

Coursera
IBM Data Science Professional Certificate

**Predicting the success of a certain type of food
establishment in Russia**

Applied Data Science Capstone
by
Susanin Dmitry

November, 2019

1. Introduction

1.1 Background

Russia is a multinational country with more than 190 nations (Figure 1). In connection with this and the huge size there is a great number of cuisines that are preferred in certain regions. Also, the popularity of a particular type of food establishment affects the income of the population, the location relative to the seas and oceans, folk traditions and customs, and hundreds of different factors.

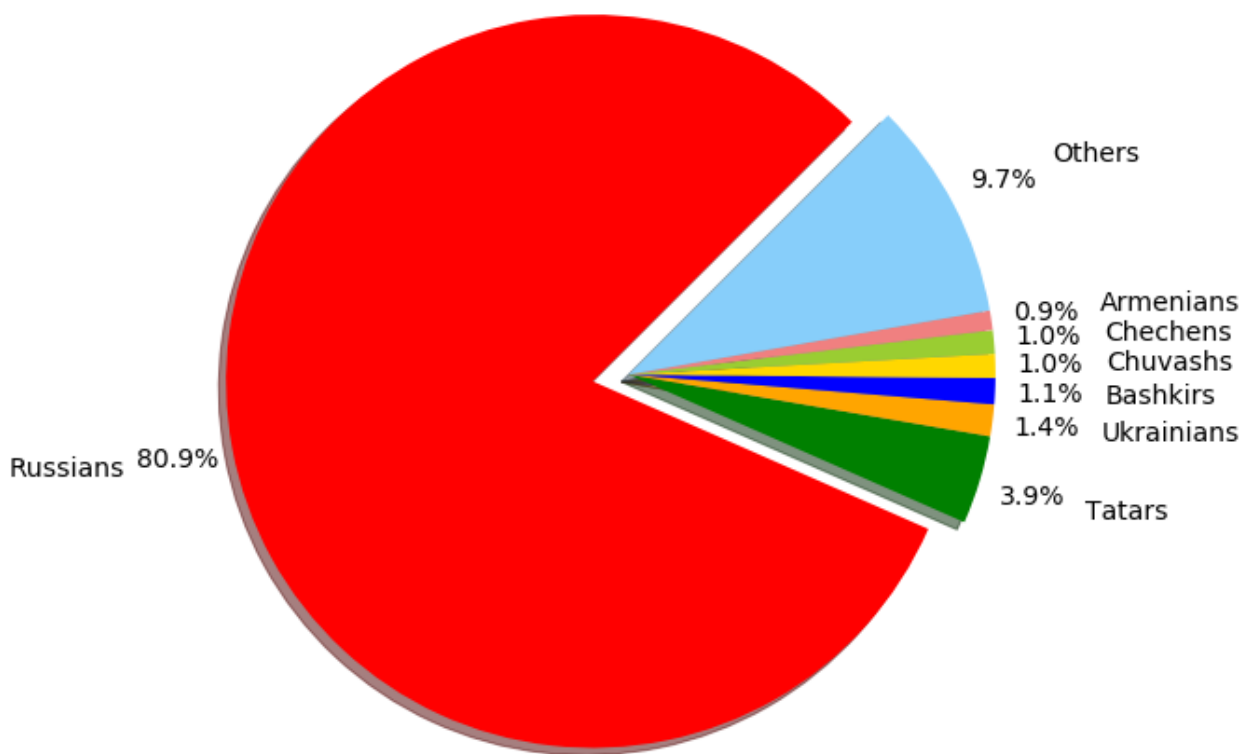


Figure 1. Ratio of main nationalities in Russia.

1.2 Problem

The main problem is that no one can predict all the features, which affects on the popularity of some type of food establishment. The only way to find out that is to check the top 3 most popular type in each region, for example.

1.3 Interest

Potential investors can get an understanding of which type of restaurant they should open, for example, for best financial indicators. This project can also be used to define which city offers the best food of some cuisine (the more restaurants of the same type, the better this cuisine).

2. Data acquisition and cleaning

2.1 Data sources

Most data was taken from wiki page https://en.wikipedia.org/wiki/List_of_cities_and_towns_in_Russia_by_population, to define the location of each city I used geolocator. I also found and downloaded additional dataset ('ru.csv', from <https://simplemaps.com/data/ru-cities>), just as a precaution (because sometimes geolocator can't find any information about some cities). I didn't use it as main dataset because it hadn't fresh info (only 2010).

2.2 Data cleaning

I didn't use such columns as Rank, Federal subject names, Federal district names, Population in 2010 and change in it between 2010 and 2017, that's why I dropped them. That was enough to get a clean dataset.

2.3 Feature selection

I am only using 300 cities, because I think that the interest of investors is directly proportional to the population of the city, so I don't need to analyze small cities (towns). That's why I got 300 samples with 5 features. It may seem that this is not enough, but do not forget that the main part of the data I will get from the results of the FourSquare API work.

