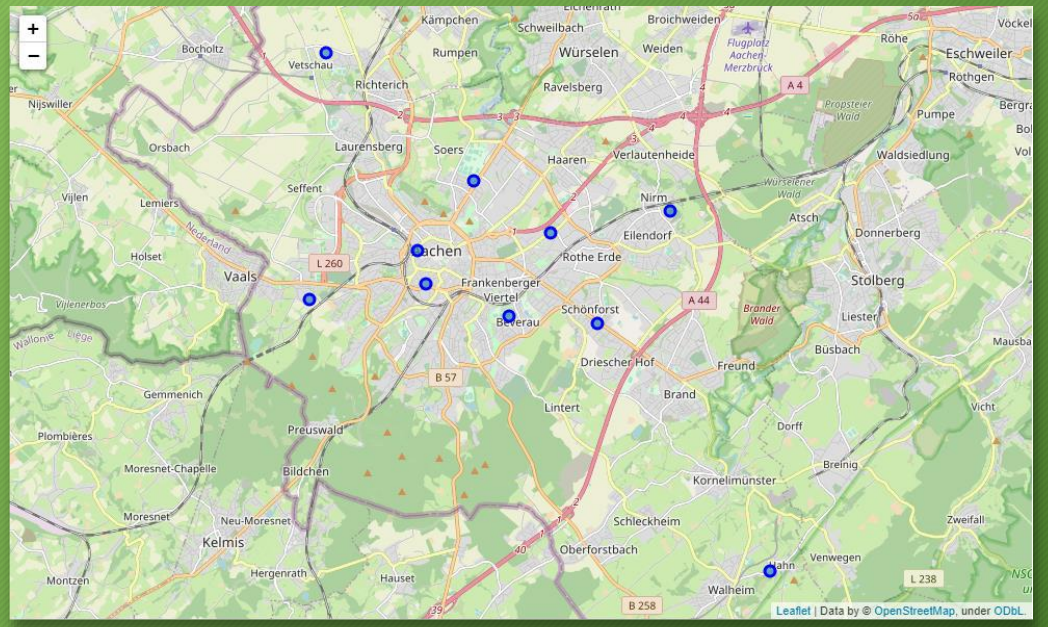
Deciding on Neighborhood to live in Aachen depending on age group

Coursera Capstone Project

Problem Statement & Data

- Find a suitable Stadtbezirke (municipalities) to live in Aachen depending on age group by analyzing venues in that region.
- Data:
  - Open data from Cybo.com
    - Collect postal codes in Aachen city
  - OpenStreetMaps
    - Collect latitude and longitudes for each postal code
  - Foursquare data
    - Collect venue information for each coordinate in a particular radius



