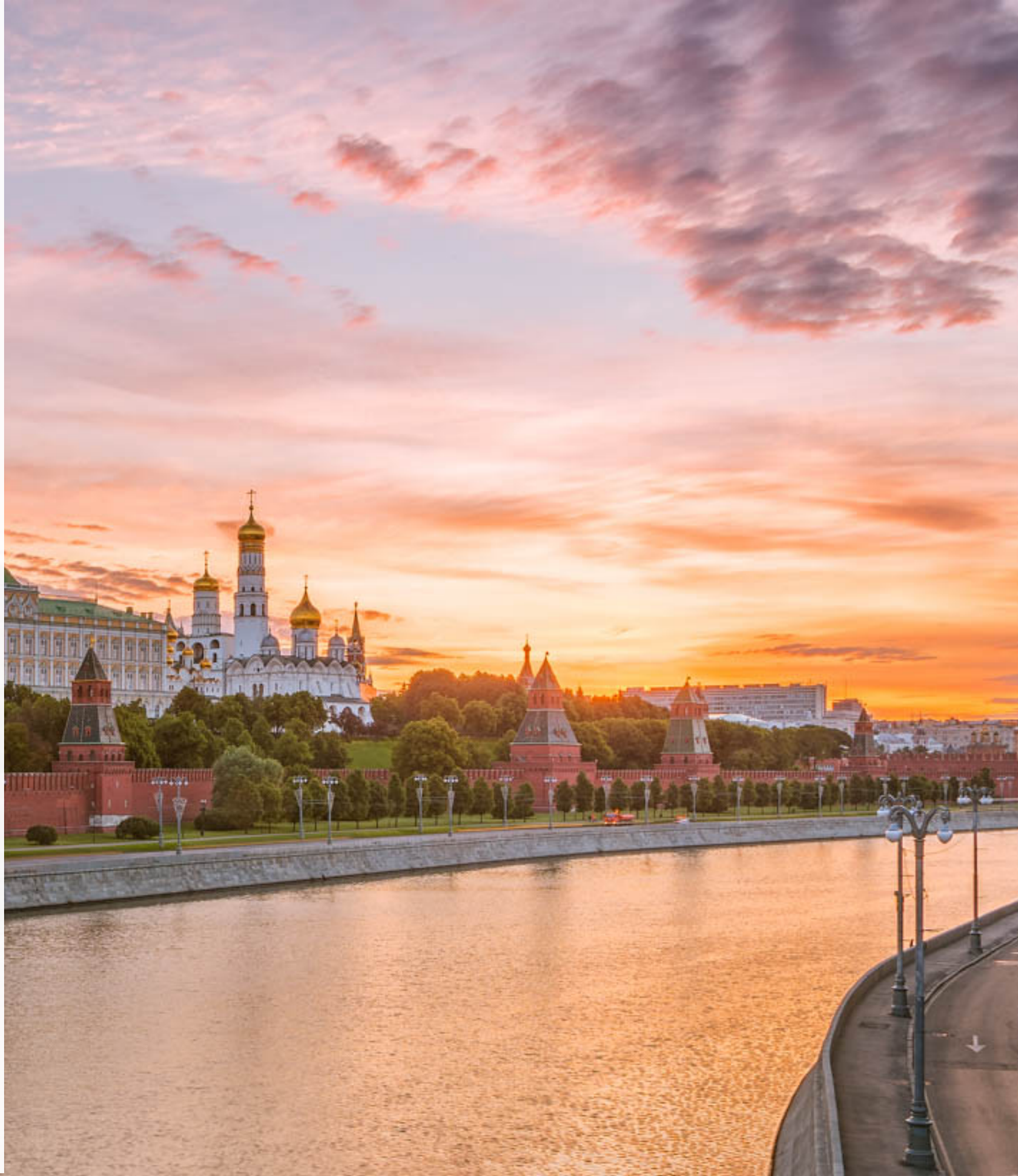
The Battle of Neigbourhoods – where to open a venue?

