IBM Data Science Capstone Project

The Battle of Neighborhoods

Predicting the right location for a new supermarket in Berlin, Germany

Kuan Yew Cheng

June 3, 2021

1. Introduction

1.1 Background

We will have the opportunity in this project to be creative and come up with an idea to leverage the Foursquare location data to explore or compare neighborhoods or cities of our choice or to come up with a problem that we can use the Foursquare location data to solve. This project is about using a data science toolset on a real-life problem and demonstrating the creation of value by applying the learned skills. The supermarket business is very lucrative in the United States because that is where households go to get their daily needs. It is estimated that shoppers made about 1.5 trips to grocery stores on average every week in 2017. In 2015, supermarkets in the United States generated an average weekly sale of about 328,390 U.S. dollars, which works out to about 31,275 U.S. dollars per checkout, per week. According to the latest figures, supermarkets in the United States generated about 582.1 billion U.S. dollars in 2015 (Cynthia, 2021). Therefore, it is advantageous for us to accurately predict the proper location to set up a supermarket or retail market.

1.2 Problem

The client's issue that we are trying to solve is that company stakeholders of supermarket chains had invested plan to expand the number of supermarkets in Berlin, Germany. With a pandemic that began around March 2020 due to the highly infectious virus known as the coronavirus, or COVID-19, the statistics showed that the grocery sales boosted. This promotes the interest of the company in the expansion of its business. The spending amid pandemics should be conservative from the perspective of the stakeholders. The aim is to determine the strategic location, where the supermarket is easily accessible for local citizens and tourists and possibly occupied with high traffic flow.

1.3 Interest

Company owners and stakeholders would be very interested in accurate prediction of the potential location of new supermarket opening, for competitive advantage, business values, and long-term strategic planning. Thus, this analysis could be useful for a group of consultants, their business specialists, and the company stakeholders.

2. Data acquisition

To consider the problem we will obtain data and other relevant information as below:

1) List of the boroughs and neighborhoods of Berlin
2) Latitudes and longitudes of the boroughs in Berlin through geocoder
3) Top venues of neighborhoods through Foursquare API

A list of boroughs and neighborhoods will be obtained from Wikipedia, which is

(https://en.wikipedia.org/wiki/Boroughs_and_neighborhoods_of_Berlin)

The postal code, borough and neighborhood data can be found in the dataset.

3. Methodology

We will use unsupervised machine learning, which is k-mean clustering algorithm to search for clusters of neighborhoods that will provide us a list of regions for consideration of the expansion of the supermarkets. The decision should be approved if the supermarket to be located near to the area where peoples frequently visited the supermarket. After cleaning up the data, we will apply k-means clustering algorithm for creating clusters of neighborhoods. The silhouette score will be applied for choosing the optimal number of clusters before the search.

3.1 Data preparation and exploration

For data preparation, we begin by creating from our dataset a list of neighborhoods and boroughs in Berlin and then insert the coordinates of each neighborhood into this table. This can be done by using the list and a python library called geocoder. Through geocoder functions, we identify the latitude and longitude values of each neighborhood. After performing the task, we add the coordinates to the table. Finally, we received the following table as shown in figure 1 which exists as pandas dataframe format.

There are 97 neighborhoods in Berlin. All 97 neighborhoods will be used in the experiment. The decision of applied all neighborhoods for analysis will possibly affect the outcome of the analysis as some neighborhoods, for example, boroughs Steglitz-Zehlendorf and Reinickendorf have located outskirts of the city. Since we want to select an optimal location to establish a new supermarket, the area at the boundary of the city should be also considered as some peoples tend to live outside of the city. In the next step, we create a visual diagram (figure 2) using Folium library to show where the neighborhoods are situated in Berlin.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	101	Mitte	Mitte	52.517885	13.404060
1	102	Mitte	Moabit	52.530102	13.342542
2	103	Mitte	Hansaviertel	52.519123	13.341872
3	104	Mitte	Tiergarten	50.340922	6.956329
4	105	Mitte	Wedding	52.550123	13.341970
5	106	Mitte	Gesundbrunnen	52.550920	13.384846
6	201	Friedrichshain-Kreuzberg	Friedrichshain	52.512215	13.450290
7	202	Friedrichshain-Kreuzberg	Kreuzberg	52.497644	13.411914
8	301	Pankow	Prenzlauer Berg	52.539847	13.428565
9	302	Pankow	Weißensee	52.554619	13.463002
10	303	Pankow	Blankenburg	51.790268	10.955199

