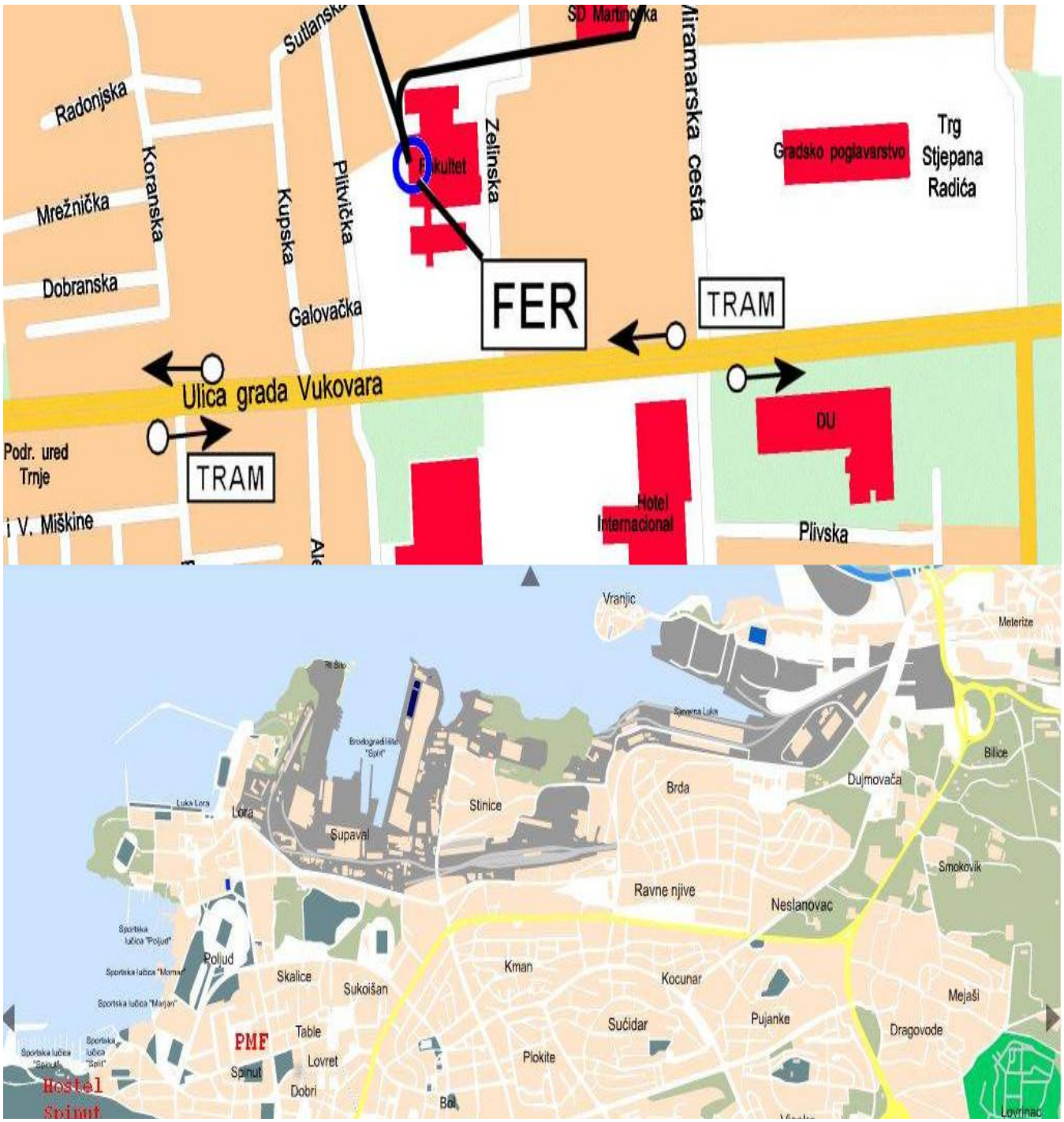
Applied Data Science Capstone Project

THE BATTLE OF NEIGHBORHOODS

Finding similar neighborhoods in Zagreb and Split, Croatia

Mile Vučković / milevucko@gmail-com, January, 2021



1 Introduction

This report is made solely for the purpose of the Coursera IBM Data science Capstone Project and it represents final delivery along with accompanying notebook and presentation.

The purpose of the project is to demonstrate knowledge of data analysis and machine learning libraries in Python, as well as acquired related skills.

This project tries to solve an imaginary, but very common problem on the Croatian real estate market. I hope it will help actors at both ends of real estate transactions, especially private traders, but also those who provide real estate services, such as real estate agencies. It includes querying Foursquare location data for the purpose of comparing city neighborhoods.

2 Business Problem

Let's first explain the niche that we are trying to fill with this solution, as well as the accompanying assumptions.

