CAPSTONE PROJECT

Opening a Five-Star hotel in A Coruña

Applied Data Science Capstone by IBM/Coursera

Gonzalo Graña

8th May, 2020

Index

Introduction

Background

A Coruña is a city located in the northwest of Spain. With an Atlantic climate (more rainy and cloudy than in other locations of Spain), the city does not apply for the typical Spanish Sun & Beach tourism model but has been traditionally working on attracting business and MICE tourism. This is important as the services around the accommodation are usually more important for this kind of tourists rather than a better location.

This means that the lodging industry in the area has been adapted to take these tourists and may have an adequate occupancy including one five-star hotel.

An opening of a new five-star hotel downtown could be beneficial for Coruña's tourism cluster when promoting the destination in the media, and a great opportunity for any hotel chain that would like to get a footprint with a luxury hotel.

Problem

The data collected should contribute to determining the location where the five-star hotel should be located to compete (in equal conditions) with the new existing hotel, and the cost of its construction for a maximum duration of 1 year.

Interest

City's tourism cluster and local authorities should be interested in this project as it could report a great impact in the city, not only in the tourists that could be attracted by an extended lodging offer but also because of the jobs this new hotel would create in Coruña. Also, a big hotel chain or investor should be interested as they would have an opportunity to lease this new property and added to their portfolio.

Data

Sources

With the following data, we should be able to figure out the location where the new five-star hotel should be built, how our hotel should be (number of rooms, number of meeting rooms, etc.) and how much the construction will cost:

- Foursquare: With its API, I should be able to retrieve all the venues located in Coruña's downtown and separate them in differentiated clusters using the KNN method. The ideal scenario would be to divide the area into two clusters (one for our hotel and the other one for our competitor's)

- Instituto Galego de Estadística: This regional's government site collects all the relevant macro data related to tourism. It will be useful to collect occupancies five-star hotels in the area (yearly and monthly) and see how stable is the tourism market and how many rooms will our hotel need to have the best efficiency.

- Turismo da Coruña: Local government site that collects information about hotels (among others). It will be useful to get data from our competitor and get an idea of the common areas, number of rooms, etc. that our hotel will need.

- Idealista.com: Most relevant Spanish websites to check rent prices.

- Hotelnewsresource.com: This website has a complete guide with the needs and estimated costs for building properties based on our needs (star rating, artwork, etc.). It will be useful to gather an average of how much will be the cost of the hotel's construction.

Formatting

Many datasets have been clean or modified to achieve the following:

- Remove unnecessary data from the data frames (e.g. min/max costs of renovations, as we already will display the average)

- Unify data language from all datasets to English

- Simplify data and get more accurate results (e.g. venues' categories)

Methodology

We will need to:

- Identify the clusters in the area where our hotel could be placed, and get to know the rent that needs to be paid in that area during the hotel's construction.

- Determine the size of our hotel. The number of rooms needed in our property will have a great impact on the costs of the construction and, in the future, about the rent.

- Calculate the construction costs.

- With all of this, we should be able to find out the total cost of our construction and present it to stakeholders.

Analysis

Where will our hotel be located?

By pulling the top venues in downtown Coruña from the Foursquare API and classifying them into 6 main categories, we begin by checking out how many clusters we have based on the venues.

