FINDING OUT THE CITIES OF NIEDERSACHSEN

Fahri Cigdem

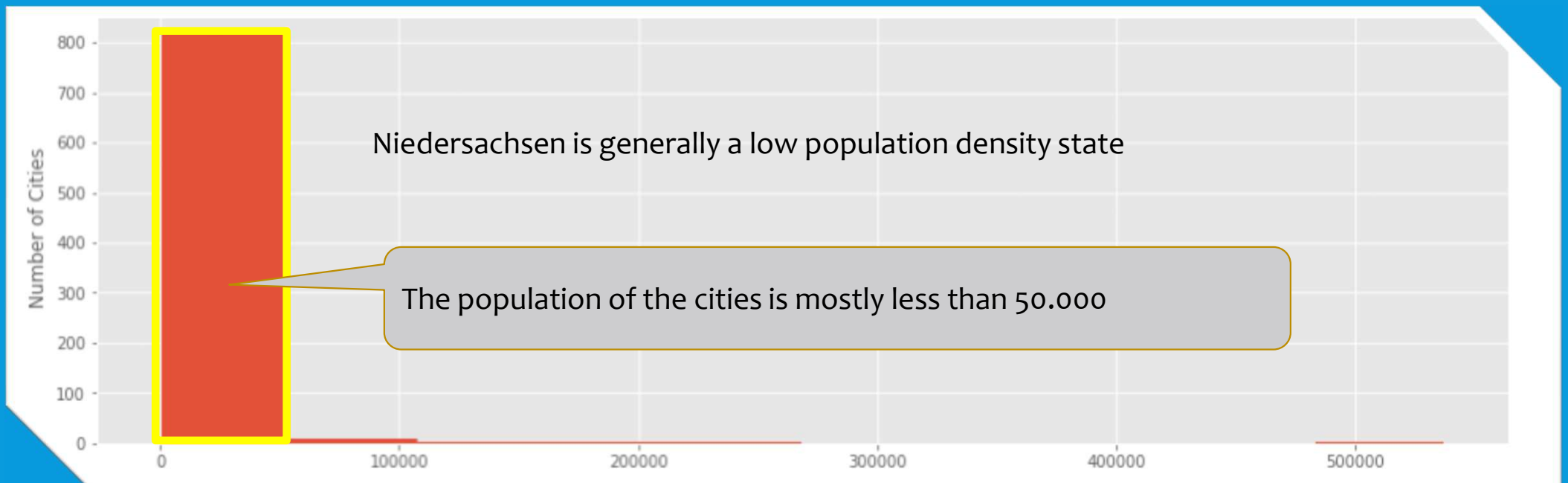August 13, 2020

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.

