Coursera

IBM Data Science Professional Certificate

Module IX, Week IV-V, Code Training Book

Final Assignment: Determining Possible Locations For Small Businesses In Moscow As A Way to Create New Workplaces After Pandemic Time

Introduction

Due to recent situation with COVID-19 the President of Russian Federation, V. Putin instated a "non-working" period at the end of March which is to run until May 31. In many regions and precisely in Moscow strict stay-at-home orders have been in place for the last two months, however as soon as Russia's capital city remains the most powerful and efficient in terms of economy and extremely high populated, stay-at-home measures were prolonged in the city until the June 14th. At the same time, according to "The Moscow Times" ministers told President Vladimir Putin that Russia's economic activity has fallen by 33% since the beginning of the coronavirus pandemic. It is important to mention that Moscow - where a lot of well-known, but also small companies are located and run their daily activities - is the biggest contributor to country's budget. But, given the actual situation, a lot of small and middle-size companies faced with unprecedented financial pressure and went bankrupt. This analysis, based on classification problem, can help individuals or households, dreaming of business set up or just trying to build small companies as a result of loosing the job, to determine, where - in other words - in which specific location in Moscow - it is a good idea to start a business or not.

In other words, I have opted for opportunity to solve the problem #2 listed in a task - identify recommendations for potential investors and businessmen if they are ready to set up a small business in Moscow as soon as a pandemic time is over.

Data Available

Small companies based in Moscow always generate "flow maps" - these maps literally show businessmen places where the demand can be generated and in what particular area it is reliable and efficient to place a shop or an office. In 80% of cases the best location is an area situated nearby Moscow metro stations. This is due to the reason that undergrounds remains the most popular, fast and cheap transport in Moscow and developing actively in the last couple of years. Taking into account this fact, I am going to look at the neighbourhoods surrounding underground stations and classify them. The fact is that some neighbourhoods are mostly residential, some have more business or commercial spaces nearby. The venues closest to a station determine why and how people use it - the is because people in Moscow tend to buy things - be it food or service - during their path to or from the nearest to their homes or offices metro stations.

In order to identify and accomplish my mission I have opted for a data presented on Wiki page: en.wikipedia.org/wiki/List_of_Moscow_Metro_stations. On this page you can find a complete list of Moscow Metro stations including the names, opening dates, types and, what is more important - coordinates. This is also the case why I have opted for this data: not so many sources in Moscow allow to generate latest geographical data for free.

In addition to the data presented above I will use Foursquare API to explore venue types located nearby each station. It will show how developed or underdeveloped is an area and, as a result, what potential business direction can be set up near different Moscow metro stations.

Methodology

To start with, I have created a data frame using the data located on the above mentioned Wikipedia page. From the whole table I use the line (metro line) number, names and coordinates fields and also ignore the cells that do not have coordinates.

	Line	English Transcription	Russian Cyrillic	Coordinates
0	1	Bulvar Rokossovskogo	Бульвар Рокоссовского	55.8148,37.7342
1	1	Cherkizovskaya	Черкизовская	55.8038,37.7448
2	1	Preobrazhenskaya Ploshchad	Преображенская площадь	55.7963,37.7151
3	1	Sokolniki	Сокольники	55.7888,37.6802
4	1	Krasnoselskaya	Красносельская	55.7801,37.6673
5	1	Komsomolskaya	Комсомольская	55.7753,37.6562
6	1	Krasnye Vorota	Красные ворота	55.7690,37.6487
7	1	Chistyye Prudy	Чистые пруды	55.7657,37.6388
8	1	Lubyanka	Лубянка	55.7597,37.6272
9	1	Okhotny Ryad	Охотный ряд	55.7577,37.6166

