Final Capstone Project

Introduction

In the following project we will analyze the state of Nevada, in this state we will focus on Las Vegas for our first analysis approach, with that information we will be able to know where should we open a restaurant, we decided to evaluate the possibility of opening a restaurant in this state because of the wide range of tourists that come to visit Las Vegas, this place is one of the most visited locations by tourists and Americans, and there might be a possibility of opening with success a restaurant.

Business problem

There are times when people just want to get out and have a good dinner in the company of friends or family, sometimes they visit the same places but they want to try something new. This is why the main reason of the project is to evaluate the possibilities o opening a brand new restaurant where friends and families can gather, but we want to know where is the right location to start the business in the city of Nevada.

Data Description

We will be analyzig the geographical data of Nevada from a dataset that has been downloaded from "https://www.kaggle.com/pavansanagapati/us-wages-via-zipcode?select=free-zipcode-database.csv", and we will only be considering the following variables:

- longitude
- latitude
- City
- Zipcode

After we have analyzed the state we will use the Foursquare API in order to find the most convenient venues throughout the state of Nevada so we can analyze where is the most strategic location to have success on the business.

Data structure

The data was extracted from the Kaggle website, it was downloaded into a csv file and that is how we began to filter the data so we could have the variables we wanted from de state of Nevada.

```
df = pd.read_csv('nevada.csv')
```

```
df.head()
```

	RecordNumber	Zipcode	ZipCodeType	City	State	LocationType	Lat	Long	Xaxis	Yaxis	Zaxis	WorldRegion
0	41350	89001	STANDARD	ALAMO	NV	PRIMARY	37.27	-115.20	-0.33	-0.71	0.60	NaN
1	41351	89001	STANDARD	RACHEL	NV	NOT ACCEPTABLE	37.27	-115.20	-0.33	-0.71	0.60	NaN
2	41352	89001	STANDARD	TEMPIUTE	NV	NOT ACCEPTABLE	37.27	-115.20	-0.33	-0.71	0.60	NaN
3	41353	89017	STANDARD	HIKO	NV	PRIMARY	37.72	-115.33	-0.33	-0.71	0.61	NaN
4	41354	89017	STANDARD	ALAMO	NV	NOT ACCEPTABLE	37.72	-115.33	-0.33	-0.71	0.61	NaN

```
df1 = df[['LocationText','City','Zipcode','Lat','Long']]
```

```
df1.head()
```

Methodology

The data will be analyzed with the tools that we learned in the past weeks about the unsupervised method of clustering with k-means an in that way we will be able to find the perfect location.

Methodology

The data was filtered so we could have the main variables and map them with the geopy library and we focused our map on Las Vegas for a first analysis approach.

```
df1.head()
```

	LocationText	City	Zipcode	Lat	Long
0	Alamo, NV	ALAMO	89001	37.27	-115.20
1	Rachel, NV	RACHEL	89001	37.27	-115.20
2	Tempiute, NV	TEMPIUTE	89001	37.27	-115.20
3	Hiko, NV	HIKO	89017	37.72	-115.33
4	Alamo, NV	ALAMO	89017	37.72	-115.33

