PREDICTING THE LOCATION FOR OPENING A NEW HOTEL ***LISBON, PORTUGAL***

Capstone Project - The Battle of Neighborhoods

1. INTRODUCTION

- A major investment must be made. Lisbon is the chosen city. A major hotel chain is expanding its' operation and the Portuguese capital is the city to go.
- With the data science project developed, we expect to choose the best location for this new building, considering not only the location of the competitors but, also, the location for all the major attractions – mainly restaurants and bars.

2. DATA

- Our core location is ‘Saldanha, Lisbon’. This is the business center for Lisbon, and the target customer is the executive one. This represent clients with business accounts, higher incomes and major plafond to spend.
- For the project, we will be addressing information from **Foursquare**. What are we interested in?
- Firstly, we want to know where the competitor hotels are located. With this, we pretend to be as far as possible from all of them.
- Also, we want to know where the main attractions are located. With it, we want to be as close as possible to most of them.

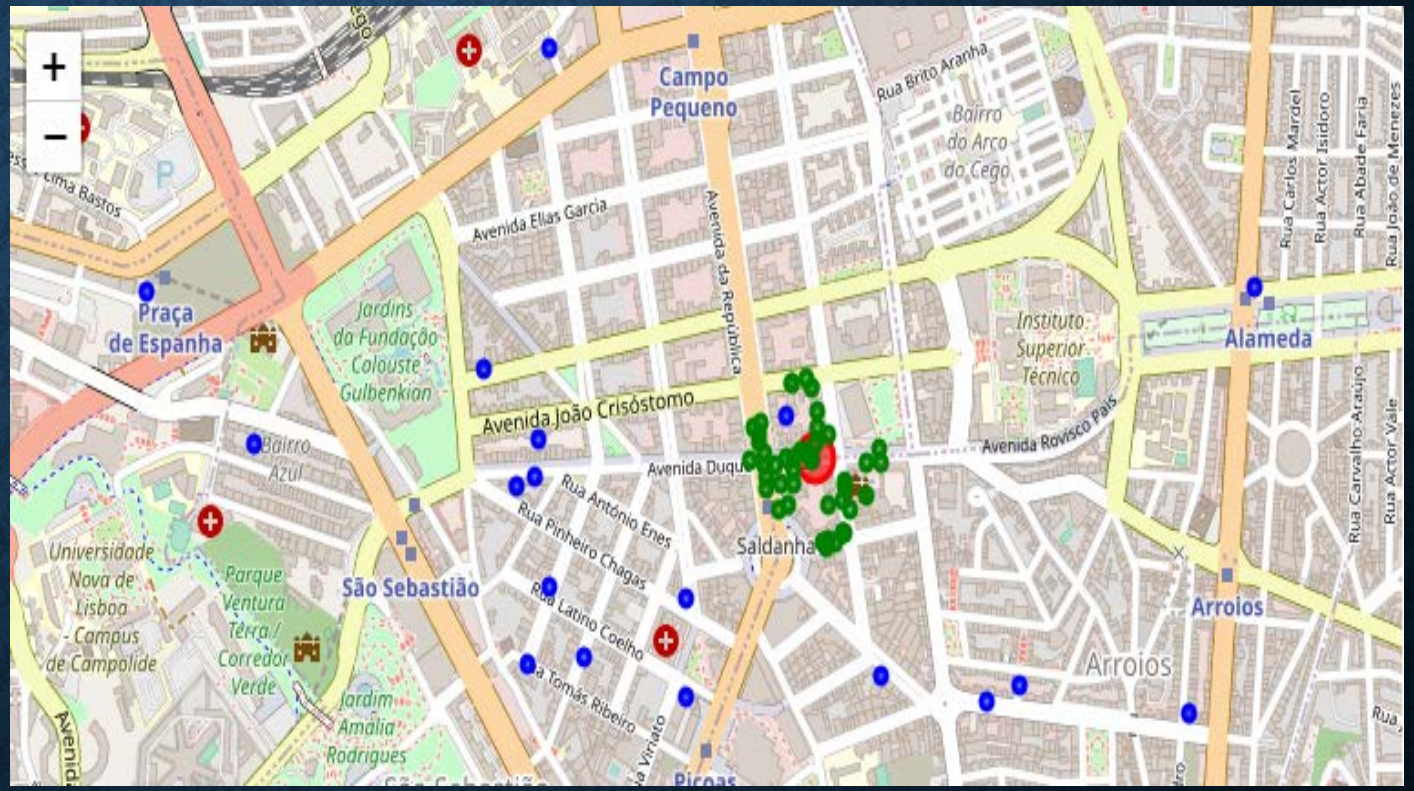
2.1. WHAT ATTRACTIONS ARE RELEVANT FOR OUR CUSTOMERS?

