**Finding the Best Places to Open a Coffee Shop in Moscow**

(IBM Capstone project on Coursera)

Sargylana Skrybykina

February 08, 2020

# Introduction

Moscow is the capital and most populous city of Russia, with approximately 12.6 million residents within city limits. The city is served by a transit network, which includes four international [airports](https://en.wikipedia.org/wiki/Airports), nine railway terminals, numerous trams, [a monorail system](https://en.wikipedia.org/wiki/Moscow_monorail) and one of the deepest underground rapid transit systems in the world, called the [Moscow Metro](https://en.wikipedia.org/wiki/Moscow_Metro), the fourth-largest in the world and largest outside [Asia](https://en.wikipedia.org/wiki/Asia) in terms of passenger numbers, and the busiest in Europe. Moscow Subway (Metro) system covers the whole urban area, hence we could use location of subway station to divide city into neighborhoods.

Coffee shops are extremely popular in Moscow. People use them to pick up coffee, meet up or to work there. The business interest is to find out the best places to open coffee shop in Moscow.

In this project, the Moscow urban area will be clustered by the most common venues surrounding each subway station. Using this data, we can define the clusters and figure out the best places to open a Coffee shop in Moscow.

# Data Collection

# Data resources

The list of subway stations and their locations are stored in [Wikipedia](https://en.wikipedia.org/wiki/List_of_Moscow_Metro_stations). Foursquare API is used to get the most common venues.

# Data collecting and preprocessing

Data Processing includes following steps

1. Re-format wiki data in tabular form for further analysis. This includes unique subway station name and location in longitude and latitude
2. Re-format longitude and latitude from wiki to work with Foursquare data
3. Get venues near subway station using Foursquare API
4. Clean and get the most common venues per subway stations

Subway locations downloaded from Wiki are visualized on Figure 1. The station that have multiple exits are merged into 1 station – the longitude and latitude are calculated as a mean.

