SCHOOL OF
ECONOMICS AND
MANAGEMENT

The Impact of Government Policies on the COVID-19 Pandemic in Scandinavia:

A Machine Learning Approach.

Anie Kdlian

Supervised by: Malgorzata Bogdan

STAN40: First Year Master Thesis in Statistics
15 ECTS

DATA SCIENCE COURSE PACKAGE

Lund University
School of Economics and Management
Department of Statistics

Spring 2024

Abstract

This thesis, "The Impact of Government Policies on the COVID-19 Pandemic in Scandinavia: A Machine Learning Approach," examines the effects of governmental policies on COVID-19 outcomes in Denmark, Finland, Norway, and Sweden. These countries, while similar in many aspects, adopted varied public health measures in response to the pandemic.

Using data pre-processing techniques and machine learning algorithms, including Ordinary Least Squares (OLS) regression, Random Forest, and Support Vector Regression (SVR), this study analyzes the effectiveness of different interventions. The results highlight the critical role of testing, strict health policies and vaccination in managing the pandemic.

The findings of this thesis reveals some important roles of testing and governments policies in managing the pandemic. Differences in vaccination and booster programs highlight the varying public health strategies, as booster doses showing more impact in Denmark and Finland than in Norway and Sweden. Sweden had a unique approach, characterized by a positive correlation between case numbers and the stringency of public health measures.

Regarding death rates, strict public health measures are associated with lower mortality. Vaccination rates does impact the case numbers, by contributing to less death rates. The distinct patterns observed across the four countries, does reflect the diverse approaches of government policies in combating the effect of the pandemic.

Overall, this research highlights the different responses of public health strategies in Scandinavia, and provides some valuable insights for public health policy and response.

Keywords: Covid-19, pandemic, Scandinavia, Sweden, Denmark, Finland, Norway, Machine Learning, ML, statistics, KNN-impute, support vector regression, Ordinary Linear Regression, Random Forest, Log-normal transformation, Box-Cox transformation, Yeo-Johnson transformation, Covid cases, Covid deaths.

Acknowledgements

I would like to thank my supervisor, Malgorzata Bogdan, for her dedication and availability, even on weekends. Her support has been invaluable.

I am also grateful to my family for their encouragement and understanding. A special thanks to my husband JVA for his constant support and patience.

Thank you all for making this thesis possible.

Contents

1 Introduction

In late 2019, an outbreak of a novel coronavirus originating from China quickly escalated into a global pandemic. The rapid increase in COVID-19 infection cases prompted countries worldwide to deploy various preventive measures to eliminate the spread of the virus (World Health Organization, 2020) [12].

Mostly, all of the European countries followed the lead of the World Health Organization's (WHO) recommendation to limit the rapid increase in the number of the infection cases. Scandinavian countries did also follow most of the recommendations of WHO, but each country had it's own adapted policy. Some called the Swedish policy as gambling (Vogel, 2020)[9], as it didn't believe in total lockdown or the military intervention, therefore Sweden had "loose" restrictions in the eyes of the world. (Kdlian, 2023)[5].

In this paper, we will further examine the Governments policies in the Scandinavian region , and discuss these policies impact on the four different countries in Scandinavia; Denmark, Finland, Norway and Sweden. These Scandinavian countries, have much historical similarities, both geographically and culturally (Wikipedia contributors,2022)[10].

Many data pre-processing methods were used, like K-Nearest Neighbours for data imputation, and for Normalization of Skewed distribution of target variables, Log-Normal, Box-Cox, and Yeo-Johnson transformations were experimented.

The analysis part starts by implementing statistical methods like Ordinary Linear Square Regression (OLS) for estimating coefficients, and continues in implementing some Machine Learning (ML) Methods Machine Learning algorithms like Random Forest and Support Vector Regression (SVR).

Overall, this report will explore the similarities and differences in predicting both, the new COVID-19 cases and deaths in the four main Scandinavian countries using algorithms implemented using Python programming language.

2 Background

2.1 The Covid-19 Virus

In late 2019, a novel virus was found in Wuhan, China. This virus, known as SARS-CoV-2, That belongs to the coronavirus family, that targets the human respiratory system, causing a range of respiratory infections (Wikipedia contributors, 2024) [11].

The symptoms of Covid-19 varies a lot. From coughing, fatigue, to even severe illness that cause death. Certain type of the populations, like the elderly and those with underlying health conditions, are at a higher risk of developing severe complications.

Many efforts from the governments, to control the spread of this novel virus have been made. Such as social distancing, mask-wearing, hand hygiene, and even total or partial lockdowns. Additionally, extensive research to find some Vaccines against the virus, That became a tool in managing the pandemic.

2.2 The Scandinavian Region

The term "Scandinavia," emerged from the collaboration between Danish and Swedish universities, in the beginning of 18th century. The term was derived from "Skåne," the southernmost region of Sweden, which is closest to Denmark and was the location of their joint events. Geographically, Scandinavia includes Denmark, Sweden, and Norway. While, culturally, the region also includes Finland, Iceland, and the Faroe Islands, that defines strong historical, cultural, and linguistic ties.

Credits: The information above is based on Kdlian (2023) [5].

3 Data source and Data Descriptions

3.1 Data Sources

Our World in Data (OWID) is a project by the non-profit Global Change Data Lab. This initiative is a collaborative effort that involves scientific researchers from the University of Oxford. The scientific content is provided by researchers from Oxford, whereas the Global Change Data Lab manages the technical aspects, including the publication and upkeep of the website and its tools.

Data Source Information

- **Source:** Our World in Data (OWID)
- **Data Set:** COVID-19 Data
- **Geographic Coverage:** The original dataset covers more than 200 countries. The focus of this study the four Scandinavian countries; Denmark, Finland, Norway and Sweden.
- **Time Period:** From January 5, 2020, to May 5, 2024.
- **Accessibility:** The complete COVID-19 dataset managed by Our World in Data is regularly updated. It encompasses a comprehensive range of data points, including confirmed cases, deaths, hospitalizations, and testing metrics.
- **Link:** OWID COVID-19 Data Repository
- **Last Accessed:** The dataset was accessed on May 19, 2024.

- **Publisher:** The COVID-19 data repository is maintained by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. Our World in Data website, includes daily updates of COVID-19 related statistics.

3.2 Attributes of Dataset

The dataset has 67 attributes. For a better understanding of the analyses that was conducted in this project, the attributes are categorized and listed as follows (source: OWID GitHub open datasets) [1]:

Confirmed COVID-19 Cases

- `total_cases`, `new_cases`, `new_cases_smoothed`
- `total_cases_per_million`, `new_cases_per_million`, `new_cases_smoothed_per_million`

Confirmed COVID-19 Deaths

- `total_deaths`, `new_deaths`, `new_deaths_smoothed`
- `total_deaths_per_million`, `new_deaths_per_million`, `new_deaths_smoothed_per_million`

Excess Mortality

- `excess_mortality`, `excess_mortality_cumulative`
- `excess_mortality_cumulative_absolute`, `excess_mortality_cumulative_per_million`

Hospital and ICU Admissions

- `icu_patients`, `icu_patients_per_million`, `hosp_patients`, `hosp_patients_per_million`
- `weekly_icu_admissions`, `weekly_icu_admissions_per_million`
- `weekly_hosp_admissions`, `weekly_hosp_admissions_per_million`

Testing and Positive Rates

- `total_tests`, `new_tests`, `total_tests_per_thousand`, `new_tests_per_thousand`
- `new_tests_smoothed`, `new_tests_smoothed_per_thousand`, `positive_rate`, `tests_per_case`, `tests_units`

Vaccinations

- total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters
- new_vaccinations, new_vaccinations_smoothed, total_vaccinations_per_hundred, people_vaccinated_per_hundred
- people_fully_vaccinated_per_hundred, total_boosters_per_hundred, new_vaccinations_smoothed_per_million, new_people_vaccinated_smoothed, new_people_vaccinated_smoothed_per_hundred

Other Indicators

- stringency_index, reproduction_rate, iso_code, continent, location, date
- population, population_density, median_age, aged_65_older, aged_70_older, gdp_per_capita
- extreme_poverty, cardiovasc_death_rate, diabetes_prevalence, female_smokers, male_smokers
- handwashing_facilities, hospital_beds_per_thousand, life_expectancy, human_development_index

3.3 Descriptions of Some Attributes

In order to do an accurate comparative analysis among the four Scandinavian countries, this study uses somr normalized attributes, to ensures feasable comparability across the countries, as their population sizes varies. More details about the data, can be found in the OxCGRT methodology documentation in the following GitHub repository, [8].

- new_cases_smoothed_per_million

  The calculation of *new cases smoothed per million* is given by the following equation:

$$\text{New cases smoothed per million} = \left(\frac{\sum_{i=0}^{6} X_{N-i}}{7 \times P} \right) \times 10^6 \tag{3.1}$$

Given that:

- X_{N-i}: The number of new COVID-19 cases reported on day $N - i$, where i is the number of days before day N ranging from 0 to 6. This represents a week's worth of data up to and including day N.
- P: The population of the country. This value is used to normalize the data per million individuals, making the data comparable across countries with different population sizes.

The equation 3.1 above, provides a smoothed metric by averaging the daily new cases over a week and then normalizing this average per million people. This method reduces the impact of daily reporting variations such as delays or under-reporting on weekends and holidays, thereby offers a better and more reliable measure for defining and comparing the rate of new infections across countries.

- `new_deaths_smoothed_per_million`

The calculation of *new deaths smoothed per million* is given by the following equation:

$$\text{New deaths smoothed per million} = \left(\frac{\sum_{i=0}^{6} D_{N-i}}{7 \times P} \right) \times 10^6 \tag{3.2}$$

Given that:

- D_{N-i}: The number of new COVID-19 related deaths reported on day $N - i$, where i ranges from 0 to 6. This accounts for a week's worth of data up to and including day N, aimed at smoothing out daily fluctuations.
- P: The population of the country. This value is crucial for normalizing the death data per million people, allowing for meaningful comparisons between nations of different sizes.

The equation (3.2) above, by averaging the daily reported deaths over a week and then adjusting this figure to a per million basis, provides a consistent and comparable measure of the pandemic's mortality impact. This smoothed approach, can minimize the variability caused by reporting delays or inconsistencies, specially on weekends and holidays, thus it provides a clearer view of trends and patterns in COVID-19 related deaths across countries.

- `stringency_index`

  The Stringency Index is a composite measure scaled from zero to one hundred, where zero indicates no stringency and one hundred represents the highest level of stringency. This index is constructed according to the nine policy indicators, to reflect the severity of government responses in order to control the spread of COVID-19. These indicators include:

  - School closures
  - Workplace closures
  - Cancellation of public events
  - Restrictions on public gatherings
  - Closures of public transport
  - Stay-at-home requirements
  - Public information campaigns
  - Restrictions on internal movements
  - International travel controls

  Each component is rated, and their average provides the overall Stringency Index score. The stringency index metric is crucial to asses government actions in response to the pandemic, and to help to understand how the different measures might impact the spread and the control of the virus.(OWID Stringency Index, 2020) [7].

- `tests_per_case`: This metric quantifies the number of tests conducted for each detected case of COVID-19, offering a perspective on the breadth of testing relative to the magnitude of the outbreak. It is defined mathematically as the reciprocal of the positive rate. Thus, the formula to compute the tests per case is:

  $$\text{Tests per Case} = \frac{1}{\text{Positive Rate}}$$

  Where the Positive Rate is the proportion of tests that return positive. This measure helps to see whether a country is testing widely enough across its population, which is crucial for controlling the spread of the virus (OWID Coronavirus Testing, 2020)[4].

4 Methods

This section outlines the methodology used to examine the impact of government policies on the COVID-19 pandemic in the Scandinavian region. The primary research question addressed is:

`What is the impact of different government policies on the pandemic in the Scandinavian Region?`

The analysis focuses on the relationship between various government policies and the metrics of new COVID-19 infections and deaths in the region.

4.1 KNN-Imputer

This part handles of missing data with using the method of the K-Nearest Neighbors (KNN) imputation technique, which utilizes five nearest neighbors for calculations. The inconsistency in data reporting, which may be because of the differences in reporting systems among the countries, and the absence of data collecting on non-working days. KNN technique estimates missing values by calculating the average of the five most similar datasets.

The formula for KNN-Imputation, when applying the mean of numerical data, is as the follow:

$$x_{\text{missing}} = \frac{1}{k} \sum_{i=1}^{k} x_{\text{nearest}_i}$$

where:

- x_{missing} is the value that needs to be imputed.
- x_{nearest_i} represents the values from the k nearest neighbors used for the imputation.
- k is the number of nearest neighbors considered, typically set to 5 in this context.

To increase the precision of this KNN method, the data was separated by country prior to applying the imputation, to assure a better accuracy (Beretta, L., Santaniello, A., 2016) [2].

4.2 Data Normalization Methods

Three different Transformations were experimented, in order to have a distribution closest to the Gaussian distribution as possible. The credits for these three methods goes to an article in "Medium.com" (Kumar, 2024) [6].

4.2.1 Log-normal Transformation

The target variable of this analysis is `"new_cases_smoothed_per_thousand"`. So, we start by plotting the distribution of the target variable. We notice that the distribution of the target variable is right-skewed in all of the four countries 36, and most of the data is located to the left. By applying a logarithmic transformation, we can normalize the skewed distributions and make it more symmetrical. Normalizing is an essential step to get a better performance when applying Machine Learning Algorithms, and it can also stabilize the variance across the range of data. So first we apply the transformation (Figure, 38):

$$x_i \longrightarrow log(1 + x_i)$$

Then, we apply the standard scaling (Z-score normalization) on the log-transformed data so the data has a mean (average) of 0 and a standard deviation of 1:

The mathematical formula for the standard scalar transformation of a data point $log(1 + x_i)$ is:

$$log(1 + x_i) \longrightarrow \frac{log(1 + x_i) - \mu}{\sigma}$$

where:

- $log(1 + x_i)$ is the log-transformed data point.
- μ is the mean of the log-transformed data.
- σ is the standard deviation of the log-transformed data.

4.2.2 Box-Cox Transformation

he Box-Cox transformation is a statistical technique used to stabilize variance and make data more normally distributed, which is crucial for many statistical modeling techniques. This method is capable to handle the positive data. It can also capable to handle zero or negative values by a slight modification known as the Box-Cox power transformation (Figure, 39).

The transformation is defined by the formula:

$$y(\lambda) = \begin{cases} \frac{(x+\alpha)^{\lambda}-1}{\lambda} & \text{if } \lambda \neq 0, \\ \log(x+\alpha) & \text{if } \lambda = 0, \end{cases}$$

where:

- x is the original data value.
- λ is the transformation parameter that is determined based on the data to maximize the likelihood of achieving a normal distribution.
- Adding a small value α (like 0.0001 in our implementation) ensures that zero and negative values do not pose computational problems.

In our application, the Box-Cox transformation was implemented to adjust the features that exhibited skewness. The transformation was applied only to numerical columns, which were strictly positive after adjusting by a small constant λ. This pre-processing step was done using the `boxcox1p` function from the SciPy library on the subsets of the countries.

