Capstone Project - The Battle of Neighborhoods

Finding the optimal location for a supermarket in Kiel, Germany

Introduction

- Background
  - Geographic location of a supermarket has a great impact on sales and earnings.
  - Important to evaluate all possible areas.

- Problem
  - Try to find the optimal site for a new supermarket in Kiel, Germany.
  - Area should meet 2 criteria:
    1. Area shoud not be crowded with existing supermarkets to minimize the effect of competitors.
    2. Close to drugstore or pharmacy to increase walk in effect.

- Interest
  - To anyone who wants to open a supermarket in Kiel.
  - They might not be physically present → an agent is time consuming and expensive.
  - Data science provides a possibility to evaluate the full area remotely and time efficient.

Data Acquisition

- Data sources:
  - Location of all supermarkets, drugstore and pharmacy within the area of interest.

    Taken from Foursqaure API

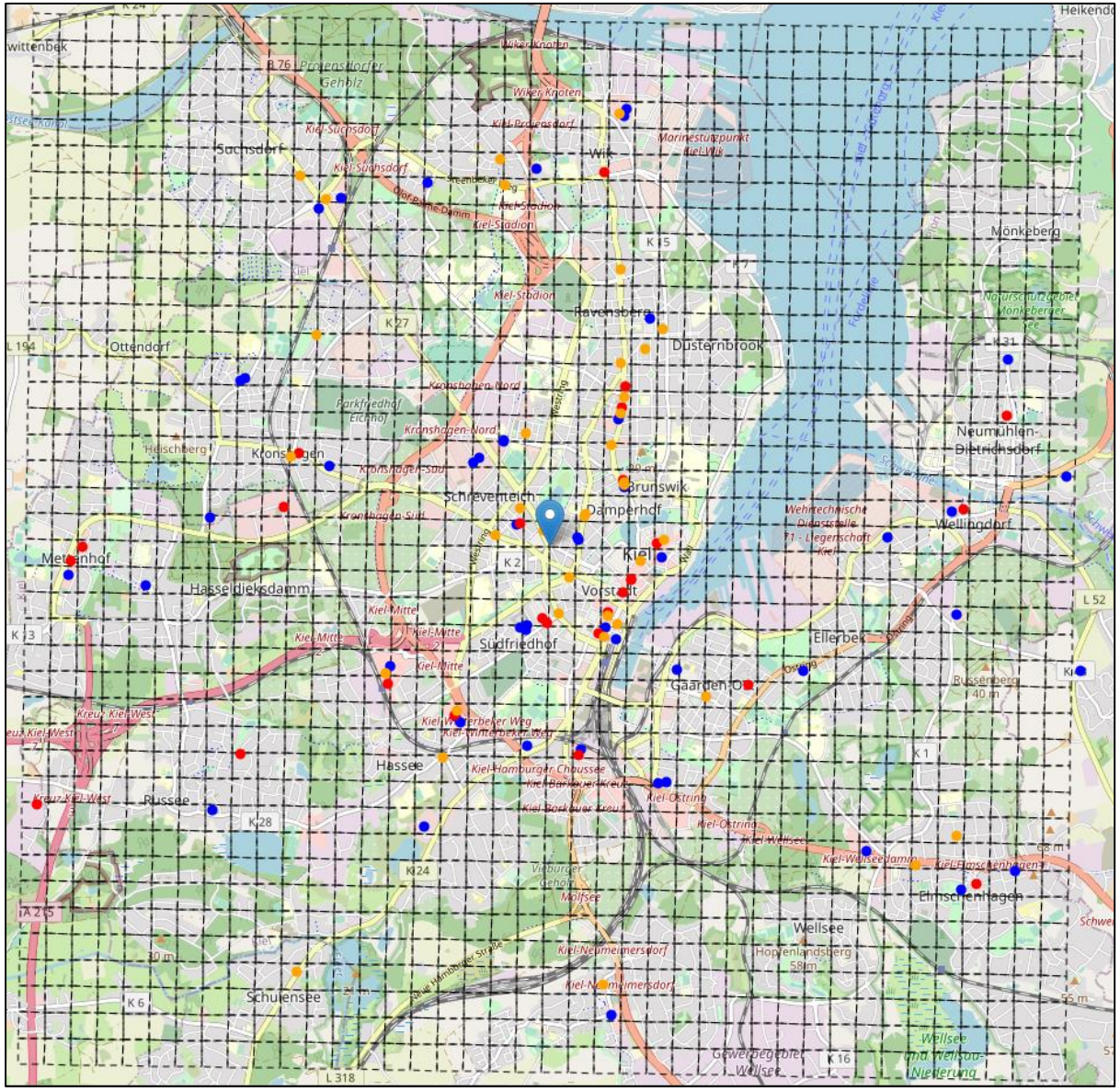
Store type:	Number:
Supermarket	49
Drugstore	27
Pharmacy	37

