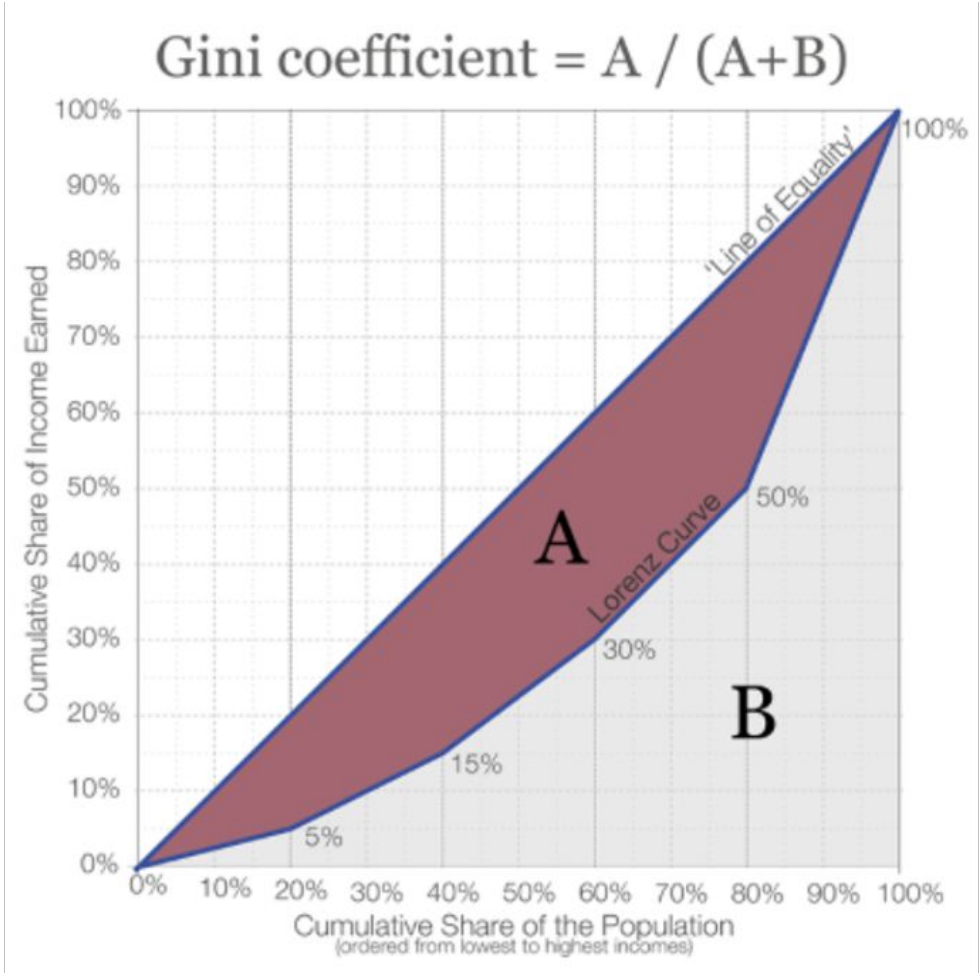
The Gini Index

Asad Sabir

What is the Gini Index

- The Gini Index, or Gini Coefficient is a measure of how close the income distribution is to perfect equality or perfect inequality. A Gini Index of 0 indicates perfect equality while a Gini Index of 100 indicates perfect inequality.

- The graph shows how the Gini Index is calculated. As inequality increases, The Lorenz curve moves further out and increases area 'A'. As inequality decreases, the Lorenz curve approaches the line of equality and decreases area ‘A’.