Figure 1. List of data generated from a website

Next step is to gather data using Foursquare API. Adding my secret parameters and using the normal (common) code give by IBM in Coursera videos, I have obtained 10 main categories - venues - located nearby Moscow metro stations.

```
Arts & Entertainment (4d4b7104d754a06370d81259)
College & University (4d4b7105d754a06372d81259)
Event (4d4b7105d754a06373d81259)
Food (4d4b7105d754a06374d81259)
Nightlife Spot (4d4b7105d754a06376d81259)
Outdoors & Recreation (4d4b7105d754a06377d81259)
Professional & Other Places (4d4b7105d754a06375d81259)
Residence (4e67e38e036454776db1fb3a)
Shop & Service (4d4b7105d754a06378d81259)
Travel & Transport (4d4b7105d754a06379d81259)
```

Figure 2. List of categories located nearby Moscow metro stations

Then, using my secret Foursquare API login and password I could calculate the total number of venues located nearby each of Moscow’s metro stations.

It made possible to create a new dataframe, that is called “stations_venues_df” and store the values in this particular table. In future steps these table will be used for further analysis, but all the data will be normalised and some columns related to specific venues such as Arts & Entertainment, College & University, Residence, Travel & Transport and Event will be dropped aiming the goal.

	Line	English Transcription	Russian Cyrillic	Coordinates	Arts & Entertainment	College & University	Event	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Residence	Shop & Service	Travel & Transport
0	1	Bulvar Rokossovskogo	Бульвар Рокоссовского	55.8148,37.7342	5	6	0	12	8	8	49	4	51	8
1	1	Cherkizovskaya	Черкизовская	55.8038,37.7448	5	18	0	14	4	19	44	5	37	10
2	1	Preobrazhenskaya Ploshchad	Преображенская площадь	55.7963,37.7151	14	22	0	34	6	30	80	14	68	15
3	1	Sokolniki	Сокольники	55.7888,37.6802	16	17	0	53	15	43	60	25	68	19
4	1	Krasnoselskaya	Красносельская	55.7801,37.6673	31	13	0	95	31	26	70	26	74	73
5	1	Komsomolskaya	Комсомольская	55.7753,37.6562	33	25	1	90	19	34	85	27	59	86
6	1	Krasnye Vorota	Красные ворота	55.7690,37.6487	47	44	1	103	38	52	103	28	68	77
7	1	Chistyye Prudy	Чистые пруды	55.7657,37.6388	68	43	0	198	118	55	102	33	79	70
8	1	Lubyanka	Лубянка	55.7597,37.6272	107	55	5	233	156	88	100	20	105	84
9	1	Okhotny Ryad	Охотный ряд	55.7577,37.6166	113	80	3	125	81	93	117	19	107	71

Figure 3. “Stations_venues_df” dataframe that stores the number of venues relating each metro station

Exploratory analysis and data cleaning & normalisation

As we can see, the most popular and frequent venues are Food, Shop & Service and Professional & Other Places. In fact, Food may contain the categories that not just sell already cooked dishes (restaurants), but also supermarkets (sometimes they sell ready-to-eat meals).

