Analysis of venues categories in the neighborhoods of Madrid - Week 2

Applied Data Science Capstone by IBM/Coursera

Introduction

- Business problem

- **How are venues distributed among the neighborhoods in Madrid?**
- **Where are the neighborhoods in which a certain venue category is specially usual?**
- **Which is the most common venue category in each neighborhood?**
- **Is there any "outlier" neighborhood in Madrid? Could this information be leveraged in any way?**

Introduction

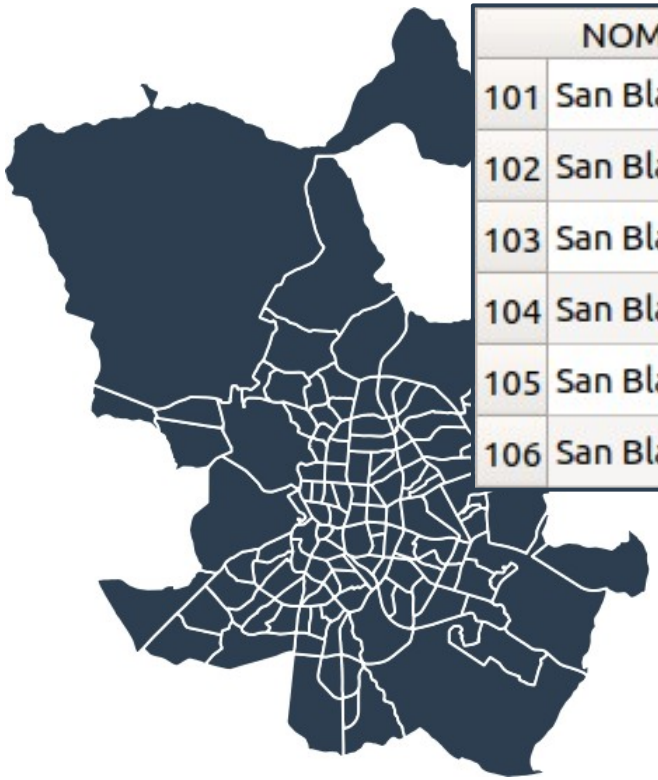
- Stakeholders

- **The stakeholders are entrepreneurs and investors from the fields of culture, leisure, catering and hostelry.**
- **If a business of a specific type is to be put into operation, where are the neighborhoods in Madrid which should be targeted?**
- **If a stakeholder plans to put a business of any type into operation in a specific neighborhood of Madrid, what business types should they consider?**
- **Is there any "special" neighborhood in Madrid regarding its venues?**

Data

- Madrid neighborhoods

- **Freely available in Madrid Open Data website.**
- **Neighborhood name, borough name, area, coordinates of their centroids, etc.**



	NOMDIS	NOMBRE	area	xcoord	ycoord
101	San Blas - C…	Simancas	2276342,4…	446981,67…	4476263,7…
102	San Blas - C…	Hellín	550555,49…	447680,48…	4475772,5…
103	San Blas - C…	Canillejas	1597706,2…	448224,08…	4477135,6…
104	San Blas - C…	Salvador	1880110,2…	446491,59…	4477373,1…
105	San Blas - C…	Rosas	9296500,8…	449590,42…	4475395,5…
106	San Blas - C…	Rejas	5015732,1…	451648,39…	4477317,1…

Data

- Venues locations

- **Extracted from Foursquare leveraging its API.**
- **Venue name, category, coordinates, etc.**

	Name	Category	Latitude	Longitude
0	La Gelateria di Angelo	Ice Cream Shop	40.397951	-3.707739
1	Parque de la Arganzuela	Park	40.398330	-3.708686
2	Restaurante Peruano Mis Tradiciones	Peruvian Restaurant	40.399816	-3.711022
3	sushi raku	Sushi Restaurant	40.404623	-3.708216
4	Le Crust Pizza Bar	Pizza Place	40.400922	-3.709890
5	AltaFit Puerta de Toledo	Gym / Fitness Center	40.405115	-3.707649

Data

- Final merged dataframe

- **Input data are pre-processed to build a final merged dataframe which shows how many venues of each category are in each neighborhood in Madrid.**
- **pandas library is leveraged to do so.**

