MegaBank

A Macro-Economic Study!
Predicting future income.

Kabuqueci Souza Alves 30.10.2023

Data

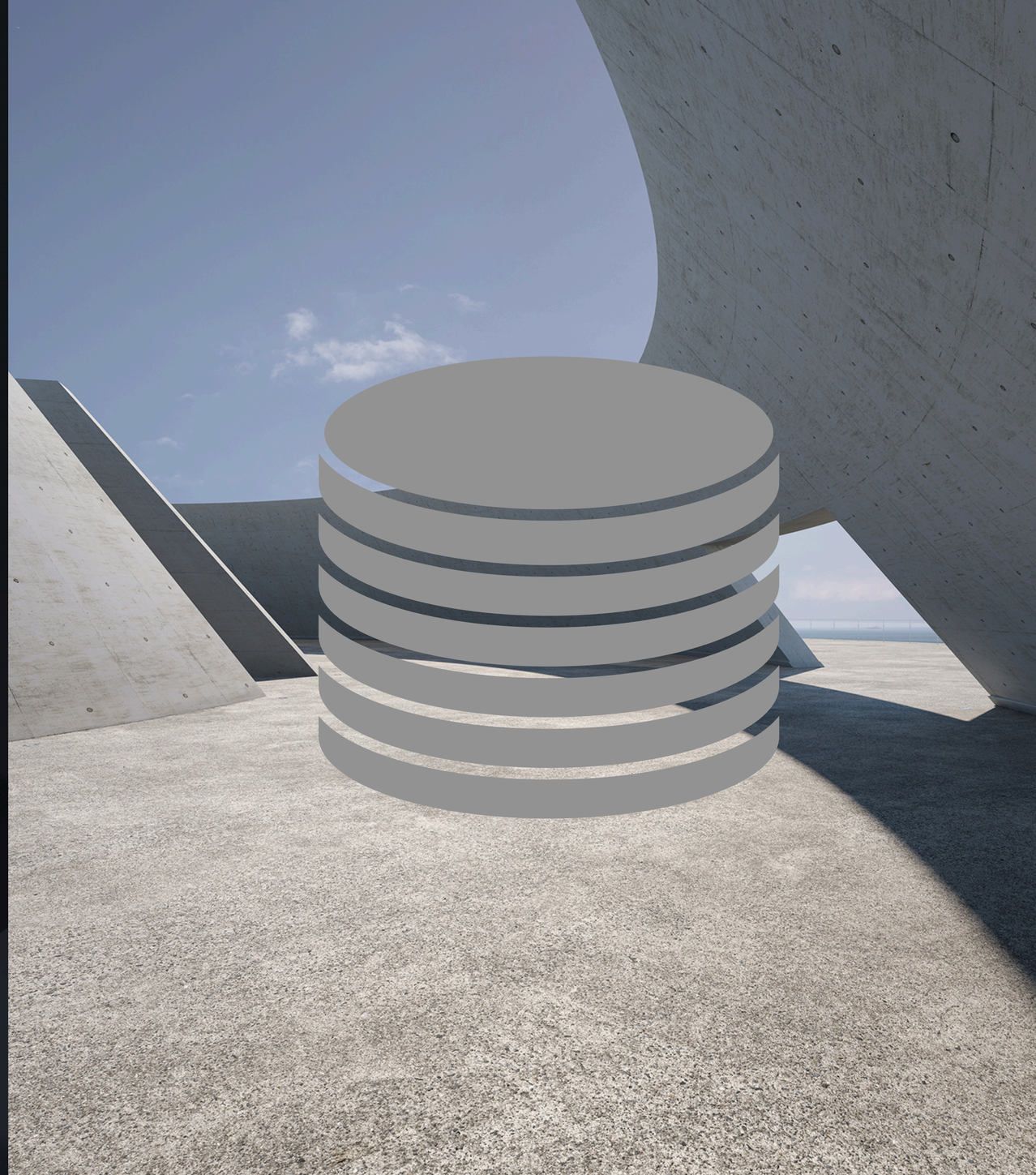
Were do our data come from?

Where do our data come from?

From the World Bank data platform!

There, I collect the needed data and calculate new variables!

You can find more in the notebook.



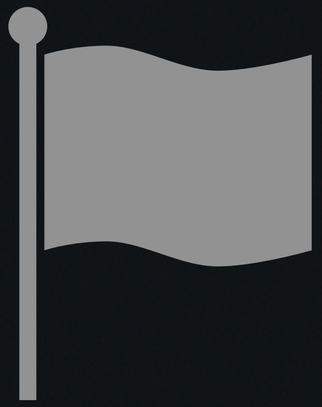
Task: 1

Preliminary analysis.