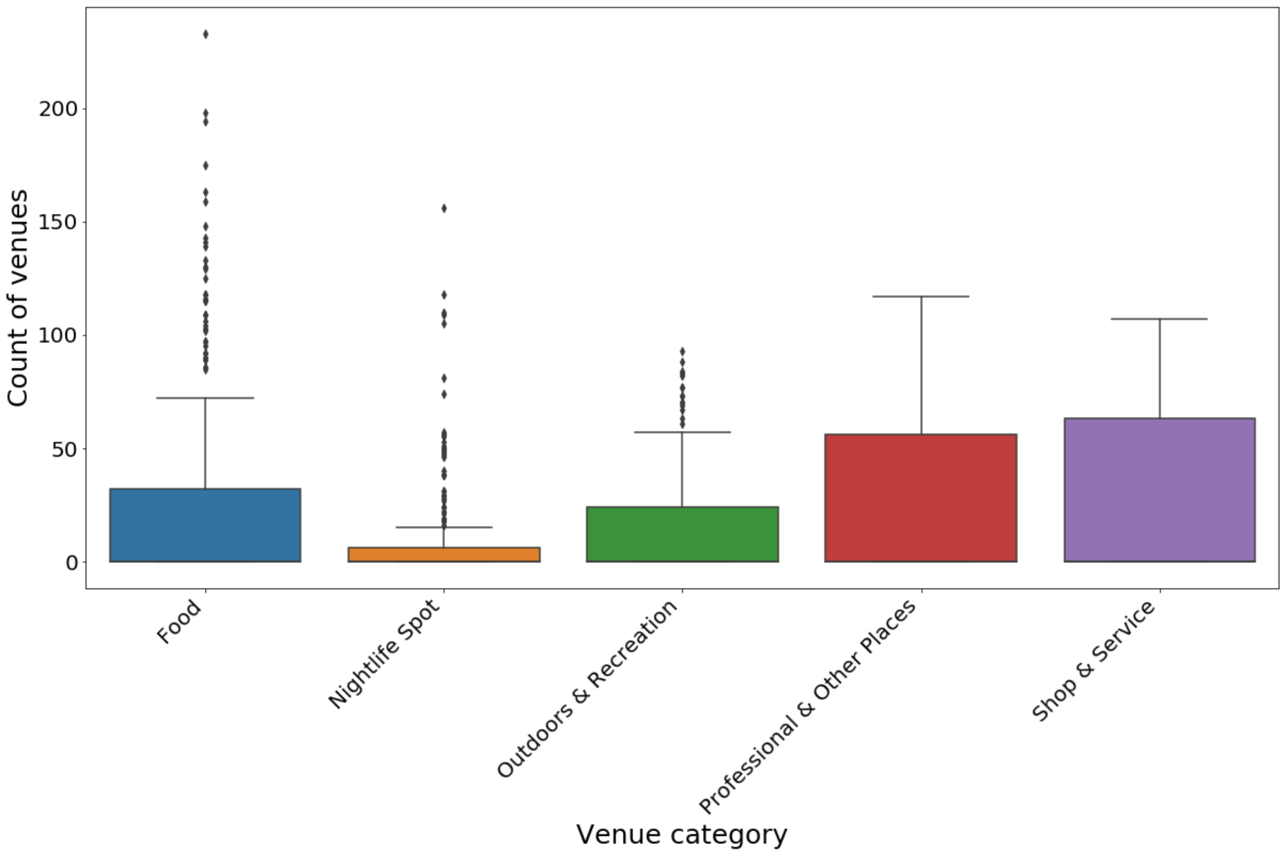

Figure 4. Count of venues in different venue categories

It has been already stated that specific venues such as Arts & Entertainment, College & University, Residence, Travel & Transport and Event were be dropped aiming the goal. As a result a new data frame called cluster_df was created, and the data visualized. Also the values were normalised using MinMaxScaler method.

	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Shop & Service
0	0.051502	0.051282	0.086022	0.418803	0.476636
1	0.060086	0.025641	0.204301	0.376068	0.345794
2	0.145923	0.038462	0.322581	0.683761	0.635514
3	0.227468	0.096154	0.462366	0.512821	0.635514
4	0.407725	0.198718	0.279570	0.598291	0.691589
5	0.386266	0.121795	0.365591	0.726496	0.551402
6	0.442060	0.243590	0.559140	0.880342	0.635514
7	0.849785	0.756410	0.591398	0.871795	0.738318
8	1.000000	1.000000	0.946237	0.854701	0.981308
9	0.536481	0.519231	1.000000	1.000000	1.000000
10	0.557940	0.314103	0.741935	0.752137	0.719626

