

Applied Data Science Capstone Project - Data Science Certification (IBM) by Mariana Ferreira

Where to set up a café in Sitges, Spain?¶

Table of contents

- 1) Introduction: Business Problem**
- 2) Data**
- 3) Methodology**
- 4) Analysis**
- 5) Results and Discussion**
- 6) Conclusion**
- 7) References**

1) Introduction: Business Problem

In this project, I will attempt to determine the most optimal location for setting up a new café in the costal town of Sitges, Spain. This project is aimed at potential business owners interested in setting up in that town.

This project will use manually acquired district/neighborhood information of Sitges as well as the Foursquare API to find future competitors and other landmarks.

Location is critical to the overall success of any business.

In the case of cafés, there are many variables to consider to maximize the chances of prosperity. Some of these are: visibility, parking, demographics, points of interest in the area (schools, hospitals, public offices, etc) and high foot traffic.

Sitges is a relatively small town, so the search will focus on finding locations that are not extremely crowded by local cafés and will favor locations as close to the city center as possible.

I will also focus on identifying the aforementioned areas of interest, since they usually signify more people going about by foot (foot traffic).

It is important to note that competition isn't necessarily a bad thing; a group of cafés can attract more people to the area over time, the key is to differentiate our business from the competition.

2) Data

Definition

Upon doing research, I discovered that, due to how small of a town Sitges is, there is no unified criteria regarding neighborhood/district division, which made it impossible to find a single source to scrape and create the database from.

Therefore, using several resources (specifically, local real estate platforms like ("Idealista" <https://www.idealista.com>) I manually created a csv file including neighborhoods' names, postcodes and coordinates, which I uploaded directly to my project.

Based on the problem I just defined , factors that will influence the decision are:

- number of and distance to existing cafés in the neighborhood
- number of points of interest (universities, schools, public offices, etc)
- distance from neighborhood to city center

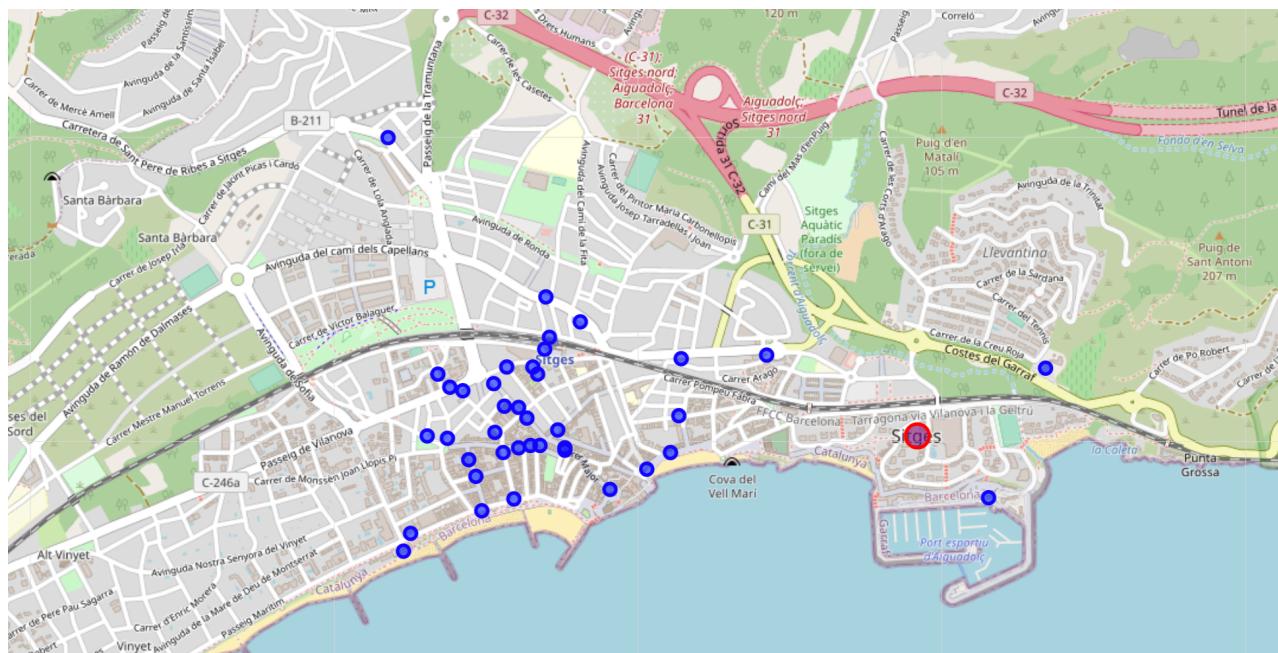
To find the number of existing cafés and landmarks and their location per neighborhood, I will be harnessing the power of the Foursquare API.