Understanding wealth inequality

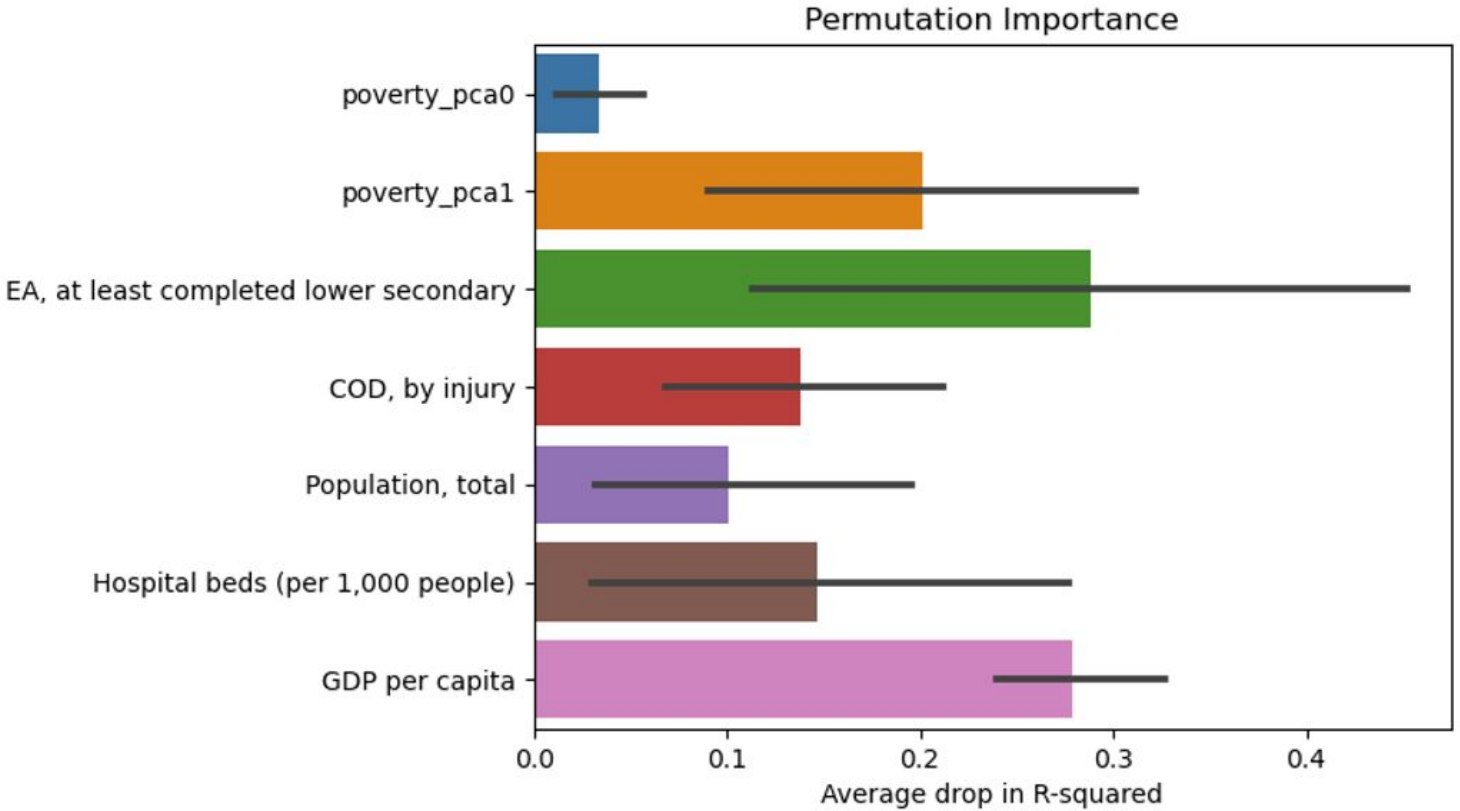
- Poverty and wealth inequality are significant global concerns that have far-reaching socio-economic implications.
- The causes, effects, and associations of income distributions in different countries are difficult to understand on a macro scale.
- By using data from the World Bank, this project aims to deepen that understanding and help inform both public opinion and policy.
- In this analysis, I mainly focused on health, education and macro-economic indicators.

The Data

The world bank has over 1400 indicators to choose from. I analyzed 27 of them and my final model uses these Seven variables:

- Poverty PCA 0 (Poverty magnitude)
- Poverty PCA 1 (Poverty severity)
- GDP per capita
- Educational attainment, at least completed lower secondary
- Cause of Death, by injury (%)
- Population, total
- Hospital beds (per 1,000 people)

Variable Importance

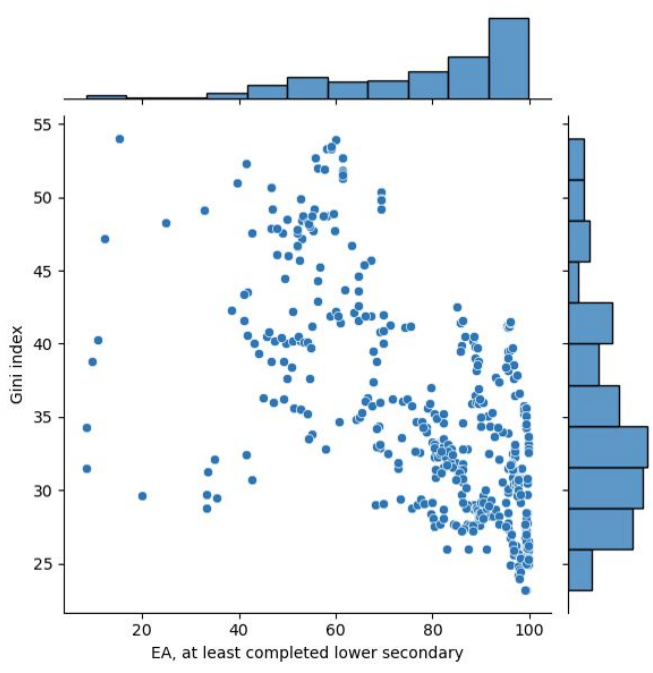


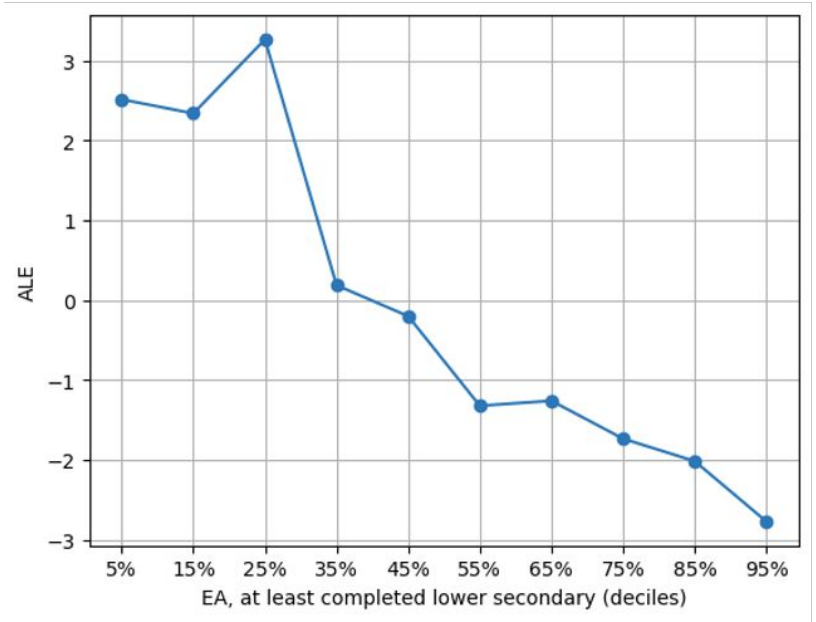
- The final model is an XGBoost with an R-squared of 0.68

Accumulated Local Effects (ALE)

- I used ALE plots to examine the relationship between variables and Gini Index.
- Accumulated local effects describe how features influence the prediction of a machine learning model on average.
- The y-axis can be interpreted as the difference in the features prediction from the average prediction on the data.
- For example, an ALE of 2 means we can expect the Gini index to be about 2 points higher than average by having that feature value.

