

IBM Data Science Professional Certificate Specialization.

Course 9 of 9:
Applied Data Science Capstone. Final Project.

Introduction

Moscow is the capital of the Russian Federation, a country of 144.5 million people.

Moscow is not only the largest city in Russia (its total area is 2,511km²), it is also the most populous city with 13.2 million residents within the city limits, 17 million within the urban area and 20 million within the metropolitan area, according to official data.





Unique Features of Moscow

1. Life in Moscow is focused in the center of the city, where almost all the businesses are blooming. The outskirts are mostly dormitory areas or manufacture zones;
2. Moscow population is extremely uneven: after new territory was joined in 2012, there are regions of Moscow with less than 1% of population on almost 40% of area;
3. Being situated between Europe and Asia, Russia has always been a multinational and multicultural country. So, it's no wonder, that Moscow attracts people from both worlds, and absorbs a lot of their views, traditions and customs.



Goal of the Project

In this project I am going to study Moscow venues from the perspective of food. I am going to inspect three major categories: bars, restaurants and fastfood venues.

I want to see if Russian people still prefer fancy restaurants, as they used to in the 90's, when so-called 'nouveaux riches' were the target auditory for newly opening places, or if they drink as much as people say, and the most popular category is any kind of bar, pub or any other drinking facility.

Or, maybe Moscow is full of fast food and street food places, just like any other megapolis, because people prefer grabbing a cup of coffee or a slice of pizza and run their errands as fast as possible.



Project objectives

- Correlation between population and number of venues
- Correlation between distance from the center and number of venues
- Most popular venues in Moscow by districts
- Clustering Moscow districts and studying its differences and similarities



Data Sources: Moscow borders and districts shapes

Moscow borders and districts shapes come from the GIS-Lab (<http://gis-lab.info>) — informal community of russian-speaking GIS/RS specialists.

The data comes in form of GeoJSON file, which is a Feature Collection, describing Moscow districts in the form of sets of coordinates, which can be used to shape polygons and multipolygons on Folium Maps, including the Choropleth Map of Population.

The second source of geospatial data is OpenStreetMap's search engine Nominatim (<https://nominatim.openstreetmap.org/>), which is a tool for searching OSM data by name and address (geocoding) and to generate synthetic addresses of OSM points.



Data Sources: Moscow population

Moscow Population data was manually collected from Wikipedia (<http://wikipedia.org>). It was put together in a csv file, containing the District Name, Borough Name and Number of People, living in each area.

This data is used to determine density and dependency of number of venues to said density.



Moscow Regions and Population

For every district and *rayon* (borough) inside this district were defined their coordinates and population. This data was collected into a dataframe.

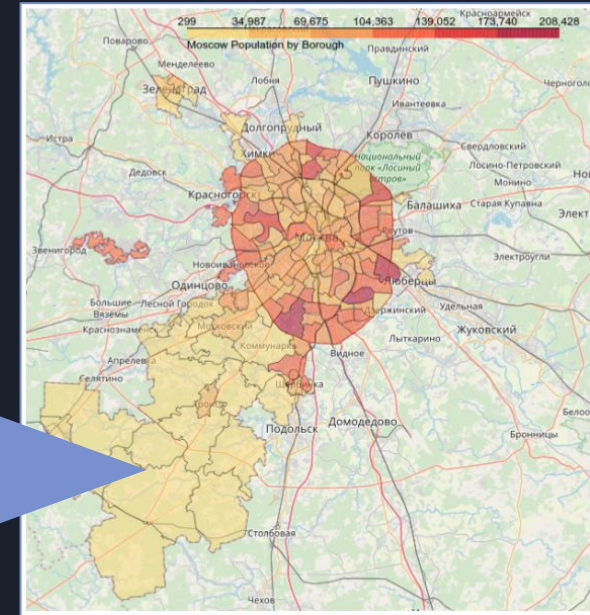
Below is the snippet from the dataframe:

	District	Borough	FullPlaceName	Latitude	Longitude	Population
0	Central Administrative Okrug	Arbat	Central Administrative Okrug, Arbat	55.751199	37.589872	25699
1	Central Administrative Okrug	Basmanny	Central Administrative Okrug, Basmanny	55.767281	37.669773	100899
2	Central Administrative Okrug	Khamovniki	Central Administrative Okrug, Khamovniki	55.740047	37.573958	97110
3	Central Administrative Okrug	Krasnoselsky	Central Administrative Okrug, Krasnoselsky	55.777447	37.654160	45229
4	Central Administrative Okrug	Meshchansky	Central Administrative Okrug, Meshchansky	55.779169	37.627755	56077

Moscow Population Choropleth Map

The color goes from light yellow for the least populated areas to dark red for the most populated.