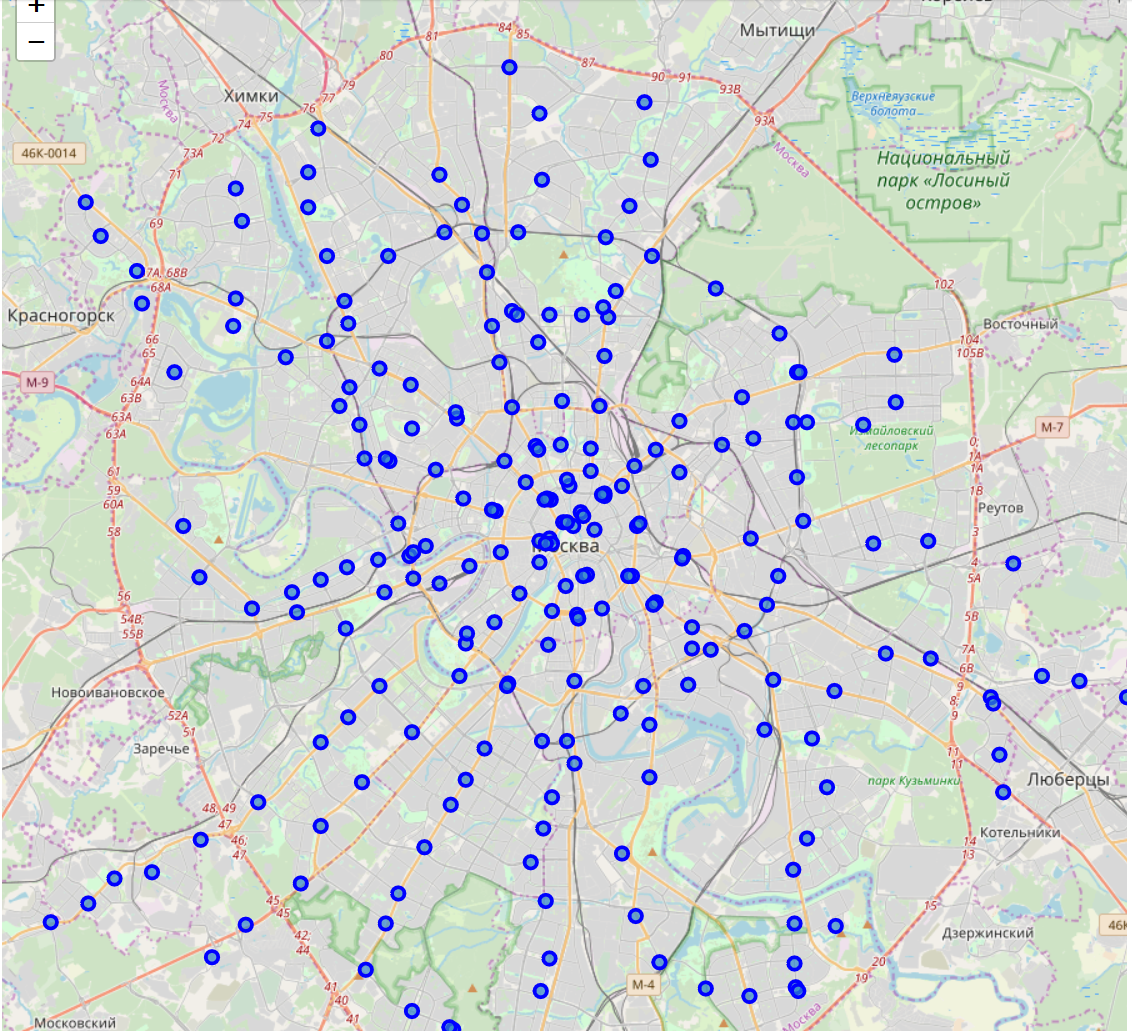


Figure Moscow Subway (Metro) Stations locations

The venues near each station have been mapped using Foursquare API calls. It contains up to 40 venues in a radius 750 meters.

The data for further analysis includes subway station name, location (latitude, longitude), venue name, venue location (longitude, latitude), and venue category. Example in the below Table 1.

Table Venues near Moscow Subway stations example data

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Index | Station | Station Latitude | Station Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
| 0 | Bulvar Rokossovskogo | 55.814722 | 37.734167 | Подружка | 55.814523 | 37.736781 | Cosmetics Shop |
| 1 | Bulvar Rokossovskogo | 55.814722 | 37.734167 | Burger King | 55.814026 | 37.733659 | Fast Food Restaurant |
| 2 | Bulvar Rokossovskogo | 55.814722 | 37.734167 | Галерея Богородское | 55.813944 | 37.734298 | Art Gallery |
| 3 | Bulvar Rokossovskogo | 55.814722 | 37.734167 | Fitlab studio | 55.810845 | 37.727964 | Gym / Fitness Center |
| 4 | Bulvar Rokossovskogo | 55.814722 | 37.734167 | Ветеринарная Аптека | 55.813464 | 37.735036 | Pet Store |

Initial data contains 232 Moscow Subway (Metro) stations and 375 unique categories, such as Cosmetics Shop, Fast Food Restaurant, Gym / Fitness Center, Art Gallery, Pet Store, Vietnamese Restaurant, Auto Workshop, Candy Store, Mobile Phone Shop, Park, Coffee Shop, Bookstore etc.

