Capstone Project

Battle of the Neighborhoods - Vegetarian Restaurant in Krakow

V. Pacheco de Freitas, 27-06-2021

Introduction

- Recent results show that over 3 million Poles no longer eat meat and 8.4% of Polish adults had followed a vegetarian diet or had eaten only fish but no other meat for a period of time and people who eat meat want to make an effort to limit their consumption.

- Warsaw, capital of the country was also ranked the 6th most vegan friendly city on the planet. Thinking of this, we want to expand this trend to Krakow, and open a vegetarian restaurant in the city, which is one of the largest cities in Poland.

- The advantages of Krakow include the fact that the city attracts tourists with its beauty, unique historic complex and cultural events.

- Considering that the analysis in the scope of our initial research relies on free available information and constraints imposed by the Foursquare API, used to track the nearby venues for the district in the city, we will limit our analysis to a few key points:

-Population in each district

-Number of restaurants per district and

-Number of vegetarian restaurants already in the area.

Data Cleaning

For this preliminary analysis we will consider:

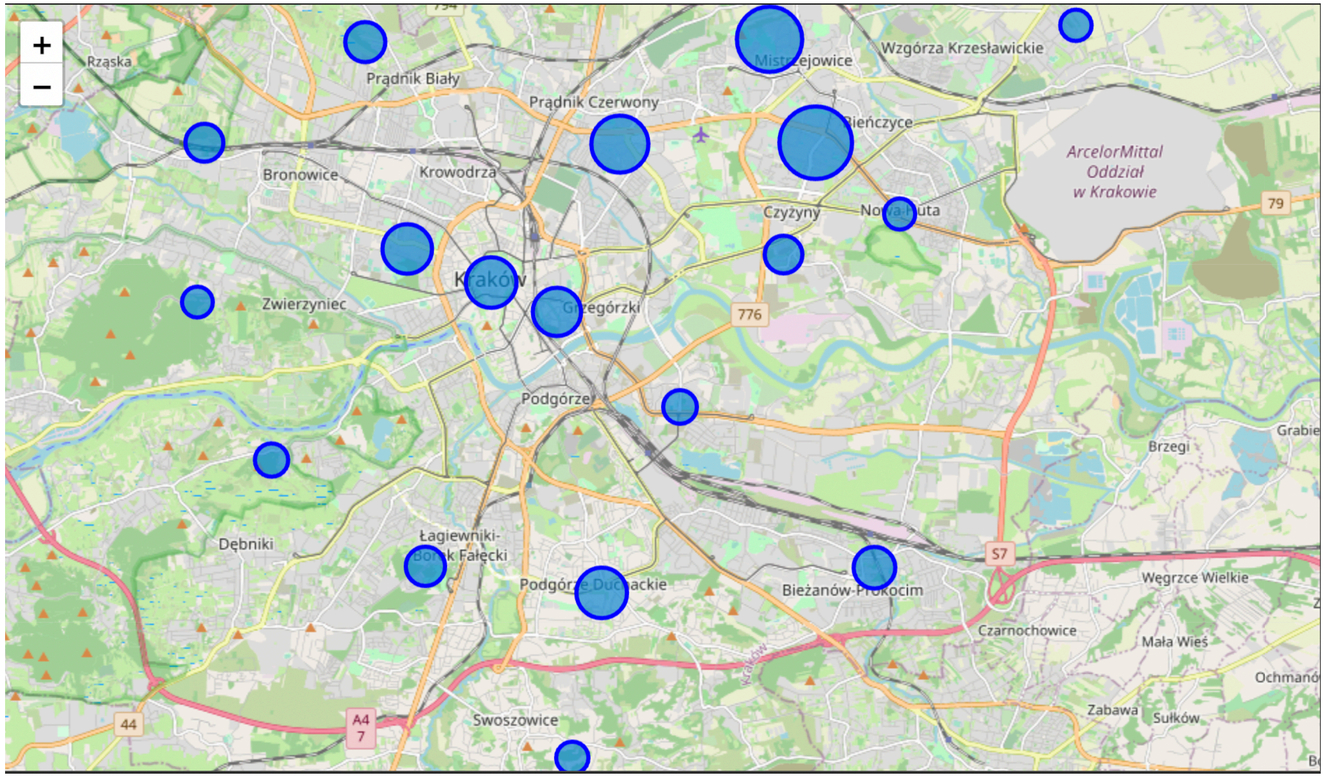
- the list of 18 districts in the Krakow area and their coordinates;-
- the population and population density in each district;
- the number of restaurantes within a certain radius from each district and the number of vegetarian restaurants in each district. This will be limited to the free information from Foursquare API.

