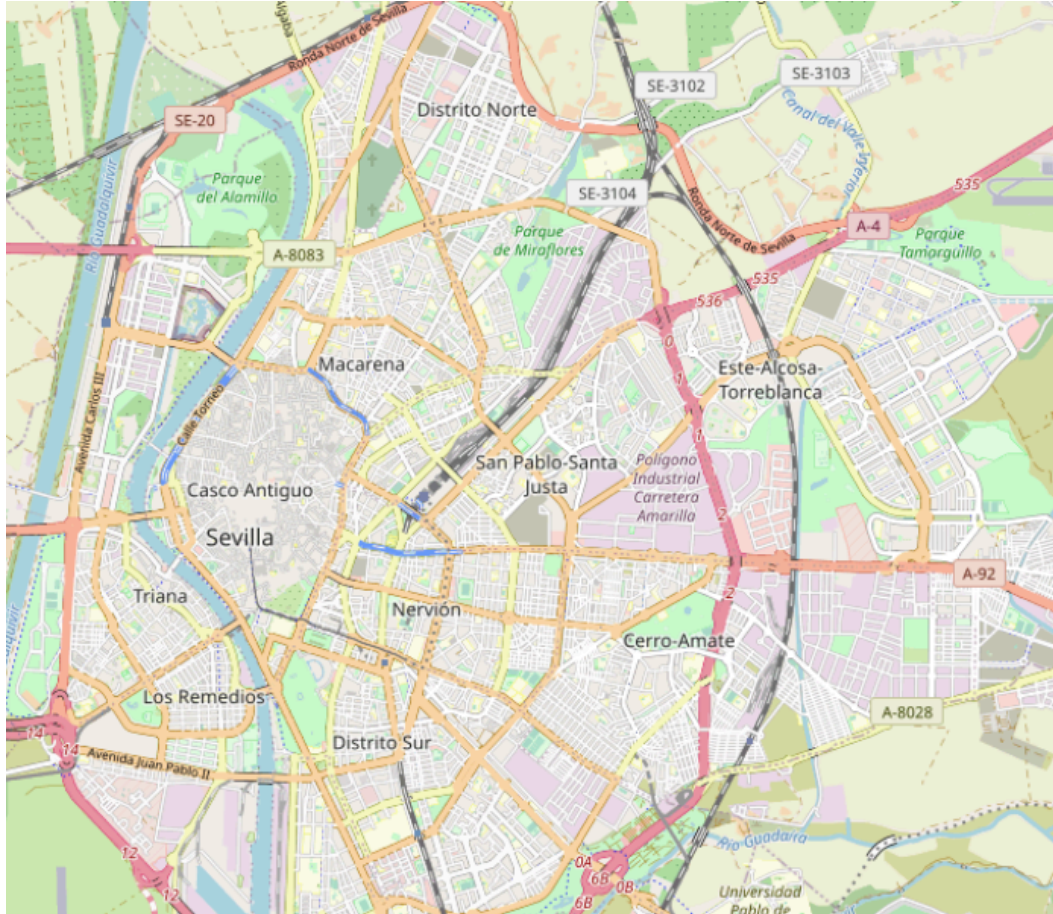# Where to start a new tapas restaurant in Seville

