

Business Case 1

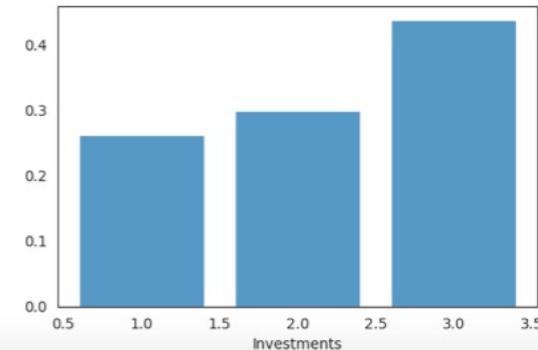
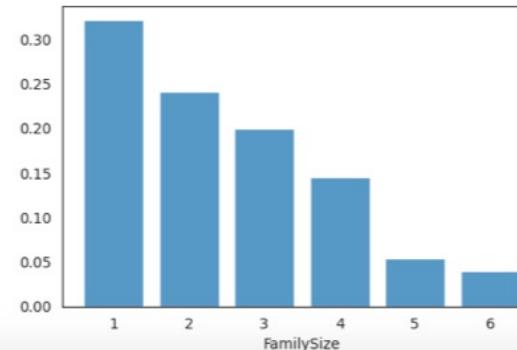
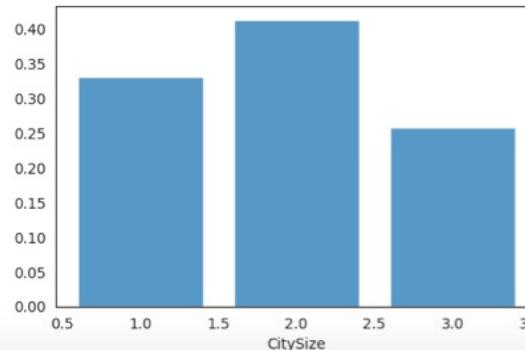
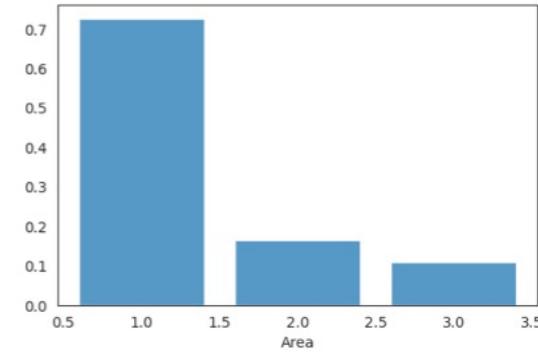
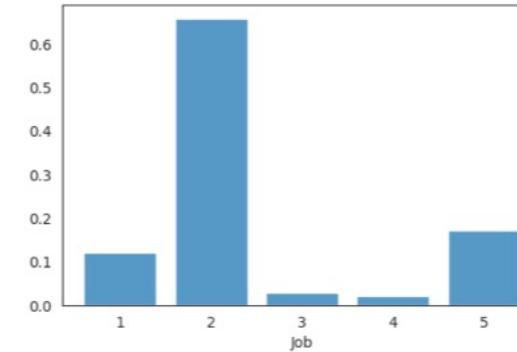
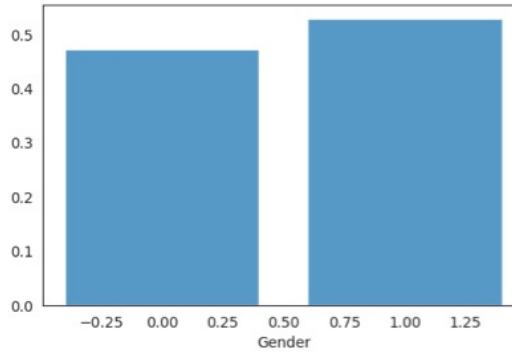
Group 14

