Establishing a kennel in Madrid

1. Introduction

Establish a dog kindergarten service in the city of Madrid

It is planned to examine the city of Madrid, in order to establish a dog kennel; So, you need to see in which parts of the city are places with greater and lesser concentration related to pets, whether they are nurseries, vets, veterinary emergencies, parks.

So we are going to cluster the places in the city and see where we can be located with the greatest probability of success.

Madrid was chosen for being the city with the most business possibilities, in terms of the size of the market in Spain, as well as being the city of Spain with the most canine pets, compared to other cities, both percentage-wise at the level of the general population. Madrid, as in number of canine pets.

2. Data acquisition and cleaning

2.1 Data source

The general idea is to bring the relevant data to examine on the map of Madrid, which areas have a better environment to care for dogs. Basically we will see that areas or neighborhoods of the city have more parks and services for pets.

The data of the places analyzed was taken from the foursquare.com API.

2.2 Data cleaning

Two calls were made to the foursquare API, to find where the city parks were located and where the pet-related businesses were located. They then became a single dataframe.

The fields where there was no information were identified, and fortunately all were located in columns that contained irrelevant information, so the complete columns were eliminated from the dataframe.

Then we additionally remove the unnecessary columns for our investigation.

3. Methodology

2.1 Exploring the data

First, additional data on the neighborhoods of Madrid were sought on official pages of the Spanish government, but unfortunately they do not have enough data, such as the location of each neighborhood with coordinates or at least postal codes, they are incomplete.

So it was decided to work only with the data available in Foursquare, and then to achieve differences in the areas through some clustering method.

After downloading the necessary data, and reading it using json reading methods, and cleaning the data using various techniques in python; Through the visualization with Folium methods, it was possible to observe the points of interest on the map of the city of Madrid.

2.2 Clustering

After trying several methods of machine learning, clustering, it was found that the most appropriate was K-means, taking into account that the resulting data frame was quite small. However, what we needed to find out was more than enough data.

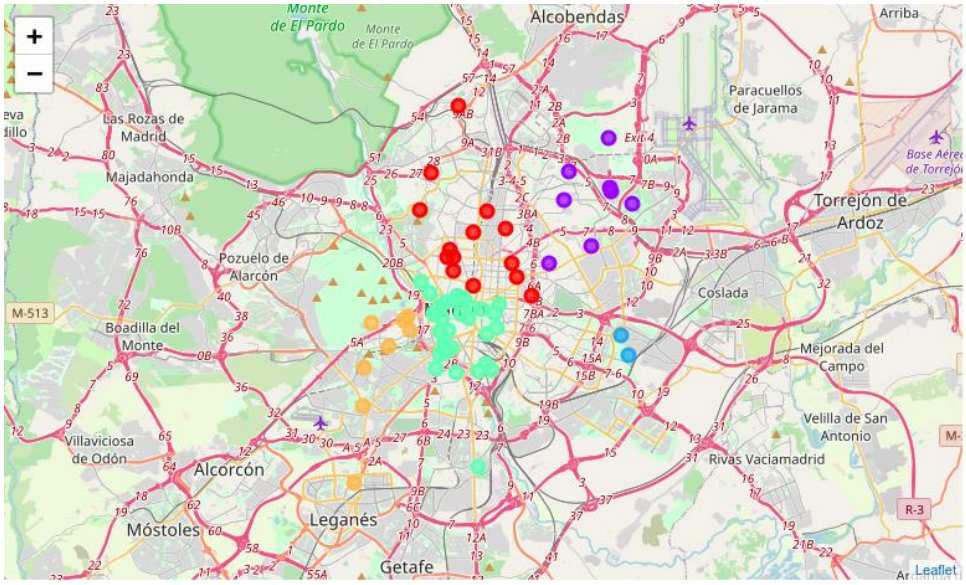
With this method, it was successfully divided into fairly clear areas and one of those areas, one of the clusters, differed more than twice from the others, leaving without a doubt that it was the appropriate one.

4. Results

After cleaning the data and merging them into a single dataframe, performing the entire clustering process, our final dataframe looked like this, well the first five rows of it:

	Km Clusters	location.lat	location.lng	name
0	3	40.415179	-3.681471	Parque del Retiro
1	3	40.420484	-3.704736	Parque Madrid Río (Bosque de Trepa)
2	3	40.417535	-3.712035	Parque Infantil Plaza De Oriente
3	3	40.417478	-3.712095	Parque infantil
4	3	40.414676	-3.714790	Parque del Emir Mohamed I

Our clustered data visualization map looked like this:



And our resulting dataframe to verify how many sites of interest there were in each cluster, in our case the city area, clearly cluster 3 stands out more than twice, in concentration of related places, above the others:

Km Clusters			
0	14	14	14
1	8	8	8
2	2	2	2
3	35	35	35
4	8	8	8

5. Discussion

After seeing this dataframe, we have no doubt that the resulting cluster number 3 contained more than twice as many places related to the business to be established; keeping in mind that this was the answer to finding our initial goal. It can also be seen in a more graphic way on our map.

6. Conclussion

We then found cluster 3 as the appropriate area to establish our kennel business, as it is surrounded by places related to the business, which enhances the chances of success.

This more specifically speaking refers to the neighborhoods within the area of "Centro de Madrid", "Arganzuela" and "Retiro".