Columns with constant values were excluded from the transformation process to prevent errors and ensure meaningful modifications to the dataset.

The effectiveness of this transformation in preparing data in handling complex datasets, is detailed in Kumar's discussion on various data transformations in machine learning applications (Kumar, 2024) [6].

4.2.3 Yeo-Johnson Transformation

The Yeo-Johnson transformation is utilized to normalize data, and to make data distributions more symmetric and thus, more suitable for analytical models that assume normality. This transformation is particularly useful in handling skewness in both positive and non-negative data, making it a good choice in pre-processing steps(Figure, 40).

The transformation formula for the Yeo-Johnson method is expressed as the follow:

$$y(\lambda) = \begin{cases} [(y+1)^{\lambda} - 1]/\lambda & \text{if } \lambda \neq 0 \text{ and } y \geq 0, \\ \log(y+1) & \text{if } \lambda = 0 \text{ and } y \geq 0, \\ -\log(-y+1) & \text{if } \lambda = 2 \text{ and } y < 0, \\ -[-(y+1)^{2-\lambda} - 1]/(2-\lambda) & \text{if } \lambda \neq 2 \text{ and } y < 0. \end{cases}$$

- y represents the original data values, adjusted by adding 1 to accommodate zero or negative values.
- λ is the transformation parameter optimized to best stabilize variance and bring data closer to a normal distribution.

5 Pre-Analysis

This section outlines the methods that were used to prepare the data for analysis. From handling missing values, normalization, and standardization. Additionally, it details the formulation of the regression model, like describing the target variables and factors.

5.1 Handling missing values

The initial step in our analysis is to remove columns that are either unnecessary or have a high percentage of missing values. After clustering the dataset by the four countries;Denmark, Finland, Norway, and Sweden, We assess each attribute's relevance.

The original dataset contains 67 attributes. To derive meaningful information, we eliminate those attributes that are missing over 95% of their data across the entire dataset. A further step is to eliminate any of the attributes that are entirely missing in one or more of the specified countries.

Following this, we eapply the KNN Imputation method to handle the remaining missing values. The KNN Imputer handles these by using the five nearest neighbors (k=5) for imputation. This method is effective, as the dataset varies in reporting systems across the four different countries.As, reporting may not be done on a daily base. Hence, possibly some data is missing on holidays or not recorded daily. By using the nearest neighbor method, we can fill these gaps with the average values from the five closest neighbors. For better accuracy for imputation, we segmented the dataset into subsets for each country before applying the imputation.

5.1.1 Data Normalization

In this part we aim to Normalized the data to avoid the high variance in data, which we noticed after plotting the attributes "The number of Covid-19 new cases per thousand" and "The New COVID-19 Deaths smoothed per thousand in each of the four Scandinavian countries.In our analysis, we experimented three different methods for an attempt for normalization across the several countries. These methods for normalization transformation are, the Log-normal transformation, Box-Cox transformation and Yeo-Johnson transformation. Overall, The Yeo-Johnson transformation has improved the normality of the datasets. More information regarding the normalization method can be found in "Methods" part of this paper, section "4.2- Data Normalization Method".

5.1.2 Standardization of variables scales

Standardization (variable scaling) is important to ensure consistency in data analysis, especially when comparing statistics across different countries. A key step in this process involves normalizing all data attributes to a common scale, specifically 'per thousand' inhabitants. Originally, the dataset included various scales, with some attributes measured per hundred, others per thousand, and some even per million inhabitants. An adjustment in these variables were made to a uniform scale, Per thousand.

5.1.3 Government different Policies

The Policies that the government did follow were different. In this paper, we will focus on three different policies. These policies are selected according to the available attributes of OWID dataset and inspired by the article available in "Nature.com", (Hale, T., Angrist, N., Goldszmidt, R. et al., 2021) [3]

- `Public Health Policies:` This policy describes nine different indicators, which are as the follow:
  - school closures.
  - workplace closures.
  - cancellation of public events.
  - restrictions on public gatherings.
  - Closures of public transport.
  - stay-at-home requirements.
  - public information campaigns.
  - restrictions on internal movements.
  - international travel controls.

- `Health system Policies:` These include measures such as testing, vaccination programs, booster shots, and the number of hospital beds.
- `Economic Policies:` The focus is on the economic indicators "Gross Domestic Product (GDP) per capita", which provides a general indication of the average living standards in a country.

These policy areas were integral to the models used in our analysis, serving as key factors in assessing impacts and outcomes.

6 Analysis

6.1 Part 1: Analysis of the Number of New Covid-19 Cases

The Number of new Covid-19 cases

in this part of the analysis, we will present the Analysis of

the number of Covid new cases. The attribute that we are using as our target variable is "new_cases_smoothed_per_thousand".

For the analysis, 70% of the data has been used for training and 30% for testing.

6.1.1 Ordinary Least Squares Regression (OLS)

First approach is about using a fundamental statistical technique to estimate the relationship between number of Covid new cases (smoothed per thousand), and some chosen variables to test against the three government policies .

You can find a detailed description about the government policies that we are testing, in the section "Pre-Analysis" of this paper under the title "Government different Policies".

Denmark Case

The regression model aims to explain variations in the smoothed number of new COVID-19 cases per thousand people in Denmark, using several predictors such as total tests per thousand, hospital beds per thousand, the stringency index, GDP per capita, total vaccinations per thousand, and total booster doses per thousand.

- The OLS model explains about 47.2% (R-squared = 0.472) of the variability in the dependent variable (new cases smoothed per thousand)(Figure 1), which is understandable, given the complexity of the dataset.
- **Coefficients:**
  - The coefficient for *Total Vaccinations per Thousand* is not statistically significant (p = 0.339). This suggests that in our model, there is no clear effect of total vaccinations per thousand on the new case rate.
  - On the other hand, we have five highly significant coefficients ($p < 0.001$). Two of these have a negative impact on the number of COVID-19 cases:
    - *GDP per Capita* (Coefficient = -3.262e-05), suggesting that wealthier areas, like Denmark, may have better management strategies leading to fewer COVID-19 cases.
    - *Hospital Beds per Thousand* (Coefficient = -1.747e-09), indicating that better healthcare availability might slightly reduce cases, though the effect size is practically negligible.
  - The remaining significant coefficients all show a positive impact on the number of COVID-19 cases:
    - *Total Tests per Thousand* (Coefficient = 0.0002), reflecting that increased testing is associated with detecting more cases.
    - *Stringency Index* (Coefficient = 0.0228), possibly indicating that stricter government measures might lead to increased case detection or reflect a reaction to rising case numbers.
    - *Total Booster Doses per Thousand* (Coefficient = 0.0014), which may suggest that regions with higher booster administration are either better at detecting cases or see transient increases in cases post-booster due to enhanced social interaction.

Finland Case

The regression model is used to explain variations in the smoothed number of new COVID-19 cases per thousand people in Finland, using predictors such as total tests per thousand, hospital beds per thousand, the stringency index, GDP per capita, total vaccinations per thousand, and total booster doses per thousand.

- The OLS model explains about 48.1% (R-squared = 0.481) of the variability in the dependent variable (new cases smoothed per thousand)(Figure, 3), which is also understandable given the complexity of the data.
- All the coefficient in the model of Finland shows statistical significant ($p < 0.001$), and are divided equally between the negative and positive impact on the number of new Covid cases.
  - *The coefficients with a negative impact on the number of new Covid cases.*
    - *GDP per Capita* (Coefficient = -6.166e-05), suggesting that higher economic levels in Finland are associated with better management and thus fewer COVID-19 cases.
    - *Hospital Beds per Thousand* (Coefficient = -4.983e-09), indicating that greater healthcare capacity may marginally reduce the number of cases, reflecting effective management and treatment capabilities, although the effect size is minimal.
    - *Total Booster Doses per Thousand* (Coefficient = -0.0027), which unexpectedly suggests a decrease in cases with more booster doses, potentially indicating the effectiveness of booster vaccinations in controlling the spread of the virus.
  - *The coefficients with a positive impact on the number of new Covid cases*
    - *Total Tests per Thousand* (Coefficient = 0.0011), demonstrating that increased testing correlates with higher detected case numbers, likely due to improved case identification.
    - *Stringency Index* (Coefficient = 0.0336), possibly showing that stricter measures and responses to rising cases can lead to increased detection or reporting of new infections.
    - *Total Vaccinations per Thousand* (Coefficient = 0.0009), This might reflect higher detection rates as healthcare engagement increases with vaccination efforts.

Norway Case

The regression model explains the variations in the smoothed number of new COVID-19 cases per thousand people in Norway.

- The OLS model explains about 43.3% (R-squared = 0.433) of the variability in the dependent variable (new cases smoothed per thousand) (Figure, 5), which is understandable, given the complexity of the factors influencing pandemic trends.
- All the coefficients in the model of Norway show statistical significance ($p < 0.001$), with a clear delineation between those that have

a positive and those that have a negative impact on the number of new COVID cases.

* *The coefficients with a negative impact on the number of new Covid cases:*
  - *GDP per Capita* (Coefficient = -2.781e-05), indicating that higher economic levels in Norway are associated with better management strategies leading to fewer COVID-19 cases.
  - *Hospital Beds per Thousand* (Coefficient = -1.545e-09), suggesting that an increased healthcare capacity may slightly reduce the number of cases, reflecting efficient healthcare system capabilities, albeit the effect size is very small.
  - *Total Booster Doses per Thousand* (Coefficient = -0.0037), strongly indicating that booster vaccinations play a significant role in reducing the spread of the virus, likely reflecting an effective vaccination program.
* *The coefficients with a positive impact on the number of new Covid cases:*
  - *Total Tests per Thousand* (Coefficient = 0.0005), showing that increased testing is associated with a higher detection of cases, which is expected as more comprehensive testing uncovers more cases.
  - *Stringency Index* (Coefficient = 0.0208), possibly indicating that tighter government restrictions and measures, although intended to control the spread, correlate with periods of higher reported cases, possibly due to increased transmission before measures take effect or increased compliance and testing during high-stringency periods.
  - *Total Vaccinations per Thousand* (Coefficient = 0.0011), which might suggest that as vaccination rates increase, so do the detected cases, perhaps due to higher interaction levels as public confidence in safety grows, or increased detection through focused testing in vaccinated populations.

Sweden Case

The regression model aims to explain variations in the smoothed number of new COVID-19 cases per thousand people in Sweden.(Figure, 7)

- The OLS model explains about 63% (R-squared = 0.630) of the variability in the dependent variable (new cases smoothed per thousand) (Figure, 7), which is considered good, given the complexity of the model in capturing pandemic dynamics in Sweden.**??**
- **Coefficients:**

* The coefficient for *Total Booster Doses per Thousand* is not statistically significant (p = 0.059), suggesting that within this model framework, booster doses do not have a clear impact on new case rates.
* The model identified three coefficients with a negative impact on the number of new COVID-19 cases, all statistically significant ($p < 0.001$):
  - *GDP per Capita* (Coefficient = -3.53e-05), indicating that higher economic levels are associated with better management and thus fewer COVID-19 cases.
  - *Hospital Beds per Thousand* (Coefficient = -1.669e-09), suggesting that superior healthcare infrastructure may marginally reduce the number of cases.
  - *Total Vaccinations per Thousand* (Coefficient = -0.0010), reflecting that, increased vaccination rates might lead to a short-term rise in reported cases due to enhanced surveillance and testing but overall suggest a negative correlation with case numbers.
* There are also two coefficients that have a positive impact on the number of new COVID-19 cases:
  - *Total Tests per Thousand* (Coefficient = 0.0023), confirming that more extensive testing is linked to higher detection of cases.
  - *Stringency Index* (Coefficient = 0.0337), which may imply that stricter government measures, while intended to control the virus, correspond with higher reported cases possibly due to increased transmission just before and compliance after measures are implemented.

6.1.2 Random Forest Model on the Number of New Covid-19 Cases

This section presents the analysis of COVID-19 case trends using Random Forest regression models for Denmark, Finland, Norway, and Sweden. The model aims to identify the most significant predictors affecting the number of new COVID-19 cases per thousand people.

Overall, the variance explained in the four countries are considered high, as R-squared value is around 0.98. This might highly likely indicate an over-fitting problem in the model of Random Forest.

In order to ensure a good model that generalizes well on a new unseen data. Many methods were implemented, like using "Cross validation" in finding the best hyper-parameters, also pruning method to reduce the risk of over-fitting, and parameter tuning using "GridSearchCV".

Denmark Case

The Random forest model in Denmark The prediction of the new cases in Denmark using Random Forest has a The Random Forest model for Denmark has the indicator R^2 value of 0.9898.

- **Feature Importance:** (Figure, 25)
  - *Total Tests per Thousand* has the highest importance score (0.3951), suggesting that testing volume is a crucial predictor of case numbers.
  - *Total Vaccinations per Thousand* follows with an importance score of 0.3259, indicating good influence, possibly capturing effects of widespread vaccination on new Covid-19 case trends.
  - *Total Boosters per Thousand* also shows substantial impact (0.2623), reflecting the role of booster campaigns in the pandemic control.
  - *Stringency Index* has a relatively minor importance (0.0167), indicating less influence compared to testing and vaccination metrics.

Finland Case

The model for Finland also shows high predictive performance with an R^2 value of 0.9891.

- **Feature Importance:** (Figure, 26)
  - *Total Boosters per Thousand* is the predictor with most impact (0.3696), showing the importance of booster doses in controlling the spread.

* *Total Tests per Thousand* remains a significant predictor (0.3398), consistent with the testing-tracing approach.
* *Stringency Index* has a considerable importance (0.2571), suggesting that government interventions had a notable impact on case numbers in Finland.

Norway Case

Norway's model has an R^2 value of 0.9837.

- **Feature Importance:** (Figure, 27)
  * *Total Vaccinations per Thousand* leads with the highest importance (0.4163), indicative of the effectiveness of vaccination.
  * *Total Tests per Thousand* (0.3349) and *Total Boosters per Thousand* (0.2297) also plays significant role, aligning with observed public health responses.

Sweden Case

Sweden's model, with an R^2 of 0.9946.(Figure, 28)

- **Feature Importance:** (Figure, 28)
  * *Stringency Index* has an unusually high importance score (0.6539), potentially reflecting the specific nature of Sweden's public health strategy.
  * *Total Tests per Thousand* also remains a crucial factor (0.2516), emphasizing the testing policy.

6.1.3 Support Vector Machine (SVR) on the Number of New Covid-19 Cases

This section presents the outcomes of the Support Vector Machine (SVR) models employing the Radial Basis Function (RBF) kernel to predict COVID-19 case trends per thousand people in Denmark, Finland, Norway, and Sweden. The models assess the impact of the various predictors and determine their importance through permutation feature importance.

Each country's SVR model was configured with an RBF kernel, optimized through Halving Grid Search to ensure effective parameter tuning. Permutation feature importance was computed to evaluate the influence of each predictor on the model's performance, to provide insights into which factors has the most significant effect on COVID-19 case trends.

Denmark Case

The SVR model for Denmark highlights several key predictors influencing COVID-19 case numbers.

