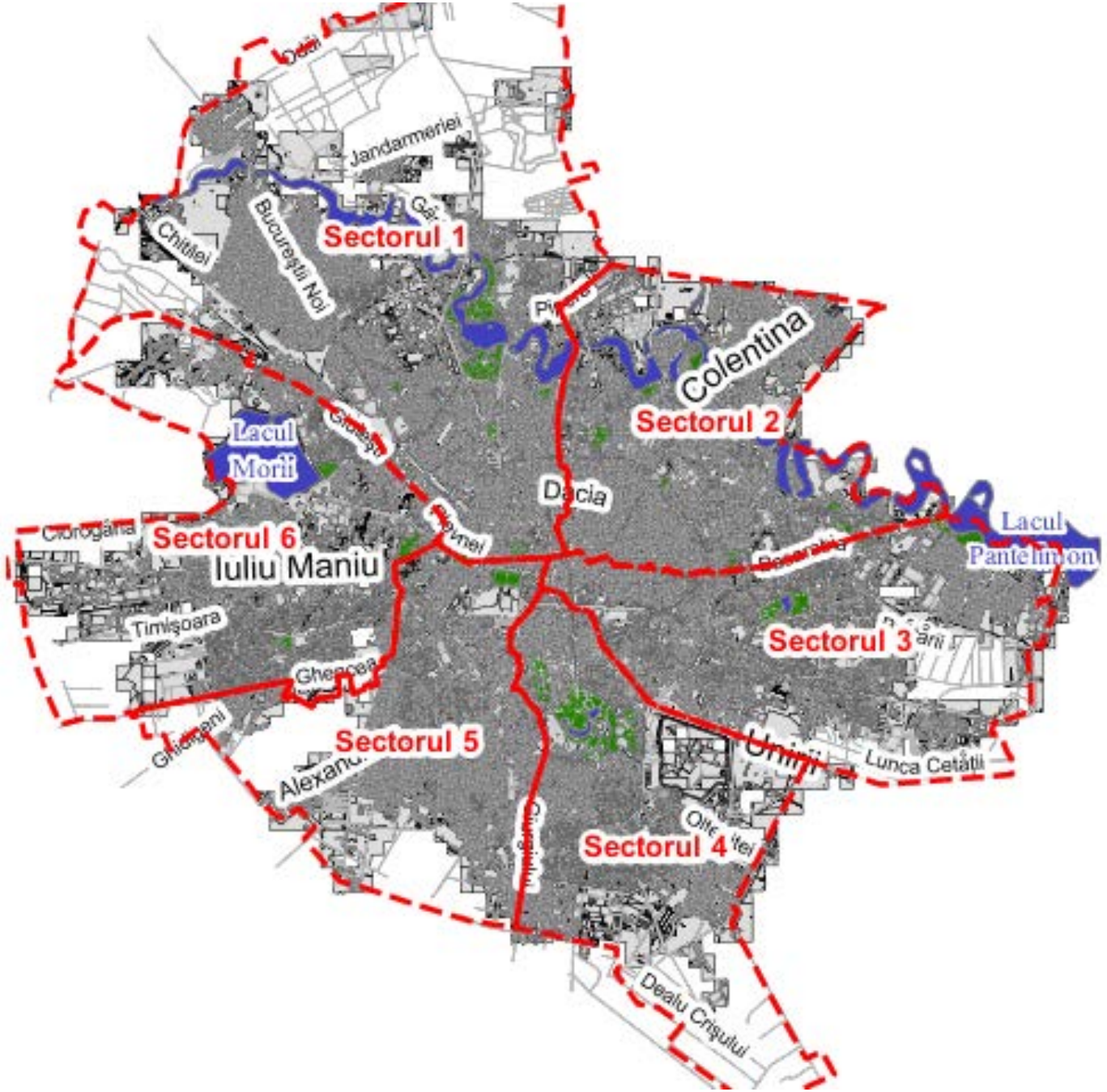
CREATING SMARTER CITY DISTRICTS IN BUCHAREST

Coursera Peer Graded Assignment - Capstone Project

Claudia Hrabac

INTRODUCTION

Background and problem



Bucharest's current administrative units

Bucharest, the capital of Romania is currently divided into **6 administrative** units, called sectors ("sectoare" in Romanian).

Bucharest's current administrative units

- The current divisions of sectors of this city date back to **august 1979!**

- There is an incentive to redefine they way Bucharest is divided, as ***the territories encompass diverse neighborhoods*** which translate into **diverse needs hard to tackle by the local administrations**.

