The Battle of the Neighborhoods – Sofia vs Bucharest

Martin Koshnicharov
January 17, 2021

1. Introduction / Business Problem

In my Business Problem I will analyse and compare **restaurants** and the **clusters of food venues** in the center of two different European cities, namely - **Sofia and Bucharest**.

As a Bulgarian it is interesting to me to compare two cities like **Sofia and Bucharest**, which are still in a transition period after the fall of the Communist regime in East Europe and despite that they are changing and there are already people from different nationalities living there, they are still **not that multicultural cities comparing to many European capitals**.

I want to analyse whether the **richness of restaurants** with cuisines from different countries is **connected to the multiculturality** of a city or whether the natural people's desire to try dishes from all over the world makes the venues of a city multicultural, before the city has become such.

Additionally, I want to analyse whether there are any restaurants with traditional national dishes from one of the two countries in the other capital city (Bulgarian restaurants in Bucharest and Romanian restaurants in Sofia). And where in the city centre is the best place for such a restaurant to be opened.

This problem can be of a great use for private entrepreneurs from any of the two countries, who want to open a restaurant with traditional national cuisines in the neighbor country. It can also help governmental or tourism organisations who want to promote their city and the diversity of restaurants and food venues it offers.

2. Data

- The first step is to allocate a main touristic sightseeing place in both cities the Court House in Sofia and the Old Town in Bucharest.
- Second, the geocoordinates of the two locations will be defined, using Google Maps.
- The next step will be to define a surface with radius of 6 km from the central spot, which will be filled with a grid of small circles of radius 300 m. The Center of the small circles will be used to extract the nearby food venues with the help of Foursquare.
- The venues will be first used to extract the restaurants and analyse them how many restaurants there are nearby, whether there are any Bulgarian/Romanian restaurants.

Depending on that, places with smaller number of Bulgarian/Romanian and with smaller number of restaurants as a whole will be highlighted as good places for a restaurant to be opened.

- Later the founded venues will be used for a more expanded clustering (food venues based) of the city centre.
- K Means Clustering will be used for the both clustering cases restaurants and fodd venues based. It will help us find places with less number of restaurants nearby and will help us compare how similar they are to each other and how multicultural they are.

The Places selected:

- Sofia, the Court House 42.695138306155314, 23.320175955491948
- Bucharest, The Old Town 44.43320239120804, 26.10238305064323

The Court House in Sofia is at the beginning of the Vitosha Boulevard, which is the main commercial street in the center of Sofia, the capital of Bulgaria. The place is abundant in posh stores, restaurants and bars

The Old Town of Bucharest, Romania is located in the center of the city and is popular for its nightlife, restaurants, bars and other venues.

3. Methodology

• Restaurants Analysis

In the first part of our analysis we will use the obtained restaurants data in order to analyse the **density of restaurants** in the given areas (**radius of 6 km from the city center**). The density of Romanian/Bulgarian restaurants would be also of interest, but **as we will later find out, there are no Romanian restaurants in Sofia and no Bulgarian restaurants in Bucharest, too.**

Heatmaps will be used to identify and visualize the areas with low restaurants density.

Areas with no more than 2 restaurants in a radius of 250 meters will be allocated as the most promising areas for future restaurant owners. The locations will be marked and clusters of those locations (K-Means clustering) will be created as a starting point for deeper exploration of the areas.

• Food Venues Analysis

The **number** of food venues, the different types of unique **categories** of venues and the **frequency** of different types of food venues near every spot (center of "neighborhood") will be analyzed. The **top 5 and top 10 of the venues** for every given address will be defined. On the basis of **the top 10 venues a K-Means clustering** will be applied in order to find the **5 most**

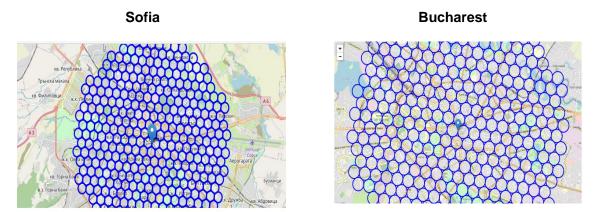
common clusters of top 10 venues in the city center. The 5 clusters from Sofia and their frequency will be compared to the five clusters in Bucharest and their frequency in order to get first impression of the atmosphere of the center of both cities.

Restaurants Analysis

As mentioned, the first step was to create a grid of circles("neighborhoods") with a radius of 300 meters in a big circle with a radius of 6 km from Sofia Court House/Bucharest the Old Town. Our grid of circles are our grid of "neighborhoods", which food venues we analyzed and clustered.

