# COVID-19 Pandemic Analysis

Presented by **Group #21**

Abbas Raza
Mishaal Usman
Rahimah Siddiqi
Syed Rayyan Jamal
Syed Rohaan Hussain

# Contents

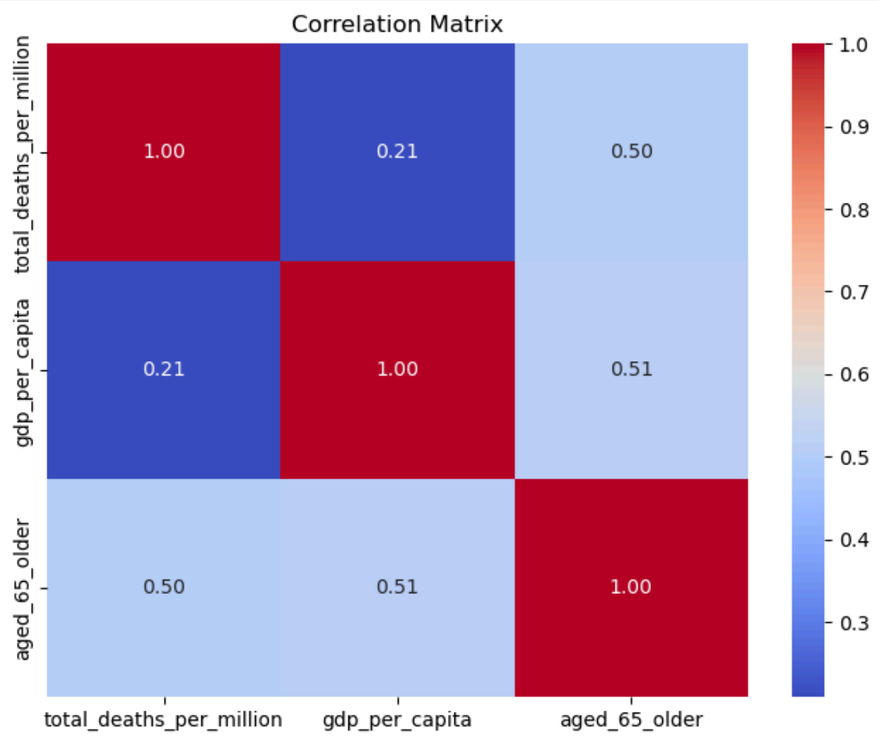
# Introduction

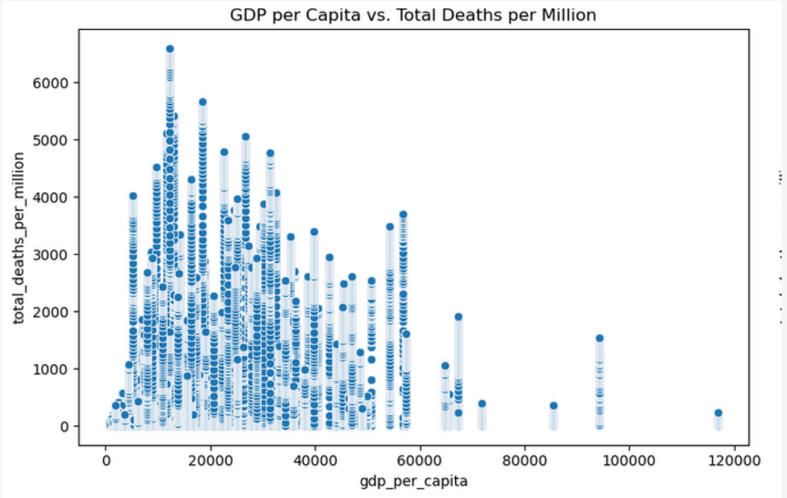
The COVID-19 Pandemic has been the most significant health crisis of the 21st century, with a considerable contrast existing in the intensity of the Pandemic's impact on different countries. Our dataset comprehensively tracks multiple economic and demographic indicators across all countries in the world on a daily basis from 2019 to 2025. We use this data to develop a machine learning model that predicts total deaths per million people based on:

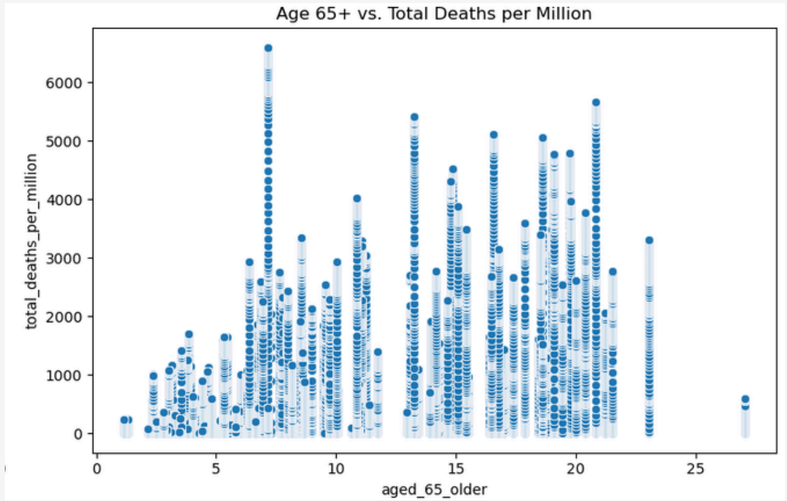
- GDP per capita
- Percentage of population aged 65 or above
- Population density
- HDI
- Hospital beds per thousand people
- Government response measures (stringency)

to answer the following research questions:

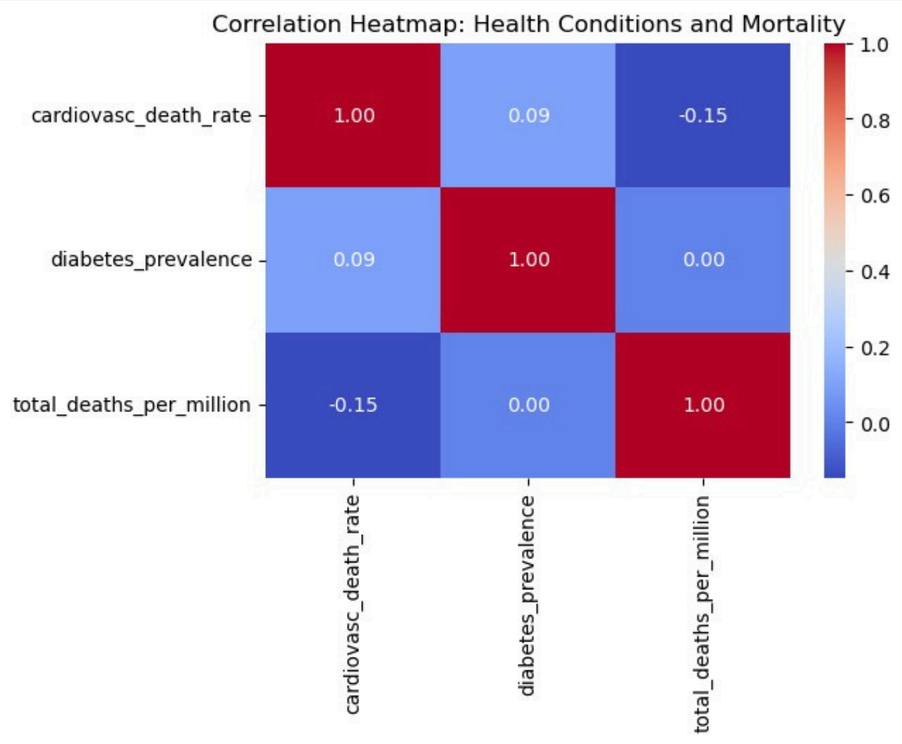
- Does higher HDI and GDP per capita lead to a lower number of cases and deaths?
- To what extent did pre-existing health conditions, such as cardiovascular disease and diabetes, contribute to a higher mortality rate among COVID-19 patients?
- How did demographic factors such as median age and proportion of elderly individuals impact mortality rates across countries?

