Analysis of Moscow southern part metro stations

Evgeny Lavriv

March 02, 2020

**Introduction**

**Background**

I live in Moscow and metro system is very vast and developed here. Lets have a look at the neighborhoods surrounding metro stations and classify them. They can be divided in mostly residential, business or commercial areas, surrounding them. Venues positioning arount station determine why and how people use it. If professional places zer not in a neighborhood its residents are likely to travel to other areas for work. This creates daily migrations of people. WIth analysis of this data it is possible to classify stations by their primary usage. It could be useful for city planners to determine where from and where to people are most likely to travel for work and leisure, plan further extension of the network and find places for new development. I would like to perform such analysis for Southern Moscow, where I live.

**Problem**

It could be very useful for people who are planning to rent an appartment in this area, to choose the metro station. It could also be useful for companies, which choose a place for their office.

**Interest**

Moscow now inhabits more than 12 million people and such analysis could have been useful for anybody from this number.

**Data acquisition and cleaning**

**Data sources**

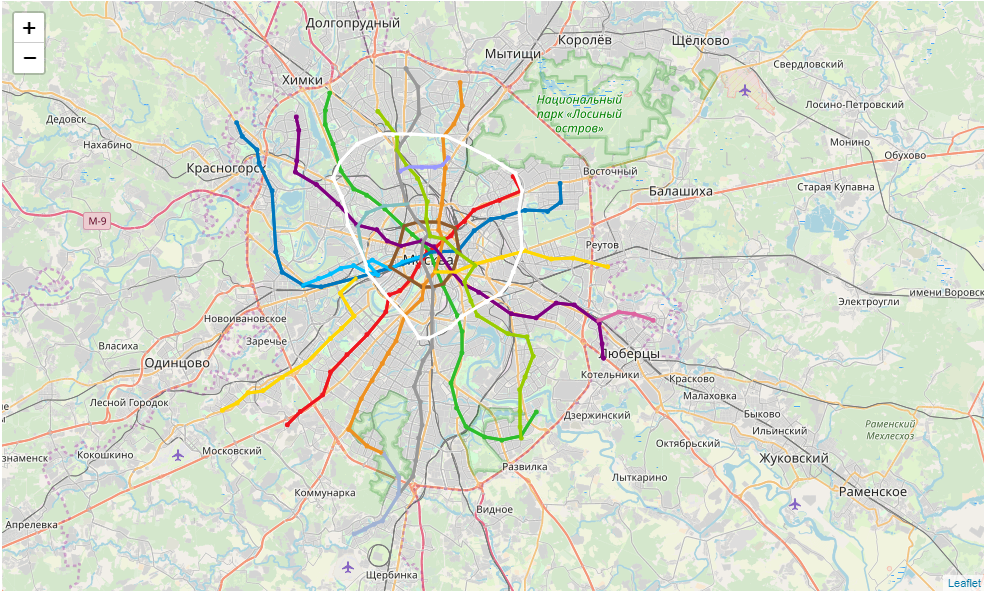
Data would be taken from Wikipedia and Foursquare API on the basis of corrdinates.

**Data cleaning**

Data is not needed to be cleaned as Wikipedia and Foursquare are quite reliable sources. Only some formatting issues wil have to be solved. We will be basing on stations coordinates and venue categories from Foursquare.

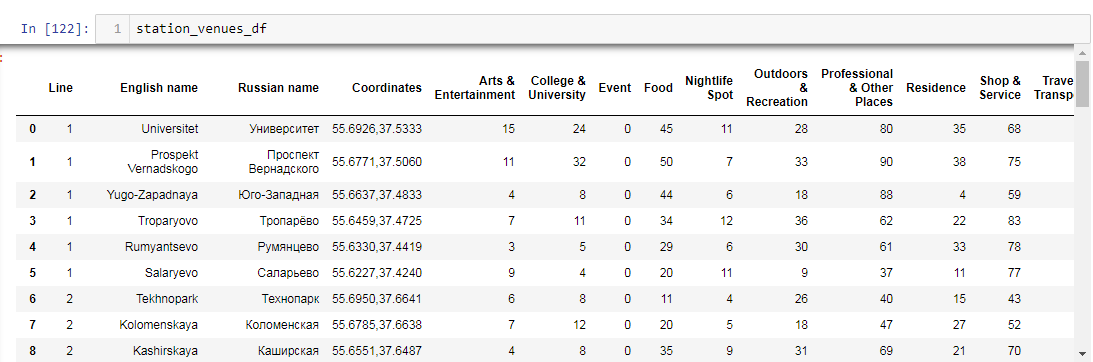
**Methodology**

First, data is visualized for stations and lines on the Map using Folium.

