# Data :

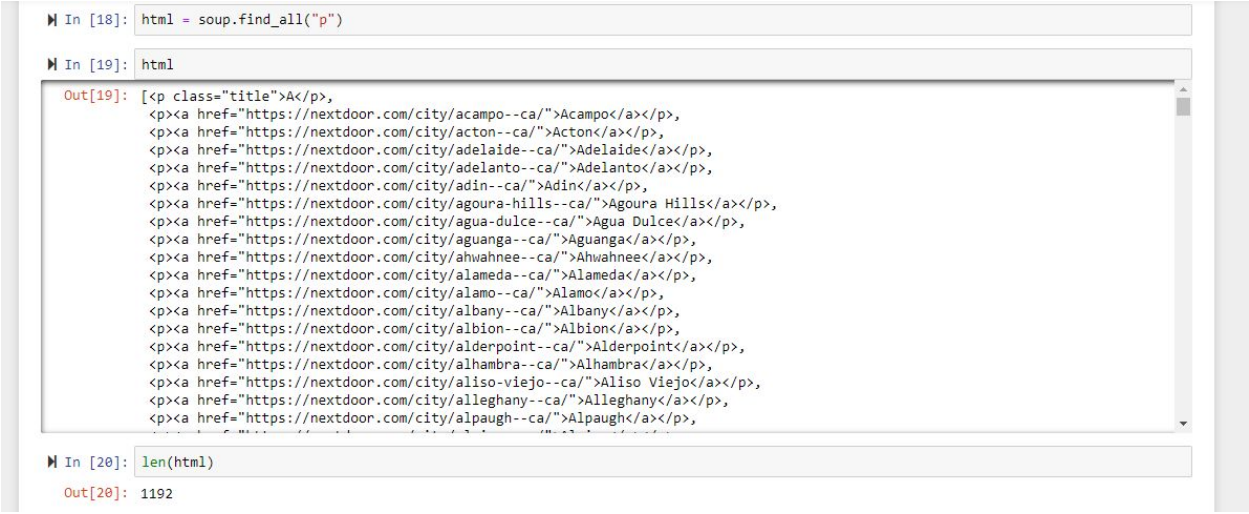
Location names is downloaded from nextdoor.com.

With the help of location names we got their latitude and longitude using python geocoder.

This data is then sent to foursquare API from which venues near the location is downloaded.

All this data are saved in csv for easy access in future references.

Link to the data ipynb :
https://github.com/Udolf15/Coursera_Capstone/blob/master/DataExtraction.ipynb



```
In [18]: html = soup.find_all("p")

In [19]: html

Out[19]: [<p class="title">A</p>,
 <p><a href="https://nextdoor.com/city/acampo--ca/">Acampo</a></p>,
 <p><a href="https://nextdoor.com/city/acton--ca/">Acton</a></p>,
 <p><a href="https://nextdoor.com/city/adelaide--ca/">Adelaide</a></p>,
 <p><a href="https://nextdoor.com/city/adelanto--ca/">Adelanto</a></p>,
 <p><a href="https://nextdoor.com/city/adin--ca/">Adin</a></p>,
 <p><a href="https://nextdoor.com/city/agoura-hills--ca/">Agoura Hills</a></p>,
 <p><a href="https://nextdoor.com/city/agua-dulce--ca/">Agua Dulce</a></p>,
 <p><a href="https://nextdoor.com/city/aguanga--ca/">Aguanga</a></p>,
 <p><a href="https://nextdoor.com/city/ahwahnee--ca/">Ahwahnee</a></p>,
 <p><a href="https://nextdoor.com/city/alameda--ca/">Alameda</a></p>,
 <p><a href="https://nextdoor.com/city/alamo--ca/">Alamo</a></p>,
 <p><a href="https://nextdoor.com/city/albany--ca/">Albany</a></p>,
 <p><a href="https://nextdoor.com/city/albion--ca/">Albion</a></p>,
 <p><a href="https://nextdoor.com/city/alderpoint--ca/">Alderpoint</a></p>,
 <p><a href="https://nextdoor.com/city/alhambra--ca/">Alhambra</a></p>,
 <p><a href="https://nextdoor.com/city/aliso-viejo--ca/">Aliso Viejo</a></p>,
 <p><a href="https://nextdoor.com/city/alleghany--ca/">Alleghany</a></p>,
 <p><a href="https://nextdoor.com/city/alpaugh--ca/">Alpaugh</a></p>,

In [20]: len(html)

Out[20]: 1192
```