Lazzarelli Chiara, Polidori Alberto,
Rizzo Giovanni, Shubhankar Kumar

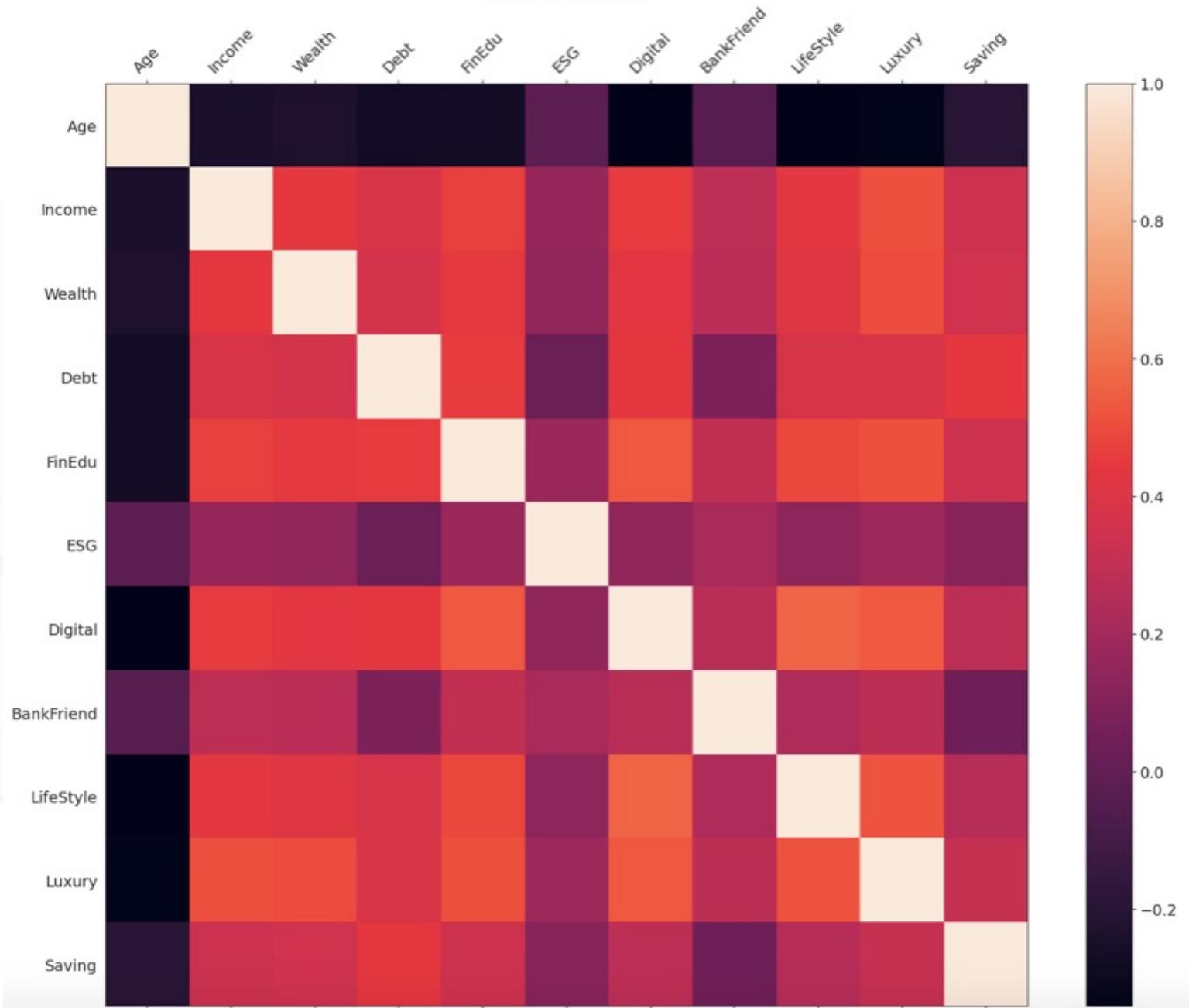


Exploratory analysis

- 17 covariates, 11 numerical and 6 categorical
- Quick look at the distribution of the categorical features:

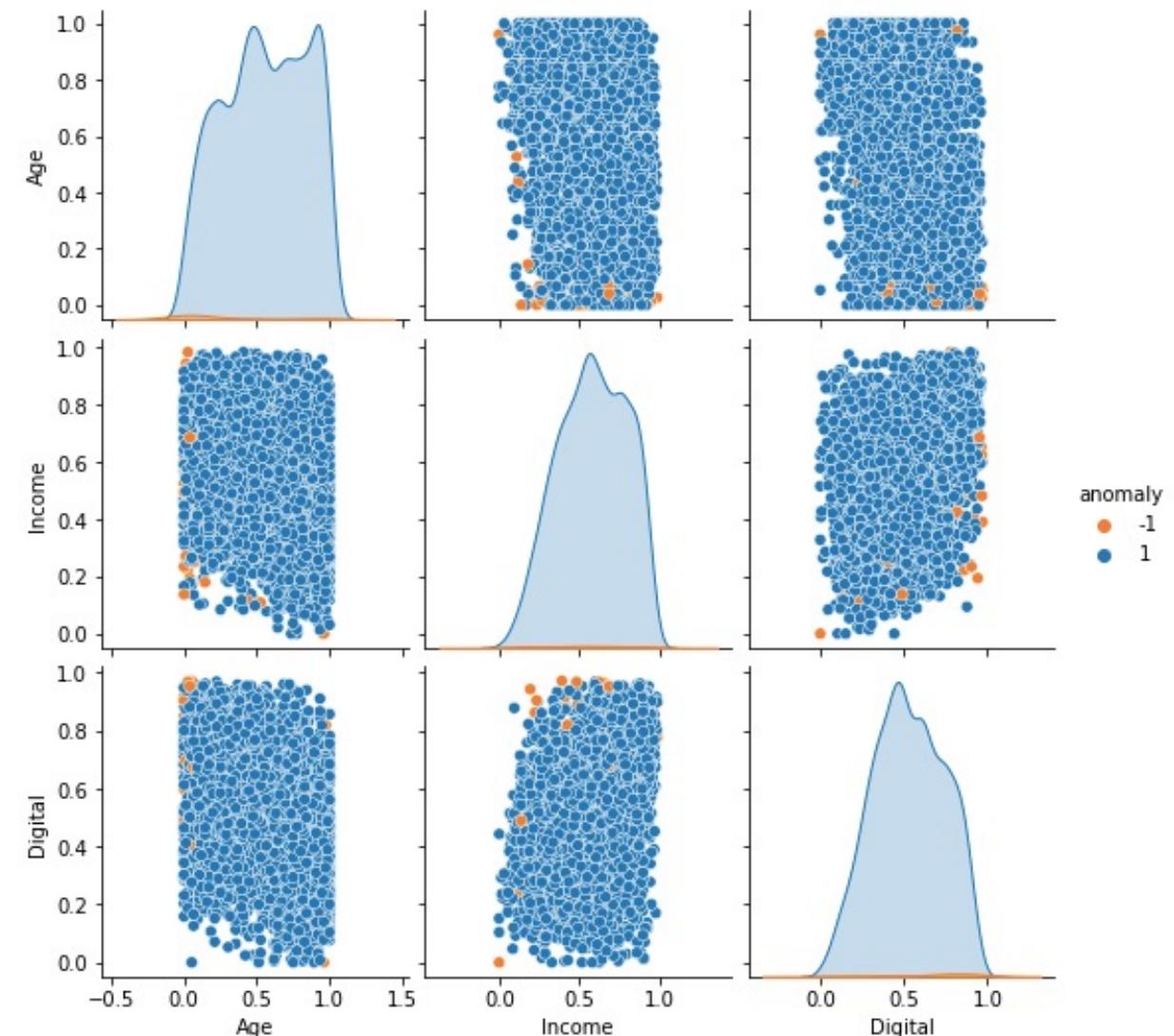


- Some of the numerical features (eg. ‘Wealth’ and ‘Income’) are highly correlated as one may expect