	Neighborhood_	Accessories Store	Adult Boutique	Airport	Airport Lounge	Airport Service	American Restaurant	Aquarium	Arcade	Arepa Restaurant
0	Abrantes	0	0	0	0	0	0	0	0	0
1	Acacias	0	0	0	0	0	0	0	0	0
2	Adelfas	0	0	0	0	0	0	0	0	0
3	Aeropuerto	2	0	1	8	5	0	0	0	0
4	Alameda de Osuna	0	0	0	0	0	0	0	0	0

Methodology/analysis

- K-Means clustering algorithm

- **K-Means clustering algorithm is run against the aforementioned final dataframe in order to cluster the neighborhoods in Madrid according to the categories of their venues. scikit-learn library is leveraged to do so.**

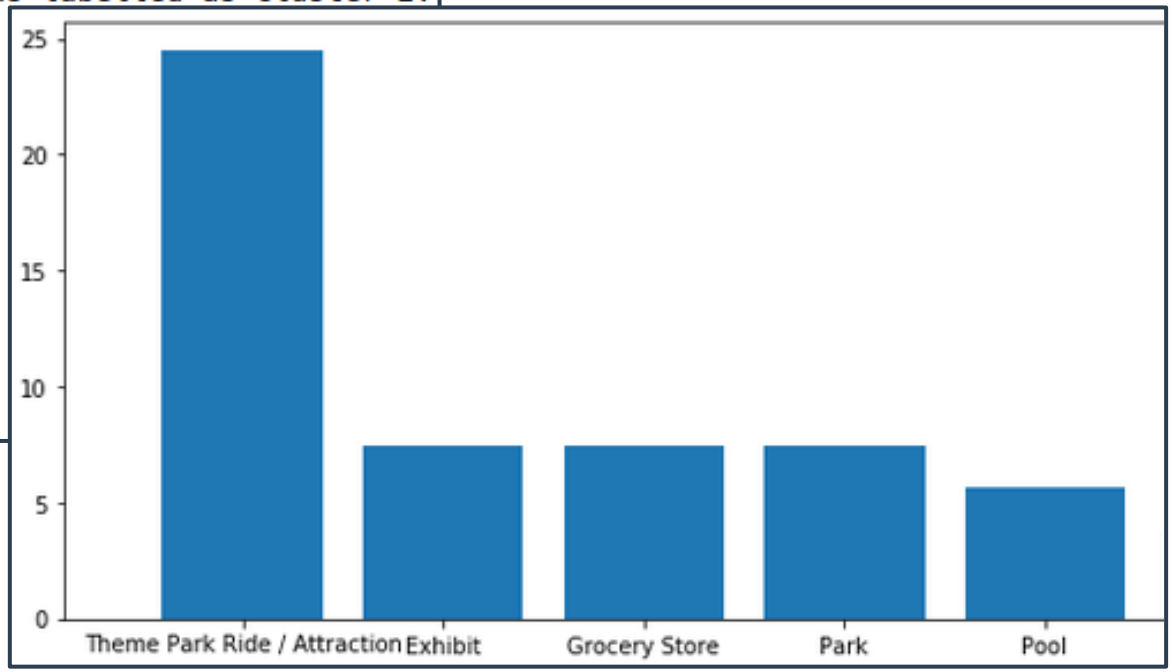
```
clusters_number = 5

# The neighborhood column is dropped
madrid_venues_neighborhood_clustering = madrid_venues_neighborhood.drop("Neighborhood_", axis = 1)

# The dataframe is normalized so that features with different magnitudes and distributions are interpreted equally
madrid_venues_neighborhood_clustering = StandardScaler().fit_transform(madrid_venues_neighborhood_clustering)

# K-Means algorithm is run
kmeans = KMeans(n_clusters = clusters_number, random_state = 0).fit(madrid_venues_neighborhood_clustering)

# Cluster labels are checked
kmeans.labels_[0:10]
```

Methodology/analysis

- Clusters inspection

- **Then, the resulting clusters are examined one by one to determine their distinctive features.**