Zagreb is the capital of Croatia with almost a million inhabitants, and is located in the inland of Croatia. The climate type in Zagreb is 'Cfb' (Moderately warm humid climate with warm summers). Just like many other CE countries, Croatia is a highly centralized country, which means that major companies, government agencies, universities, and other vital institutions are located in Zagreb.

Split is the second largest town in Croatia, and is situated south, on the Adriatic coast. Climate type in Split is 'Csa' (Mediterranean climate with hot summers). Its main economic branch is tourism. As of 2017, the port of Split ranks as the largest passenger port in the Adriatic, and the 11th largest port in the Mediterranean, with annual passenger volume of approximately 5 million.

Because of all of the above, a large number of people from Zagreb want to buy, or trade, a real estate in Split, especially when they reach pension so that they can enjoy in the warmer

winter climate of the Adriatic coastline. On the other hand, a large number of people from Split that are not involved in tourism are drawn to Zagreb, either due to larger labor market, number and strength of employers, cheaper real estates, or other reasons.

The main idea of this project is to create clusters of neighborhoods of these two cities, so that we can compare them and provide insight as a service for actors on the real estate market. Top venues of neighborhoods are going to be used as data for neighborhood comparison.

3 Data

The following data are used in this project:

1. List of the neighborhoods in Zagreb and Split.
2. Geocoordinates of the neighborhoods in Zagreb and Split.
3. Top venues of neighborhoods in Zagreb and Split.

3.1 Neighborhoods in Zagreb and Split

Neighborhoods in Split are loaded from the excel sheet, picked up from the city's official website (https://www.split.hr/DesktopModules/Bring2mind/DMX/API/Entries/Download?language=hr-HR&Command=Core_Download&EntryId=6080&PortalId=0)

As for Zagreb's neighborhoods, the data is collected from Wikipedia (https://hr.wikipedia.org/wiki/Zagreba%C4%8Dke_gradske_%C4%8Detvrti)

3.1 Geocoordinates of the neighborhoods in Zagreb and Split

The latitude and longitude of the neighborhoods are retrieved using a geopy library. Neighborhood address is used as the input parameter for geocoordinate retrieval.

3.2 Top venues of neighborhoods in Zagreb and Split

The venue data is collected using Foursquare REST API. Number of collected venues is limited to 1000 per neighborhood. When collecting venues, a 800 meter radius is used around the center coordinates of the neighborhood.

