The UFO Route 2019 Capstone Project

The UFO Route Company

Email: manuel.alcantara@pe.ibm.com

Website: ibm.com

Tel: [Telephone]

Address, City, ST ZIP

TABLE OF CONTENTS

INTRODUCTION / BUSINESS PROBLEM

We are a company that offer travel experiences to anywhere in the world where sightings of unidentified flying objects have been documented.

Our regular clients are professional or amateur photographers, researchers, scientists, graphic journalists, writers and students.

Our trips are made in 5 countries in which unexplained phenomena have been reported in the sky, these are:

- Australia
- Canada
- Germany
- United Kingdom
- United States America

For the following season, the marketing department proposed to investigate the documented sightings near **Area 51**, in the state of Nevada in the United States of America.

This submission will eventually become your Introduction / Business Problem section in your final report. So I recommend that you push the report (having your Introduction / Business Problem section only for now) to your Github repository and submit a link to it.

DESCRIPTION OF THE DATA

📊 **Kaggle.com**

Next, using Kaggle.com datasets, we study and identify the geographic areas with the largest number of reported occurrences.

From our research, we identified 2 cities in the **State of Nevada** with the highest concentration of sightings, these are:

- 🎸 Reno
- 🎰 Las Vegas

🗺️ **Foursquare.com**

After our travelers finish the tour around the sighting areas and using Foursquare.com, we find hotels, restaurants and other interesting places to visit.

This submission will eventually become your Data section in your final report. So I recommend that you push the report (having your Data section) to your Github repository and submit a link to it.

2.1. Introduction where you discuss the business problem and who would be interested in this project

We can group our needs into two main objectives, which are:

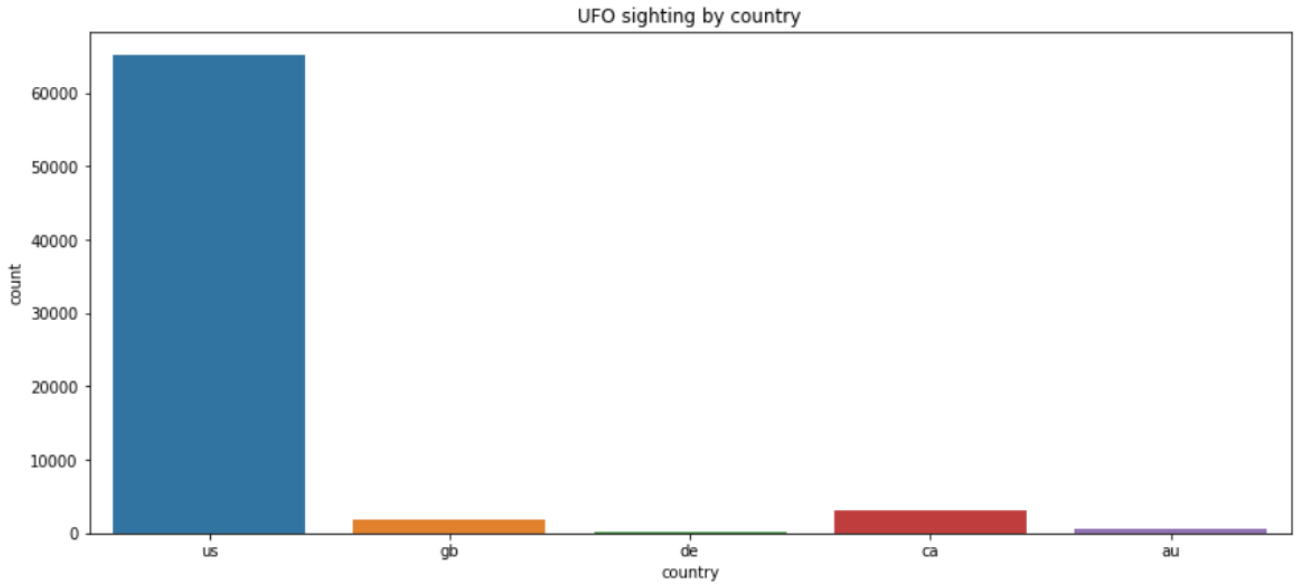
1. Due to the interest of our clients to take pictures of the sky of the state of Nevada (where is the famous Area 51), we want to find the cities of Nevada near the sighting points.
2. After finding the neighborhoods closest to the sighting points, we need to locate hotels and restaurants.