Figure 1: First 10 Neighborhoods with latitude and longitude

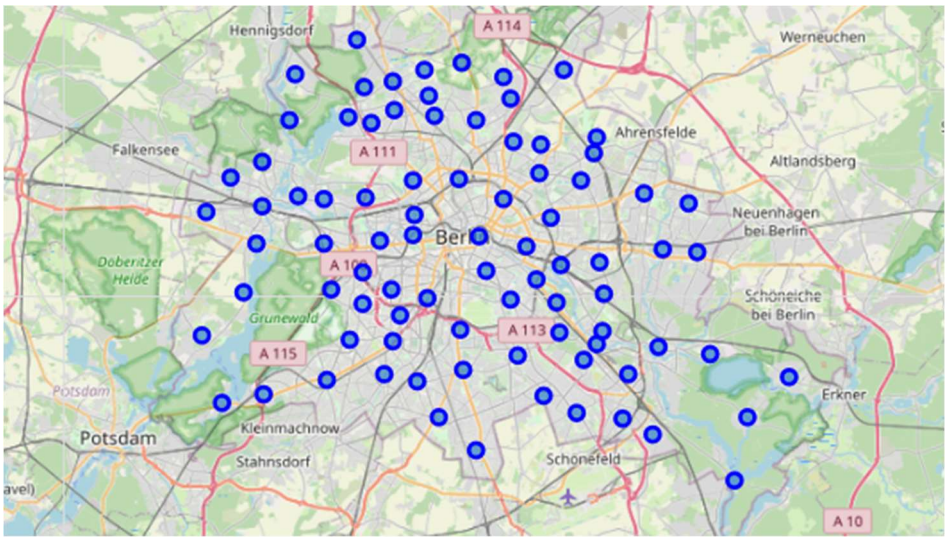

Figure 2: The location of all 97 neighborhoods in Berlin, Germany

Next, the neighborhoods were further explored. It means that the information of venues was collected for each neighborhood via Foursquare API. The venue data from Foursquare is retrieved in JSON format. We limited the maximum of 100 venues for each neighborhood. Venues are collected within a radius of 1000 meters from the neighborhood coordinates. The collected data is shown in figure 3.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Mitte	52.517885	13.40406	Bronzestatue "Heiliger St. Georg im Kampf mit ...	52.516290	13.405558	Outdoor Sculpture
1	Mitte	52.517885	13.40406	Designpanoptikum - surreales Museum für indust...	52.516941	13.406072	Museum
2	Mitte	52.517885	13.40406	Kuppelumgang Berliner Dom	52.518966	13.400981	Scenic Lookout
3	Mitte	52.517885	13.40406	Tigertörtchen	52.517150	13.407926	Cupcake Shop
4	Mitte	52.517885	13.40406	Radisson Blu	52.519561	13.402857	Hotel

Figure 3: The first five rows of collected data of the first neighborhoods via Foursquare API

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Adlershof	9	9	9	9	9	9
Alt-Hohenschönhausen	9	9	9	9	9	9
Alt-Treptow	23	23	23	23	23	23
Altglienicke	1	1	1	1	1	1
Baumschulenweg	4	4	4	4	4	4
Blankenburg	6	6	6	6	6	6
Blankenfelde	4	4	4	4	4	4
Bohnsdorf	4	4	4	4	4	4
Borsigwalde	4	4	4	4	4	4

Figure 4: Overview of the dataset with added venue data

For analyzing the neighborhoods in Berlin, we focus on venue categories. We use the one-hot encoding in this case. This method creates dummy variables for categories before the dataset is applied in the machine learning technique. In addition, we create a function to sort the venues in descending order for each neighborhood, which shows the top ten most common venues for each neighborhood (shown in table 1).