  - Coordinates of Kiel City Center (lat.:54.3232927, long.:10.1227652)

    From: www.gps-latitude-longitude.com

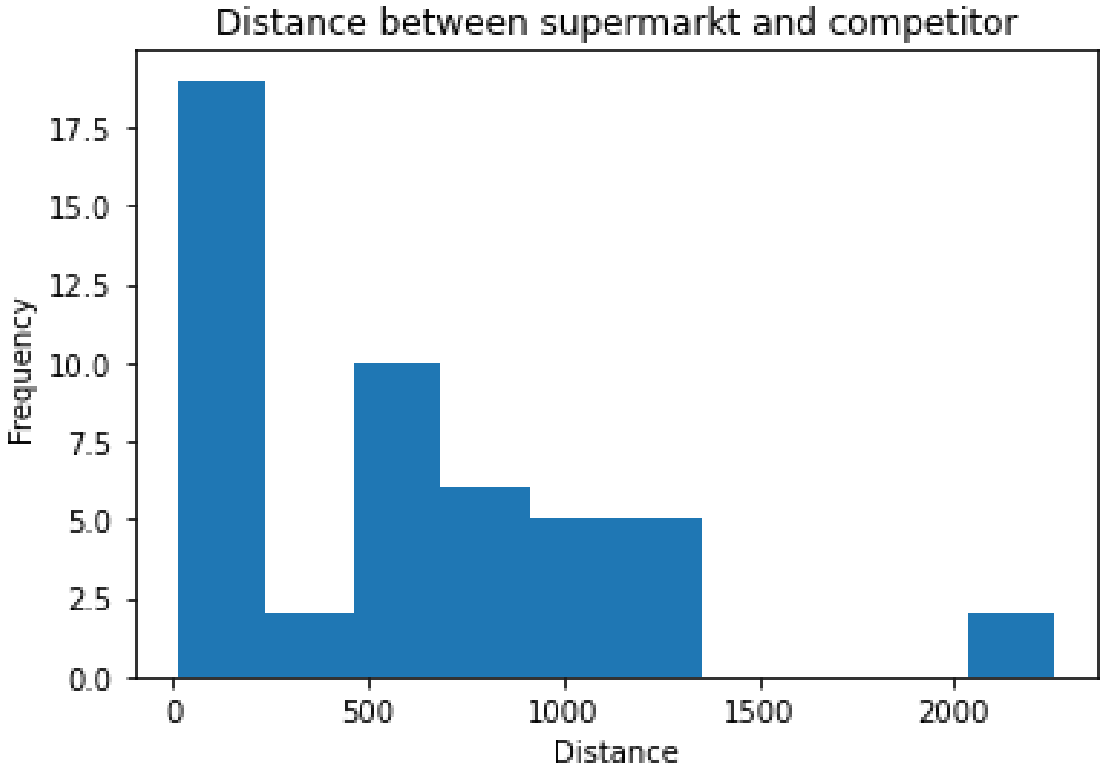
- Area of interest:
  - 5km x 5km around the city center of Kiel.
  - Regular spaced grid was created with a cell size of 250m x 250m.