As you can see, population in Moscow is extremely uneven, especially in the newly joined regions, which used to be Moscow suburbs until 2012





Moscow Venues from Foursquare

Data, collected from the Foursquare, included the name of the venue, its category and coordinates. A dataframe called [moscow_venues] was put together to keep the data together:

	FullPlaceName	Latitude	Longitude	Venue	VenueLatitude	VenueLongitude	Category
0	Central Administrative Okrug, Arbat	55.751199	37.589872	Corner Café & Kitchen	55.751496	37.586757	Japanese Restaurant
1	Central Administrative Okrug, Arbat	55.751199	37.589872	Театр им. Вахтангова	55.749569	37.591638	Theater
2	Central Administrative Okrug, Arbat	55.751199	37.589872	Obedbufet (Обедбуфет)	55.752268	37.592275	Buffet
3	Central Administrative Okrug, Arbat	55.751199	37.589872	Buffalo's	55.751840	37.587376	Wings Joint
4	Central Administrative Okrug, Arbat	55.751199	37.589872	Кофемания	55.752094	37.588102	Coffee Shop

There were more than 2580 venues total. As the goal of the study was to learn about food venues, the list of venues was restricted to these categories.



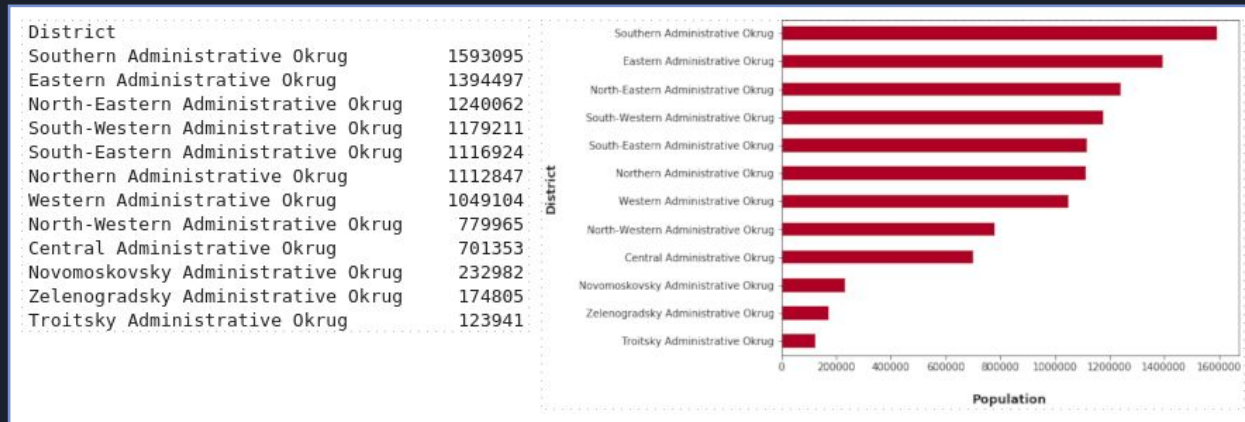
Venues Categories

All the venues were divided by three general categories: bar, restaurant and fast food.

There were total of 322 restaurants, 431 fast food venues and 98 bars found in Moscow.

Moscow Population by Districts

As you can see, some districts are way more populated than others:





Moscow Population: Unevenness

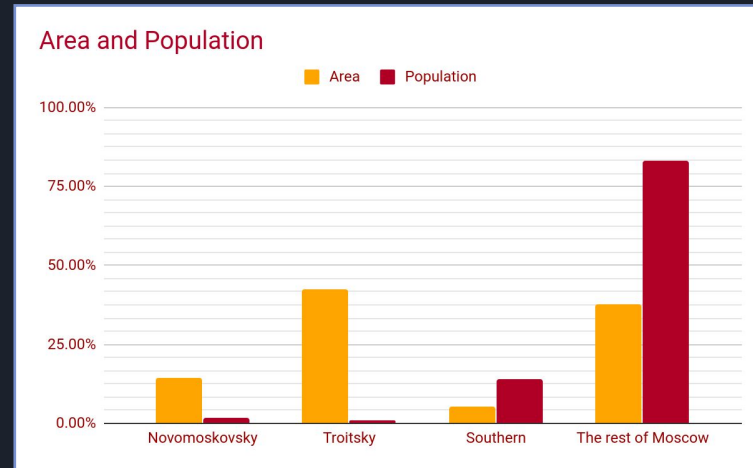
Moscow population is extremely uneven. The whole part of Moscow, which was joined in 2012 (bright-yellow color on the map) is way less populated than the old part of Moscow.

In Novomoskovsky Administrative Okrug (361.4 m², 14.4% of Moscow area) live only 232,982 people (1.85% of Moscow population). In Troitsky Administrative Okrug (1,084.3 m², 42.5% of Moscow area) live only 123,941 people (0.99% of Moscow population).

For comparison: in Southern Administrative Okrug (131,773 m², 5.3% of Moscow area) live 1,593,095 people (14.14% of Moscow population)

Moscow Population: Comparison of several districts

Comparison between Novomoskovsky, Troitsky, Southern districts and the rest of Moscow:





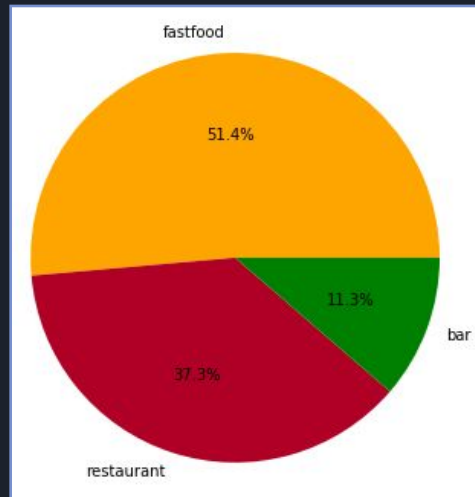
Venue Categories

Top 5 most popular venue categories:

Bar		Fast food		Restaurant	
Wine Shop	21	Coffee Shop	99	Restaurant	44
Bar	19	Cafe	83	Sushi Restaurant	43
Beer Bar	11	Bakery	45	Italian Restaurant	33
Wine Bar	9	Fast Food Restaurant	43	Caucasian Restaurant	31
Cocktail Bar	7	Pizza Place	41	Middle East Restaurant	31

Venue Categories: Ratio

Ratio between fastfood, bar, restaurants:





Venue Categories:

Common categories among districts

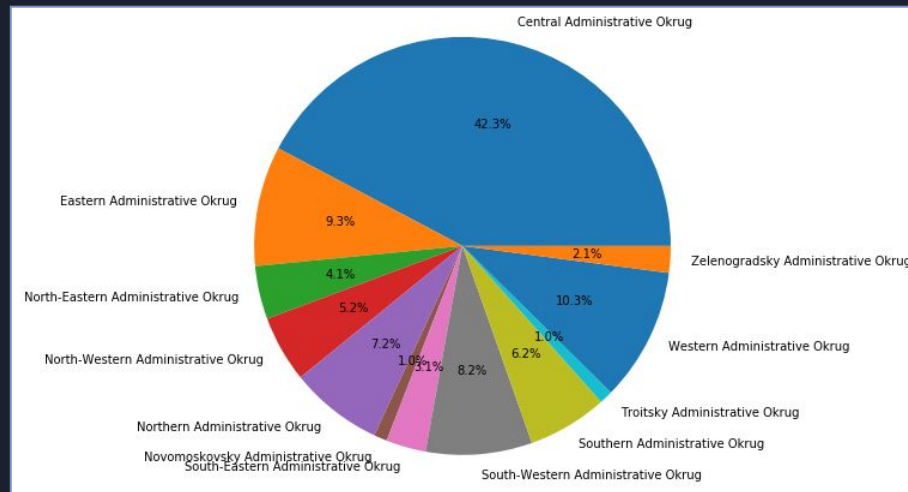
Number of fast food venues, restaurants and bars in each *okrug* (district):

District	Fastfood	Restaurant	Bar
Central Administrative Okrug	114	109	41
Eastern Administrative Okrug	39	32	9
North-Eastern Administrative Okrug	45	32	4
North-Western Administrative	21	14	5
Northern Administrative Okrug	51	34	7
<u>Novomoskovsky</u> Administrative Okrug	4	2	1
South-Eastern Administrative Okrug	33	18	2
South-Western Administrative Okrug	39	22	8
Southern Administrative Okrug	32	27	7
Troitsky Administrative Okrug	7	2	1
Western Administrative Okrug	45	29	10
<u>Zelenogradsky</u> Administrative Okrug	8	3	2

Venue Categories:

Bars

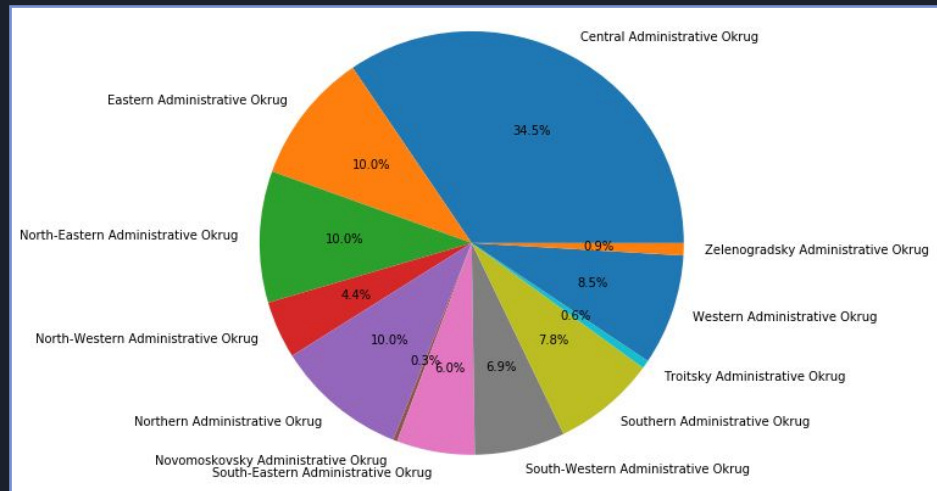
Bars ratio in each district



Venue Categories:

Restaurants

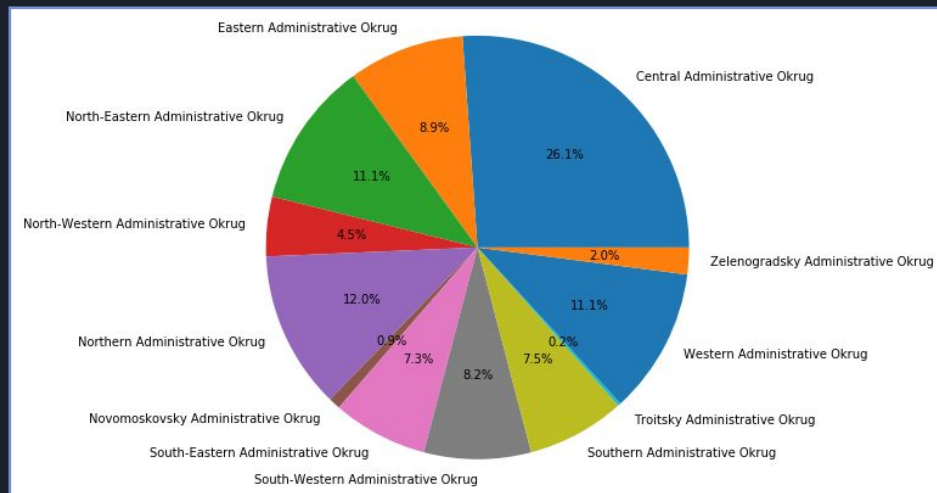
Restaurants ratio in each district



Venue Categories:

Fast food

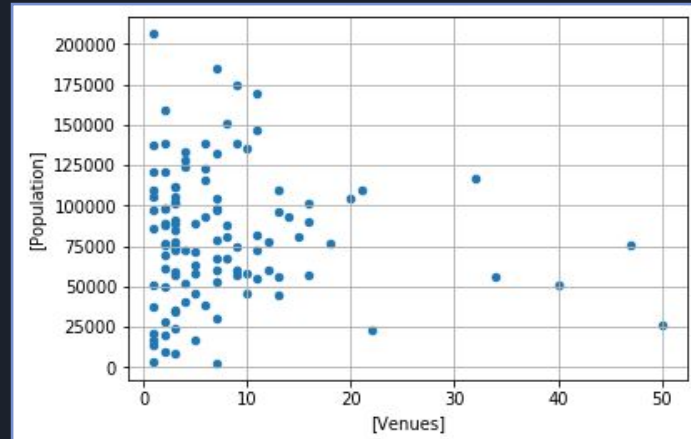
Fast food venues ratio in each district



Number of Venues and Population Correlation

According to the research there is no correlation between population and number of venues:

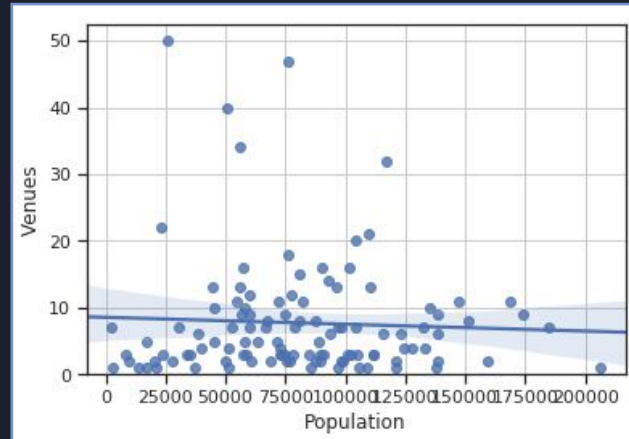
`np.corrcoef` Result: `-0.04201284243983189`



Number of Venues and Population Correlation

According to the research there is no correlation between population and number of venues:

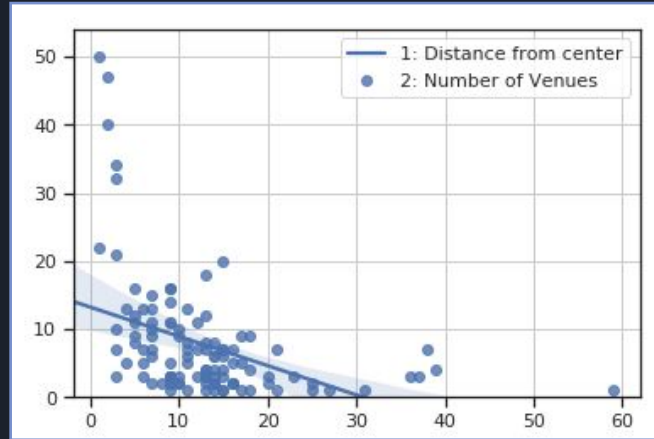
`sns.regplot`



Number of Venues and Distance from Center Correlation

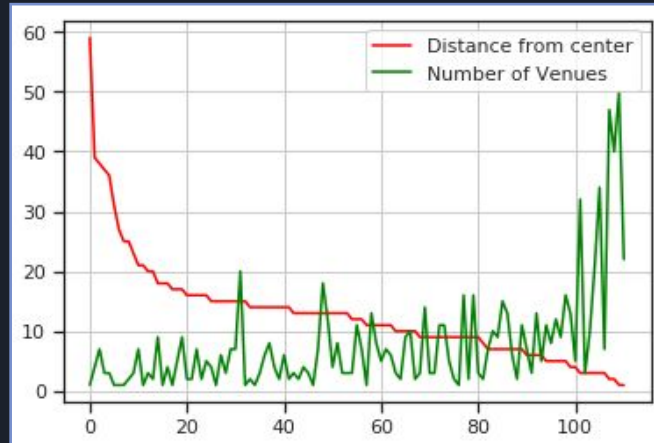
According to the research there is correlation between distance from center and number of venues

```
sns.regplot
```



Number of Venues and Distance from Center Correlation

According to the research there is correlation between distance from center and number of venues





K-Means Clustering

Choice explanation

K-Means is a type of unsupervised learning, which gives pretty accurate results and does not take a lot of time and resources.