- We are also particularly interested in **rural areas.** We would also prefer locations **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.
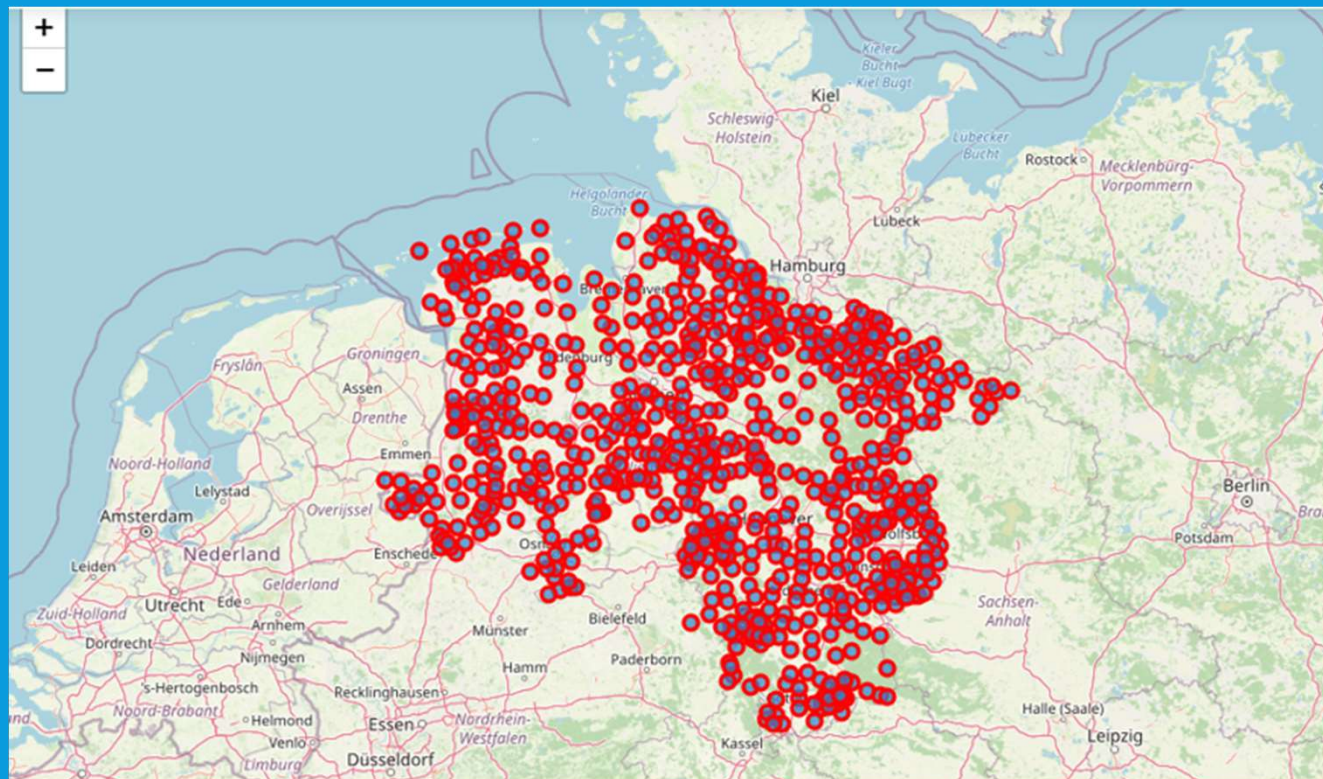
DATA SOURCES

- 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**

DATA ACQUISITION & FEATURE SELECTION

- 839 cities (settlement unit) have been identified in the state of Niedersachsen.

Niedersachsen is generally a low population density state

The population of the cities is mostly less than 50.000
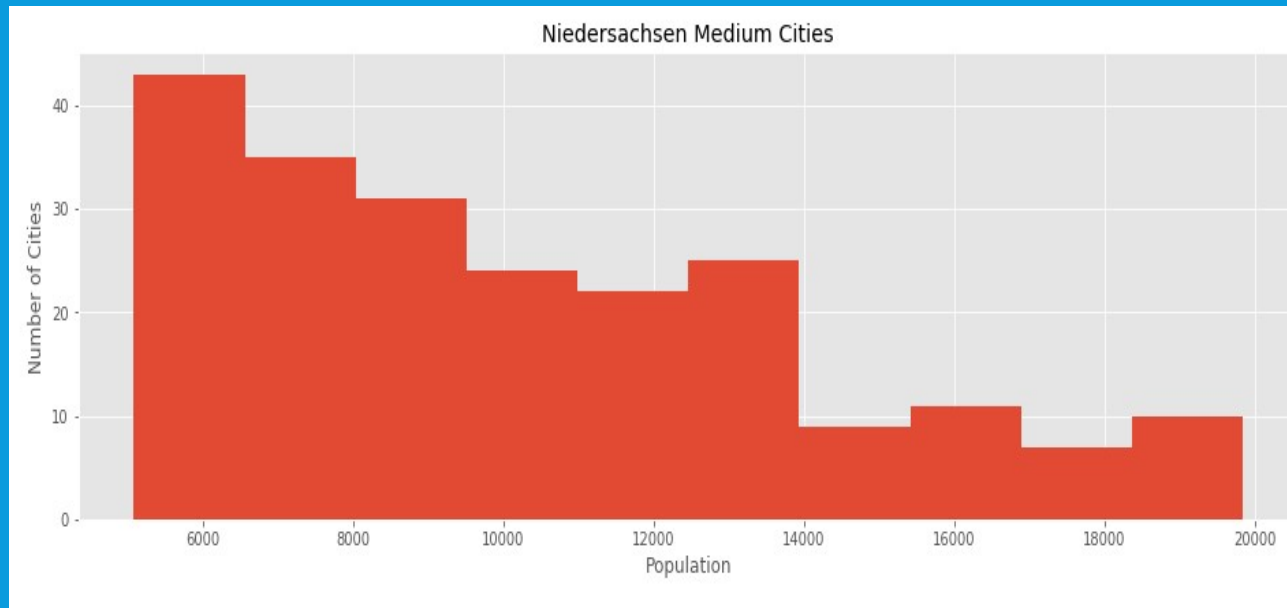
DATA ACQUISITION & FEATURE SELECTION

