The Battle of Neighborhoods

IBM Data Science Professional Certificate – Capstone Project

Opening a new ice cream shop in Rome, Italy

Chiara Picardi, December 2020

Introduction

Background

Rome is the capital city and a special comune of Italy as well as the capital of the Lazio region. With 2,860,009 residents in 1,285 km^2 (496.1 sq mi), it is the country's most populated comune and the third most populous city in the European Union by population within city limits. In 2019, Rome was the 11th most visited city in the world with 10.1 million tourists, the third most visited in the European Union, and the most popular tourist destination in Italy and its historic centre is listed by UNESCO as a World Heritage Site.

Business problem

A client famous for his high quality ice cream (gelato) wants to open a new ice cream shop in Rome. He is not sure in which neighborhood is better to open the shop due to the high competition. Italian gelato is very well known and it is easy to imagine that in a city like Rome the ice cream shops are very popular. In this project the neighborhoods of Rome will be examined in order to determine which are the neighborhoods with a low presence of ice cream shops and then less competition.

Target audience

This project is useful for investors interested in opening an ice cream shops in Rome.

Data

The list of neighborhood in Rome is available on wikipedia: https://en.wikipedia.org/wiki/Category:Quarters_of_Rome. The wikipedia page is used to obtain the relevant information, using web scraping techniques, Python requests and BeautifulSoup packages. The list of neighborhood in Rome is converted in a dataframe. The first 5 samples are reported in figure 1.

	Neighborhood
0	► Rome Q. XXIII Alessandrino (7 P)
1	► Rome Q. XXV Appio Claudio (1 C, 4 P)
2	► Rome Q. IX Appio-Latino (18 P)
3	► Rome Q. XXVI Appio-Pignatelli (2 P)
4	► Rome Q. XX Ardeatino (13 P)

Figure 1: Data retrieved from wikipedia

Methodology

Once obtained the original data frame, in order to solve the business problem we need:

- Longitude and Latitude for each neighborhood
- Ice cream shops venue data
- Different cluster of neighborhoods characterizing the frequency of ice cream shops

Geocoder package

The longitude and the latitude for each neighborhood is obtained using the Python Geocoder package and added to the original dataframe. The first five samples of the new dataframe are reported in figure 2.

	Neighborhood	Latitude	Longitude
0	► Rome Q. XXIII Alessandrino (7 P)	41.87139	12.57974
1	► Rome Q. XXV Appio Claudio (1 C, 4 P)	41.98425	12.71421
2	► Rome Q. IX Appio-Latino (18 P)	41.87461	12.51333
3	► Rome Q. XXVI Appio-Pignatelli (2 P)	41.84326	12.54058
4	► Rome Q. XX Ardeatino (13 P)	41.53654	12.56257

Figure 2: Dataset creating using Geocoder

Foursquare API

Foursquare API is used to retrieve information of existing ice cream shops in Rome. The frequency of ice cream shops for the first five neighborhood in Rome is reported in figure 3:

	Neighborhoods	Ice Cream Shop
0	► Rome Q. I Flaminio (18 P)	0.10
1	► Rome Q. II Parioli (16 P)	0.05
2	► Rome Q. III Pinciano (3 C, 20 P)	0.07
3	► Rome Q. IV Salario (5 P)	0.08
4	► Rome Q. IX Appio-Latino (18 P)	0.09

Figure 3: Dataset created using Foursquare API

K-means Clustering

The machine learning algorithm called K-means Clustering is used to perform a segmentation in the dataset. K-means Clustering is a simple and popular unsupervised algorithms that can be used to makes segmentations.

Segmentation is a practice of divide a feature into groups with similar characteristics. Therefore one can get some insights about the characteristics of the data. The dataset is segmented using the frequency of ice cream shops into four different classes. In figure 4,5,6 and 7 neighborhoods belonging to each class and the ice cream shops frequency are reported.

