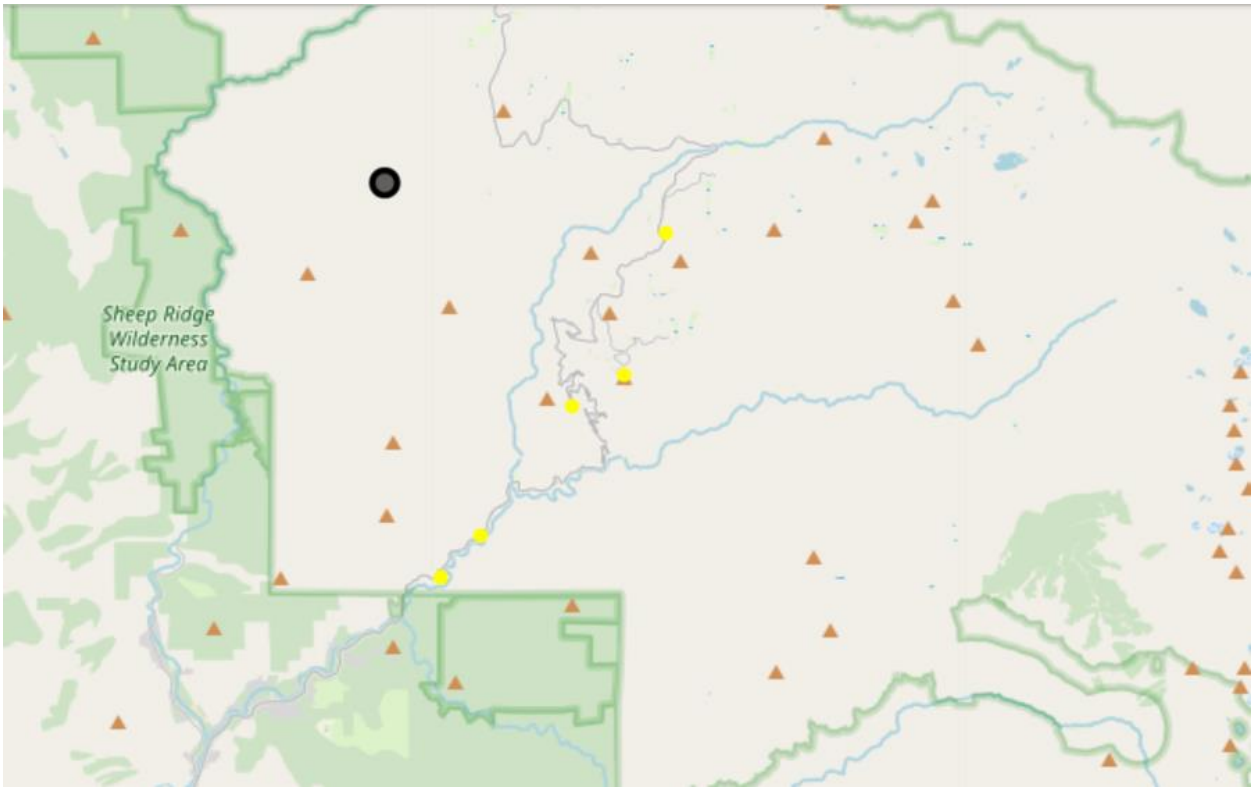
```
array(['BBQ Joint', 'Bakery', 'Deli / Bodega', 'Coffee Shop'],
      dtype=object)
```



