

# Analysis of Pharmacies Locations in Bucharest

by  
Emilian Zaman  
21.10.2019

## 1. Introduction – Business Problems

In this project we will try to identify the best locations to open a new pharmacy in Bucharest, Romania. This town is crowded with pharmacies, so it will be quite challenging to find a good spot that will not be surrounded with lots and lots of this kind of stores on a 500 meters radius.

We will use our data science skills to try to find the best locations in this city also based on a proximity criteria to the city center

## 2. Data

I will use exclusively google api, to extract locations of all pharmacies in the Bucharest. I've tried foursquare in other projects and it has not all the venues in Bucharest, so this will be a real problem when we are talking about accuracy of data.

I will define circles with 500m radius and I will extract all pharmacies in those circles by using each of those circle centers as a point of origin in google api query. In this way I will compose a pandas data frame with locations (latitude and longitude) of all pharmacies.

This is a sample of the data frame:

Name of the pharmacy	Google Rating	Latitude	Longitude
Belladonna	3.1	44.479836	26.100635
Sensiblu	3.5	44.478684	26.102669

We have a total of 642 rows and 4 futures. We have the name of location, google rating and latitude and longitude.

Fortunately the data comes very clean from parsing the json file so we don't need to do any cleaning.

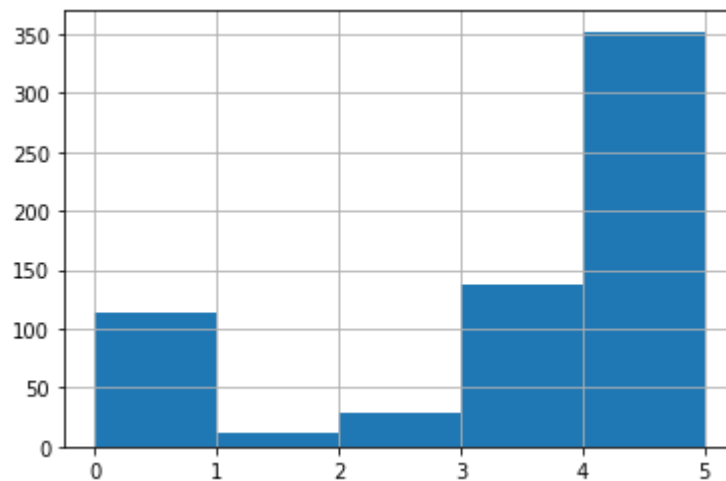
## 3. Methodology

In the first step we gathered the data by using google api and after that parsing the json file. After we will plot a heat map to see if we spot some empty spaces around city center. Next we will try to identify a few promising areas we will focus on those to determine the best spot for a future place of our investor pharmacy.