Screenshot of the manually obtained neighborhood database:

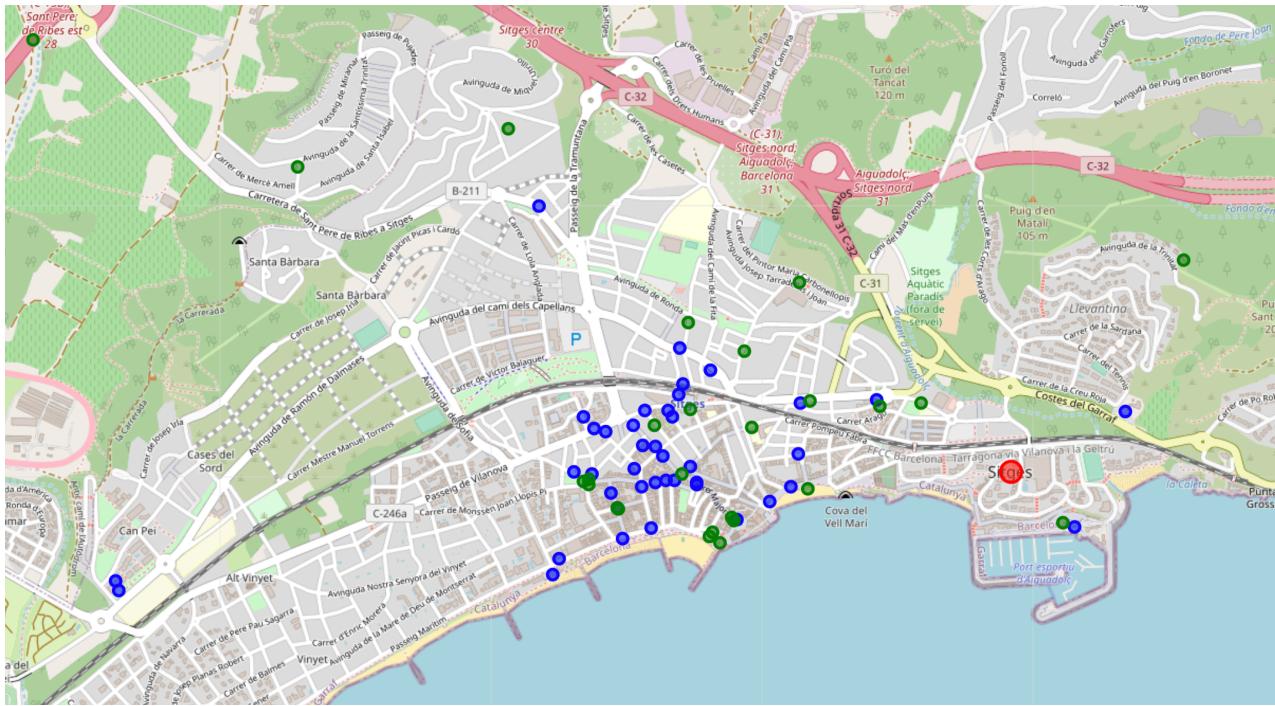
	Neighborhood	Postal Code	Latitude	Longitude
0	Downtown Sitges/Sant Sebastià	08870	41.2372	1.8059
1	Aiguadolç	08870	41.2348	1.8254
2	Urbanització Levantina	08870	41.2403	1.8249
3	Urbanització Montgavina	08870	41.2383	1.8313
4	Quint Mar	08870	41.2495	1.8661
5	Urbanització Vallpineda	08870	41.2459	1.7962
6	Poble Sec	08870	41.2439	1.8088
7	Els Molins	08870	41.2392	1.8170
8	La Plana	08870	41.2387	1.8097
9	Vinyet	08870	41.2344	1.7911
10	Urbanització Rocamar	08870	41.2352	1.7817
11	Terramar	08870	41.2279	1.7822

Then I proceeded to use the Foursquare API to get the top 100 cafés in Sitges within a radius of 3000 meters from Downtown.

