



# **WORLD DEVELOPMENT INDICATORS (WDI) EXPLORATORY STUDY**

## Contents

Executive Summary .....	3
Introduction .....	4
Methodology.....	6
Access .....	6
<i>Investigate and Prepare</i> .....	6
<i>Body and Analysis</i> .....	9
Descriptive analytics .....	9
Variables Overview .....	9
Explanatory analysis .....	11
Multilinear Regression .....	17
Correlation matrix.....	17
Descriptive statistics.....	19
Overall model - All countries .....	20
Logistic regression.....	27
Optimization of the Model .....	28
Assessment of the model.....	29
Model interpretation .....	32
Forecasting .....	33
GDP per Capita Trend analysis.....	33
Forecasting Model.....	34
Conclusion and Recommendation .....	39
References .....	41
Appendix 1 .....	43
<i>Investigate and prepare</i> .....	43
Descriptive analysis .....	44

## **Executive Summary**

The "World Development Indicators" dataset was examined and analyzed by the researchers using Microsoft Excel and SAS Visual Analytics to generate insights on factors contributing to economic performance and elements in monetary policy indicated by GDPperCapita and Developed as Global North/South is found to be too simplistic and inconclusive to divide the economy geographically. Three key models: Multilinear Regression, Logistic Regression, and Forecasting were created with assisting analyses like Descriptive statistics and Correlation matrixes in SAS Viya. The key insights derived from both multilinear regression and forecasting models are that health expenditure is key driver of GDPperCapita growth in both developed and developing economies. Therefore, to increase GDPperCapita by 10% in three years, monetary policy to increase budget allocation on healthcare should be increased. Besides, Logistic regression explains that GDPperCapita and Value-added manufacturing are significantly to economic development, represented by Developed variable. The conclusion summarizes the 10 insights and recommended to document additional data points like unemployment rate, Foreign Direct Investment and more.

## Introduction

This report aims to study the economic performance of the 20 selected countries over the span of 50 years (Figure 1). Several business analytics applications are deployed in SAS Viya, including descriptive analytics, explanatory analysis, multilinear regression, logistic regression, and forecasting to understand (a) the contributing factors of economic development and (b) the role of monetary policy in improving the nations' performance; to collectively present key interesting insights.

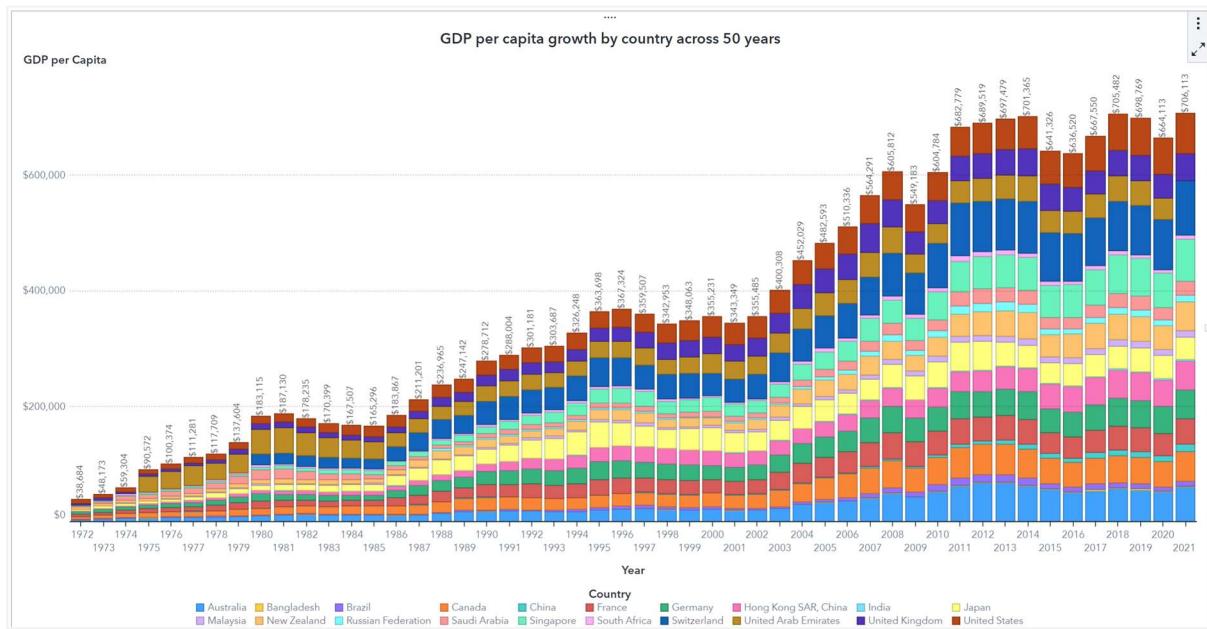


Figure 1: GDP per capita growth by countries in 50 years – Observation

This analysis introduced a new categorical variable named "Global North/South", where the world is divided geographically into Global North and South by The Brandt line, as shown in Diagram 1 below (Royal Geographical Society 2022). The Brandt line has been in place for more than 40 years to section the global community into a wealthy Global North and poor Global South in the tropical regions (Solarz 2019). However, its relevancy has been questionable following the rise of UAE, Saudi Arabia and the Asian Tigers with growing middle- and upper-class consumers in the last two decades (Guttal 2016). The validation of this category is investigated in the report.



Diagram 1: The Brandt line that divides Global North and South

## Methodology

### Access

The given partial set of World Development Indicators (WDI) data was first accessed in Microsoft Excel to investigate and prepare the data before being uploaded to SAS Viya.

### Investigate and Prepare

#### Excel

Several data investigation and preparation steps were taken to improve data readability and interpretability:

1. Replaced the columns marked (...) to missing data
2. Added the Global North/South column; This column is populated according to the references given ([here](#)):

Country	CountryCode	Global North/South
Australia	AUS	Global North
Canada	CAN	Global North
China	CHN	Global South
France	FRA	Global North
Germany	DEU	Global North
Hong Kong SAR, China	HKG	Global South
India	IND	Global South
Japan	JPN	Global North
Malaysia	MYS	Global South
New Zealand	NZL	Global North
Russian Federation	RUS	Global North
Saudi Arabia	SAU	Global South
Singapore	SGP	Global North
South Africa	ZAF	Global South
Switzerland	CHE	Global North
United States	USA	Global North
United Kingdom	GBR	Global North
Bangladesh	BGD	Global South
Brazil	BRA	Global South
United Arab Emirates	ARE	Global South

3. Identified the columns with missing data points across 50 years (sorted by highest missing observation in descending order):

<b>Data points</b>	<b>Missing observation</b>
HlthExCapita_33	620
PrimaryEnrol_3	599
DeplntRate_35	453
LenIntRate_91	395
TaxGDP%_135	382
Labor_89	360
MfgValGDP_96	247
ServicesAdd_130	217
kWhCapita_45	158
NatIncome_5	142
IncomeCapita_7	142
Inflation_83	107
PricelIndex_24	101
AirPassengers_15	73
AirFreight_14	71
GDP_65	28
GDPCapita_67	28
AgriLand_9	20
LandArea_90	20
Country	0
CountryCode	0
Year	0
Population_110	0
Developed	0
Global North/South	0

## SAS Viya

The cleansed data in Excel were then uploaded to and accessed by SAS Viya using the "Develop SAS Code" feature.

Data points were transformed from categories to measures and vice versa.

Duplicated data items with "Average aggregation type" and new calculated data items are created to enable further analysis.

New calculated items:

New calculated items	Explanation	Formula	Type
Year	Derived from "Years" to be used in forecasting	Calculated by Date operator	Date
Labour percentage of the population	Derived from "Labour_89" and "Population_110" to explain the percentage of the working population in each country	(Labour_89/Population_110) * 100	Measure
AgriLand (% of total LandArea)	Derived from "AgriLand_9" and "LandArea_90" to explain the percentage of the agricultural land relative to the total land area in each country	(AgriLand_9/LandArea_90) * 100	Measure

Please refer to the appendix for further information.

## Body and Analysis

### Descriptive analytics

#### Variables Overview

It was pointed out that numerous variables were missing data in the data preparation stage. It is worth investigating to ensure these variables are representative in their respective segments to predict GDPperCapita growth, Economic development (Developed) and forecasting the future trend.

Figure 2 shows that Electric power consumption (kWhCapita\_45) data were not recorded since 2015. However, Moss et al. (2020) suggest that electricity consumption strongly correlates with GDP growth, albeit highly dependent on the context. Similarly, it is noted that Current health expenditure per capita (HlthExCapita\_33) data was not captured until 2000.

Besides, the figure below shows that the LandArea variable seems to have remained constant over the years (Refer appendix 1).

Interestingly, inflation of consumer prices (Inflation\_83) shows that the fluctuations between 1989 and 1994 were exceptionally high.



Figure 2: Data availability across years – to examine variance in data documentation

### *Inflation Fluctuations*

Figure 3 shows the inflation box plot created to investigate the fluctuations. It shows that Brazil (a developing country = 0) and Russian Federation (a developed country = 1) stood out for having high upper whisker values, particularly Brazil. Brazil's inflation in consumer prices breakthrough 4-digit in 1989 and remained at the higher end for five years. It explains that the country faced hyperinflation from January 1989 to June 1994 caused by skyrocketed budget deficits and excessive money growth (Araujo & Junior 2018). Likewise, Russian Federation's inflation rate peaked in 1993 at 874% when it was first tracked, signifying an extremely low buying power for a start, aggravated the fluctuations of the inflation rate for the five years discussed above.

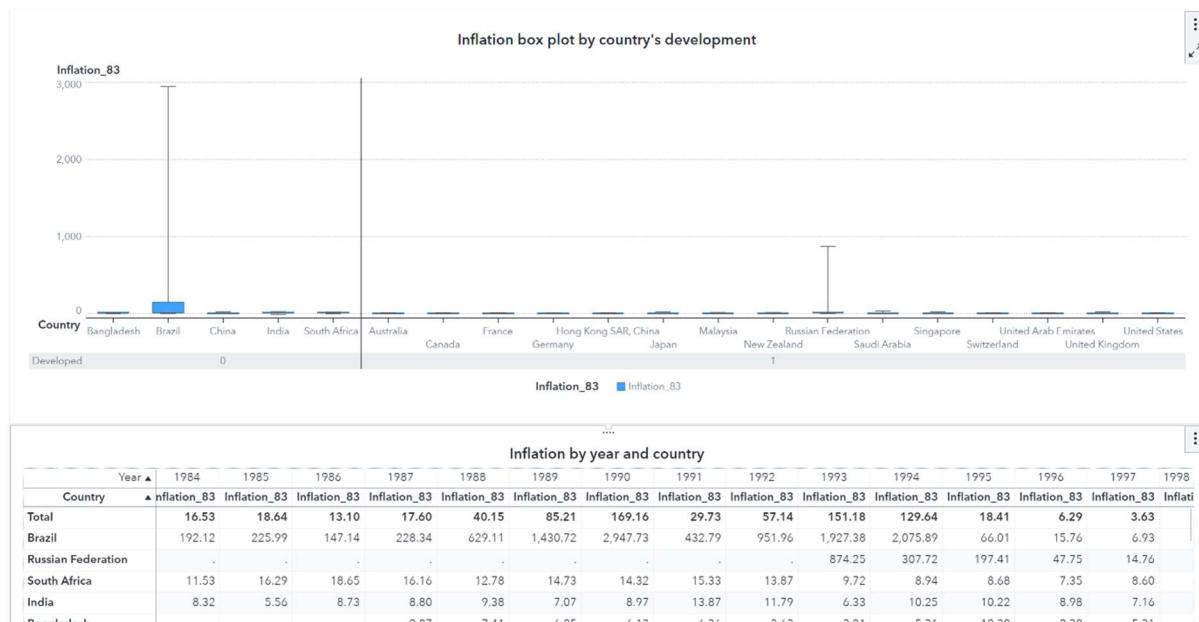


Figure 3: Inflation box plot

## *Explanatory analysis*

### *GDP measurement*

The Developed variable in this analysis is categorized according to the World Population Review's (2022) definition, where a developed country has a relatively mature economy compared to a developing country using the Human Development Index (HDI). In this analysis, the variable Developed is categorized according to HDI by their human development, where an HDI score above 0.80 is considered a developed economy (World Population Review 2022).

Figure 4 compared GDP per Capita against Developed and Global North/South variables and confirmed that the Brandt line (indicated by the Global North/South) is too simplistic and inconclusive to section the world into North and South. It is supported by the findings in Figure 4 where the last five years (from 2017 to 2021) “GDPperCapita” sorted in descending order, Hong Kong SAR, China (GDPperCapita: \$47,763) and United Arab Emirates (GDPperCapita: \$40,868) in Global South have higher “GDPperCapita” compared to some of the Global North countries such as Japan (GDPperCapita: \$39,645) and Russian Federation (GDPperCapita: \$11,360).

According to Elistia (2018), human development (HDI) affects economic growth, as shown in GDPperCapita. The split of economies by HDI also aligns with “GDPperCapita”, where the split occurs at \$10,700, where nations with greater values are considered developed (Figures 4). The relationship between Developed and “GDPpercapita” is tested in Logistic Regression to analyze their relationships.

Filters: 2017; 2021

1972 2017 to 2021 2021

Comparing characteristics of Global North/South and Developed/Developing						
Developed	Global North/South	Country	GDP per Capita	Income per Capita	GDP_65	NatIncome_5
Total			\$34,768	\$26,691	3.2T	2.6T
1	Global North	Switzerland	\$87,127	\$63,686	747B	544B
		Singapore	\$65,473	\$46,100	368B	261B
		United States	\$64,026	\$53,466	21T	18T
		Australia	\$55,521	\$42,103	1.4T	1.1T
		Germany	\$47,295	\$39,118	3.9T	3.2T
		Canada	\$46,663	\$36,854	1.8T	1.4T
		New Zealand	\$43,888	\$34,941	219B	173B
		United Kingdom	\$43,202	\$35,154	2.9T	2.3T
		France	\$40,702	\$33,324	2.7T	2.2T
		Japan	\$39,645	\$31,125	5T	3.9T
0	Global South	Russian Federation	\$11,360	\$8,624	1.6T	1.2T
		Hong Kong SAR, China	\$47,763	-	356B	-
		United Arab Emirates	\$40,868	\$37,512	396B	363B
		Saudi Arabia	\$22,454	\$18,146	769B	617B
		Malaysia	\$10,971	\$7,631	351B	242B
		China	\$10,366	\$7,303	15T	10T
		Brazil	\$8,458	\$7,433	1.8T	1.6T
		South Africa	\$6,594	\$5,316	386B	309B
		Bangladesh	\$2,152	\$1,767	351B	287B
		India	\$2,052	\$1,748	2.8T	2.4T

Figure 4: Economies categorization by Developed and Global North/South - Descriptive statistics



Diagram 2: Country's development by GDPperCapita

### Agricultural land use ratio

Agricultural land is compared against the total land area to highlight the countries' land use ratio for agriculture over the years. Figure 5 indicates that Agricultural lands in developing countries (blue line in the line chart) accounted for an enormous proportion of the land area. It trended upward recently and reached 60% last year (2021). Developed countries, on the other hand, have less proportion of agricultural

land. While it may be due to their natural locations, this proportion has decreased further over the years. This trend can be interpreted as developed countries have shifted their focus away from agricultural products and tapped into other land developing plans (e.g., transportation infrastructure). On the other hand, developing countries have increased their investment in the agriculture industry by increasing their Agricultural land.