Applied Data Sciece Capstone

Álvaro Romero Muñiz
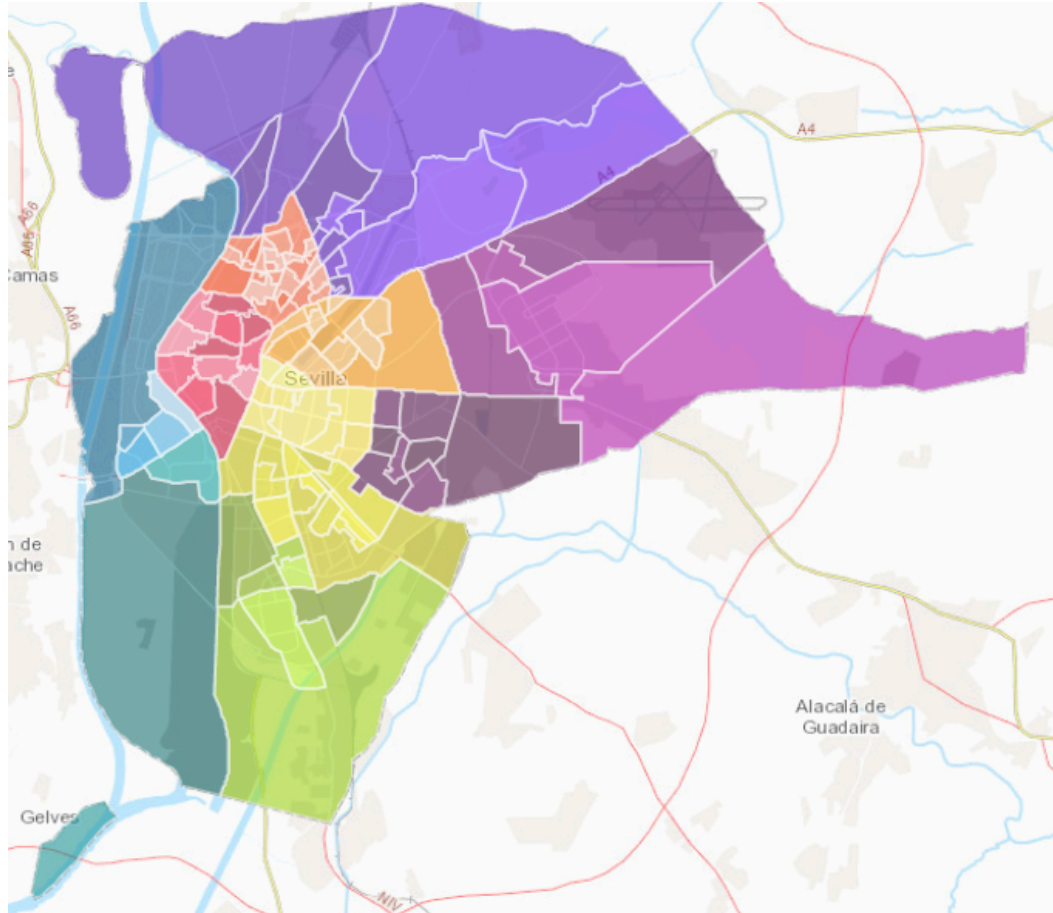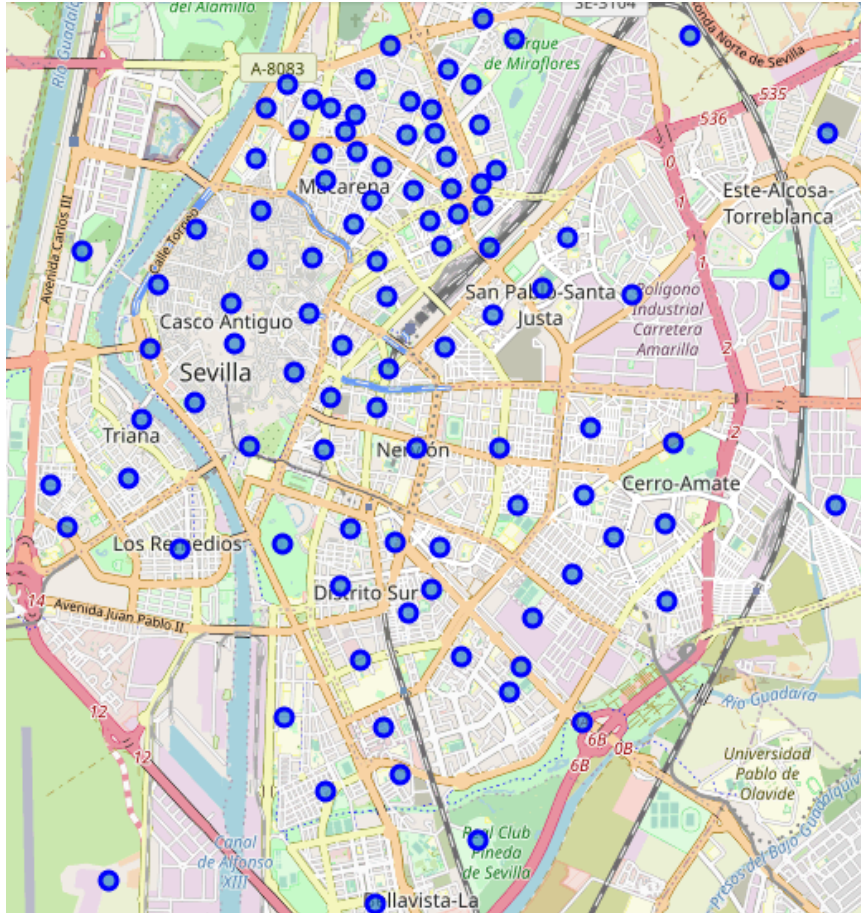
28 June 2021

INTRODUCTION

- 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

- 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. "The more similar it is, the easier it will be for my restaurant to succeed"

DATA SOURCE

- The Seville city hall has digital resources of their spatial planning (urbanismosevilla.org)

