Applied Data Science Capstone
by IBM and COURSERA

Finding Out the Cities of Niedersachsen

Muhammed Alperen Bingöl

18.10.2020

1. Introduction: Business Problem

In this project we will try to find optimal locations for the investments. Specifically, this report will be targeted to stakeholders interested in opening stores in **Niedersachsen (Lower Saxony)**, Germany. Niedersachsen is a Land (German state) in northwestern Germany. It is the second-largest state by land area, and fourth largest in population (7.9 million) among the 16 Länder federated as the Federal Republic of Germany.

Since there are lots of information about big cities, we will try to detect **locations that are not already crowded with shops**. We are also particularly interested in **rural areas**. We would also preferlocations **with middle level population**.

We will use our data science powers to generate more information about promising cities based on these criteria. Advantages of each area will then be clearly expressed so that best possible locations can be chosen by stakeholders.

2. Data Acquisition and Cleaning

1. Data sources and Data Cleaning

Based on definition of our problem, factors that will influence our decision are:

- number and types of existing venues in the city (any type of venue)
- number of and distance to big cities in the vicinity if any
- population density of area

We decided to use locations, centered around city center, to define our area.

Following data sources will be needed to extract/generate the required information:

- Approximate addresses of centers of those areas will be obtained using **geopy.geocoders**
- Number of venues, their type and location in every city will be obtained using **Foursquare API**
- List of cities and their populations of Niedersachsen state will be obtained using **www.citypopulation.de**

Geographical Location data for all cities are obtained through geopy.geocoders. Detailed information about venues in all cities will be examined. And we will direct our study based on this information. Furthermore, there have been doubts about the correctness of the positions of some cities. There were cities in other states with the same name, even in other countries. Then they have been cleared from dataset.

2. Feature selection

839 cities (settlement unit) have been identified in the state of Niedersachsen. City refers to all kinds of settlements. Some of these cities are the size of large cities. However, Niedersachsen is generally a low population density state. Therefore, Niedersachsen usually has small settlements.

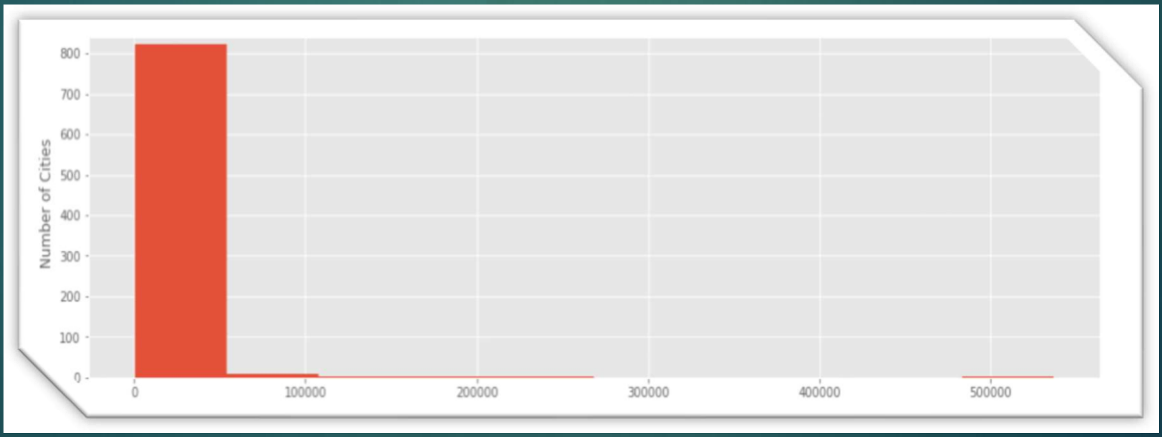
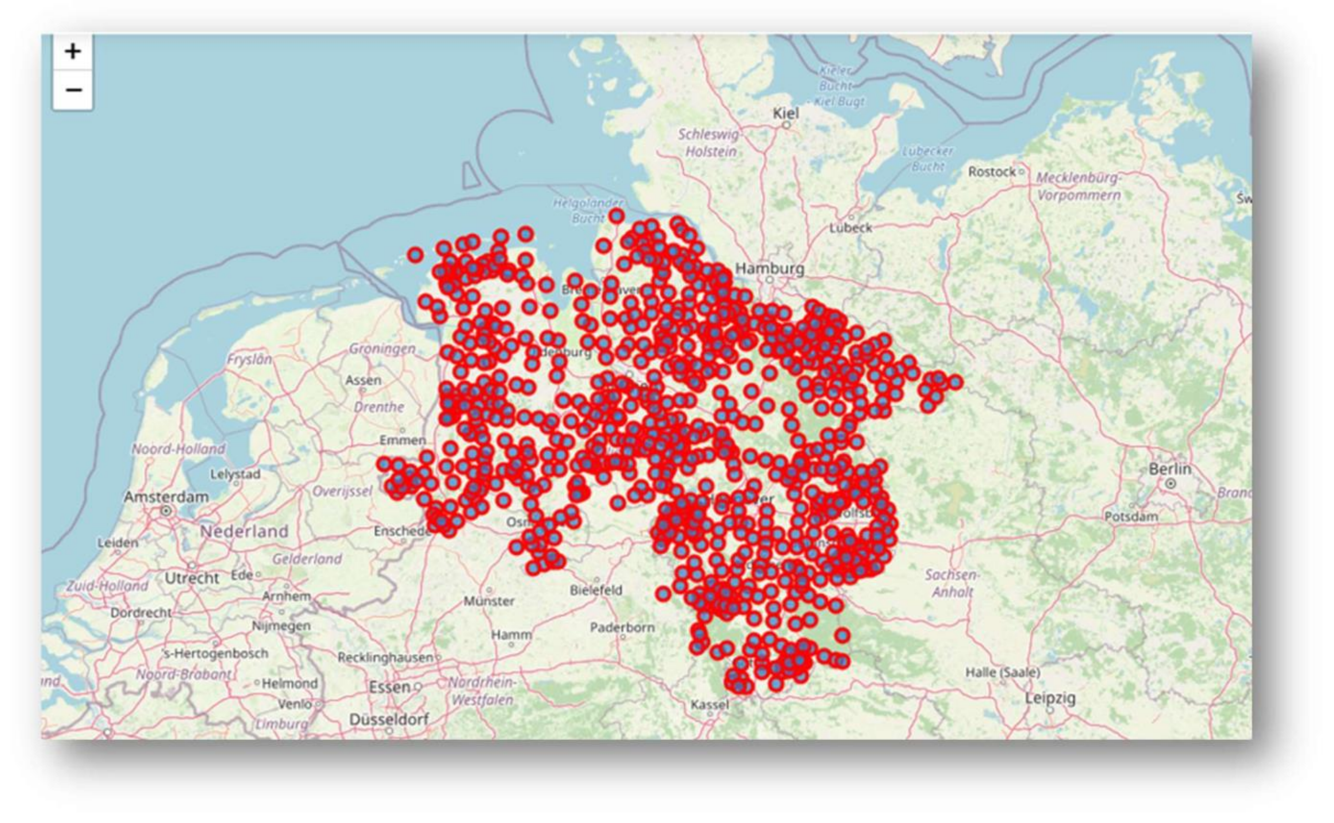


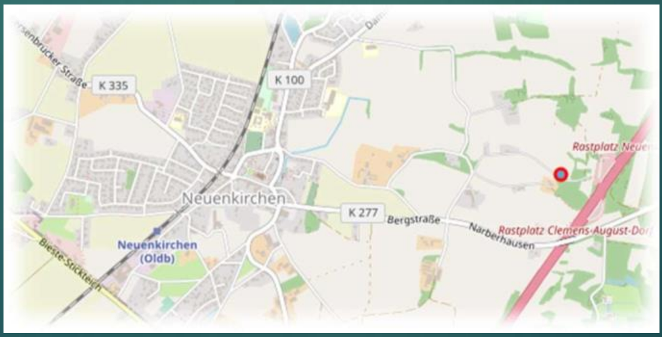
Figure 1. Histogram of the population of cities in Niedersachsen province.

In this histogram, most of the cities are small. In other words, the population of the cities is usually less than 50000.



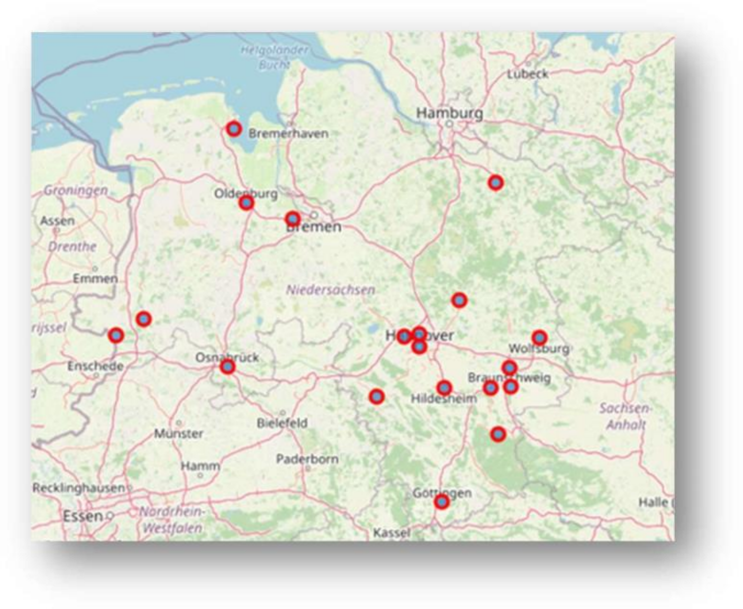
Map 1. All small or large cities in Niedersachsen province

All cities are mapped. This includes very small settlements. So, we can get an overview of the state of Niedersachsen. However, when we examine the map closely, it is seen that the city centers are not located at the center of the residential areas. This is especially valid for large populated areas. The information we will obtain through Foursquare is based on coordinates. Information of the venues can be only obtained around these coordinates. When these coordinates are at the wrong point of the cities, we cannot get the correct information. Therefore, in our study, not large, but medium-sized cities will be examined. These kinds of cities are already intended to be made in investments.



Map 2. A place where the real center of the city and the Geopy data does not overlap

In this map down we see the major cities in Niedersachsen. The locations of the cities on this map correspond to the population density map. It is a useful map in terms of controlling the location information we have acquired ourselves.



Map 2. Areas where population density is very high

217 medium-sized cities of Niedersachsen have been identified. These cities have a population of between 5,000 and 20,000. We already have the location and population information of these cities.

