Project: The Battle of Neighborhoods

Introduction

Salamanca is a Spanish university city. Due to the large number of students, most of the services available in the city are oriented to meet the needs of university students.

Above all, there are many food establishments.

However, food from establishments close to university areas does not offer good quality food and managers are not aware of creating businesses that respect sustainability and provide ecological products.

That's why Pepe, a citizen of Salamanca wants to embark on a sustainable business that sells organic products and food made from them.

Objetive

The aim of this work is to provide Pepe with a tool that, using machine learning techniques, provides Pepe with information about the services available in the city so that he can choose the area that best suits him to start his business.

Data

In order to solve the problem, we use the data provided by the web foursquare for the coordinates of the city of Salamanca.

In the first place the coordinates of the city center will be obtained.

Then, a query in the Foursquare website is done to obtain the top sites in Salamanca with their geographical coordinates.

Methodology

From the coordinates of the city of Salamanca, data was collected from the top Sites from FourSquare Website.

The variables we will use will be:

- Postal code
- Name of the venue
- Category
- Latitude

- Longitude

Based on this information, frequency distribution analyses were carried out to see what type of venues are most present in each area of the city.

Based on these frequencies, a cluster k-means analysis was performed in order to see which areas of the city are more similar in terms of the number of venues of each type that exist.

Finally, this information is represented in a map.

Results

The analysis of frequency distribution shows that the type of venue in each area of the city of Salamanca is: 37001 and 37002 the most frequent kind of venue is tapas restaurant, in the 37003 postal code, the 50% of the venues are parks and the 50% are museums, in the 37004 postal code, there are plazas and burguers, in the 37005 there are bars and restaurants, however, in the 37007 and 37008 postal code there are mainly hotels.

This result can be seen in the following tables:

```
----37001----
            venue  freq
0  Tapas Restaurant   0.3
1          Wine Bar   0.1
2        Restaurant   0.1
3         Gastropub   0.1
4              Park   0.1



----37002----
                    venue  freq
0         Tapas Restaurant  0.09
1                     Café  0.09
2  Mediterranean Restaurant  0.09
3                    Hotel  0.05
4             Gourmet Shop  0.05
```

```
----37003----
                venue  freq
0               Park    0.5
1             Museum    0.5
2          Hotel Bar    0.0
3           Irish Pub   0.0
4  Italian Restaurant   0.0


----37004----
                venue  freq
0              Plaza    0.5
1       Burger Joint   0.5
2               Park    0.0
3          Irish Pub   0.0
4  Italian Restaurant   0.0


----37005----
                venue  freq
0                 Bar  0.25
1     Tapas Restaurant  0.25
2  Japanese Restaurant  0.25
3               Diner  0.25
4         Art Gallery  0.00


----37007----
                venue  freq
0              Hotel  0.33
1        Coffee Shop  0.33
2              Plaza  0.33
3          Irish Pub  0.00
4  Italian Restaurant  0.00


----37008----
                venue  freq
0              Hotel  0.29
1         Art Museum  0.14
2  Spanish Restaurant  0.14
3               Café  0.07
4  Monument / Landmark  0.07
```

The cluster analysis showed that there are two clusters, the first is formed by the 37004 and 37007 postal codes. The second group is formed by the rest of postal codes.

In the following map the top venues are shown.

Discussion/Conclusion

From the information obtained through the analysis we have seen that in the areas 37001 and 37002 there are many venues type bar and restaurant.

The best areas to open an organic food business is 37003 and 37004 as there are park, museums and squares and are the most central areas.