Next, some main statistics will be reviewed here!

Data Insight!

58



118

97%

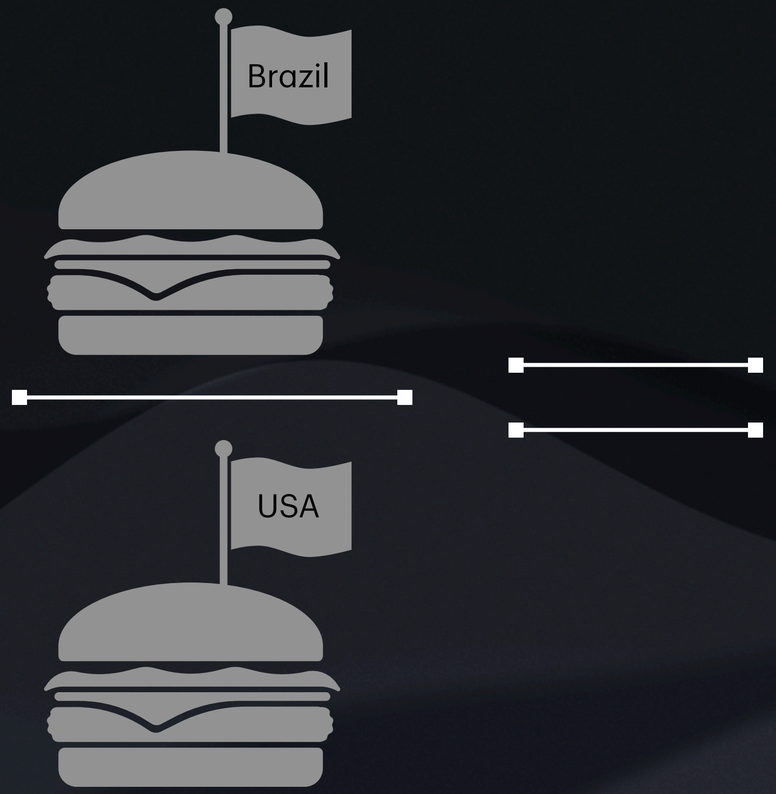
20

The $PPP

unity and an explanation of why (not why not) is an excellent method to compare countries.

Why is the PPP useful?

For comparing the standard of living in different countries, it represents the total number of goods and services that can be purchased with a given currency.

