Data Science Capstone by IBM/Coursera

Table of contents

Introduction: Business Problem

In this project, we will try to find the optimal location for a supermarket. This report will be of interest to food retailers for opening new supermarkets in Moscow. Since there are a huge number of supermarkets in Moscow, we will try to find out from the existing offers for renting out real estate, those where opening a store is optimal. First of all, we are interested in offers, near which, where the placement of supermarkets nearby is minimal. We will also choose from offers where the rental rate does not exceed the median. We will use our data science capabilities to select a few of the most promising proposals based on these criteria, so that stakeholders can choose the best possible end location.

Data

Based on the definition of our problem, the factors that will influence our decision are: the number of existing supermarkets in the neighborhood the number and distance to the supermarkets in the neighborhood Our research uses the following data sources:

- data on rental offers for premises with an area of 260 to 500 square meters located on the 1st floor of buildings.

Link

https://www.cian.ru/cat.php?building_class_type%5B0%5D=1&building_class_type%5B1%5D=2&building_class_type%5B2%5D=3&building_class_type%5B3%5D=4&building_class_type%5B4%5D=8&condition_type_f%5B0%5D=2&condition_type_f%5B1%5D=5&condition_type_f%5B2%5D=6&contract%5B0%5D=1&deal_type=rent&engine_version=2&is_first_floor=1&m2=1&maxarea=500&minarea=260&offer_type=offices&office_type%5B0%5D=2&office_type%5B1%5D=5&office_type%5B2%5D=11&placement_type%5B0%5D=2&region=1&zerocom=0

- the number of supermarkets, their location in each area will be obtained using the Foursquare API

Foursquare

Now that we have our location candidates, let's use the Foursquare API to get information about the supermarket in each neighborhood.

We are interested in establishments of the supermarket category. We are not interested in establishments of a different format for the sale of food.

Categories are taken from the link https://developer.foursquare.com/docs/build-with-foursquare/categories/

Now we have all the supermarkets in the immediate vicinity of every rental offer. This completes the data collection stage - we are now ready to use this data for analysis to prepare a report on the optimal locations for the new supermarket

Methodology

In this project, we will focus our efforts on identifying areas of Moscow with a low density of supermarkets, especially those where there are few of them.

- At the first stage, we collected the necessary data from the database of real estate rental offers
- The second step in our analysis will be to calculate and study the density of supermarkets within a 1 km radius of the potential location and identify several promising areas with a low number of supermarkets and focus our attention on these areas.
- At the third and final stage we will focus on the most promising areas that meet some basic requirements: we will consider locations with no supermarkets at a distance of 100 meters, the number of supermarkets not exceeding 1 quantile of the average number of supermarkets for a sample of about 200 premises offered for rent.

