

Data science capstone project report

New Cafe place in Casablanca



By: Ansari Youssef

Casablanca is the largest city of Morocco. Located in the central-western part of Morocco bordering the Atlantic Ocean, it is the largest city in the Maghreb region and the eighth-largest in the Arab world. Casablanca is Morocco's chief port and one of the largest financial centers in Africa. According to the 2014 population estimate, the city has a population of about 3.35 million in the urban area and over 4.27 million in the Greater Casablanca.

The statistics above makes this city perfect for opening a new coffee place!

Introduction:

In the city of Casablanca, if an entrepreneur is looking to open new big cafe place, the question is, where would you recommend that they open it? The background of the problem is that in order for a café to be profitable, there must be enough customers, and in order to have enough customers, it is not worth setting up a café in the immediate proximity of existing ones, as well the location must ideal to access, it should be near the middle of the city and close to the main highway.

Let's also make sure that audience is explicitly defined to be the local restaurant entrepreneurs in Casa and they should care about this problem because the location of the new cafe has a significant impact on the expected returns.

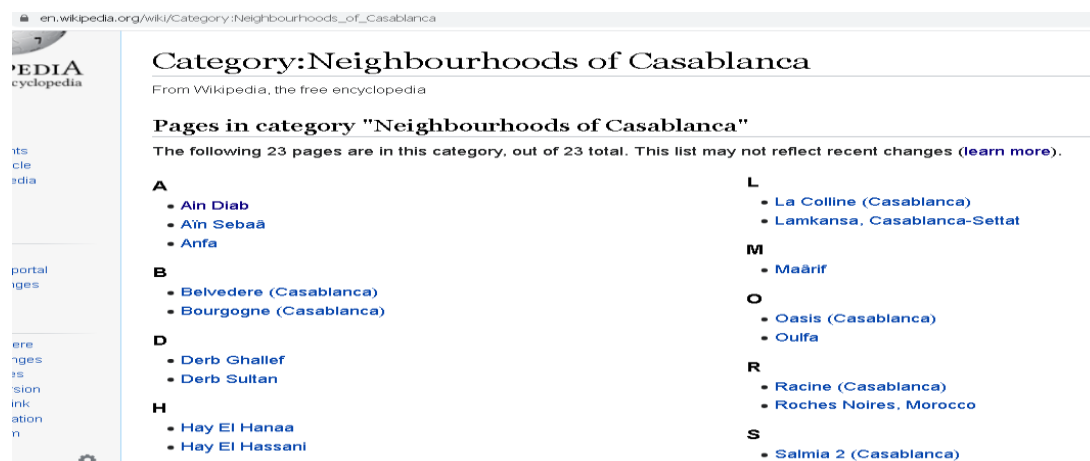
Data Acquisition:

Extracting Data from sources and cleaning it:

1-Casablanca Neighborhoods Data:

We will scrape the names of the neighborhoods of Casablanca from the following Wikipedia page:

https://en.wikipedia.org/wiki/Category:Neighbourhoods_of_Casablanca .



en.wikipedia.org/wiki/Category:Neighbourhoods_of_Casablanca

Category:Neighbourhoods of Casablanca

From Wikipedia, the free encyclopedia

Pages in category "Neighbourhoods of Casablanca"

The following 23 pages are in this category, out of 23 total. This list may not reflect recent changes ([learn more](#)).

A	L
<ul style="list-style-type: none">Ain Diab	<ul style="list-style-type: none">La Colline (Casablanca)
<ul style="list-style-type: none">A'in Sebaā	<ul style="list-style-type: none">Lamkansa, Casablanca-Settat
<ul style="list-style-type: none">Anfa	M
B	<ul style="list-style-type: none">Maarif
<ul style="list-style-type: none">Belvedere (Casablanca)	O
<ul style="list-style-type: none">Bourgogne (Casablanca)	<ul style="list-style-type: none">Oasis (Casablanca)
D	<ul style="list-style-type: none">Oulfa
<ul style="list-style-type: none">Derb Ghallef	R
<ul style="list-style-type: none">Derb Sultan	<ul style="list-style-type: none">Racine (Casablanca)
H	<ul style="list-style-type: none">Roches Noires, Morocco
<ul style="list-style-type: none">Hay El Hanaa	S
<ul style="list-style-type: none">Hay El Hassani	<ul style="list-style-type: none">Salmia 2 (Casablanca)
<ul style="list-style-type: none">Hay Salama	