The problem with redefining these units

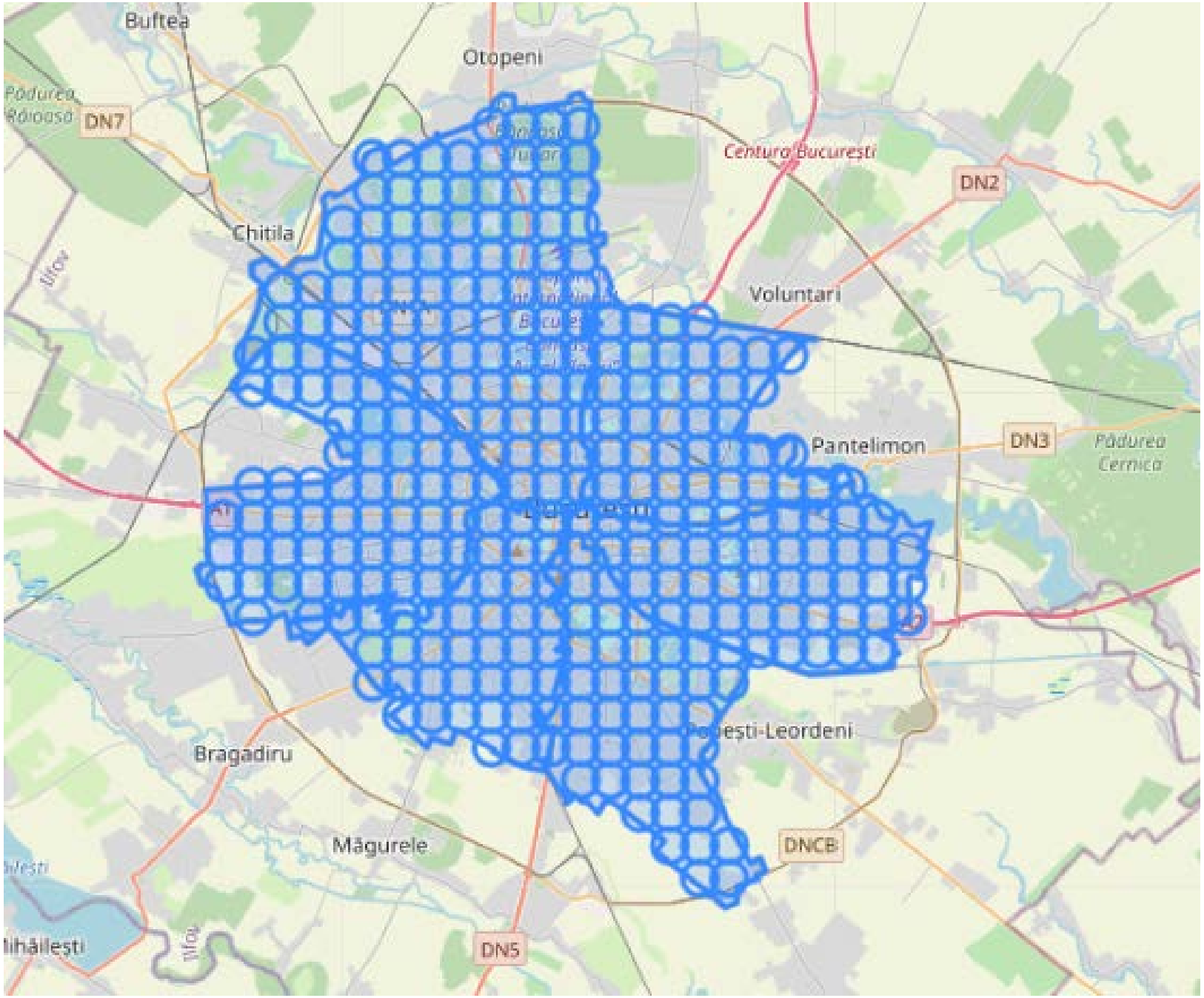
- Dividing a territory into coherent divisions demands **taking into account a large number of factors** regarding what venues are present in the district, what kinds of services are operating and at what level of quality, etc.
- My goal is to use **machine learning algorithms** to try to divide the territory of Bucharest in a more efficient way, with neighborhoods defined by the types of restaurants, parks, museums present.

DATA

Presenting the data used in this project

Foursquare data extraction strategy

- Foursquare venue data will be leveraged in order to help build new clusters of neighborhoods which could potentially replace the present administrative divisions of Bucharest.



Foursquare data extraction strategy

First step is to define the city's limits and select data for venues inside the city.

