IBM Data Science

Capstone Project Report

Vitaly Kamentsev

2019

Diversity of Ethnic Cuisines in Europe. Analysis of City Venues

1. Introduction

Diversity of ethnic cuisines in the world is immense. Moreover, it is easy to find many different ethnic restaurants in any big city. This market is constantly growing and gets more and more complex. European ethnic foods was worth around EUR4.12bn in 2016. The market becomes more and more competitive because of world globalization process and increasing mobility of people around the globe.

Looking into Europe as a very diverse region, some ethnic cuisine trends are obvious but some of them are not. For example, you can find Italian restaurants in any city or town of Europe. Japanese food also took significant popularity over last decades. But what about other ethnic cuisines like Turkish, Vietnamese, Hungarian or Russian? How easy to find it in different parts of big Europe?

2. Questions to answer

Aim of this project is to answer these questions:

1. Take some particular cuisine as a parameter and collect data of its presence in main cities of Europe.
2. Detect geographical trends.
3. Compare popularity of ethnic cuisines around Europe.

Target audience:

- Professional researchers
- Tourists and food lovers

3. Data Description

I used following data in my project:

1. List of largest cities in Europe: wiki page

There are list of main cities in Europe:

Largest cities [edit]

Note: The cities are sorted by official population. Cities in **bold** are capital cities of their respective countries. Only cities with more than 1 million people are listed.

	City	Country	Official population	Date	2011 Eurostat population[1]	Image	Location	Ref.
1	Istanbul[a]	Turkey	15,029,231	31 December 2017			41.013611°N 28.955°E	[2]
2	**Moscow**[b]	Russia	13,197,596	1 January 2017			55.75°N 37.616667°E	[3]
3	**London**	United Kingdom	9,126,366	31 December 2018	8,173,941		51.507222°N 0.1275°W	[4]

Two main features I am interested in are City Name and Population. Population is needed to normalize statistics and avoid difference between largest and smallest cities from the list.

2. Google Geocoding web service

This is main source of geographic coordinates. Assume, that it returns coordinates of city center if city name is requested.

3. List of restaurants of specific national cuisine: Foursquare API

I used it to get list of restaurants of specific ethnic cuisines.

4. Methodology

In general, I used following steps to find and transform data for the analysis:

1. Download wiki table of largest cities of Europe, extract it and build dataset of cities and its population.
2. Enrich the dataset with latitude and longitude of each city with help of Google Geocoding API.
3. For each city in the list extract list of venues of specific category (for example, Turkish Restaurants).
4. Merge Cities and Venues datasets, group and clean to get statistics by city.
5. Normalize statistics by city population.
6. Make visualization on Europe map.

Now, step by step with details:

1. Download wiki table and parse it

For html parsing I used BeautifulSoup framework. I extracted table of cities and took two columns of it: City Name and Population.

Here is a result set (top 10 records):

	cityName	population
0	Istanbul	15029231
1	Moscow	13197596
2	London	9126366
3	Saint Petersburg	5381736
4	Berlin	3748148
5	Madrid	3223334
6	Kyiv	2950819
7	Rome	2857321
8	Paris	2140526
9	Bucharest	2106144

2. Enrich the dataset with latitude and longitude of each city

I am going to visualize my results on Europe map, so geo coordinates are required.

Google Geocoding API provides method to obtain geo coordinates of the place by its text name. I called it for each city from the Cities dataset from step 1 and got results as JSON with geometry data. I parsed it and extracted latitude and longitude. These values were added to Cities. Here is a result set (top 10 records):

	cityName	population	latitude	longitude
0	Istanbul	15029231	41.008238	28.978359
1	Moscow	13197596	55.755826	37.617300
2	London	9126366	51.507351	-0.127758
3	Saint Petersburg	5381736	59.934280	30.335099
4	Berlin	3748148	52.520007	13.404954
5	Madrid	3223334	40.416775	-3.703790
6	Kyiv	2950819	50.450100	30.523400
7	Rome	2857321	41.902783	12.496365
8	Paris	2140526	48.856614	2.352222
9	Bucharest	2106144	44.426767	26.102538

3. For each city in the list extract list of venues of specific category

As I am analyzing presence of some particular cuisines in different cities I need to get list of venues for each of the cities from my list. There is Foursquare API for this purpose.