2-Foursquare Data:

We will get the location data of each neighborhood using the foursquare API, it would look like this:

Out[12]:

	Neighborhood	Latitude	Longitude
0	Ain Diab,Casablanca	33.592288	-7.676419
1	Ain Sebaâ,Casa	33.608591	-7.524144
2	Anfa,Casa	33.592632	-7.672177
3	Belvedere (Casablanca),Casa	33.591293	-7.590353
4	Bourgogne (Casablanca),Casa	33.597893	-7.638162
5	Derb Ghallef,Casa	33.570614	-7.630057
6	Derb Sultan,Casa	33.572170	-7.595850
7	Hav El Hana.Casa	33.577079	-7.666133

We'll use the API once again to get the venues of each neighborhood; we end up with the following data frame:

NearbyVenues							
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Ain Diab,Casablanca	33.592288	-7.676419	Basmane Restaurant	33.591577	-7.677553	Moroccan Restaurant
1	Ain Diab,Casablanca	33.592288	-7.676419	Le Boudoir	33.593305	-7.677625	Lounge
2	Ain Diab,Casablanca	33.592288	-7.676419	Sinatra	33.595661	-7.677022	Italian Restaurant
3	Ain Diab,Casablanca	33.592288	-7.676419	Sun Beach - CCC	33.595406	-7.678437	Pool
4	Ain Diab,Casablanca	33.592288	-7.676419	Azur Hotel Casablanca	33.595360	-7.676824	Hotel
5	Ain Diab,Casablanca	33.592288	-7.676419	Hotel Val D'Anfa	33.591664	-7.678328	Hotel
6	Ain Diab,Casablanca	33.592288	-7.676419	Hotel Club Val d'Anfa	33.591684	-7.678351	Hotel
7	Ain Diab,Casablanca	33.592288	-7.676419	Armstrong Legend	33.595402	-7.677080	Jazz Club
8	Ain Diab,Casablanca	33.592288	-7.676419	Le Squat	33.593145	-7.677817	Nightclub

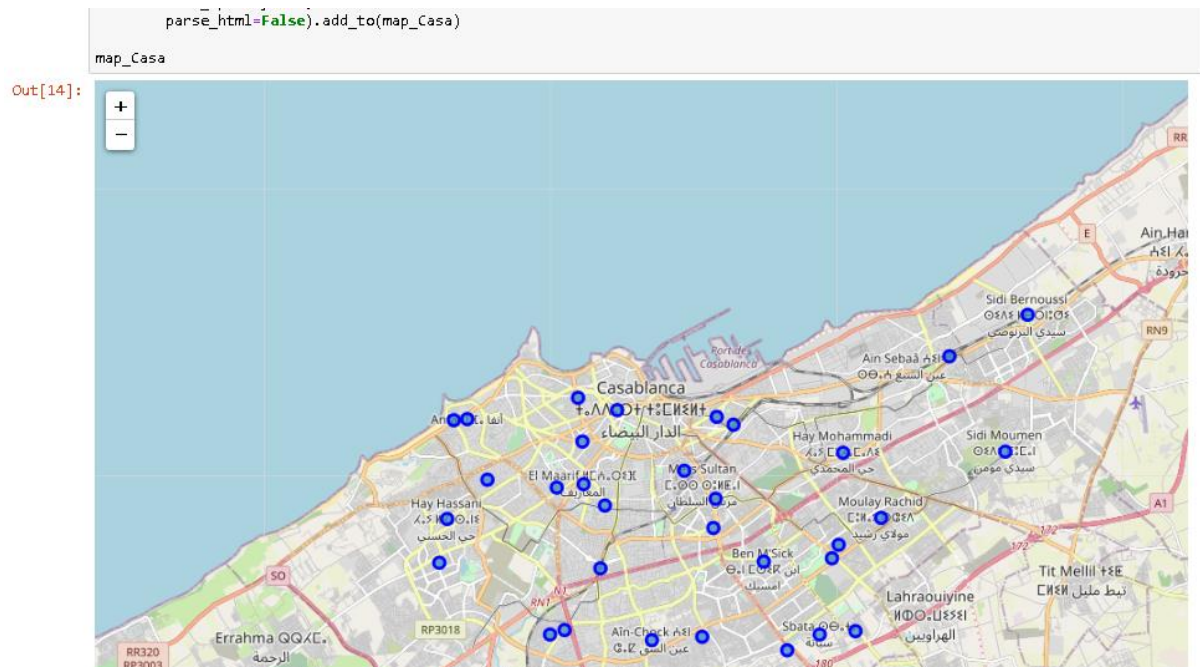
We'll use this data frame to draw maps that will help us choose an ideal place to open our new coffee place, based on the locations of other coffees and their areas of influence.