Mapping data

As the information was already filtered, we mapped all the zipcodes that we had on our data.
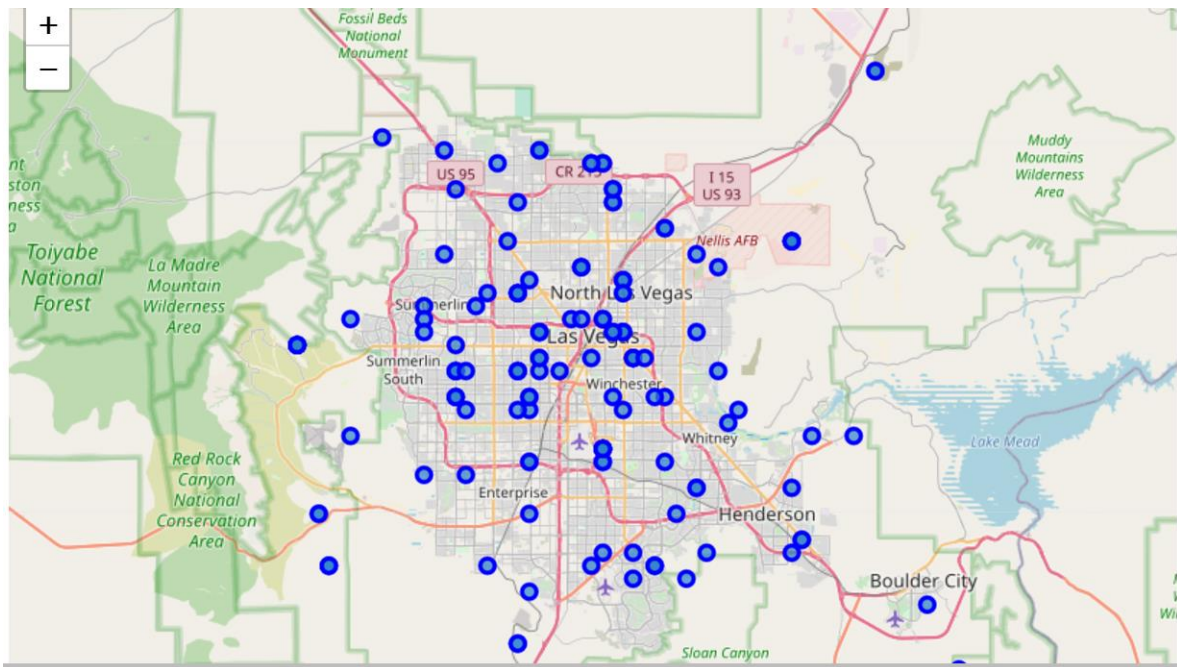
Foursquare API

By this point we used the Foursquare API to manage the data and to gather the venues that were located in the state of Nevada with their own coordinates so we could be able to get them on the map and have our cluster analysis. The categories that we looked for were restaurants, bars, taverns and sports bars.

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	BOULDER CITY	35.96	-114.83	Blue Breed Grill	35.960842	-114.830060	Bar
1	BOULDER CITY	35.96	-114.83	The Dirty Irish Saloon	35.962753	-114.833336	Bar
2	BOULDER CITY	35.96	-114.83	Committed Care Experts Carpet Cleaning	35.959109	-114.835441	Home Service
3	CARSON CITY	39.14	-119.71	Carson River Park	39.141567	-119.706068	Park
4	CARSON CITY MALL	39.14	-119.71	Carson River Park	39.141567	-119.706068	Park

Restaurant data

Now the main idea is to gather the information we have about restaurants and Sports Bar restaurants and group them by city so we can have an idea of how many of this 2 categories can we find on every city of Nevada.

Number of restaurants	
City	
ARDEN	5
BORDERTOWN	19
BOULDER CITY	2
CITIBANK	10
CITIBANK NV	9

| Number of Sports Bar restaurants	
City	
BORDERTOWN	9
BOULDER CITY	2
CITIBANK NV	2
CITY CENTER	1
COLD SPRINGS	9