Analysis

## Distribution of the average number of supermarkets
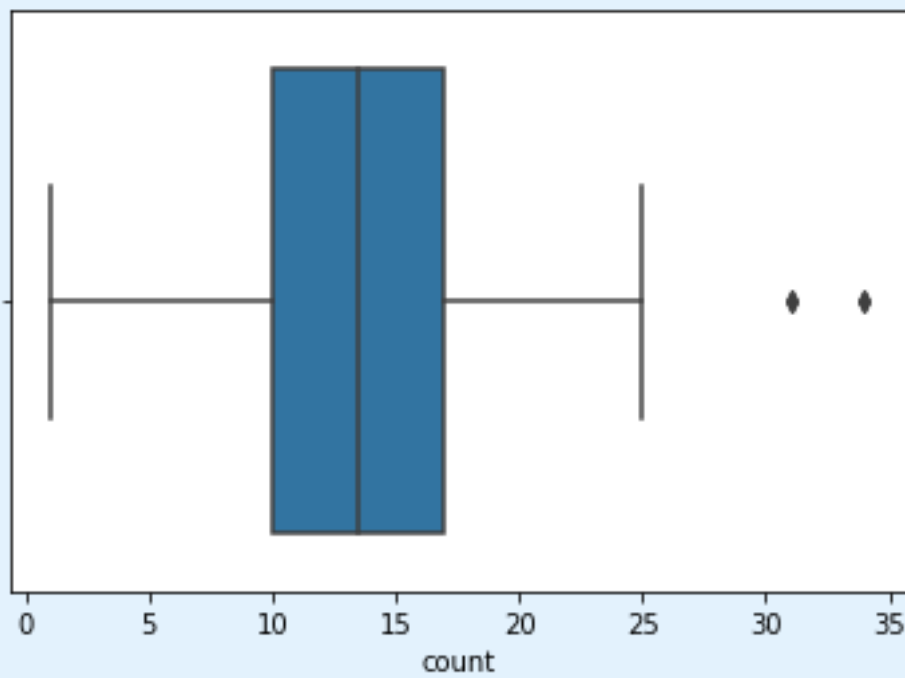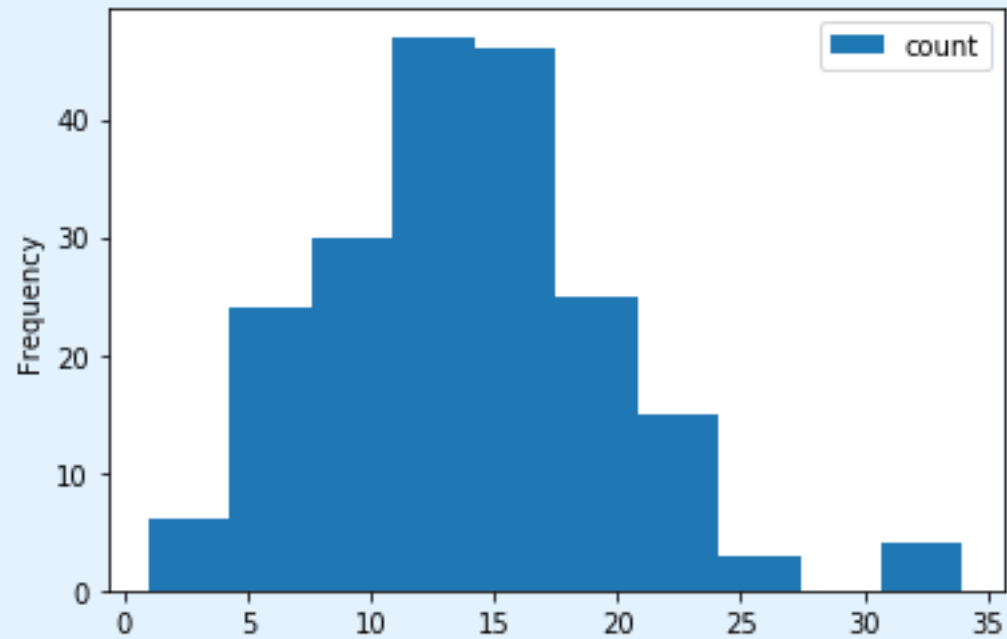
