Capstone Project - The UFO route 2019

Manuel David Alcantara

Table of Contents

I. Week 1

1. Description of the problem and a discussion of the background. (15 marks)

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.

I. Week 1

2. Description of the data and how it will be used to solve the problem. (15 marks)

📊 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.

II. Week 2

1. This notebook on your Github repository. (15 marks) ¶

In [1]:
```
import numpy as np # library to handle data in a vectorized manner

import pandas as pd # library for data analsysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json # library to handle JSON files

#! pip install geopy
#!conda install -c conda-forge geopy --yes # uncomment this line if you haven't completed the Foursquare API lab
from geopy.geocoders import Nominatim # convert an address into latitude and longitude values

import requests # library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors
import matplotlib.pyplot as plt
```

II. Week 2

2. Full report with the following components (15 marks): ¶

Next, I describe the components of the report:

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

II. Week 2

2.2.1. 📊 UFO sighting data from Kaggle.com

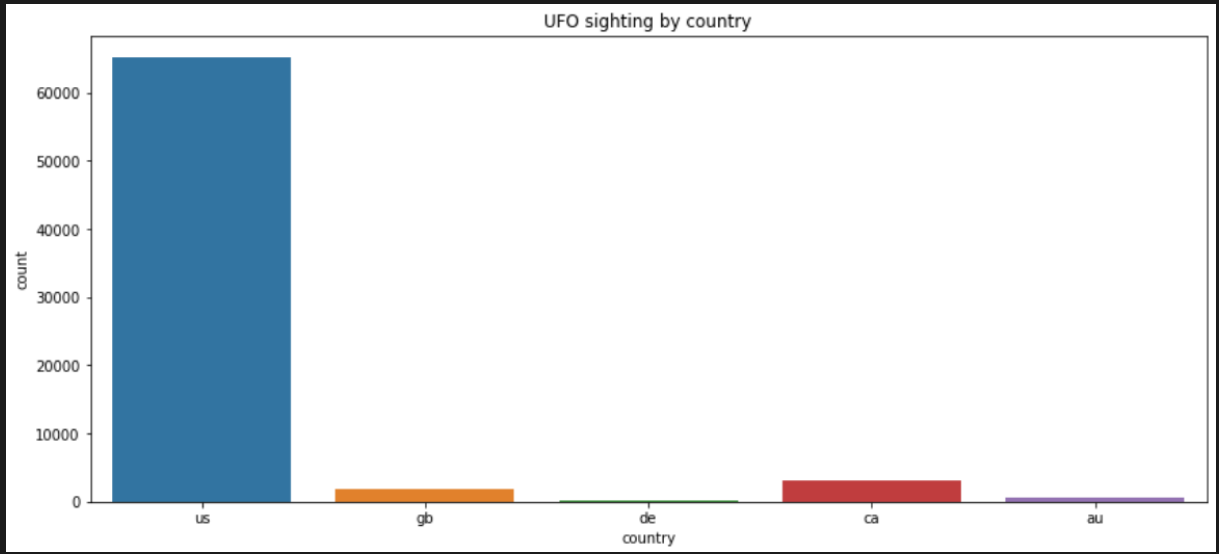
https://www.kaggle.com/NUFORC/ufo-sightings

	Date_time	city	state/province	country	UFO_shape	length_of_encounter_seconds	described_duration_of_encounter	description	date_documented	latitude	longitude
0	10/10/1949 20:30	san marcos	tx	us	cylinder	2700	45 minutes	This event took place in early fall around 194...	4/27/2004	29.8830556	-97.941111
1	10/10/1949 21:00	lackland afb	tx	NaN	light	7200	1-2 hrs	1949 Lackland AFB&#44 TX. Lights racing acros...	12/16/2005	29.38421	-98.581082
2	10/10/1955 17:00	chester (uk/england)	NaN	gb	circle	20	20 seconds	Green/Orange circular disc over Chester&#44 En...	1/21/2008	53.2	-2.916667
3	10/10/1956 21:00	edna	tx	us	circle	20	1/2 hour	My older brother and twin sister were leaving ...	1/17/2004	28.9783333	-96.645833
4	10/10/1960 20:00	kaneohe	hi	us	light	900	15 minutes	AS a Marine 1st Lt. flying an FJ4B fighter/att...	1/22/2004	21.4180556	-157.803611