Using maps is ideal for this problem because we're going to be able to visualize the other coffee places, their influence area, and helps us choose an ideal place in the middle of the city that is close to the main route, because we want our new coffee place to be easy to get to.

Methodology:

After getting the data, we will begin our analysis:

We'll start with drawing a map of the city with the neighborhoods on it; we will do this to make sure we covered all of the big neighborhoods of the city in our data.



This looks good; all of our neighborhoods are correctly in place so our data was perfectly cleaned and prepared.

Now we'll get the important nearby venues of each neighborhood into a new data frame:

NearbyVenues							
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Ain Diab,Casablanca	33.592288	-7.676419	Basmane Restaurant	33.591577	-7.677553	Moroccan Restaurant
1	Ain Diab,Casablanca	33.592288	-7.676419	Le Boudoir	33.593305	-7.677625	Lounge
2	Ain Diab,Casablanca	33.592288	-7.676419	Sinatra	33.595661	-7.677022	Italian Restaurant
3	Ain Diab,Casablanca	33.592288	-7.676419	Sun Beach - CCC	33.595406	-7.678437	Pool
4	Ain Diab,Casablanca	33.592288	-7.676419	Azur Hotel Casablanca	33.595360	-7.676824	Hotel
5	Ain Diab,Casablanca	33.592288	-7.676419	Hotel Val D'Anfa	33.591664	-7.678328	Hotel
6	Ain Diab,Casablanca	33.592288	-7.676419	Hotel Club Val d'Anfa	33.591684	-7.678351	Hotel
7	Ain Diab,Casablanca	33.592288	-7.676419	Amstrong Legend	33.595402	-7.677080	Jazz Club
8	Ain Diab,Casablanca	33.592288	-7.676419	Le Squat	33.593145	-7.677817	Nightclub

Let's count how many venues in each of the neighborhoods:

```
: NearbyVenues[['Neighborhood', 'Venue']].groupby('Neighborhood').count()
```

```
:
```

Venue	
Neighborhood	
Ain Diab,Casablanca	42
Al Fida casablanca	6
Anfa,Casa	20
Ain Sebaâ,Casa	4
Belvedere (Casablanca),Casa	5
Bourgogne (Casablanca),Casa	6
Derb Ghallef,Casa	5
Derb Sultan,Casa	3
Habbous,Casa	8
Hay El Hana,Casa	6
Hay El Hassani,Casa	5
Hav Salama,Casablanca	4

The above data frame helps us recognize the popular neighborhoods in the city.