It was chosen as a main approach in this study, because it works with unlabeled data (which can be easily obtained with one hot encoding method from the data already present), and suits perfectly the goal to see similarity between different Moscow regions, according to what venue categories are present in each region.



K-Means Clustering

Frequency of occurrence

The mean value was taken to see the frequency of occurrence of each category:

	Borough	American Restaurant	Argentinian Restaurant	Asian Restaurant	BBQ Joint	Bakery	Bar	Beer Bar	Belgian Restaurant	Bistro	...
0	Aeroporto	0.00	0.0	0.00	0.00	0.333333	0.00	0.0	0.0	0.0	...
1	Akademichesky	0.00	0.0	0.00	0.00	0.153846	0.00	0.0	0.0	0.0	...
2	Alexeyevsky	0.00	0.0	0.00	0.00	0.000000	0.00	0.0	0.0	0.0	...
3	Altufyevsky	0.00	0.0	0.00	0.00	0.000000	0.00	0.0	0.0	0.0	...
4	Arbat	0.02	0.0	0.02	0.02	0.100000	0.02	0.0	0.0	0.0	...



K-Means Clustering

Top 10 venues

Then the top 10 venues for each borough were found:

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th
0	Aeroporto	Coffee Shop	Bakery	Wine Shop	Burger Joint	Donut Shop	
1	Akademichesky	Sushi Restaurant	Bakery	Pizza Place	Cocktail Bar	Pub	
2	Alexeyevsky	Pizza Place	Diner	German Restaurant	Cocktail Bar	Coffee Shop	
3	Altufyevsky	Pizza Place	Café	Wings Joint	Coffee Shop	Creperie	
4	Arbat	Coffee Shop	Bakery	Burger Joint	Italian Restaurant	Turkish Restaurant	



K-Means Clustering

Number of clusters

As there are not too many venues found by Foursquare free API, the logical decision was to limit the number of clusters to 5: less number is not informative enough, while larger number of clusters would be too meticulous.

After all clusters were determined, a [m_merged] dataframe looks like this:

	District	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Central Administrative Okrug	Arbat	55.751199	37.589872	3	Coffee Shop	Bakery	Burger Joint	Italian Restaurant	T...
50	Central Administrative Okrug	Basmanny	55.767281	37.669773	3	Coffee Shop	Irish Pub	Mexican Restaurant	Donut Shop	Cr...
53	Central Administrative Okrug	Khamovniki	55.740047	37.573958	3	Pizza Place	Pastry Shop	Caucasian Restaurant	Bakery	M...
60	Central Administrative Okrug	Krasnoselsky	55.777447	37.654160	3	Hotel Bar	Bakery	Coffee Shop	Asian Restaurant	Middle E...

K-Means Clustering

Clusters on the map

Map Color Legend:

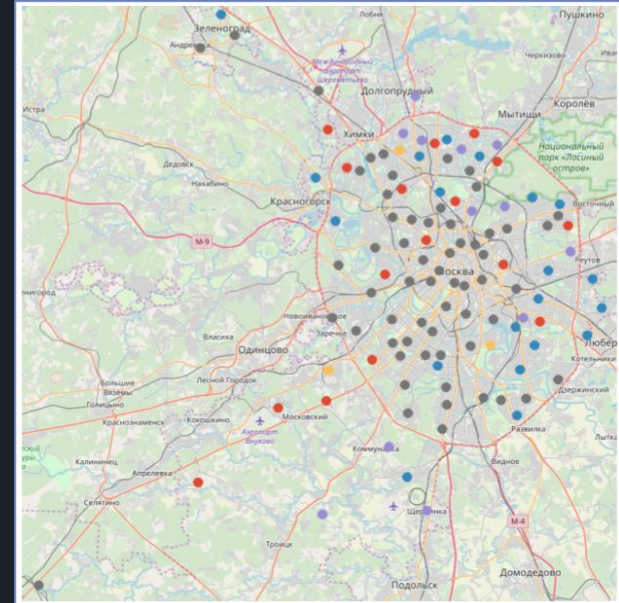
● Cluster 1

● Cluster 2

● Cluster 3

● Cluster 4

● Cluster 5





Clusters Examination

Categories of Venues in Clusters

Total Number of Categories:

Among top-10 most popular venues,
there are 16 categories.

	Category	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Total
0	Café	16.0	NaN	NaN	30.0	NaN	46.0
1	Dessert Shop	13.0	15.0	11.0	34.0	3.0	76.0
2	Creperie	12.0	17.0	12.0	29.0	3.0	73.0
3	Diner	12.0	13.0	11.0	25.0	3.0	64.0
4	Cupcake Shop	11.0	16.0	12.0	32.0	3.0	74.0
5	Czech Restaurant	11.0	14.0	12.0	26.0	3.0	66.0
6	Donut Shop	10.0	11.0	NaN	23.0	3.0	47.0
7	Wings Joint	9.0	12.0	10.0	NaN	3.0	34.0
8	Coffee Shop	6.0	14.0	12.0	44.0	3.0	79.0
9	Restaurant	5.0	NaN	NaN	21.0	NaN	26.0
10	Sushi Restaurant	NaN	16.0	NaN	NaN	NaN	16.0
11	Fast Food Restaurant	NaN	13.0	NaN	19.0	NaN	32.0
12	Pizza Place	NaN	NaN	12.0	NaN	NaN	12.0
13	Cocktail Bar	NaN	NaN	8.0	NaN	NaN	8.0
14	German Restaurant	NaN	NaN	5.0	NaN	NaN	5.0
15	Dumpling Restaurant	NaN	NaN	NaN	NaN	3.0	3.0
16	Eastern European Restaurant	NaN	NaN	NaN	NaN	3.0	3.0

