Where to start a new tapas restaurant in Seville

Applied Data Science capstone
Álvaro Romero Muñiz
28 June 2021

1 **NTRODUCTION**

1.1The place

The city of Seville (Andalusia, Spain) is well know by its gastronomy. You can have a good glass of wine and a tapa anywhere. Seville is divided by the Guadalquivir River. Most of the city is on the east side of the river, while Triana, La Cartuja and Los Remedios are on the west side. Triana is the oldest neighbourhood of the west bank of the river and it is very different from the Ancient District.

Fig 1: Map of Seville

1.2 The problem

The owner of a restaurant in Triana wants to start another business in the Casco Antiguo (Acinet Districton) on the other side of the river, but does not know which area may be the best. He thinks that all the neighbourhoods are similar but, the more similar it is, the easier it will be for their restaurant to succeed.

2 DATA ADQUISITION AND CLEANING

2.1 The source

The Seville city hall has digital resources of their spatial planning. A .kml archive can be downloaded with the districts and neighbourhoods of the city, that should be transformed into points (centroids of polygons) to get the coordinates of each one. There are 108 different areas, some of them contain industrial parks or farmlands and they will not be considered in the study.

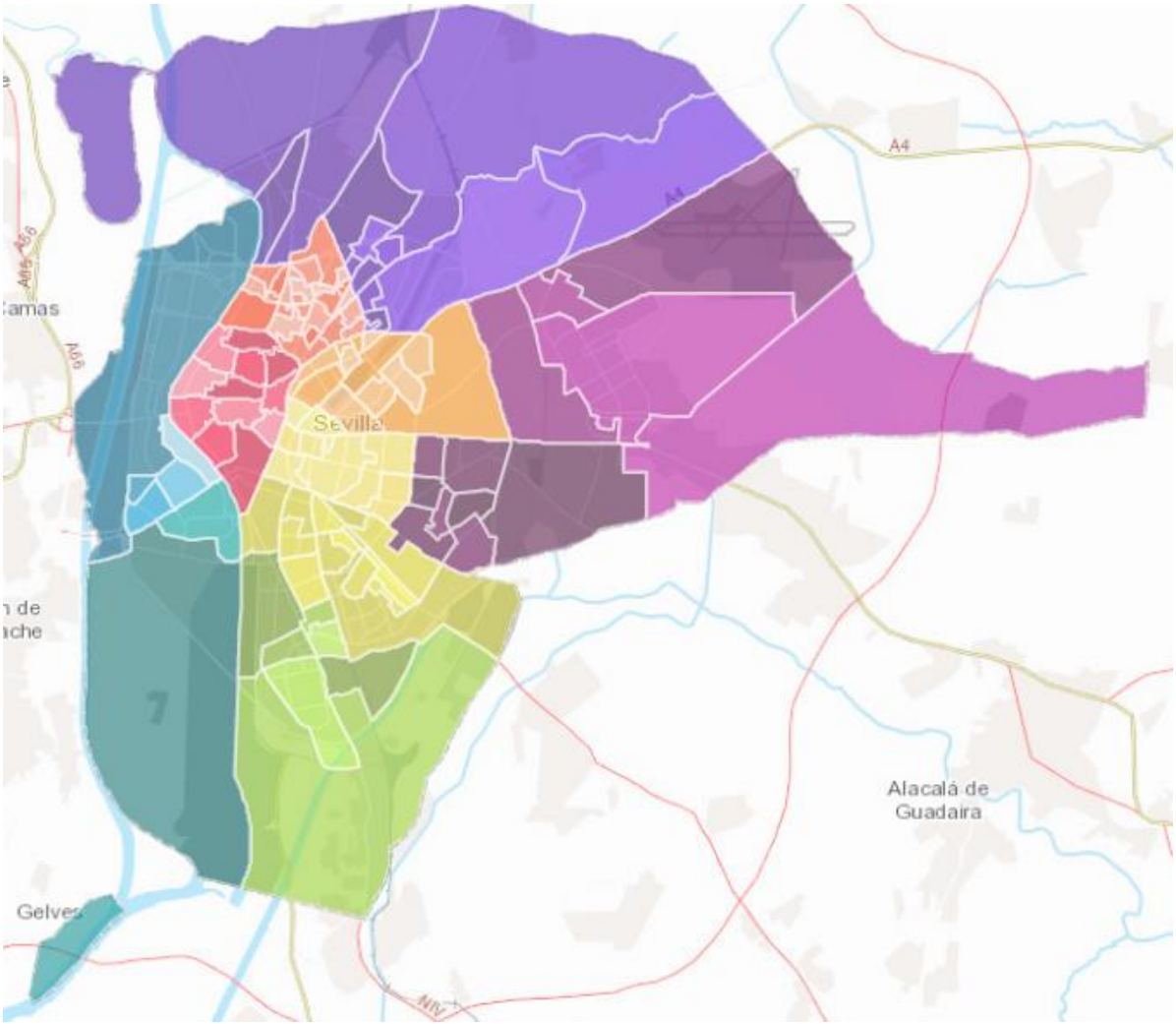

Fig 2: Neighborhoods of Seville (as polygons)

2.2 Data cleaning

The archive with the polygons is converted to a .shp file and the centroids of the neighborhoods are obtained using the method centroid of the GeoPandas Python library. The result is stored in a comma separated value file to use in the analysis with the Jupyter Notebook.