First, we work with metrics in meters. Later we convert the coordinates into latitude/longitude degrees to be shown on Folium map. So we use functions to convert between WGS84 spherical coordinate system (latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters).

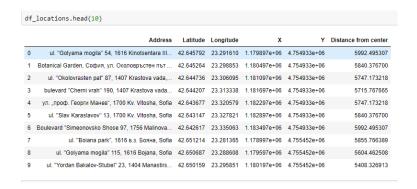
Next we create our hexagonal grid of cells.



Google Maps

We now have the coordinates of the centers which we will use to extract the venues nearby.

We use Google Maps API to get approximate addresses of those locations and save them in a data frame.



Foursquare

Having the locations, we use our Foursquare API to get the food venues and from them we extract the restaurants in every neighborhood.

We get the following results:

Sofia:

Total number of restaurants: 867
Total number of Romanian restaurants: 0
Percentage of Romanian restaurants: 0.00%

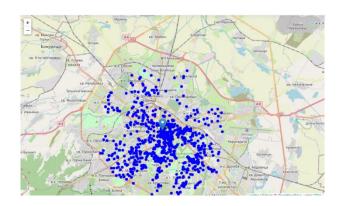
Average number of restaurants in neighborhood: 2.3873626373626373

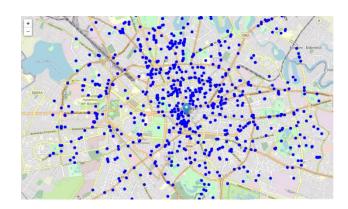
Bucharest:

Total number of restaurants: 955 Total number of Bulgarian restaurants: 0 Percentage of Bulgarian restaurants: 0.00%

Average number of restaurants in neighborhood: 2.6263736263736264

And we plot all the restaurants on the map of Sofia and Bucharest



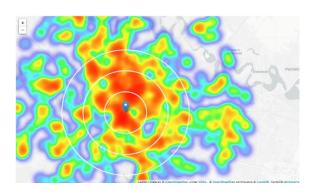


We continue with the processing and analysis of data in order to prepare it for the heatmaps. So we count the the **number of restaurants in every area** with a radius of 300m. With this data in hand we plot the heatmaps with the density of restaurants.

Sofia

11 (1249) 13 (147) 14 (176) 15 (176) 16 (176) 16 (176) 17 (176) 18 (1

Bucharest

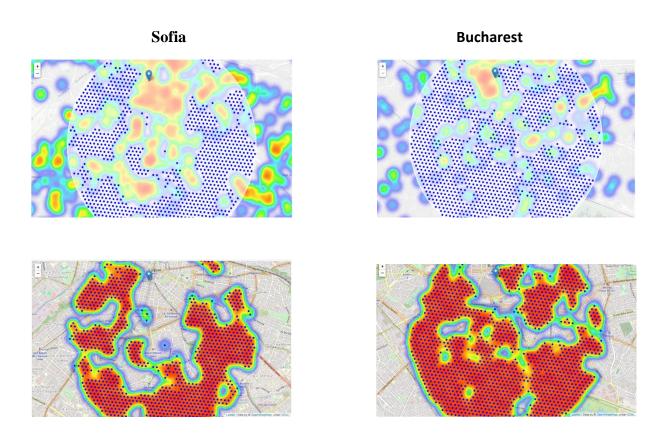


We further define a new, more narrow region of interest with radius 2500 meters, which will include low-restaurant-count South - East from Sofia's Court House and Bucharest's Old Town, which seem much more appealing than the part on the Nord – East on both locations. We then

create a new denser grid of "neighborhoods" only 50m in radius (100m "neighborhood" center from one another).

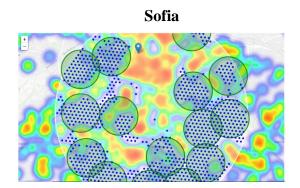
We then filter those locations: we're interested only in locations with no more than two restaurants in radius of 250 meters, and no Romanian/Bulgarian restaurants in a given radius, but as already metioned, since we don't have Romanian/Bulgarian restaurants in Sofia/Bucharest according to our data, this is not a factor any more.

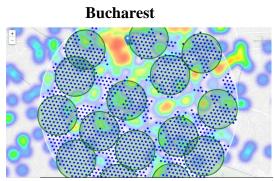
And we plot these locations on a map:



For Sofia these are 1048 and for Bucharest 1439 locations.