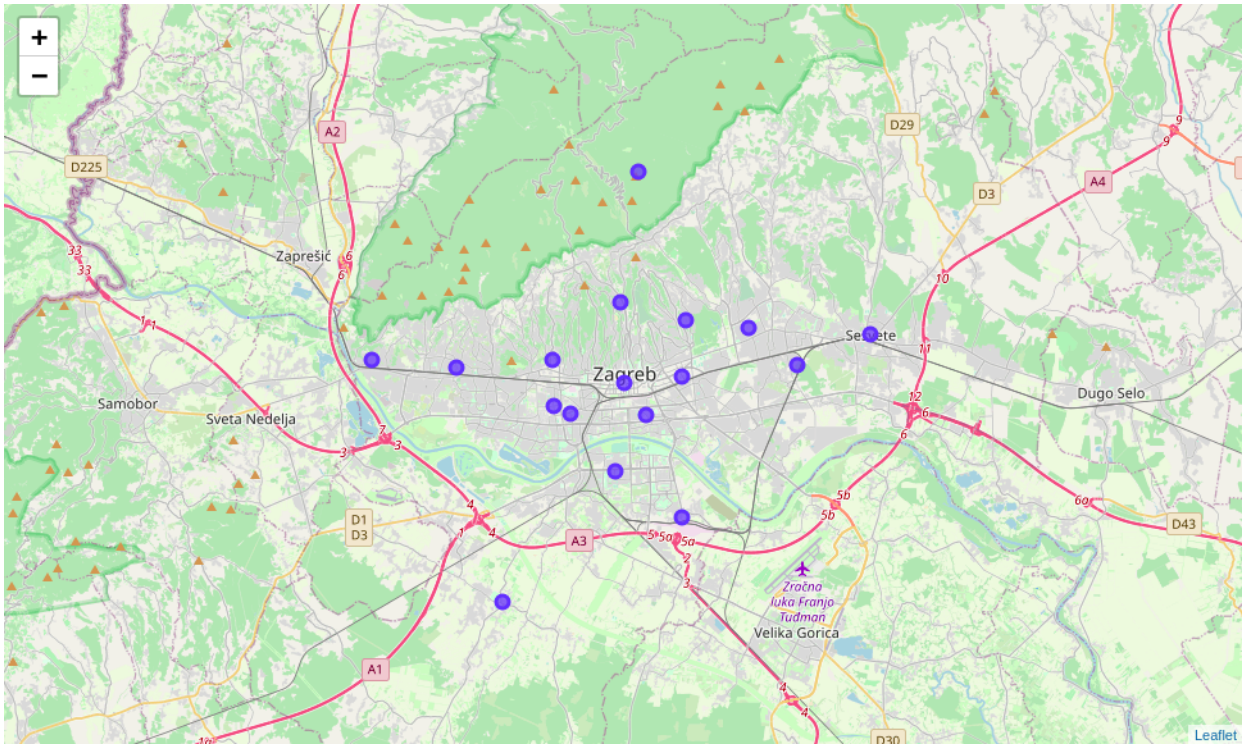
3 Methodology

After cleaning up the data, specifically K Means machine learning for creating clusters of neighborhoods is used. Silhouette score metrics are applied for choosing the optimal number of clusters. The final considerations are made using that optimal cluster number.

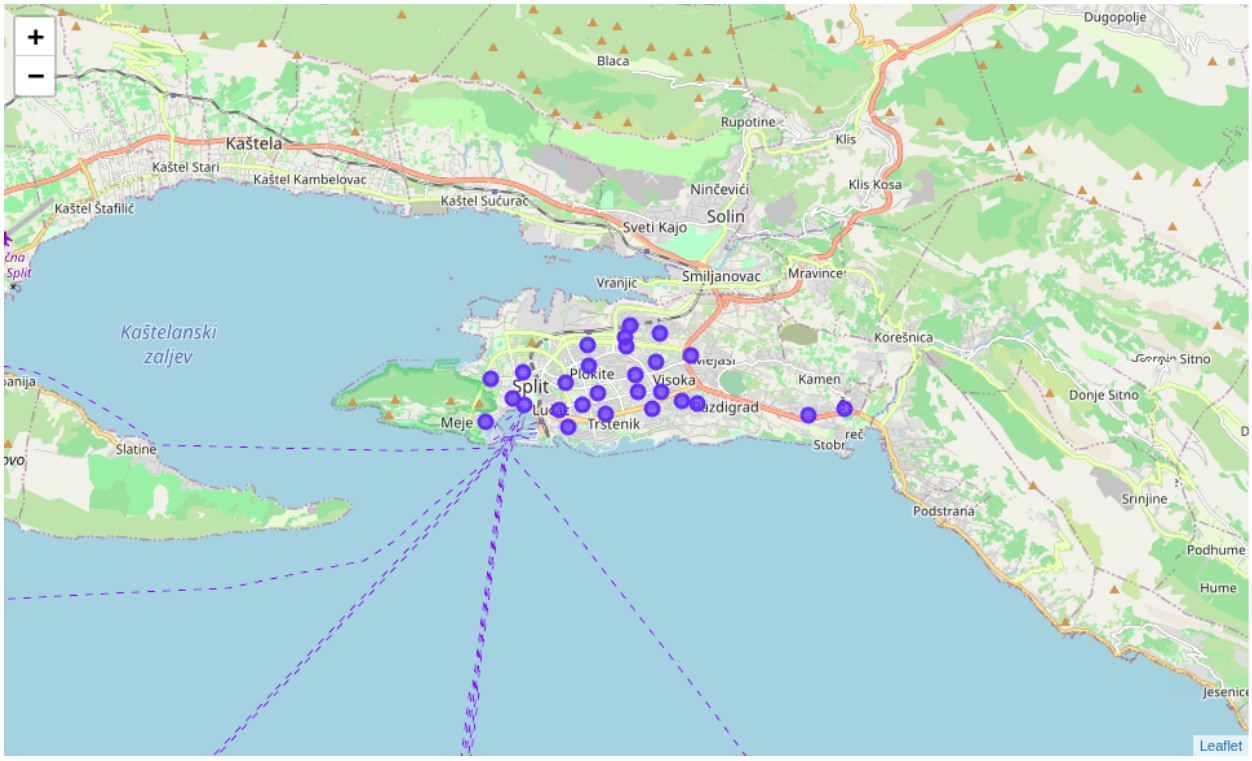
So, first thing's first: tidying up the data and inserting geocoordinates into pandas dataframe with the help of Python geocode library. We will end up with an overall of 44 neighborhoods in both cities. Let’s show them in a dataframe and then on a map using folium library.

