**A. Introduction**

**A.1. Description & Disscusion of the Background**

Madrid is the capital of bars, it is one of the hospitality epicenters of the world. Experts attribute this tendency to the personality of the average citizen of Madrid who is extroverted, cosmopolitan and given to going out of the house for his leisure and recreation. It is one of the attractions that tourists come looking for when they visit this city. For this reason we are studying the possibility of opening a bar specially oriented to international tourism, so we want to identify in which areas the most important and most visited places in Madrid are located, since this will depend on whether we have more or fewer clients and, consequently, more or less profits. Anyone who would like to invest in a Bar in Madrid could be interested in this project.

### A.2. Data Description

Based on definition of our problem, factors that will influence our decision are:

* Number of most popular places in the neighborhoods.

We´ll obtein the different data from the next origins:

* Capture the Madrid´s neighborhood: <https://www.geonames.org/postalcode-search.html?q=madrid&country=ES>
* Location of neighborhood will be determined by coordinates added using geocoder
* Number of 100 venues in Madrid, its localitation and its type using Foursquare API.

**B. Methodology**

We will develop next phases for assure the success of the project:

### B.1 Business understanding (Objectives and requirements from a non-technical perspective)

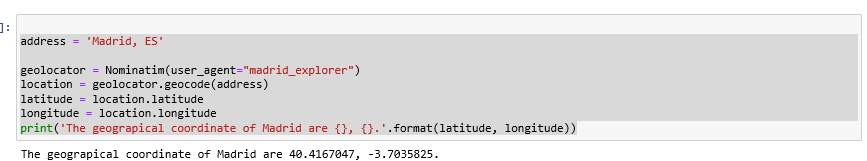
Our business objective is to open a bar in Madrid oriented to tourists and that is successful.

### B.2 Understanding the data (Familiarize yourself with the data with your business goals in mind)

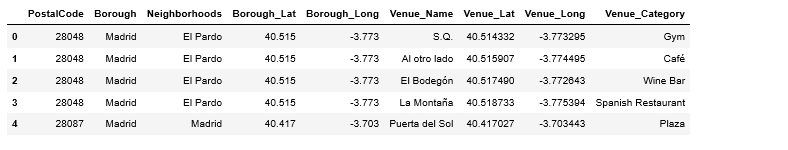
* Initial data collection: We have collected the data from the sources:
  + <https://www.geonames.org/postalcode-search.html?q=madrid&country=ES>



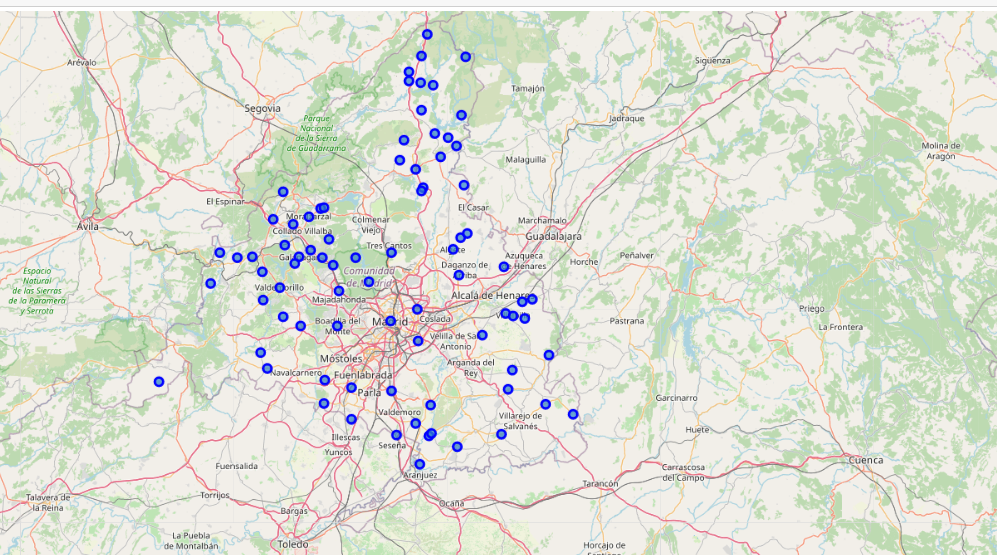
* + Geocoder: In oder to obtain longidudes and latitudes and be able to draw them into a map using Folium library