We find out that all venues can be summarized in one cluster, but since all venues can be consolidated in one cluster but we want to compete in the same conditions as the existing hotel, we will divide venues into two clusters. Now we can have both clusters, and the location of our competitor (displayed by a marker). Our cluster will be the one inside the **red points** in the map below.
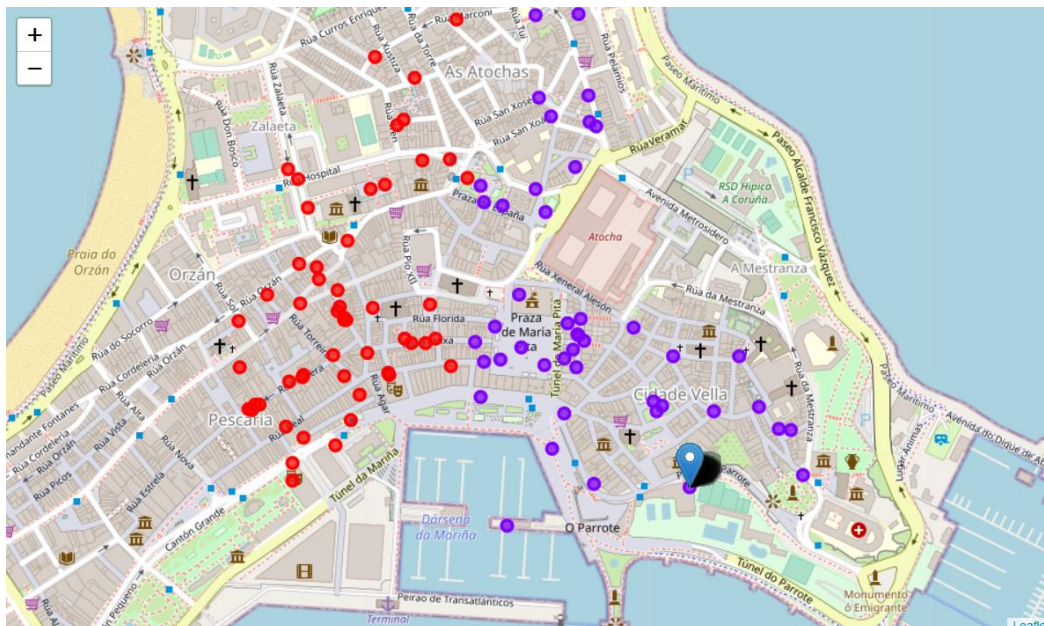
Just to get a better idea if we have a big disadvantage in any venue's category and the total amount of venues, we can compare both clusters as shown in the following graphs.
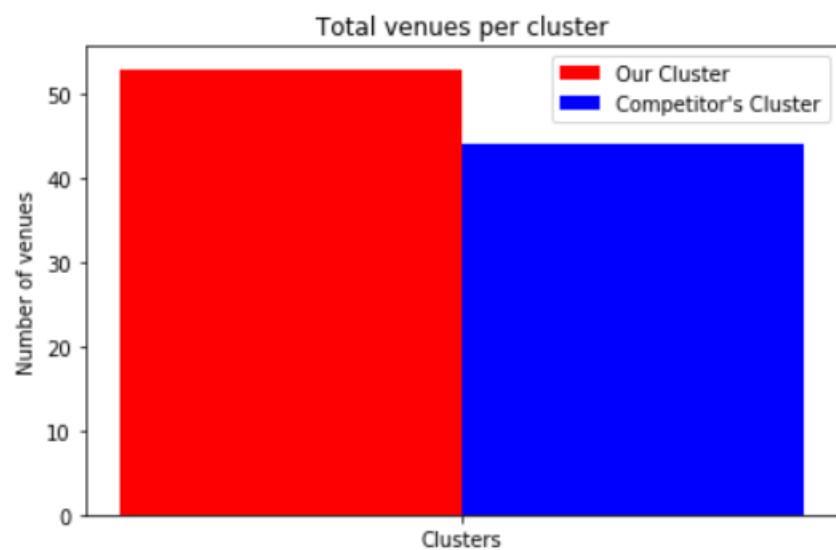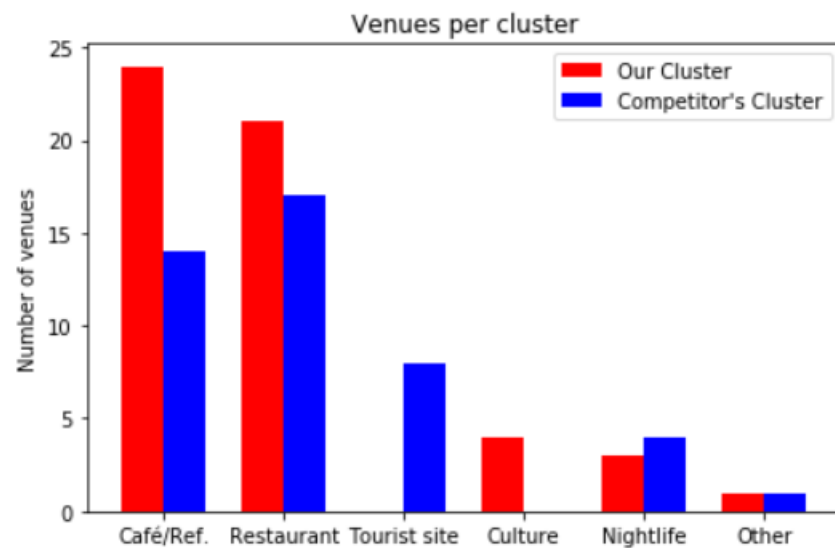
What will be the size of our hotel?