**Getting the Latitude and Longitude of the Neighborhood list**

```
In [49]: dataFrame = pd.DataFrame(columns=['Neighborhood','Latitude','Longitude'])

geolocator = Nominatim()
i=0
for place in finalList:
    print(place)
    while True:
        try:
            geo = geolocator.geocode(place+" california")
            if geo != None:
                dataFrame.loc[i] = [place , geo.latitude, geo.longitude]
                i = i+1
            break
        except Exception as e:
            print(e)
```

**Saving the Data**

```
In [51]: dataFrame.to_csv('final_California_csv')

In [ ]:
```

Link to the data : https://github.com/Udolf15/Coursera_Capstone/tree/master/data

# Methodology :

**1 : Convert location names into latitude and longitude.**

**2 : Then use the foursquare api to get nearby places.**

**3 : Then process it to get the nearby places as attributes using dummy variables.**

**4 : After that using k means to get clusters having same properties.**

**5 : Using folium library to draw the cluster on map.**

**List of Top 10 venues for every neighborhood**

In [97]:
```
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
neighborhoods_venues_sorted_cal = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted_cal['Neighborhood'] = cal_grouped['Neighborhood']

for ind in np.arange(cal_grouped.shape[0]):
    neighborhoods_venues_sorted_cal.iloc[ind, 1:] = return_most_common_venues(cal_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted_cal.head()
```

Out[97]:

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Acampo | Winery | Business Service | Women's Store | Food Service | Filipino Restaurant | Fish & Chips Shop | Fish Market | Flea Market | Flower Shop | Fondue Restaurant |
| 1 | Adelaide | Hotel | Cocktail Bar | Bar | Theater | Speakeasy | Art Gallery | Thai Restaurant | Indian Restaurant | Hostel | Spa |
| 2 | Adelanto | Business Service | Lawyer | Thai Restaurant | Food | Concert Hall | Fish & Chips Shop | Fish Market | Flea Market | Flower Shop | Fondue Restaurant |
| 3 | Adin | Burger Joint | Market | Women's Store | Food Service | Filipino Restaurant | Fish & Chips Shop | Fish Market | Flea Market | Flower Shop | Fondue Restaurant |
| 4 | Agoura Hills | Home Service | Thai Restaurant | Women's Store | Food Court | Filipino Restaurant | Fish & Chips Shop | Fish Market | Flea Market | Flower Shop | Fondue Restaurant |

In [98]:
```
cal_grouped.head()
```

Out[98]:

## Applying Kmeans on the data

In [99]:
```
kclusters = 5

cal_grouped_clustering = cal_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(cal_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

Out[99]:
```
array([3, 1, 1, 1, 4, 1, 1, 1, 1, 1])
```

In [100]:
```
kmeans.labels_
```

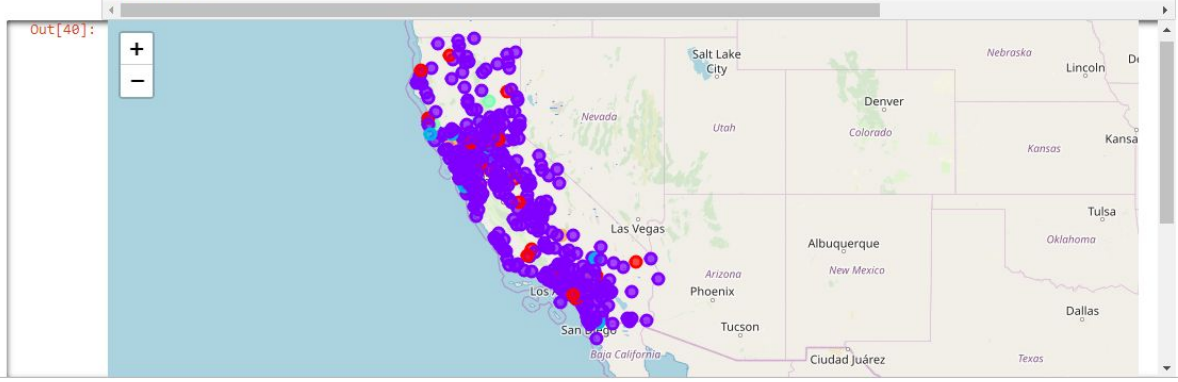
Out[100]:
```
array([3, 1, 1, 1, 4, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       4, 1, 1, 1, 3, 1, 1, 1, 1, 1, 0, 1, 1, 1, 3, 1, 1, 1, 1, 1, 4,
       1, 1, 1, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 4, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 3, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 3, 1, 1,
       1, 1, 1, 1, 1, 3, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1,
       1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 3, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 4, 1, 1, 2, 1, 1, 1,
       1, 4, 4, 4, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 2, 1, 1, 0, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1,
       1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 4, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1,
       1])
```

## Representing the 5 Clusters on map

In [40]:
```
map_clusters = folium.Map(location=[34, -118], zoom_start=4)

# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i+x+(i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(neighborhood_merged['Latitude'], neighborhood_merged['Longitude'], neighborh
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```

```
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```

Out[40]:




# 5 Custers :

## Cluster 1

```
In [119]: neighborhood_merged.loc[neighborhood_merged['Cluster Labels'] == 0, neighborhood_merged.columns[[1] + list(range(5, neighborhood_
```

Out[119]:

| | 1st Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue | Cluster Labels | Latitude | Longitude |
|---|---|---|---|---|---|---|---|---|---|---|
| 22 | Park | Fish & Chips Shop | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | Food | 0 | 38.174918 | -122.260804 |
| 76 | Bed & Breakfast | Food Truck | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | Food | 0 | 34.267886 | -116.848598 |
| 144 | Home Service | Café | Women's Store | Food Court | Fish & Chips Shop | Fish Market | Flea Market | 0 | 33.864429 | -118.053932 |
| 179 | Sporting Goods Shop | Fish & Chips Shop | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | Food | 0 | 37.216492 | -121.739417 |
| 205 | Park | Food Service | Fish & Chips Shop | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | 0 | 37.914774 | -121.594005 |
| 221 | Hotel | Food Court | Filipino Restaurant | Fish & Chips Shop | Fish Market | Flea Market | Flower Shop | 0 | 39.206009 | -120.837717 |
| 246 | Baseball Stadium | Food Service | Fish & Chips Shop | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | 0 | 38.694591 | -122.021368 |

## Cluster 3

```
In [117]: neighborhood_merged.loc[neighborhood_merged['Cluster Labels'] == 2, neighborhood_merged.columns[[1] + list(range(5, neighborhood_
```

Out[117]:

| | 1st Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue | Cluster Labels | Latitude | Longitude |
|---|---|---|---|---|---|---|---|---|---|---|
| 139 | Grocery Store | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | Food | Food & Drink Shop | 2 | 38.533412 | -123.085463 |
| 188 | Grocery Store | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | Food | Food & Drink Shop | 2 | 34.863417 | -116.868032 |
| 217 | Grocery Store | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | Food | Food & Drink Shop | 2 | 32.644225 | -116.781409 |
| 324 | Grocery Store | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | Food | Food & Drink Shop | 2 | 38.295747 | -121.244392 |
| 369 | Grocery Store | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | Food | Food & Drink Shop | 2 | 38.965089 | -122.837168 |
| 433 | Grocery Store | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | Food | Food & Drink Shop | 2 | 37.271054 | -122.306581 |
| 457 | Grocery Store | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | Food | Food & Drink Shop | 2 | 38.970185 | -123.688073 |

## Cluster 2

