Choosing the best place for new coffe shop

Aleksey Drabovich

March 12 2020

1. Introduction

1.1 Background

If you want to open a coffee shop, you will encounter many difficulties. Among them, choose the best place for this. It is not a secret that the best place for new coffe shop is where a lot of people are and lack of places to get coffe. There are many different location data providers, but how not to get confused in a lot of information? Can machine learning be useful to complete this non-trivial task?

1.2 Problem

This project aims to predict the best place for coffe shop based on location data. In this project we'll try to supply business owner with such insights, as:

- Get a visual representation of venues distribution in the selected city.

- See how they are grouped into clusters in places of particular popularity

- Determine the centers of these clusters.

- For each cluster you calculate how many coffe-shops are already there.

1.3 Interest

Obviously coffee shop keepers and businessmen who want to open their own coffee shop will be interested in the results of this project.

2. Data source

In this document we will use Foursquare location data. Foursquare location data is data describing places and venues, such as their geographical location, their category, working hours, full address, and so on, such that for a given location given in the form of its geographical coordinates (or latitude and

longitude values) one is able to determine what types of venues exist within a defined radius from that location.

Using the Foursquare API, we can search for specific type of venues or stores around a given location. And for a given location you will be able to tell how many of each venue category exist and how each surrounding venue is reviewed by other people.

As parameters for building the model, geographical coordinates of Brest city venues were selected within a radius of 3000 meters from its center. Such a distance made it possible to cover the entire historical center and the surrounding areas, potentially interesting for placing coffee shops.

Using the parameters indicated above, we were able to obtain information about 54 places of rest. This is a fairly small amount of data, since Forsquare is not very popular in Belarus. Nevertheless, this study will be useful as additional information for starting a new business.

3. Exploratory Data Analysis

Having visualized the distribution of venues on the map, the regularities in their distribution are not obvious. Such visualization does not provide an understanding of the category of venues and the centers around which they are placed.
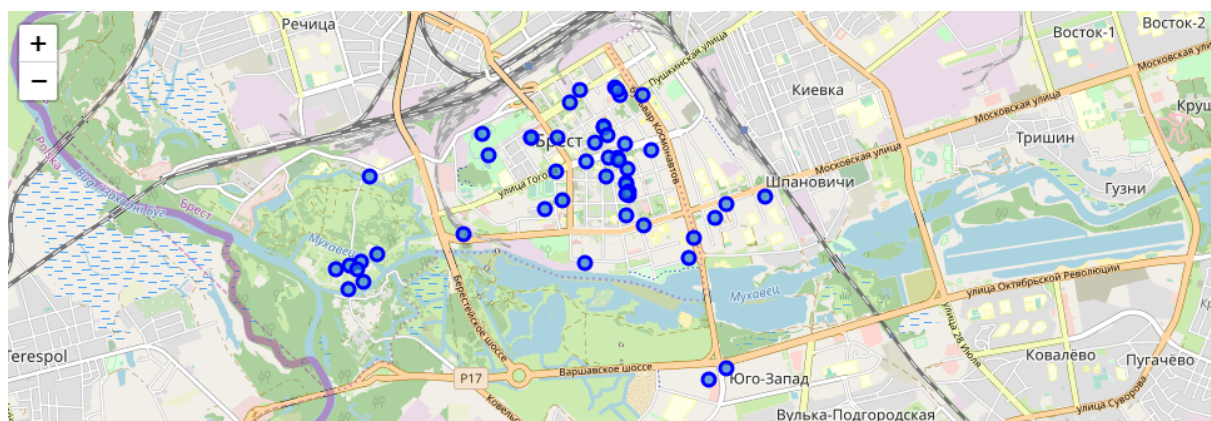

Our task now is to determine the centers of congestion and determine the number of existing coffee shops around these centers. A large number of venues with a small number or absence of coffee shops will inform us of a potentially advantageous placement. But firs we'll create a feature dataset with a row for aech venue with columns Venue Latitude and Venue Longitude:

	Venue Latitude	Venue Longitude
0	52.093937	23.681345
1	52.097556	23.693544
2	52.091322	23.694616
3	52.091152	23.684870
4	52.094807	23.691402

Now we fit the K-means model with this dataset. K-Means is a type of partitioning clustering, that is, it divides the data into K non-overlapping subsets or clusters without any cluster internal structure or labels.We will take K=10 as Brest is not a big city. It will give us not very big clusters with high venues density. As a result for each venue we will get its cluster label and for each cluster we will get its center:

	Labels	Center Latitude	Center Longitude
0	0	52.095343	23.692038
1	1	52.083799	23.657573
2	2	52.074038	23.706963
3	3	52.115494	23.687284
4	4	52.094699	23.683584
5	5	52.086793	23.706993
6	6	52.112134	23.666705
7	7	52.090093	23.694585
8	8	52.087809	23.685638
9	9	52.090892	23.674078

So now that we have the labels and their centers data generated and the KMeans model initialized, let's plot them and see what the clusters look like.
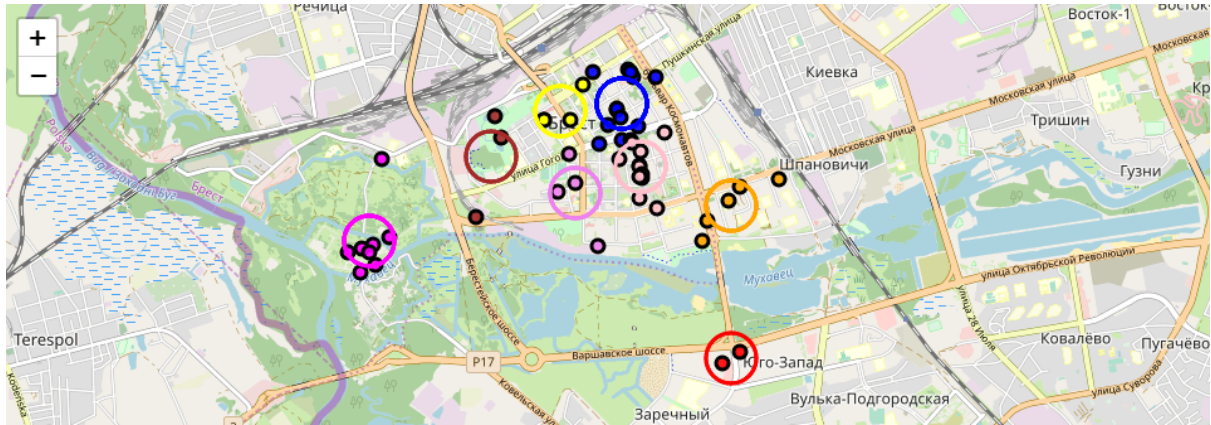
This time the map looks more informative. Clusters became apparent and cluster centers are circled by spherical lines.Now we will calculate total number of venues in the cluster, number of coffe shops and cofe shop - venue ratio, wich will tell us the level of competition

Labels	Coffee Shop	Other	Total	Ratio
0	3	9	12	0.33
1	0	8	8	0.00
2	0	2	2	0.00
3	0	2	2	0.00
4	0	4	4	0.00
5	0	5	5	0.00
6	0	1	1	0.00
7	4	8	12	0.50
8	1	3	4	0.33
9	0	3	3	0.00