- As stated, our target customer is the executive one. So, it's relevant to consider specific categories of attractions for this profile. Do we want to consider amusement parks and beaches? We don't want to…

Arts & Crafts Store
Asian Restaurant
Bakery
Bar
Bistro
Boutique
Building
Café
Casino
Indian Restaurant
Italian Restaurant
Music Venue
Office
Portuguese Restaurant
Restaurant
Spa
Sushi Restaurant
Deli / Bodega

2.2. MAPPING THE LOCATIONS

- The next figure may show us three things (to be confirmed further ahead in this report):
  - The business center for the city – in red;
  - The competitor hotels marked in blue – it is possible to verify a great dispersion of them around the business center;
  - The attractions' location in green – there are many of them, so the Foursquare API is showing us more of 30 places closest to the business center;

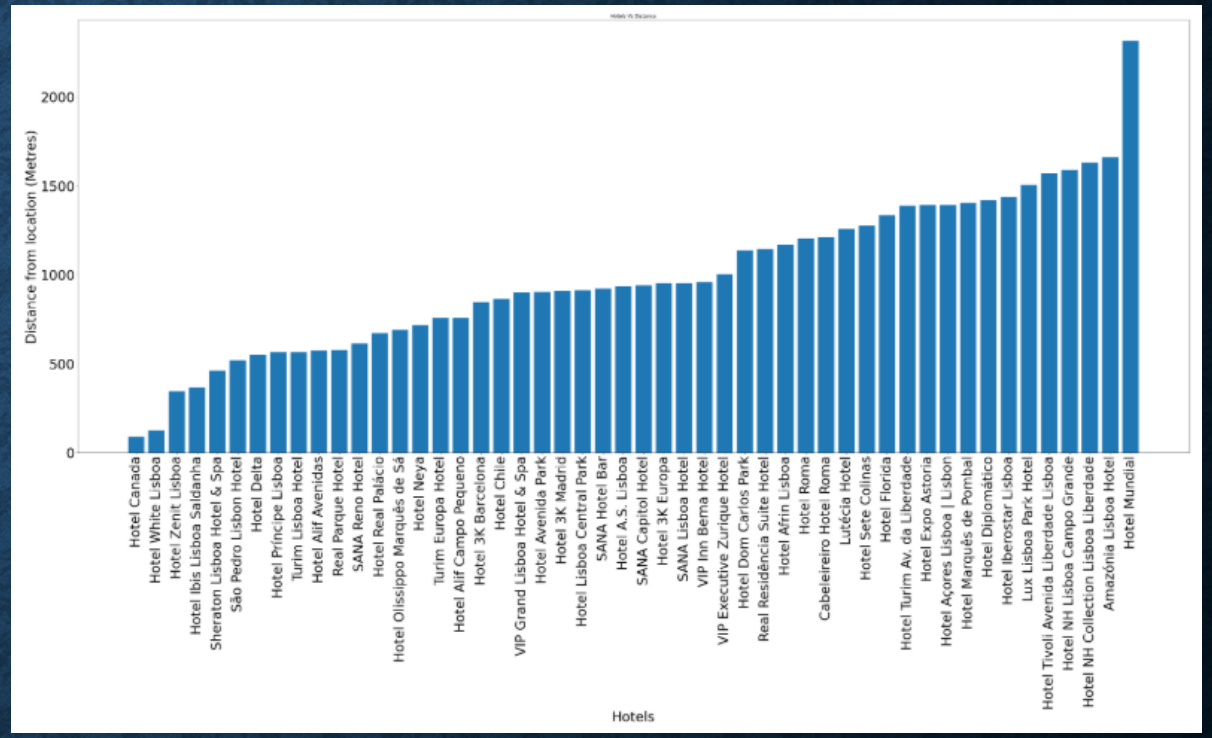


3. METHODOLOGY

- We want to look at the data gathered
  - How far are the competitors hotels from the city business center?
  - How close are the main attractions to this center, considering the relevant categories?
    - How are they clustered, considering the distance from the center?
    - Considering the categories chosen, how many of each category can we find in a short distance?

3.1. THE COMPETITORS

- We want to look at the data gathered. **How far are the hotel from the city center?**
- At first, our algorithm tells us that the average distance between hotels and the business center is **988 meters**. It's quite far, indeed.
- And if we want to 'cluster' the distance from the center it is possible to understand the real distance distribution