```
In [118]: neighborhood_merged.loc[neighborhood_merged['Cluster Labels'] == 1, neighborhood_merged.columns[[1] + list(range(5, neighborhood_
```

Out[118]:

| | 1st Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue | Cluster Labels | Latitude | Longitude |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Hotel | Speakeasy | Art Gallery | Thai Restaurant | Indian Restaurant | Hostel | Spa | 1 | 37.787483 | -122.412150 |
| 2 | Business Service | Concert Hall | Fish & Chips Shop | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | 1 | 34.582770 | -117.409215 |
| 3 | Burger Joint | Filipino Restaurant | Fish & Chips Shop | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | 1 | 41.193780 | -120.945245 |
| 5 | Bakery | Convenience Store | Grocery Store | Café | Food Court | Fish Market | Flea Market | 1 | 34.496382 | -118.325635 |
| 6 | Intersection | Fish & Chips Shop | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | Food | 1 | 33.442809 | -116.865024 |

## Cluster 5

In [115]:
```
neighborhood_merged.loc[neighborhood_merged['Cluster Labels'] == 4, neighborhood_merged.columns[[1] + list(range(5, neighborhood_
```

Out[115]:

| | 1st Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue | Cluster Labels | Latitude | Longitude |
|---|---|---|---|---|---|---|---|---|---|---|
| 4 | Home Service | Filipino Restaurant | Fish & Chips Shop | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | 4 | 34.136395 | -118.774535 |
| 66 | Home Service | Fish & Chips Shop | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | Food | 4 | 35.652175 | -118.322581 |
| 87 | Home Service | Fish & Chips Shop | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | Food | 4 | 32.664798 | -117.015962 |
| 104 | Home Service | Fish & Chips Shop | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | Food | 4 | 34.144664 | -118.644097 |
| 321 | Home Service | Fish & Chips Shop | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | Food | 4 | 38.610516 | -122.881341 |
| 331 | Home Service | Fish & Chips Shop | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | Food | 4 | 34.164091 | -118.657837 |
| 332 | Home Service | Filipino Restaurant | Fish & Chips Shop | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | 4 | 34.128344 | -117.208651 |

## Cluster 4

In [116]:
```
neighborhood_merged.loc[neighborhood_merged['Cluster Labels'] == 3, neighborhood_merged.columns[[1] + list(range(5, neighborhood_
```

Out[116]:

| | 1st Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue | Cluster Labels | Latitude | Longitude |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Winery | Filipino Restaurant | Fish & Chips Shop | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | 3 | 38.174640 | -121.278559 |
| 70 | Business Service | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | Food | Food & Drink Shop | 3 | 40.002107 | -121.306633 |
| 80 | Business Service | Fish & Chips Shop | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | Food | 3 | 39.412388 | -121.712751 |
| 166 | Business Service | Fish Market | Flea Market | Flower Shop | Fondue Restaurant | Food | Food & Drink Shop | 3 | 39.264895 | -123.591122 |

# Discussion :

### Analysis of Clusters

#### Cluster 1

Cluster 1 represents location that are near to Park, Church, National Park, Stadium that is good for adventure loving people.

#### Cluster 2

Cluster 2 represents location related to mexican restaurant, Fast food chains, bar that is it is good for gourmand people or those who wants to open a food restaurant.

#### Cluster 3

Cluster 3 represents locations that are near to Grocery store or market that is good for those who want to store near themselves, open stores or to sell their porduct in market.

#### Cluster 4

Cluster 4 represent place where most of people use business services so if someone wants to setup business services for other businesses it is the right choice.

#### Cluster 5

Cluster 5 represent place where most of people use home services so if someone wants to setup home services or shops it is the right choice.

# Result :

From the above discussion we can easily find the location or neighborhood which will be suitable for living as per need of a person whether it may be a immigrant or a business person.