Clusters

By this point we have one dataframe with all the information about the venues and the coordinates of each of them with their corresponding city and we want to know how many clusters should we do to make the model optimal and we get as a result 2 by the elbow method.
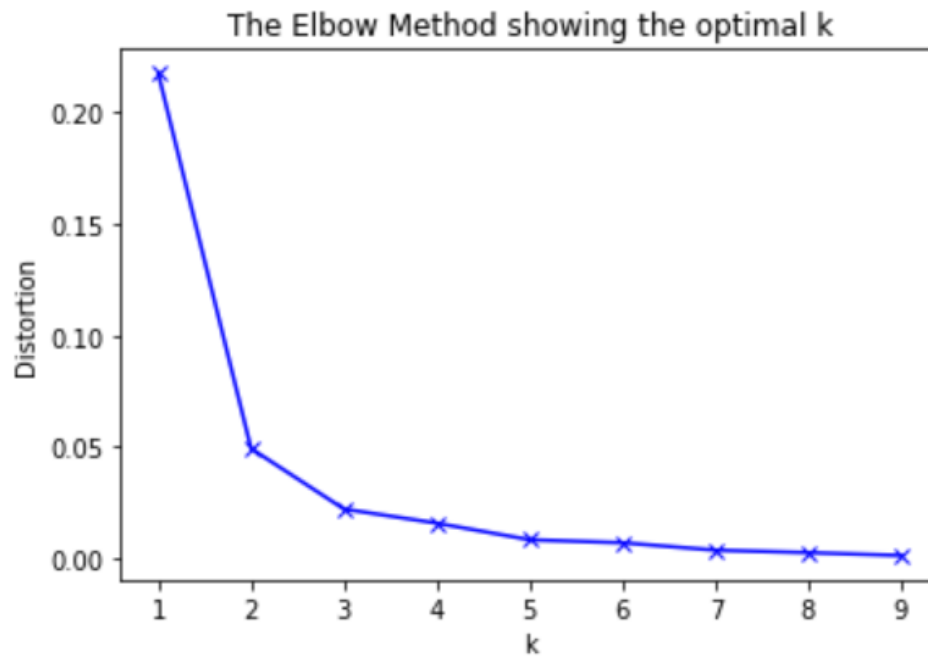
Results

Clusters are organized in 2 different regions, some of them are concentrated in Las Vegas and the others are near Reno, Goldhill, Highlands.
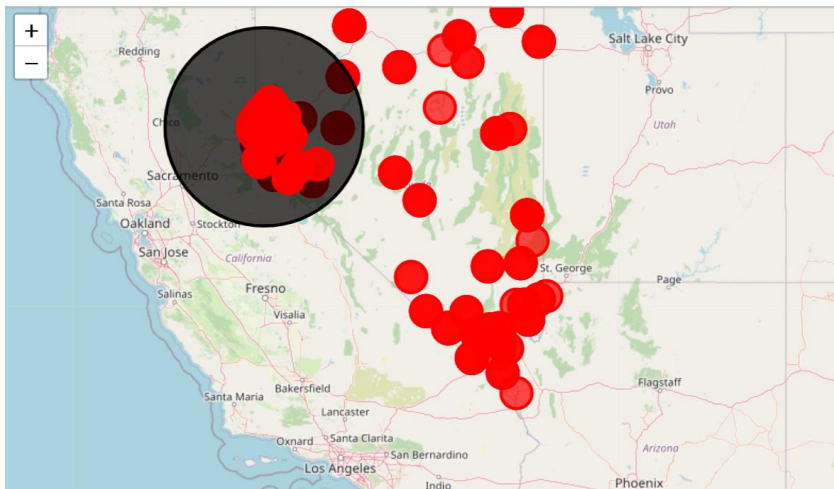
Discussion

The information that was retrieved from the csv and the Foursquare API were essential in the way that we were able to visualize and map the possible areas where the restaurant could be opened with success, clusters are concentrated in 2 different regions and there is a window of opportunity to start a new business near the areas that were previously mentioned.

Conclusions

As we introduce the clusters and the zipcodes of all the cities in Nevada, we can see that we can have an opportunity if we manage to open a restaurant near Reno, Goldhill, Highlands and those cities or neighbourhoods, and there is also a possibility of success in Las Vegas where the rest of the clusters are, if we zoom out on the map we can see that the majority of the clusters are in the multiple cities that I just named, but in Las Vegas there are a lot more of other clusters that might give us an idea of where is most of the population in Nevada.