We use the Foursquare API to search the venues in each district and to sort the total number of restaurants and total number of vegetarian or vegan restaurants in each district.

Using this in comparison to the population in each district, we apply an unsupervised learning technic (KMeans) to cluster the districts with regards to our feature selection in order to help us to assess the best neighborhood for the new restaurant.

Data Cleaning

Districts in our sample and some of the main venues in each of them, according to Foursquare



District	Stare Miasto	Grzegórzki	Prądnik Czerwony
1	Café	Bar	Food & Drink Shop
2	Hotel	Hotel	Bus Station
3	Italian Restaurant	Food Truck	Pizza Place
District	**Prądnik Biały**	**Krowodrza**	**Bronowice**
1	Coffe Shop	Café	Flower Shop
2	Grocery Store	Park	Train Station
3	Italian Restaurant	Bistro	Flea Market
District	**Zwierzyniec**	**Dębnik**	**Łagiewniki-Borek Fałęcki**
1	Zoo	Scenic Lookout	Shopping Mall
2	Grocery Store	Restaurant	Bus Station
3	Restaurant	Pizza Place	Gym
District	**Swoszowice**	**Podgórze Duchackie**	**Bieżanów-Prokocim**
1	Shoe Store	Supermarket	Convenience Store
2	Memorial Site	Convenience Store	Gym/Fitness Center
3	Liquor Store	Gas Station	Platform
District	**Podgórze**	**Czyżyny**	**Mistrzejowice**
1	Tram Station	Bus Stop	Tram Station
2	Bus Station	Supermarket	Park
3	Supermarket	Department Store	Sushi Restaurant
District	**Bieńczyce**	**Wzgórza Krzesławickie**	**Nowa Huta**
1	Supermarket	Construction & Landscaping	Park
2	Fast Food Restaurant	Diner	Bus Station
3	Tram Station	Pet Store	Café

Methodology

- Compute the number of restaurants and the number of vegetarian restaurants
- Compute the density of restaurants and density of vegetarian restaurants, which number of restaurants/area.

