# 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.

## 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.

## 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.

At the beginning data will be taken from de csv that I retrieved from the Kaggle website, I will filter it into a new data frame where we can access to the initial variables of coordinates and cities so we can map them with the geopy library and then focus primarly on the zipcodes that are contained in Nevada, the main dataset contained information about all the states of the United States, but since we will only be working with the information about Nevada I will filter the data, after all of that is done I will use the Foursquare API to manage the data of restaurants that are in Nevada, I will look for bars, taverns and restaurants I ordered the data so I could see the number of restaurants by city and the number of Sports bar restaurants because I decided that I could use the information of both categories, after that I got the tables joint with the main data frame and used the describe function to get general statistical information about the variables and at the end in order to use the cluster library I eliminated the values that were repeated all over the dataframe and used the ones that had estimated population greater than 0, after that I ran the elbow method and got 2 clusters as a result and mapped them.

## 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.