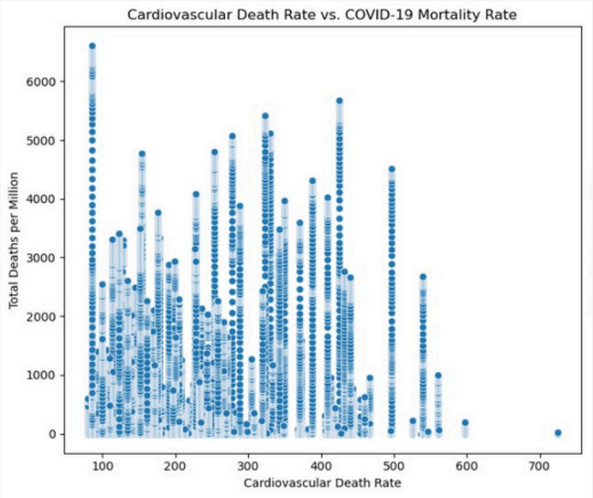


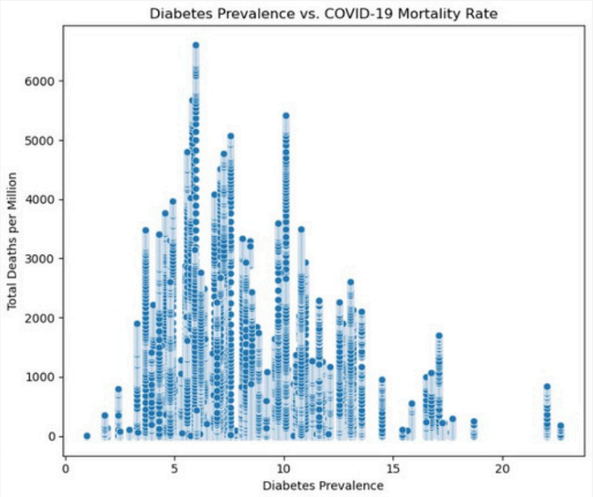




The weak positive correlation between gdp_per_capita and total_deaths_per_million (0.21) and moderate positive correlation between aged_65_older and total_deaths_per_million (0.50) provide strong support for our first and third research questions