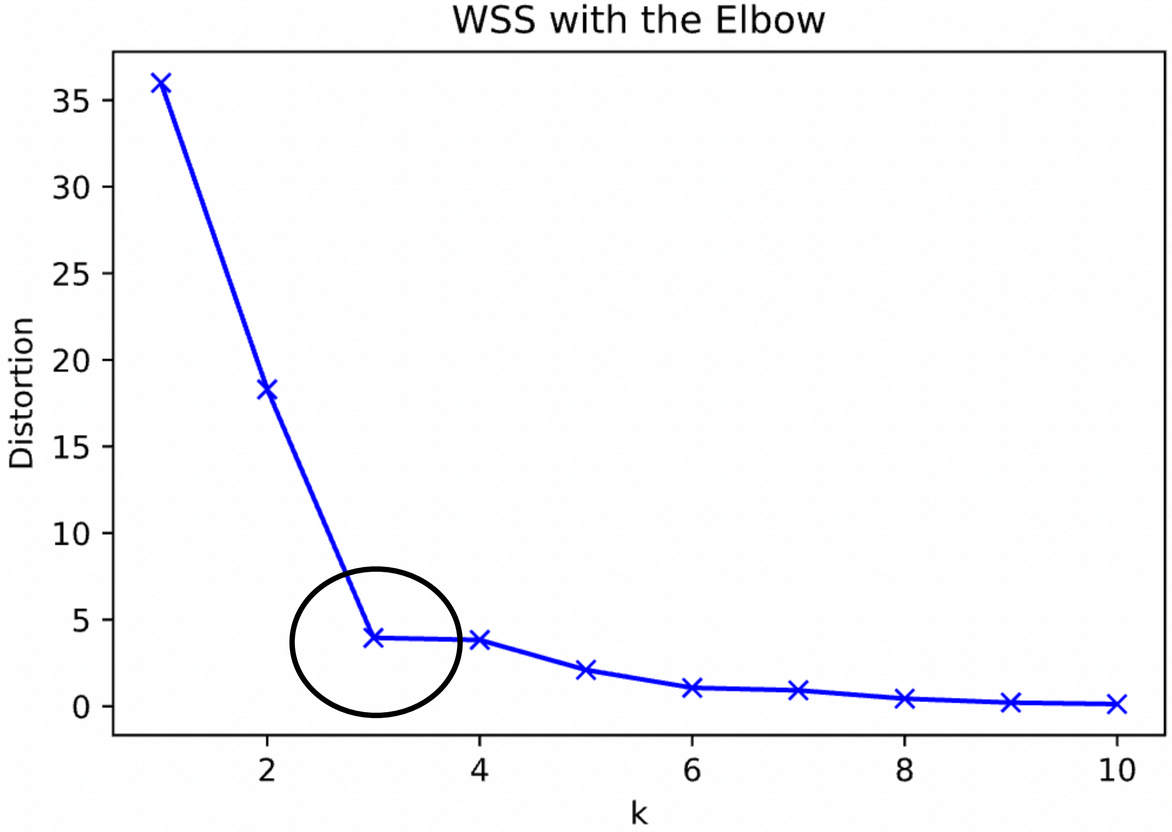
-After normalising this data we want to use the unsupervised learning method KMeans to cluster our data based on these characteristics. A cluster refers to a collection of data points aggregated together because of certain similarities.