With that new dataframe, I have created a map to visualize all cafés.



Then, using the same method with Foursquare API, I located points of interest within the same radius and mapped them along with the cafés, to get a sense of where most venues were located.



This map shows what my previous research suggested; areas where there are hospitals, schools, post offices and/or parks, tend to be preferred when it comes to setting up a café.

3) Methodology

In this project I've directed my efforts towards detecting specific locations around areas where there are points of interest, particularly those with low number of cafés nearby.

In these first steps I have collected the required data: neighborhoods' definition and location of cafés and points of interest per neighborhood.

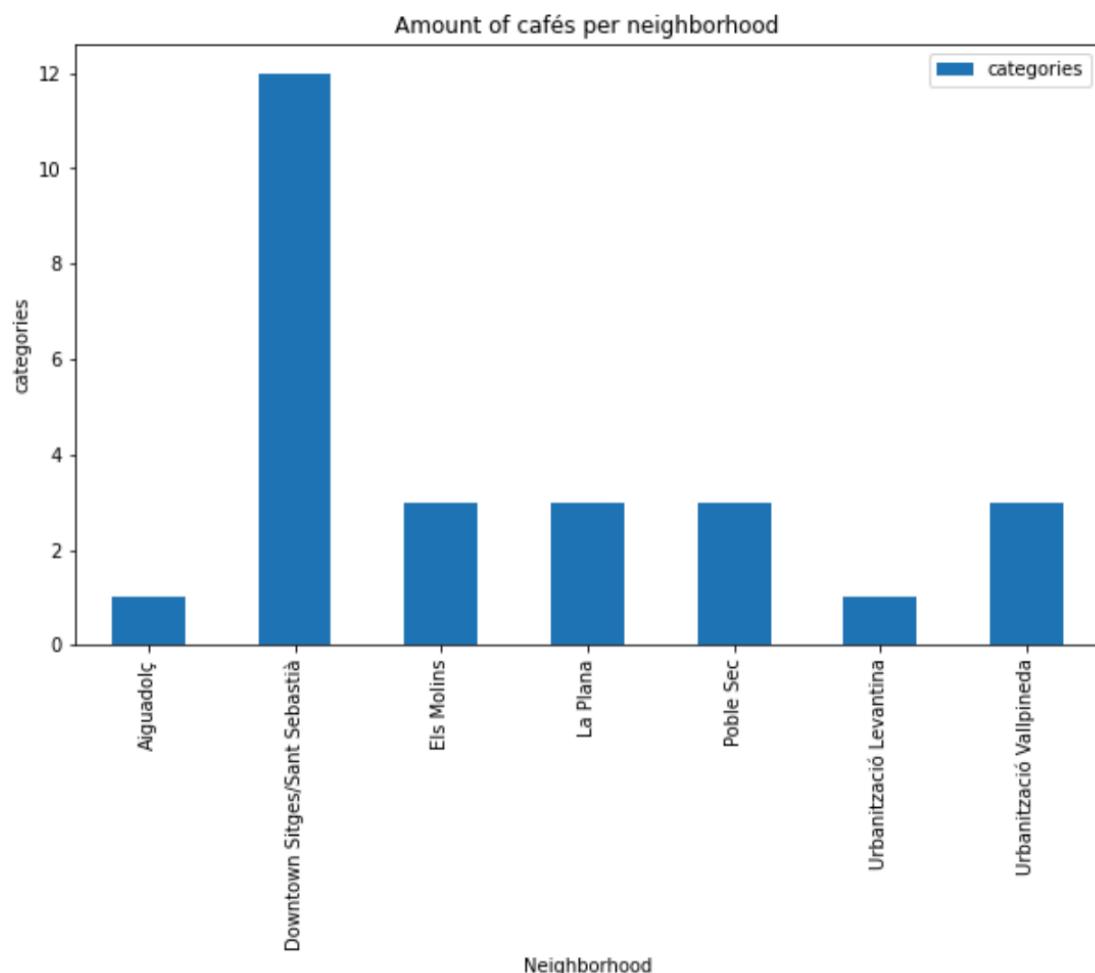
The next step in our analysis was determining 'venue frequency' across different areas of Sitges and focus my attention on areas where there are at least two points of interest and less than four cafés (competition).