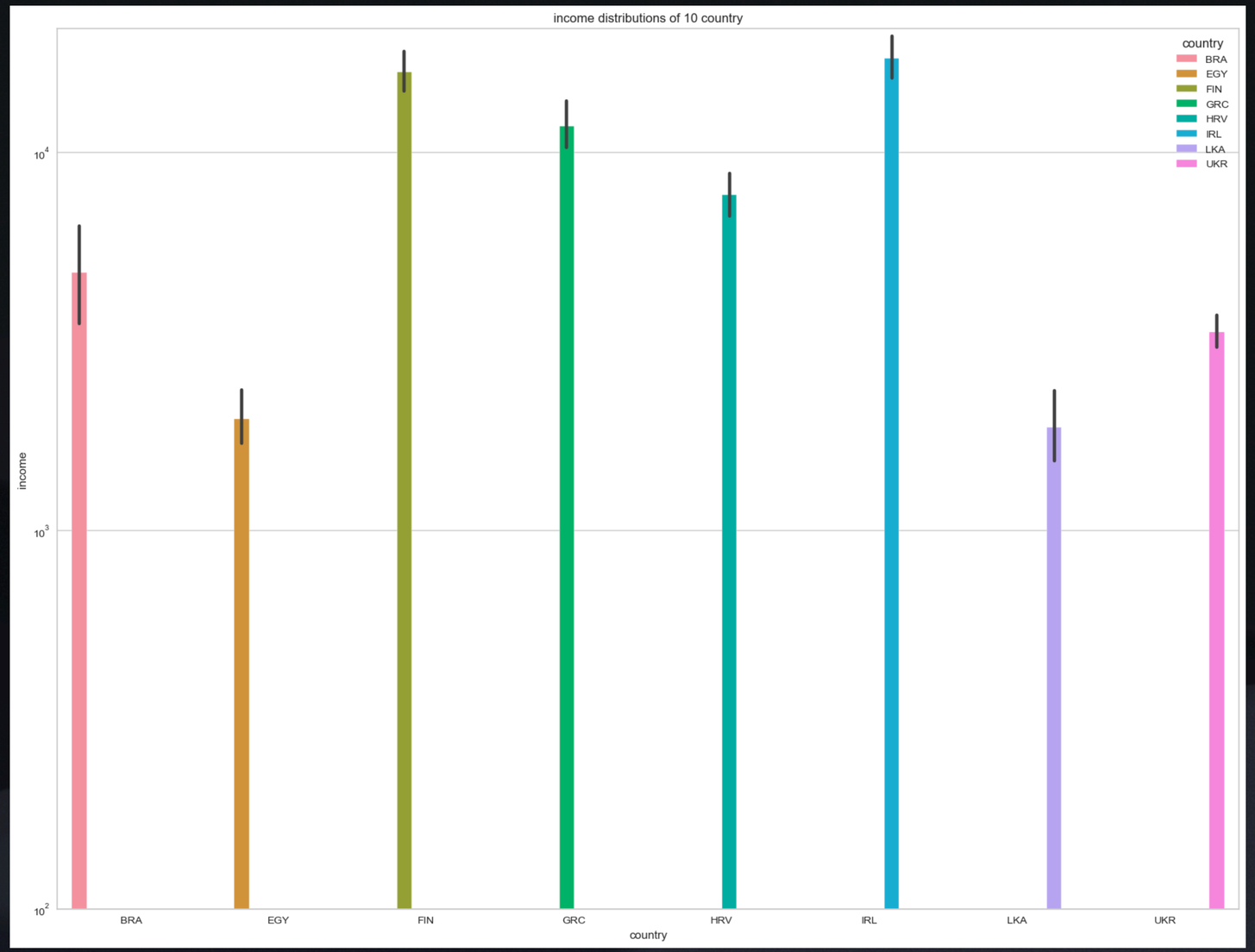


$US 4.27

Country income distributions.

From the ten select countries, I can see:

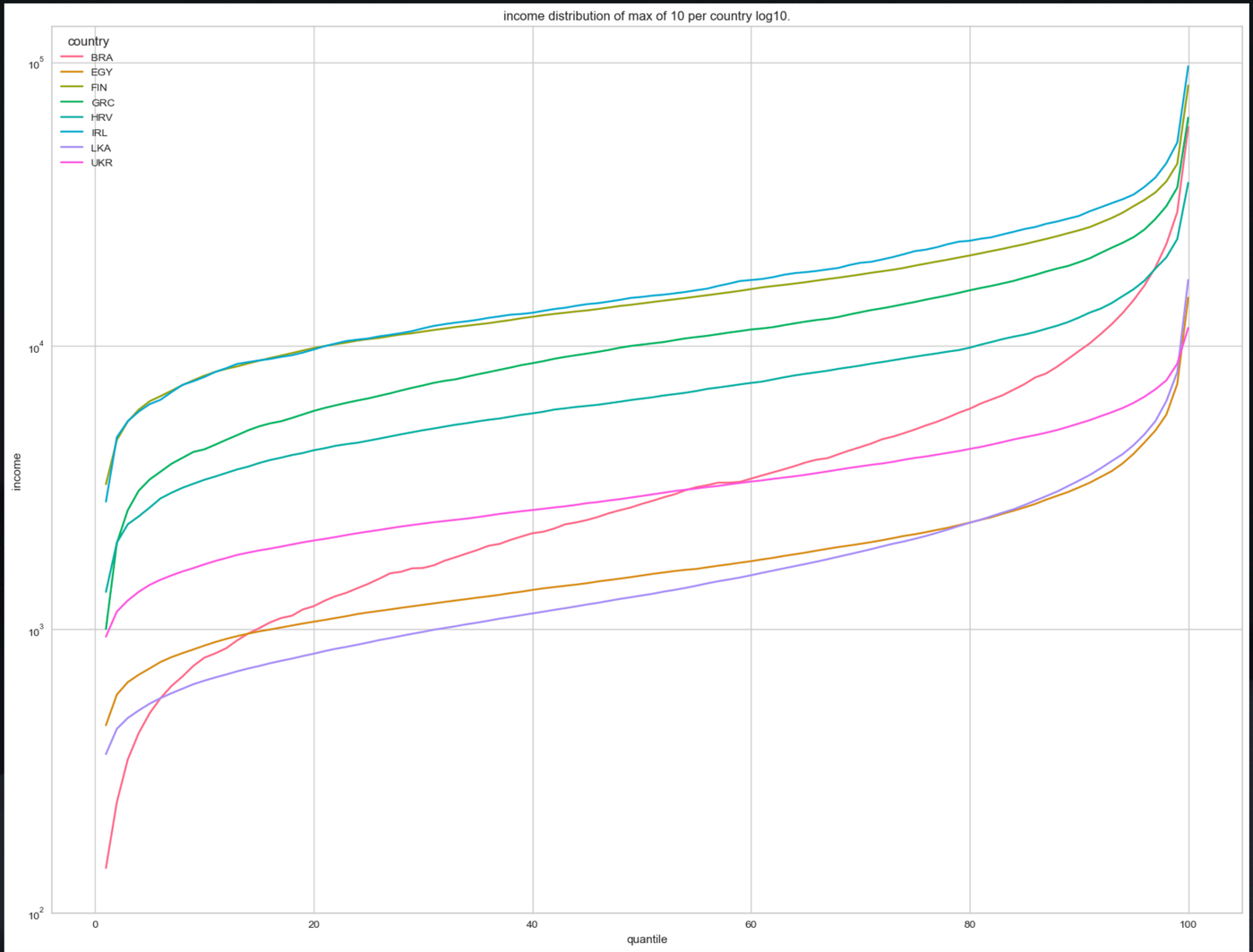
How some of the higher and lower income ones compare to my country, Brazil, where Ireland is at the top and Sri Lanka the lowest.