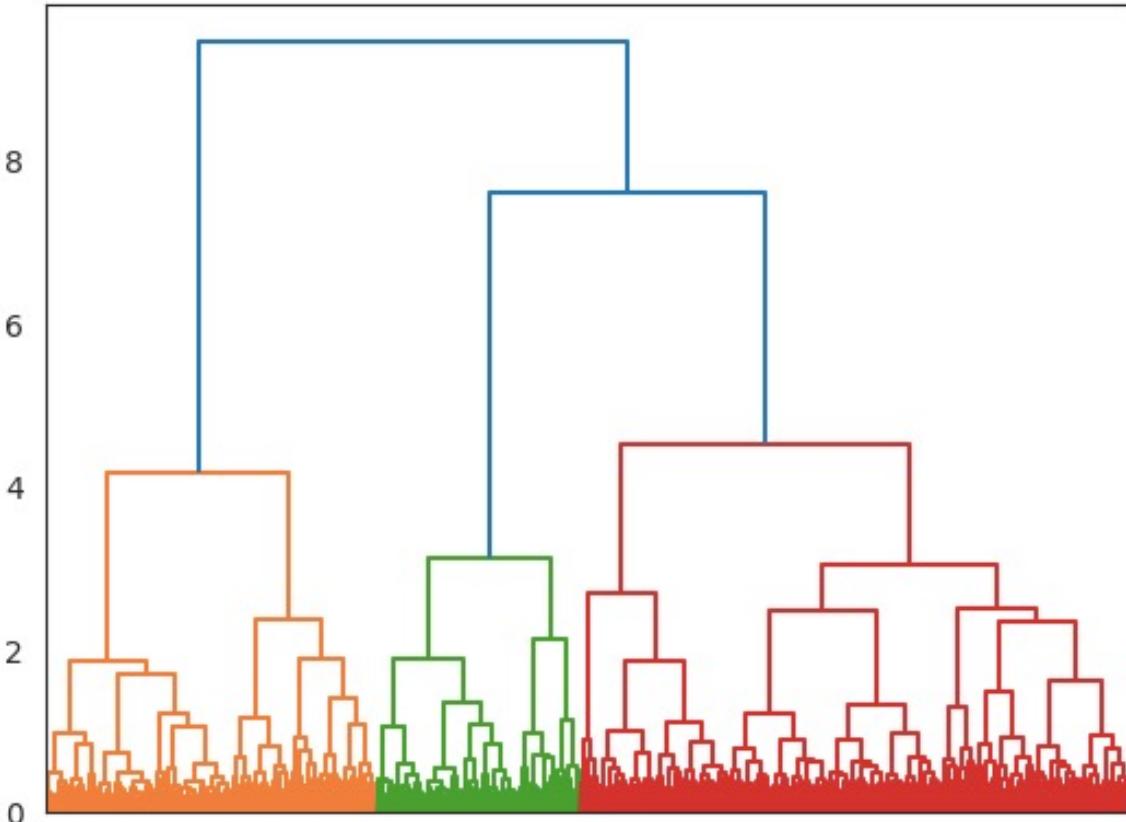
Outlier analysis

- Our goal is to keep the data in their totality, we just want to check that there are not fat-finger errors
- Isolation forest with the hypothesis that 1% of the data is outlier
- No strong evidence, we kept all the data

