Coursera_Capstone,The-Battle-of-Neighborhoods

How to Open Most Attractive Student Cafe ?

Andrey Skurikhin - Data Science Capstone - IBM Data Science Professional Certificate on Coursera

1. Introduction

1.1. Background

There are many people who attempted to open cafe or restaurant business. Sometimes they have success, but most cases are failed. The most common problem is bad prediction of risk factors which have negative influence. However, we have data science methods, which might predict and reduce most negative cases.

1.2. Problem

Imagine that you want to open most successful busuness ever been. As an Example, you want to open a student cafe network. In the near future, this set cafe is expected to be the most popular and best-known in the world. But today, you have limited amount of money, so it seems to be enough only for one cafe opening. The problem we aim to solve is to analyze good locations in the major Russian cities the best places to open our first cafe. In this case, we expect to return invested money quickly. Our main target are students who learn in top Russian cities.

1.3. Interest

The student's interest in good cafe near the learning site is ovbious. If we have some creativity, we can found a modern place for relaxing and discussions.

2. Data section

2.1. Data Sources

We will use the FourSquare API to collect data about locations of 1.Russian colleges and Universities and 2.Popular Cafes in 6 major Russian cities which are: Moscow, Saint

Petersburg, Novosibirsk, Yekaterinburg, Kazan and Nizny Novgorod. These are one of the most populated Russian cities and we can hope that they have many colleges and cafes.

2.2. Data Cleaning

We try to examine the Cafe/College density for each city and then we check such colleges that haven't popular cafes in 10 minutes walk. If we cluster such places, we can conclude that the largest cluster could be the best place to open our first Cafe.

2.3. Feature Selection

We need to select the current set of cafes as well as set of colleges in each city. Then, we'll try to find proportions and intersections of each data, apply filtering, clustering and find the best place based on these features.

We need the following features for select: Latitudes and longitudes of colleges and Latitudes and longitudes of opened cafes.

3. Methodology

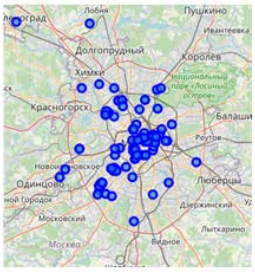
3.1. Calculating College to Café Ratio

First, we need to collect all colleges and cafes in selected top-6 Russian Cities. According to Foursquare API, Cafe CATEGORY ID has code 4bf58dd8d48988d16d941735 and College CATEGORY ID has code 4d4b7105d754a06372d81259. So the part of code for selecting Colleges and Cafes in python will be the following:

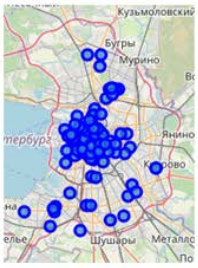
```
cities = ["Moscow, RU", 'Saint Petersburg, RU', 'Novosibirsk, RU', 'Yekaterinburg, RU', 'Kazan, RU', 'Nizhny Novgorod, RU']
url_cafe = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&near={}&limit={}&categoryId={}'.format("4bf58dd8d48988d16d941735") # Cafe CATEGORY ID
url_college = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&near={}&limit={}&categoryId={}'.format("4d4b7105d754a06372d81259") # College CATEGORY ID
```

The set of Colleges for cities is the following:

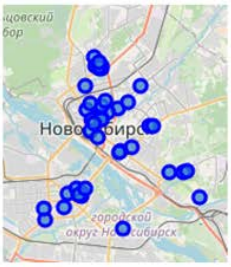
Moscow



Saint-Petersburg



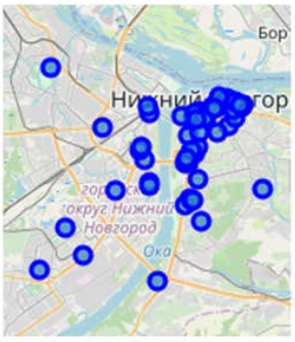
Novosibirsk



Yekaterinburg

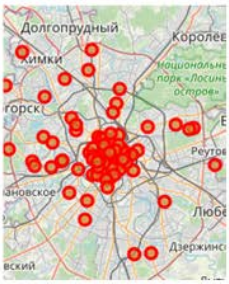
Kazan

Nizhniy Novgorod



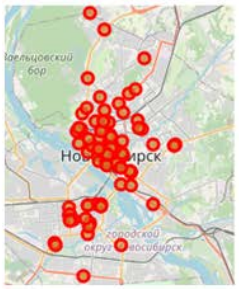
The set of Cafes for cities is the following:

Moscow

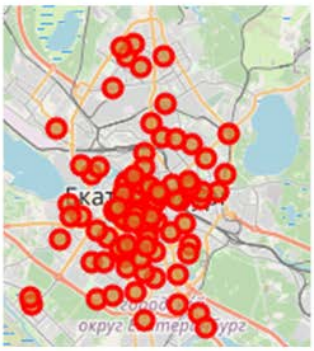


Saint-Petersburg

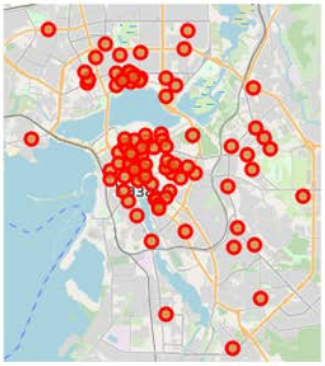
Novosibirsk



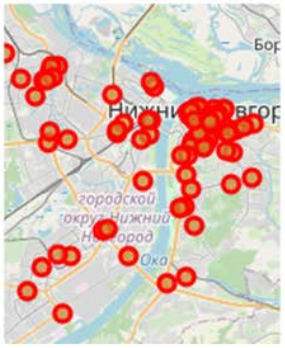
Yekaterinburg



Kazan



Nizhniy Novgorod



The total number of Colleges and College/Café ratio is the following:

- Total number of college places in Moscow, RU = 224
- Cafes per college in Moscow, RU = 1.375
- Total number of college places in Saint Petersburg, RU = 220 Cafes per college in Saint Petersburg, RU = 1.327
- Total number of college places in Novosibirsk, RU = 46
- Cafes per college in Novosibirsk, RU = 2.435
- Total number of college places in Yekaterinburg, RU = 76
- Cafes per college in Yekaterinburg, RU = 1.618
- Total number of college places in Kazan, RU = 100
- Cafes per college in Kazan, RU = 1.35
- Total number of college places in Nizhny Novgorod, RU = 63
- Cafes per college in Nizhny Novgorod, RU = 1.905

Anti-leaders for densities are Moscow, St.Petersburg and Kazan

3.2. Calculating College to Café Distance and Selecting colleges without any café nearby

Based on collected results, we need to find distance between colleges and nearest cafes. If we have at least one café in 10 minute walk near college, this place could be marked as high risk. In opposite, if there no well-known cafes near college, this place is good for café placement.

Let's restrict distance by 10 minutes walk from college and count number of cafes there. We have to calculate Euclidean distance based on Latitude and longitude. At first iteration, we consider city latitudes similar and neglect by small metric difference caused by it.

So the part of code for selecting Colleges and Cafes in python will be the following:

```
for latcl, lngcl, labelcl  in zip(college[city]['Lat'], college[city]['Lng'], college[city]['Name']):
    numcafe = 0
    for latcf, lngcf, labelcf  in zip(cafe[city]['Lat'], cafe[city]['Lng'], cafe[city]['Name']):
        dist=sqrt((latcl-latcf)*(latcl-latcf)+(lngcl-lngcf)*(lngcl-lngcf))
        if (dist < 0.01):
            numcafe = numcafe + 1
```

The results for absolute number and percent of such colleges are the following:

Total number of college places no cafe in Moscow = 56
Percent of college places no cafe in Moscow = 25.0%
Total number of college places no cafe in St Petersburg = 38
Percent of college places no cafe in St.Petersburg = 17.2%
Total number of college places no cafe in Novosibirsk = 11
Percent of college places no cafe in Novosibirsk = 23.9%
Total number of college places no cafe in Yekaterinburg = 9
Percent of college places no cafe in Yekaterinburg = 11.8%
Total number of college places no cafe in Kazan, RU = 17
Percent of college places no cafe in Kazan, RU = 17%
Total number of college places no cafe in Nizhny Novgorod = 6
Percent of college places no cafe in Nizhny Novgorod = 9.5%