All small or large cities in Niedersachsen province

DATA ACQUISITION & FEATURE SELECTION

Cities to be covered by the study

DATA ACQUISITION & FEATURE SELECTION

Population distribution of medium-sized cities

DATA ACQUISITION & FEATURE SELECTION

	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

A section of venue data from Foursquare

# DATA ACQUISITION & FEATURE SELECTION

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

METHODOLOGY

- In first step we have collected the required **data: location and Population Data for each city then type (category) of every venues**

- Second step in our analysis will be calculation and exploration of **'venue density'**

- In third and final step we will create **clusters of cities**

EXPLORATORY DATA ANALYSIS

Venue Counts and Population for 217 Cities
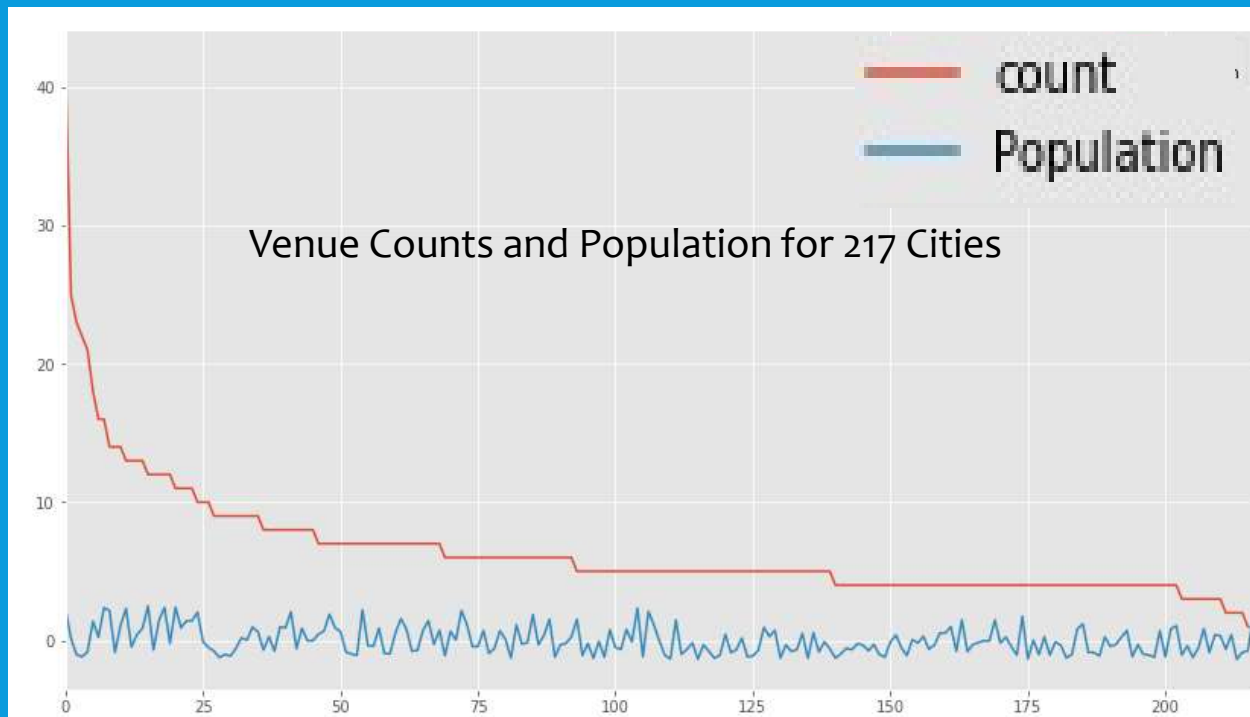
	City	count	Population
100	Hemmingen	42	18974.0
1	Adendorf	25	10853.0
33	Bispingen	23	6410.0
42	Braunlage	22	5795.0
125	Lemwerder	21	7122.0
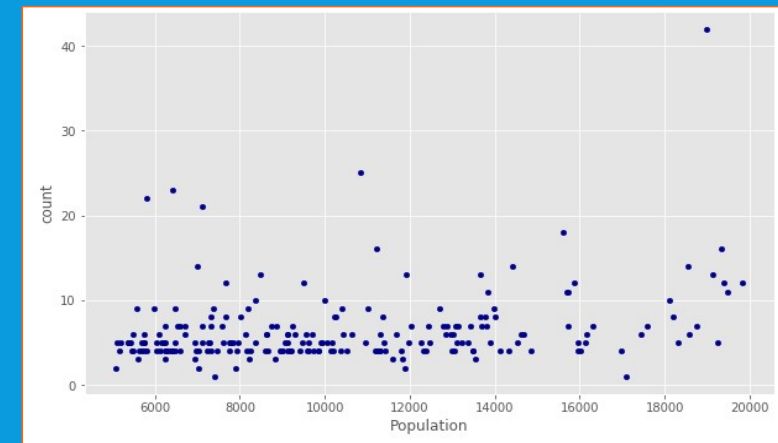
There is no regression between Venue Counts and Population