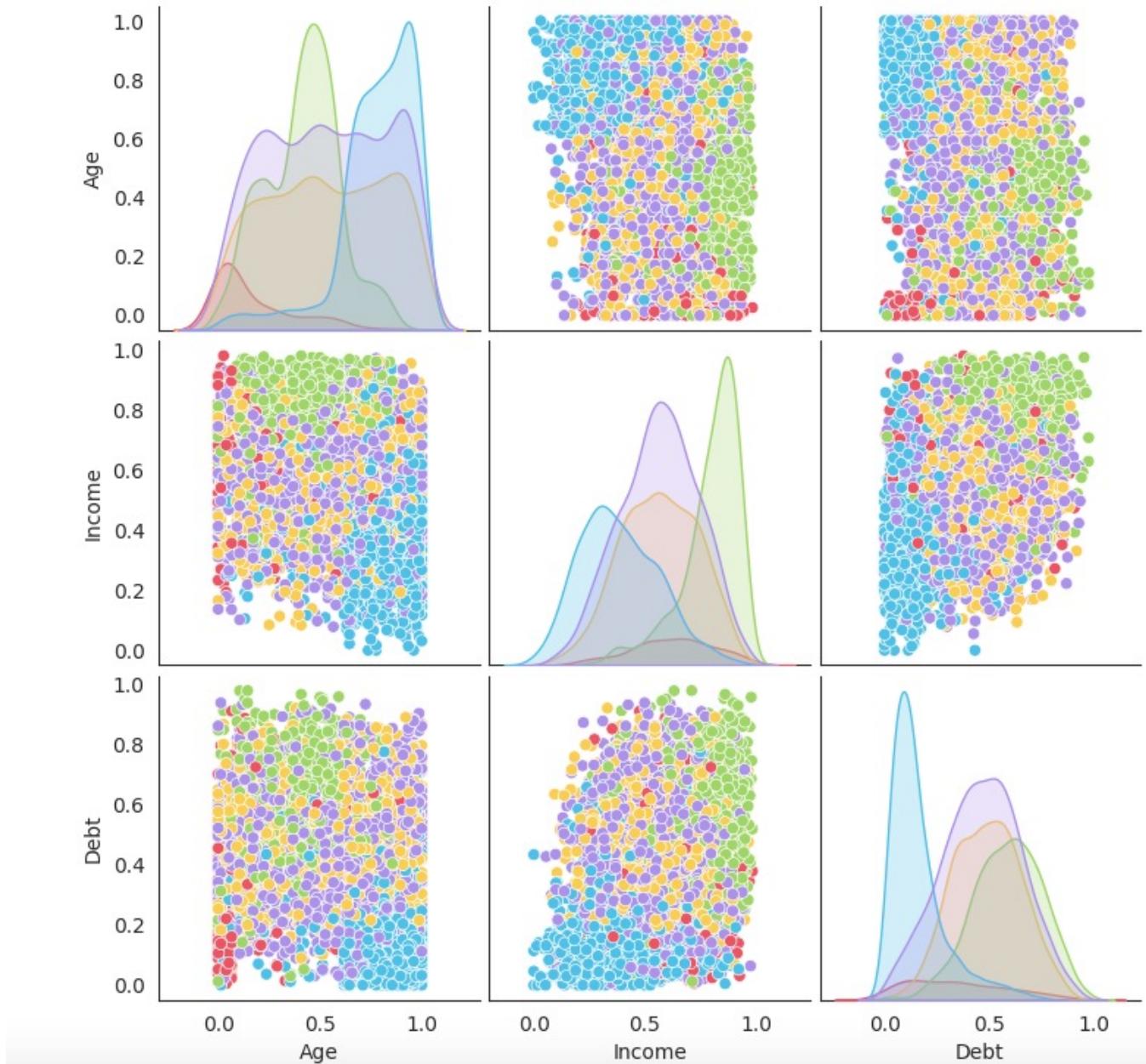


Cluster Model– Hierarchical (with categorical)

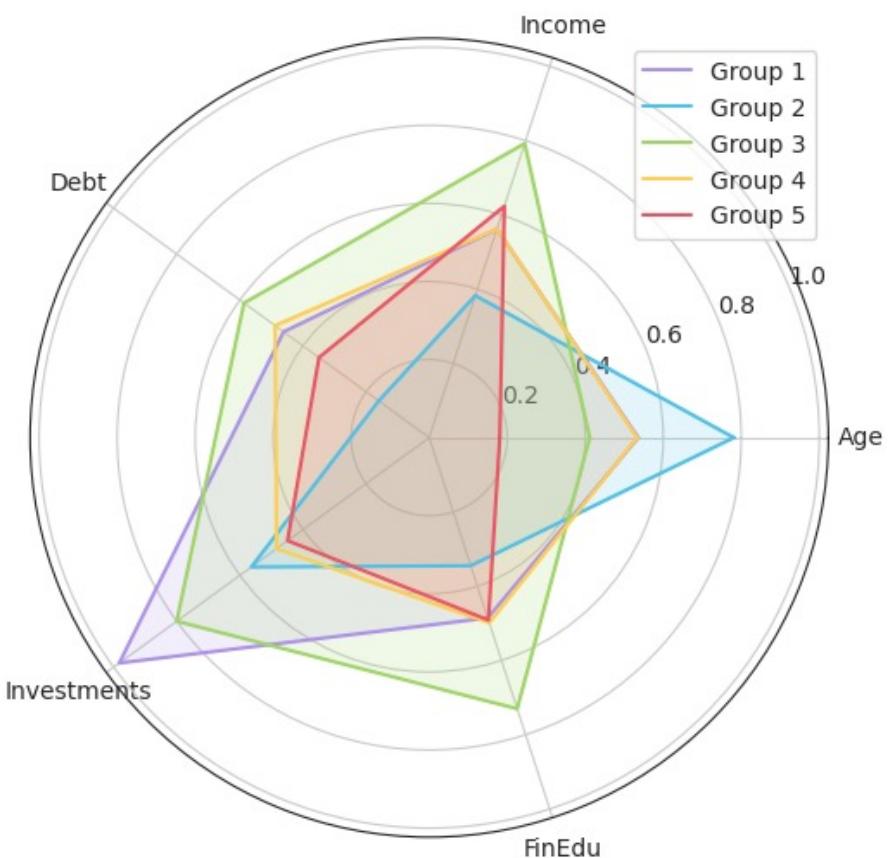
- Both numerical and categorical covariates
- We removed ‘Gender’ and ‘Area’
- Gower distance
- Ward linkage
- Cut at 5 in order to catch more distinctive elements