Table 1: First 4 neighborhoods with the top ten most common venues

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Adlershof	Supermarket	Pizza Place	Italian Restaurant	Steakhouse	Shopping Mall	Greek Restaurant	Trattoria/Osteria	Drugstore	Diner	Dessert Shop
Alt-Hohenschönhausen	Tram Station	Indian Restaurant	Greek Restaurant	Big Box Store	Supermarket	Discount Store	Post Office	Drugstore	Falafel Restaurant	Farmers Market
Alt-Treptow	Italian Restaurant	Bakery	Playground	Tapas Restaurant	Garden Center	Boat or Ferry	Electronics Store	Mexican Restaurant	Juice Bar	Drugstore
Altglienicke	Home Service	Yoga Studio	Ethiopian Restaurant	Food Court	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market	Fish & Chips Shop	Fast Food Restaurant
Baumschulenweg	Supermarket	Ice Cream Shop	Asian Restaurant	Event Space	Fountain	Food Court	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market

3.2 Clustering algorithm

After the preparation of our dataset, we perform clustering to investigate the structure of the data. For this, an unsupervised machine learning technique will be used based on the K-means algorithm. Clustering is one of the most common exploratory data analysis techniques used to get an intuition about the structure of the data. It can be defined as the task of identifying subgroups in the data such that data points in the same subgroup (cluster) are very similar while data points in different clusters are very different (Dabbura, 2018). For K-means clustering, we first need to decide on the number of clusters that we want to use. To avoid the trial-and-error approach, the silhouette score was applied. Figure 5 shows the silhouette scores for a range of cluster variations.

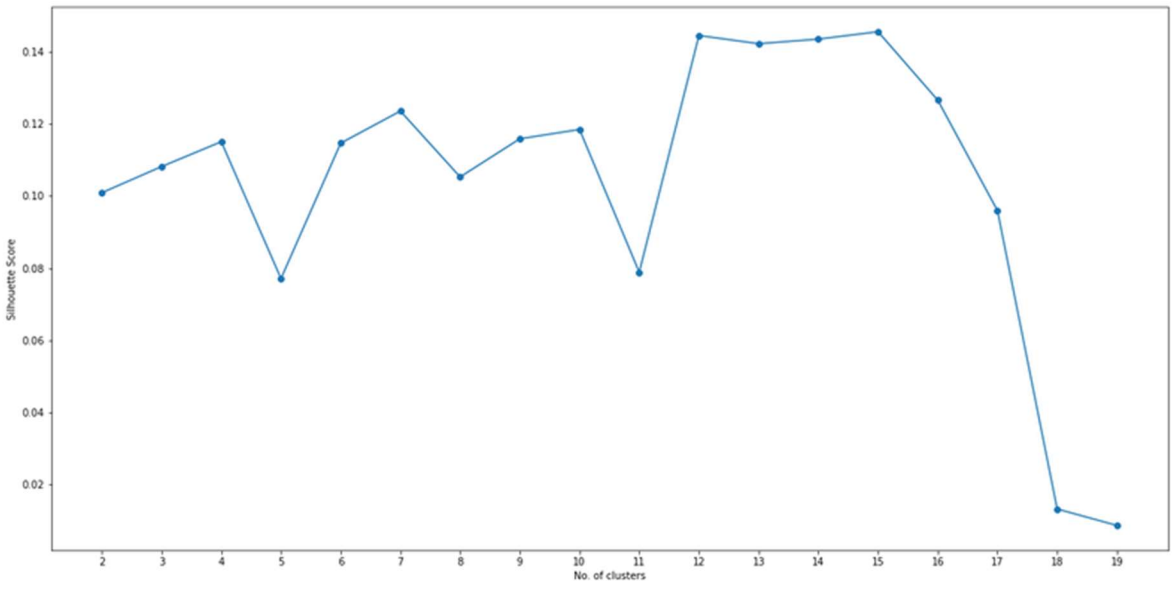

Figure 5: Silhouette score of different number of clusters applied to K-means algorithm

From the figure 5, It is clear that the optimal number of clusters to be used is 12 since it obtained the highest silhouette score. Furthermore, we apply the K-means clustering algorithm with the parameter of 12 as the number of clusters. When the algorithm finished computing, we add the cluster labels to the dataset as shown in table 2.