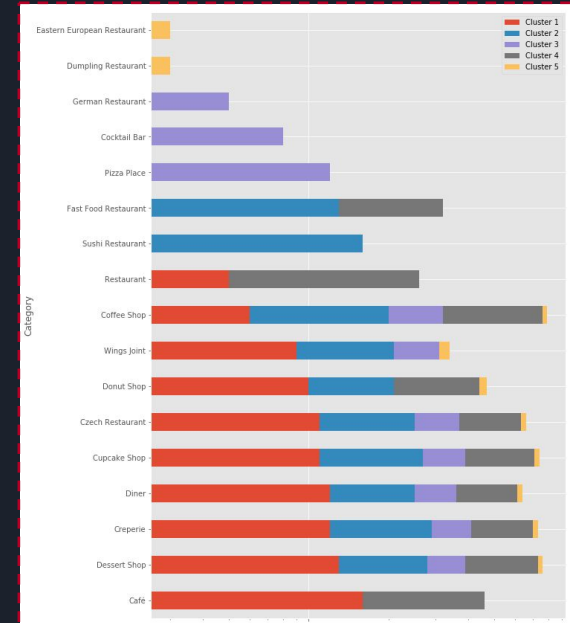
Clusters Examination

Categories of Venues in Clusters

Categories Ratio in Clusters

The following graph allows us to see how different categories of venues are spread among clusters.

It becomes obvious why cluster 5 was highlighted by the algorithm, as it contains two unique categories, though the cluster has only 3 points on the map..



Clusters Examination

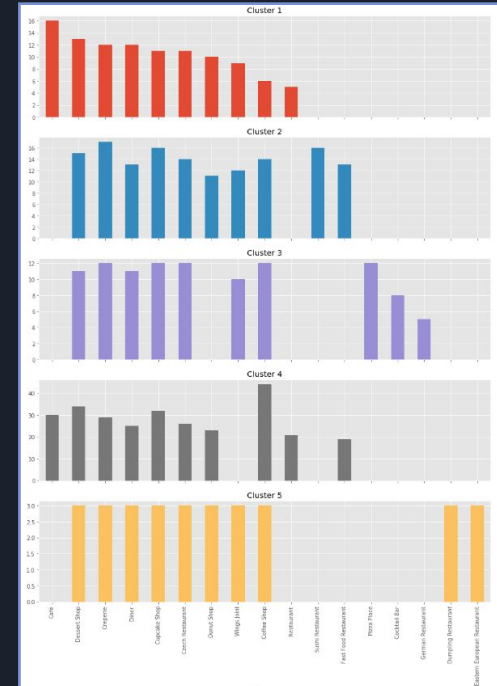
Categories of Venues in Clusters

Number of Venues in Each Cluster

The following illustration shows categories of each cluster in numerical terms.

Each bar is the number of categories in each cluster.

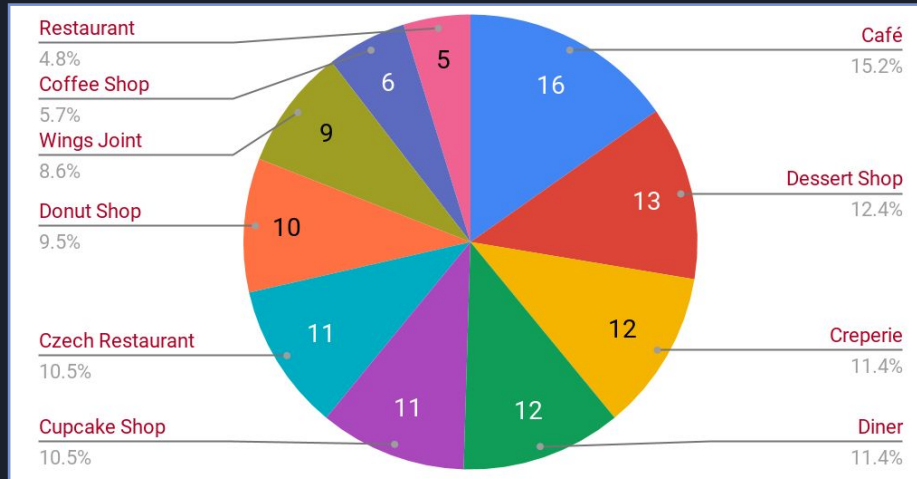
It shows that almost all clusters have unique categories.



Clusters Examination

Individual Clusters Examination

Cluster 1



It seems like people in this cluster are seriously into sweet small bites.

