Deciding on Neighborhood to live in Aachen depending on age group

1. Background

Aachen is an historic city in Germany. Its borders touch Belgium and Netherlands, which makes it an important place where people from three countries come to live and work. Moreover, it is a student city with three Universities spread out all over the city. Because of the huge number of students, it is hard to find an appropriate place in Aachen to live. In this report it has been discussed how to find a place of your choice in Aachen.

2. Problem Statement

The problem statement is to find a suitable Stadtbezirke (municipalities) to live in Aachen depending on age group by analyzing venues in that region. As different age groups have different lifestyle and requirements, in general they tend to choose specific neighborhoods, which have all the amenities that support their lifestyle. The categories of age groups, that will be considered here are Students (18-30), Working People (31-50), Older citizens (50+).

3. Data

The data to be used for this problem will be as follows:

1. Open data from cybo.com
   - it was used to collect postal codes in Aachen city and coordinates will be acquired to retrieve foursquare location data.
2. foursquare location data
   - The information of different business venues along with public services like parks, hospitals, universities.
   - All the venue information can be used to cluster the municipalities into different requirements for specific age groups. e.g.: people in the age group 15-30 like clubbing and going to universities. Therefore, the areas having more clubs and university buildings will choose that place for their residence. Similarly, older people above age 50 will like to have park and medical facilities close to their residence.

4. Data Collection & Cleaning

1. Collecting location information in Aachen

The Postal Code information for Aachen city was collected from https://postal-codes.cybo.com/germany/aachen/#listcodes, which gives the population and area corresponding to each postal code in Aachen. Subsequently each postal code area is used to calculate the radius, that we need

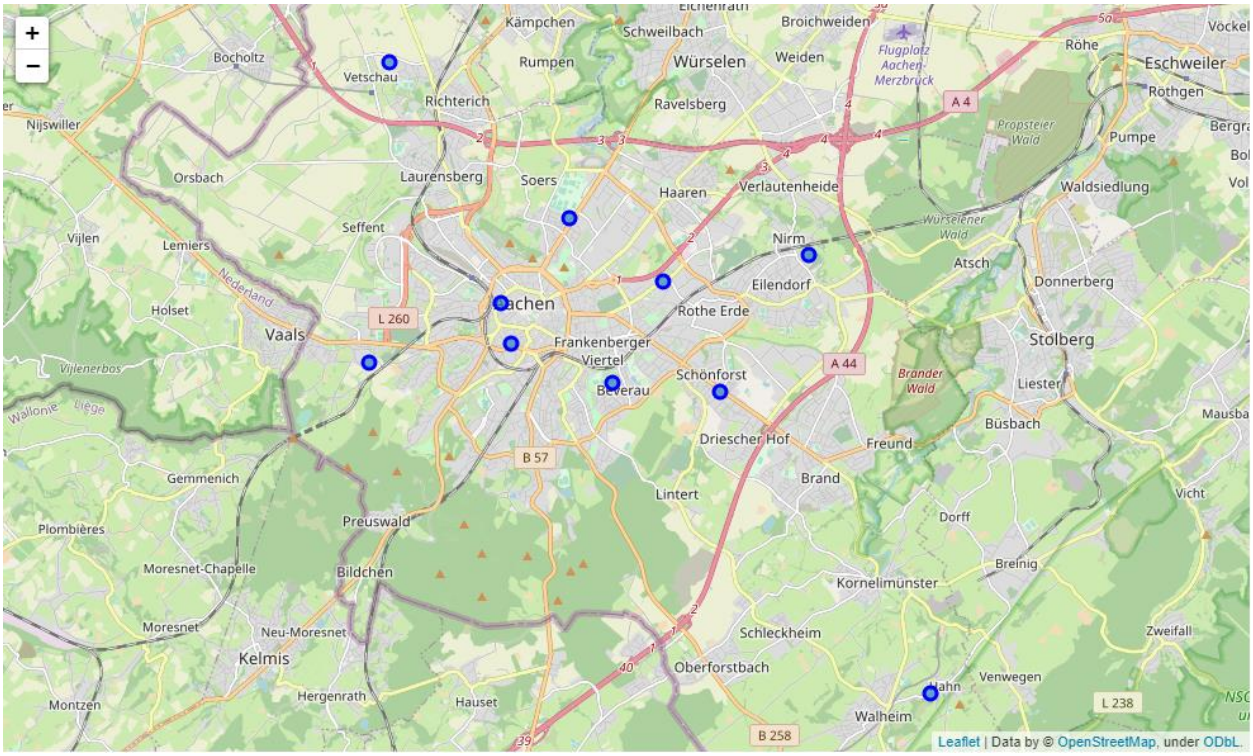
to know the area in which we will look for the venues for the corresponding postal code (We assume the region covered by each postal code as circular for easier data acquisition). The latitude and longitude for each postal code was collected from OpenStreetMaps using geocoder Nominatim API.