2.2. Data where you describe the data that will be used to solve the problem and the source of the data

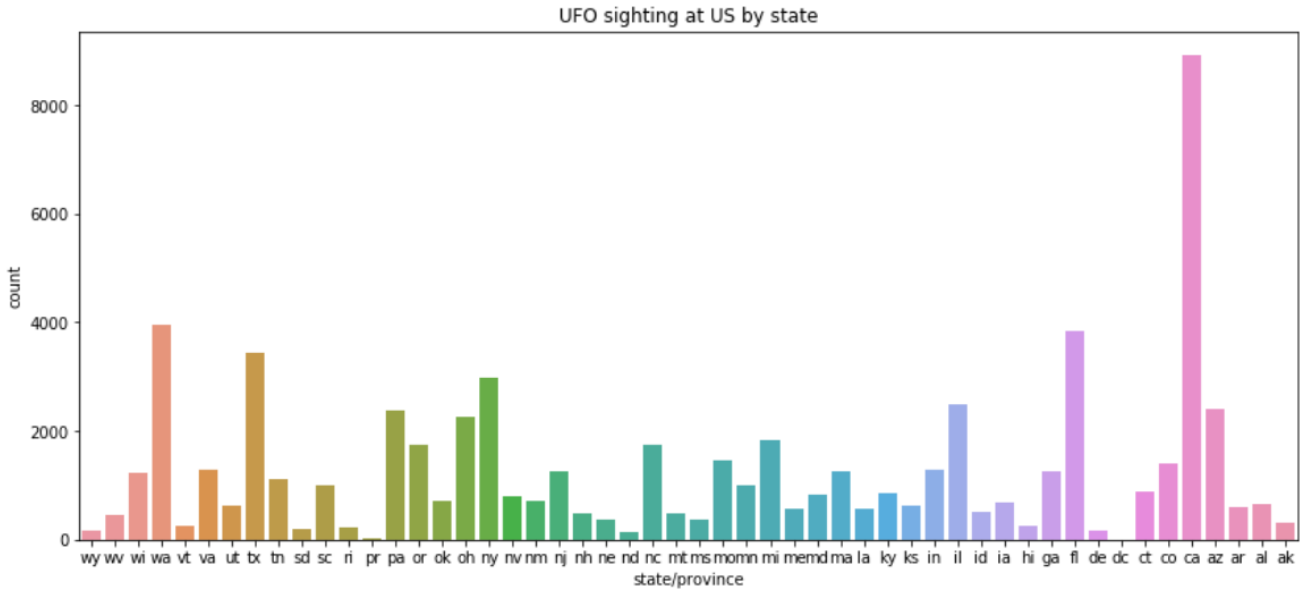
We use data from Kaggle.com and Foursquare.com

2.2.1. 📊 UFO sighting data from Kaggle.com

https://www.kaggle.com/NUFORC/ufo-sighting



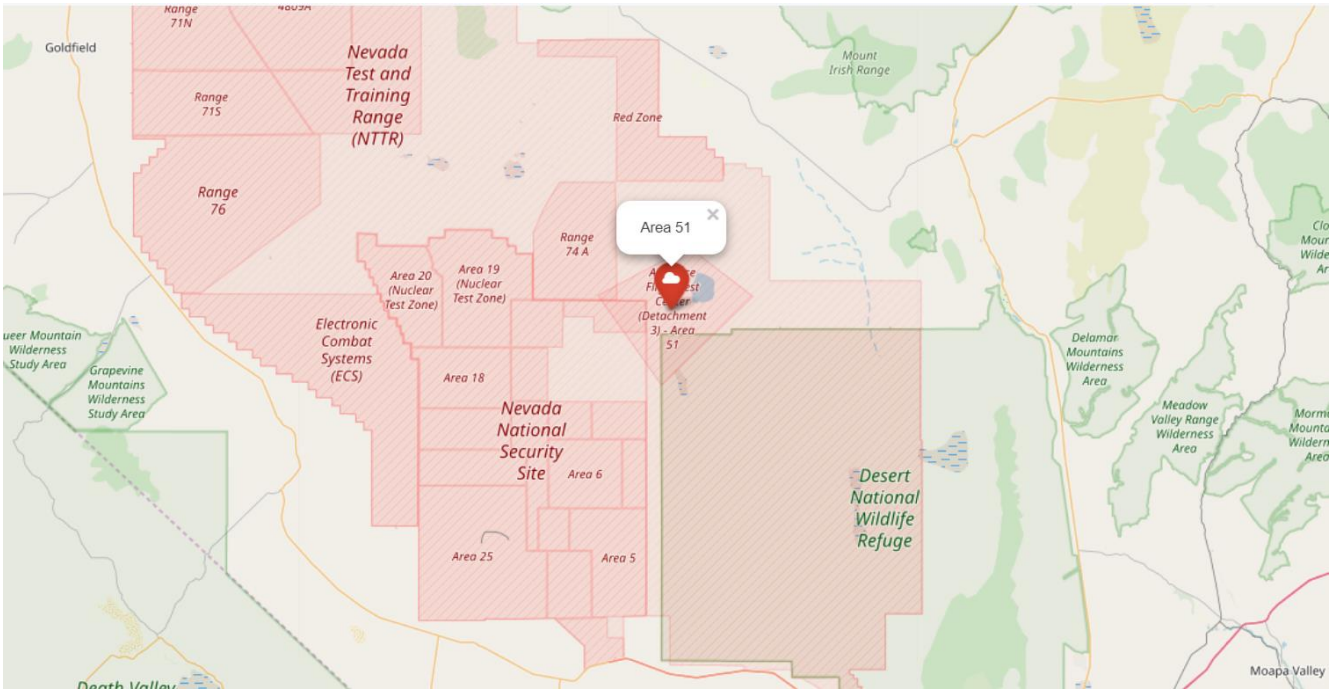
Rank	Country	Frecuency
1.	United States America (US)	65114
2.	Canada (CA)	3000
3.	United Kingdom (GB)	1905
4.	Australia (AU)	538
5.	Germany (DE)	105


UFO sighting at US by state
count
8000
6000
4000
2000
0
wy wv wi wa vt va ut tx tn sd sc ri pr pa or ok oh ny nv nm nj nh ne nd nc mt ms momn mi memd ma la ky ks in il id ia hi ga fl de dc ct co ca az ar al ak
state/province