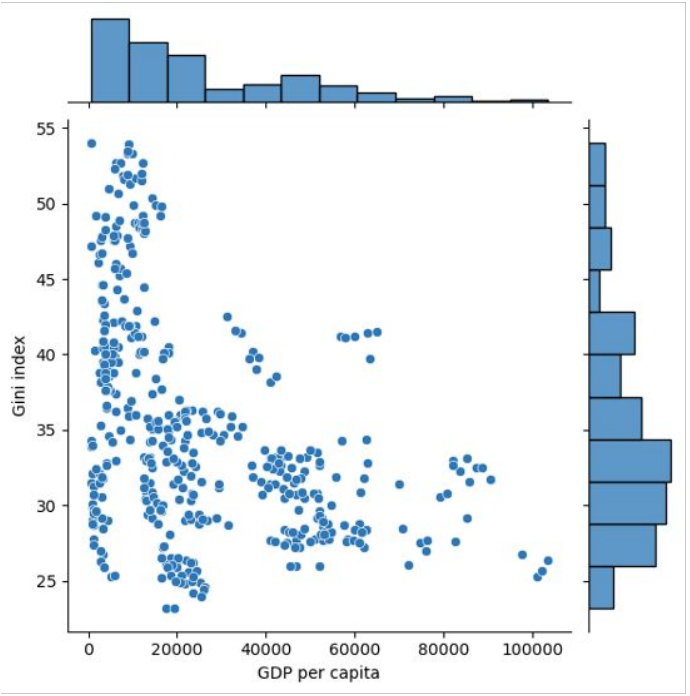
Educational Attainment, at least completed secondary

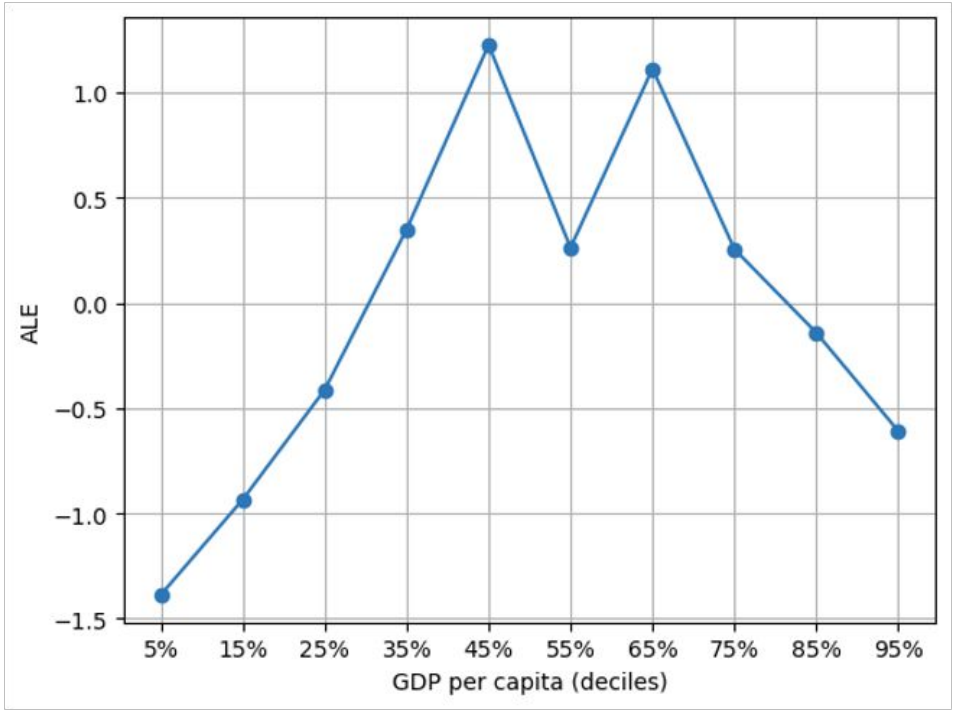




- This indicator is a measure of the percentage of population ages 25 and over that attained or completed lower secondary education.
- Countries at higher levels are associated with a lower Gini index and lower wealth inequality.
- The data for education is mapped to the International Standard Classification of Education (ISCED) to ensure the comparability of education programs at the international level.
- The median value for this indicator is at 85% and the 75th percentile is at 96%. This shows that even for countries where this measure is high, the association is still strong and increasing the educational attainment rate could help decrease wealth inequality.

