**Capstone Project**

**THE BATTLE OF NEIGHBORHOODS**

1. **DESCRIPTION OF THE PROBLEM AND BACKGROUND**

This report aims to analyze the city of Madrid and establish, based on different criteria, the best place to install a dental clinic.

Madrid is the capital of Spain, so it is an pretty big city and has a very abundant population, being one of the most important capitals of Europe.

It is also a city that has a great tourist attraction, so many areas of the capital are occupied by businesses dedicated to attract tourists.

In this sense, it should be noted that, when looking for a suitable location for a dental clinic, it is necessary to look for an area with a high number of fixed inhabitants, that do not vary over time and that can be established as fixed business customers, so it is necessary to avoid the most tourist areas and look for residential areas but with a high economic activity.

Therefore, the aspects to study when evaluating an adequate location for a dental clinic would be:

- Number of inhabitants

- Per capita income of the inhabitants

- Existing businesses in the surroundings

- Adequate communication with the rest of the city

-Touristic attractions nearby

1. **DESCRIPTION OF THE DATA**

To solve the problem, the following data is needed:

- Districts Madrid data(code, name, coordinates, area, habitants...): This information has been obtained from different websites, and has been stored as shp and csv files.

- Diferent bussines in the neighbourhood: thanks to Foursquare API, we can obtain information about the venues in each neighborhood

1. **APPROACH**

- Collect the districts of Madrid info

- Use FoursquareApi to find all venues in the different districts

- Finally, considering the venues, the communications and the number of inhabitants choose the best localization for the dental clinic.

1. **LIBRARIES REQUIRED**

- numpy: library to handle data in a vectorized manner

- pandas: library for data analsysis

- json: library to handle JSON files

- json\_normalize module to tranform JSON file into a pandas dataframe

- geopy: library to locate the coordinates of addresses, cities ...

- requests: library to handle requests

- matplotlib: library to plot

- sklearn: a machine learning library

- folium: a map rendering library

- geopandas: library to work with geospatial data

1. **DATA COLLECTION**

- Distritos.csv: File with the districts information

- Distritos.geojson: File with the districts geospatial data

- ParadasMetro.geojson: File with the underground stations geospatial data

- https://api.foursquare.com: API to get de venues in the area

1. **ANALYSIS**

The process to do the analysis has been as follows:

First, the map with the districts of Madrid has been downloaded in shp format .Geojson has been transformed and a map of the area has been generated with folium.

Secondly, the population, area and population density data of each district in .csv format have been obtained. This information has been combined with the district layer and a chloropleth type map with folium has been generated.   
 Subway stations have been incorporated into this map, which had previously been downloaded in shp format and transformed to geojson.

Subsequently, with the foursquare API, the existing businesses in each area have been obtained and the 10 most common of each of them have been filtered

Finally, with sklearn, a clustering of the analyzed districts has been done, with the commented data 5 groups of districts have been obtained. Among them, the most appropriate group has been chosen according to the conditions initially mentioned. Within that group the district that has been considered the best has been chosen, although all those included in the selected cluster would have been valid.

1. **CONCLUSION**

To choose the district in which to locate the dental clinic, we look for a residential area, with a high number of residents and with a high commercial activity. Likewise, it must be a well-connected district, for which the metro stops of the city of Madrid have been analyzed.

As indicated, the appropriate areas would be those included in cluster 3, since they are residential areas and not heavily influenced by tourism.

Within cluster 3, districts with a greater number of inhabitants, and therefore more apt to install the new business would be Carabanchel or Ciudad lineal.

Since Carabanchel is better communicated, since it has a greater number of subway stops, in addition to a commercial activity more compatible with the new business, it is considered the most appropriate district to install a new dental clinic.