

Presentation

The Battle of Neighbourhoods

Moscow

The Battle of multicultural gastronomic variety

Introduction

Moscow Metro has more than 250 stations and due to constant road problems and massive traffic jams is the best way of transportation so if you decide to go out metro is often the best way to get to a place. But which metro station is the best if you want to have a nice restaurant experience with diverse chose of restaurants with many different national and international cuisines?

The main purpose of this project is to determine which neighbourhoods surrounding which metro stations in Moscow offers the biggest number of restaurants with the largest variety of national and international cuisines. In other words we are looking for a metro station that can offer us the most multicultural gastronomic diversity experience and choice when deciding on a restaurant.

This information can be used by both tourists and locals to chose a metro station from where they will have the biggest number of choices for a restaurant and can assist business people in opening a new restaurant and deciding which national or international cuisine to go for and where to open it.

Data

We will use Foursquare location data with basic descriptive statistics as well as some unsupervised machine learning with cluster analysis to classify stations and derive the information and conclusions we need.

The base list of metro stations and their geographical coordinates are scraped from this Wikipedia page: https://en.wikipedia.org/wiki/List_of_Moscow_Metro_stations

The goal is then to merge the two data sources and to get them into a pandas DataFrame which would show all metro stations with their latitude and longitude from the Wikipedia page and for each metro station show all associated food venues like that:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bulvar Rokossovskogo	55.8148	37.7342	Burger King	55.814026	37.733659	Fast Food Restaurant
1	Bulvar Rokossovskogo	55.8148	37.7342	Broker Coffee	55.813895	37.733526	Coffee Shop
2	Bulvar Rokossovskogo	55.8148	37.7342	шаурма	55.816166	37.730582	Shawarma Place
3	Bulvar Rokossovskogo	55.8148	37.7342	Китайская Кухня "Лотос"	55.813580	37.732730	Chinese Restaurant
4	Bulvar Rokossovskogo	55.8148	37.7342	Подсолнухи Art&Food	55.816065	37.736457	Food Court

Methodology

Initially we will use just basic statistical summary analysis with max , min counts of venues per each station neighbourhood. We will display this resulting analysis apart from simple summary table by overlaying the results on the map using Folium library.

Each circle will be mapped to metro station coordinates and then we will use 'colour' library to create a gradient list of colours from blue to red. These gradient colours will correspond to smaller number of unique venues – more blue to high number of unique venues – more red, with green and orange colours in the middle corresponding to various degree of medium numbers of venues.

We will then use Kmean cluster analysis from 'sklearn.cluster' library with onehot encoding transformation to try to identify if there are any meaningful and explainable clusters among venues in relation to different national/international cuisine mixes.

Each cluster like in the previous min/max analysis will be mapped using Folium maps to metro station coordinates and then we will use another 'colour' library from 'matplotlib.colors' to just create a random unique colour for each cluster so we can clearly see them and visually identify.