	Neighborhood	Ice Cream Shop	Category	Latitude	Longitude
29	► Rome Q. XXX San Basilio (3 P)	0.1	0	41.90454	12.48909
6	► Rome Q. VI Tiburtino (1 C, 10 P)	0.1	0	41.89699	12.51100
0	► Rome Q. I Flaminio (18 P)	0.1	0	41.92093	12.47104

Figure 4: Neighborhoods in Category 0

	Neighborhood	Ice Cream Shop	Category	Latitude	Longitude
27	► Rome Q. XXVII Primavalle (10 P)	0.021739	1	41.92044	12.413030
17	► Rome Q. XVII Trieste (15 P)	0.000000	1	41.61325	12.506000
31	► Rome Q. XXXII Europa (22 P)	0.000000	1	41.72399	13.008330
28	► Rome Q. XXVIII Monte Sacro Alto (3 P)	0.012821	1	41.94691	12.550420
30	► Rome Q. XXXI Giuliano-Dalmata (6 P)	0.000000	1	41.81505	12.504710
25	► Rome Q. XXV Appio Claudio (1 C, 4 P)	0.000000	1	41.98425	12.714210
26	► Rome Q. XXVI Appio-Pignatelli (2 P)	0.000000	1	41.84326	12.540580
34	► Rome Q. XXXV Lido di Castel Fusano (1 C, 2 P)	0.000000	1	41.73477	12.315330
24	► Rome Q. XXIX Ponte Mammolo (4 P)	0.000000	1	41.93480	12.569220
23	► Rome Q. XXIV Don Bosco (14 P)	0.000000	1	41.70693	12.684590
19	► Rome Q. XX Ardeatino (13 P)	0.000000	1	41.53654	12.562570
13	► Rome Q. XIV Trionfale (12 P)	0.000000	1	41.98198	12.408469
12	► Rome Q. XIII Aurelio (9 P)	0.000000	1	41.98844	12.717120
8	► Rome Q. VIII Tuscolano (15 P)	0.000000	1	41.74937	12.856980

Figure 5: Neighborhoods in Category 1

	Neighborhood	Ice Cream Shop	Category	Latitude	Longitude
22	► Rome Q. XXIII Alessandrino (7 P)	0.043478	2	41.871390	12.579740
16	► Rome Q. XVI Monte Sacro (8 P)	0.060000	2	41.933192	12.533585
14	► Rome Q. XIX Prenestino-Centocelle (7 P)	0.047619	2	41.884120	12.566190
21	► Rome Q. XXII Collatino (4 P)	0.030000	2	41.907430	12.550710
20	► Rome Q. XXI Pietralata (10 P)	0.040000	2	41.920490	12.546500
32	► Rome Q. XXXIII Lido di Ostia Ponente (1 C)	0.045977	2	41.732350	12.276390
7	► Rome Q. VII Prenestino-Labicano (16 P)	0.060000	2	41.885700	12.535210
33	► Rome Q. XXXIV Lido di Ostia Levante (1 C, ...	0.045977	2	41.732350	12.276390
2	► Rome Q. III Pinciano (3 C, 20 P)	0.060000	2	41.918860	12.484100

Figure 6: Neighborhoods in Category 2

	Neighborhood	Ice Cream Shop	Category	Latitude	Longitude
10	► Rome Q. XI Portuense (18 P)	0.080000	3	41.85283	12.45724
18	► Rome Q. XVIII Tor di Quinto (6 P)	0.070000	3	41.94000	12.48039
3	► Rome Q. IV Salario (5 P)	0.080000	3	41.91331	12.50217
4	► Rome Q. IX Appio-Latino (20 P)	0.080000	3	41.87461	12.51333
5	► Rome Q. V Nomentano (15 P)	0.080000	3	41.91017	12.50223
1	► Rome Q. II Parioli (18 P)	0.070000	3	41.92422	12.49075
15	► Rome Q. XV Della Vittoria (26 P)	0.076923	3	41.45669	12.65157
9	► Rome Q. X Ostiense (23 P)	0.080000	3	41.87451	12.48077
11	► Rome Q. XII Gianicolense (15 P)	0.080000	3	41.87294	12.46486

Figure 7: Neighborhoods in Category 3

Results

Figure 8 shows the different clusters on the map. The location of the different categories can be easily found reading the map and the respective legend reported in the caption. For each category the mean frequency of ice cream shops is computed and stored in an array. Using this mean a bar chart is produced representing the mean frequency of ice cream shops

corresponding to each category. The resulted bar chart is reported in figure 9. Neighborhoods in category 0 have already lots of ice cream shops, while ones in category 1 have very few ice cream shops. Neighborhoods in category 2 have medium presence of ice cream shops while in category 3 there is a moderate presence of ice cream shops.