EXPLORATORY DATA ANALYSIS

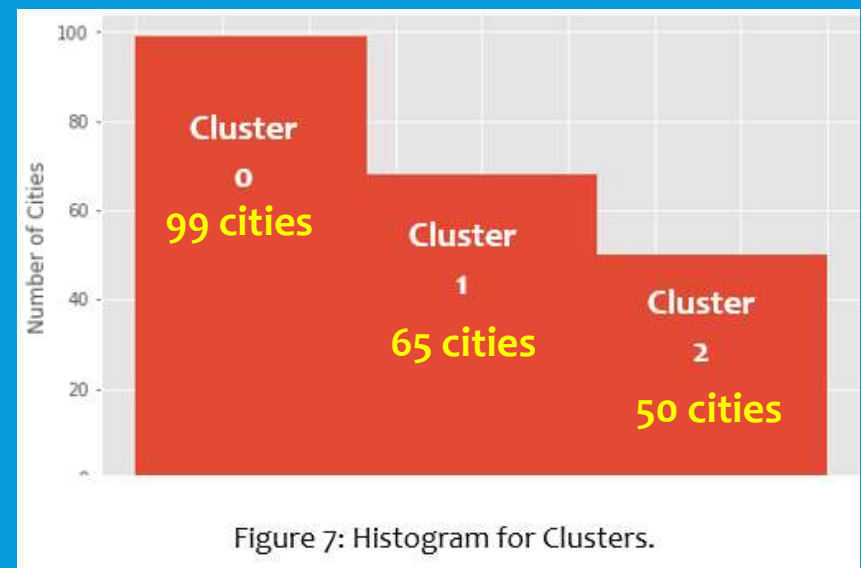|	City	City Latitude	City Longitude	City Population	ATM	Airport	Alternative Healer	American Restaurant	Apres Ski Bar	Art Gallery	...	Trail	Train Station
0	Adelebsen	51.579484	9.752448	6245.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.4
1	Adendorf	53.281748	10.439299	10853.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0
2	Aerzen	52.049607	9.263816	10524.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0
3	Ahlerstedt	53.406983	9.452321	5451.0	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

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	
0	Adelebsen	Train Station	Liquor Store	Electronics Store	Supermarket			...
1	Adendorf	Supermarket	Shopping Mall	Gym	Hardware Store	Golf Course	Furniture / Home Store	...
2	Aerzen	Hotel	Golf Course	Italian Restaurant	Liquor Store			...
3	Ahlerstedt	Supermarket	Construction & Landscaping	Hotel	Gas Station			...
4	Alfeld (Leine)	Supermarket	Bakery	Italian Restaurant	Fast Food Restaurant	Big Box Store	Eastern European Restaurant	...