	Neighborhood	Town	latitude	longitude
0	Donji grad	Zagreb	45.809878	15.977081
1	Gornji grad - Medveščak	Zagreb	45.839926	15.975096
2	Kruge	Zagreb	45.797630	15.988700
3	Maksimir	Zagreb	45.833312	16.010151
4	Peščenica - Žitnjak	Zagreb	45.812034	16.008027
5	Novi Zagreb - istok	Zagreb	45.759520	16.007969
6	Novi Zagreb - zapad	Zagreb	45.776532	15.972081
7	Trešnjevka - sjever	Zagreb	45.801108	15.938832
8	Trešnjevka - jug	Zagreb	45.798288	15.947751
9	Črnomerec	Zagreb	45.818301	15.938581
10	Gornja Dubrava	Zagreb	45.830290	16.043665
11	Donja Dubrava	Zagreb	45.816409	16.069855
12	Stenjevec	Zagreb	45.815391	15.887082
13	Podsused - Vrapče	Zagreb	45.818406	15.841420
14	Podsljeme	Zagreb	45.888748	15.984549
15	Sesvete	Zagreb	45.827605	16.108984
16	Brezovica	Zagreb	45.727649	15.911798
17	GK BAČVICE	Split	43.503720	16.449910
18	GK BLATINE ŠKRAPE	Split	43.506350	16.459912
19	GK BOL	Split	43.512204	16.449302
20	GK BRDA	Split	43.523212	16.466332
21	GK GRAD	Split	43.508028	16.438180
22	GK GRIPE	Split	43.508042	16.453580
23	GK KMAN	Split	43.519505	16.454961
24	GK KOCUNAR	Split	43.519219	16.465257
25	GK LOKVE	Split	43.510136	16.458023
26	GK LOVRET	Split	43.514318	16.437930
27	GK LUČAC MANUŠ	Split	43.506923	16.447698
28	GK MEJE	Split	43.504774	16.428016
29	GK MEJAŠI	Split	43.517509	16.482567
30	GK MERTOJAK	Split	43.508746	16.480214
31	GK NESLANOVAC	Split	43.521584	16.474297
32	GK PLOKITE	Split	43.515554	16.455609
33	GK PUJANKE	Split	43.516156	16.473191
34	GK RAVNE NJIVE	Split	43.520962	16.464976
35	GK SIROBUJA	Split	43.505995	16.513706
36	GK SPINUT	Split	43.512971	16.429533
37	GK SPLIT 3	Split	43.510466	16.468661
38	GK SUĆIDAR	Split	43.513754	16.467775
39	GK ŠINE	Split	43.507287	16.523303
40	GK TRSTENIK	Split	43.507303	16.472339
41	GK VAROŠ	Split	43.509172	16.435229
42	GK VISOKA	Split	43.510468	16.474535
43	GK ŽNJAN	Split	43.508112	16.484383

Neighborhoods of Zagreb.



Neighborhoods of Split.



Now is the time to use Foursquare REST API to get venues in order to augment neighborhood data. The data is collected and here are just few records to serve as an example:

	Town	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Zagreb	Donji grad	45.809878	15.977081	Trg Nikole Šubića Zrinskog | Zrinjevac	45.810244	15.978109	Plaza
1	Zagreb	Donji grad	45.809878	15.977081	Quahwa	45.811030	15.975471	Café
2	Zagreb	Donji grad	45.809878	15.977081	Heritage Croatian Food	45.810810	15.979922	Comfort Food Restaurant
3	Zagreb	Donji grad	45.809878	15.977081	Buzz Bar Zagreb	45.808939	15.974639	Bar
4	Zagreb	Donji grad	45.809878	15.977081	Korica	45.807549	15.974863	Bakery
...	...	...	...	...	...	...	...	...

There are 118 unique categories of venues in Zagreb, and 99 unique categories in Split. Overall, there are 157 unique categories in both Zagreb and Split.

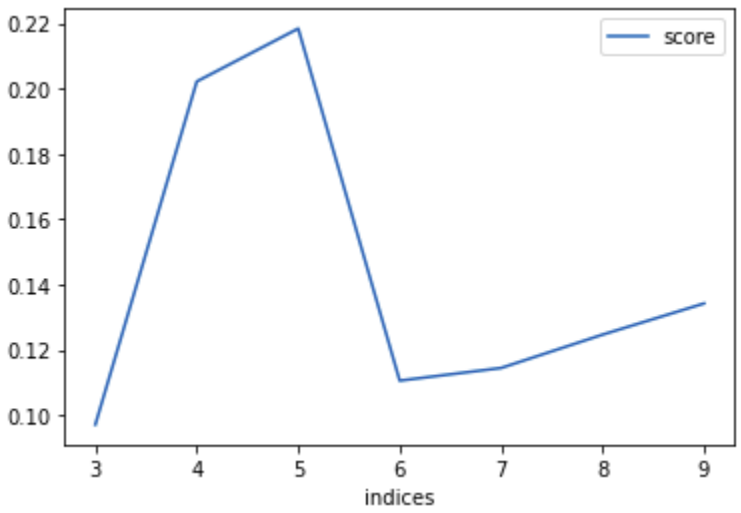
In order to analyze neighbourhoods, we'll apply a function that gets ten top venues for each neighborhood:

	Town	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	Split	GK BAČVICE	Café	Bar	Mediterranean Restaurant	Pizza Place	Hotel	Restaurant	Beach	Nightclub	Cocktail Bar	Bistro
1	Split	GK BLATINE ŠKRAPE	Café	Bar	Restaurant	Coffee Shop	Mediterranean Restaurant	Pizza Place	Hotel	Eastern European Restaurant	Beach	Spa
2	Split	GK BOL	Hotel	Hostel	Grocery Store	Bar	Café	Pizza Place	Restaurant	Farmers Market	Bistro	Fast Food Restaurant
3	Split	GK BRDA	Grocery Store	Big Box Store	Cafeteria	Mediterranean Restaurant	Café	Electronics Store	Bar	Women's Store	Farmers Market	Falafel Restaurant
4	Split	GK GRAD	Mediterranean Restaurant	Boat or Ferry	Restaurant	Pizza Place	Hostel	Bar	Italian Restaurant	Plaza	Ice Cream Shop	Café

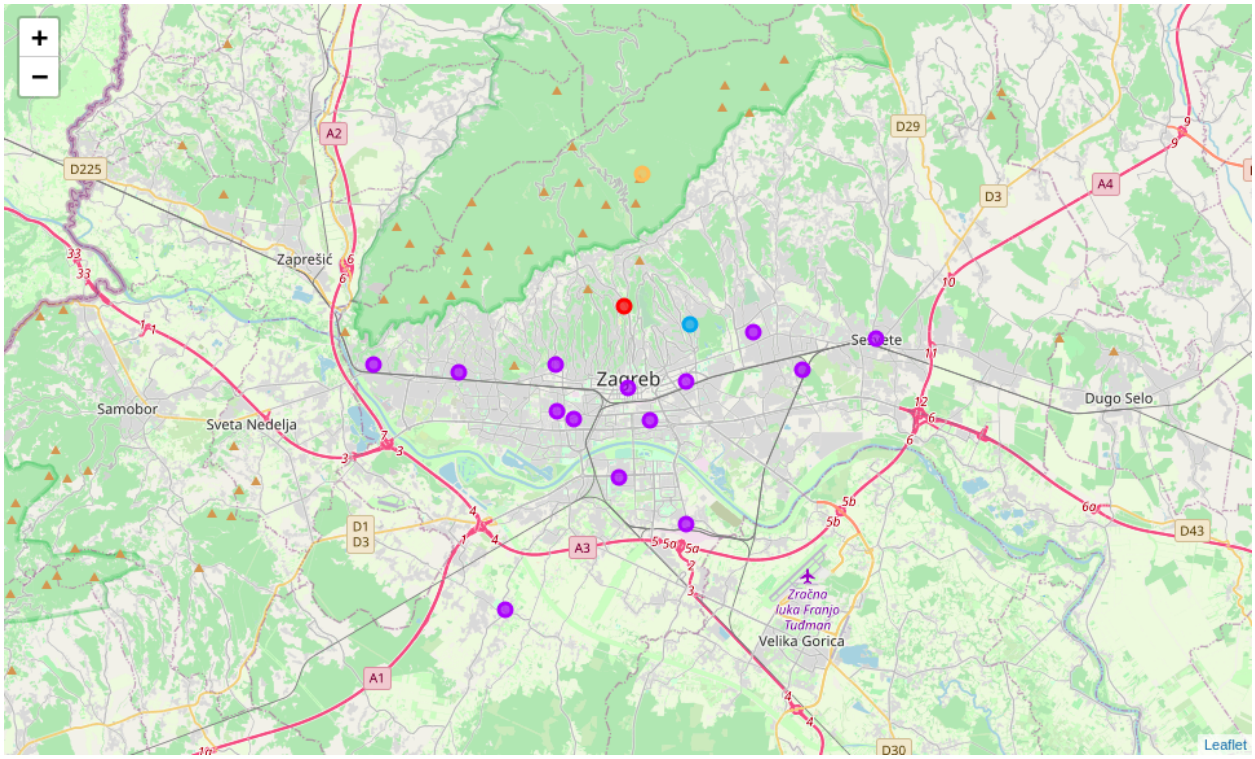
We then use one-hot encoding technique in order to convert categorical into numerical values so they can be used by K means, which is a form of unsupervised machine learning algorithm, or even more narrow - clustering algorithm.

Next, we'll find the optimal number of clusters in range between 3 and 10 since the K means requires the number of clusters as an input parameter. Optimal number is found using silhouette_score metrics. So, let's plot the metrics.