Foursquare Venue Categories

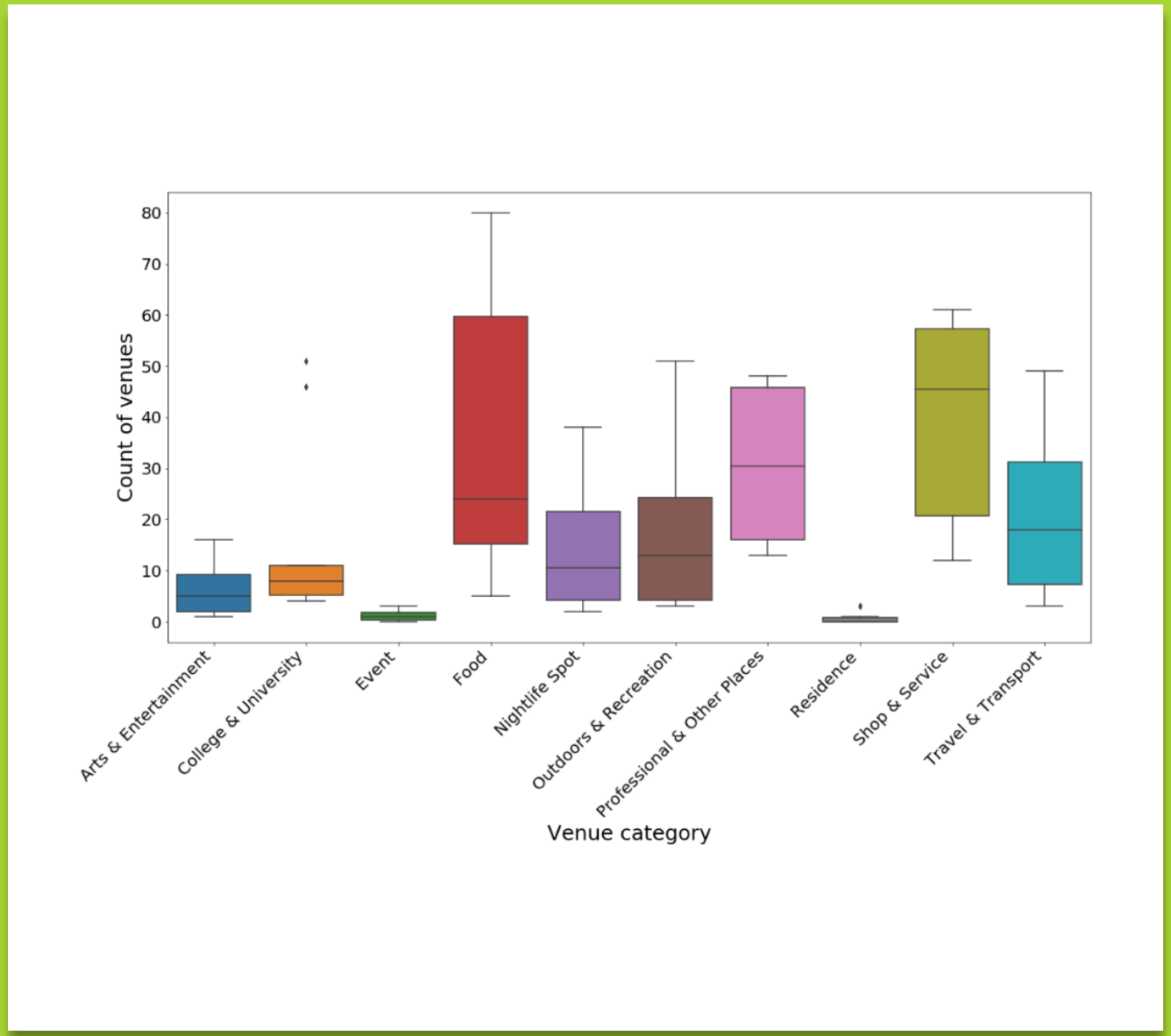
Category Name	Category ID
Arts & Entertainment	4d4b7104d754a06370d81259
College & University	4d4b7105d754a06372d81259
Event	4d4b7105d754a06373d81259
Food	4d4b7105d754a06374d81259
Nightlife Spot	4d4b7105d754a06376d81259
Outdoors & Recreation	4d4b7105d754a06377d81259
Professional & Other Places	4d4b7105d754a06375d81259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

Counts of Venue Categories per 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

