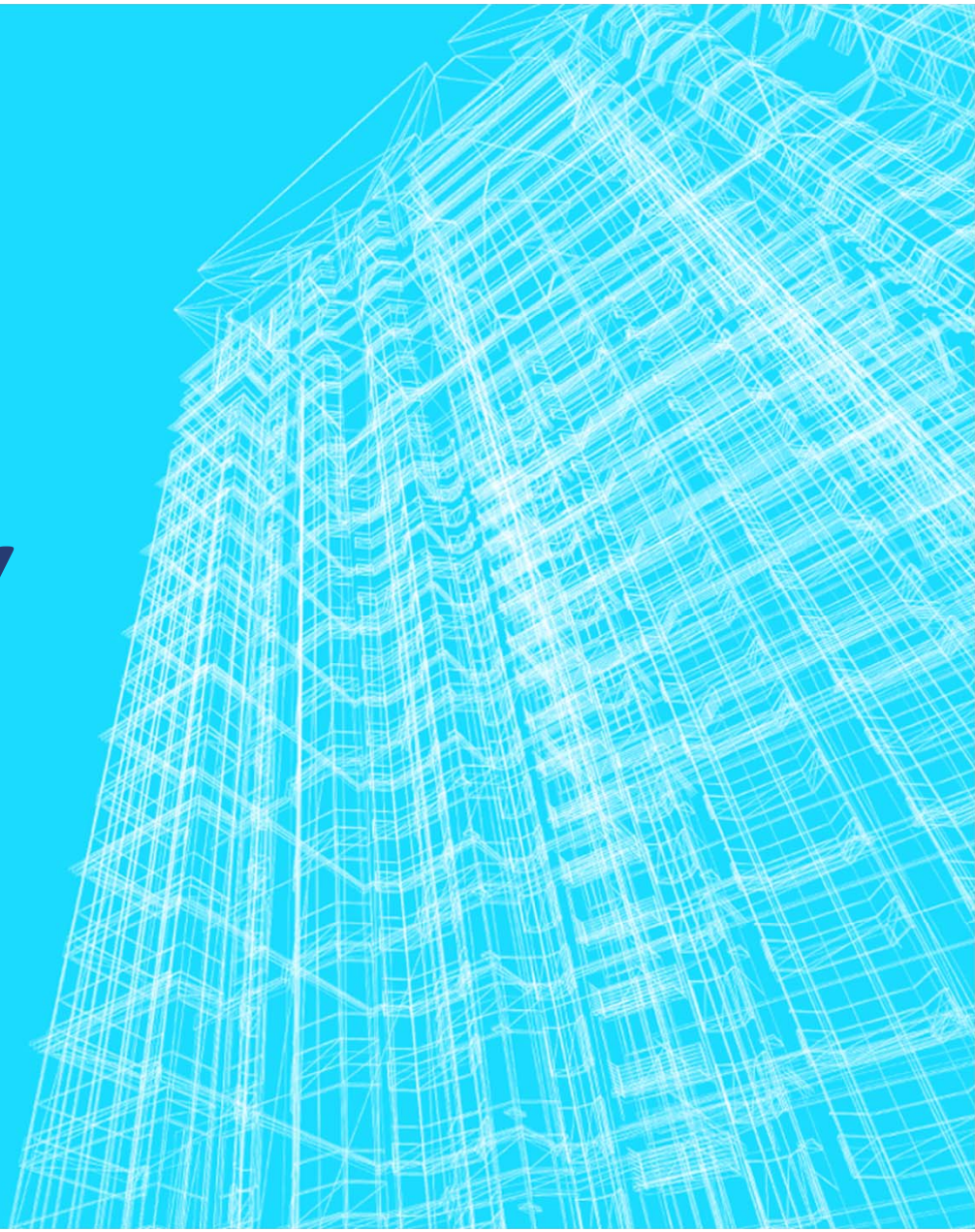
VENUES DATA ANALYSIS OF MOSCOW CITY

Capstone Project
by Alex Pr



INTRODUCTION

Background

Moscow is one of the largest metropolises with a population of more than 15 million people, covers an area around 2561.5 km² with an average density of inheritance of 4924.96 people / km² . It is really diverse city with great difference among its districts.