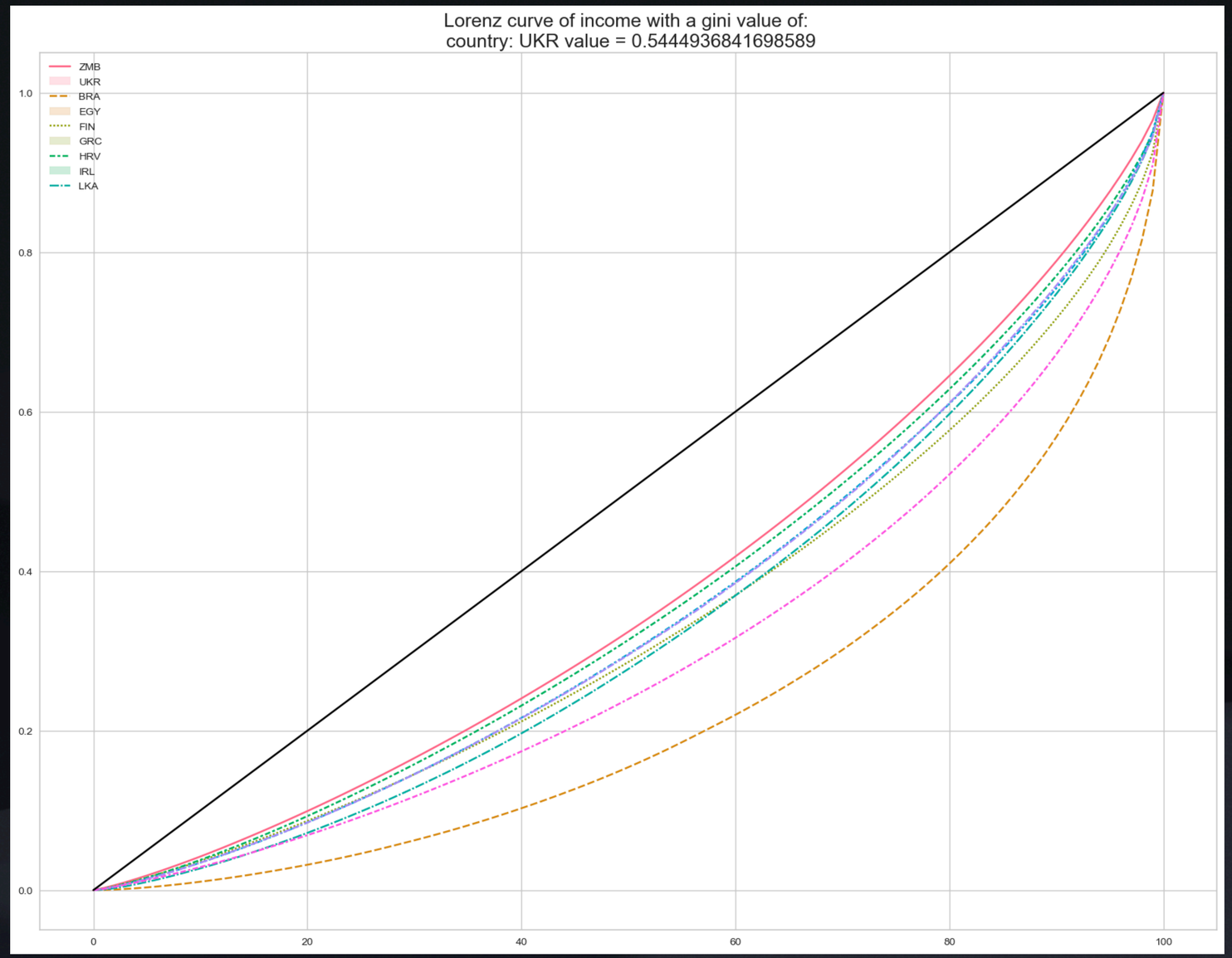
A graph showing how diverse countries are in terms of income distribution.

