# A PROJECT REPORT

**on**

**Covid-19 impact analysis**
**Project ID-18**

**Submitted to**

## KIIT Deemed to be University

**In Partial Fulfillment of the Requirement for the Award of**

**BACHELOR'S DEGREE IN**
**INFORMATION TECHNOLOGY**

**BY**
**Group ID- 2**


| | |
|---|---|
| **Rabneet Singh Nanhra** | 2105221 |
| **Raul Das** | 2105224 |
| **Debanandan Pradhan** | 2105272 |
| **Soumyadeep Saha** | 2105246 |
| **Ankush Bandyopadhyay** | 2105444 |

**UNDER THE GUIDANCE OF**
**Mr.Prabhu Prasad Dev (Assistant Professor)**

**SCHOOL OF COMPUTER ENGINEERING**
**KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY**

**BHUBANESWAR, ODISHA - 751024**
**May 2024**

# Acknowledgement

We are profoundly grateful to **Mr.Prabhu Prasad Dev (Assistant Professor)** for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion.

Rabneet Singh Nanhra
Raul Das
Debanandan Pradhan
Soumyadeep Saha
Ankush Bandyopadhyay

# ABSTRACT

The global prevalence of COVID-19 patients is rapidly increasing, creating a significant challenge in providing appropriate treatment and care. A key priority during the pandemic is to identify clinical indicators that can predict the progression of the disease to severe conditions. This is essential for effectively stratifying risk and allocating resources optimally.To triage patients effectively, it is crucial to classify the severity of COVID-19 infections into different levels, ranging from mild to critical, based on their symptoms. Various symptomatic parameters play a pivotal role in assessing the severity of the infection. Given the rapid spread and transmissibility of COVID-19, the implementation of telemonitoring systems for COVID-19 patients has become essential. These systems facilitate remote data exchange among healthcare providers, patients, and stakeholders. By enabling continuous remote monitoring, telemonitoring interventions can help mitigate risk and provide timely medical interventions. In this paper, we present an exploratory data analysis of symptoms, comorbidities, and other relevant parameters related to COVID-19. Additionally, we evaluate different machine learning algorithms for detecting case severity. This analysis aims to identify patterns and correlations that can assist in predicting the severity of the disease. We propose a truthfulness-based system for detecting the severity of COVID-19 cases. This system utilizes various indicators and parameters to accurately stratify the risk levels for anticipated moderate and severe COVID-19 patients.Finally, we outline a telemonitoring model for COVID-19 patients which ensures remote and continuous monitoring of case severity progression. This model facilitates the implementation of appropriate risk mitigation strategies and timely interventions based on the evolving severity of the disease.

In summary, this research paper focuses on exploring data related to symptoms, comorbidities, and other relevant parameters of COVID-19. It evaluates different machine learning algorithms for case severity detection and proposes a truthfulness-based system for risk stratification. Additionally, a telemonitoring model is outlined to enable remote and continuous monitoring of COVID-19 patients, aiding in the implementation of appropriate risk mitigation strategies.

# Contents



\

# Chapter 1

# Introduction

In December 2019, the world witnessed the emergence of a novel corona virus in Wuhan, China, sparking a global pandemic that has since affected millions of lives worldwide. This new virus, named SARS-CoV-2, belongs to the same family of viruses as severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS). Its rapid spread, primarily through respiratory droplets, has led to a surge in severe cases and fatalities across the globe.

Under the electron microscope, SARS-CoV-2 exhibits a distinctive crown-like appearance, attributed to the presence of glycoprotein spikes on its surface. As a member of the beta-corona virus genus, it is believed to have originated from animal reservoirs, potentially bats or other animals, before jumping to humans. The virus manifests with a broad spectrum of symptoms, ranging from mild respiratory discomfort to life-threatening acute respiratory distress syndrome (ARDS).

Efforts to mitigate the impact of COVID-19 focus on understanding the factors influencing disease severity. While age and underlying health conditions are known to increase susceptibility to severe illness, the intricate interplay between various comorbidities adds complexity to the disease progression. Researchers strive to unravel these complexities to devise more effective diagnostic and therapeutic interventions.