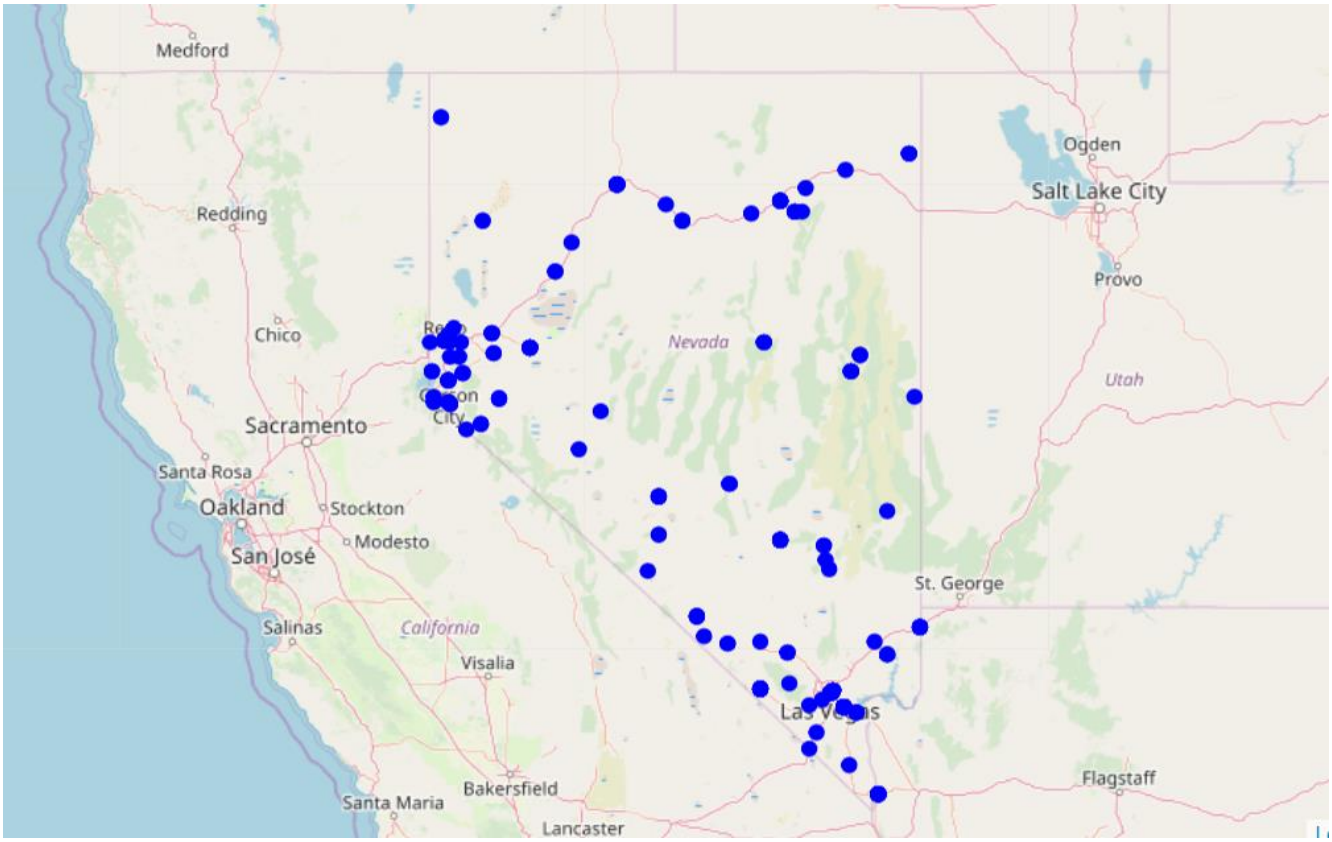

Medford
Redding
Chico
Reno
Nevada
Carson City
Sacramento
Santa Rosa
Oakland
Stockton
Modesto
San José
Salinas
California
Visalia
Las Vegas
Ogden
Salt Lake City
Provo
Utah
St. George

Area 51, Nevada

In [14]:
```
# Area 51
lat_51=37.237080
lon_51=-115.871727

a51_map = folium.Map(location=[lat_51,lon_51], zoom_start=9)
folium.Marker(
    location=[lat_51,lon_51],
    popup= 'Area 51',
    icon=folium.Icon(icon='cloud', color='red')
).add_to(a51_map)

a51_map
```



	Date_time	city	state/province	country	UFO_shape	length_of_encounter_seconds	described_duration_of_encounter	description	date_documented	latitude	longitude
0	10/10/1998 17:30	las vegas	nv	us	cigar	600	10 minutes	White&#44 vertical cigar shape floating around...	11/1/1998	36.175	-115.136389
1	10/10/1998 17:30	las vegas	nv	us	circle	2700	45min.	Ufo sighting in las vegas near Area51	8/30/1999	36.175	-115.136389
2	10/10/1999 21:00	rachel	nv	us	light	10800	3 hours	Bright lights with incredible agility seen fro...	5/24/2005	37.6447222	-115.742778
3	10/10/2008 04:00	carlin	nv	us	disk	600	10 mins	Ufo sighting in Carlin&#44 NV&#44 at the pilot...	7/19/2010	40.7138889	-116.103056
4	10/11/1999 21:00	winnemucca	nv	us	cigar	2	seconds	oblong&#44 extremely large and bright object ...	10/19/1999	40.9730556	-117.734722



Clustering Nevada locations

In [22]:
```
data_ufo_nevada.head()
```

Out[22]:

	country	state/province	city	UFO_shape	length_of_encounter_seconds	latitude	longitude
0	us	nv	las vegas	cigar	600.0	36.175000	-115.136389
1	us	nv	las vegas	circle	2700.0	36.175000	-115.136389
2	us	nv	rachel	light	10800.0	37.644722	-115.742778
3	us	nv	carlin	disk	600.0	40.713889	-116.103056
4	us	nv	winnemucca	cigar	2.0	40.973056	-117.734722

