Applied data science capstone - The battle of Neighbourhoods in the city of Bari

Final report week 5 - Data Science by IBM

Introduction and Business problem

In this final project I focused on parsing and segmenting Bari's neighbours data in order to find the best location in which open a new international food restaurant. Bari is the largest city of Puglia, Italy and it's known for its food tradition, in particular seafood and fresh seafood. Since its food tradition there are very few international options that are mostly Greek or middle-east food restaurant. In this work I simulate to be an investor who wants to open a Mexican or Thai restaurant.
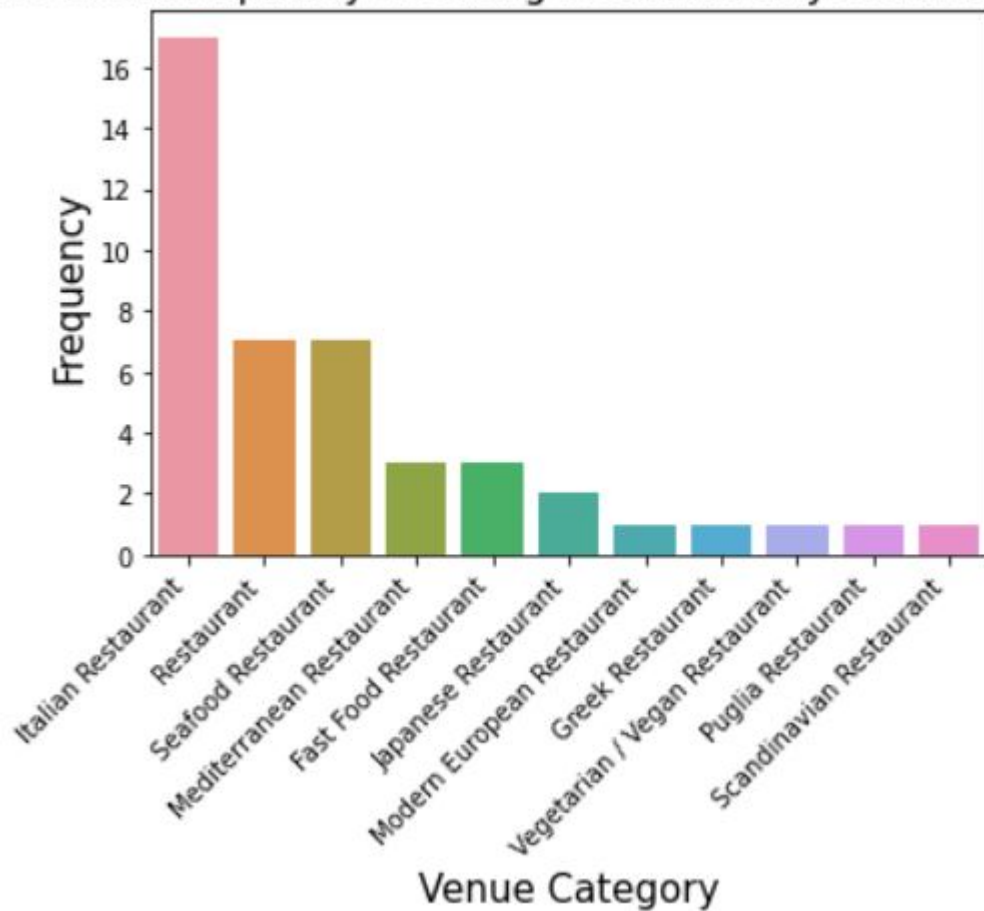
Data collection and description

Data were acquired from https://en.wikipedia.org/wiki/Bari#Quarters (https://en.wikipedia.org/wiki/Bari#Quarters). After exploring a bit the data and using foursquare API to collect informations about venues present over Bari's neighbourhood I focused mostly on restaurants and I create a dataframe that collect infos about Neighbours, their coordinates (latitude and longitude) and the most common venues that chracterize those boroughs.

