How to find the best area to place a new venue in Copenhagen using Python.

Example: Cafe

The problem

If you want to open a new cafe in Copenhagen, how do you find the best place, considering:

- How many café's are in the area
- How many people live in the area

The solution

Get

Get Venue data

- Make the user input venue type, in this example cafe
- Get administrative units roder of Copenhagen, with coordinates in csv.
- Make function to iterate over rode coordinates, and pass the coordinates to the foursquare api

Cluster

Cluster data

Use k-means algorithm to cluster the results, using the elbow method to define best number of k's

Plot

Use folium

Plot clusters in map with borough limits

Plot clusters in in folium and add population numbers from roder

The data

For population stats

Get csv file with administrative units (roder) and corresponding population numbers from the municipality of Copenhagen's open databank

For venues data

Use Foursquare API to get results

For mapping and population data

From the municipality of Copenhagen's open databank get the following

- Geojson with borough limits
- Geojson with administrative units (roder) limits

Getting coordinates and population stats

- First the csv file is loaded to a DataFrame, and cleaned so that it only contains rode nr, and population stats

- Merge the above DataFrame with a geojson containing roder

- Create DataFrame with first coordinates extracted from polygon in geojson

	rode_nr	Personer
0	1	549
1	2	698
2	3	385
3	4	295

	rode_nr	geometry	Personer
0	17	(POLYGON ((12.5826025317837 55.67674174792494,...	0.0
1	107	(POLYGON ((12.56159677231597 55.69635404179332...	303.0
2	164	(POLYGON ((12.54193932257353 55.70658806248724...	4586.0
3	2	(POLYGON ((12.56625392688763 55.67856006861916...	698.0
4	3	(POLYGON ((12.5684420463914 55.67670315220757,...	385.0

	lat	lng
0	55.676742	12.582603
1	55.696354	12.561597
2	55.706588	12.541939
3	55.678560	12.566254
4	55.676703	12.568442