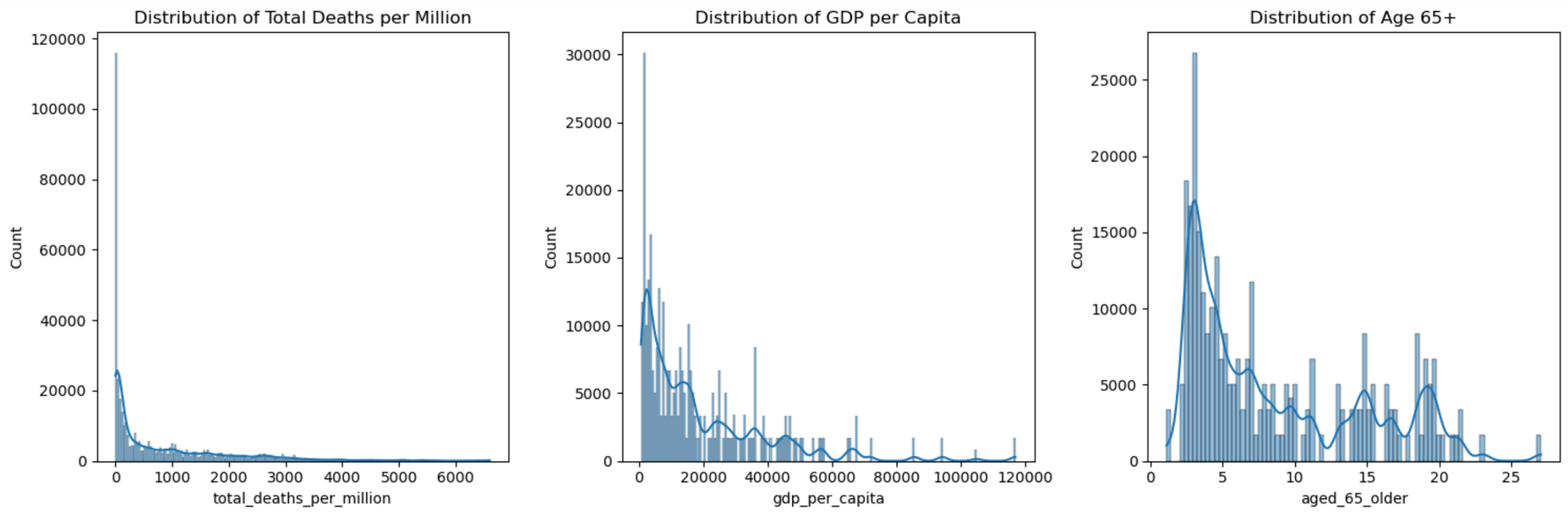


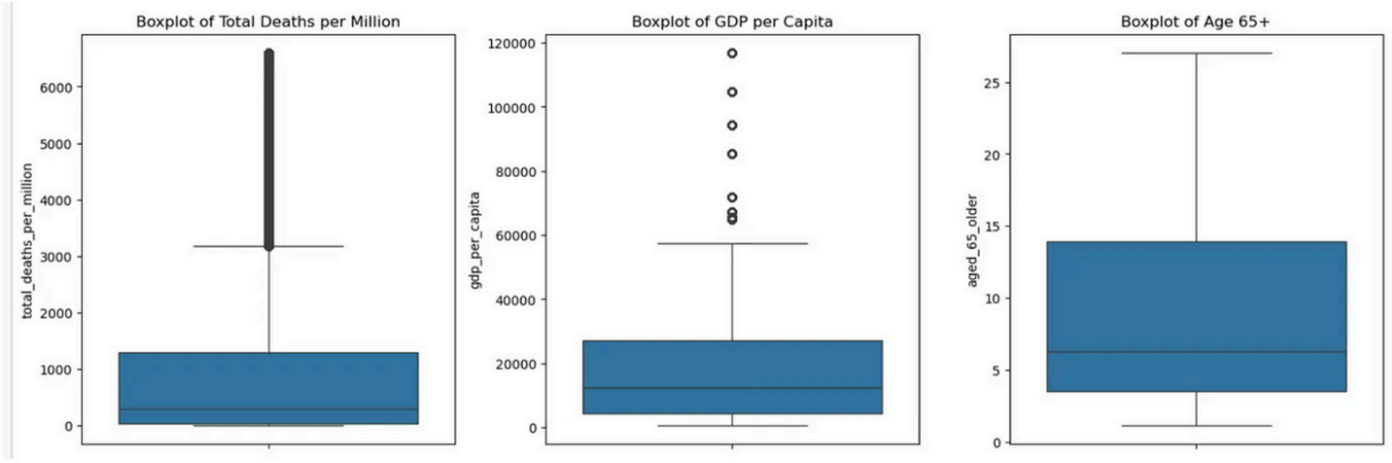



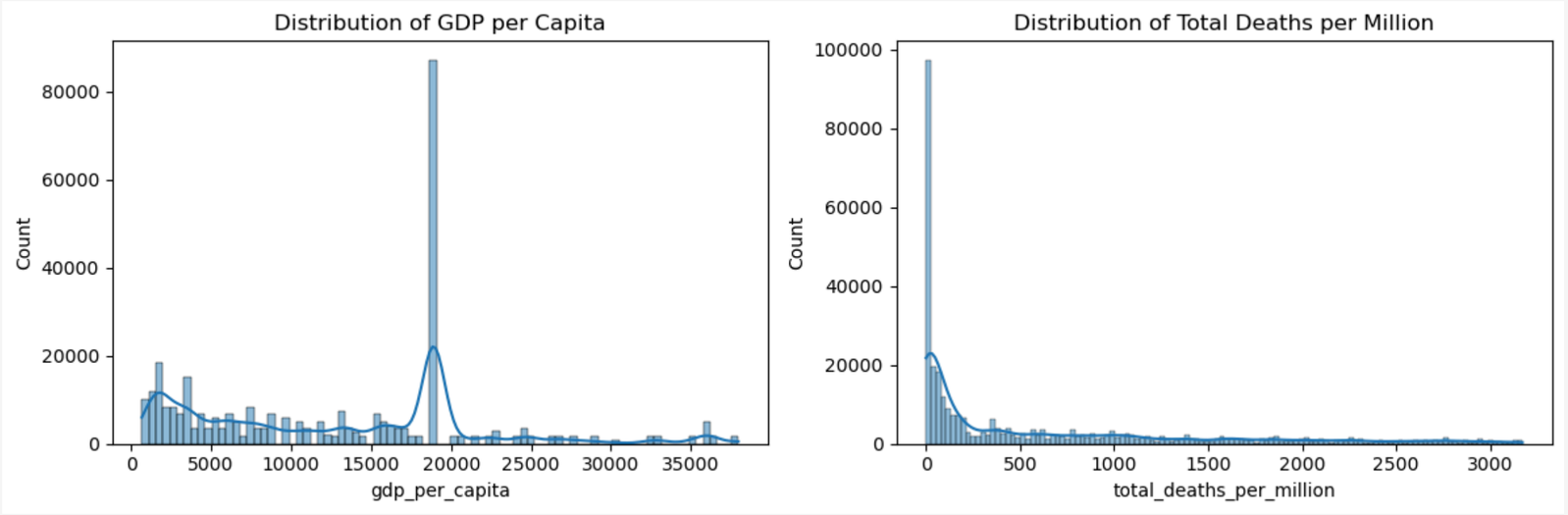

The absence of correlation between diabetes_prevalence and total_deaths_per_million (0.00) and negative correlation between cardiovasc_death_rate and total_deaths_per_million (-0.15), leading us to discard our second research questions