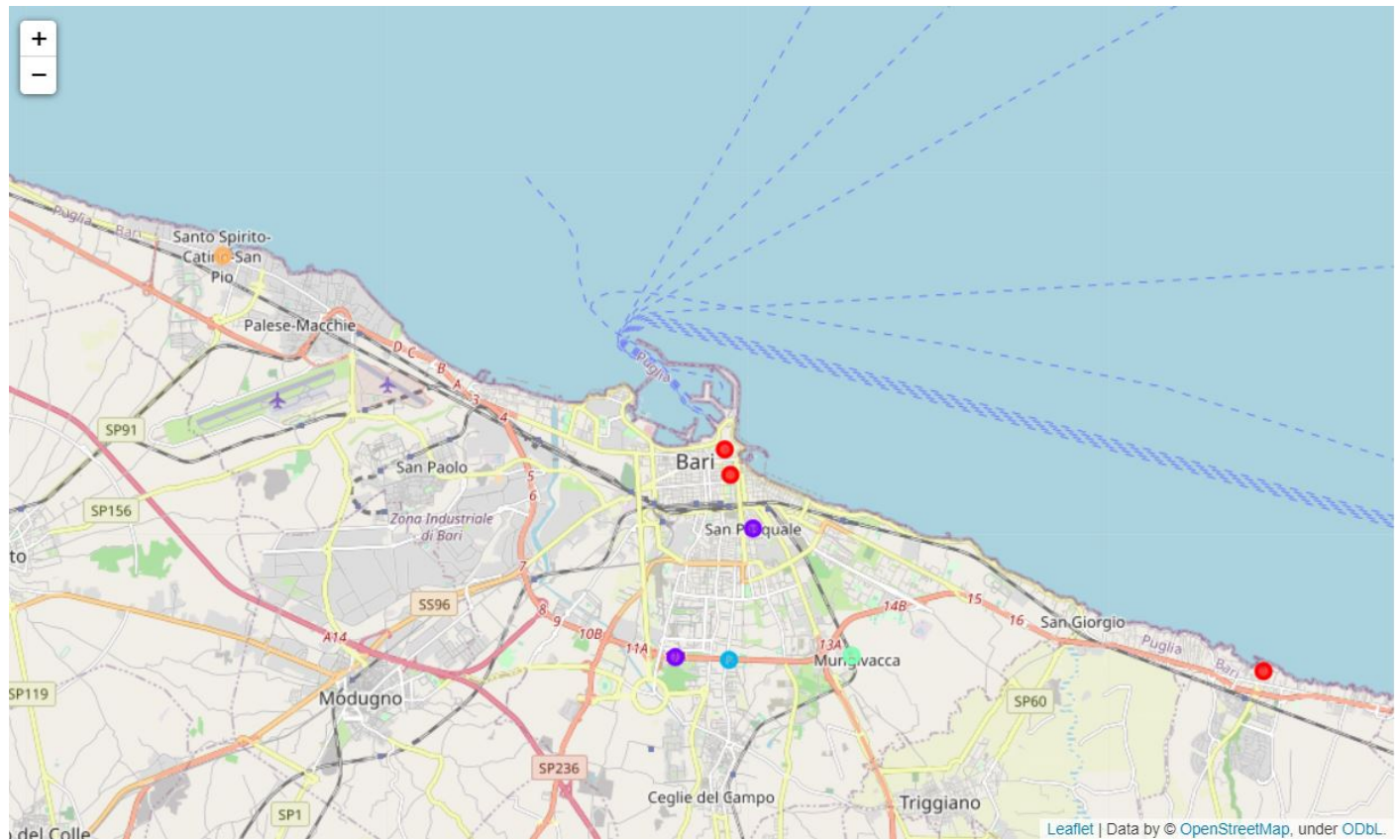|	index	Quarters	city	location	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7t Cc Ve
0	0	Murat	Bari	Bari,Murat	41.123301	16.870601	0.0	Italian Restaurant	Restaurant	Seafood Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Puglia Restaurant	M Eu Re
1	0	San Nicola	Bari	Bari,San Nicola	41.128054	16.869299	0.0	Italian Restaurant	Seafood Restaurant	Restaurant	Mediterranean Restaurant	Japanese Restaurant	Greek Restaurant	Ve Ve Re
5	0	Torre a mare	Bari	Bari,Torre a mare	41.087498	17.000413	0.0	Italian Restaurant	Seafood Restaurant	Vegetarian / Vegan Restaurant	Scandinavian Restaurant	Restaurant	Puglia Restaurant	M Eu Re
6	1	Poggiofranco	Bari	Bari,Poggiofranco	41.090048	16.857068	1.0	Italian Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Scandinavian Restaurant	Restaurant	Puglia Restaurant	M Eu Re
8	1	Carrassi	Bari	Bari,Carrassi	41.089559	16.870042	2.0	Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Scandinavian Restaurant	Puglia Restaurant	Modern European Restaurant	M Re
9	1	San Pasquale	Bari	Bari,San Pasquale	41.113456	16.876145	1.0	Italian Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Scandinavian Restaurant	Restaurant	Pu Re
10	1	Mungivacca	Bari	Bari,Mungivacca	41.090225	16.900127	3.0	Scandinavian Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Restaurant	Puglia Restaurant	M Eu Re
20	4	Catino	Bari	Bari,Catino	41.163358	16.746813	4.0	Seafood Restaurant	Restaurant	Vegetarian / Vegan Restaurant	Scandinavian Restaurant	Puglia Restaurant	Modern European Restaurant	M Re