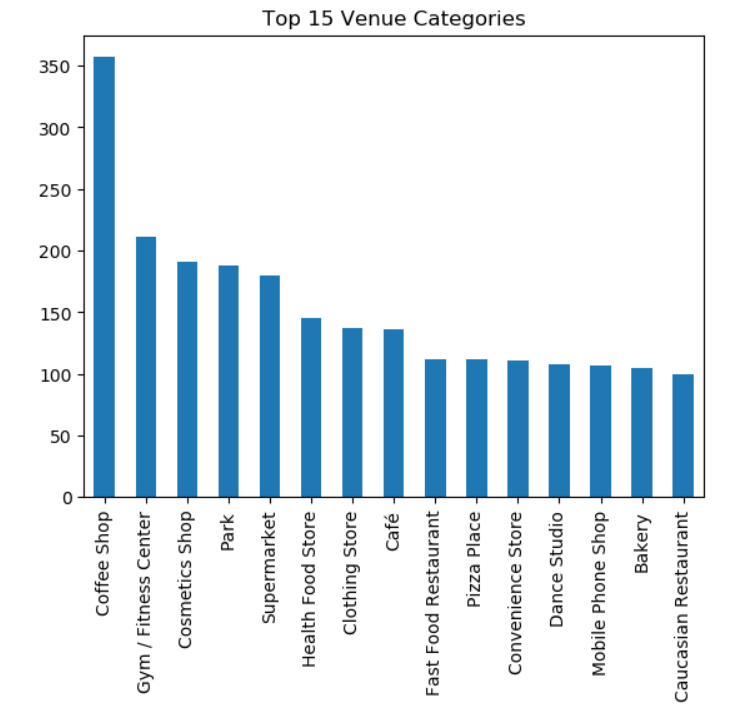


Figure Top 15 Popular Venue Categories

There are stations with just few registered venues in Foursquare (Figure 3). These stations are surrounded mostly by park, field, and forest. These are not considerable for opening the Coffee Shop; hence these are filtered out from further analysis as outliers.

Furthermore, some Venue Categories are very rare - these are excluded from further analysis as well.

Final data includes 223 stations, 157 unique categories, and 7084 total venues.



Figure Less demand locations

Most common venues for each station are identified

Table Most Common Venues data frame example

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Station | **Latitude** | **Longitude** | **1st Most Common Venue** | **2nd Most Common Venue** | **3rd Most Common Venue** | **4th Most Common Venue** | **5th Most Common Venue** |
| Aeroport | 55.80028 | 37.53278 | Coffee Shop | Cosmetics Shop | Pet Store | Gourmet Shop | Hockey Arena |
| Akademicheskaya | 55.68778 | 37.57333 | Park | Coffee Shop | Bakery | Sushi Restaurant | Beer Store |
| Aleksandrovsky Sad | 55.7525 | 37.60861 | Plaza | Concert Hall | Park | Art Gallery | Restaurant |
| Alekseyevskaya | 55.80889 | 37.63889 | Coffee Shop | Hobby Shop | Health Food Store | Furniture / Home Store | Auto Workshop |
| Alma-Atinskaya | 55.6325 | 37.76611 | Supermarket | Convenience Store | Cosmetics Shop | Concert Hall | Pharmacy |

# Clustering Moscow using K-mean algorithm

# K-mean algorithm

**K -mean** is a simple unsupervised machine learning algorithm is used for partitioning clustering. It groups the data into non-overlapping subsets (clusters) without any cluster-internal structure. K is a specified number of clusters. The **elbow method** is **used** to determine the optimal number of clusters. The elbow method plots the value of the cost function produced by different values of k. As you know, if k increases, average distortion will decrease, each cluster will have fewer constituent instances, and the instances will be closer to their respective centroids. However, the improvements in average distortion will decline as k increases. The value of k at which improvement in distortion declines the most is called the elbow, at which we should stop dividing the data into further clusters.