Table 1: Postal Codes in Aachen with Latitude, Longitude and radius according to their area

	postcode	city	state	population	area	lat	lng	radius
0	52062	Aachen	North Rhine-Westphalia	8,152	1.576 km²	50.776348	6.077642	708
1	52064	Aachen	North Rhine-Westphalia	10,981	2.017 km²	50.769547	6.080325	801
2	52066	Aachen	North Rhine-Westphalia	20,958	7.5 km²	50.762723	6.107663	1545
3	52068	Aachen	North Rhine-Westphalia	22,000	5.6 km²	50.780074	6.121386	1335
4	52070	Aachen	North Rhine-Westphalia	26,939	9.1 km²	50.791070	6.096156	1702
5	52072	Aachen	North Rhine-Westphalia	21,712	19.3 km²	50.817683	6.047362	2479
6	52074	Aachen	North Rhine-Westphalia	36,959	33.4 km²	50.766230	6.041954	3261
7	52076	Aachen	North Rhine-Westphalia	24,587	55.1 km²	50.709571	6.193844	4189
8	52078	Aachen	North Rhine-Westphalia	30,463	12.5 km²	50.761230	6.136976	1995
9	52080	Aachen	North Rhine-Westphalia	24,343	15.1 km²	50.784631	6.160787	2192

The following map shows the coordinates of postal codes on Aachen city.

Figure 1: Aachen Postal Codes



2. Collecting Foursquare data

To start with, we collect the high-level categories of venues available on Foursquare, which are as follows with their corresponding category IDs.

Table 2 Foursquare Top-Level Venue Categories

Category Name	Category ID
Arts & Entertainment	4d4b7104d754a06370d81259
College & University	4d4b7105d754a06372d81259
Event	4d4b7105d754a06373d81259
Food	4d4b7105d754a06374d81259
Nightlife Spot	4d4b7105d754a06376d81259
Outdoors & Recreation	4d4b7105d754a06377d81259
Professional & Other Places	4d4b7105d754a06375d81259
Residence	4e67e38e036454776db1fb3a
Shop & Service	4d4b7105d754a06378d81259
Travel & Transport	4d4b7105d754a06379d81259

The above list was used to collect count of venues in a category corresponding to each location within their radii respectively. The following table shows the collected data.