Data Acquisition



Data Analysis: Criterion 1- Competitor

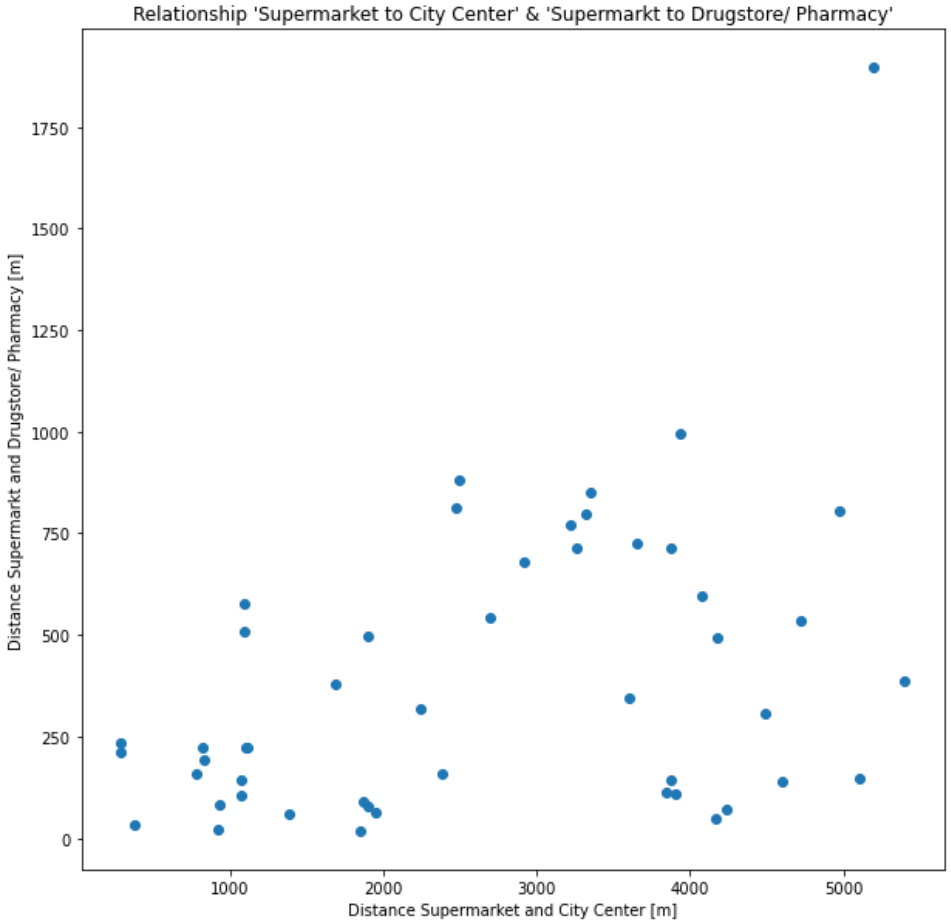
- The distance between existing supermarket and its nearest competitor was analyzed.



- Average distance: 566m.
- Minimum distance: 10m
- Maximum distance: 2256m

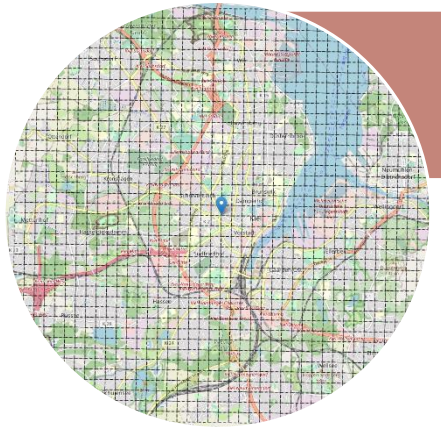
Data Analysis: Criterion 2- Walk in Effect

- The relationship between the ‚distance supermarket to city center‘ and ‚distance supermarket to neareast drugstore/pharmacy‘ was analyzed.



- Polynomial function might fit the datapoints best.
- Up to approx. 3000m to the city center the distance to the nearest drugstore pharmcy is increasing.
- To 4400m to the city center the distance to the nearest pharmacy/drugstore is decreasing.
- Beyond 4400m the datapoints are diffuse.

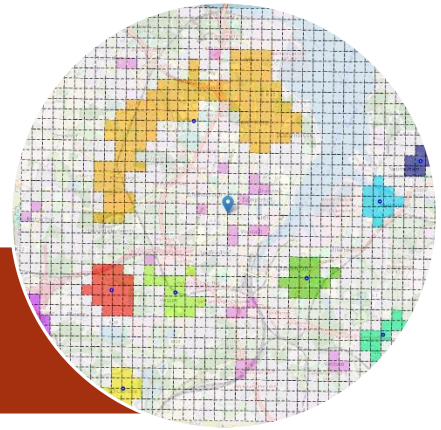
Data Modelling



Beginning: Each cell could be the spot.

