## Effect of economic growth on health

# Introduction

Health and economy are closely interconnected, and numerous studies have established a strong relationship between the two sectors. Although the relationship seems complex, the countries with higher GDP per capita tend to exhibit better health outcomes. Economic growth, GDP, investment, and employment contribute to the performance and quality of the economic system. Also, understanding the dynamics between economic productivity, economic growth, and health outcomes is crucial for policymakers.

## Background information:

In this report, GDP per capita is used as an indicator of economic development and infant mortality rate as a health indicator.

The infant mortality rate is one of the closely related health indicators with the economic conditions. Infant Mortality is the death of an infant before turning one year old. Infant mortality rate measures the number of deaths among infants under one year of age per 1,000 live births. It provides valuable insights into the overall health and well-being of a population, as they demonstrate the efficiency of healthcare systems, accessibility to maternity care, dietary habits, environmental conditions, and other socioeconomic factors.

In the past, nations with higher GDP per capita have been able to devote more funds to public health initiatives, educational initiatives, and hospital facilities, leading to better health results for their populations. These nations frequently offer more accessibility to healthcare resources, higher-quality medical care, and cutting-edge medical technology for disease prevention, diagnosis, and treatment. While improvements in health outcomes are often a result of economic development, there can be differences and exceptions depending on the circumstances of a given country and historical events.

# Research objectives:

## The goals of the research and the related questions are outlined in this section.

# 1. Investigating whether there are geographical variations in the relationship between GDP per capita and infant mortality.

# 2. Spotting breaks in the previously mentioned trend and relating them to events in history.

# 3. Investigating different health indicators.

# Data Identification:

The World Bank data consists of demographic and other statistical data related to Population, Employment, Health, GDP, Energy Consumption, etc. for most of the countries from the year 1960 to 2022. This allows us to look at variation between countries and over time. The categories are called indicators and are each defined by a code.

The following indicators have been chosen for analysis:

1. NY.GDP.PCAP.CD - GDP per capita
2. SP.DYN.IMRT.IN - Infant mortality rate
3. SP.DYN.LE00.IN - Life expectancy at birth, total (years
4. SP.DYN.CBRT.IN - Birth rate, crude (per 1,000 people
5. SP.DYN.CDRT.IN - Death rate crude (per 1,000 people)
6. NY.GDP.PCAP.CD - GDP per capita
7. SP.DYN.IMRT.IN - Infant mortality rate

# Exploratory Data Analysis

## Dimensions:

In this analysis, there are two different data-frames. The first data frame consists of GDP per capita, Infant mortality rate, date, and country as the columns. It has 13156 rows which constitutes data of 266 countries over the time from 1960 to 2022.

For the second data-frame, the length of each column is different. Life expectancy, birth rate, death rate are the columns considered for the same number of countries along with infant mortality against GDP per capita.

# Descriptive analysis

Here are the definitions of the variables:

1. Life expectancy is the average number of years that a person is expected to live.
2. Death rate is the number of deaths per 1,000 people in a population in a year.
3. Birth rate is the number of births per 1,000 people in a population in a year.
4. Infant mortality rate is the number of deaths of infants under the age of 1 per 1,000 live births.
5. GDP per capita is the total GDP of a country divided by the population of that country.

The descriptive analysis of the data reveals valuable insights into the economic conditions across the countries and over time. The bar graphs depict the descriptive statistics of each input (Life expectancy, Death rate, Birth rate, Infant Mortality rate) and target variable (GDP per capita)

The analysis reveals interesting patterns in the data. The graph illustrating life expectancy demonstrates relatively low variance making it more consistent. Also, the quartile values of this variable are very close to each other, suggesting most countries might have similar life expectancy. This analysis is to be further analyzed.

Although, the range of values vary proportionally for birth rate, the death rate, GDP, Infant mortality rate accommodate wide range of values, the maximum value is much greater than the minimum value. It is significant to take note of the existence of outliers with extraordinarily high or low values, which may influence particular models. It is vital to remember that these outliers represent the reality of the nations that succeed or fail in these metrics. Keeping these outliers can offer important insights on the differences across nations.