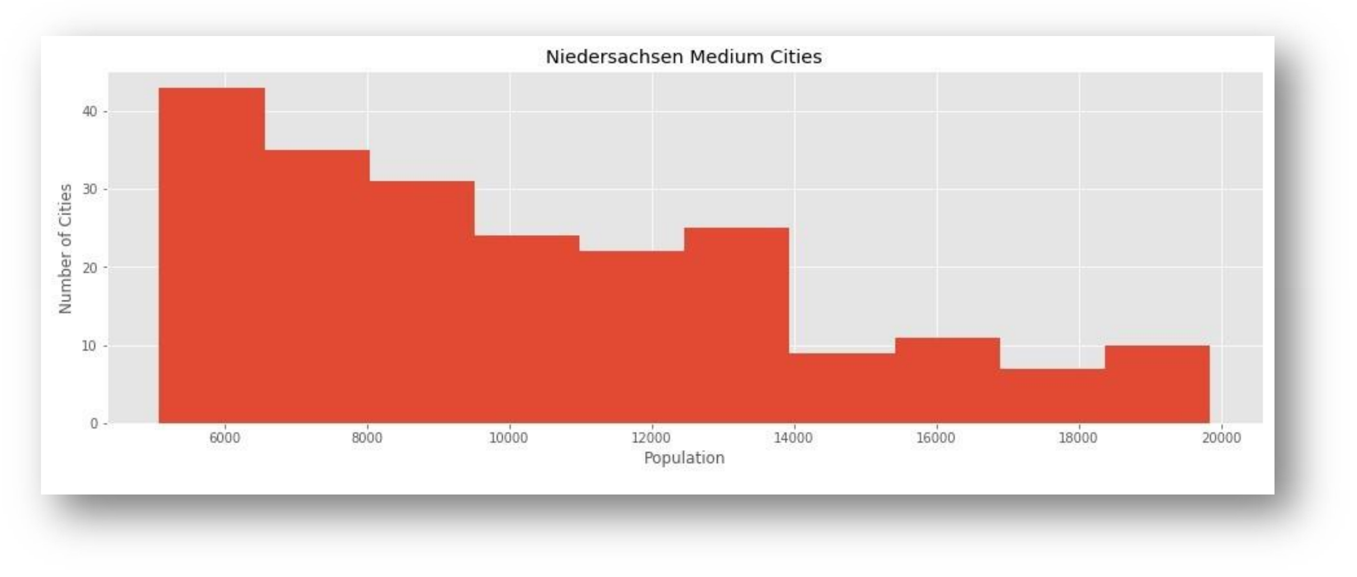
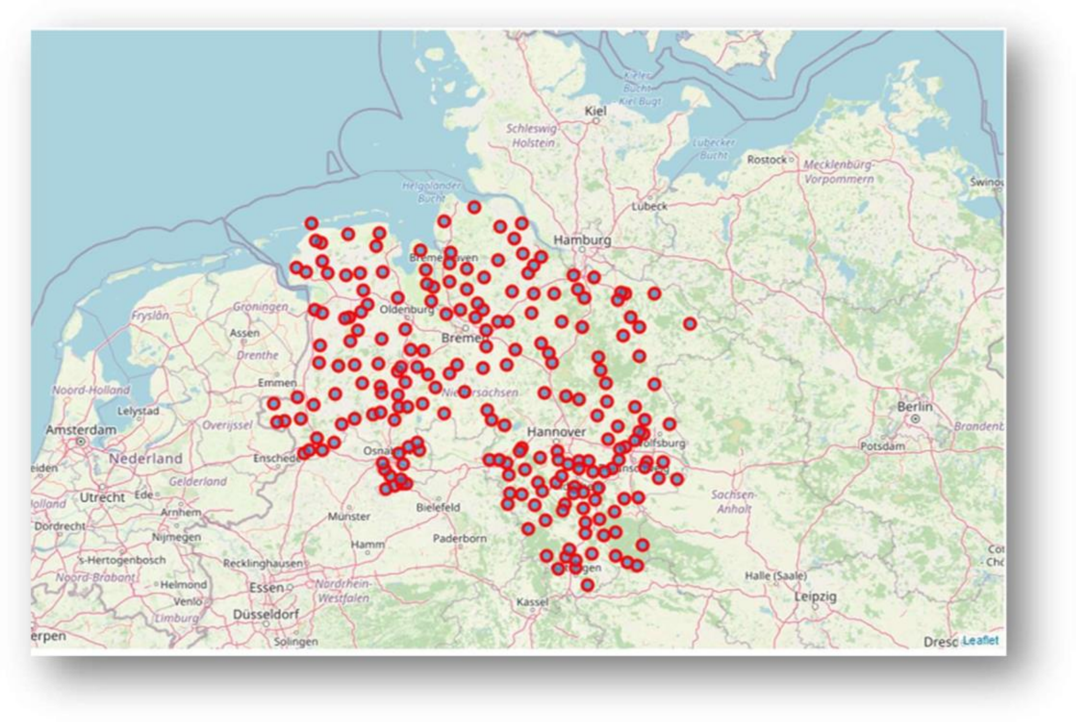


Figure 2. Population distribution of medium-sized cities

As can be seen, the cities we are studying have usually a population of between 5000 and 15000. For example, about 150 cities have a population between 5000 and 15000. The remaining 70 cities have a population of between 15000 and 20000. However, cities with a population of between 15000 and 20000 still have a significant number. In this study, all these cities with medium-sized population will be examined. Thus, we have used our Niedersachsen demographic information in our study.

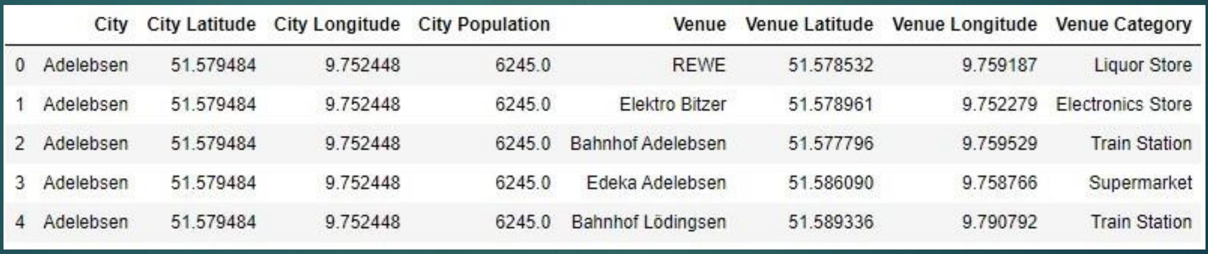


Map 3. Cities to be covered by the study

It is clear from this map that we can get an idea of the whole state of Niedersachsen by examining the medium-sized cities. As it seems, almost the entire state of Niedersachsen is covered. However, it is striking that the surroundings of large cities appear empty on the map. Because there are densely populated settlements around big cities.

In this study, shops, namely venues, in cities with medium-sized population are examined. It is desired to reach the information about what kind of businesses are there in these cities. We will access this information via Foursquare.

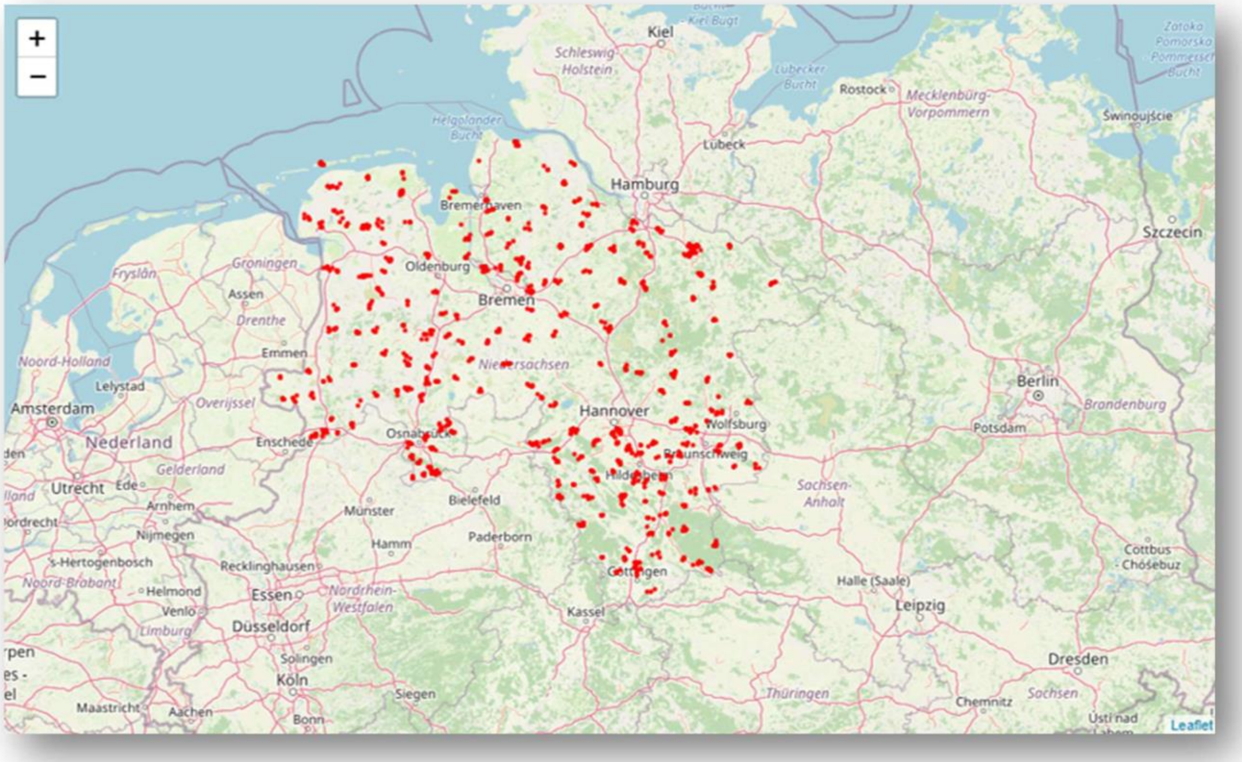
Table 1.A section of venue data from Foursquare



	City	City Latitude	City Longitude	City Population	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Adelebsen	51.579484	9.752448	6245.0	REWE	51.578532	9.759187	Liquor Store
1	Adelebsen	51.579484	9.752448	6245.0	Elektro Bitzer	51.578961	9.752279	Electronics Store
2	Adelebsen	51.579484	9.752448	6245.0	Bahnhof Adelebsen	51.577796	9.759529	Train Station
3	Adelebsen	51.579484	9.752448	6245.0	Edeka Adelebsen	51.586090	9.758766	Supermarket
4	Adelebsen	51.579484	9.752448	6245.0	Bahnhof Lödingsen	51.589336	9.790792	Train Station

Thus, we learned about 1398 Venue. We now know their locations, the cities they are in, the locations of these cities and the populations of these cities. This information constitutes the basis of our work. In this study, a conclusion will be made based on the information.

We have 1398 Venues in all middle cities in Niedersachsen (in 3 km of the location points)



Map 4. All venues in medium-sized cities

3. Methodology

In this project we will direct our efforts on understanding rural areas of Niedersachsen that have low venue density, particularly those with medium number of populations. We will limit our analysis to area ~3km around city center. Considering the target settlements, the city will be covered at a significant rate with this 3 km distance. To be clear, these criteria have been created by taking into account the possibilities offered by Foursquare. Because more than 100 venue information could not be obtained around a point. Therefore, residential units where the number of venues does not exceed 100 were targeted. Therefore, scanning an area of 3 km radius is enough to cover medium-sized cities.

In first step we have collected the required **data:** **location and Population** **Data for each city then** **type (category) of every venues** **within 3km from each city center** (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of **'venue density'** across different areas of Niedersachsen - we will use **Scatter Plots and Histograms** to identify a few promising areas close to center with high number of venues in general.