# Data Analysis

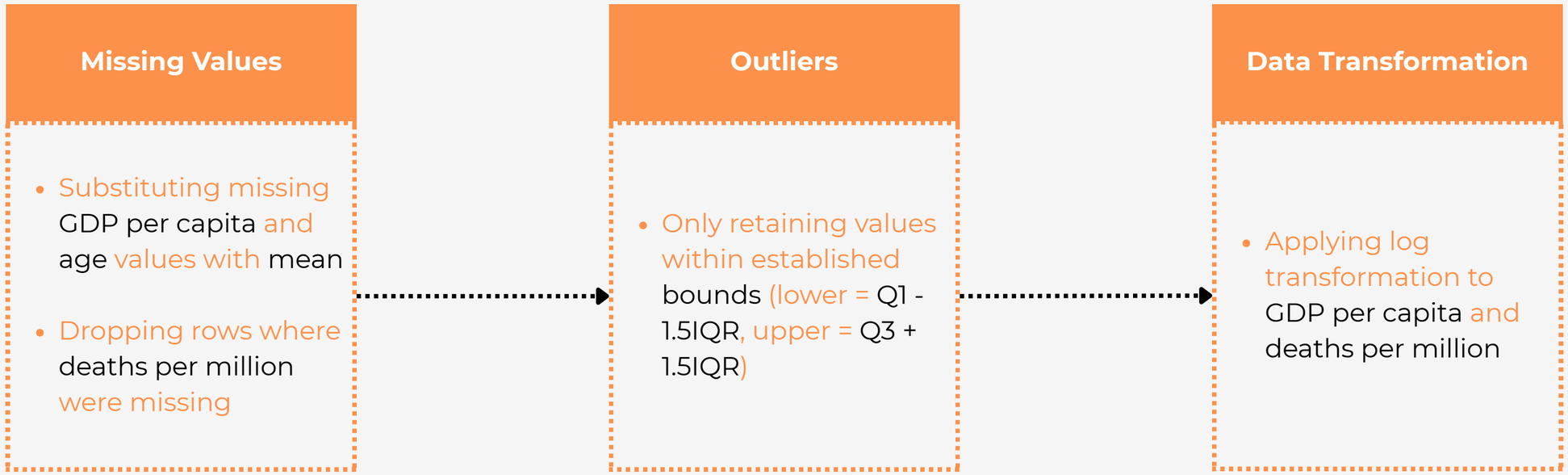



Total deaths per million: strong right skew, with most countries experiencing lower death rates
GDP per capita: right-skewed distribution, with most countries clustered at lower to mid-range values
Population aged 65 or above: multimodal distribution, with different demographic stages for different countries