The file with the coordinates of the neighborhoods can be read with pandas and the venues in each area are gotten using foursquare. There are 908 results with 136 different categories, but only 55 neighborhoods have 10 or more venues. That will be the sample used in the algorithm.

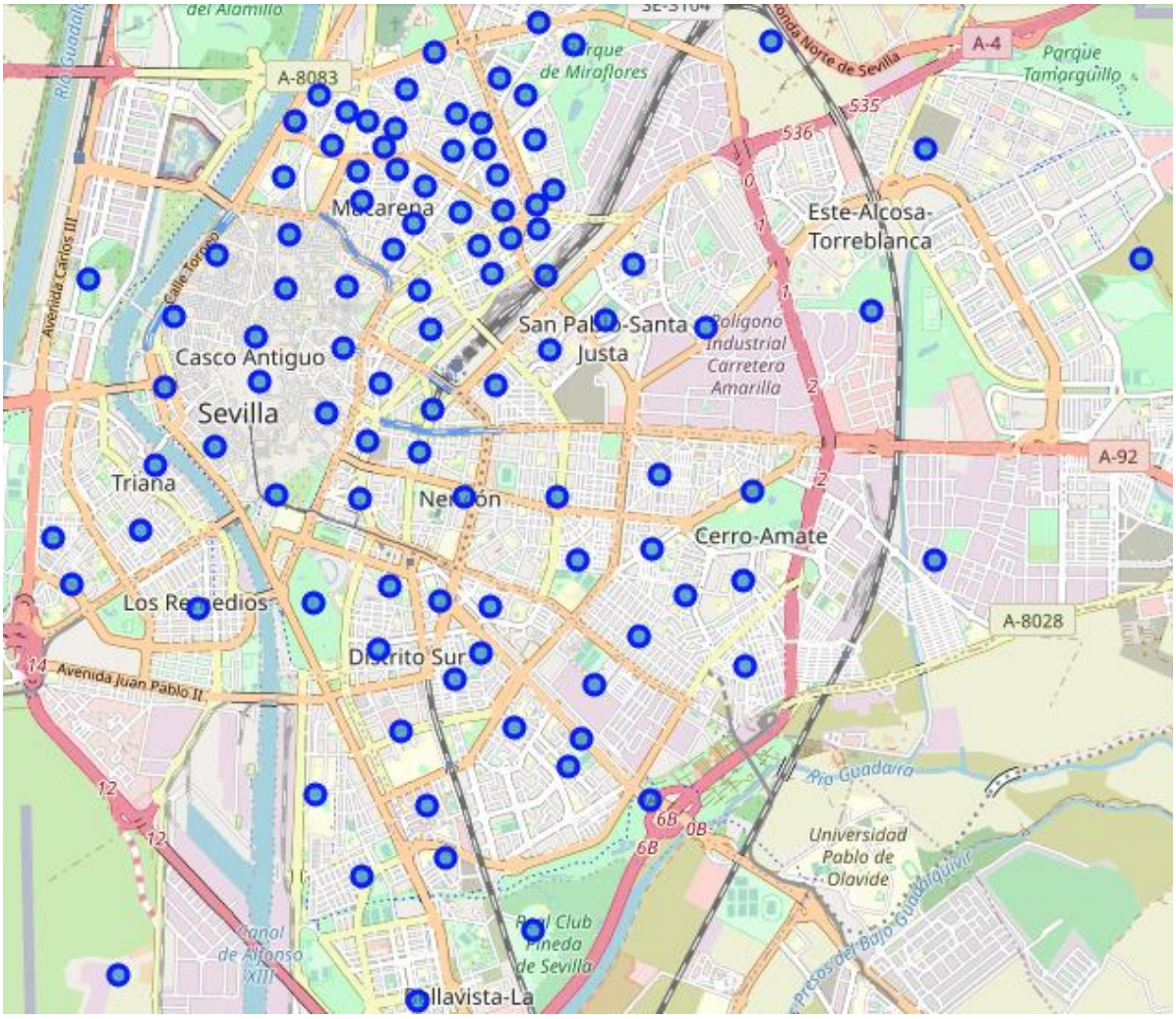

Fig 3: Centroids of the neighborhoods

A dataframe is creaed with the neighborhoods of the sample and the most common venues. It is transformed using dummy variables instead of the feature name.

