

# Finding the right location for a new restaurant

---

Roberto Puglisi

11/2020

## 1. Introduction

Our Client, one of the most famous and awarded chefs in Italy, is considering starting a new business in Sicily. More specifically, his business idea is to open a new restaurant in one of the top summer locations on the Sicilian coasts.

The main requirements are:

- The restaurant should be close (<1km range) to one of the top 50 beaches in Sicily (as a driver for the most touristic places).
- The location should be determined considering the density of other restaurants in the area, excluding indirect competitors such as fast food, wine bars, pubs, beach bars, bars, café, and other not direct competitors.
- Locations in minor islands are considered in scope.

The Client is asking the support of a data-driven analysis in order to make the final choice for the location.

## 2. Data sources

Data sources used in the analysis:

- Public data related to famous beaches and summer locations in Sicily, they will be scraped from the webpage: <https://www.marcopolo.tv/sicilia-spiagge>
- Geo-Coordinates: retrieved using the Python library “geopy”.
- Public data related to venues: retrieved using Foursquare API.

After the data collection stage, data cleaning will be needed in order to:

- Managing missing and incorrect information (e.g. location names).
- Managing duplicate values
- Other data cleaning operations

Feature engineering will be needed for statistical analysis to:

- group in the same class types of venues with similar category name, and
- create a residual class needed for statistical analysis.

### 3. Methodology

The following list of locations was collected from the above mentioned public source. It contains the top 50 summer locations in Sicily, which were requested by the Client as a minimum requirement.

1. Mondello, Palermo
2. Isola delle Femmine
3. Cala Capreria, Riserva Zingaro
4. La Praiola, Terrasini
5. Magaggiari, Cinisi
6. Cefalù
7. Pollina Torre Conca, Palermo
8. Punta Spalmatore, Ustica
9. Cala Junco, Panarea
10. Pollara, Salina
11. Cala Croce, Lampedusa
12. Tindari Marinello, Messina
13. Vendicari, Noto
14. Baia di SantAntonio, Milazzo
15. Lipari, Lipari
16. Isola Bella, Taormina
17. Santa Tecla, Acireale
18. San Marco, Calatabiano
19. Caldura
20. Foce del Simeto, Catania
21. Marina di Cottone, Catania
22. Calamosche, Siracusa
23. Fontane Bianche, Siracusa
24. San Lorenzo, Siracusa
25. Eloro Pizzuta, Siracusa
26. Santa Maria del Focallo, Ragusa
27. Marina di Modica
28. Marina di Ragusa
29. Playa Grande, Ragusa
30. Punta Secca, Ragusa
31. Randello, Ragusa
32. Pozzallo, Ragusa
33. Punta Braccetto, Ragusa
34. Porto Palo di Menfi, Agrigento
35. Seccagrande, Ribera
36. Scala dei Turchi, Agrigento
37. Isola dei Conigli, Lampedusa
38. Cala Pozzolana, Linosa
39. Sciacca, Sciacca
40. Playa Carratois, Portopalo di Capopassero
41. Spiaggia Isola di Capo Passero
42. Cala Vincenzina, Marina di Palma
43. San Vito lo Capo
44. Scopello, Trapani

45. Scala dei Turchi, Agrigento
46. Porta Bottegghelle, Trapani
47. Cala Rossa, Favignana
48. Cala Levante, Pantelleria
49. Cala Capreria, Trapani
50. Castellammare del Golfo, Trapani

The map below shows the geolocation of each item in the list:

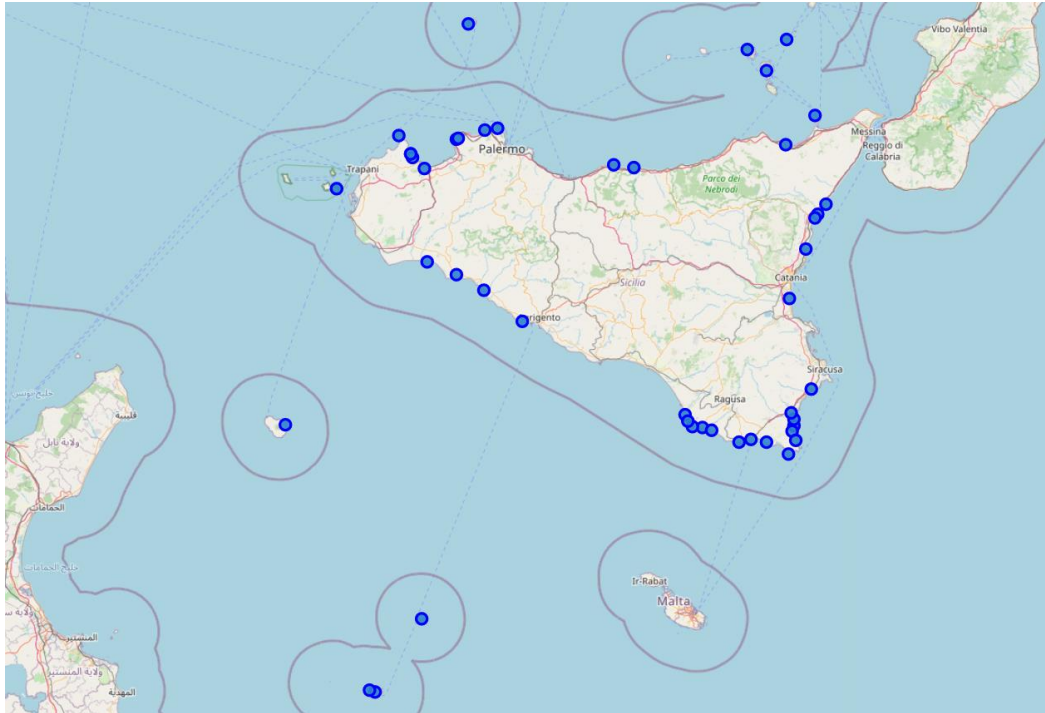


Figure 1 Geo-location

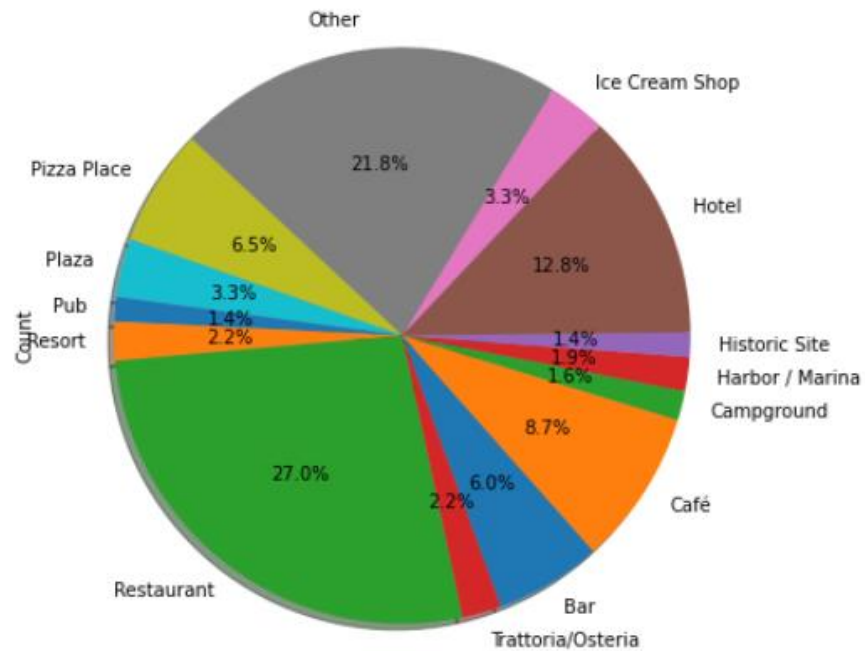
Using the Foursquare API we collected data about venues located around each point on the map into a 1km range.

The list below represents the Venue Categories, after grouping categories with similar names in the same class.

Excluding the residual category “Other”, the first 10 most popular venue categories account for about 62 % of the total.

**Restaurants account for about 27% of the total venues.**

Type	Count
Bar	22
Café	32
Campground	6
Harbor / Marina	7
Historic Site	5
Hotel	47
Ice Cream Shop	12
Other	80
Pizza Place	24
Plaza	12
Pub	5
Resort	8
<b>Restaurant</b>	<b>99</b>
Trattoria/Osteria	8
<b>TOTAL</b>	<b>367</b>



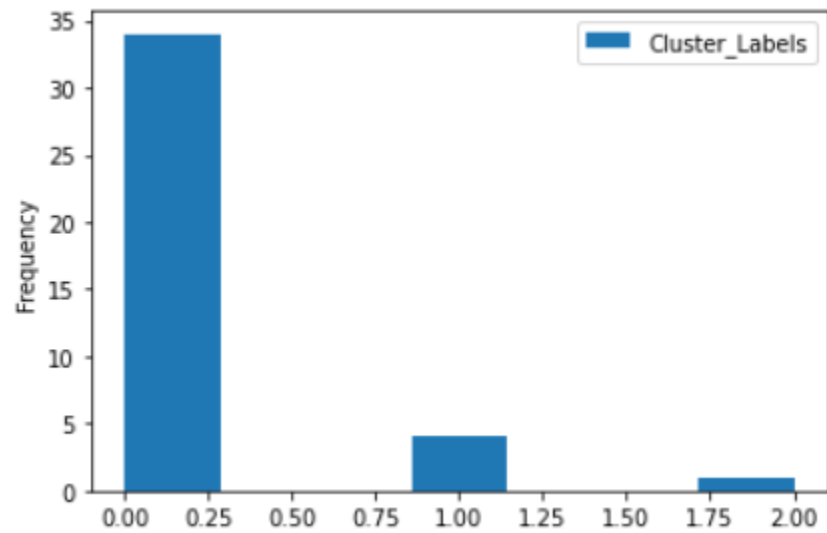
For each location, an analysis has been conducted to understand the top 10 most popular venues.

This was achieved by applying the “one hot encoding technique”, which allows to encode categorical features as a one-hot numeric array. After that each feature “venue1, venue, venue n” has been ranked.