Check if coordinates covers Copenhagen

By plotting coordinates from DataFrame in folium

Obtain venues data from Foursquare API

- Run the coordinates trhu function 'getNear' calling the Foursqaure API

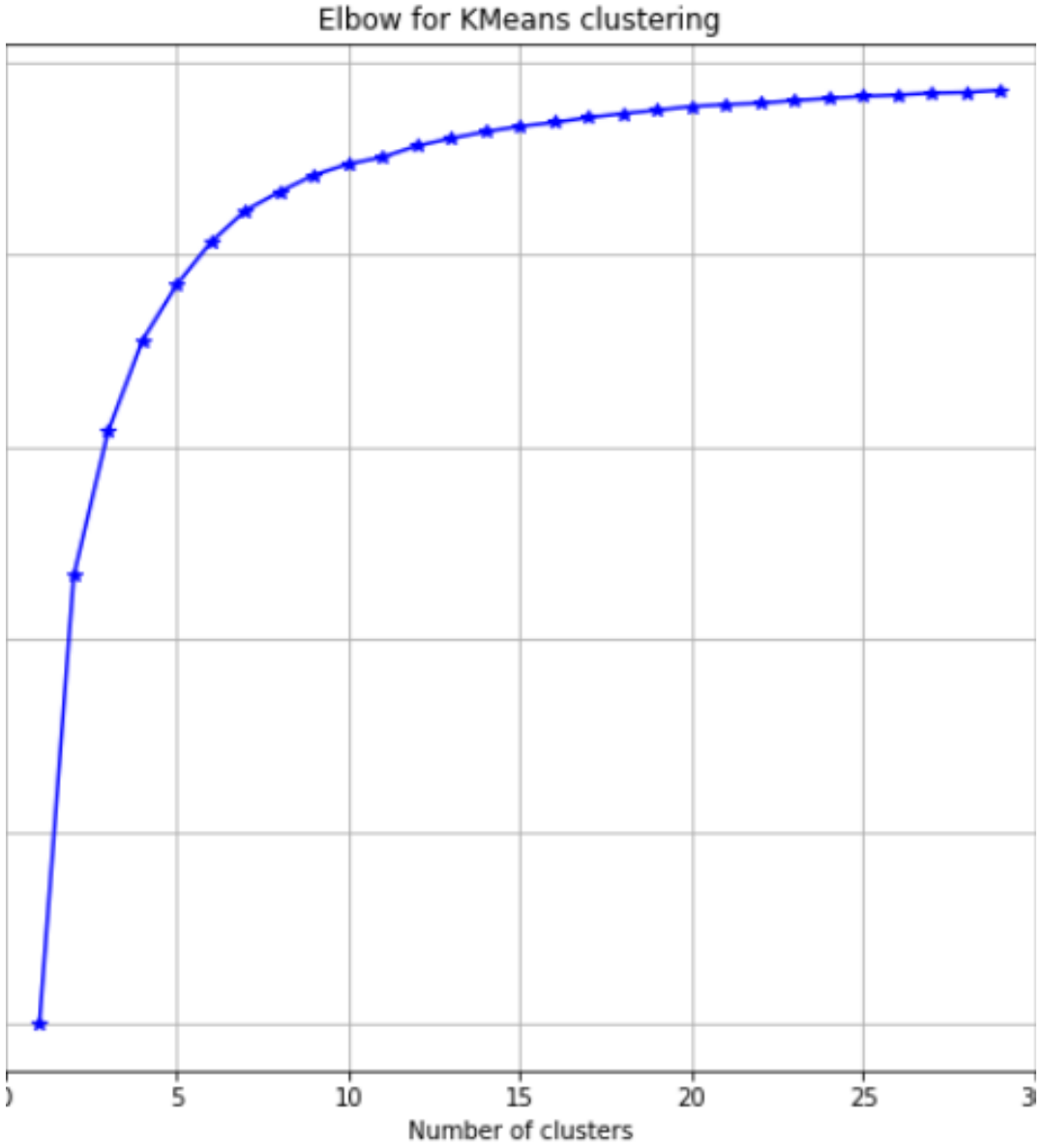
- Remove duplicates

	13Z Café	55.664292	12.622599
	13Z Café	55.664292	12.622599
13	13Z Café	55.664292	12.622599

	Name	lat	lng
	13Z Café	55.664292	12.622599
	3'erens IsCafe	55.654568	12.649390
	Aliva Foods Concept Store & Cafe	55.703021	12.584680
	Allehånde Café	55.647969	12.649141
4	Almasa Cafe	55.661127	12.604182

Cluster data via KMeans

- Make and run function 'plot_elbow' to determine best number of k's
- Plot function
- Determine best number of k's

