

Data Science Capstone Project

Research of the best restaurant location in Minsk

Anastasia Gerasimovich

15 July 2019

1. Introduction

1.1 Background

Minsk is a business and cultural center of Belarus. Food industry in Minsk develops really fast. Cafes and restaurants of all kinds appear one after another. What is wonderful, most of fast food places people are used to see in other spots of the world still are nowhere to find in Minsk: no Subway or Starbucks, but there are still tens of places you can choose from.

1.2 Problem

Over the last 5 years the number of restaurants has trippled, the growing market makes investors become more interested in the location. The aim of this project is to specify most prospective locations in Minsk to open a restaurant.

1.3 Interest

Obviously, investors shall be very interested in accurate prediction of profitable restaurant locations to obtain competitive advantage and business values.

2. Data acquisition and cleaning

2.1 Data sources

The research will be based on analysis of most popular restaurants according to Foursquare information. Number of restaurants and their geographical location, as well as distance from Minsk center will be used in the research.

2.2 Data cleaning

The data returned from Foursquare API was sorted according to the information associated with the location of restaurants. For sake of simplicity all unrequired columns were dropped.

The main problem with the data that I faced was insufficient number of restaurants registered in Minsk according to foursquare information. Unfortunately, I did not manage to find any available information source with free API for Minsk venues. That is why the main assumption of the research is that Foursquare data set is discoverable and sufficient.

3. Exploratory Data Analysis

In order to satisfy the aim of the research I needed to arrange data into clusters. Among clustering techniques K-means clustering was chosen. The K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. This principle exactly matches the research goals.

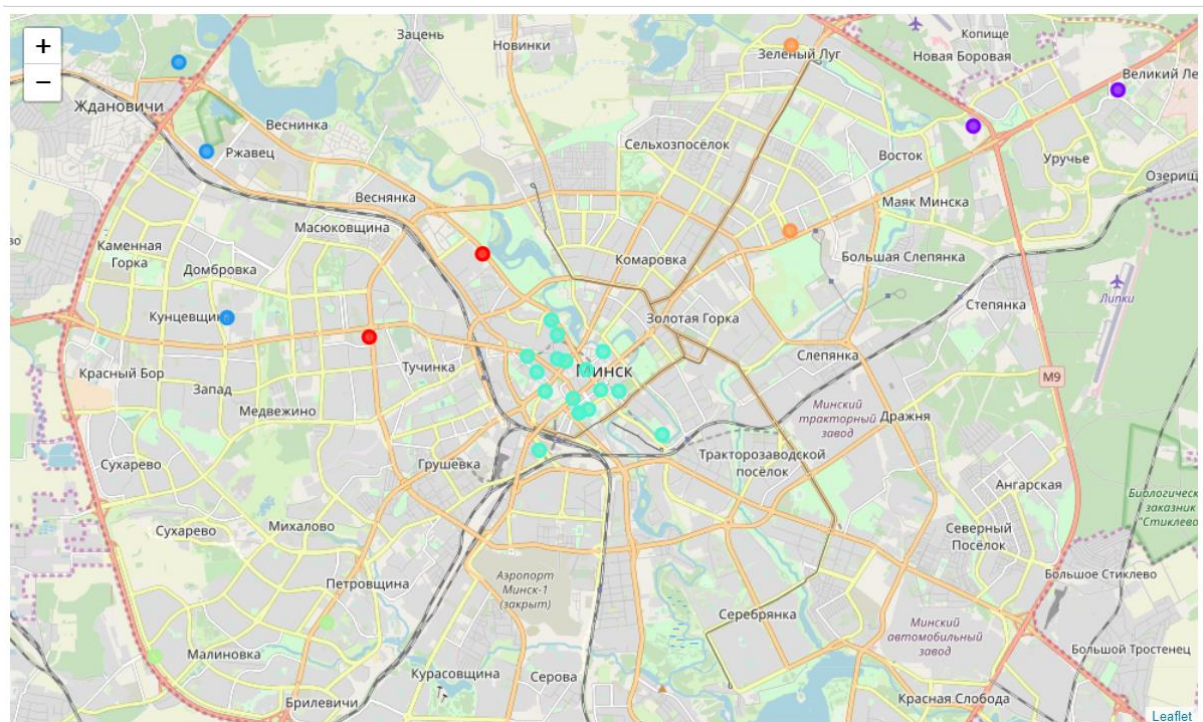


Figure 1. Clustering restaurants location on Minsk map

Having all available locations grouped into clusters next step was to define the cluster with the highest locations density. This cluster is supposed to be located in a cultural center with high publicity.

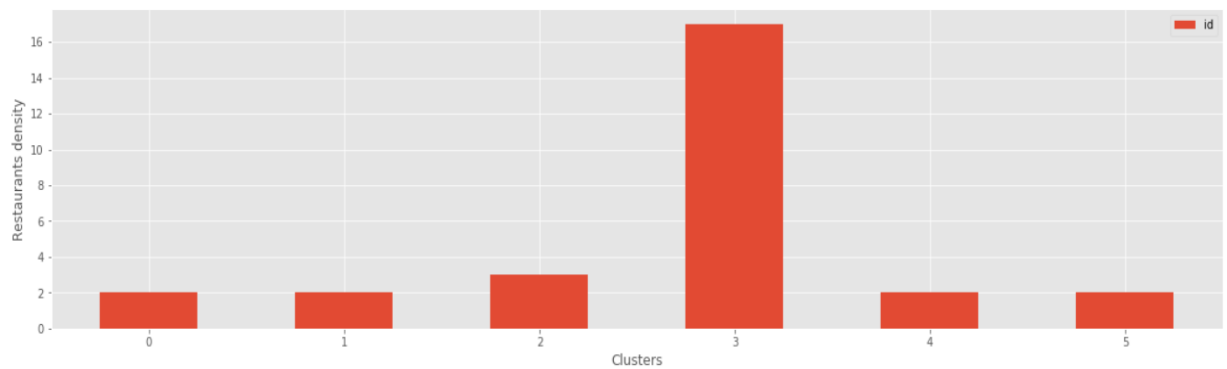


Figure 2. Clusters density

As shown on Figure 2 the highest density of restaurants appears in cluster 3. Therefore I defined the restaurants which is located in a largest distance from Minsk center to set the radius of search for a new location. According to the results, the new restaurant should be located not more than 1916 meters away from Minsk center.

5. Conclusions

In this study, I analyzed the location of restaurants in Minsk in order to find territories with high level of publicity, which is essential for a profitable restaurant business. I built clustering model and analyzed given data according to several parameters: geographical position, distance from Minsk center, density. This model can be very useful for potential investors when choosing a place to open a new restaurant.