Another requirement I've put is to fetch venues only for specific category. Category means type of cuisine in this case. There is a list of categories provided in Foursquare developer documentation.

Foursquare provides search API method for fetching venues by categories. I parsed the JSON results and extracted venue name, coordinates and category.

Here is a result set for all cities combined (for category: "Dutch Restaurant"):

	cityName	venueName	venueLatitude	venueLongitude	category
0	Istanbul	Prime Fıtnıs	41.008236	28.978360	Dutch Restaurant
1	Istanbul	Sofana Ristorant	41.010390	28.975569	Dutch Restaurant
2	Istanbul	Hasanpasha Konagi	41.007720	28.958768	Dutch Restaurant
3	London	My Old Dutch	51.517422	-0.122062	Creperie
4	London	Loccas	51.519842	-0.163475	Dutch Restaurant
5	Saint Petersburg	ска мразь	59.928168	30.359470	Dutch Restaurant
6	Kyiv	"Ибсен" ресторан (Ибсен)	50.437442	30.534536	Dutch Restaurant
7	Kharkiv	Ukrayna	49.993500	36.230385	Dutch Restaurant
8	Milan	GELATERIA AMBROSIANA	45.464084	9.188858	Dutch Restaurant
9	Brussels	Thon Hotel Restaurant	50.855912	4.356881	Dutch Restaurant

4. Merge Cities and Venues datasets, group and clean to get statistics by city

Now I have raw data of venues. Next step is to merge Venues data set with Cities and group it by city name to obtain statistics – how many of venues of given type are in each city.

Here is a result set (top 10 records):

	cityName	population	latitude	longitude	count
0	Istanbul	15029231	41.008238	28.978359	3.0
1	Moscow	13197596	55.755826	37.617300	0.0
2	London	9126366	51.507351	-0.127758	2.0
3	Saint Petersburg	5381736	59.934280	30.335099	1.0
4	Berlin	3748148	52.520007	13.404954	0.0
5	Madrid	3223334	40.416775	-3.703790	0.0
6	Kyiv	2950819	50.450100	30.523400	1.0
7	Rome	2857321	41.902783	12.496365	0.0
8	Paris	2140526	48.856614	2.352222	0.0
9	Bucharest	2106144	44.426767	26.102538	0.0

5. Normalize statistics by city population

Additional step is to normalize statistics by city population (make proportion) and avoid NaN values by replacing it with zeros in case if there are no such restaurants in the city at all.

6. Make visualization on Europe map

Visualization is the main part of this part. I implemented to ways of visualization – Europe map with “bubbles” (main one) and word cloud (additional one).