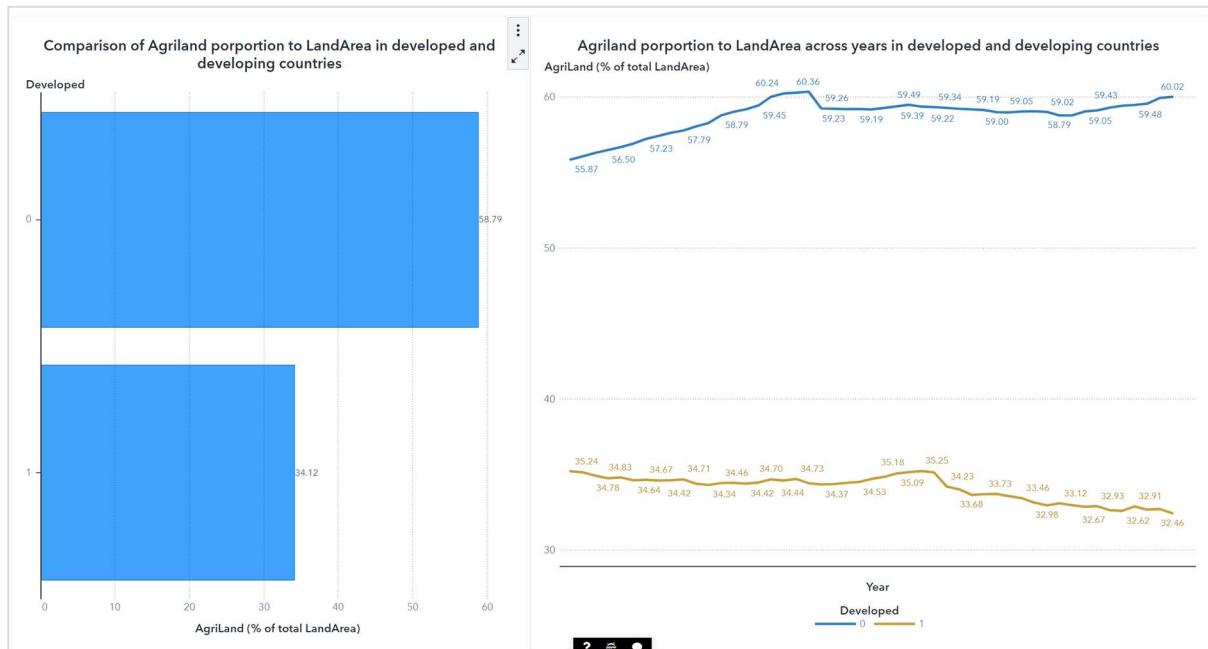
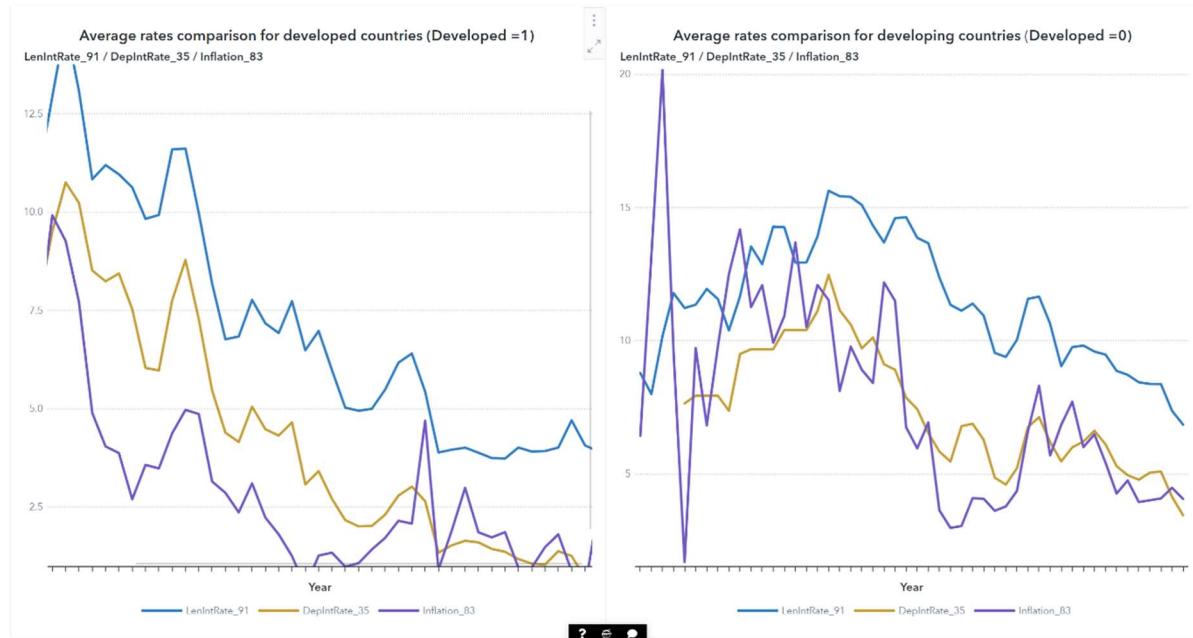


Figure 5: Agricultural land use in developed and developing countries throughout the years

## *Comparison of Deposit rate, lending interest rate and inflation rate in developed and developing countries*

In figure 6, the values for Brazil and Russia are excluded since their corresponding average rates differ dramatically from other countries and might disrupt a clear interpretation. The line chart on the left indicates that developed countries have pursued a robust monetary policy to ensure a stable and predictive growth environment for businesses. In contrast, the developing countries' macroeconomic policies show the opposite. Developed countries have adjusted their lending and deposit interest rates according to the inflation rate. A Higher deposit rate than the inflation rate has been maintained over the years in developed countries, resulting in people not losing their value of money because of inflation. Therefore people will not move their money out of the country for higher yields.



*Figure 6. Comparison of Financial rates in developed and developing countries*

### *Comparison of GDP sources in developed and developing countries*

The bubble chart indicates that, on average, the proportion of the GDP from manufacturing activities is higher in developing countries (18% of GDP) than in developed countries (15% of GDP), indicating the different pace of industrial revolutionization where developing economies are involved in labour-intensive, raw or processed physical production. In contrast, tax revenues and value-added services accounted for a high ratio of GDP accounted for developed countries. It can be interpreted that developed countries have more sophisticated tax systems focused on a fair distribution of wealth among citizens. Furthermore, technological advancements put them at a competitive advantage to excel service sector.

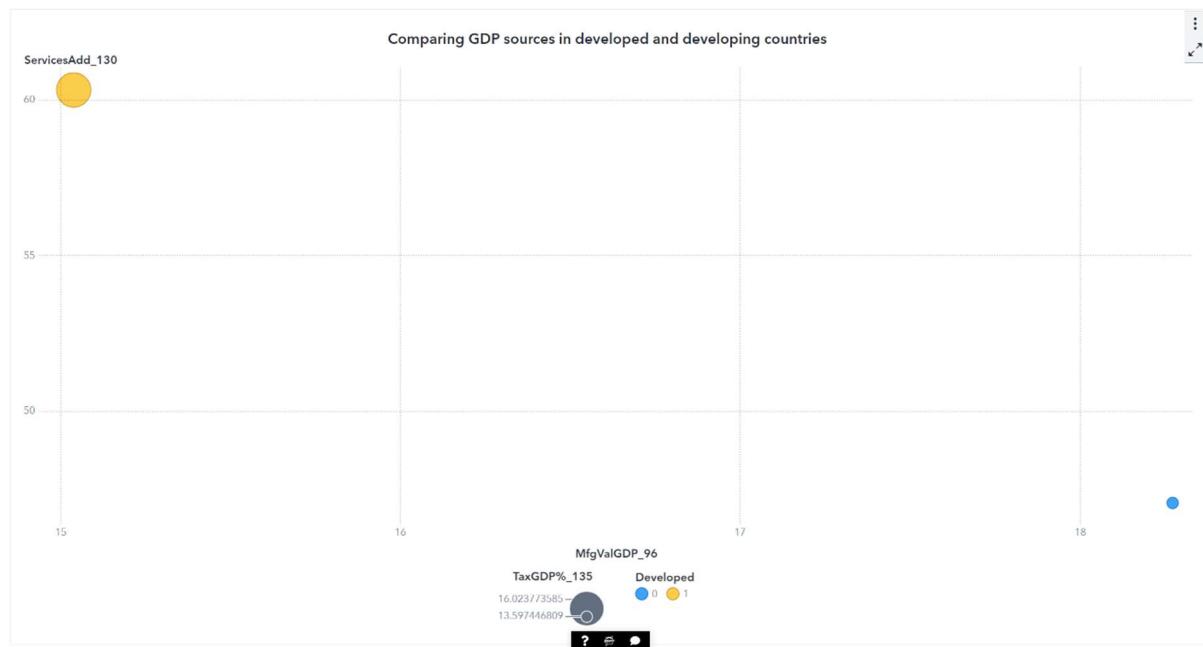


Figure 7. GDP sources comparison

### *Labours Contribution*

Figure 8 indicates that the average percentage of the working population is higher in developed countries (53% to 44%), suggesting stable employment over the years and most likely a less unemployment rate in developed countries. The average number of labours in developing countries is significantly higher than in developed countries, which indicates that labours in developing countries provide less value than their counterparts. In the past ten years, regardless of the time value of money, dividing the corresponding average GDP by the average number of labourers indicates that each labour in developing countries has contributed \$1200 to the GDP of their countries. On the other hand, each labour in developed countries has contributed \$8500.

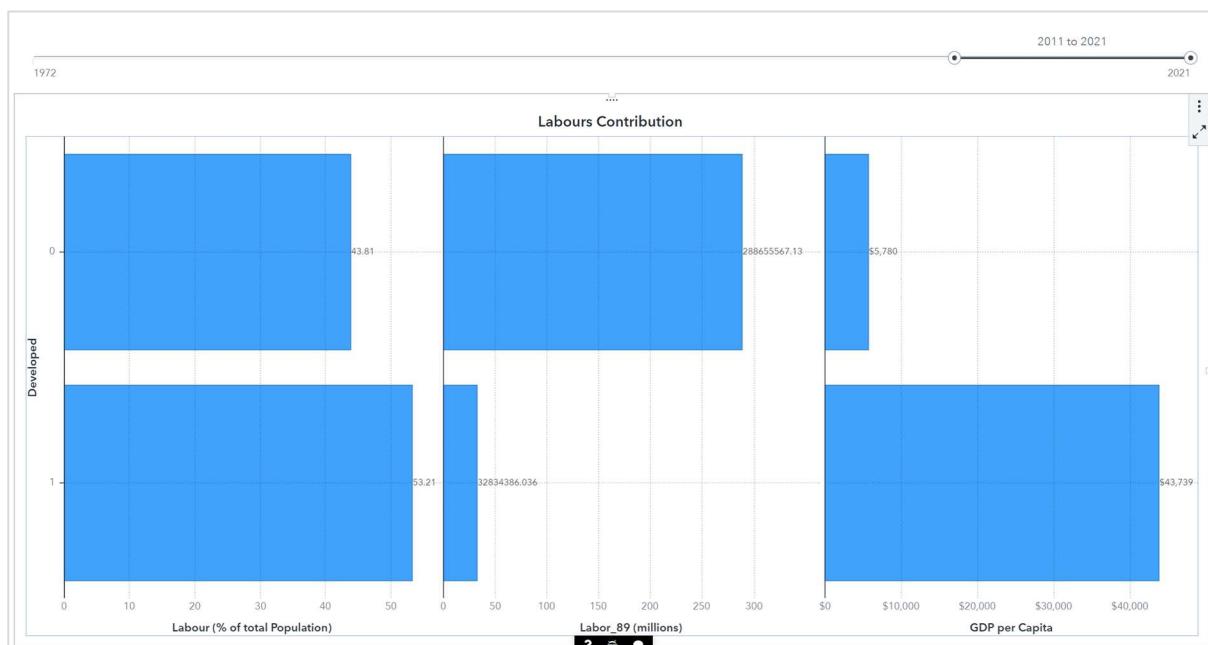


Figure 8.: Labour contribution in the last ten years

## Multilinear Regression

### Correlation matrix

All measure variables (exclusions are listed in the below Exclusion table) are investigated in a correlation matrix to examine their relationship.

“GDPperCapita” is most commonly used for measuring real economic growth as it considers the total value of goods and services produced by the economy and human capital (Lepenies & Gaines 2016). Hence, “GDPperCapita” and its variation are removed from the matrix as it would be used as the response for multilinear regression to examine the contributing factors. Similarly, Net National Income (NNI-NatIncome) that derives from GDP with added net receipts from abroad would be excluded. The (7) full exclusion table for contributing factors is as below:

Excluded variables	Rationale
GDPCapita	To be used as a Target/Response variable
GDP	Used as an input for GDPperCapita
Population	Used as an input for GDPperCapita
TaxGDP%	Derived from GDP
MfgValGDP	Derived from GDP
ServicesAdd	Derived from GDP
NatIncome	Used GDP as input; GDP is more commonly used.
IncomeCapita	Derived from NetIncome and Population
LandArea	Constant across time

The variables with high correlations will be scrutinized and removed to ensure they are independent for the following regression analyses.

The number of missing data points in these variables was considered in this step to allow a sufficient sampling size. Seven variables with low to moderate relationships were identified in the correlation as below (sorted ascendingly): -

No.	Remained	Removed
1	AirFreight	AirPassengers
2	HlthExCapita	
3	Inflation	DepIntRate
4	kWhCapita	
5	LenIntRate	
6	Pricelndex	
7	PrimaryEnrol	
8	AgriLand	

- Labor (correlated with Population used in GDPperCapita)

Figure 9 Correlation matrix shows that the variables with the highest correlation at 0.59 were “HlthExCapita” and “kWhCapita”, indicating a moderate relationship. Hence, all the variables are appropriate for the following analyses.

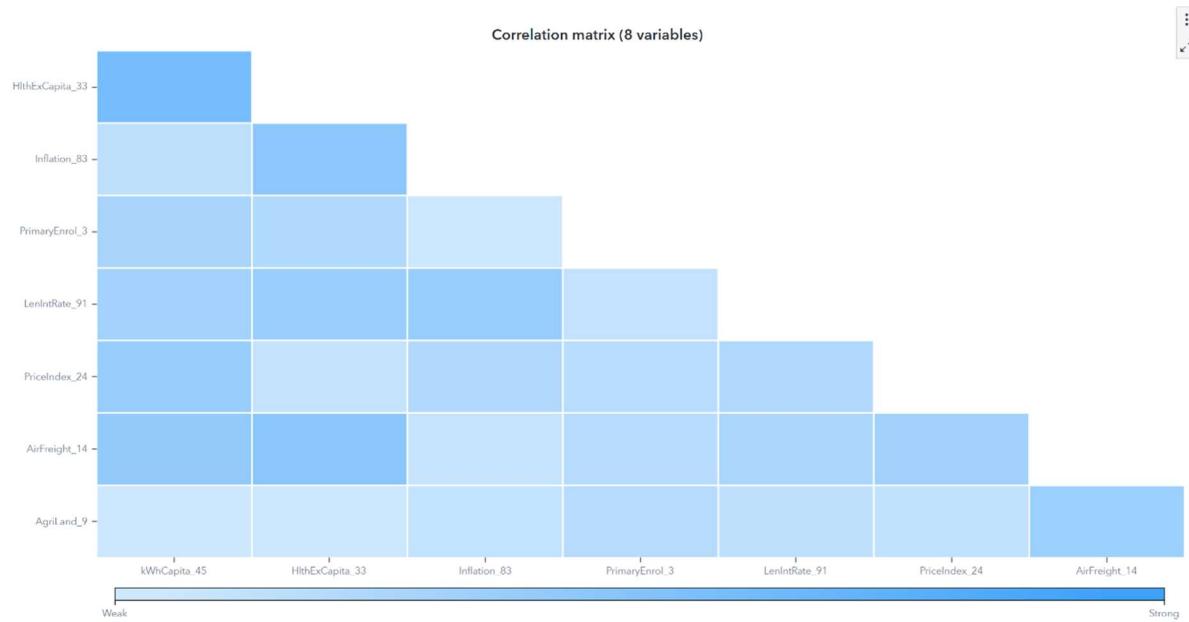


Figure 9: Correlation Matrix

### *Descriptive statistics*

Most of the explanatory factors, including the response "GDPperCapita", have a strong right-skewed tendency. A regression model's variables should be logarithmically transformed when there is a non-linear relationship between the independent and dependent variables (Ondruskova & Vitkova 2017, p.631). Thus, natural logarithm is applied in the model to give these factors a more symmetric or normal shape, a simplified interpretation and, significantly, to improve the overall model (as denoted in R-square value).

Hence, 'HlthExCapita', 'PrimaryEnrol', 'LenIntRate', 'Inflation', 'AirFreight', 'AgriLand', and 'GDPperCapita' variables are transformed to natural logarithm format as shown in the table below (sorted in descending order by skewness):

<b>Continuous effects</b>	<b>Mean</b>	<b>Coefficient of variation</b>	<b>Skewness</b>	<b>Missing data</b>
Inflation	19.2	798.7	14.0	107.0
LenIntRate	10.5	94.9	4.0	395.0
AirFreight	3,775.1	173.6	3.7	71.0
AgriLand	1,190,222.5	133.4	1.3	20.0
HlthExCapita	2,507.7	100.4	1.1	620.0
*kWhCapita	5,595.2	71.9	0.6	158.0
*PriceIndex	75.6	49.3	0.1	101.0
PrimaryEnrol	94.0	8.3	- 2.4	599.0

<b>Response effect</b>	<b>Mean</b>	<b>Coefficient of variation</b>	<b>Skewness</b>	<b>Missing data</b>
GDPperCapita	19,163.5	97.8	1.1	28.0

Note: Natural logarithms are not applied to variables marked (\*)

Source: *Descriptive data recorded from measure details in SAS*

A control button for Developed was added to the model to distinguish the contributing factors between the two grouped economies. The performance of these models will be assessed individually in the arranged order: (a) All countries, (b) Developed countries and (c) Developing countries.

Moreover, due to the high number of missing data points, 'information missingness' is applied in the models where the used variables are imputed with observed means and 'new variables' that indicate missingness are created alongside.

### Overall model - All countries

In Figure 10, the eight variables identified above will be tested in multilinear regression. It is observed that the three variables (Ln Inflation, Ln PrimaryEnrol, and Ln LenIntRate) with P-value on the borderline or greater than 0.01 (marked in blue and light purple bars) are removed, and their contributions to the adjusted R-square improvement are minimal.

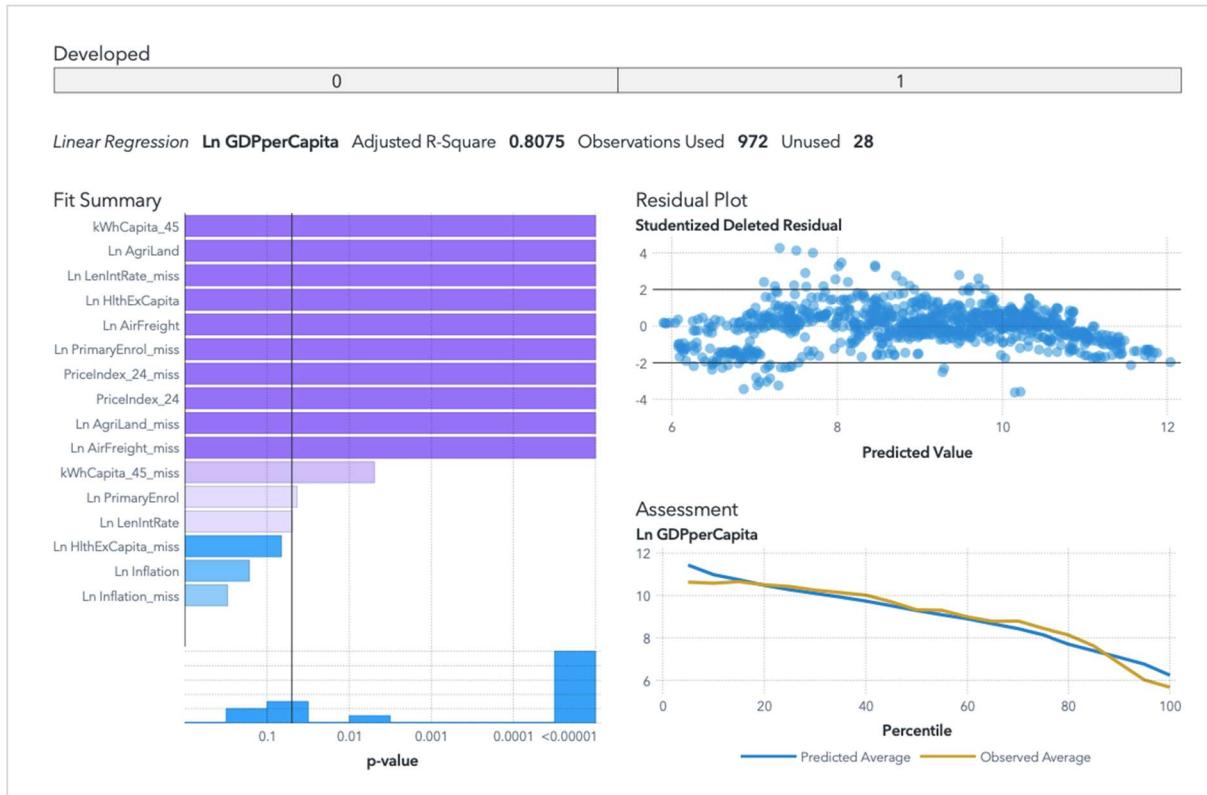


Figure 10. Fit summary of multilinear regression for all eight variables

Therefore, in Figure 11, the five statistically significant variables remained in this model: “kWhCapita”, “Ln AgriLand”, “Ln HlthExCapital”, “PricelIndex”, and “Ln AirFreight”. The model recorded total observations of 972 with the application of information missingness.