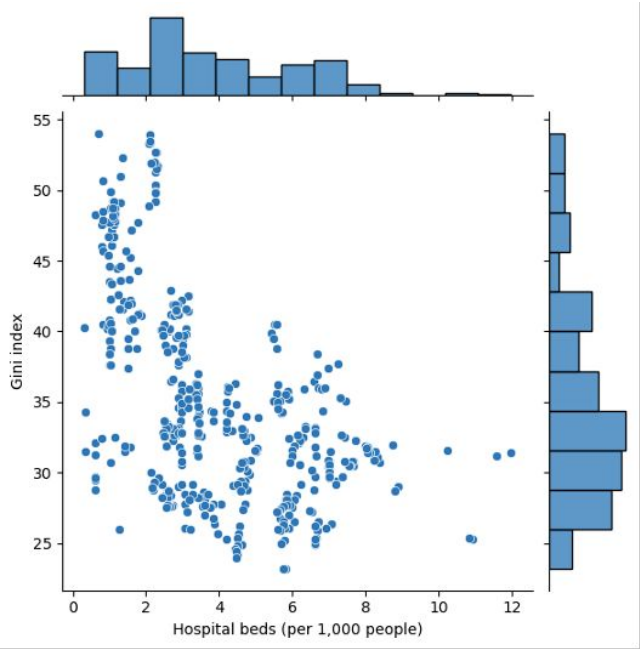
GDP per capita

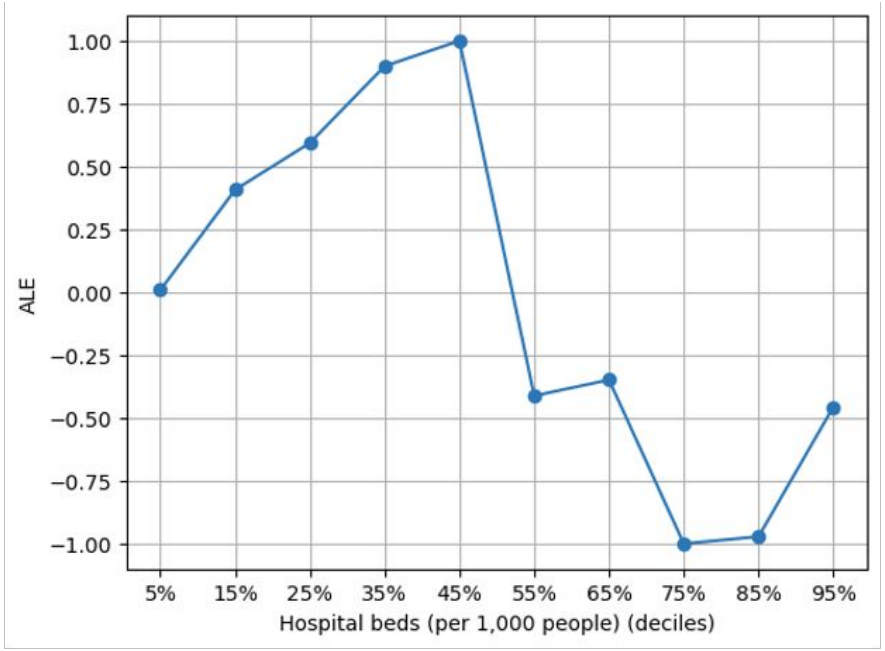




- GDP per capita represents the value of all the goods and services produced by a country (GDP) divided by its population.
- This indicator is often used to measure the general prosperity of a country.
- The ALE shows that increasing GDP per capita to about the median level, increases wealth inequality. However, after that point, the trend inverses.

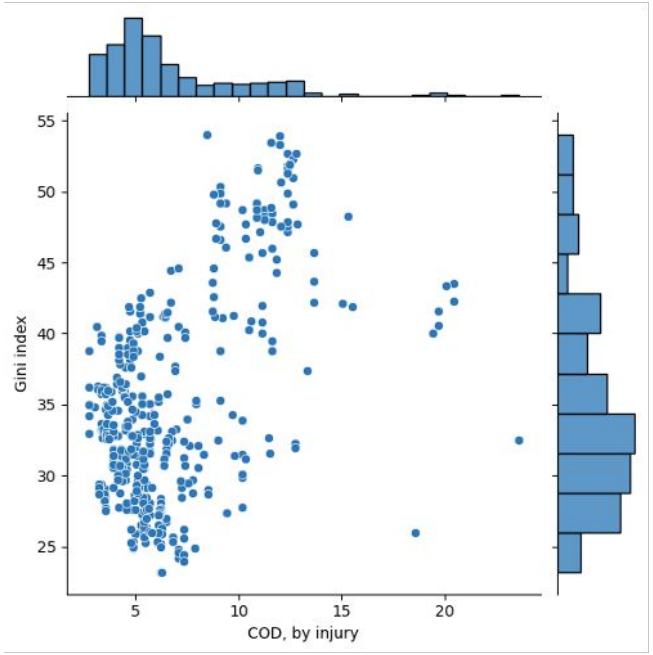
Hospital Beds (per 1,000 people)