Anti-Leaders for colleges without cafes are Moscow and Novosibirsk

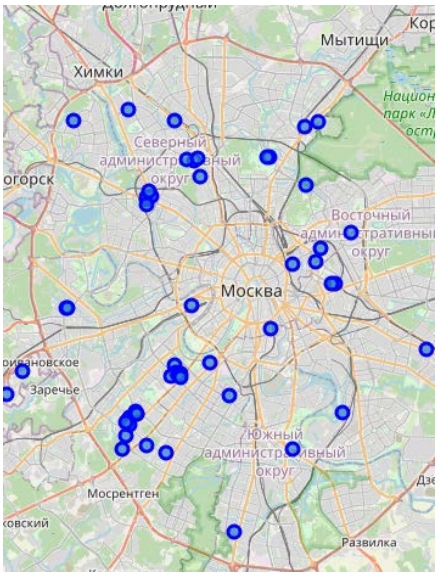
3.3. Selecting city for next stage

Based on Anti-Leader top in both cases, we need to select Moscow as a candidate for café placement

3.4. Filtering Data

In selected city (Moscow) we need to remove from map such colleges that have at least 1 café nearby (10-mintes walk). The result is following:

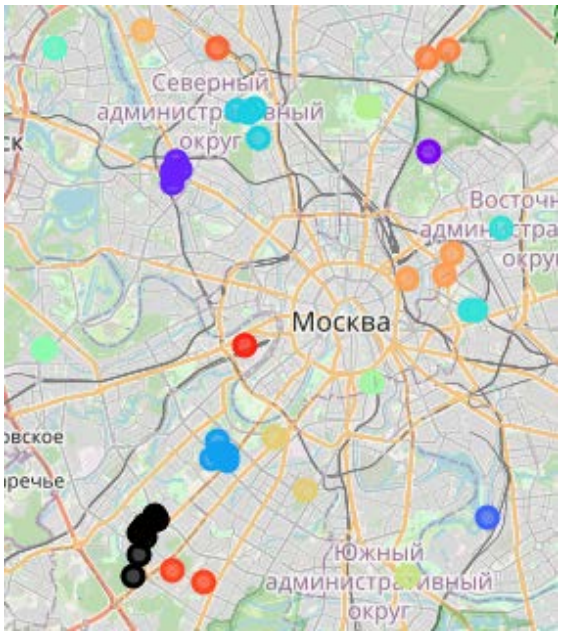
We can see that colleges without café aren't spreading randomly, but group in some



agglomerations. So if we cluster these agglomerations and find the set with maximum number of elements, we get good starting point for café placement.

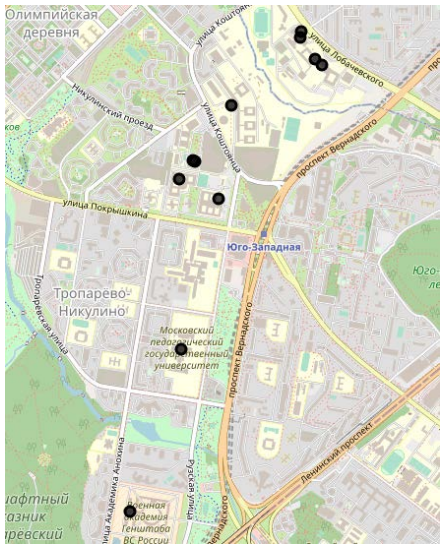
3.5. Clustering Data

Let's cluster remained colleges in groups based on distance using k-means method. The number of clusters is defined empirically as total number of colleges without cafes divided by 2 The results of clustering is shown below. Maximum cluster is marked by black



So the best cluster for placing our café is situated in south-west part of Moscow.

This place is also known as Olympic Village. Let's zoom it:



4. Results

As a summary, we obtained the following results:

- found cities with least college-to-restaurant density

- found cities with most percent of colleges without cafes

- based on these criteria, selected city for future investigations

- clustered remained colleges and found the best site for café placement.

5. Discussion:

The analysis of café placement is quite simple. For next investigations we can pay into attention the following facts:

- number of rich and poor students in each college

- number of rich and poor people in neighborhood

- gender and nationality of students to select the best menu

6. Conclusion:

The machine learning and Data Science methods are good for this business task, and, after a little improvement, could be applied to real business problems.