

Clustering sea cities of Spain

Alla Sidko

February 26, 2020

1. Introduction

1.1 Background

Every year or a few times a year almost all people have the same problem and the same dream. All the life people wait for vacation for rest and for relax. And the best way to relax is to go to the sea. All of us, at least once in a lifetime, dream about lying on the beautiful beach under the sun. Thus, every time we decide to go to sea, we have a problem with the choice of country and city.

As a rule, we want to go to a new place, to watch and try something new, but we don't know where our rest would be the best and would not cost so expensive. If you decide to travel to the Europe, one of the most popular sea country is Spain. Summer resorts and beaches were the first to be developed in Spain, and today, generate the most income for the Spanish economy. The mild climate during the whole year and the extensive sandy beaches of the Mediterranean and Atlantic Ocean as well as of its two archipelagoes have been attracting tourists from all of the world.

The most popular Spanish mainland coasts are on its Mediterranean side. There are two most visitable beach sides - Costa Brava and Costa Daurada, which are not so far from Barcelona. But for this work I want to explore Costa Brava. I'll get 11 popular towns in this area and provide clustering of each of these towns to specify which town is the most suitable for different people and their needs.

1.2 Problem

There are a lot of websites and videos with information about these towns, but often it is only general description without information about accommodation, stores, beaches and other facilities. And there is a problem to choose the most appropriate town, that can satisfy your

needs and possibilities. This project aims to describe 11 similar sea towns in Spain and which town would be suitable for different needs.

1.3 Interest

This information would be useful for people, who decide to go to the sea and can not choose a city. I've chosen such groups of people: family with children, couples, aged people, friends, people who want calm vacation or vice versa. I'm going to investigate which city is the most suitable for each group.

2. Data

2.1 Data sources

To solve this problem I will use the Foursquare location data. Foursquare is the most trusted, independent location data platform for understanding how people move through the real world. With Places Database, I can access precise, up-to-date community-sourced venue data. To gain my purpose I need data for "Palamos", "Calella de Palafrugell", "Llafranc", "Tamariu", "Pals", "Begur", "Cadaques", "Lloret de Mar", "Tossa de Mar", "Blanes", "l'Escala". In every city I will explore their hotels, apartments, restaurants, cafes, groceries, markets, beaches, bus stations, train station, parks, night life, playgrounds. I separately got all categories that I need for my exploration and joined into one table. So I choose such categories:

1. Hotel
2. Food
3. Food&Drink Shop
4. Bus Stop
5. Bus Station
6. Train Station
7. Nightlife
8. Outdoors&Recreation
9. Arts&Entertainment

2.2 Data Selection

I got data about venues for each of these cities in radius 2000 m. Then I stored it into dataframe and have such columns: *'id', 'name', 'categories', 'referralId', 'hasPerk', 'location.address', 'location.lat', 'location.lng', 'location.labeledLatLngs', 'location.distance', 'location.postalCode', 'location.cc', 'location.city', 'location.state', 'location.country', 'location.formattedAddress', 'location.crossStreet', 'venuePage.id'*. Column *categories* is in JSON format, so I select only name and filter the category for each row. After that I keep columns that include venue name, and anything that is associated with location and clean column names by keeping only last term. I suppose, that for my project aim I don't need columns as *"postalCode", "cc", "crossStreet", "state", "country", "labeledLatLngs", "distance", "address"*, so I decided to delete them. Also I added one column with names of main categories. At the end I have this list of columns:

Column Name	Type
<i>name</i>	<i>object</i>
<i>categories</i>	<i>object</i>
<i>lat</i>	<i>float64</i>
<i>lng</i>	<i>float64</i>
<i>city</i>	<i>object</i>
<i>formatted Address</i>	<i>object</i>
<i>main_categories</i>	<i>object</i>
<i>id</i>	<i>object</i>

2.3 Data cleaning

Some rows can have *None* value in column “categories”, so I have to delete them first. Because for my exploration it is the main field and it can not be *None*. Column “city” has *NuN* values

and can have incorrect names of city, so I replaced all *NaN* value in the column 'city' by city name and replace all incorrect name of city by actual name.

Also, some data have *NaN* value in another columns, but it don't have influence on my research and I can leave such rows in tables. The last, I deleted all rows with category 'City', because it contains only name of city that I explore, so I don't need them.

As a result, I have 11 tables with venues of selected categories for each of towns. After data cleaning and data preparation I have such amount data:

- Palamos - 64 rows \times 8 columns
- Calella de Palafrugell - 17 rows \times 8 columns
- Llafranc - 29 rows \times 8 columns
- Tamariu - 22 rows \times 8 columns
- Pals - 16 rows \times 8 columns
- Begur - 44 rows \times 8 columns
- Cadaques - 45 rows \times 8 columns
- Lloret de Mar - 89 rows \times 8 columns
- Tossa de Mar - 61 rows \times 8 columns
- Blanes - 85 rows \times 8 columns
- l'Escala - 102 rows \times 8 columns

And for analysys I concat these tables with all of venues of these towns to one dataframe. In all 574 rows and 8 columns.