Exploratory data analysis (EDA) and machine learning (ML) algorithms serve as valuable tools in uncovering patterns and predicting disease outcomes. By scrutinizing patient symptoms and comorbidities, researchers can develop sophisticated decision support systems to categorize disease severity and monitor patient health remotely. These advancements hold promise for enhancing our understanding of COVID-19 and improving patient care strategies in the face of this global health crisis.

# Chapter 2

# Literature Review

With the quickly evolving information, the author generates a scope of case casualty proportion gauges for novel coronavirus that contrasts considerably to a great extent. The author intends to give robust evaluations, representing ascertainment inclinations and censoring. The individual case information was utilised to investigate the time between the beginning of comorbidities/symptoms and the result (discharge or death). The authors (Verity et al., 2019) got age-delineated evaluations of the case casualty proportion by relating the total case distribution to the detected aggregate death rate in China, considering a steady age-based attack rate and adjustment for demography and area, and age-based under ascertainment. The authors evaluated the case fatality proportion from the line-list information (on the individual basis) on 1334-cases recognised from China's exterior. Utilising information on the predominance of PCR-affirmed cases (worldwide occupants) who visited China, they acquired age-delineated assessments of the disease fatality proportion. Moreover, information on age-separated seriousness in a subset of 3665-cases from China was utilised to evaluate the extent of COVID-19 victims, who will probably be hospitalised.

# Chapter 3

# Methodology

Methodology: Exploring the Relationship Between COVID-19 Spread and Happiness (using XML Data)
This project investigates the potential link between the spread of COVID-19 and national happiness levels. Python will be our primary tool for data acquisition (assumed to be completed), manipulation, analysis, and visualization.

Data Sources:
Confirmed COVID-19 Cases: We'll utilize a publicly available dataset on confirmed COVID-19 cases by country, potentially from Johns Hopkins University ([Resource 1]). This dataset is expected to be in XML format (covid19_Confirmed_dataset.xml).
Resource 1: Johns Hopkins University COVID-19 data repository can potentially be found through a web search for "[Johns Hopkins University COVID-19 data]"

National Happiness Scores: We'll employ data on national happiness scores, potentially obtained from the World Happiness Report ([Resource 2]). This dataset is also expected to be in XML format (worldwide_happiness_report.xml).
Resource 2: World Happiness Report website can be found through a web search for "[World Happiness Report]"

Jupyter Notebook: The analysis will be conducted within a Jupyter notebook source file named
Covid19_data_analysis_notebook.ipynb.
Data Preprocessing:
Loading Data: We'll use libraries like xml.etree.ElementTree or pandas.read_xml() to parse and load the XML datasets into Pandas DataFrames.

Merging Datasets: The key step will be merging the confirmed cases and happiness data based on a common identifier, likely "Country" or "Country/Region". We'll achieve this using the pandas.merge() function.

Data Cleaning: We'll explore the merged dataset for missing values, inconsistencies (e.g., invalid data types), and outliers. Techniques like imputation, removal, or winsorization might be applied to address these issues.

Feature Engineering: We might consider creating new features based on the existing data, such as daily case growth rate or cumulative confirmed cases.

Data Analysis:

Exploratory Data Analysis (EDA): We'll employ various visualization techniques (histograms, scatter plots) to understand the distribution of variables and examine potential relationships between confirmed COVID-19 cases and happiness scores. Descriptive statistics (mean, median, standard deviation) will also be used to summarize the data.

Correlation Analysis: We'll calculate correlation coefficients (e.g., Pearson correlation) to assess the strength and direction of the linear relationship between COVID-19 metrics (confirmed cases, death rate, daily case growth) and national happiness scores.

Hypothesis Testing: We might conduct statistical hypothesis tests (e.g., t-tests) to compare happiness scores between groups with high vs. low COVID-19 cases. This will help us determine if there's a statistically significant difference in happiness levels based on COVID-19 prevalence.

Data Visualization:

Interactive Visualizations: We can utilize libraries like Plotly or Bokeh to create interactive visualizations within the Jupyter notebook, allowing users to explore trends dynamically. (Optional, based on project scope)

Static Visualizations: We'll generate static plots (scatter plots, heatmaps, bar charts) using Matplotlib or Seaborn to depict key findings from the analysis.