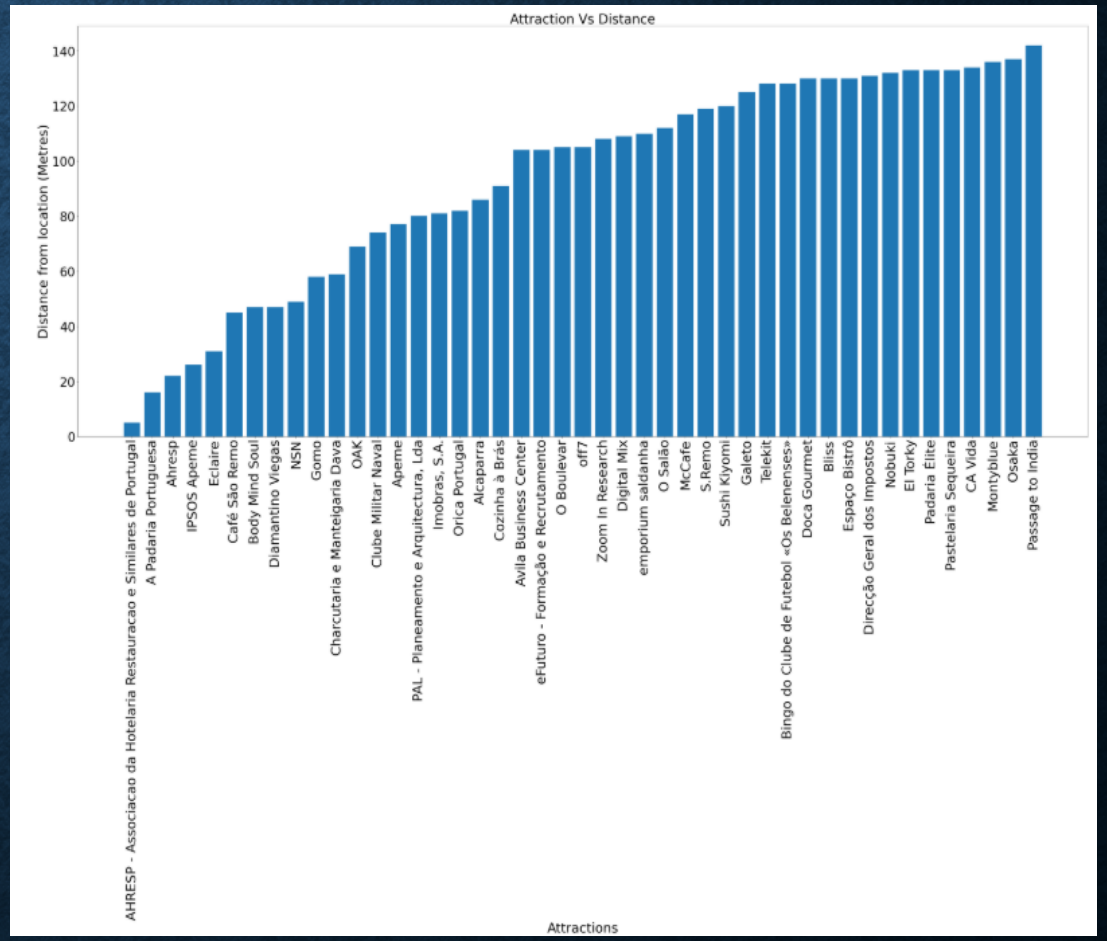


Distance range (meters)	Count
0-30	4
30-60	7
60-100	8
100-250	26
+250	0

3.2. THE ATTRACTIONS (1/2)