- **Permutation Feature Importance:** (Figure, 29)
  - *Stringency Index* shows the highest importance measures on case trends.
  - *Total Tests per Thousand*, *Total Vaccinations per Thousand* and *Total Boosters per Thousand* also play significant roles, reflecting the importance of testing and vaccination in controlling the spread of Covid-19 cases in Denmark.

Finland Case

Finland's SVM model demonstrates some different pattern,in feature importance, it underlines the role of vaccination and boosters with number of Covid-19 new cases.

- **Feature Importance:** (Figure, 29)
  - *Total Boosters per Thousand* and *Total Vaccinations per Thousand* are prominent, indicating the critical role of vaccination campaigns.
  - *Stringency Index* and *Total Tests per Thousand* indicators, highlight the significance of stringent measures and widespread testing.

Norway Case

Norway's SVR model identifies different aspects that influence the spread of Covid infection, with a strong focus on first testing and then vaccination.

- **Feature Importance:** (Figure, 30)
  - *Total Tests per Thousand* emerges as the most influential factor, followed by
  - *Total Vaccinations per Thousand*, indicating effective public health response.

Sweden

Sweden's model underscores the importance of testing and public health policies.

- **Feature Importance:** (Figure, 30)
  - *Stringency Index* and *Total Tests per Thousand* has high impact, reflecting Sweden's unique approach to managing the pandemic.

6.2 Part 2: Analysis of the Number of New Covid-19 Deaths

In this section of the analysis, we explore the factors affecting the number of new COVID-19 deaths, smoothed per thousand people, using the OWID COVID-19 dataset. This model tries to identify the contribution of different variables to mortality rates across different countries.

Ordinary Least Squares Regression (OLS) the Number of New Covid-19 Deaths

Denmark Case

The regression model aims to explain variations in the smoothed number of new COVID-19 deaths per thousand people in Denmark. Predictors include total tests per thousand, hospital beds per thousand, the stringency index, GDP per capita, total vaccinations per thousand, and total booster doses per thousand.

- The OLS model explains about 33.4% (R-squared = 0.334) of the variability in the dependent variable (new deaths smoothed per thousand), showing a moderate explanatory power (Figure 13).
- **Coefficients:**
  - Significant coefficients with a negative impact on the number of new COVID-19 deaths include:
    - *GDP per Capita* (Coefficient = -4.424e-05, $p < 0.001$), indicates that higher economic levels correlate with lower death rates.
    - *Hospital Beds per Thousand* (Coefficient = -2.369e-09, $p < 0.001$), suggests that better healthcare infrastructure helps to reduce mortality rates.
  - Coefficients with a positive impact:
    - *Total Tests per Thousand* (Coefficient = 9.876e-05, $p < 0.001$), implies that increased testing may correlate with higher recorded death rates, possibly due to better reporting and case confirmation.
    - *Stringency Index* (Coefficient = 0.0303, $p < 0.001$), which could reflect that stricter measures are often a response to severe outbreaks, hence higher mortality rates during such periods.
    - *Total Booster Doses per Thousand* (Coefficient = 0.0039, $p < 0.001$), suggests that increased booster coverage might temporally associate with mortality due to various factors like initial vaccine program roll-out to high-risk groups.

Finland Case

The regression model for Finland shows how several predictors relate to the number of new COVID-19 deaths per thousand people.

- The OLS model accounts for about 38.7% (R-squared = 0.387) of the variability in the dependent variable (new deaths smoothed per thousand) (Figure, 15), indicating strong model performance.
- **Coefficients:**
  - Negative impacts:
    - *Hospital Beds per Thousand* (Coefficient = -5.149e-09, $p < 0.001$), shows significant healthcare capacity impact on reducing death rates.
    - *GDP per Capita* (Coefficient = -6.372e-05, $p < 0.001$), supports the idea that higher economic standards improve survival rates.
  - Positive impacts:
    - *Stringency Index* (Coefficient = 0.0384, $p < 0.001$), perhaps indicates that higher stringency levels relate to periods with more deaths.
    - *Total Vaccinations per Thousand* (Coefficient = 0.0012, $p < 0.001$), possibly reflects that increased interactions to vulnerable populations.

Norway Case

This model explores the impact of various factors on the number of new COVID-19 deaths per thousand people in Norway.

- The OLS model explains 12.4% (R-squared = 0.124) of the variability in the dependent variable (new deaths smoothed per thousand) (Figure, 17), indicating limited explanatory power.
- **Coefficients:**
  - Negative impacts:
    - *Hospital Beds per Thousand* (Coefficient = -1.261e-09, $p < 0.001$), shows the importance of healthcare availability in mortality reduction.
    - *GDP per Capita* (Coefficient = -2.27e-05, $p < 0.001$), correlates the higher economic status with better health outcomes.
  - Positive impacts:
    - *Stringency Index* (Coefficient = 0.0107, $p < 0.001$), which may reflects stricter measures during higher mortality phases.

Sweden Case

Analyzing the factors affecting new COVID-19 deaths per thousand people in Sweden through regression modeling.

- The OLS model explains 26.5% (R-squared = 0.265) of the variability in the dependent variable (new deaths smoothed per thousand) (Figure, 19), showing some level of explanation.
- **Coefficients:**
  * Negative impacts:
    - *Hospital Beds per Thousand* (Coefficient = -1.485e-09, $p < 0.001$), indicating the critical role of healthcare capacity.
    - *GDP per Capita* (Coefficient = -3.14e-05, $p < 0.001$), reinforcing the impact of economic well-being on health outcomes.
  * Positive impacts:
    - *Stringency Index* (Coefficient = 0.0505, $p < 0.001$), suggests increased deaths may coincide with the implementation of stringent measures during severe epidemic waves.
    - *Total Vaccinations per Thousand* (Coefficient = 0.0015, $p < 0.001$), reflects complex dynamics, possibly it is related to early vaccine distribution among high-risk groups.

6.2.1 Random Forest Model on the Number of New Covid-19 Deaths

This section presents the analysis of COVID-19 death trends using Random Forest regression models for Denmark, Finland, Norway, and Sweden. The models aim to identify the most significant predictors affecting the number of new COVID-19 deaths per thousand people.

While the Random Forest models demonstrate high R^2 values, suggesting good model fit, there is a risk of over-fitting given the complexity of the pandemic data. Careful measures such as cross-validation, pruning, and hyperparameter tuning using GridSearchCV have been implemented to enhance model robustness and generalizability.

Denmark Case

The Random Forest model for Denmark shows that certain predictors are particularly influential in explaining the death trends due to COVID-19.(Figure, 31)

- **Feature Importance:**

* *Stringency Index* is the most significant predictor (Importance: 0.425914), indicating the impact of government measures on mortality rates.
* *Total Vaccinations per Thousand* (Importance: 0.255245) and *Total Tests per Thousand* (Importance: 0.195389) also significantly affect the death rates, underlining the roles of widespread testing and vaccination efforts.
* *Total Boosters per Thousand* shows considerable importance (0.123452), suggesting the effectiveness of booster campaigns in reducing severe cases and deaths.

Finland Case

The model for Finland identifies similar patterns in feature importance as seen in the Danish model, indicating consistent factors influencing death rates across these nations.(Figure, 31)

- **Feature Importance:**
  * *Stringency Index*, *Total Vaccinations per Thousand*, *Total Tests per Thousand*, and *Total Boosters per Thousand* mirror the importance rankings observed in Denmark, emphasizing the critical role of health policy and vaccination in managing the pandemic's impact.

Norway Case

Norway's Random Forest model also highlights the key drivers behind the COVID-19 death rates with an R^2 value (Figure, 32).

- **Feature Importance:**
  * *Total Vaccinations per Thousand* emerges as the top predictor (Importance: 0.416313), followed by *Total Tests per Thousand* (Importance: 0.334853) and *Total Boosters per Thousand* (Importance: 0.229705).
  * These factors suggest that vaccine rollout and testing coverage are central to understanding and controlling COVID-19 mortality.

Sweden Case

Sweden's model, shows a good fit with high R^2 values potentially indicating over-fitting.(Figure, 32)

- **Feature Importance:**

    * *Stringency Index* dominates (Importance: 0.653916), significantly more than in other countries, possibly due to Sweden's unique approach to pandemic restrictions and measures.
    * *Total Tests per Thousand* and *Total Vaccinations per Thousand* remain crucial, highlighting the different trends in Sweden's pandemic response.

6.2.2 Support Vector Machine (SVR) the Number of New Covid-19 Deaths

This section presents the outcomes of the Support Vector Machine (SVR) models to predict COVID-19 deaths trends per thousand people in Denmark, Finland, Norway, and Sweden. The models assess the impact of the various predictors and determine their importance through permutation feature importance, a technique suitable for non-linear SVR models.

Each country's SVR model was configured with an RBF kernel, optimized through Halving Grid Search to ensure effective parameter tuning. Then, permutation feature importance was computed to evaluate the influence of each predictor on the model's performance, to provide insights into which factors has the most significant effect on COVID-19 death trends.

Denmark Case

The SVR model for Denmark underscores the importance of several predictors in influencing COVID-19 death numbers.The model have an R-squared = 0.513.

- **Permutation Feature Importance:**
    * *Total Boosters per Thousand* has the highest importance, which emphasis the critical role of booster vaccinations in managing death rates.
    * *Stringency Index* and *Total Tests per Thousand* also has a significant importance, which reflects the importance of government interventions and extensive testing in reducing death cases.

Finland Case

Finland's model indicates a slightly different pattern of feature importance, it highlights the impact of public care policies on death rates. The model have an R-squared = 0.725.

- **Feature Importance:**

    * *Total Boosters per Thousand* and *Total Vaccinations per Thousand* are the most important predictors, underlining the importance of vaccination efforts in reducing COVID-19 deaths.
    * *Stringency Index* and *Total Tests per Thousand*, suggests that stringent measures and robust testing are vital in the fight against the pandemic.

Norway Case

Norway's SVR model reveals a strong dependency on certain health measures to manage death trends effectively. The model have an R-squared = 0.483.

- **Feature Importance:**
    * *Total Tests per Thousand* and *Total Vaccinations per Thousand* are the most important factors, which underlines the importance of testing and vaccination programs.

Sweden Case

The SVR model for Sweden in managing death trends, has an R-squared = 0.581.

- **Feature Importance:**
    * *Stringency Index* and *Total Tests per Thousand* are the most important features. Which indicates the success of Sweden's unique public health strategies.

7 Results: Comparative Analysis

7.1 FIRST PART: Results on the Number of Covid-19 Cases

7.1.1 Comparative Analysis of OLS Regression Models for COVID-19 New cases Across Countries

The Ordinary Least Squares (OLS) regression analysis gives an insight into the impact of various factors on the smoothed number of new COVID-19 cases per thousand across Denmark, Finland, Norway, and Sweden. By comparing these models, we can identify commonalities and distinctions in how different variables influence COVID-19 case trends within these nations.

Similarities

Across all four countries, the results shows consistent patterns with certain predictors. For instance, the coefficients for *Total Tests per Thousand* and *Stringency Index* are positive in each country, indicating that increased testing and stricter government measures generally correspond with a higher number of reported COVID-19 cases. This could be attributed to more comprehensive detection and the lag effect of stricter measures. Additionally, both *Hospital Beds per Thousand* and *GDP per Capita* consistently show negative coefficients, suggesting that better healthcare infrastructure and higher economic status are associated with fewer COVID-19 cases. These indicators may reflect the effective management and resilience of healthcare systems and economies in combating the pandemic.

Differences

There are notable differences among the countries in the effects of vaccination rates. Sweden stands out with a negative coefficient for *Total Vaccinations per Thousand*, which contrasts with the positive coefficients observed in Denmark, Finland, and Norway. This might indicate different effectiveness of vaccination campaigns or variations in health care strategies and population behavior post-vaccination. Furthermore, the impact of *Total Booster Doses per Thousand* shows positive importtance in Denmark, implying an increase in cases, potentially due to higher social interaction that follows booster vaccination. In contrast, this predictor has a negative effect in Finland, Norway, and Sweden, suggesting that booster doses may help reduce case numbers by enhancing immunity in these populations.

Overall Results

The models for all four countries has a R-squared values ranging from 43.3% to 63%, which highlights the impact of the predictors on the number of new COVID-19 cases. These findings underscore the importance of good healthcare systems, economic stability, and proactive testing and vaccination policies in managing the spread of COVID-19.

7.1.2 Comparative Analysis of Random Forest Models for COVID-19 New cases Across Countries

This part presents the results of Random Forest models applied to assess the impact of several predictors on the number of new COVID-19 cases per thousand in the four Scandinavian countries: Denmark, Finland, Norway, and Sweden. The models offer insights into the importance of various factors affecting case numbers during the pandemic.

Similarities Across the Countries

The Random Forest results revealed common significant predictors across all four countries. For instance:

– **Total Tests per Thousand** consistently shown as an important predictor, underscoring the critical role of testing in identifying case numbers across the board.

– **Stringency Index** also is shown as an important indicator in all of the four countries, suggesting that government interventions and policies played an important role in the dynamics of the pandemic's spread.

Differences

While there are common factors, each country also had some unique patterns in how other variables influenced COVID-19 case numbers:

- In **Denmark**, the most important predictors after total tests were *total vaccinations per thousand* and *total boosters per thousand*, which highlights the impact of vaccination campaigns.
- **Finland** showed a big reliance on *total boosters per thousand*, which indicates a robust booster campaign influence, followed by *stringency measures*.
- For **Norway**, *total vaccinations per thousand* has the most impact, followed closely by *total tests per thousand* and *total boosters per thousand*.
- **Sweden** uniquely displayed the *Stringency Index* as the predominant feature, followed by *total tests per thousand*, which may reflect specific public health policies.

Overall Results

The Random Forest models achieved high R-squared values around 0.98 in all countries, which might indicate a strong fit or even an overfitting problem to the data. However, each country's model underscores different aspects of the pandemic response, from testing and stringency to vaccination rates, revealing tailored responses and outcomes based on governments strategies and healthcare capabilities. These results provide valuable insights into the effectiveness of various strategies in each country to manage the pandemic.

7.1.3 Comparative Analysis of Support Vector Regression for COVID- 19 New cases Across Countries

The Support Vector Regrssion (SVR) Method was used to understand how different factors influence the new COVID-19 cases per thousand in Denmark, Finland, Norway, and Sweden. This method helps in underlying variables that are most significant in predicting case trends in these countries.

Similarities

The SVR results show that in all four countries, testing rates and government response stringency consistently have a strong influence on the number of reported COVID-19 cases. Higher testing rates lead to more cases being detected, while more stringent measures likely reflect government response to increasing cases.

Differences

Notable differences are observed in the impact of vaccinations and booster doses. While booster doses are important predictors in Denmark and Finland, showing a positive correlation with case numbers, their influence is less significant in Norway and Sweden. Furthermore, Sweden shows a unique trend where the stringency of government measures is the most influential factor, suggesting a particularly robust governmental response to the pandemic compared to the other countries.

Overall Results

The SVR models for Denmark, Finland, Norway, and Sweden highlight the importance of testing and government policies in controlling the spread of COVID-19. While each country exhibits unique method in the effectiveness of vaccinations and healthcare resources in managing the pandemic, Sweden distinctively shows that increases in case numbers lead to more stringent public health measures, alongside testing and vaccination strategies.