IBM Coursera Capstone Project

Problem Background

- Moscow is the **economic centre** with the high density of population and non-declining number of incoming labour from overall Russia as well as other countries appeals new businesses in both traditional and innovative sectors.
- We **majorly** address problem to the **local small or middle-size businesse**s that are concentrated on **'venue'-type** businesses and where questions as following can arise:
  - ✓ Based on average income of the area, which city do we choose?
  - ✓ What is the situation in borrows and whom can we target?
  - ✓ What venues types are in the area and which ones are most popular?

... and so many more. Therefore, we to help answer questions that certainly arise when there exists a consideration of establishing a new venue to enrich strategical aims based on data.



Data to crunch

- **Forsquare API** to get venues and details of a chosen city
- As up-to-date granular data for income distribution in Moscow was not available, therefore we use average house renting prices per m^2 from **Domofond.ru** within districts
- We use **Wikipedia** page on Moscow population data by Districts
- When the venue type is chosen we analyse the use sorted **CIAN** data on available commercial areas to rent

EXTRACTED AND CLEANED DATA EXAMPLES:

	Borough	District Name
0	Западный	Филёвский Парк
1	Зеленоградский	Матушкино
2	Западный	Внуково
3	Зеленоградский	Савёлки
4	Зеленоградский	Силино
...	...	...
120	Восточный	
121	Восточный	
122	Восточный	
123	Юго-Восточный	
124	Западный	

125 rows × 2 columns

	District Name	ATM	Adult Boutique	American Restaurant	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Art Studio	Arts & Crafts Store	...	Video Store	Viet Res
0	Академический	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000	...	0.0	0.0
1	Алексеевский	0.0	0.0	0.000000	0.045455	0.0	0.000000	0.000000	0.0	0.000000	...	0.0	0.0
2	Алтуфьевский	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000	...	0.0	0.0
3	Арбат	0.0	0.0	0.010101	0.010101	0.0	0.020202	0.010101	0.0	0.010101	...	0.0	0.0
4	Аэропорт	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000	...	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
88	Южное Тушино	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000	...	0.0	0.0
89	Южнопортовый	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000	...	0.0	0.0
90	Якиманка	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.016667	...	0.0	0.0
91	Ярославс												
92	Ясенево												

	Borough	District Name	Rent	Latitude	Longitude	Area	Population	Population density
108	Южный	Зябликово	789	55.621900	37.742900	438	133278.0	30428.77
120	Восточный	Новокосино	765	55.740167	37.861725	360	107907.0	29974.17
11	Юго-Западный	Ломоносовский	976	55.678778	37.533239	334	88320.0	26443.11
52	Северный	Восточное Дегунино	780	55.880100	37.557600	377	98923.0	26239.52
69	Северо-Восточный	Бибирево	789	55.893800	37.611100	645	160163.0	24831.47
46	Северный	Бескудниковский	826	55.865900	37.553300	330	79603.0	24122.12
55	Юго-Западный	Зюзино	889	55.653100	37.598400	545	126815.0	23268.81
78	Северо-Восточный	Северное Медведково	833	55.888000	37.645500	566	127819.0	22582.86
111	Восточный	Новогиреево	799	55.748154	37.804108	445	98415.0	22115.73
76	Северо-Восточный	Южное Медведково	778	55.871000	37.638300	388	85698.0	22087.11
53	Северный	Савёловский	988	55.801600	37.564700	270	59287.0	21958.15
39	Юго-Западный	Коньково	921	55.643414	37.530588	718	156389.0	21781.20
101	Юго-Восточный	Марьино	782	55.652664	37.744774	1191	253943.0	21321.83
100	Южный	Орехово-Борисово Южное	913	55.604264	37.733132	694	147789.0	21295.24
64	Южный	Чертаново Северное	857	55.634300	37.603500	540	114548.0	21212.59