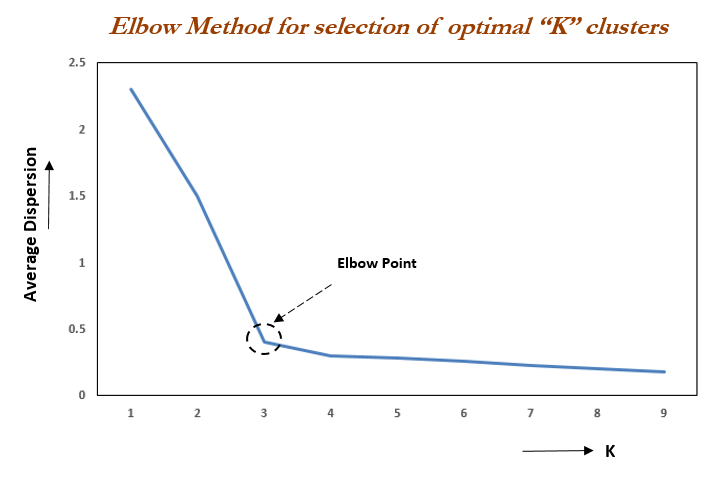


Figure Elbow method to determine the optimal number of clusters (k)

# In this project k is projected as 6.

# 

Figure Defining k using yellobrick KElbowVisualizer library

# Observation and results

K-mean algorithm has grouped Moscow in 6 clusters – you can see on Figure 6.