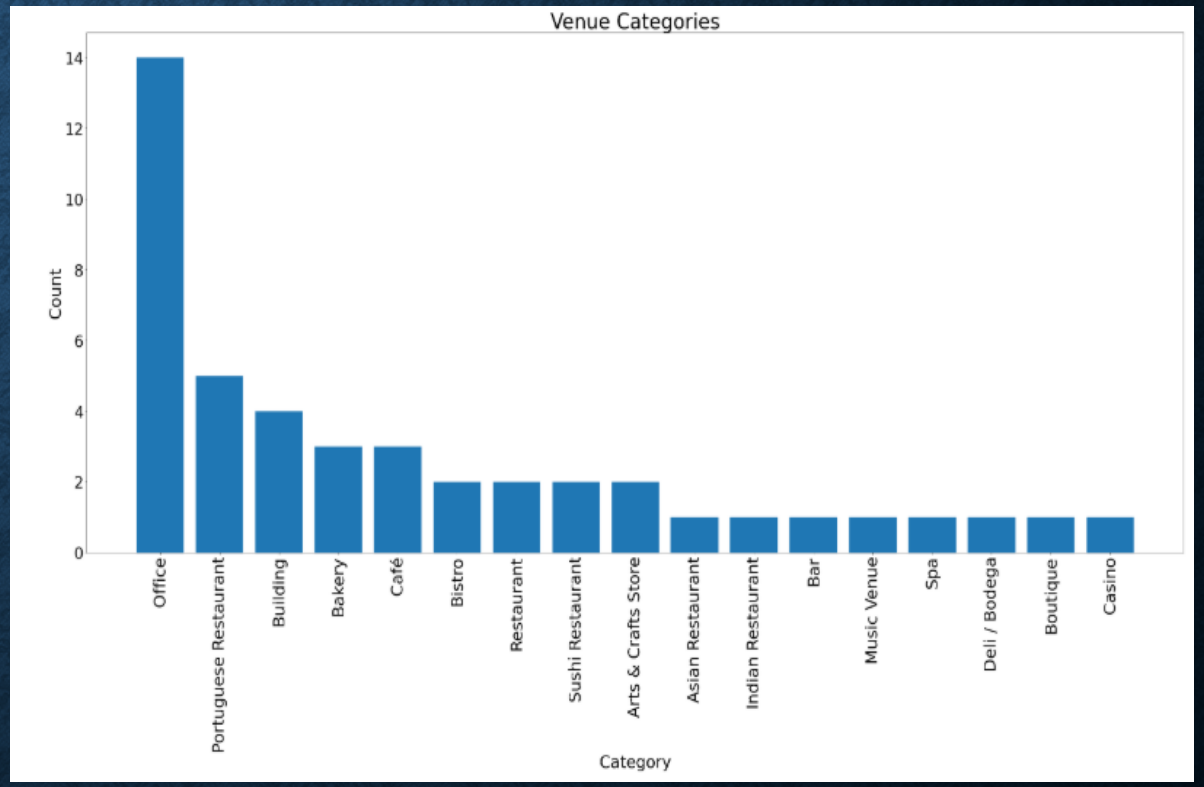
- Here, the algorithm tells us that the average attraction distance to the city center are 94 meters.