Ethical Considerations:
We'll acknowledge any limitations of the data and potential biases that might influence the results.

We'll discuss data privacy and anonymity if the datasets contain personally identifiable information (PII). We'll avoid disclosing such information.

Expected Outcomes:
This analysis aims to reveal any potential correlation between COVID-19 spread and national happiness levels.

The findings will be presented in a clear and concise manner, highlighting statistical results and visualizations. We'll acknowledge limitations of the study and suggest potential areas for further investigation.

Note: This methodology provides a general framework. Specific implementations and analyses might be adjusted based on the chosen datasets, the structure of the XML files, and the evolving nature of the COVID-19 pandemic.

# Chapter 4

# Standards Adopted

Several practices are followed to ensure accuracy and reliability when examining COVID-19 data:

Data import and preprocessing: The COVID-19 dataset is imported using the Pandas library in Python. Appropriate preliminary steps are taken to ensure data quality, such as handling missing values and removing irrelevant columns.

Aggregation: COVID-19 data is aggregated by splitting data by country to obtain the number of confirmed cases over time. This aggregation facilitates national surveillance of the spread of the disease.

Determination of key metrics: Key metrics such as the highest number of cases are calculated for each country. These indicators provide an understanding of the severity of the outbreak in different regions.

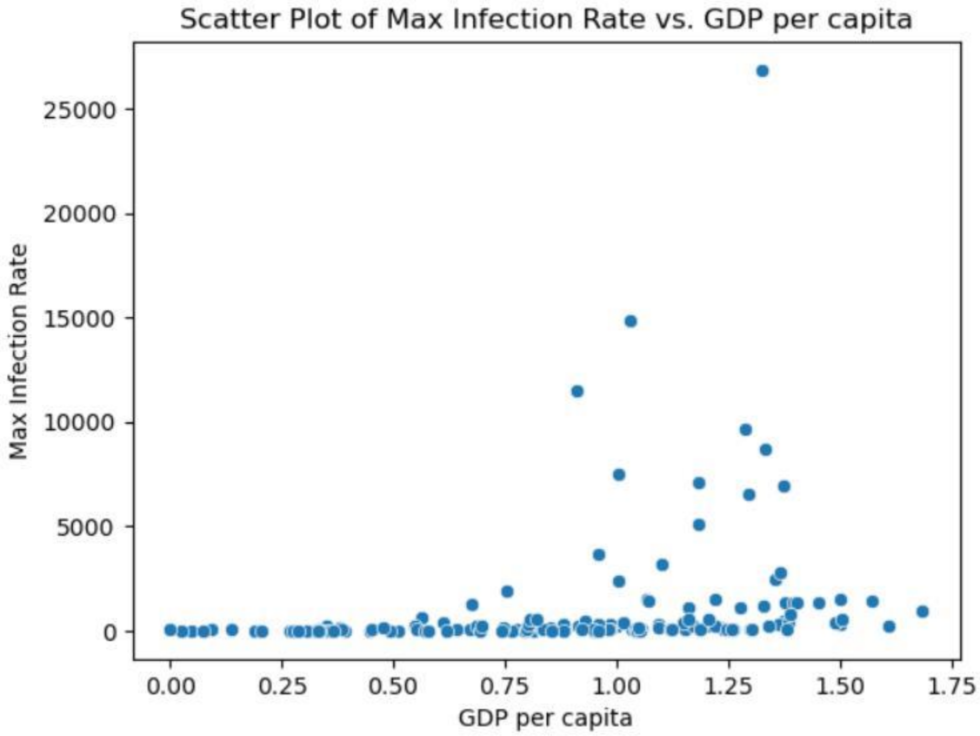
Integration with World Happiness Report: COVID-19 data is integrated with the World Happiness Report dataset to explore potential correlates of the virus spread, such as GDP per capita, social support, and life expectancy correlation.