Problem statement

In such place business competition is really high. To start a new business you should have information about city venues, population and expenses for real estate. You face to multiple variables choice and Data Science can offer a solution. Owners of social facilities are expected to prefer boroughs with a high population density. Investors will prefer areas with low housing costs and low competitiveness. By means of Exploratory Data Analysis and k-means Clustering we visualize potential areas and venue type, that can be successful .

Interest

Investors to new social facilities and owners of existing ones to open a new one.

In my research, I will try to determine the optimal places for the location of Auto Workshops in Moscow boroughs, taking into account the number of people, the cost of real estate and the density of other facilities.

DATA SET

City data

- Name, type of the each Moscow Borough
- Population, housing area, average housing price for each Moscow Borough
- Geographical coordinates and shape of the each Moscow Borough

Borough_Name	District_Name	Borough_Type	ATO_Borough_Cc	TMO_District_C	Borough_Area	rough_Populatic	Populatic	rough_Housing_Ai	using_Are	Latitude	Longitude	orough_Housing_Pric
Академический	ЮЗАО	Муниципальный округ	45293554	45397000	5.83	109387	18762	2467.00	22.70	55.69	37.58	199999.00
Алексеевский	СВАО	Муниципальный округ	45280552	45349000	5.29	80534	15223	1607.90	20.50	55.81	37.65	199474.00
Алтуфьевский	СВАО	Муниципальный округ	45280554	45350000	3.25	57596	17721	839.30	15.50	55.88	37.58	138021.00
Арбат	ЦАО	Муниципальный округ	45286552	45374000	2.11	36125	17120	731.00	26.00	55.75	37.59	438568.00
Аэропорт	САО	Муниципальный округ	45277553	45333000	4.58	79486	17355	1939.70	25.90	55.80	37.53	234544.00
Бабушкинский	СВАО	Муниципальный округ	45280556	45351000	5.07	88537	17462	1586.30	18.50	55.87	37.66	164324.00
Басманный	ЦАО	Муниципальный округ	45286555	45375000	8.37	110694	13225	1991.80	18.40	55.78	37.69	302021.00
Беговой	САО	Муниципальный округ	45277556	45334000	5.56	42781	7694	791.10	18.80	55.78	37.57	261402.00
Бескудниковский	САО	Муниципальный округ	45277559	45335000	3.30	79603	24122	1391.70	18.40	55.86	37.56	158398.00
Бибирево	СВАО	Муниципальный округ	45280558	45352000	6.45	160163	24831	2521.80	15.80	55.88	37.60	140533.00
Бирюлёво Восточное	ЮАО	Муниципальный округ	45296553	45911000	14.77	155863	10552	2122.20	14.70	55.59	37.66	124645.00
Бирюлёво Западное	ЮАО	Муниципальный округ	45296555	45912000	8.51	88672	10419	1183.20	13.20	55.59	37.64	109421.00
Богородское	ВАО	Муниципальный округ	45263552	45301000	10.24	109324	10676	1744.10	16.90	55.82	37.71	178577.00
Братеево	ЮАО	Муниципальный округ	45296557	45913000	7.63	110021	14419	1585.40	15.50	55.64	37.76	136300.00
Бутырский	СВАО	Муниципальный округ	45280561	45353000	5.04	71458	14178	1236.20	18.30	55.81	37.59	182641.00
Вешняки	ВАО	Муниципальный округ	45263555	45302000	10.72	122285	11407	1976.80	16.20	55.73	37.82	147352.00
Внуково	ЗАО	Муниципальный округ	45268552	45317000	17.42	25471	1462	416.60	17.80	55.61	37.30	113399.00
Войковский	САО	Муниципальный округ	45277565	45336000	6.61	70729	10700	1531.00	23.10	55.82	37.49	207242.00
Восточное Дегунино	САО	Муниципальный округ	45277568	45337000	3.77	98923	26239	1592.50	16.70	55.88	37.56	146300.00

Fig.1 represents DF city data ready for analysis.