🔴↔🔵 K-Means

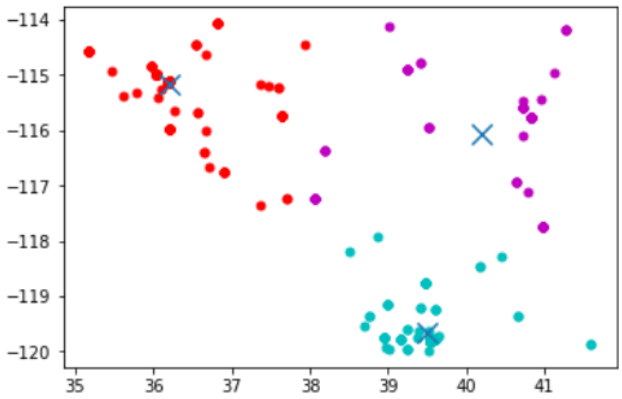
In [26]:
```
#
n_clusters = 3

kmeans=KMeans(n_clusters=n_clusters)
kmeans=kmeans.fit(X)
labels=kmeans.predict(X)

print(kmeans)
```

```
KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=300,
    n_clusters=3, n_init=10, n_jobs=1, precompute_distances='auto',
    random_state=None, tol=0.0001, verbose=0)
```

In [27]:
```
centroids=kmeans.cluster_centers_
colors=['m.','r.','c.','b.','y.']

print(centroids)
```

```
[[  40.18513207 -116.0772632 ]
 [  36.21324354 -115.16784237]
 [  39.49137706 -119.66651597]]
```

In [28]:
```
#Plot chart without Folium map
for i in range(len(X)):
    plt.plot(X[i][0],X[i][1], colors[labels[i]], markersize=10)

# centroids
plt.scatter(centroids[:,0], centroids[:,1], marker='x', s=150, linewidths=5, zorder=10)

# Show plot
plt.show()
```

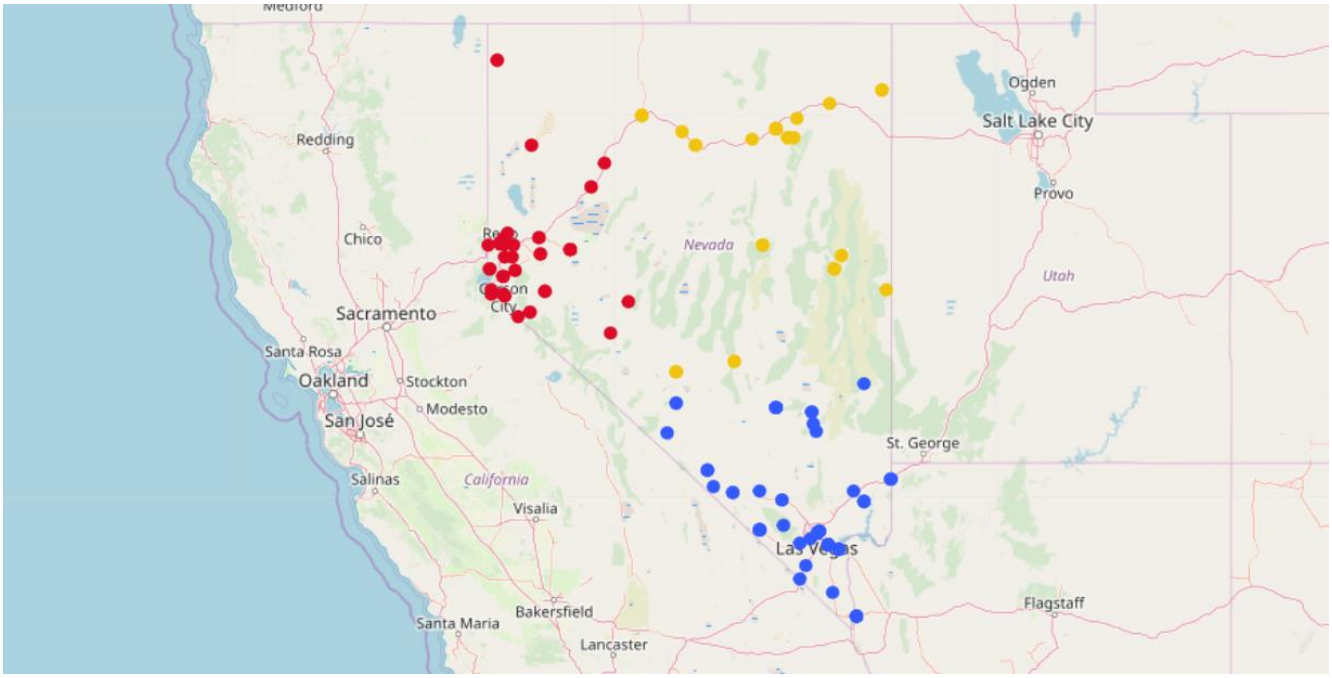


🔰 Mapping Cluster Table

In [31]:
```
data_ufo_nevada_clusters = pd.DataFrame([data_ufo_nevada_ll.index,data_ufo_nevada_ll['latitude'],
                                         data_ufo_nevada_ll['longitude'],labels]).T
data_ufo_nevada_clusters.columns = ['last_index', 'latitude', 'longitude', 'cluster_group']

data_ufo_nevada_clusters['last_index'] = data_ufo_nevada_clusters.last_index.astype(int)
data_ufo_nevada_clusters['cluster_group'] = data_ufo_nevada_clusters.cluster_group.astype(int)

data_ufo_nevada_clusters.head()
```