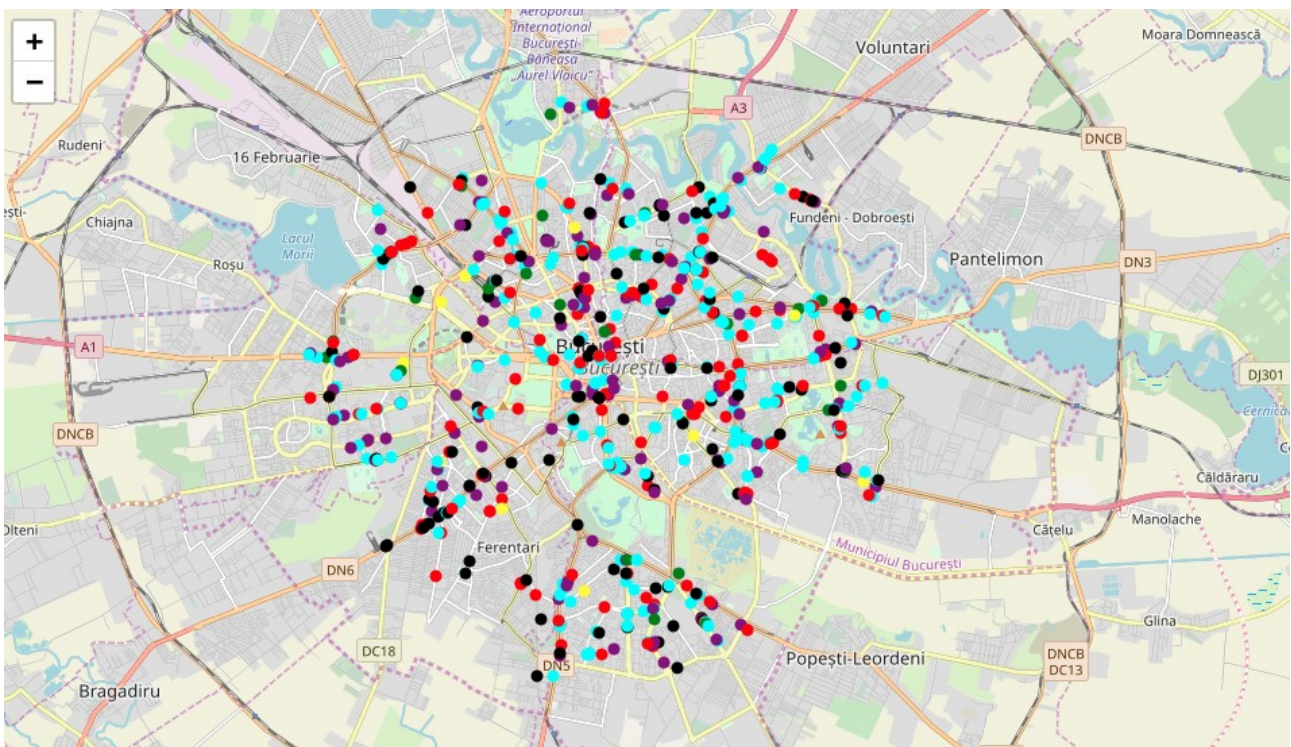
#### 4. Exploratory Data Analysis

First we know that we have 642 places in entire city. At first it seems quite an average number. The city has a population of 1.883.425 in 2016 according to wikipedia, so we have one pharmacy for 2933 people. For comparison in Paris we have one per 2200 of inhabitants, so 642 pharmacies in Bucharest are quite a lot. But we will still going to find if there are any places not covered by this 642.

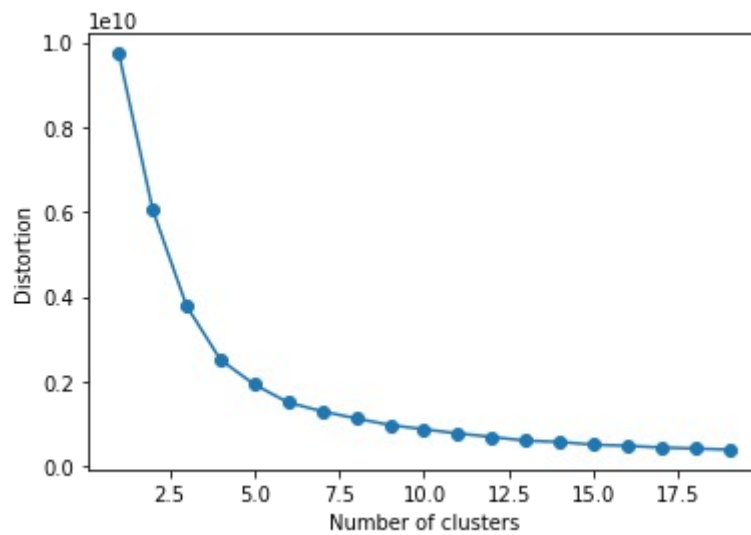
Second we draw a histogram to see how the rating is distributed. The conclusion is that most of them are 5 stars rated, second place with 4 stars rated, third place no rating. Poor rating like one or two stars are insignificant.