SCATTER PLOT BY GROUPS



RADAR PLOT AND SUMMARY AVERAGE TABLE

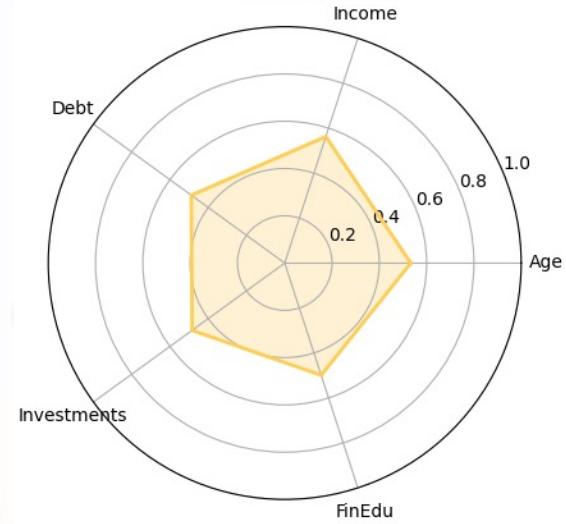
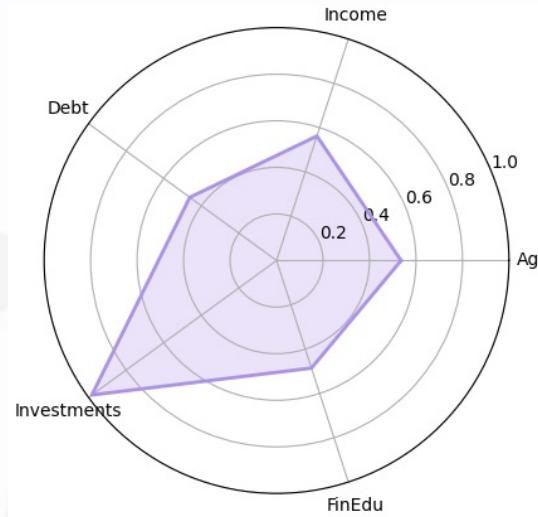


GROUP	AGE	GENDER	FAM SIZE	INCOME
1	60	0.55	2.74	0.56
2	78	0.57	2.23	0.38
3	50	0.44	1.98	0.79
4	60	0.50	2.82	0.56
5	32	0.72	2.30	0.62
GROUP	WEALTH	DEBT	FIN EDU	INVESTING
1	0.56	0.46	0.49	2.95
2	0.38	0.16	0.34	1.70
3	0.80	0.59	0.73	2.40
4	0.57	0.49	0.50	1.45
5	0.58	0.35	0.49	1.34

GROUPS 1 & 4

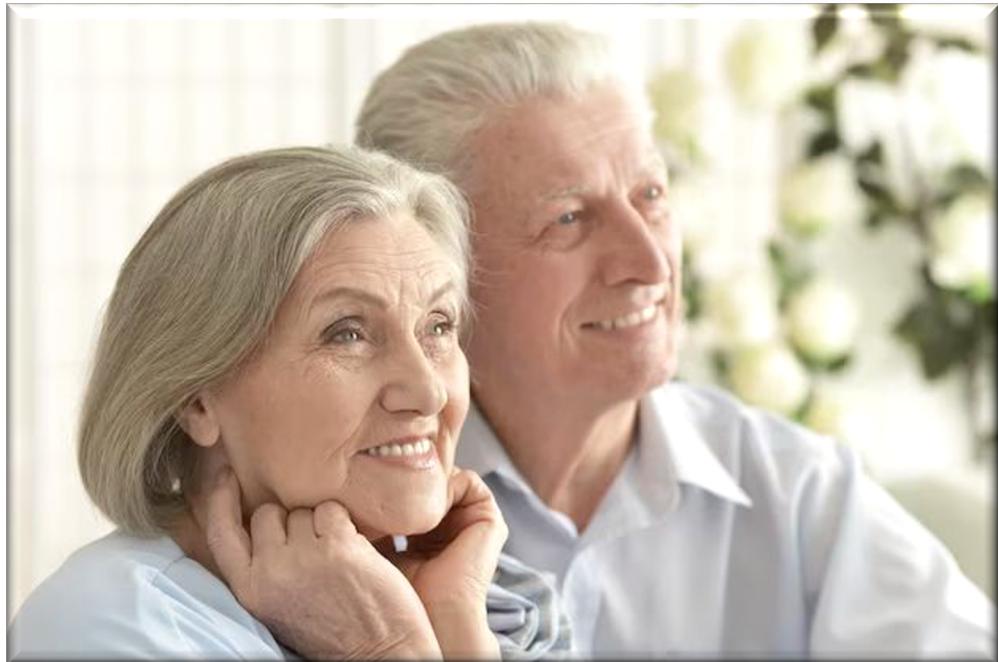


office worker with family

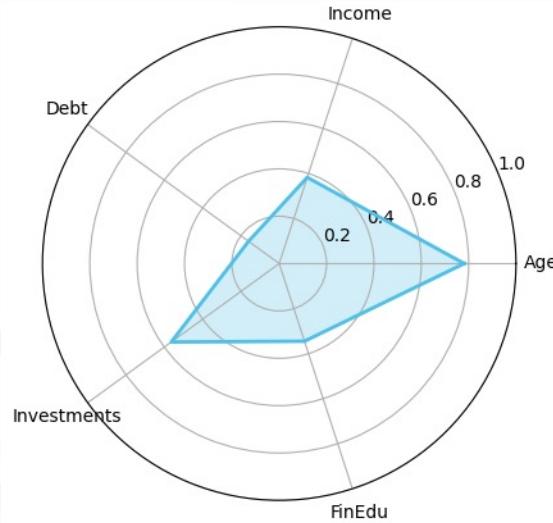


- people of all ages
- higher average number of family members
- standard wages and wealth
- 1 & 4 deeply differ in their investment approach

