

Capstone Project – The Battle of Neighborhoods (Week 1)

Description of Data

The data used to carry out the analysis has been obtained through different techniques:

- Reading of datasets stored in the form of csv files.
- Information extraction via Web Scrapping.
- Obtaining geospatial information using the Foursquare API.

Using this techniques, we have obtained this properties:

1. **Latitude and longitude for each town¹:** to locate each town in the Balearic Islands map.
2. **Population for each town²:** we use this metric, combined with others (occupation estimation, tourist relevant towns, etc), to get the volume of potential customers for each town.
3. **Lenght for each town (km²)³:** to analyze different aspects. For example, to analyze the relation between the number of fast food restaurants for each km² in every town.
4. **Occupation estimation for each town (for each quartile of the year):⁴** consists of the estimation of the number of tourist customers that will go to every town in the dataset. It shows how many tourist customers are expected to be in each quarter of the year. Even it is an estimation, is good metric because, since 2014, almost every estimation produced was similar to the real facts of even where wost (so the occupation was bigger). Using this information and the population size for each town, we can have an idea of the volume of potential customers we could have in every town in the dataset.
5. **If a town is considered tourist relevant:** if a town is not considered tourist relevant, makes no sense to think that we can set up a restaurant expecting to have tourist customers. Our only customers will be the population from the town (or from other towns).
6. **Number of fast food restaurants foe each town:** obtained using the Foursquare API. Using this information we can analyze the competition in every town in the Balearic Islands, etc.

1 Source: <https://www.businessintelligence.info/resources/assets/listado-longitud-latitud-municipios-espana.html>

2 Source: https://es.wikipedia.org/wiki/Anexo:Municipios_y_comarcas_de_las_Islas_Baleares

3 Source: <https://www.infolaso.com/geografia/114-geografia-de-espana/1162-extension-geografica-de-los-municipios-de-baleares-balears.html>

4 Source: https://ibestat.caib.es/ibestat/estadistiques/614884d6-737a-401d-a8c3-a35519b8fec9/6b9fd2d8-30ca-4373-b13f-3818b206fc51/es/I208013_m020.px

We also use a gejson file of the Balearic Islands separating between the municipalities it has.

Usage of Data

All this information will be used to:

- Process and generate new data.
- Apply different clustering algorithms:
 - K-Means
 - DBSCAN
- Understand the context visualizing data using tools like Folium and choropleth maps.

Using all this techniques, we expect we answer all the questions we have written in the previous section.