Table 2: The dataframe modifed after appending the cluster labels for each neighborhood

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	101	Mitte	Mitte	52.517885	13.404060	1.0	German Restaurant	Museum	Hotel	History Museum	Café
1	102	Mitte	Moabit	52.530102	13.342542	1.0	Café	German Restaurant	Burger Joint	Gym / Fitness Center	Hostel
2	103	Mitte	Hansaviertel	52.519123	13.341872	1.0	Café	Art Museum	Pedestrian Plaza	Irish Pub	Metro Station
3	104	Mitte	Tiergarten	50.340922	6.956329	1.0	Hotel	Rental Car Location	Trail	Restaurant	Eastern European Restaurant
4	105	Mitte	Wedding	52.550123	13.341970	7.0	Supermarket	Pharmacy	Ice Cream Shop	Tennis Court	Track
5	106	Mitte	Gesundbrunnen	52.550920	13.384846	1.0	Turkish Restaurant	Drugstore	Supermarket	Clothing Store	Hotel

We visualize the clusters on the map using Folium library for better interpretation.

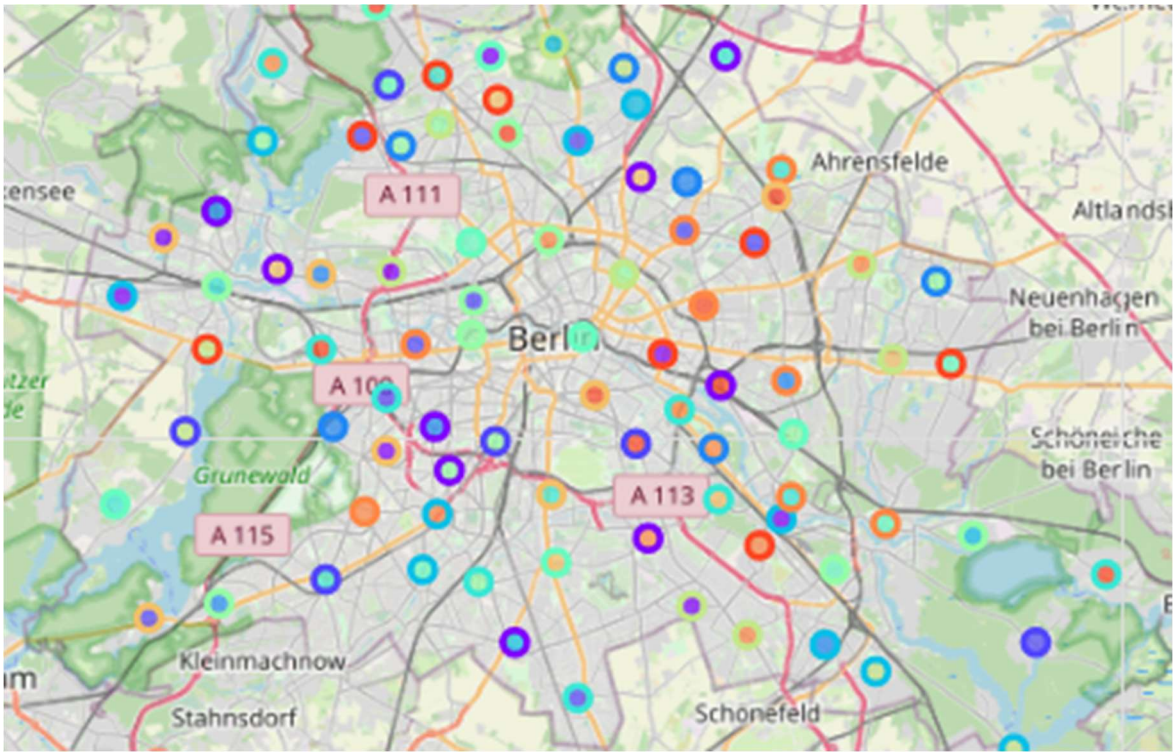

Figure 6: The location of the clusters after running the K-means algorithm

3.3 Limitation

The experiment has some limitations that should be mentioned in this report. First, when collecting venues via Foursquare API, a 1000-meter radius is used around the center coordinates of the neighborhoods. The number of collected venues is limited to 100 per neighborhoods. Second, four of the neighborhoods were not successfully assigned with cluster labels due to some unknown technical issue with Kmeans function. One of the possible reasons is that these places are outliers of the dataset.

4. Results

4.1 Examine clusters

By reviewing at the cluster tables, we can see that cluster 8 is the one that we are the most interested in.