DATA SET

Venues data

By using API of Foursquare was obtained data about venues type and location. It takes special procedure and many iteration to obtain it about the all city districts.

Cell_id	Venue_Id	Borough_Name	Venue_Name	Venue_Latitude	Venue_Longitude	Venue_Category_Name
55.7020821…	511629f5e4b051a081439bf5	Очаково-Матвеевское	"Aminevskoe hotel" restaurant	55.703032	37.454590	Hotel
55.8350558…	5023841de4b0e6fe1a411c7d	Ростокино	"Cosmos 2" Hotel	55.836780	37.665548	Hotel
55.8277624…	505f30d2e4b0d9a2f19a319d	Покровское-Стрешнево	"Karaoke&Bar G-Voice"	55.827876	37.409241	Karaoke Bar
55.6864545…	4efb158da17cdc15b40b98fc	Очаково-Матвеевское	"MOON"	55.686766	37.414477	Furniture / Home Store
55.7213688…	5905a5870123587260ffe1d5	Южнопортовый	"Mime" Film Company (Мим Кинокомпания)	55.722946	37.679820	Film Studio
55.7488985…	5083dcc4e4b0ba1a3249d19f	Вешняки	"Red House" Клуб-Сауна	55.746088	37.838734	Sauna / Steam Room
55.7454108…	50eadc9de4b02662c430d51c	Новокосино	"Александр"	55.744217	37.877648	Department Store
55.7366957…	4eb12a04b63434fc86fa3310	Дорогомилово	"Аргумент - кафе"	55.738145	37.532077	Restaurant
55.7143244…	53a02544498e62c556da1f3f	Хамовники	"Банкет Холл" Лужники	55.715131	37.547142	Russian Restaurant
55.8692166…	5299878d11d2d1319ecea89f	Северное Тушино	"Бегемотики"	55.870727	37.440701	Kids Store
55.7623045…	50162ce6e4b01bcdb30b45e0	Крылатское	"Беговая дорожка" в Крылатском	55.762294	37.416648	Athletics & Sports
55.6249294…	4d877bec99b78cfaf7f5f91f	Орехово-Борисово Севе…	"Борисовский" билиардная	55.624427	37.709809	Bar
55.7949991…	503ccbf9e4b0708fceeb8ad1	Строгино	"Веселуха"	55.795756	37.405038	Dance Studio
55.8866119…	50420be2e4b0b5223de4c8a5	Дмитровский	"Волчий лес" / "Wolf Wood"	55.885273	37.528364	Café
55.6367977…	4f2c1f33e4b0ecad92a8352c	Коньково	"Гермес"	55.639274	37.544578	Convenience Store
55.6645507…	4f6a1b18e4b0ed0504f11293	Марьино	"Городская аптека"	55.662385	37.773821	Pharmacy
55.8777268…	50fbfea6e4b09f8ff7c27c93	Куркино	"Золотые Дуги"	55.880515	37.396922	American Restaurant
55.7902398…	4d43cae40349224b7365f34e	Восточное Измайлово	"Измайловский СДС" Филиал ГУП "Мосзеленх…	55.793075	37.823913	Flower Shop
55.7110205…	56b5e6ed498e16a72e900561	Даниловский	"Комус"	55.709422	37.657847	Paper / Office Supplies Store
55.8952978…	5558da32498ed73c64236d90	Лианозово	"Лавочки"	55.896766	37.580660	Park
55.8951981…	4ead5cf729c2a9bb97952c9e	Дмитровский	"Левый Берег" торговый центр	55.895344	37.503386	Shopping Mall
55.6521319…	4ea54de79adff6343ad6ff45	Тропарёво-Никулино	"Леди & Бродяга"	55.651273	37.470040	Pet Store
55.6833684…	51f7c3b0498e305d9ef6b5b2	Некрасовка	"Магнит"	55.683751	37.928274	Supermarket
55.8798507…	541c4831498e76f1b432ffee	Ярославский	"Магнит"	55.878228	37.729744	Supermarket
55.6628188…	51bea6bf498ea7d17efe1403	Люблино	"Мекона" Сервис	55.661802	37.807258	Auto Workshop