3. Methodology

In this project we will direct our efforts on exploration sea cities of Costa Brava side, particularly specify where is better to travel. We will limit our analysis to area ~2km around city center. In first step when we have all the required data we make data preparation for next analysis.

Second step in our analysis will be clustering of our 11 sea towns of Spain to define groups by specified categories.

In third and final step we will focus on clusters, where there are the most venues in our categories. Then we will explore towns separately in chosen clusters and compare them to make decision and conclusion about the best town for vacation.

4. Analysis

I discovered that Begur and Pals are not located near the sea, so we can not explore them.

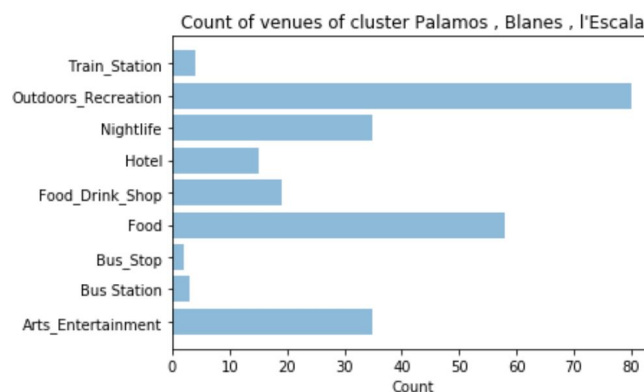
4.1 Data Preparation

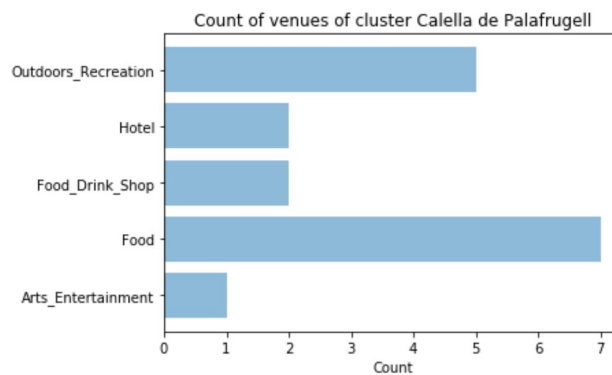
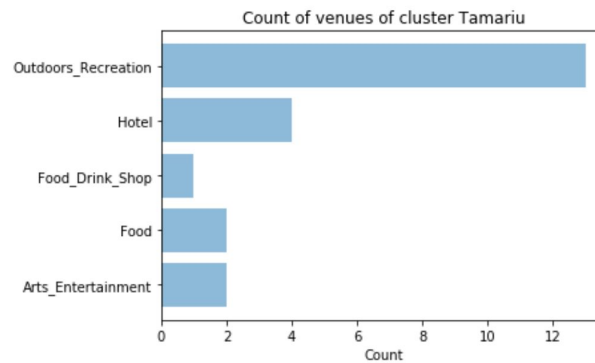
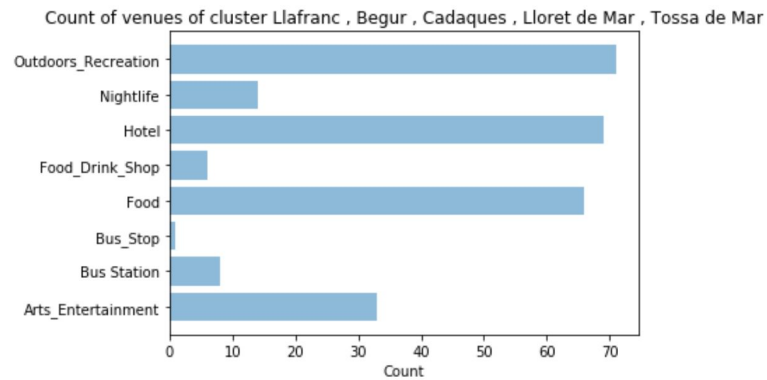
Before clustering let's make data preparation. First we make dummies for column "categories" and add column "city" and move it to the first column. As a result we have table with 574 rows and 139 columns. Next I grouped table by cities and get mean value.

4.2 K-means Clustering

For my exploration I used such machine learning algorithm as K-means clustering to have some groups of chosen towns. I have chosen number of clusters - 5. Dropped column "city" for fitting clustering. Run clustering with random state = 0 and checked cluster labels generated for each row in the dataframe. After that I added received clustering labels to venues in general dataframe.

As a result, I have such clusters and bar plots for our five clusters with amount of specific main categories.





Tree towns: Pals, Tamariu, and Calella de Palafrugell are clusters, that have very little amount of venues comparing with first two clusters. So let's explore only cities that first and second cluster contain.