In third and final step we will create **clusters of locations**: we will take into consideration locations with **different kinds of Venue Category**, and we want locations **with high population**. We will present map of all such locations but also create clusters (using **k-means clustering**) of those locations to identify zones / cities and search for optimal location determination by stakeholders.

4. Exploratory Data Analysis

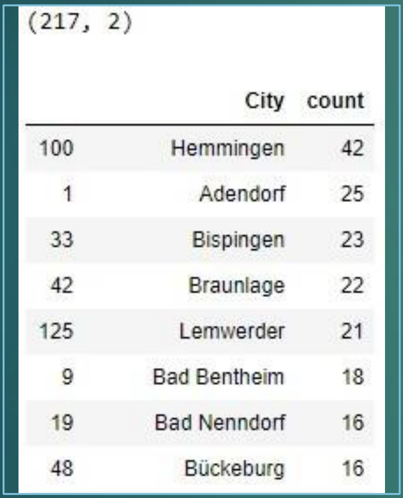
1. Calculation of targetvariable

We perform some basic explanatory data analysis and derive some additional info from our raw data.

4.2 Relationship between the Venue Counts and Population

The Count of Venues for each Middle-City in Niedersachsen

Table 2. A section of Venue Data from Foursquare (the original table consists of 217 lines)



```
(217, 2)
```

	City	count
100	Hemmingen	42
1	Adendorf	25
33	Bispingen	23
42	Braunlage	22
125	Lemwerder	21
9	Bad Bentheim	18
19	Bad Nenndorf	16
48	Bückeburg	16

We now know the Venue Numbers for each City. For example, there are 42 Venue in Hemmingen. Hemmingen is a highly populated city. Therefore, the number of Venue is also very high. I wonder if this applies to all cities. So, does a city's population increase in proportion to its Venue Number? The best way to find the answer to this question is to compare the population of each city with the number of Venues. It is convenient to do this comparison with line plot. For this, of course, applying normalization to the population column will give a better result. We see the relationship between Population and Venue Counts.

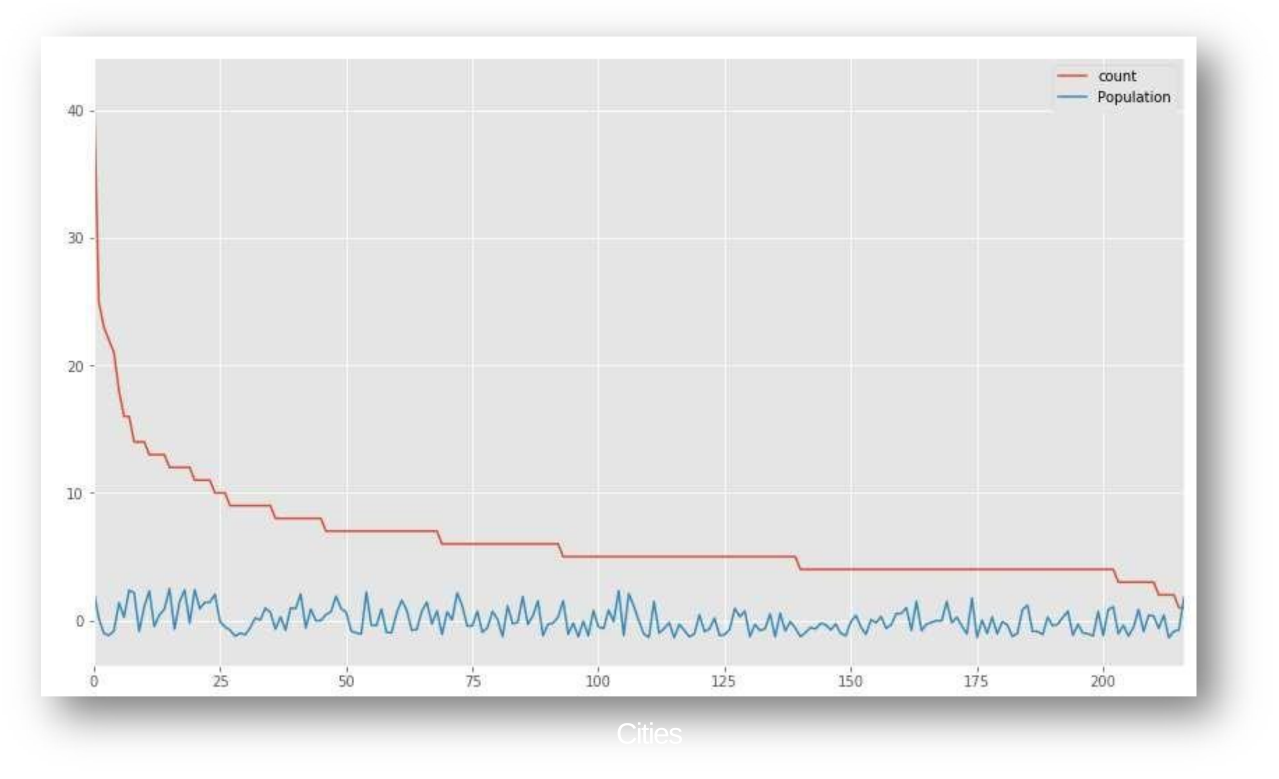


Figure 3: Venue Count and Population for 217 Cities.

It appears that there is no connection between Venue Counts and Population in the Line Plot. In other words, with the decrease in the number of Venues, the population number does not decrease regularly. Based on the data we have; we cannot establish a relationship between the population and the Venue Count. An important reason for this is that, in my opinion, the information received from Foursquare is not at a perfect level. Normally, as the population increases, the number of Venue there should also increase. Another reason may be the accuracy of our locations, that is, our latitude and longitude information. The latitude and longitude information specified in the information in the hand does not always indicate the center of the city in terms of population density. Latitude Longitude information sometimes shows the geographical center of the city, not population center. Therefore, this point can sometimes be a location in the middle of agricultural lands. Therefore, information received from Foursquare may not show the reality enough. However, we will ignore this condition in this study. Because our primary goal is to follow a certain methodology. We are going to see the result together.

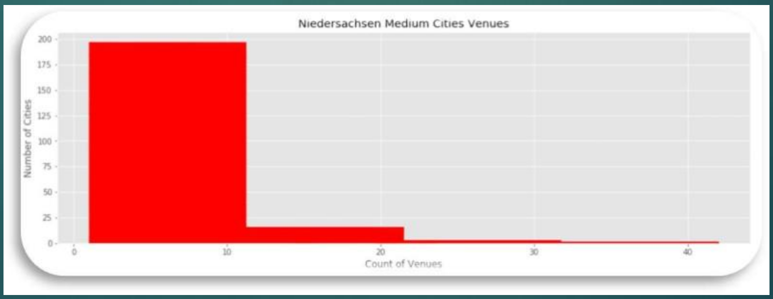


Figure 4: Histogram for Venue Counts.

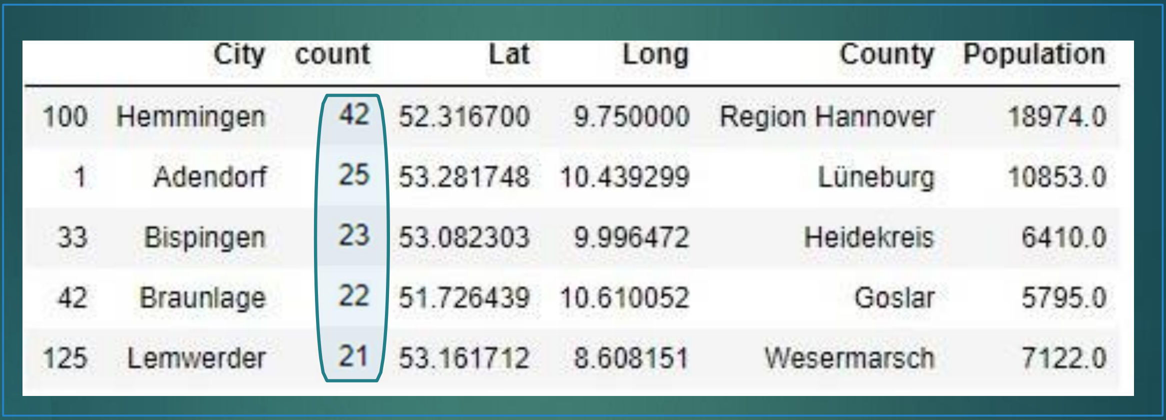
As seen in the histogram plot, there are less than 10 venues in most cities. The biggest reason for this is that the cities we studied are generally settlements with small population density. Our aim is to invest exactly in like these regions. Therefore, the obtained Venue Number data does not logically completely conflict with the population data. Conclusively, it should be kept in mind that there is no linear regression between the population and Venue numbers.

4.3 Clustering Test based on Venue Count and Population

We have a new Data Frame that shows the Venue Counts and Population

Maybe we can cluster Cities based on Population or Venue Counts. We know that there is no linear relationship between population and Venue Count. This does not mean that there is no link between them. So, we will try to make a clustering based on these two pieces of information. Our aim is to see if a meaningful photo will come out.

Table 3. The Total Venue Counts for each City (the original table consists of 217 lines)



	City	count	Lat	Long	County	Population
100	Hemmingen	42	52.316700	9.750000	Region Hannover	18974.0
1	Adendorf	25	53.281748	10.439299	Lüneburg	10853.0
33	Bispingen	23	53.082303	9.996472	Heidekreis	6410.0
42	Braunlage	22	51.726439	10.610052	Goslar	5795.0
125	Lemwerder	21	53.161712	8.608151	Wesermarsch	7122.0

And now, we have 4 type of City "Based on Venue Count and Populations"

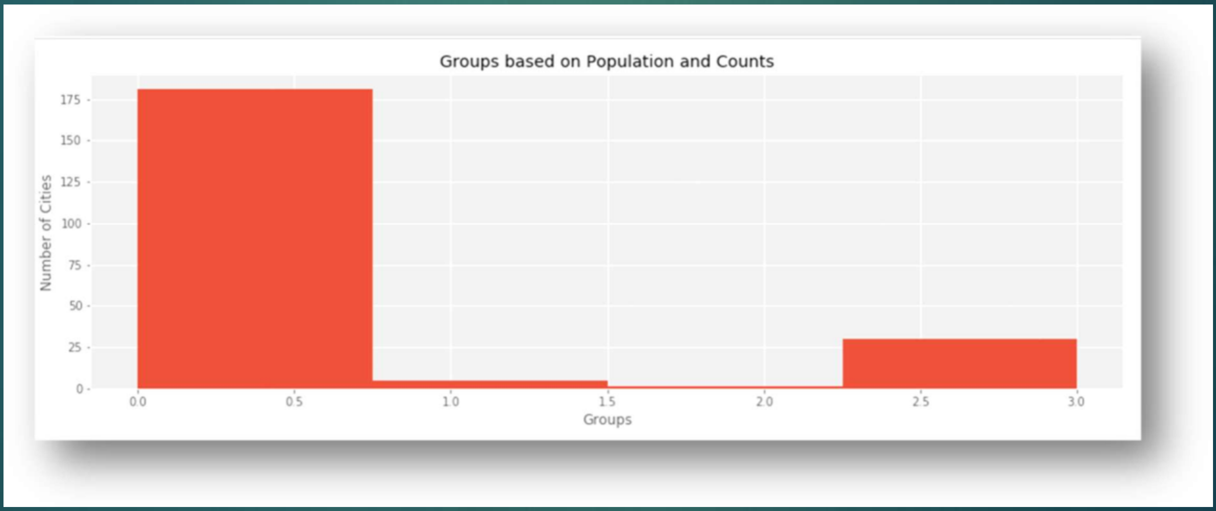


Figure 5: Histogram for Groups based on Population und Venue Count.

