Opening a new office building for an engineering company in Madrid, Spain.

IBM Applied Data Science Capstone

Raúl Cabanas Contreras

June 2020

Business Problem

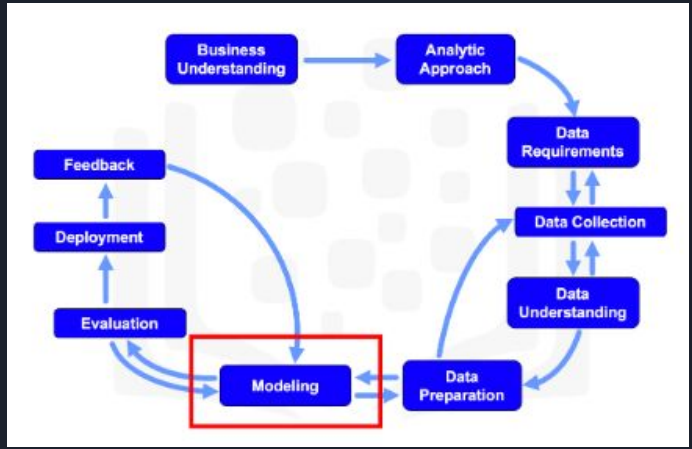
- An engineering company, in Spain, that currently have one office building in Madrid, and because the business increase, they have to establish a new location to place more employees.
- The current location is in the district of Charmartin. Employees surveys about the current location are very good.
- Objective: determinate which other districts in Madrid are equivalent to Charmartin district in order to maintain the employees satisfaction.

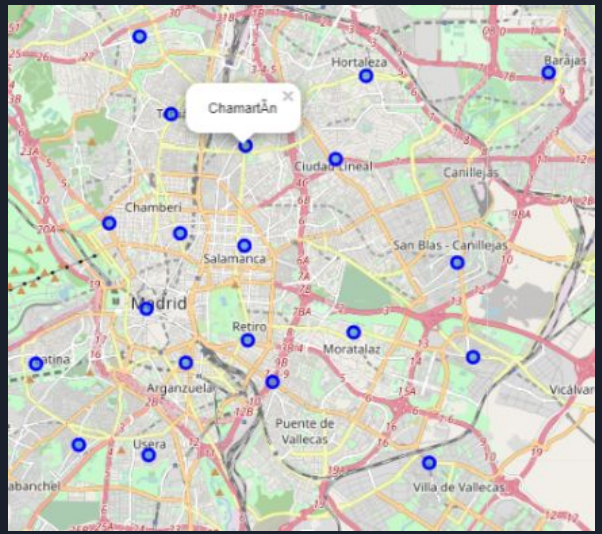
Data

- We will use the following data:
  - Districts information of Madrid: The source found is the government official data web page in Spain: https://datos.gob.es/en
  - Venue data from Foursquare API.
  - Geocoder package for latitude and longitude of Madrid.

Methodology

- Load the districts dataframe and data wrangling
- Locate Madrid districts in a map
- Get the information of Madrid position with Geocoder
- Use Foursquare API to get venue data
- Data wrangling to get a dataframe with the most important venues per district and its population density and area.
- Perform clustering of the data by using K-Means clustering.
- Visualize de cluster using Folium
- Analyze the results