Fig.2 represents DF venues data ready for analysis.

METHODOLOGY

Exploratory Data Analysis

We have following key features in Moscow Boroughs dataset:

- District - name of the Moscow District in which Borough is belong to
- Area - area of the Moscow Borough in square kilometers
- Population_Density - population density of the Moscow Borough
- Housing_Area & Housing_Price of the Moscow Borough

For features analyzing were used:

- descriptive statistical analysis
- categorical variables analysis
- correlation analysis

Fig.3 shows basic statistics for all features.

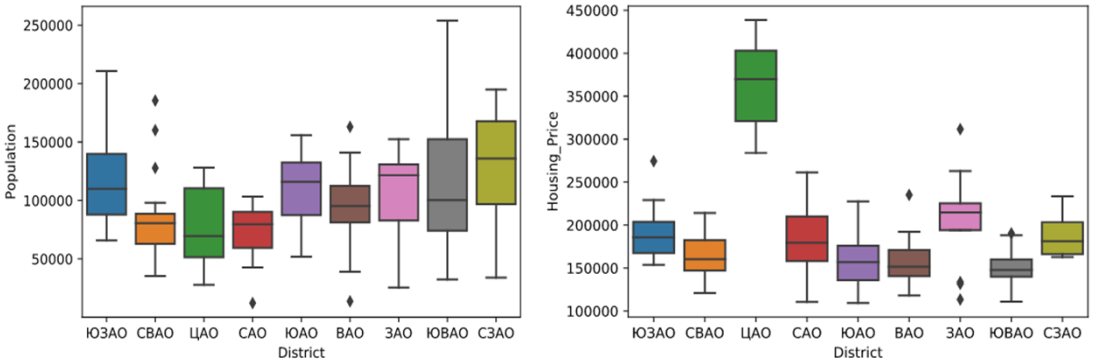
	Area	Population_Density	Housing_Area	Population	Housing_Price
count	120.000000	120.000000	120.000000	120.000000	120.000000
mean	8.706417	13426.608333	1775.684167	99847.608333	190037.316667
std	4.927028	5956.551611	815.978445	44024.992123	66182.885601
min	2.110000	559.000000	69.900000	12194.000000	109421.000000
25%	5.395000	9745.750000	1244.450000	71821.750000	147339.000000
50%	7.680000	13266.000000	1709.450000	93892.000000	168172.500000
75%	10.282500	17151.000000	2206.600000	126545.750000	210978.000000
max	27.570000	30428.000000	4523.000000	253943.000000	438568.000000