Another important observation is that the distribution of the quartiles proposes that the adult death rate and Infant mortality rate are similar and are very close to that of GDP quartile distribution. These hypotheses need to be further analyzed by statistical models to conclude on the effect of these input variables on the GDP, the indicator of economy.

A picture containing text, screenshot, diagram, plot

Description automatically generated

A picture containing text, screenshot, diagram, plot

Description automatically generatedA picture containing text, screenshot, diagram, plot

Description automatically generated

A picture containing text, screenshot, diagram, plot

Description automatically generated A picture containing text, screenshot, diagram, plot

Description automatically generated

## Null Values:

The GDP per capita and Infant mortality rate columns has 3602 and 4162 NaNs respectively. Life expectancy, birth rate, death rate are the columns considered have 892, 721, 739, 3602 and 4162.

A colorful pie chart with white text

Description automatically generated with medium confidence

## Handling Missing Values

The missing values need to be handled before passing the data through any model to avoid incorrectness of the model results. Some models do not accept the NaNs. Data Imputation is performed to handle missing values because in our case the missing values are missing at random (MAR). Interpolation is the process of calculating the values of missing data by considering the values of nearby observations or data points. Interpolation techniques can infer plausible hypotheses about the missing values by making use of the currently available data points and their relationships.

One of the popular techniques for filling missing values in a dataset is linear interpolation. This strategy is simple to use and appropriate when the data exhibits a distinct trend or linear pattern. Therefore, the assumption here is that there is a linear relationship between adjacent data points. With linear interpolation, the missing values are imputed with values that are along a straight line connecting the observed data points adjacent to them.

# Analysis

## Objective 1

The first objective of the analysis is to investigate if there is any difference in the relationship between GDP per capita and infant mortality rate over the period of 1960-2022 for 266 different countries based on the availability of the information.

### Linearity Check

The scatter plot of Infant mortality rate vs GDP per capita indicates exponential spread. This can be seen in the plot below.

A picture containing text, screenshot, diagram, line

Description automatically generated

### Transformation of data

Since the spread looks exponential, it is only wise to check if the log transformation of the data brings out the linearity amongst the input and target variable.

A diagram of a scatter plot

Description automatically generated with low confidence

### Normality Check

The variables do not seem to have normality before transformation. Since the transformed data is linear, the histogram was drawn for the log transformed data and found that the normal distribution holds good. There is a slight left skewness observed from the histogram of Infant mortality rate. Most of the regression models expect normality of data.

A picture containing text, screenshot, display, diagram

Description automatically generated A picture containing diagram, screenshot, plot, text

Description automatically generated

A graph of infant mortality rate

Description automatically generated with low confidence A picture containing diagram, screenshot, text, plot

Description automatically generated

### Mixed effects model

The data needs to be treated as grouped, based on the ‘country’ to compare the kind of relationship. When there is a hierarchical structure to the data, such as when data is nested within groups, mixed-effects models are frequently used. The statistics are nested inside countries in case of both infant mortality rate and GDP per capita. In this situation, a mixed-effects model would take into account the potential that there are regional differences in the link between GDP per capita and infant mortality rate.

y = Xβ + Zu + ε

where:

y is the response variable

X is the design matrix for the fixed effects

β is the vector of fixed effects

Z is the design matrix for the random effects

u is the vector of random effects

ε is the error term

### Hypothesis testing

It is clear from the model results that the p-values for different countries and different years are different. Further this hypothesis is tested. Generally, the relationship between GDP per capita and infant mortality can vary between different countries. This is because there are several factors that can contribute to infant mortality, such as the quality of healthcare, access to healthcare, and the distribution of wealth. The analysis is performed to draw conclusions about the kind of relationship.

In this report, the Wald test is conducted on the data for testing the hypothesis. The Wald test is a statistical test that is used to test the hypotheses about parameters that have been estimated by maximum likelihood as mixed effect linear models use the same approach. This test is used to check whether a parameter is equal to a hypothesized value called null hypothesis. The null hypothesis for the Wald test is that the parameter is equal to the hypothesized value, and the alternative hypothesis is that the parameter is not equal to the hypothesized value.