Let's see how many venues we have per venue category:

```
NearbyVenues['Venue Category'].value_counts()
```

Café	52
Hotel	19
Fast Food Restaurant	16
Italian Restaurant	15
Coffee Shop	10
Pizza Place	8
Diner	8
Lounge	6
Restaurant	6
Bakery	6
Shopping Mall	5
Sandwich Place	5
Nightclub	5
Pub	5
Moroccan Restaurant	5
French Restaurant	5
Sushi Restaurant	5
Middle Eastern Restaurant	4
Tram Station	4
Pool Hall	3

We notice that Café category is dominating, that's way choosing an ideal place for the new coffee place is a must!

Let's create a data frame of the top 10 venues for each neighborhood:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Ain Diab,Casablanca	Café	Hotel	Nightclub	Fast Food Restaurant	Restaurant	Lounge	Italian Restaurant	Tea Room	Cocktail Bar	Doner Restaurant
1	Al Fida casablanca	Fast Food Restaurant	Department Store	Flea Market	Supermarket	Soccer Field	General Entertainment	Eastern European Restaurant	Fish & Chips Shop	Food	Food Court
2	Anfa,Casa	Café	Italian Restaurant	Lounge	Nightclub	Yoga Studio	Hotel	Hookah Bar	Middle Eastern Restaurant	Fast Food Restaurant	Pub
3	Ain Sebaâ,Casa	Diner	Café	Coffee Shop	Italian Restaurant	BBQ Joint	Gym	Flea Market	Food	Food Court	French Restaurant
4	Belvedere (Casablanca),Casa	Café	Fast Food Restaurant	Hotel	Tram Station	Pizza Place	Yoga Studio	Fish & Chips Shop	Flea Market	Food	Food Court
5	Bourgogne (Casablanca),Casa	Middle Eastern Restaurant	Pizza Place	Bakery	Shopping Mall	Café	Yoga Studio	General Entertainment	Fish & Chips Shop	Flea Market	Food
6	Derb Ghallef,Casa	Print Shop	Café	Fast Food Restaurant	Flea Market	Gym	Eastern European Restaurant	Fish & Chips Shop	Food	Food Court	French Restaurant
7	Derb Sultan,Casa	Fast Food Restaurant	Pharmacy	Tea Room	Yoga Studio	General Entertainment	Fish & Chips Shop	Flea Market	Food	Food Court	French Restaurant

The above data frame above gives us some insights, For example: Aid Diab is a neighborhood where opening a new coffee place is definitely a bad idea, since the most common venue in this neighborhood are Coffee places!

Since we are interested in Coffee places, let’s create a new data frame that contains this category only:

```
cafe_places=NearbyVenues[NearbyVenues['Venue Category'].isin(('Café','Coffee Shop'))].reset_index(drop=True)
cafe_places
```

