OPENING A NEW COFFEE HOUSE IN CLUJ-NAPOCA

Applied Data Science Capstone

Dobrei Gabriel

Introduction: Business Problem

- find an optimal location to open a Coffee House/Place in central part **Cluj-Napoca**, **Romania**.

- there are lots of Coffee Houses in Cluj-Napoca

- detect **locations that are not already crowded with Coffee Houses**.

- KPI: **as close to city center as possible**.

- KPI: density of coffee houses

Data

Factors that will influence our decision are:

- number of existing Coffee Houses/places in the neighborhood

- number of and distance to Coffee Houses in the neighborhood, if any

- distance from city center

- Cluj-Napoca city center information was retrieved with geo locator at: 46.769379, 23.5899542

Methodology

- the environment - Jupyter Notebook,

- main libraries *beautifulsoup4, lxml, geopy, folium*, *Pandas, Numpy*,

- Data was retrived via Foursquare geolocator

- 4 sets of data were created:

  - *dataframe_filtered_red — 500 m radius from city center*

  - *dataframe_filtered_yellow — 750 m radius from city center*

  - *dataframe_filtered_green — 1250 m radius from city center*

  - *dataframe_filtered_test — 5000 m radius from city center*

Results – distribution of Coffe Places

- Distribution of Coffee Places in test zone – 5000 m radius from city centre
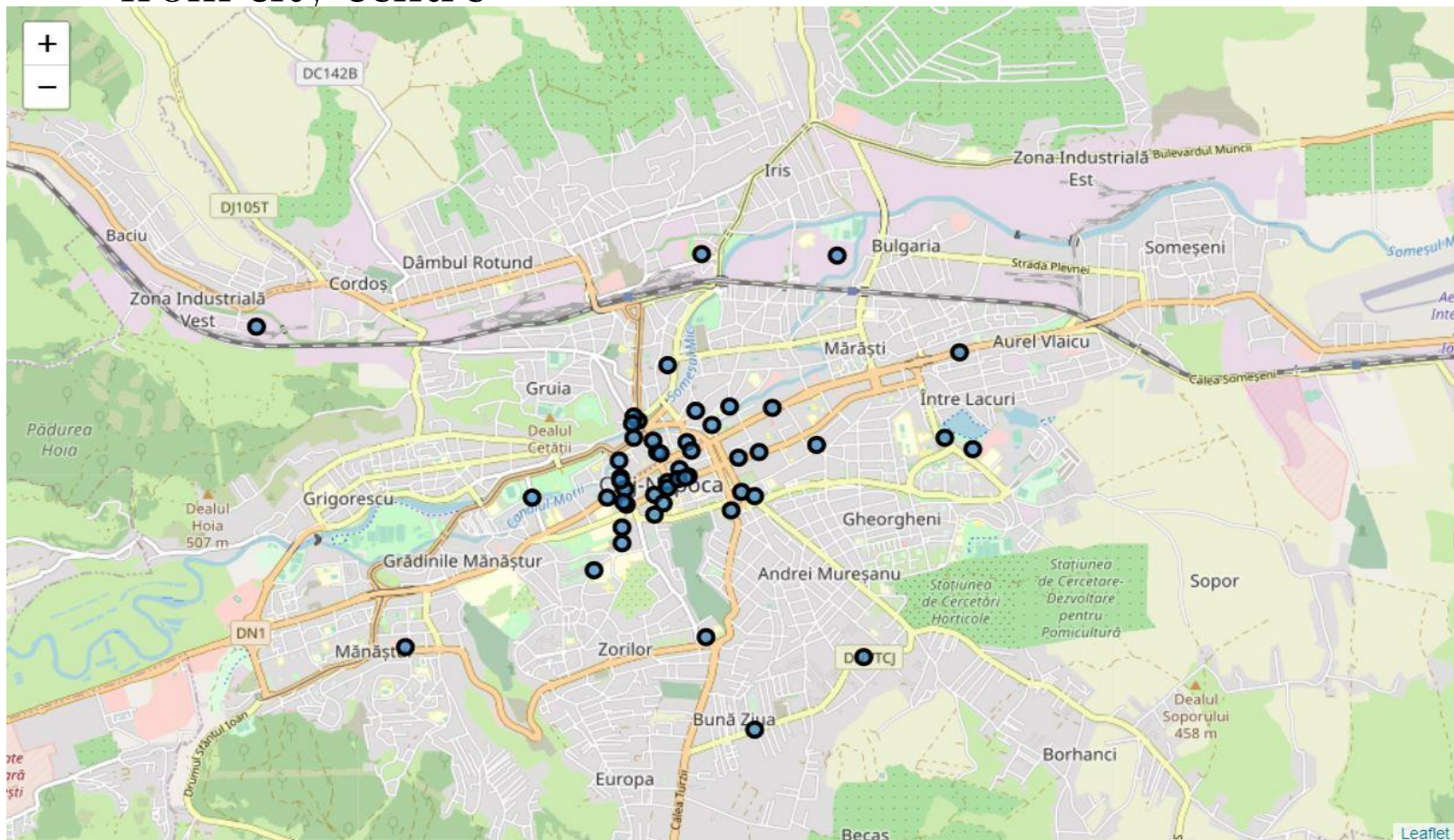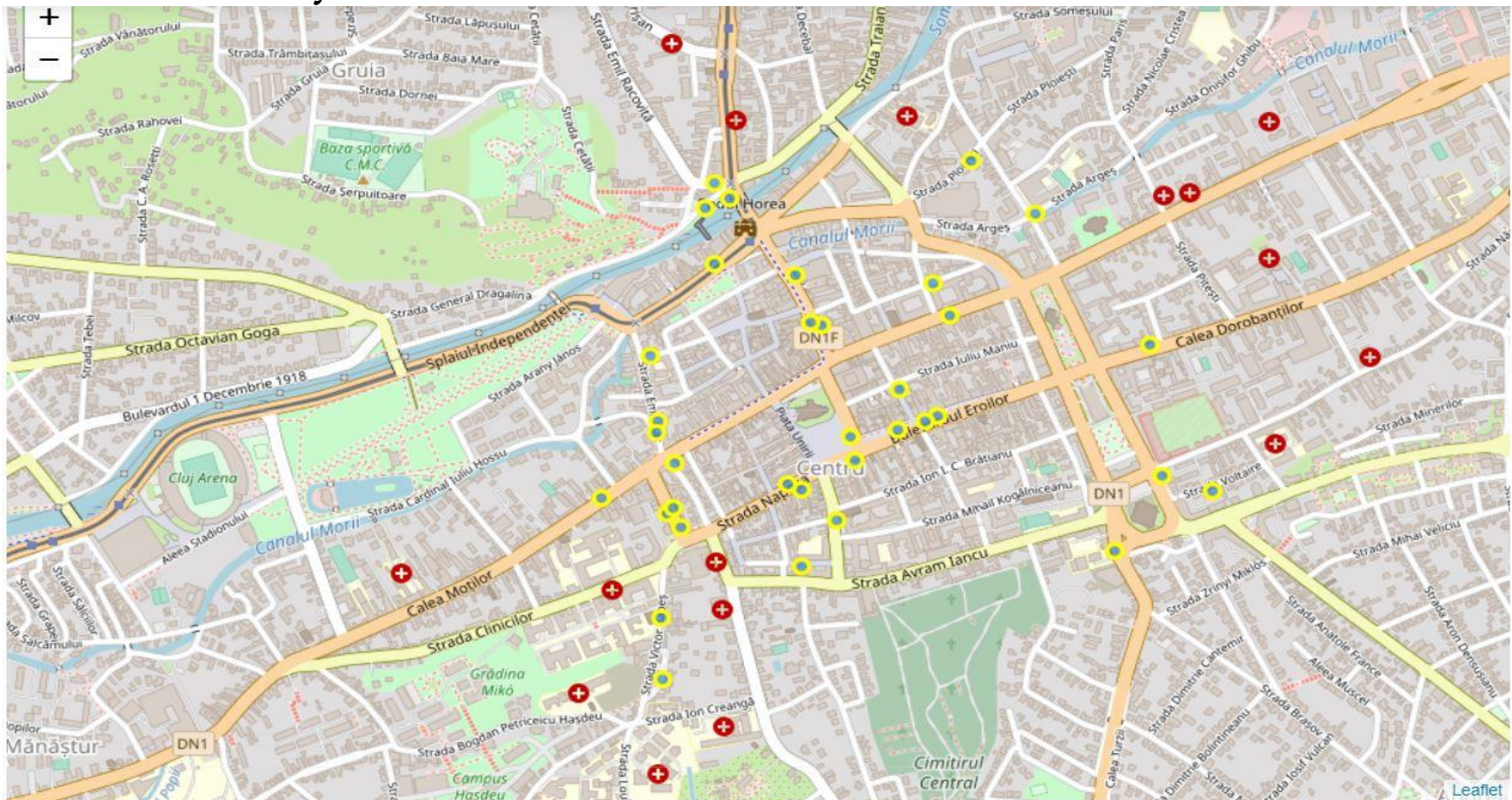
Results – distribution of Coffe Places