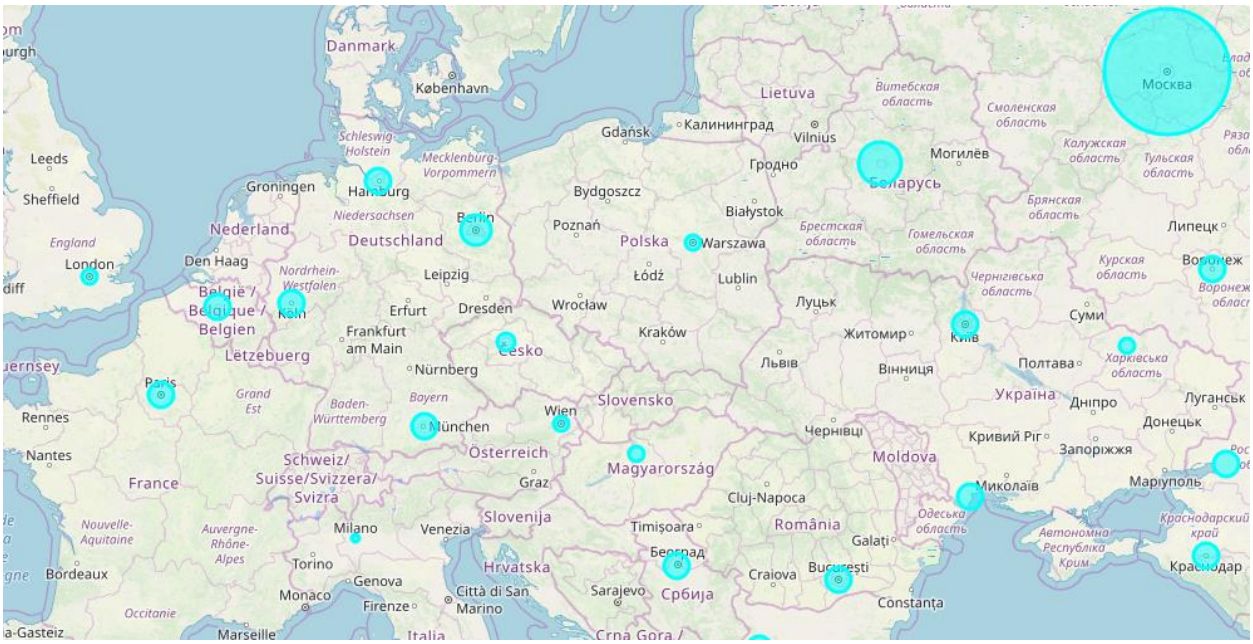
Few examples of these visualizations can be found in Results section.

5. Results

Let’s test the methodology on the real data.

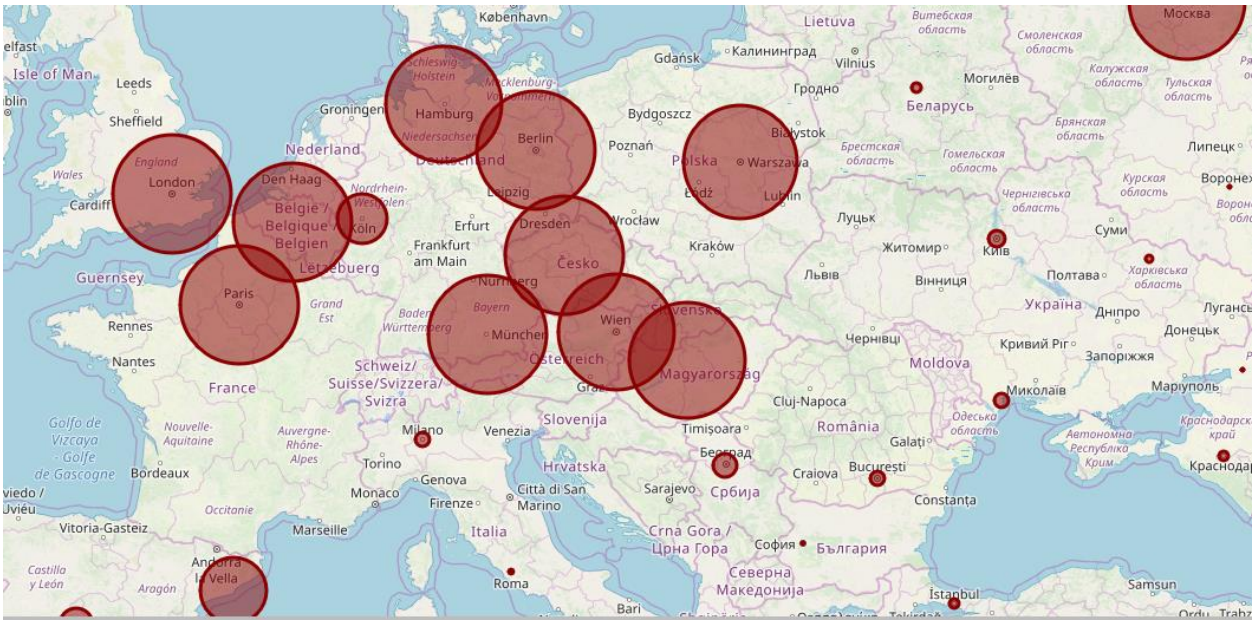
I took 3 ethnic categories for my result: Russian, Vietnamese and Spanish.