	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

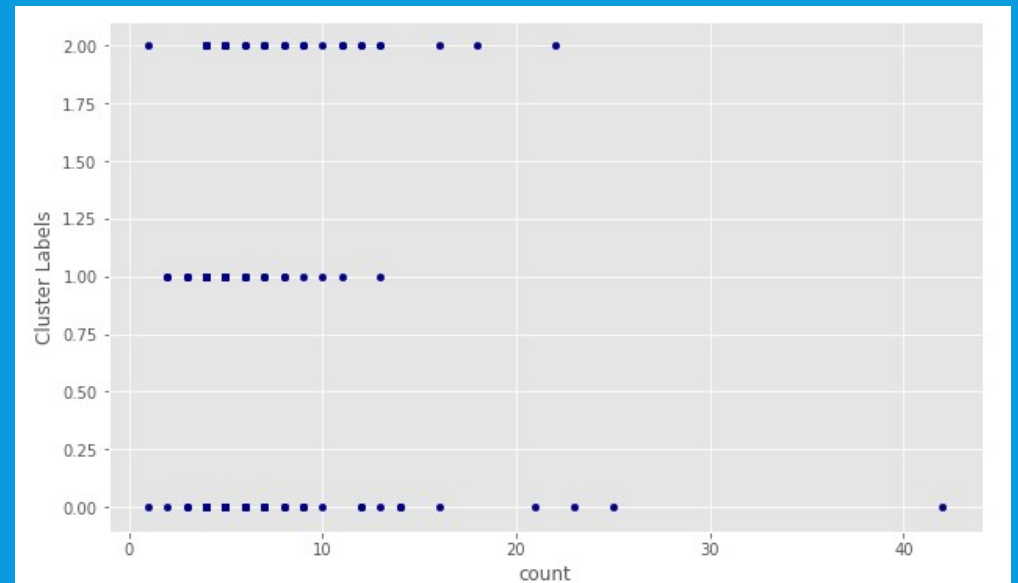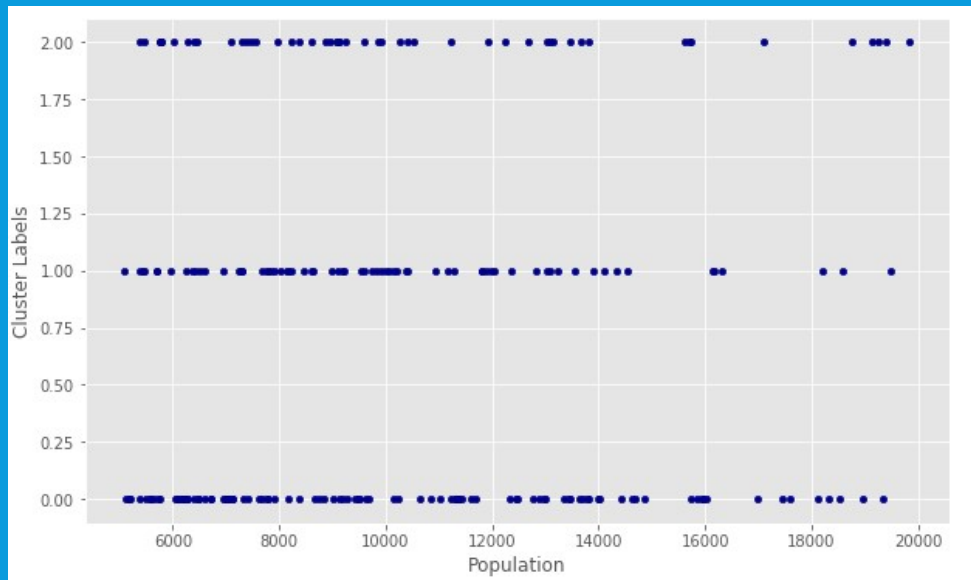**Looking at this new table, we see that the ratio of Train stations in Adelebsen to other venues is 40%**

CLUSTERING

	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

Cluster 0 — 99 cities
Cluster 1 — 65 cities
Cluster 2 — 50 cities

Figure 7: Histogram for Clusters.

