Clustering Neighborhood in Nicaragua

1. Introduction

a. A description of the problem and a discussion of the background.

The computational advance in the field of data science has provided the opportunity to answer different questions from our environment using databases and machine learning models. Companies and governments invest in research in data science to improve their income statements (companies) and the well-being of their inhabitants (governments).

The objective of our project is to carry out a clustering of the different neighborhoods of Nicaragua based on the types of businesses and places of recreation found in each of the neighborhoods. Our project will allow public institutions to know the internal structure of each cluster to take public policy measures to guarantee the economic growth of the country. It will also allow this clusterization to be offered to private agencies with the aim of attracting private investment to the country. The private company will be able to carry out in its analysis what are the types of businesses that are in certain clusters and thus invest to complement the offer.

2. A description of the data and how it will be used to solve the problem.

Firstly, we will build a database that contains all the departments of Nicaragua with each of its neighborhoods (municipalities), in each of the neighborhood we will locate their respective latitude and longitude.

We don't have this databases already , so we will extract these data from the page https://www.geodatos.net/. That is a webpage that provides latitude and longitude of countries around the world.

In our database we extracted 139 neighborhoods of Nicaragua from 17 departments.

```
In [5]: df = pd.read_excel('geo_nic.xlsx')
        df.head()
```

Out[5]:

	Borough	Neighborhood	Latitude	Longitude
0	Managua	Managua	12.13282	-86.25040
1	Managua	Ticuantepe	12.02263	-86.20493
2	Managua	San Rafael del Sur	11.84854	-86.43839
3	Managua	El Crucero	11.99008	-86.30954
4	Managua	Tipitapa	12.19732	-86.09706

Once we have the latitude and longitude of the differents neighborhoods in Nicaragua we compile business and leisure data for each neighborhoods in Nicaragua, ***we will use the data provided by Foursquare***.

Foursquare API is a location data provider that describe places and venues around the world. Is able to determine what types of venues exist with a defined radius from that location.Also we have access to photos, comments and tips.

```
In [37]: nearby_venues
```

Out[37]:

	name	categories	lat	lng
0	Termales Tipitapa	Hot Spring	12.202835	-86.091691
1	Sopas Mirna	Soup Place	12.191147	-86.100832
2	Asados Guadalupe	BBQ Joint	12.197969	-86.094788
3	Pollo Estrella	Fried Chicken Joint	12.201215	-86.097057
4	Restaurante Silva	Seafood Restaurant	12.201381	-86.096420
5	Bar El Chanchito	Restaurant	12.163399	-86.117496

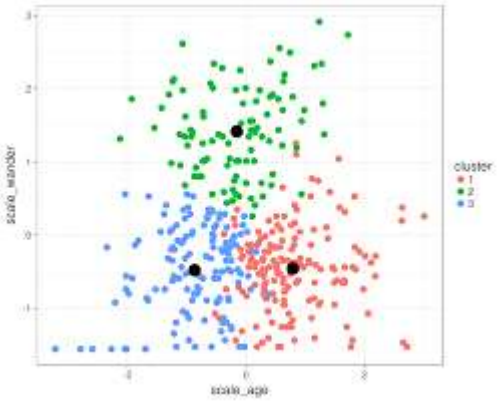
3. Methodology

We will use a machine learning tecnhique called ‘Cluster analysis’.

Cluster analysis or clustering ***is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups.***

We will cluster the differents neighborhoods in Nicaragua based on the differents places in the neighborhood. Doing this, we will discover inside patterns in each cluster. To make this cluster análisis we will use k-means.

K-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster. In other words ***K-means is a simple unsupervised machine learning algorithm that groups data into a specified number (k) of clusters.***

