

Where is the best place to open a restaurant in Lisbon (with Foursquare API)

Nicolas Pogeant

Introduction:

Starting a new business is always complicated but even more if you start in a completely new city, a city that you don't know and that you don't have the time to discover. Therefore, this project has to aim someone in this situation, a lack of knowledge about a city and its fundamentals.

The aim of this project is to advise someone that would open a restaurant in Lisbon by using Foursquare API and data analysis. Lisbon is a city with many tourists every year so already taken by food businesses. Depending on your budget, you would be able to open everywhere but the issue would be to get people to your restaurant so competition with others has to be taken. To be able to understand that, targeting the best place in the city is a must. Therefore, we will look at the density of the different places of the city, from hotel to restaurant, passing by bar/café and others.

Data section:

For this Capstone Project, I will mainly use data from the Foursquare API, doing request and scrapping, cleaning what I get. I will use also data from the wikipedia page of Lisbon . The issue is to determine where to open a restaurant, then I will use the location, latitude, longitude, type of venue of the different neighborhood of Lisbon to understand if the place is prompt to that.

Methodology:

I started by representing every neighborhood of Lisbon from nothing.

For that, I used the geopy library to be able to locate each neighborhood from their name. There are 15 main neighborhoods in Lisbon:

	Latitude	Longitude
- Alcântara	Alcântara	38.703113 -9.180685
- Alfama	Alfama	38.712287 -9.129858
- Mouraria	Mouraria	38.714991 -9.134842
- Principe Real	Principe Real	38.716407 -9.148270
- Ajuda	Ajuda	38.712327 -9.201241
- Beato	Beato	38.732622 -9.110240
- Belém	Belém	38.697769 -9.209432
- Santos-o-Velho	Santos-o-Velho	38.707323 -9.157380
- Chiado	Chiado	38.710330 -9.140519
- Sao Bento	Sao Bento	39.526646 -8.790177
- Santo Condestavel	Santo Condestavel	38.719582 -9.168407
- Sao Sebastiao da Pedreira	Sao Sebastiao da Pedreira	38.730150 -9.152138
- Sao Francisco Xavier	Sao Francisco Xavier	38.710503 -9.213808
- Arroios	Arroios	38.731932 -9.134246
- Penha de Franca	Penha de Franca	38.726161 -9.126913

```
for neigh in Lisbon_neigh_adress:
    geolocator = Nominatim(user_agent="foursquare_agent")
    location = geolocator.geocode(neigh)
    latitude = location.latitude
    longitude = location.longitude

    rows.append([latitude, longitude])

df_lisb = pd.DataFrame(rows, columns=["Latitude", "Longitude"], index=[Lisbon_neigh])
df_lisb
```

The cell aim to collect every location data as we can see.

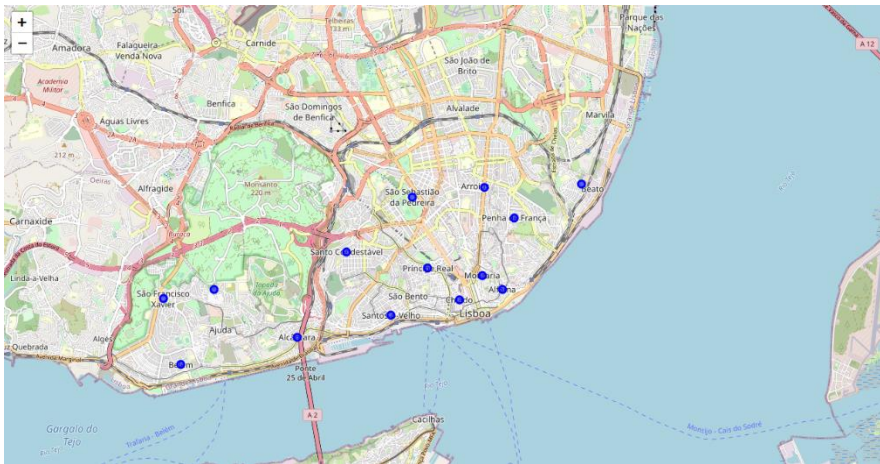
After that, I decided to represent these neighborhoods on a map, using folium

```
lisb_map = folium.Map(location=[38.734825, -9.143337], zoom_start=13)

for Latitude, Longitude in zip(df_lisb.Latitude, df_lisb.Longitude):
    folium.features.CircleMarker(
        [Latitude, Longitude],
        radius=5,
        fill=True,
        color='blue',
        fill_color='blue',
        fill_opacity=0.6
    ).add_to(lisb_map)

# display map
lisb_map
```

That is what we obtain :



Now that we lay the foundation, lets use the Foursquare API to get information of the venues in Lisbon.