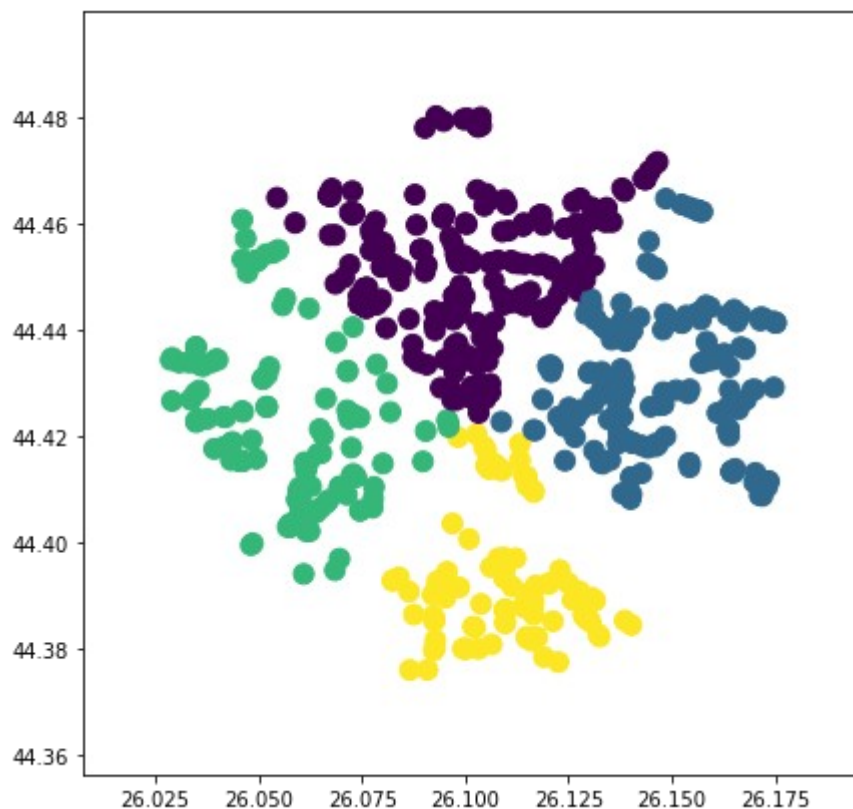
After that we will plot the position on the map using folium



As you can see, besides green area where are parks or nature reserve mostly everything is covered. We will try to do some clustering to see it better. First I tried Kmeans.



I determine the optimum number of clusters is 4 and this is the result:

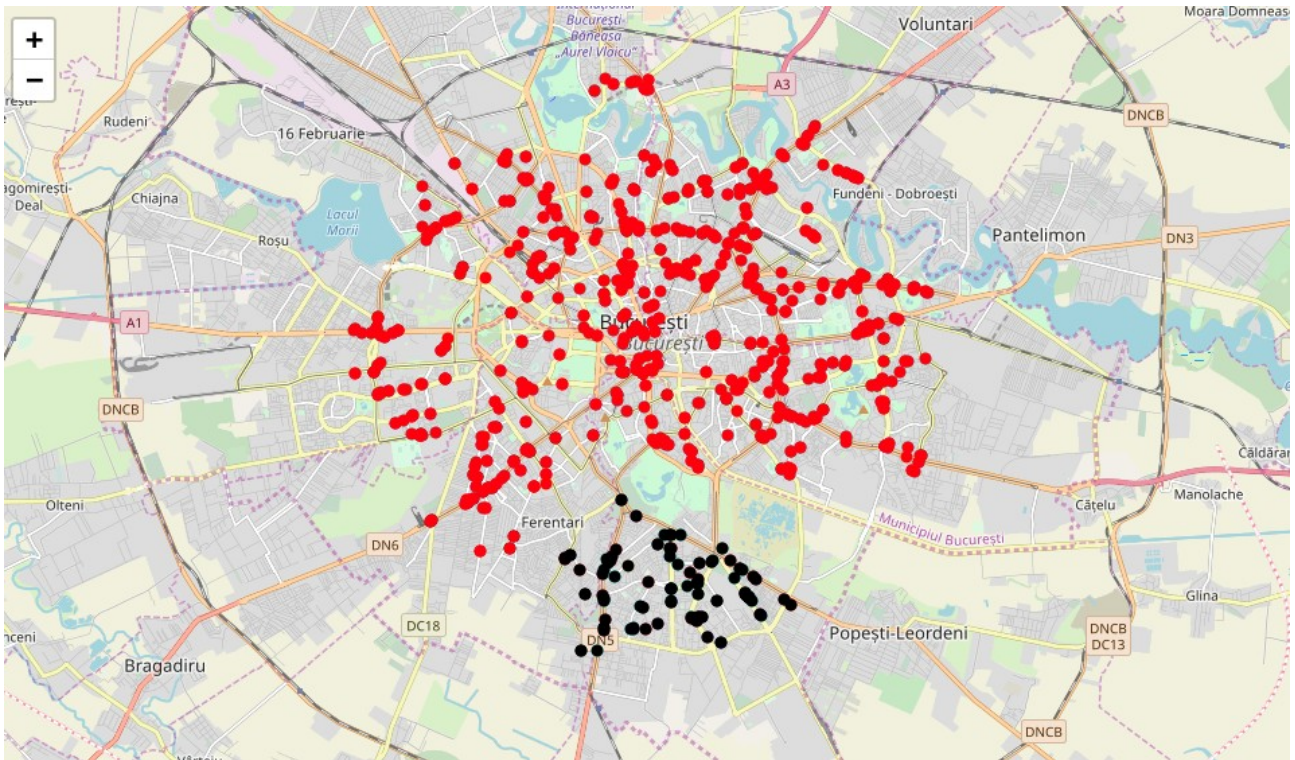


I wasn't very happy with the result so I tried something from DBSCAN family, like HDBSCAN.

