Unsupervised Learning Final Exam
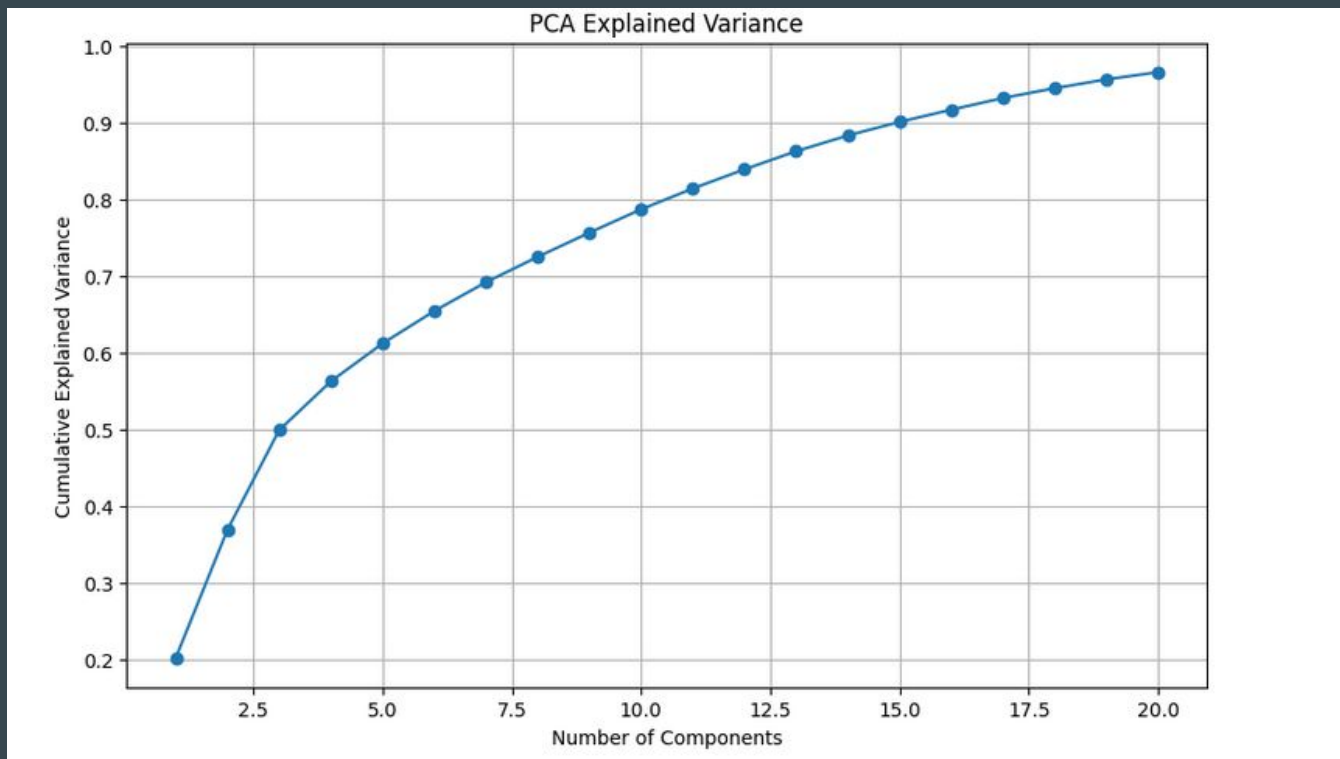
●●●

US Counties clustering analysis

- Clustering analysis of US Counties using socio-economic data

- Goal is to identify counties with similar characteristics
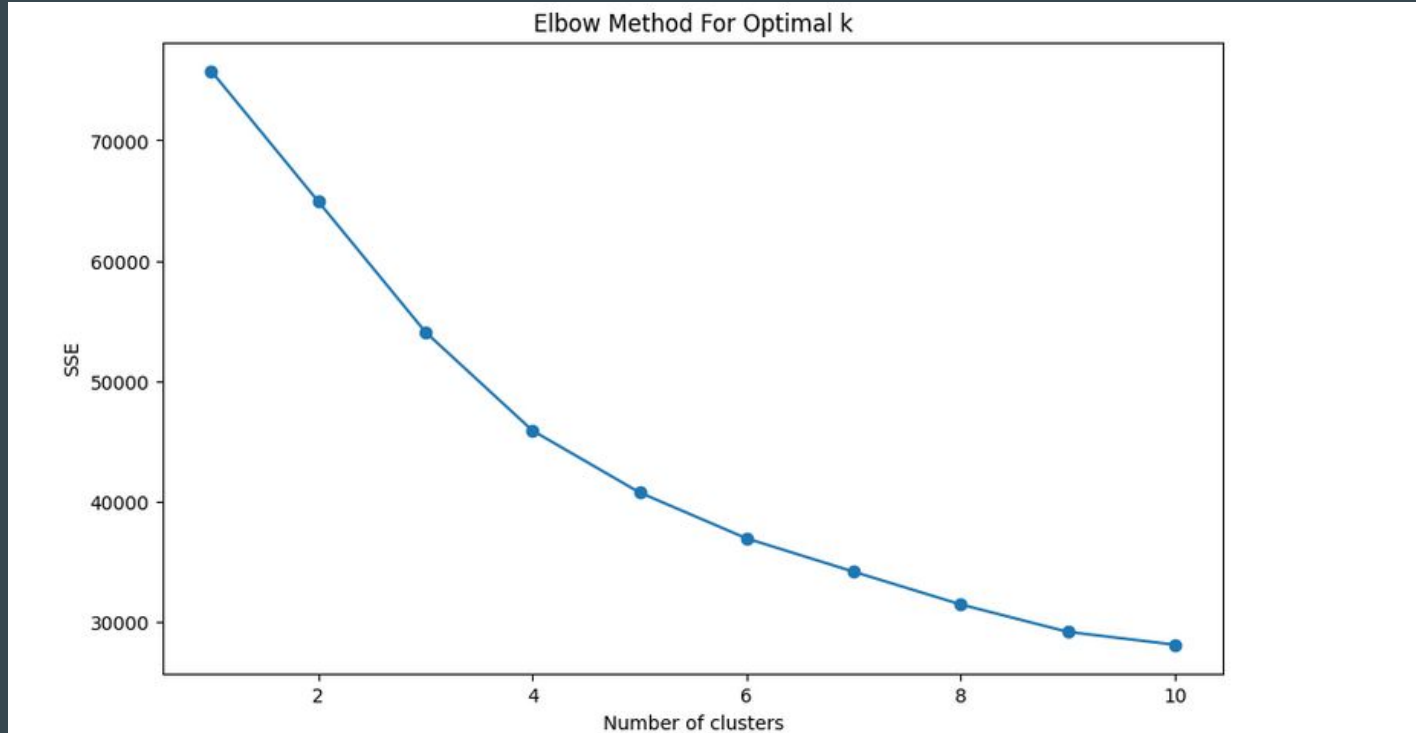
What's the data look like ?

	CountyId	State	County	TotalPop	Men	Women	Hispanic	White	Black	Native	...	Walk	OtherTransp	WorkAtHome	MeanCommute	E
0	1001	Alabama	Autauga County	55036	26899	28137	2.7	75.4	18.9	0.3	...	0.6	1.3	2.5	25.8	
1	1003	Alabama	Baldwin County	203360	99527	103833	4.4	83.1	9.5	0.8	...	0.8	1.1	5.6	27.0	
2	1005	Alabama	Barbour County	26201	13976	12225	4.2	45.7	47.8	0.2	...	2.2	1.7	1.3	23.4	
3	1007	Alabama	Bibb County	22580	12251	10329	2.4	74.6	22.0	0.4	...	0.3	1.7	1.5	30.0	
4	1009	Alabama	Blount County	57667	28490	29177	9.0	87.4	1.5	0.3	...	0.4	0.4	2.1	35.0	

