

IBM Applied Data Science Capstone

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

Raúl Cabanas Contreras
June 2020



Introduction

For many companies the decision of where to place a new office building for the company is not an easy process. They have to take into account many factors that are very important such as prize, venues, quick access for clients, employees preferences... and so on. This decisión have to be taken by analyzing data and discusión with all the parts involved.

One of the big factors, is the location according the point of view of the employees. Many studies show that one of the most valuable factors of people to work in a company is its location.

People want to live near the place of work, so they could minimize the transport time. This make the location of the company a key factor to be eligible for the best employees.

Business Problem

In this project we are going to analyze the case of 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 north of Madrid, in the district of Charmartin as is shown in the above map. Employees surveys have been done about the current location, and the results are very goods. Most of the employees rate the current location with a grade 5/5.

For the decision many factors will be reviewed, but this report is focused in determinate which other districts in Madrid are equivalent to Charmartin district in order to maintain the employees satisfaction.

Data

In order to work in the project, 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>
Here we can find many sata set, and it is possible to download the files in json or csv formats.
In the .csv dataset downloaded, we can find information about the 20 districts in madrid: name, latitude, longitude, population, area and district id.
- Venue data from Foursquare API, which as one of the largest database of places and it is used by many developers.

In this project, we will use many Data Science skills, such as loading csv file, data cleaning, data wrangling, working with API, and machine learning algorithm.