

Introduction

The question I am answering is how economics and social conditions impact a country's educational well-being. This analysis can be used by domestic and international governments (policy makers, grassroots, civilians, etc.) to identify the most pressing issues involved with a country's education. They can then tailor their decisions and policies to relieve any correlational educational distress.

Definitions for this project:

- * Economic well-being will be measured in PISA score (feature: rating)
- * PISA - Programme for International Student Assessment, is a triennial international assessment that measures 15-year-old students' literacy skills in reading, mathematics, and science. It is organized by the Organisation for Economic Co-operation and Development (OECD) to evaluate how well education systems prepare students for real-world challenges. The assessment focuses on how well students can apply their knowledge, rather than just what they have memorized.
- * PISA score range: 0 to 1000 points (highest possible score)
- * PISA score mean: 500 points
- * PISA score standard deviation: 100 points

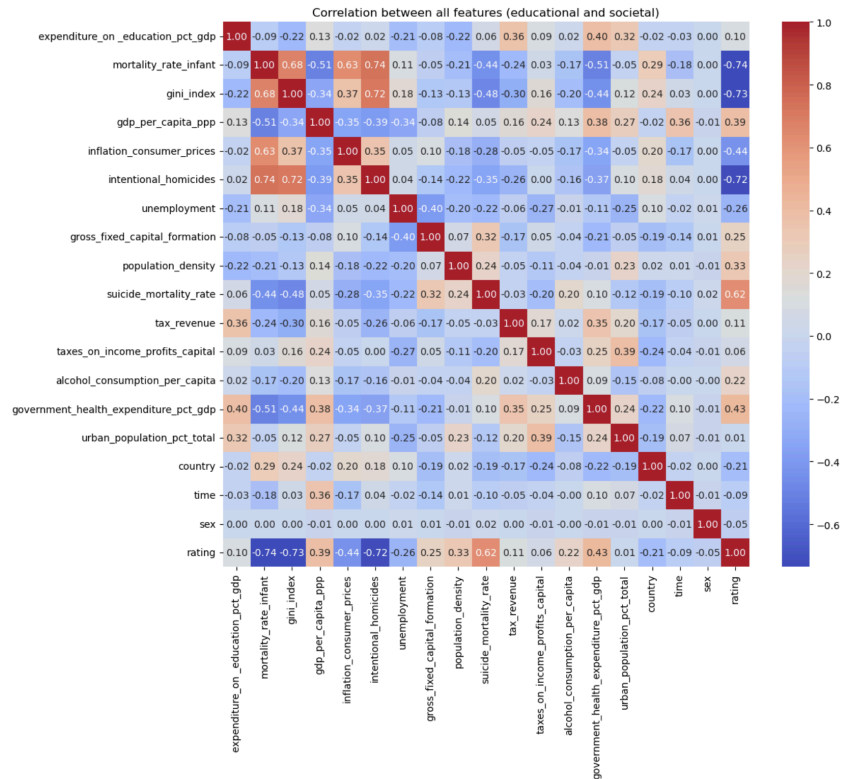
Dataset

This is an Economics and Education dataset that was obtained through a deleted post by Walasse Tomaz on Kaggle. I mapped out the correlations between the data features seaborn and the built-in pandas dataframe '.corr' function. Additionally, I used its data to train my linear regression model so that it can make its own "correlations" on the data, using its coefficients, which can also be graphically interpreted. Finally, I made some graphical interpretations of the data using standard matplotlib graphs.

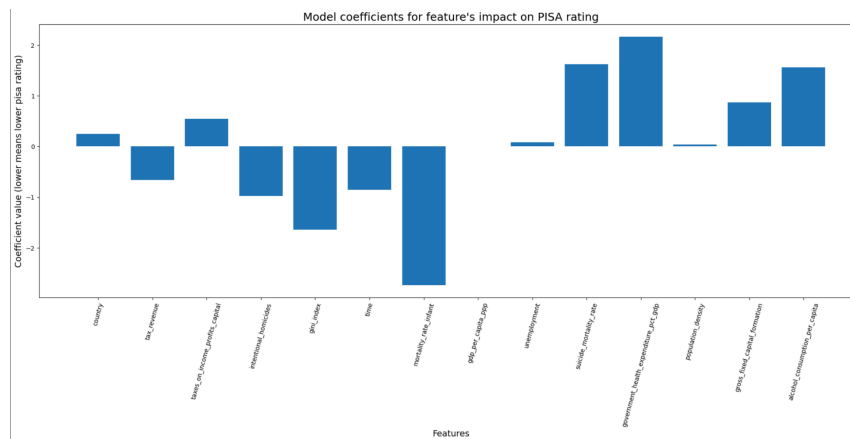
Results

As expected, there are **negative correlations** between infant mortality rate, Gini index, and intentional homicides with PISA scores, suggesting countries with higher PISA scores (our measure of educational well-being) tend to have fewer occurrences of these negative features. Also, as expected, there are **positive correlations** between PISA scores and government health and education expenditure as a percentage of GDP, suggesting that the more a government invests in health and education, the higher their scores will grow.