Figure 5. “Cluster_df” dataframe that stores needed types and number of venues

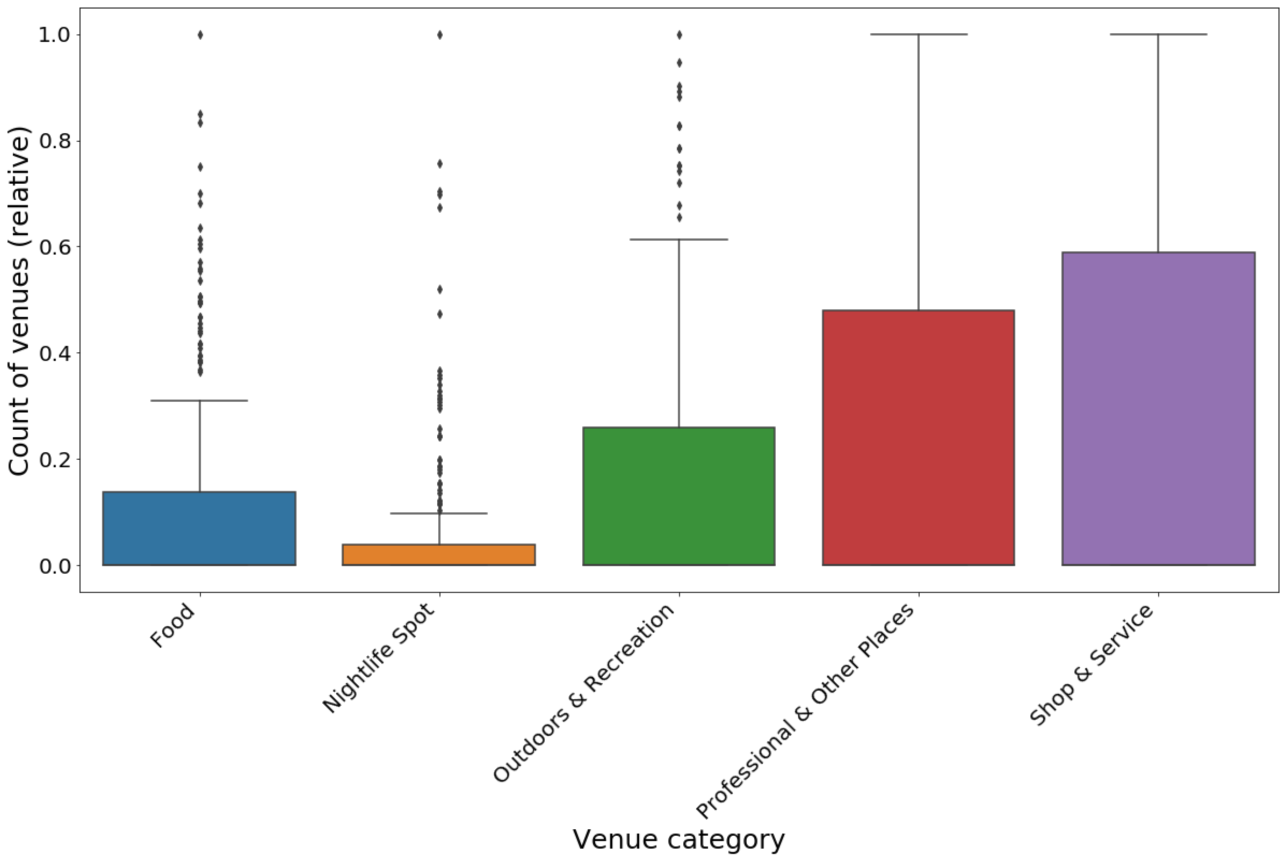

Figure 6. Count of venues in different categories (normalized using MinMaxScaler method)

Then I initiated Clustering procedure with 4 clusters and plotted the results on a map.

Results

We can see different clusters shown on a map. Blue points represent the situation that different (almost all) types of venues are around nearby by an underground station - here is is not the best decision to set up business due to fierce competition. Professional and other services are the most frequent venue here. Red points shows almost the same results. Both these points are located in the city-centre, so the results are quite obvious. Green points show lower marks but reflect good scores in Shop and Service. Yellow marks represents mostly by Food category.