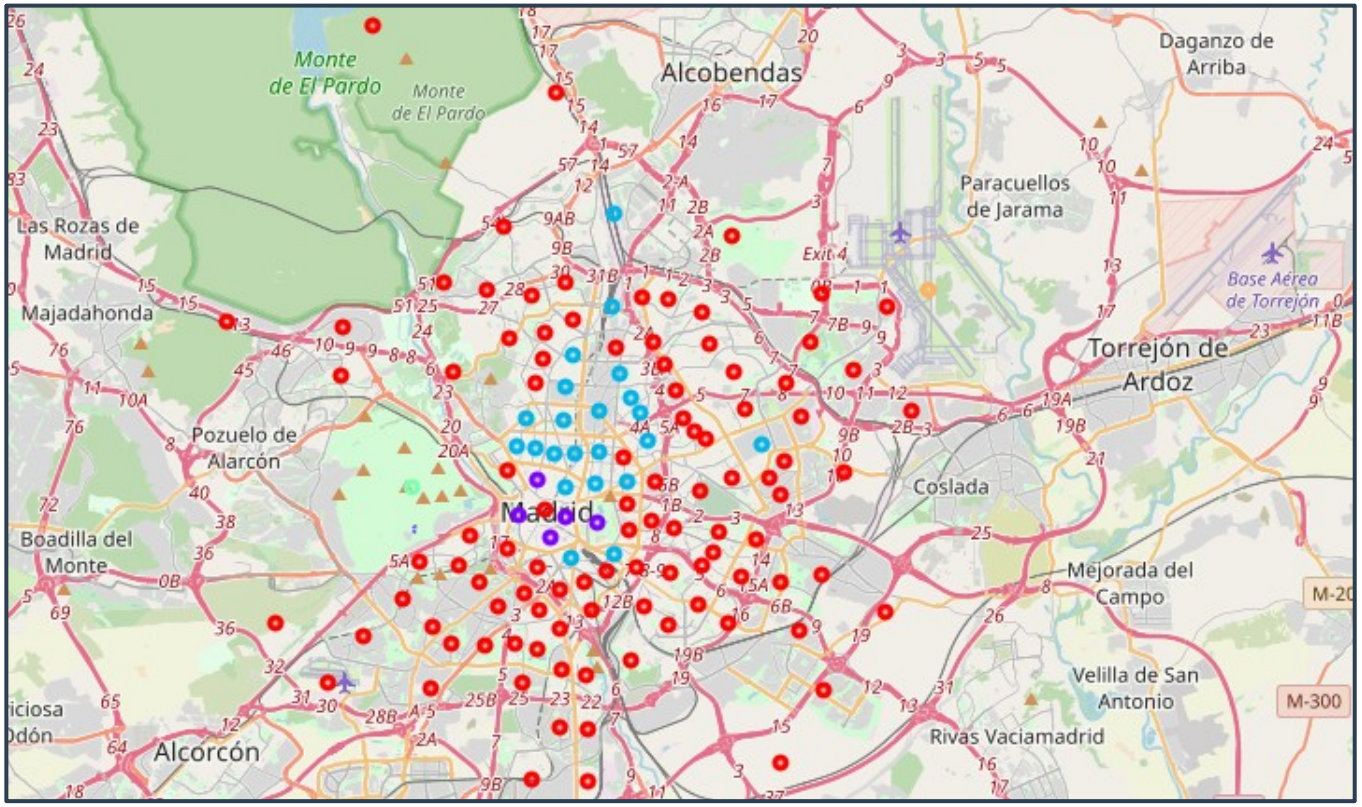
```
**************************************************** CLUSTER 2 ***

There are 22 neighborhoods labelled as Cluster 2 (17.1 %).

There are 976 venues in the neighborhoods labelled as Cluster 2.

                Count  Percentage
Bakery             69         7.1
Café               69         7.1
Coffee Shop        60         6.1
Hotel              54         5.5
Supermarket        42         4.3
Burger Joint       34         3.5
Plaza              31         3.2
Sandwich Place     30         3.1
Grocery Store      29         3.0
Brewery            21         2.2
```