Table 3: Count of Venue Categories in each postal code

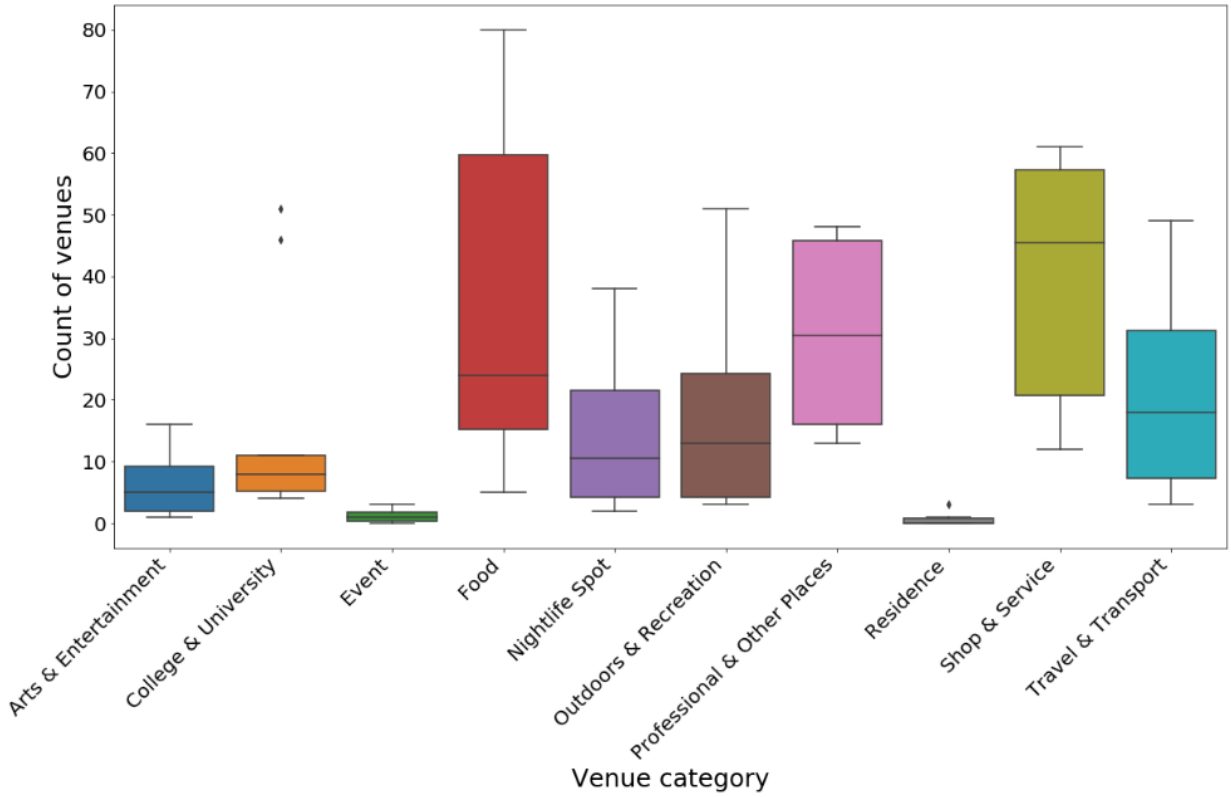
	postcode	population	lat	lng	Arts & Entertainment	College & University	Event	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Residence	Shop & Service	Travel & Transport
0	52062,Aachen	8,152	50.776348	6.077642	10	46	0	62	30	15	48	0	59	18
1	52064,Aachen	10,981	50.769547	6.080325	12	11	1	66	22	22	46	0	61	32
2	52066,Aachen	20,958	50.762723	6.107663	7	9	2	53	16	25	43	1	60	29
3	52068,Aachen	22,000	50.780074	6.121386	2	5	1	16	4	11	18	0	46	18
4	52070,Aachen	26,939	50.791070	6.096156	7	11	1	29	20	30	45	1	45	49
5	52072,Aachen	21,712	50.817683	6.047362	1	6	0	5	5	4	13	0	14	6
6	52074,Aachen	36,959	50.766230	6.041954	16	51	3	80	38	51	48	3	52	49
7	52076,Aachen	24,587	50.709571	6.193844	2	5	0	15	5	5	14	0	16	3
8	52078,Aachen	30,463	50.761230	6.136976	3	7	1	19	4	4	16	0	35	7
9	52080,Aachen	24,343	50.784631	6.160787	2	4	2	9	2	3	16	0	12	8

This information is analyzed to find suitable locations for each age group.