I used my credentials and then create a API request URL in a loop to get every venues from every neighborhoods:

```
venues_list=[]
for name, x, y in zip(Lisbon_neigh, df_lisb.Latitude, df_lisb.Longitude):

    # create the API request URL
    url = 'https://api.foursquare.com/v2/venues/explore?client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
        CLIENT_ID,
        CLIENT_SECRET,
        VERSION,
        x,
        y,
        radius,
        LIMIT)

    # make the GET request
    results = requests.get(url).json()["response"]["groups"][0]["items"]

    # return only relevant information for each nearby venue
    venues_list.append([
        name,
        x,
        y,
        v['venue']['name'],
        v['venue']['location']['lat'],
        v['venue']['location']['lng'],
        v['venue']['categories'][0]['name'] for v in results])

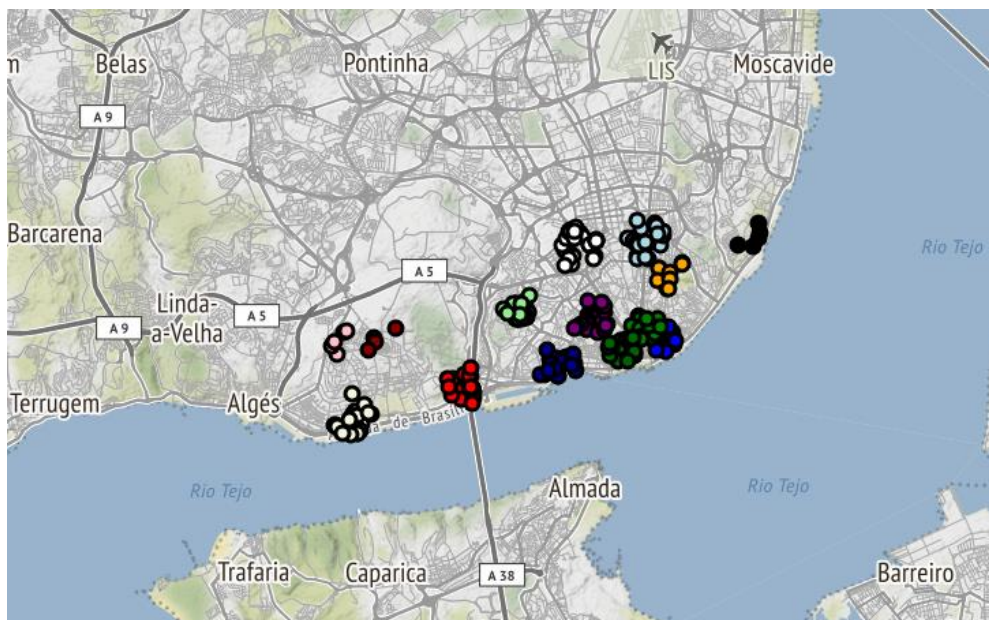
nearby_venues = pd.DataFrame([item for venue in venues_list for item in venue])
nearby_venues.columns = ['Neighborhood',
                        'Neighborhood Latitude',
                        'Neighborhood Longitude',
                        'Venue',
                        'VenueLatitude',
                        'VenueLongitude',
                        'VenueCategory']
nearby_venues = nearby_venues.drop(columns=['Neighborhood Latitude', 'Neighborhood Longitude'])
nearby_venues.tail(10)
```

Thanks to the API, we have now essential information to understand a little bit of the issue.

