

Capstone Project

CLUSTERING OF CLINTON STORES

Business Problem

- ▶ The company Clinton is specializing in selling clothes in the region Germany, Austria and Switzerland and currently operates more than 200 stores in the given countries. The expansion department is planning to leverage the low rent prices due to corona crisis and open additional stores.
- ▶ In order to make a better decision on the particular locations, the expansion department needs additional analysis of the existing stores based on the venues which are located in close proximity to the stores. That is why they asked for a separate classification of the existing stores in order to be implemented into their normal store classification system (store type, store size, rent, brand portfolio, etc.).

Description Data

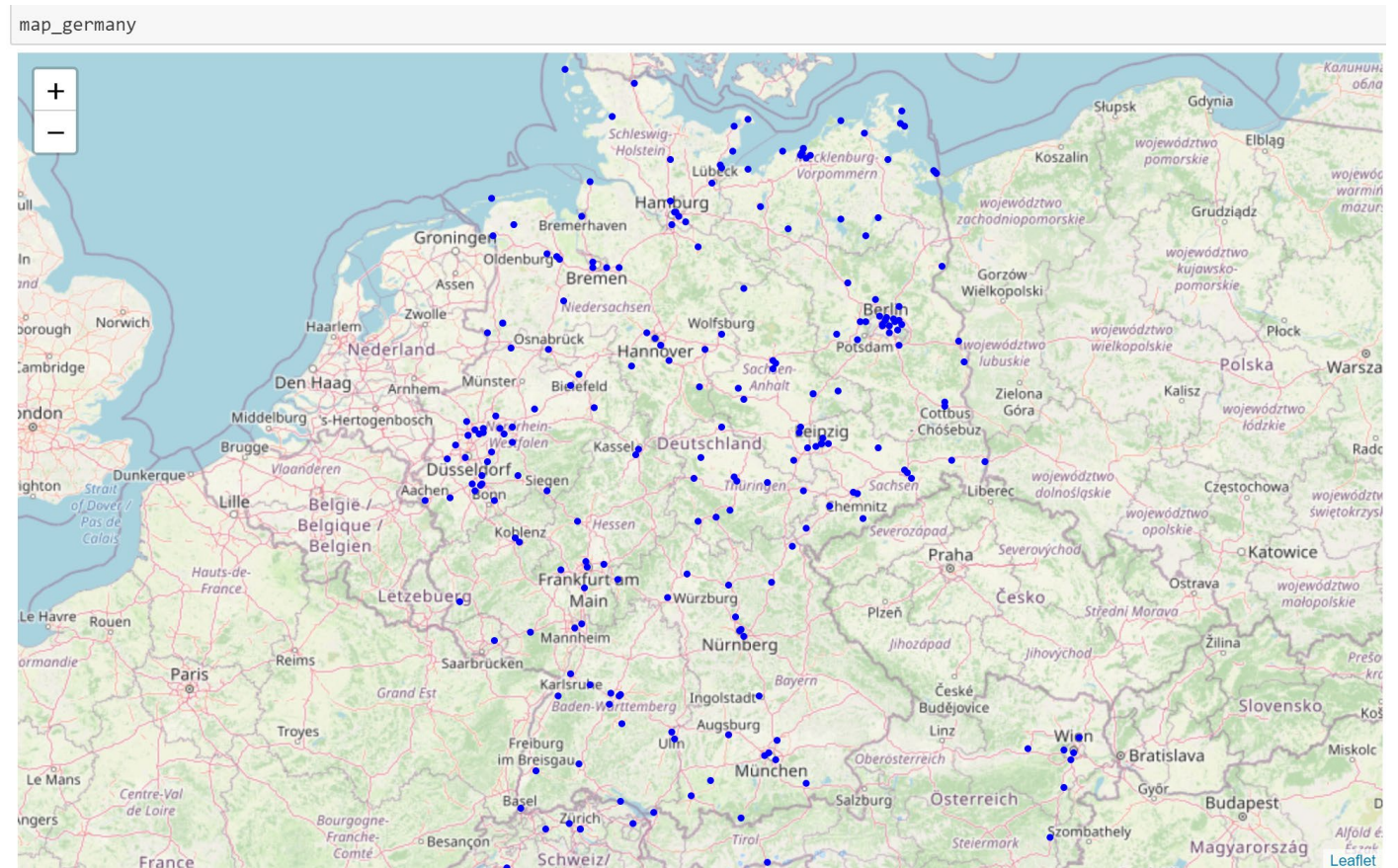
- ▶ For creating of additional classification, the locations of the existing stores will be used. The addresses of the stores will be scraped at the official website of the company in the section „Storefinder“
<https://www.campdavid-soccx.de/storefinder/>
- ▶ Afterwards the store addresses will be translated into geographical coordinates in order to be used for foursquare queries enabling getting the venues which are placed in close proximity to the stores.
- ▶ Having the data on the close venues from foursquare, the further analysis will be done using K-Means in order to classify the existing stores according to the wishes of the expansion department.