Distribution by rental price per square meter
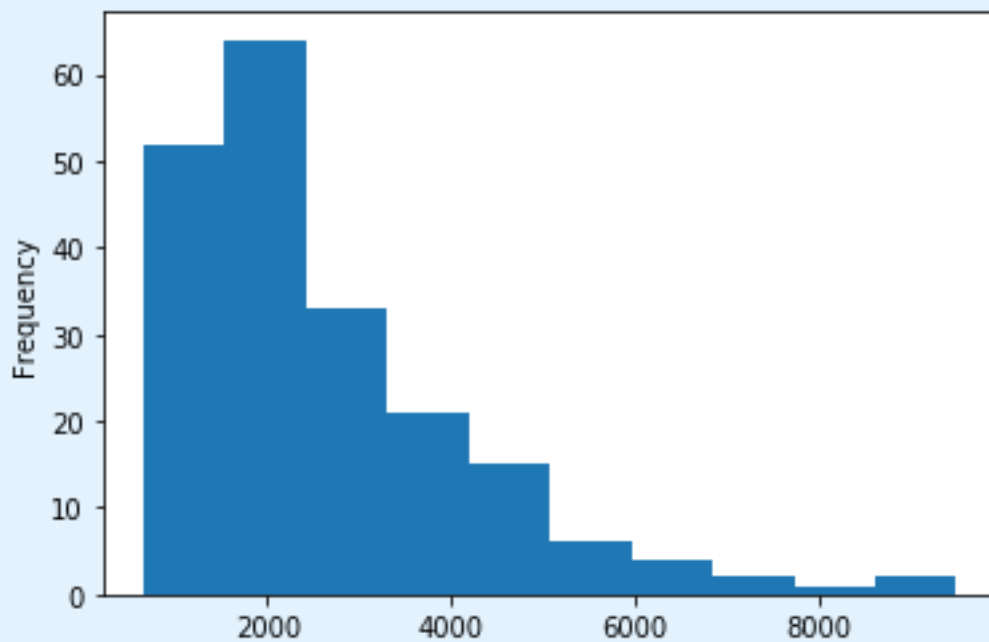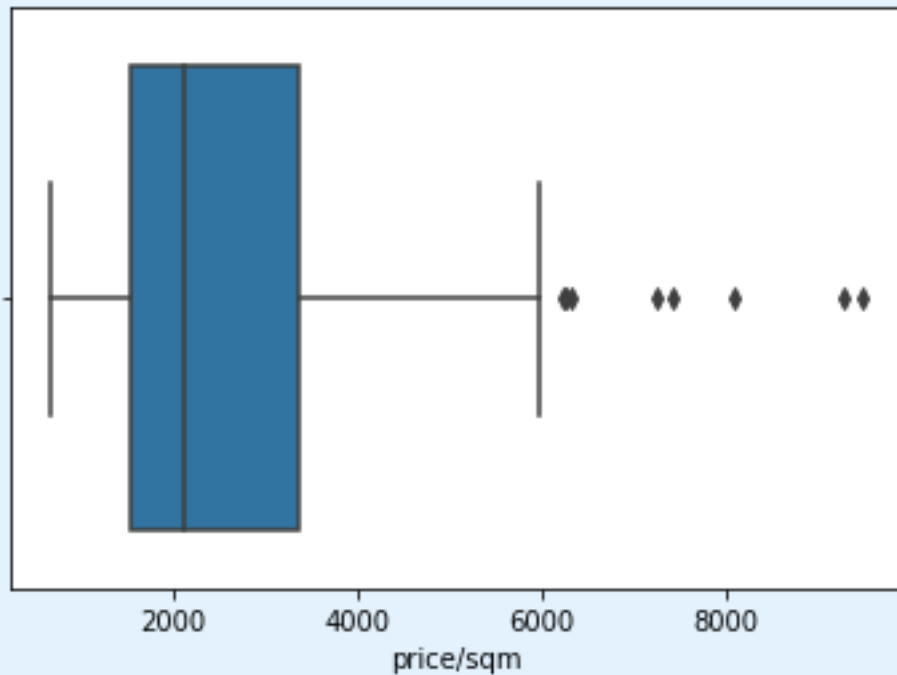
Optimization