- Distribution of Coffee Places in red zone – 500 m radius from city centre
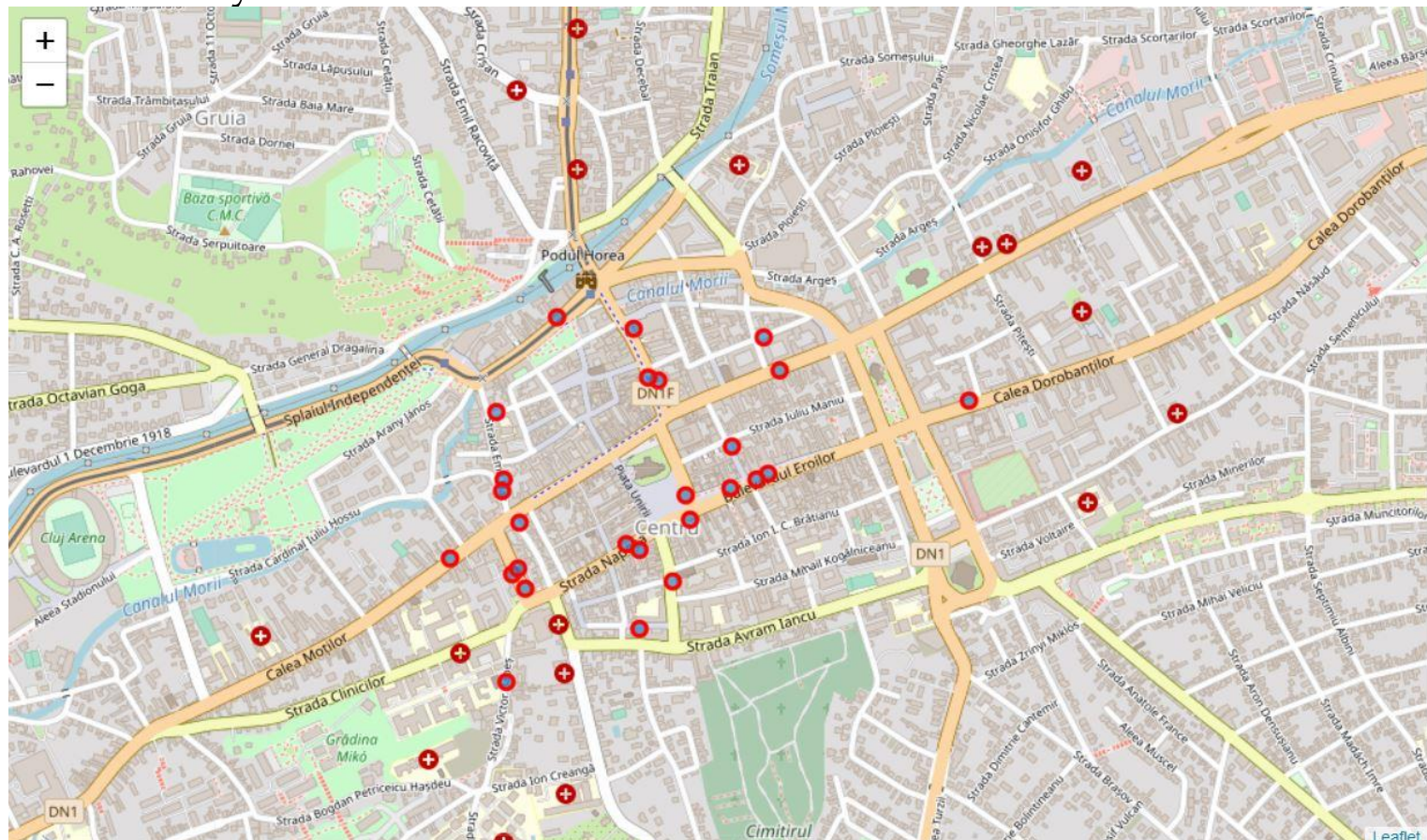