THE BATTLE OF NEIGHBORHOODS

This report contains the study performed for the Capstone Project Week 4 and is structured as follows: Section 1 presents the business problem. Finally, Section 2 describes the data that is going to be analyzed.

1. INTRODUCTION / BUSINESS PROBLEM

An important gym group is willing to open a new gym in Madrid and they want to find the best place to open the club. Currently more and more people finds specialized places to practice sports and some neighborhoods can be crowded, so opening a new gym without analyzing the current venues in the city and the potential customers can be risky.

Additionally, our customer has transmitted to us that they're age target group is people aged between 20 and 39 years. Moreover, we will rely on income data per district to refine the final decision.

Consequently, the company came to us to provide them with useful information about Madrid districts and neighborhoods so they can base the decision of the new gym location in reliable data.

2. DATA

The data used to solve this problem came from four sources:

1. **District and neighborhood names from Madrid**

   Source: https://www.madrid.es/

   The columns extracted for this study are :

   - Hood number : Unique number of the hood
   - Hood: Name of the hood
   - Neighborhood: Name of the neighborhood

Here is an extract of the first rows of this dataset:

Codigo de ba	Codigo de di	Nombre de b	Nombre acer	Superficie (n	Perimetro (n
1	1	PALACIO	PALACIO	1471085	5754
1	2	IMPERIAL	IMPERIAL	967500	4557
1	3	PACIFICO	PACÍFICO	750065	4005
1	4	RECOLETOS	RECOLETOS	870857	3927
1	5	EL VISO	EL VISO	1708046	5269
1	6	BELLAS VISTA	BELLAS VISTA	716261	3443

2. **Geographical data**

   Source: https://www.123coordenadas.com

This web was used to calculate coordinates for each neighborhood and obtain a CSV file with the following columns:

- Longitude : Geographical longitude
- Latitude : Geographical latitude

Here is an extract of the first rows of this dataset:

Nombre	Latitud	Longitud
ZOFIO ,MADRID	403.798.077	-371.521.158.874.426
VISTA ALEGRE ,MADRID	403.887.883	-37.400.441
VINATEROS ,MADRID	404.051.965	-36.415.467
VENTAS ,MADRID	40.430.831	-36.632.802
VALVERDE ,MADRID	405.011.401	-367.859.163.522.496
VALLEHERMOSO ,MADRID	404.430.572	-371.168.099.208.445
VALDEZARZA ,MADRID	4.046.530.915	-371.695.805.610.135

3. **Population data**

Source : https://www.madrid.es/

An Excel file with the population of Madrid classified in districts and age ranges was obtained and used to determine the number of persons with ages between 20 and 39 years per district. This data is stored in a column called Population.

Here is an extract of the first rows of this dataset:

Edad	2020
TOTAL	**141.527**
De 0 a 4 años	6.308
De 5 a 9 años	6.256
De 10 a 14 años	6.030
De 15 a 19 años	5.819
De 20 a 24 años	6.076
De 25 a 29 años	7.805
De 30 a 34 años	9.470

4. **Income data**

Source : https://www.madrid.es/

An Excel file with the income per person of Madrid classified in districts was obtained and used to determine the average income for each district. This data is stored in a column called Income.

Here is an extract of the first rows of this dataset:

Distrito / Barrio	Renta media por persona	Renta media por hogar
Ciudad de Madrid	**15.930**	**40.195**
01. Centro	**16.711**	**33.473**
011. Palacio	18.254	36.357
012. Embajadores	13.454	27.655
013. Cortes	19.431	37.725
014. Justicia	21.570	43.045
015. Universidad	16.869	33.209

Using these four sources a dataframe is created. The first 5 rows of this dataframe is depicted in Figure 1.

	Hood Number	Hood	Neighborhood	Latitude	Longitude	Population	Income
0	2	ARGANZUELA	PALOS DE MOGUER	40.403638	-3.695289	40000.769957	17738
1	2	ARGANZUELA	DELICIAS	40.397292	-3.689495	40000.769957	17738
2	2	ARGANZUELA	CHOPERA	40.394893	-3.699705	40000.769957	17738
3	2	ARGANZUELA	IMPERIAL	40.406929	-3.717321	40000.769957	17738
4	2	ARGANZUELA	ATOCHA	40.405204	-3.687930	40000.769957	17738

Figure 1. Dataframe head