II. Week 2

2.2.1. 📊 UFO sighting data from Kaggle.com

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

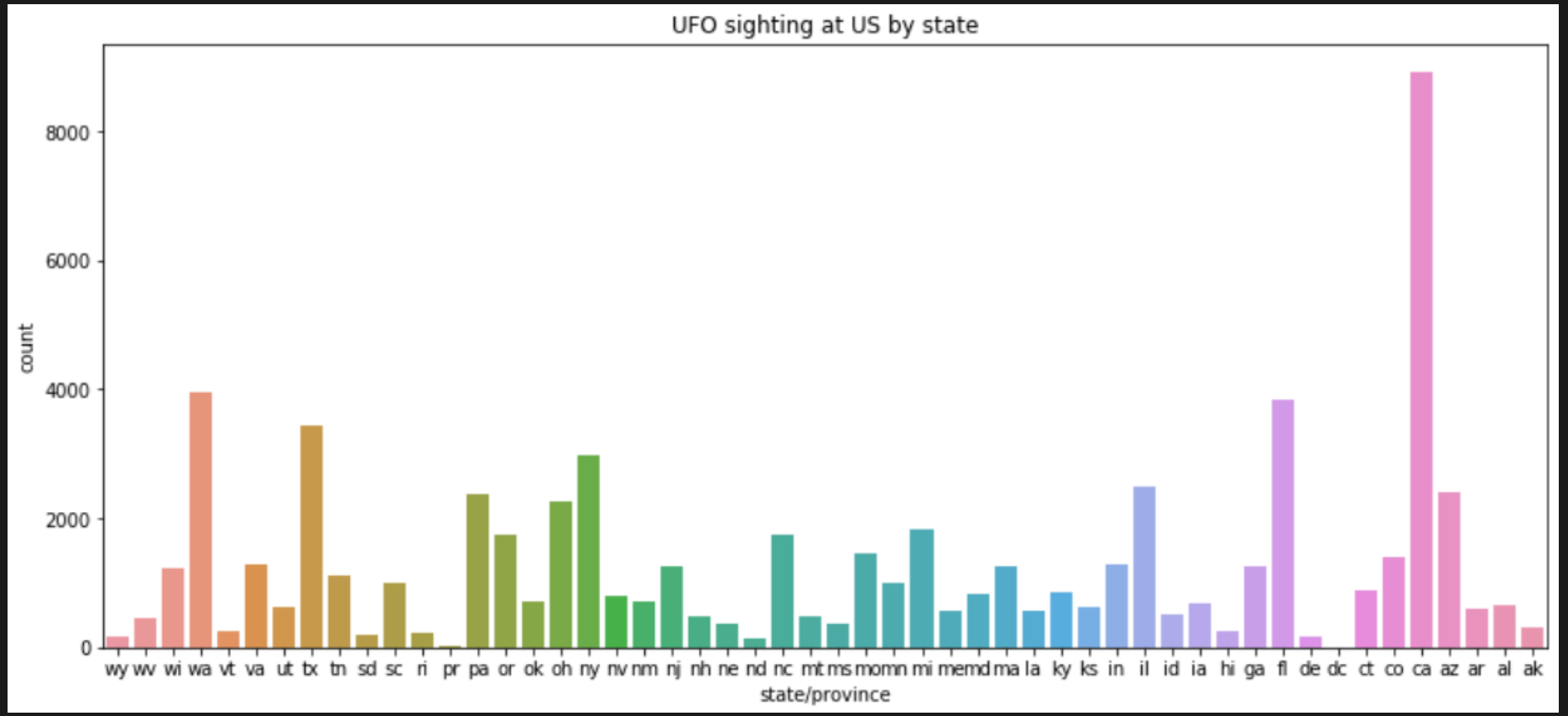


Rank	Country	Frecuency
1.	United States America (US)	65114
2.	Canada (CA)	3000
3.	United Kingdom (GB)	1905
4.	Australia (AU)	538
5.	Germany (DE)	105

II. Week 2

2.2.1. 📊 UFO sighting data from Kaggle.com

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



II. Week 2

2.2.1. 📊 UFO sighting data from Kaggle.com

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

Nevada State, US

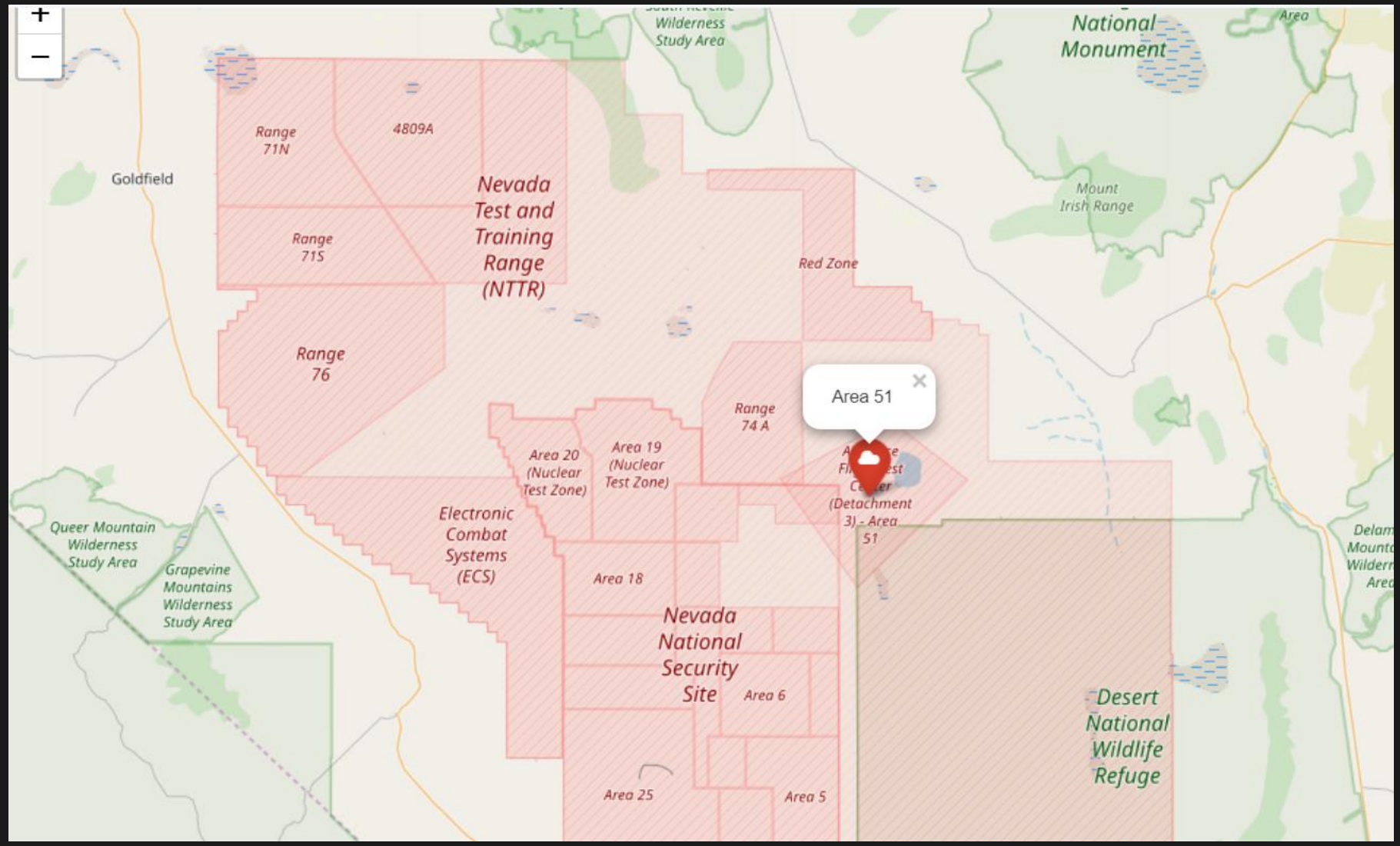
```
# Nevada
lat_NV=39.876019
lon_NV=-117.224121
```