Quite interesting…



3.2. THE ATTRACTIONS (2/2)

- This information show us the major concentration of the attractions, considering the categories we have defined.
- A bunch of offices are close by, but also there are some Portuguese restaurants, bakeries, cafes and bistros... We even may find a casino.



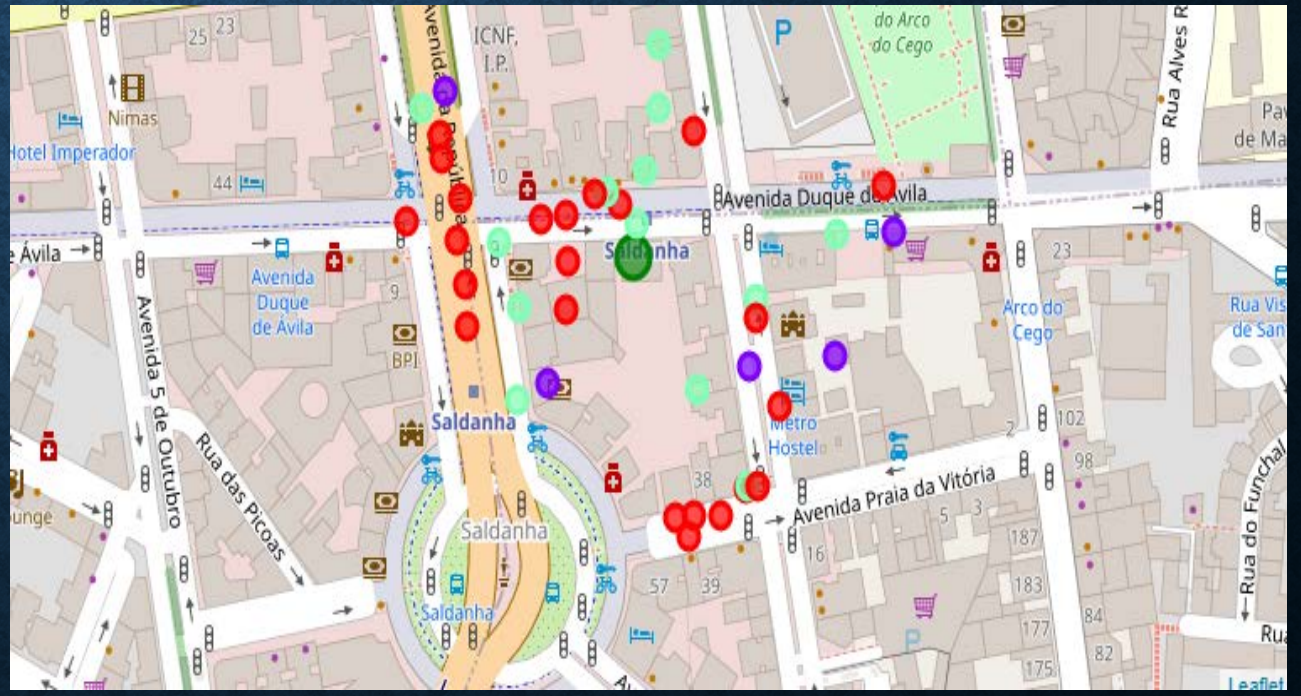
3. METHODOLOGY

- It is important here to draw some conclusions from our data analysis:
  - We have been able to find 50 hotels (at least) within the 5km radius from the business center. However, on average, they are 750 meters far from the center;
  - We have also been able to find 32 attractions (after filtering for the relevant attraction categories). But, they are, on average, 87 meters away from the center;
  - Considering the attractions, we have to point out that the most common attractions are offices (relevant for business travelers), portuguese restaurantes (what else…?), cafes, bakerys and relevant buildings. We have a bunch of those a few meters away from the center.

The potential is proven.