As seen in the histogram, no meaningful grouping occurred. Therefore, the population information we have is not suitable to be used for the clustering of cities in Niedersachsen. At least we can say this according to the information we have. Whether this reflects the truth is not clear. To understand this, we must firmly trust the accuracy of the Venue number in each city. Therefore, in our study, we operate with the information we have. Given the information we have, population information is not useful in clustering cities in terms of Venue. We can confirm this once with a scatter plot.

We see Populations and Venue Countstogether

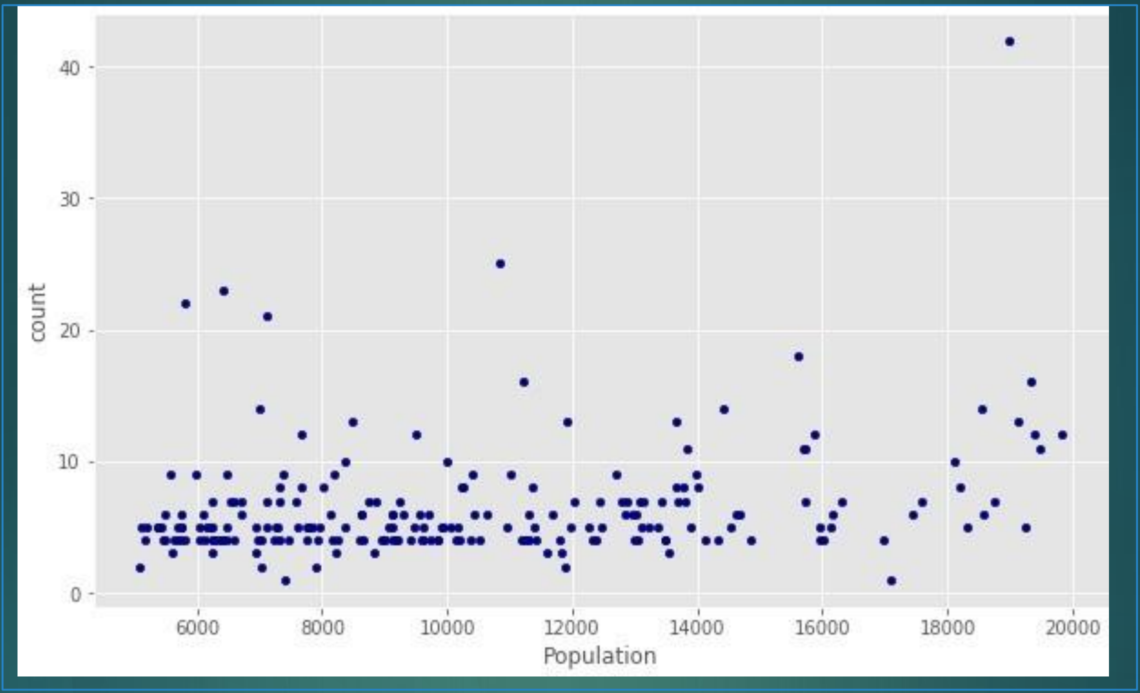


Figure 6: Scatter Plot for Population and Venue Count of each City.

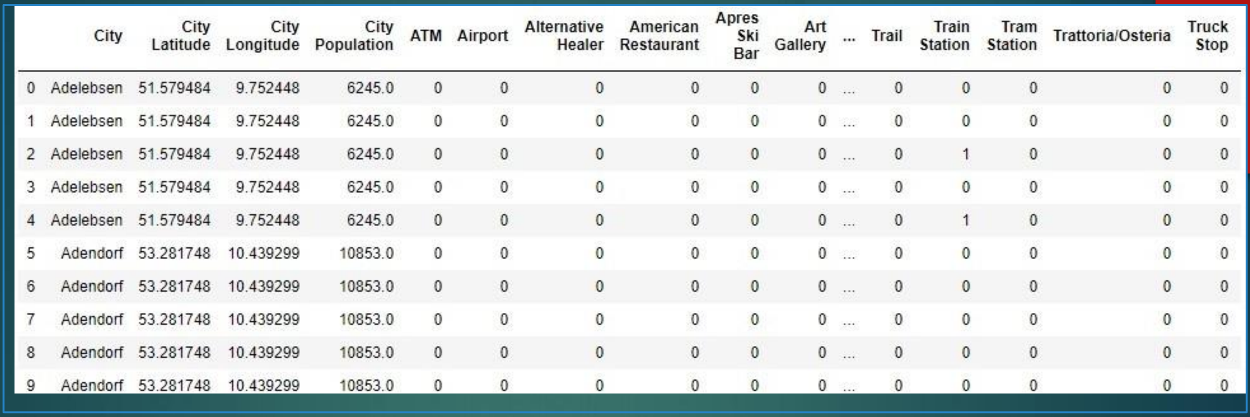
The situation that appears in the Scatter Plot also supports our conclusion. In other words, the distinguishing features of cities are not the number of venues data and population data we have. So, we have to follow another method in order to determine the distinctive features of cities. At this point, we will take a look at the category information of the Venues we have.

4.4 Exploring the Categories of Venues: There are 184 unique categories.

We have around 1400 venue data for more than 200 cities in Niedersachsen. We examine the category data of these 1400 venues. The category of the venues that we have knowledge of is 184. So, there are 184 categories. Drugstore, Nightclub, Pet store, Historic Site, Bakery, French Restaurant and Tea Room are some examples of them.

We see here Categories of all 1398 Venues: We create a new data frame to see how much venue in which city and which category.

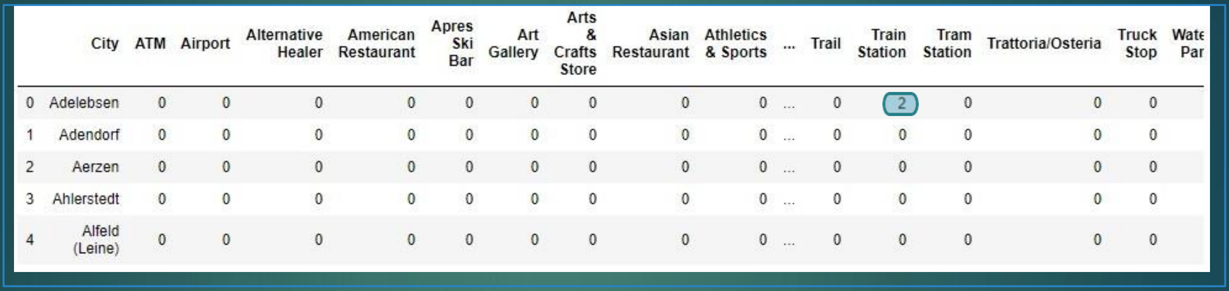
Table 4. Venues and Categories (the original table consists of 1398 lines and 188 Columns)



	City	City Latitude	City Longitude	City Population	ATM	Airport	Alternative Healer	American Restaurant	Apres Ski Bar	Art Gallery	...	Trail	Train Station	Tram Station	Trattoria/Osteria	Truck Stop
0	Adelebsen	51.579484	9.752448	6245.0	0	0	0	0	0	0	...	0	0	0	0	0
1	Adelebsen	51.579484	9.752448	6245.0	0	0	0	0	0	0	...	0	0	0	0	0
2	Adelebsen	51.579484	9.752448	6245.0	0	0	0	0	0	0	...	0	1	0	0	0
3	Adelebsen	51.579484	9.752448	6245.0	0	0	0	0	0	0	...	0	0	0	0	0
4	Adelebsen	51.579484	9.752448	6245.0	0	0	0	0	0	0	...	0	1	0	0	0
5	Adendorf	53.281748	10.439299	10853.0	0	0	0	0	0	0	...	0	0	0	0	0
6	Adendorf	53.281748	10.439299	10853.0	0	0	0	0	0	0	...	0	0	0	0	0
7	Adendorf	53.281748	10.439299	10853.0	0	0	0	0	0	0	...	0	0	0	0	0
8	Adendorf	53.281748	10.439299	10853.0	0	0	0	0	0	0	...	0	0	0	0	0
9	Adendorf	53.281748	10.439299	10853.0	0	0	0	0	0	0	...	0	0	0	0	0

We see Cities and their Venue Counts based on their Categories:

Table 5. Cities and Venues Categories (the original table consists of 217 lines and 185 Columns)

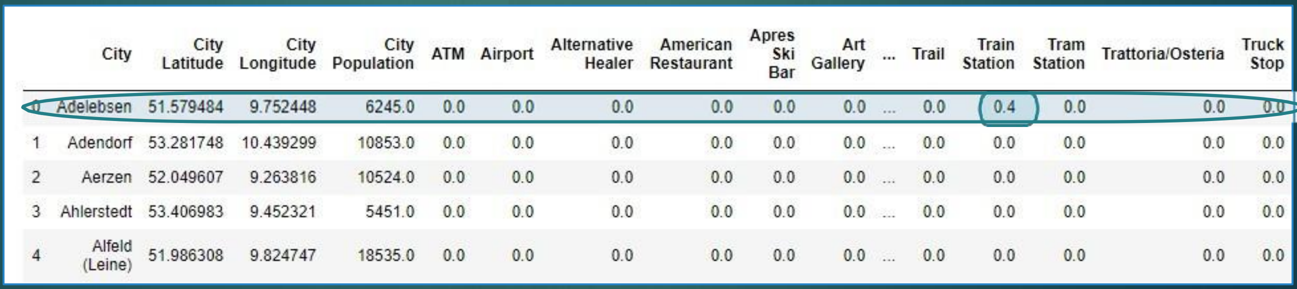


	City	ATM	Airport	Alternative Healer	American Restaurant	Apres Ski Bar	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	...	Trail	Train Station	Tram Station	Trattoria/Osteria	Truck Stop	Wate Par
0	Adelebsen	0	0	0	0	0	0	0	0	0	...	0	2	0	0	0	
1	Adendorf	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
2	Aerzen	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
3	Ahlerstedt	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
4	Alfeld (Leine)	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	

Yes, in this data frame, we see how many of which kind of venue there are in which city. For example, there are 2 train stations in Adelebsen. But there is no Asian Restaurant there. There is a train station in Bad Fallingbostel. Yes, I have been at Bad Fallinbostel for a while and I personally confirm this information.

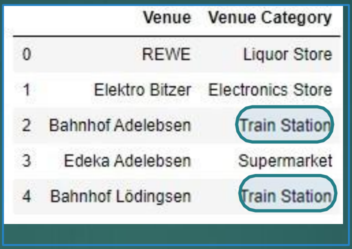
However, we noticed that the Venue Count data was not satisfactory enough as a distinguishing feature in the previous operations. So, we need new and more useful information here. This information is the venue rates. Which venue is there at what rate? Accordingly, the general characteristics of the cities can be understood. And a grouping can be made based on this information. We will make our investment decisions based on this grouping.

