The Battle of Neighborhoods – Report

By Daniel Chamorro

Introduction

The "ÉXITO" group is a Colombian multinational which has a huge network of supermarkets in South America. In this case, the company's management has opened a call to data scientists interested in solving the following case study:

The management of "ÉXITO" group in the city of Bogotá, capital of Colombia, has detected that it is losing market share due to the appearance and positioning of new competitors. The supermarket networks "D1", "ARA" and "OXXO" have been identified as the strongest competitors at the moment. The management of the company to face this situation, has made the decision to expand its supermarket network in the city of Bogotá. The invitation to all interested parties is to determine in which localities in the city of Bogotá the supermarket network should be strengthened, considering the number of existing supermarkets of the detected competitors and those of the company "ÉXITO".

Data

Because the city of Bogotá is divided into 20 localities, the first set of data relates the number, name, zip code, area, population and population density of each of the localities. This information will be extracted from Wikipedia, which can be seen in the following link: https://es.wikipedia.org/wiki/Anexo:Localidades_de_Bogot%C3%A1.

Now to have information on each of the supermarkets, both their own and those of the competitors, it will be necessary to use the Foursquare API. The most relevant information to extract for this problem is the number of supermarkets in each of the networks, which will be obtained through calls to the API with their respective parameters.

When all the information is extracted, it will be organized in DataFrames, with which a Clustering model will be proposed. The objective of the model will be to cluster the localities of the city of Bogotá in clusters depending on the number of supermarkets identified by each of the networks ("D1", "ARA", "OXXO" and "ÉXITO"). Subsequently, an analysis of each of the clusters obtained will be carried out in order to suggest to the company's management "ÉXITO" in which localities the new supermarkets should be placed.

Note: in this project the name of real companies is taken, information about the location of their supermarkets, simply for educational purposes. Similarly, the situation described is not real and only seeks to represent a case study in the field of data science.

Methodology

The methodology used in this project can be divided into three steps:

1. Extract and transform the data

To begin, information related to the localities that make up the city of Bogotá is extracted. As this information is extracted from a web page (Wikipedia) it is necessary to convert it into a DataFrame to be able to manipulate it in a better way. Subsequently, the geographical coordinates of each locality are extracted

through the 'geopy' library. Now that you have the geographical coordinates of each locality, the existing supermarkets are searched for each of the networks. This information is also deposited in a DataFrame to facilitate its manipulation.

2. Clustering model

Knowing how many supermarkets there are in each locality for each of the networks, a clustering model is proposed to group the localities into clusters of interest. After running the model, the 20 locations have been grouped into 7 clusters. The results can be visualized through a map.

3. Cluster analysis

Now that the clusters obtained are known, the total market share of each of the networks is calculated. Similarly, the market share of each of the networks in each cluster is calculated, this in order to identify the most interesting cluster.

Results

After running the clustering model, the 20 locations in the city of Bogotá have been classified into 7 clusters. In the following map we can visualize how the grouping has been:

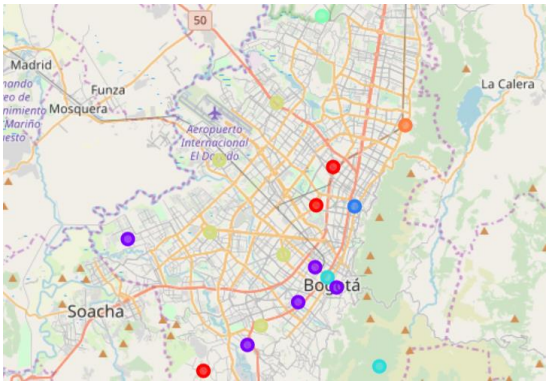

Figure 1

Additionaly, in the following table we can see how many supermarkets each of the networks has in each of the clusters:

Cluster	D1	Ara	Oxxo	Exito
0	25	8	12	37
1	11	3	12	11
2	20	10	20	15
3	2	14	22	46
4	24	5	6	9
5	24	16	8	38
6	29	12	16	22

Table 1

With this information, you can calculate the total market share percentage for each of the networks, obtaining the following results:

D1	Ara	Oxxo	Exito
28,3%	14,3%	20,1%	37,3%

Table 2

Similarly, the market share of each of the networks in each cluster was calculated. The results can be displayed in the following bar graphs:

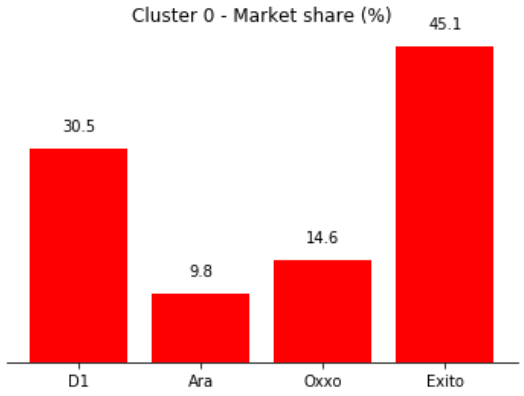

Figure 2

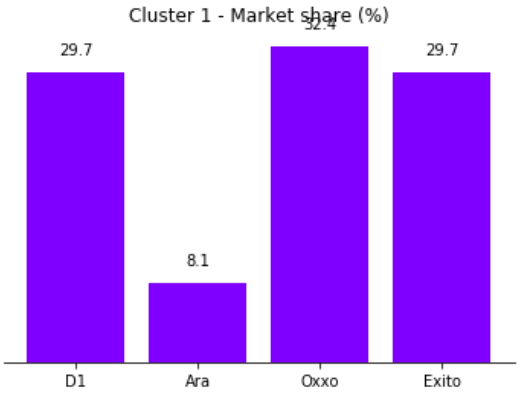

Figure 3

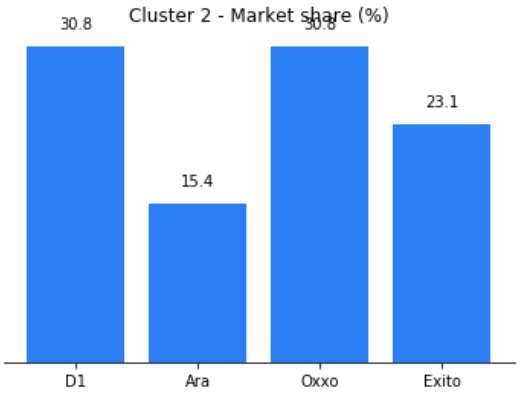

Figure 4

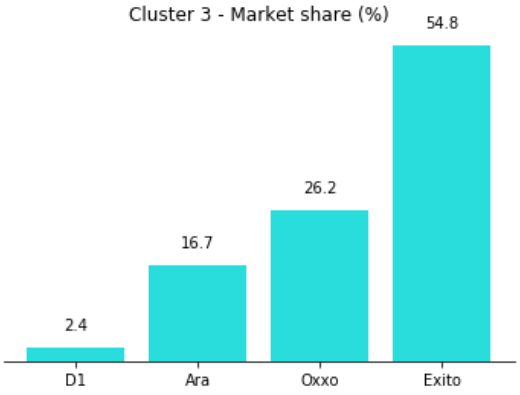

Figure 5

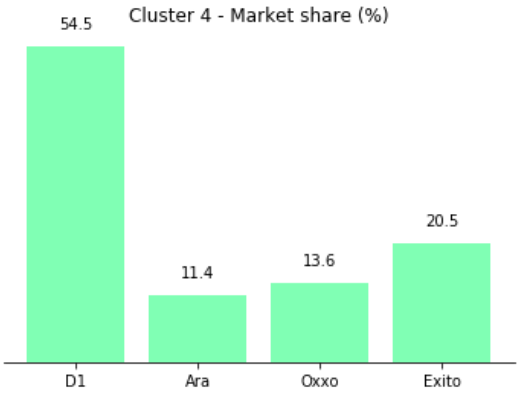

Figure 6

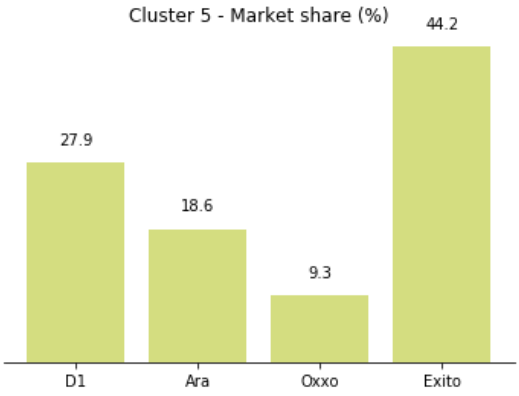

Figure 7

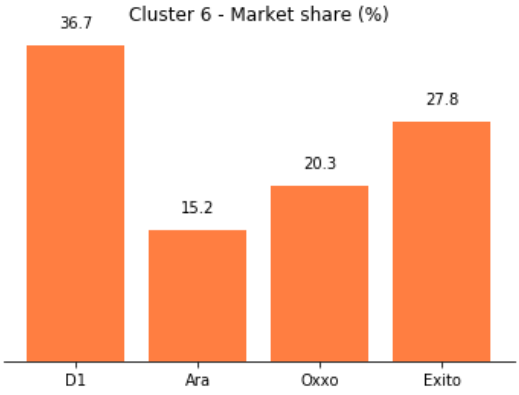

Figure 8

Discussion

As can be seen in Table 2, we can verify that the "EXITO" group is the supermarket network with the largest market share with 37.3%. However, the "D1" network is becoming its most important competitor

since it occupies 28.3% of the market. Additionally, between the "ARA" and "OXXO" networks, they account for 34.4% of the market share, slightly more than a third of the entire market.

Now, if we analyze the market share of each network in each of the clusters obtained, we can see that the group "EXITO" dominates the market in clusters 0, 3, 5, covering 11 localities in the city of Bogotá. However, we can also see that the domain of the remaining clusters is shared between the "D1" and "OXXO" networks, with the "ARA" network being the one with the lowest presence.

Since the management of the "EXITO" group wishes to know in which localities it should concentrate its efforts to recover the market share due to the arrival of the new networks, I would recommend that they should locate new supermarkets in the localities of: "Tunjuelito", "Bosa" and "Antonio Nariño" where they only have one supermarket. Similarly, I can suggest that they should locate new supermarkets in the localities of "Chapinero" and "Usaquén" in order to regain control of the market share. Finally, I consider that where new supermarkets must be located urgently is in the locality of "Suba", given that the "D1" network has positioned itself with 54% of the market in this locality.

Conclusion

As we have seen throughout this report, the data extracted from Wikipedia and the Foursquare API, about the localities that make up the city of Bogotá in Colombia and the supermarket networks "D1", "ARA", "OXXO" and "EXITO" have been used to propose a clustering model. The proposed model has served to make some suggestions of which are the localities that need more attention from the management of the "EXITO" group, in order to locate more supermarkets that allow them to recover their market share. It is expected that the recommendations made in the "Discussion" section will be sufficient to solve the identified business problem.