We will use a grid of small neighborhoods within the city in order to make the exploration calls for venues.

METHODOLOGY

Pre-processing, PCA, Clustering Algorithms

Data Extraction

- After extracting venue data through Foursquare's API, the dataset resulted from this process included data for each venue found, each line representing a venue.

	Neighborhood index	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	43	44.341705	26.153622	Baza Steaua Bucuresti(Berceni)	44.341298	26.153184	Soccer Field
1	90	44.358354	26.121788	Jumbo	44.361887	26.124094	Toy / Game Store
2	90	44.358354	26.121788	Auchan	44.361260	26.122536	Department Store
3	90	44.358354	26.121788	Hasco Fashion	44.360040	26.123019	Shopping Mall
4	90	44.358354	26.121788	Orange store	44.361326	26.122492	Electronics Store

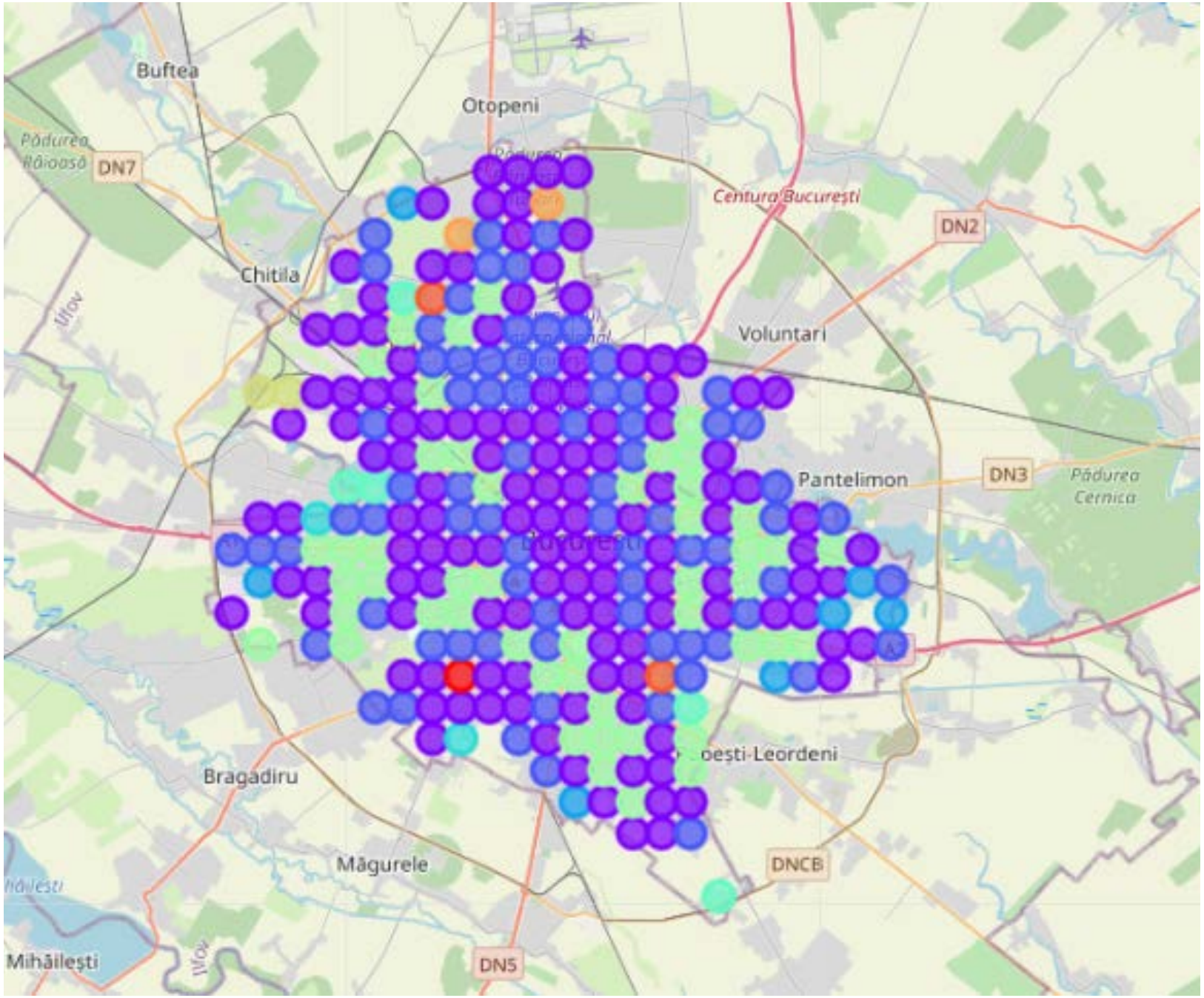
Data processing

- In order to use DBSCAN clustering algorithm, I have decided to add the geographical coordinates of each neighborhood in the analysis, and to reduce the dimensions of existing variables.
- I have decided to reduce the 302 columns representing venues to 10 principal components, using PCA Analysis.

