**CLUSTERING**

*Math and wealth variables (family wealth and ESCS status)*

k-Means clustering: Ho provato a runnare kmeans su ciascun paese con valori diversi di k per osservare il valore ottimale di k rispetto alla diminuzione di “entropia” (i.e. somma delle distanze interne di ciascun cluster).

Chart, scatter chart

Description automatically generated

Si può notare che dopo 5 clusters la decrescita di entropia si riduce (il grafico è dell’Italia ma questo andamento è comune per tutti i paesi), per questo motivo ho scelto di tenere k = 5 (per il quale si ha una diminuzione di entropia del 60%).

Dopodichè andando ad analizzare i cluster si può notare come essi raggruppino studenti con features correlate (i.e. cluster 1 sei bravo e ricco, cluster 2 un po' meno bravo e più povero ecc..), andando quindi a confermare un risultato che avevamo già trovato precedentemente.

Inoltre si conferma anche il trend per cui i cluster “peggiori ” siano quelli contenenti la percentuale più alta di immigrati 🡪 per esempio il cluster peggiore del Lussemburgo è formato all’80% da immigrati.Diagram

Description automatically generated