Methodolgy

Once selected only restaurant venues I decided to plot a histogram that could highlight the frequency of every restaurant type across all the neighbourhoods

I used k-means clustering in order to group neighbours by restaurant types and numbers and then I used folium package to plot those clusters.

Once clusters have been created I analyzed each of the identified groups in order to highlight clusters characteristics.

```
In [52]: bari_merged_rest.loc[bari_merged_rest['Cluster Labels'] == 0, bari_merged_rest.columns[[1] + list(range(7, bari_merged_rest.shape[1]))]]
```

Out[52]:

	Quarters	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
0	Murat	Italian Restaurant	Restaurant	Seafood Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Puglia Restaurant	Modern European Restaurant	Mediterranean Restaurant	Scandinavian Restaurant	Japanese Restaurant
1	San Nicola	Italian Restaurant	Seafood Restaurant	Restaurant	Mediterranean Restaurant	Japanese Restaurant	Greek Restaurant	Vegetarian / Vegan Restaurant	Scandinavian Restaurant	Puglia Restaurant	Modern European Restaurant
5	Torre a mare	Italian Restaurant	Seafood Restaurant	Vegetarian / Vegan Restaurant	Scandinavian Restaurant	Restaurant	Puglia Restaurant	Modern European Restaurant	Mediterranean Restaurant	Japanese Restaurant	Greek Restaurant

Discussion

After cleaning and parsing data about Restaurant clustering around the city of Bari we can see that there are few neighbours that contain the majority of restaurant of the entire area. The following table highlights the total number of restaurants per neighbours.

```
In [63]:  bari_Venues_restaurant

Out[63]:  Neighborhood
          Carrassi          1
          Catino            2
          Mungivacca        2
          Murat            15
          Poggiofranco      2
          San Nicola       14
          San Pasquale      3
          Torre a mare      5
          Name: Venue Category, dtype: int64
```

Top 3

We can state that "Murat" and "San Nicola" are the top two neighbours if we consider the number of operating restaurants. This is consistent with the fact that these are two central neighbours located in the turistic area of the city.

The third neighbours in absolute number of restaurant is "Torre a Mare" which is a small centre near Bari that is famous for its seafood restaurant which are characterized by their view. All of the restaurant are located by the sea so it is very popular if you wanna experience a meal in front of the blue adriatic sea.

These three neighbours represent cluster 0 and of course they are the best choice if an investor want to open a new activity. The centrality of these with respect to the city and their touristic attraction that characterize these borough are the two major reason that attract someone to invest on.

Top 4th and 5th

The cluster number 1 respresent a very peculiar group. This is composed by "Poggiofranco" and "San Pasquale" that are the Univeristy neighbours. These represent good neighbours marked by young people in which one can experience vegetarian, vegan and international food.

Cluster n.3

Cluster n 3 represents a very peculiar one. It is located at city borders and it's characterized by fast food located on the highway.

Conclusion

Considering the conclusion, if I am and investor and I want to open an international food restaurant (e.g Mexican or Thai etc.) the first thing that I have to ask myself is if I want to open a fast food or a restaurant. If I choose fast food option I can state that Mungivacca is the best neighbours in which invest on. If I want to open a restaurant in a very populated, touristic and easily accessable neighbours I can look at cluster one and in particular "Murat" and "San Nicola" neighbours, located at city center.