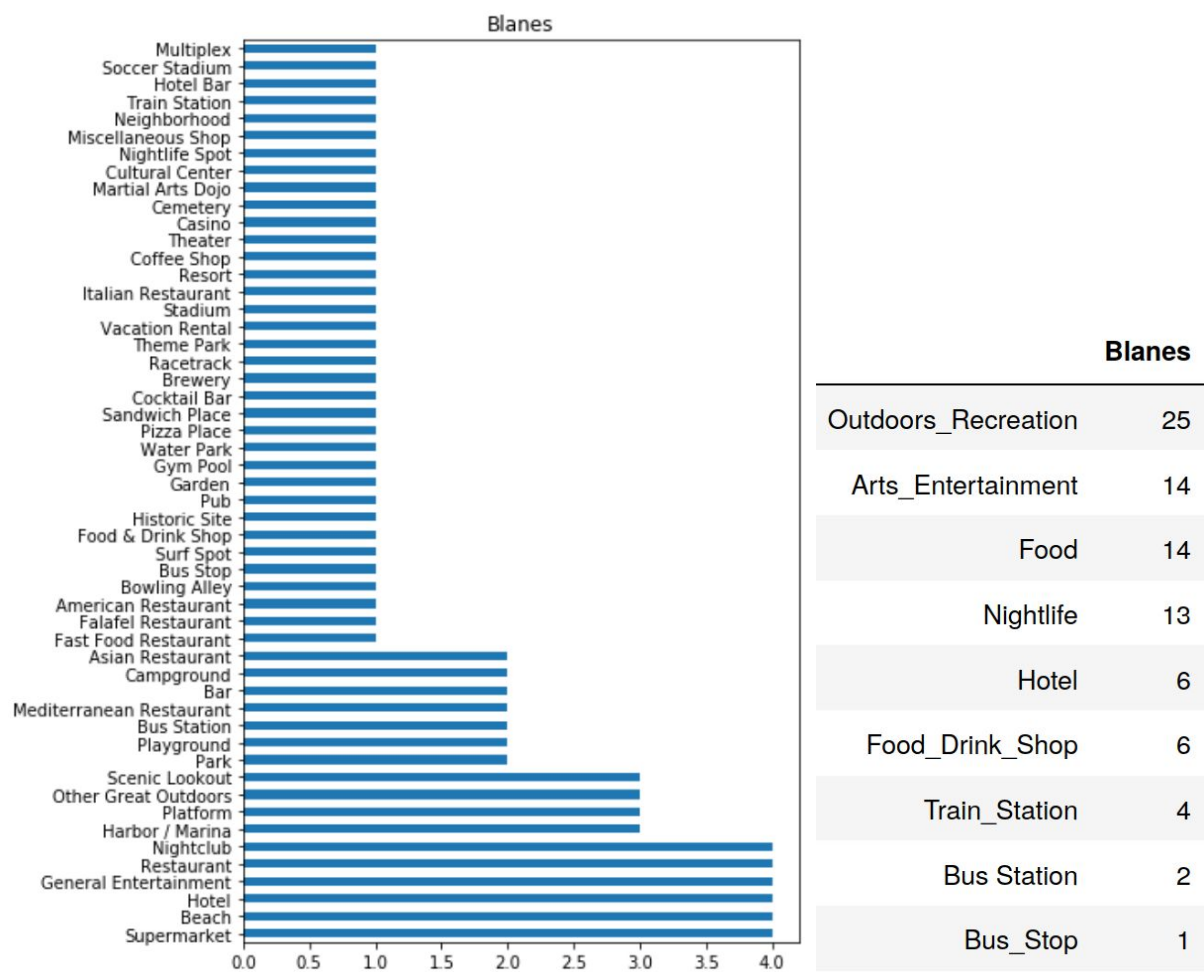
First cluster

I got first cluster with Palamos, Blanes, and l'Escala and plot bars with main categories for each city separately. In all three cities we have almost the same number of Recreations and Outdoors approximately **25**.

About 15 nightlife. In Palamos and Blanes 15 food places and in l'Escala 30 such places. There are not so much hotels - around 5 in each.