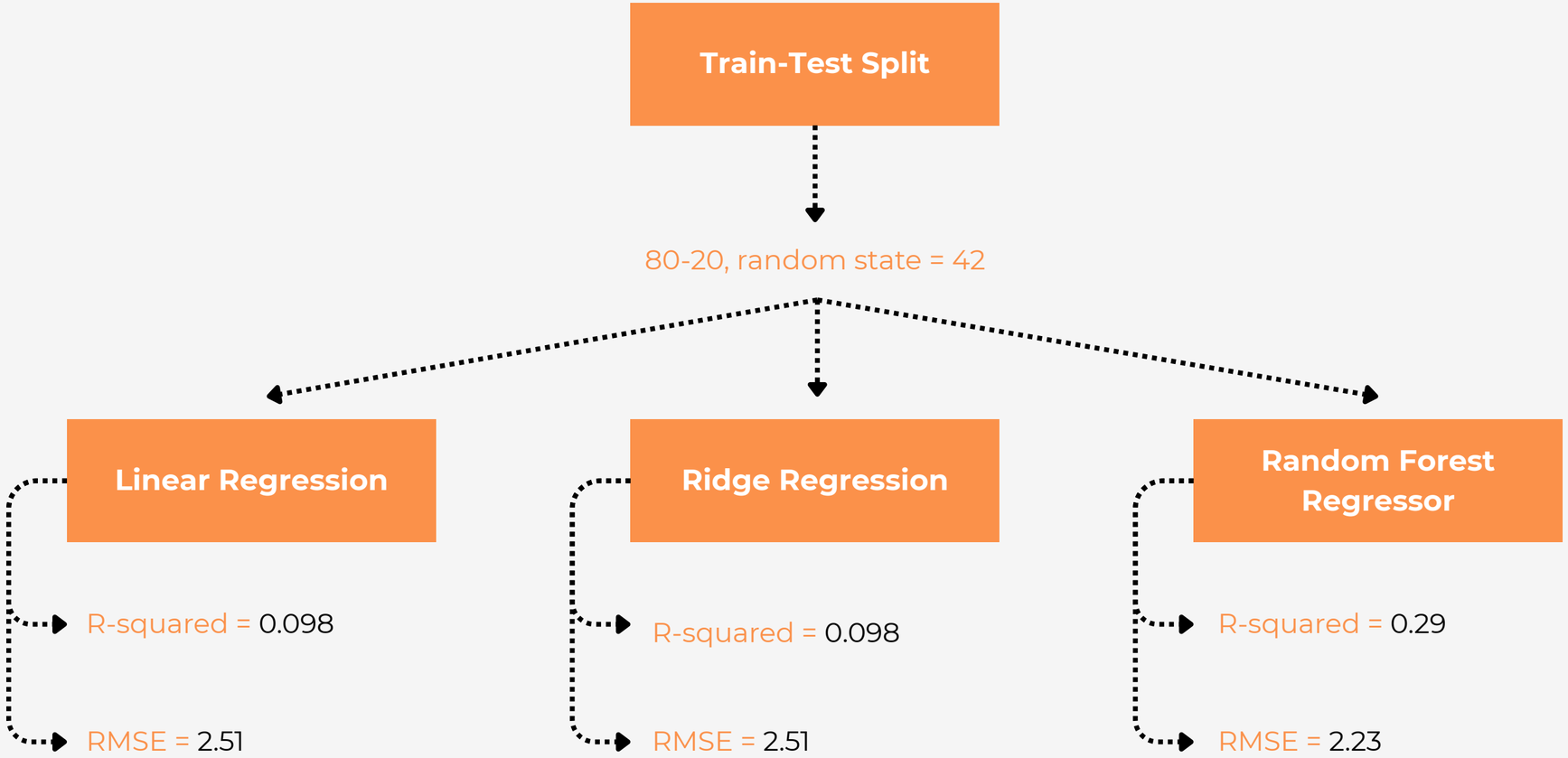

Boxplot of Total Deaths per Million
total_deaths_per_million
0
1000
2000
3000
4000
5000
6000
Boxplot of GDP per Capita
gdp_per_capita
0
20000
40000
60000
80000
100000
120000
Boxplot of Age 65+
aged_65_older
0
5
10
15
20
25


Distribution of GDP per Capita
Count
0
20000
40000
60000
80000
0
5000
10000
15000
20000
25000
30000
35000
gdp_per_capita
Distribution of Total Deaths per Million
Count
0
20000
40000
60000
80000
100000
0
500
1000
1500
2000
2500
3000
total_deaths_per_million