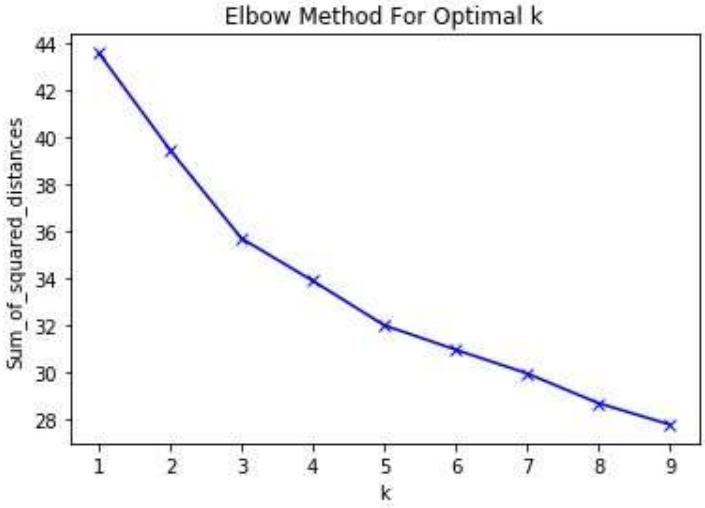


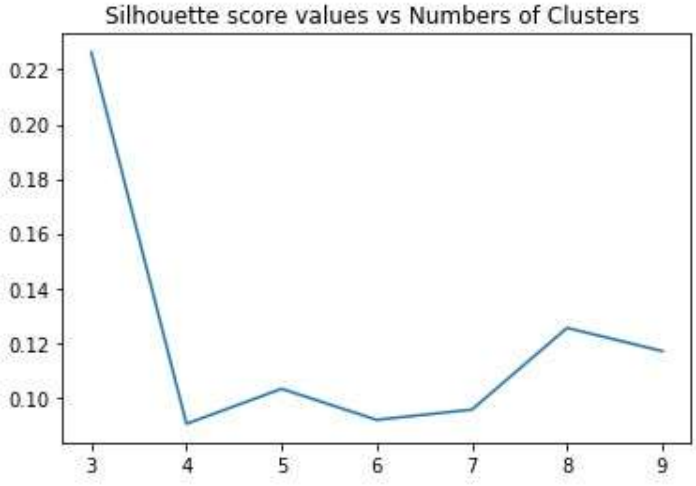
To determine the number of cluster we will use the Elbow method. The elbow method runs k-means clustering on the dataset for a range of values for k and then for each value of k computes an average score for all clusters.

4. Results

There are neighborhoods where there are no leisure places according to the Foursquare database, in addition, these neighborhoods are of small territorial extension. These municipalities were eliminated from the analysis cluster and presented as a cluster where there are no businesses and there is an opportunity to invest because there is an unmet market.

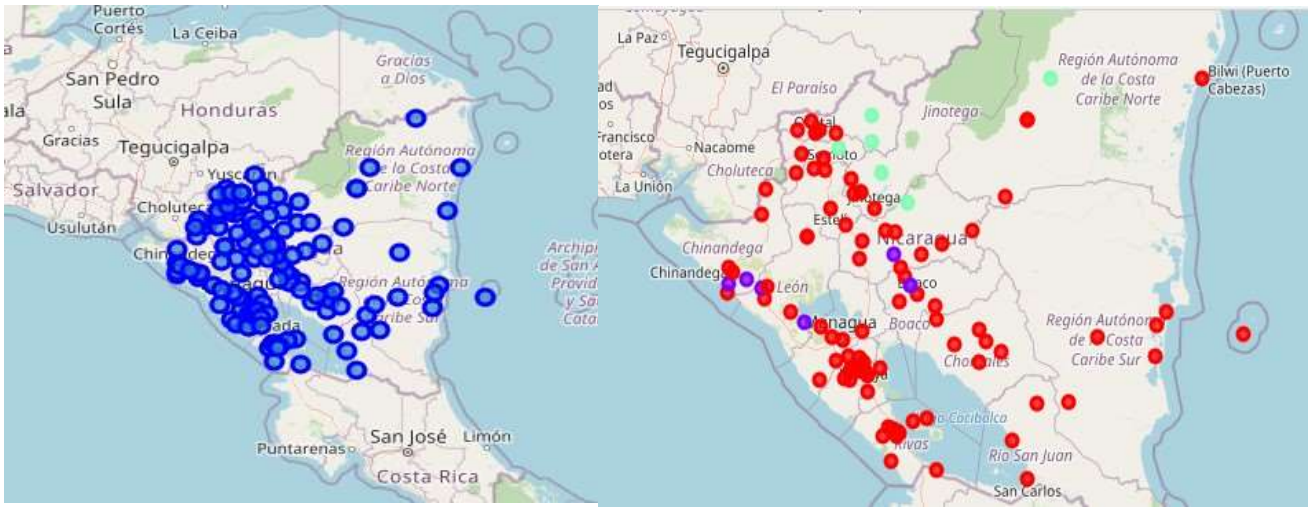
The Elbor method says the number of optimal clusters is 3. So, we run our k-means model with three different clusters k=3






```
Optimal number of components is:
3
```

We plot the differents neighborhoods in a map of Nicaragua, we will visualize the differents clusters. We create the two maps using Folium library.