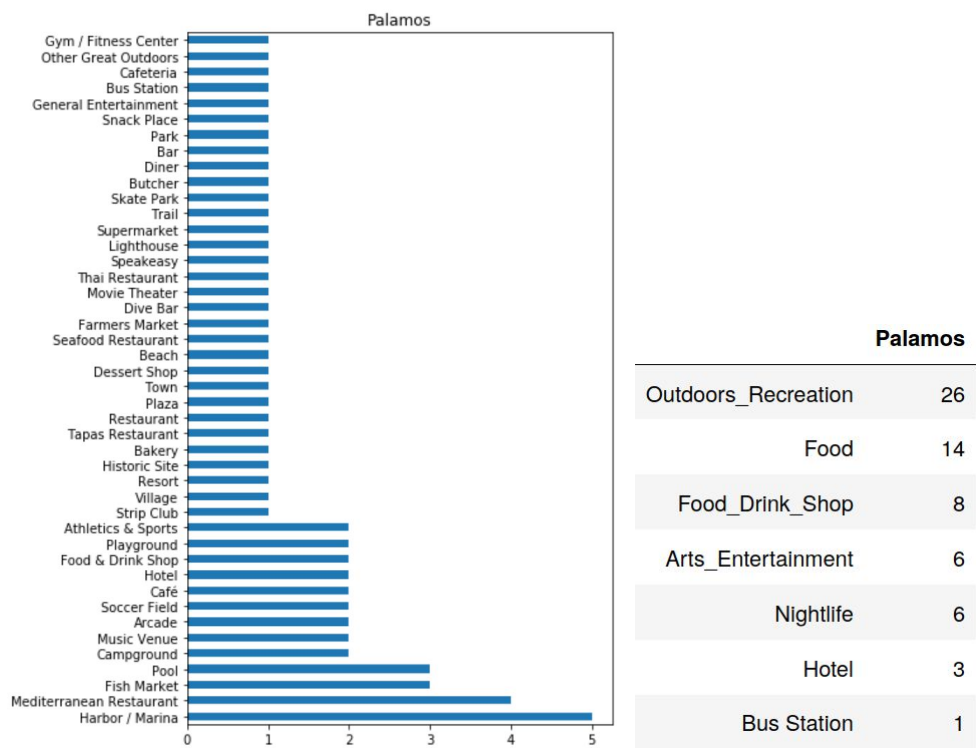
Blanes

Only in Blanes are 5 Train stations. Blanes has a lot of different venues, activities. But, in this town there are the most supermarkets, beaches, hotels, restaurants, night clubs and other entertainments. So this city will be suitable for people who love active rest, night life. It can be good for friends and couples. Also for family with children, because there are some playgrounds.