5 rows × 37 columns

PCA Analysis

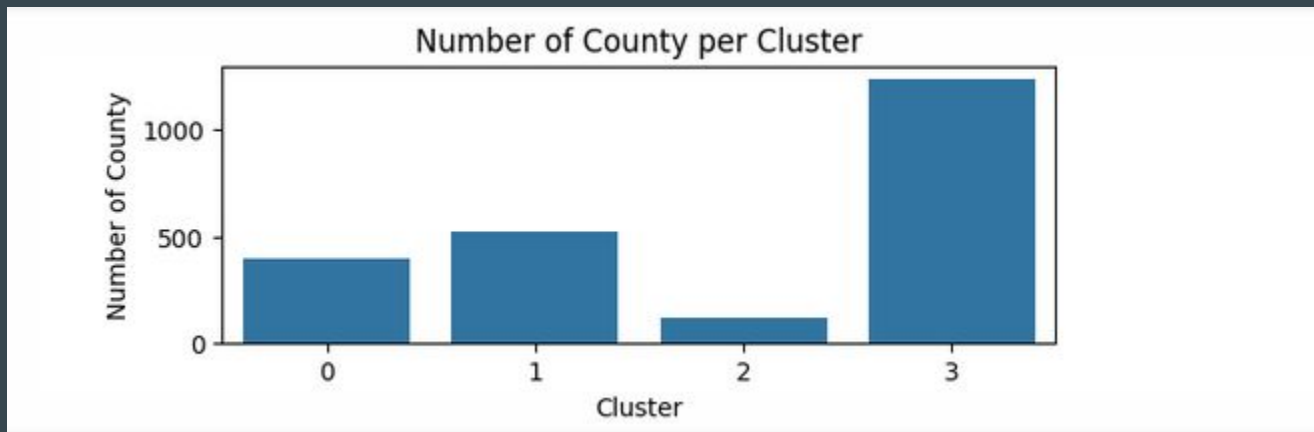