This is an example of some venues in Penha de Franca:

	Neighborhood	Venue	VenueLatitude	VenueLongitude	VenueCategory
314	Penha de Franca	Miradouro do Monte Agudo	38.726119	-9.131500	Scenic Lookout
315	Penha de Franca	Lidl	38.727171	-9.125546	Supermarket
316	Penha de Franca	Café Esplanada do Miradouro do Monte Agudo	38.726286	-9.131437	Café
317	Penha de Franca	Casinha dos Doces	38.727133	-9.125544	Coffee Shop
318	Penha de Franca	Pizza Hut	38.725426	-9.128492	Pizza Place
319	Penha de Franca	Olha Que Dois	38.723539	-9.129371	Portuguese Restaurant
320	Penha de Franca	Davito	38.722125	-9.129159	Italian Restaurant
321	Penha de Franca	Tebas	38.723779	-9.131739	Coffee Shop
322	Penha de Franca	Fábrica Do Físico	38.724036	-9.128460	Gym
323	Penha de Franca	gelados.come	38.722203	-9.128805	Ice Cream Shop

After creating another map with folium, we get to understand where are the main places in the capital of Portugal:



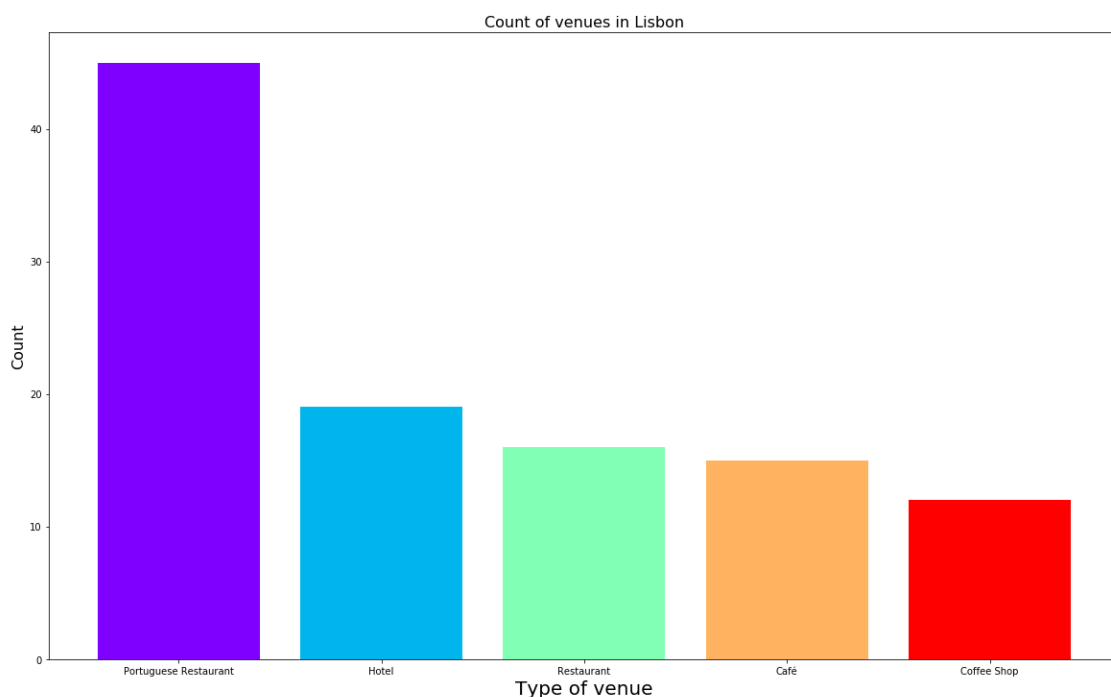
Therefore, the center of the city is where venues are the most, the west side is less garnished.

Now that we have venues information, we will be able to analyze a bit.

```
df2 = nearby_venues['VenueCategory'].value_counts().sort_values(ascending=False).head(5)
df2
```