Methodology

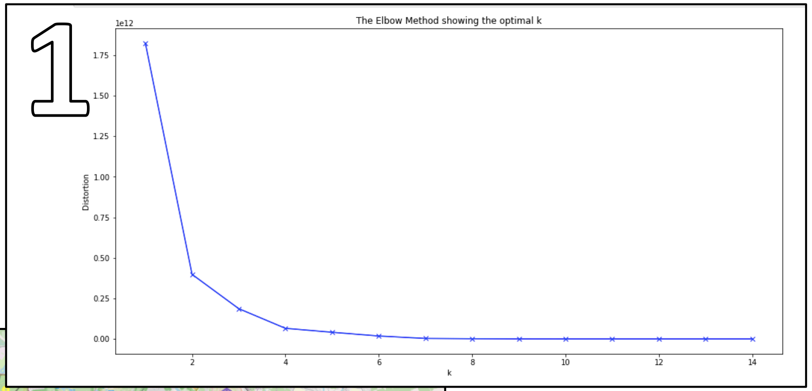
- Idea is to get overall data on renting for living and analyse it as an alternative to the income data. Then we want to pick the most lucrative city for analysis. We concentrate on Moscow (3rd place) as the two highest seem to be ranked so due to some values in the average that upscaled the overall pricing.
- For the sake of better understanding we want check how many districts are there in each borough. We further proceed with the highest average renting price per borough.
- Then we change the granularity to the level of districts to be more precise in our analysis and draw a heatmap with average renting prices per district. Then we explore population density to get a better understanding of how many customers we can serve
- We get venue data and first look at the overall picture, e.g. what is present more frequently among. We proceed with clustering of this big groups
- At the end, we choose one particular venue based on analysis and own interest and add pricing of the commercial renting rent for the business building and analyse which location would be the optimal for opening a pharmacy based on given data and conditions. We cluster and build population density heatmap with these clusters. Optimal k for clustering is obtained via elbow-method

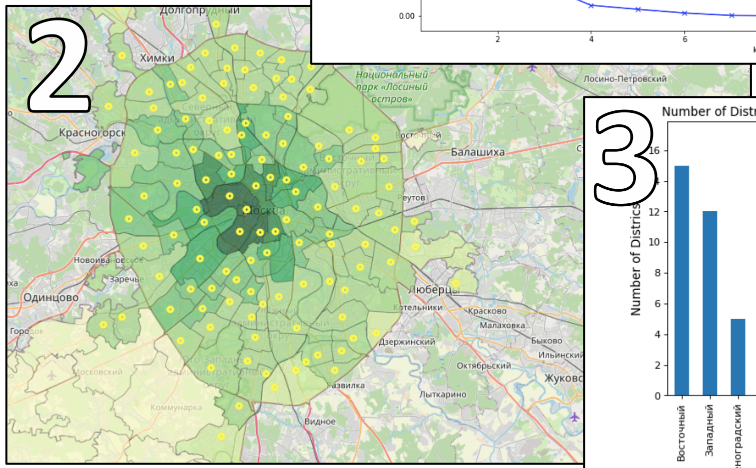
OUTCOMES EXAMPLES:

