Capstone Project - The Battle of the Neighborhoods¶

Applied Data Science Capstone by IBM/Coursera

Table of contents

- Introduction: Business Problem
- Data
- Methodology
- Analysis
- Results and Discussion
- Conclusion

Introduction: Business Problem

In this project we will try to find an optimal location for a warehouse. Specifically, this report will be targeted to management interested in opening new facilities to optimize logistics to chain supermarkets **"Perekrestok"** within **South-Eastern and Southern Administrative Okrugs of Moscow, Russia.** "Okrug" means "district" and is something in between Neighborhood and Borough by similarity.

Since there are lots of stores we need to deliver on a regular basis we will try to detect **locations that allow us to optimize routes and minimize expenses**.

We will use our data science powers to generate a few most promising locations based on this criteria. All distances between optimal location and warehouses available will be calculated so that the best possible final candidates can be chosen by management.

Data

Based on definition of our problem, factors that will influence our decision are:

- number and locations of delivery points (stores) in area
- number of warehouses we should set to supply stores effectively and reduce costs
- number of warehouses available for rent around optimum points.

- distance of a store to optimal warehouse

We decided to use an area of 5 km radius around each district center, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

- coordinates of district centers will be obtained using **GeoPy service**
- number of chain stores and location in every area will be obtained using **Foursquare API**
- centers of optimal locations will be generated algorithmically using clustering algorithm
- list of warehouses for rent will be obtained using **Yandex Realty service:** https://realty.yandex.ru/?from=yatab&utm_source=tab-yandex-glavnaya&utm_content=web_yatab

Methodology

In this project we will direct our efforts on researching areas of **Moscow, South-Eastern Administrative Okrug and Southern Administrative Okrug.** Within the borders of these districts there are a number of chain supermarkets, and warehouse alternatives.

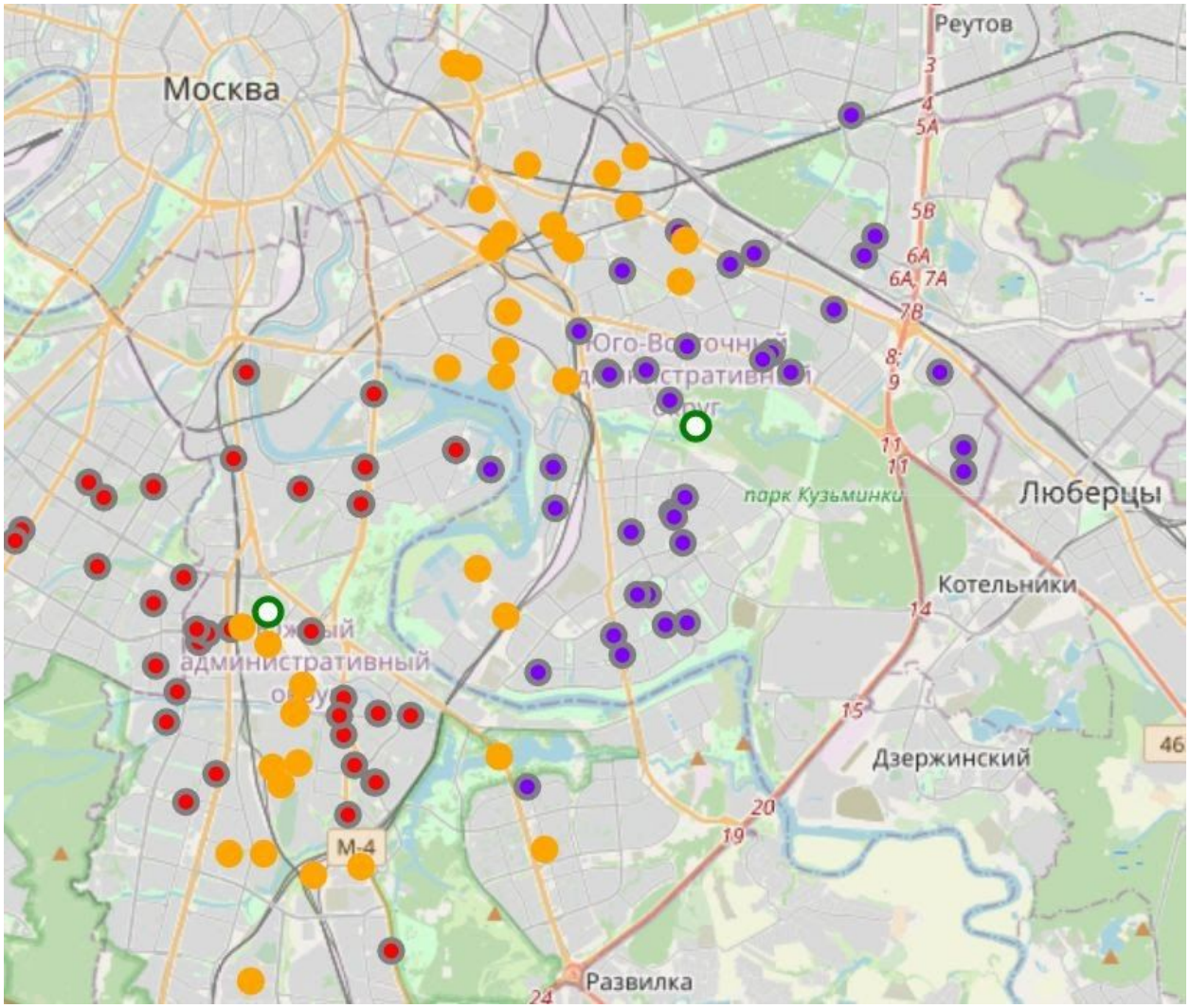
In first step we have collected the location of every supermarket within 5km from each district center. Then we separate them in a few clusters (using k-means clustering) for more effective logistics.