7.2 SECOND PART: Results on the Number of Covid-19 Deaths

7.2.1 Comparative Analysis of OLS Regression Regression for COVID- 19 Deaths Across Scandinavia

Ordinary Least Squares (OLS) regression was used to explore the factors affecting the smoothed number of new COVID-19 deaths per thousand in the Scandinavian countries of Denmark, Finland, Norway, and Sweden. The analysis allows us to find both common and unique influences of various predictors on death rates due to COVID-19 across these nations.

Similarities

Across Denmark, Finland, Norway, and Sweden, certain variables consistently did influence the death rates. Notably, *Hospital Beds per Thousand* had a statistically significant negative coefficient in all of the countries, suggesting that better hospital infrastructure is associated with lower death rates. Additionally, *Stringency Index*, representing the severity of government health care policies, showed a positive relationship with the death rates, indicating that higher stringency might be a response to severe outbreaks.

Differences

Despite these similarities, the countries also displayed distinct patterns:

- In Denmark and Sweden, *GDP per Capita* presented a significant negative effect, highlighting that higher economic status might help in relieving death rates more effectively compared to Finland and Norway where its impact was not significant.
- The effect of *Total Vaccinations per Thousand* varied, with Denmark showing a small but significant negative coefficient, suggesting some effectiveness of vaccination in reducing death rates. Contrastingly, in Sweden, this variable was not a significant predictor.
- *Total Tests per Thousand* had varying impacts, with a positive coefficient in Norway, indicating that increased testing correlates with higher reported death rates possibly due to better detection of COVID-19 related deaths.

Overall Results

The R-squared values from these models suggest that while the predictors explain a substantial portion of the variance in COVID-19 death rates, there are still other factors that play a role in these outcomes. The models show that healthcare infrastructure and government policies are key in controlling the death rates from the pandemic. Besides, reveal how economic and social factors interact with public health measures.

7.2.2 Comparative Analysis of Random Forest Models for COVID-19 Deaths Across Scandinavia

The Random Forest analysis explores the factors that influence COVID-19 death rates in Denmark, Finland, Norway, and Sweden. By examining these models, we can identify both common and country-specific observations .

Similarities

The analysis shows that the stringency index, and the number of vaccinations are consistently important across all four countries. These factors indicate that strict policies and high vaccination rates are effective in managing the mortality rates of COVID-19.

Differences

While testing and vaccination are crucial across the board, the impact of booster doses varies among countries. Denmark and Finland see significant influence from booster doses, suggesting that these boosters play a key role in reducing death rates.

Overall Results

The Random Forest models achieve high R-squared values, suggesting a good fit that captures a substantial portion of the variability in death rates. This highlights the importance of stringency measure and vaccinations to manage the pandemic's mortality impact.

7.2.3 Comparative Analysis of SVR for COVID-19 Deaths Across Scandinavia

This analysis examines the effectiveness of the Support Vector Machine (SVR) model in predicting COVID-19 death rates in Denmark, Finland, Norway, and Sweden. The study reveals how various factors contribute to the mortality outcomes associated with the pandemic across these countries.

Similarities

The SVR models used for each country, shows that certain factors like total tests per thousand and stringency of government policies play a consistent role in all countries. High stringency index values and extensive testing are commonly important factors across Denmark, Finland, Norway, and Sweden, highlighting their critical roles in managing the pandemic's impact.

Differences

While there are notable differences also emerge in the impact of vaccination rates and healthcare capacity. For instance, total boosters per thousand has varying levels of importance, suggesting differences in how booster vaccination campaigns influence death rates across these nations.

Overall Results

The SVR models has varying R-squared values, indicating different levels of model accuracy in predicting death rates across the countries. Finland and Sweden show higher explanatory power, suggesting that the models fit better in these countries compared to Denmark and Norway. This variation underlines the complex nature of pandemic dynamics and the influence of localized public health strategies.

8 Discussion and Findings

While the models vary, they all emphasize the significant role of testing and public health policies (Stringency Index)in managing the pandemic, and the spread of the Covid-19 virus. Differences in vaccination and booster effects, suggest varying levels of vaccine roll-out effectiveness and public health strategies among these countries. For example, booster doses have a more impact in Denmark and Finland compared to Norway and Sweden, indicating different health responses to vaccination campaigns. Consequently, the implementation of effective public and health policies plays a crucial role in mitigating the spread of COVID-19.

A notable observation from the analysis is that Sweden demonstrated a distinct approach in managing the rise in COVID-19 cases. As, the data shows a significant positive correlation between the number of cases and the stringency of public health measures implemented. This suggests that Sweden actively intensified its governmental response as case numbers increased, reflecting a proactive strategy in curbing the spread of the virus.

In terms of death rates, the models indicate that

the stringency of government measures often shows a strong association with lower death rates, suggesting effective containment and mitigation strategies can significantly influence mortality outcomes.

Additionally, vaccination rates, while varying in their impact on case numbers, generally help in reducing death rates, underscoring their importance in public health strategies across Scandinavia. The distinct patterns in each country highlight the different approaches and varied effectiveness of Governments policies in combating the pandemic's effects.

8 Bibliography

[1] owid/covid-19-data:. https://github.com/owid/covid-19-data/tree/master/public/data.

[2] Lorenzo Beretta and Antonio Santaniello. Nearest neighbor imputation algorithms: A critical evaluation. https://doi.org/10.1186/s12911-016-0318-z, 2016.

[3] Thomas Hale, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, Emily Cameron-Blake, Laura Hallas, Saptarshi Majumdar, and Helen Tatlow. A global panel database of pandemic policies (oxford covid-19 government response tracker). *Nature Human Behaviour*, 2021.

[4] Our World in Data. Coronavirus testing. https://ourworldindata.org/coronavirus-testing, 2023. Accessed: 2023-05-14.

[5] Kdlian, Anie. The impact of "loose " restrictions in Sweden during the COVID-19 virus , compared to the neighbouring Scandinavian countries. http://lup.lub.lu.se/student-papers/record/9111768, 2023. Student Paper.

[6] Vinod Kumar. Data transformations in ml: Different transformations in machine learning. https://medium.com/@vinodkumargr/09-data-transformations-in-ml-different-transformations-in-machine-learning-log-tr 2019. Accessed: 2024-05-14.

[7] Our World in Data. Covid-19 stringency index. https://ourworldindata.org/covid-stringency-index, 2024.

[8] Oxford COVID-19 Government Response Tracker. Oxcgrt documentation: Index methodology. https://github.com/OxCGRT/covid-policy-tracker/blob/master/documentation/index_methodology.md, 2024.

[9] Gretchen Vogel. Sweden's gamble. https://www.science.org/doi/10.1126/science.370.6513.159, 2020.

[10] Wikipedia contributors. Scandinavia — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Scandinavia&oldid=1105651968, 2022. [Online; accessed 29-August-2022].

[11] Wikipedia contributors. Coronavirus — Wikipedia, the free encyclopedia, 2024. [Online; accessed 3-June-2024].

[12] World Health Organization and others. WHO Director-General's opening remarks at the media briefing on COVID-19-11 March 2020. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-mar 2022.

Appendix of Graphs

I. OLS Regression Result for Covid New Cases in Denmark

```
Country: Denmark
Model Summary:
                            OLS Regression Results
==========================================================================================
Dep. Variable:     new_cases_smoothed_per_thousand   R-squared:                       0.472
Model:                                         OLS   Adj. R-squared:                  0.470
Method:                              Least Squares   F-statistic:                     246.0
Date:                             Tue, 21 May 2024   Prob (F-statistic):          4.55e-151
Time:                                     07:42:01   Log-Likelihood:                -1264.2
No. Observations:                             1108   AIC:                             2538.
Df Residuals:                                 1103   BIC:                             2563.
Df Model:                                        4
Covariance Type:                         nonrobust
==================================================================================================
                                     coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------------
total_tests_per_thousand           0.0002   6.53e-06     27.147      0.000       0.000       0.000
hospital_beds_per_thousand     -1.747e-09   2.31e-10     -7.577      0.000    -2.2e-09   -1.29e-09
stringency_index                   0.0228      0.003      6.938      0.000       0.016       0.029
gdp_per_capita                 -3.262e-05   4.31e-06     -7.577      0.000   -4.11e-05   -2.42e-05
total_vaccinations_per_thousand 9.567e-05      0.000      0.956      0.339      -0.000       0.000
total_boosters_per_thousand        0.0014      0.000      6.017      0.000       0.001       0.002
==============================================================================
Omnibus:                       78.152   Durbin-Watson:                   2.110
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               92.799
Skew:                           0.700   Prob(JB):                     7.06e-21
Kurtosis:                       2.776   Cond. No.                     1.61e+19
==============================================================================
```

Figure 1: Denmark:OLS Model for number of Covid Cases Summary and Coefficients (Part 1)

```
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 9.38e-27. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
Mean Squared Error: 0.570543513233118
Coefficients:
total_tests_per_thousand           1.771991e-04
hospital_beds_per_thousand        -1.747144e-09
stringency_index                   2.278336e-02
gdp_per_capita                    -3.262443e-05
total_vaccinations_per_thousand    9.567282e-05
total_boosters_per_thousand        1.425863e-03
dtype: float64
```

Figure 2: Denmark :OLS Coefficients Detail for number of Covid Cases (Part 2)

II. OLS Regression Result for Covid New Cases in Finland

```
Country: Finland
Model Summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:     new_cases_smoothed_per_thousand   R-squared:                       0.481
Model:                                         OLS   Adj. R-squared:                  0.479
Method:                              Least Squares   F-statistic:                     255.9
Date:                             Tue, 21 May 2024   Prob (F-statistic):          1.59e-155
Time:                                     07:42:01   Log-Likelihood:                -1283.1
No. Observations:                             1108   AIC:                             2576.
Df Residuals:                                 1103   BIC:                             2601.
Df Model:                                        4
Covariance Type:                         nonrobust
====================================================================================================
                                        coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------------
total_tests_per_thousand              0.0011      0.000      6.377      0.000       0.001       0.001
hospital_beds_per_thousand        -4.983e-09   2.52e-10    -19.755      0.000   -5.48e-09   -4.49e-09
stringency_index                      0.0336      0.003     12.608      0.000       0.028       0.039
gdp_per_capita                    -6.166e-05   3.12e-06    -19.755      0.000   -6.78e-05   -5.55e-05
total_vaccinations_per_thousand       0.0009      0.000      5.423      0.000       0.001       0.001
total_boosters_per_thousand          -0.0027      0.000    -13.916      0.000      -0.003      -0.002
==============================================================================
Omnibus:                      527.635   Durbin-Watson:                   2.001
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               76.295
Skew:                           0.304   Prob(JB):                     2.71e-17
Kurtosis:                       1.867   Cond. No.                     6.70e+18
==============================================================================
```

Figure 3: Finland:OLS Model for number of Covid Cases Summary and Coefficients (Part 1)

```
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 4.07e-26. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
Mean Squared Error: 0.5941776680178193
Coefficients:
total_tests_per_thousand           1.141430e-03
hospital_beds_per_thousand        -4.982855e-09
stringency_index                   3.358438e-02
gdp_per_capita                    -6.165633e-05
total_vaccinations_per_thousand    9.158593e-04
total_boosters_per_thousand       -2.716702e-03
dtype: float64
```

Figure 4: Finland:OLS Coefficients Detail for number of Covid Cases (Part 2)

III. OLS Regression Result for Covid New Cases in Norway

```
Country: Norway
Model Summary:
                            OLS Regression Results
===========================================================================================
Dep. Variable:     new_cases_smoothed_per_thousand   R-squared:                       0.433
Model:                                         OLS   Adj. R-squared:                  0.431
Method:                              Least Squares   F-statistic:                     210.7
Date:                             Tue, 21 May 2024   Prob (F-statistic):          2.65e-134
Time:                                     07:42:01   Log-Likelihood:                -1137.7
No. Observations:                             1108   AIC:                             2285.
Df Residuals:                                 1103   BIC:                             2310.
Df Model:                                        4
Covariance Type:                         nonrobust
===================================================================================================
                                     coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------------------
total_tests_per_thousand           0.0005      0.000      3.642      0.000       0.000       0.001
hospital_beds_per_thousand      -1.545e-09   9.22e-11    -16.763      0.000   -1.73e-09   -1.36e-09
stringency_index                   0.0208      0.002     10.763      0.000       0.017       0.025
gdp_per_capita                  -2.781e-05   1.66e-06    -16.763      0.000   -3.11e-05   -2.46e-05
total_vaccinations_per_thousand    0.0011      0.000      7.482      0.000       0.001       0.001
total_boosters_per_thousand       -0.0037      0.000    -21.082      0.000      -0.004      -0.003
==============================================================================
Omnibus:                      314.464   Durbin-Watson:                   2.070
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              832.229
Skew:                           1.473   Prob(JB):                    1.92e-181
Kurtosis:                       6.058   Cond. No.                     6.83e+18
==============================================================================
```

Figure 5: Norway:OLS Model for number of Covid Cases Summary and Coefficients (Part 1)

```
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 9.99e-26. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
Mean Squared Error: 0.44057591824032083
Coefficients:
total_tests_per_thousand           5.294143e-04
hospital_beds_per_thousand        -1.545242e-09
stringency_index                   2.075982e-02
gdp_per_capita                    -2.781438e-05
total_vaccinations_per_thousand    1.116485e-03
total_boosters_per_thousand       -3.677396e-03
dtype: float64
```

Figure 6: Norway:OLS Coefficients Detail for number of Covid Cases (Part 2)

IV. OLS Regression Result for Covid New Cases in Sweden

```
Country: Sweden
Model Summary:
                                 OLS Regression Results
==========================================================================================
Dep. Variable:     new_cases_smoothed_per_thousand   R-squared:                       0.630
Model:                                         OLS   Adj. R-squared:                  0.628
Method:                              Least Squares   F-statistic:                     469.1
Date:                             Tue, 21 May 2024   Prob (F-statistic):         3.52e-236
Time:                                     07:42:01   Log-Likelihood:               -975.44
No. Observations:                             1108   AIC:                             1961.
Df Residuals:                                 1103   BIC:                             1986.
Df Model:                                        4
Covariance Type:                         nonrobust
==================================================================================================
                                    coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------------
total_tests_per_thousand          0.0023      0.000     16.739      0.000       0.002       0.003
hospital_beds_per_thousand    -1.669e-09   2.43e-10     -6.859      0.000   -2.15e-09   -1.19e-09
stringency_index                  0.0337      0.002     17.546      0.000       0.030       0.037
gdp_per_capita                 -3.53e-05   5.15e-06     -6.859      0.000   -4.54e-05   -2.52e-05
total_vaccinations_per_thousand  -0.0010      0.000     -7.721      0.000      -0.001      -0.001
total_boosters_per_thousand      -0.0006      0.000     -1.888      0.059      -0.001    2.37e-05
==============================================================================
Omnibus:                       34.303   Durbin-Watson:                   2.002
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               67.496
Skew:                           0.187   Prob(JB):                     2.21e-15
Kurtosis:                       4.150   Cond. No.                     1.93e+19
==============================================================================
```