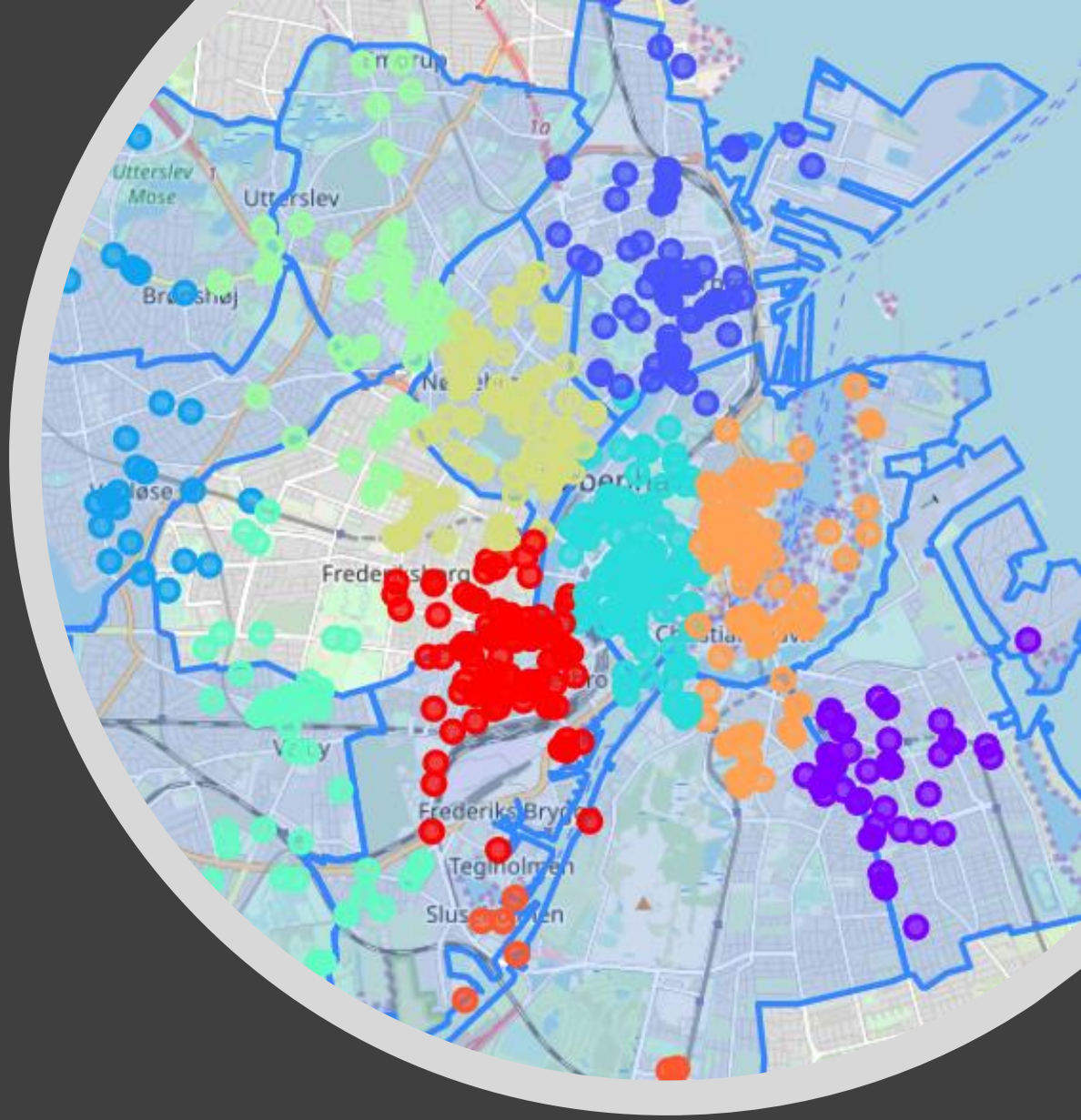


Pot the results on Copenhagen map with borough limits

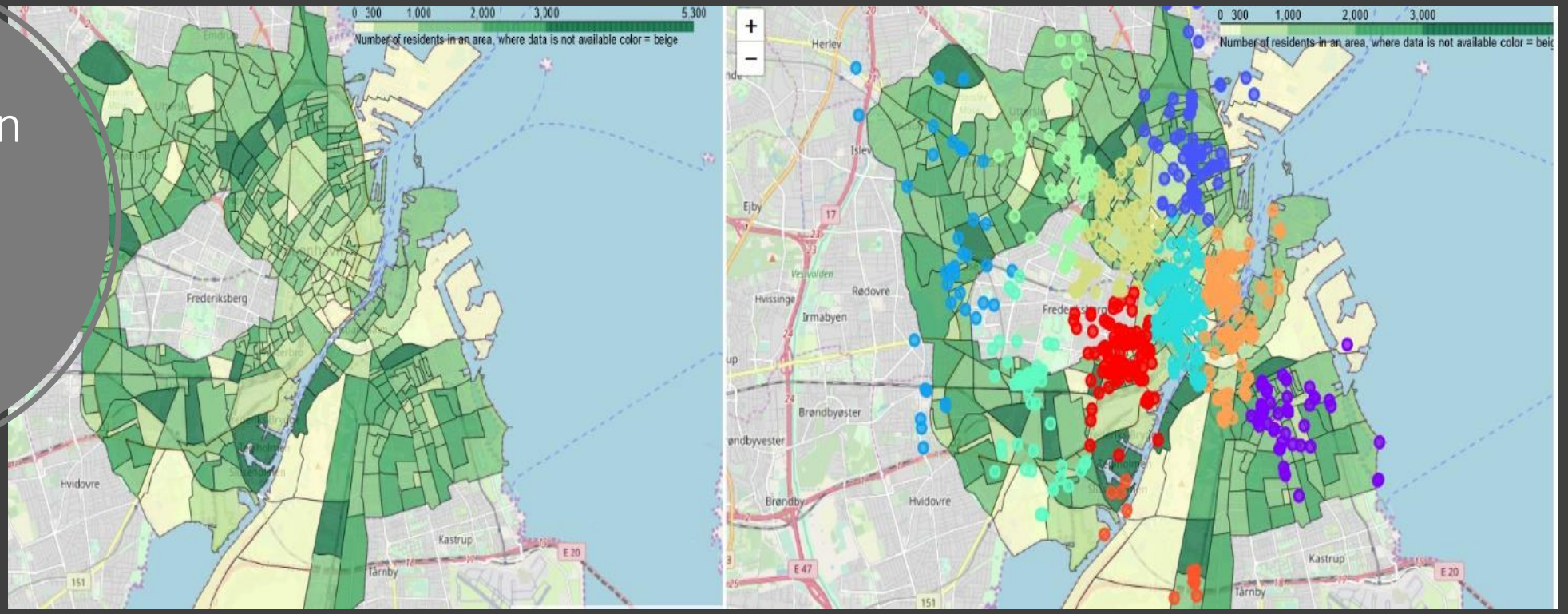
10 clusters

Concentrated at the center

Most cafés are situated in the most popular boroughs



Plot cafe’s on map with population stats



- Most cafe's are not necessary in areas with most inhabitants
- In the outskirts of the city there are areas with small numbers of cafes, but high population stats
- Shows possibilities for good places to open a new café

Comments and recommendations

Main findings

- High competition in inner city boroughs, but these boroughs does not contain the highest number of inhabitants
- Some outer city areas contain good population stats, and low rates of competition

Recommendations

- Check out the area, consider the allure of the area, for example parks, educational facilities (lot of customers) , traffic flow etc.