5. Exploratory Analysis

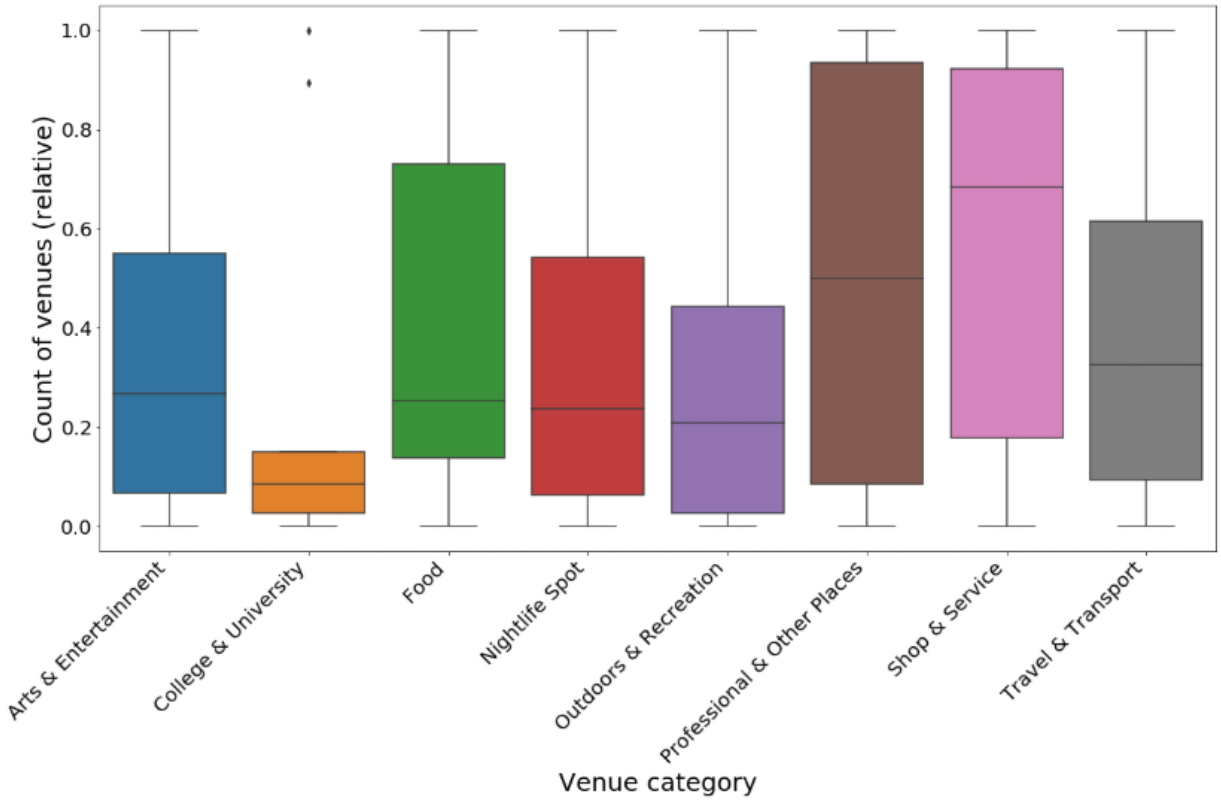
The information in table 3 in the last chapter is used to analyze which information we need for the analysis and which are irrelevant. The following box plot of all the counts gives us idea of relevancy of the data.

Figure 2: Box Plot for Count of Venues for Categories



It is clear from the plot that the Event and Residence values are too less to be used in the clustering task. Hence, we drop those columns. The venue counts are then normalized for better clustering. Following figure shows box plot for normalized data where we can clearly see the comparisons between the venue counts. Now this data can be used to cluster the similar regions in Aachen city, which can be assigned to different age groups depending on the category of venues.