Then step in our analysis we define warehouses available for rent within our districts using **Yandex Realty** service.

To accurately calculate distances between warehouses and centroids we use functions to convert between WGS84 spherical coordinate system (latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters).

Finallinaly we will focus on most promising warehouse locations according to how close they are to optimum cluster centers. We will present a map for optimal search facility location.

Analysis



We prepared data effectively so analysis is quite straightforward.

When we look at the map we see, that warehouses (orange dots) spread is not even through our area. So while there are some candidates really close to one supermarket optimum center (left white dot with in center of red dots cluster), the other center has only distant options (another white dot with grape dots around).

Next we need to get the exact distances of each warehouse to each optimum.

As mentioned above, to accurately calculate distances between warehouses and centers we use functions to convert between WGS84 spherical coordinate system

(latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters).

The results are shown here:

	Address	Distance_C1
0	Moskva, Kashirskij proezd, 17s1	616.917562
1	Moskva, 2-j Kotljakovskij pereulok, 1s69	645.654343
2	Moskva, 1-j Kotljakovskij pereulok, 13	1653.896744
3	Moskva, 1-j Kotljakovskij pereulok, 12s2	2110.589138
4	Moskva, Kantemirovskaja ulitsa, 65	2137.561909

Here we have 2 options within 650 m to optimum center which is good. Red cluster which is mainly in Southern Administrative Okrug.

	Address	Distance_C2
0	Moskva, Shossejnaja ul, d. 2A	2823.710375
1	Moskva, 2-j Vjazovskij proezd, 16	2978.868355
2	Moskva, 1-j Vjazovskij proezd, 4k1	3832.524355
3	Moskva, Juzhnoportovaja ulitsa, 40	4114.997046
4	Moskva, Juzhnoportovaja ulitsa, 36s1	4182.554675

The other one seems not so good. The closest option is almost 3000 m to the second optimum center.

Results and Discussion

Our analysis shows that there are **80 supermarkets** that we divided optimally for two clusters in future logistic purposes.

We also got **38 warehouse** candidates for the moment that met all formal basic conditions (fee, engineer systems, construction conditions).

Mapping and future calculations showed that candidates are spread irregularly. Which can be caused by the city specific zone: living area, recreational or industry which is out of our control. Or just matter of supply up to date that could change sometime.

Resulting with two good options for one cluster and far options (about 3000 m closest) to another.

Purpose of this analysis was to only provide info on best available for rent warehouse locations. Which was solved only partially because of factors that are out of our control.

Recommended locations should therefore be considered only as a starting point for more detailed analysis with other factors taken into account and all other relevant conditions met.

Conclusion

Purpose of this project was to identify optimal locations for warehouse rent in the area of **Moscow, South-Eastern and Southern Administrative Okrugs** in order to narrow down the search of available candidates on the realty market.

By getting Foursquare data we first identified locations of all **“Perekrestok”** supermarkets in the area.

Clustering of those locations was then performed in order to create even groups around 40 (+/-5) for optimal logistics and find their centers. Which we used like the optimum location to have a warehouse. Clustering resulted in two exactly equal groups.

Mapped results and calculations showed irregular distribution of warehouse candidates, which resulted in completion of our optimization task partially.

Final decisions on warehouse rent will be made by management based on specific characteristics of facilities and locations closest to each optimum center.