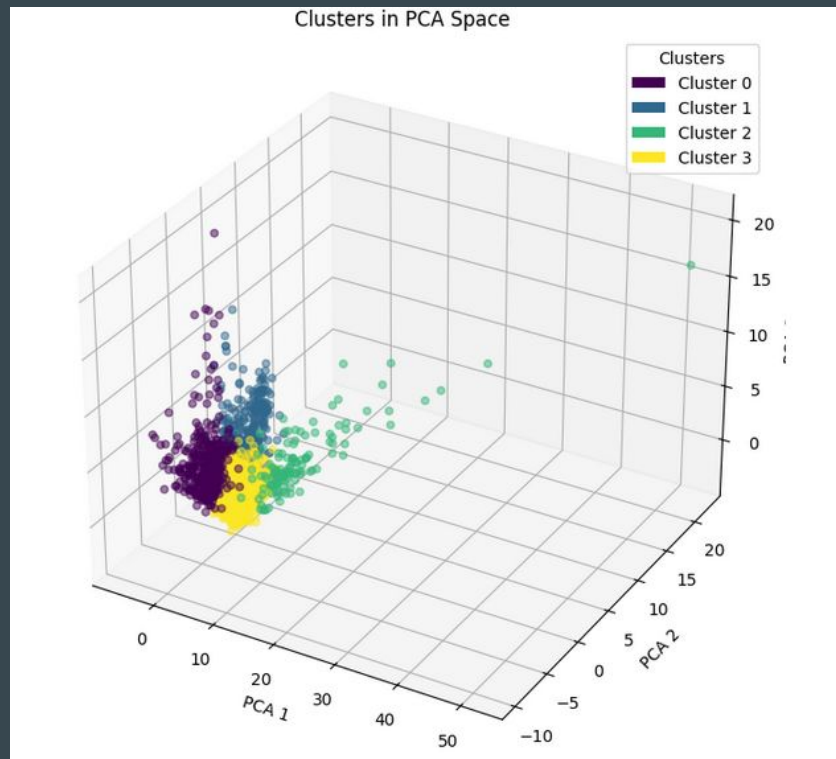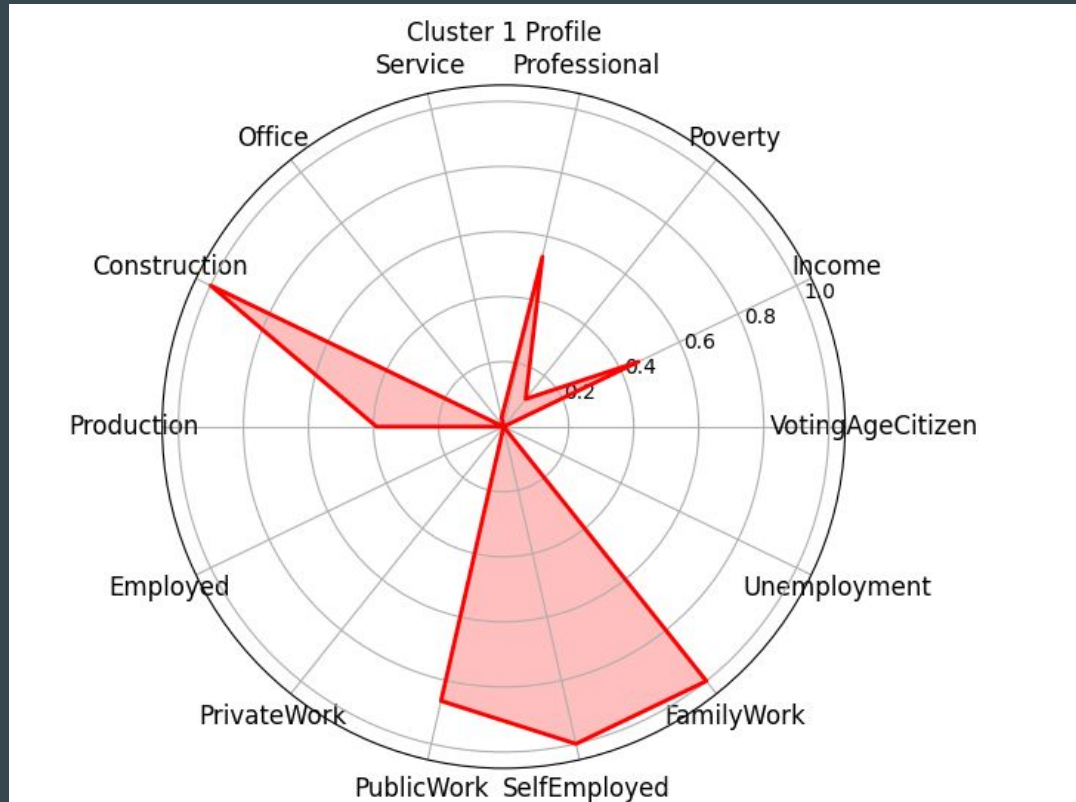# Clustering analysis