The Adjusted R-square of the overall model stood at 0.7656, indicating that 76.56% of the variability observed in Ln GDPperCapita is explained by the regression model, suggesting a good fit for the model. The F-value = 318.13 and the corresponding p-value ( $\text{Pr} > \text{F}$ )  $< 0.00001$ , showing that this model is statistically significant, and Ln GDPperCapita is statistically dependent on the five selected variables.



Figure 11, Multilinear regression for all five variables

In overall economies, 'Ln HlthExCapita' is the most important variable for Ln "GDPperCapita", which exhibits a coefficient of 0.36, explaining the importance of healthcare budget allocation in all nations to provide better health opportunities to improve human capital that crucially related to a nation's productivity.

On the other hand, the coefficient of 'Ln AgriLand' at -0.11 indicates a negative relationship with Ln GDPperCapita. This can be expressed that the higher the "GDPperCapita", the lesser the dependency on the agricultural product (shown by the agricultural land use).

### Multilinear regression equation:

$$\text{Ln GDPperCapita} = 5.05 - 0.12 * \text{Ln AgriLand} + 0.19 * \text{Ln AriFreight} + 0.36 * \text{Ln HlthExCapita} + 0.01 * \text{PriceIndex} + 0.0002 * \text{kWhCapita}$$

Developed	<input type="checkbox"/> 0	<input checked="" type="checkbox"/> 1				
Linear Regression Ln GDPperCapita	Adjusted R-Square	0.7656	Observations Used	972	Unused	28
Dimensions	Overall ANOVA	Fit Statistics	Parameter Estimates	Type III Test	Assessment	Assessment Statistics
Parameter		Estimate	Standard Error	t Value	Pr >  t	
Intercept		5.049875	0.250669	20.14563	<0.0001	
Ln AgriLand		-0.11632	0.007949	-14.6326	<0.0001	
Ln AirFreight		0.189749	0.020712	9.161423	<0.0001	
Ln HlthExCapita		0.359991	0.025839	13.932	<0.0001	
PricelIndex_24		0.006851	0.00114	6.006983	<0.0001	
kWhCapita_45		0.000198	8.158E-6	24.29799	<0.0001	

Figure 12: Parameter Estimates for All countries

### (a) Developed countries

When developed = 1, the adjusted R-square revised downwards slightly to 0.7537, which means the model explains 75.37% of the observations in “Ln GDPperCapita”, and the strength of association remained high.

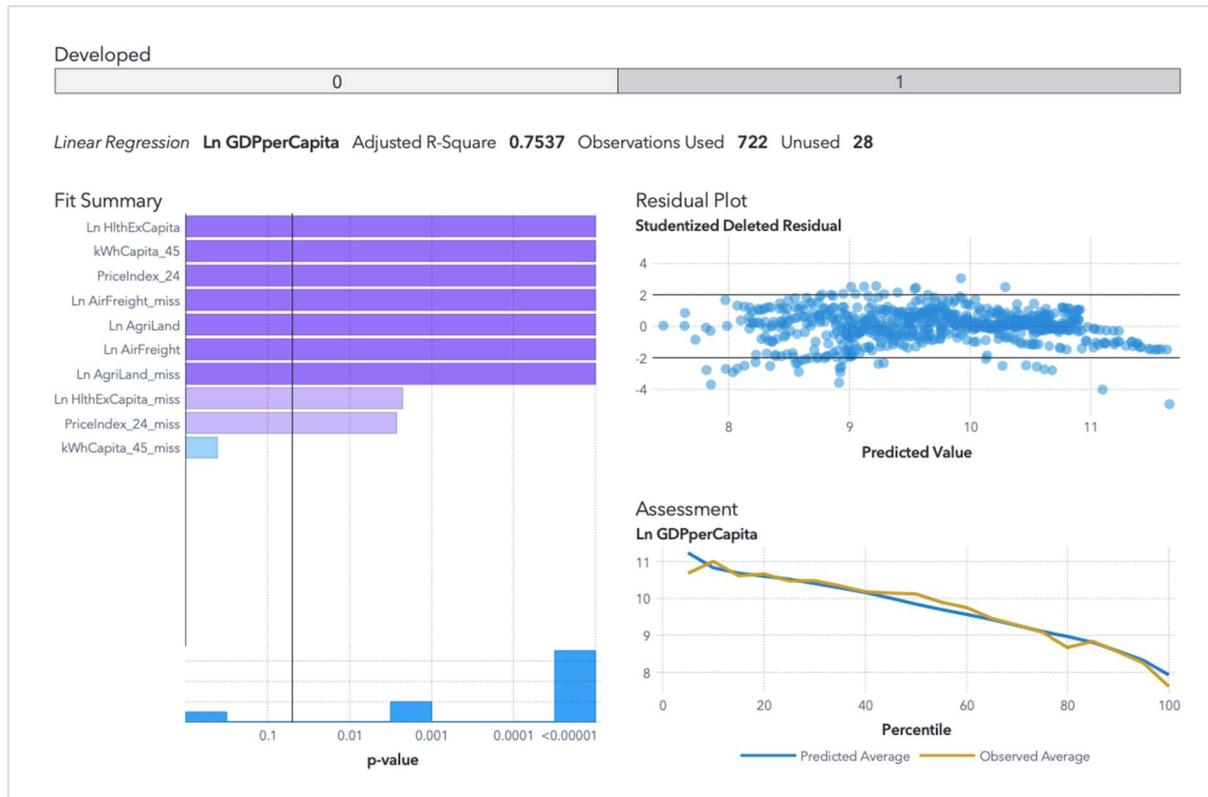


Figure 13. Multilinear regression for the developed group

Figure 14 ANOVA table indicates that the F-value = 221.68 and the corresponding p-value ( $\text{Pr} > \text{F}$ )  $< 0.00001$ , explaining this model for developed countries is statistically significant.

Developed	
0	1
<b>Linear Regression Ln GDPperCapita</b>	
Adjusted R-Square 0.7537 Observations Used 722 Unused 28	
Dimensions	Overall ANOVA
Source	Deg Freedom
Model	10
Error	711
Corrected Total	721
Sum of Squares	519.9512
Mean Square	51.99512
F Value	221.6803
Pr > F	<0.00001
R-Square	0.757156

Figure 14: ANOVA result for the developed group

In developed countries, 'Ln HlthExCapita' is the most crucial variable, which exhibits a coefficient of 0.44. Health expenditure is emphasized for developed countries as the relationship between human capital productivity and "GDPperCapita" is established (Hitiris and Posnett, 1992). The 'Ln AgriLand' coefficient at -0.04 indicates a negative relationship with GDPperCapita. Advanced economies tend to move away from focusing on agricultural development to improve economic growth.

### Multilinear regression equation:

$$\text{Ln GDPperCapita} = 4.38 - 0.04 * \text{Ln AgriLand} + 0.11 * \text{Ln AriFreight} + 0.44 * \text{Ln HlthExCapita} + 0.01 * \text{PriceIndex} + 0.000095 * \text{kWhCapita}$$

Developed																																				
	0      1																																			
<b>Linear Regression Ln GDPperCapita</b>																																				
Adjusted R-Square	0.7537																																			
Observations Used	722																																			
Unused	28																																			
.....																																				
<a href="#">Dimensions</a> <a href="#">Overall ANOVA</a> <a href="#">Fit Statistics</a> <b>Parameter Estimates</b> <a href="#">Type III Test</a> <a href="#">Assessment</a> <a href="#">Assessment Statistics</a>																																				
<table border="1"> <thead> <tr> <th>Parameter</th> <th>Estimate</th> <th>Standard Error</th> <th>t Value</th> <th>Pr &gt;  t </th> </tr> </thead> <tbody> <tr> <td>Intercept</td> <td>4.377485</td> <td>0.244953</td> <td>17.87073</td> <td>&lt;0.00001</td> </tr> <tr> <td>Ln AgriLand</td> <td>-0.04026</td> <td>0.005838</td> <td>-6.89694</td> <td>&lt;0.00001</td> </tr> <tr> <td>Ln AirFreight</td> <td>0.107674</td> <td>0.016158</td> <td>6.663905</td> <td>&lt;0.00001</td> </tr> <tr> <td>Ln HlthExCapita</td> <td>0.441278</td> <td>0.028043</td> <td>15.73595</td> <td>&lt;0.00001</td> </tr> <tr> <td>PriceIndex_24</td> <td>0.013639</td> <td>0.000984</td> <td>13.86518</td> <td>&lt;0.00001</td> </tr> <tr> <td>kWhCapita_45</td> <td>0.000095</td> <td>6.673E-6</td> <td>14.21341</td> <td>&lt;0.00001</td> </tr> </tbody> </table>		Parameter	Estimate	Standard Error	t Value	Pr >  t	Intercept	4.377485	0.244953	17.87073	<0.00001	Ln AgriLand	-0.04026	0.005838	-6.89694	<0.00001	Ln AirFreight	0.107674	0.016158	6.663905	<0.00001	Ln HlthExCapita	0.441278	0.028043	15.73595	<0.00001	PriceIndex_24	0.013639	0.000984	13.86518	<0.00001	kWhCapita_45	0.000095	6.673E-6	14.21341	<0.00001
Parameter	Estimate	Standard Error	t Value	Pr >  t																																
Intercept	4.377485	0.244953	17.87073	<0.00001																																
Ln AgriLand	-0.04026	0.005838	-6.89694	<0.00001																																
Ln AirFreight	0.107674	0.016158	6.663905	<0.00001																																
Ln HlthExCapita	0.441278	0.028043	15.73595	<0.00001																																
PriceIndex_24	0.013639	0.000984	13.86518	<0.00001																																
kWhCapita_45	0.000095	6.673E-6	14.21341	<0.00001																																

Figure 15: Parameter Estimates for the developed group

### (b) Developing countries

When Developing=0, The adjusted R-square is 0.8394, which means the model explains 83.94% variation in “Ln GDPperCapita”. The strength of association is high and suggests a better fit model compared to the All countries (0.7656) and Developed countries (0.7537) models above.



Figure 16: Multilinear regression for developing group

Figure 17 ANOVA table indicates that the F-value = 131.17 and the corresponding p-value ( $\text{Pr} > \text{F}$ )  $< 0.00001$ , explaining this model for developing countries is statistically significant. “Ln GDPperCapita” is statistically dependent on the five selected variables.

Developed		Linear Regression Ln GDPperCapita		Adjusted R-Square	0.8394	Observations Used	250
Dimensions	Overall ANOVA	Fit Statistics	Parameter Estimates	Type III Test	Assessment	Assessment Statistics	
Source		Deg Freedom		Sum of Squares		Mean Square	
Model		10		371.2869		37.12869	F Value
Error		239		67.64704		0.283042	Pr > F
Corrected Total		249		438.934		.	R-Square
							<0.00001 0.845883

Figure 17: ANOVA result for developing group

In developing countries, “Ln AirFreight” has the highest coefficient of 0.33, indicating that it is the critical driver for “Ln GDPperCaptia” in developing economies. Air Freight is indispensable in stimulating economic growth by liberating and improving air cargo networks to enable global trade (International Civil Aviation Organization 2015).

Similarly, the coefficient of 'Ln AgriLand' has a negative impact on “Ln GDPperCapita” that suggests the developing economies should shift their focus away from agriculture to other developments.

### Multilinear regression equation:

$$\text{Ln GDPperCapita} = 5.57 - 0.17 \cdot \text{Ln AgriLand} + 0.33 \cdot \text{Ln AirFreight} + 0.24 \cdot \text{Ln HlthExCapita} + 0.000944 \cdot \text{PriceIndex} + 0.000492 \cdot \text{kWhCapita}$$

Developed				
	0 1			
<i>Linear Regression Ln GDPperCapita</i> Adjusted R-Square <b>0.8394</b> Observations Used 250				
Dimensions Overall ANOVA Fit Statistics Parameter Estimates Type III Test Assessment Assessment Statistics				
Parameter	Estimate	Standard Error	t Value	Pr >  t
Intercept	5.572312	0.453048	12.29961	<0.0001
Ln AgriLand	-0.17229	0.046037	-3.74233	0.00023
Ln AirFreight	0.33119	0.046472	7.125066	<0.0001
Ln HlthExCapita	0.242845	0.045147	5.378953	<0.0001
PriceIndex_24	0.000944	0.001547	0.610074	0.54239
kWhCapita_45	0.000492	0.000029	16.92305	<0.0001

Figure 18: Parameter Estimates for developing group

## Logistic regression

A logistic regression model has been built to predict whether a country is developed. No meaningful categorical predictors were present, and all the continuous variables used in the model were picked based on the correlation matrix to avoid using highly correlated predictors together.

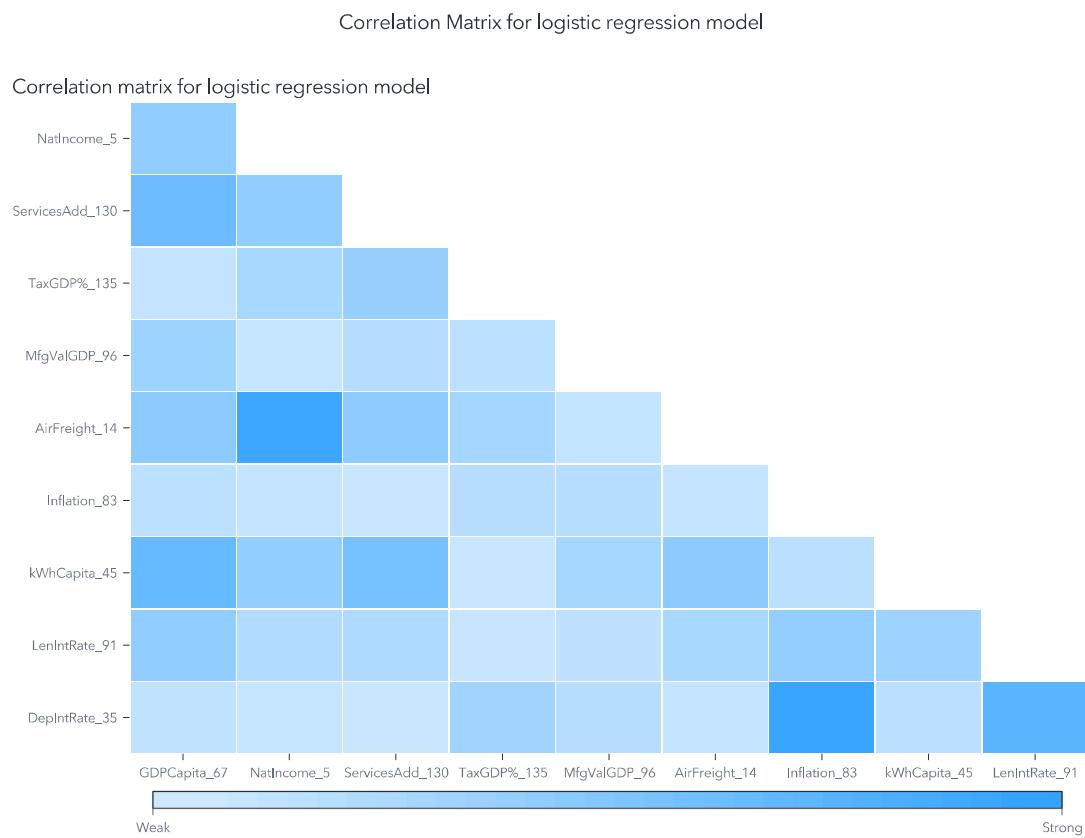


Figure 19. Correlation matrix

Due to many missing values, two variables representing health expenditure per capita and primary school enrolment rate were not used. The selected variables are as follows:

Selected variables for the Logistic regression model
MfgValGDP_96
GDPCapita_67
kWhCapita_45

TaxGDP%_135
AirFreight_14
ServicesAdd_130
Inflation_83

Due to the relatively large number of missing values and the small number of observations, SAS did not permit to apply partition to the model. the Primary model is as follows:

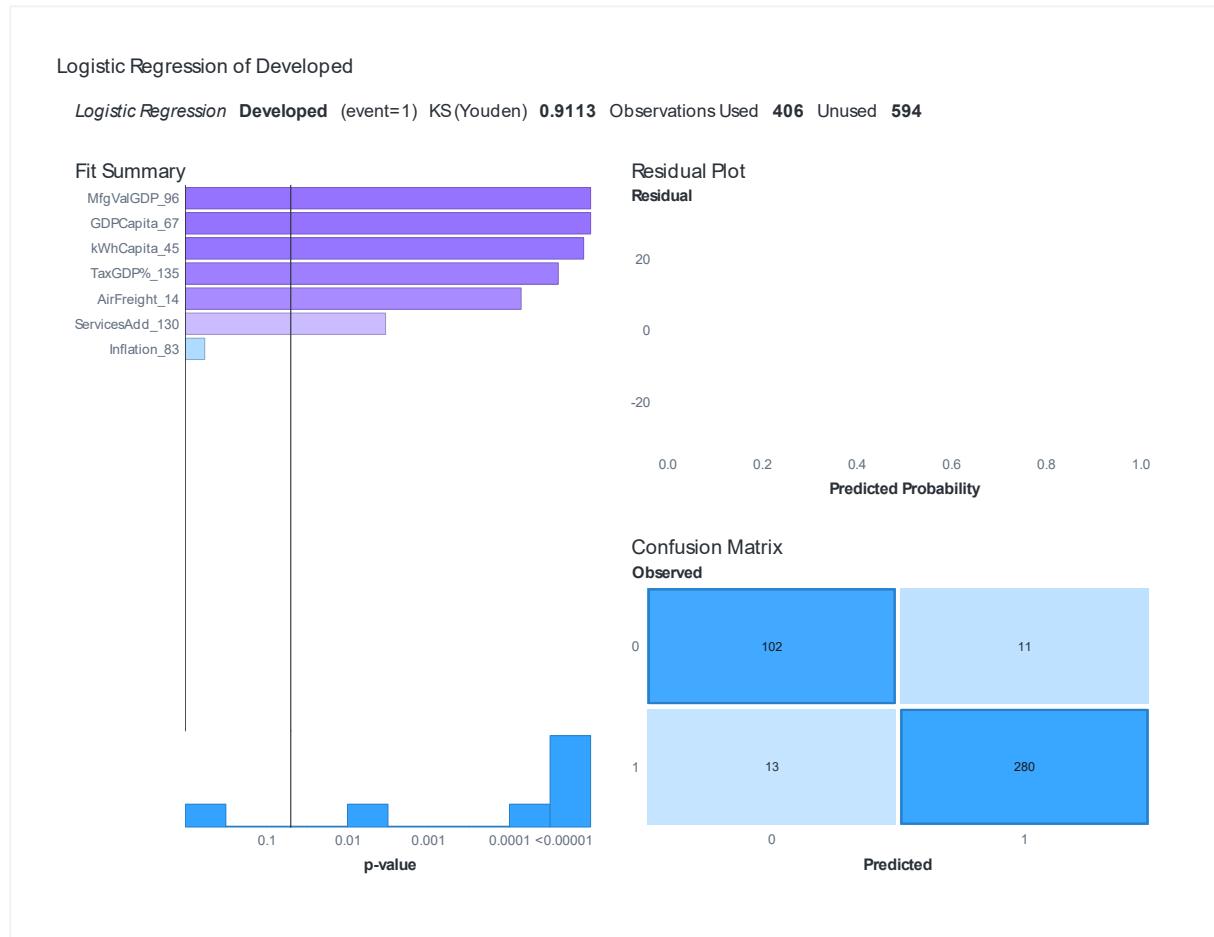


Figure 20. The primary logistic regression model

### Optimization of the Model

The KS (Youden) in the primary model is 0.9113, with 594 unused observations. First, the "Informative missingness" function with the backward "Variable selection method" is tested to optimize the accuracy of the model. Since it reduced the model's total accuracy, the function had no benefit. Also, using half of the

observations based on speculation may result in an unrealistic model. Thus, the use of the "Informative missingness" function is rejected.

The significance level is specified at 0.01, and predictors that do not meet this significant level are eliminated from the model. Thus, the five predictors for the final model are "GDPCapita\_67", "MfgValGDP\_96", "AirFreight\_14", "kWhCapita\_45", and "TaxGDP%\_135".

From the residual plot and the influence plot outliers are excluded. This action contributed to the model's accuracy and increased the KS(Youden) index to 0.9177.

Since the accuracy of the model, not the prediction of one specified event, is the main objective, based on the ROC graph, the cutoff is specified at 0.59 to produce a less misclassification rate.

#### *Assessment of the model*

The final model overview is as follows:

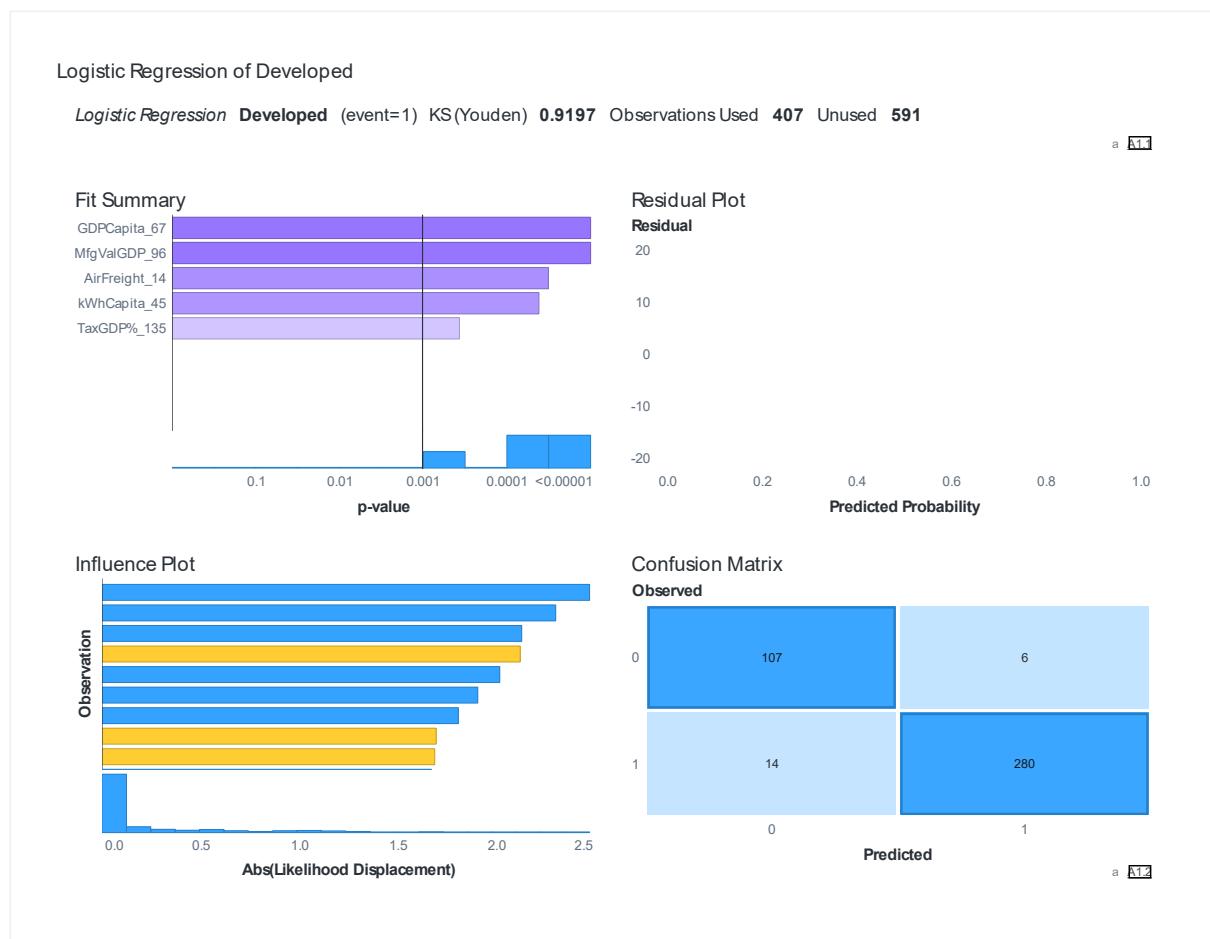


Figure 21. The final logistic regression model

The KS(Youden) of 0.9197 indicates a robust predictive model.

The misclassification rate is calculated as follows:

$$\text{Model's Accuracy} = 1 - \text{Misclassification rate} \times 100 = (1 - 0.049) \times 100 = 95.1\%$$

The confusion matrix indicates that 94.69% of the cases are correctly classified in developing countries, and the model's accuracy is 95.24% in the prediction of developed countries.

Classifier	Result
Accuracy	95.1%
Sensitivity (True positive rate)	95.2%
Specificity (True negative rate)	94.6%

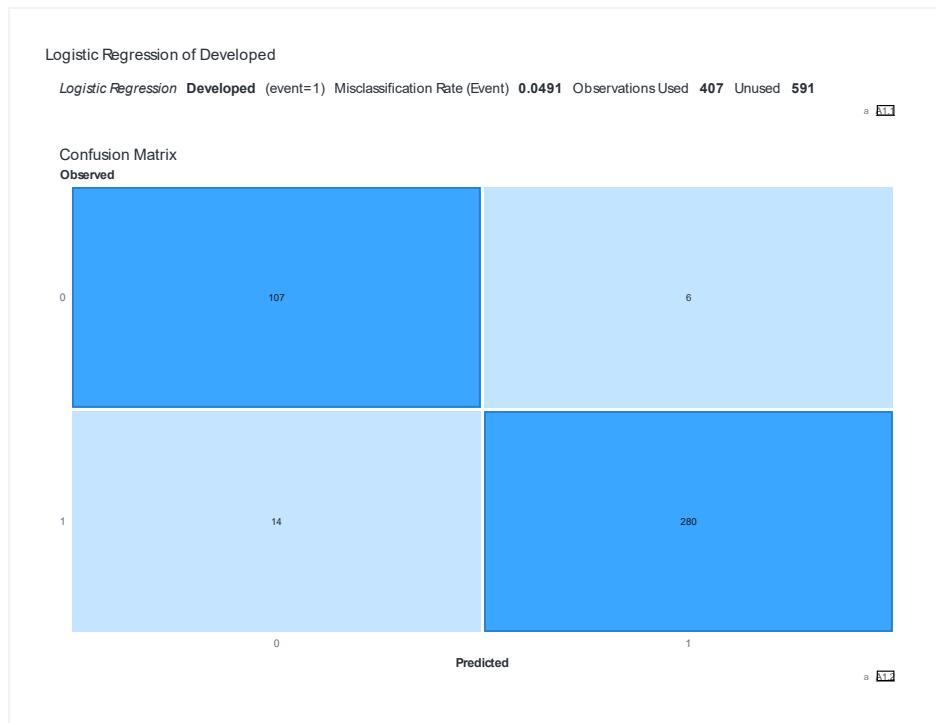


Figure 22. Confusion matrix of the final model

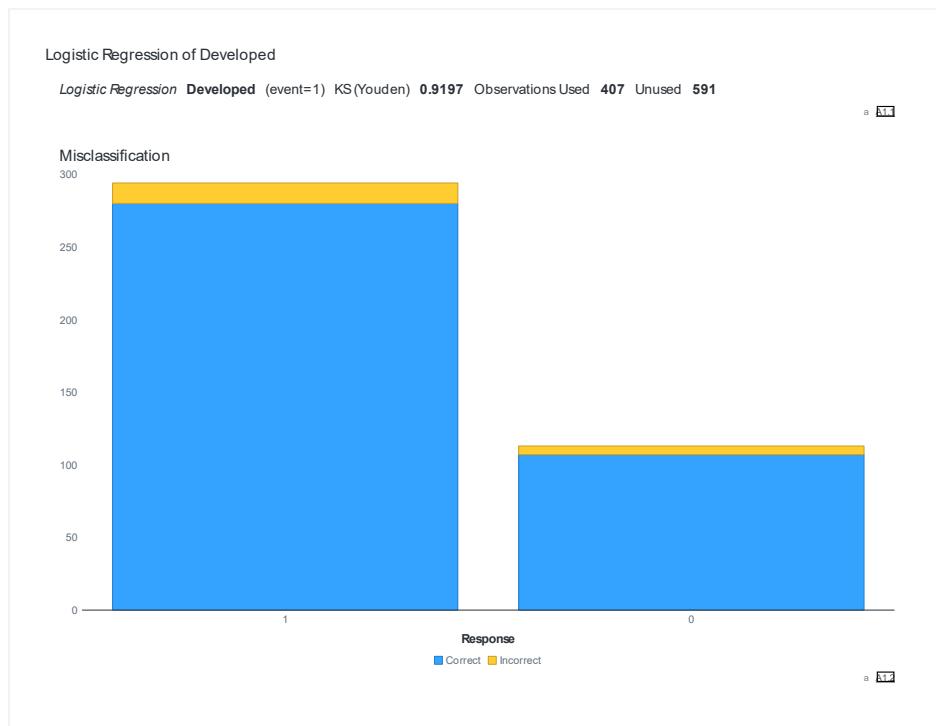


Figure 23. Misclassification table

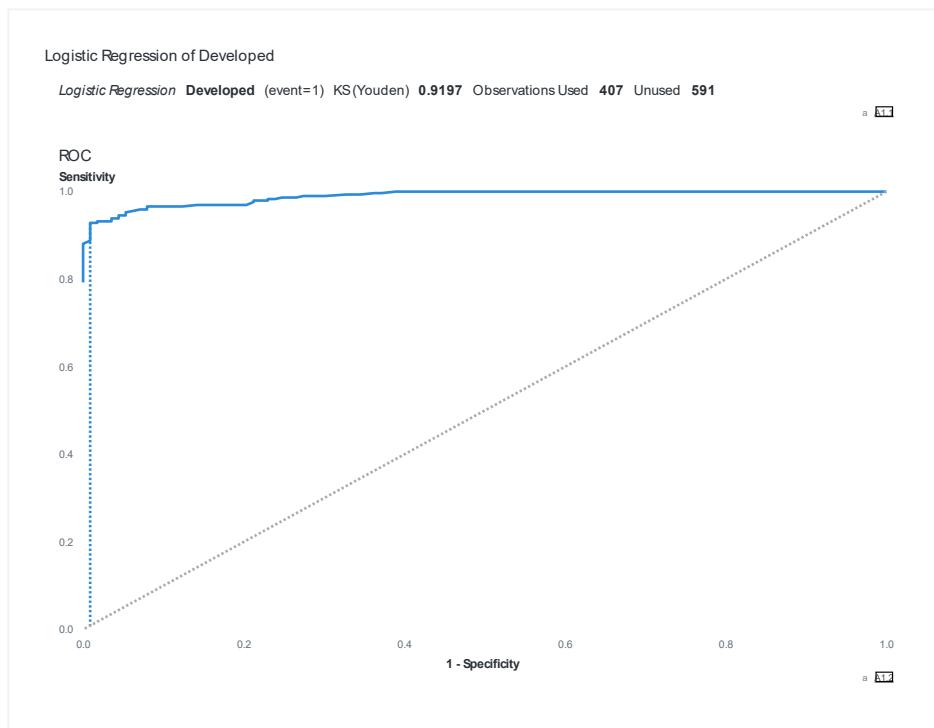


Figure 24. The ROC graph of the final model

### *Model interpretation*

The logistic regression model can predict whether a country is developed based on GDP per Capita, manufacturing and tax income percentage of GDP, electricity consumption per capita and Air freight volume in any given country.

## Forecasting

### GDP per Capita Trend analysis

Figure 25 shows the overall GDP growth has an upward mild-cyclical pattern that follows the economic cycle, which encompasses 4 phases (expansion, peak, contraction, and trough) affected by several economic factors (Mehta 2022). The average actual GDP per Capita in 2021 was \$37,164.

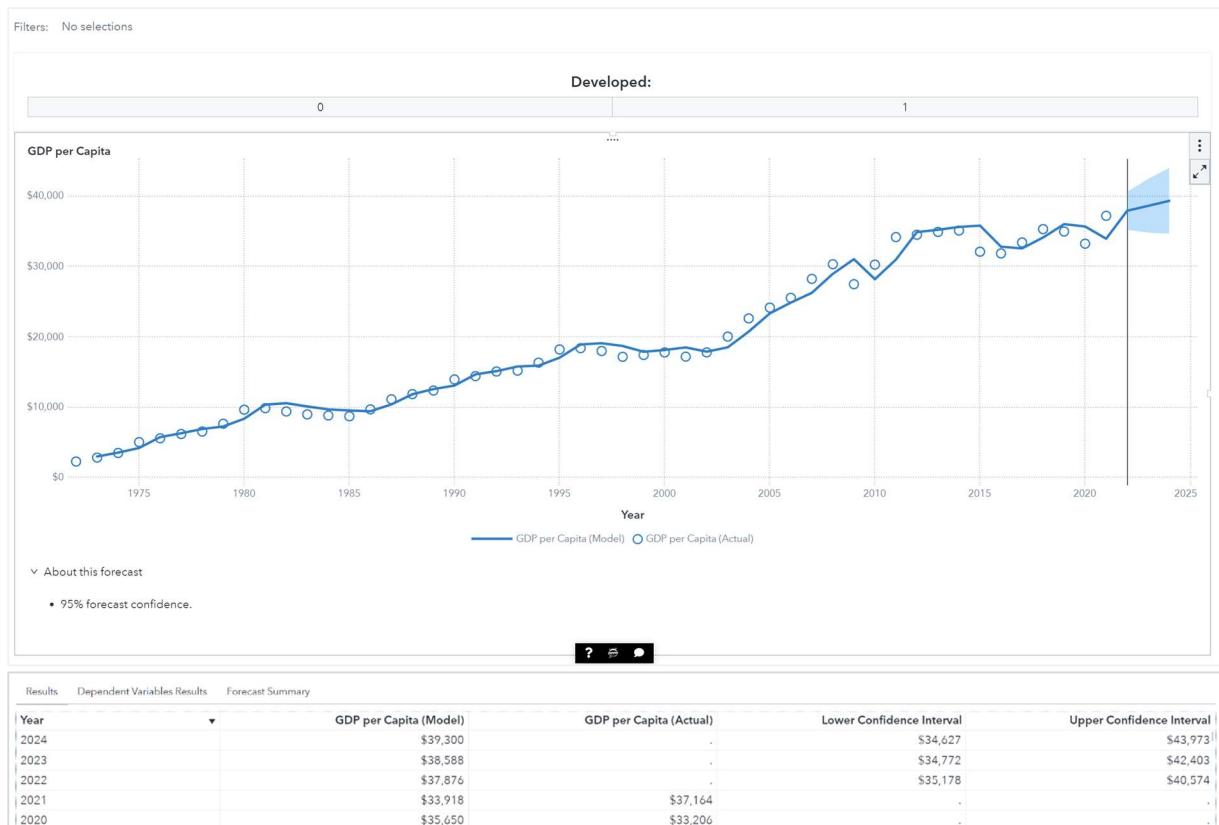


Figure 25: Forecasting model without underlying factors

## *Forecasting Model*

### *(a) All countries*

The consumer price index (PriceIndex\_24) variable is eliminated from the eight remaining variables outlined in the Correlation matrix as it is usually used as a government benchmark to determine the inflation rate. Hence, six variables were used as the underlying factors to predict the GDP per capita growth in the effort to strategize monetary policy.