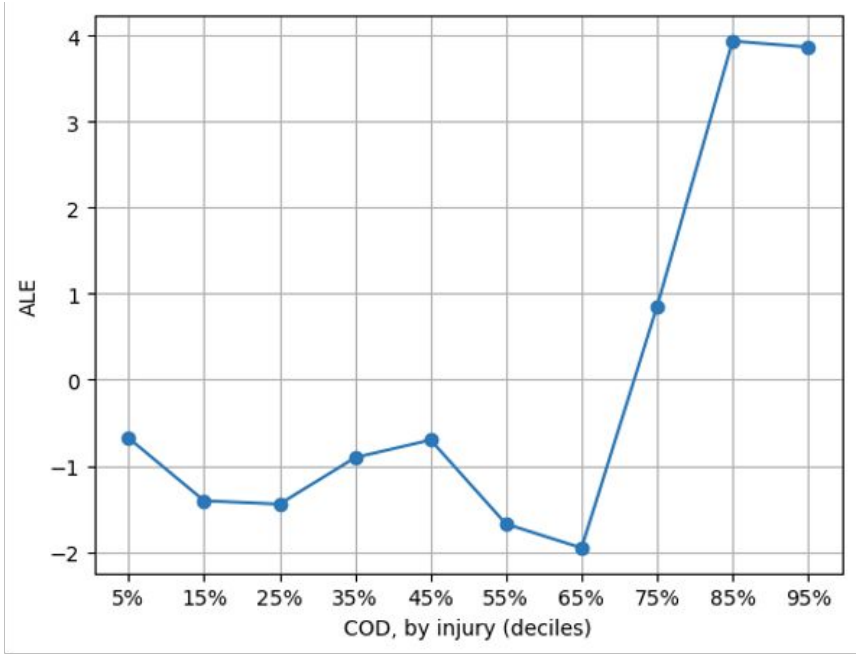




- Hospital beds per 1,000 people is one of many key indicators tracked by the World Health Organization (WHO) to evaluate health systems.
- This feature is difficult to interpret on its own as it reflects both demand and supply side factors.
- For example, a low value could represent a healthy population or the inadequacy of a health system to provide enough beds. Whereas a high value could be because of an epidemic.

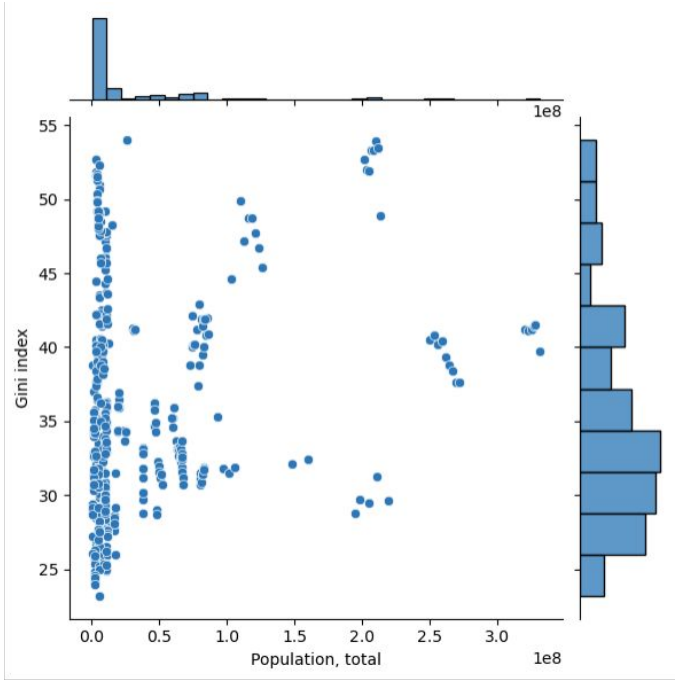
Cause of Death, by injury

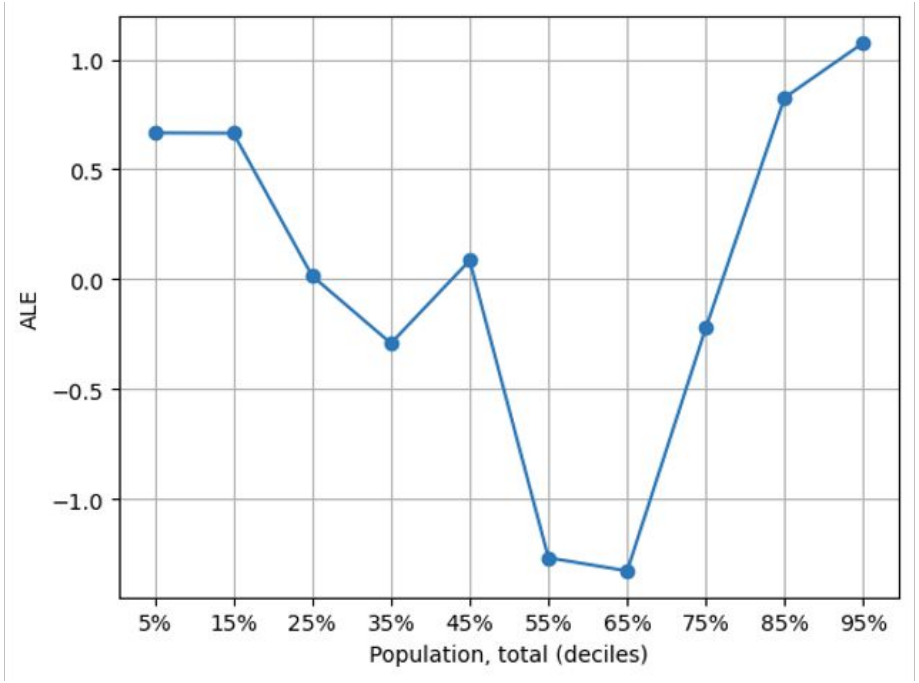




- Cause of Death, by injury, refers to the percent share of all deaths for all ages by intentional or unintentional injuries.
- The ALE plot seems to plateau or even decrease with this indicator until the 65th percentile mark which is about 7.4%. After that we see a sharp increase in the predicted Gini index.