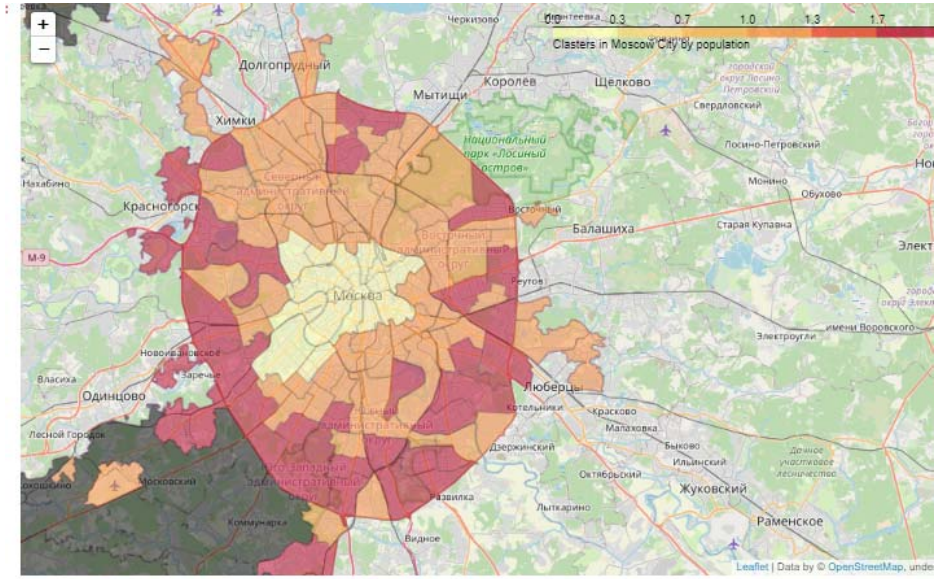
METHODOLOGY

Categorical variables analysis

Relationship between categorical feature 'District' and key criteria is shown using boxplots visualization.

We can see that the distributions of Population between Boroughs in the different Districts have an overlap, most populated Boroughs are placed in 'ЮЗАО', 'ЮАО', 'СЗАО' and 'ЗАО' Districts. It is shown on a map.





METHODOLOGY

Correlation analysis

- Correlation between 'Area', 'Population_Density' and 'Population' is statistically significant, although the linear relationship isn't extremely strong.
- Correlation between 'Housing_Are' and 'Population' is statistically highly significant, and the linear relationship is extremely strong.
- Correlation between 'Area', 'Population_Density', 'Housing_Area' and 'Housing_Price' is not statistically significant, although the linear relationship isn't strong.
- Correlation between 'Area' to 'Population_Density' is statistically hugely significant, and the linear relationship is extremely strong.
- So we can exclude 'Population_Density' from our considerations.

	Area	Population_Density	Housing_Area	Population	Housing_Price
Area	1.000000	-0.585991	0.344188	0.380587	-0.154996
Population_Density	-0.585991	1.000000	0.289456	0.338621	-0.101348
Housing_Area	0.344188	0.289456	1.000000	0.887856	-0.016971
Population	0.380587	0.338621	0.887856	1.000000	-0.195774
Housing_Price	-0.154996	-0.101348	-0.016971	-0.195774	1.000000

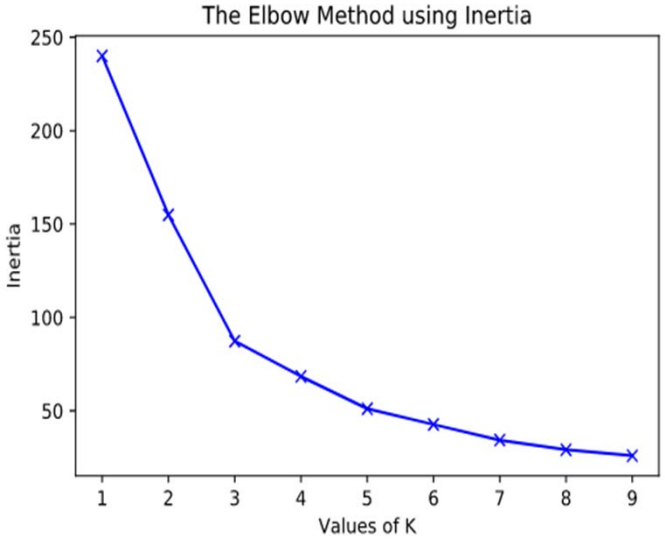
Fig.7 shows correlation between features.