Table 6. Cities and Venue Rates (the original table consists of 217 lines and 188 Columns)



	City	City Latitude	City Longitude	City Population	ATM	Airport	Alternative Healer	American Restaurant	Apres Ski Bar	Art Gallery	...	Trail	Train Station	Tram Station	Trattoria/Osteria	Truck Stop
0	Adelebsen	51.579484	9.752448	6245.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.4	0.0	0.0	0.0
1	Adendorf	53.281748	10.439299	10853.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
2	Aerzen	52.049607	9.263816	10524.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
3	Ahlerstedt	53.406983	9.452321	5451.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
4	Alfeld (Leine)	51.986308	9.824747	18535.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0

Table 7. The Venues of Adelebsen



	Venue	Venue Category
0	REWE	Liquor Store
1	Elektro Bitzer	Electronics Store
2	Bahnhof Adelebsen	Train Station
3	Edeka Adelebsen	Supermarket
4	Bahnhof Lödingsen	Train Station

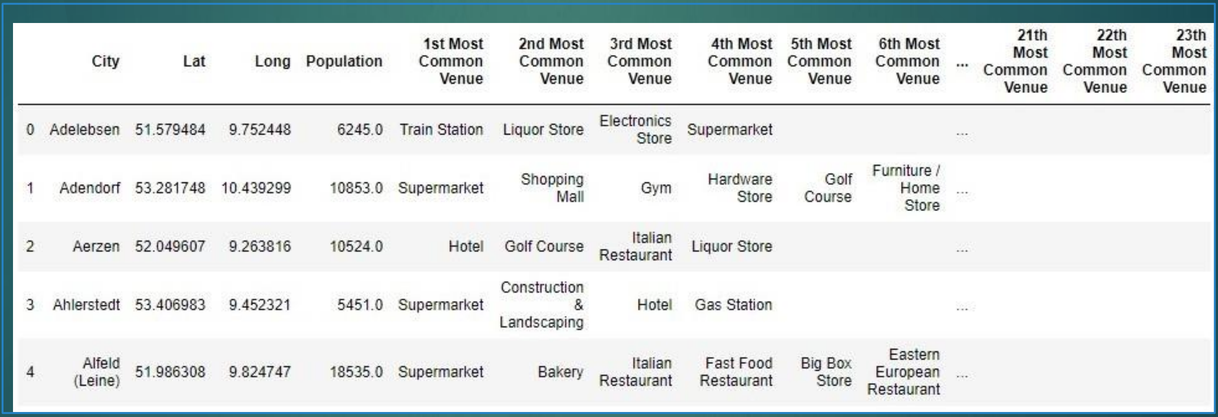
Looking at this new table, we see that the ratio of Train stations in Adelebsen to other venues in Adelebsen is 40%. In a city with two train stations, it is expected that there will be a lot of venues. However, we know that this is not the case here in reality. Normally, there are many venues in Adelebsen. However, The Foursquare informs us that this venue is. According to the information we have, there are a total of 5 venues in the city, but, there are at least 30. However, we base our work on Foursquare information. And we will conclude accordingly. Therefore, we are trying to apply the right method. However, it seems difficult for us to come to an exact grouping conclusion. As a result, we will continue our analysis with a Kmeans algorithm based on category data.

%Modeling: Clustering

5.1. Most Common Venues

We list the popular venues before clustering because that gives us an idea about the cities.

Table 8. The Frequency Order of Venue Types (the original table: 217 lines and 34 Columns)



	City	Lat	Long	Population	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	...	21th Most Common Venue	22th Most Common Venue	23th Most Common Venue
0	Adelebsen	51.579484	9.752448	6245.0	Train Station	Liquor Store	Electronics Store	Supermarket			...			
1	Adendorf	53.281748	10.439299	10853.0	Supermarket	Shopping Mall	Gym	Hardware Store	Golf Course	Furniture / Home Store	...			
2	Aerzen	52.049607	9.263816	10524.0	Hotel	Golf Course	Italian Restaurant	Liquor Store			...			
3	Ahlerstedt	53.406983	9.452321	5451.0	Supermarket	Construction & Landscaping	Hotel	Gas Station			...			
4	Alfeld (Leine)	51.986308	9.824747	18535.0	Supermarket	Bakery	Italian Restaurant	Fast Food Restaurant	Big Box Store	Eastern European Restaurant	...			

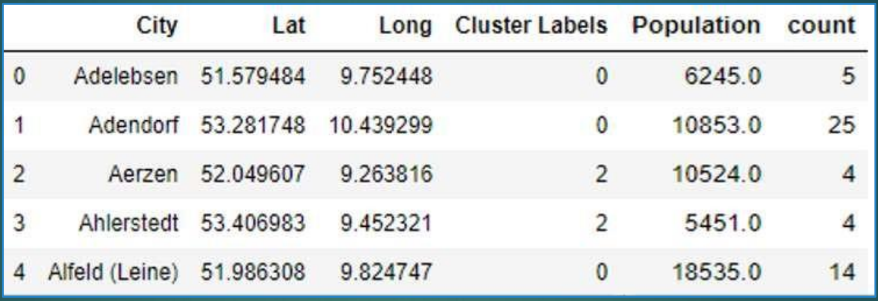
The popular venue data for each city appears in the list. The Venues of only 25 cities have more than 10 categories. Most of them have less than 10 categories. This Dataset is good for understanding the characteristic features of cities. However, this is not very important as our grouping process is not based on this dataset. We did not use this table for clustering. We put it here just to see the general situation. However, word clouds will be produced based on these data.

5.2. Clustering

We will cluster the cities based on the rate of the venues in each city. Relationship between Clusters and Venue Counts and Populations: We are going to see it via 1Histograms and 2 Scatter Plots

For that we need a new Data Frame like this:

Table 9. Clusters, Population and Venue Count for each City (the original table: 217 lines)



	City	Lat	Long	Cluster Labels	Population	count
0	Adelebsen	51.579484	9.752448	0	6245.0	5
1	Adendorf	53.281748	10.439299	0	10853.0	25
2	Aerzen	52.049607	9.263816	2	10524.0	4
3	Ahlerstedt	53.406983	9.452321	2	5451.0	4
4	Alfeld (Leine)	51.986308	9.824747	0	18535.0	14

5.2.1. The Distribution of Cities:

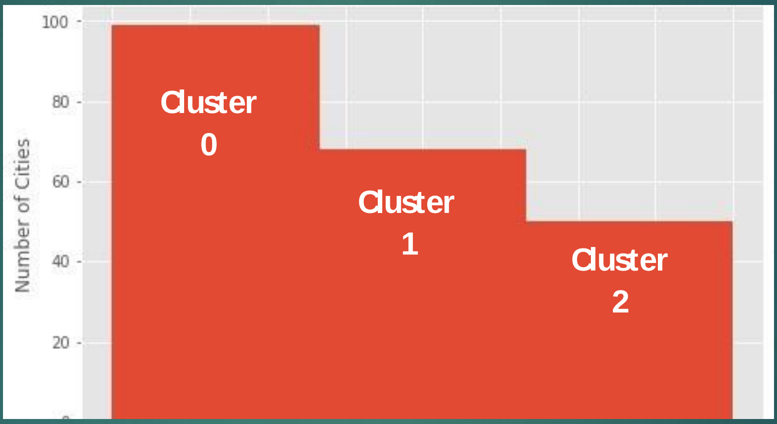


Figure 7: Histogram for Clusters.

As a result of clustering based on venue ratios, groups of 99, 65 and 50 cities have been formed.

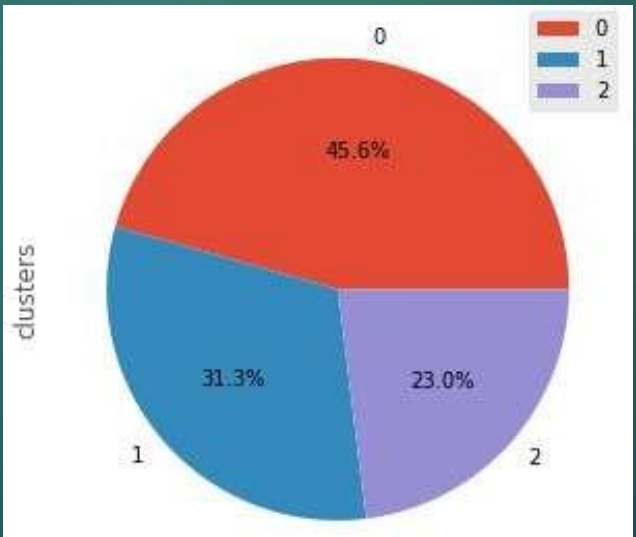


Figure 8: Cluster 0, cluster 1and cluster 2.

4.2.2 Relationship between the Clusters and Population:

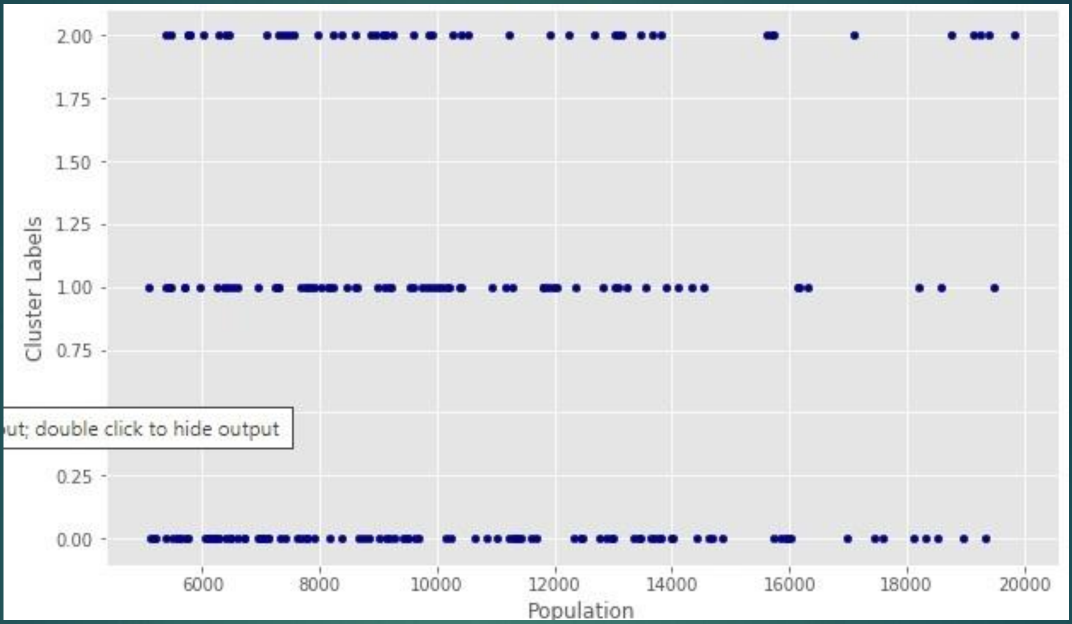


Figure 9: Scatter Plot for clusters and Population (each point is a city).