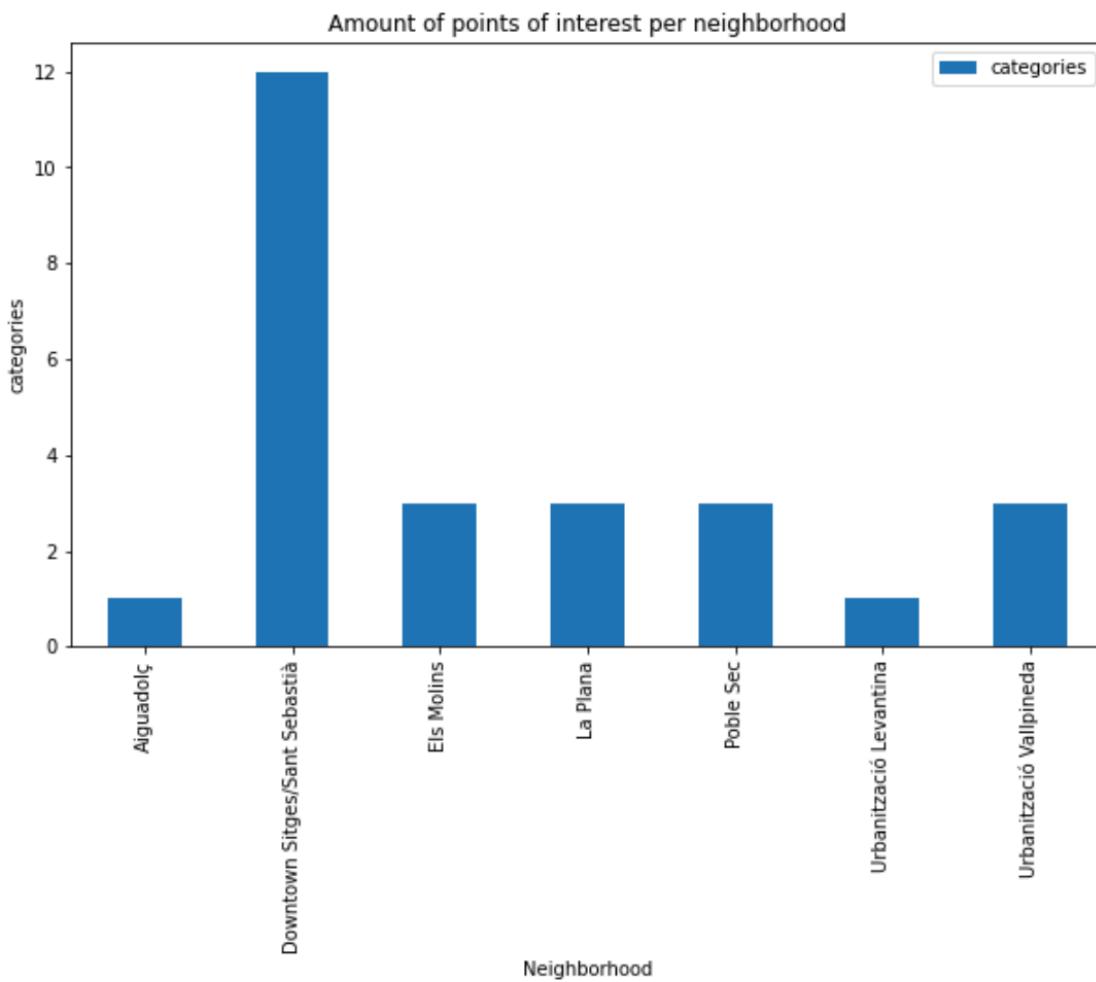
In the third and final step, I have focused on most promising areas (based on the previously discussed criteria) and within those, mark on the map, specific locations where we could set up our business.