3.3. CLUSTERING

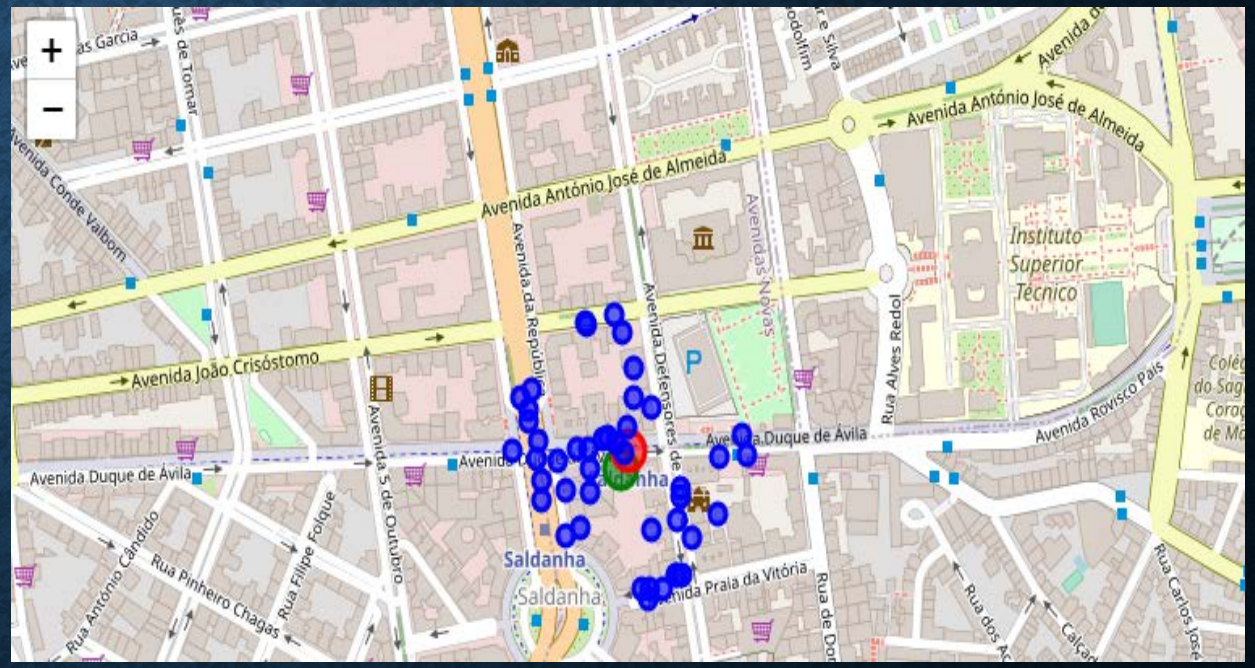
- The definition of our clusters can be perceived in this figure, when compared to the city business center.



3.4. MAPPING THE RESULTS

- And, for it, we obtained a proposed location for our hotel in 38.73501,-9.14402. This is as close to the center as it gets.
- In the above Folium map, we may encounter the business city center - red - as the hotel proposed location - green - and all the attractions marked in blue.

It's quite amazing. It couldn't get any better.



4. RESULTS

- We have been able to pin point the perfect location for our hotel: **38.73501, -9.14402**.
- We, after all, should place almost at the exact city business center, since the main attractions are quite close to it.

5. DISCUSSION

- In this project, I've considered a specific niche potential customer. I was not looking for general tourist, but targeting the business visitors to the city. Because of this, when considering attractions, it was important to select the specific ones to this kind of travelers - offices and buildings, but also cafes, restaurants and formal venues. It's important to state that, targeting a different kind of customer, we would probably get different results and hotel proposed location.
- Also, it would be relevant to argue that the limited group of results when 'acquiring' attractions from Foursquare API may condition the clustering results and, with it, the expected result for the proposed hotel location.
- Finally, in a real world approach to this problem, many other factors and conditionings would have to be considered. Let's face an example: we cannot place the hotel wherever we want... Is there an existing building on the proposed location? Is it possible to build something new there? It would be an amazing and never ending discussion...

6. CONCLUSION

- As stated, in this simplistic approach to this problem, we have been able to find the perfect location for our hotel.
- It's a real world application to our problem and has proven the data science application to this same real world problem.

Our mission is complete.