Results

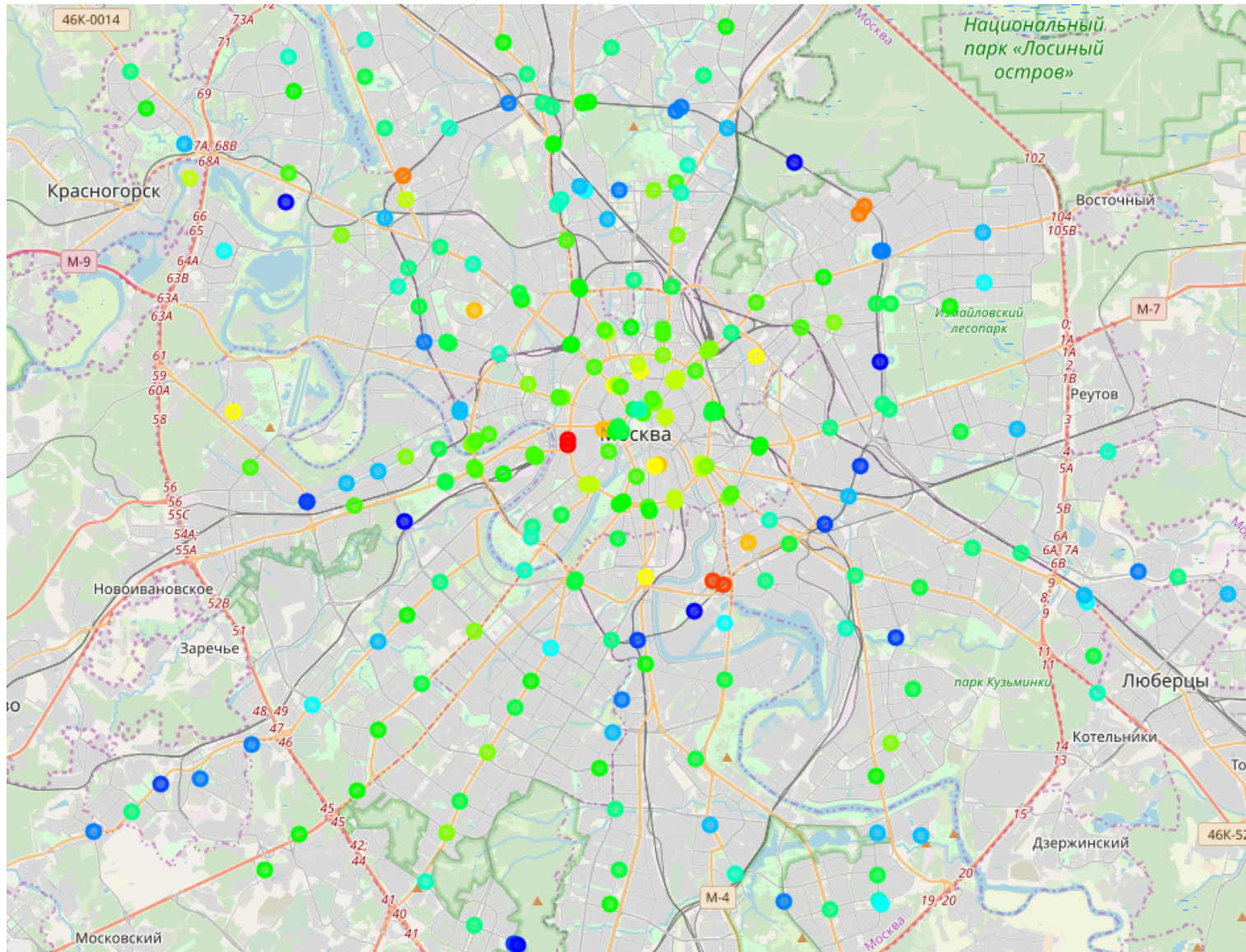


Image 1

Results

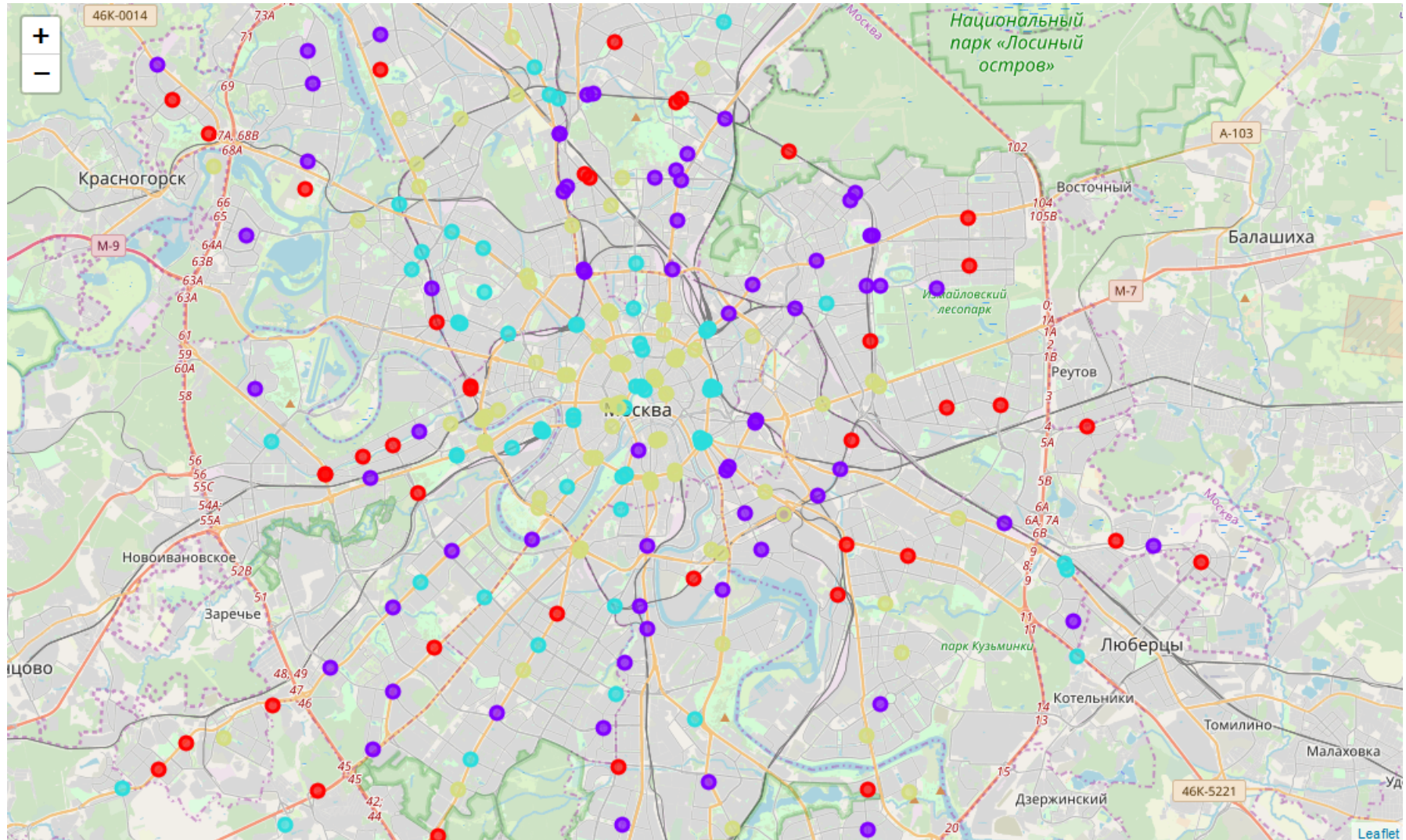


Image 2

Results

The first image (image 1) demonstrates counts of unique national/international offers showing ethnic diversity which was discussed in the introduction section. As discussed in Methodology section the image shows gradient colours from Red = maximum diversity to Blue = minimum diversity, with Green-Yellow colours representing different levels of medium diversity. As we can see the centre of Moscow mostly represented by high to medium diversity offers where as on the peripheries we see more 'Blue' stations indicating low cultural cuisine diversity. This again can be used as a business opportunity for finding a new area for a specific ethnic/national cuisine. For further more in-depth review interactive map can be downloaded here:

https://htmlpreview.github.io/?https://github.com/Serge-ds/Coursera_Capstone/blob/master/GradientMap1.html

Moving on to Kmeans cluster analysis (image 2), as mentioned in the Methodology section, we decided to run it using 'sklearn.cluster' library with onehot encoding, selecting k=4 clusters. Four clusters were selected using common sense, trial and error and empirical knowledge as the most suitable number of clusters that can have meaningful interpretation.

The results are displayed in the image below and we can immediately see that centre of Moscow is dominated by two clusters: Pale Yellow and Light Blue. The peripheries of Moscow are dominated by two other clusters Red and Purple.

It turns out that that Light Blue cluster offers the most diverse combination of restaurants with the biggest choice of all nationality groups; Pale Yellow on the other hand offers a slightly smaller selection still large enough but with a slight focus on Asian foods.

As for the other two clusters which dominate the peripheries of the city one of them – Red – offers small selection of venues mostly including Sushi restaurants and Japanese restaurants where as Purple cluster offers slightly larger selection than the Red but focus is more on Middle Eastern food.

For further more in-depth review interactive map can be downloaded here:

https://htmlpreview.github.io/?https://github.com/Serge-ds/Coursera_Capstone/blob/master/ClusterMap1.html

Discussion

Both maps and the corresponding analysis clearly demonstrate that the centre of Moscow is well serviced with a great variety of international food available at a walking distance from most central metro stations. However, the peripheries are serviced not as well as the centre. This might be expected of course but it gives us an opportunity to see where to go if we want a great food diversity when we go out and where to look for business opportunities when looking to open a new restaurant venue.

Conclusion

Although Foursquare data is limited it can still provide some insights into food venues location and combining this data with other data sources we were able to conduct meaningful analysis which resulted in more or less clear classification of Moscow neighbourhoods in terms of cultural diversity and business opportunities when looking to open a new restaurant venue.