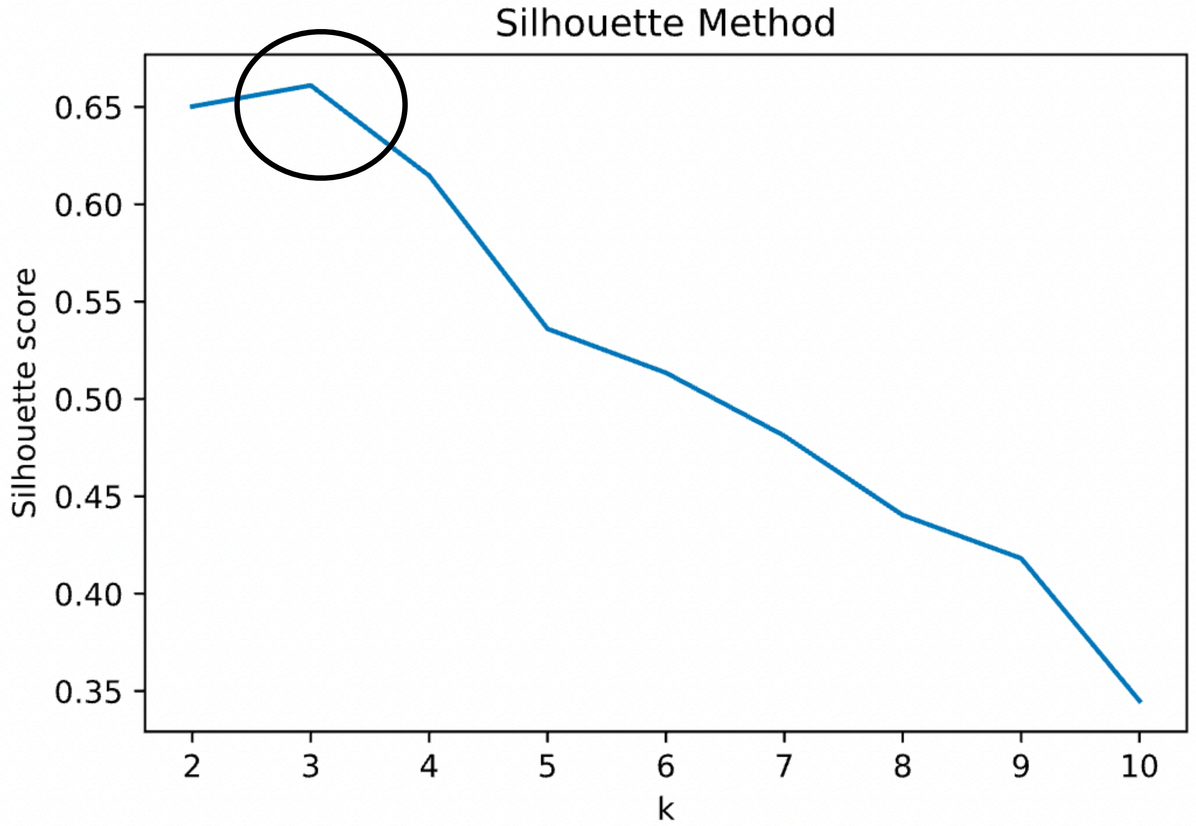
- We need to select the optimal K

K Selection

- **Elbow Method:** Within-Cluster-Sum of Squared Errors (WSS) for different values of k and choose the k for which WSS starts to diminish

