Analysing Madrid

Coursera Capstone Project

Analysing Madrid districts for opening a cafeteria

- Starting a new business project is a challenge for stakeholders and shareholders.

- Geographic location plays a major role in decisions to start a new business..

- Knowing the reviews provided by the Foursquare API is a step further and an extra help for a venture.

Data acquisition and cleaning

- COVID data are taken from https://datos.comunidad.madrid

- Districts and neighborhoods information https://datos.gob.es

- Prominents Madrid's data https://datos.madrid.es

- Irrelevant data such as area codes were deleted.

- The district dataset contains 21 features corresponding to the districts of Madrid.

Datasets and preprocessing

- The data for venues, neighborhoods and district was extracted using the Foursquare API.

- The Madrid COVID dataset is distributed by districts.
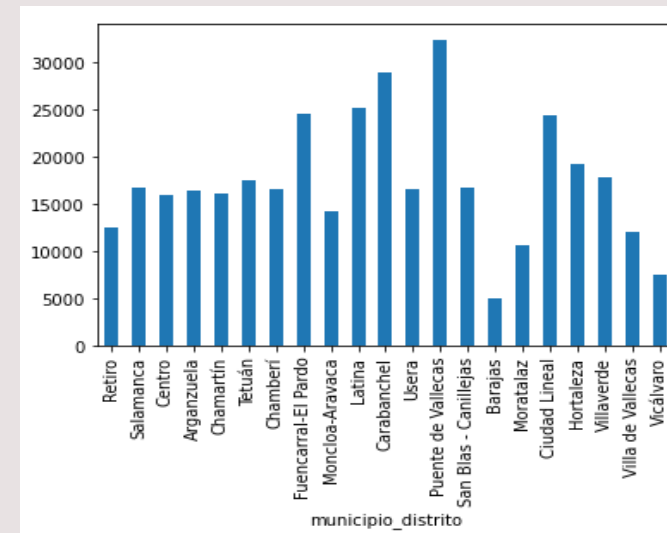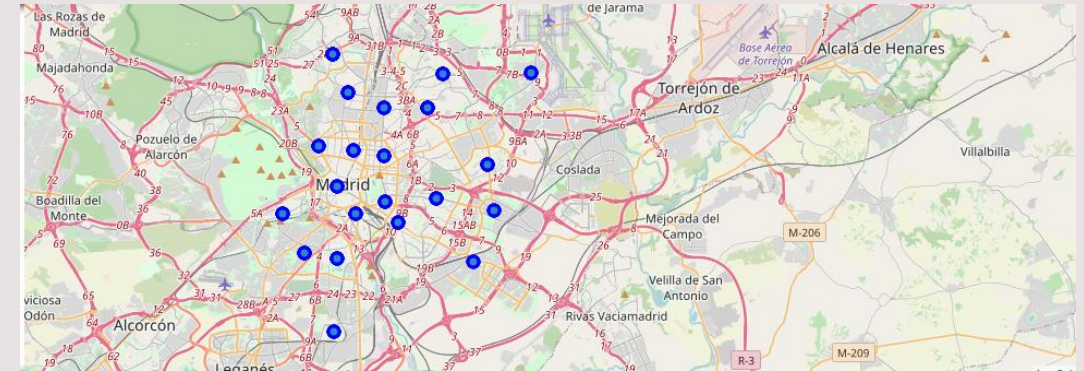
Dataset Clustering

- Clustering technique is used to identify the best and most popular areas to start the new business.

- The optimal number of clusters is obtained by using the Silhouette Score.
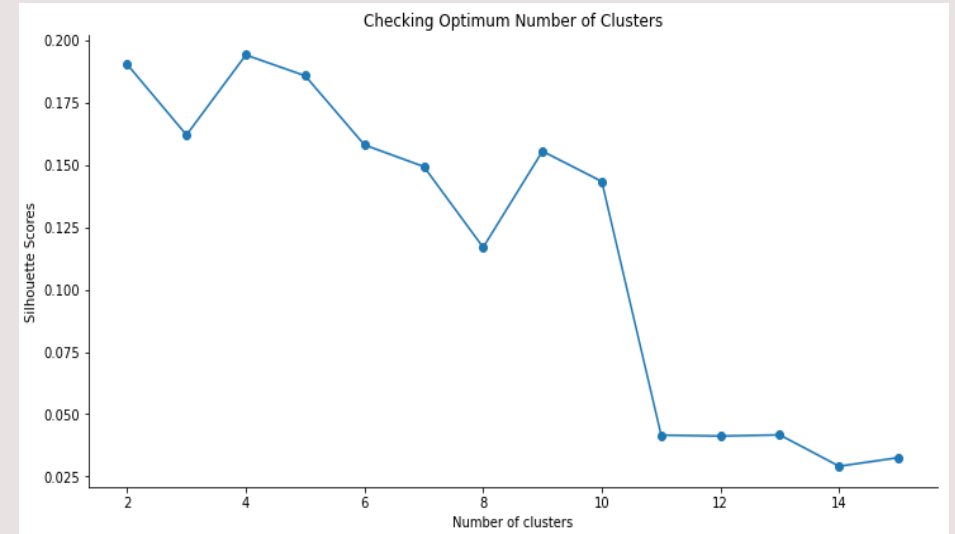
Dataset Clustering