Methodology/analysis

- Final map

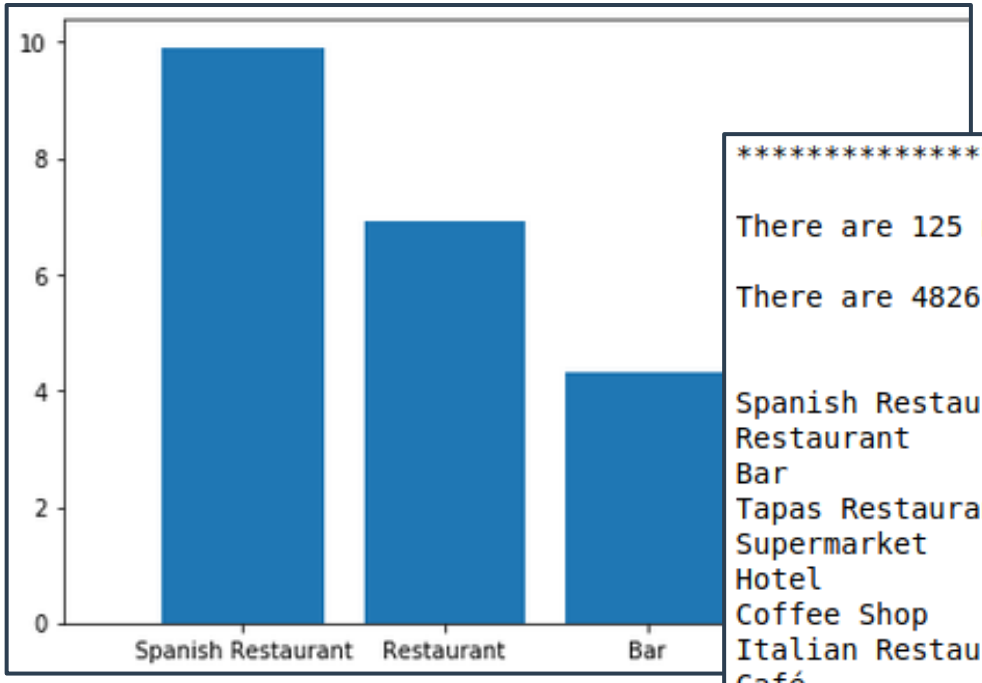
- **Finally, the neighborhoods in Madrid are displayed in a map and their icons are coloured according to their clusters. This way, the distribution of the clusters in Madrid can be easily visualized. folium library is leveraged to do so.**



Results/conclusions

- First approach

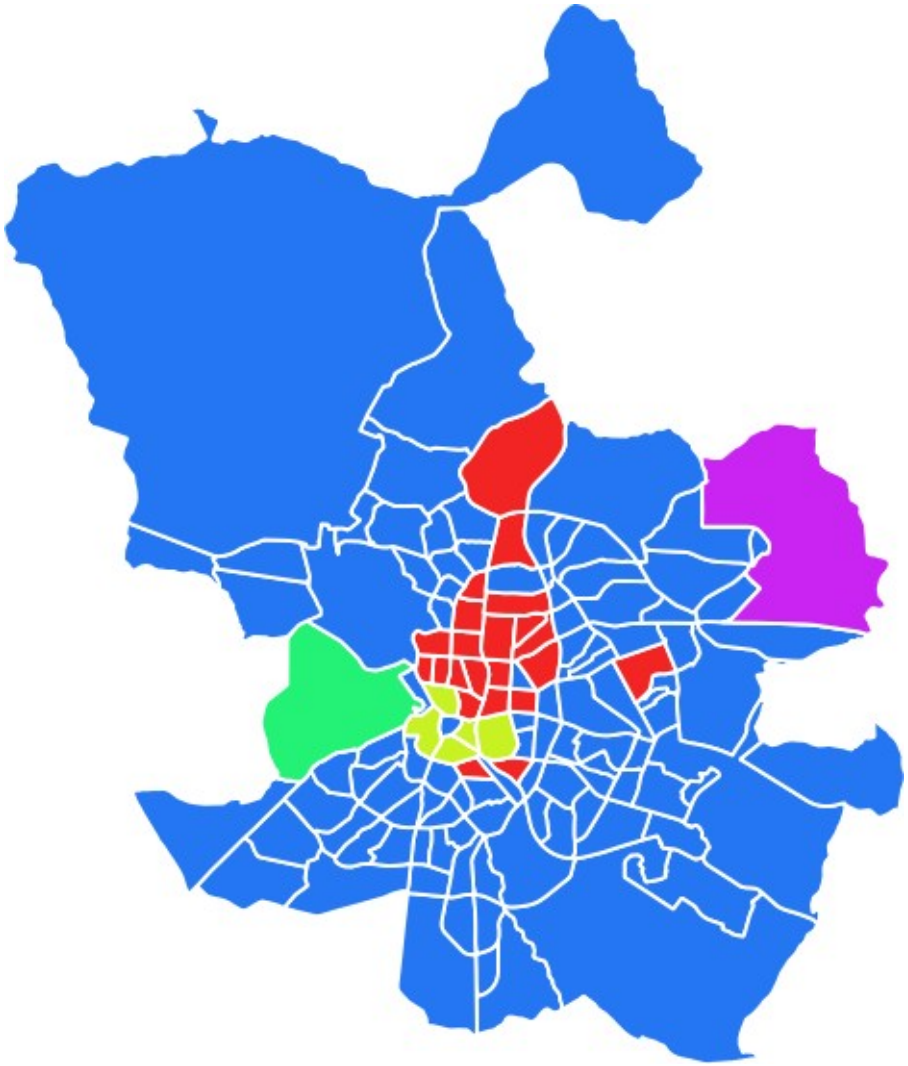
- **It seems that <u>restaurants are ubiquitous in Madrid!</u> On one side, this could be the first conclusion. On the other side, this forced to adopt a second approach extricating the restaurants from the analysis in order to get further details.**




```
********************************************************* CLUSTER 1 ****

There are 125 neighborhoods labelled as Cluster 1 (96.9 %).

There are 4826 venues in the neighborhoods labelled as Cluster 1.

                      Count  Percentage
Spanish Restaurant      478         9.9
Restaurant              332         6.9
Bar                     209         4.3
Tapas Restaurant        176         3.6
Supermarket             136         2.8
Hotel                   135         2.8
Coffee Shop             125         2.6
Italian Restaurant      125         2.6
Café                    124         2.6
Bakery                  124         2.6
```

