Prime location for Spanish Lenguage Academy in Madrid, Spain

IBM DATA SCIENCE PROFESSIONAL CERTIFICATION

INDEX

INTRODUCTION

- **Problem description:**

Optimal location for a Spanish as a Second Language Academy in Madrid, Spain. In order to do so an analytical approach will be used with advanced machine learning, using clustering to solve the problem.

- **Data presentation:**

There are two databases that will be accessed to do this project:

A. **Foursquare API:** most common venues per neighbourhood in the city and to understand where people might be interested to attend Spanish classes.

B. **Madrid City Hall's We portal**. Iimmigrant information per country and nationality in Madrid.

- **Target audience:**

This project is for both immigrants and tourists that want to learn Spanish this is why there will be a cross validation between the immigrants and where they live and the most popular places in Madrid.

DATA

City Hall: Nationalities and neighborhoods

	Country of Procedence	Total Ciudad de Madrid	Centro	Arganzuela	Retiro	Salamanca	Chamartin	Tetuán	Chamberí	Fuencarral-El Pardo	...
0	Rumanía	45036.0	815.0	754.0	480.0	753.0	680.0	1468.0	597.0	1830.0	...
1	China	37276.0	1508.0	1356.0	564.0	755.0	652.0	1988.0	816.0	1733.0	...
2	Ecuador	23953.0	647.0	741.0	265.0	619.0	380.0	1395.0	453.0	632.0	...
3	Venezuela	23359.0	1563.0	913.0	638.0	1564.0	933.0	1310.0	794.0	1428.0	...
4	Colombia	22618.0	998.0	717.0	483.0	803.0	551.0	822.0	659.0	999.0	...

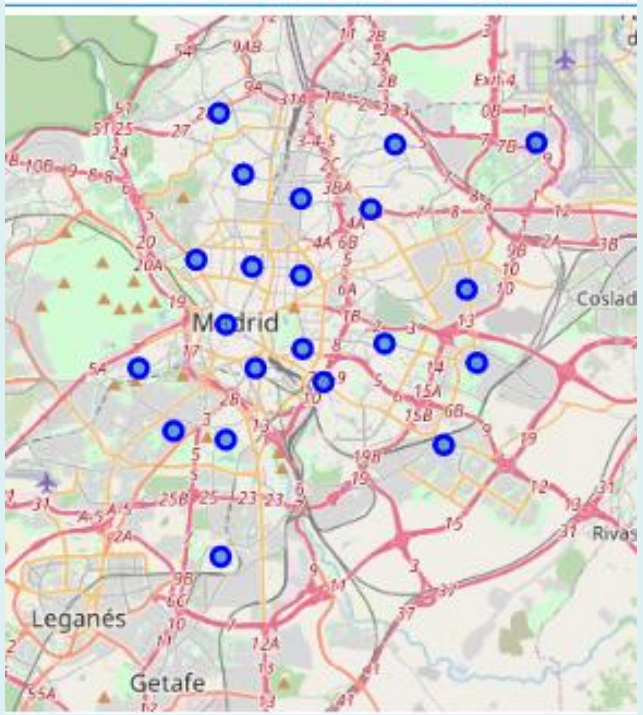
Foursquare API: Madrid Venues by neighborhood

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Centro	40.415347	-3.707371	La Taberna de Mister Pinkleton	40.414536	-3.708108	Other Nightlife
1	Centro	40.415347	-3.707371	The Hat Madrid	40.414343	-3.707120	Hotel
2	Centro	40.415347	-3.707371	Plaza Mayor	40.415527	-3.707506	Plaza
3	Centro	40.415347	-3.707371	Plaza Menor	40.414192	-3.708494	Lounge
4	Centro	40.415347	-3.707371	Bodegas Ricla	40.414266	-3.708077	Wine Bar