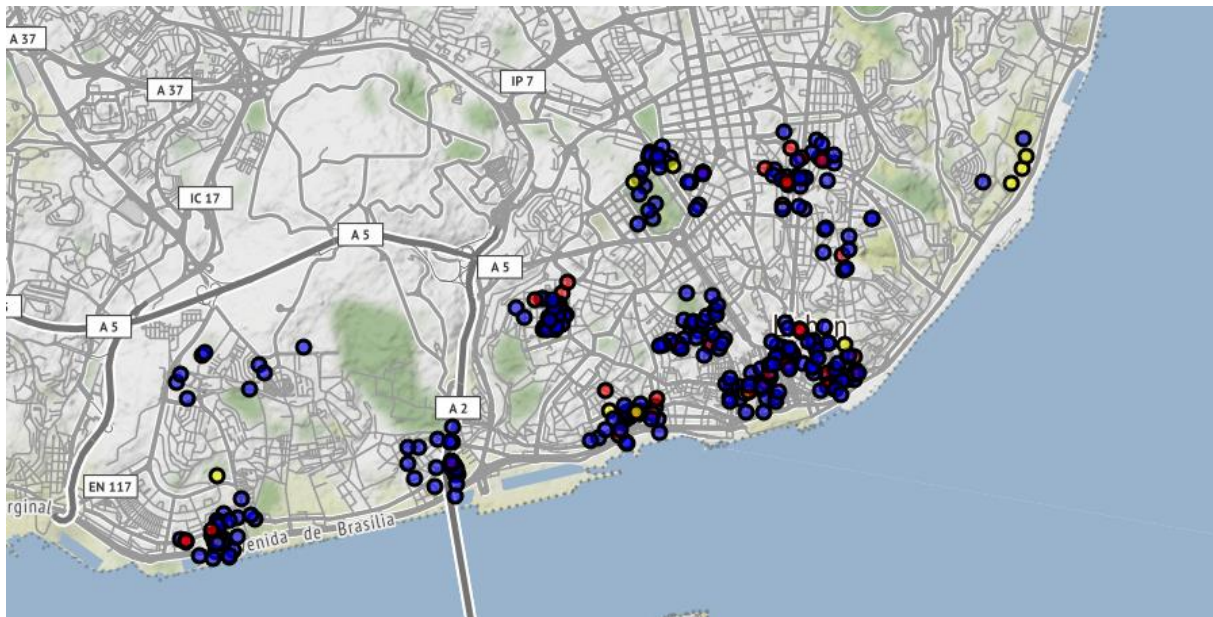
```
Portuguese Restaurant    45
Hotel                    19
Restaurant               16
Café                    15
Coffee Shop              12
Name: VenueCategory, dtype: int64
```

From a value_counts, we get to understand that most of the venues are restaurant or food shop.



The issue become harder because to open a restaurant in a place like this is a real competition where you have to attract people to come to your place. So, if you want to deal with competition, either you create something new, something special, either you try to go where competition is less present. I chose the second one, looking to a place where competitors are fewest because it will let the restaurant to become a unique place and allow it to grow without being confronted to other big places already installed.

What Lisbon looks like in term of restaurants:



What is not blue, represents restaurant, pure restaurant. They are everywhere so the business problem is a real one. To better understand, where to open, we will need to create clusters and see where are the smallest clusters.

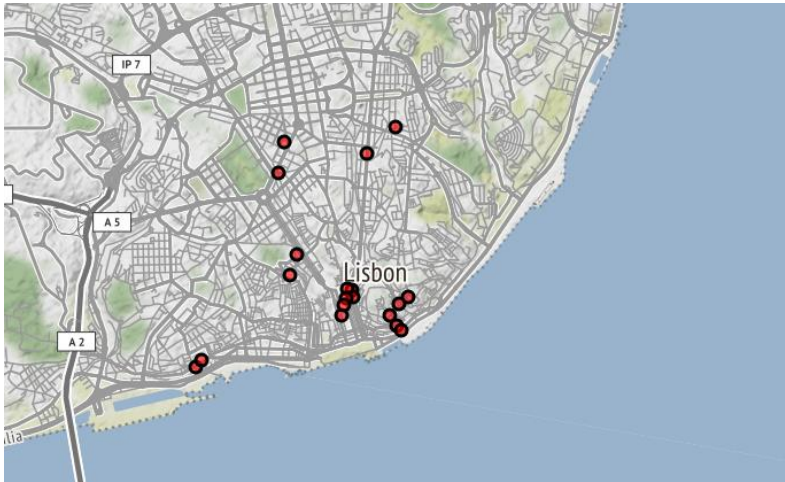
Before, lets see where are an important aspect with a restaurant business are, the hotels. Thanks to hotels, tourists can go to your place, so where are hotels, there are people with the want to eat.