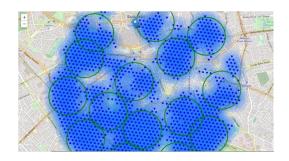
We then **cluster** these areas in order to get **centers of areas with good locations for further analysis**. Highlighting these zones, their centers and getting their addresses is the final part of our Restaurants Analysis.













Food Venues Analysis

We have the neighborhoods, the coordinates, the venues' names and category which we obtained at the beginning (the examples are taken from the data for Sofia):

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	ul. "Okolovrasten pat" 87, 1407 Krastova vada,	42.644736	23.306095	Atspas	42.643090	23.308539	Modern European Restaurant
1	bulevard "Cherni vrah" 190, 1407 Krastova vada	42.644207	23.313338	Mandjarnika	42.646186	23.314297	Restaurant
2	bulevard "Cherni vrah" 190, 1407 Krastova vada	42.644207	23.313338	Angeliki's Greek Taverna	42.646165	23.314415	Greek Restaurant
3	ul. "Slav Karaslavov" 13, 1700 Kv. Vitosha, Sofia	42.643147	23.327821	Neighbors	42.643944	23.330162	Italian Restaurant
4	Boulevard "Simeonovsko Shose 97, 1756 Malinova	42.642617	23.335063	Делфинчето	42.643694	23.337646	Seafood Restaurant

We then check the number of venues in every neighborhood and save this for later use.

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
"Slivnitsa" Metro Station, 1360 Moderno predgradie, Sofia	3	3	3	3	3	3
14 Tsar Osvoboditel Blvd., 5th floor, 1000 Old City Center, Sofia	16	16	16	16	16	16
62A, 1303 zh.k. Zona B-5-3, Sofia	17	17	17	17	17	17
Balkankar AD, 1618 g.k. Slavia, Sofia	5	5	5	5	5	5
Block E2, 1404 Manastirski Livadi, Sofia	5	5	5	5	5	5
Boulevard "Akademik Ivan Evstratiev Geshov " 19, 1431 Sofia Center, Sofia	4	4	4	4	4	4
Boulevard "Professor Tsvetan Lazarov" 13, 1574 SPZ Slatina, Sofia	4	4	4	4	4	4
Boulevard "Professor Tsvetan Lazarov" 14, 1784 м. Къро, Sofia	1	1	1	1	1	1
Boulevard "Shipchenski Prohod" 69, 1574 g.k. Hristo Smirnenski, Sofia	7	7	7	7	7	7
Boulevard "Simeonovsko Shose 4, 1700 zh.g. Zoopark, Sofia	3	3	3	3	3	3

We then obtain the number of the unique categories of venues and the frequency of every given category. Next we group rows by neighborhood and take the mean of the frequency of occurrence of each category.

	Neighborhood	Afghan Restaurant	American Restaurant	Asian Restaurant	BBQ Joint	Bagel Shop	Bakery	Bistro	Breakfast Spot	Buffet	Bulgarian Restaurant	Burger Joint	Burrito Place	Cafeteria	Café	F
("Slivnitsa" Metro Station, 1360 Moderno predgr	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0000	0.0	0.0	0.000000	0.0000	0.0	0.0	0.000000	
1	14 Tsar Osvoboditel Blvd., 5th floor, 1000 Old	0.0	0.0	0.0	0.062500	0.0	0.125000	0.0625	0.0	0.0	0.000000	0.0625	0.0	0.0	0.000000	
2	62A, 1303 zh.k. Zona B-5- 3, Sofia	0.0	0.0	0.0	0.058824	0.0	0.058824	0.0000	0.0	0.0	0.058824	0.0000	0.0	0.0	0.294118	
3	Balkankar AD, 1618 g.k. Slavia, Sofia	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0000	0.0	0.0	0.000000	0.0000	0.0	0.0	0.000000	
4	Block E2, 1404 Manastirski Livadi, Sofia	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0000	0.0	0.0	0.000000	0.0000	0.0	0.0	0.200000	

We then print each neighborhood along with the top 5 most common venues:

```
----"Slivnitsa" Metro Station, 1360 Moderno predgradie, Sofia----
venue freq

0 Restaurant 0.33

1 Eastern European Restaurant 0.33

2 Comfort Food Restaurant 0.03

3 Modern European Restaurant 0.00

4 Paella Restaurant 0.00

----14 Tsar Osvoboditel Blvd., 5th floor, 1000 Old City Center, Sofia----
venue freq

0 Restaurant 0.19

1 Modern European Restaurant 0.19

2 Italian Restaurant 0.12

3 Bakery 0.12

4 Burger Joint 0.06

----62A, 1303 zh.k. Zona B-5-3, Sofia----
venue freq
```