1. We filter rental offers that are outside of 1 quantile
2. Remove offers with a value exceeding median price
3. Filter these locations: we are only interested in locations with no supermarkets within a radius of 100 meters

Creating a map with rental offers and existing supermarkets
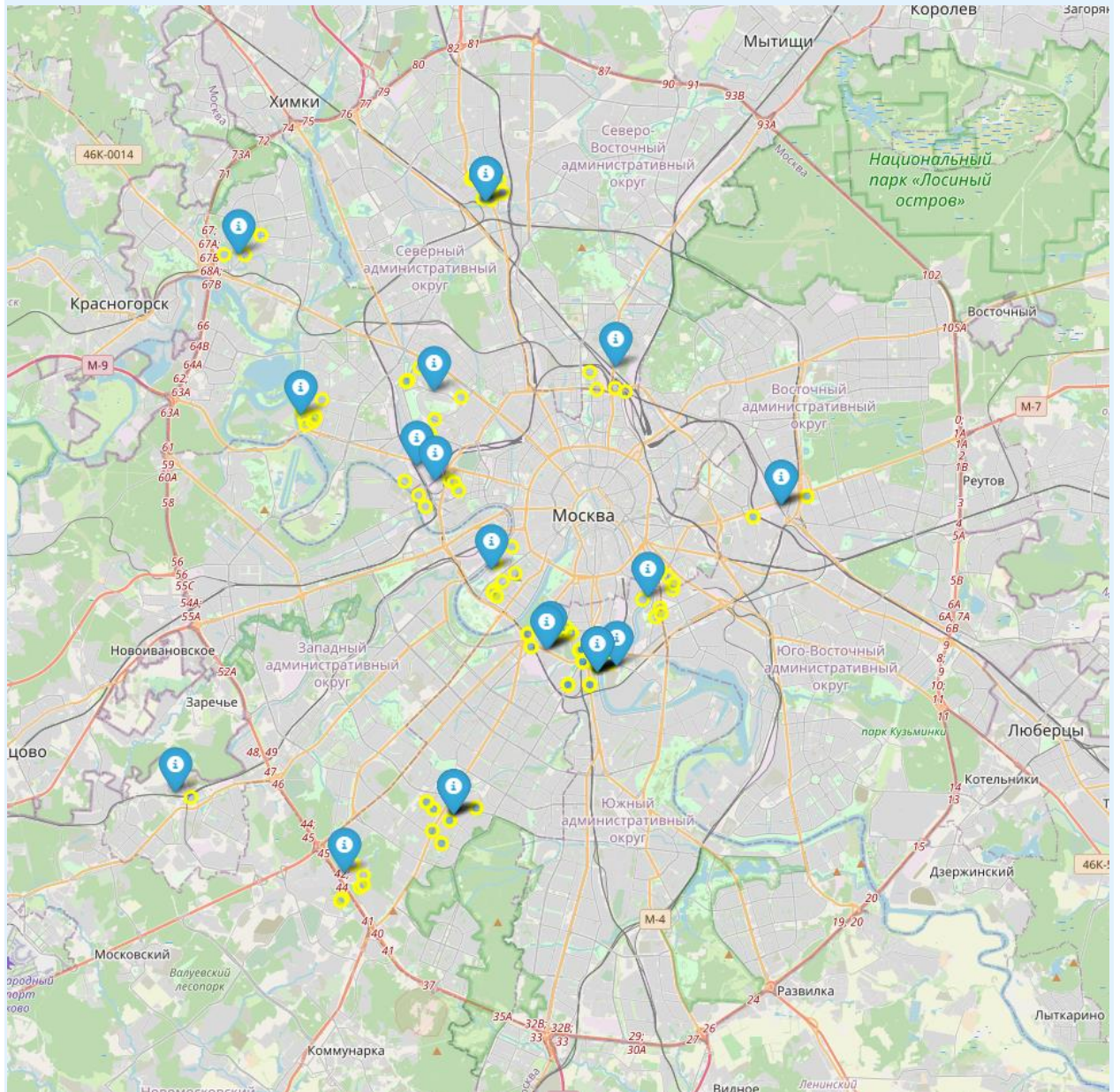
This concludes our analysis. We have created **26** addresses containing places with a small number of supermarkets, no other supermarkets within a 100 m radius. The location of supermarkets is very irregular, and their centers/addresses should only be considered as a starting point for exploring the surrounding area in search of potential locations for a new supermarket.

Results and Discussion

Our analysis shows a huge number of supermarkets in Moscow, but there are pockets of low supermarket density. The result of all this is **26** potential offers for new supermarkets depending on the number and distance to existing sites . This, of course, does not mean that these areas are actually the optimal places for a new supermarket. The purpose of this analysis was to provide information for further consideration of suitable proposals. Therefore, the recommended zones should

only be considered as a starting point for a more detailed analysis, which may eventually lead to a location where not only there will be no competition nearby, but other factors will be taken into account and all other relevant conditions will be met.

Conclusion

The goal of this project was to identify areas with a small number of supermarkets within Moscow's borders to help stakeholders narrow down the search for the optimal location for opening a new supermarket. By calculating the distribution of supermarket densities based on Foursquare data, we have created a collection of locations that meet some of the basic requirements for existing nearby supermarkets. The final decision on the optimal location of the restaurant will be made by the stakeholders based on the specific characteristics of the facilities