That is a dataframe with hotels only :

```
[99]: lisbon_hotel = nearby_venues.loc[nearby_venues['VenueCategory'] == 'Hotel']
lisbon_hotel
```

```
[99]:
```

	Neighborhood	Venue	VenueLatitude	VenueLongitude	VenueCategory
34	Alfama	Santiago de Alfama	38.711449	-9.131231	Hotel
36	Alfama	Memmo Alfama	38.710304	-9.130199	Hotel
40	Alfama	Hotel convento do Salvador	38.712807	-9.129892	Hotel
49	Alfama	Alfama Lounge Suites	38.713654	-9.128511	Hotel
53	Alfama	Hotel Palacete Chafariz de'l Rei	38.709806	-9.129534	Hotel
68	Mouraria	Hotel Portugal	38.714400	-9.136865	Hotel
73	Mouraria	Tesouro da Baixa	38.714640	-9.137485	Hotel
77	Mouraria	Lisboa Tejo Hotel	38.713620	-9.136734	Hotel
79	Mouraria	The Beautique Hotel	38.713519	-9.137683	Hotel
95	Principe Real	Memmo Principe Real Hotel	38.716161	-9.146148	Hotel
113	Principe Real	Sofitel Lisbon Liberdade	38.718660	-9.145059	Hotel
174	Santos-o-Velho	Janelas Verdes	38.705433	-9.160109	Hotel
188	Santos-o-Velho	Hotel York House	38.706224	-9.159258	Hotel
204	Chiado	Brown 's Hotel Central	38.711454	-9.138375	Hotel
218	Chiado	Hotel Santa Justa	38.712692	-9.138045	Hotel
270	Sao Sebastiao da Pedreira	Sheraton Lisboa Hotel & Spa	38.731753	-9.146995	Hotel
271	Sao Sebastiao da Pedreira	Eurostars Lisboa Parque	38.728153	-9.147768	Hotel
284	Arroios	Vila Garden Guest House	38.730409	-9.134692	Hotel
310	Arroios	My Place - Lisbon Lounge Suite	38.733399	-9.130302	Hotel



When we look at the map, we see that the heart of Lisbon is concentrated with a lot of hotels. That would be a great point to open there because it would allow many people to come. Therefore, we need to see if competitors are not too much in this area. Because even if you have more customers in your area, if you have more competitors, you will not necessarily have a better recipe at the end of the day.

We have to create clusters from restaurant information thanks to the Foursquare API to understand where would be the perfect place to open a restaurant.

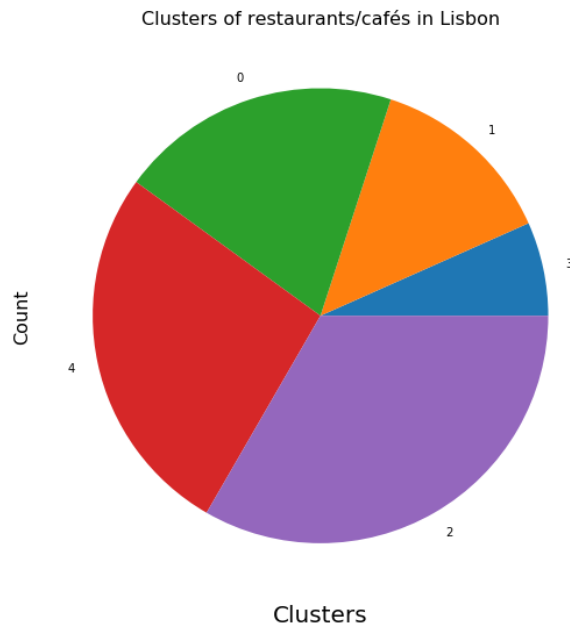
```
array = ['Restaurant', 'Portuguese Restaurant', 'Café']
lisb_rest = nearby_venues.loc[nearby_venues['VenueCategory'].isin(array)]
lisb_rest
```

	Neighborhood	Venue	VenueLatitude	VenueLongitude	VenueCategory
7	Alcântara	1300 Taberna	38.702888	-9.178968	Restaurant
8	Alcântara	Rio Maravilha	38.701798	-9.178076	Portuguese Restaurant
15	Alcântara	Cantina LX Factory	38.703228	-9.178896	Portuguese Restaurant
16	Alcântara	chef nino	38.703059	-9.178802	Café
17	Alcântara	Village Underground	38.700861	-9.178150	Café
...
306	Arroios	Franguinho Real	38.730520	-9.136216	Portuguese Restaurant
308	Arroios	O Telheirinho	38.730796	-9.136367	Portuguese Restaurant
313	Arroios	Restaurante Quionga	38.732946	-9.131793	Portuguese Restaurant
316	Penha de Franca	Café Esplanada do Miradouro do Monte Agudo	38.726286	-9.131437	Café
319	Penha de Franca	Olha Que Dois	38.723539	-9.129371	Portuguese Restaurant