Finally, locations have been divided into **3 clusters** through the implementation of an **unsupervised machine learning technique**. For this purpose, the **K-Means** algorithm has been applied, which is specifically used for Clustering.

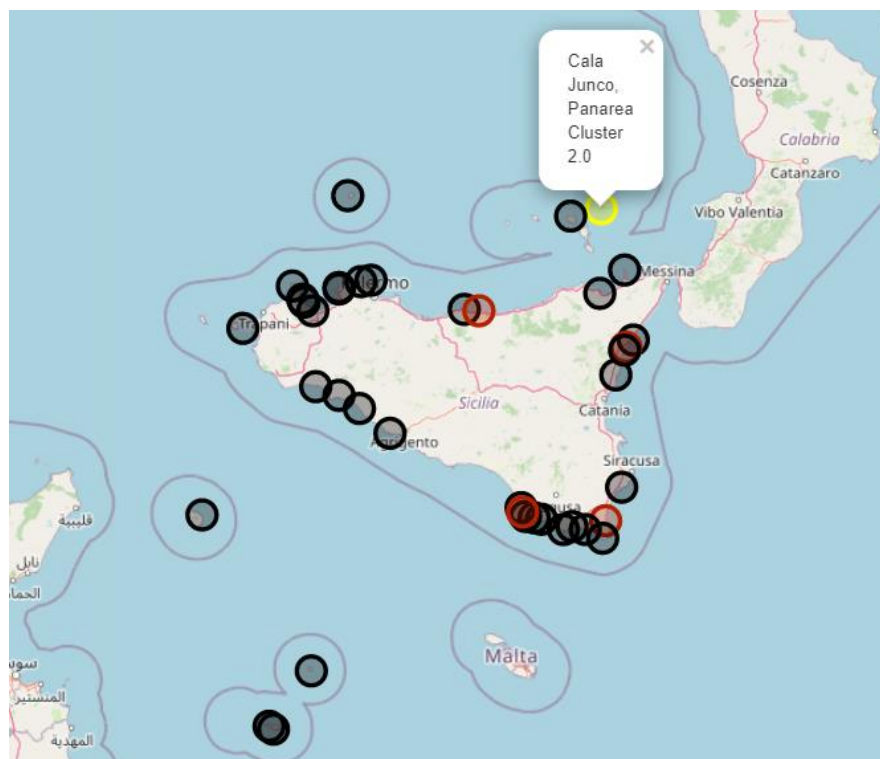
## 4. Results

Locations have been divided into 3 clusters as reported below:



Graphically, the 3 clusters are highlighted below with colors:

- Cluster 0: **Black**
- Cluster 1: **Red**
- Cluster 2: **Yellow**



## 5. Discussion

The algorithm calculated clusters based on qualitative data only (10 top popular venues for each location). Our interpretation of the results suggests that:

- Locations included in Cluster 0 (**Black**), are popular locations and very well known by tourists. Services, including restaurants, are present and developed. Opening a new restaurant in these locations would mean entering in a competitive environment. Therefore, this choice for a new restaurant should be carefully reviewed and motivated by further analysis. We may define this cluster as “**Mass tourism**”.
- Locations included in Cluster 1 (**Red**), are popular locations known by the majority of tourists. Not every service is offered in this area, as they are mainly organized for camping and short stays. Food & Beverage businesses are also limited. Therefore a premium offer of a world class restaurant in these locations may not be the right move. We may define this cluster as “**Outdoor tourism**”.
- The only one location included in Cluster 3 (**Yellow**), is popular but at the same does not attract as many tourists as locations in the other two clusters, mainly due to:
  - The place located on one of the minor Sicilian islands (Panarea), which is only reachable by boat
  - Panarea is an exclusive location, costs of hotels and restaurants are on average much higher than other locations in scope of this analysis.

We may define this cluster as “**Luxury tourism**”.

## 6. Conclusions

Given that our client is a top class chef, our analysis suggests that a proper location for opening a new restaurant may be Panarea, included in **cluster 2 (Luxury tourism)**.

Data show that no restaurants are operating in the area, except for low quality restaurants (Trattoria) and some local fast food.

Place	1st most popular venue type	2nd most popular venue type	3rd most popular venue type	4th most popular venue type	5th most popular venue type	6th most popular venue type	7th most popular venue type	8th most popular venue type	9th most popular venue type	10th most popular venue type
Cala Junco, Panarea	Bay	Trattoria/Osteria	Coffee Shop	Gastropub	Furniture / Home Store	Fried Chicken Joint	Food Truck	Food Stand	Food Court	Food & Drink Shop

In modern times, Panarea has become a fashionable vacation spot and in 2011, it was described by W magazine as "the epicenter of the chicest summer scene in the Mediterranean."

However, given the strict regulations in the area, getting authorization for a new business may require a significant financial effort.

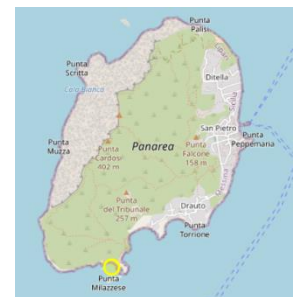






Figure 2 - View of Panarea