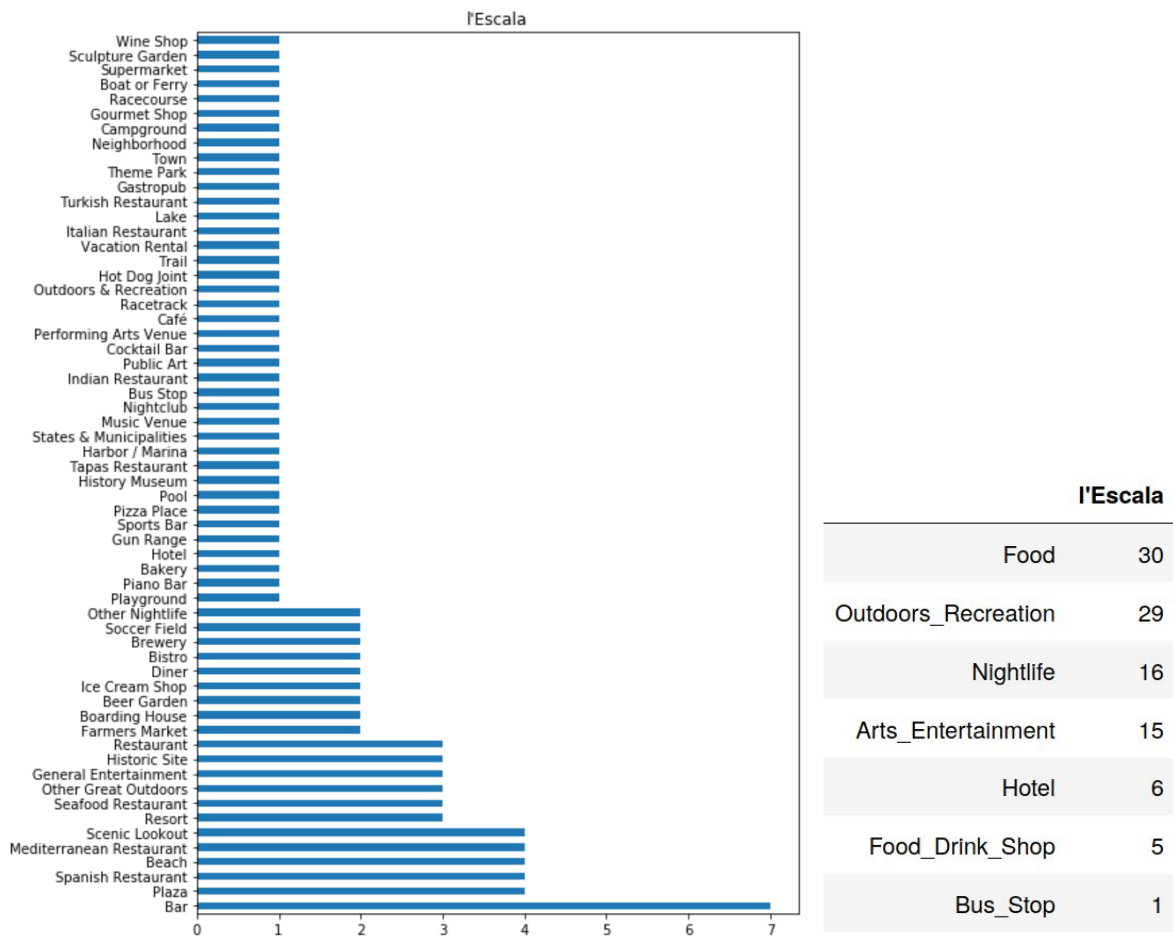
Palamos

Palamos is reach by variety of outdoors and recreations. There are not a lot of hotels and nightlife. In this town are harbours, beaches, different shops, cafes, restaurants, playgrounds. So, such town is suitable for families with children and for old-aged people.



l'Escala

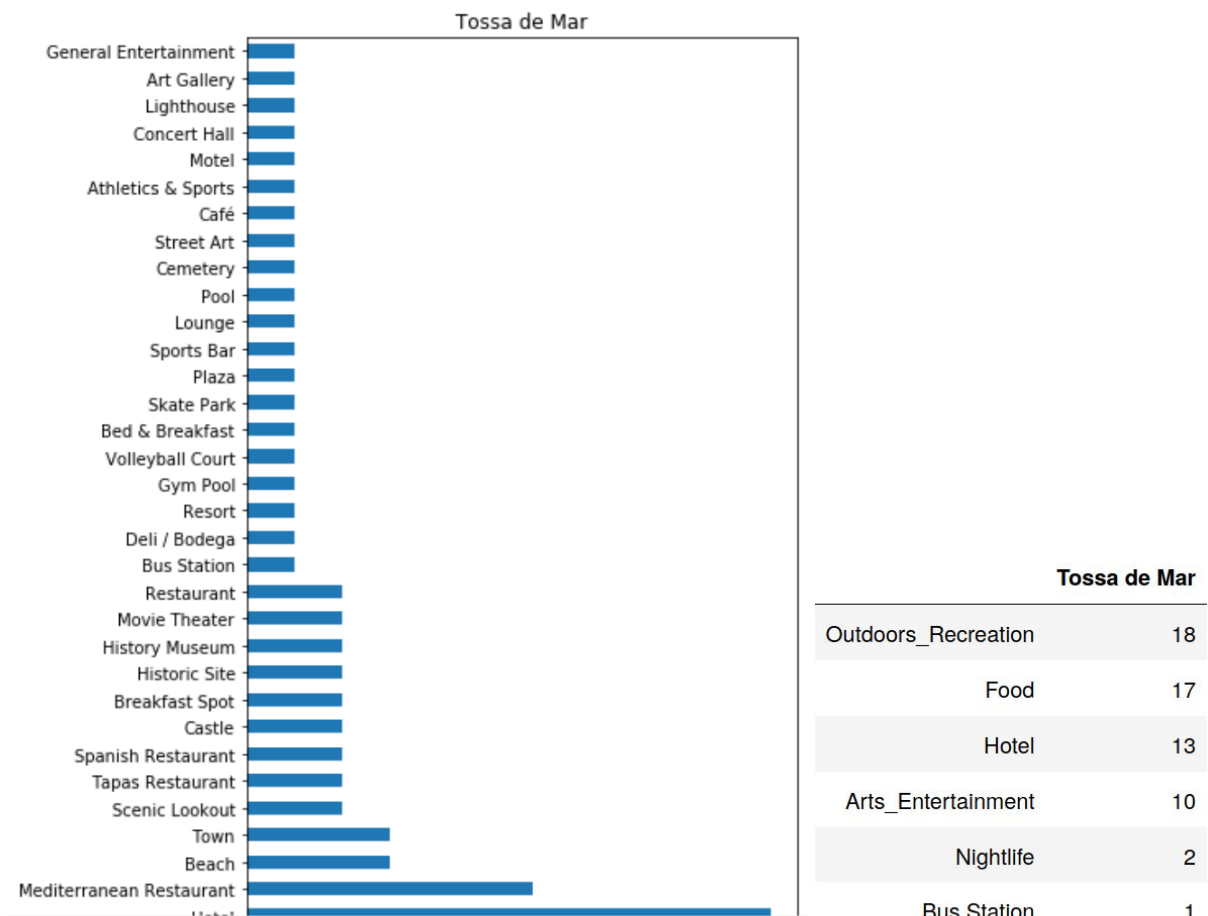
In l'Escala we can find variety of food places and outdoors&recreations. There is not so big of choice of hotels. There are a lot of bars, restaurants, beaches, cafes, historic sites. So we can say, that l'Escala is gastronomic town, and we can find what to do there. I guess, it will be good for young people, for teams, friends.