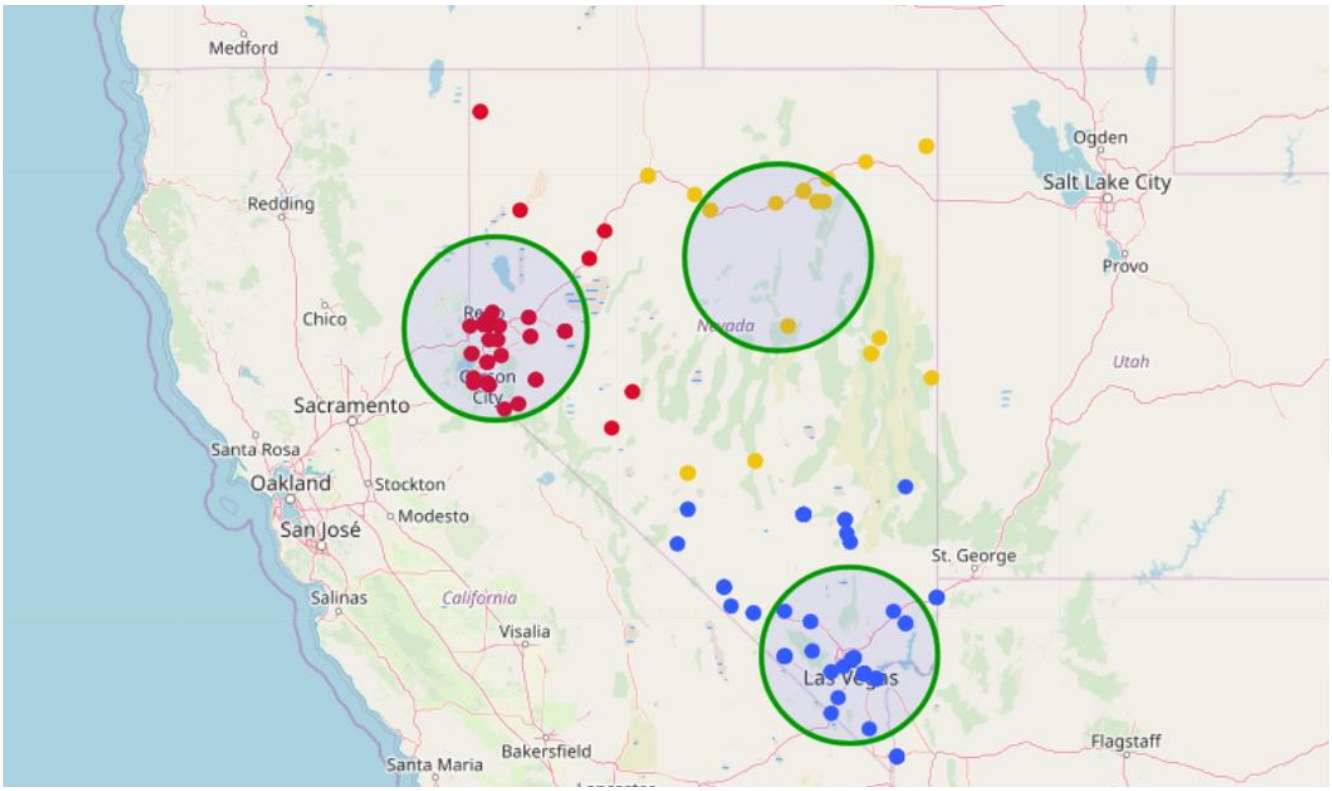
Out[31]:

	last_index	latitude	longitude	cluster_group
0	0	36.175000	-115.136389	1
1	1	36.175000	-115.136389	1
2	2	37.644722	-115.742778	1
3	3	40.713889	-116.103056	0
4	4	40.973056	-117.734722	0