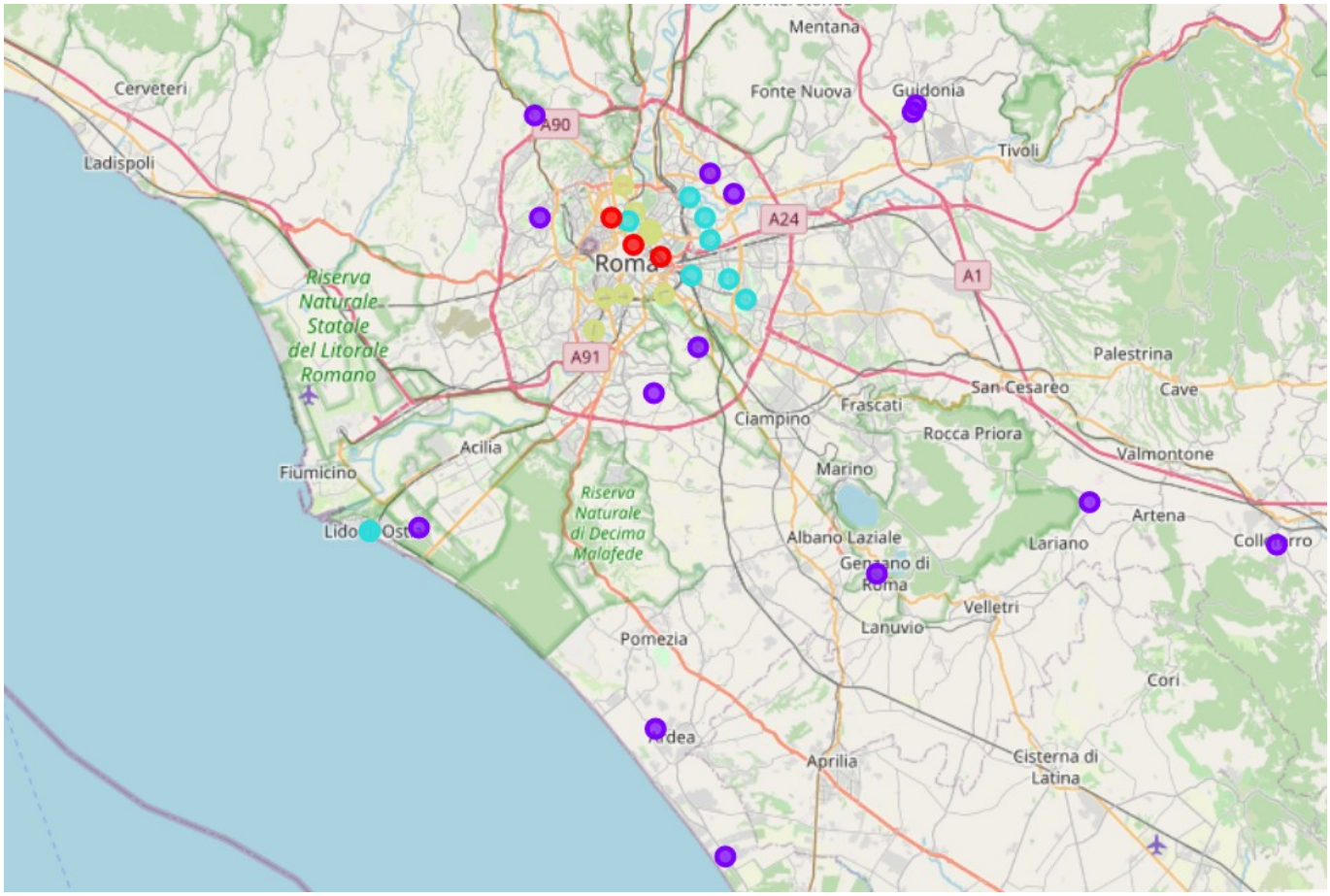

Figure 8: Categories represented on the map. Category 0 is red, Category 1 is in purple, Category 2 is in light blue and Category 3 is in light green.

