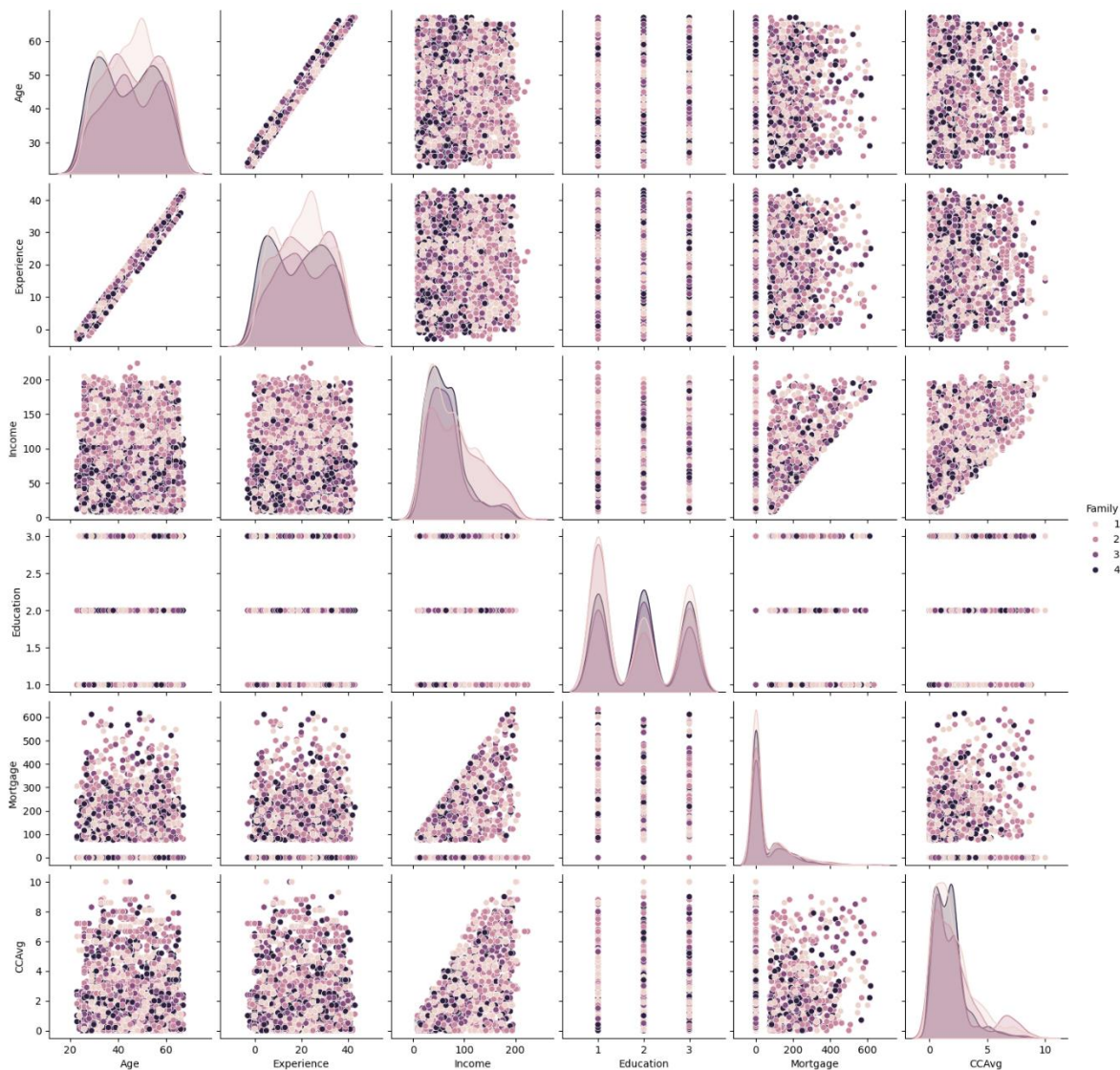


### Reporte 3 Clasificación

Para esta práctica, trabajé con un archivo obtenido de Kaggle (<https://www.kaggle.com/datasets/luisenriquesguerrero/creditos-personales-actualizado>), una base de datos donde muestra a clientes de un banco, ordenados por su ID, edad, experiencia laboral, ingreso y entre otros datos, la relación entre sus ingresos y gastos y si tienen algún préstamo, en resumen, un pequeño balance de su situación económica.

Para hacer más entendible la clasificación, omití algunas variables de la base original, considerando a los dependientes familiares como el punto de referencia para agrupar los datos restantes, (Edad, experiencia laboral, Ingreso, Nivel educativo, deuda hipotecaria y gastos), bajo la asunción de que dependiendo los familiares que tuvieran como dependientes serían los causantes de las otras variables, claro que eso no es un dato del todo correcto.

Relación de la variable “Dependientes familiares” con el resto, ya mencionadas



Definitivamente no hay una separación clara en el resto de las variables a causa de los dependientes familiares,

Mediante una regresión logística, hice la clasificación de mis datos y obtuve lo siguiente

precision		recall	f1-score	support	
1	0.29	0.54	0.38	362	
2	0.45	0.17	0.24	346	
3	0.00	0.00	0.00	258	
4	0.28	0.45	0.34	284	
accuracy			0.30	1250	
macro avg		0.25	0.29	0.24	1250
weighted avg		0.27	0.30	0.26	1250
		[[194 45 0 123] [180 58 0 108] [142 13 0 103] [142 14 0 128]]			
Precisión: 0.304					

Tomando en cuenta el gráfico previo al resultado, creo que es una precisión adecuada, al ser una cantidad de variables muy mezclada entre si.