- A .kml file 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

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 venues in each area are gotten using foursquare

- There are 908 results with 136 different categories, but only 55 neighborhoods have five or more venues

- A dataframe is creaed with the 55 neighborhoods of the sample and the most common venues. It is transformed using dummy variables instead of the feature name

MODEL

- The algorithm of k-means clustering is run the dataframe of dummy variables and different number of clusters

- Classify the different neighborhoods without additional information. With more than seven clusters some remain empty

- 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()
```
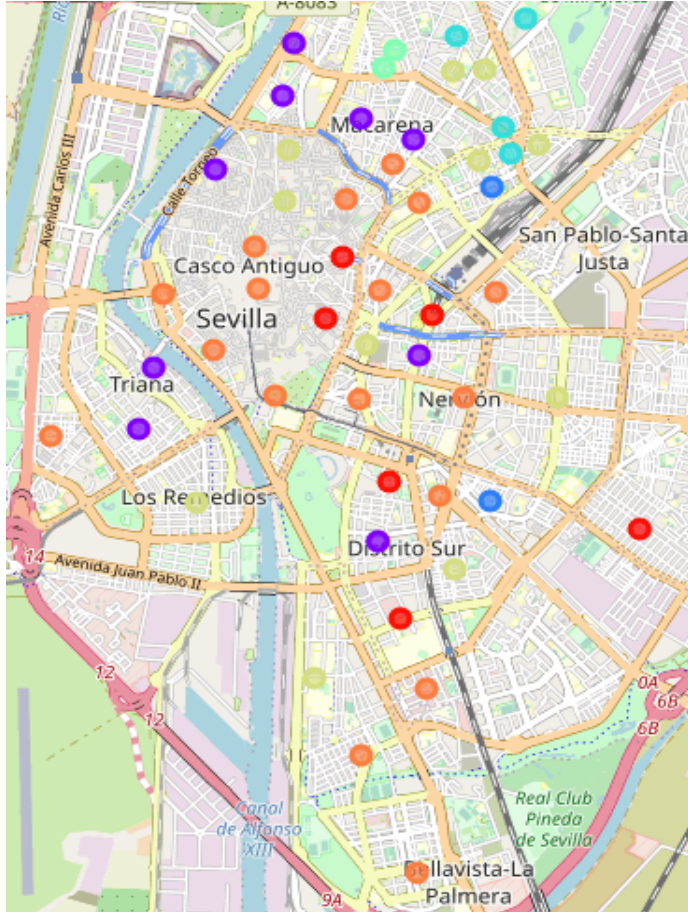
RESULTS

Name	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Retiro Obrero	Coffee Shop	Tapas Restaurant	Lounge	Ice Cream Shop	Spanish Restaurant	Wine Bar	Farmers Market	Food	Fish Market	Fish & Chips Shop
El Porvenir	Tapas Restaurant	Spanish Restaurant	Restaurant	Gym / Fitness Center	Grocery Store	Mexican Restaurant	Tennis Court	Ice Cream Shop	Bar	Café
Doctor Barraquer, G. Renfe, Policlínico	Tapas Restaurant	Supermarket	Sandwich Place	Beer Garden	Pub	Spanish Restaurant	Event Space	Fish Market	Fish & Chips Shop	Fast Food Restaurant
Macarena 3 Huertas, Macarena 5	Tapas Restaurant	Coffee Shop	Fast Food Restaurant	Spanish Restaurant	Café	Wine Bar	Event Space	Food	Fish Market	Fish & Chips Shop
Triana Casco Antiguo	Tapas Restaurant	Spanish Restaurant	Restaurant	Bar	Brewery	Deli / Bodega	Lounge	Market	Fish & Chips Shop	Snack Place
San Lorenzo	Tapas Restaurant	Gay Bar	Spanish Restaurant	Gym, Fitness Center	Mexican Restaurant	Hotel	Café	Indian Restaurant	Bar	Nightclub
Triana Este	Tapas Restaurant	Seafood Restaurant	Gastropub	Chinese Restaurant	Spanish Restaurant	Pub	Fish Market	Fish & Chips Shop	Fast Food Restaurant	Farmers Market
León XIII, Los Naranjos	Tapas Restaurant	Coffee Shop	Spanish Restaurant	Café	Wine Bar	Event Space	Food	Fish Market	Fish & Chips Shop	Fast Food Restaurant
Huerta del Pilar	Tapas Restaurant	Spanish Restaurant	Park	Beer Bar	Mediterranean Restaurant	Restaurant	Pub	South American Restaurant	Bar	Mexican Restaurant

CONCLUSION

- 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

- 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