Second cluster

I got second cluster with Llafranc, Begur, Cadaques, Lloret de Mar, and Tossa de Mar and plot bars with main categories for each city separately. There are not a lot of venues in Llafranc and Begur, so we can stop explore these two towns. Cadaques has foods, outdoors recreation and hotels, but Lloret de Mar has much more hotels(around 30). In Tossa de Mar there are a lot of outdoors recreations.

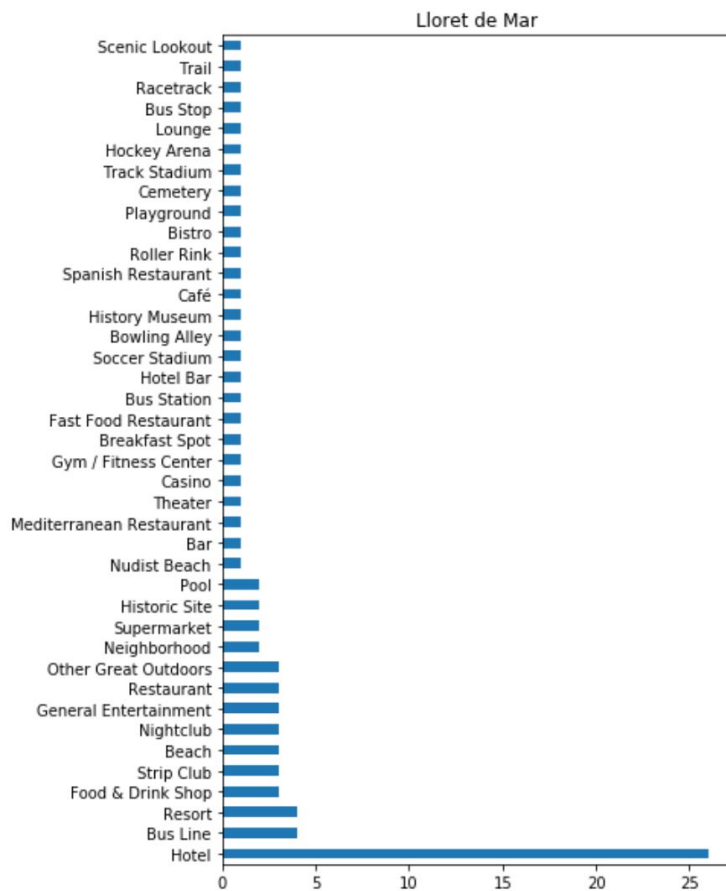
Tossa de Mar



We can see, that here nightlife and entertainment are not so developed. Tossa de Mar has enough hotels, Mediterranean restaurants and beaches. Here is Castle, Historic Site, Museums. This town is suitable for calm and quiet rest. For people who don't love nightlife, noisy places and want to relax and be with nature. It is good for couples or old-aged people.

Lloret de Mar

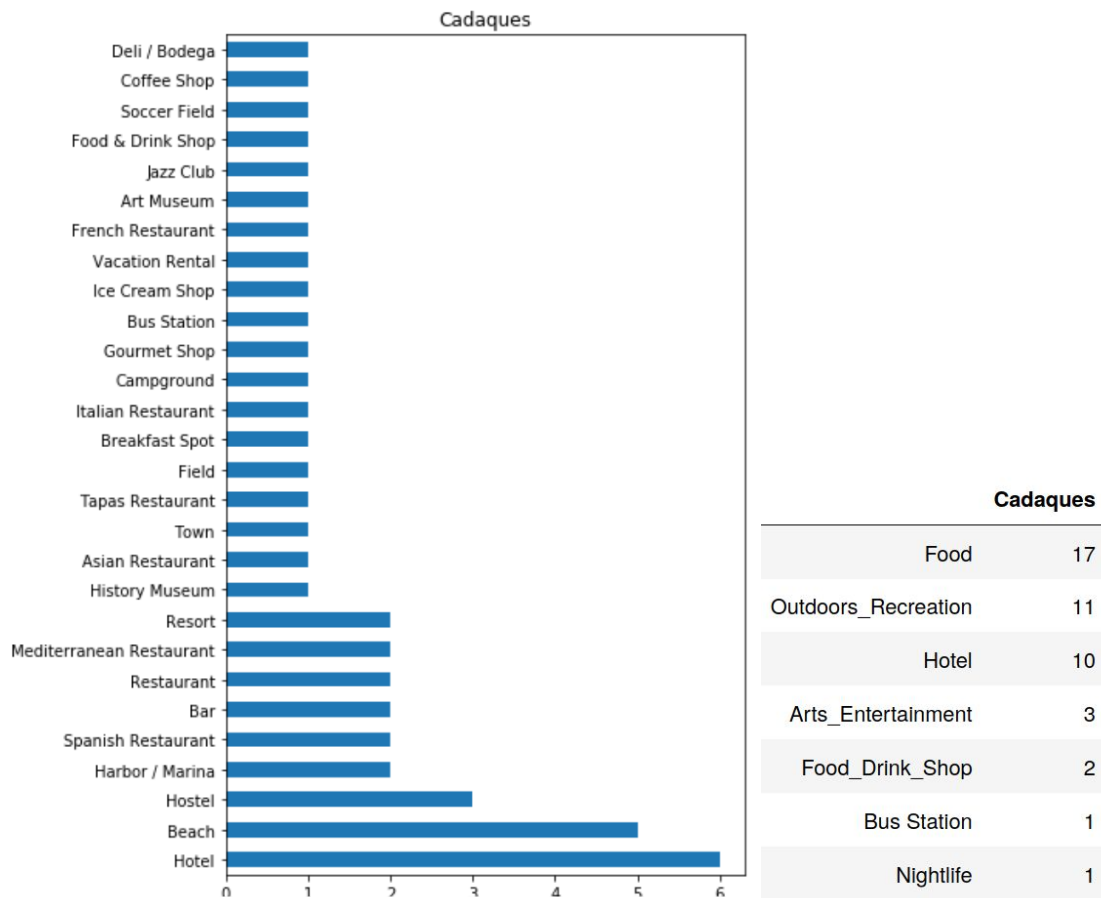
Lloret de Mar has big choice of hotels and accomodations and all necessary facilities for comfortable rest. Except hotels there are shops, beaches, restaurants, neighborhoods, museum, bus station and stop. So it is also suitable for active vacation, for families and friends.