Figure 7: Sweden:OLS Model for number of Covid Cases Summary and Coefficients (Part 1)

```
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 6.57e-27. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
Mean Squared Error: 0.34063747255453347
Coefficients:
total_tests_per_thousand           2.258266e-03
hospital_beds_per_thousand        -1.669099e-09
stringency_index                   3.370298e-02
gdp_per_capita                    -3.529866e-05
total_vaccinations_per_thousand   -9.745576e-04
total_boosters_per_thousand       -6.006648e-04
dtype: float64
```

Figure 8: Sweden:OLS Coefficients Detail for number of Covid Cases (Part 2)

V. OLS Regression Feature Importance for Covid New Cases

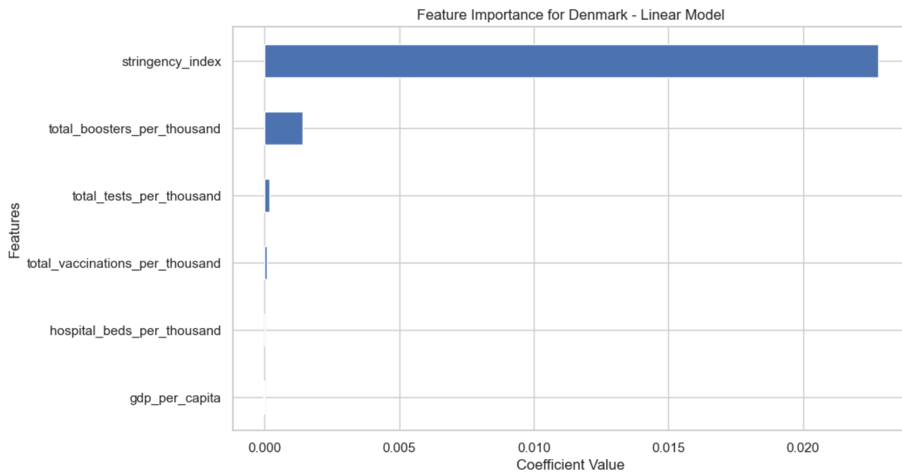

Figure 9: Feature Importance of OLS Model in Denmark for the Number of New Cases

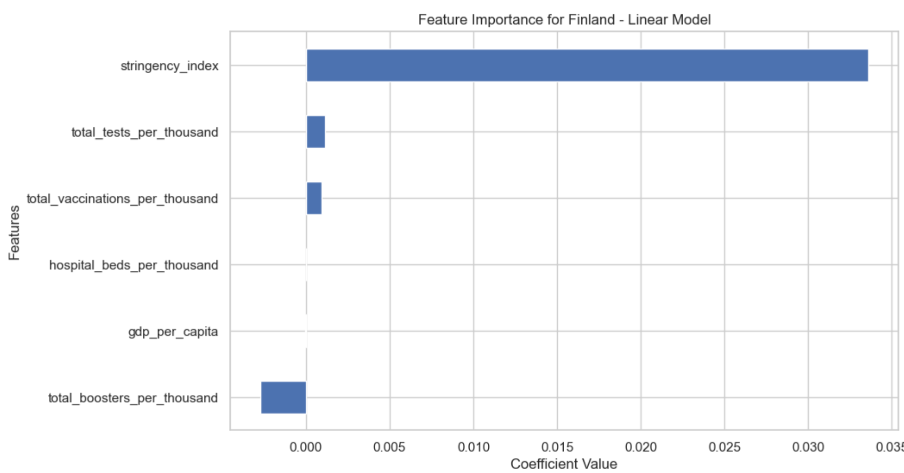

Figure 10: Feature Importance of OLS Model in Finland for the Number of New Cases

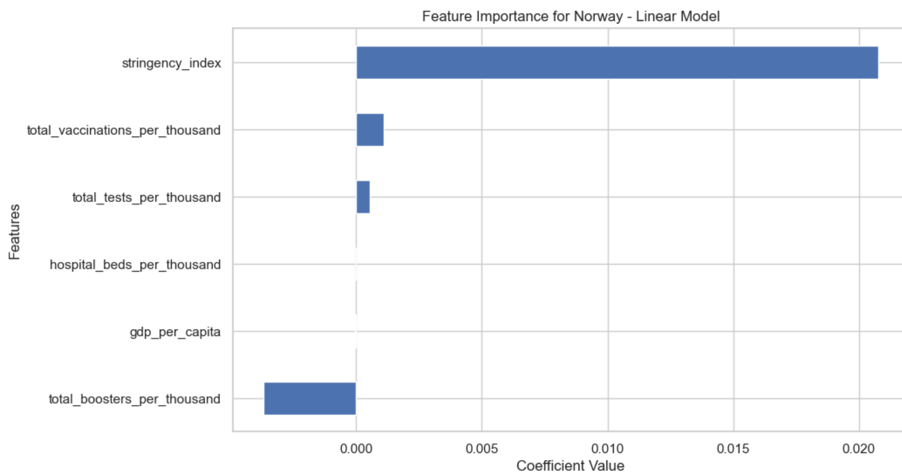

Figure 11: 'Feature Importance of OLS Model in Norway for the Number of New Cases'

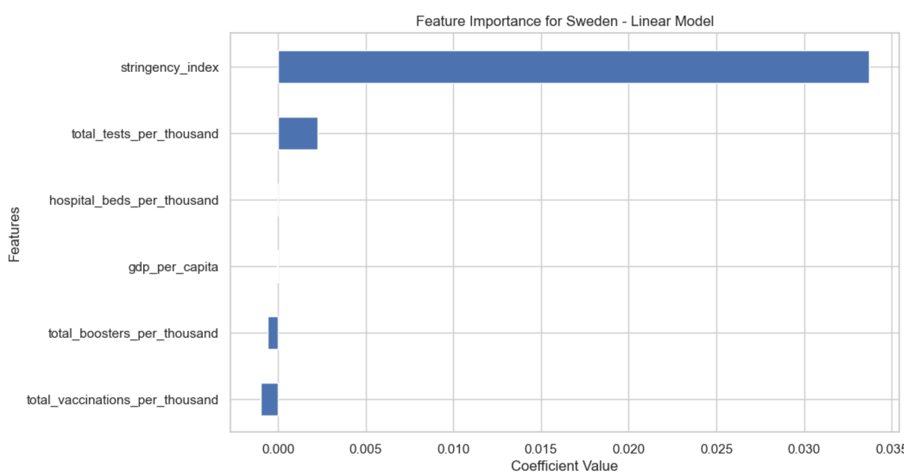

Figure 12: 'Feature Importance of OLS Model in Sweden for the Number of New Cases'

VI. OLS Regression Result for Covid New Deaths in Denmark

```
Country: Denmark
Model Summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:     new_deaths_smoothed_per_thousand   R-squared:                       0.334
Model:                                  OLS   Adj. R-squared:                  0.332
Method:                       Least Squares   F-statistic:                     138.6
Date:                      Tue, 21 May 2024   Prob (F-statistic):           5.49e-96
Time:                              07:43:01   Log-Likelihood:                -1291.8
No. Observations:                      1108   AIC:                             2594.
Df Residuals:                          1103   BIC:                             2619.
Df Model:                                 4
Covariance Type:                  nonrobust
====================================================================================================
                                       coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------------
total_tests_per_thousand          9.876e-05   6.69e-06     14.758      0.000    8.56e-05       0.000
hospital_beds_per_thousand       -2.369e-09   2.36e-10    -10.022      0.000   -2.83e-09   -1.91e-09
stringency_index                     0.0303      0.003      9.009      0.000       0.024       0.037
gdp_per_capita                   -4.424e-05   4.41e-06    -10.022      0.000   -5.29e-05   -3.56e-05
total_vaccinations_per_thousand     -0.0003      0.000     -2.445      0.015      -0.000   -4.96e-05
total_boosters_per_thousand          0.0039      0.000     16.126      0.000       0.003       0.004
==============================================================================
Omnibus:                      122.838   Durbin-Watson:                   2.032
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               52.642
Skew:                           0.338   Prob(JB):                     3.71e-12
Kurtosis:                       2.173   Cond. No.                     1.61e+19
==============================================================================
```

Figure 13: Denmark: OLS Model for Number of Covid Deaths Summary and Coefficients (Part 1)

```
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 9.38e-27. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
Mean Squared Error: 0.6137312654543283
Coefficients:
total_tests_per_thousand           9.876291e-05
hospital_beds_per_thousand        -2.369225e-09
stringency_index                   3.033130e-02
gdp_per_capita                    -4.424056e-05
total_vaccinations_per_thousand   -2.509636e-04
total_boosters_per_thousand        3.918066e-03
dtype: float64
```

Figure 14: Denmark: OLS Coefficients Detail for Number of Covid Deaths (Part 2)

VII. OLS Regression Result for Covid New Deaths in Finland

```
Country: Finland
Model Summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:     new_deaths_smoothed_per_thousand   R-squared:                       0.387
Model:                                          OLS   Adj. R-squared:                  0.385
Method:                               Least Squares   F-statistic:                     174.3
Date:                              Tue, 21 May 2024   Prob (F-statistic):          1.00e-115
Time:                                      07:43:01   Log-Likelihood:                -1383.6
No. Observations:                              1108   AIC:                             2777.
Df Residuals:                                  1103   BIC:                             2802.
Df Model:                                         4
Covariance Type:                          nonrobust
=========================================================================================================
                                      coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------------------------
total_tests_per_thousand            0.0001      0.000      0.688      0.492      -0.000       0.001
hospital_beds_per_thousand      -5.149e-09   2.76e-10    -18.645      0.000   -5.69e-09   -4.61e-09
stringency_index                    0.0384      0.003     13.170      0.000       0.033       0.044
gdp_per_capita                  -6.372e-05   3.42e-06    -18.645      0.000   -7.04e-05    -5.7e-05
total_vaccinations_per_thousand     0.0012      0.000      6.229      0.000       0.001       0.002
total_boosters_per_thousand     -8.122e-05      0.000     -0.380      0.704      -0.001       0.000
==============================================================================
Omnibus:                      209.497   Durbin-Watson:                   2.057
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               63.312
Skew:                          -0.339   Prob(JB):                     1.79e-14
Kurtosis:                       2.046   Cond. No.                     6.70e+18
==============================================================================
```

Figure 15: Finland: OLS Model for Number of Covid Deaths Summary and Coefficients (Part 1)

```
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 4.07e-26. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
Mean Squared Error: 0.6502739972739217
Coefficients:
total_tests_per_thousand           1.347437e-04
hospital_beds_per_thousand        -5.149392e-09
stringency_index                   3.841436e-02
gdp_per_capita                    -6.371700e-05
total_vaccinations_per_thousand    1.151950e-03
total_boosters_per_thousand       -8.121641e-05
dtype: float64
```

Figure 16: Finland: OLS Coefficients Detail for Number of Covid Deaths (Part 2)

VIII. OLS Regression Result for Covid New Deaths in Norway

```
Country: Norway
Model Summary:
                            OLS Regression Results
==========================================================================================
Dep. Variable:     new_deaths_smoothed_per_thousand   R-squared:                       0.124
Model:                                          OLS   Adj. R-squared:                  0.121
Method:                               Least Squares   F-statistic:                     39.04
Date:                              Tue, 21 May 2024   Prob (F-statistic):           1.33e-30
Time:                                      07:43:01   Log-Likelihood:                -1316.3
No. Observations:                              1108   AIC:                             2643.
Df Residuals:                                  1103   BIC:                             2668.
Df Model:                                         4
Covariance Type:                          nonrobust
===============================================================================================
                                  coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------------
total_tests_per_thousand        0.0004      0.000      2.071      0.039    1.86e-05      0.001
hospital_beds_per_thousand  -1.261e-09   1.08e-10    -11.643      0.000   -1.47e-09   -1.05e-09
stringency_index                0.0107      0.002      4.730      0.000       0.006      0.015
gdp_per_capita               -2.27e-05   1.95e-06    -11.643      0.000   -2.65e-05   -1.89e-05
total_vaccinations_per_thousand 0.0004      0.000      2.149      0.032    3.28e-05      0.001
total_boosters_per_thousand    -0.0006      0.000     -3.121      0.002      -0.001     -0.000
==============================================================================
Omnibus:                       71.363   Durbin-Watson:                   2.052
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               80.495
Skew:                           0.640   Prob(JB):                     3.32e-18
Kurtosis:                       2.676   Cond. No.                     6.83e+18
==============================================================================
```

Figure 17: Norway: OLS Model for Number of Covid Deaths Summary and Coefficients (Part 1)

```
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 9.99e-26. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
Mean Squared Error: 0.6201814673320226
Coefficients:
total_tests_per_thousand           3.537577e-04
hospital_beds_per_thousand        -1.261062e-09
stringency_index                   1.071912e-02
gdp_per_capita                    -2.269913e-05
total_vaccinations_per_thousand    3.768487e-04
total_boosters_per_thousand       -6.395938e-04
dtype: float64
```

Figure 18: Norway: OLS Coefficients Detail for Number of Covid Deaths (Part 2)

IX. OLS Regression Result for Covid New Deaths in Sweden

```
Country: Sweden
Model Summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:     new_deaths_smoothed_per_thousand   R-squared:                       0.265
Model:                                          OLS   Adj. R-squared:                  0.262
Method:                               Least Squares   F-statistic:                     99.38
Date:                              Tue, 21 May 2024   Prob (F-statistic):           2.83e-72
Time:                                      07:43:01   Log-Likelihood:                -1417.0
No. Observations:                              1108   AIC:                             2844.
Df Residuals:                                  1103   BIC:                             2869.
Df Model:                                         4
Covariance Type:                          nonrobust
====================================================================================================
                                      coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------------
total_tests_per_thousand           -0.0007      0.000     -3.628      0.000      -0.001      -0.000
hospital_beds_per_thousand      -1.485e-09   3.63e-10     -4.096      0.000    -2.2e-09   -7.74e-10
stringency_index                    0.0505      0.003     17.642      0.000       0.045       0.056
gdp_per_capita                   -3.14e-05   7.67e-06     -4.096      0.000   -4.64e-05   -1.64e-05
total_vaccinations_per_thousand     0.0015      0.000      8.004      0.000       0.001       0.002
total_boosters_per_thousand        -0.0012      0.000     -2.488      0.013      -0.002      -0.000
==============================================================================
Omnibus:                       90.000   Durbin-Watson:                   2.034
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               30.479
Skew:                           0.080   Prob(JB):                     2.41e-07
Kurtosis:                       2.204   Cond. No.                     1.93e+19
==============================================================================
```

Figure 19: Sweden: OLS Model for Number of Covid Deaths Summary and Coefficients (Part 1)

```
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 6.57e-27. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
Mean Squared Error: 0.8136636449697074
Coefficients:
total_tests_per_thousand          -7.291326e-04
hospital_beds_per_thousand        -1.484939e-09
stringency_index                   5.048220e-02
gdp_per_capita                    -3.140398e-05
total_vaccinations_per_thousand    1.504855e-03
total_boosters_per_thousand       -1.179306e-03
dtype: float64
```

Figure 20: Sweden: OLS Coefficients Detail for Number of Covid Deaths (Part 2)