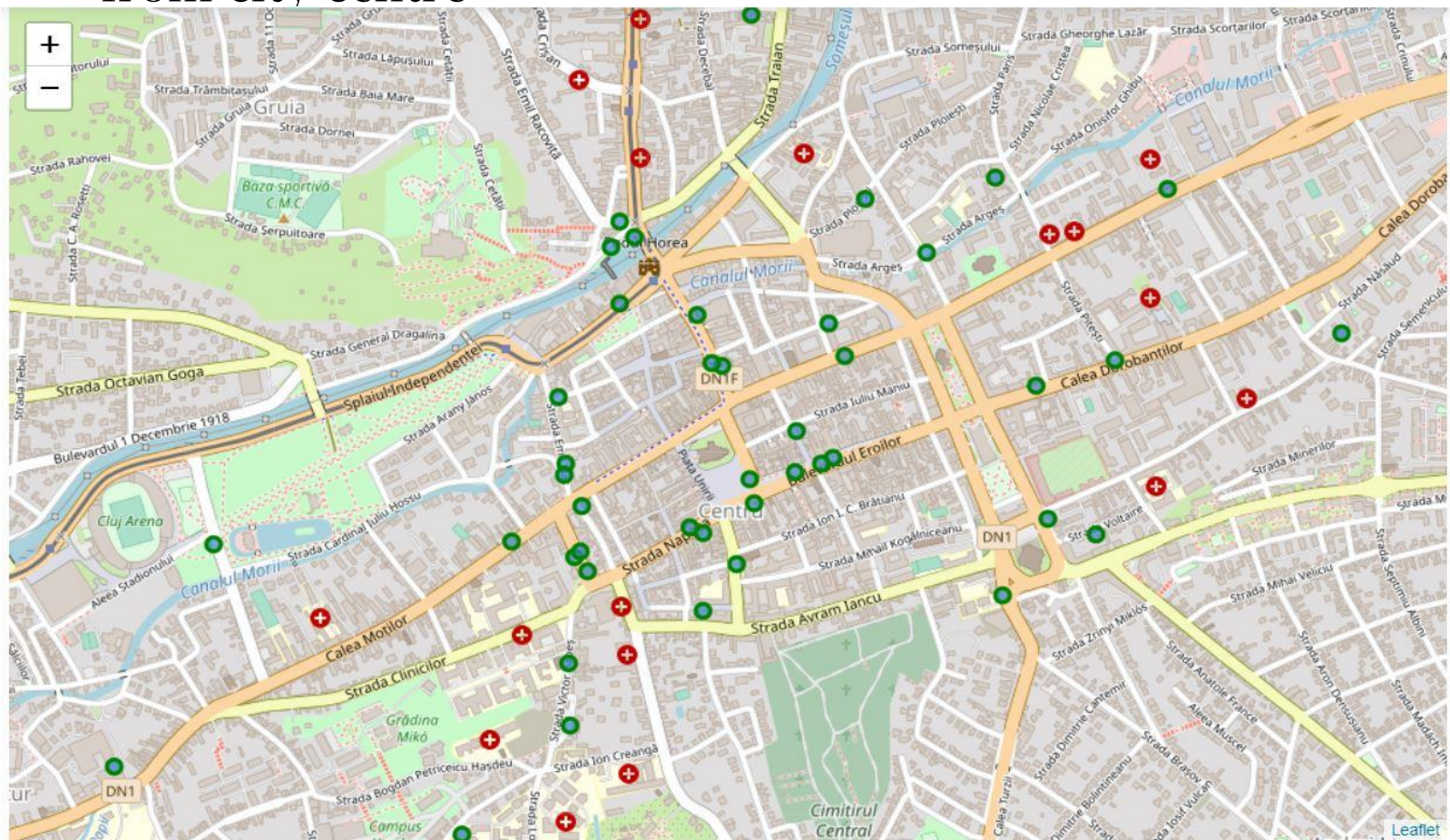
Results – distribution of Coffe Places