4.1.1 Cluster 1

The first cluster (Cluster label 0) is outskirt where retail market is not really represented. Restaurants and public transport are at the top.

	Postal Code	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
14	307	0.0	Asian Restaurant	Turkish Restaurant	Tram Station	Yoga Studio	Ethiopian Restaurant	Food Court	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market
70	1001	0.0	Asian Restaurant	Windmill	German Restaurant	Supermarket	Tram Station	Fish & Chips Shop	Falafel Restaurant	Farmers Market	Fast Food Restaurant	Yoga Studio
83	1101	0.0	Tram Station	Indian Restaurant	Greek Restaurant	Big Box Store	Supermarket	Discount Store	Post Office	Drugstore	Falafel Restaurant	Farmers Market

4.1.2 Cluster 2

Cluster 2 (Cluster label 1) is the biggest cluster with a number of 57 neighborhoods. This is where we see lots of commercial activities related venues (coffee shops, restaurants, hotels, museums, pubs, etc...)

	Postal Code	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	101	1.0	German Restaurant	Museum	Hotel	History Museum	Café	Exhibit	Restaurant	Fountain	Plaza	Drugstore
1	102	1.0	Café	German Restaurant	Burger Joint	Gym / Fitness Center	Hostel	Bar	Doner Restaurant	Hotel	Clothing Store	Drugstore
2	103	1.0	Café	Art Museum	Pedestrian Plaza	Irish Pub	Metro Station	Bus Stop	Mediterranean Restaurant	Farmers Market	Liquor Store	Bistro
3	104	1.0	Hotel	Rental Car Location	Trail	Restaurant	Eastern European Restaurant	Steakhouse	Fish Market	Fish & Chips Shop	Fast Food Restaurant	Electronics Store
5	106	1.0	Turkish Restaurant	Drugstore	Supermarket	Clothing Store	Hotel	Bookstore	Bar	Trail	Garden	Big Box Store
6	201	1.0	Coffee Shop	Bookstore	Bagel Shop	Pub	Doner Restaurant	Café	Middle Eastern Restaurant	Hostel	Vegetarian / Vegan Restaurant	Plaza
7	202	1.0	Café	Italian Restaurant	Bakery	Coffee Shop	Waterfront	German Restaurant	Turkish Restaurant	Bar	Nightclub	Kurdish Restaurant
8	301	1.0	Café	Beer Bar	Cocktail Bar	Park	Organic Grocery	Coffee Shop	Bakery	Falafel Restaurant	Concert Hall	Candy Store
9	302	1.0	German Restaurant	Hotel	Tram Station	Memorial Site	Playground	Vietnamese Restaurant	Beach	Park	Exhibit	Falafel Restaurant
10	303	1.0	Hotel	German Restaurant	Plaza	Supermarket	Park	Café	Yoga Studio	Farmers Market	Exhibit	Falafel Restaurant
15	308	1.0	Café	Miscellaneous Shop	Automotive Shop	Event Space	Fountain	Food Court	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market

4.1.3 Cluster 3

Cluster 3 (Cluster label 2) contains only one neighborhood where service sector is the most common venue, but behind that, fish market and fast-food restaurant are on the list. This is mainly area with minor commercial area which surrounded by neighborhood.

	Postal Code	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
79	1101	2.0	Photography Studio	Grocery Store	Yoga Studio	Food Court	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market	Fish & Chips Shop	Fast Food Restaurant

4.1.4 Cluster 4

Cluster 4 (Cluster label 3) contains only one neighborhood which is quite similar to cluster 3. This is mainly area with minor commercial area which surrounded by neighborhood.

	Postal Code	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
34	507	3.0	Playground	Yoga Studio	Ethiopian Restaurant	Food Court	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market	Fish & Chips Shop	Fast Food Restaurant

4.1.5 Cluster 5

Cluster 5 (Cluster label 4) contains only one neighborhood which is quite similar to clusters 3 and 4. This is another area with minor commercial area which surrounded by neighborhood.

	Postal Code	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
67	913	4.0	Hotel	Yoga Studio	Ethiopian Restaurant	Food Court	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market	Fish & Chips Shop	Fast Food Restaurant

4.1.6 Cluster 6

Cluster 6 (Cluster label 5) contains only one neighborhood which is quite similar to clusters 3, 4 and 5. This is again another region with minor commercial area which surrounded by neighborhood.