The Wald test statistic is calculated as follows:

W = (ˆθ - θ₀) / SE(ˆθ)

where:

ˆθ is the estimated value of the parameter.

θ₀ is the hypothesized value of the parameter.

SE(ˆθ) is the standard error of the estimated value of the parameter.

|  |  |  |
| --- | --- | --- |
|  | chi2 | p>chi2 |
| date | [148.2587592162855] | 3.1757476668253152e-09 |
| C(country) | [2708.4714920357915] | 0.0 |
| log\_GDP\_per\_capita:date | [1195.0735434270662] | 5.4261618293174605e-210 |

The null hypothesis is that the relationship between GDP per capita and infant mortality rate is stable among different countries over time. This means that the relationship between these two variables is the same for given 266 countries, regardless of the year.

The alternate hypothesis is that the relationship between GDP per capita and infant mortality rate is different among different countries over time. This indicates that not all countries have the same relationship between these two factors, and that relationship may evolve over time.

The p-values of the test are lower than 0.05 which indicates that the alternate hypothesis is supported and the null hypothesis is rejected. This directly means that the relationship between GDP per capita and infant mortality rate is different among different countries over time. There are a few possible justifications for this.

One possibility is that different countries have different healthcare systems, infrastructure, policies and investments. Countries with better healthcare systems might be able to reduce infant mortality rates even in countries with low GDP per capita.

Furthermore, it's likely that there is a nonlinear relationship(mostly exponential) between the infant mortality rate and GDP per capita. This suggests that the relationship between these two factors may alter when GDP per capita increases.

The findings of the Wald test point to the complexity of the relationship between infant mortality and GDP per capita. The log\_GDP\_per\_capita:date variable, nation, and date are all significant predictors of infant mortality rate, according to the Wald test results. This relationship can be influenced by a few factors, and it may change over time depending on the country. Understanding these elements can help us create policies and initiatives that will lower infant mortality.

## Objective 2

The second objective of the analysis is to investigate if there are any break in the trend of relationship between GDP per capita and infant mortality rate over the period of 1960-2022.

Using the mixed effect model results obtained from the previous analysis, following observations are made. The trends are concluded based on the coefficients of each year in the model results.

1. The period from 1961-1964 seem to indicate a negative trend line with the GDP per capita. There could be many explanations to this. One of the popular historical event being, cold war(1961-1964) which happened between the United states and soviet union. The intense friction between the countries indicated huge spending on military and research which had its negative impacts on the economy and infrastructure. The Cold War had far-reaching effects on global economy. Them being the huge contributors to the world economy(GDP per capita), there is a possibility that the model results could be biased because the countries in the western Europe had a sustained growth.
2. The year 1965 saw a significant inflation pressure in the United States.
3. Starting in March 2000, the dot-com bubble began to deflate as investors started to realize the unsustainable nature of the valuations. Many internet companies faced financial difficulties and saw their stock prices decline rapidly. Numerous dot-com companies filed for bankruptcy or closed altogether. We can see decrease in the coefficient value from late 1990s to early 2000s.
4. In 2002, the global economy started to recover from the downturn experienced in the previous years. Many countries implemented fiscal and monetary policies to stimulate growth, leading to improved economic conditions. This was a significant global recovery.
5. The trendline declined continuously from 2019-2021, the reason could be the economic slowdown and the popular COVID-19 global pandemic outburst, Lockdowns, travel restrictions, and social distancing measures led to disruptions in supply chains, business closures, and a sharp decline in economic activity. This combined with the war situation between Ukraine and Russia impacted the world economy.

## Objective 3

In this objective, the idea is to explore at least three different health indicators that might have their effects on GDP per capita, the economic indicator. The considered indicators for the comparison are life expectancy, birth rate, death rate along with infant mortality rate. The goal is to compare the indicators to identify the extent of influence of each variable by modeling the data.

### Random Forest Algorithm

To model the data, we are using random forest algorithm because it does not assume if the input variables are collinear and are having normal distribution. The variables are assumed to be independent and do not have strong correlations with each other.