GROUP 2



retired couple

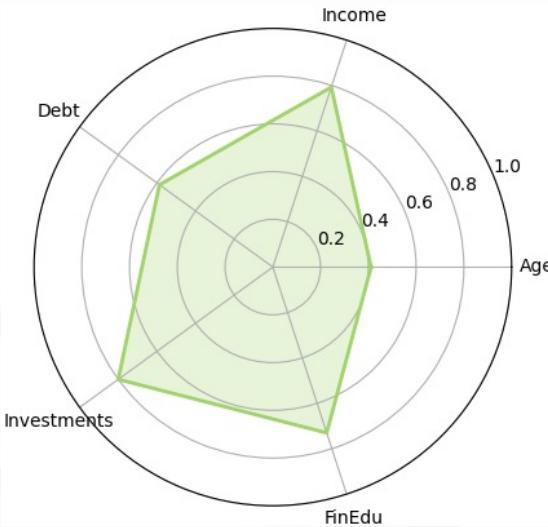


- older people
- most of them retired
- little or no debts

GROUP 3

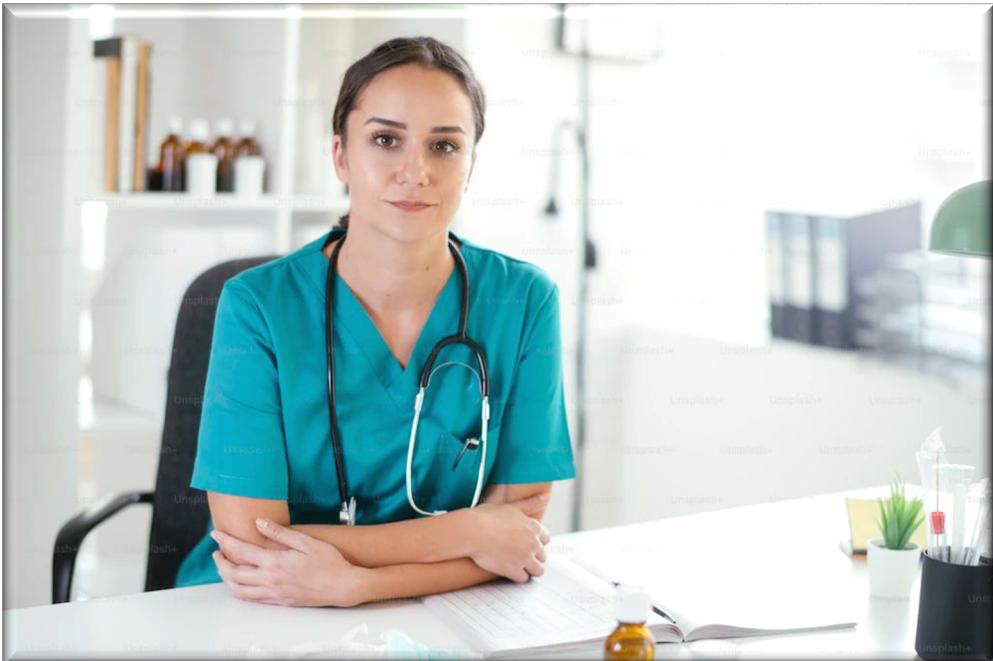


manager

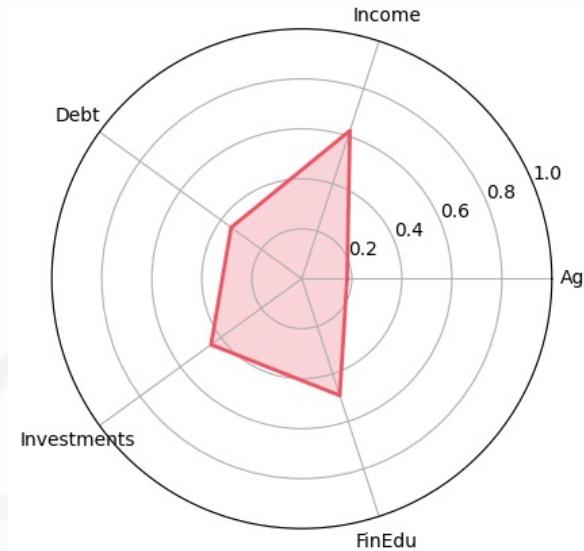


- very wealthy people, 50yrs average
- high income
- high financial education
- higher debts but it's ok since they're likely to pay them back