II. Week 2

Area 51, Nevada

2.2.1. 📊 UFO sighting data from Kaggle.com

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

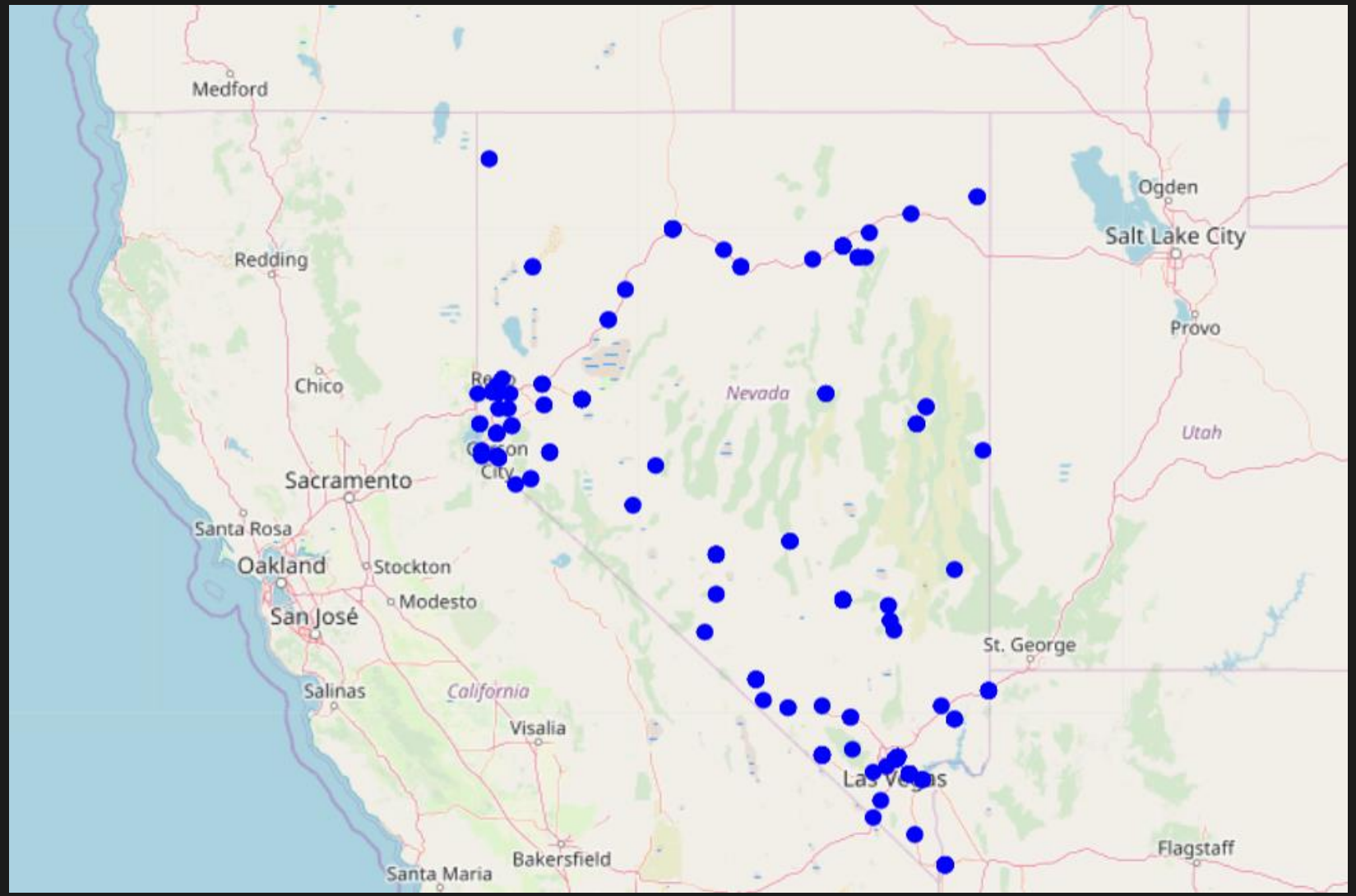


II. Week 2

2.2.1. 📊 UFO sighting data from Kaggle.com

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

UFO sighting at Nevada State



II. Week 2

2.2.1. 📊 UFO sighting data from Kaggle.com

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

Clustering Nevada locations

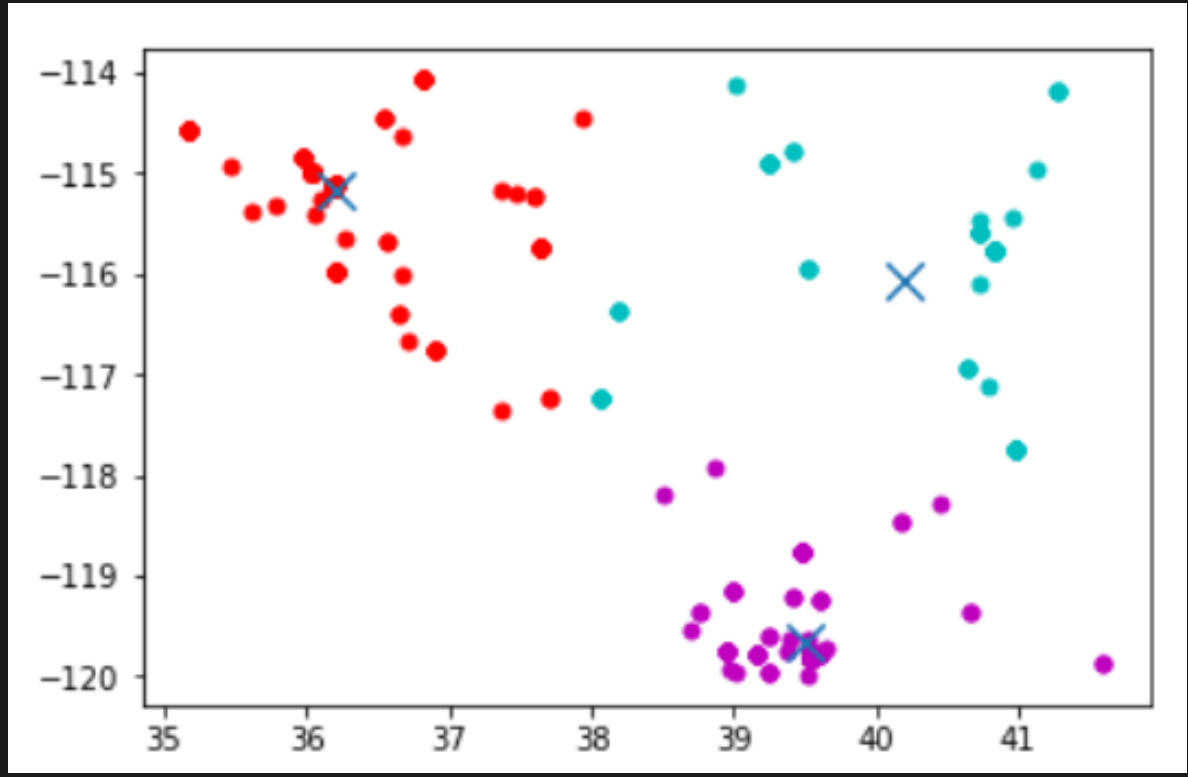
🔴↔🔵 K-Means

```
#
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)
```

