1. INTRODUCTION

I always wanted to open a cafe bar by myself. The problem is that I don't want that my cafe bar is on a random place in a random city. I want a cafe bar in one of the biggest European cities, and not just a regular one placed in the middle of nowhere, I want it to be placed at the city center.

So, I did a research on the list of the most populated cities in Europe based on the number of similar contents that serve coffee in the center area of the city.

I would like to present my research.

2. IMPORTING AND REDEFINING THE DATASET

First step was importing all of the libraries needed. Then I imported the dataset I wanted from wikipedia. I used a data from the wikipedia page with list of the biggest European cities („https://en.wikipedia.org/wiki/List_of_European_cities_by_population_within_city_limits"), and when I did a little arrangement it looked like this:

	Rank	City	Country	Population
0	1	Istanbul[a]	Turkey	15,519,267[2]
1	2	Moscow[b]	Russia	12615279
2	3	London	United Kingdom	9126366
3	4	Saint Petersburg	Russia	5383890
4	5	Berlin	Germany	3748148
5	6	Madrid	Spain	3223334
6	7	Kiev	Ukraine	2950800
7	8	Rome	Italy	2844750
8	9	Paris	France	2140526
9	10	Bucharest	Romania	2106144
10	11	Minsk	Belarus	1982444
11	12	Hamburg	Germany	1930996
12	13	Vienna	Austria	1899055
13	14	Warsaw	Poland	1802237
14	15	Budapest	Hungary	1768073
15	16	Barcelona	Spain	1636762
16	18	Munich	Germany	1471508
17	17	Kharkiv	Ukraine	1451132
18	19	Milan	Italy	1389834
19	20	Prague	Czech Republic	1308632
20	21	Nizhny Novgorod	Russia	1259013
21	22	Kazan	Russia	1243500
22	23	Sofia	Bulgaria	1238438
23	24	Birmingham	United Kingdom	1223000
24	25	Brussels	Belgium	1211035
25	27	Samara	Russia	1170910
26	28	Belgrade	Serbia	1166763
27	29	Ufa	Russia	1121429
28	30	Rostov-on-Don	Russia	1119875
29	31	Cologne	Germany	1085664
30	32	TekirdaÄ□	Turkey	1055412
31	33	Voronezh	Russia	1054537
32	34	Perm	Russia	1051583
33	35	Volgograd	Russia	1013533
34	36	Odessa	Ukraine	1011494

3. GETTING LOCATION DATA FOR THE CITIES

If you noted that there's no city ranked on no. 26 (not the 26th index value), it's a mistake by the Wikipedia database, I didn't just clear it away.

The second thing I did was discovering the coordinates of each city from our list. I did it using the geographical libraries from python. After writing some codes and arranging the dataframe we got this (you'll see only the first ten rows).

	Rank	City	Country	Population	latitude	longitude
0	1	Istanbul	Turkey	15519267	41.009633	28.965165
1	2	Moscow	Russia	12615279	55.479205	37.327330
2	3	London	United Kingdom	9126366	51.507322	-0.127647
3	4	Saint Petersburg	Russia	5383890	59.960674	30.158655
4	5	Berlin	Germany	3748148	52.517037	13.388860
5	6	Madrid	Spain	3223334	40.416705	-3.703582
6	7	Kiev	Ukraine	2950800	50.450034	30.524136
7	8	Rome	Italy	2844750	41.893320	12.482932
8	9	Paris	France	2140526	48.856697	2.351462
9	10	Bucharest	Romania	2106144	44.436141	26.102720

4. GETTING THE DATA FOR ALL THE CONTENTS BASED IN THE CENTRE OF EACH CITY

Now that we have a dataframe filled with all the things about the cities that we need for our project, the next thing is to get the best data about the contents in the city center. I did it using the Foursquare API from which I got data about all the contents that are based strictly in the center of each city from our list.

Here you can see a sample of the list containig the information about the city, the name of the content, its coordinates and a category to which it belongs.

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Istanbul	41.009633	28.965165	Bayezid Camii	41.009736	28.964500	Mosque
1	Istanbul	41.009633	28.965165	Day Day Pastanesi	41.009440	28.968023	Bakery
2	Istanbul	41.009633	28.965165	Kapalıçarşı	41.009725	28.968315	Historic Site
3	Istanbul	41.009633	28.965165	Nuruosmaniye Camii	41.010151	28.970701	Mosque
4	Istanbul	41.009633	28.965165	Cevahir Bedesteni (Oldbazaar)	41.010894	28.969446	Antique Shop
...	...	...	...	...	...	...	...
3061	Odessa	46.487319	30.739278	Детская площадка у "Немо"	46.477106	30.764376	Playground
3062	Odessa	46.487319	30.739278	Wine Story - Винфорт	46.463149	30.741597	Wine Shop
3063	Odessa	46.487319	30.739278	Make My Cake Cafe	46.463345	30.754717	Dessert Shop
3064	Odessa	46.487319	30.739278	Дельфинарий Немо / Nemo Dolphinarium (Дельфіна...	46.476975	30.765269	Aquarium
3065	Odessa	46.487319	30.739278	Одесса, Пляж Аркадия, Центральная Аллея	46.473283	30.764131	Beach

5. LETS CONCLUDE THE FIRST PART OF THE PROJECT (WEEK 1)

The data is now ready and from now on it's just about coding. I clustered cities in two groups based on the contents that are the most common there. Further proccesing was made on the group that had a small number of objects that serve coffee in the center of the city. Based on the processing, I determined which of the biggest European cities has the smallest amount of similar kind objects and defined which city is the best to open a new cafe bar in its city center. See you in the second part.