After that we put that into a *pandas* data frame, sort the venues in a descending order, create a new data frame and display the top 10 venues for each neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	"Slivnitsa" Metro Station, 1360 Moderno predgr	Restaurant	Comfort Food Restaurant	Eastern European Restaurant	Wings Joint	Czech Restaurant	Diner	Doner Restaurant	Donut Shop	Ethiopian Restaurant	Falafel Restaurant
1	14 Tsar Osvoboditel Blvd., 5th floor, 1000 Old	Modern European Restaurant	Restaurant	Bakery	Italian Restaurant	Vegetarian / Vegan Restaurant	BBQ Joint	Bistro	Soup Place	Burger Joint	Sandwich Place
2	62A, 1303 zh.k. Zona B-5-3, Sofia	Café	Eastern European Restaurant	Pizza Place	Food Court	Donut Shop	Salad Place	Fast Food Restaurant	Bulgarian Restaurant	Bakery	BBQ Joint
3	Balkankar AD, 1618 g.k. Slavia, Sofia	Restaurant	Fast Food Restaurant	Wings Joint	Diner	Doner Restaurant	Donut Shop	Eastern European Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fish & Chips Shop
4	Block E2, 1404 Manastirski Livadi, Sofia	Pizza Place	Vegetarian / Vegan Restaurant	Food Truck	Café	Falafel Restaurant	Deli / Bodega	Diner	Doner Restaurant	Donut Shop	Eastern European Restaurant

We use ${\bf k}$ – **Means Clustering** to put the neighborhoods in 5 clusters depending on the top 10 venues for each neighborhood.

We create a new data frame that includes the cluster and the top 10 venues for each neighborhood.

				,								
:		Address	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
	2	ul. "Okolovrasten pat" 87, 1407 Krastova vada,	42.644736	23.306095	4	Modern European Restaurant	Wings Joint	Fast Food Restaurant	Diner	Doner Restaurant	Donut Shop	Eastern European Restaurant
	3	bulevard "Cherni vrah" 190, 1407 Krastova vada	42.644207	23.313338	1	Greek Restaurant	Restaurant	Czech Restaurant	Deli / Bodega	Diner	Doner Restaurant	Donut Shop
	5	ul. "Slav Karaslavov" 13, 1700 Kv. Vitosha, Sofia	42.643147	23.327821	4	Italian Restaurant	Wings Joint	Fish & Chips Shop	Diner	Doner Restaurant	Donut Shop	Eastern European Restaurant
	6	Boulevard "Simeonovsko Shose 97, 1756 Malinova	42.642617	23.335063	4	Bakery	Café	Fast Food Restaurant	Eastern European Restaurant	BBQ Joint	Chinese Restaurant	Seafood Restaurant

Finally, we visualize the resulting clusters:

Sofia

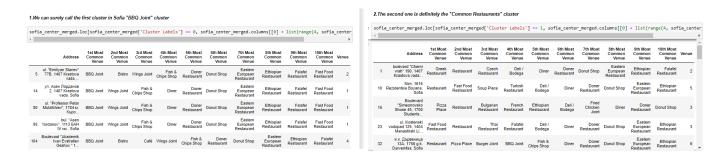


Bucharest



We then analyze every particular cluster and give them names depending on the frequency of venues in it.

Example from the data for Sofia:



We got the following results for both cities:

Sofia: the predominant cluster in Sofia is luckily the fourth one (orange) - the "Rich on International Restaurants" cluster, followed by the "Common Restaurants" (dark blue) and the "Cafes" (light blue) cluster. And again luckily much - much less neighborhoods fall in the "BBQ Joint" (red) and "The Fast Food" (green) clusters.

Bucharest: The predominant cluster for Bucharest is rich on international restaurants, but not that well defined. There are a lot of spots with predominant common restaurants, Romanian restaurant and a lot of different food venues, too. That is why we called it a "Mixed" cluster (green). And it is not that much dominating the area like the predominant one in Sofia. The second one is the "Cafes" cluster (light blue) with a small edge over the "Restaurants Cluster" (purple cluster) with mainly common restaurants and some other types like Romanian, East –European restaurants, Hungarian and others. The less often to find are the "Pizza Place" (orange) and the "Italian Restaurants" (red) clusters.