The plot here shows how the income changes by every exponent.

The higher the income class quantile, the higher the income.



The Lorenz curve for these countries.



Far from Equality!

The Countries close to the line have less unequal.

The greater the area under the line, the more the Inequality!

Task: 2

Calculating new variables!

Thanks to the research from B.Milanovic:

Global Inequality of Opportunity How much of our income is determined by where we live?

Research paper calcualtions.

This research indicates that for predicting income, we must have some country variables:

The mean income, Gini index, and the individual classes for parents and children.It describes an equation that helps to calculate the Elasticity Coefficient.

$$ln(Ychild) = \alpha + \rho j \; ln(Yparent) + \varepsilon$$

Task: 3

Statistical Models and Machine Learning

For Income Prediction.

Anova Analysis

An ANOVA with the individual's country as the only explanatory variable.

Country on income

Variables	F	PR(>F)
C(country)	49217.464227	0.0
Residual		

Effects

Variables	SSeffect
C(country)	2.92%
mj	3.80%
gini	0.00%
c_i_parent	4.79%
gdpppp	0.00%
Residual	88.48%

As a result, the mean income, parent income, and country have the most influence on the dependent variable income. But this is only part of the story.

Linear Regression

Including as explanatory variables only the mean income and the Gini index of the individual's.

Model log 2

Variables	Score
Gini	77.72...
Mean Income	
Parent Income	

Model 1.

Variables	d Coef_ f	Score
Gini	-3.92...	49.37...
Mean Income	1.00...	

Model log 1

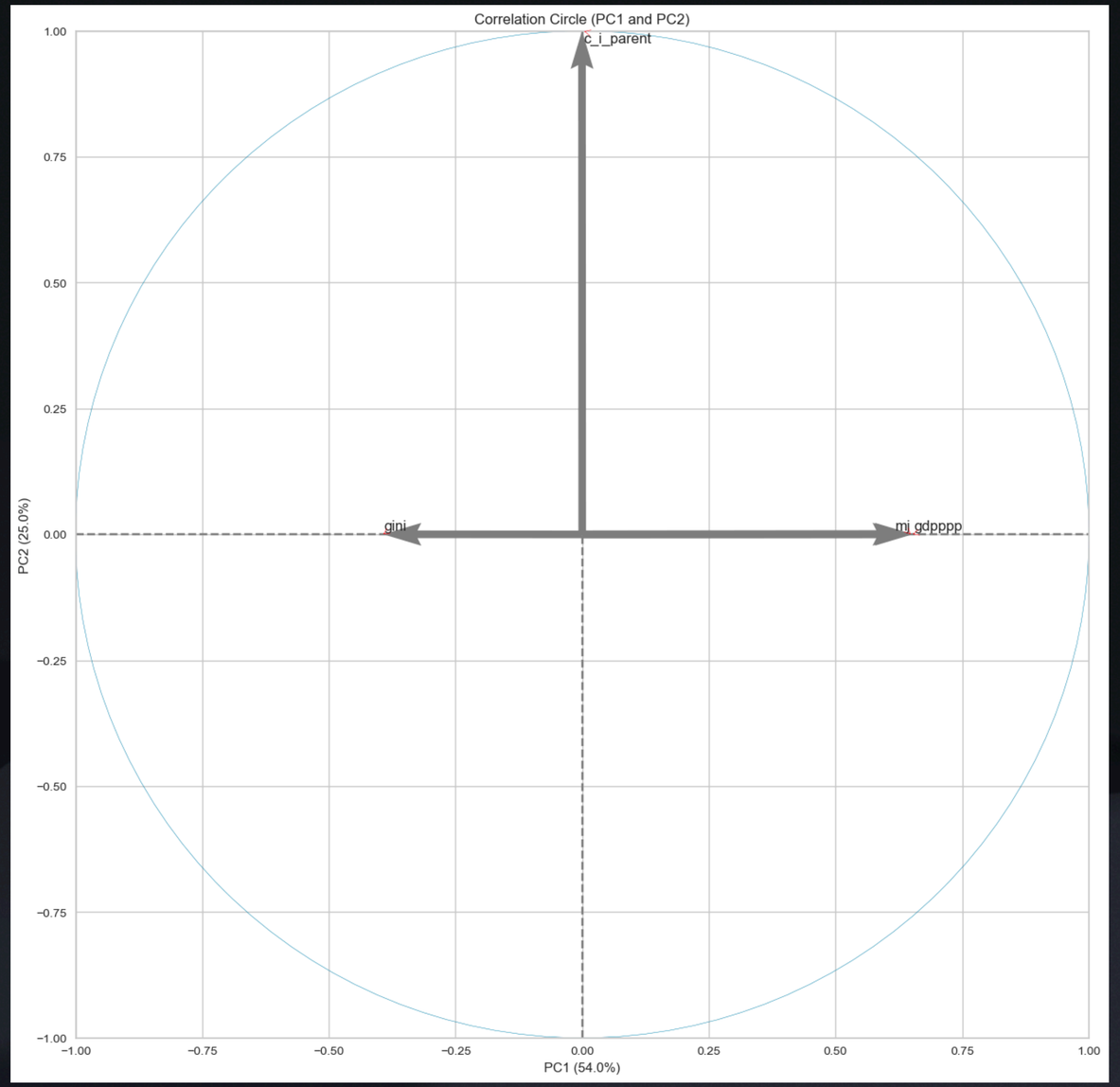
Variables	d Coef_ f	Score
Gini	-0.017...	72.67...
Mean Income	0.98...	