* + Fousquare API: In oder to obtain the top 100 venues in Madrid for each neigborhood toward their latitudes and longidues:



* Data description.
* Data exploration including its visualization on a map through Folium:



* Data quality verification: Checking the quantity of missing values

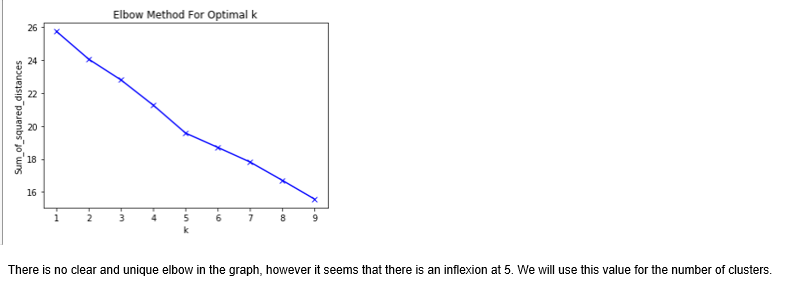
### B.3 Preparing the data (Get the mineable view or dataset)

* Selection of data
* Data cleaning: Handling duplicates and missing valuesConstrucción de datos
* Data construction
* Data integration
* Data formatting

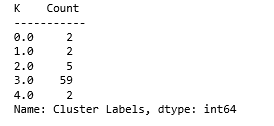
### B.4 Modeling (Apply data mining techniques to datasets)

* Selection of the modeling technique: Cluster through the k-means algorithm
* Evaluation design
* Model construction

The selected modeling technique is an unsupervised k-means algorithm. Firstly, the elbow technique has been applied to determine the optimal number of clusters.



Although the graph is not entirely explanatory, we have selected 5 clusters, distributed most of observations in one of them.

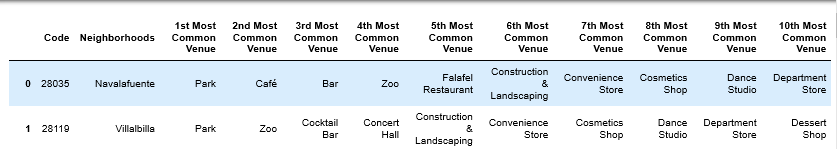


**C. Analyzing Clusters:**

We try to analyze the different cluster in order the venues they contain:

### C.1 Cluster 0

This cluster would be referred to an industrial zone



### C.2 Cluster 1

This cluster would be referred to a turism zone.

### 

### C.3 Cluster 2

This cluster would be referred to a residencial and turism zone.

### 

### C.4 Cluster 3

This cluster would be referred to a residencial zone.

### 

### C.5 Cluster 4

This cluster would be referred to a residencial and industrial zone.

### 

**D. Recomendations:**

Taking into account our final objective, we believe that the cluster where our bar should be integrated would be at number == 1, since it is the one that best represents tourism within the city, determined above all by Hotels, Tapas Restaurants, Concert Hall and Shops. For that, our recommendation if you are planed opening a bar in Madrid focused in international tourism, is that you open it inside de cluser 1.

**E. Conclusions:**

Opening a tourism-oriented bar in the capital of bar tourism is always good business. But, for this business to be as successful as possible, the following factors should be taken into account:

* Choose the best neighborhood for it. Based on our model, we recommend you choose the cluster neighborhoods == 1
* Take into account the costs of premises in said neighborhood, since tourist areas usually have a fairly high rental Price.