· **We cluster the cities based on the rate of the venues in each city**

CLUSTERING

· Relationship with Population and Total Venue Count:

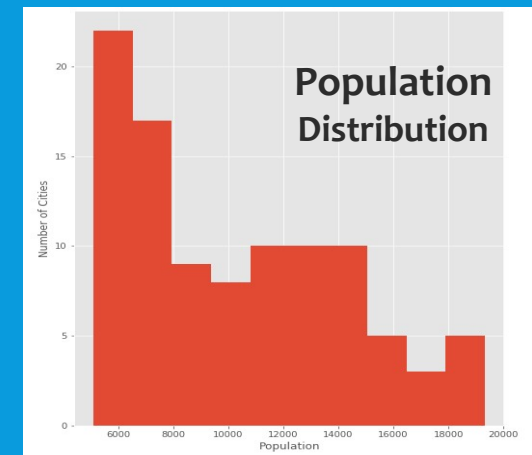 **· there was no significant correlation**

CLUSTERING

Green: Cluster 0, Blue: Cluster 1, Red: Cluster 2

The geographical features of the cities don't have influence on the grouping
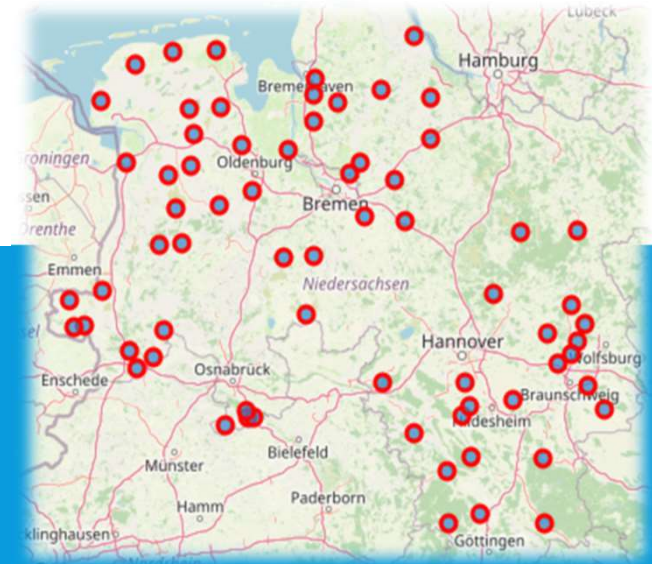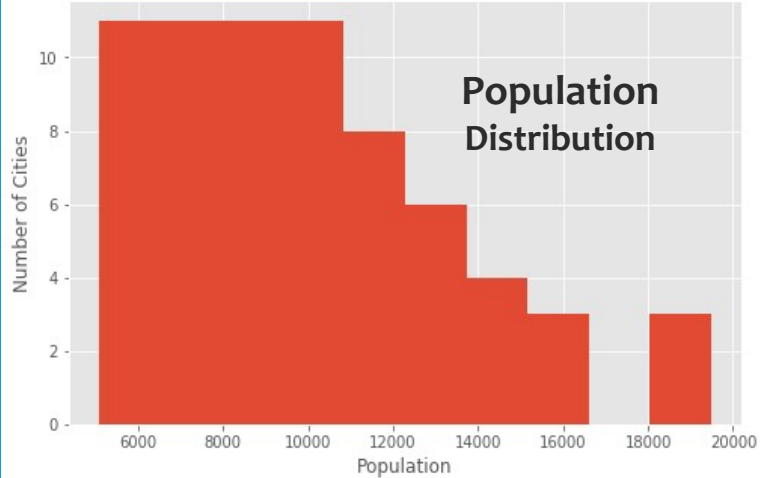
CLUSTER 0