Elbow Method For Optimal k
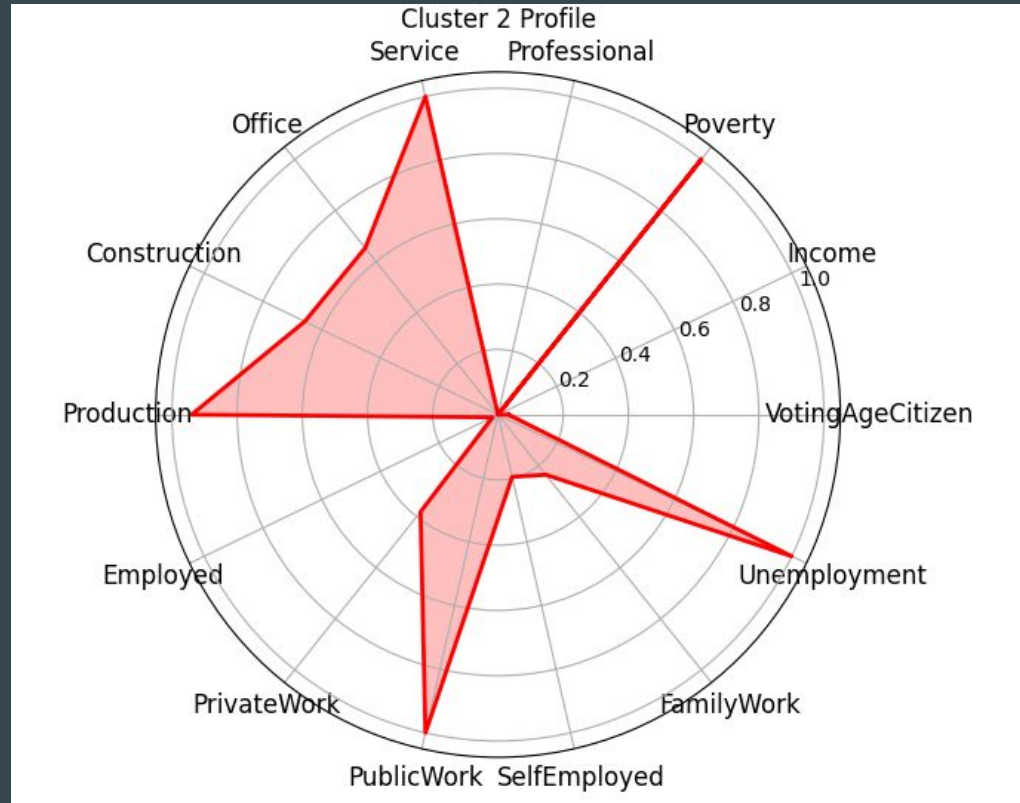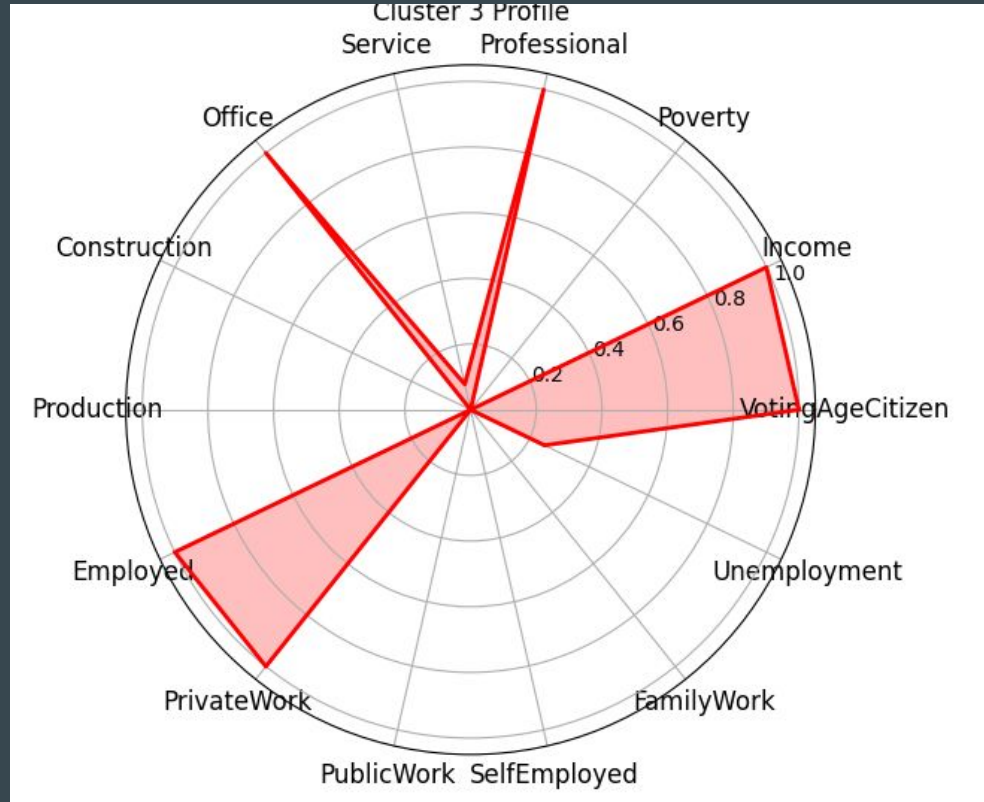
Clustering analysis