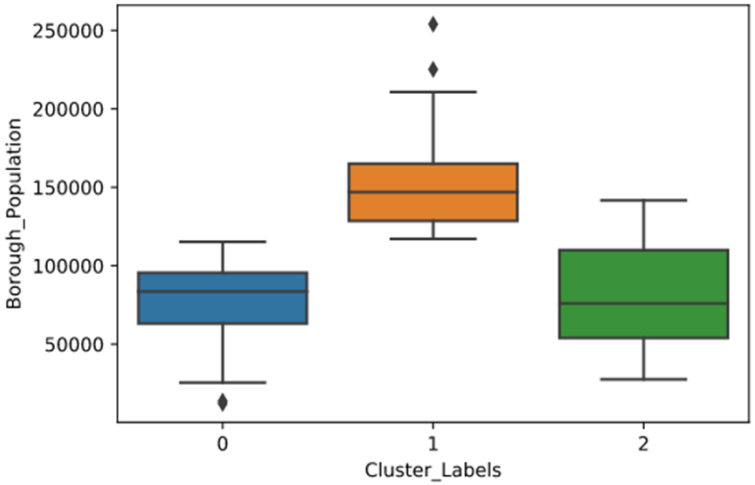
METHODOLOGY

Clustering

It is segmentation with K-Means clustering to detect Boroughs that have highest population and smallest housing price.

- For determining right number of clusters elbow method is used. Picture shows elbow method using Inertia with elbows at 3 and 5 centroid. It is used 3 centroid in my research.





METHODOLOGY

Clustering

There are 3 clusters:

- “1" Cluster - characterized by low mean population (78538 people per Borough), relatively high mean housing price (173695 rubles/m²) and low population density (10328 people/km²)
- “2" Cluster - characterized by highest mean population (153187 people per Borough), smallest mean housing price (160741 rubles/m²) and highest population density (13312 people/km²)
- “0" Cluster - characterized by low mean population (79805 people per Borough), highest mean housing price (333794 rubles/m²) and low population density (10533 people/km²)

```
# Print clusters
Moscow_Clustering_df.head()
```

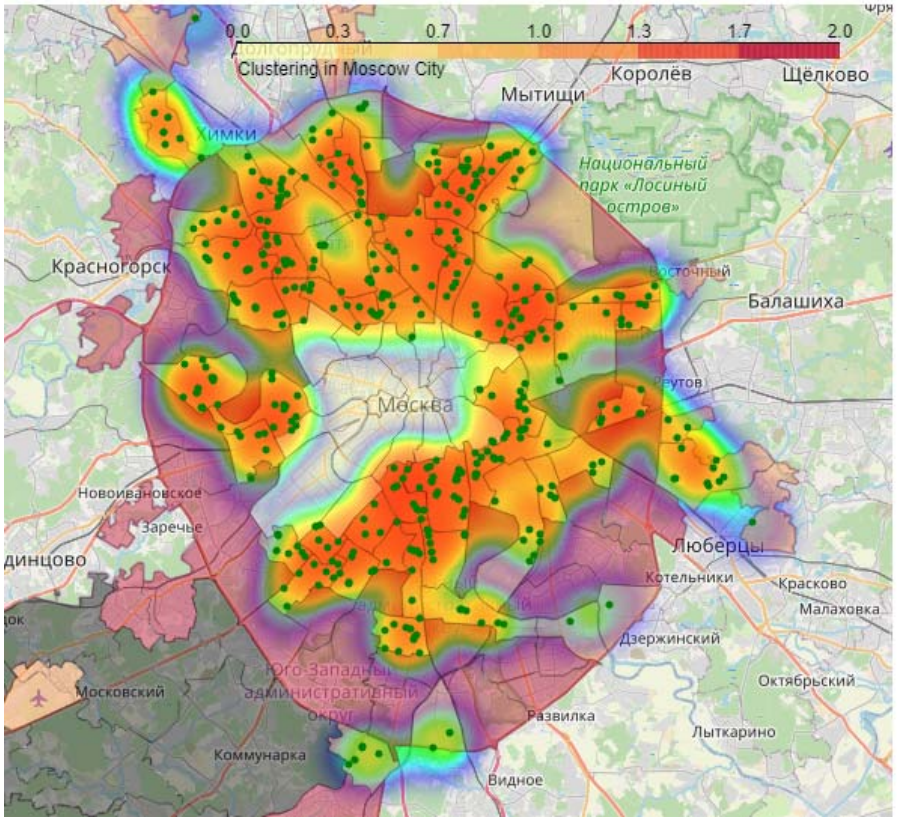
	Cluster_Labels	Population_Mean	Housing_Price_Mean	Population_Sum	Population_%	Borough_Count	Area_Sum	Area_%	Population_Density
0	0	79805.666667	333794.866667	1197085	9.990934	15	113.65	10.877992	10533.084030
1	1	78538.901408	173695.070423	5576262	46.539773	71	539.87	51.673574	10328.897698
2	2	153187.235294	160741.323529	5208366	43.469294	34	391.25	37.448434	13312.117572