	Cluster Labels	City	Lat	Long	Population	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	0	Adelebsen	51.579484	9.752448	6245.0	Train Station	Liquor Store	Electronics Store
1	0	Adendorf	53.281748	10.439299	10853.0	Supermarket	Shopping Mall	Gym
4	0	Alfeld (Leine)	51.986308	9.824747	18535.0	Supermarket	Bakery	Italian Restaurant
5	0	Algermissen	52.251407	9.967904	7918.0	Construction & Landscaping	Mobile Phone Shop	Liquor Store
7	0	Apen	53.221426	7.809734	11692.0	Construction & Landscaping	Insurance Office	Diner

the original table has 99 lines

Population Distribution

CLUSTER 1

Cluster Labels	City	Lat	Long	Population	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
21	1	Bad Rothenfelde	52.111020	8.161528	8470.0	Supermarket	Hotel	Ice Cream Shop
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
45	1	Bunde	53.184017	7.269545	7675.0	Supermarket	Border Crossing	Ice Cream Shop
...	...	...	...	...	...	...	...

the original table has 65 lines

Population Distribution

CLUSTER 2

|	Cluster Labels	City	Lat	Long	Population	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
2	2	Aerzen	52.049607	9.263816	10524.0	Hotel	Golf Course	Italian Restaurant
3	2	Ahlerstedt	53.406983	9.452321	5451.0	Supermarket	Construction & Landscaping	Hotel
6	2	Ankum	52.542303	7.868022	7568.0	Supermarket	Hotel	Liquor Store
9	2	Bad Bentheim	52.302479	7.160592	15609.0	Supermarket	German Restaurant	Hotel
10	2	Bad Bevensen	53.080941	10.574481	9248.0	Golf Course	Hot Spring	Hotel
12	2	Bad Fallingbostel	52.865448	9.693020	11926.0	Supermarket	Hotel	Fast Food Restaurant
14	2	Bad Grund (Harz)	51.785465	10.212901	8238.0	Historic Site	Hotel	Train Station
16	2	Bad Laer	52.101414	8.086517	9110.0	Hotel	Supermarket	Gym / Fitness Center

Population Distribution
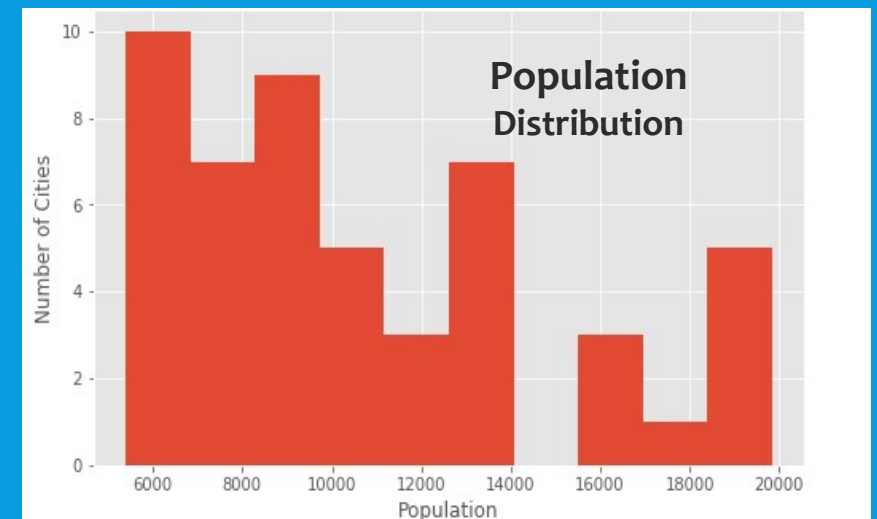
SOLUTION TO THE PROBLEMS

- 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.

- As a result, we will reveal group characteristics based on these word clouds.