76 rows × 5 columns

With this dataframe, we will be able to have clusters thanks to the kmeans method.

```
kclusters = 5
lisbon_rest_cluster = lisb_rest.drop(['Venue', 'VenueCategory', 'Neighborhood'], 1)
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(lisbon_rest_cluster)
kmeans.labels_[0:10]
array([1, 1, 1, 1, 1, 1, 1, 1, 2, 2], dtype=int32)
```



This is what we obtain after a pyplot creation. We understand that 5 main clusters are present in Lisbon. What interested us is the cluster 3. That is a cluster with few restaurant, so lets see where it is.

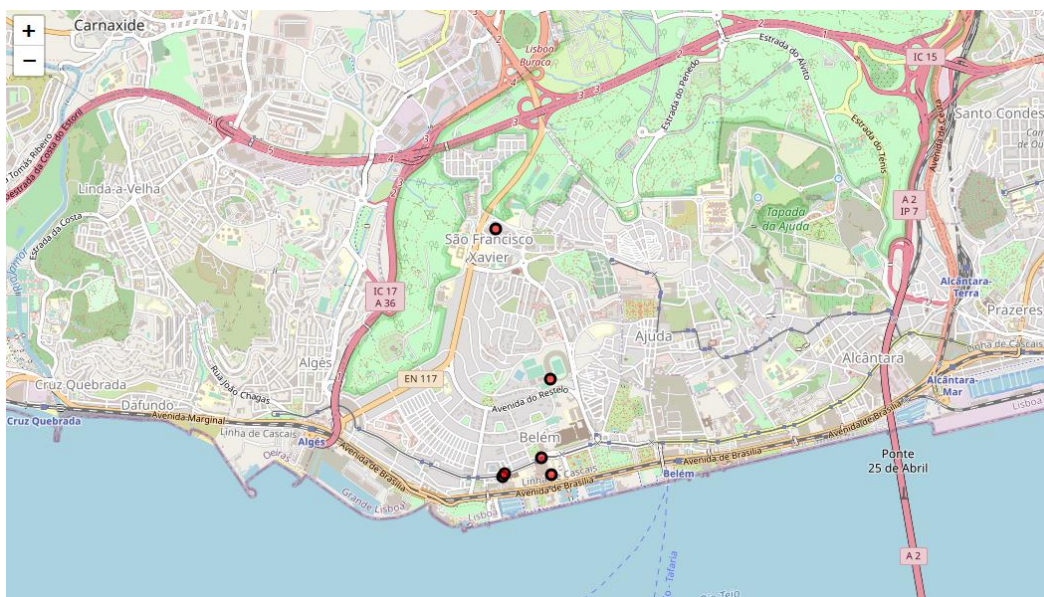
I used another dataframe to select only venues of the cluster:

```
[207]: Cluster_targeted = lisbon_rest_cluster.loc[lisbon_rest_cluster['Cluster Labels'] == 3]
Cluster_targeted
```

```
[207]:
```

	Cluster Labels	VenueLatitude	VenueLongitude
142	3	38.695090	-9.212624
145	3	38.695259	-9.212497
148	3	38.695248	-9.208428
157	3	38.696385	-9.209303
158	3	38.701688	-9.208539
282	3	38.711753	-9.213203

We get a map with few venues. These are all restaurants and we can see that competition is low.



Thanks to the map of the hotel before, we can recommend to a business man to target the cost, near the hotel, not in the heart of Lisbon but in the west side of the city where competition is low.

Discussion:

This analysis can be taken as a beginning of a eventual accurate analysis determining with more than just few factors the way to open a restaurant in a city like Lisbon. In this analysis, the main factor was the concentration of other restaurant in the area but factors as the rent price of areas in the city or the type of restaurant would be more things to add to the analysis and make the results more pertinent.

Conclusion:

The food business is hard when it comes to aim the right target in a place you don't know. Using information on venues in the city allow the understanding of the sector. The heart of Lisbon is already largely taken by restaurants and would be too hard to become a competitor as strong as the place that are already implanted. Therefore, a right choice would be to target a low competitive area to start, because people would come for you and that would allow the business to grow and maybe become unique.