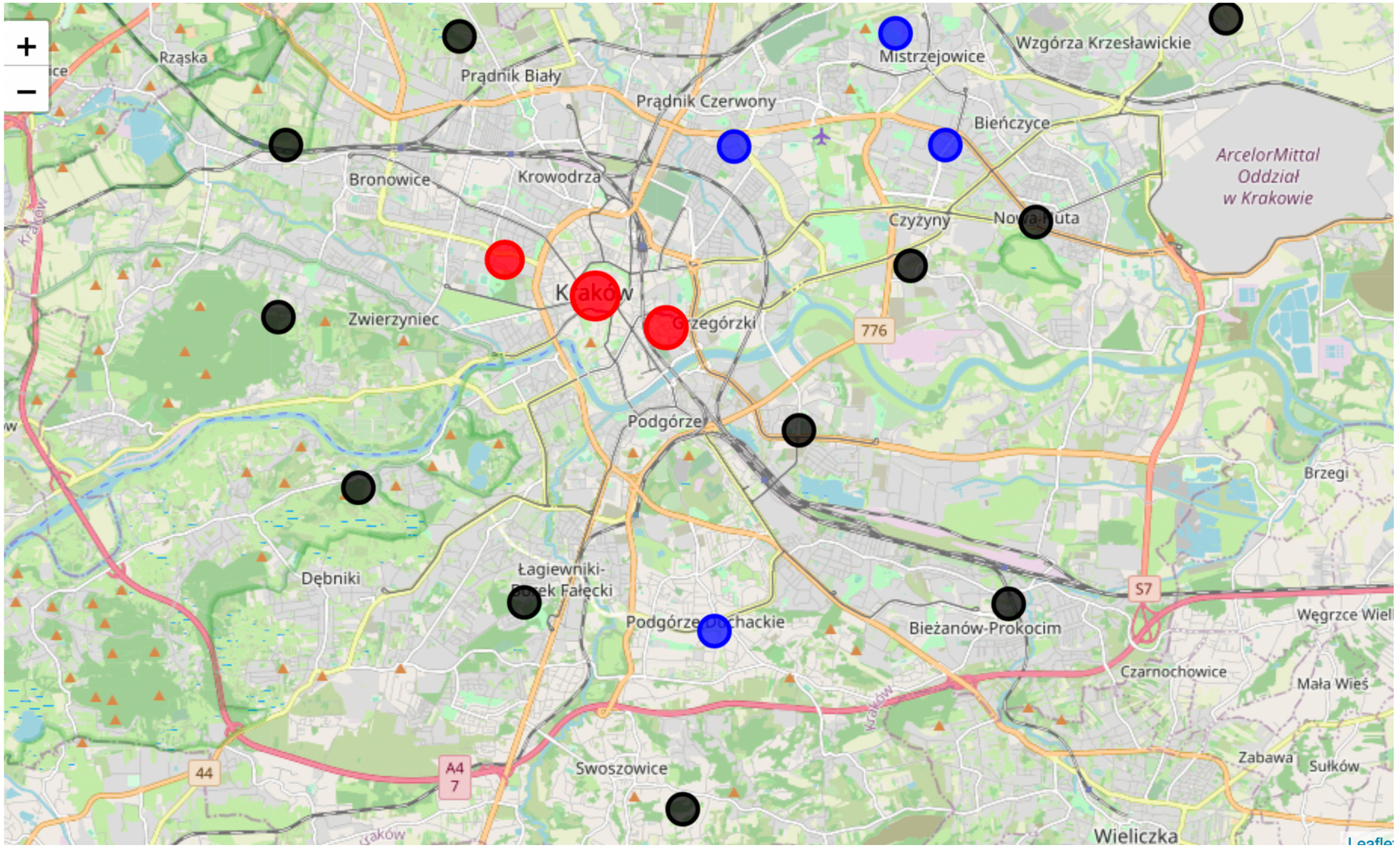
- **Silhouette Method:** A high value indicates that the point is placed in the correct cluster. This method reaches its global maximum at the optimal k.