Now that we know the location of our hotel and we made sure we won't have a disadvantage on the services point of view, we need to find out how big our hotel should be to reduce our costs as much as possible. In a business hotel, its size is usually measured by the number of rooms it has, so we will try to get that number.

First, we load yearly occupancy data for 5-stars hotels in the area (the data will be really accurate as our competitor is the only 5-stars hotel). The yearly occupancy is not so relevant about getting the number of rooms we need but is a good indicator to measure if the market is stable, and we should trust data that will be gathered ahead.

Year	Occupancy Rate
2014	57.60%
2015	60.30%
2016	62.10%
2017	61.70%
2018	60.20%

Now, we will check the monthly occupancy during the last 3 years.

Month	2017	2018	2019
January	29.20%	33.40%	35%
February	34.10%	38.90%	39.10%
March	39.10%	41.30%	39.60%
April	50.80%	47.30%	48.20%
May	43.10%	46.70%	53.30%
June	51.70%	52.80%	56.10%
July	71.80%	70.30%	71.60%
August	82.30%	83%	83%
September	58.40%	57.50%	61.30%
October	50.30%	50.40%	53.30%

November	42%	48.50%	45.20%
December	37.50%	42.20%	43.40%

In the table, we see that A Coruña has a strong high season. As high season will bring big profits to the hotel, we don't want to miss the opportunity to have it on 100% full occupancy, so **we will need to have just the 83% rooms that our current competitor has.**

That way, we will not have empty rooms and we will have fewer costs during our construction. Also, future investors will have to pay less rent, taxes, maintenance costs, etc.

Then, we need to figure out much space our competitor is using in common areas and adapt it to the size we just calculated. Here are the meeting rooms and fitness facilities from our competitor and their size:

Common Areas	Size (SqMtr)
Victoria	441
Miramar	180
Mindanao	98.26
Isla	83.43
Finisterre	85.36

So we can determine the size of our competitor's hotel:

Total size of competitor's common areas: 888.05 m2

Total size of competitor's room areas are: 3449 m2

What will be the cost of our hotel?

According to Idealista, one of the biggest real state sites in Spain, the price as to rent per May 3rd, 2020 is 9.2 EUR/m2 so we can calculate **the rent**. Now we know the rent during the next 12 months **will have a fixed cost of 397,412.52 EUR**.

About the construction costs, from Hotel News Resource site we have an example of the cost (in USD) to build/renovate an upscale hotel as the one we want. So we can now load the data and adapt the costs as per the size of our hotel. So the **building renovation cost**, with the criteria select, is **4,093,987.68 EUR.**