No.	<b>Underlying factors used</b>
1	AirFreight
2	HlthExCapita
3	Inflation
4	kWhCapita
5	LenIntRate
6	PrimaryEnrol
7	AgriLand

This model has 95% forecast confidence which is used to predict the GDP per capita growth in the coming three years (set forecast horizon to 3) from 2022 to 2024. The optimal algorithm chosen by the SAS model is ARIMA.

The model will also evaluate the underlying factors if they are contributing to the accuracy of the forecast. Overall, current health expenditure per capita and inflation are found to increase the accuracy of the forecast as it narrows the confidence interval. As such, the remaining variables were not found to contribute to the accuracy and were not added to the model.

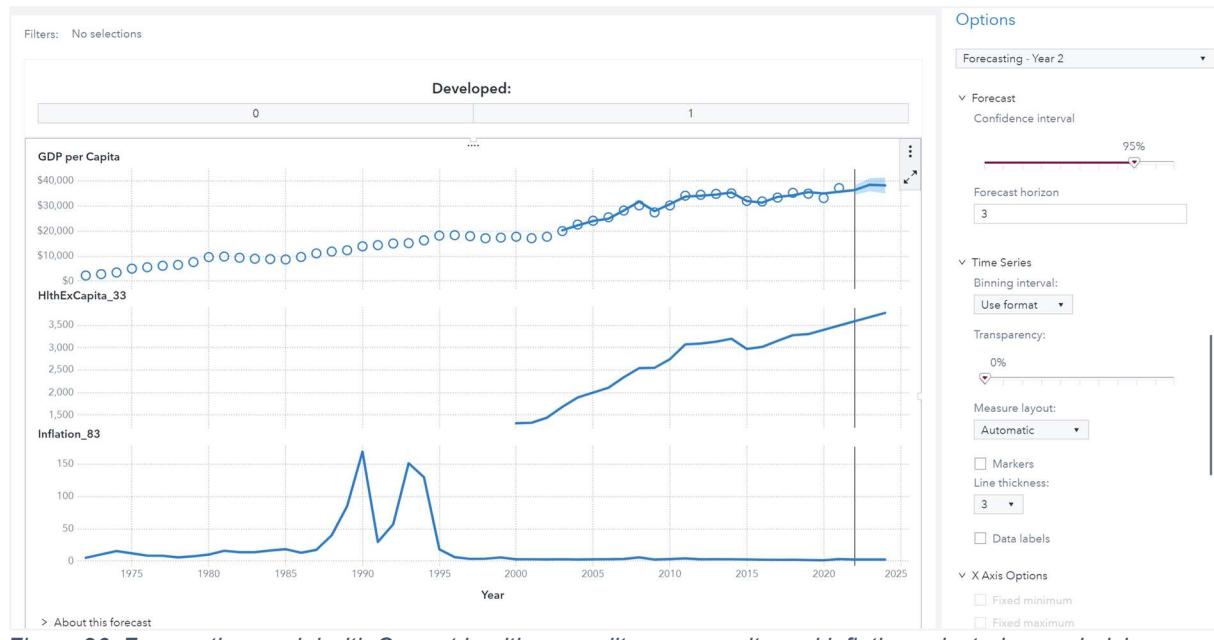


Figure 26: Forecasting model with Current health expenditure per capita and inflation selected as underlying factors – All countries

Besides, What-If Analysis was added to perform goal-seeking analysis to forecast GDP per capita by an increment of 10%. The original forecast was indicated by the black lines, while the blue lines show the new targeted forecasts. The model explains that to increase the “GDPperCapita”, current health expenditure per capita and inflation have to be increased. For example, to achieve a “GDPperCapita” of \$39,976 in 2022, elements in the monetary policy, such as current health expenditure per capita and inflation rate for consumer prices, have to revise to \$3,594 and 9.86, respectively, if both the developed and developing economies are managed together.

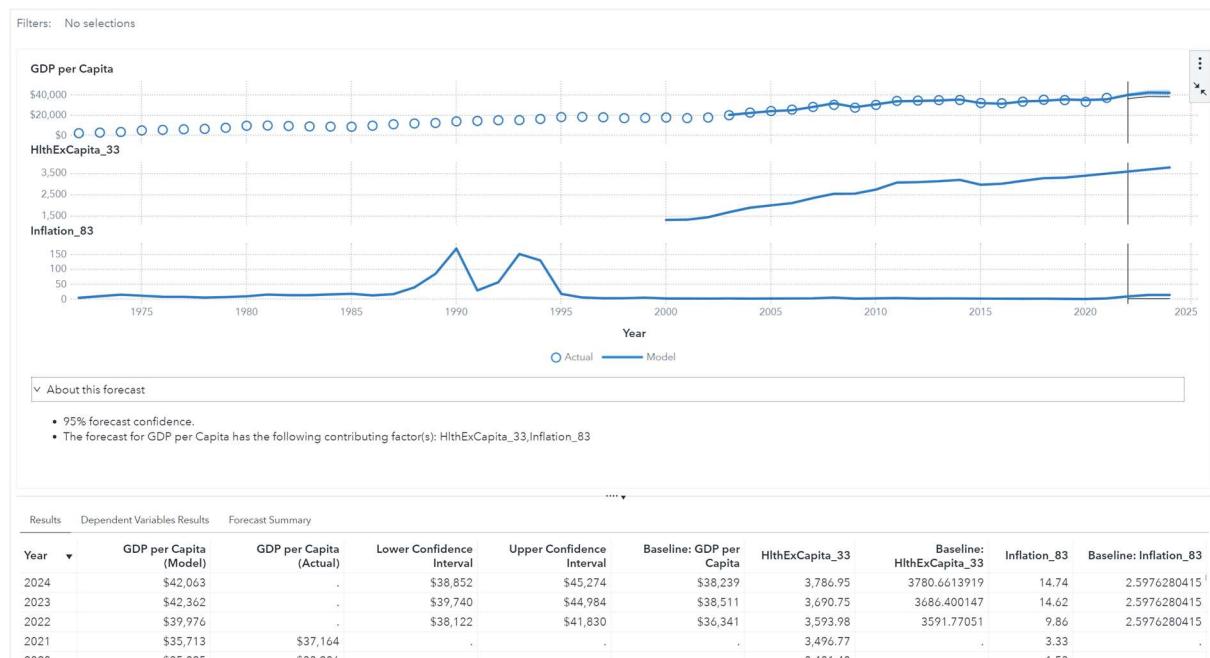


Figure 27. Forecasting model with Current health expenditure per capita and inflation selected as underlying factors with Goal seeking analysis of 10% increment – All countries

### (b) Developed countries

When Developed =1, it shows that current health expenditure per capita in the developed economies model would improve the model's accuracy and narrow the confidence interval. To increase the “GDPperCapita” by 10% within the developed countries, it is required to increase health expenditure per capita. It aligns with the findings in multilinear regression, where healthcare is of utmost priority for developed nations.

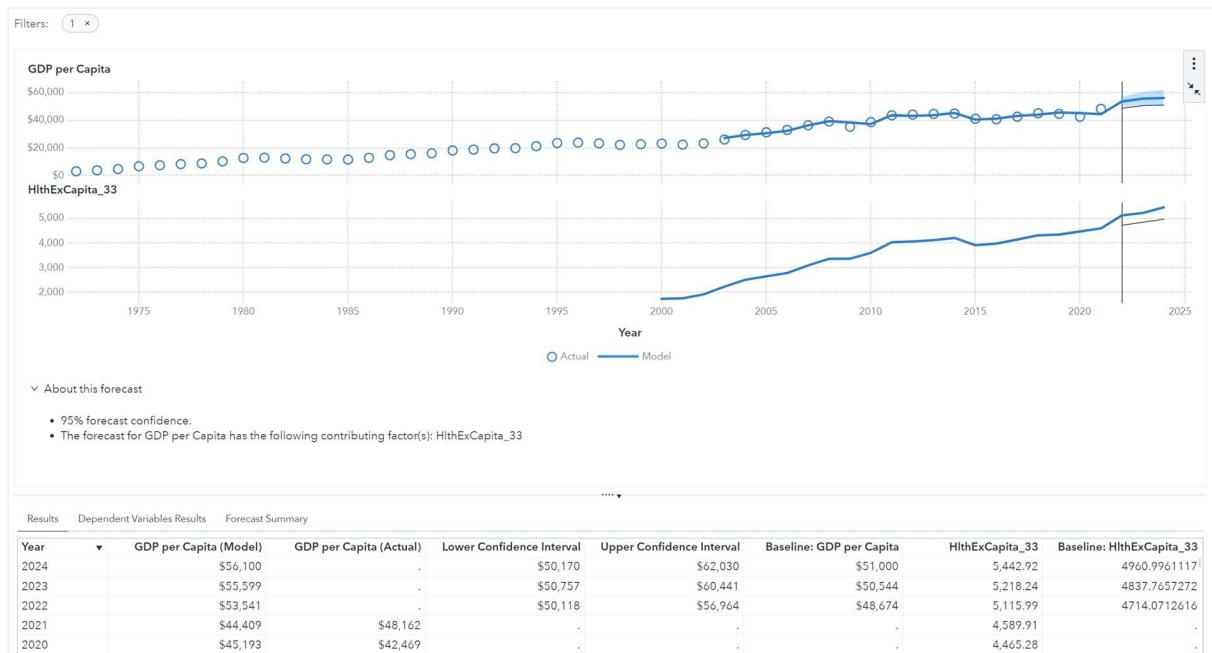


Figure 28. Forecasting model with Current health expenditure per capita as selected underlying factor with Goal seeking analysis of 10% increment – Developed countries

### (c) Developing countries

Similarly, when Developed=0, the model selects current health expenditure per capita, which would contribute to the model's accuracy. It stresses again the healthcare spending to achieve a 10% increment in GDPperCapita.



Figure 29: Forecasting model with Current health expenditure per capita as selected underlying factor with Goal seeking analysis of 10% increment – Developing countries

## **Conclusion and Recommendation**

The key insights from the explanatory, predictive and prescriptive analysis of the WDI dataset are as follows:

1. Global North/South indicator is obsolete to categorizing economies. Instead, the developed/developing indicator based on the Human Development Index (HDI) is more relevant and inclusive.
2. Developing countries put greater emphasis on agricultural industry as evidently in their agricultural land use ratio while developed countries shifted their focus away to tap on other land development opportunities
3. Developed countries tend to pursue a more disciplined macroeconomic policies by maintaining a stable deposit and interest rate corresponding inflation rate over the years. In contrast, in adjusting those rates, developing countries' financial policies were not constant concerning the inflation rate.
4. Developed countries have high proportion of GDP constitutes from value-added services and tax gains. On the other hand, in developing countries, manufacturing activities' revenues account a higher percentage of their GDP.
5. In the past ten years, the percentage of the working population was higher in developed countries.
6. In the past ten years, on average, each Labourer in developed countries has generated seven times more wealth than their counterparts in developing countries
7. By inserting the values of GDP per capita, electricity consumption per citizen, air freight load and the percentage of manufacturing and tax revenue of GDP, the presented logistic regression model can predict whether any given country is developed with 95% accuracy.
8. Health Expenditure is the key driver for economy growth in 3 years' forecasts for both developed and developing economies respectively.
9. To achieve 10% increment in GDPperCapita in 3 years forecast, health expenditure per capita and inflation have to be increased to meet the target for overall economies.
10. In both developed and developing countries, agriculture is projected to be shifted to other fields to improve the economic.

The lack of some data items has obstructed the analysis from further investigations. It was noted the dataset contains numerous missing data points which potentially leads to inconclusive analyses, with the most significant one being current health expenditure per capita that consists of 620 missing data points. Nevertheless, this variable was termed as a crucial factor in various models affecting economic growth and priority should be focus to the completeness in data collection to produce an insightful report.

Besides, the table below shows the additional data that can be collected to enhance future reporting: -

Trade of balance
Exchange rate
Foreign Direct Investment
Government total Expenditure
Unemployment rate
Literacy rate
Population age
Taxation rate and policy in countries
Agricultural production

For example, the lack of information on the population's age distribution and the criteria in which each country identifies a person as a worker has prevented us from determining how countries' aging trend contributes to its economic outcomes.

## References

- Araujo, JP & Junior, MR 2018, 'Price setting in Brazil from 1989 to 2007: Evidence on hyperinflation and stable', *Banco Central do Brasil*.
- Elistia, S 2018, 'The correlation of the Human Development Index (HDI) towards Economic Growth (Gdp Per Capita) in 10 Asean member countries', vol 2, no. 2, pp 40-46.
- Guttal, S 2016, *Interrogating the relevance of the Global North-South divide*, Focus on The Global South, viewed 8 November 2022,  
<[https://www.cetri.be/IMG/pdf/shalmali\\_guttal\\_23000\\_eng\\_3-2.pdf](https://www.cetri.be/IMG/pdf/shalmali_guttal_23000_eng_3-2.pdf)>
- Hitiris, T & Posnett, J 1992, 'The determinants and effects of health expenditure in developed countries', *Journal of Health Economics*.
- International Civil Aviation Organization 2015, *Impact of air cargo services on economic development*, viewed 14 November 2022.<[https://www.icao.int/Security/aircargo/Documents/AirCargo\\_Economic\\_Development.pdf](https://www.icao.int/Security/aircargo/Documents/AirCargo_Economic_Development.pdf)>
- Lepenies, P & Gaines, J 2016, 'The power of a single number: a political history of GDP', *Columbia University Press*, New York.
- Mehta S 2022, 'Stages of the Economic Cycle', *Financial Edge*, viewed 8 November 2022, <[https://www.fe.training/free-resources/asset-management/stages-of-the-economic-cycle/#:~:text=There%20are%20four%20stages%20in,stops%20decreasing%20and%20starts%20increasing\).>](https://www.fe.training/free-resources/asset-management/stages-of-the-economic-cycle/#:~:text=There%20are%20four%20stages%20in,stops%20decreasing%20and%20starts%20increasing).>)
- Moss, T et al. 2020, 'The Modern Energy Minimum: The case for a new global electricity consumption threshold', *Energy for Growth Hub*, viewed 8 November 2022,  
<<https://www.rockefellerfoundation.org/case-study/the-modern-energy-minimum/>>
- Ondruskova, E & Vitzkova, E 2017, of the construction company sales development by sales development by using time series analysis method - linear regression model with, *Section Environmental Economics*.
- Royal Geographical Society 2022, 'The Global North/South Divide', *Royal Geographical Society*, 8 November 2022,<<https://www.rgs.org/CMSPages/GetFile.aspx?nodeguid=9c1ce781-9117-4741-af0a-a6a8b75f32b4&lang=en-GB>>
- Solarz, MW 2019, *The Global North-South Atlas : Mapping Global Change*, 1st edn, Routledge, London, viewed 9 November 2022,  
<<https://www.routledge.com/The-Global-North-South-Atlas-Mapping-Global-Change/Solarz/p/book/9781138588844>>

World Population Review 2022, *Developed Countries List 2022*, viewed 8 November 2022, <<https://worldpopulationreview.com/country-rankings/developed-countries>>

## Appendix 1

### Investigate and prepare

The table below shows the finalized categorization of each variable:

Data points	
HlthExCapita_33	Measure
PrimaryEnrol_3	Measure
DepIntRate_35	Measure
LenIntRate_91	Measure
TaxGDP%_135	Measure
Labor_89	Measure
MfgValGDP_96	Measure
ServicesAdd_130	Measure
kWhCapita_45	Measure
NatIncome_5	Measure
IncomeCapita_7	Measure
Inflation_83	Measure
PriceIndex_24	Measure
AirPassengers_15	Measure
AirFreight_14	Measure
GDP_65	Measure
GDPCapita_67	Measure
AgriLand_9	Measure
LandArea_90	Measure
Population_110	Measure
Labour percentage of population	Measure
Agriland percentage of total land	Measure
Country	Category
CountryCode	Category
Year	Category
Developed	Category
Global North/South	Category
Date	Date

To ensure accuracy two variables (GDPperCapita and Income per Capita) are created to cross-check with the existing numbers:

New calculated items	Formulas	Existing items
GDPperCapita	GDP_65/Population_110	GDPCapita_67
Income per Capita	NatIncome_5/ Population_110	IncomeCapita_7

## Descriptive analysis

### Land area investigation

As indicated in Figure 6 Land area distribution across country, the growth largely remained constant across for each country across the years, with Russian Federation leads the way. Therefore, it should be eliminated in the following analyses as a contributing factor nor can be manipulated in monetary policy to sustain growth.