X. OLS Regression Feature Importance for Covid New Deaths

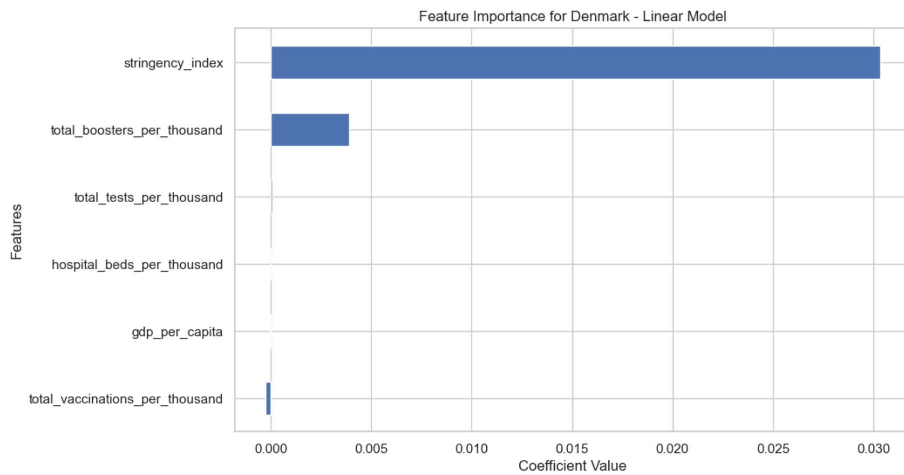

Figure 21: 'Feature Importance of OLS Model in Denmark for the Number of New Deaths'

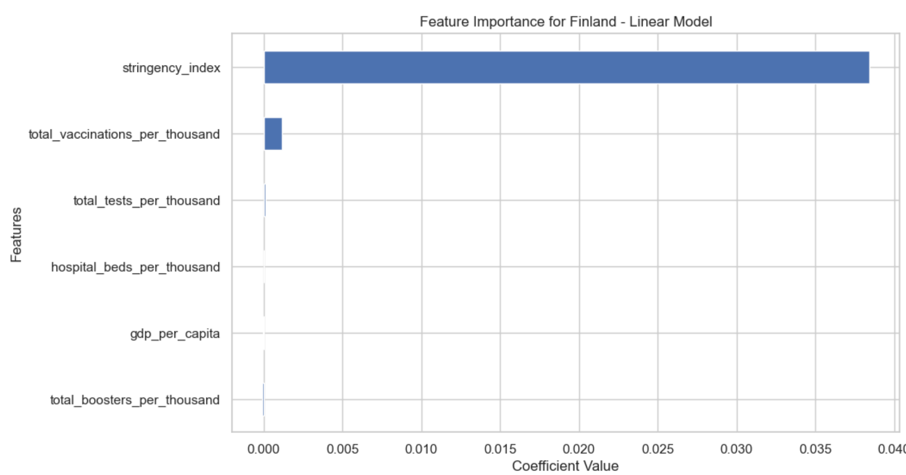

Figure 22: 'Feature Importance of OLS Model in Finland for the Number of New Deaths'

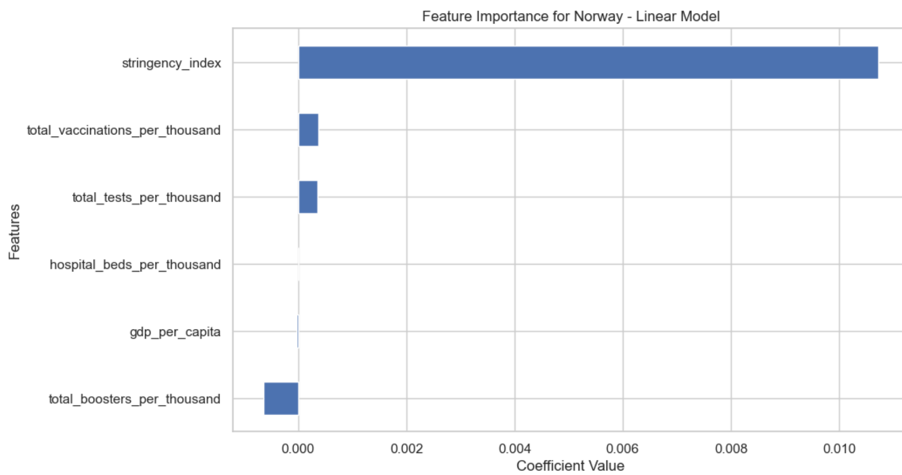

Figure 23: 'Feature Importance of OLS Model in Norway for the Number of New Deaths'

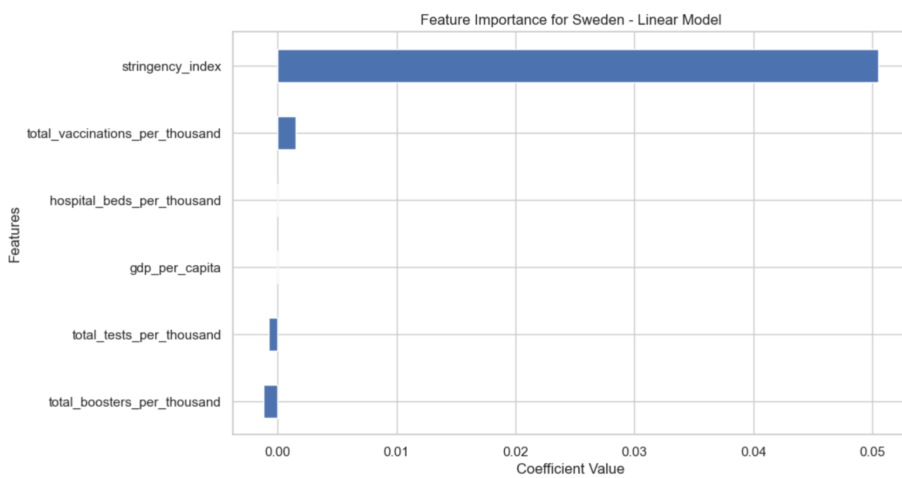

Figure 24: 'Feature Importance of OLS Model in Sweden for the Number of New Deaths'

XI. Random Forest Model Results for the Number of COVID Cases in Each of the Scandinavian Countries

```
Country: Denmark
Random Forest MSE: 0.010961906730313 73
Random Forest R-squared: 0.9897738111222808
Feature Importance:
                                 importance
total_tests_per_thousand           0.395111
total_vaccinations_per_thousand    0.325895
total_boosters_per_thousand        0.262285
stringency_index                   0.016709
hospital_beds_per_thousand         0.000000
gdp_per_capita                     0.000000
```

Figure 25: Random Forest Model Results for the Number of COVID Cases in Denmark

```
Country: Finland
Random Forest MSE: 0.011363121037163968
Random Forest R-squared: 0.9890648878619096
Feature Importance:
                                 importance
total_boosters_per_thousand        0.369582
total_tests_per_thousand           0.339774
stringency_index                   0.250760
total_vaccinations_per_thousand    0.039884
hospital_beds_per_thousand         0.000000
gdp_per_capita                     0.000000
```

Figure 26: Random Forest Model Results for the Number of COVID Cases in Finland

```
Country: Norway
Random Forest MSE: 0.01301090465273621
Random Forest R-squared: 0.9832792959720845
Feature Importance:
                                  importance
total_vaccinations_per_thousand     0.416313
total_tests_per_thousand            0.334853
total_boosters_per_thousand         0.229705
stringency_index                    0.019128
hospital_beds_per_thousand          0.000000
gdp_per_capita                      0.000000
```

Figure 27: Random Forest Model Results for the Number of COVID Cases in Norway

```
Country: Sweden
Random Forest MSE: 0.0051406364436877554
Random Forest R-squared: 0.9945774848351111
Feature Importance:
                                  importance
stringency_index                    0.653916
total_tests_per_thousand            0.251585
total_vaccinations_per_thousand     0.067511
total_boosters_per_thousand         0.026988
hospital_beds_per_thousand          0.000000
gdp_per_capita                      0.000000
```

Figure 28: Random Forest Model Results for the Number of COVID Cases in Sweden

XII. SVR Model Results for the Number of COVID Cases in Each of the Scandinavian Countries

```
Country: Denmark
Permutation Feature Importance:
                                  importance_mean  importance_std
stringency_index                        14.613830        0.677829
total_vaccinations_per_thousand          2.901465        0.153297
total_boosters_per_thousand              2.399659        0.124799
total_tests_per_thousand                 1.613949        0.083834
hospital_beds_per_thousand               0.000000        0.000000
gdp_per_capita                           0.000000        0.000000
Processing countries:  50%|█████     | 2/4 [00:28<00:26, 13.23s/it]

Country: Finland
Permutation Feature Importance:
                                  importance_mean  importance_std
total_boosters_per_thousand              2.308593        0.109834
total_vaccinations_per_thousand          1.654640        0.063427
total_tests_per_thousand                 1.446501        0.079811
stringency_index                         0.957048        0.050407
hospital_beds_per_thousand               0.000000        0.000000
gdp_per_capita                           0.000000        0.000000
Processing countries:  75%|███████▌  | 3/4 [00:44<00:14, 14.35s/it]
```

Figure 29: Support Vector Machine (SVM) results for COVID-19 cases in Denmark and Finland.

```
Country: Norway
Permutation Feature Importance:
                                  importance_mean  importance_std
total_tests_per_thousand                13.472765        0.428766
total_vaccinations_per_thousand          1.549968        0.033902
total_boosters_per_thousand              1.177950        0.099795
stringency_index                         0.861064        0.050155
hospital_beds_per_thousand               0.000000        0.000000
gdp_per_capita                           0.000000        0.000000
Processing countries: 100%|██████████| 4/4 [00:54<00:00, 13.53s/it]

Country: Sweden
Permutation Feature Importance:
                                  importance_mean  importance_std
stringency_index                         2.589713        0.084098
total_tests_per_thousand                 2.226373        0.133703
total_boosters_per_thousand              1.393001        0.106876
total_vaccinations_per_thousand          0.620338        0.045320
hospital_beds_per_thousand               0.000000        0.000000
gdp_per_capita                           0.000000        0.000000
```

Figure 30: Support Vector Machine (SVM) results for COVID-19 cases in Norway and Sweden.

XIII. Random Forest Model Results for the Number of COVID Deaths in Each of the Scandinavian Countries

```
Country: Denmark
Random Forest MSE: 0.01313503244740504
Random Forest R-squared: 0.9874819262072623
Best Parameters: {'max_depth': 20, 'min_samples_leaf': 1, 'n_estimators': 200}
Feature Importance:
                                 importance
stringency_index                   0.425914
total_vaccinations_per_thousand    0.255245
total_tests_per_thousand           0.195389
total_boosters_per_thousand        0.123452
hospital_beds_per_thousand         0.000000
gdp_per_capita                     0.000000

Country: Finland
Random Forest MSE: 0.01313503244740504
Random Forest R-squared: 0.9874819262072623
Best Parameters: {'max_depth': 20, 'min_samples_leaf': 1, 'n_estimators': 200}
Feature Importance:
                                 importance
stringency_index                   0.425914
total_vaccinations_per_thousand    0.255245
total_tests_per_thousand           0.195389
total_boosters_per_thousand        0.123452
hospital_beds_per_thousand         0.000000
gdp_per_capita                     0.000000
```

Figure 31: Random Forest results for COVID-19 death rates in Denmark and Finland

```
Country: Norway
Random Forest MSE: 0.01313503244740504
Random Forest R-squared: 0.9874819262072623
Best Parameters: {'max_depth': 20, 'min_samples_leaf': 1, 'n_estimators': 200}
Feature Importance:
                                 importance
stringency_index                   0.425914
total_vaccinations_per_thousand    0.255245
total_tests_per_thousand           0.195389
total_boosters_per_thousand        0.123452
hospital_beds_per_thousand         0.000000
gdp_per_capita                     0.000000

Country: Sweden
Random Forest MSE: 0.01313503244740504
Random Forest R-squared: 0.9874819262072623
Best Parameters: {'max_depth': 20, 'min_samples_leaf': 1, 'n_estimators': 200}
Feature Importance:
                                 importance
stringency_index                   0.425914
total_vaccinations_per_thousand    0.255245
total_tests_per_thousand           0.195389
total_boosters_per_thousand        0.123452
hospital_beds_per_thousand         0.000000
gdp_per_capita                     0.000000
```

Figure 32: Random Forest results for COVID-19 death rates in Norway and Sweden

XIV. Stringency Index Plots in the Scandinavian Countries

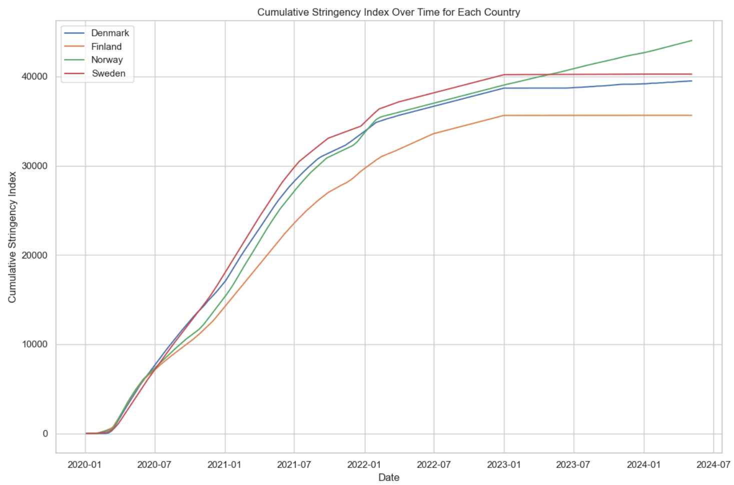

Figure 33: Cumulative Stringency Index Over Time for Each Country.

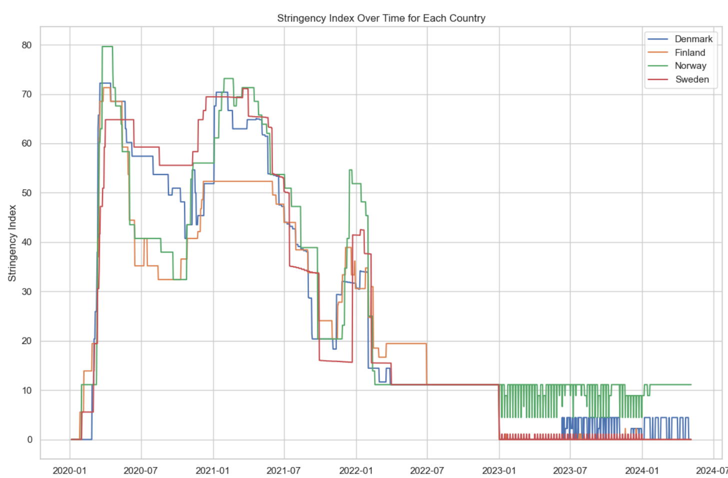

Figure 34: Stringency Index Over Time for Each Scandinavian Country.