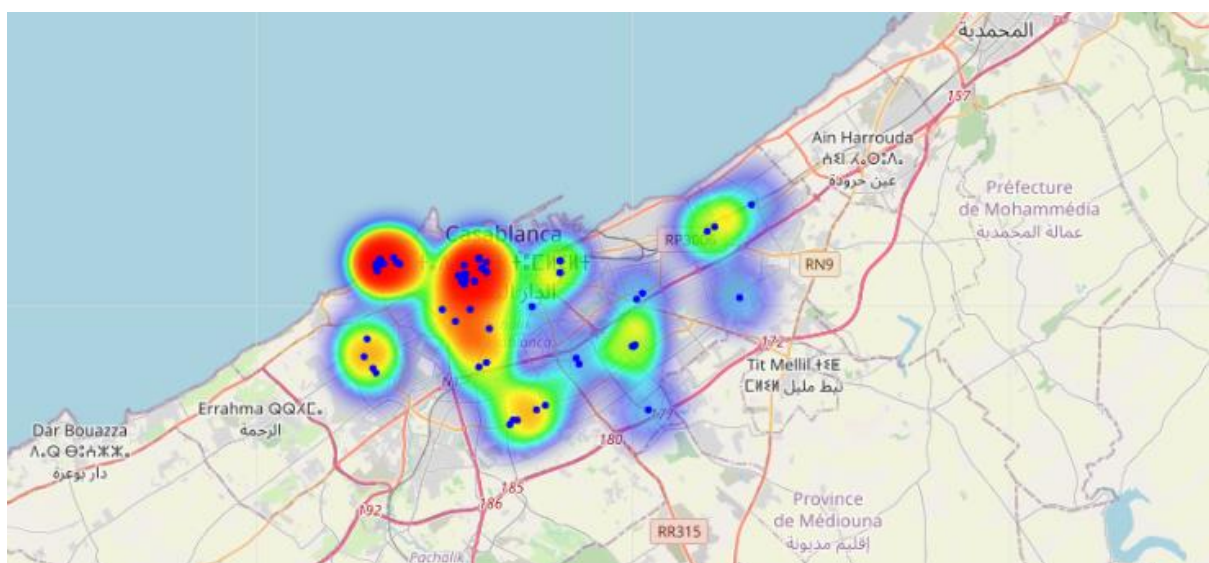
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Ain Diab,Casablanca	33.592288	-7.676419	Le Crépuscule	33.593492	-7.679094	Café
1	Ain Diab,Casablanca	33.592288	-7.676419	Gossip Beach	33.594370	-7.678767	Café
2	Ain Diab,Casablanca	33.592288	-7.676419	Cafe Calypso	33.593473	-7.678322	Café
3	Ain Diab,Casablanca	33.592288	-7.676419	Hooka Blanca	33.591470	-7.678073	Café
4	Ain Diab,Casablanca	33.592288	-7.676419	Tropicana	33.596276	-7.676693	Café
5	Ain Diab,Casablanca	33.592288	-7.676419	XS Lounge Casablanca	33.594576	-7.675280	Café
6	Ain Diab,Casablanca	33.592288	-7.676419	Tropicana Terrasse	33.596304	-7.676723	Café
7	Ain Sebaâ,Casa	33.608591	-7.524144	CAFE ADANA	33.608835	-7.520845	Café
8	Ain Sebaâ,Casa	33.608591	-7.524144	Half Moon	33.607218	-7.524362	Coffee Shop
9	Anfa,Casa	33.592632	-7.672177	Roosevelt Café	33.594718	-7.667885	Café
10	Anfa,Casa	33.592632	-7.672177	Theo'reme Beach	33.594801	-7.667846	Café
11	Anfa,Casa	33.592632	-7.672177	XS Lounge Casablanca	33.594576	-7.675280	Café
12	Anfa,Casa	33.592632	-7.672177	Hediard Café	33.595128	-7.669210	Café
13	Anfa,Casa	33.592632	-7.672177	Coffeeshop Ain Diab	33.596760	-7.670433	Café
14	Belvedere (Casablanca),Casa	33.591293	-7.590353	Adi Ice Café	33.590917	-7.592887	Café
15	Bourgogne (Casablanca),Casa	33.597893	-7.638162	Lugano	33.594135	-7.637739	Café
16	Derb Ghallef,Casa	33.570614	-7.630057	Brioche Dorée	33.569532	-7.625870	Café
17	Hay El Hassani,Casa	33.567046	-7.678489	café zima	33.565235	-7.683097	Café
18	Hay Salama,Casablanca	33.560505	-7.558101	Le Secret	33.562695	-7.558556	Café
19	Hay Salama,Casablanca	33.560505	-7.558101	Brumm's	33.562305	-7.558053	Café