	Postal Code	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
89	1201	5.0	Insurance Office	Lawyer	Yoga Studio	Ethiopian Restaurant	Food Court	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market	Fish & Chips Shop

4.1.7 Cluster 7

Cluster 7 (Cluster label 6) is neighborhood where public travel rated top common venue, behind that are restaurants and yoga studios. These are mainly areas with family houses and apartments, a vibrant area which is a potential place to set up a business, where this area lacks of grocery stores.

	Postal Code	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
13	306	6.0	Playground	Bus Stop	Restaurant	Yoga Studio	Ethiopian Restaurant	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market	Fish & Chips Shop
31	504	6.0	Bus Stop	Supermarket	Restaurant	Eastern European Restaurant	Yoga Studio	Event Space	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market
56	902	6.0	Bus Stop	Supermarket	Light Rail Station	Yoga Studio	French Restaurant	Food Court	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market
88	1201	6.0	Bus Stop	Supermarket	Italian Restaurant	Restaurant	Yoga Studio	Ethiopian Restaurant	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market

4.1.8 Cluster 8

Cluster 8 (Cluster label 7) is neighborhood where we can see a variety of supermarket established in the cluster and plenty of food related venues (ice cream shop, Asian

restaurants, Italian restaurants, pizza places, etc..). There is also lots of commercial activities going around due to shopping malls, drugstore and movie theater for example.

	Postal Code	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
4	105	7.0	Supermarket	Pharmacy	Ice Cream Shop	Tennis Court	Track	Big Box Store	Gas Station	Park	Farmers Market	Exhibit
11	304	7.0	Supermarket	Fried Chicken Joint	Chinese Restaurant	Fish Market	Tram Station	Food Court	Food & Drink Shop	Flower Shop	Fishing Store	Fish & Chips Shop
12	305	7.0	Supermarket	Bus Stop	Skating Rink	Yoga Studio	Event Space	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market	Fish & Chips Shop
17	310	7.0	Supermarket	Garden Center	Drugstore	Organic Grocery	Yoga Studio	Event Space	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market
35	508	7.0	Supermarket	Snack Place	Drugstore	Yoga Studio	Ethiopian Restaurant	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market	Fish & Chips Shop
39	603	7.0	Supermarket	Drugstore	Movie Theater	Fast Food Restaurant	Park	German Restaurant	Bakery	Yoga Studio	Exhibit	Food & Drink Shop
43	607	7.0	Supermarket	Harbor / Marina	Indian Restaurant	Chinese Restaurant	Bank	Bakery	Austrian Restaurant	Liquor Store	Post Office	Farmers Market
47	704	7.0	Supermarket	Ice Cream Shop	Park	Steakhouse	Chinese Restaurant	Greek Restaurant	Diner	Exhibit	Food & Drink Shop	Dance Studio
53	804	7.0	Supermarket	Ice Cream Shop	Italian Restaurant	Drugstore	Bank	Diner	Falafel Restaurant	Fountain	Food Court	Deli / Bodega
57	903	7.0	Supermarket	Ice Cream Shop	Asian Restaurant	Event Space	Fountain	Food Court	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market
61	907	7.0	Supermarket	Pizza Place	Italian Restaurant	Steakhouse	Shopping Mall	Greek Restaurant	Trattoria/Osteria	Drugstore	Diner	Dessert Shop
62	908	7.0	Miscellaneous Shop	Supermarket	Insurance Office	Italian Restaurant	Event Space	Fountain	Food Court	Food & Drink Shop	Flower Shop	Fishing Store
73	1004	7.0	Supermarket	Hotel	Plaza	Park	Light Rail Station	Ethiopian Restaurant	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market
80	1101	7.0	Supermarket	Café	Italian Restaurant	German Restaurant	Seafood Restaurant	Yoga Studio	Event Space	Food & Drink Shop	Flower Shop	Fishing Store
82	1101	7.0	Supermarket	Movie Theater	Gym / Fitness Center	Train Station	Tram Station	Yoga Studio	Falafel Restaurant	Event Space	Exhibit	Fast Food Restaurant
84	1101	7.0	Supermarket	Italian Restaurant	Sushi Restaurant	Fishing Store	Drugstore	Construction & Landscaping	Café	Plaza	Park	Doner Restaurant
92	1201	7.0	Supermarket	Drugstore	Gym / Fitness Center	Italian Restaurant	Light Rail Station	Grocery Store	Fast Food Restaurant	Hardware Store	Shopping Mall	Warehouse Store
96	1201	7.0	Motorcycle Shop	Supermarket	Go Kart Track	Bakery	Yoga Studio	Event Space	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market