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

print(centroids)
```

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



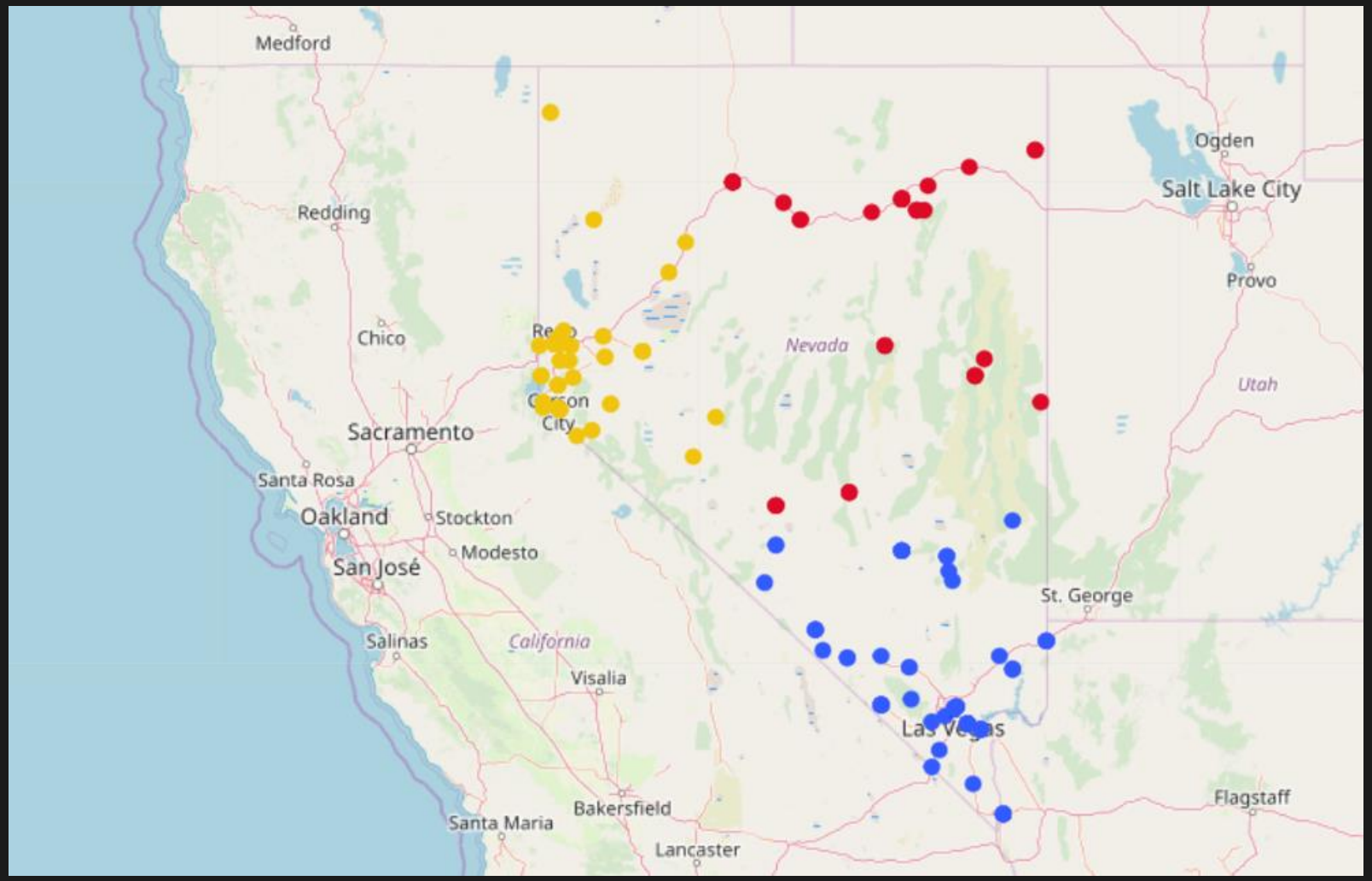
II. Week 2

2.2.1. 📊 UFO sighting data from Kaggle.com

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