Figure 30: Land area by country and year – to examine variance in data documentation

# MBAS903\_A2\_P1

Creation Date: Wednesday, November 16, 2022 05:11:20 PM

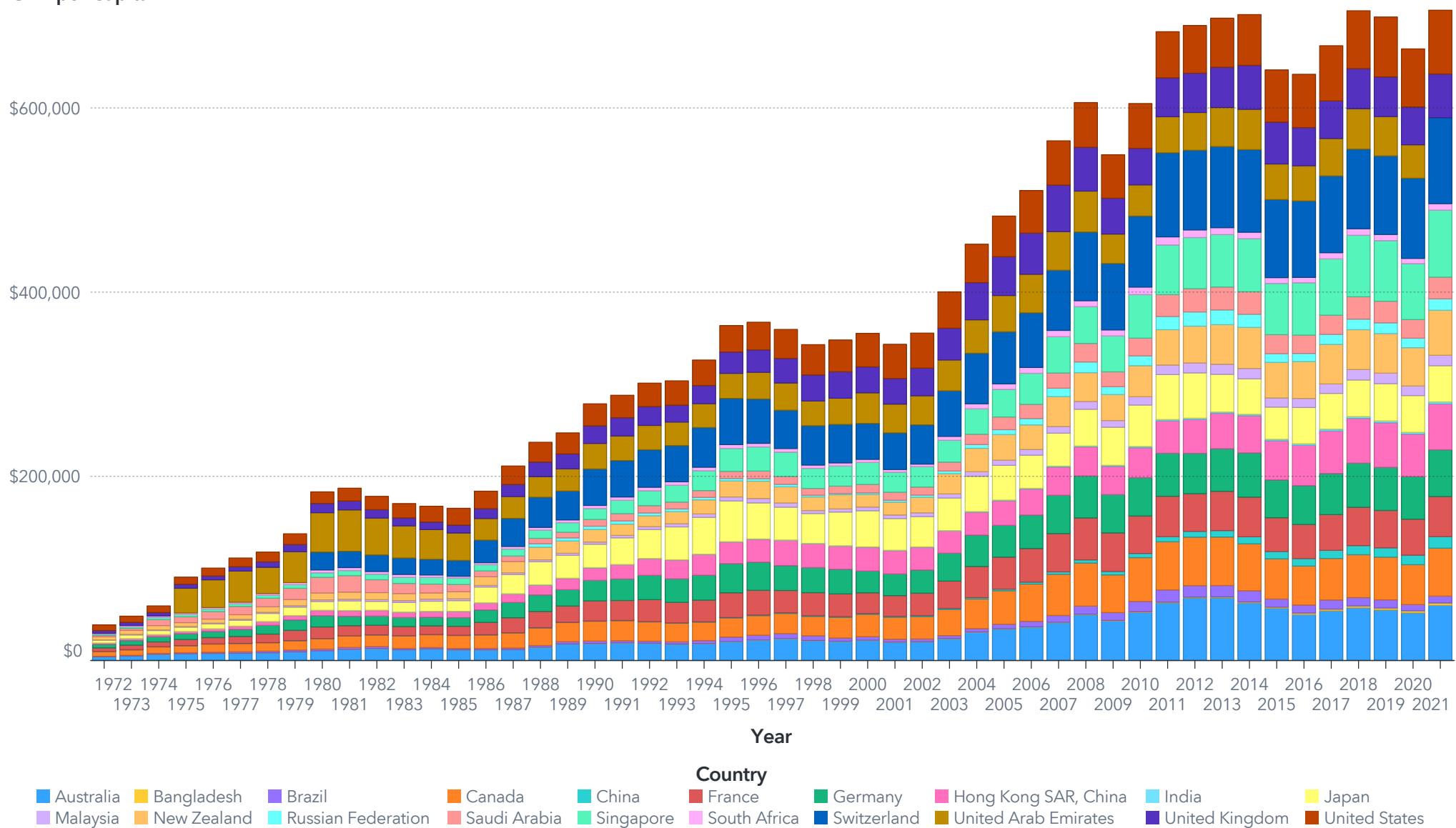
Author: kml957@uowmail.edu.au

Appendix

# Observation

## Observation: GDP per capita growth by countries in 50 years

GDP per Capita



## Data by year

Developed:

0

1

### Data availability by year

kWhCapita\_45

0

HlthExCapita\_33

0

Inflation\_83

0

PrimaryEnrol\_3

0

LenIntRate\_91

0

AirFreight\_14

0

PricelIndex\_24

0

Population\_110 (millions)

0

LandArea\_90 (millions)

0

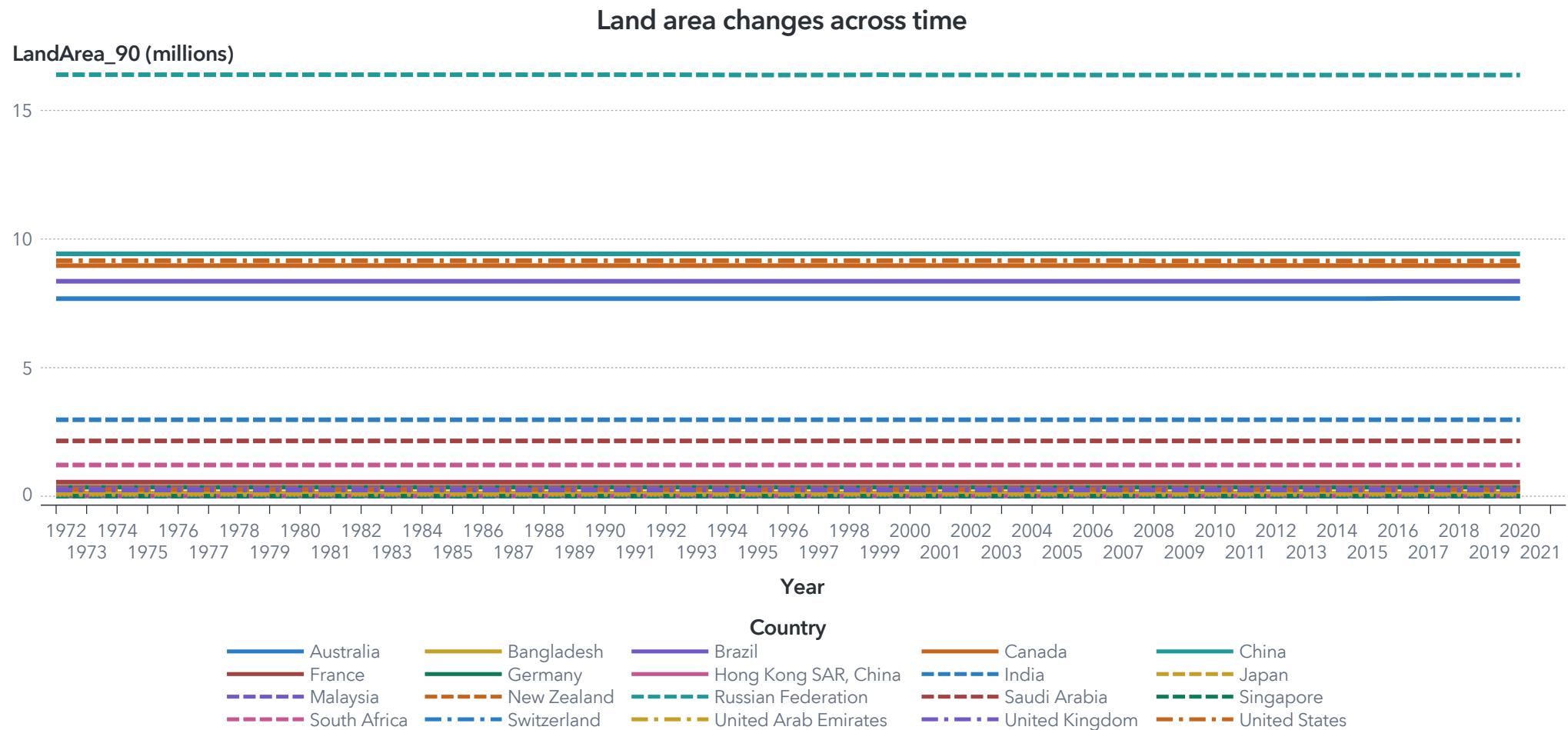
AgriLand\_9 (millions)

0.00

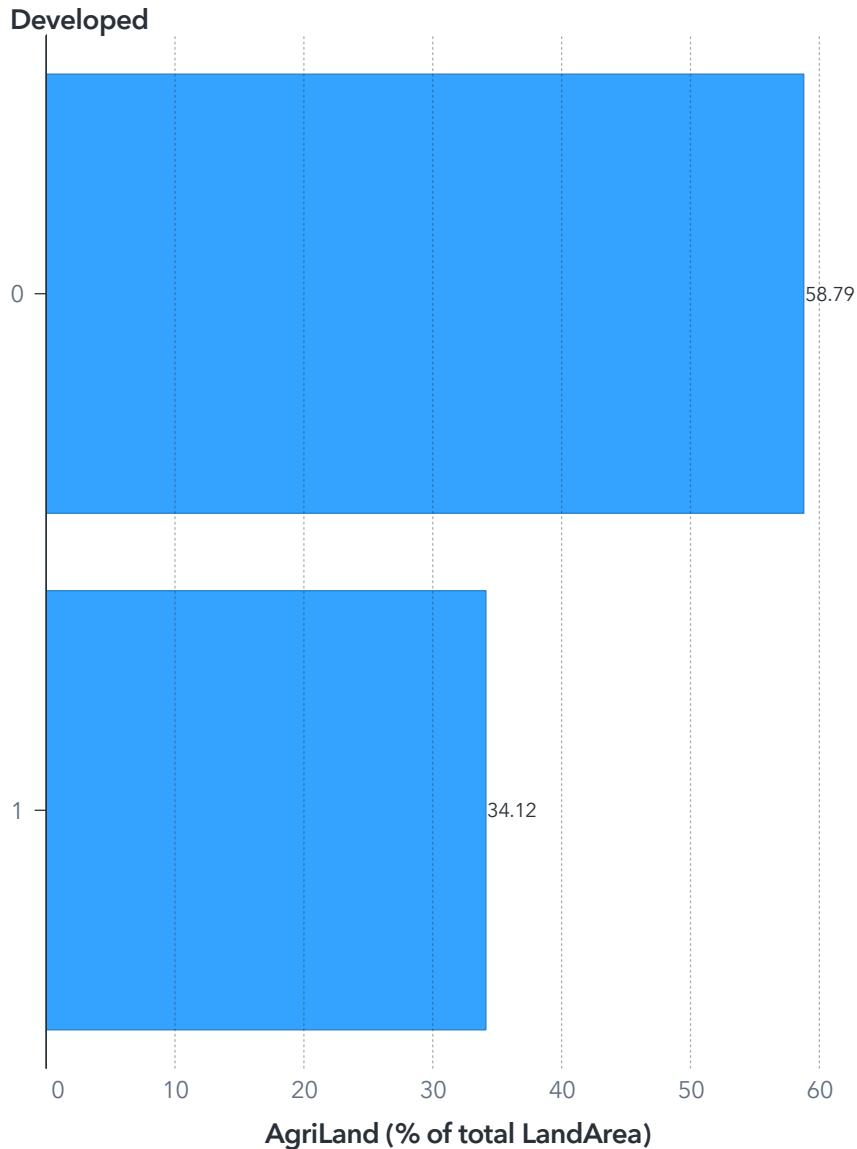
1972 1974 1976 1978 1980 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2018 2020  
1973 1975 1977 1979 1981 1983 1985 1987 1989 1991 1993 1995 1997 1999 2001 2003 2005 2007 2009 2011 2013 2015 2017 2019 2021

Year

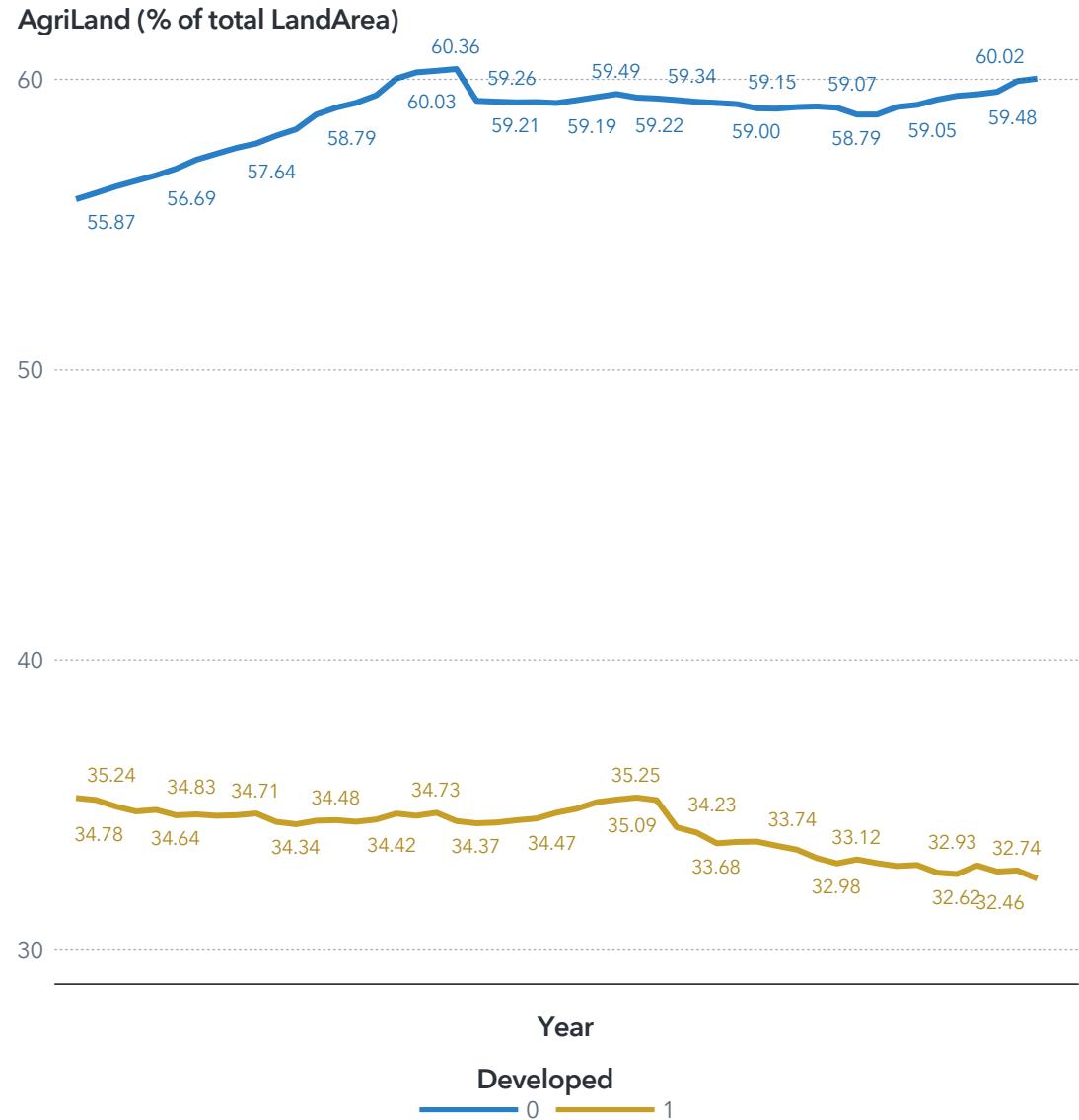
## Land Area



### Comparison of Agriland porportion to LandArea in developed and developing countries

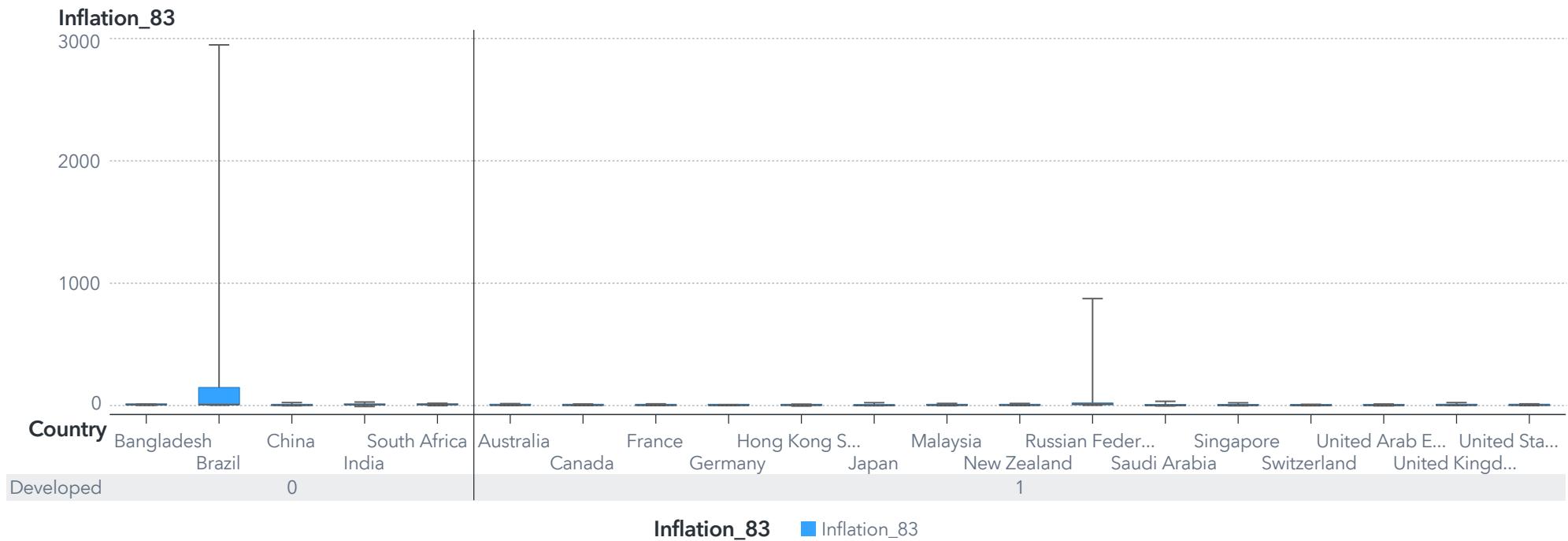


### Agriland porportion to LandArea across years in developed and developing countries



## Inflation

### Inflation box plot by country's development



### Inflation by year and country

Year ▲	Total	1972	1973	1974	1975	1976	1977	1978
Country ▲	Inflation_83 ▼	Inflation_83						
Total	19.241970885	5.2735714286	10.57	15.693571429	12.191428571	8.6271428571	8.4642857143	5.9885714286
Brazil	286.87463415	.	.	.	.	.	.	.
Russian Federation	61.16137931	.	.	.	.	.	.	.
South Africa	8.9246	6.43	9.43	11.72	13.43	11.02	11.15	11.14
India	7.7952	6.44	16.94	28.6	5.75	-7.63	8.31	2.52

