Finding the best location for opening a new restaurant

in Cleveland, Ohio, USA

Background

- two important aspects when looking for the location for a new restaurant:

  1. How many people live nearby the restaurant?
  2. How many other restaurants are already located nearby?

- Potential restaurant owners have huge interest as the location has a huge influence on how successful the restaurant will be (!)

Data Acquisition

▶ ZIP Codes of Cleveland: https://zipcode.org/city/OH/CLEVELAND

▶ Geo coordinates for ZIP Codes: https://gist.github.com/erichurst/7882666

▶ Population for ZIP Codes: https://www.zipdatamaps.com/zipcodes-cleveland-oh

▶ Venues/Restaurants for ZIP Codes: https://api.foursquare.com (Foursquare API)

Data Cleaning

▶ scraped websites and merged the information into a single Dataframe

▶ ZIP Codes that were not assigned coordinates were dropped

▶ ZIP Codes that were not assigned population number were dropped

▶ when no number of restaurants was assgined to a ZIP Code, the value was set to zero

▶ Population and Number of Restaurants was normalized using min/max-method

▶ Resulting DataFrame:

|	Zip Code	Latitude	Longitude	Population	Population_normalized	Restaurant	Restaurant_norm
0	44102	41.479174	-81.740603	45014	1.000000	12.0	0.444444
1	44103	41.519415	-81.642123	18123	0.324160	9.0	0.333333
2	44104	41.482230	-81.626784	22640	0.437684	3.0	0.111111
3	44105	41.449476	-81.630289	40089	0.876222	6.0	0.222222
4	44106	41.505341	-81.605432	26896	0.544648	23.0	0.851852

Data Analysis (Step 1)

▶ used Folium to visualize the geographical distribution of ZIP Codes:
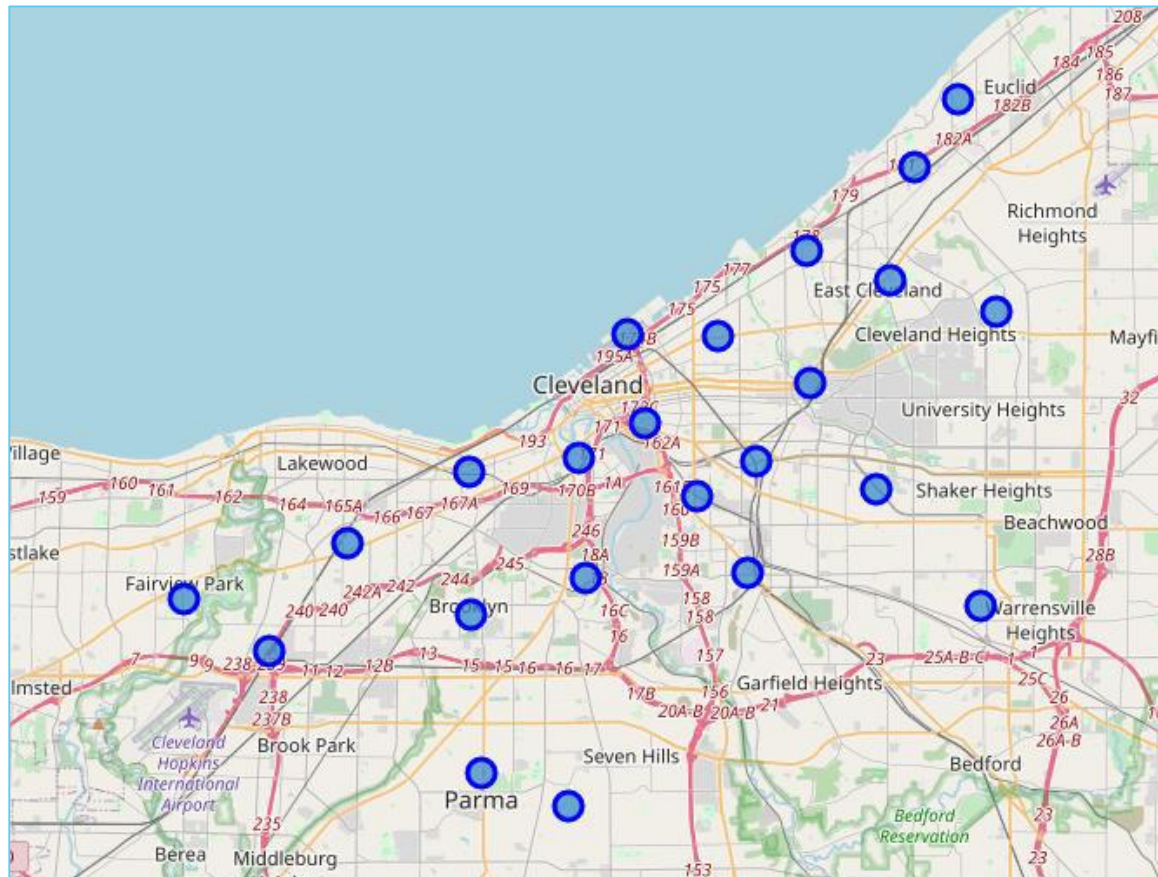
Data Analysis (Step 2)