4.1.9 Cluster 9

Cluster 9 (Cluster label 8) contains only one neighborhood where service sector and gastronomy sector are the most common venue, but behind that, fish market and fast-food restaurant are on the list. This is mainly area with minor commercial area which surrounded by neighborhood.

	Postal Code	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
33	506	8.0	Bakery	Flower Shop	Ethiopian Restaurant	Fountain	Food Court	Food & Drink Shop	Fishing Store	Fish Market	Fish & Chips Shop	Fast Food Restaurant

4.1.10 Cluster 10

The cluster 10 (Cluster label 9) is a neighborhood which contains three neighborhoods. Here we see hostel and historic site at the top. It is possibly a tourist attraction region which is another ideal and interesting site to invest a new supermarket for tourists. Restaurants and public transport are at the top.

	Postal Code	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
75	1101	9.0	Hostel	Historic Site	Yoga Studio	Ethiopian Restaurant	Food Court	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market	Fish & Chips Shop
78	1101	9.0	Hostel	Historic Site	Yoga Studio	Ethiopian Restaurant	Food Court	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market	Fish & Chips Shop
86	1201	9.0	Hostel	Insurance Office	Nature Preserve	Liquor Store	Yoga Studio	Event Space	Food Court	Food & Drink Shop	Flower Shop	Fishing Store

4.1.11 Cluster 11

Cluster 11 (Cluster label 10) contains two neighborhoods where service sector and gastronomy sector and football field are the most common venue, but behind that, fast-food restaurant is on the list. These are mainly areas with minor commercial area which surrounded by neighborhood.

	Postal Code	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
49	706	10.0	Soccer Field	Supermarket	Bus Stop	Doner Restaurant	Miscellaneous Shop	Yoga Studio	Exhibit	Food Court	Food & Drink Shop	Flower Shop
81	1101	10.0	Indian Restaurant	Supermarket	Soccer Field	German Restaurant	Fountain	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market	Fish & Chips Shop

4.1.12 Cluster 12

Cluster 12 (Cluster label 11) contains only one neighborhood which is quite similar to clusters 3, 4 and 5. This is again another location with minor commercial area which surrounded by neighborhood.

	Postal Code	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
60	906	11.0	Home Service	Yoga Studio	Ethiopian Restaurant	Food Court	Food & Drink Shop	Flower Shop	Fishing Store	Fish Market	Fish & Chips Shop	Fast Food Restaurant

5. Discussion

Based on that, we can advise the stakeholders to consider the neighborhoods from clusters the potential location for the new supermarket. Clusters 1 and 7 for a family-friendly supermarket that suitable for families to do grocery shopping. Cluster 10 for the supermarket that targets visitors and tourists, which is equipped with express counters and compact interior design, that speeds up the purchase experience. Cluster 2 and 8 for a supermarket with luxury facilities and high-class ingredients that aiming the hype customers, which can compete with other supermarket brands in this cluster. These satisfy the original criteria that the set-up location should be easily accessible for local citizens and tourists and is possibly occupied with high traffic flow.

6. Conclusion

In this study, the process of finding an answer for the simulated business problem is discussed. The experiment was performed based on the data science tools and the use of Python libraries such as Pandas, Numpy, Scikit-learn, Folium, Geocoder, and others. Data were collected from only one source and in different formats. For clustering 97 neighborhoods in Berlin, a machine learning technique was used. The outcome of the results provided insight for drafting several recommendations to the mentioned business problem.

7. References

The Jupyter notebook can be found on GitHub:

Cynthia, E. (2021) *50 Best Supermarket Business ideas You Can Start in 2021*.

Available at: https://www.profitableventure.com/supermarket-business-ideas/ (Accessed: 2 June 2021).

Dabbura, I. (2018) *K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks*. Available at: https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a (Accessed: 2 June 2021).