🔰 Mapping Cluster Table

	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	2
4	4	40.973056	-117.734722	2



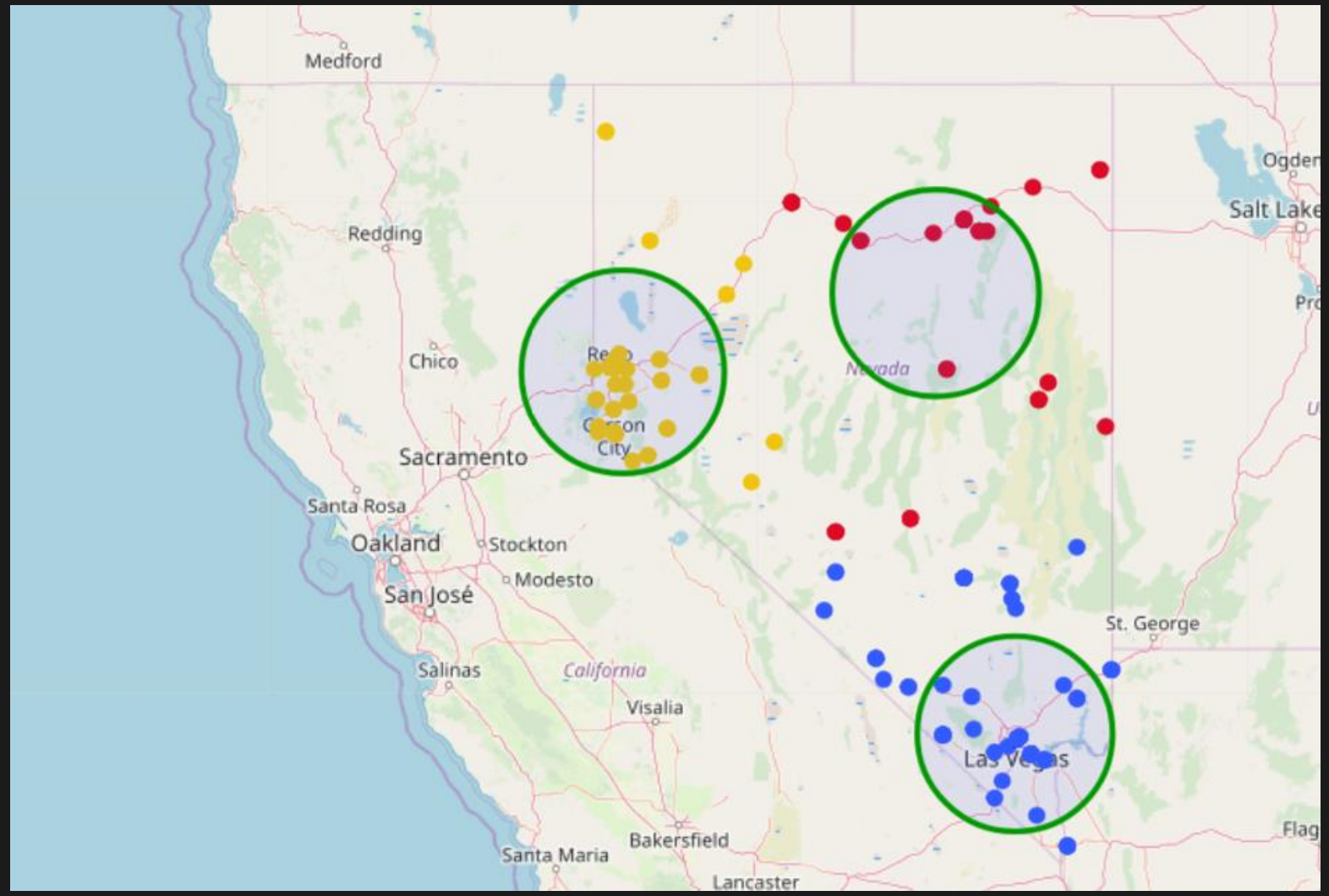
II. Week 2

2.2.1. 📊 UFO sighting data from Kaggle.com

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

🎯 Mapping Centroids in Cluster Table