▶ added population to map: the higher the circle's radius, the higher the population in the according ZIP Code
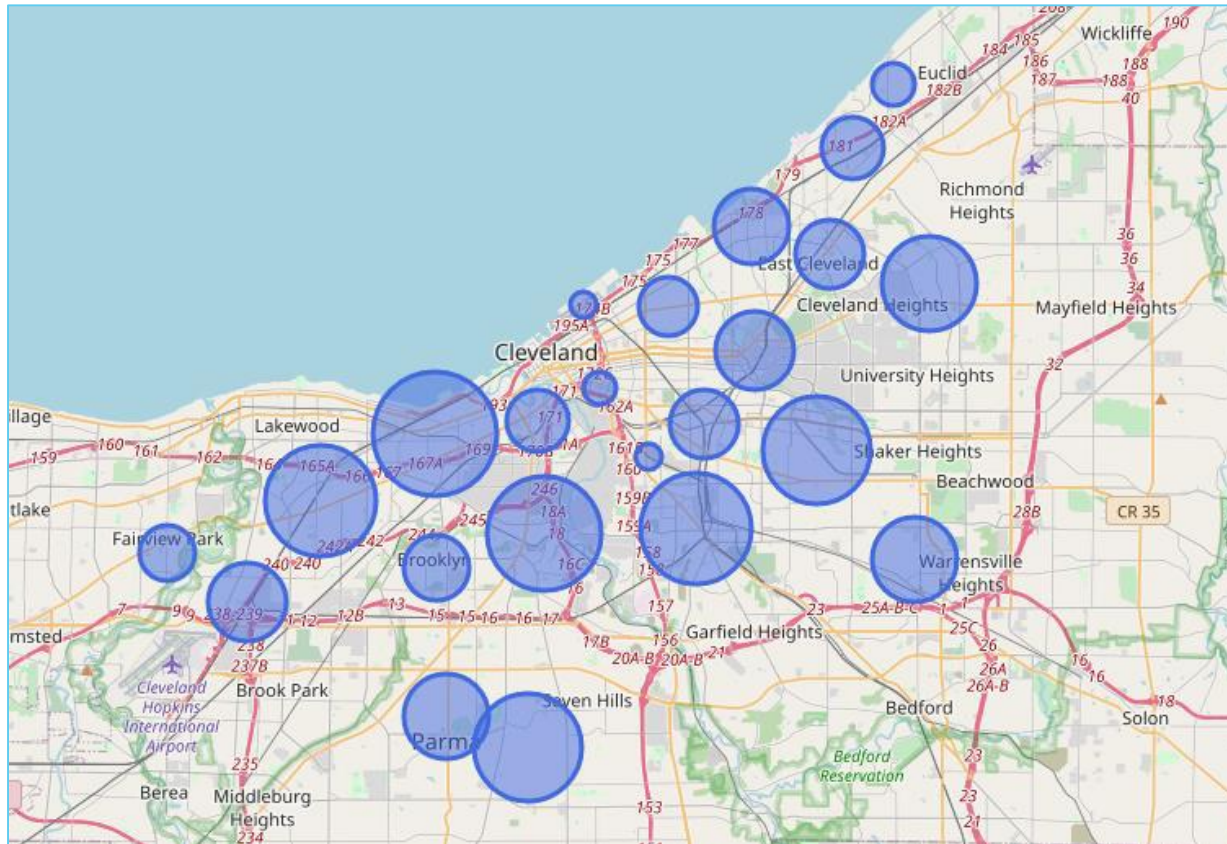
Data Analysis (Step 3)