- The number of clusters used is 4, being cluster number 1 the best option because it has more districts and the best venues in cafeteria and coffee shop.
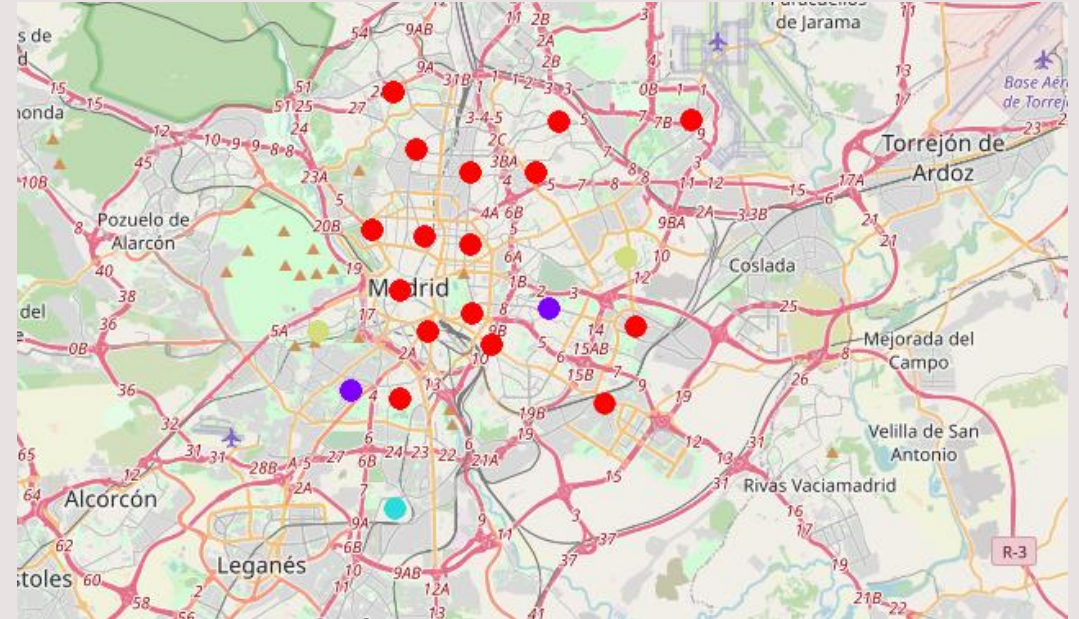
Dataset clustering

- Puente de Vallecas and Barajas are the most popular districts for coffee shops and cafes.

- The most touristic districts offer a wide range of culinary options.

- Other districts such as Centro and Moncloa-Aravaca offer similar restaurant-type offerings, which is indicative of a culinary trend in the area.
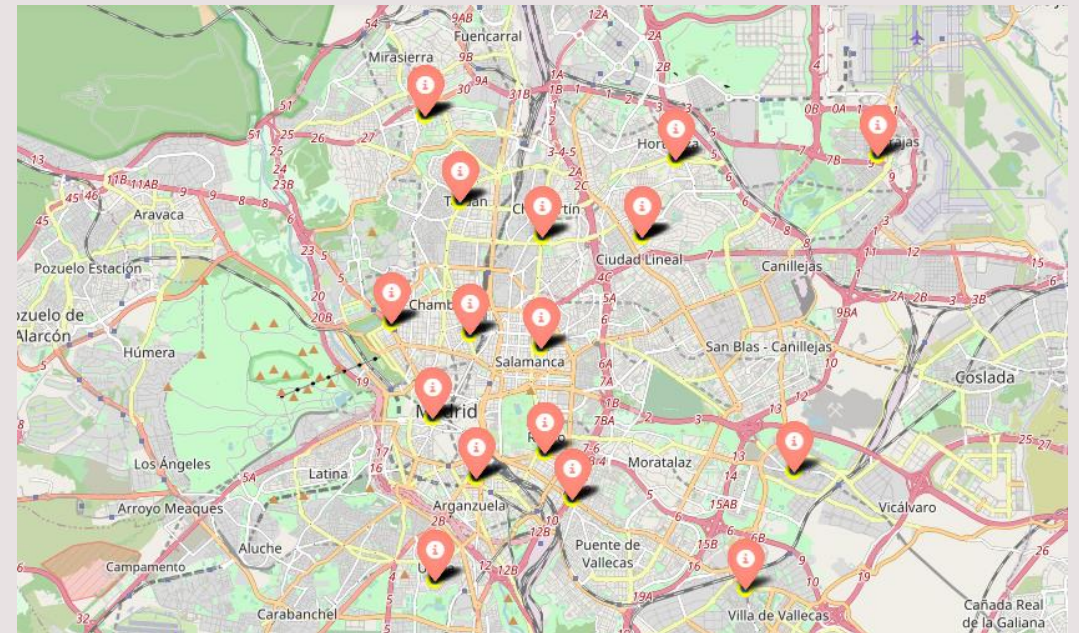
Conclusión and future directions

Clustering model shows trends and commercial offers using the Foursquare API.

Future stakeholders will decide the best location also taking into account the COVID data by district whose average is 17382.23 cases.

The downtown district has 15821 cases below average and is a good location for coffee shops.

The Barajas district with 4907 cases is popular with coffee shop venues.

Other forward-looking ideas that influence decision making are: costs, transportation, safety, security, etc.