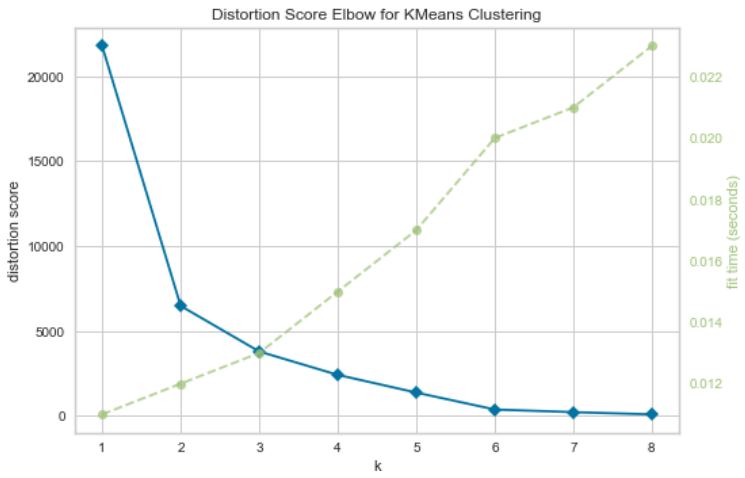
Figure 3: Box plot for normalized venue counts



6. Clustering

To separate the postal codes into different clusters on similarity in the available venues, we need to know how many clusters are possible from the available data. One of the effective methods to find the number of clusters is elbow method. The following graph helps us to find the elbow and select the number of clusters.

Figure 4: Elbow method to find the number of Clusters



In this graph we can se there is a sharp change after 2 clusters, but according to our age group we can consider 3 clusters as our choice. 3 clusters still have acceptable distortion score which is not too high or too low. Now after running KMeans clustering on the data we found the following clusters.

Table 4: Clusters generated by KMeans

Arts & Entertainment	College & University	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Shop & Service	Travel & Transport	Cluster	PostalCode
0.600000	0.893617	0.760000	0.777778	0.250000	1.000000	0.959184	0.326087	2	52062,Aachen
0.733333	0.148936	0.813333	0.555556	0.395833	0.942857	1.000000	0.630435	0	52064,Aachen
0.400000	0.106383	0.640000	0.388889	0.458333	0.857143	0.979592	0.565217	0	52066,Aachen
0.066667	0.021277	0.146667	0.055556	0.166667	0.142857	0.693878	0.326087	1	52068,Aachen
0.400000	0.148936	0.320000	0.500000	0.562500	0.914286	0.673469	1.000000	0	52070,Aachen
0.000000	0.042553	0.000000	0.083333	0.020833	0.000000	0.040816	0.065217	1	52072,Aachen
1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.816327	1.000000	2	52074,Aachen
0.066667	0.021277	0.133333	0.083333	0.041667	0.028571	0.081633	0.000000	1	52076,Aachen
0.133333	0.063830	0.186667	0.055556	0.020833	0.085714	0.469388	0.086957	1	52078,Aachen
0.066667	0.000000	0.053333	0.000000	0.000000	0.085714	0.000000	0.108696	1	52080,Aachen

We can analyze each cluster separately and see which age group we can assign to that cluster.