▶ added the restaurant density to map: the higher the fill opacity of a circle, the higher the amount of restaurants nearby the ZIP Code it represents
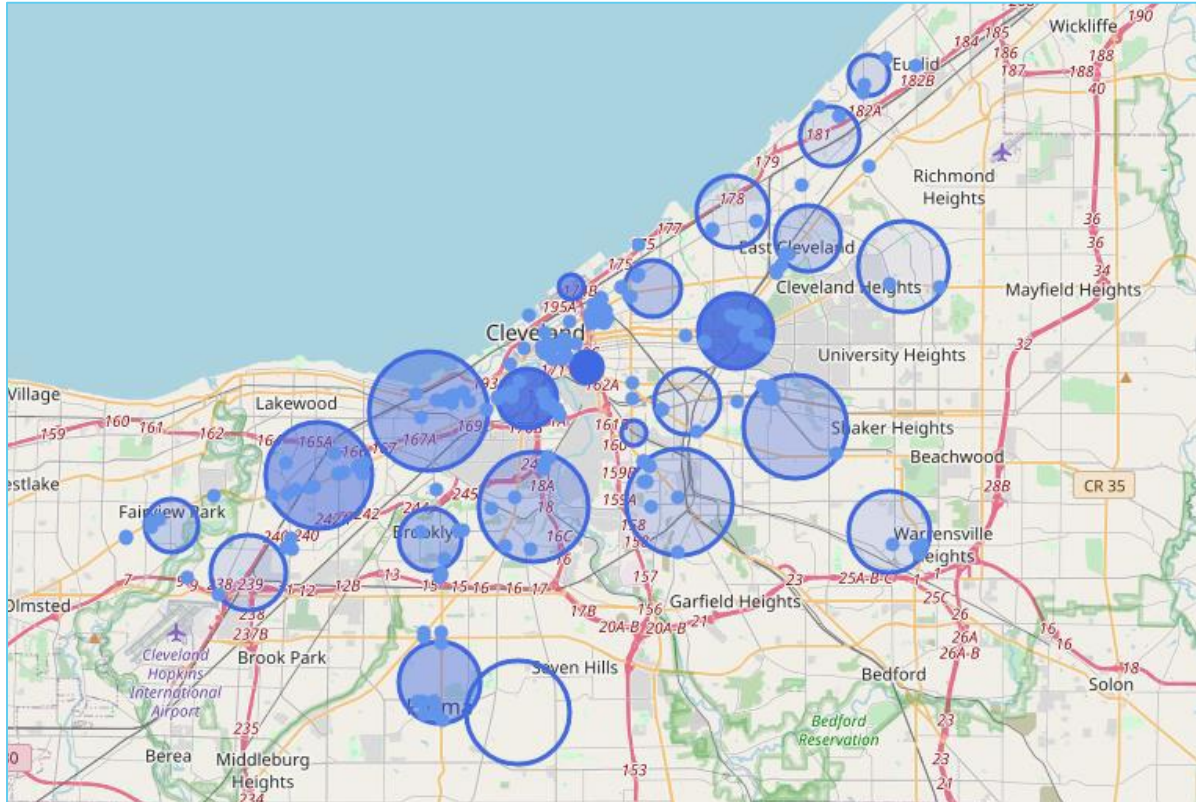
Clustering Algorithm