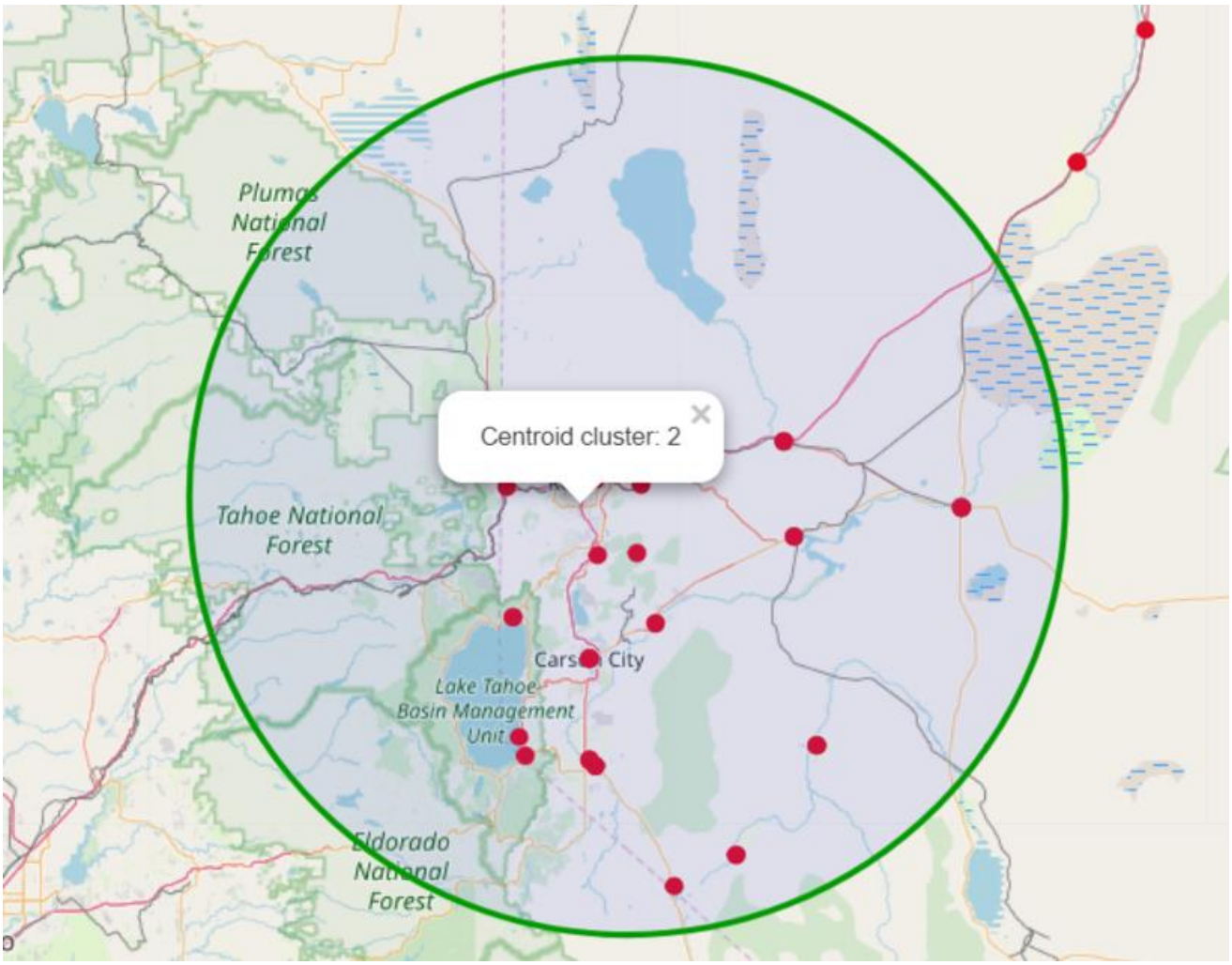
🎯 Mapping Centroids in Cluster Table

In [35]:
```
n=0
for row in centroids:
    la=row[0]
    lo=row[1]
    label2= 'Centroid cluster: {}'.format(n)
    label2 = folium.Popup(label2, parse_html=True)
    folium.Circle(
        [la, lo],
        radius=100000,
        popup=label2,
        color= '#009900',
        fill=True,
        fill_color= rainbow[cluster-1],#'#009900',
        fill_opacity=0.1
    ).add_to(ufo_map_nevada_map)
    n=n+1

ufo_map_nevada_map
```



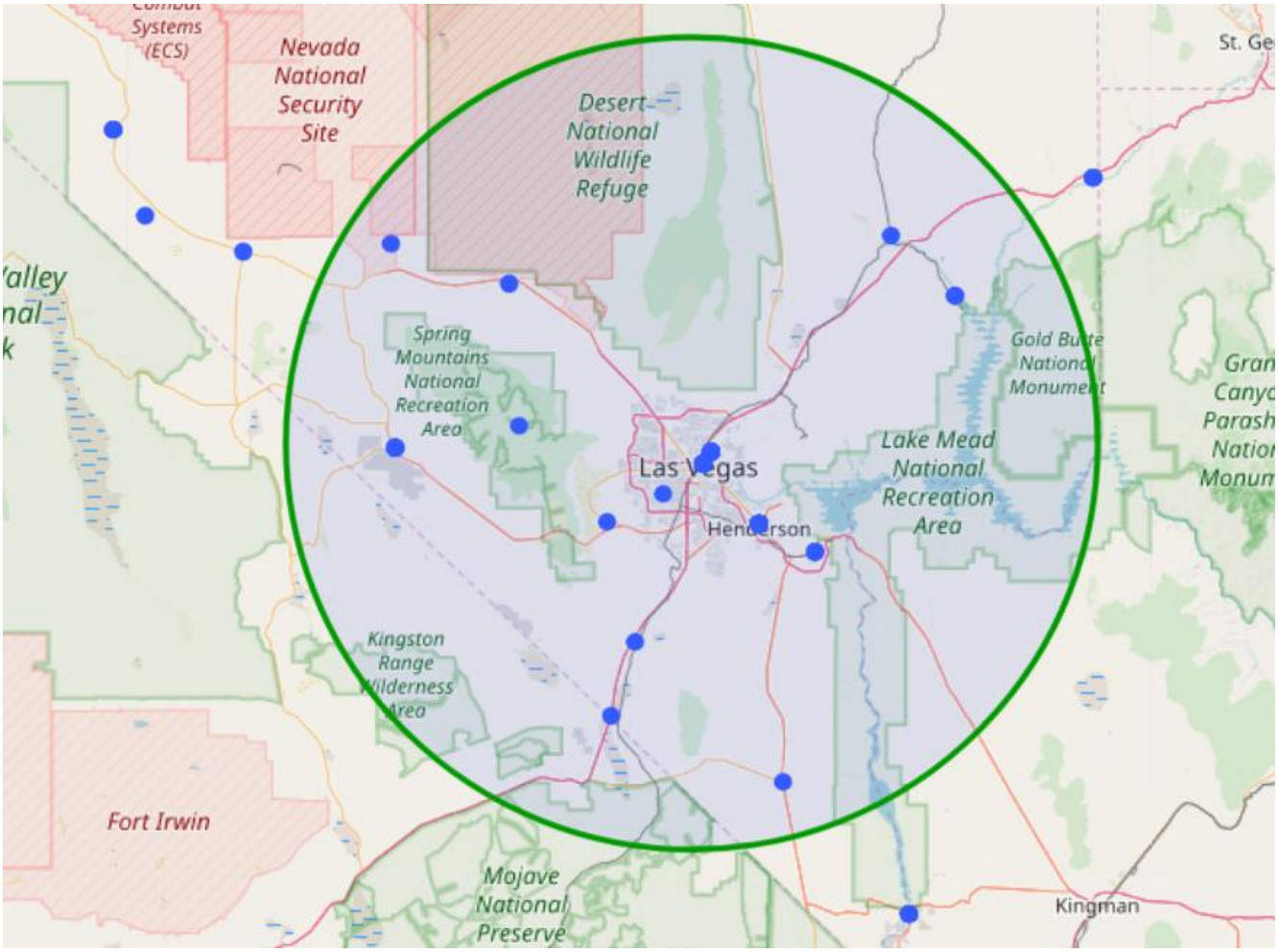
Reno City, NV

In [36]:
```
# Reno
lat_Re=39.52927
lon_Re=-119.8136744

ufo_map_nevada_map.location=[lat_Re,lon_Re]
ufo_map_nevada_map.zoom_start=9
ufo_map_nevada_map
```