PCA

Variance Ratio.

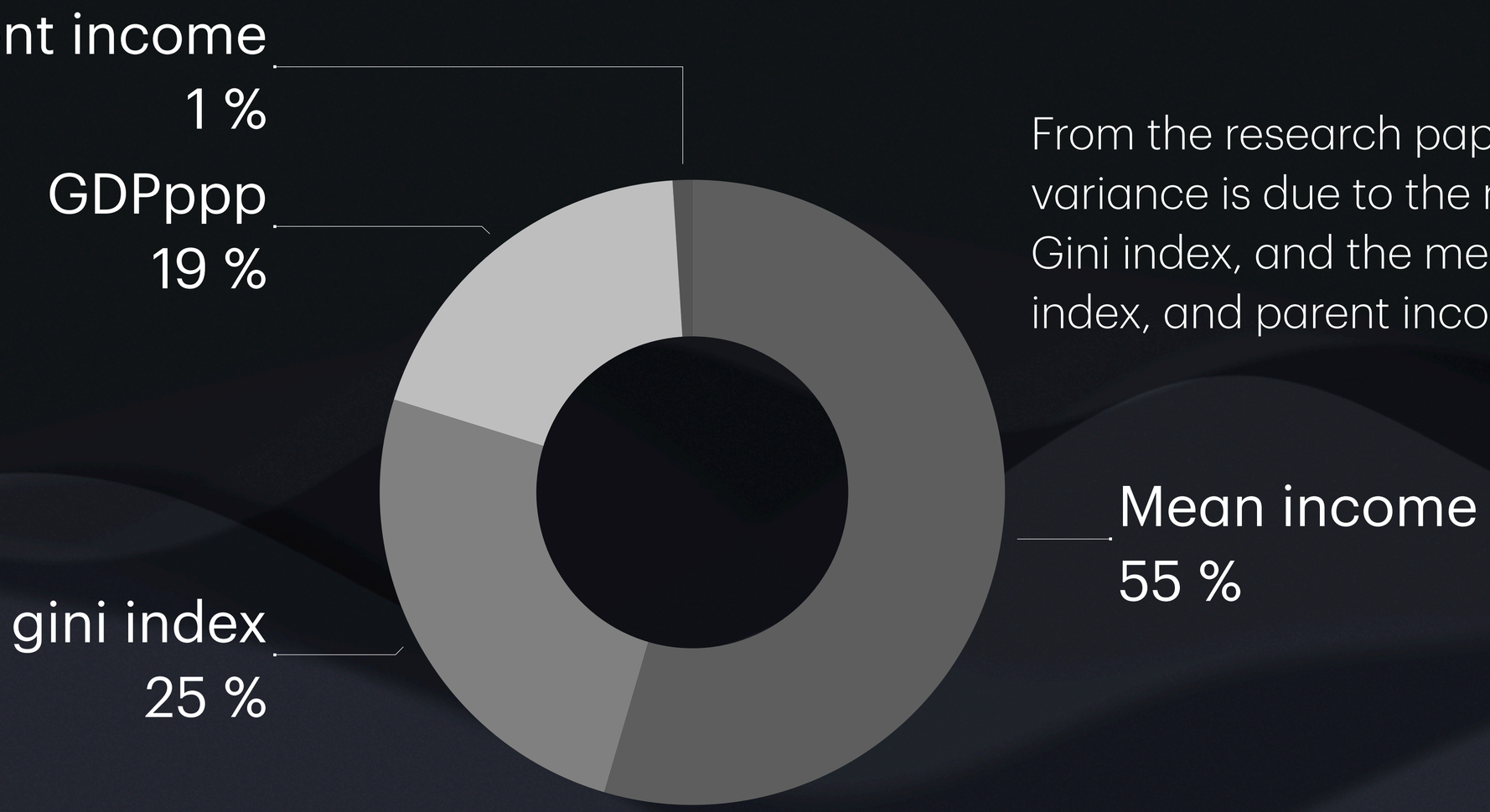
The parent income is enough to correlate with all the influential variables;

