Coursera Capstone

IBM Applied Data Science Capstone

Franchise Development in Belarus

By: Yakuta Aleh

January 2020

Introduction

Belarus has only recently opened to the West, it is a developing and promising country, both for foreign and domestic franchises.

For many buyers, the presence of a famous brand known outside of our country (McDonald's, CFS, Dodo Pizza) will be a significant motivation to visit this institution.

Of course, as with any business decision, the opening of a new point and further development strategy requires serious consideration. The choice of the next city for development and the formation of a plan to increase points in the cities of our country is one of the most difficult and important decisions that will determine whether the business will be successful and how effective the development will be.

Business problem

The purpose of this project is to analyze the cities of Belarus and institutions located on their territory to determine a strategy for developing their network in a specific area in Belarus. Using the methodology of data science and machine learning, using the clustering method, this project aims to provide a solution that answers the business question: how to build a strategy for developing a network of establishments of a certain type in medium and small cities of Belarus, in simple words, where it is worth opening a new point first of all?

The target audience of this project

This project is especially useful for entrepreneurs and investors who want to open or invest in the development of their franchise or buying an existing one and developing it in our country. Franchises such as CFS, burger king, H&M are already beginning to develop the regional centers of our country, and soon they will probably pay attention to medium and small cities.
The country needs the development of large networks, which will create competition and fruitfully affect the domestic market and the development of entrepreneurship in general.

Data

To solve the problem, we need the following data:

- Required data
  - ➢ List of cities in Belarus
  - ➢ Latitude and longitude of cities
  - ➢ Population data for different years, salary for different years
- Data sources
  - ➢ Data from the website of the National Statistical Committee (https://www.belstat.gov.by/)
  - ➢ Geocoder package for obtaining latitude and longitude coordinates

➢ Foursquare API for receiving data on institutions of various types in cities

Data sources and methods for their extraction:

The website of the National Statistics Committee contains annual newsletters that contain information on cities, their population, and average wages in each city. We convert this data to Excel format, because Data is available for download only.

After that, it will be necessary to clear the data, also calculate the coefficients of increase or decrease in the population and wages, as well as the coefficients of comparing the population and wages between cities.

The Python Geocoder package will provide us with the latitude and longitude coordinates of cities.

After that, we will use the Foursquare API to obtain data on establishments located in these settlements and all the information necessary for analysis.

Methodology

We get information about the list of cities of Belarus from the website of the Statistical Committee of the Republic of Belarus. The site contains newsletters with the necessary information in PDF format. Open the necessary files and add to our source file Excel. We add information on the population for 2019 and 2016, as well as the salary for 2019 and 2015.

For analysis, we will use only medium and small cities, as the capital and all cities of the regional type are too large, so they are not of interest to us for analysis.

To correctly compare population information, we introduce additional coefficients. Firstly, the trend coefficient - it shows whether the population is increasing or decreasing as a percentage. We also add the city size factor, dividing our sample depending on the value from 0.1 to 1.0 using the data on the maximum value.

To analyze information about wages, we will transfer all values to USD, because In Belarus, inflation was very high during these years, then we add the trend coefficient - it shows whether wages increase or decrease as a percentage. Add the salary ratio - the ratio of salary to average salary as of January 1, 2019.

We get the latitude and longitude coordinates using Geocoder, for each locality. Next, we visualize all the cities on the map of Belarus using the Folium package. Thus, we verify the correctness of the data obtained.

Next, we will use the Foursquare API to get all the places and their types that are in these cities.

We use data filtering by one of the types.

Clustering cities using the k-means method.

We will group cities in 5 clusters. We display the cities on the map and use different colors depending on the resulting cluster.