🎰 Las Vegas City, NV

In [37]:
```
# Las Vegas
lat_VE=36.1662859
lon_VE=-115.149225

ufo_map_nevada_map.location=[lat_VE,lon_VE]
ufo_map_nevada_map.zoom_start=9
ufo_map_nevada_map
```



2.2.2. Reno & Las Vegas data from Foursquare.com

Define Foursquare Credentials and Version

In [38]:
```
#CLIENT_ID = 'your-client-ID' # your Foursquare ID
#CLIENT_SECRET = 'your-client-secret' # your Foursquare Secret

CLIENT_ID = 'ZFKLYJXKXTAYKUCHPGQ0J10WY4F0D4XCDUX3P0KW3BHA3GKI' # your Foursquare ID
CLIENT_SECRET = '3ZTDEP4OA0D0PSZRA4EPZENOX3BI2NSNACSA3KGWI25JWE1E' # your Foursquare Secret
VERSION = '20180605' # Foursquare API version

print('My credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
My credentails:
CLIENT_ID: ZFKLYJXKXTAYKUCHPGQ0J10WY4F0D4XCDUX3P0KW3BHA3GKI
CLIENT_SECRET:3ZTDEP4OA0D0PSZRA4EPZENOX3BI2NSNACSA3KGWI25JWE1E
```

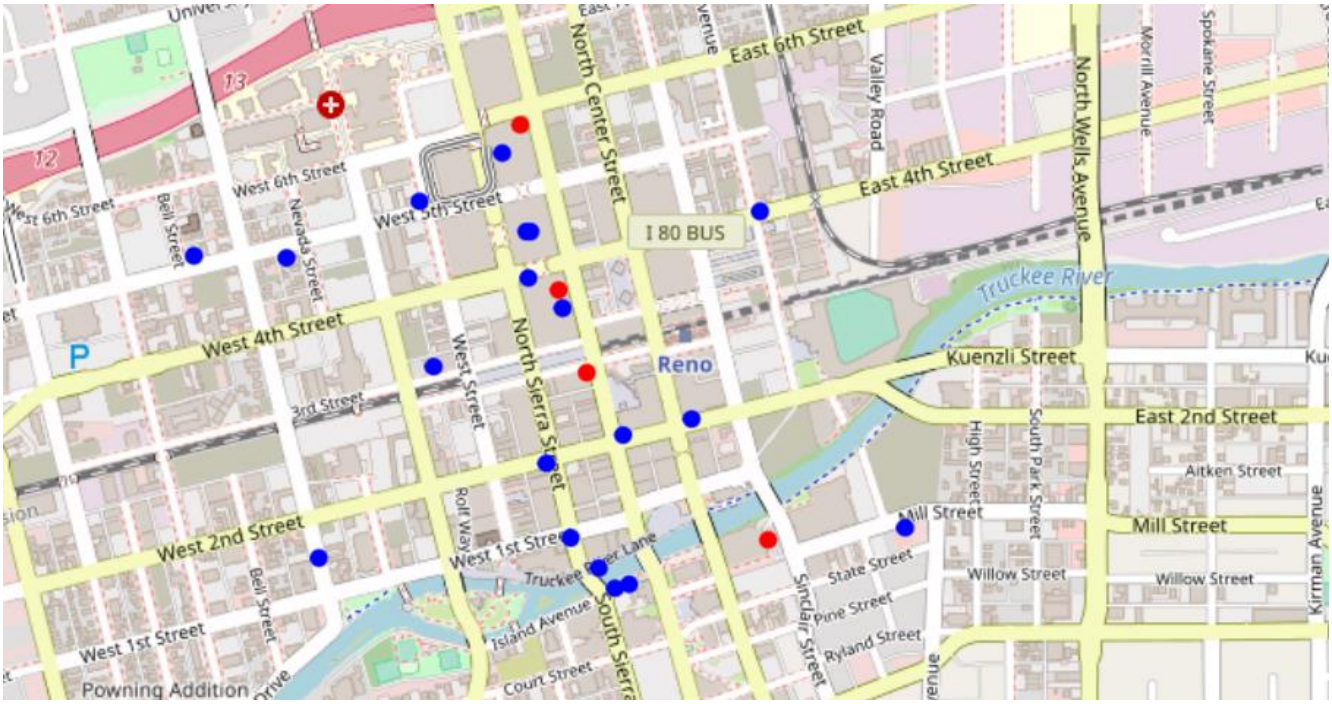
Looking for recommended hotels and restaurants in the cities of Reno and Las Vegas