Data frame: Neighborhoods coordinates

```
coord_df.head()
```

	Neighborhood	Latitude	Longitude
0	Centro	40.415347	-3.707371
1	Arganzuela	40.402733	-3.695403
2	Retiro	40.408072	-3.676729
3	Salamanca	40.43	-3.677778
4	Chamartin	40.453333	-3.6775



MACHINE LEARNING: CLUSTERING

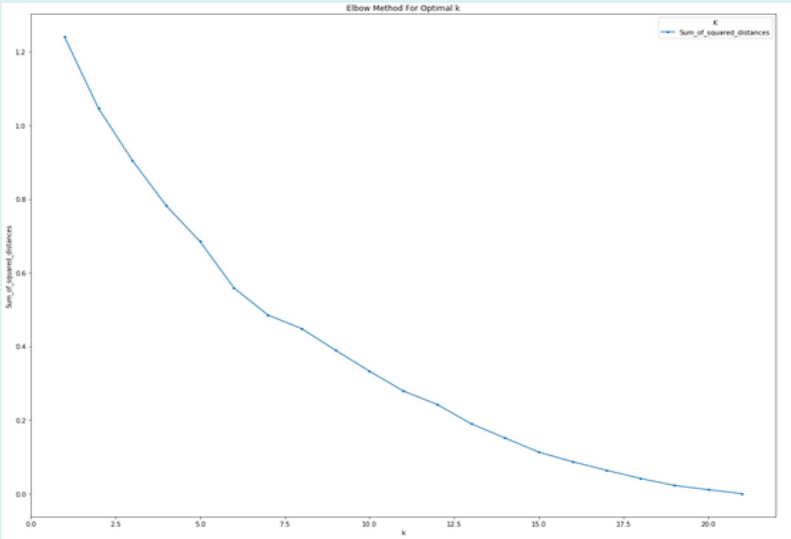
Clustering: K-means algorithms

- Iterative method until it converges (Careful: it might converge to a local minimum, this is why the process needs to be run several times over the date)
- It is very important to normalize the data
- Calculates the distance between the centroids and the points and finds similarities this way:

$$Distance(x_1, x_2) = \sqrt{\sum_{i=0}^{n} (x_{1i} - x_{2i})^2}$$

Convergence

- In this case the model converves when k=5 because:
- For this value, there is an elbow point in the gaph that shows the mean distance of data points in the cluster to the centroid vs. the value of k.