4) Analysis

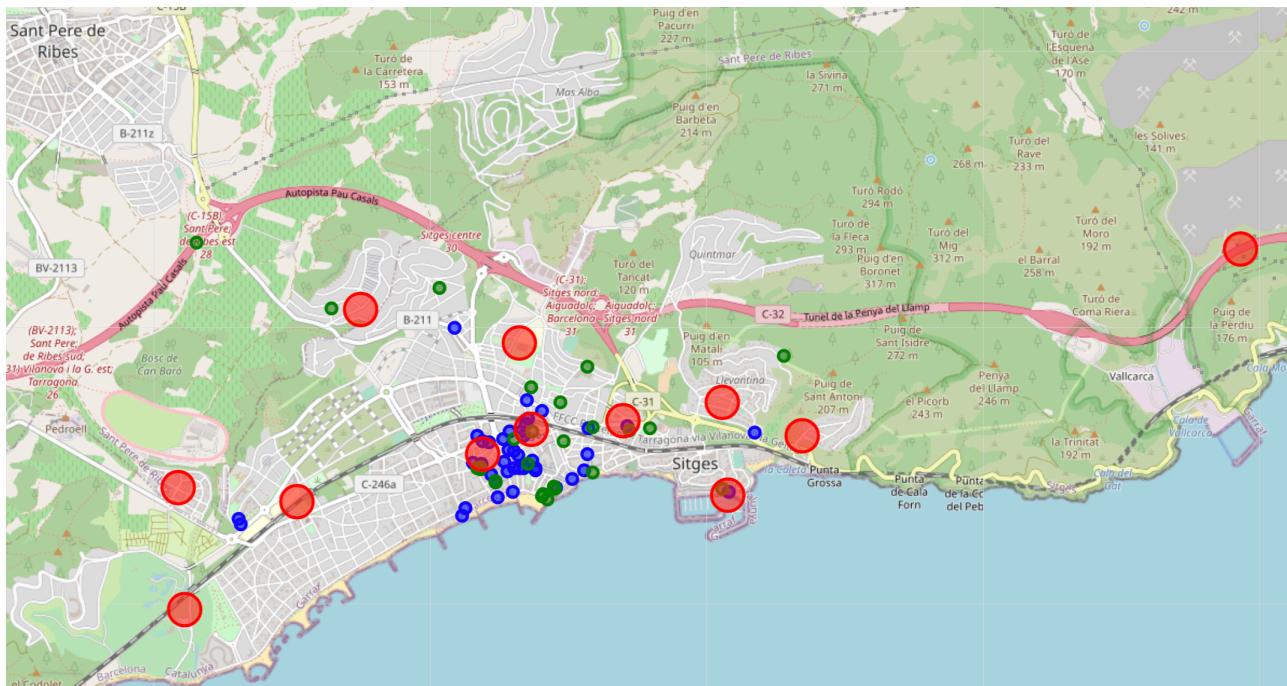
I have proceeded to merge all dataframes to create one comprehensive dataset featuring neighborhoods and all venues; cafés and points of interest.

Then I measured the frequency of both cafés and points of interest per neighborhood and created bar charts to visualize which neighborhoods feature the most venues.