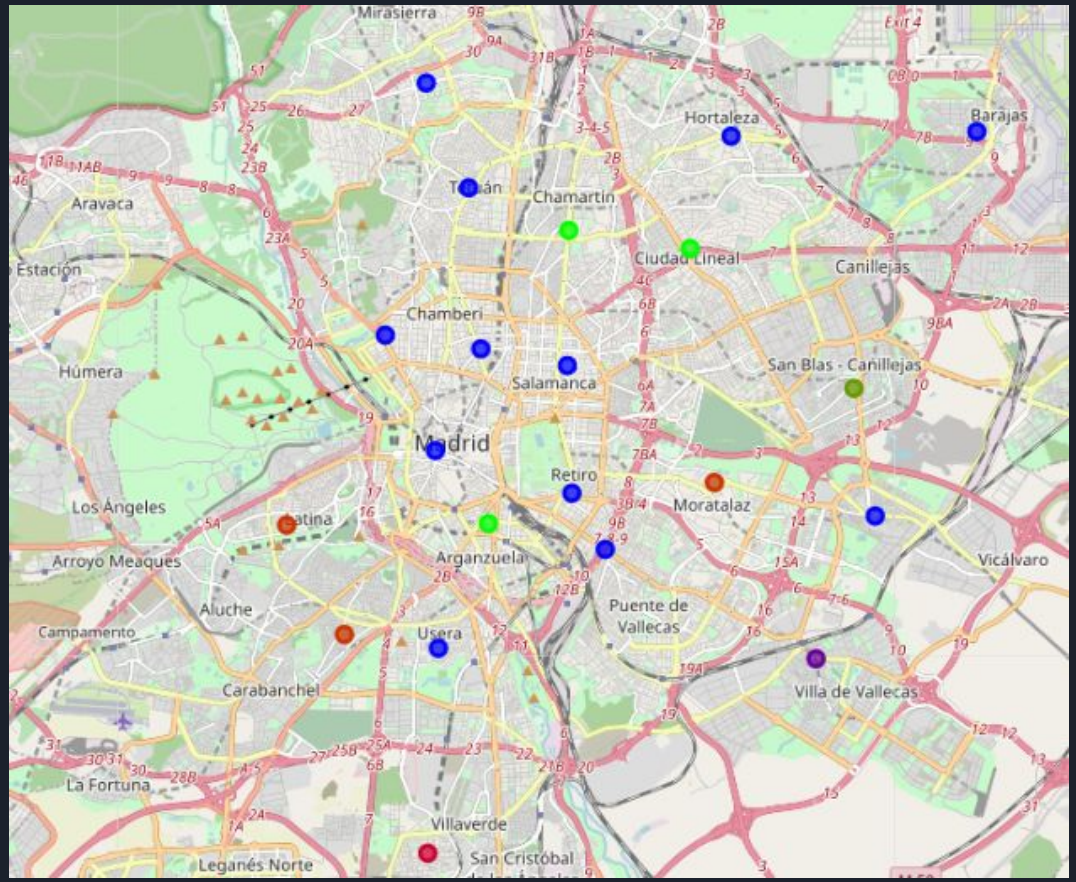




Results

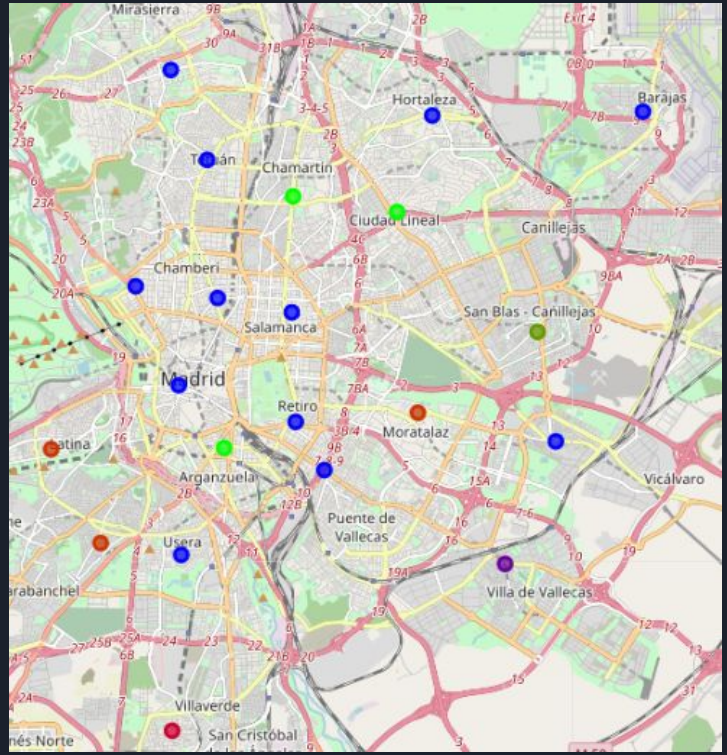
- K-means clustering algorithm by finally setting up the value of k=6

- Districts that are similar to the district target, Chamartin, are Ciudad Lineal, and Arganzuela. (Green)



Discussion

- Most of the districts can be included in cluster number 1 (blue) and are places more related to residential areas
- The second big group is the red one, which include outer districts like Latina, Carabanchel or Moratalaz with similar characteristics
- Villa de Vallecas, San Cristobal and San Blas Canillejas districts with unique characteristics. The three of them are far from the city centre
- Green cluster, which contains the districts of Chamartin, Ciudad Lineal, and Arganzuela



Conclusion

01 The two districts that are similar according the output data of the K-means algorithm are: Ciudad Lineal and Arganzuela.

02 Both of these districts have similar characteristics to Charmartin and placing the new offices in any of these two districts, the satisfaction of employees based on the venues and the population density should be reached.

THANK YOU