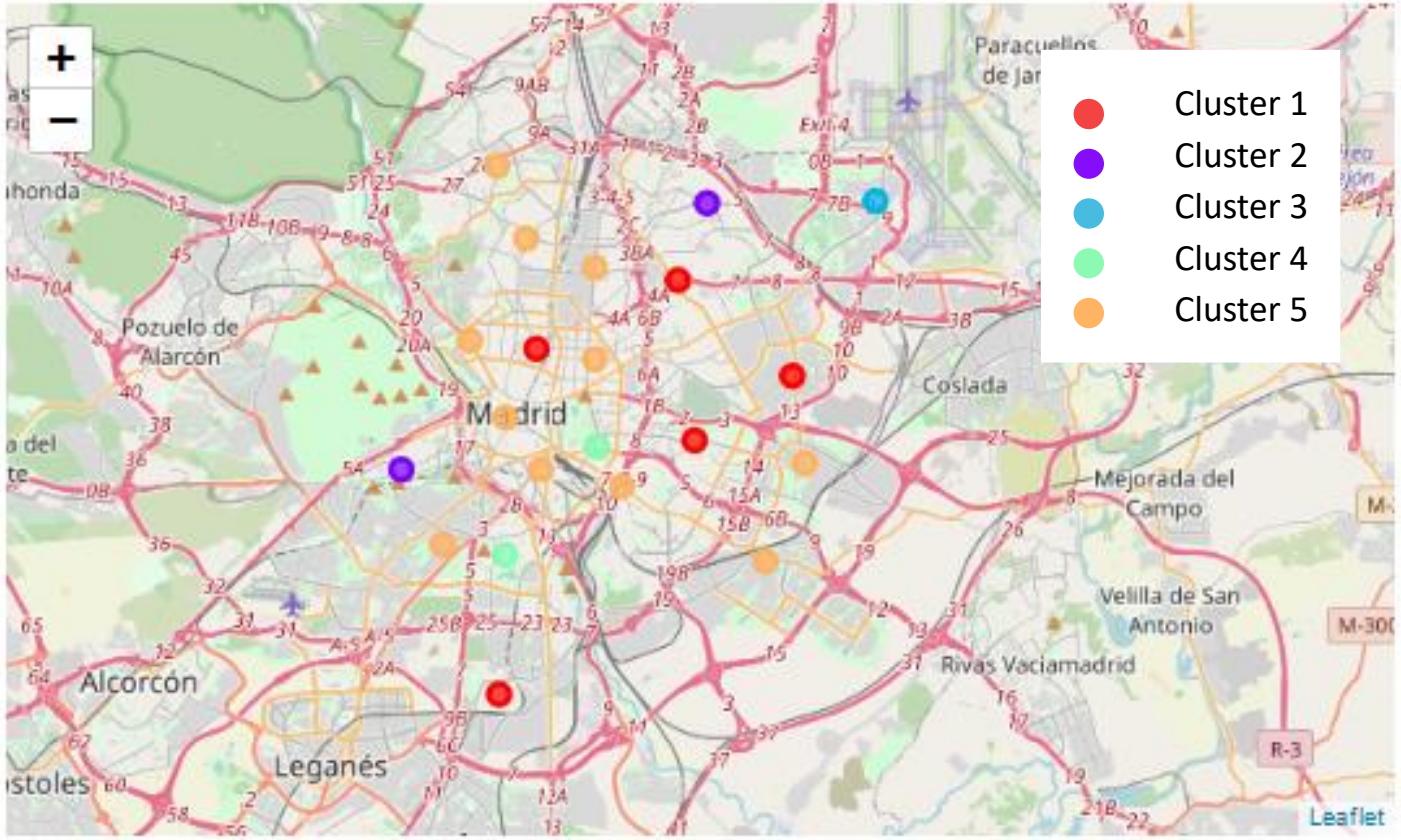


As it can be seen on the graph, k=5 is the optimal value of K as it is a little higher and off trend than the rest of the values.

RESULTS

Clustering results

- **Cluster 1**: centric, contains mostly west European nationalities: French, Portuguese, Italian and British and most of the cultural venues such as Plazas or Museums.
- **Cluster 2**: outskirts of town
- **Cluster 3**: Airport neighborhood, far away from the cultural venues and from general leisure cultural events.
- **Cluster 4**: South east of Madrid with a high Spanish speaking foreigner population. (Latino neighbourhoods)
- **Cluster 5**: centric and large number of neighborhoods, contains great variety of nationalities.



CONCLUSION

CLUSTER 1

	Country of Procedence	Salamanca
19	Reino Unido	550.0
13	Portugal	695.0
6	Italia	1817.0
14	Francia	968.0
16	Brasil	431.0

Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
Villa de Vallecas	0	Bar	Plaza	Spanish Restaurant	Grocery Store	Soccer Field	Platform	Diner
Retiro	0	Spanish Restaurant	Bar	Museum	Supermarket	Tapas Restaurant	Gym	Italian Restaurant
Ciudad Lineal	0	Spanish Restaurant	Gastropub	Supermarket	Argentinian Restaurant	Restaurant	Burger Joint	Bakery
Salamanca	0	Spanish Restaurant	Restaurant	Mediterranean Restaurant	Seafood Restaurant	Burger Joint	Tapas Restaurant	Bakery
Tetuán	0	Spanish Restaurant	Supermarket	Grocery Store	Brazilian Restaurant	Chinese Restaurant	Bakery	Coffee Shop


Thank You!