**Life Expectancy** - Life expectancy is a statistical measure of the average number of years a person is expected to live, typically based on factors such as their birth year, gender, and socioeconomic conditions.

**Birth Rate** - Birth rate refers to the number of live births per 1,000 individuals in each population over a specific time.

**Death Rate** - Death rate refers to the number of deaths per 1,000 individuals in each population over a specific time.

Based on the definitions we can infer that the features are not causal even though the correlation coefficients indicate high dependency. The correlation index of 0.8 between birth rate and life expectancy suggests a strong positive relationship between the two variables, but it doesn't necessarily indicate a causal relationship.

A picture containing text, screenshot, parallel, square

Description automatically generated

Random forest is an ensemble learning algorithm that consists of a large number of decision trees. Each decision tree is trained on a random subset of the training data, and the predictions of the individual trees are then combined to make a final prediction.

Random forest is chosen to perform this task because it is one of the robust and accurate algorithm which can provide easily interpretable outputs. There was no need to perform any transformations on the input variables.

Results:

Mean Squared Error: 90670960.06879058

R2 Score: 0.689924596906659

Feature Importance

0 Life expectancy at birth, total (years) 0.412439

3 Infant mortality rate 0.314949

1 Birth rate, crude (per 1,000 people) 0.143539

2 Death rate crude (per 1,000 people) 0.129073

The trained model produced the following variable importance corresponding to each input vector. he variable importance scores tell you how much each feature contributes to the accuracy of the model. The features with the highest feature importance scores are the most important features for predicting GDP per capita. The Life expectancy being the most influencing feature is followed by infant mortality rate. The influence of birth and death rates of the adults are the least influencing factors on GDP per capita.

A picture containing screenshot, text, line, diagram

Description automatically generated

In this case, life expectancy is the most important feature for predicting GDP per capita. This is because life expectancy is a measure of the overall health of a population. A country with a high life expectancy is likely to have a healthy population, which can lead to a higher GDP per capita.

The infant mortality rate is the second most important feature for predicting GDP per capita. This is because infant mortality is a measure of the health of newborns. A country with a high infant mortality rate is likely to have a less healthy population, which can lead to a lower GDP per capita.

The birth and death rates of adults are the least influencing factors for predicting GDP per capita. This is because the birth and death rates of adults are not as closely linked to the overall health of a population as the life expectancy and infant mortality rate.

Overall, the variable importance scores show that the life expectancy and infant mortality rate are the most important features for predicting GDP per capita. These features can be used to identify countries with a high potential for economic growth.

## Conclusion

The connection between health and economic growth of a country is perhaps bidirectional. On the one hand, better health infrastructure boosts the economy by producing healthier people who can work. Healthy people are less prone to illness and thus work hard to help build a strong economy. A healthy mind and body give bandwidth to not only think and come up with new ideas for starting the business and making investments but to work towards it.

Perhaps on the other hand, the better the economy, the better the health infrastructure due to growth in investment of healthcare. The feasibility and accessibility of healthcare systems also increases that eventually pushes the country towards healthy living. Additionally, economic developments can influence other sectors like education and medical research which improves the country’s health. This dependency of these two sectors on one another leads to a cycle.

The model helped to derive the fact that this relationship can vary between countries and between specific time periods. The infant mortality rate being one of the important health indicators remains to be inversely proportional to the GDP per capita but directly proportional to time. We are trying to advance in both the field year by year but there are several historical factors that influenced this fact and broke the trend.

Also, through the models we were able to conclude that the death rate, birth rate, life expectancy, and infant mortality rate are all crucially related with GDP per capita. The life expectancy and infant mortality rate are the most important features for predicting GDP per capita. A country with a high life expectancy is likely to have a healthy population, which can lead to a higher GDP per capita.

Overall, the conclusions from both models show that the health of a population is a significant factor in determining GDP per capita. Countries with a healthy population are likely to have a higher GDP per capita. The relationship between health and economy is complex and bidirectional. However, both health and economy are important for individual well-being and for the overall success of a society.