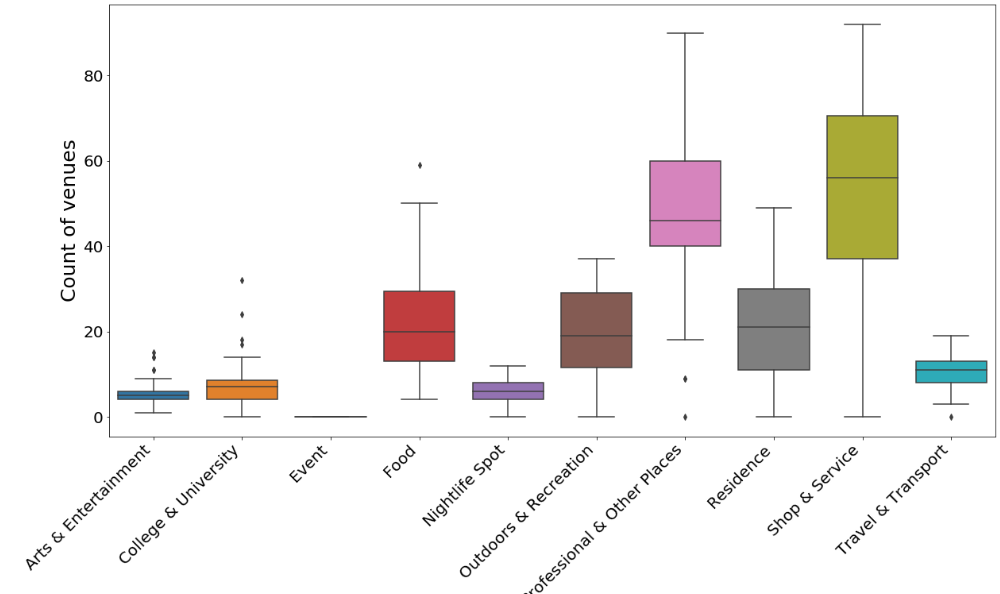


Next, we exclude all the stations that are not in southern part of the city.

After data is being taken from Foursquare about the number of venue around each station.

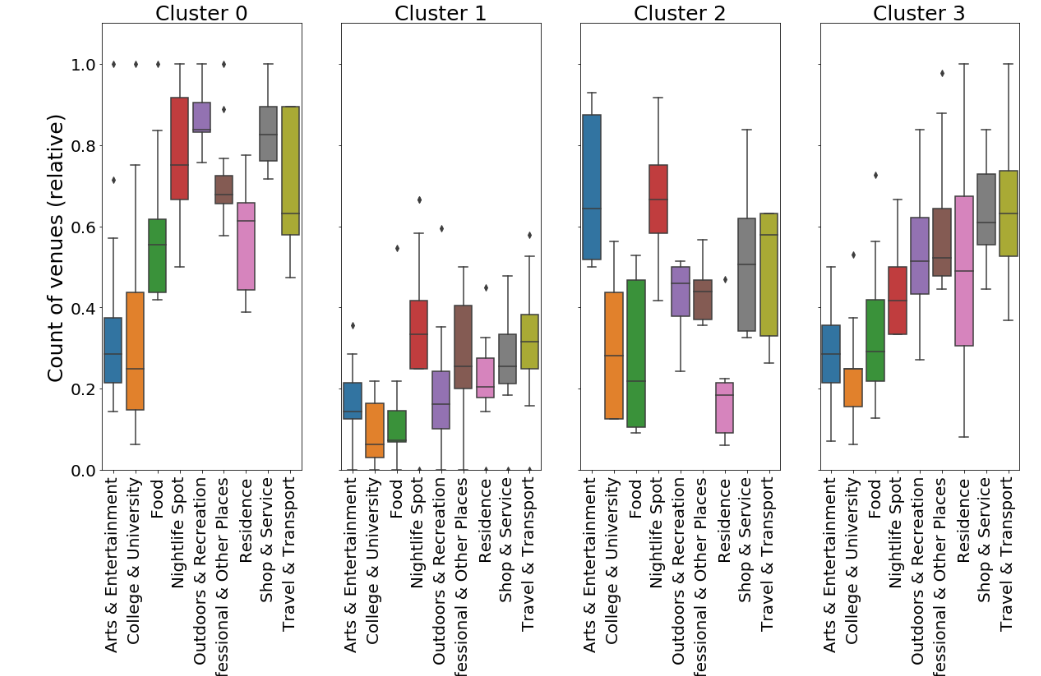


We analyze if any categories are empty using visualization.

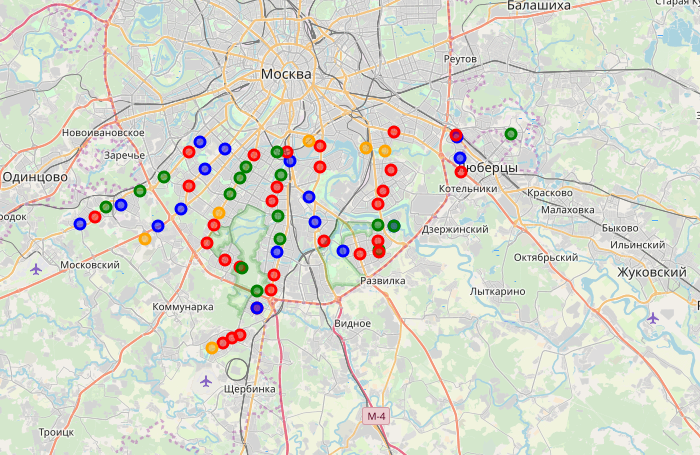


It is visible that Events category is empty and we can drop it easily.

After exclusion of Events we can categorize data on venues and stations and visualize 4 categories.



Also, we can plot them on the map.



**Results**

Here is how we can characterize the clusters by looking at venue scores

Cluster 0 (Blue) is leading in nightlife, shops and transport. This is the most diversely developed part of the city.

Cluster 1 (Green) is low in all categories. Those areas seem to be in industrial area.

Cluster 2 (Orange) is high in Arts, Entertainment and Food. It seems that this is a part of a town where people come to have fun.

Cluster 3 (Red) is leading in residences, shops and transport. Those are the places where people live.

Plotting the clusters on a map shows us that clusters are randomly situated around the town and there are no certain rules.

**Discussion**

To be fair, Foursquare data isn’t all-encompassing. The highest number of venues are in the Food and Shop & Service categories. Data doesn’t take into account a venue’s size (e.g. a university building attracts a lot more people that a hot dog stand – each of them is still one Foursquare “venue”).

**Conclusion**

Foursquare data is limited but can provide insights into a city’s development. This data could be combined with other sources (e.g. city data on number of residents) to provide more accurate results.