	principal component 1	principal component 2	principal component 3	principal component 4	principal component 5	principal component 6	principal component 7	principal component 8	principal component 9	principal component 10	LAT	LON
0	-1.188686	0.222272	0.185851	-0.477468	0.412854	0.132388	0.396173	0.589247	-0.067252	0.210014	44.366679	26.100566
1	-1.212115	0.334354	0.317480	-0.415647	0.424840	0.363880	0.297824	0.678941	-0.143923	0.116646	44.366679	26.111177
2	-1.254399	0.131917	0.490719	-0.490620	0.549470	-0.021312	0.510665	0.616766	-0.012034	-0.347121	44.366679	26.121788
3	-1.183395	0.255866	0.326575	-0.437731	0.412390	0.226118	0.367707	0.627006	-0.063849	0.094994	44.366679	26.132399
4	-0.290958	0.028854	0.392725	0.139747	-0.540154	-0.335909	1.080192	0.728426	-0.612113	-0.459225	44.366679	26.143010
...	...	...	...	...	...	...	...	...	...	...	...	...
267	-1.174844	0.238706	0.266944	-0.387348	0.312172	0.204702	0.593901	0.661348	-0.163267	0.176935	44.533175	26.089954
268	-1.426632	0.208617	0.485592	-0.605706	0.623290	0.118338	0.687633	0.776213	-0.630681	0.329806	44.533175	26.100566
269	-1.078632	0.433739	1.240629	-0.869439	0.305050	0.116106	0.918119	1.015051	-0.221664	-0.119576	44.358354	26.121788
270	-1.481491	0.310229	0.788381	-0.605131	0.564863	0.108351	0.577781	0.811739	-0.276436	0.040928	44.358354	26.132399
271	-1.288875	0.272317	0.468213	-0.490964	0.428280	0.305833	0.359454	0.748325	-0.119022	0.052103	44.358354	26.143010