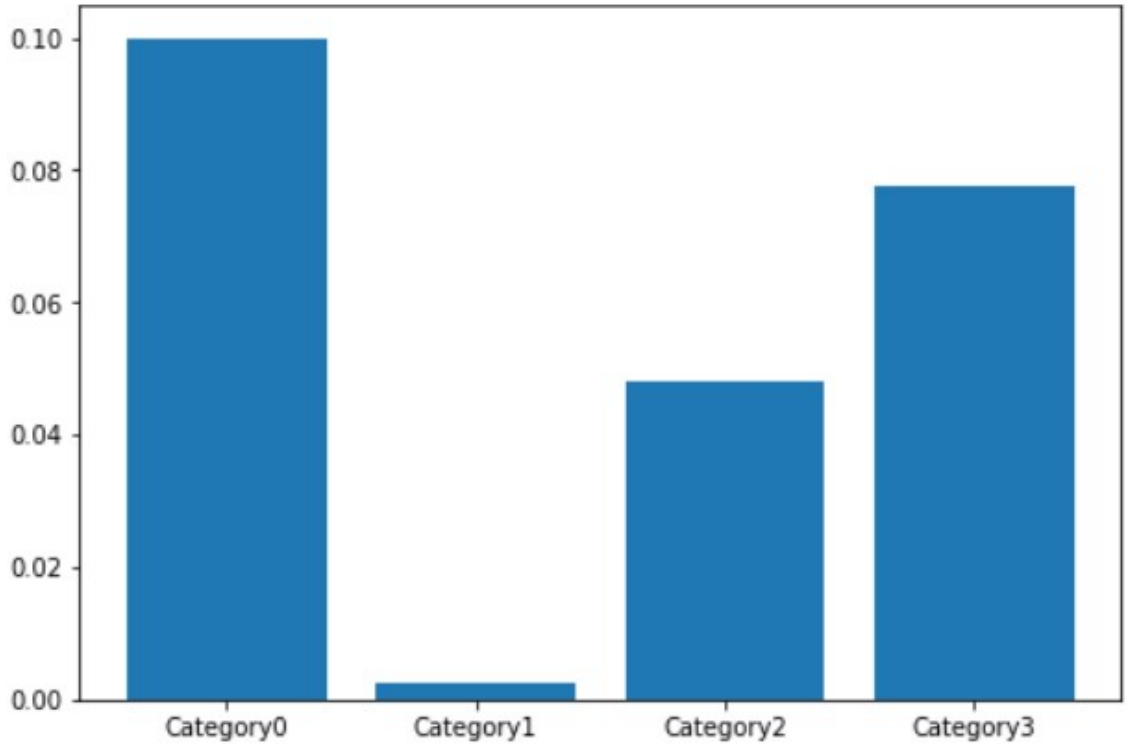

Figure 9: Mean frequency of ice cream shops is reported for each category

Discussion

From the results I can suggest that open a new ice cream shop in a neighborhood belonging to category 0 is not probably a good idea considering the high presence of these kind of shops. In neighborhoods of category 1 there is a very low competition but it is probably a risk considering that the population could not be interested in ice cream shops. Neighborhoods belonging to category 2 are certaintly a good choice for a new ice cream shop because the copetition is not very high and the population is already familiar with ice cream shops. Neighborhoods in category 3 can also be considered but the competition will be higher respect to the ones in the previous category. Taking in account all these factors probably the best choice for opening a new ice cream shop will be any neighborhood belonging to category 2.

There are some limitations in the study proposed. Cost of renting and buying in different neighborhoods is not considered and also presence of interesting landmarks in the areas are not taked in account. Considering these factors can be very important to complete this preliminary study and solve the business problem in a better way.

Conclusion

In this data science project, I showed how to explore venues using Foursquare API and how to get latitudes and longitudes using Geocoder. I chose the ice cream shop category to explore Foursquare venues in the city of Rome. I applied the Machine Learning algorithm K-means Clustering and I made segmentations of the types of ice cream shops. Therefore, It was possible to observe in the ‘Folium map‘ the locations of the shops in each of the clusters created. In the end, I examined the results determing the best places to open an ice cream shops in Rome.