DBSCAN algorithms are nicer than Kmeans because you don't have to use any empirical methods to determine the number of cluster, like elbow method. It finds by himself the number of clusters. The tradeoff is in computational power it is more intensive than Kmeans. But in our case we have a

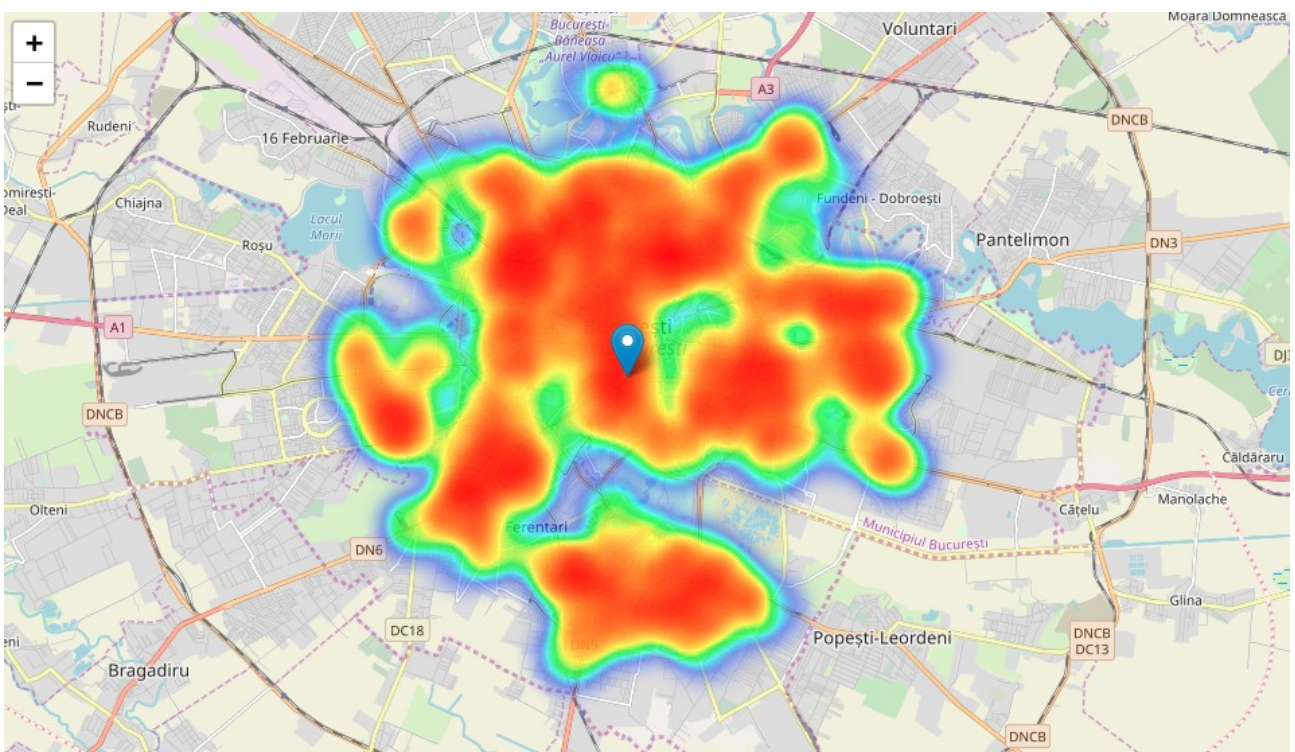


small dataset, so we can use without any problem the DBSCAN and for this particular situation we will use HDBSCAN which performs better for geo-location data.



This is the result. The space between clusters is formed by one park and one natural reserve in the east and in the west by Ferentari Neighborhood which is a pretty tough neighborhood in Bucharest, this seems to be the major reason for the lack of pharmacies in this area. Our job is to determine the blank spots near the center so we will investigate further.

