Where to open a Veggie Center

Thomas Pfau

13. April 2020

Contents

Background

- We are a group of people dreaming of a wholesome nutrition for all citizens of Dresden to
  - Save the health of residents but the environment too
  - Cultivate regional vegetables
  - Avoid long shipping distances
  - Prevent the use of too much plant protection
- Our group likes to start with one shop and hopefuly becoming more and more popular
- Like to grow at the end

The right environment is important

- The first shop has to become a well-known place to bye wholesome nutrition
- Should be located in the neighborhood of venues for people that are looking for such a shop
- What groups of people would buy in our shop
- Wich venues does this people like

Data sources

- All venue data was retrieved from Foursquare
  - A great source to get interesting information
  - Venues are categorized and tagged
  - Data is proofed by many individuals
- To find groups of peoples in relation to organic food we used a study of Georg-August-Universität Göttingen ('Target groups for organic food: an overview')

Cleaning data

- Because of limitations by the Foursquare sandbox account there was some additional work needed:
  - Splitting the map into sections
  - Optimize the number of sections to avoid the daily limit
- Some corrections was made
  - Dropping wrong countries
  - Dropping wrong federal states
  - Correct the spelling of Dresden
- Creation of a higher level of categorization (domains) of the venues
- At the end there are 8.490 venues and 16 domains left

Target

- we are looking for a list of weighted geographic coordinates
- The quality of a site depends on the kind of the surrounding venues
- The algorithm has to find coordinates as center points of a bunch of interesting near located venues
- The quality of each found center can be weighted using the weigths of the cluster members

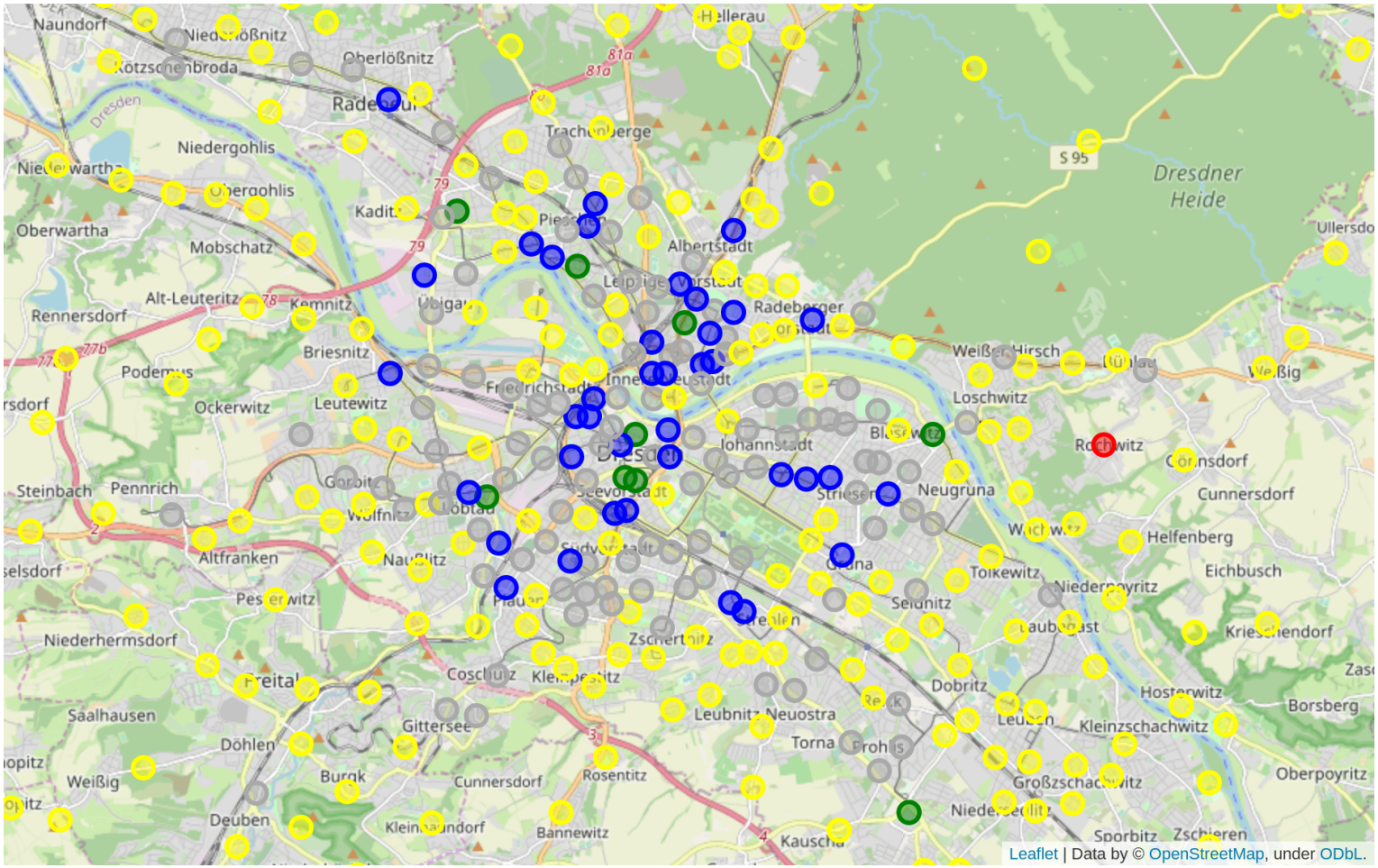
Relationship between the venues and the shop sites

- It is very important to find the right weights for the different kinds of venues
- Different weights of the venue domains
- Positive and negative weighted domains show the impact
- The sum of the weights of the surrounding venues is the quality of the site

K-Means

- This problem is related to explore clusters of venues
- K-Means clustering algorithm seems to be a very good choice
- It is fast and doesn't need labeled data
- We had to find the right 'K'
- For 8490 venues the value of 'K' was set to 424
- The calculation rans surprisingly fast: under a minute

Overview of the weighted centers of interest



Conclusions

- Analyzed the distribution of venues in relation to find promising sites
- Identified all interesting venues in the area using Foursquare as data source
- Data has to be cleaned and preparedto be useful
- It was needed to build a higher level category
- Using K-Means as cluster algorithm it was possible to process the whole area at once with very high performance finding 424 centers of interest

Future directions

- As result we got nine excellent sites and many good ones
- Seems to be sufficient
- Using Google Maps I recognized usefulness of the results
- Future improvements
  - Extenting the weights with more data
  - rental prices
  - number of residents