Existing stores

Using the Folium maps, one can see the current situation with the opened stores.

Most of the stores are situated in Germany, and a small part in Austria and Switzerland.

We can see the characteristic placement within main centers of settlement in Germany.



Foursquare Venues

For all the stores we exported the venues which are situated in close proximity to them (200 m and less), using the Foursquare API.

In total we received 2789 venues for further analysis.

```
print(cad_venues.shape)
cad_venues.head()
```

(2789, 7)

jt[18]:

	Adresse Voll	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Aachener Straße 1253, 50858 Köln	50.937966	6.835605	FC-FanShop	50.937860	6.835595	Sporting Goods Shop
1	Aachener Straße 1253, 50858 Köln	50.937966	6.835605	Apple Rhein-Center	50.937148	6.835336	Electronics Store
2	Aachener Straße 1253, 50858 Köln	50.937966	6.835605	Hollister	50.937880	6.835327	Clothing Store
3	Aachener Straße 1253, 50858 Köln	50.937966	6.835605	Food Lounge Rhein-Center	50.937214	6.836101	Food Court
4	Aachener Straße 1253, 50858 Köln	50.937966	6.835605	Starbucks	50.937892	6.835435	Coffee Shop

Grouping the venues

Using One-hot grouping, we managed to find out top 5 groups of venues for each store.

This will be our main dataframe to cluster the existing stores into groups, according to the task.