XV. The New Covid Cases and Deaths (smoothed per thousand) Plots in each of the Scandinavian Countries

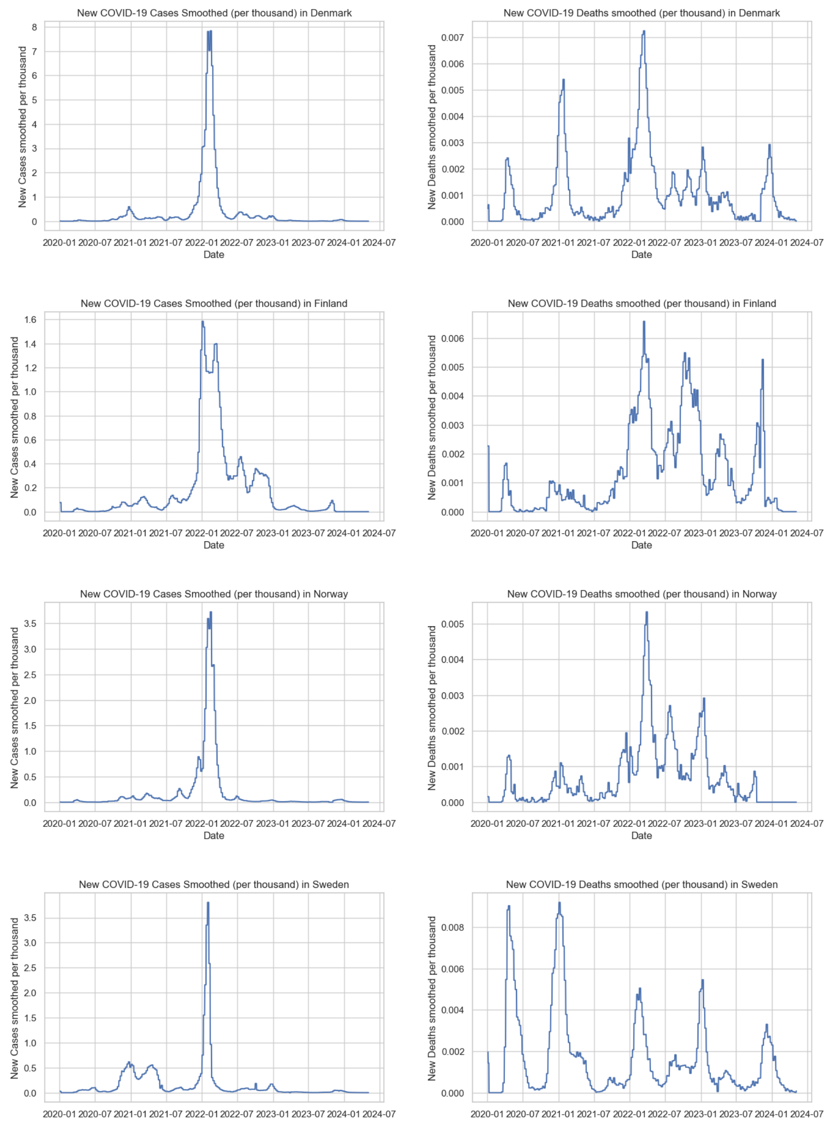

Figure 35: New COVID-19 Cases and Deaths Smoothed per Thousand in Denmark, Finland, Norway, and Sweden.

XVI. The distribution of the Number of New Covid Cases in Each of The Scandinavian Countries

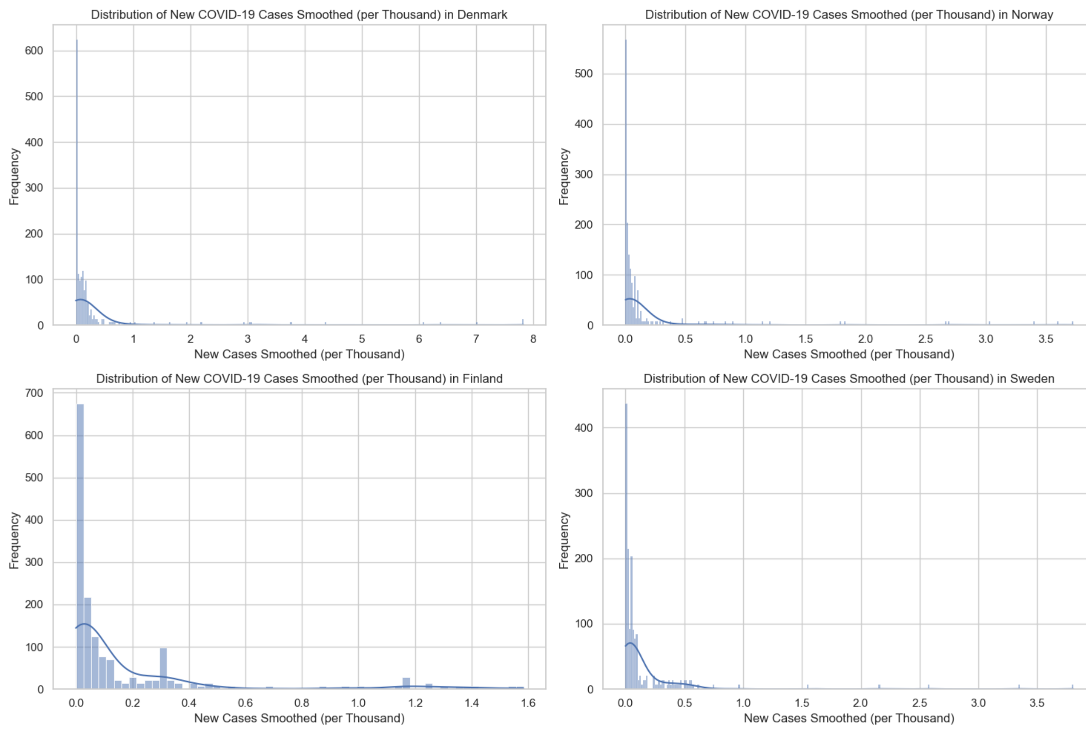

Figure 36: Distribution of COVID-19 New Cases (smoothed per thousand) in Denmark, Finland, Norway, and Sweden.

XVII. The distribution of the Number of New Covid Deaths in Each of The Scandinavian Countries

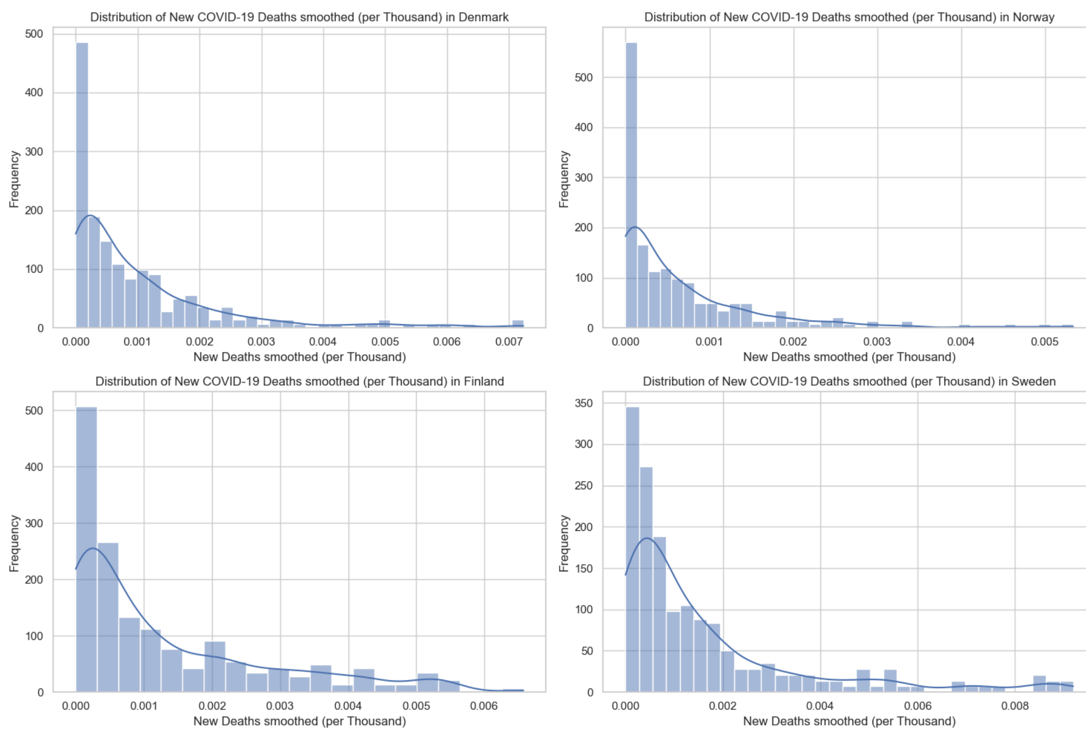

Figure 37: Distribution of COVID-19 New Deaths (smoothed per thousand) in Denmark, Finland, Norway, and Sweden

XVIII. The Log-Normal Transformation of The Number of Covid New Cases and Deaths

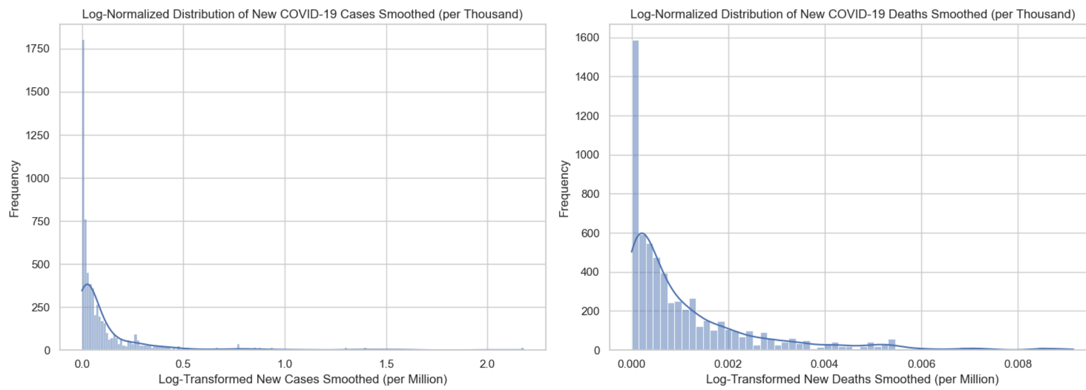

Figure 38: Log-Normalized Distribution of New COVID-19 Cases and New Covid-19 Deaths Smoothed (per Thousand)

XIX. The Box-Cox Transformation of The Number of Covid New Cases and Deaths

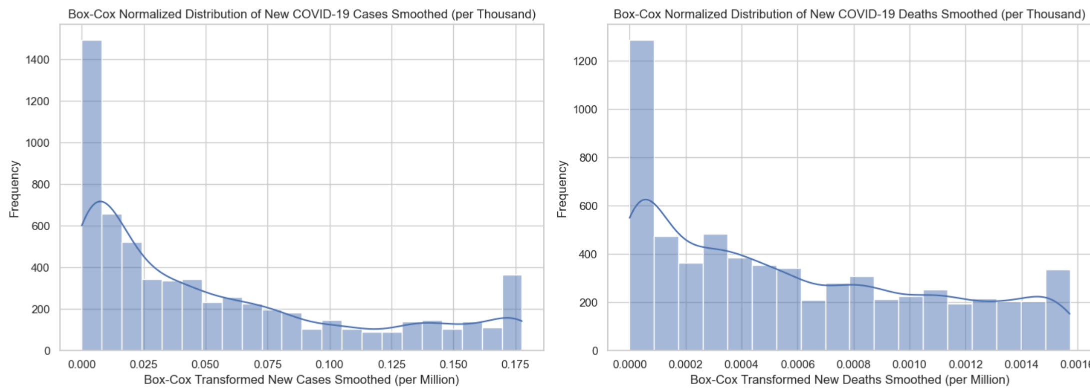

Figure 39: Box-Cox Transformed Distribution of New COVID-19 Cases and New Covid-19 Deaths Smoothed (per Thousand)

XX. The Yeo-Johnson Transformation of The Number of Covid New Cases and Deaths

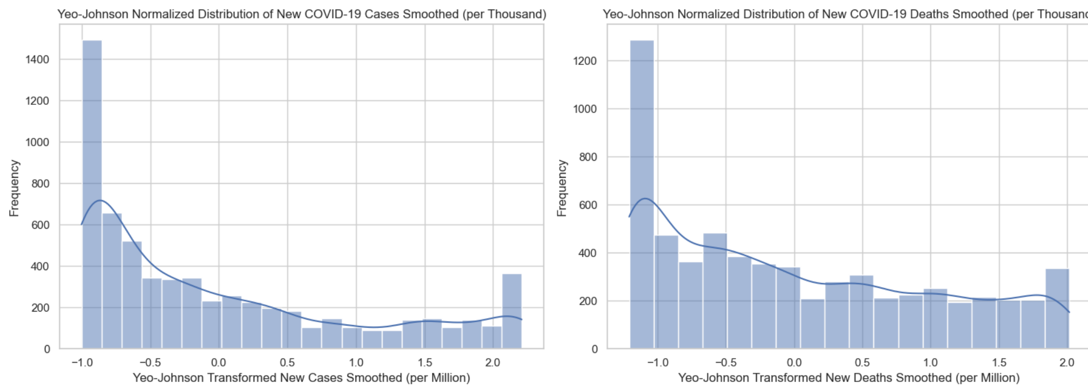

Figure 40: Yeo-Johnson Transformed Distribution of New COVID-19 Cases and New Covid-19 Deaths Smoothed (per Thousand)