- The investor will determine the fields of entrepreneurship based on these.
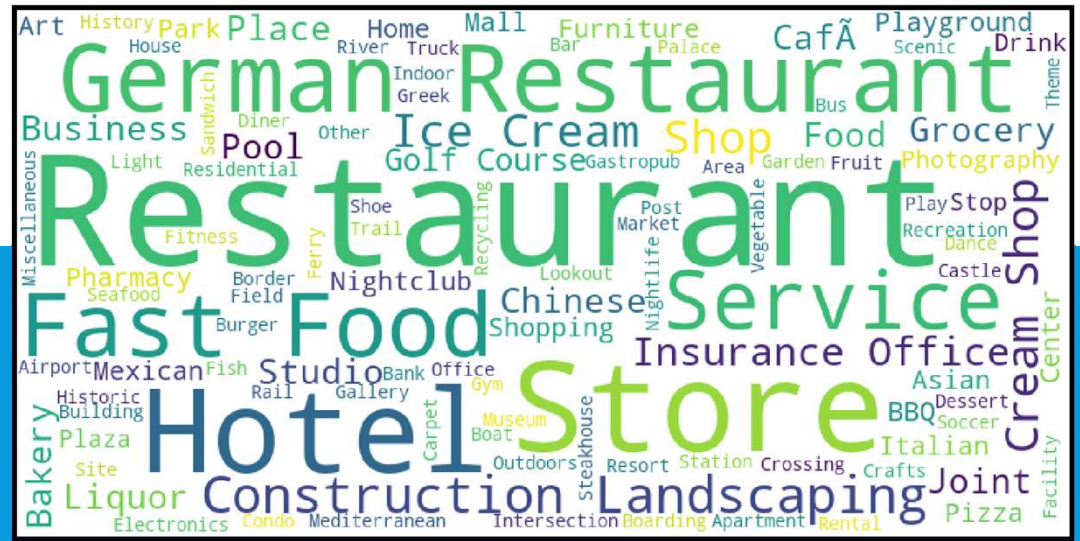
CLUSTER 0
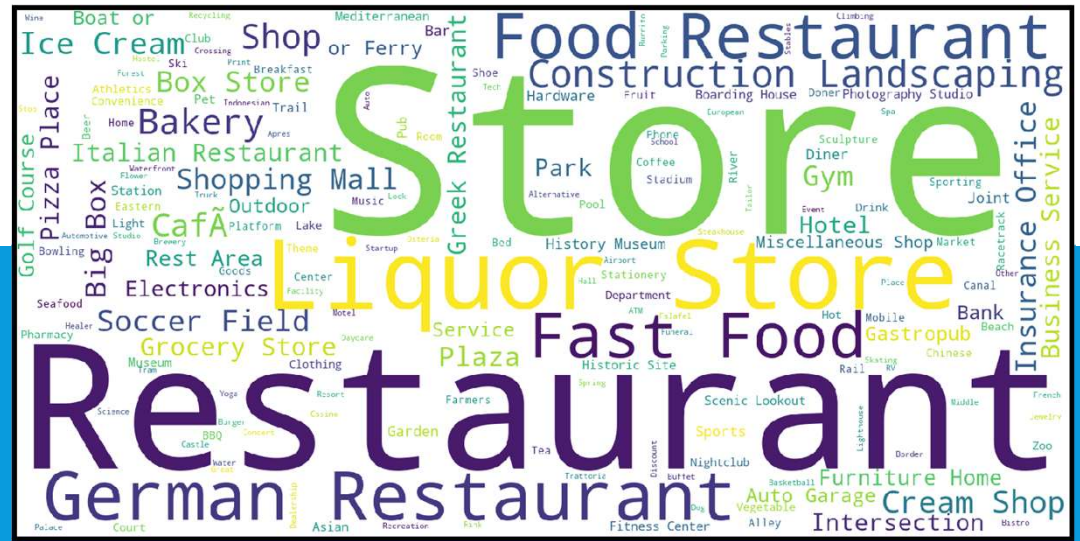
Store has become the most prominent venue of the group in cluster number 1. This becomes even more evident when considering that the ratio of Liquor Store should be added to this. Therefore, these cities generally have the feature of being settled. As it turns out, the Hotel couldn't find a place for 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.

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.

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.

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.

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.**