4. Results

Our analysis showed that surprisingly there are no Romanian/Bulgarian restaurants in Sofia/Bucharest in a circle with a radius of 6km from the Court House/Old Town. I further conducted an Internet research which also gave the same results. For both cities I found only one sign of a venue with a Bulgarian dishes in Bucharest which was approximately 4-5 km from the Old Town and it advertisement was not very clear (it stayed that it is a fast food venue). So I am not sure if it can be even found on Foursquare.

I think that these are good news for eventual stakeholders/entrepreneurs who want to open such a restaurant in one of the both countries. Although the national dishes on the Balkans are similar there is still big richness and a variety of cuisines in every country and the citizens of these both capitals will be happy to try some of the national dishes of their neighbor country. I think that the main reason that there are still not such restaurants is because of the similarity of both national dishes, second because there are not many expats from the neighbor countries in the both cities and the third - the economic situation of both countries in the transition period.

Our analysis can be very useful for future stakeholders, because we highlighted very appealing areas of the city for such an investment. After defining the zone with a 6km radius around the chosen spot, we extracted data for all the food venues and restaurants in it. Using heatmaps to show us where the areas with high density of restaurants are, we defined new areas south - east from the chosen central spots. Those areas have low density of restaurants - less than two in a radius of 250m. The initial idea that the areas should also have no Romanian/Bulgarian restaurant in a radius of 400m didn't need further analysis, because there weren't such restaurants in the whole central area we chose.

We clustered these zones with low density of restaurants and got 15 zones/clusters that are the perfect places for a new Romanian/Bulgarian restaurant. The addresses of these new best "neighborhoods" were recorded which will be the perfect starting point for further analysis of the "neighborhood" and the choice of a restaurant place.

We used the obtained data of food venues in order to cluster the city center and define areas with different predominant venues, tastes, national dishes, which help us get some initial picture for the atmosphere of the city. We found that the predominant cluster for the city center of Sofia is the one rich on international restaurants, followed by the cluster rich on common restaurants and caffees. At the back are the BBQ joint and the Fast food clusters, which paints a very nice first impression for the city center of Sofia.

The predominant cluster for Bucharest is also rich on international restaurants, but not that well defined. There are a lot of spots with predominantly common restaurants, Romanian restaurants and a lot of different food venues, too. That is why we called it a "Mixed" cluster. And it is not that much dominating the area like the predominant one in Sofia. The second one is the "Cafes" cluster with a small edge over the "Restaurants" cluster with mainly common restaurants and some other types like Romanian, East - European, Hungarian and others. The less often to find are the "Pizza Place" and the "Italian Restaurants" clusters. But as a whole, the division of clusters in Bucharest shows less dominance of one of them, as already mentioned. Another particular feature of Bucharest is the very high number of Italian Restaurants which is pretty obvious.

5. Discussion

As already mentioned, the analysis is only a starting point for a further more thorough analysis of suitable places for Romanian/Bulgarian restaurants. The defined 15 clusters of neighborhoods with low density of restaurants were highlighted as circles, but actually they have an irregular form and the found addresses are only the starting point for further analysis of the area. There can be many reasons why there are not that many restaurants in these zones, which can make them unsuitable for an investment.

Another topic which should be discussed is the clustering of the food venues. Since the "K-Means" function performs unsupervised Machine Learning process, we can get slightly different results, every time we perform it. Especially when we run it on not that easily separable data like the food venues in Bucharest (that is why we have this "Mixed" cluster). The data for Sofia is pretty stable and gives almost the same results, but the one for Bucharest gives slightly different clustering every time it is run. The main separation of clustering and their dominance in the city center is still recognizable, but one should not be surprised if she doesn't get the results discussed and the graphs showed to 100%.

As already discussed, too, I was a little bit surprised that there are no Romanian restaurants in Sofia and no Bulgarian restaurants in Bucharest center, but I think that this is mainly due to reasons mentioned in the "Results" section. Both Balkan countries and its capitals faced a big transformation in the last thirty years and I am sure that the lack of such restaurants is actually a very good niche for future stakeholders and such venues would be very profitable if their place, interior and service are on the needed level.

The next step in the further analysis of the center of both cities is to analyze all the venues in it (not only the food venues). This will help us see the full picture of it and discover their atmosphere.

6. Conclusion

Our analysis is a good starting point for a business plan for a new restaurant or for state administration or tourism companies which want to promote both cities. We got 15 addresses in each city, which are a good starting point for analysis of potential neighborhoods for a future Romanian or Bulgarian restaurant. We analyzed the "flavor" of both city-centers and highlighted spots where we can relax and enjoy a nice meal after a long tourist walk along the city. Next time we visit one of the both cities - either as a tourist or on a business trip, we know where we can find delicious food.