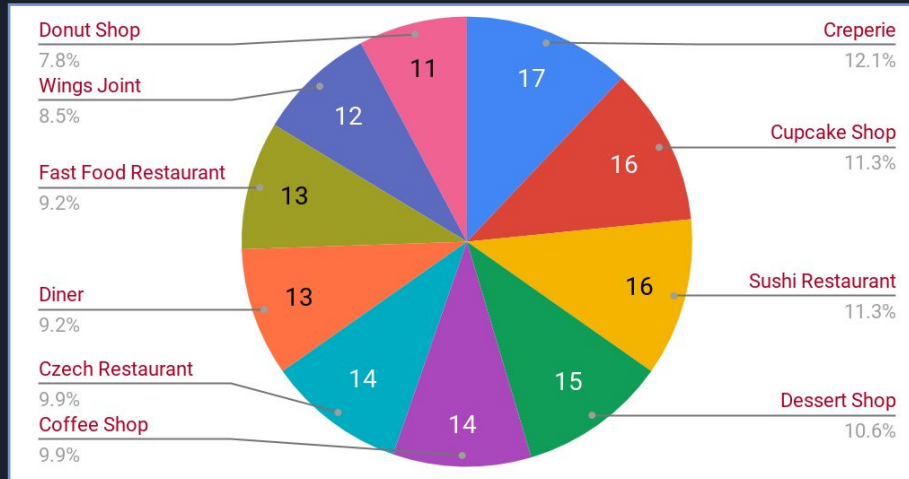
There are some restaurants, like Czech restaurants, and Wings Joints, yet most of these venues are places where you can get something sweet: cupcakes, donuts, crepes.

Further study shows that Moscow is very fond of sweets.

Clusters Examination

Individual Clusters Examination

Cluster 2



Cluster 2 points are mostly situated on the East of Moscow.

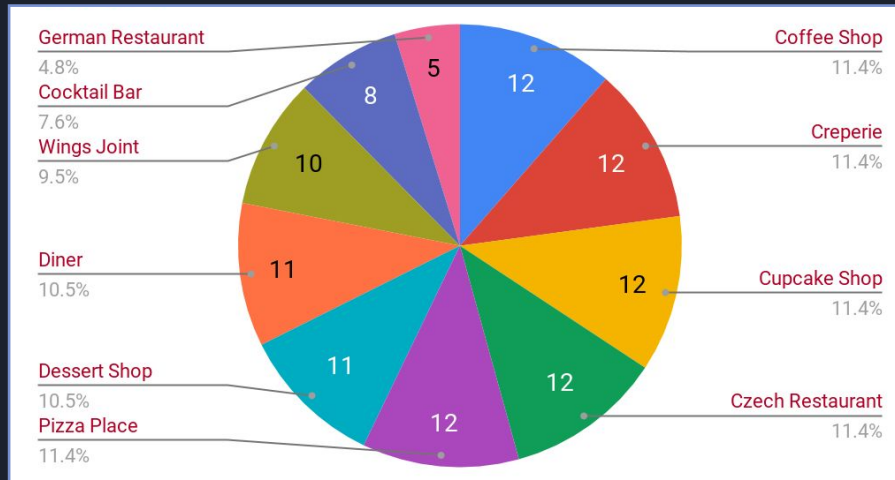
The popularity of Creperies is not surprising as well: crepes, or blinis, are considered Russian national meal, so people do love it here.

Sushi Restaurants are the third popular type of restaurant in Moscow, somehow Russian people fell in love with Japanese cuisine, and are not going to let it go.

Clusters Examination

Individual Clusters Examination

Cluster 3



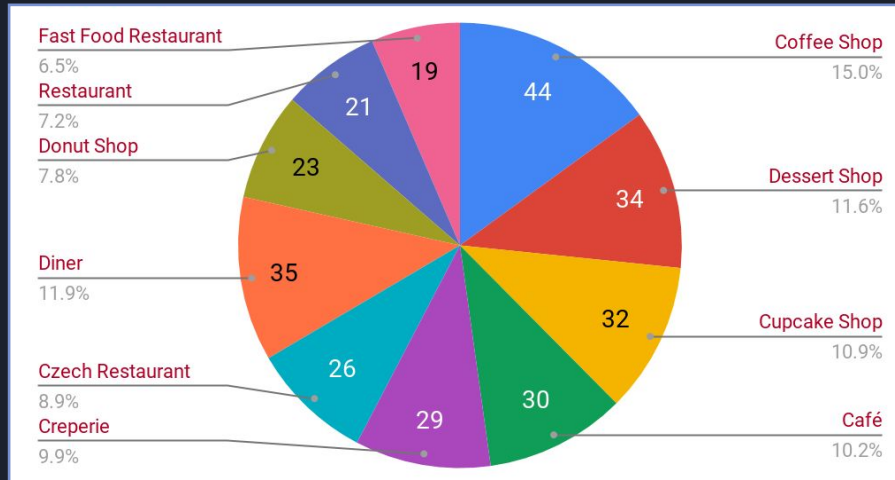
The most of points of Cluster 3 are situated on the North of Moscow. It is a part of the city, mostly built in the 1960 - 1970's, and historically artists and aviators used to live there.

Interestingly, the only 3 points of this cluster on the South of Moscow are nearby the Ostafyevo Airport, which is a class "B" international airport. renovated and opened for civilian flights in 2000 on the grounds of a former military airbase.