- Kmeans-algorithm to cluster the ZIP Codes into four different groups (based on the population and amount of restaurants)
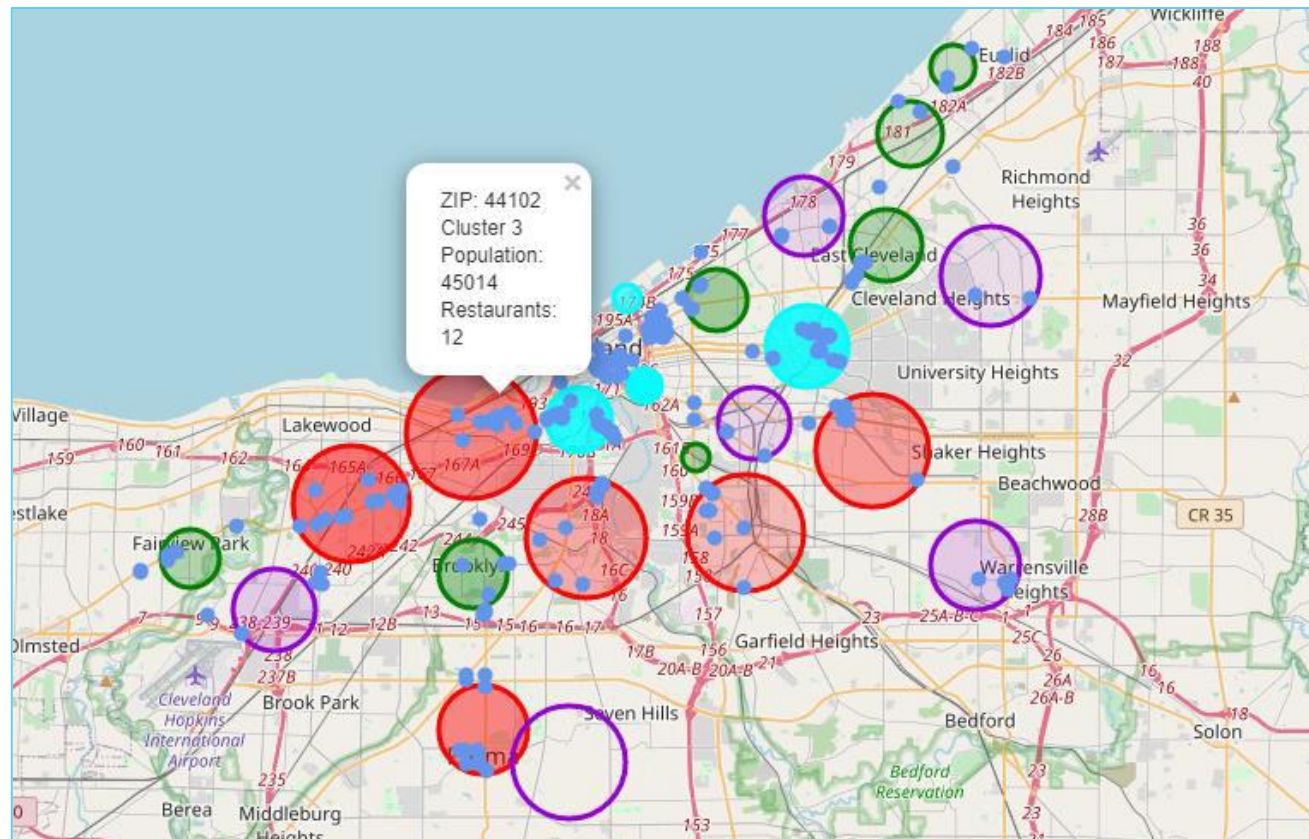
Results

- Cluster 0 (violet): medium high population / low restauraunt density

- Cluster 1 (blue): low population / high restaurant density

- Cluster 2 (green): medium low population / medium low restaurant density

- Cluster 3 (red): high population / medium low restaurant density

Conclusion and next steps

- analyzed different ZIP Codes in Cleveland, Ohio, based on their population and restaurants nearby

- Data analysis and Clusters give a first indicator of which locations might be suitable for opening a new restaurant in Cleveland

- for final decision-making more importants factors need to be analyzed (which type of restaurants are nearby?, which group of people lives nearby? etc.)

- This analysis gives a first hint on which location might be worth analyzing them in more detail