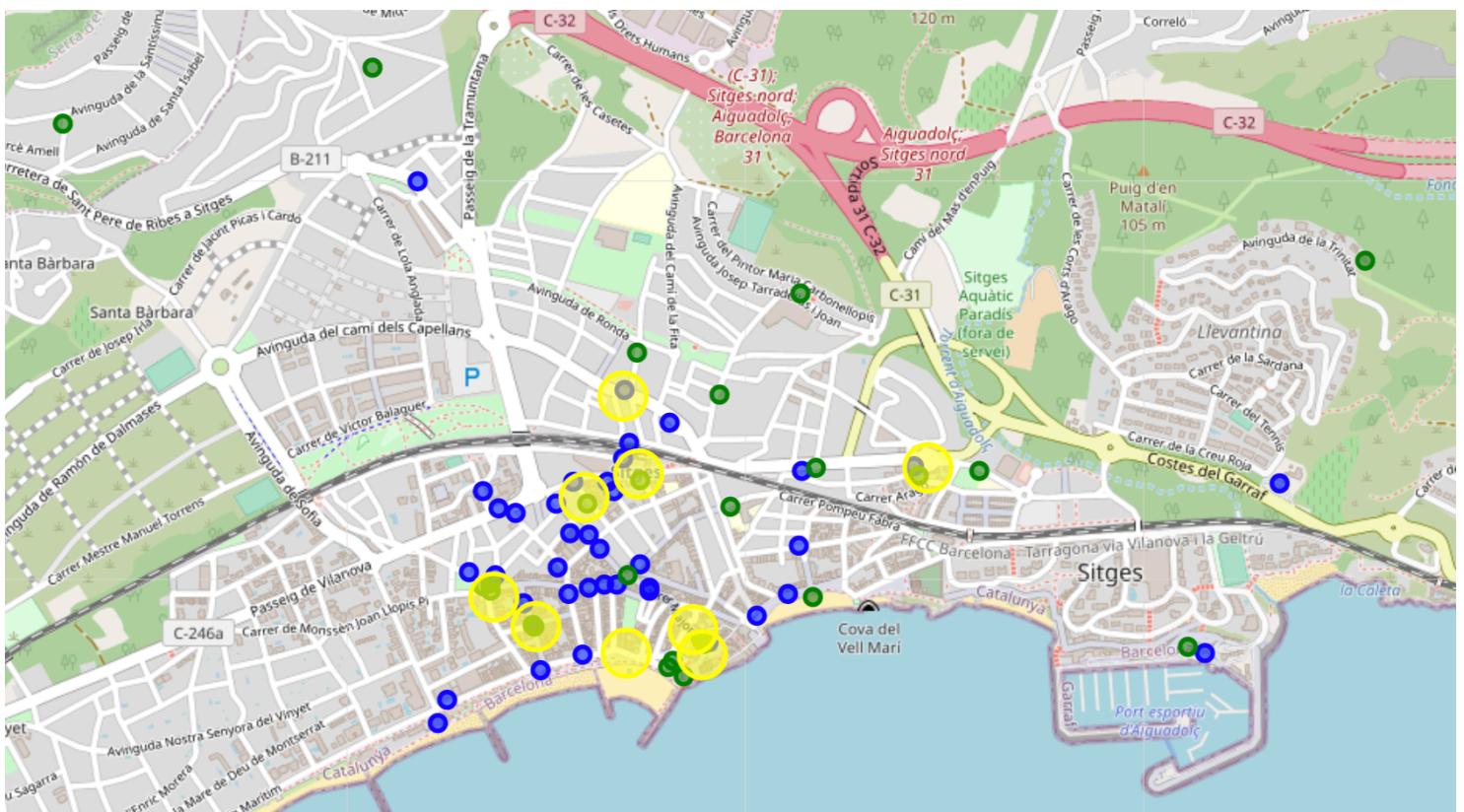
Apart from the charts, I mapped neighborhoods, cafes and neighborhoods as seen below.



Based off of these charts and the map, I was able to determine which neighborhoods feature the highest venue density.

I then created a dataframe with potential locations and their coordinates, which I imported directly to the project.

After that, I created a final map featuring those potential locations for setting up the business, which are marked in yellow.



5) Results and Discussion

Sitges is a small town in Spain, which posed a challenge in terms of finding information regarding the neighborhoods and areas where it was not already crowded by many other cafés.

I manually created a database featuring all neighborhoods by analyzing local real estate and tourism information.

I then used the Foursquare API to find all cafés in town, as well as points of interest where there is potentially high foot traffic, which is one of the factors which contributes to a successful cafés.

Later, I created a map as well as bar charts with all the aforementioned venues, to determine the neighborhoods with the highest venue density.

After that, I discovered that most venues are concentrated around Downtown and La Plana. Other neighborhoods are mostly residential areas, which feature less points of interest and less foot traffic and are therefore, less likely to be the best options for setting up a new, successful business.

My analysis shows that despite the limitations, there are still areas near points of interest (with foot traffic), where we could set up a new café. The result of this analysis is nine potential candidates.

These areas are, mostly around Downtown and La Plana, near at least two points of interest and around less than four competitors.

6) Conclusion

The purpose of this project was to identify Sitges' areas close to the center, with a low number of cafés and potentially high foot traffic, in order to aid future business owners in narrowing down the search for optimal location for a new café.

By using Foursquare data I have determined venue density distribution across town. After creating maps and charts, I have identified the neighborhoods with the most potential.

I have then created a database of locations which satisfy our requirements, to be used as starting points for final exploration.

Final decision on optimal café location will be made by future business owners based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration proximity to major roads, real estate availability, prices, distinctive characteristics of the competitors and the new business, among others.

7) References

<https://www.idealista.com>

<https://www.sevenmiles.com.au/editorial/cafe-location/>

<https://articles.bplans.com/how-to-choose-a-business-location/>

<https://www.sitges-tourist-guide.com/en/areas/sitges-districts.html>