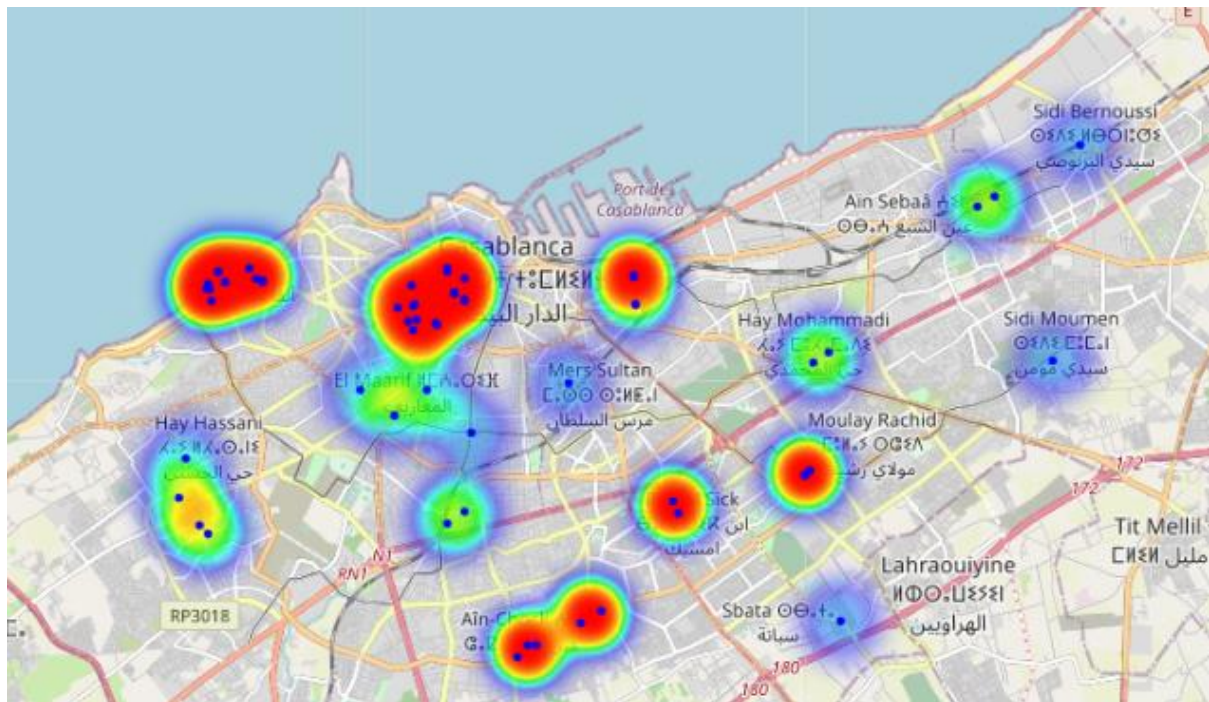
Results:

We start by drawing a map of the `café_places` data frame that contains major coffee places in the city:



There are Many Coffee places all over the city! Let's create a heat map to make it easier for us to visualize the areas with existing coffee places as well as their influence; this will help us choose the ideal location for opening our new coffee place!



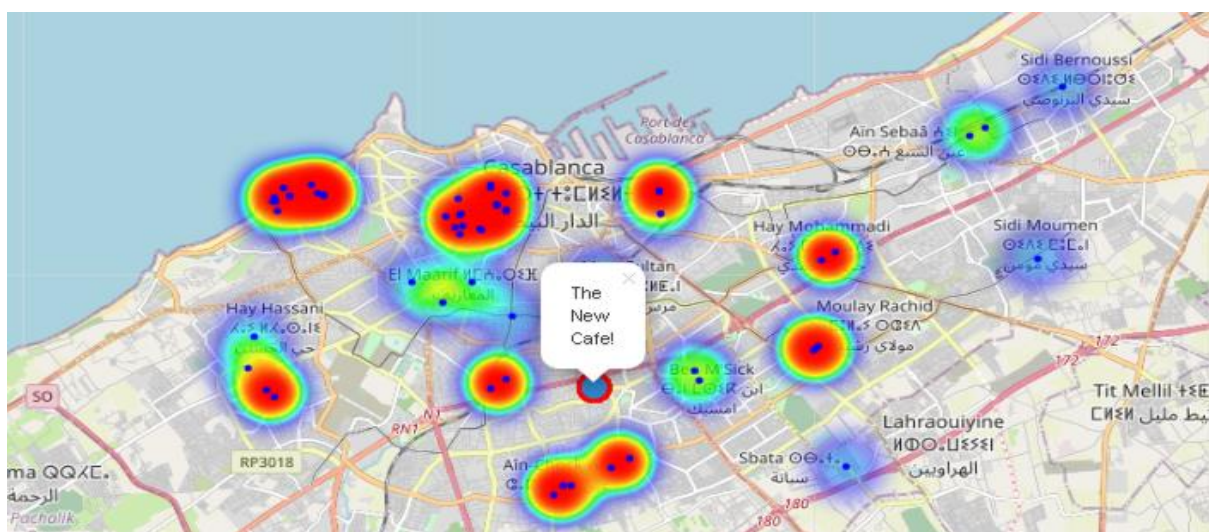


Discussion:

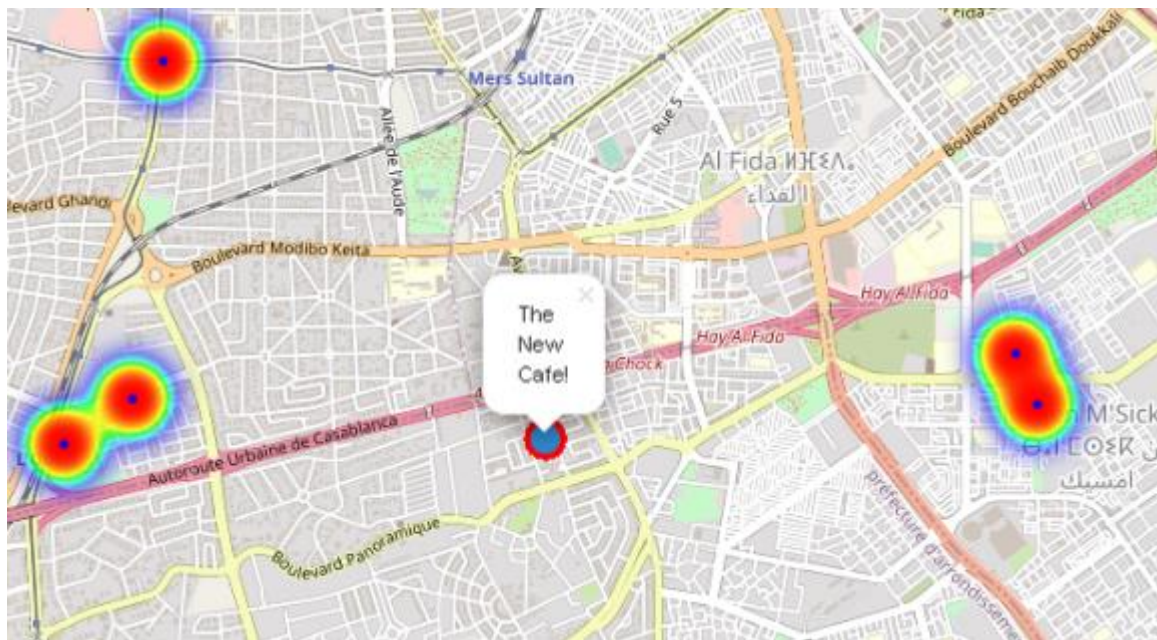
Judging from the heat map, we can tell that neighborhoods that are close to the beach and in the middle of the city as well aren't ideal to open our new coffee place, because they are full, on the other hand we can see that some areas in the middle of the city between heated areas are available for us to choose from.

For example, a great location would be: Ain-Chok, Boulevard Panoramique ,

Since it is in the middle of the city , far away from the heated areas , and close to the main high way of the city , a bonus that costumers , we can see the location in the following map :



We can see how close it is to the highway in the following map:



The highway is the red route!

There are other possible locations, but based on the analysis we did, this one meets our criteria best!

Conclusion:

In this project, we worked on data from multiple sources, merged it together and gotten valuable insights from it. We did meet the goal that we set at the start, that is finding an ideal place for the new coffee place. After our descriptive analysis, we used mainly the library folium to generate maps from our data in order for us to visualize it and choose the idea location.

This project shows how location data could be valuable for entrepreneurs when starting a new business, this is just an example, and the problems that could be solved using this technique are limitless.