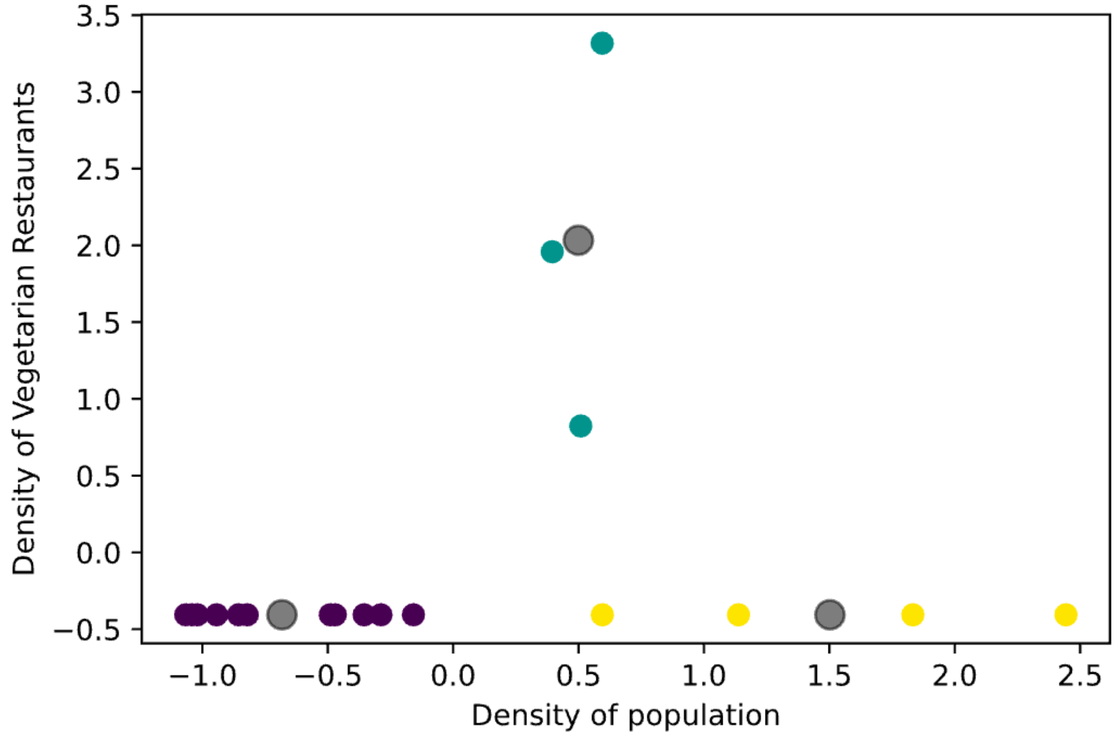
KMeans with k=3

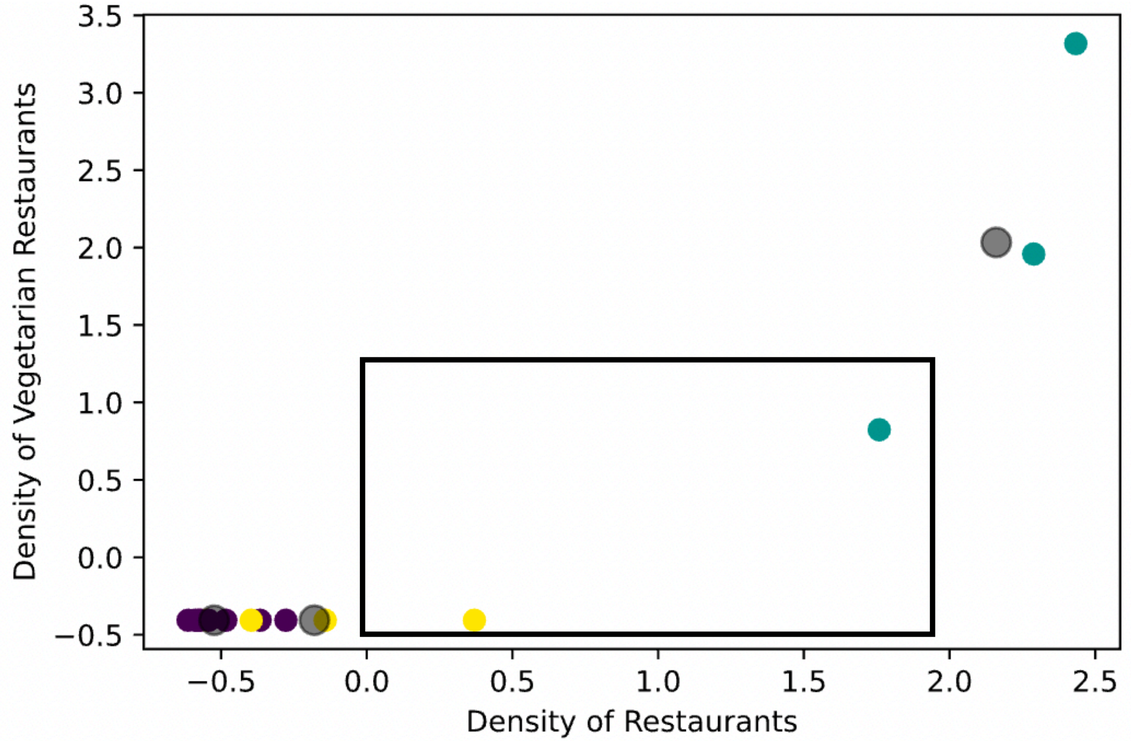
Density of restaurants in each cluster/district



Discussion and Final Selection

- We want a balance between population density and restaurant density
- We try to avoid the high restaurant density regions, since it would be more competitive





Conclusions

- Ultimately we choose **Krowodrza** as our district, considering that it has a high population density and it is close to the city center, universities, museums and touristic areas
- Also, it has a more moderate restaurant density than the other districts in the same cluster