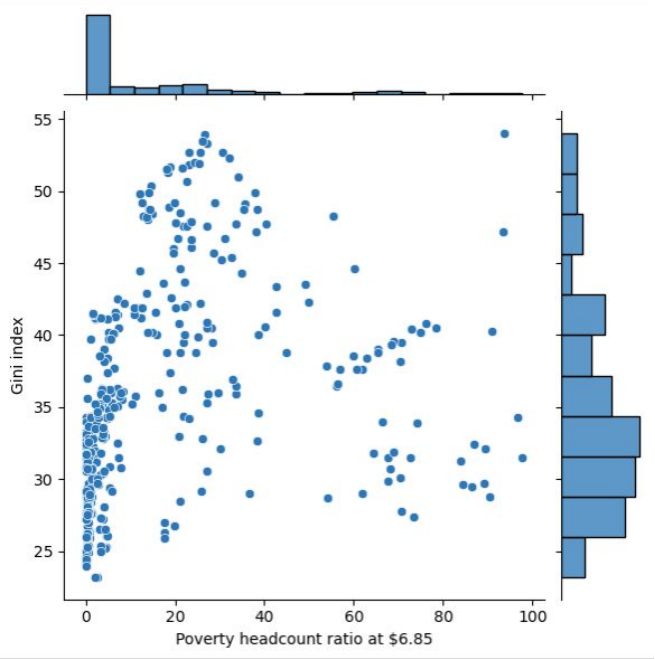
Population, total

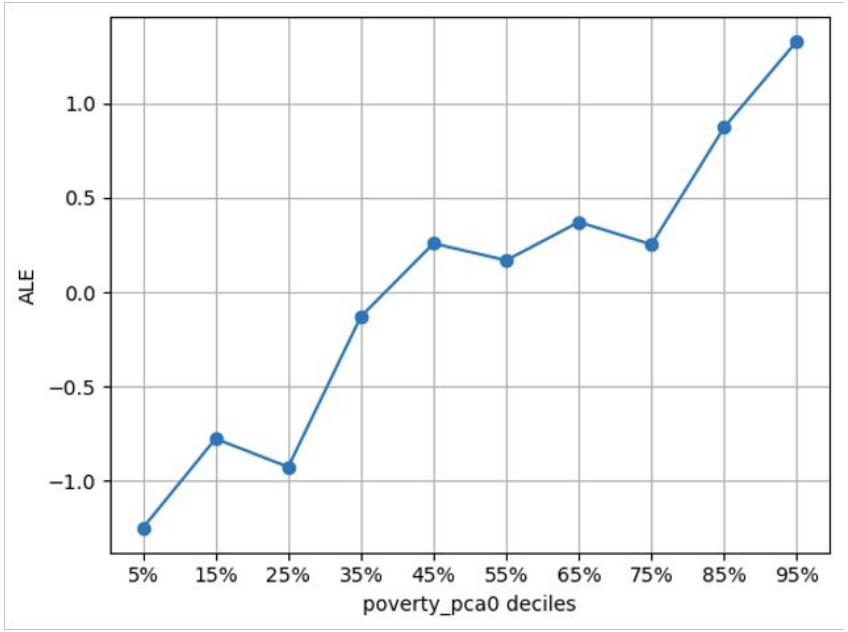




- This indicator counts all residents regardless of legal status or citizenship.
- The ALE plot suggests that there is an optimal level of population that keeps the Gini index low. A population higher or lower than that level, increases the Gini index.
- That level, which seems close to the 55th to 65th percentile mark, is around 11,000,000 people.

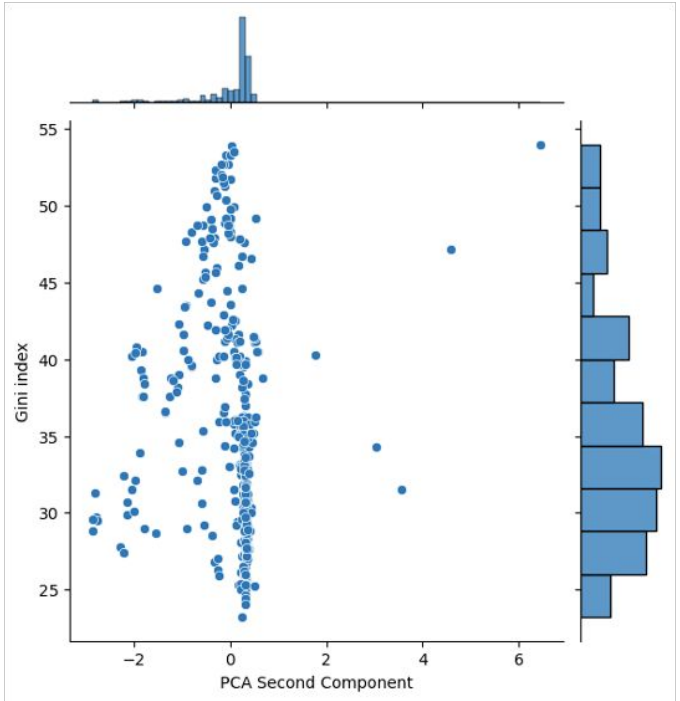
Average level of poverty (Poverty PCA 0)