Russian cuisine across Europe:



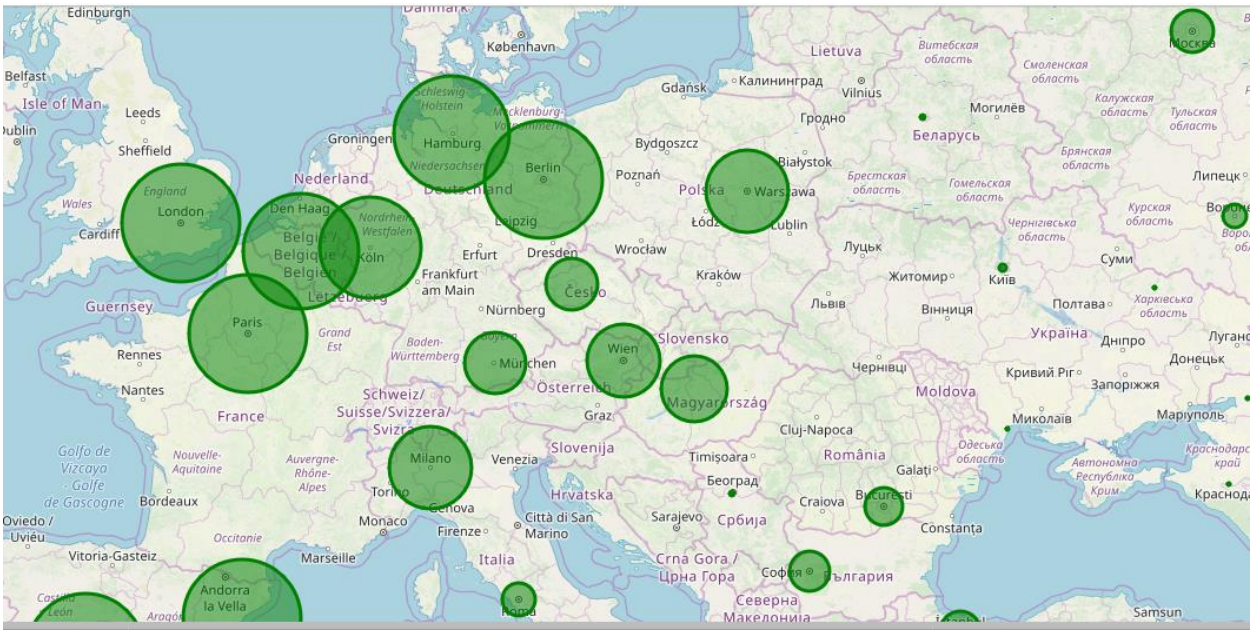
Obviously, Russian restaurants are mostly presented in former USSR countries. I would say, though, that it is not that hard to find them in Western Europe as well.

Vietnamese cuisine across Europe:



Surprisingly, Vietnamese cuisine is quite popular in Europe, but mostly in multi-cultural capitals such as London, Berlin, Moscow, Paris. It is almost not presented in Easter Europe except Moscow which is exceptional case due to its size and multi-cultural nature.

Spanish cuisine across Europe:



Besides Spain, Spanish restaurants are presented mostly in north part of Europe. I believe that warm character of this cuisine is very attractive in more cold countries.

Additional visualization – word clouds

Another very interesting idea is to visualize most popular types of restaurants for every city from my list. I used Word Cloud diagrams for it.

Here are few examples of it.

Istanbul:

Moscow:

Berlin:

Madrid:

6. Conclusion

As result I got very interesting data of what cuisines are preferred by citizens and visitors of main cities of Europe. These data and the way I visualized it for sure will be helpful for researchers as well as for tourists and food lovers around the world.