# Data Cleaning



# Defining Model

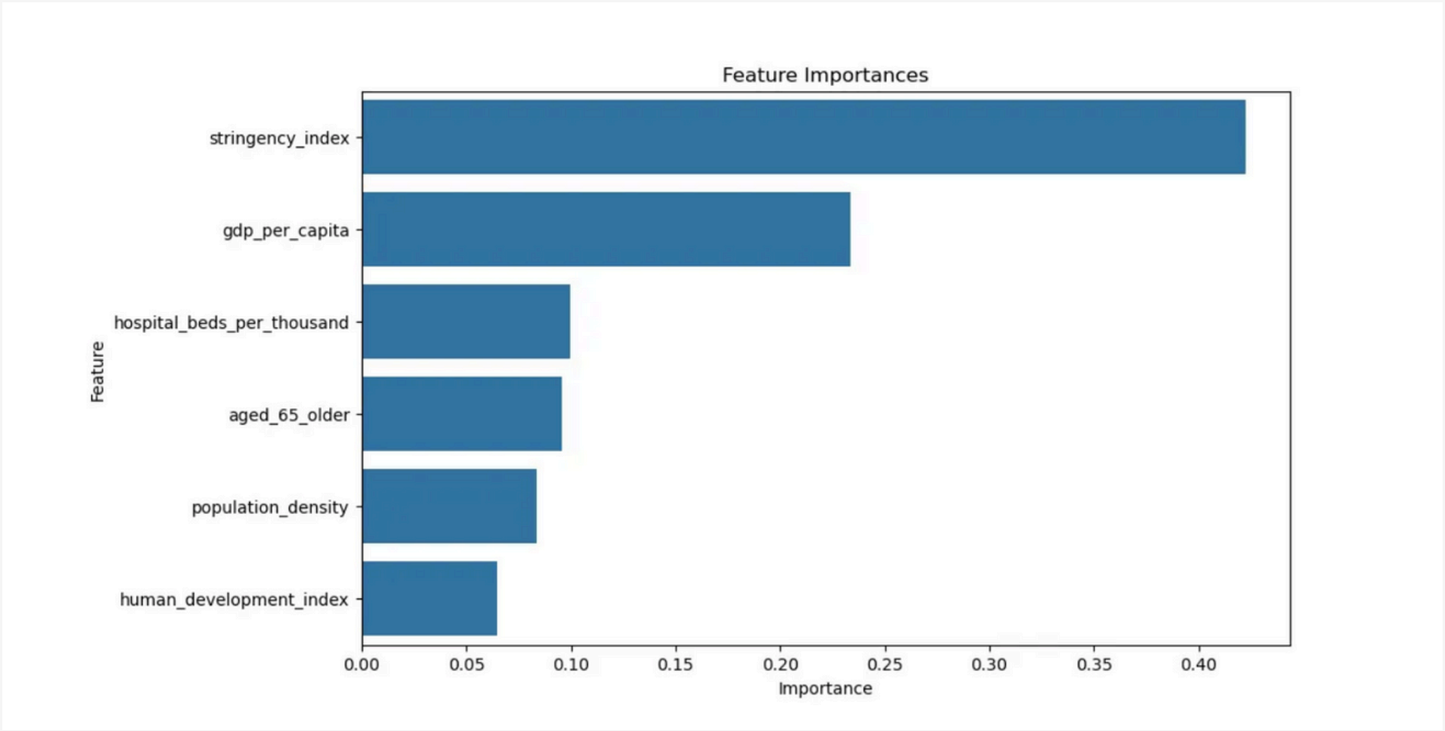



Best model due to lowest RSME

Including the following additional features to boost the robustness of the model:

- hospital_beds_per_thousand
- human_development_index
- population_density
- stringency_index

the target remains total_deaths_per_million as before




R-squared: 0.92

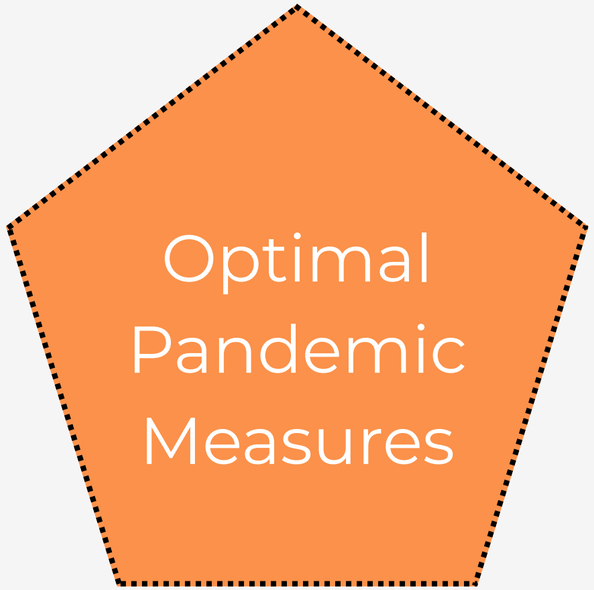
# Key Findings

| Rank | Feature | Importance |
|---|---|---|
| 1 | Stringency Index | 0.422433 (42.2%) |
| 2 | GDP Per Capita | 0.233479 (23.3%) |
| 3 | Hospital Beds Per Thousand | 0.095572 (9.6%) |
| 4 | Aged 65 Older | 0.095916 (9.6%) |
| 5 | Population Density | 0.083601 (8.4%) |
| 6 | Human Development Index | 0.064998 (6.5%) |

- Limited predictive power of basic models: The low R-squared values (around 9.8) for linear models indicated that simple GDP and age relationships cannot adequately explain COVID-19 mortality patterns
- Complex factor interplay: The superior performance of the Random Forest model (R-squared 0.92) suggests that mortality prediction requires accounting for complex, non-linear interactions between multiple factors
- Policy matters most: The high importance of the stringency index (42.2%) suggests that government policy responses were the single most influential factor in determining mortality outcomes
- Age vulnerability confirmed: The substantial importance of the aged 65+ population (23.3%) aligns with clinical observations of higher COVID-19 risk in older demographics
- Economic factors are less critical than expected: GDP per capita's relatively modest contribution (9.6%) challenges assumptions that economic development was a primary protective factor

Prioritise rapid government action

Develop specialised strategies for vulnerable demographics




Address density-related transmission risks

Focus on economic indicators in tandem with demographic indicators

Collectively establish linkages between all metrics

# Limitations

Temporal Dynamics

Variable Selection

Geographic Factors

Model Exploration

Causality vs. Correlation

# Thank You