7. Results

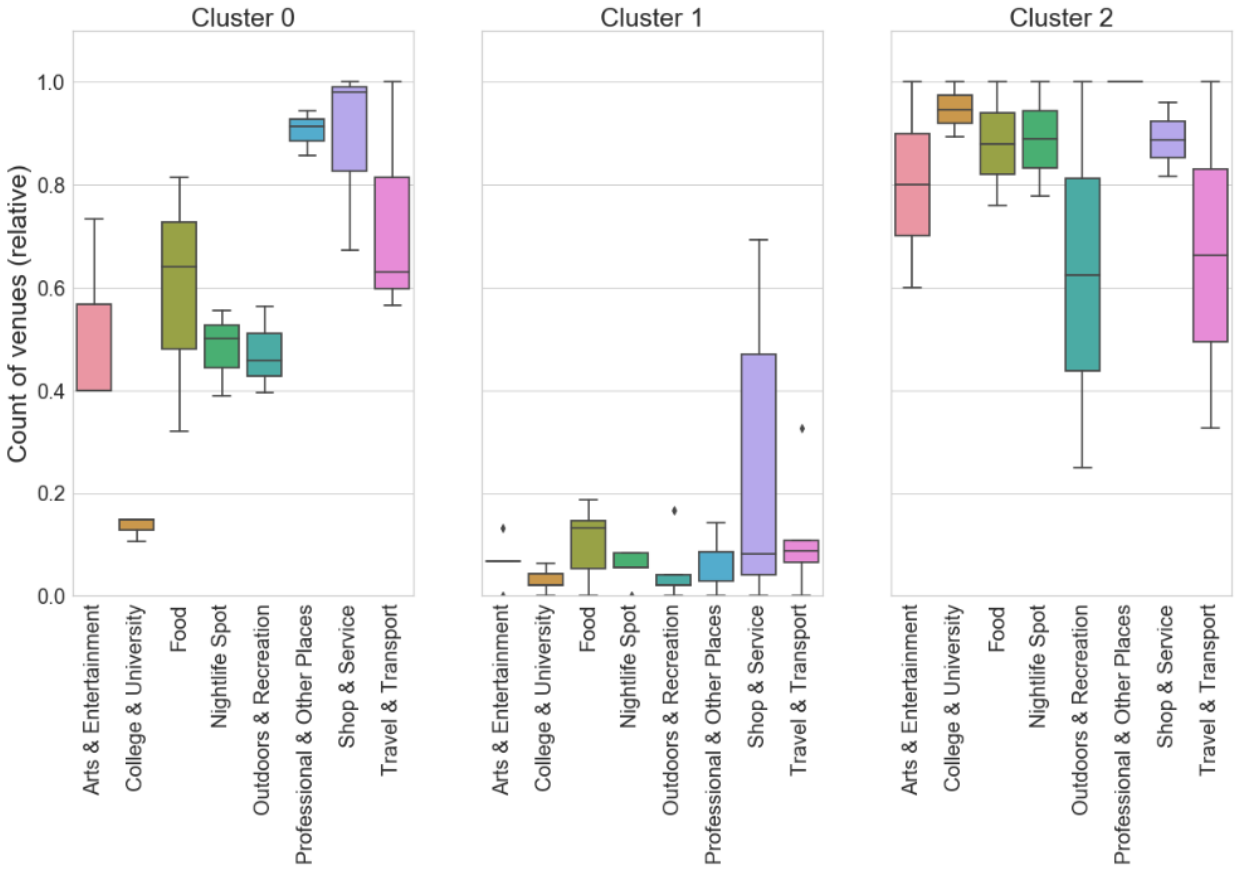
The following are three clusters representing suitable location to live for 3 different age groups. We will analyze Figure 5 to decide the assignment.

- **Cluster 0 (52064, 52066, 52070)**
  These regions have less venues for College and University. But high number of venues in category professional & other places along with decent number of night spots. So, this place might be a better choice for middle aged people, who are working professionals and are engaged into moderate amount of night life.

- **Cluster 1 (52068, 52072, 52076, 52078, 52080)**
  These regions are low in number for all category venues except for shops and service which are low to moderate in number. This shows that there will be less noise in these regions which might be perfect for older citizens.

- **Cluster 2 (52062, 52074)**
  These regions are highly populated with all kinds of venues especially college & university. Hence, this cluster is best for young adults, who are mostly students. This cluster also has most Night life spots, which is desired by young citizens

Figure 5: Properties of different Clusters



8. Conclusion and Discussion

From the above analysis we concluded that we can separate the city into three regions which can be chosen by three different age groups. The following table shows the assignments.

Age Group	Postal Codes
Students (18-30)	52062, 52074
Working People (31-50)	52064, 52066, 52070
Older citizens (50+)	52068, 52072, 52076, 52078, 52080

These results are based on assumptions that the mentioned age groups have some specific needs. The data from Foursquare is also not enough to decide on lifestyle requirements. It also doesn't give information about apartment prices, which plays an important role in choosing living places.

This analysis can be refined by collecting data from more resources and accurately finding information for each postal code. Also, the pricing for venues in categories Food, Nightlife Spot, Shop & Services. However, this analysis was a good overview on distribution of amenities in the city of Aachen.