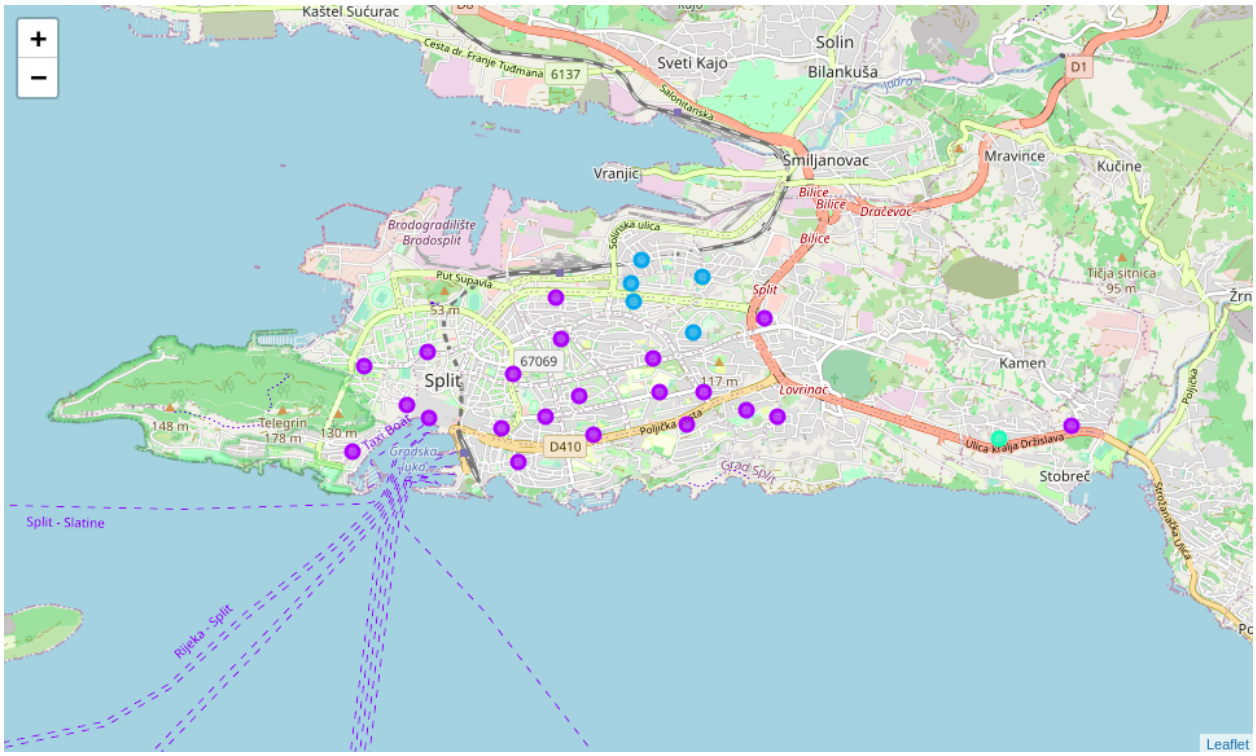
Image showing number 5 as optimal number of clusters.



Once again, KMeans class from sklearn python library is used, with the same *random_state* input parameter, and this time input is set for 5 clusters. The labels are finally added to pandas dataframe. Let's visualize neighborhood clusters in Zagreb:



And, now let's visualize neighborhood clusters of Split:



4 Discussion

Let's dive one last time into venues in our neighborhood clusters.

Cluster 1

```
df_merged.loc[df_merged['Cluster Labels'] == 0, :]
```

	Neighborhood	Town	latitude	longitude	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
1	Gornji grad - Medveščak	Zagreb	45.839926	15.975096	Light Rail Station	Eastern European Restaurant	Field	Pub	Electronics Store	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Eye Doctor	Exhibit