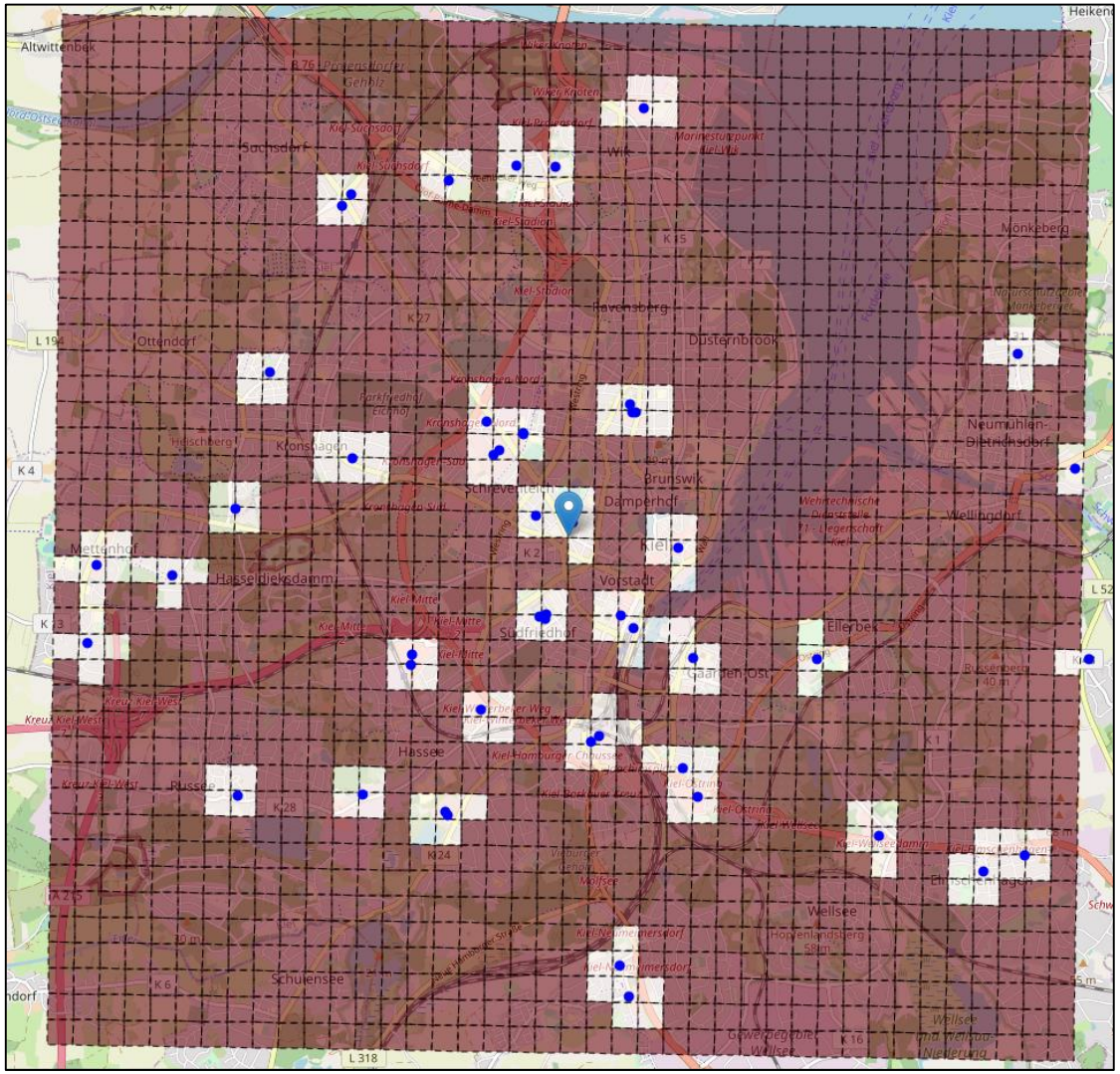
Modelling steps:

1. Define detailed criterion 1.
   → Minimum distance to next competitor.
2. Evaluate cells fulfill criterion 1.
3. Define detailed criterion 2.
   → Polynomial regression.
4. Evaluate cells fullfill criterion 2.
5. Clustering of remaining cells.



Result: Map showing clusters of possible spots.

Data Modelling: Criterion 1



- Due to data analysis the minimum distance to the next competitor was stet to 300m.