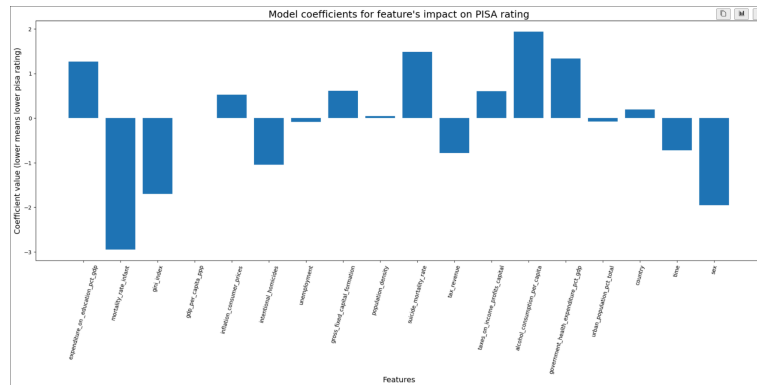
What was surprising was the **positive correlation** between PISA scores and suicide mortality rates. This can be the result of added educational pressure, higher population density, or lack of support within countries during the 2003 and 2018 periods. It goes to show that solely growing a test score metric doesn't account for all the issues within a society. Also surprising was the **negative correlation** between PISA scores and time, suggesting that as time went on, scores actually went lower. This can be the result of changes to the test or worsening educational conditions during the 2003 and 2018 periods, though we really don't know.



I have trained models both on features that were hand-crafted and kitchen sink. Neither model didn't deviated from the correlation that much, with the hand-crafted model slightly more accurate. This can be the result of overfitting or simply that the predictive information is already given through standard correlation trends. (Note: the high positive correlation to alcohol consumption can be ignored because most of its data was missing. Further mentioned in the limitations section.

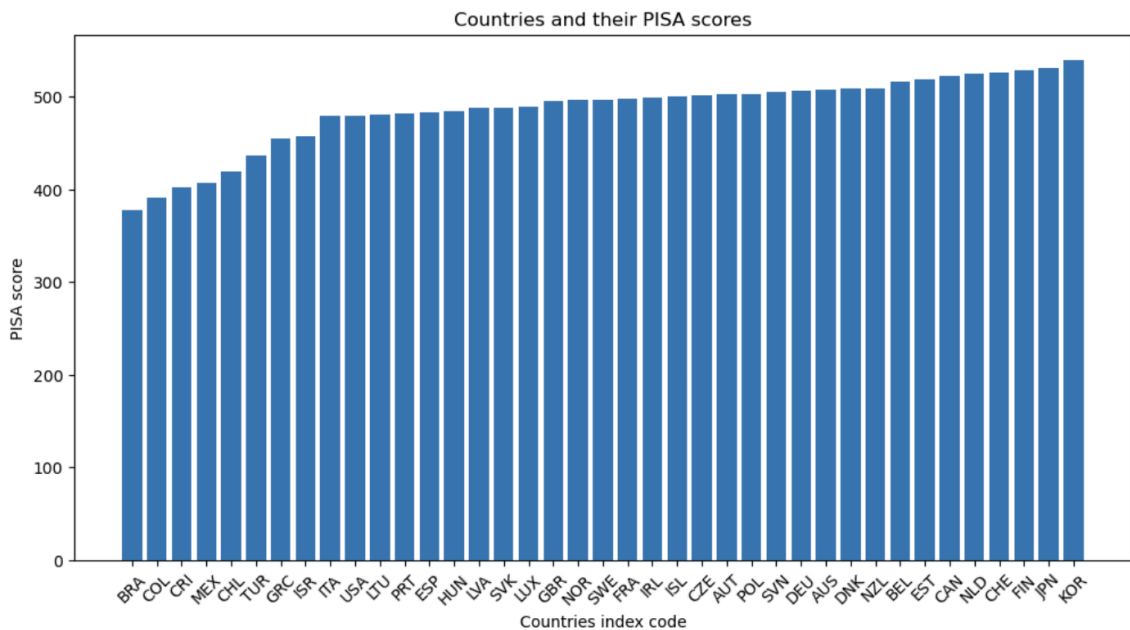


Hand-crafted model



Kitchen sink model

Unsurprisingly, more developed countries have a higher PISA score and vice versa. This model can provide insight into the countries that can have better educational policies: Brazil, Colombia, Costa Rica, Mexico, Chile, Turkey, Greece, and Israel, as their scores visibly appear lower than the others.



Limitations

- * PISA scores are individualized by subject (math, science, and reading), whereas this dataset only includes the mean score per country across the 5 tests taken from 2003 to 2018.
- * This data set also only includes 39 of the 79 countries that have taken the PISA exam from 2003 to 2018
- * Missing values are replaced by the mean values within each column. This impacts data validity, especially with the measurement of alcohol consumption per capita (Feature: alcohol_consumption_per_capita), which has most of its values missing.
- * There are other correlations not accounted for in the data/models. High government spending on health can be related to having more wealth, which correlates with higher education.

* There are simply so many features and government-related information that aren't present in the data, which can explain the data's results.

Conclusion

In trying to decipher which economic and social conditions impact educational well-being within a country, I further supported the importance of a quality education within countries, as countries with higher PISA scores tend to have lower rates of infant mortality, Gini index, and intentional homicides. I've also found that increasing government expenditure on health and education also boosts this score, and thus educational wellbeing. That said, many unaccounted confounding variables should be considered.

Unfortunately, my models failed to predict outside of standard correlations, and we can only get predictions on the PISA scores.