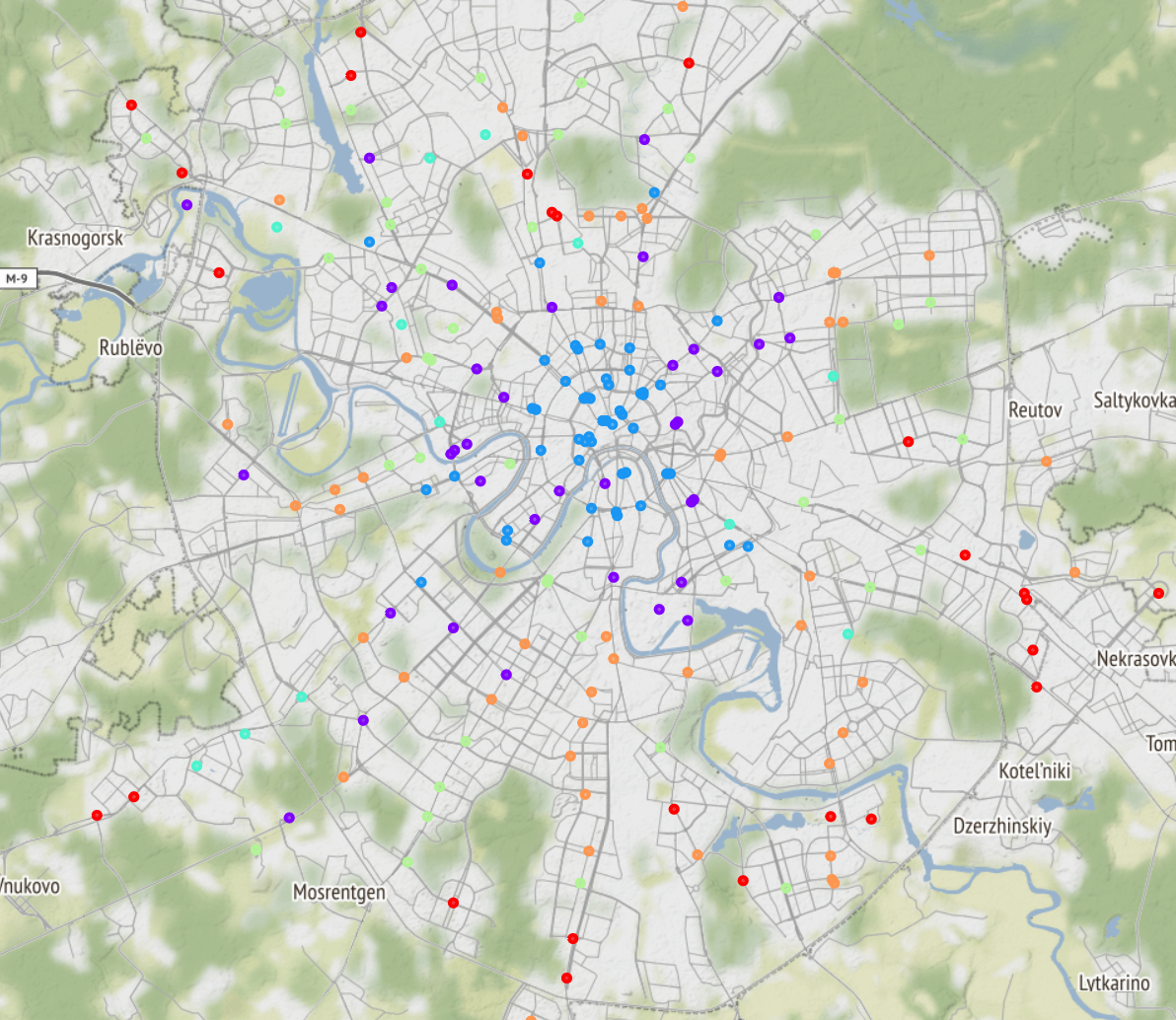


Figure Clustering Moscow using K-mean algorithm: Cluster 0 - Red, Cluster 1 - Violet, Cluster 2 - Blue, Cluster 3 - Mint , Cluster 4 - Green, Cluster 5 - Orange

Cluster 0 (Red)  is a living area that is far away from the center of Moscow. It is surrounded mostly by supermarkets, pharmacies. Most likely new residential neighborhoods with developing infrastructure.

Cluster 1 (Violet) is a business area. It is full of entertainments such as gym, SPA, and lots of coffee shops.

Cluster 2 (Blue)  is not only geographical center of Moscow, but also cultural and shopping center. It attracts tons of tourists and locals. It is surrounded by hotels, restaurants, coffee shops and historical sights.

Cluster 3 (Mint) is most likely industrial area. Less venue categories, mostly auto repair shops.

Cluster 4 (Green) is a local business area surrounded by mostly local cosmetic, and clothing stores.

Cluster 5 (Orange)  is a mix of local business and residential area. It has good infrastructure and mostly surrounded by parks, gyms, fitness centers, and restaurants