Kiel City Center

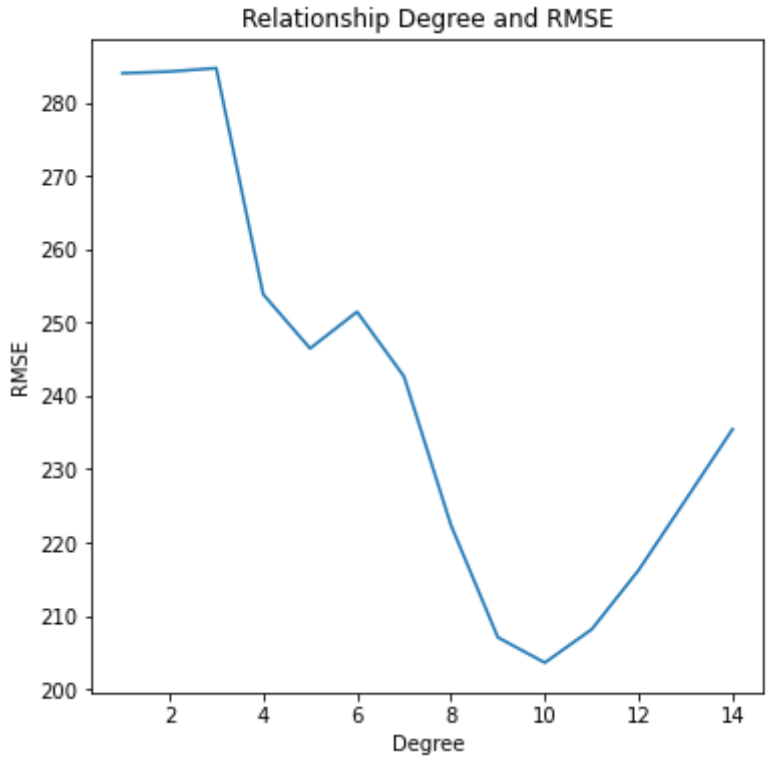
Supermarket

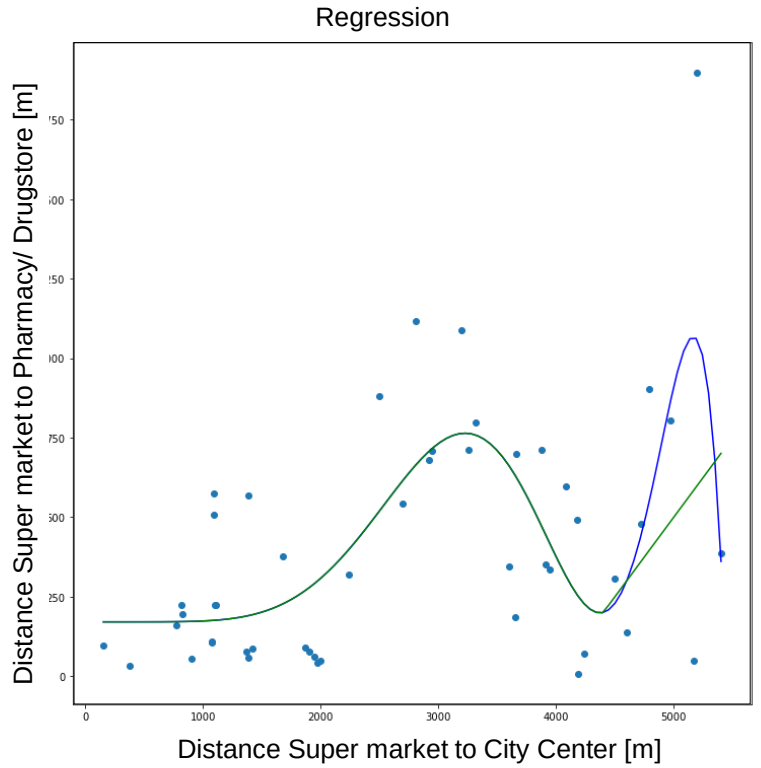
Dropped cells

Kept cells

Data Modelling: Criterion 2 - Regression

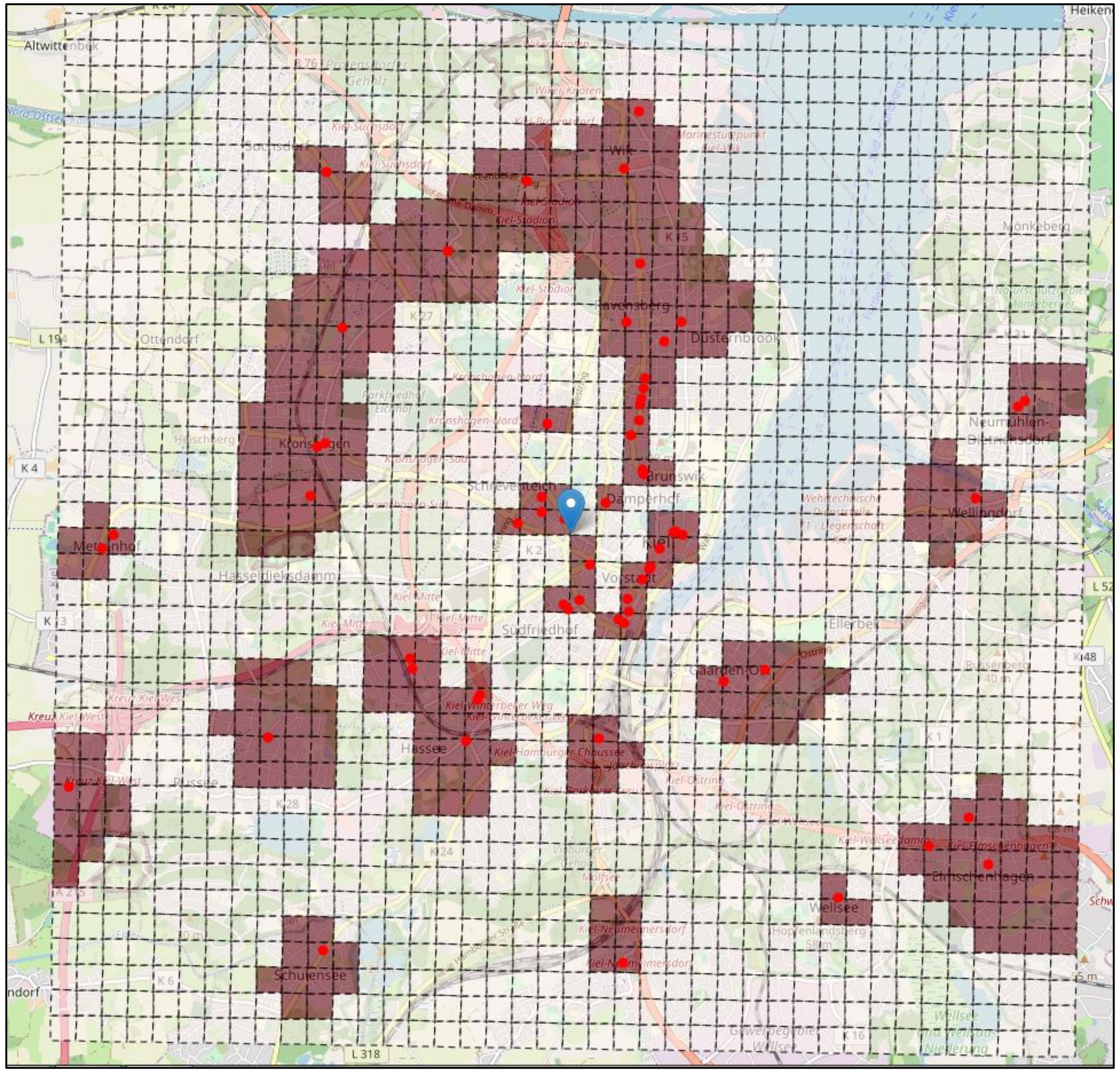
- Polynomial linear regression was used to fit the data points.
- The best calculated degree (min RMSE): 10.
- Comparision data points vs. regression deg. 10: Overfitting above 4400m from the city center.
- Lower degrees: Less accurate fit below 4400m.
- Final function: Below 4400m polynomial regression with degree10, above linear function.