Looking for	Reno	Las Vegas
Hotels	▫	▫
Restaurants	▫	▫

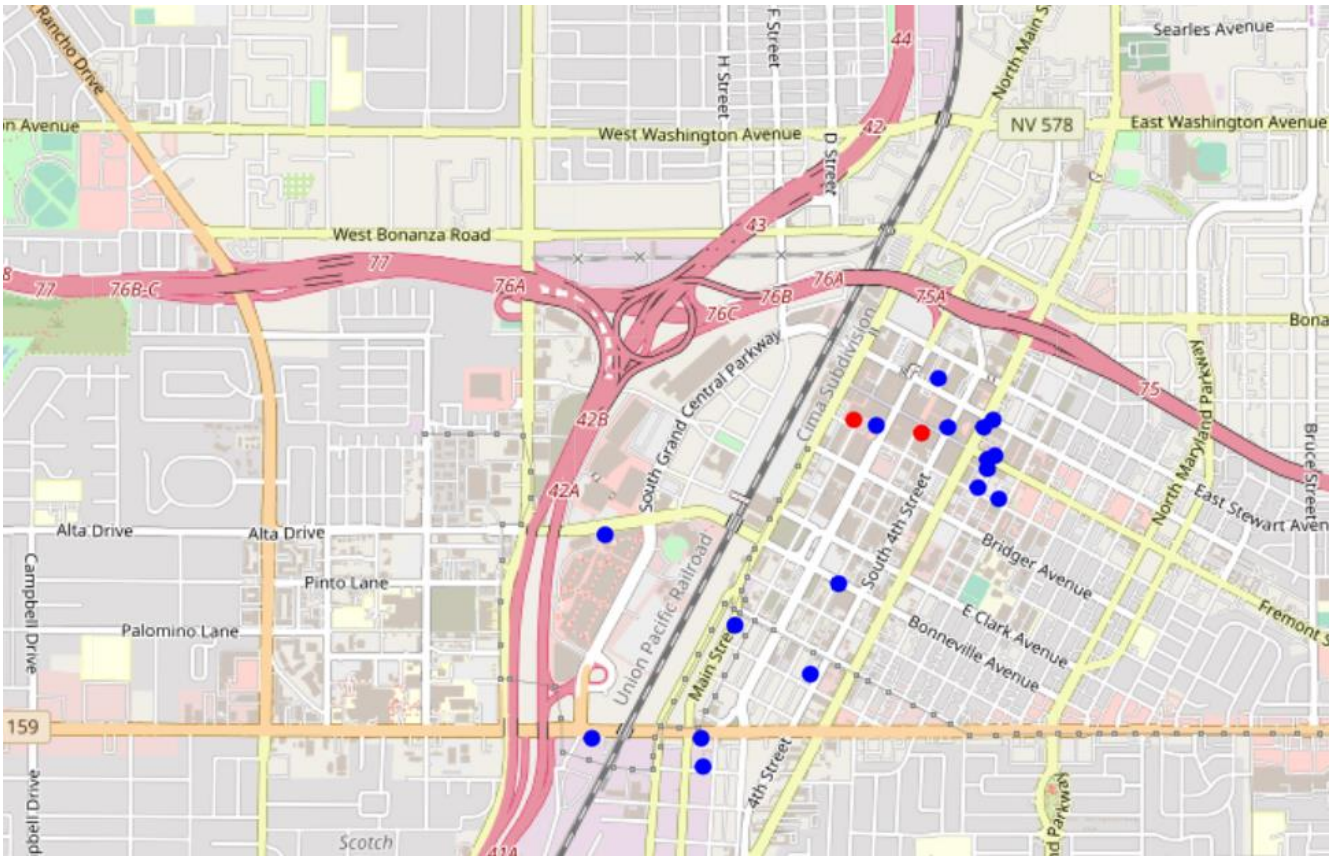
Reno City, NV

In [39]:
```
# Manu Code
city1 = 'Reno, NV'

geolocator = Nominatim(user_agent='0')
location_RE = geolocator.geocode(city1)
latitude_RE = location_RE.latitude
longitude_RE = location_RE.longitude
print('The geograpical coordinate of Reno City are {}, {}.'.format(latitude_RE, longitude_RE))
```

```
The geograpical coordinate of Reno City are 39.52927, -119.8136744.
```



Las Vegas City, NV



Looking for	Reno	Las Vegas
Hotels	4	2
Restaurants	25	17

2.3. Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, and what machine learnings were used and why.

1. Business understanding:

During this stage the need of the business was identified which was to identify the hotels and restaurants near the places of sightings in the state of Nevada in the United States of America.

2. Analytic approach:

For this stage it has only been considered to use the dataset about sightings of UFOs of Kaggle and the data of the cities obtained in Foursquare. From both sources, only data concerning the state of Nevada has been analyzed.

3. Data requirements:

The data has been imported from its origins in CSV formats.

4. Data collection:

In addition to the data obtained in Kaggle and Foursquare, other data sources were not necessary.

5. Data understanding:

Considering the imported data, the most valuable were the coordinates (latitude and longitude) of the described places, either of the places of sightings or of the hotels and restaurants located.

6. Data preparation:

The rows with null values were eliminated and the data types of each column were adapted to be able to process them.

7. Modeling:

Para identificar la densidad de las ubicaciones se utilizó el método de agrupamiento k-means.

2.4. Results section where you discuss the results.

In the analysis of the data, the following were identified:

- The country with the highest number of sightings is the United States of America.
- The state of the USA with the highest number of sightings is California with 13%, followed by the state of Washington with 6.09%, Florida with 5.89%. Nevada only represents 1.23% of the total.

- Of the data of sightings corresponding to Nevada, when executing K-means with 3 clusters, 2 centroids were located near the cities of Reno and Las Vegas, the third centroid appeared in the northeast of Nevada, far from any city.
- For Reno city, with a radius of 2000, 4 hotels and 25 restaurants were found.
- For Las Vegas city, with a radius of 2000, 2 hotels and 17 restaurants were found.

2.5. Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.

- I recommend running the K-Means command with a larger number of clusters to identify a larger number of centroids and, consequently, a larger number of cities.
- I also recommend expanding the value of the range of data obtained on Foursquare because we obtained a small number of hotels.

2.6. Conclusion section where you conclude the report.

- The data obtained in Kaggle are not necessarily reliable, however, they allow to study in various scientific fields.
- The data obtained from Foursquare allow to find necessary places for travelers.
- Area 51 is a military location that does not allow to be visited by tourists, however, you can visit the surrounding areas and other tourist sites.

3. Your choice of a presentation or blogpost. (10 marks)

See the presentation file.