4.2.3 Relationship between the Clusters and Venue Count:

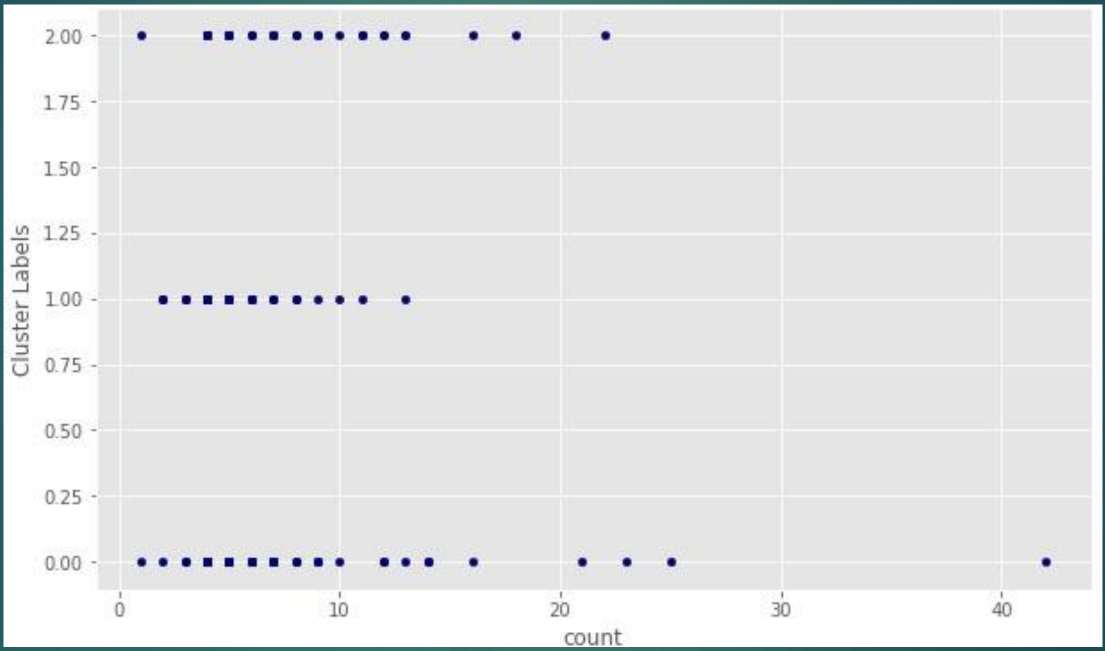
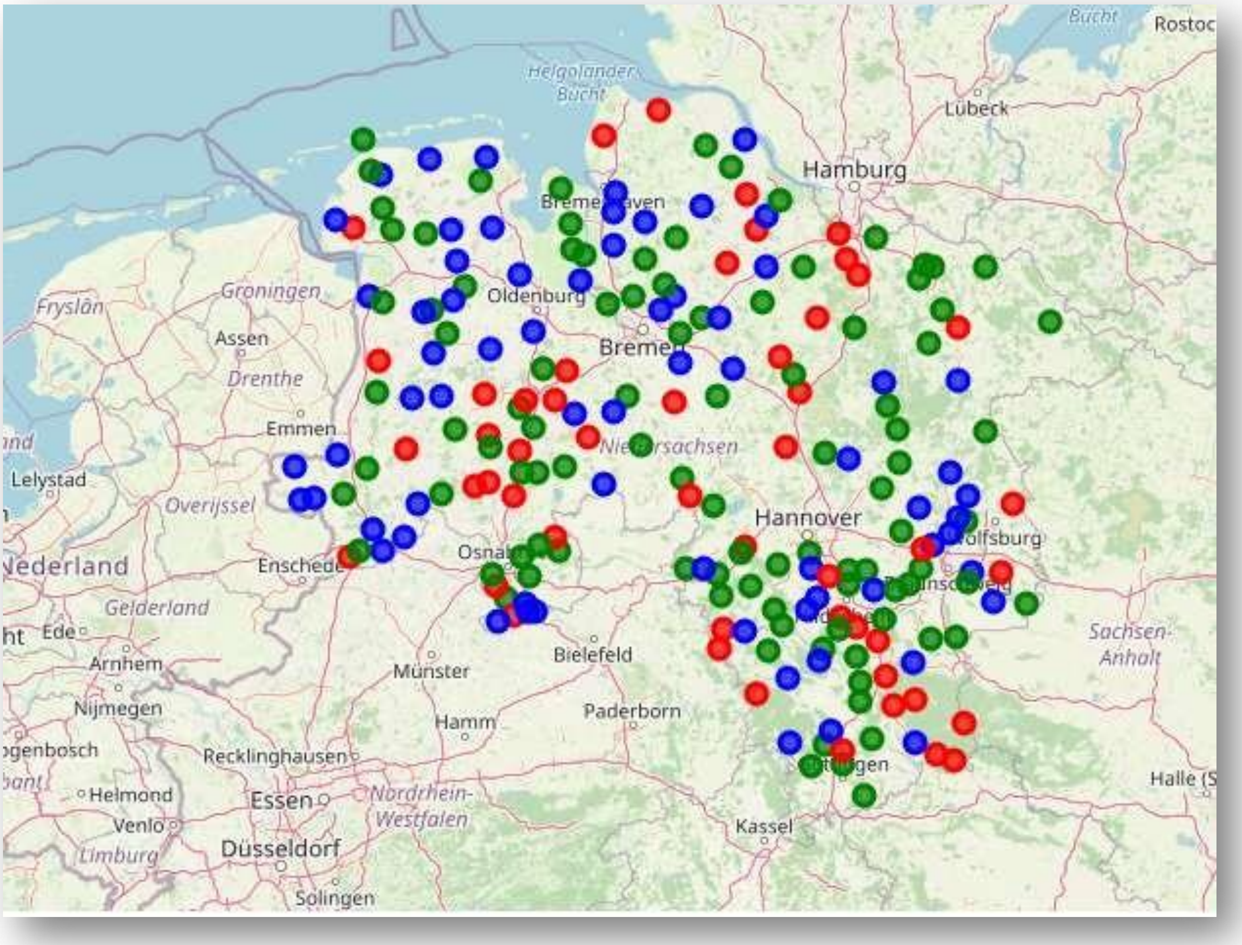


Figure 10: Scatter Plot for clusters and Venue Counts (each point is a city).

As a result, there was no significant correlation between population and venue counts and clustering. Now we see the clusters of cities on the map. Perhaps the geographical features of the cities may have influenced the grouping.

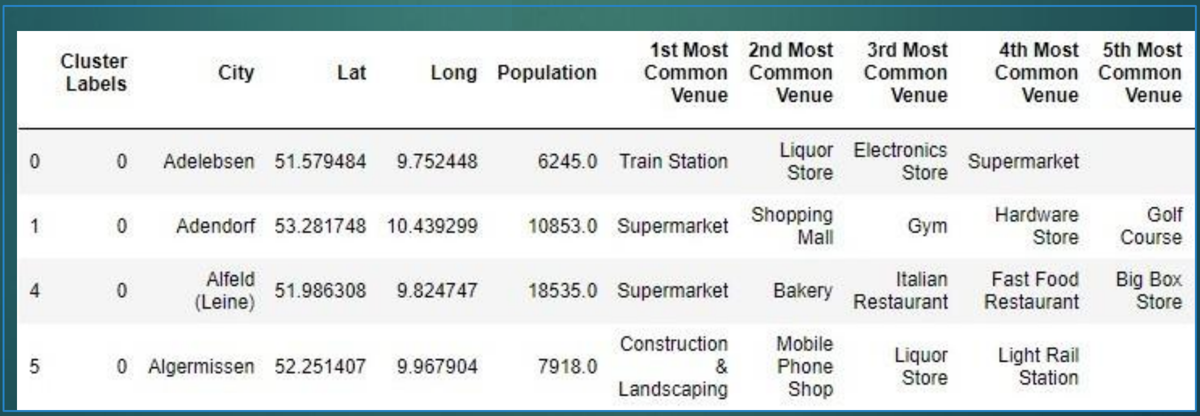


Map 4. Clusters on Map (Green: Cluster 0, Blue: Cluster 1, Red: Cluster 2)

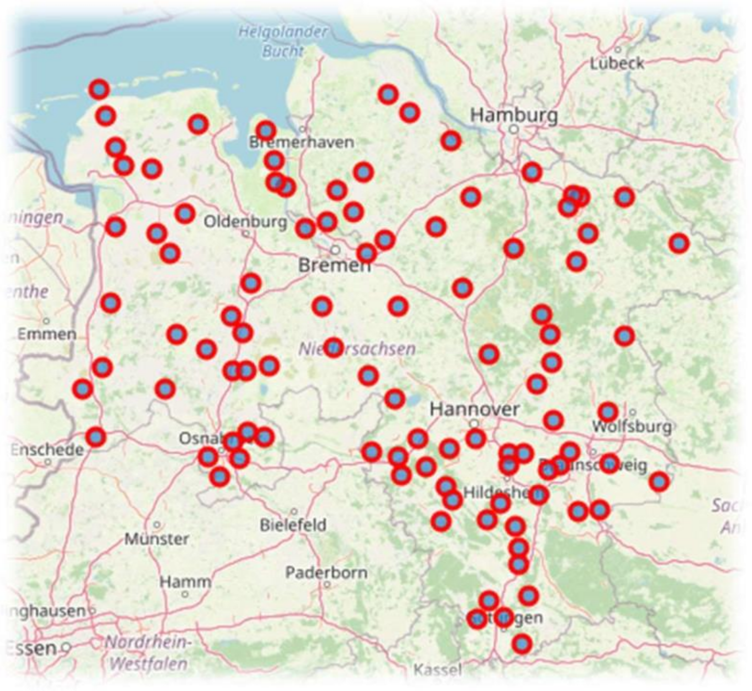
3. Now we see our City Groups

1. Cluster 0

Table 9. Cluster 0 (the original table: 99 lines)



	Cluster Labels	City	Lat	Long	Population	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	0	Adelebsen	51.579484	9.752448	6245.0	Train Station	Liquor Store	Electronics Store	Supermarket	
1	0	Adendorf	53.281748	10.439299	10853.0	Supermarket	Shopping Mall	Gym	Hardware Store	Golf Course
4	0	Alfeld (Leine)	51.986308	9.824747	18535.0	Supermarket	Bakery	Italian Restaurant	Fast Food Restaurant	Big Box Store
5	0	Algermissen	52.251407	9.967904	7918.0	Construction & Landscaping	Mobile Phone Shop	Liquor Store	Light Rail Station	