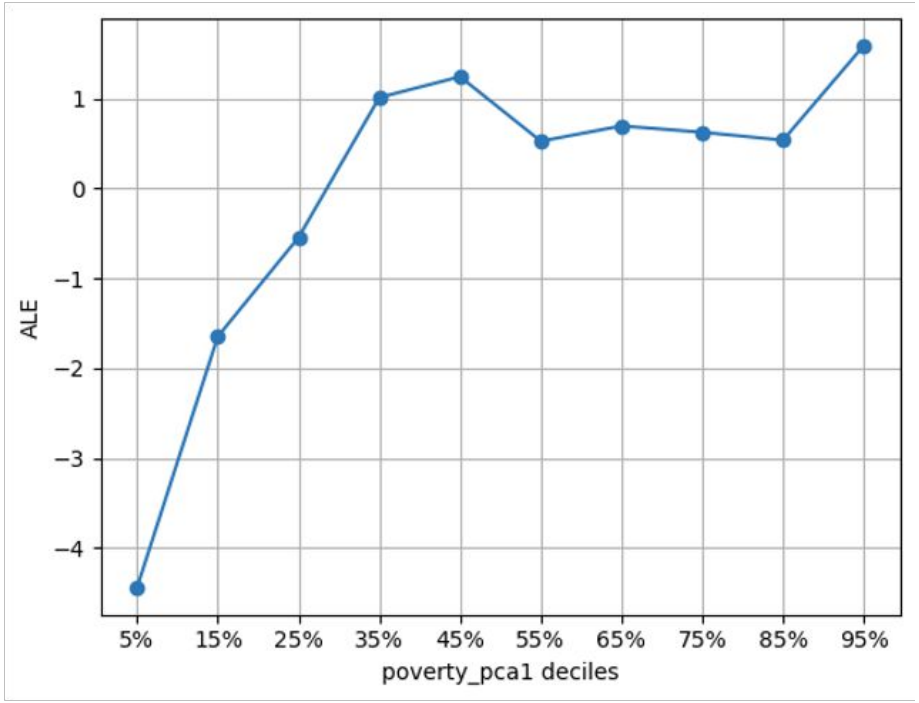




Increase in levels of poverty is correlated with an increase in Gini index and in wealth inequality.

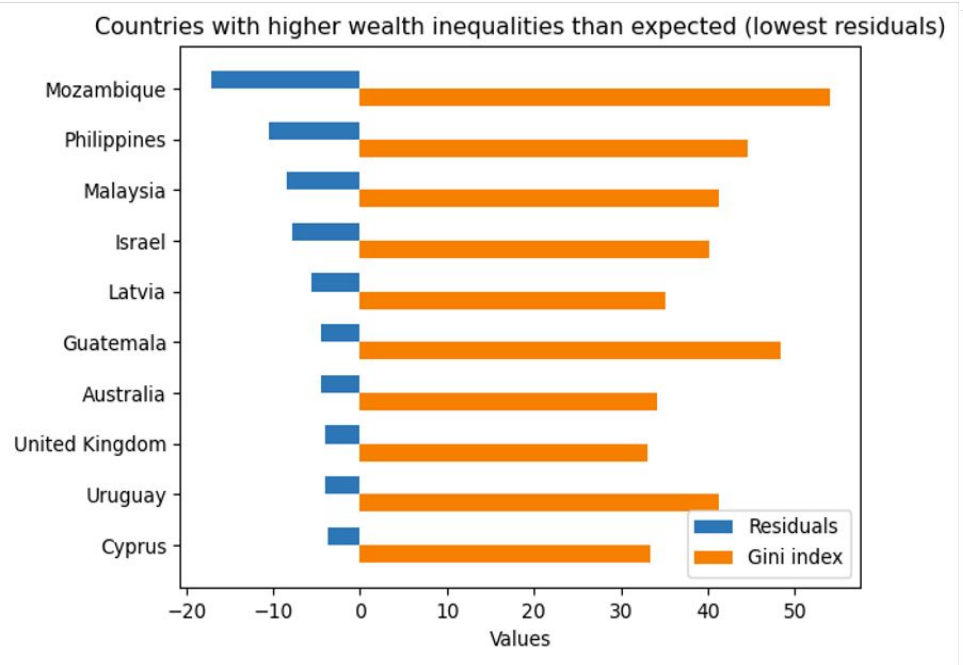
Severity of Poverty (Poverty PCA 1)