Results

The k-means clustering results show that we can classify neighborhoods by 5

City analysis clusters for the type of Restaurant establishments, depending on the attractiveness for opening a new institution of this type:

- Cluster 3 (green): high salary, high strength, positive growth rates.
- Cluster 4 (orange): average salary, high strength, positive growth rates.
- Cluster 0 (red): cities with a small number but not the lowest salary.
- Cluster 2 (blue): unattractive cities, with a small number.
- Cluster 1 (purple): indicates cities that already have establishments of this type
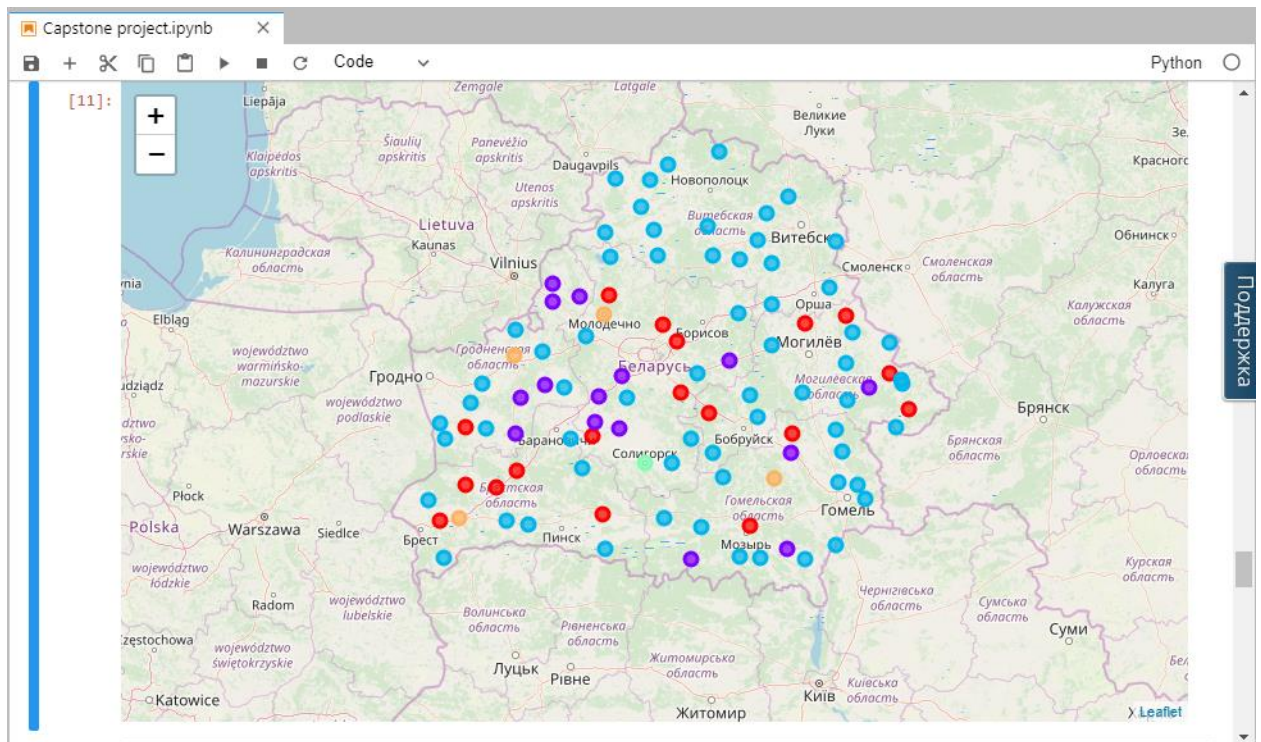
Discussion

It can be noted that, first of all, in order to open a new franchise point, it is necessary to pay attention to cities where there are no establishments of this type, but at the same time a high population and relatively high wages, clusters 3 and 4.

Next, pay attention to the middle-income cities of clusters 0 and 2. Avoid cities from cluster 1, as there are already establishments of this type and competition will not always positively affect the further development of the business.

On the other hand, this project shows how little data we have on open access cities.

Limitations and suggestions for future research

In this project, we considered only one type of institution, we also used data on population and wages for 2 years. For a deeper analysis, you can add data, for example, for 10 years. You can also use data of a different nature, for example, the number of working-age population in the city.

We used the Foursquare API data, as we used the free version, we have limitations on the number of API calls. I would also like to draw attention to the low popularity of Foursquare in Belarus, which may affect the reliability of information about the availability of establishments in cities.

Conclusion

In this project, we went through the process of identifying a business problem, determined the necessary list of data, engaged in their extraction and purification, and then used machine learning by clustering data in 5 clusters based on their similarities. In conclusion, we provided our recommendations based on an analysis of the resulting clusters.

The results of the analysis will help the entrepreneur to determine the most profitable and promising way to develop his own franchise in the cities of Belarus.