Moscow Metro Stations Analysis

Introduction

- Moscow has one of the largest metro in the world.
- Approximately 5 million people use the Moscow metro per day.
- This project aims to select metro stations in the vicinity of which it is better to open a new business.
- This analysis will be useful for businessmen or managers of large businesses for a more informed decision-making on location to open a new sales point.

Data acquisition and cleaning

- A list of Moscow Metro stations with coordinates can be found on Wikipedia.
- Passenger traffic data scraped from the site of the company “MosOblReklama”.
- I also used the service Foursquare to list venues near each metro station.
- Stations, located closer than 250 meters to each other are combined into one station.
- In total, 178 entries of metro stations and 119 of them contain passenger traffic data.

Center vs. Rest

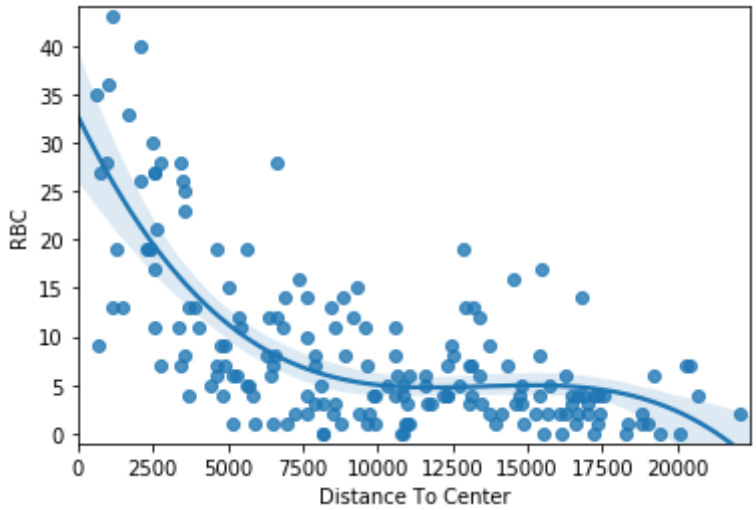
- Historically, Moscow has a circular structure with the Kremlin in the center.
- Therefore, I decided to add parameter “Distance To Center”.
- Then I split the metro stations into two groups (those that are within a radius of 3 km to the center and the rest)

Let's compare the average values of two groups.

Category	In The Center	Out Of Center	Difference
Hotel	1.173913	0.200000	0.973913
Art Gallery	1.217391	0.077419	1.139972
Theater	1.565217	0.174194	1.391024
Bakery	1.913043	0.509677	1.403366
Plaza	1.782609	0.225806	1.556802
Coffee Shop	5.173913	1.587097	3.586816
Bar / Pub	6.521739	1.109677	5.412062
Restaurant	12.086957	3.735484	8.351473
Passengers Per Day	84.996143	67.610735	17.385408
Venue Count	65.652174	30.012903	35.639271

Of the categories of venues, there are 3 that are very dominant in the center: **Restaurant**, **Bar / Pub**, **Coffee Shop**. These three categories are similar in meaning and I decided to combine these three categories into one and name it **RBC** by the first letters of each category.

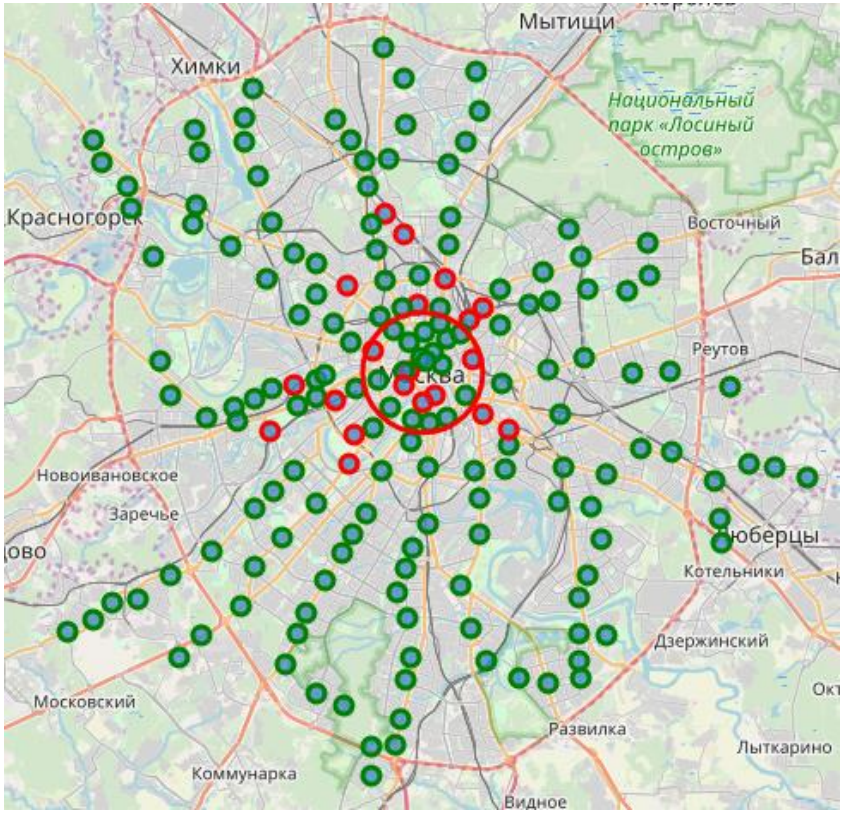
RBC & Distance To Center



Cubic polynomial regression is best suited.

R^2 = **0.54**

Map



Stations marked in red are more preferable for opening a business category RBC.

Conclusion

- Stakeholders should use the list of metro stations based on the cubic regression between the number of venues of category RBC (Restaurant, Bar / Pub, Coffee Shop) and the distance to the center.
- Final decision on optimal location will be made by stakeholders based on specific characteristics of metro station.
- It also seems to me to be useful to use the ratio between "Venue Count" and "Passengers Per Day".