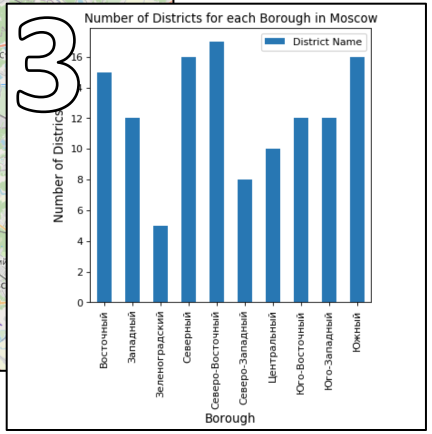
1. Elbow method

2. Average renting price heatmap

1. # of Districts per Borough



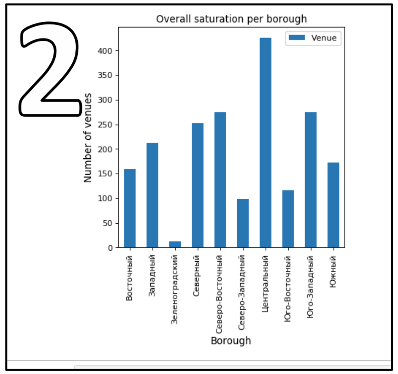


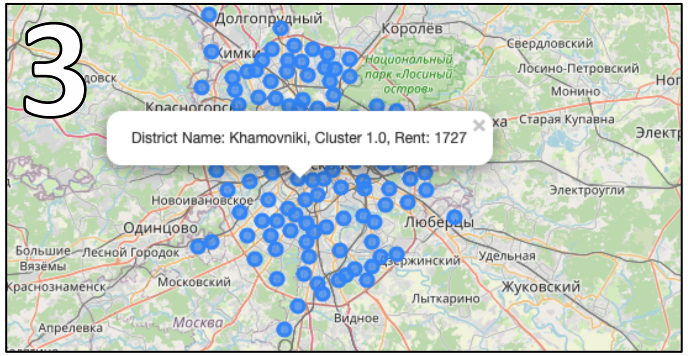


Outcome for general case

1

	Cluster Labels	District 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
	1	Академический	Coffee Shop	Pharmacy	Health Food Store	Sporting Goods Shop	Beer Store	Dance Studio	Park	Optical Shop	Bookstore	Sushi Restaurant
1	1	Алексеевский	Recording Studio	Supermarket	Pizza Place	Auto Workshop	Historic Site	Gym / Fitness Center	Mobile Phone Shop	Food & Drink Shop	Office	Farmers Market
2	0	Алтуфьевский	Supermarket	Lake	Gym / Fitness Center	Hotel	Pub	Dry Cleaner	Café	Grocery Store	Cafeteria	Salon / Barbershop
3	1	Арбат	Coffee Shop	Hotel	Yoga Studio	Hostel	Bakery	Museum	Caucasian Restaurant	Gym / Fitness Center	Italian Restaurant	Flower Shop
4	1	Аэропорт	Coffee Shop	Salon / Barbershop	Martial Arts Dojo	Smoke Shop	Gourmet Shop	Clothing Store	Pet Store	Pharmacy	Asian Restaurant	Caucasian Restaurant
...	...	...	...	...	...	...	...	...	...	...	...	...
88	1	Южное Тушино	Pizza Place	Flower Shop	Sushi Restaurant	Convenience Store	Park	Pub	Coffee Shop	Food & Drink Shop	Movie Theater	Electronics Store
89	1	Южнопортовый	Film Studio	Electronics Store	Dance Studio	Boutique	Bakery	Baby Store	Fried Chicken Joint	Café	Fast Food Restaurant	Farmers Market
90	1	Якиманка	Coffee Shop	Bakery	Gym / Fitness Center	Shoe Store	Bridal Shop	Donut Shop	Pub	Mobile Phone Shop	Karaoke Bar	Cosmetics Shop
91	1	Ярославский	Park	Fountain	Pizza Place	Shopping Mall	Café	Auto Workshop	Coffee Shop	Photography Studio	Concert Hall	Bus Stop
92	1	Ясенево	Fast Food Restaurant	Blini House	Pharmacy	Supermarket	Bookstore	Sporting Goods Shop	Burger Joint	Shopping Mall	Clothing Store	Coffee Shop