- Distribution of Coffee Places in yellow zone – 750 m radius from city centre

Results – distribution of Coffe Places

- Distribution of Coffee Places in yellow zone – 1250 m radius from city centre

Results – subcategories of venues

Counting the unique values in our data frames on the column Category

```
In [58]:   1  print(dataframe_filtered_test['categories'].value_counts())
```

```
Coffee Shop          32
Café                  9
Bar                   3
Pub                   1
College Rec Center    1
Cocktail Bar          1
Beach Bar             1
Casino                1
Chocolate Shop        1
Name: categories, dtype: int64
```

```
In [54]:   1  print(dataframe_filtered_green['categories'].value_counts())
```

```
Coffee Shop          29
Café                  7
Bar                   2
Casino                1
College Rec Center    1
Cocktail Bar          1
Beach Bar             1
Chocolate Shop        1
Name: categories, dtype: int64
```

```
In [55]:   1  print(dataframe_filtered_yellow['categories'].value_counts())
```

```
Coffee Shop          25
Café                  7
College Rec Center    1
Cocktail Bar          1
Bar                   1
Name: categories, dtype: int64
```

```
In [59]:   1  print(dataframe_filtered_red['categories'].value_counts())
```

```
Coffee Shop     17
Café             7
Cocktail Bar     1
Bar              1
Name: categories, dtype: int64
```

Results – Specific density of venues

Tier		Area (skm)	all venues	Coffee Shop	Cafe	Bar	Chocolate Shop	Collage Rec Center	Pub	Beach Bar	Cocktail Bar	Casino
test	number	78.5	50	32	9	3	1	1	1	1	1	1
	density		0.63	0.40	0.11	0.03	0.01	0.01	0.01	0.01	0.01	0.01
green	number	4.2	43	29	7	2	1	1	-	1	1	1
	density		10.23	6.9	1.66	0.47	0.23	0.23	-	0.23	0.23	0.23
yellow	number	1.75	35	25	7	1	-	1	-	-	1	-
	density		20	14.28	4	0.57	-	0.57	-	-	0.57	-
red	number	0.8	25	17	7	1	-	-	-	-	1	-
	density		31.25	21.25	8.75	1.25	-	-	-	-	1.25	-

Discussions

- Foursquare limitations - In Romania and subsequently in Cluj-Napoca Foursquare usage and accuracy is a problem

- Cluj –Napoca has medium coverage of venues but seriously lack ratings and review

- Other Limitations include:

- accuracy: we have seen many incomplete rows, some missing address, and missing city while we even had some having been pointed in the wrong places.

- we do not have any venues showing in distant parts of the city.

Recomandations

- avoid opening any type of venue related to Coffee House in the red zone of Cluj-Napoca due to very high density in the area

- if a good opportunity presents itself we recommend to focus on opening a Café in the yellow zone.

- Coffee Shop density in yellow area is still high enough to make investment risky due to high competition,

- in the green area we recommend any type of Coffee House since the overall density is low.

- we recommend opening a Café in the green area , the competition in the sector is very low 1.6/square km.