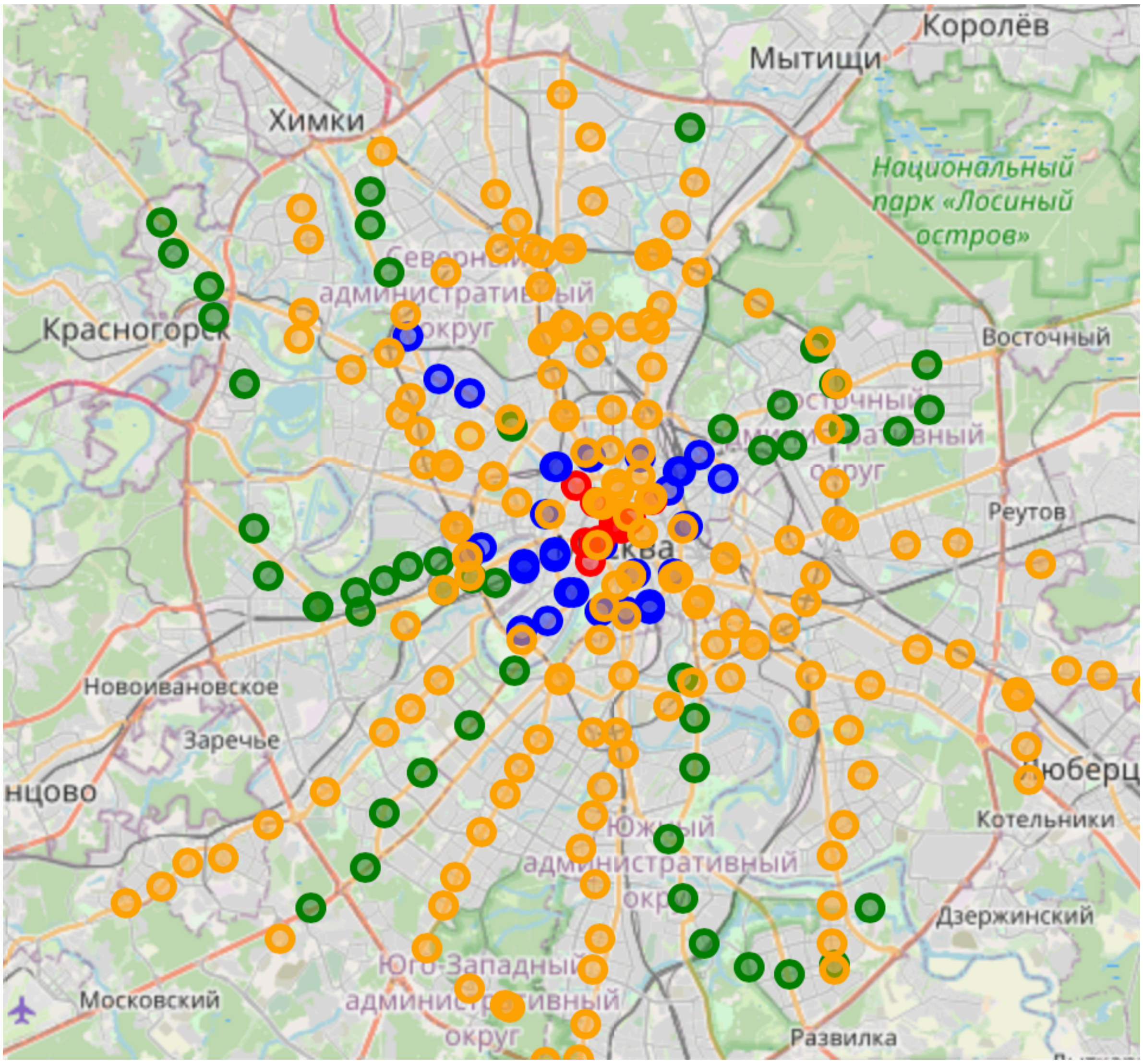

Figure 7. Map with the results obtained

Discussion

The estimation is limited. The next step of the analysis could be an analysis of sub-categories downloaded via Foursquare to estimate Moscow districts with specific type of venues to be opened there.

Conclusion

Foursquare data - is a good analytical instrument that can be used in different marketing strategies and by lots of businessmen. Understanding of a code can simplify and cut down costs without a need to pay consulting companies for any type of specific research.