As a result we will print out beautiful report. Such report, as well as cluster centers location will be perfect start point for a businessmen who want to open new coffe shop and looking for the best place for it.

```
              Cluster 0                              Cluster 1                              Cluster 2                              Cluster 3
*************************************      *************************************      *************************************      *************************************
            venue  freq                             venue  freq                             venue  freq                             venue  freq
0       Coffee Shop  0.23                0        Historic Site  0.50          0       Gourmet Shop  0.25        0   Belarusian Restaurant  0.5
1   Asian Restaurant  0.08               1       History Museum  0.25          1        Shoe Store  0.25        1                   Hotel  0.5
2           Garden  0.08                 2              Church  0.12           2               Café  0.25       2           National Park  0.0
3             Park  0.08                 3   Monument / Landmark  0.12         3  Other Great Outdoors  0.25     3               Hotel Bar  0.0
4     National Park  0.08                4      Asian Restaurant  0.00         4     Asian Restaurant  0.00      4       Italian Restaurant  0.0
-----------------------------------      -----------------------------------      -----------------------------------      -----------------------------------
   Coffee Shop  Other  Total  Ratio         Coffee Shop  Other  Total  Ratio         Coffee Shop  Other  Total  Ratio         Coffee Shop  Other  Total  Ratio
0           3     10     13    0.3        1           0      8      8    0.0        2           0      4      4    0.0        3           0      2      2    0.0
*************************************      *************************************      *************************************      *************************************
```

```
              Cluster 4                              Cluster 5                              Cluster 6                              Cluster 7
*************************************      *************************************      *************************************      *************************************
            venue  freq                             venue  freq                             venue  freq                             venue  freq
0             Park  0.17                 0              Hotel  0.5            0      Shopping Mall  1.0          0              Bus Stop  1.0
1             Pool  0.17                 1       Shopping Mall  0.5          1   Asian Restaurant  0.0          1      Asian Restaurant  0.0
2            Hotel  0.17                 2    Asian Restaurant  0.0          2              Hotel  0.0          2  Other Great Outdoors  0.0
3       Coffee Shop  0.17                3          Shoe Store  0.0          3         Shoe Store  0.0          3    Italian Restaurant  0.0
4            Plaza  0.17                 4               Road  0.0           4               Road  0.0          4 Mediterranean Restaurant  0.0
-----------------------------------      -----------------------------------      -----------------------------------      -----------------------------------
   Coffee Shop  Other  Total  Ratio         Coffee Shop  Other  Total  Ratio         Coffee Shop  Other  Total  Ratio         Coffee Shop  Other  Total  Ratio
4           1      5      6    0.2        5           0      2      2    0.0        6           0      1      1    0.0        7           0      2      2    0.0
*************************************      *************************************      *************************************      *************************************
```

```
              Cluster 8                              Cluster 9
*************************************      *************************************
                       venue  freq                           venue  freq
0                 Coffee Shop  0.31     0  Eastern European Restaurant  0.33
1                   Steakhouse  0.08    1                      Museum  0.33
2   Modern European Restaurant  0.08    2                        Park  0.33
3          Gym / Fitness Center  0.08   3        Other Great Outdoors  0.00
4                         Gym  0.08     4          Italian Restaurant  0.00
-----------------------------------      -----------------------------------
   Coffee Shop  Other  Total  Ratio         Coffee Shop  Other  Total  Ratio
8           4      9     13    0.44       9           0      3      3    0.0
*************************************      *************************************
```

4. Results

As noted above, Brest is not a big city. Foresquare data is not enough for a full analysis to open a new cafe. However, the results helped determine the best place. According to this study, the best cluster for a cafe is cluster 0 - with the largest number of venues and a relatively low density of cafe shops. Clusters 2 and 4 with a low density of coffee shops and with thematically suitable venues in it may also be promising.

5. Future directions

Of course, we did not conduct a full analysis. The Foursquare dataset is much more powerful. We could additionally take user ratings into account in order to understand the level of competitors. In addition, we could make an analysis of the clusters and determine their thematic affiliation - sports, cultural, etc. Let us leave this for future versions.

6. Conclusions

In this study, I analized location data that could be helpful fo those who want to open their firs caffe shop as well as for those, who allready own coffe shop

and want to be in the know of the competitive world. Thank you!