To do this we will proceed of using a heatmap representation in folium:

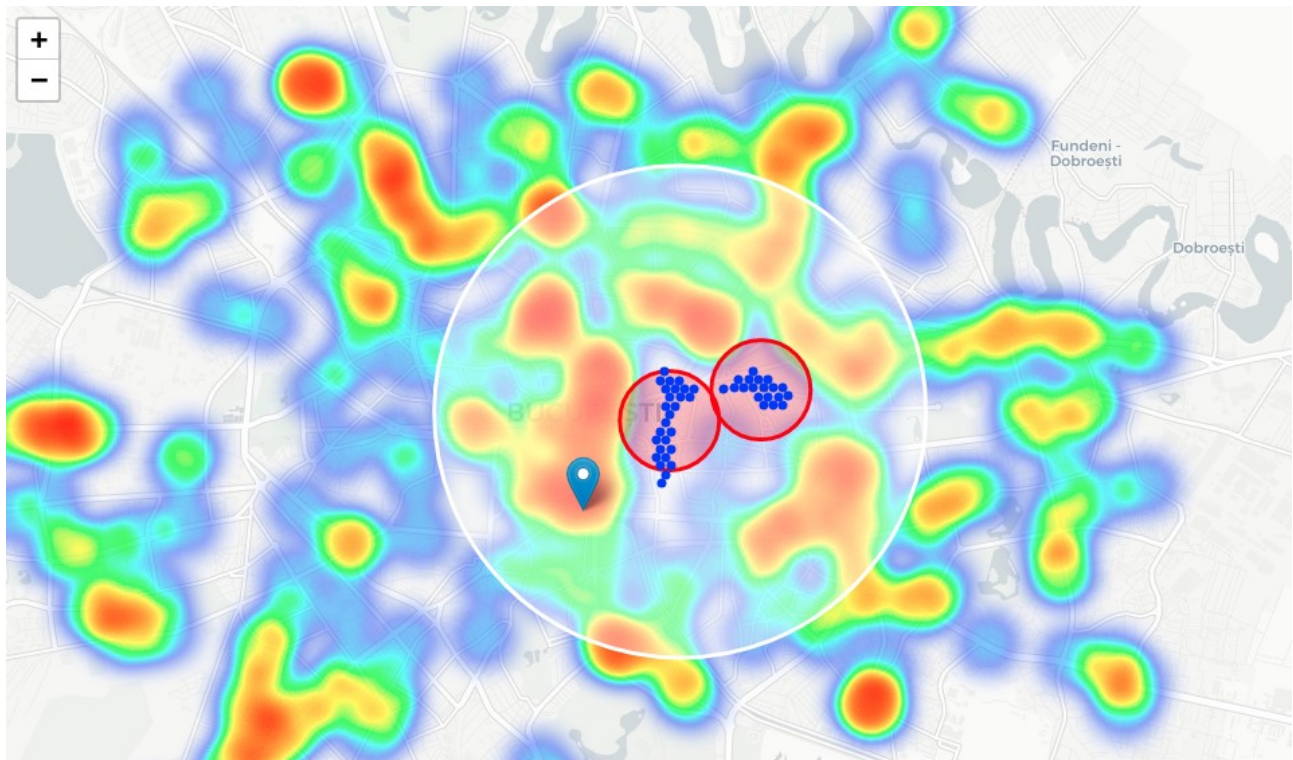




We spot an entire empty area in the east of the city center. We will investigate this one. We define a circle of 1500m radius around this area and we will try to find all the spots inside this area from whom in 500m we will not find any pharmacy.

To do this we define circle areas, each of them with 1km in diameter. For each area we will find how many pharmacies are inside.

After we find the spots we plot them and we try to determine the number of cluster, this time using Kmeans.



## 5. Conclusions

We identified 46 potential spots to place a new pharmacy. Those two red circles have centers with following addresses:

Strada Alexandru Romano 4, București 030167 => 1.3km from Bucharest Center

Strada Rondă 8, București 030167 => 2.2km from Bucharest Center

I think this is a pretty good result. We found two areas, closed to city center, one by 1.3km and the other by 2.2km. We propose to start the search for potential locations to rent or buy from those two addresses above.

## 6. Future directions

One step for improvement is a price classification of those pharmacies, like places with low prices, medium prices and high prices. This could be done by scraping the pharmacies web sites and gather informations about products prices. In this way we can see what is the economic power of the neighborhoods and what kind of price will be implemented in future pharmacy.