Capstone Project - The Battle of Neighborhoods

Finding the optimal location for a supermarket in Kiel, Germany

1. Introduktion

1.1. Background

The geographic location of a supermarket has a great impact on sales and earnings. Therefore it is important to evaluate all possible areas and find the best location before opening a new store. Walk-in traffic has a positive effect on earnings, while competitors nearby will have a negative effect.

1.2. Problem and Discussion of the background

In this project we will try to find the optimal site for a new supermarket in Kiel, Germany. The new supermarket should be in an area which is not already crowded with other supermarkets to minimize the effect of competitors. At the same time it should be close to a pharmacy or drugstore. The latter could cause a higher walk- in traffic and increase the revenue. The aim of the project is to have a map showing the best areas for an opening of a new supermarket based on the given conditions by using data science.

1.3. Interest

This study will be of interest to anyone who wants to open a supermarket in Kiel, for example supermarket chains. Possible stakeholders might not be physically present to search for the right place. An agent might be needed to evaluate the full area, which is expensive and time consuming. Data science provides a possibility to evaluate the full area remotely and time efficient.

2. Data

2.1. Data sources

For the study all supermarket, drugstore and pharmacy locations within the area of interest are needed. The information is gathered by using Foursqaure API.

2.2. Data formatting

The first step was to define the city center of Kiel. According to 'www.gps-latitude-longitude.com' the latitude is 54.3232927 and the longitude is 10.1227652.

The area of interest is 5km x 5km around the city center. A regular spaced grid was created with a cell size of 250m x 250m around the city center. The retrieved API data is transformed into python dataframes. Post data import data was cleaned and formatted. Only the important information such as 'name' of the store, address (street, postcode, city and state), latitude and longitude were kept. In

the end three dataframes exits, one for all supermarkets, one for all drugstores and one for all pharmacies.

	name	address	lat	lng	postalCode	city	state
0	REWE	Weißenburgstraße 15	54.325208	10.117955	24116	Kiel	Schleswig-Holstein
1	sky	Knooper Weg 41-43	54.323887	10.127047	24103	Kiel	Schleswig-Holstein
2	REWE	Knooper Weg 41-43	54.324142	10.126881	24103	Kiel	Schleswig-Holstein
3	REWE	Holstenstr. 1-11	54.322404	10.139209	24103	Kiel	Schleswig-Holstein
4	ALDI NORD	Torfmoorkamp	54.354451	10.104890	24106	Kiel	Schleswig-Holstein

Table 1: Top of supermarket dataframe.

	name	address	lat	lng	postalCode	city	state
0	Rossmann	Weißenburgstr. 21-29	54.325340	10.118420	24116	Kiel	Schleswig-Holstein
1	Rossmann	Schönberger Straße 32-34	54.326501	10.183425	24148	Kiel	Schleswig-Holstein
2	Rossmann	Am Ihlberg 4	54.301307	10.047539	24109	Melsdorf	Schleswig-Holstein
3	dm-drogerie markt	Herzog-Friedrich-Str. 30-42	54.320519	10.134703	24103	Kiel	Schleswig-Holstein
4	Rossmann	Kurt-Schumacher-Platz 15	54.322130	10.052540	24109	Kiel	Schleswig-Holstein

Table 2: Top of drugstore dataframe.

	name	address	lat	lng	postalCode	city	state
0	Forellen-Apotheke	Hamburger Landstr. 26a	54.286980	10.085543	24113	Molfsee	Schleswig-Holstein
1	West-Apotheke	Ringstr. 64	54.317645	10.124075	24103	Kiel	Schleswig-Holstein
2	Kronen-Apotheke	NaN	54.328792	10.133755	24105	Kiel	Schleswig-Holstein
3	Doc Morris Apotheke	NaN	54.360407	10.133042	24106	Kiel	Schleswig-Holstein
4	Impuls Apotheke	Sophienblatt 36	54.315578	10.130846	24103	Kiel	Schleswig-Holstein

Table 3: Top of pharmacy dataframe.