3. Methodology

Main idea of a project is to find out which type of food establishment is the most popular in every city of Russia. As I wrote above, I was using wiki, geolocator and additional dataset to get data about population and location. Then, I made a call to FourSquare API to get information about food establishments in every city, and saved all results into a new dataframe. After that, there was a code to find out which type is the most popular in every city. So, I got a dataset with 300 samples and 7

columns (because I think Top 3 types by number of opened establishments in city is enough).

That's enough data to draw a map of the most popular food establishment types for Russian cities. Next step was to make this. I used a Folium package and with randomizing colors for every type of establishment get the following picture (Figure 2). Map legend is interactive, user can make visible or hide some types and can zoom in/zoom out if it's necessary.

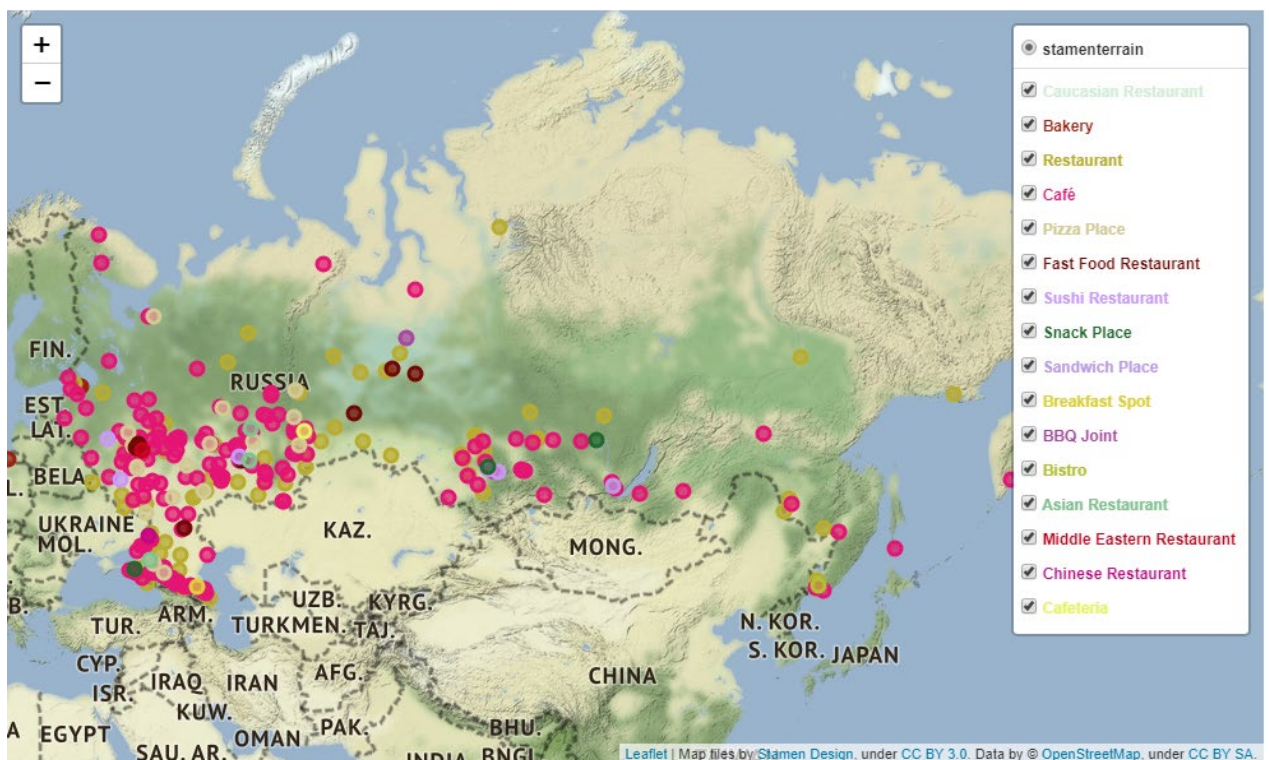


Figure 2. Map of the most popular food establishment types for Russian cities with colored markers.

4. Results

The result of this program is the map of the most popular food establishment types in Russian cities. It shows potential investor, that the most popular type in Russia is café.

5. Discussion

There is also a little interactive code, which recommend you which type you should open in certain city or in what city you should open certain type. It's obvious

that restaurants are popular in big cities, probably because of greater income of citizens. Fast food also not so popular, but there is a lot of Bistro type establishments (French type of a small restaurants with a simple meals).

All of that means, that Russian people do not like to spend a lot of time in food establishments. For us, restaurants are places where you can sit, relax, meet your friends and discuss some issues, but not to eat and that's it.

6. Conclusion

In conclusion, I would like to note that this project allows us to give an impression of what type of food establishments should be considered for opening. At the same time, the project should not be determined one hundred percent and serves only as a demonstration and a hint for which type is worth a closer look, since there are dozens of other factors that deserve additional research in other works.

Thanks for watching!