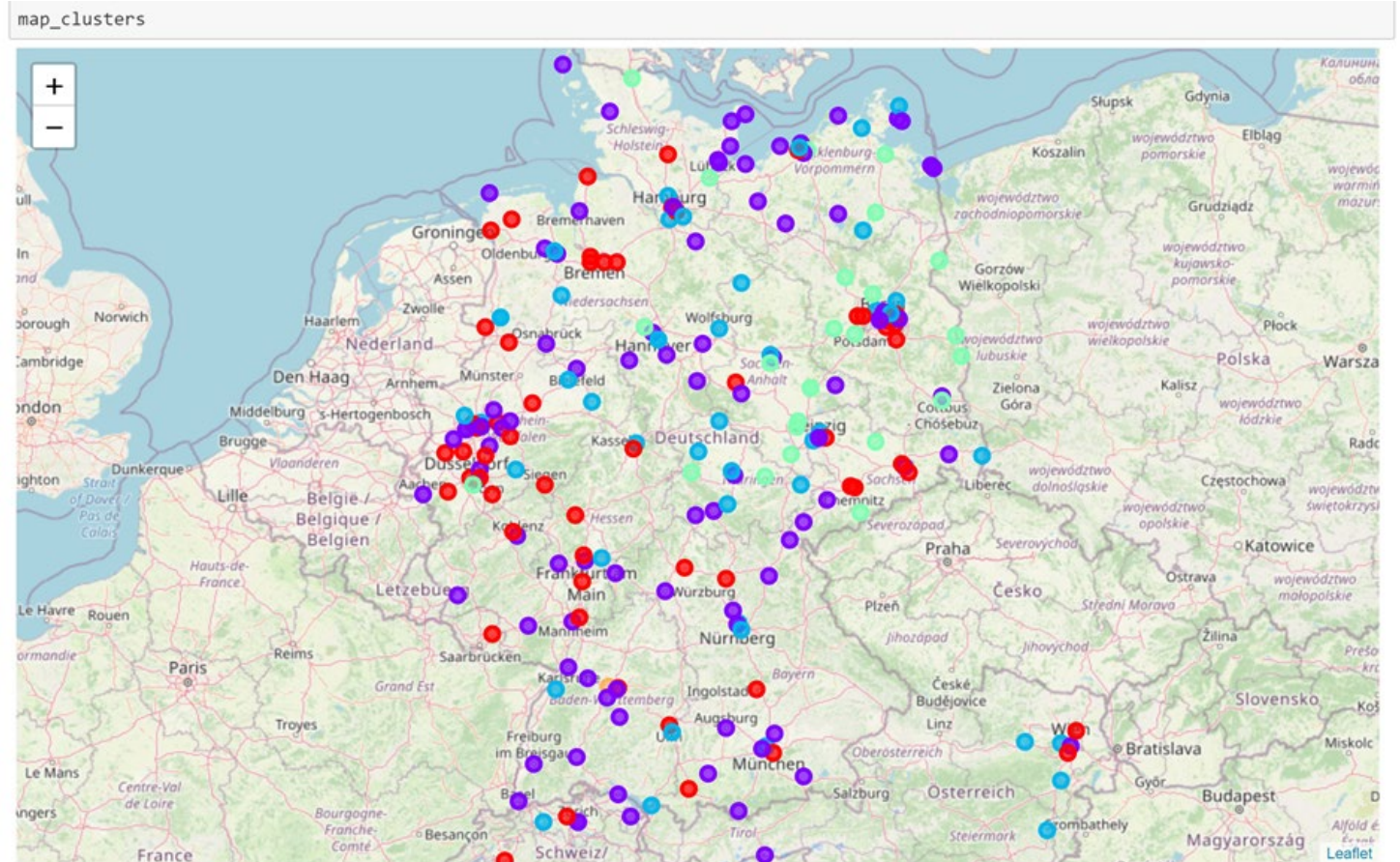
```
cad_venues_sorted.head()
```

27]:

	Adresse Voll	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	AG-Weser-Str. 3, 28237 Bremen	Drugstore	Coffee Shop	Multiplex	American Restaurant	Snack Place
1	Aachener Straße 1253, 50858 Köln	Clothing Store	Coffee Shop	Mobile Phone Shop	Electronics Store	Bakery
2	Achternstrasse 15-16, 26122 Oldenburg	Café	Nightclub	Clothing Store	Italian Restaurant	Drugstore
3	Adolfsreihe 6, 26548 Norderney	Café	Italian Restaurant	Tourist Information Center	Hotel	Drugstore
4	Albert-Schweitzer-Gasse 6, 1140 Wien	Italian Restaurant	Clothing Store	Café	Men's Store	Drugstore

Clustering

- Using K-Means we clustered the existing stores in 5 groups according to the venues that are placed in the close proximity to them.
- You can see the spread of the clusters on the map, done using Folium.



5 Clusters

► **Cluster 0 (marked red on the map)**

Most of the stores within cluster 2 are placed in close proximity to other clothing stores in the city areas where people normally go to buy clothes or places where you normally like to spend time at: bowling clubs, food courts , etc. You can find there also shopping malls and different types of catering as a normal part of the city surroundings.

► **Cluster 1 (marked violet on the map)**

Most of the stores within cluster 2 are placed close to supermarkets and big box stores which has less to do with selling clothes, but rather with food products, technical appliances, autos, electronics, etc. The catering industry is presented mostly with fast food / take away restaurants. These areas have a large number of people passing them by each day, but not normally treated as a place to shop clothes or enjoy your tame staying there.

► **Cluster 2 (marked blue on the map)**

Most of the stores within cluster 2 are placed in close proximity to drugstores and sport/yoga studios. The customers passing by will definitely have a large percentage of people having a healthy lifestyle.

► **Cluster 3 (marked green on the map)**

Most of the stores within cluster 3 are placed within popular resort areas at the North Sea and the Baltic Sea as well as mountain resorts in Switzerland and Austria. The nearby venues are mostly hotels, high-end restaurants and recreational areas.

► **Cluster 4 (marked orange on the map)**

The store in cluster 4 is rather an outlier being a mixture of clusters 0, 1 and 2.

Conclusion

- ▶ We used the Foursquare API and K-Means Clustering for the given Analysis. The output clusters make total sense and will be used for further analysis by the expansion team as well as for the scenario analysis of the potential locations for the future stores.
- ▶ These clusters can be also used by the assortment and merchandising teams in order to better plan the brand portfolio and product groups presented in the stores.