Cluster 2

```
df_merged.loc[df_merged['Cluster Labels'] == 1, :]
```

	Neighborhood	Town	latitude	longitude	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 Cor
0	Donji grad	Zagreb	45.809878	15.977081	Café	Plaza	Restaurant	Bar	Bistro	Dessert Shop	Theater	Mediterr Restı
2	Kruge	Zagreb	45.797630	15.988700	Café	Bar	Concert Hall	Restaurant	BBQ Joint	Resort	Supermarket	Pizza
4	Peščenica - Žitnjak	Zagreb	45.812034	16.008027	Café	Restaurant	Bar	Grocery Store	Dessert Shop	Mediterranean Restaurant	Bus Station	Superr
5	Novi Zagreb - istok	Zagreb	45.759520	16.007969	Bakery	Sushi Restaurant	Hotel	Restaurant	Train	Diner	Soccer Field	Paper / Supplies
6	Novi Zagreb - zapad	Zagreb	45.776532	15.972081	Café	Bakery	Bar	Gym	Fast Food Restaurant	Soccer Field	Furniture / Home Store	Food &
7	Trešnjevka - sjever	Zagreb	45.801108	15.938832	Bar	Café	Bakery	Grocery Store	Liquor Store	Restaurant	Shop & Service	Pizza
8	Trešnjevka - jug	Zagreb	45.798288	15.947751	Café	Bar	Bakery	Electronics Store	Shop & Service	Pizza Place	Drugstore	Grocery
9	Črnomerec	Zagreb	45.818301	15.938581	Café	Supermarket	Grocery Store	BBQ Joint	Pizza Place	Dog Run	Bakery	
10	Gornja Dubrava	Zagreb	45.830290	16.043665	Café	Light Rail Station	Bar	Pizza Place	Clothing Store	Supermarket	Pharmacy	Restı
11	Donja Dubrava	Zagreb	45.816409	16.069855	Electronics Store	Train Station	Supermarket	Bus Station	Bakery	Women's Store	Fast Food Restaurant	Fa M
12	Sjenjevec	Zagreb	45.815391	15.887082	Café	Pizza Place	Grocery Store	Bar	Soccer Field	Supermarket	Restaurant	Sport
13	Podsused - Vrapče	Zagreb	45.818406	15.841420	Café	Hookah Bar	Bakery	Pub	Park	Bus Station	Brewery	Grocery
15	Sesvete	Zagreb	45.827605	16.108984	Café	Cosmetics Shop	BBQ Joint	Big Box Store	Furniture / Home Store	Grocery Store	Electronics Store	E Eur Restı
16	Brezovica	Zagreb	45.727649	15.911798	Stables	Historic Site	Food & Drink Shop	Pharmacy	Bar	Women's Store	Electronics Store	Fa M
17	GK BAČVICE	Split	43.503720	16.449910	Café	Bar	Mediterranean Restaurant	Pizza Place	Hotel	Restaurant	Beach	Nig
18	GK BLATINE ŠKRAPE	Split	43.506250	16.459912	Café	Bar	Restaurant	Coffee Shop	Mediterranean Restaurant	Pizza Place	Hotel	E Eur Restı
19	GK BOL	Split	43.512204	16.449302	Hotel	Hostel	Grocery Store	Bar	Café	Pizza Place	Restaurant	Fa M
21	GK GRAD	Split	43.508028	16.438180	Mediterranean Restaurant	Boat or Ferry	Restaurant	Pizza Place	Hostel	Bar	Italian Restaurant	
22	GK GRIPE	Split	43.508042	16.453580	Café	Nightclub	Fast Food Restaurant	Hotel	Bar	Restaurant	Bistro	Mediterr Restı
23	GK KMAN	Split	43.519505	16.454961	Hotel	Fast Food Restaurant	Café	Bar	Mobile Phone Shop	Grocery Store	Multiplex	Cal
25	GK LOKVE	Split	43.510136	16.458023	Café	Bar	Coffee Shop	Eastern European Restaurant	Bistro	Fast Food Restaurant	Mediterranean Restaurant	Fa M
26	GK LOVRET	Split	43.514218	16.437930	Mediterranean Restaurant	Restaurant	Grocery Store	Café	Hostel	Bar	Hotel	
27	GK LUČAC MANUŠ	Split	43.506923	16.447698	Mediterranean Restaurant	Café	Bar	Nightclub	Hostel	Hotel	Restaurant	Restı
28	GK MEJE	Split	43.504774	16.428016	Mediterranean Restaurant	Restaurant	Hotel	Café	Pizza Place	Italian Restaurant	Harbor / Marina	
29	GK MEJAŠI	Split	43.517509	16.462567	Fast Food Restaurant	Clothing Store	Grocery Store	Supermarket	Shopping Mall	Café	Toy / Game Store	Pet
30	GK MERTOJAK	Split	43.508746	16.460214	Café	Fast Food Restaurant	Bar	Irish Pub	Park	Grocery Store	Beach	Mediterr Restı
32	GK PLOKITE	Split	43.515554	16.455609	Hotel	Café	Fast Food Restaurant	Bar	Pizza Place	Basketball Stadium	Farmers Market	Music
36	GK SPINUT	Split	43.512971	16.429533	Restaurant	Café	Grocery Store	Burger Joint	Ice Cream Shop	Mediterranean Restaurant	Plaza	Pizza
37	GK SPLIT 2	Split	43.510466	16.468661	Bar	Coffee Shop	Café	Restaurant	Supermarket	Park	Drugstore	Grocery
38	GK SUČIDAR	Split	43.512754	16.467775	Bar	Café	Grocery Store	Coffee Shop	Gym	Shopping Mall	Big Box Store	Elect
39	GK ŠINE	Split	43.507287	16.523302	Beach	Restaurant	Pizza Place	Italian Restaurant	Mediterranean Restaurant	Seafood Restaurant	Café	Campg
40	GK TRSTENIK	Split	43.507302	16.472329	Bar	Fast Food Restaurant	Restaurant	Grocery Store	Beach	Hotel	Cocktail Bar	
41	GK VAROŠ	Split	43.509172	16.435229	Mediterranean Restaurant	Hostel	Café	Pizza Place	Bar	Restaurant	Grocery Store	
42	GK VISOKA	Split	43.510468	16.474525	Fast Food Restaurant	Grocery Store	Bar	Restaurant	Bus Station	Shopping Mall	Café	Superr
43	GK ŽNJAN	Split	43.508112	16.484383	Grocery Store	Café	Hotel	Park	Restaurant	Pizza Place	Athletics & Sports	Mediterr Restı