Results/conclusions

- Second approach



- **Parks, groceries stores and supermarkets**
- **Cafés and coffee shops, bakeries and hotels**
- **Hotels and plazas**
- **Theme park rides and attractions**
- **Airport-related venues**

Results/conclusions

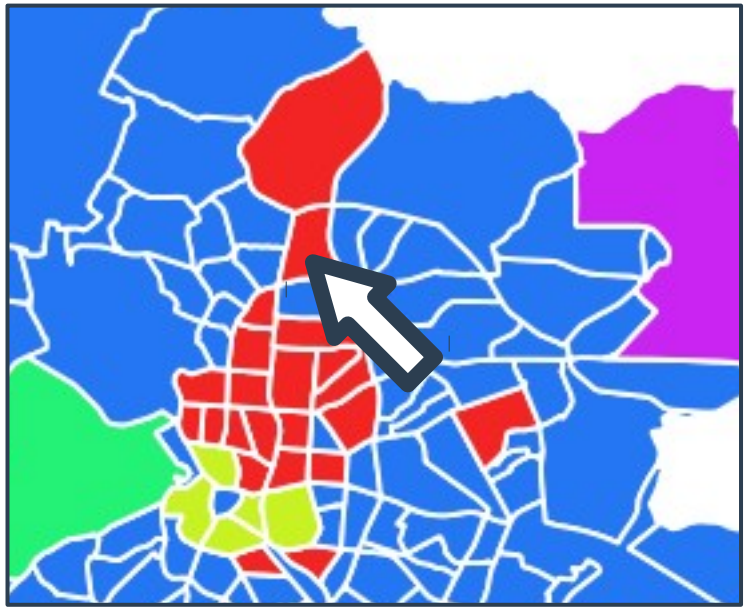
- Second approach

- **In the old town of Madrid (i. e., the neighborhoods in Centro borough) the top venues categories are clearly hotels and plazas. This area is full of historical buildings, monuments, cultural highlights and so on. See yellow colored areas below.**

Results/conclusions

- Second approach

- **In the down town of Madrid (i. e., all boroughs inside the "Central Almond"), the top venues categories are coffee shops and cafés, bakeries and hotels. Though the population is slowly decreasing in this area, here is concentrated most of the economic activity. See red colored areas below.**



Results/conclusions

- Second approach

- **In the suburbs of Madrid (i. e., all boroughs outside the "Central Almond"), the top venues categories are parks, groceries stores and supermarkets. They are some times "bedroom communities", i. e., their inhabitants work in other areas of the city, and come back home only to rest. See blue colored areas below.**

Results/conclusions

- Second approach

- **Casa de Campo is one "outlier" neighborhood. It is very clearly characterized by the presence of theme park rides, attraction and exhibits. This neighborhood borders Casa de Campo, the largest public park in Madrid, which houses Madrid Amusement Park as well as Madrid Zoo. See green colored areas below.**

Results/conclusions

- Second approach

- **<u>Aeropuerto</u> is another "outlier" neighborhood. It is very clearly characterized by the presence of <u>venues linked to airports</u>. This neighborhood includes the territory around Aeropuerto Adolfo Suárez Madrid-Barajas, the main airport of Madrid. See purple colored areas below.**