Introduction

Culiacán is a city in Mexico located in the state of Sinaloa. For many years, sushi restaurants have become really popular. The popularity of these restaurants is so great that the inauguration of a new restaurant of this type is common. Many of these new restaurants close their doors soon, this is caused by different factors, but, one of them and very important is the location.

The objective of this project is to analyze and select the best locations in the city of Culiacan, to open a new sushi restaurant. Using data science methodology and machine learning techniques as a clustering.

Data

The following data will be used:

- List of neighborhoods in Culiacan.
- The latitude and longitude coordinates of those neighborhoods.
- Businesses around neighborhoods, particularly data related to restaurants.

The information of the neighborhoods in Culiacan will be extracted from the page: https://micodigopostal.org/sinaloa/culiacan/. On this page you will find a table with all the neighborhoods and their postal codes. With the help of the BeautifulSoup library, the table with the neighborhoods was scraped.

For the data of the latitude and longitude of these neighborhoods, the Google Maps API will be used.

Finally, to find the places that are in each of the neighborhoods, the foursquare api will be used.

Methodology

The following steps were carried out to carry out this project:

1. The list of neighborhoods in the city of Culiacán was obtained. This information is obtained from the page https://micodigopostal.org/sinaloa/culiacan/. To do the scraping to the page with the library, the requests were made a request and with the help of the beautifulsoup library the information was extracted from the neighborhood table.
2. Once a data frame with the names of the neighborhoods was obtained, the geographic coordinates of each of them were obtained. To do this, the Google Maps API is used, which allows you to search for an address (name of the neighborhood) and receive the geographical coordinates.
3. The Folium library is used to visualize the map of the city with markers that indicate the coordinates of the different neighborhoods.
4. The next step was to use the Foursquare API to get the 100 best places that are within a radius of 2000 meters. A query was made with the application for each neighborhood using the geographic coordinates (latitude and longitude).
5. Foursquare will return the place data in JSON format and we will extract the place name, place category, latitude and longitude of the place within a data frame.
6. Then, we will analyze each neighborhood grouping the rows by neighbor and taking the average of the frequency of occurrence of each place category. Since we are analyzing "sushi restaurant" data, we will filter the income category "sushi restaurants" for neighborhoods.
7. To finish, a clusters was performed on the data using k-means. 3 different groups were selected. Being an unsupervised learning algorithm, it is enough to give the number of clusters and the algorithm to group them in the best way.
8. Clusters are grouped according to their frequency of occurrence for "sushi restaurants".

Results

The neighborhoods are grouped as follows:

- Cluster 0 (red): Neighbourhoods with low number of sushi restaurants.
- Cluster 1 (purple): Neighbourhoods with moderate number to existence of sushi restaurants.
- Cluster 2 (green): Neighbourhoods with high concentration of sushi restaurants.