```
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
```

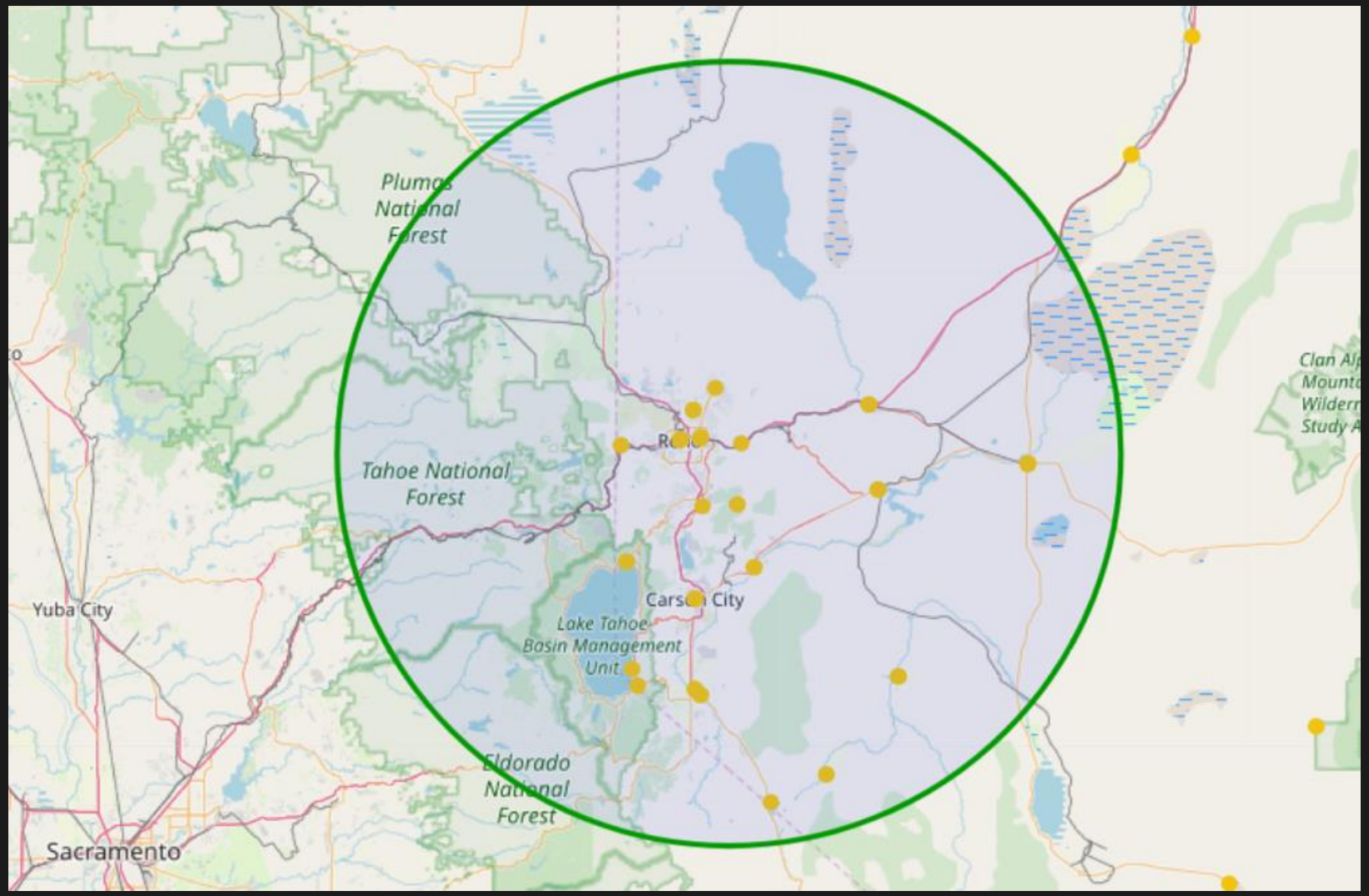


II. Week 2

2.2.1. 📊 UFO sighting data from Kaggle.com

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

🛩 Reno City, NV

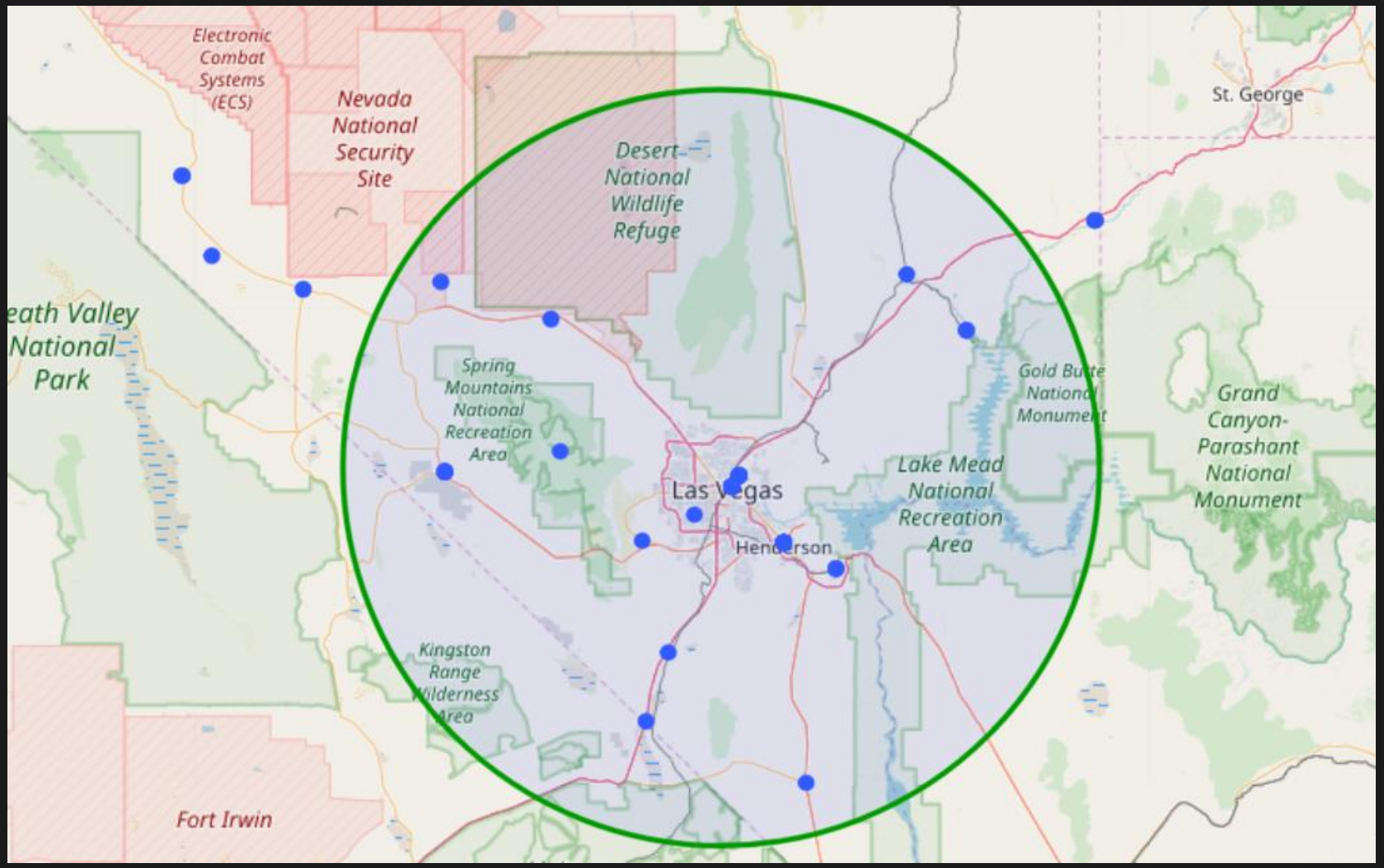


II. Week 2

2.2.1. 📊 UFO sighting data from Kaggle.com

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

🌇 Las Vegas City, NV



II. Week 2

2.2.1. UFO sighting data from Kaggle.com

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

2.2.2. Reno & Las Vegas data from Foursquare.com

