

Introduction

It's a common idea for students to open a coffee point near their universities. In fact, it is #1 idea student come up with. It seems quite simple and requires minimal investments - because you do not need to rent a huge location and equipment is cheap comparing to other types of venues.

Usually, they simply find the closest place to the university available and then rent a corner to set a bar counter with a couple of chairs. The only thing left is to find a barista and now you are ready.

But the core mistake is lack of analysis, the majority do not perform research in order to understand whether the place is good or bad. That is why it might happen that no customers would appear. Reasons differ - sometimes it depends on competitors, sometimes on place that is not popular enough.

As a result, most of them bankrupt less than in a year and loose their investments.

What we are going to go is to discover Moscow universities in order to find a location that would be great to set a coffee point there.

This research would help those who wants to start their business in a student age ang give them understandable metrics to properly evaluate their ideas.

Business Problem

As soon as we are going to find a great place for coffee point near one of universities, we have to develop some hypothesis about metrics, that would help us to evaluate a specific location and rank it properly.

The task of the research is to measure three most important metrics for opening a coffee point for answering the following questions:

1) Is a place popular?

The answer is crucial for understanding whether or not you should even think about opening a venue here.

Will there be enough customers for your business?

2) How many competitors will you have?

That is important because you would compete with them for customer's attention.

3) Is a place expensive?

As soon as equipment's cost is constant from place to place, the main cost you will suffer from is rent price - the cheaper the better.

Data

I used the following data:

1) Foursquare API

I will use it for discovering venues in cafe category that already exists near universities.

We will try to compare them to find the place that has many students, but no cafe venues nearby.

2) Russian Federal State Statistics service

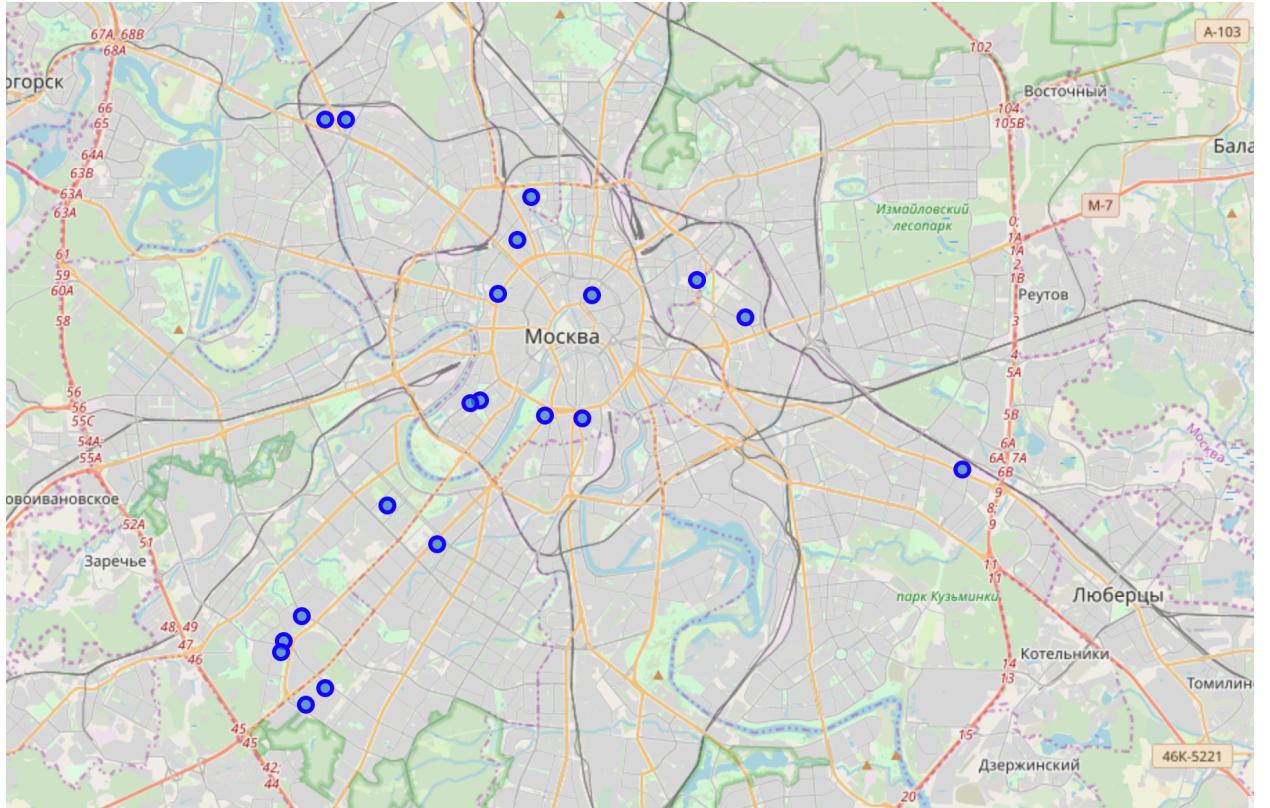
It is a federal service that collect statistic data about every aspect of economical life in Russia, so it is very useful when it comes to getting offciant and accurate data.