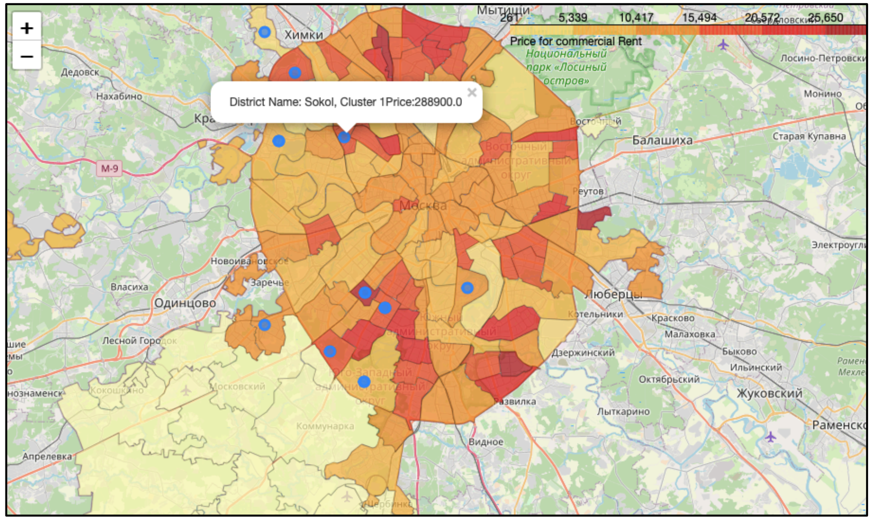


OUTCOMES EXAMPLES:

1. ***Most common venues***
2. ***Venue saturation per borough***
3. ***Cluster map***

By analyzing data obtained we can see that the pharmacy is present in both clusters (**Cluster 0**: Shops&Malls&All for living, **Cluster 1**: Being active and socializing). However, based on given data market is not oversaturated with Pharmacies, therefore we target it further as the business to open

Outcome for Pharmacy Case



The best suit seems to be the district Sokol as it has **high population density**, **pricing** is **reasonable**, and **area** of the commercial part is **spacious**. Sokol is also the district where the **mean rent** is slightly **above average**, therefore we assume that ability to pay of customers is also there. District is **not oversaturated** with pharmacies and seems to be the perfect match for the new spot and the distance to the metro is only **3 min** walk.

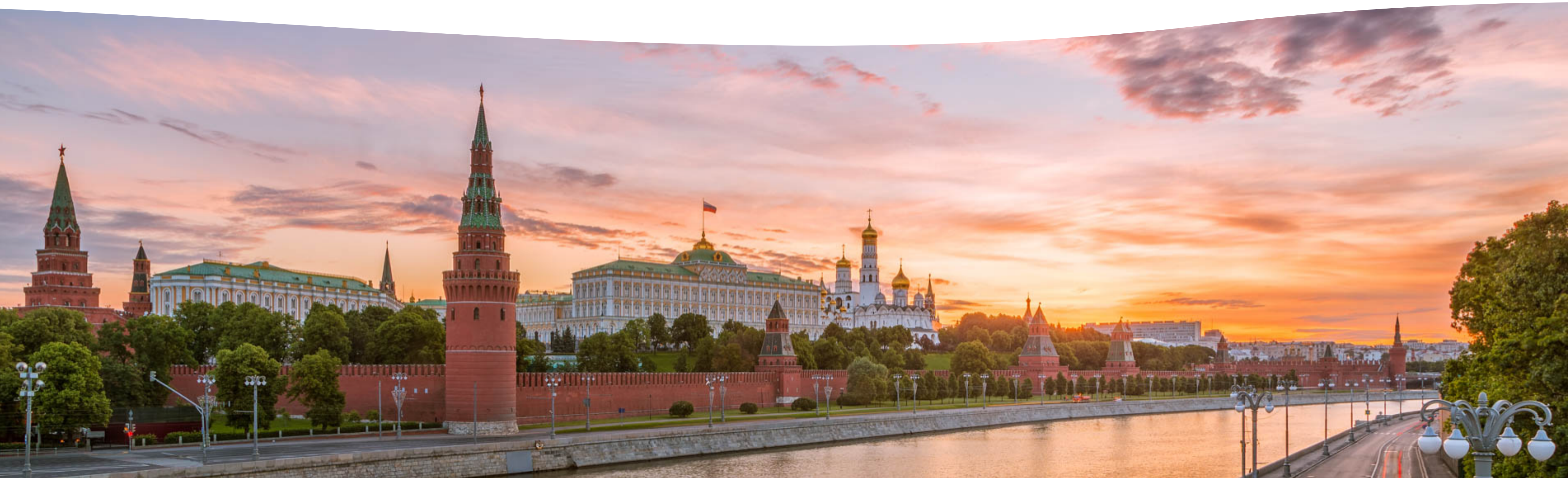
	District Name	Venue Latitude	Venue Longitude	Area of Com. Rent	Metro	Price	Borough	Rent	Area	Population	Population density
5	Тверской	55.773321	37.598090	106.0	м. Лубянка (4 мин пешком)	1100001.0	Центральный	1500	727	77864.0	10710.32
7	Тверской	55.773321	37.598090	114.0	м. Китай-город (7 мин пешком)	848999.0	Центральный	1500	727	77864.0	10710.32
10	Тверской	55.776630	37.607171	106.0	м. Лубянка (4 мин пешком)	1100001.0	Центральный	1500	727	77864.0	10710.32
12	Тверской	55.776630	37.607171	114.0	м. Китай-город (7 мин пешком)	848999.0	Центральный	1500	727	77864.0	10710.32
14	Якиманка	55.732885	37.613876	78.0	м. Новокузнецкая (2 мин пешком)	1250001.0	Центральный	1562	480	27672.0	5765.00
15	Якиманка	55.732885	37.613876	145.0	м. Третьяковская (7 мин пешком)	899000.0	Центральный	1562	480	27672.0	5765.00