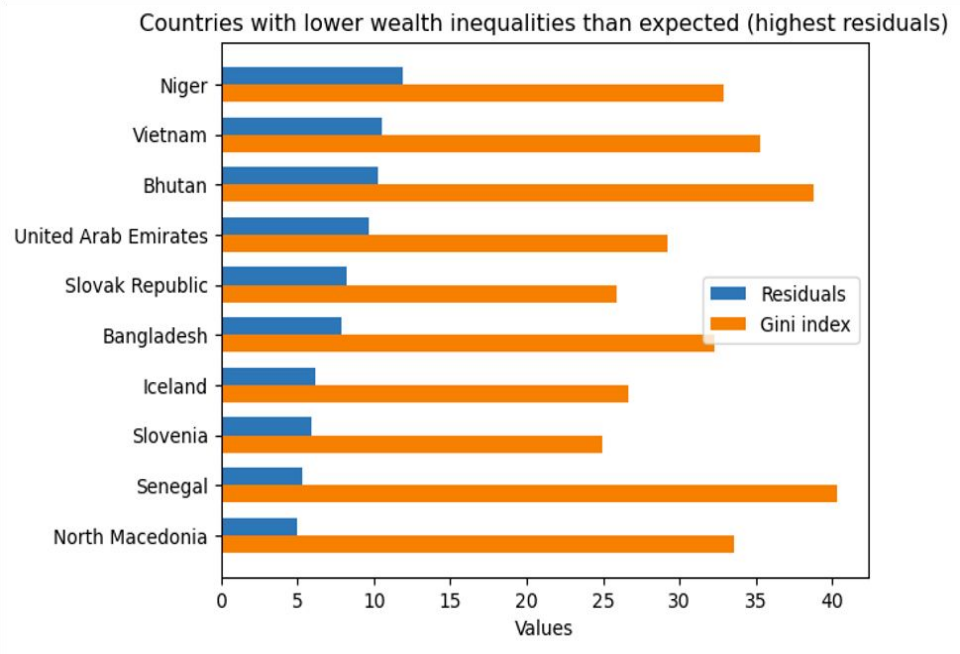




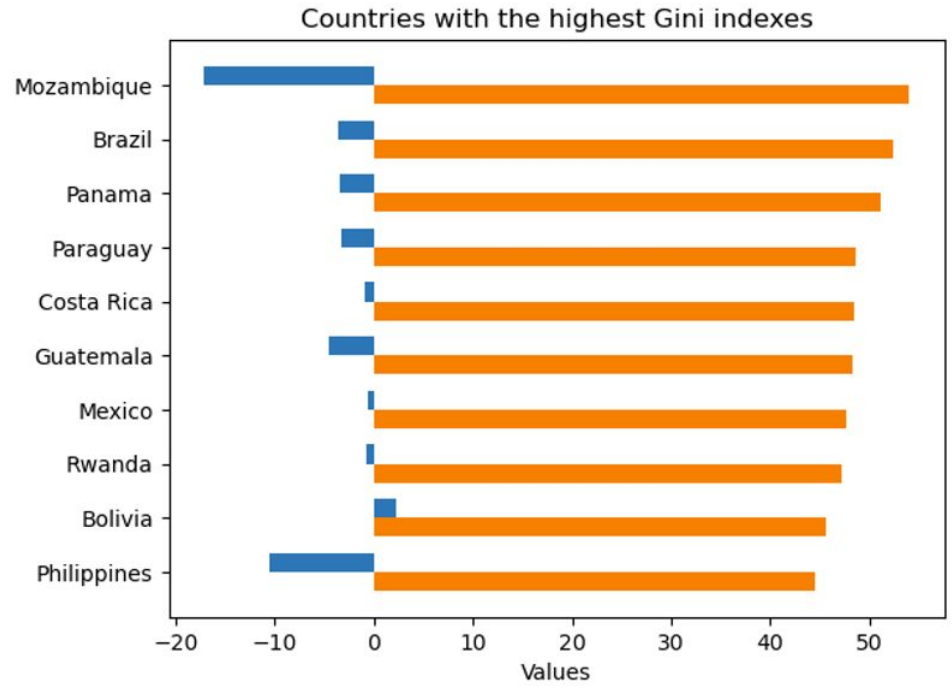
- Increase in the severity of poverty is correlated with an increase in Gini index and in wealth inequality up to the 35th percentile, after which the Gini index plateaus.

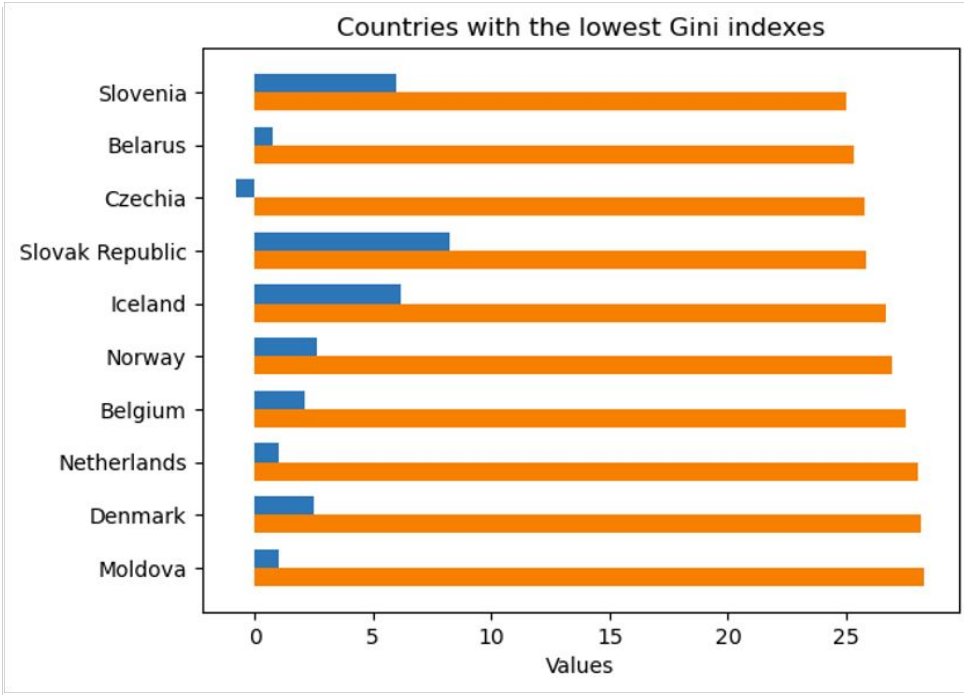
Countries where the model fails





Outlier Countries





What does it all mean?

- While my current analysis focuses on associations, it refrains from making strong causal claims.
- The strongest predictor in my analysis was educational attainment at an early stage which had a strong negative correlation with the Gini Index
- The relationships observed with population and GDP per capita are intriguing and opposite to each other.
- Two of the health indicators, ‘Cause of Death, by injury’ and ‘Hospital beds (per 1,000 people)’ emerged as significant but presented challenges in interpretation.