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

Looking for	Reno	Las Vegas
Hotels	▫	▫
Restaurants	▫	▫

II. Week 2

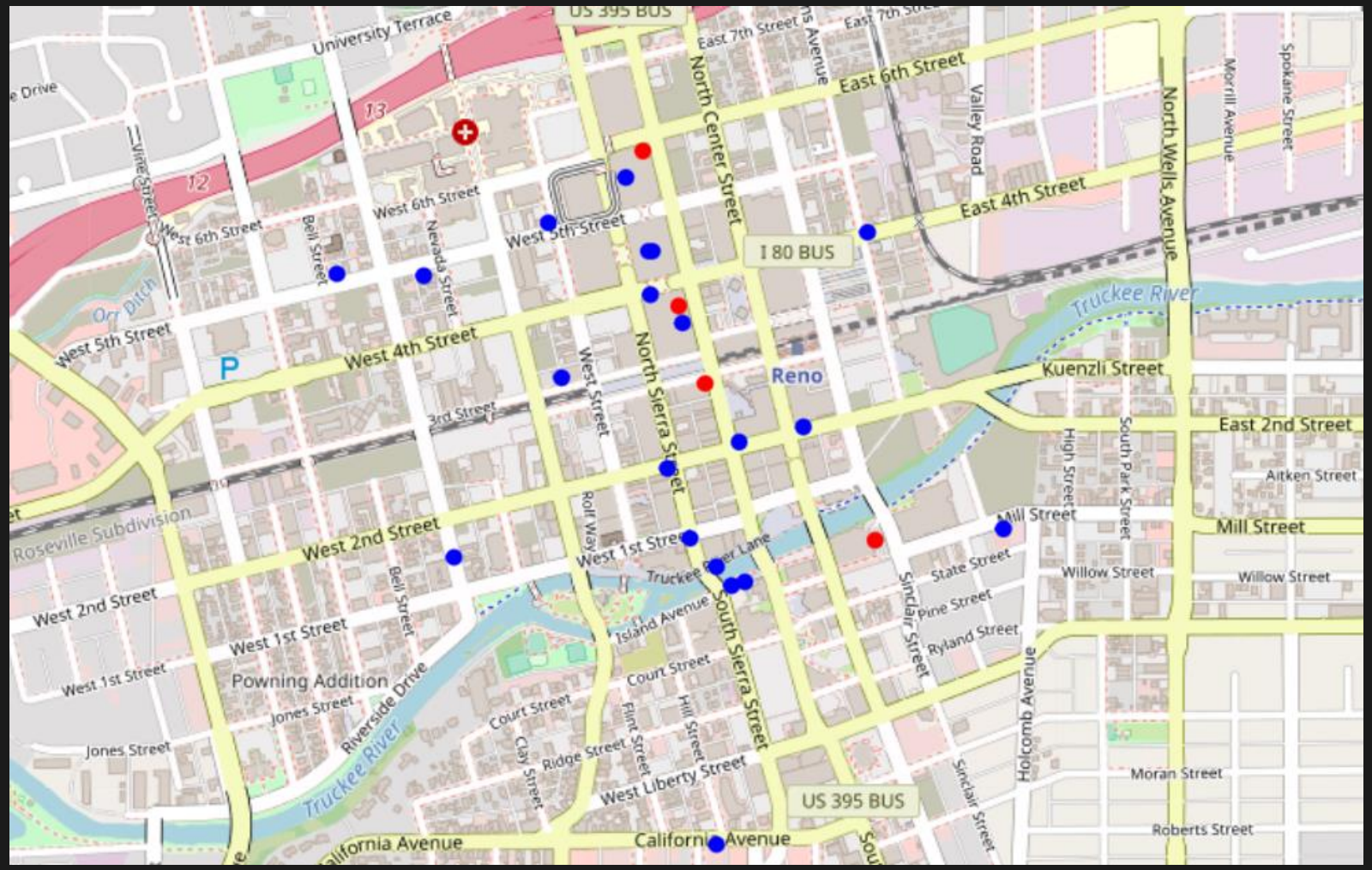
2.2.1. 📊 UFO sighting data from Kaggle.com

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

Reno City, NV

Recommended hotels in Reno City, NV 🔴

Recommended restaurants in Reno City, NV 🔵



II. Week 2

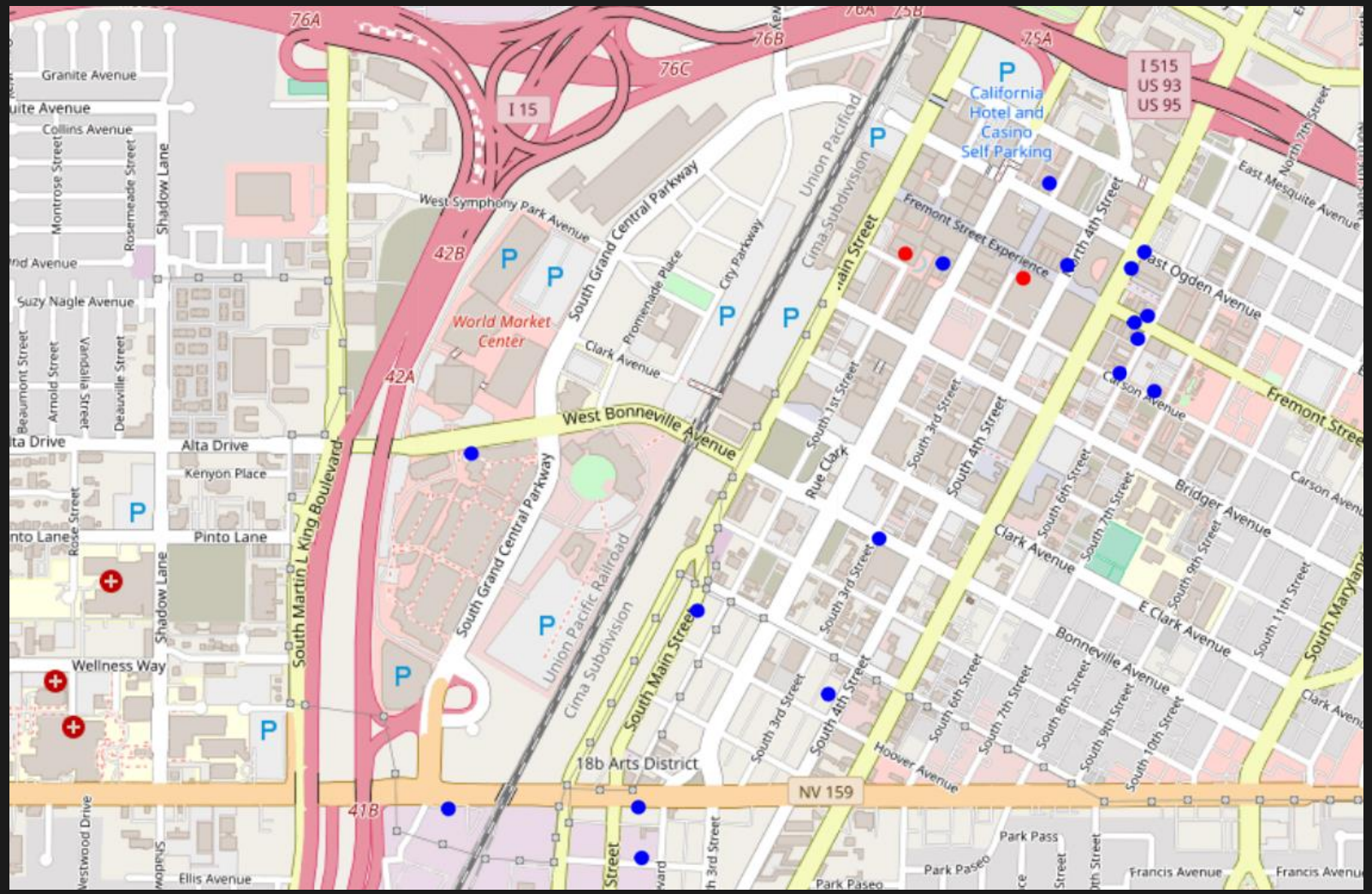
2.2.1. 📊 UFO sighting data from Kaggle.com

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

🎰 Las Vegas City, NV

🎰🔴 Recommended hotels in Las Vegas City, NV

🍽️🎰🔵 Recommended restaurants in Las Vegas City, NV



II. Week 2

2.2.1. 📊 UFO sighting data from Kaggle.com

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

2.2.2. 🗺️ Reno & Las Vegas data from Foursquare.com

Looking for	🎿 Reno	🎰 Las Vegas
🏨 Hotels 🔴	4	2
🍽️ Restaurants 🔵	25	17

II. Week 2

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.

II. Week 2

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.

II. Week 2

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.

THANK YOU!

Manuel David Alcantara