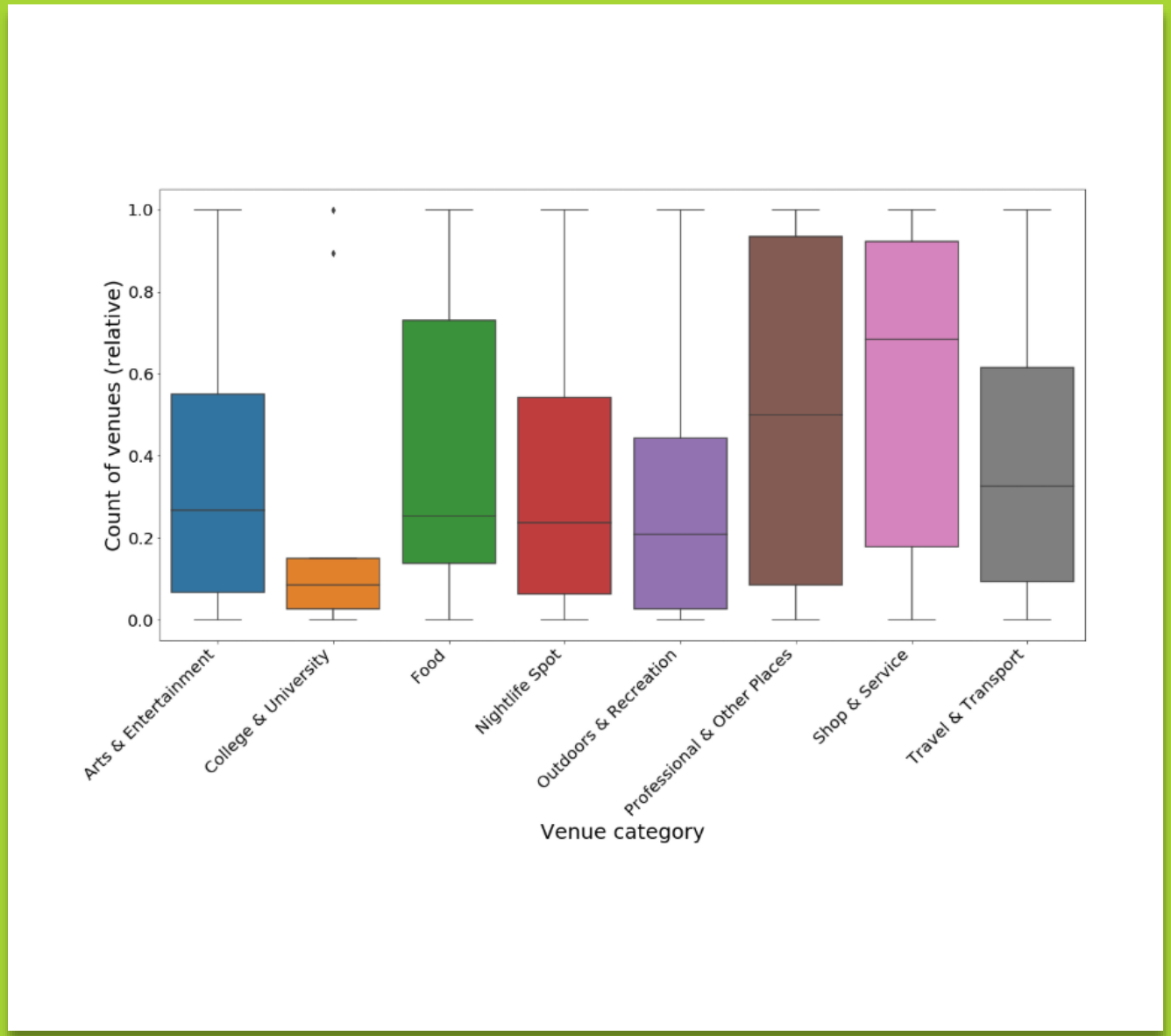
Counts of Venues in Plot

- This plot shows count of different venue categories in Aachen.
- There are substantial number of food venues followed by shops and services.
- Good number of Nightlife Spots because it is a Student City.
- Also enough professional venues which is good for working people.
- Very few Event and Residence venues.



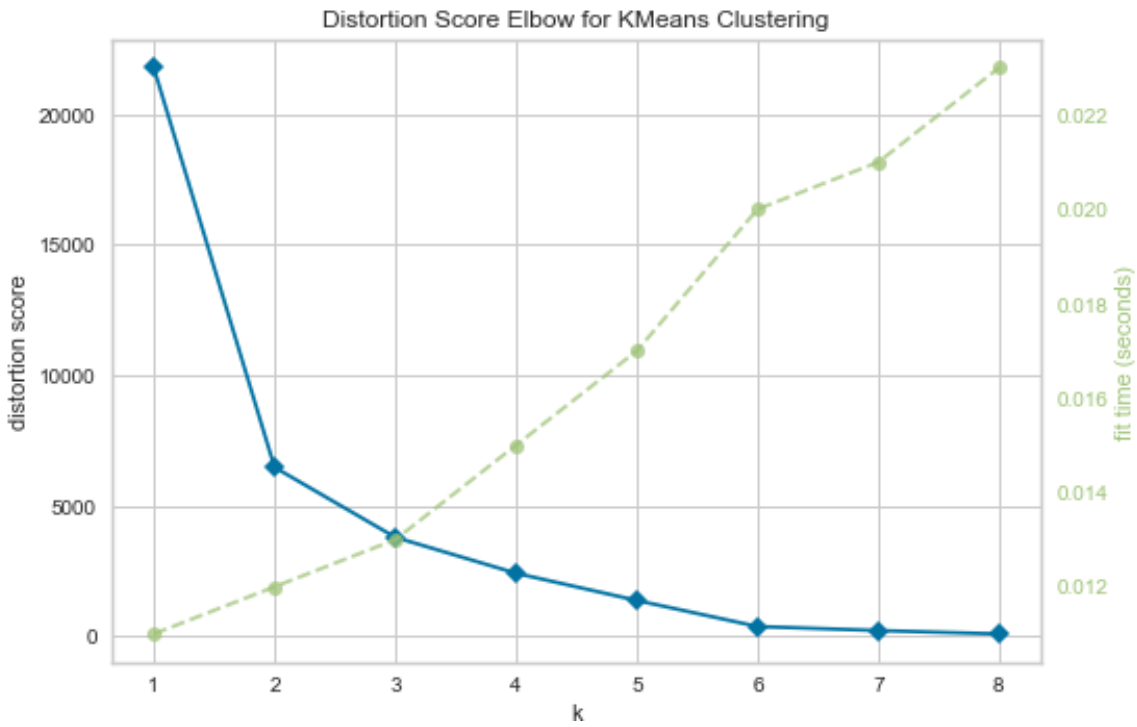
Normalized Count Plot

- The venue counts are normalized.
- The fewer number of venue categories are removed for better analysis.
- After normalization it is clear that there are good distribution of professional places and shops & services.