Cluster 3

```
df_merged.loc[df_merged['Cluster Labels'] == 2, :]
```

	Neighborhood	Town	latitude	longitude	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	1
3	Maksimir	Zagreb	45.833312	16.010151	Restaurant	Grocery Store	Steakhouse	Bus Station	Café	Falafel Restaurant	Eye Doctor	Exhibit	Escape Room	E
20	GK BRDA	Split	43.523212	16.466332	Grocery Store	Big Box Store	Cafeteria	Mediterranean Restaurant	Café	Electronics Store	Bar	Women's Store	Farmers Market	R
24	GK KOCUNAR	Split	43.519219	16.465257	Café	Bar	Grocery Store	Big Box Store	Bus Station	Cafeteria	Electronics Store	Women's Store	Eye Doctor	R
31	GK NESLANOVAC	Split	43.521584	16.474297	Grocery Store	Bar	Electronics Store	Big Box Store	Pet Store	Shopping Mall	Bakery	Toy / Game Store	Fast Food Restaurant	
33	GK PUJANKE	Split	43.516156	16.473191	Grocery Store	Electronics Store	Irish Pub	Dance Studio	Big Box Store	Bus Station	Bar	Gym	Farmers Market	R
34	GK RAVNE NJIVE	Split	43.520962	16.464976	Big Box Store	Cafeteria	Café	Electronics Store	Bar	Grocery Store	Women's Store	Exhibit	Fast Food Restaurant	

Cluster 4

```
df_merged.loc[df_merged['Cluster Labels'] == 3, :]
```

	Neighborhood	Town	latitude	longitude	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	10
35	GK SIROBUJA	Split	43.505995	16.513706	Italian Restaurant	Pizza Place	Mediterranean Restaurant	Grocery Store	Women's Store	Eastern European Restaurant	Farmers Market	Falafel Restaurant	Eye Doctor	Exhi

Cluster 5

```
df_merged.loc[df_merged['Cluster Labels'] == 4, :]
```

	Neighborhood	Town	latitude	longitude	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
14	Podsljeme	Zagreb	45.888748	15.984549	Trail	Café	Women's Store	Farmers Market	Falafel Restaurant	Eye Doctor	Exhibit	Escape Room	Electronics Store	Eastern European Restaurant

Clusters '1' and '5' represent neighborhoods that are unique to Zagreb, with each cluster containing only one neighborhood. The same can be said for cluster '4', except that it contains a neighborhood of Split.

Neighborhoods in clusters '1' and '5' lie furthest on the north of Zagreb. In cluster '1', the most common venues fall in the category of 'Light Rail Station', and in cluster '5', in the

'Trail' category. 'Podsljeme' neighborhood, from cluster '5' is a part of Zagreb with family houses, green areas, but with not much social activities going on.

Neighborhood 'Sirobuja' from cluster '4' is on the outer ring of Split.

Cluster '3' represent mostly neighborhoods of Split, with the exception of one neighborhood of Zagreb, 'Maksimir'. Those neighborhoods of Split are part of an industrial area, so it's no wonder that some of the most common venues there are categorized as some kind of stores, etc.

It's obvious that there is a major overlapping of neighborhoods of Split and Zagreb in cluster '2'. Of all the neighborhoods in both towns, more than half of them belong to this cluster. In Split, we can see different types of neighborhoods as we go from the centar of the town (and its famous Roman palace that is usually crowded with tourists) further to north and east. Neighborhoods in cluster '2' contain cafés, restaurants, bar's, etc. In Split, those are areas with lots of tourists, and in Zagreb, those are neighborhoods where lots of people work and socialize.

5 Conclusion

Even though this kind of approach is highly sensitive to Foursquare venue categories, there are interesting insights that can be quickly obtained about similarities and dissimilarities of neighborhoods of Zagreb and Split that could potentially serve real estate traders. Since, for instance, venue category ‘restaurant’ and ‘mediteranian restaurant’ are completely different venue types, the project could be upgraded with finer comparison of the venues, and therefore, city neighborhoods.