Lloret de Mar	
Hotel	30
Outdoors_Recreation	17
Arts_Entertainment	13
Food	10
Nightlife	9
Bus Station	5
Food_Drink_Shop	4
Bus_Stop	1

Cadaques

Cadaques has not a lot of venues, no nightlife. Only some food places, hotels and outdoors&recreations. And we can make a conclusion, that it is for calm rest. There are Hotels, beaches, restaurants, some bars, museum.



But after that, I did diagram with all of these towns to have possibility compare them among main categories. For this purpose, write chosen towns to the list. Next, I created new dataframe, where columns are names of main categories and indexes names of defined cities. First, fill all cells by zeros. After, I filled new table with actual data for each city. Remember, that cities can have not all categories, so in such cells will be zeros. And I plotted bar with all cities. And we can compare them. First, I created array from data frame and plot all bars.

	Hotel	Food	Food_Drink_Shop	Bus Station	Nightlife	Outdoors_Recreation	Arts_Entertainment	Train_Station	Bus_Stop
Blanes	6	14	6	2	13	25	14	4	1
Tossa de Mar	13	17	0	1	2	18	10	0	0
Lloret de Mar	30	10	4	5	9	17	13	0	1
l'Escala	6	30	5	0	16	29	15	0	1
Palamos	3	14	8	1	6	26	6	0	0
Cadaques	10	17	2	1	1	11	3	0	0

As we can see at *Figure 1* , such categories as Food, Hotel, Outdoors Recreation, Art Entertainments, and Nightlife are in all towns that we choose earlier.

In **Tossa de Mar** we did not find any food and drink shops. In **l'Esala** there is no Bus Station. Train Stations are only in **Blanes**. In **Tossa de Mar**, **Palamos**, and **Cadaques** there are no bus stops, in the other towns there are the same amount of bus stops.

- The biggest number of hotels is in **Lloret de Mar**.
- The biggest number of Food places is in **l'Esala**.
- The biggest number of Food and Drink shops is **Blanes** and **Palamos**.
- The biggest number of Nightlife is in **Blanes** and **l'Esala**.
- The biggest number of Outdoors and Recreation is in **Blanes**, **l'Esala**, and **Palamos**.
- The biggest number of Art Entertainment is in **Blanes**, **Lloret**, and **l'Esala**.