Map 5. Cluster0

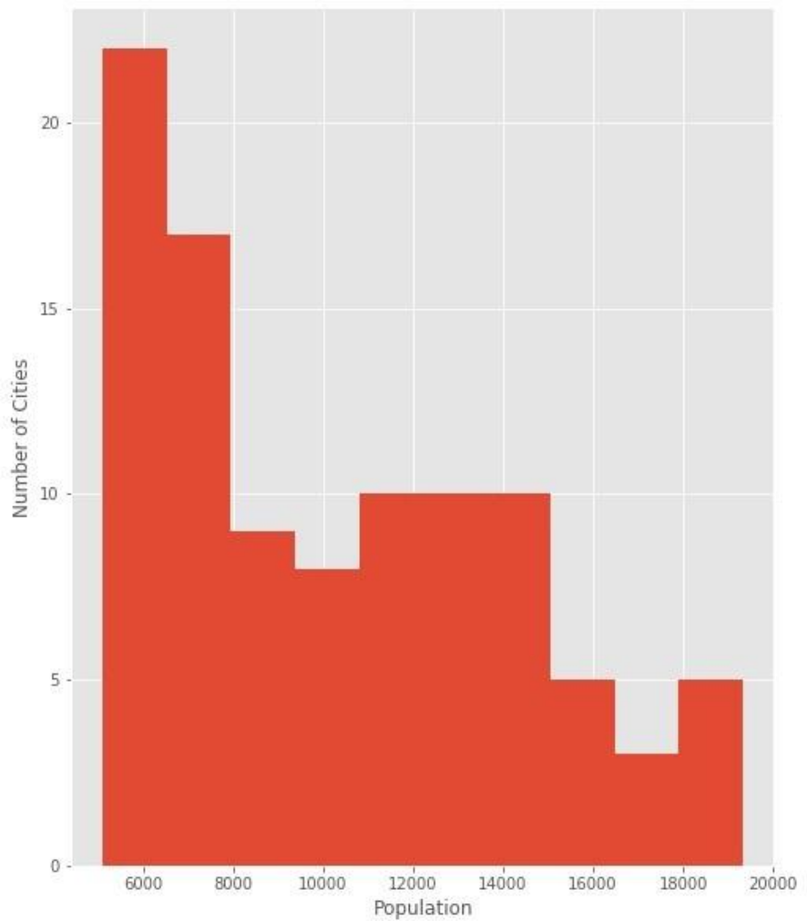
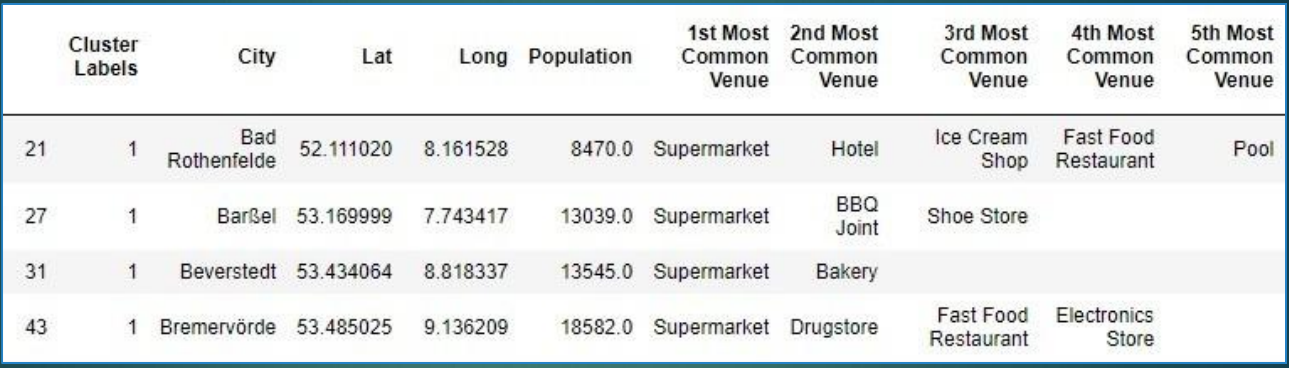


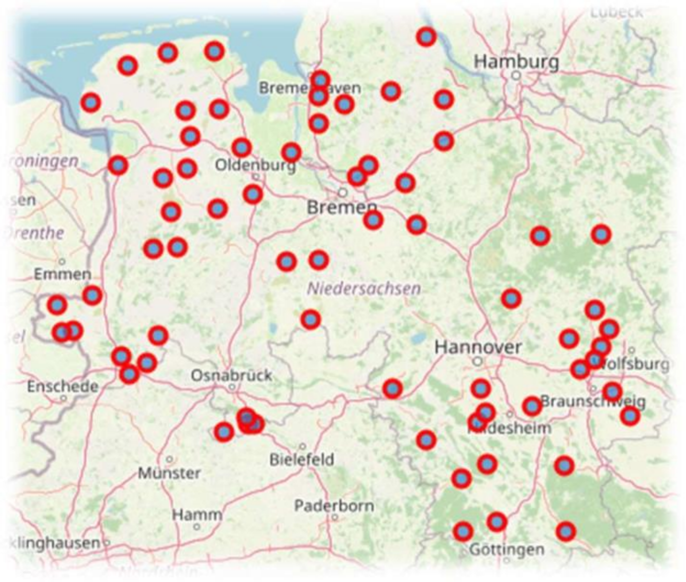
Figure 10: Histogram for Population Distribution in cluster 0

4.3.2. Cluster 1

Table 9. Cluster 1(the original table: 65 lines)



	Cluster Labels	City	Lat	Long	Population	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
21	1	Bad Rothenfelde	52.111020	8.161528	8470.0	Supermarket	Hotel	Ice Cream Shop	Fast Food Restaurant	Pool
27	1	Barßel	53.169999	7.743417	13039.0	Supermarket	BBQ Joint	Shoe Store		
31	1	Beverstedt	53.434064	8.818337	13545.0	Supermarket	Bakery			
43	1	Bremervörde	53.485025	9.136209	18582.0	Supermarket	Drugstore	Fast Food Restaurant	Electronics Store	



Map 6. Cluster 1

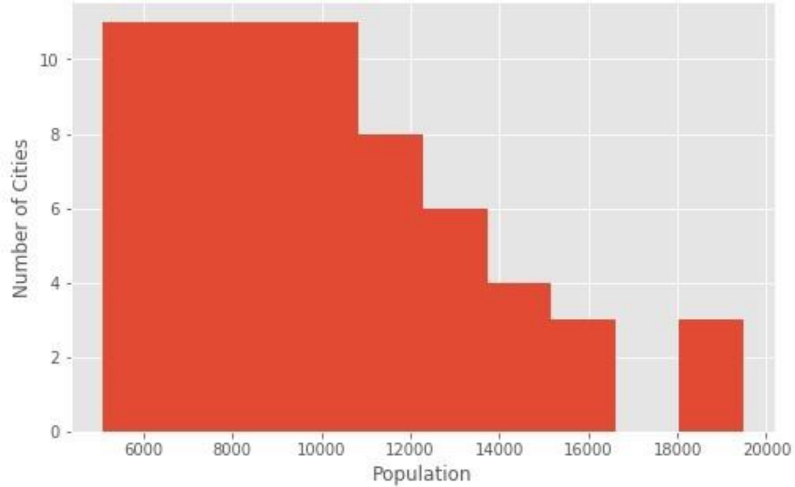
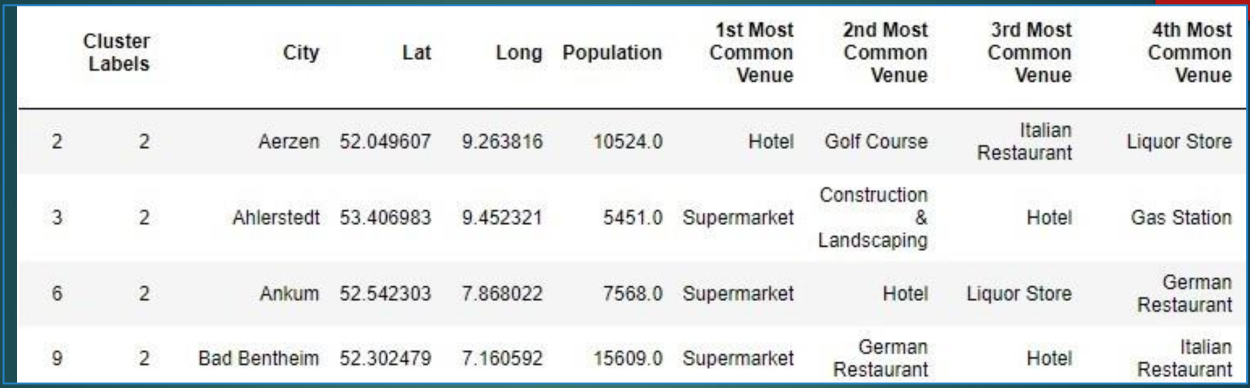


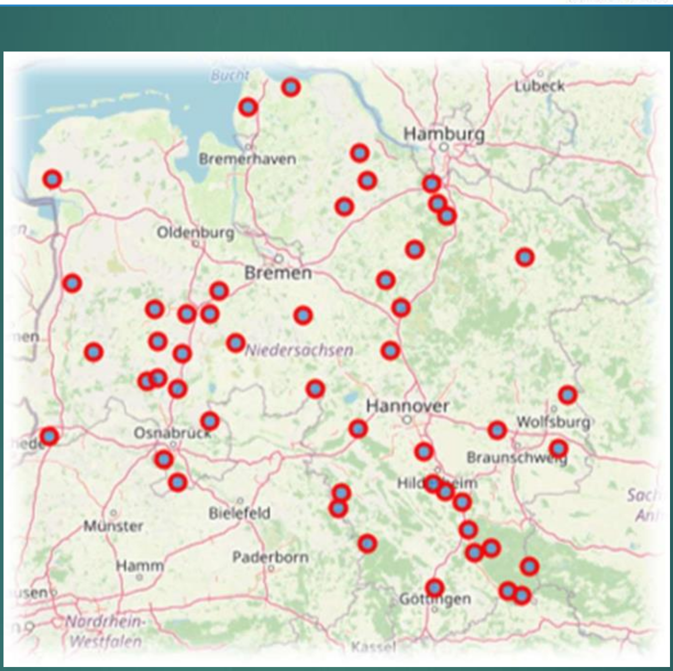
Figure 11: Histogram for Population Distribution in cluster 1

4.3.3. Cluster 2

Table 9. Cluster 2 (the original table: 50 lines)



	Cluster Labels	City	Lat	Long	Population	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
2	2	Aerzen	52.049607	9.263816	10524.0	Hotel	Golf Course	Italian Restaurant	Liquor Store
3	2	Ahlerstedt	53.406983	9.452321	5451.0	Supermarket	Construction & Landscaping	Hotel	Gas Station
6	2	Ankum	52.542303	7.868022	7568.0	Supermarket	Hotel	Liquor Store	German Restaurant
9	2	Bad Bentheim	52.302479	7.160592	15609.0	Supermarket	German Restaurant	Hotel	Italian Restaurant