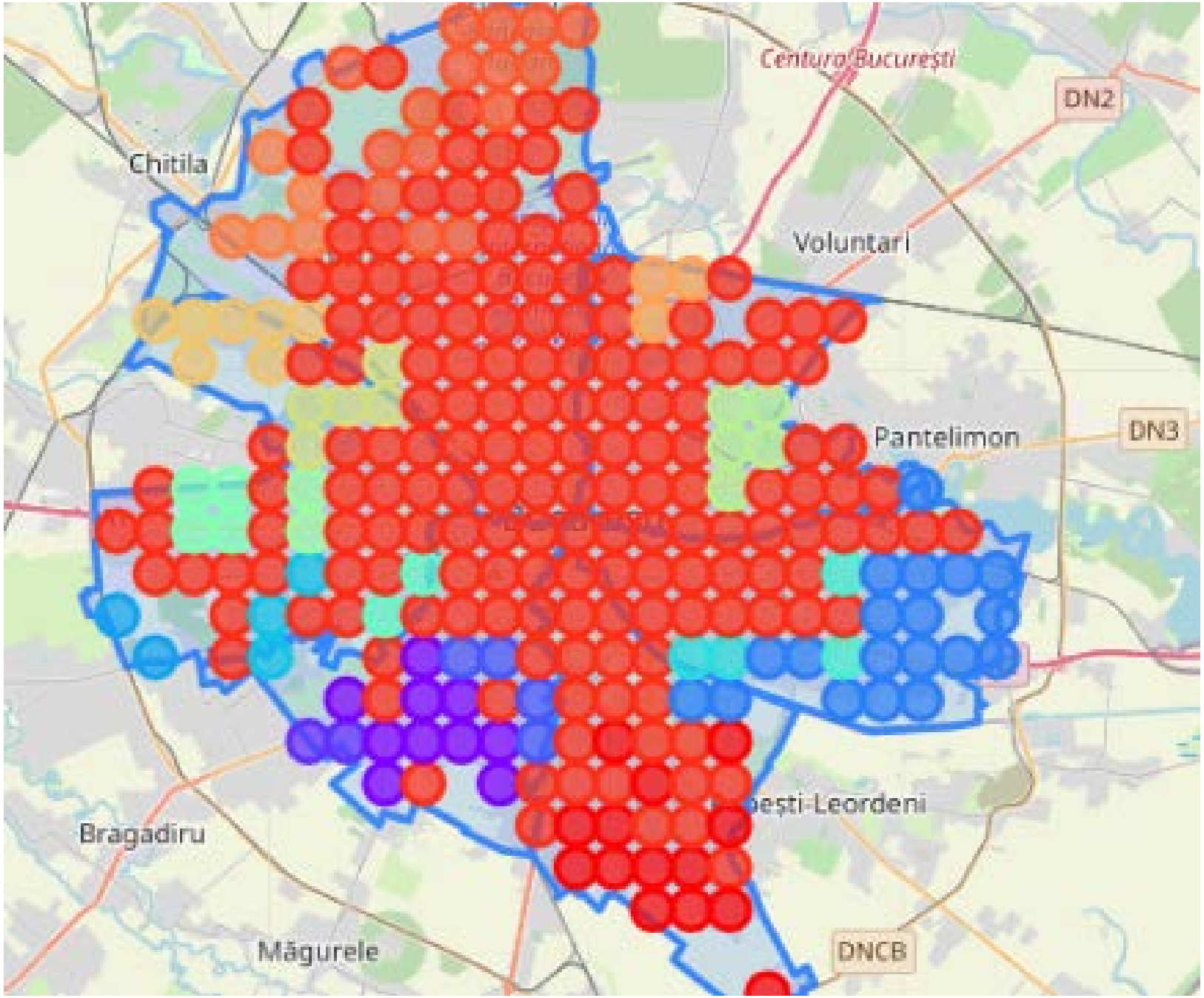
RESULTS

KMeans, DBSCAN



Kmeans results

As seen here, we need to take into account the latitudes and longitudes of each neighborhood and give them a higher weight in the analysis.



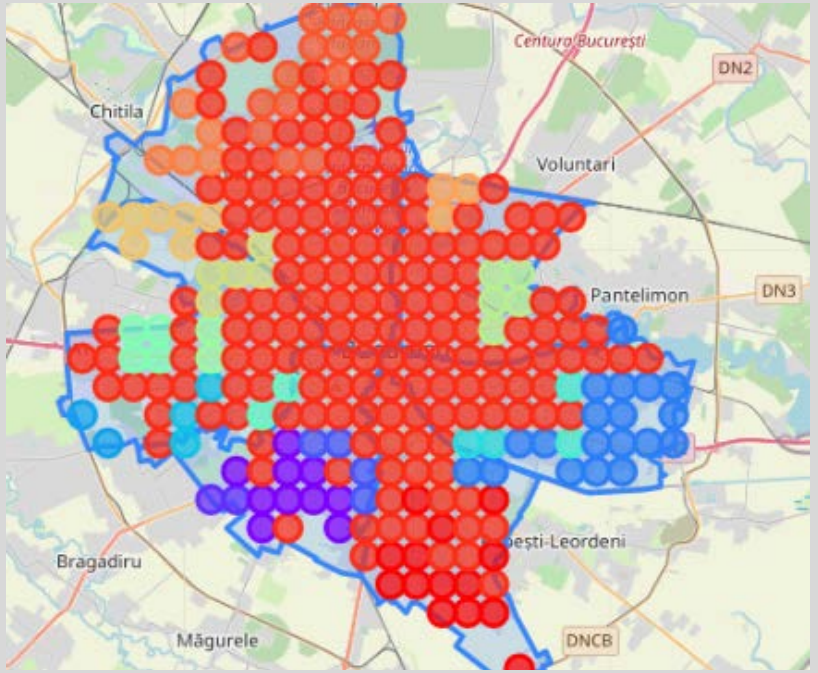
DBSCAN

A total number of clusters of 19, distributed in such a way that they can arguably become “new neighborhoods”.

DISCUSSION AND CONCLUSION

Discussion and conclusion

As one could expect, the newer areas where Bucharest has expanded over the last years, have a distinct composition of local venues, which in turn will be reflected in the needs of their residents.



Discussion and conclusion

- Local administrations are struggling every year to tackle problems in these big areas, and to balance their budget in a way that is fair for every part of the sector.
- This problem could be greatly improved by creating **smaller districts**, **centered around neighborhoods which present similar problems**.
- For future projects, there are lots of additional data which can be added in order to create smarter city districts, and more efficient local administrations.