By plotting all stores (supermarkets, pharmacies and drugstores) it was observed that 3 stores (1 pharmacy, 1 drugstore and 1 supermarket) where outside the defined area of interest. They were discarded for further analysis. Finally 49 supermarkets, 37 pharmacies and 27 drugstores were found and displayed by using folium (see figure 1).
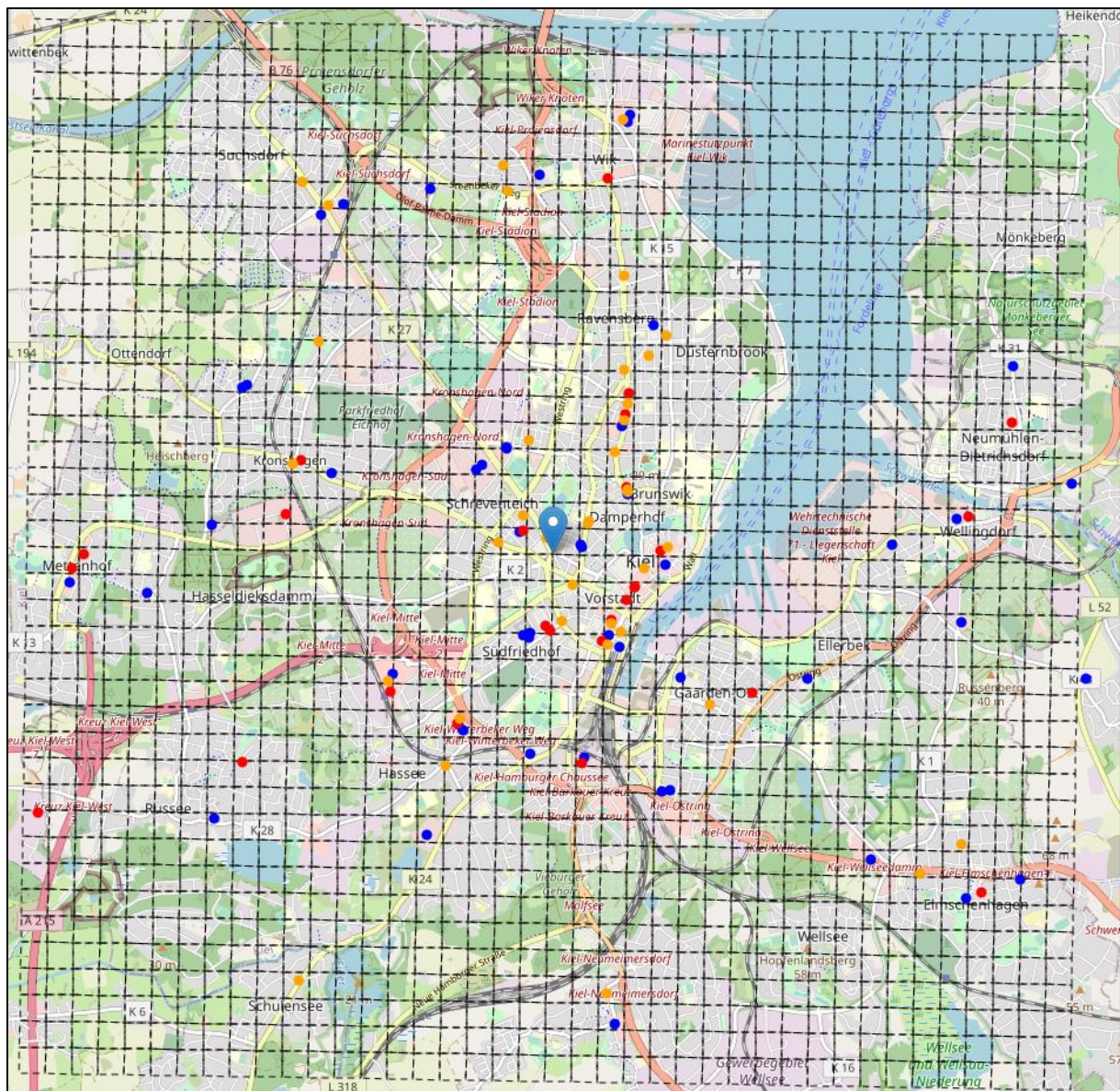
Figure 1: The Grid shows the area of interest around the city center of Kiel. Blue: supermarket, red: drugstore and orange: pharmacy.

2.3. Data use

The prepared data can be used to analyze which cell/neighborhood is optimal to open a new supermarket dependent on the two criteria: 1. Limit effect of competitors and 2. Higher walk in traffic due to an optimal distance to the next drugstore/pharmacy. By calculated the distance between an existing supermarket and its nearest competitor give information about the actual regional setting. A distance may be derived from this relationship to define which distance to a competitor is good for criteria 1.

For a higher walk in traffic the distance between supermarket and the next drugstore/ pharmacy will be analyzed. People living further away from the city center might be used to longer distances. Therefore the relationship between the distance from an existing 'supermarket to city center' and the distance between 'supermarket and drugstore/pharmacy' will be analyzed.

Finally all cells which meet both criteria will be clustered to have a map showing optimal areas for a new supermarket.