So finally, we can sum both amounts and determine how much will **our hotel's total cost during its construction: 4,491,400.20 EUR**. To see the impact of each feature in the construction, we can see below a bar chart with its breakdown.
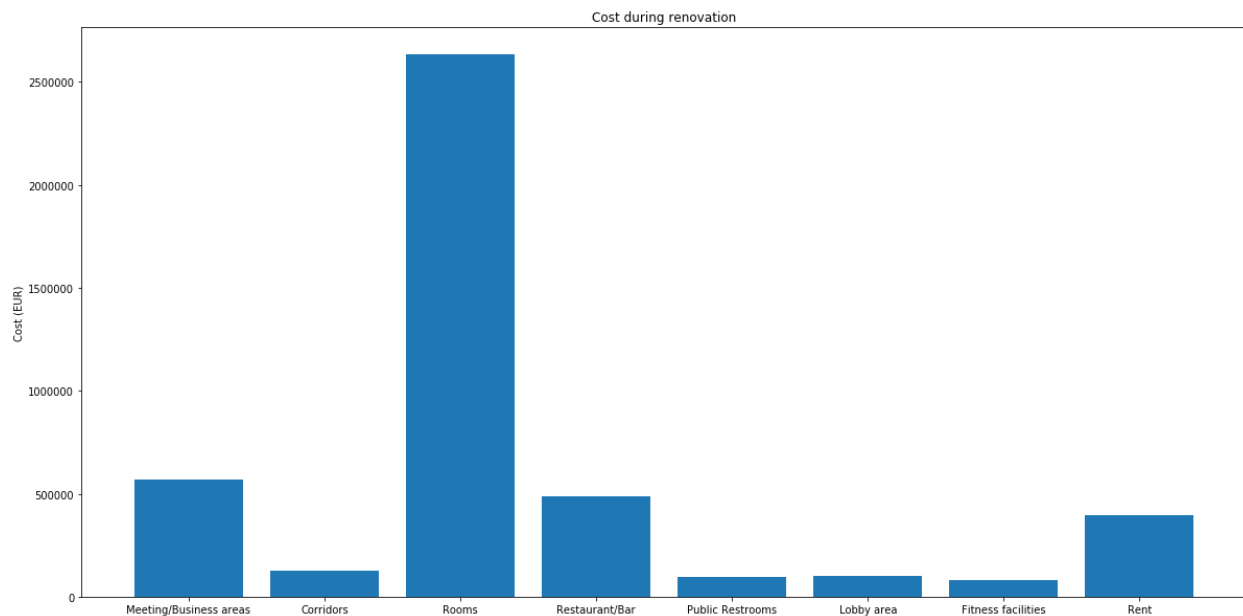
Results and Discussion

The purpose of this project was to find the best location in downtown A Coruña to build a five-star hotel and calculate the estimation of the costs, before renting/selling it to a hotel chain or investor.

Knowing that there is only another five-star hotel in A Coruña and that venues are the most important services around a business hotel, we located the main venues in the area and divided them into two clusters: One that will belong to our hotel and our competitor.

However, we needed to double-check the balance of those venues so we displayed them on different charts. The result was that our cluster would have a bigger amount of venues, while its distribution is somehow equal. The lack of tourist sites could be a problem in our neighbor and we weren't able to measure the impact, but it is being compensated with the number of cultural places we will have around.

Then, we needed to figure out the costs of our hotel which will be based on size. So, to determine the size of the hotel we pulled data about the occupancies (average and maximum) in the last years for our competitor. As we found out that the market was stable and the maximum occupancy during high season was 83%, we assumed that our hotel should have a maximum of 83% the number of rooms (and therefore, common areas) our competitor has. This was the most important data gathered, as this assisted us to determine:

- How much will be the rent during the construction
- How much will the construction itself cost

Since we are having a better efficiency than our competitor, we should be able to decide in the future not just if we want to compete with them in terms of quality but also in price (we will have fewer maintenance costs as we will have fewer rooms/common areas).

Conclusion

The data in the previous section has been pulled for building a luxury property, which has a little bit better quality than an average five-star hotel in Spain (and, consequently, better than our competitor), so there is still room to reduce our costs. Our analysis says that if our investors decide we should lower the costs, we should start by the rooms as they are absorbing most of our expenses.

Also and to complement that, we should work on a follow-up program to this project to assist in attracting stakeholders to this acquisition.

Mainly, this should consist of a feasibility study where we could calculate the time needed for the ROI, the average daily rate that we will need, how much of a positive impact it could have in the image (therefore the increase of reservations) in the acquirer chain, etc.