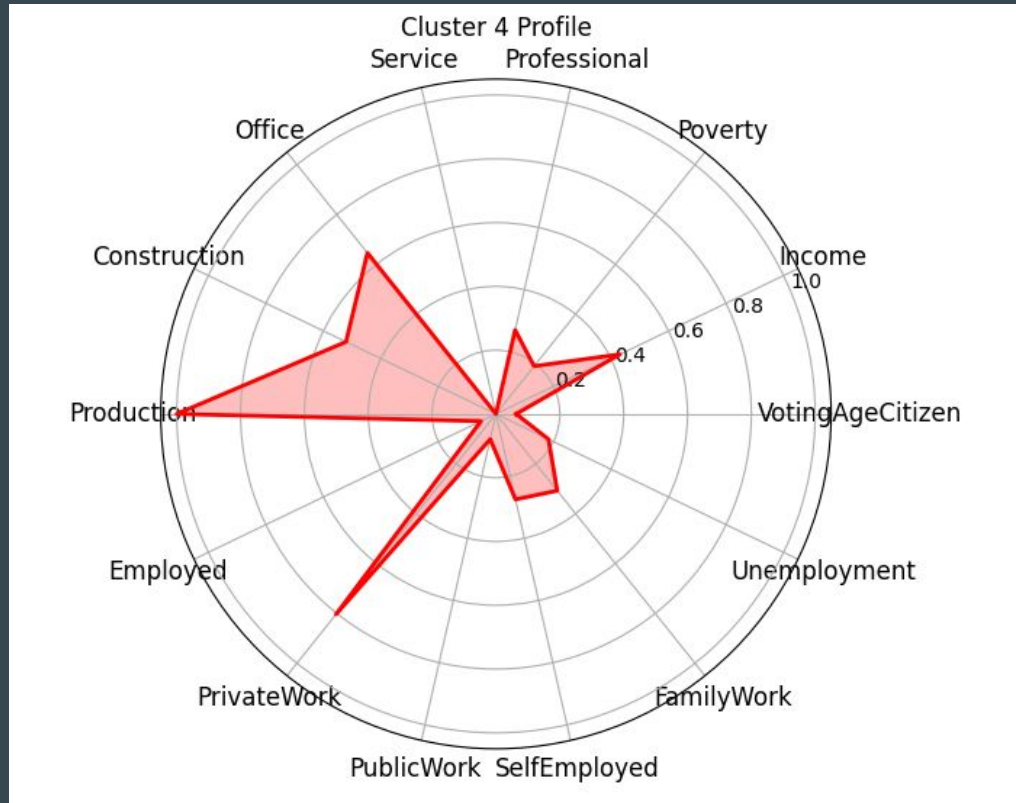# Clustering analysis

Clustering analysis

Clustering analysis

Clustering analysis

Clustering analysis

Conclusion

- After reducing our dataset to 7 principal components we have been able to cluster US counties into 4 group.

- Areas of improvement
  - Increase number of PCs
  - Other clustering methods