Clusters Examination

Individual Clusters Examination

Cluster 4



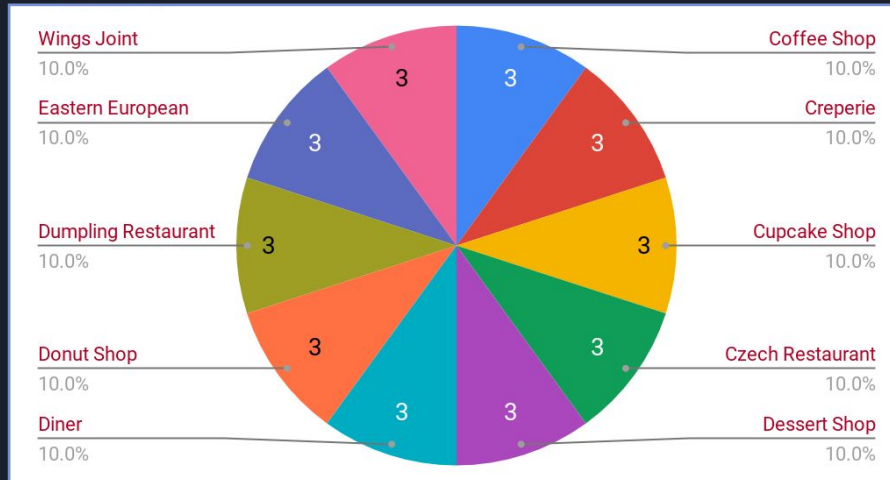
Cluster 4 points are spread all over Moscow, which seem to be quite logical: Moscow loves its coffee. There are coffee shops literally in every block, so it's no wonder it's the most popular category in the largest cluster.

Russia is famous for its cold winters, and even though winter in Moscow is usually not as bad as in the most of the country, still, it takes a lot of coffee to get through many months of gloomy grey mornings.

Clusters Examination

Individual Clusters Examination

Cluster 5



This cluster is very small, it has only three points on the map, so it looks more like a residue.

Interesting that it contains two unique categories for other clusters: Dumpling Restaurant and Eastern European Restaurant.



Results

There are 12 districts (or *okrugs*), 146 boroughs (or *rayons*) and settlements in Moscow.

Each district sometimes seems very different from another, but this study showed how similar these regions might be concerning food.

All Moscow venues, acquired from the Foursquare API, were split into three categories: bars, fast food venues and restaurants.

It appeared that most of venues in Moscow are fast food venues (along with cafes, coffee shops, creperies, etc.) (51.4%), and the least venues are bars (11.3%), and this ratio keeps the same among all the districts.



Results

Moscow population is around 11.92 million people. After joining the large territory in 2012, Moscow are grew almost by half, with no significant change in population, so one of goals of this study was to find out if number of venues in each district depends on the population. Surprisingly enough, there was no correlation between these numbers.

Moscow is very disperse in terms of concentration of all kinds of venues, except for the city center. So another goal was to find out how the distance from the center of Moscow affects the number of venues. Initial assumption was confirmed: venue density significantly decreases with distance from the center.



Results

The main goal of this study was to split Moscow into clusters and compare them.

- There were 5 clusters, chosen by the top-10 most popular venue categories.
- There were categories, common for all clusters, such as coffee shops and creperies, but some categories appeared to be unique for some clusters, like German restaurant and Cocktail bar (cluster 5), or Sushi restaurant (cluster 2).
- The most popular category in Moscow is Coffee Shop (it was the most popular category in 3 clusters of 5).
- There were clusters, situated mostly in working zones or dormitory areas (such as cluster 2), or spread all over the city (cluster 4).



Results

Moscow life is focused in the center, and the outskirts are used as dormitory areas, or manufacturing zones, where people prefer to eat on the go.

Russian people prefer cozy inexpensive places where they can have a coffee and desserts, instead of fancy restaurants. Speaking of restaurants, people in Moscow prefer sushi restaurants and Czech restaurants.

There are a lot of places, where there are too few venues of any kind, and it can be a good business strategy to place a fast food venue, if you want to create a very popular place, or local bar, if you prefer filling the niche.



Conclusion

The goal of this research was to determine what and where does Moscow eat and drink. Is it true, that Russian people stick to their traditions and don't accept new streams? Is it true, that people in Russia drink a lot, and prefer bars to any other kind of venue? Is it true, that outside of the Moscow center there's almost nowhere to grab a bite.

Also, there was a task to split Moscow regions in clusters and see what they have in common, and what makes them different.

Well, all the goals were achieved.



Conclusion

Russian people do stick to their traditions, they still prefer creperies and dumpling shops to burgers or udon shops.

And among all the cuisines of the world they chose Czech and German cuisine, as something, that's closer to their taste and habits.

But they also love trying new things, and if they love something, they love it deeply like coffee shops and sushi restaurants.

And no, Russians don't drink as much as Hollywood movies show: there are very few bars in Moscow comparing to other kinds of venues, and the only category that made it to the top-10 was not vodka bar, it was a cocktail bar.



Conclusion

Yes, in the outskirts of Moscow there might be from very few to no restaurants, but there's always a coffee shop or a cafe where you can grab a bite, so you'll never stay hungry, wherever you are.

So, welcome to Moscow!



Thank you for your attention!