In the second map, points of the same color are in the same cluster. We can visualize the three differents colors: red, green and purple.

5. Discussion

Among the three clusters we find a majority cluster which reflects the similarity between the neighborhoods of Nicaragua. This cluster presents hotels, restaurants and bars.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Managua	Breakfast Spot	Mexican Restaurant	Pet Store	Chinese Restaurant	Big Box Store	Hotel	Sports Bar	Snack Place	Department Store	Diner
1	Ticuantepe	Caribbean Restaurant	BBQ Joint	Clothing Store	Market	Bar	Dive Bar	Diner	Dessert Shop	Department Store	Cuban Restaurant
2	San Rafael del Sur	Soccer Stadium	Ice Cream Shop	Mobile Phone Shop	Snack Place	Market	Pizza Place	Forest	Food Service	Chinese Restaurant	Church
3	El Crucero	City	Restaurant	Coffee Shop	Yoga Studio	Donut Shop	Church	Clothing Store	Comfort Food Restaurant	Concert Hall	Convenience Store
4	Tipitapa	Seafood Restaurant	Fruit & Vegetable Store	Restaurant	BBQ Joint	Yoga Studio	Concert Hall	Dessert Shop	Department Store	Cuban Restaurant	Convenience Store

In the second cluster we find small neighborhoods with parks, convenience stores and churches.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
18	Nagarote	Park	Restaurant	Yoga Studio	Caribbean Restaurant	Church	City	Clothing Store	Coffee Shop	Comfort Food Restaurant	Concert Hall
24	Quezalguaque	Park	Yoga Studio	Donut Shop	Church	City	Clothing Store	Coffee Shop	Comfort Food Restaurant	Concert Hall	Convenience Store
28	Chichigalpa	Grocery Store	Park	Donut Shop	Church	City	Clothing Store	Coffee Shop	Comfort Food Restaurant	Concert Hall	Convenience Store
30	El Realejo	Park	Yoga Studio	Donut Shop	Church	City	Clothing Store	Coffee Shop	Comfort Food Restaurant	Concert Hall	Convenience Store
87	Santa Lucia	Park	Yoga Studio	Donut Shop	Church	City	Clothing Store	Coffee Shop	Comfort Food Restaurant	Concert Hall	Convenience Store
111	San Dionisio	Park	Yoga Studio	Donut Shop	Church	City	Clothing Store	Coffee Shop	Comfort Food Restaurant	Concert Hall	Convenience Store

Finaly, the third cluster has neighborhoods from the north of the country. The neighborhoods has mountains, churches, cafes and yoga studies.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
69	Telpaneca	Mountain	Yoga Studio	Donut Shop	Church	City	Clothing Store	Coffee Shop	Comfort Food Restaurant	Concert Hall	Convenience Store
79	Murra	Mountain	Yoga Studio	Donut Shop	Church	City	Clothing Store	Coffee Shop	Comfort Food Restaurant	Concert Hall	Convenience Store
80	Quilalí	Hotel	Mountain	Yoga Studio	Drugstore	City	Clothing Store	Coffee Shop	Comfort Food Restaurant	Concert Hall	Convenience Store
103	Santa María de Pantasma	Pet Store	Mountain	Yoga Studio	Donut Shop	City	Clothing Store	Coffee Shop	Comfort Food Restaurant	Concert Hall	Convenience Store
113	La Dalia	Mountain	Yoga Studio	Donut Shop	Church	City	Clothing Store	Coffee Shop	Comfort Food Restaurant	Concert Hall	Convenience Store
122	Bonanza	Mountain	Yoga Studio	Donut Shop	Church	City	Clothing Store	Coffee Shop	Comfort Food Restaurant	Concert Hall	Convenience Store

6. Conclusion

In this work we carry out a cluster analysis on the different neighborhoods of Nicaragua based on the different businesses and leisure places present in each neighborhood. The data of the different businesses in each neighborhood were extracted from the Foursquare database.

This analysis was performed in 139 neighborhoods in Nicaragua, and clustered with the K-means algorithm. We performed the Elbow method and determined that the optimal number was 3 clusters.

In the first cluster we find 89 neighborhoods where we find mostly hotels and restaurants. In the second cluster we have 6 neighborhoods in which we find small towns with churches and Parks. Finally, in the last cluster we find 6 neighborhoods wuth mountainous areas, churches and yoga studios.