Map 7. Cluster 2

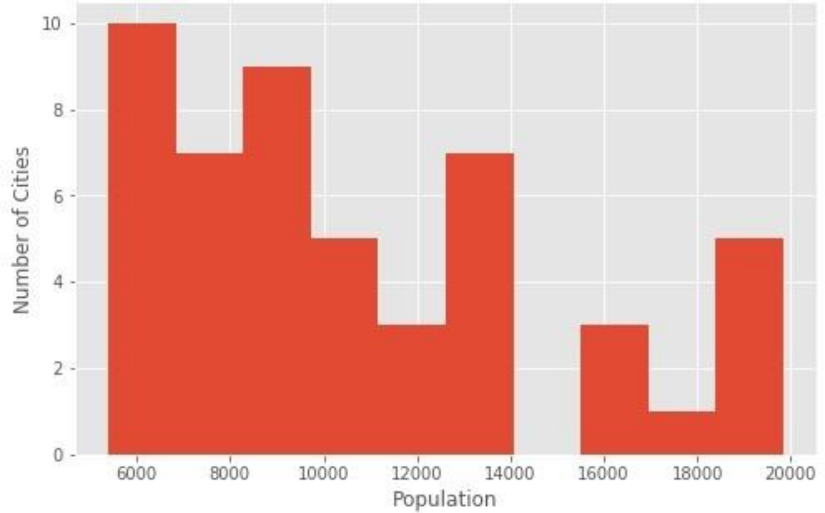


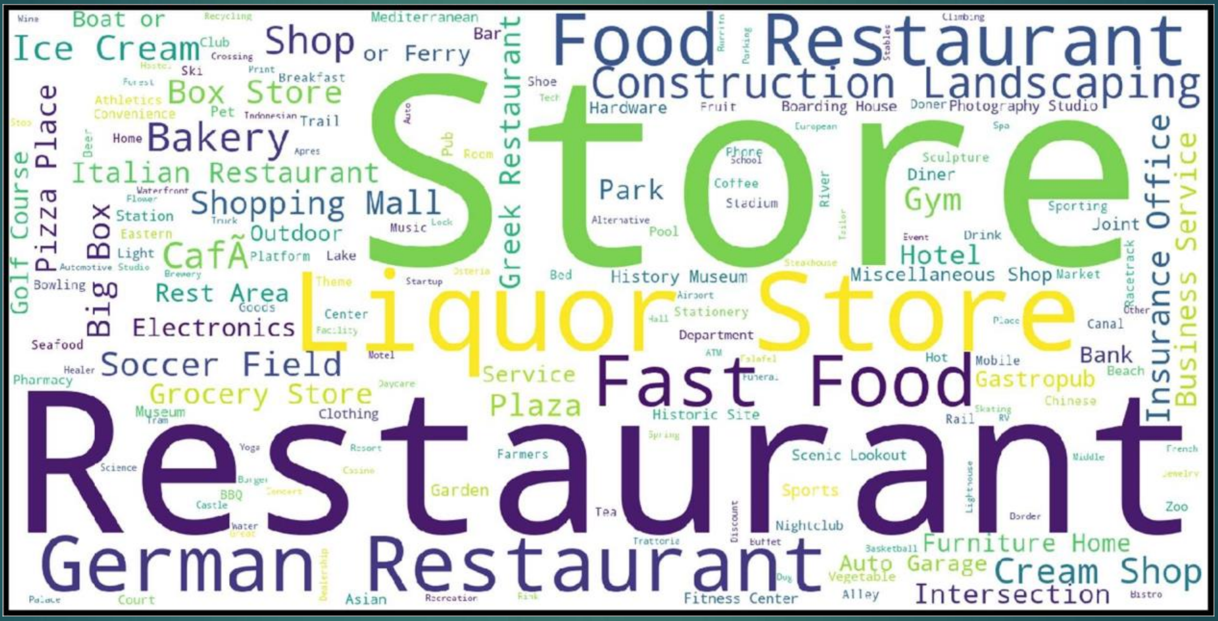
Figure 12: Histogram for Population Distribution in cluster 2

4. Solution to the Problems

Currently, we have 3 different groups. These groups are determined according to the venue distribution in the cities. The grouping process was made using the Kmeans algorithm over 184 different categories belonging to 217 cities. We will now examine these three different groups. We have to find the features that distinguish these groups from each other. In this way, the investor will get an idea of which type of investments to make in which group. To do this, we need to examine what kind of venues are in each city in the groups. For this, we need a new data frame for each group. We will be able to see all types of categories in each city in these data frames. So, we will see not just the top ten categories, but all of them. Then we will create word cloud by converting them into text files. There will be one-word cloud for each group. As a result, we will reveal group characteristics based on these word clouds. The investor will determine the fields of entrepreneurship based on these.

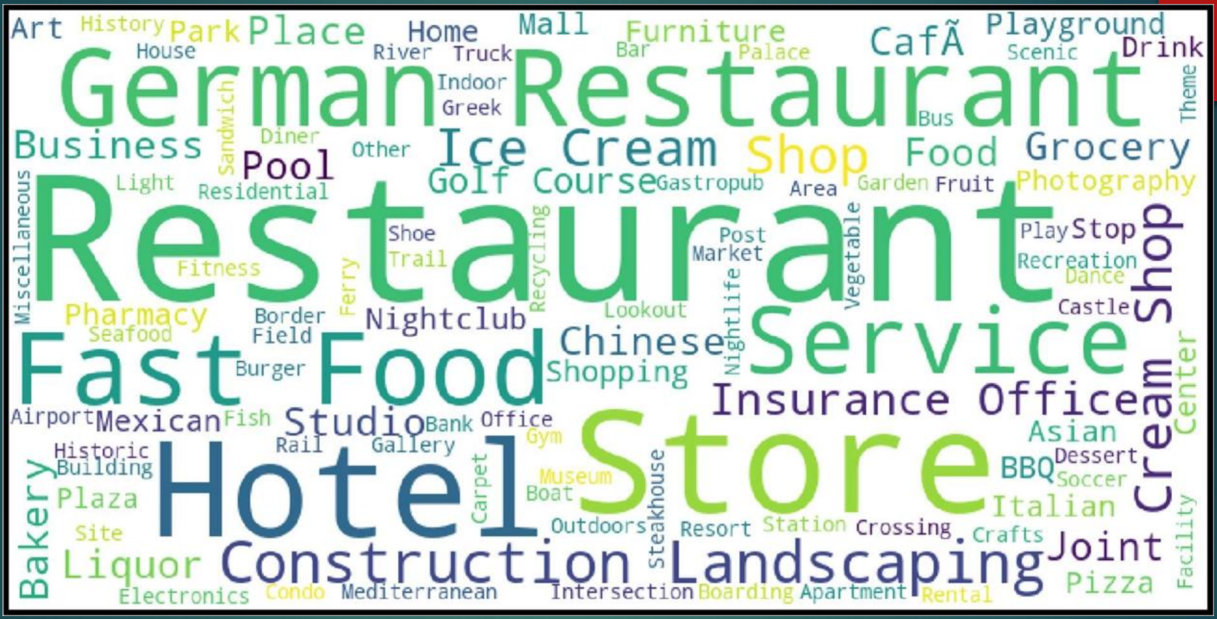
While the word clouds are being prepared, words that may give wrong ideas have been removed. For example, normally the supermarket is Venue that comes first in each group. However, corrections have been made in the texts in order to provide distinction between the Groups.

4.1. Word Cloud for Cluster 0



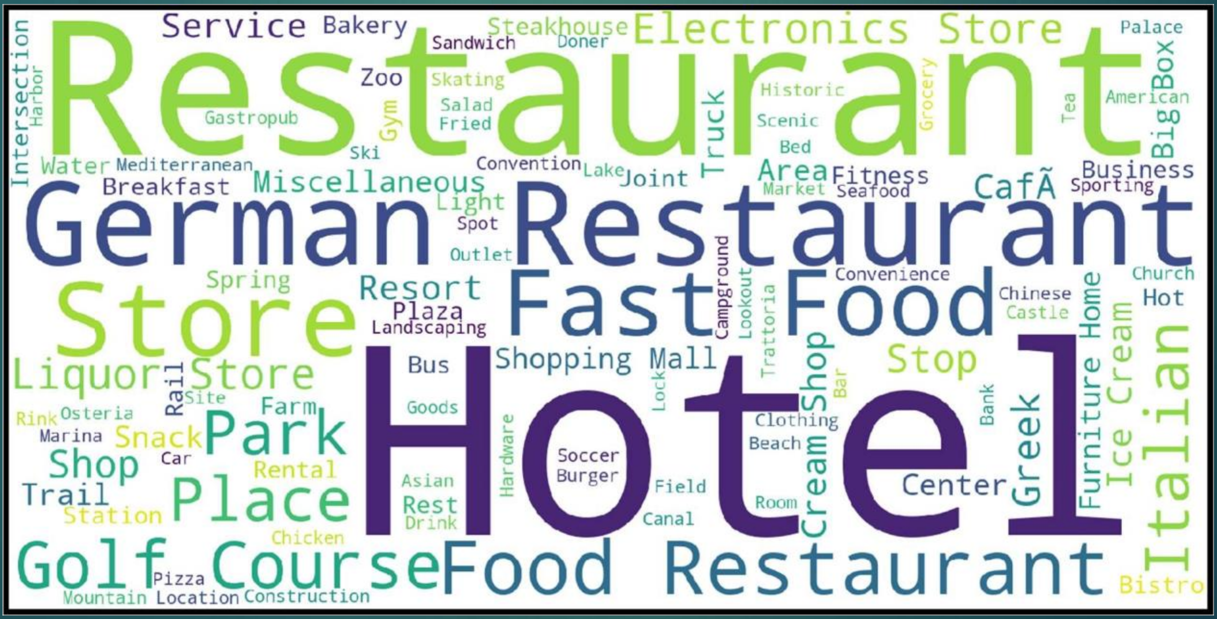
Store has become the most prominent venue of the group in cluster number 1. This becomes even more evident when considering that the ratio of Liquor Store should be added to this. Therefore, these cities generally have the feature of being settled. As it turns out, the Hotel couldn't find a place for itself here. This means that the cities in question are not actually very colorful. Restaurant and German Restaurant have similar rates in all groups. Therefore, these Venues also does not provide much opportunity for categorization. Therefore, investors should make new investments according to the needs of ordinary daily life in these cities.

4.2. Word Cloud for Cluster 1



2.Cluster is marked by Restaurant. Other eating and drinking places also support this. Therefore, it can be thought that social life and common activities are intense in these cities. Therefore, it is considered appropriate to make investments in the entertainment sector in these cities. The social liveliness of the city in this group can continue at night. At least this is the group with such cities. Therefore, it can be accepted that the cities with high visibility of young people are in this group. It is considered that choosing the target audience of young people can be a profitable choice for this group.

4.3. Word Cloud for Cluster 2



In the 3rd Cluster, the most prominent is Venue Hotel. It can be said that these cities are lively and changeable. It might be thought that new people outside of the city come here often. Therefore, businesses such as gift shops or those that will come from outside of the city will be the right choice. Shopping centers can be among the right choices for these cities. It is considered that restaurant-cafe style initiatives that reflect regional historical characteristics towards touristic historical areas will also be appropriate. These are places that can be visited by holidaymakers. It can be said that the cities where marine supplies for swimming, sailing, and rowing enthusiasts can be marketed are predominantly in this group.

5. Results and Discussion

It has been observed that the changes in the population of cities have no effect on categorization. Therefore, venue distribution in the cities belonging to each group has been tried to be analyzed. Based on this, the evaluations are given under the groups.

As a result: Venues in more than 200 cities have been studied. The population situation in these cities and the total number of venues in these cities have also been evaluated. However, it was concluded that the most important factor in classification is the variety of venues in cities. However, it is worth noting that: most of the cities we studied were cities of less than 10000 inhabitants. In such cities, people do not need applications such as Foursquare when determining the restaurant, they will go to. Therefore, it should not be ignored that there is a weakness in terms of data richness.

6. Conclusion

The aim of this study was to study the medium-sized cities in the German state of Niedersachsen. Based on this analysis, it was aimed to see which investments could be made in which cities. The study was completed at a level where a certain evaluation can be made based on the information we have.

As a result: It is evaluated that investors can make a profit by taking into account the evaluations made about each group. It cannot be denied that we have some commercial knowledge of Niedersachsen with this study.