Capstone Project – The Battle Of Neighborhoods

Data acquisition and cleaning

Data sources

To carry out this project, the following information is necessary:

- Spatial information about neighborhoods. These data were obtained through the portal of the Valencia City Council. Files that described the administrative distribution of neighborhoods were searched. link

- Information on the location and price of rental apartments. Data from an already preprocessed collection of Airbnb were used. These data continue the price, the exact location and the neighborhood to which each one belongs, among other data. link

- Information about the places of interest in each neighborhood. Searches were made through the Foursquare API.

As an example we obtain the following tables, in the next point the procedure will be detailed.

	coddistbar	Neighborhood	Latitude	Longitude
0	171	BENIFARAIG	39.52564	-0.38462
1	161	BENICALAP	39.49301	-0.3910

Table 1. Fragment of data on the position of neighborhoods.

name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price
VALENCIA HISTORIC HOUSE 50M BEACH	7093832	Francisca	POBLATS MARITIMS	LA MALVA-ROSA	39.47553	-0.32461	Entire home/apt	150

Table 2. Excerpt from neighborhood rental data.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
AIORA	39.46573	-0.34336	Jardines de Ayora	39.46814	-0.34340	Playground
AIORA	39.46573	-0.34336	Bar Mochuelos	39.46614	-0.34794	Mediterranean Restaurant

Table 3. Fragment of the venues of the neighborhoods.

Data cleaning

First of all, the files that contain the position of the neighborhoods are in geojson format, this specific file uses the UTM 30 coordinate format for spatial location. These were projected into the WGS84 format because Foursquare and the "folium" map display library work with that coordinate format. This was accomplished through two libraries:

- pyproj, used to project individual coordinates.
- geopandas, used to create a new geojson file with the appropriate format.

Next, reference coordinates for each neighborhood were obtained from the central position obtained through the "shapely" library. These data will be used later to obtain the places of interest.

Regarding rental housing, it was grouped by neighborhoods and both the average rental price and the number of apartments per neighborhood were obtained.

At this point, both dataframes were joined, obtaining in the same table the location data of the neighborhood and the average value of the rental price and the number of homes per neighborhood.

It was discovered that the Benifaraig neighborhood lacked any rental housing, because the table with the grouped description of rental housing had one row less than the table with information on neighborhoods. It was decided to dispense with that neighborhood for the rest of the project.

The following figures represent the price per neighborhood, and the number of houses per neighborhood respectively. The darker the color, the greater the value.

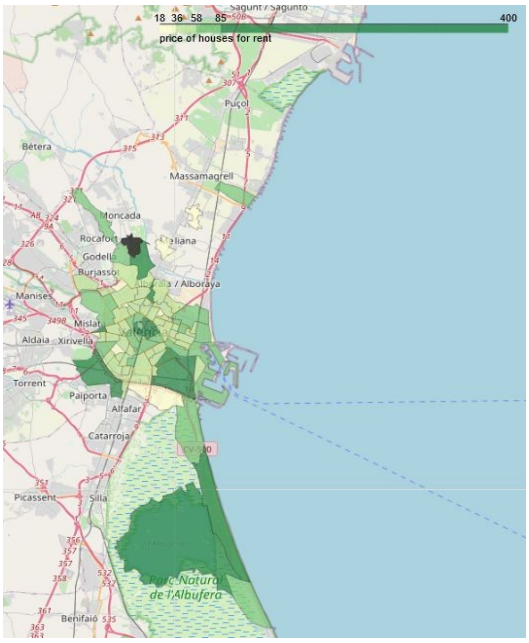

Figure 1. Map of Valencia with the price colored by neighborhoods.

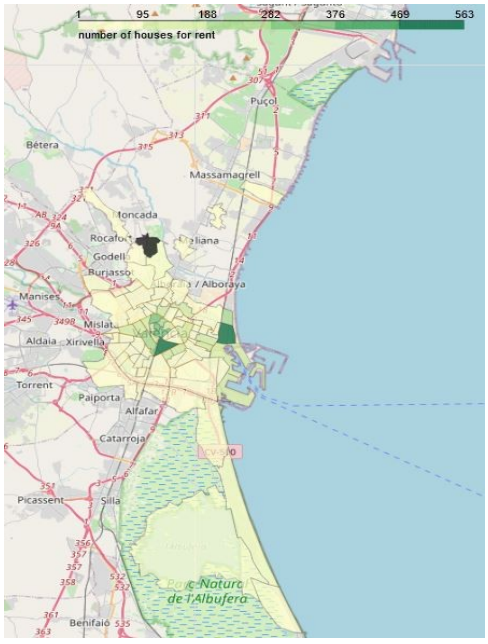

Figure 2. Map of Valencia with the number of tourist rental houses colored by neighborhoods.

The previous choropleth maps were obtained through the "folium" library.

To finish this section, the data from Foursquare was obtained by launching an exploration request in the reference coordinate of each neighborhood with a radius of 500 meters.