GROUP 5



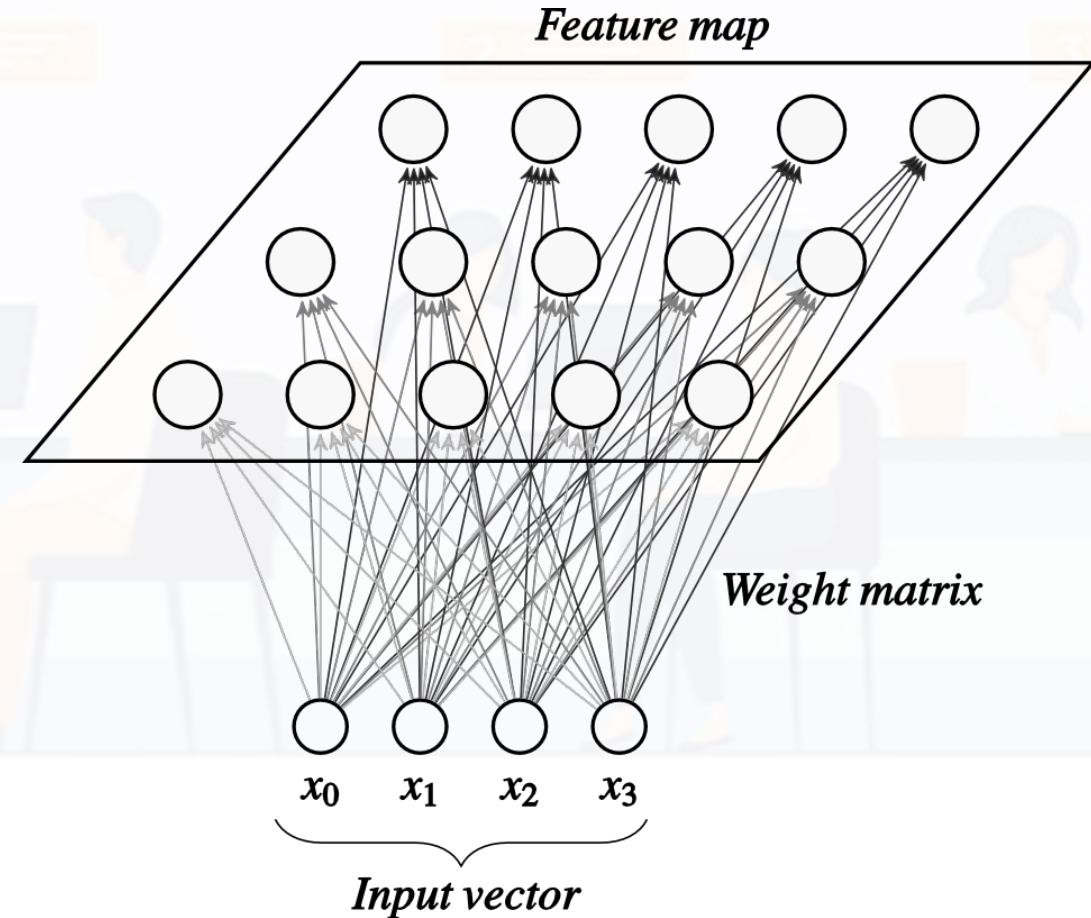
nurse



- very young people, 30yrs average
- quite good income
- mostly females (70%, notice that the gender has not been used for clustering)
- probably ex college students

SOM as a final check

- Basic 2x2 neurons lattice grid
- All 4 neurons as clusters
- It's just a check: we want only to see if the NN catches patterns similar to our hierarchical clusters



SUMMARY AVERAGE TABLE

SOM - C	AGE	FAM SIZE	INCOME	WEALTH	FIN EDU	INVESTING
1	0.57	2.76	0.53	0.54	0.46	1.26
2	0.40	1.98	0.74	0.73	0.67	2.40
3	0.53	2.90	0.56	0.56	0.49	2.83
4	0.80	2.08	0.39	0.39	0.35	1.80

We can find similar data separation pattern between SOM and Hierarchical Clustering:

- C1 very similar to G4
- C2 is a mixture of G3 and G5
- C3 same characteristics of G1
- C4 is retired people like G2