This one is for obtaining data about rent prices through the years. Also it might help to find a place with a cheapest rent price.

3) Wikipedia

To pick top-20 universities's and their web-cites and to find out how many students each of them has.

All we are looking for is a dataset of universities, their locations and number of students.



name	code	latitude	longitude	rent	students	coffee	place	food
Lomonosov Moscow State University	MSU	55.703092	37.530772	47100	40000	0	1	0
Higher School of Economics	HSE	55.761751	37.632456	60000	40000	8	0	14
MGIMO University	MGIMO	55.671999	37.488288	40300	6000	2	0	3
The Russian Presidential Academy of National E...	RANEPA	55.665107	37.479167	37700	6500	4	1	2
RUDN University	RUDN	55.652011	37.499527	37700	31500	0	0	5
Plekhanov Russian University of Economics	REA	55.727510	37.627605	37700	10800	5	0	9
Bauman Moscow State Technological University	BMSTU†	55.766029	37.684518	47100	19000	2	1	8
Moscow Pedagogical State University	MGPU	55.732533	37.576423	115600	26000	1	3	5
Kutafin Moscow State Law University	MSAL	55.762204	37.585919	49400	16000	2	2	10
Sechenov First Moscow State Medical University	1stMED	55.731687	37.571696	49500	15000	1	3	10
Synergy University	Synergy	55.811087	37.510359	41100	18000	2	0	4
Moscow State University of Railway Engineering	MIIT	55.789306	37.602056	43500	17000	0	2	2
MIREA†Ч Russian Technological University	MIREA	55.661905	37.477617	37700	40000	4	3	9
Russian State University for the Humanities	RSUH	55.777178	37.595495	50500	25000	9	1	15
National University of Oil and Gas	GUBKIN	55.692282	37.555094	36200	11000	2	2	4
Pirogov Russian National Research Medical Univ...	Pirogov†	55.647256	37.490287	41100	9000	0	1	0
Moscow Power Engineering Institute	MPEI†	55.755593	37.708721	42000	15000	2	0	6
Moscow Aviation Institute	MAI	55.811030	37.500065	41100	22000	1	1	5
National University of Science and Technology ...	MISIS	55.728425	37.609076	74500	15000	1	0	2
The State University of Management	GUU	55.713249	37.816150	28400	17000	3	1	2

Metology

1. Collecting initial data

First of my activities was to google top-20 Moscow Universities, which were listed in an excel file, where I also put their codes in order to simplify reader's understanding.

Next, I searched for their locations and added latitudes and longitudes to the file.

Also I used Wikipedia to find out how many students does each university have.

Then I used resources of Federal State Statistics service to discover the rent price in areas universities belong to.

These data were also added to the list.

As the result, I got an excel file, which I converted to a .csv, filled with the following data about every university:

- 1) Full name
- 2) Short name or code
- 3) Latitude and longitude
- 4) Rent price for nearest ares
- 5) Number of students

2. Gathering additional data

The next step was to discover the fact, whether the particular university is already surrounded with coffee-like venues or not.

I used Foursquare API to get a list of nearest venues for every university, then I grouped them by their category.

There was no surprise that coffee shop was the most common category, but also there were many other food-related venues.

I decided, that they could also affect coffee business, because they attract customers in not-morning time.

The other guess was that parks or malls nearby are good for coffee business, because it is a great place to put a point somewhere student like to rest.

So, as the result, the following data were added to my frame - for each university, the number of coffee-like venues nearby, food-related venues and number of just useful places for starting a coffee-point.

3. Clustering

I used K-means algorythm from SciKit library, because my aim was to categorize all universities depending on their coffee-business ranking and I have no idea what metric is really critual for the choice.

So, my first step was to run it with number of clusters equals to two (is the place good or bad), but results were disappointing, because the algorythm just separated bad place from not that bad places.