- Data points
- Polynomial regression degree (10)
- Final function

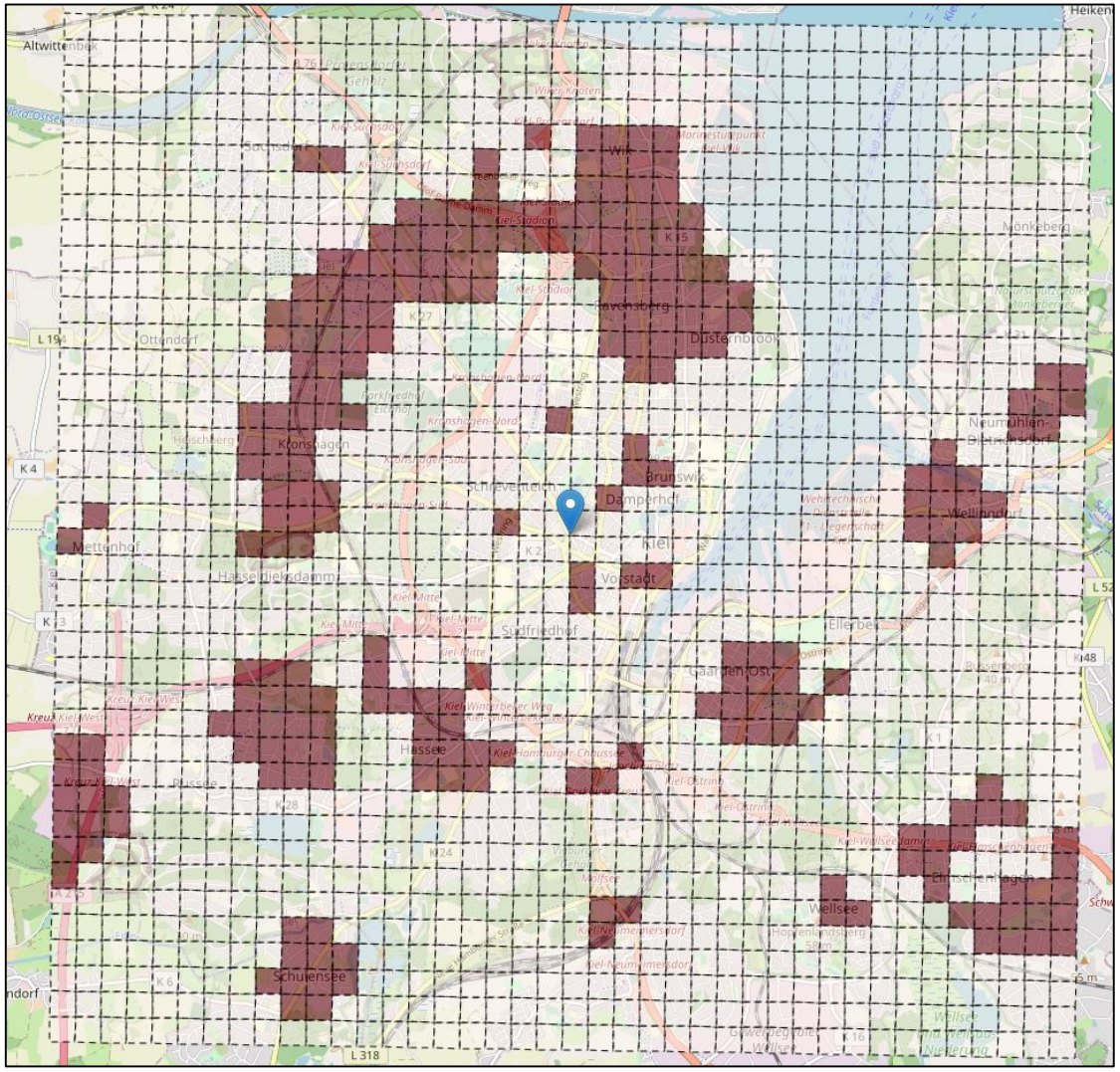
Data Modelling: Criterion 2



- The relationship between the ‚distance supermarket to city center‘ and ‚distance supermarket to neareast drugstore/pharmacy‘ could be described using a function of polynomial regression as well as linear regression for distances higher 4400m.