XXI. OLS Regression results Comparison between The number of New Cases and Deaths for each Coefficient

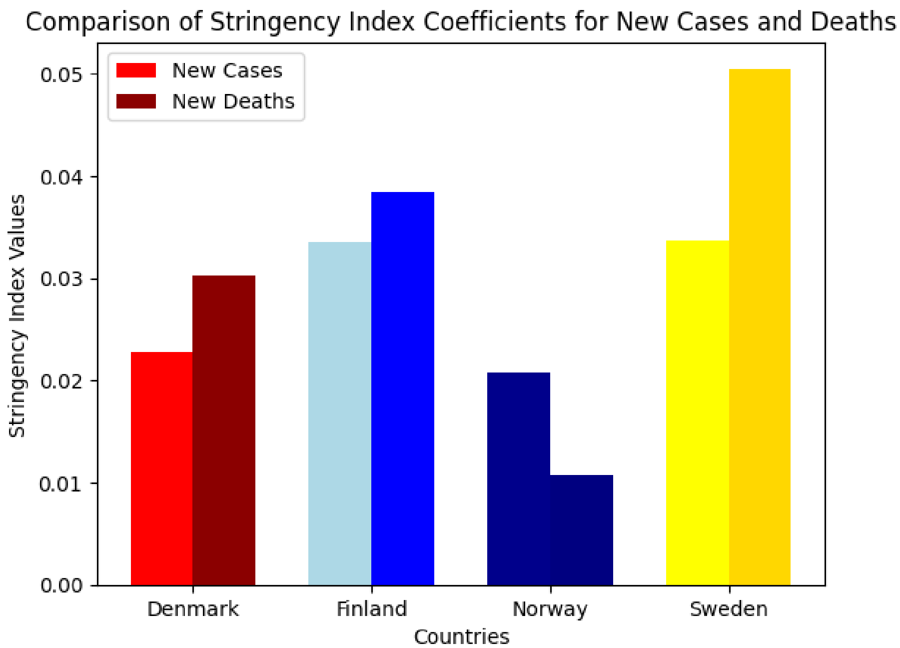

Figure 41: Comparison of Stringency Index Coefficients for New Cases and Deaths

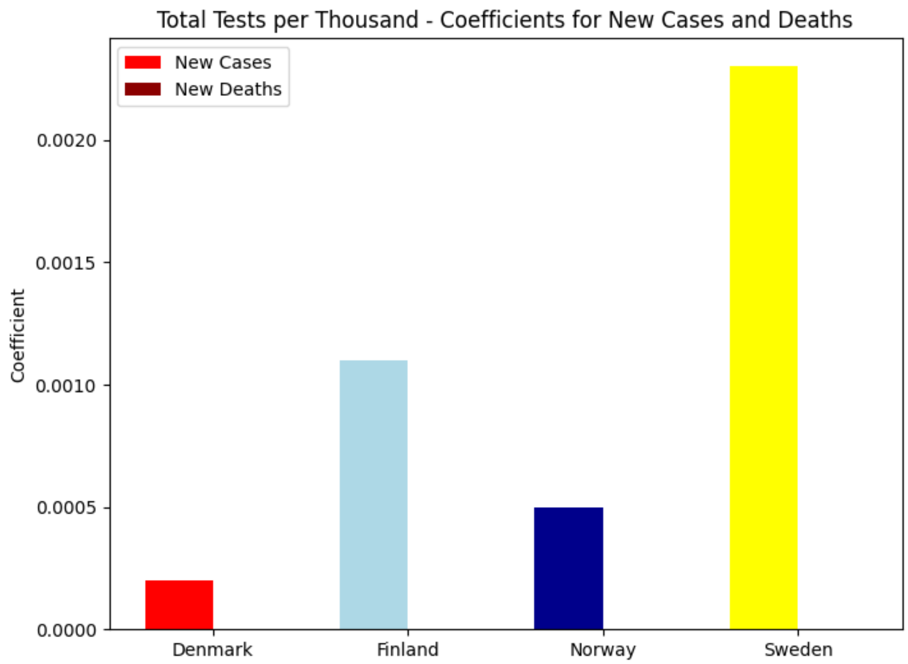

Figure 42: Coefficients for New Cases and Deaths - Total Tests per Thousand

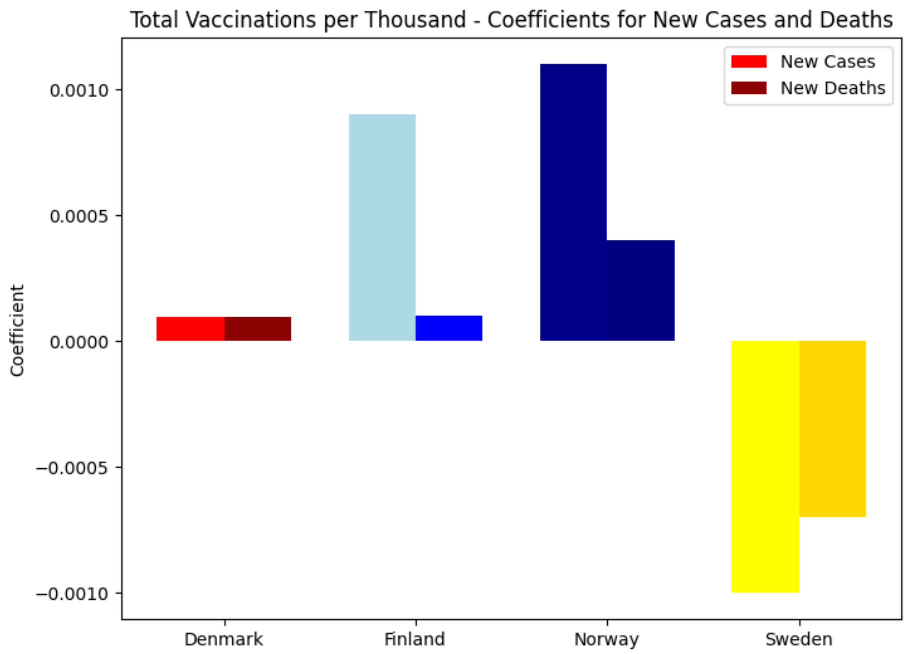

Figure 43: Coefficients for New Cases and Deaths - Total Vaccination per Thousand

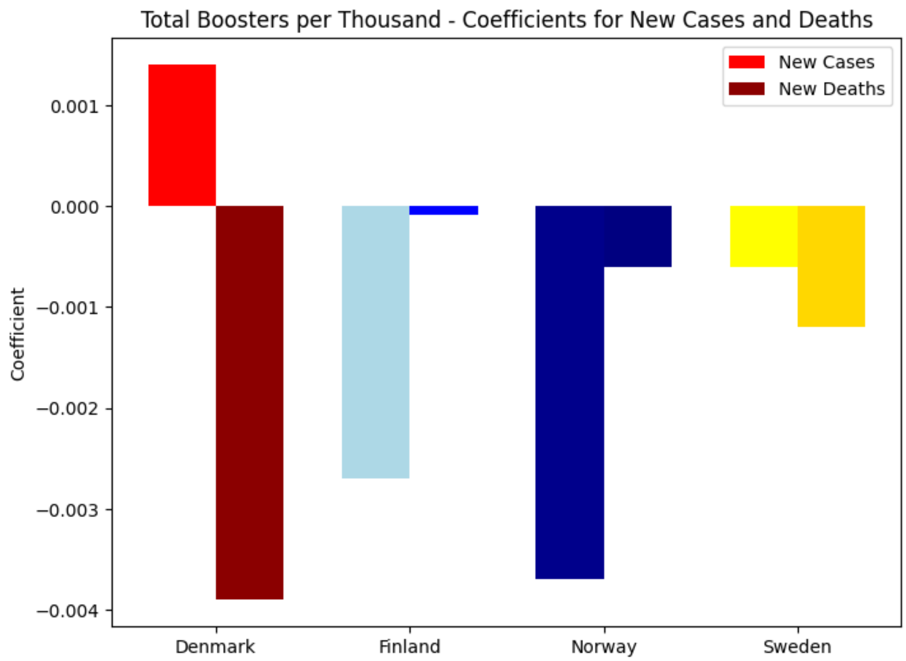

Figure 44: Coefficients for New Cases and Deaths - Total Boosters per Thousand

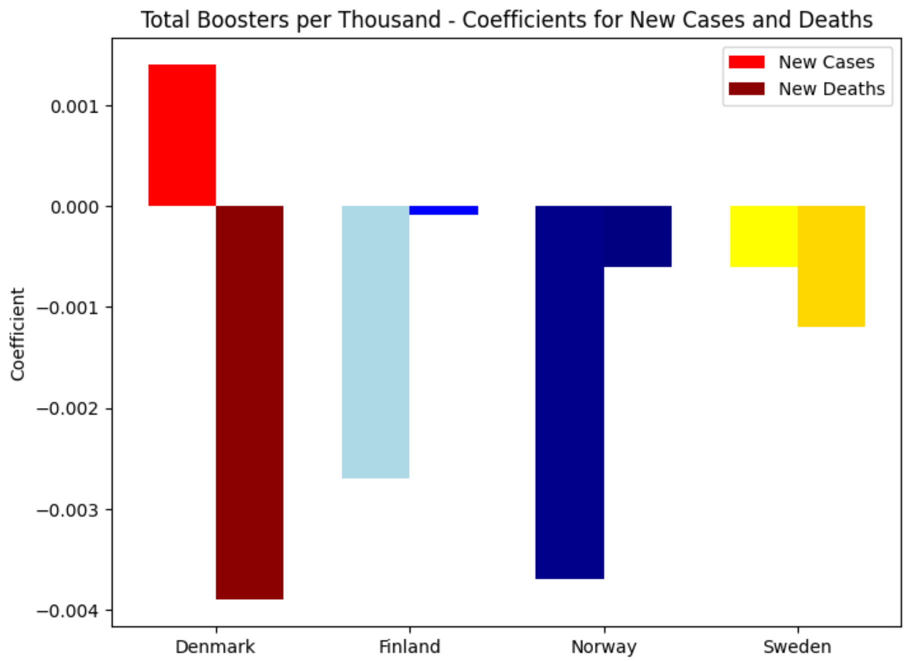

Figure 45: Coefficients for New Cases and Deaths - Total Boosters per Thousand

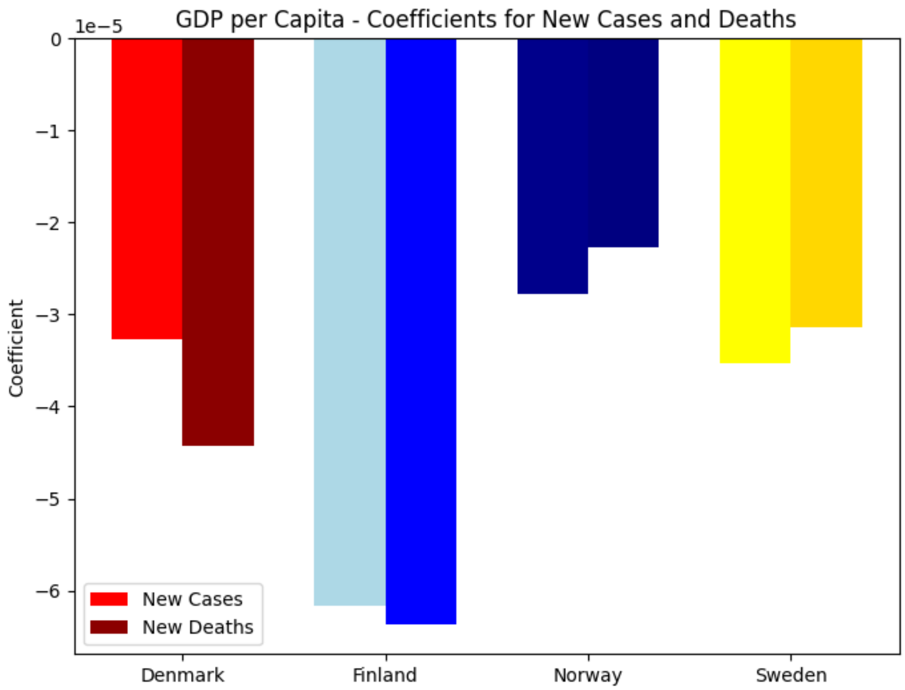

Figure 46: Coefficients for New Cases and Deaths - GDP per Capita

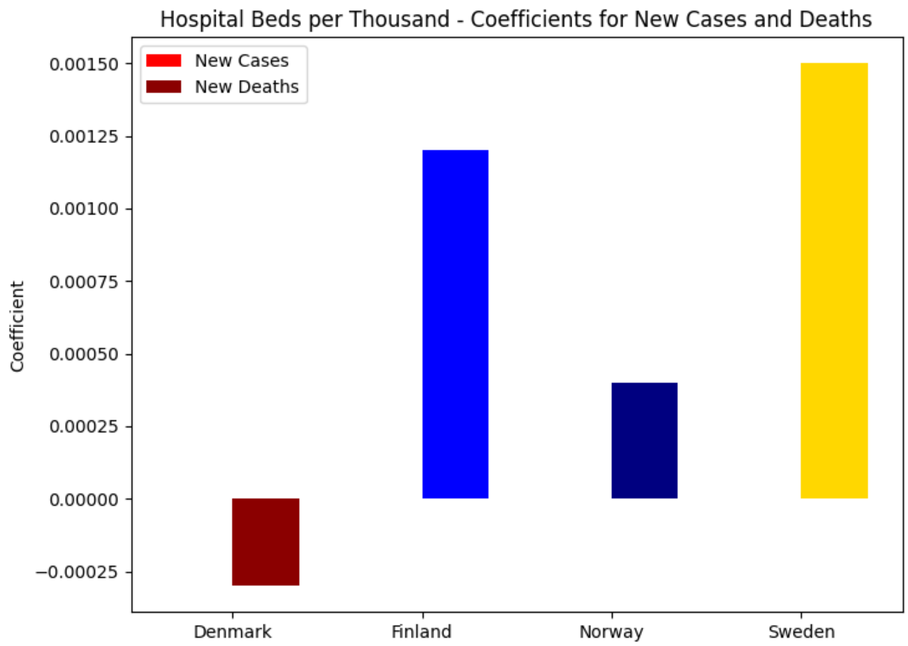

Figure 47: Coefficients for New Cases and Deaths - Hospital Beds per Thousand

XXII. Random Forest results feature importance Comparison between The four Countries for the Number of New Covid Cases

Feature	Denmark	Finland	Norway	Sweden
Total Tests per Thousand	0.427827	0.333264	0.323189	0.258239
Total Vaccinations per Thousand	0.284460	0.057090	0.421269	0.063641
Total Boosters per Thousand	0.269400	0.362449	0.236587	0.030535
Stringency Index	0.018313	0.247197	0.018955	0.647585
Hospital Beds per Thousand	0.000000	0.000000	0.000000	0.000000
GDP per Capita	0.000000	0.000000	0.000000	0.000000

Table 1: Feature Importance Values from Random Forest Analysis for Predicting New COVID-19 Cases Per Thousand in Scandinavian Countries.

XXIII. Random Forest results feature importance Comparison between The four Countries for the Number of New Covid Deaths

Feature	Denmark	Finland	Norway	Sweden
Stringency Index	0.425914	0.425914	0.425914	0.425914
Total Vaccinations per Thousand	0.255245	0.255245	0.255245	0.255245
Total Tests per Thousand	0.195389	0.195389	0.195389	0.195389
Total Boosters per Thousand	0.123452	0.123452	0.123452	0.123452
Hospital Beds per Thousand	0.000000	0.000000	0.000000	0.000000
GDP per Capita	0.000000	0.000000	0.000000	0.000000

Table 2: Feature Importance Values from Random Forest Analysis for Predicting New COVID-19 Deaths Per Thousand in Scandinavian Countries.

XXIV. SVR results feature importance Comparison between The four Countries for the Number of New Covid Cases

Feature	Denmark	Finland	Norway	Sweden
Total Boosters per Thousand	2.419003	2.419003	2.419003	2.419003
Stringency Index	2.281651	2.281651	2.281651	2.281651
Total Tests per Thousand	1.991590	1.991590	1.991590	1.991590
Total Vaccinations per Thousand	0.573666	0.573666	0.573666	0.573666

Table 3: Mean permutation feature importance values from SVR, predicting the number of Covid deaths in Scandinavian Countries.

XXV. SVR results feature importance Comparison between The four Countries for the Number of New Covid Deaths

Feature	Denmark	Finland	Norway	Sweden
Total Boosters per Thousand	2.419003	2.419003	2.419003	2.419003
Stringency Index	2.281651	2.281651	2.281651	2.281651
Total Tests per Thousand	1.991590	1.991590	1.991590	1.991590
Total Vaccinations per Thousand	0.573666	0.573666	0.573666	0.573666

Table 4: Mean permutation feature importance values from SVR, predicting the number of Covid deaths in Scandinavian Countries.