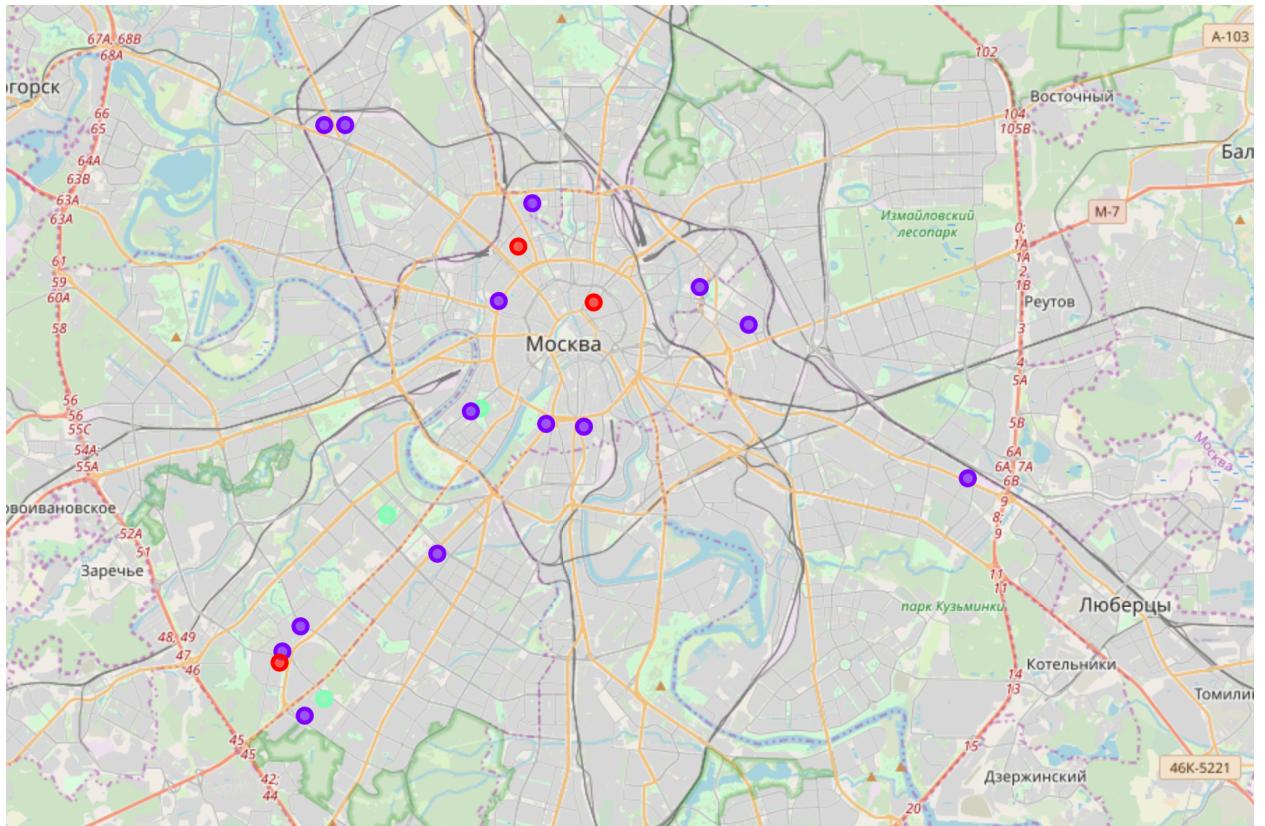
As soon as I was looking for good places, I re-run the process with other numbers of clusters, but I got only 20 universities, so the greatest result that can be properly evaluated was acchieved on the number three.

4. Evaluation

Some of these activities were already performed in the previous section - I tried to change parameters until results satisfied me, so in this section my main goal was to check if results can be implemented.

In other word, can they be used in real world? I got three groups from the clustering algorythm and they are clearly seen as "bad", "not bad" and "good". Let's look closer.

Results



So, the output of my work was the list of the universities, divided into three groups:

1. Bad guys (RED) - these universities are crowded with coffee shop, and the rent price is high. The worst places to start a regular coffee shop business.

Here they are - Higher School of Economics, MIREA - Russian Technological University and Russian State University for the Humanities.

Are they similar in their fields? No, but they are similarly bad for business.

2. Regular guys - not great, not terrible. There are some food- or coffee-related venues there, but not that much.

There are too many of them to list, you can check them in the notebook.

3. Bull's eye (GREEN) - places that are as perfect as they could be among the universities.

The coolest thing there is that they are not clearly divided by the neighborhood - some of them are pretty close to the bad ones, but belongs to the different clusters.

Lomonosov Moscow State University, RUDN University and Moscow Pedagogical State University are our winners here.

But let's have a closer look to the list of our winners - in fact, MSU is surrounded by big park (and that is good), but it's prohibited to put any not-university-related business there.

So, this option can be crossed out.

Also, we have MGPU here - #1 in rent prices. Thus, if you want to start a business, you'd better pick the RUDN - it's reasonably priced, it's allowed to place a venue there, there almost no competitors for you.

Discussion

Despite the fact that we had chosen a winner in the previous section, it required not only ML, but also my own understanding of goals I want to achieve.

What were the reasons? First of all, algorythm had no idea that you can not place a coffee-point near the MSU. So, to get rid of mistakes of that kind, you have to discover it while collecting the data.

Also, the model almost ignored rent price (maybe it thought that with a great place you would hanlde it, who knows xD)

But what is good is that at least bad places were separated correctly - there was no doubt for me that place with high rent prace and big number of competitors is not a good idea for place. So the further steps might be adding more parameters to wider the perspective.

Conclusion

So, is the problem solved? Yes, it is - our target was to find a good place to start a business in order not to make a mistake.

And with the help of ML and little analysis the winner was chosen - RUDN University.

I know that the location is not the only criteria to make a prediction about your business, but it really increases your chances and, what is the most important, could save you from definitely wrong choises.

Now it's your turn.