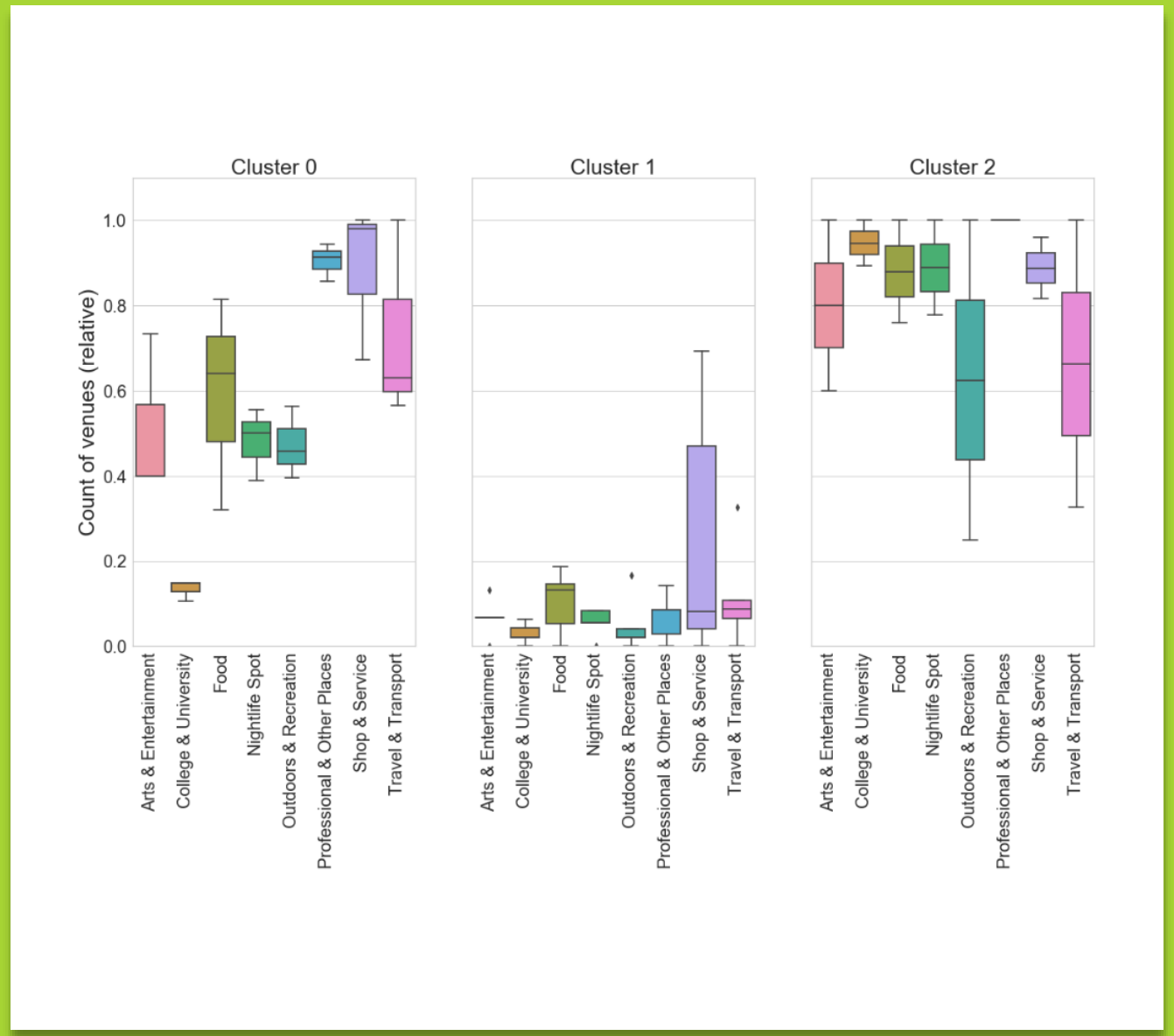
Elbow Method

- Elbow method showing the number of clusters good for clustering algorithm KMeans.
- Here there is significant change in distortion score with 2 clusters.
- But depending on our age groups and no significant change in distortion score we can choose to separate the postal codes into 3 clusters using Kmeans algorithm.



Clusters

- There are three clusters
- Cluster 0
  - These regions have less venues for College and University. But high number of venues in category professional & other places along with average number of night spots.
- Cluster 1
  - These regions are low in number for all category venues except for shops and service which are low to moderate in number.
- Cluster 2
  - These regions are highly populated with all kinds of venues especially college & university and Nightlife Spots.



Conclusion

- Depending on the Analysis on the previous slide the Clusters can be assigned to the age groups as in the table.
- 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,

Age Group	Postal Codes
Students (18-30)	52062, 52074 (Cluster 2)
Working People (31-50)	52064, 52066, 52070 (Cluster 0)
Older citizens (50+)	52068, 52072, 52076, 52078, 52080 (Cluster 1)