Cluster 3

```
dataframe_cluster3.categories.unique()
```

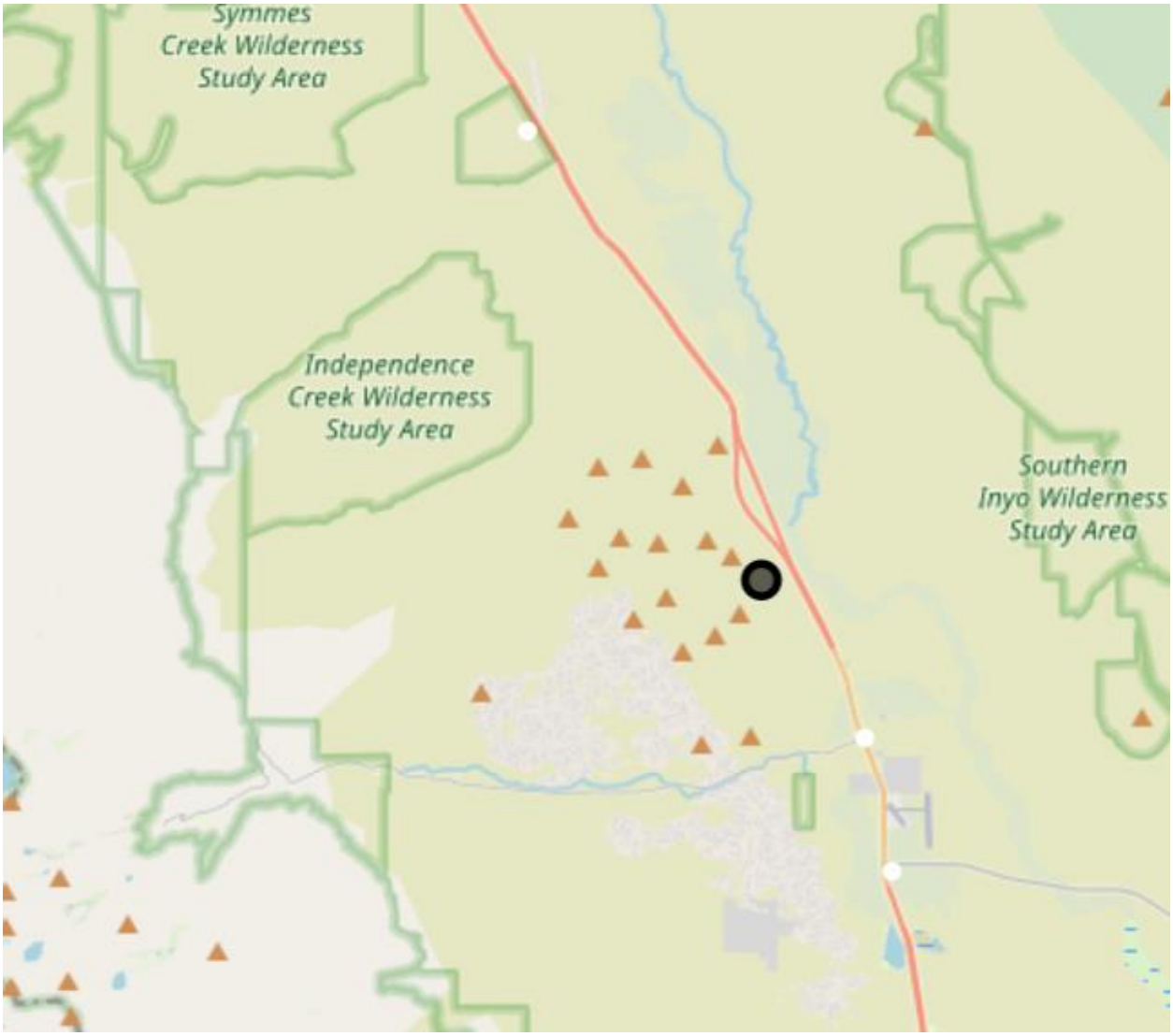
```
array(['Tree', 'National Park', 'Scenic Lookout', 'Mountain',
       'Tourist Information Center'], dtype=object)
```



Cluster4

```
dataframe_cluster4.categories.unique()
```

```
array(['Historic Site', 'American Restaurant',
       'Tourist Information Center'], dtype=object)
```



5. DISCUSSION

After the analysis, we have obtained two options to suggest to the customer who plans to move to California City. Cluster 1 is the best option where many restaurants, shopping mall, theatres, shops etc are available. Cluster0 is the second best option where few number of restaurants and shops, pizza place are available. Cluster 2 ,3,4 can be ignored as there are not much facilities available.

6. CONCLUSION

We can easily understand from the maps generated as part of analysis. Cluster 1 has maximum facilities including air transport. All facilities available for living and facilities to find job is

mostly in cluster1. College Basketball court shows the location of educational institutions. Someone who migrates to California can choose Fresno(city) as best place to live and can earn money by doing part time or full time jobs. Restaurants, pubs,theatres,coffee shops reveals the cities in cluster 1 is best place for residence as all these facilities are established near residential areas .

Cluster 0 can be second best option as the cities in cluster 0 has air transport(from map) and many shops and resturants. These facilities shows the cities in cluster 0 is also another option to be chosen as area of residence and to start a living.

Cluster 2,3,4 has some scenic,mountain, historic places which cannot be chosen as a place of living or to settle down. So these options can be ignored.

This project is performed on limited data available from Foursquare. If a good amount of data is available which has the interesting and useful venues, weather condition of the areas etc, we would be able to provide more idea and suggestions. We will be able to get customer satisfaction more if more information is provided.