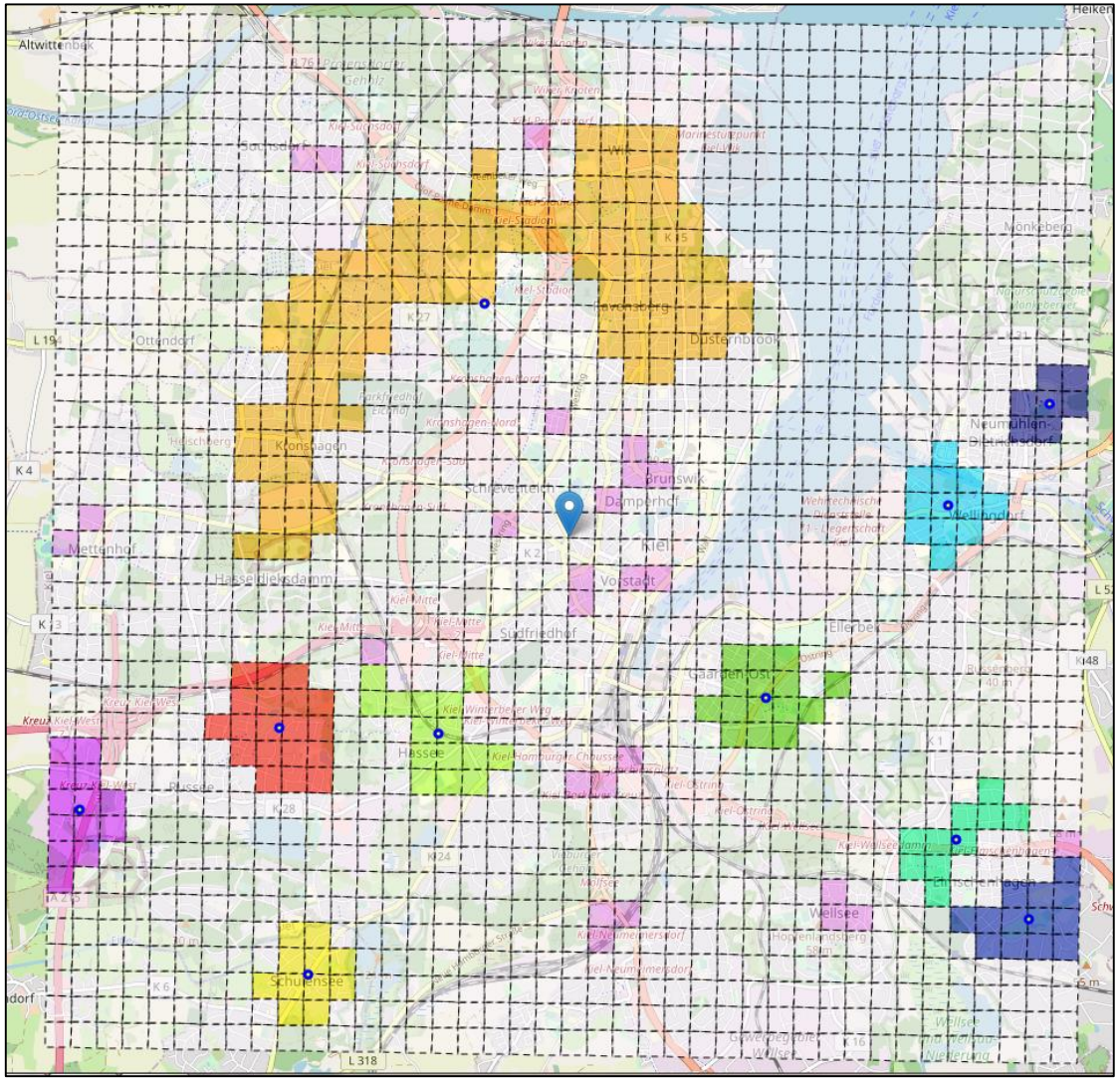
Kiel City Center
Drugstore/ Pharmacy
Dropped cells
Kept cells

Data Modelling: Clustering



- Cells which fullfill both criteria were the input for clustering.
- Around the city center neighborhoods which meet both criteria were incoherently displaced.
- It was decided to use DBSCAN to use the advantage of noise detection.
- DBSCAN parameter:
  - Radius of neighborhood: 400m.
  - Min. of data points within a neighborhood: 5.

Final Result



- 11 clusters were defined.
- 26 cells were defined as noise (rosa cell).
- The larges area (cluster) covers approx. 7,1km².
- All other areas are between 0,25km² and 1,2km² large.

Conclusion

- An area of 5km x 5km around the city center was analyzed to find the ideal location for a new supermarket.
- It was possible to define cells/ neighborhoods which were suitable for a new supermarket based on the defined 2 criteria.
- With the help of DBSCAN 11 clusters were defined to search for the best spot.
- Additional criteria might create a more detailed map.
  - Example for additional criterial:
  - Amount of people living in each neighborhood
  - Average income
  - Type of neighborhood
  - Distance to next station
- By using data science it was possible to evaluate the full area remotely and time efficient.