Table Most common venues in clusters (5 examples to each cluster)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Neighborhood** | **Cl** | **1st Most Common Venue** | **2nd Most Common Venue** | **3rd Most Common Venue** | **4th Most Common Venue** | **5th Most Common Venue** |
| Alma-Atinskaya | 0 | Supermarket | Convenience Store | Cosmetics Shop | Concert Hall | Pharmacy |
| Annino | 0 | Auto Workshop | Fast Food Restaurant | Supermarket | Baby Store | Electronics Store |
| Babushkinskaya | 0 | Supermarket | Fast Food Restaurant | Cosmetics Shop | Gym / Fitness | Gym |
| Belomorskaya | 0 | Pizza Place | Sporting Goods | Supermarket | Coffee Shop | Gym / Fitness |
| Borisovo | 0 | Supermarket | Park | Convenience Store | Brewery | Bus Stop |
| Aeroport | 1 | Coffee Shop | Cosmetics Shop | Pet Store | Gourmet Shop | Hockey Arena |
| Alekseyevskaya | 1 | Coffee Shop | Hobby Shop | Health Food Store | Furniture / Home Store | Auto Workshop |
| Avtozavodskaya | 1 | Coffee Shop | Beer Store | Dance Studio | Brewery | Ice Cream Shop |
| Baumanskaya | 1 | Coffee Shop | Dance Studio | Bar | Caucasian Restaurant | Dessert Shop |
| Begovaya | 1 | Coffee Shop | Supermarket | Convenience Store | Hookah Bar | Other Repair Shop |
| Aleksandrovsky Sad | 2 | Plaza | Concert Hall | Park | Art Gallery | Restaurant |
| Arbatskaya | 2 | Bakery | Concert Hall | Art Gallery | Coffee Shop | Plaza |
| Barrikadnaya | 2 | Zoo Exhibit | Bakery | Caucasian Restaurant | Coffee Shop | Park |
| Belorusskaya | 2 | Coffee Shop | Seafood Restaurant | Steakhouse | Gym / Fitness | Restaurant |
| Biblioteka Imeni Lenina | 2 | Concert Hall | Plaza | History Museum | Art Gallery | Coffee Shop |
| Butyrskaya | 3 | Flower Shop | Café | Bus Stop | Ice Cream Shop | Soccer Field |
| Govorovo | 3 | Bus Stop | Fast Food Restaurant | Lingerie Store | Gym | Supermarket |
| Koptevo | 3 | Bus Stop | Liquor Store | Pet Store | Auto Workshop | Soccer Field |
| Likhobory | 3 | Bus Stop | Cafeteria | Convenience Store | Wine Shop | Pool |
| Ozyornaya | 3 | Café | Auto Workshop | Supermarket | Pizza Place | Plaza |
| Bagrationovskaya | 4 | Electronics Store | Cosmetics Shop | Middle Eastern Restaurant | Other Repair Shop | Café |
| Baltiyskaya | 4 | Coffee Shop | Clothing Store | Cosmetics Shop | Lingerie Store | Sporting Goods |
| Belyayevo | 4 | Coffee Shop | Baby Store | Cosmetics Shop | Gym | Supermarket |
| Bibirevo | 4 | Fast Food Restaurant | Clothing Store | Italian Restaurant | Furniture / Home Store | Gastropub |
| Bulvar Rokossovskogo | 4 | Bus Line | Dance Studio | Convenience Store | Sushi Restaurant | Brewery |
| Akademicheskaya | 5 | Park | Coffee Shop | Bakery | Sushi Restaurant | Beer Store |
| Altufyevo | 5 | Pharmacy | Health Food Store | Supermarket | Sushi Restaurant | Pizza Place |
| Aviamotornaya | 5 | Café | Gym / Fitness | Caucasian Restaurant | Pharmacy | Pizza Place |
| Bratislavskaya | 5 | Middle Eastern Restaurant | Pizza Place | Gym / Fitness | Beer Store | Kids Store |
| Bulvar Dmitriya Donskogo | 5 | Park | Health Food Store | Middle Eastern Restaurant | Restaurant | Cosmetics Shop |

# 

Figure 1st Most Common Venues

# 

Figure 2nd Most Common Venues

# Conclusion

In the first glance, the good places to open new coffee shops are Cluster 1, Cluster 2, and Cluster 5. Cluster 2 which is in the center of Moscow is the best place to open coffee shop, but the rent could be expensive. Cluster 1 is a business area and there is high traffic on weekdays. Cluster 5 is a mix of local business and residential areas, with high traffic all days.

I would rate cluster 5 as the best candidate. But also, I would consider cluster 2 (the center of Moscow). The next question to consider would be a rent fee and a traffic analysis.

# References

# List of Moscow Subway Stations - https://en.wikipedia.org/wiki/List\_of\_Moscow\_Metro\_stations

# Information about Moscow city - <https://en.wikipedia.org/wiki/Moscow>

# Foursquare Developer API - https://developer.foursquare.com/

# K-mean elbow method - <https://www.oreilly.com/library/view/statistics-for-machine/9781788295758/c71ea970-0f3c-4973-8d3a-b09a7a6553c1.xhtml>

# Sci-kit elbow method - <https://www.scikit-yb.org/en/latest/api/cluster/elbow.html>

# The project code on GitHub, data collection (week1) -<https://github.com/Sargylana/Clustering_Moscow_Data/blob/master/IBM%20Moscow_Data.ipynb>

# The project code on GitHub, data analysis (week 2) - <https://github.com/Sargylana/Clustering_Moscow/blob/master/IBM%20Moscow%20Clustering_WK2.ipynb>