	District Name	Venue Latitude	Venue Longitude	Area of Com. Rent	Metro	Price	Borough	Rent	Area	Population	Population density
0	Нагатинский Затон	55.683017	37.681513	85.0	м. Коломенская (3 мин пешком)	450005.0	Южный	918	980	120954.0	12342.24
1	Северное Тушино	55.859719	37.432788	65.0	м. Первомайская (14 мин пешком)	124995.0	Северо-Западный	827	940	165762.0	17634.26
2	Сокол	55.803670	37.509224	81.0	м. Войковская (3 мин на машине)	149000.0	Северный	952	372	59507.0	15996.51
3	Дорогомилово	55.737393	37.524963	100.0	м. Кутузовская (5 мин на машине)	449000.0	Западный	1172	795	76093.0	9571.45
4	Тверской	55.773321	37.598090	72.0	м. Белорусская (2 мин пешком)	480000.0	Центральный	1500	727	77864.0	10710.32
6	Тверской	55.773321	37.598090	86.0	м. Белорусская (2 мин пешком)	599899.0	Центральный	1500	727	77864.0	10710.32
8	Тверской	55.773321	37.598090	133.0	м. Менделеевская (2 мин пешком)	288900.0	Центральный	1500	727	77864.0	10710.32
9	Тверской	55.776630	37.607171	72.0	м. Белорусская (2 мин пешком)	480000.0	Центральный	1500	727	77864.0	10710.32
11	Тверской	55.776630	37.607171	86.0	м. Белорусская (2 мин пешком)	599899.0	Центральный	1500	727	77864.0	10710.32
13	Тверской	55.776630	37.607171	133.0	м. Менделеевская (2 мин пешком)	288900.0	Центральный	1500	727	77864.0	10710.32

Here we obtain two clusters, first - with expensive prices and second - with average prices.

Discussion for improvements

- First and foremost obtained results are highly dependant of the quality and completeness of data which in some cases is under a big question mark.
- Although we obtained a neat and nice result, it could have been better whether we used dynamic data to show the trends in, for instance, changes in rents and therefore, could have made a better prediction.
- Moreover, whether we had access to fiscal data of the pharmacies we could have done a more rigorous competitor analysis and may have developed some set of better actions
- The granularity of districts although is not at an extremely high level but is still not enough to make a precise analysis
- One could have used rankings for the pharmacies to obtain understanding of performances nearby
- K-means is although a good clustering algorithm but there exist others that can help to obtain more precise results



Conclusion

Overall this project overlooked at the **initial** steps to be undertaken for conservative local businesses to halt usage of purely the 'guts' feeling but also to utilize the technologies to **enforce a better decision making based on data** which is sooner or later will become inevitable part of each business