# GDP exploratory

1972 to 2021

1972

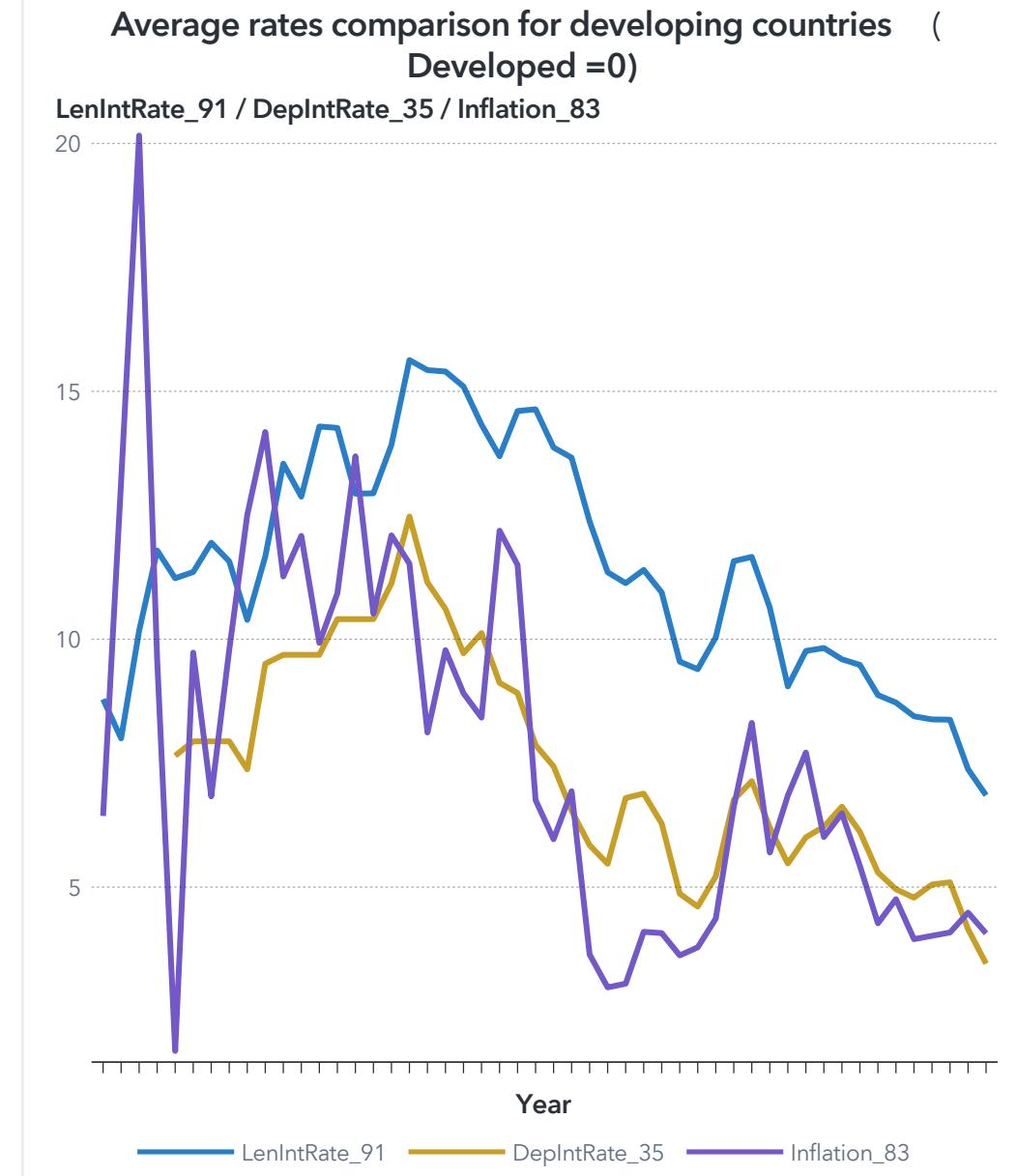
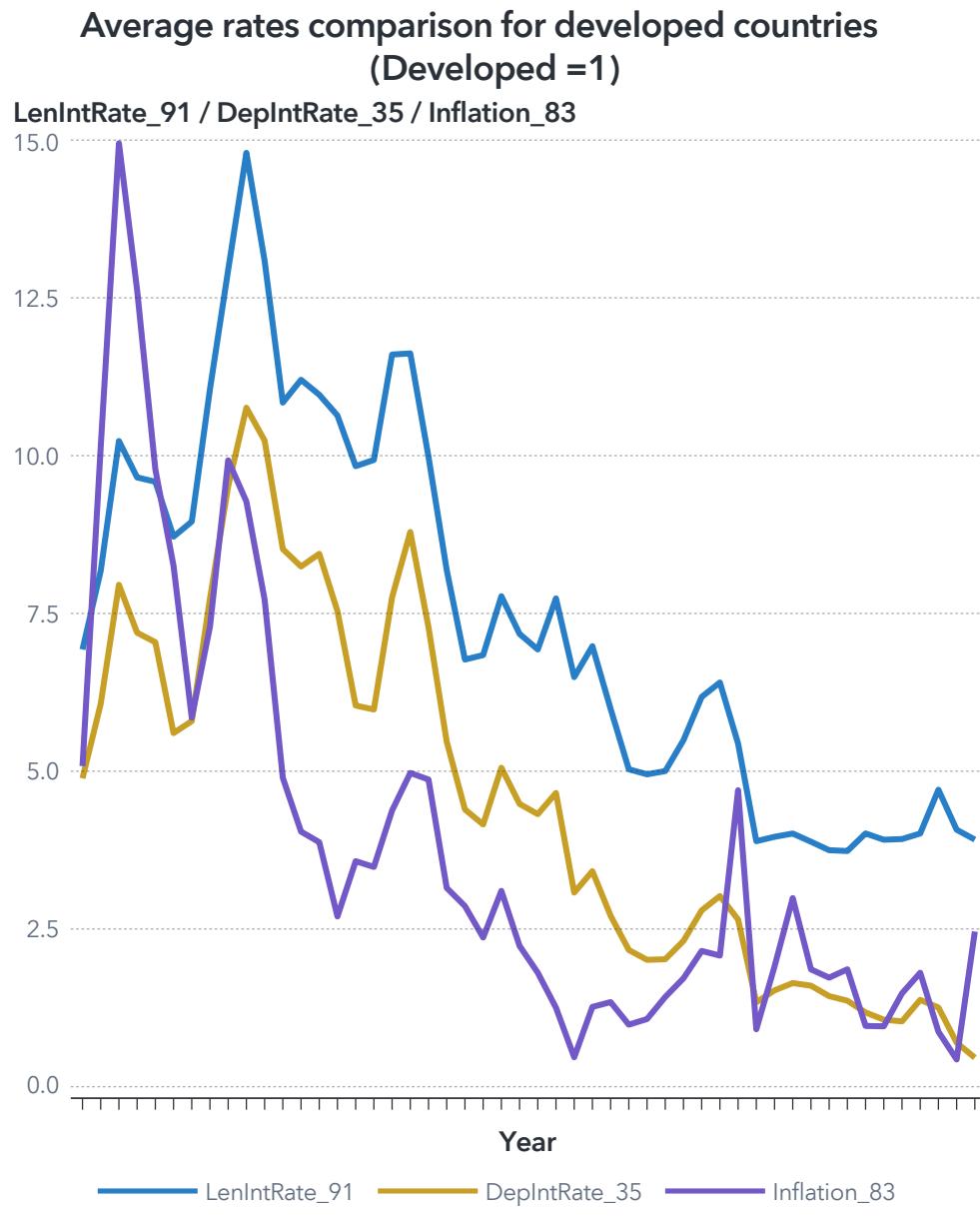
2021

## Comparing characteristics of Global North/South and Developed/Developing

Developed	▲	Global North/South	▲	Country	▲	GDP per Capita	▼	Income per Capita	GDP_65	NatIncome_5
Total						\$19,165		\$14,847	1.4T	1.2T
0	Global South	Brazil				\$4,750		\$4,050	877B	743B
		South Africa				\$4,274		\$3,343	198B	155B
		China				\$2,567		\$1,819	3.5T	2.4T
		India				\$723		\$604	854B	704B
		Bangladesh				\$611		\$537	87B	75B
	Global North	Switzerland				\$53,395		\$52,512	409B	411B
		United States				\$32,984		\$27,349	9.6T	8T
		Japan				\$28,211		\$21,899	3.5T	2.8T
		Australia				\$27,923		\$21,366	592B	448B
		Canada				\$26,773		\$21,227	865B	681B
1	Global North	Germany				\$26,665		\$21,939	2.2T	1.8T
		Singapore				\$26,146		\$20,310	124B	95B
		United Kingdom				\$25,604		\$20,881	1.6T	1.3T
		France				\$24,923		\$20,776	1.6T	1.3T
		New Zealand				\$20,352		\$15,883	86B	66B
	Global South	Russian Federation				\$6,905		\$5,461	1T	786B
		United Arab Emirates				\$33,702		\$34,271	154B	253B

聽 A1.1

## Average rates comparison



## Comparing GDP sources in developed and developing countries

ServicesAdd\_130



55

50

16

17

18

MfgValGDP\_96

TaxGDP%\_135

16.023773585  
13.597446809

Developed

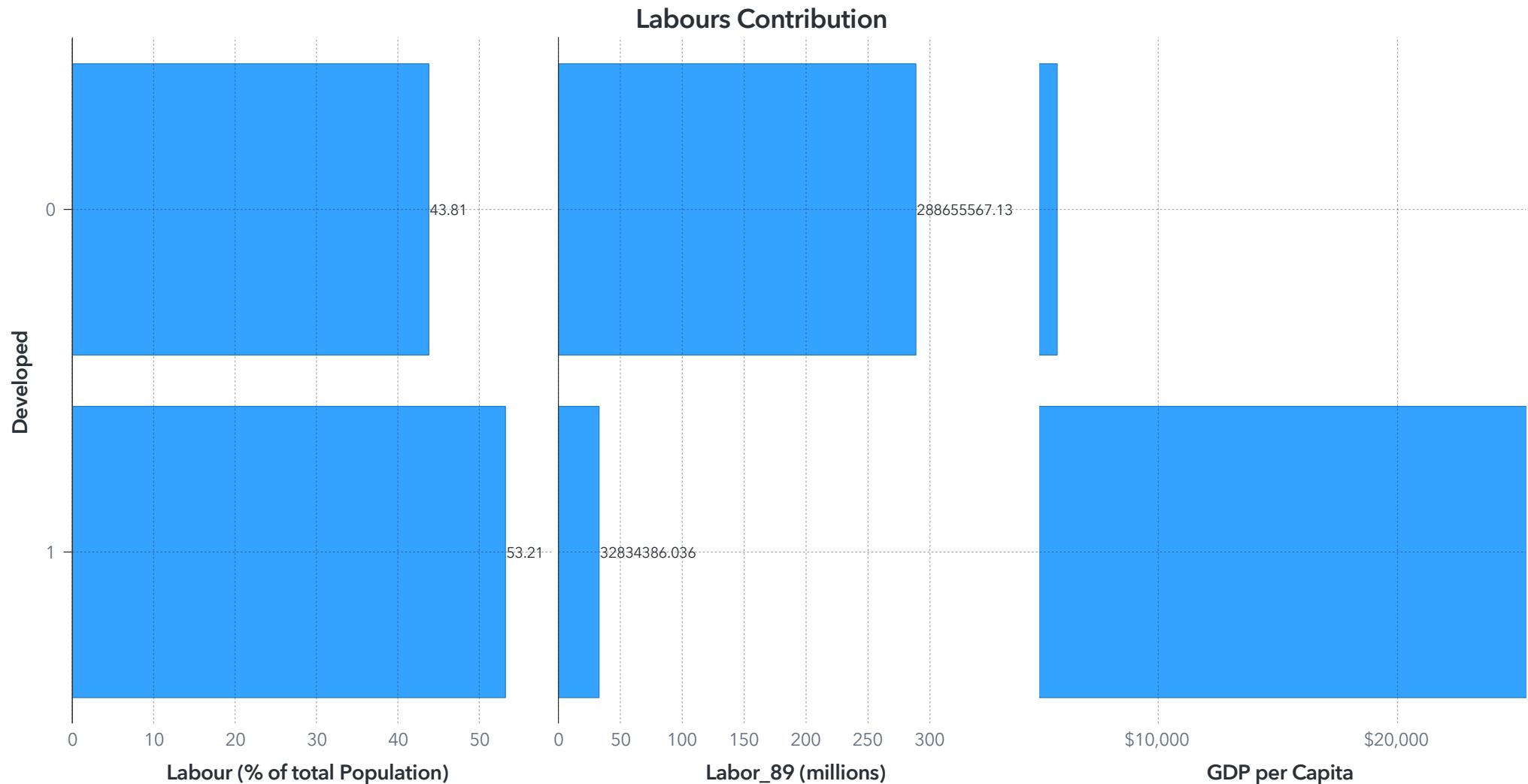
0 1

## Labour contribution

2012 to 2021

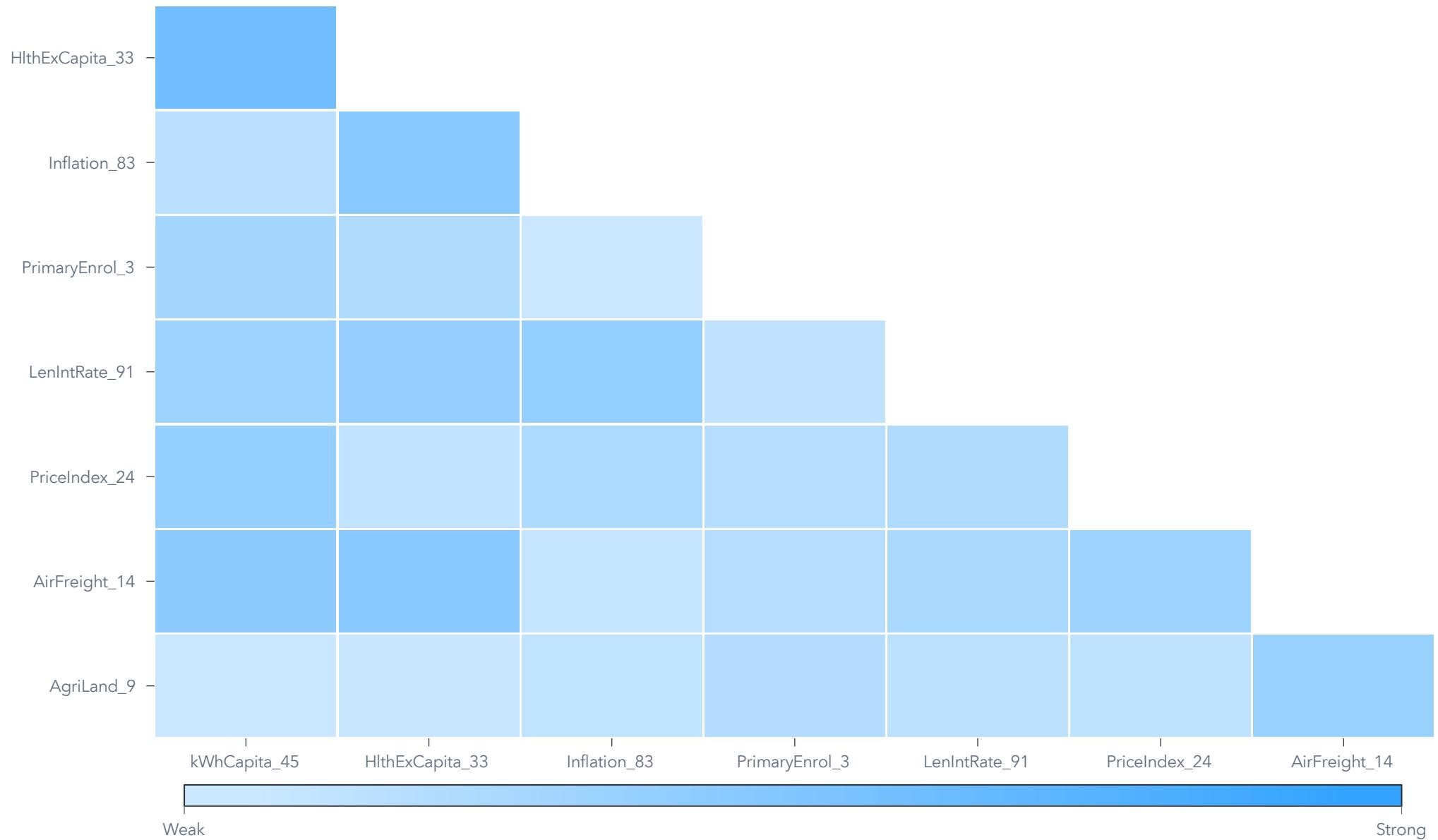
1972

2021



## Correlation for Multilinear Regression

**Correlation matrix (8 variables)**



# Multilinear Regression

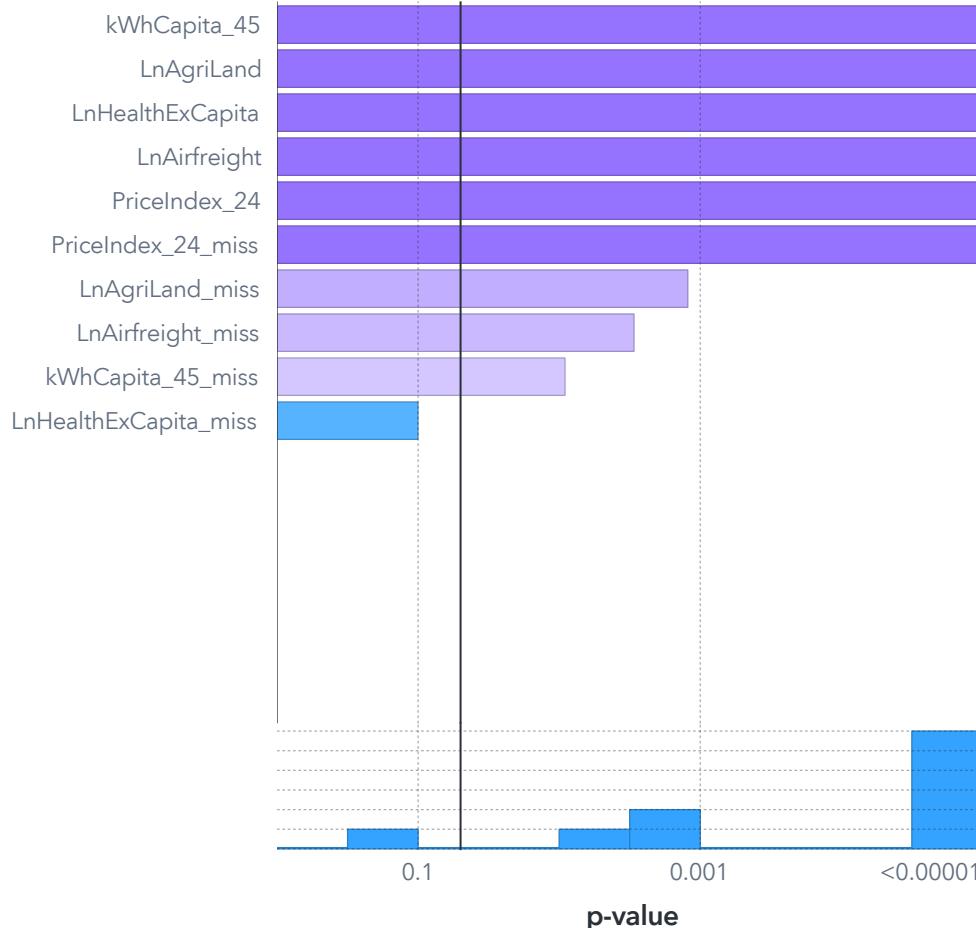
Developed:

0

1

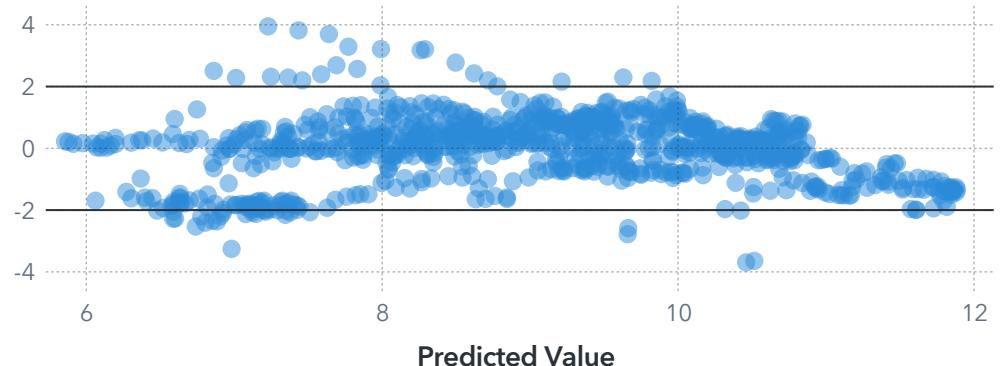
Linear Regression **LnGDPpercapita** Adjusted R-Square **0.7657** Observations Used **972** Unused **28**

## Fit Summary



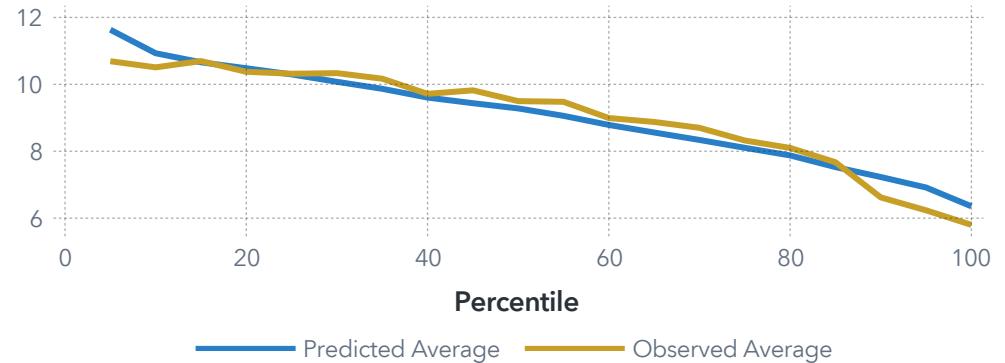
## Residual Plot

### Studentized Deleted Residual



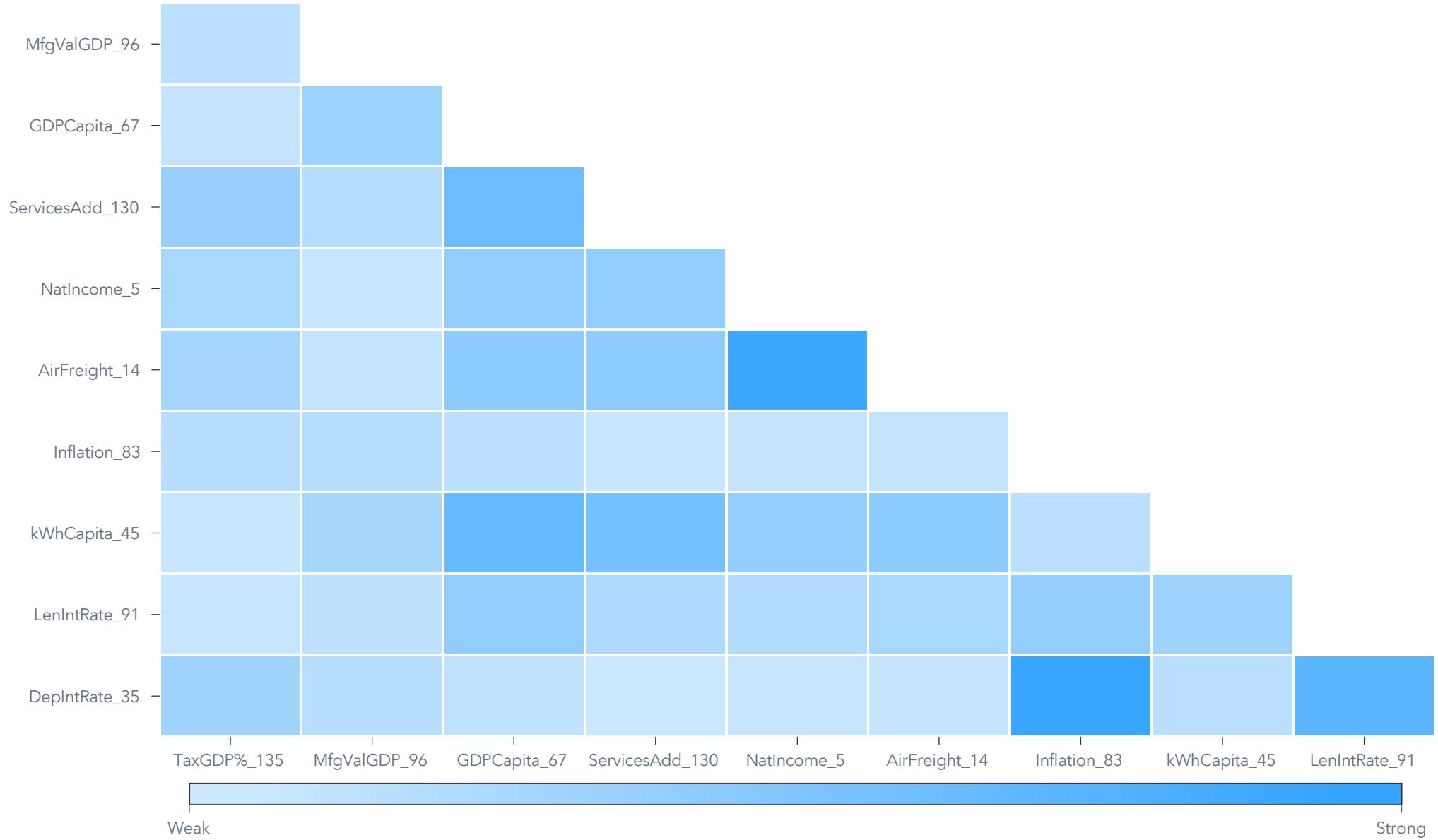
## Assessment

### LnGDPpercapita



## Correlation matrix for Log Refression

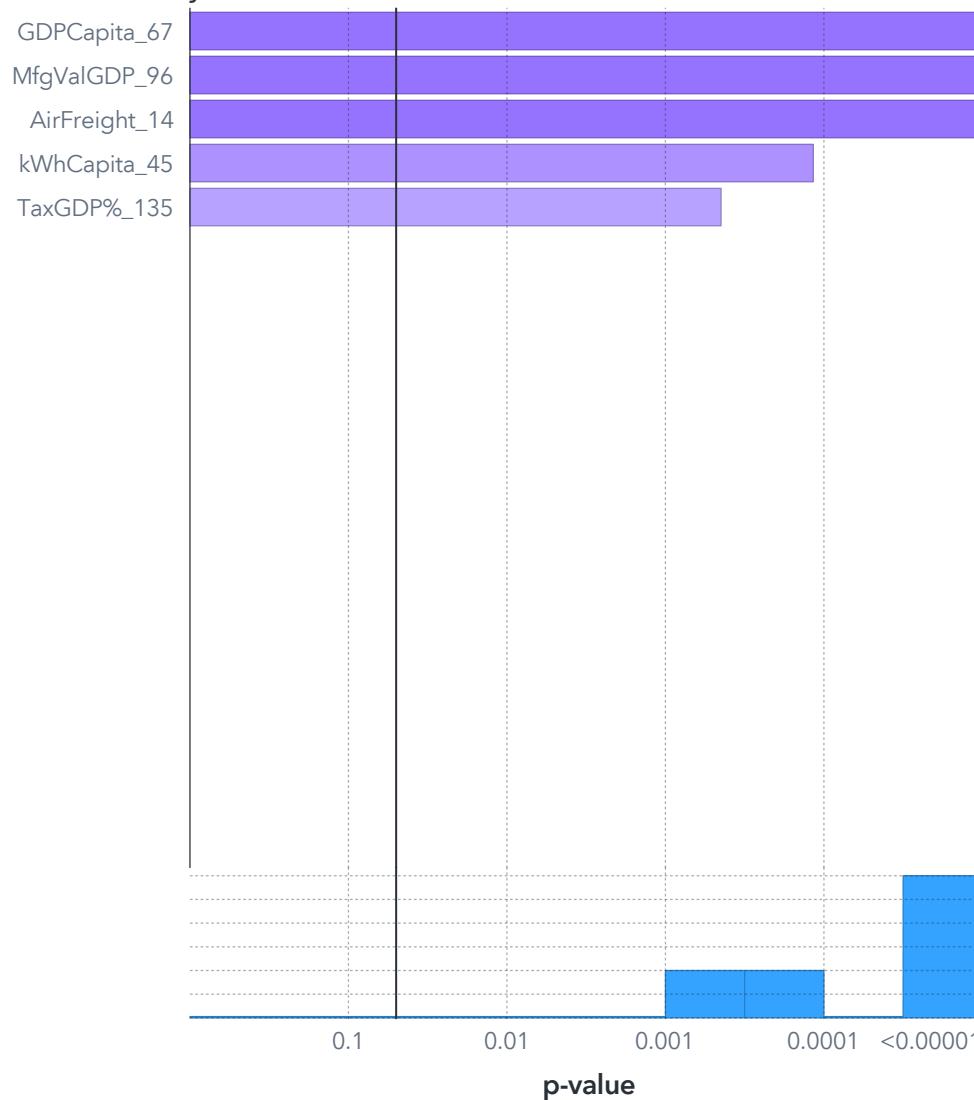
Untitled



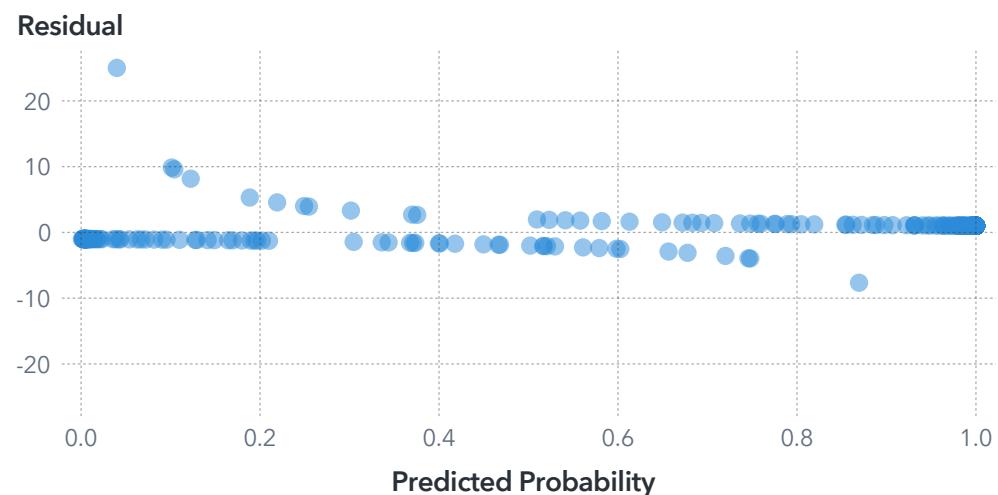
# Log Regression

Logistic Regression   **Developed** (event=1)   KS (Youden)   **0.9101**   Observations Used   **409**   Unused   **591**

## Fit Summary

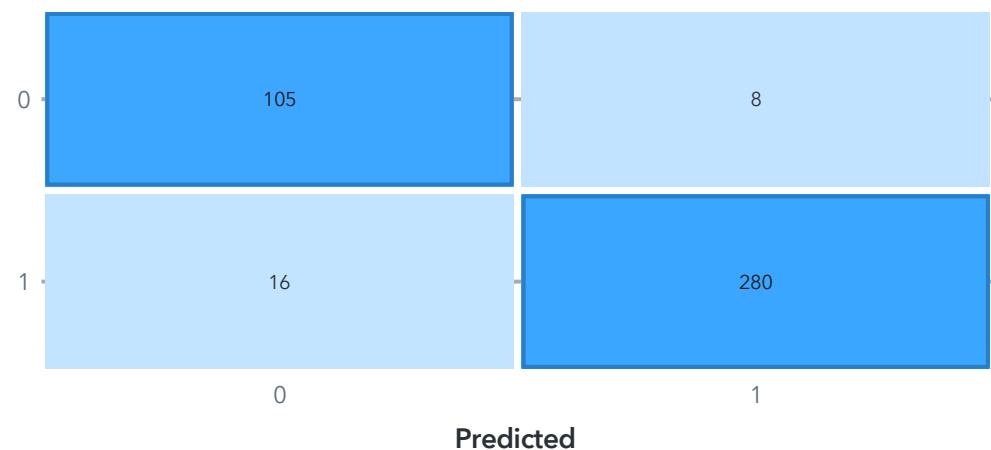


## Residual Plot



## Confusion Matrix

### Observed



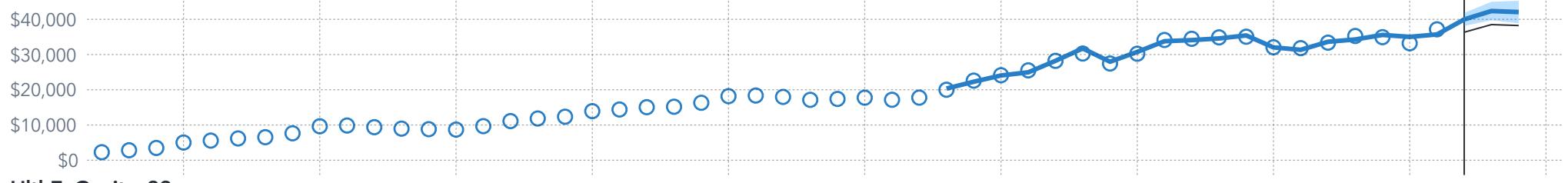
## Forecasting

Developed:

0

1

GDP per Capita



HlthExCapita\_33



Inflation\_83



○ Actual — Model

95% forecast confidence.

The forecast for GDP per Capita has the following contributing factor(s): HlthExCapita\_33, Inflation\_83

## Appendix

### A1.1 Comparing characteristics of Global North/South and Developed/Developing

---

Filters: Year BetweenInclusive(1972; 2021)

### A2.1 Average rates comparison for developed countries (Developed =1)

---

Filters: ( Country In('Australia'; 'Bangladesh'; 'Canada'; 'China'; 'France'; 'Germany'; 'Hong Kong SAR, China'; 'India'; 'Japan'; 'Malaysia'; 'New Zealand'; 'Saudi Arabia'; 'Singapore'; 'South Africa'; 'Switzerland'; 'United Arab Emirates'; 'United Kingdom'; 'United States') ) OR Country Missing  
( Developed In(1) ) OR Developed Missing

### A2.2 Average rates comparison for developing countries (Developed =0)

---

Filters: ( Country In('Australia'; 'Bangladesh'; 'Canada'; 'China'; 'France'; 'Germany'; 'Hong Kong SAR, China'; 'India'; 'Japan'; 'Malaysia'; 'New Zealand'; 'Saudi Arabia'; 'Singapore'; 'South Africa'; 'Switzerland'; 'United Arab Emirates'; 'United Kingdom'; 'United States') ) OR Country Missing  
( Developed In(0) ) OR Developed Missing

### A3.1 Labours Contribution

---

Filters: ( Year BetweenInclusive(2011; 2021) ) OR Year Missing