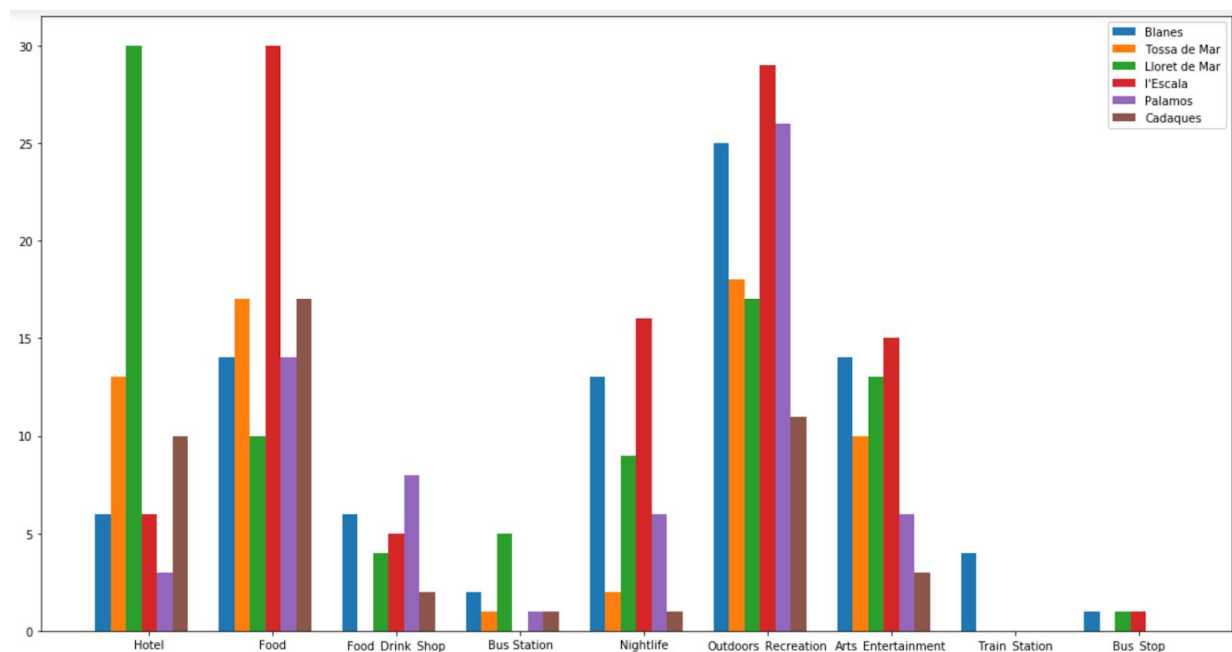


Figure 1. Comparing towns by main categories

Comparing by categories

And now the last time I compared these towns by separate category.

Beach

There are most beaches in Cadaques, Blanes and l'Escala.

Beach	
Cadaques	5
Blanes	4
l'Escala	4
Lloret de Mar	3
Tamariu	3
Tossa de Mar	3
Llafranc	2
Palamos	1
Begur	1

Hotels

There are most hotels in Lloret de Mar and Tossa de Mar.

Hotel	
Lloret de Mar	26
Tossa de Mar	11
Begur	10
Cadaques	6
Llafranc	5
Blanes	4
Palamos	2
Calella de Palafrugell	2
Tamariu	1
Pals	1
l'Escala	1

Restaurants

There are Restaurants in l'Escala, Blanes and Tossa de Mar.

Restaurant	
l'Escala	18
Blanes	12
Tossa de Mar	12
Cadaques	10
Begur	10
Palamos	8
Lloret de Mar	6
Llafranc	3
Calella de Palafrugell	3
Tamariu	2
Pals	2

Nightlife

There are Nightlife in l'Escala and Blanes. And there are not nightlife in Begur, Tossa de Mar, Pals, and Cadaques.

Nightlife	
l'Escala	16
Blanes	13
Lloret de Mar	9
Palamos	6
Begur	2
Tossa de Mar	2
Pals	1
Cadaques	1

Food and Drink shops

There are Shops in Palamos, Blanes, l'Escala, and Lloret de Mar.

Food_Drink_Shop	
Palamos	8
Blanes	6
l'Escala	5
Lloret de Mar	4
Cadaques	2
Calella de Palafrugell	2
Pals	1
Tamariu	1

Playgrounds

There are more Playgrounds in Palamos and Blanes.

Playground	
Palamos	2
Blanes	2
Llafranc	1
Lloret de Mar	1
l'Escala	1

5. Results and Discussion

My analysis shows, that in the Costa Brava sea side there are towns, for different kind of people. And anyone can find something interesting and suitable. So, I've chosen such groups of people: family with children, couples, aged people, friends, people who want calm vacation or vice versa. I investigated which city is the most suitable for each group.

We have seen, that some towns has more hotels, some has more nightlife, some has more restaurants etc.

So, here my observations, that I have seen:

1. Family with children - Blanes, Palamos, Llore de Mar
2. Couples - Tossa de Mar, Blanes
3. Aged people - Palamos, Tossa de Mar
4. Friends - l'Escala, Blanes, Lloret de Mar
5. People, who want calm vacation - Cadaques, Tossa de Mar,
6. People, who love active rest - Blanes, Lloret de Mar

6. Conclusion

Purpose of this project was to identify sea cities of Spain, cluster them and specify which city is suitable for some groups of people. The problem was that there are a lot of websites and videos with information about these towns, but often it is only general description without information about accommodation, stores, beaches and other facilities. And there is a problem to choose the most appropriate town, that can satisfy our needs and possibilities. First step was to get data about venues for chosen towns using Foursquare api. Next step was data cleaning and filtering, where we get only data that we need for further analysis. After that, I made k-means clustering and specify that Begur and Pals are not located near the sea. Tamariu and Calella de Palafrugell have only small amount of venues, so such cities I left and explore only others. I compared all left seven towns and made some observation.

I can underline my top towns, that are leaders in most categories and could be suitable for a lot of groups of people. So they are:

1. Blanes
2. Lloret de Mar
3. Tossa de Mar.