RESULTS

The result of my research consists of:

- List of the optimal Boroughs for the location of facilities centers, according to the main criterias
  - high population of the borough
  - low cost of real estate in the borough
- List of the other competitive facilities in the each Borough from the optimal list
- Interactive choropleth map and heatmap with other competitive facilities in the each Borough.

In our case there are 422 venues of "Auto Workshop" of all 20864 venues in Moscow City. There are 419 venues of all Auto Workshop in one Cluster. That means there is a reason for it!



DISCUSSION

This project helps an Investor to make an optimal DataDriven choice for the interesting location and type of venue to be demanded in this location. It can be used to make competitive venue analysis.

```
neighbourhoods_venues_sorted.head(10)
```

	Borough_Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Академический	Pharmacy	Coffee Shop	Park	Auto Workshop	Bakery	Health Food Store	Wine Shop	Shoe Store	Dance Studio	Supermarket
1	Алексеевский	Auto Workshop	Park	Supermarket	Pizza Place	Hotel	Food & Drink Shop	Coffee Shop	Pet Store	Convenience Store	Pharmacy
2	Алтуфьевский	Supermarket	Auto Workshop	Light Rail Station	Bus Station	Health Food Store	Pizza Place	Eastern European Restaurant	Shoe Store	Pedestrian Plaza	Park
3	Арбат	Coffee Shop	Bakery	Hostel	Hotel	Museum	Concert Hall	Plaza	Gym / Fitness Center	Caucasian Restaurant	Bar
4	Аэропорт	Coffee Shop	Café	Cosmetics Shop	Pharmacy	Park	Wine Shop	Bakery	Salon / Barbershop	Food & Drink Shop	Italian Restaurant
5	Бабушкинский	Park	Pharmacy	Gym	Supermarket	Bus Stop	Gym / Fitness Center	Baby Store	Food & Drink Shop	Fast Food Restaurant	Café
6	Басманный	Coffee Shop	Café	Caucasian Restaurant	Dance Studio	Bar	Bookstore	Gym / Fitness Center	Art Gallery	Beer Bar	Clothing Store
7	Беговой	Coffee Shop	Dance Studio	Gym / Fitness Center	Café	Restaurant	Bar	Hotel	Nightclub	Sandwich Place	Pizza Place
8	Бескудниковский	Bus Stop	Bus Line	Pizza Place	Supermarket	Bookstore	Pharmacy	Gym	Japanese Restaurant	Shop & Service	Eastern European Restaurant
9	Бибирево	Supermarket	Park	Bus Stop	Pharmacy	Gym	Sushi Restaurant	Health Food Store	Gym / Fitness Center	Soccer Field	Fast Food Restaurant