3 MODEL

The algorithm of k-means clustering is run the dataframe of dummy variables and different number of clusters. This method is able to classify the different neighborhoods without additional information. With more than seven some clusters remain empty so, the most accurate classification is that of kcluster = 7.

```
[91]: # set number of clusters
      kclusters = 7
      seville_grouped_clustering = seville_grouped.drop('Neighborhood', 1)

      # run k-means clustering
      kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(seville_grouped_clustering)

      # check cluster labels generated for each row in the dataframe
      kmeans.labels_[0:10]

      ...

[93]: # Insert the labels
      neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

      # Merge the dataframe with the results to the original
      seville_merged = neighborhoods[neighborhoods['Name'].isin(list(seville_filtered.index))]
      seville_merged = seville_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Name')

      seville_merged.head()
```

Fig 4: Cells that run the algorithm

4 RESULTS

The cluster that cointains Triana (Tirana Casco Antiguo and Triana Este) is shown with the name of the neighborhood and the five most common venues resulted in each one

Table 1: Most common venues of the cluster that contains Triana

Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Retiro Obrero	Coffee Shop	Tapas Restaurant	Lounge	Ice Cream Shop	Spanish Restaurant
El Porvenir	Tapas Restaurant	Spanish Restaurant	Restaurant	Gym/Fitness Center	Grocery Store
Doctor Barraquer, G. Renfe, Policlínico	Tapas Restaurant	Supermarket	Sandwich Place	Beer Garden	Pub
Macarena 3 & 5, Huertas	Tapas Restaurant	Coffee Shop	Fast Food Restaurant	Spanish Restaurant	Café
Triana Casco Antiguo	Tapas Restaurant	Spanish Restaurant	Restaurant	Bar	Brewery
San Lorenzo	Tapas Restaurant	Gay Bar	Spanish Restaurant	Gym / Fitness Center	Mexican Restaurant
Triana Este	Tapas Restaurant	Seafood Restaurant	Gastropub	Chinese Restaurant	Spanish Restaurant
León XIII, Los Naranjos	Tapas Restaurant	Coffee Shop	Spanish Restaurant	Café	Wine Bar
Huerta del Pilar	Tapas Restaurant	Spanish Restaurant	Park	Beer Bar	Mediterranean Restaurant

5 DISCUSSION

As stated in the introduction, the districts of Triana and Casco Antiguo are very different. In the second one, the vast majority of neighborhoods belong to a different group. The distribution shows that the orange ones are more tourist areas, the green ones are new or remodeled residential areas and the purple ones are traditional residential neighborhoods. There are many areas similar to Triana but, most of them are outside the Ancient District.

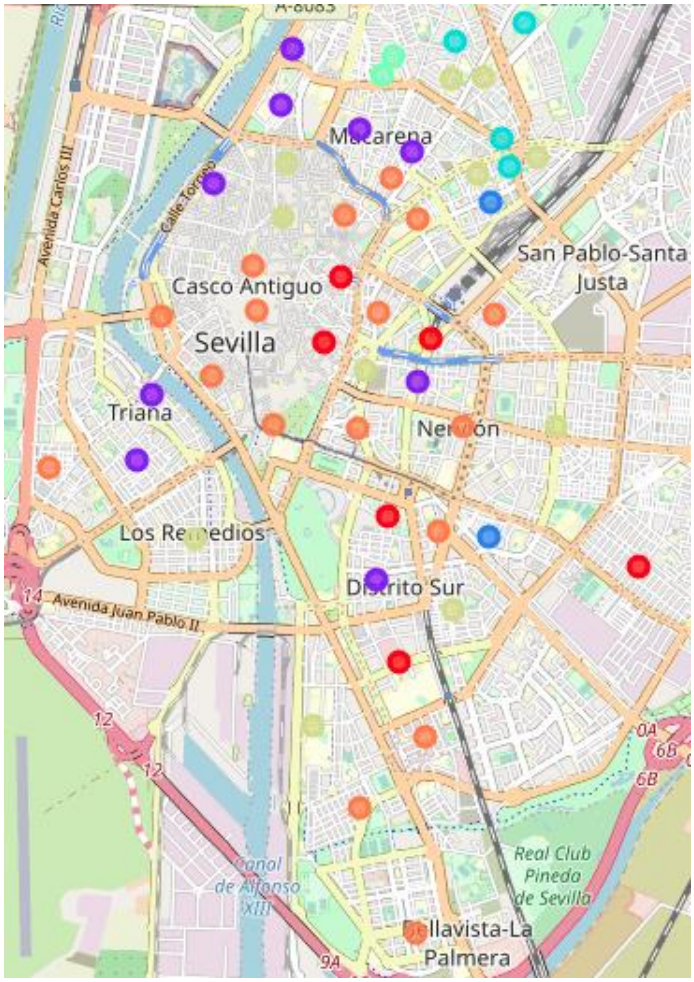

Fig 5: Neighborhoods

6 CONCLUSION

The study shows that there is only one similar location to open the new bar. It is in the cluster that contains Triana and is inside the Ancient District: **San Lorenzo**. So this area will be the best to search commercial premises for rent.