this could explain why we had a lower effect in our ANOVA complex table, as this effect is distributed from many variables and is not concentrated on a single one.



How much are the overall influencies

Variables influences



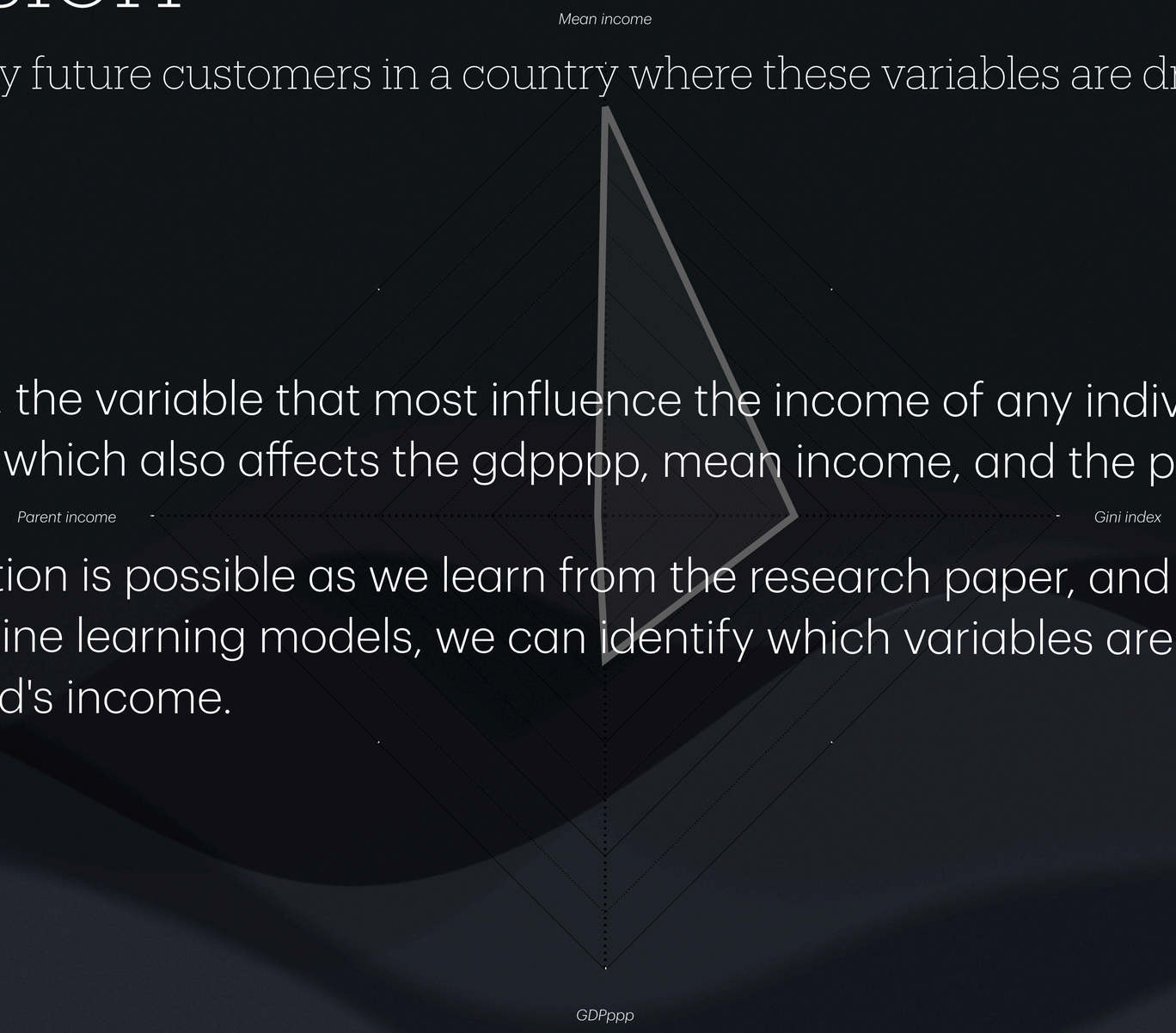
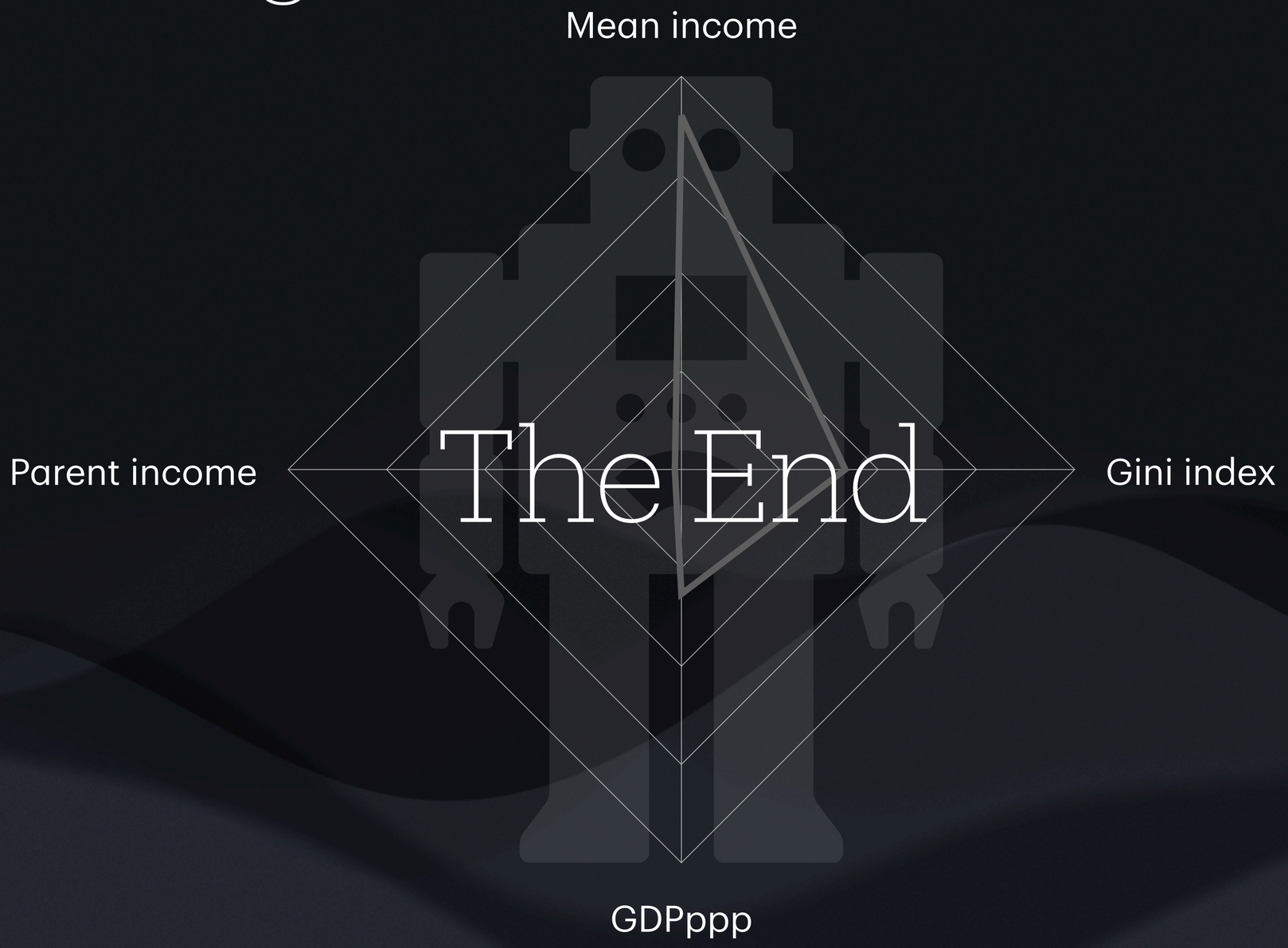
From the research paper, 60% of the variance is due to the mean income and Gini index, and the mean income, Gini index, and parent income explain 80%.

Conclusion

We can then identify future customers in a country where these variables are drive economic forces.

From this analysis, the variable that most influence the income of any individual is the countries at first, then the gini, which also affects the gdpppp, mean income, and the parent income.

An income prediction is possible as we learn from the research paper, and by applying statistical models and machine learning models, we can identify which variables are most influential in determining a child's income.



Model-2-log