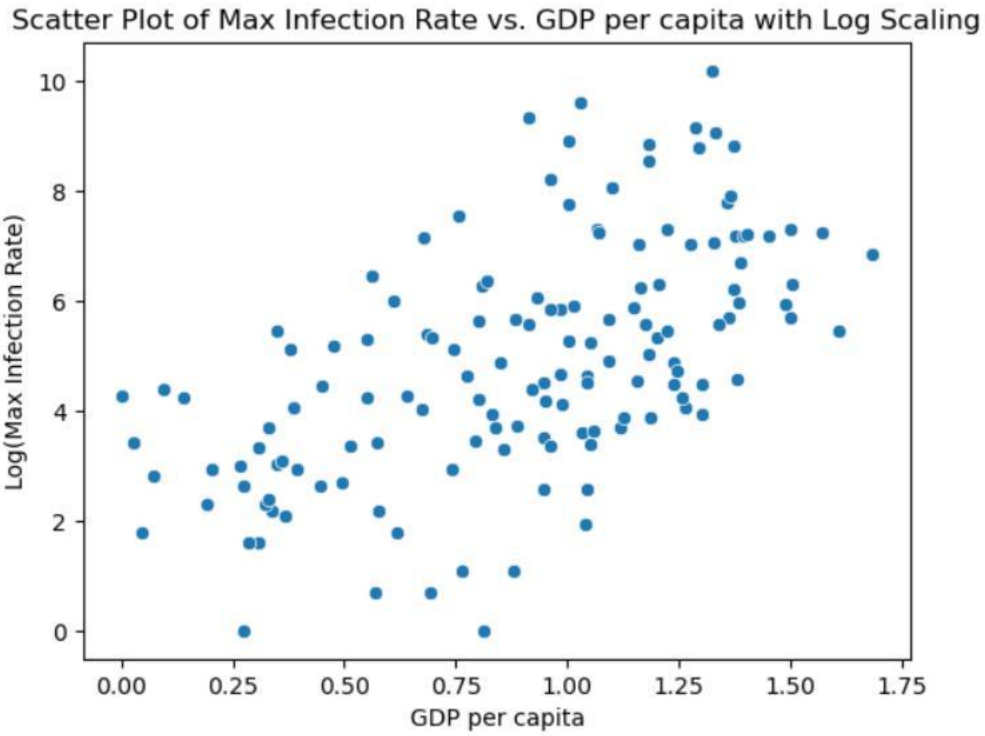
Correlation Analysis: The correlation matrix is calculated to determine the relationship between top disease and health indicators. This analysis helps understand the potential impact of economic factors on the spread of COVID-19.

Visualization: Survey results are visualized using charts and graphs to make the research more understandable and accessible to a wider audience. Visual representation aids in conveying complex information effectively and facilitates better comprehension of the findings.

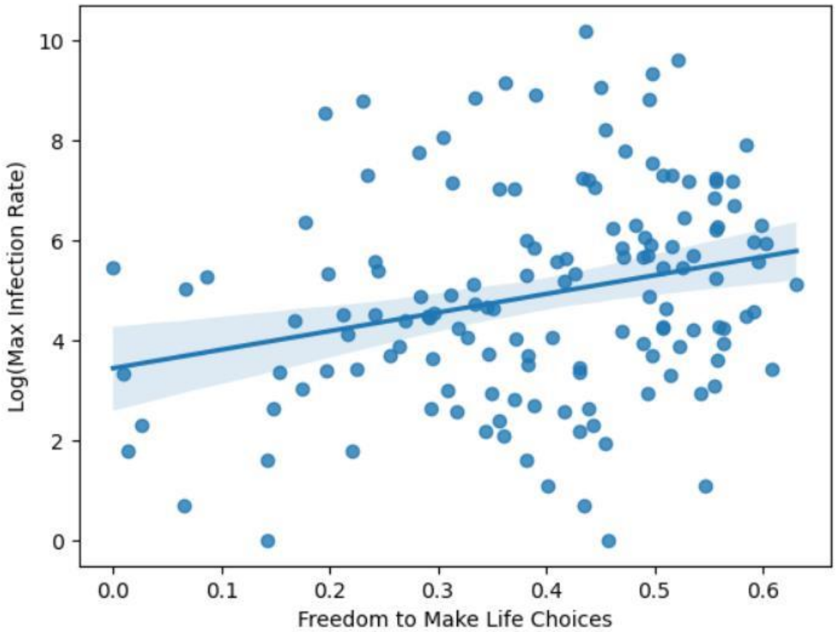
# Chapter 5

# Experimental Results

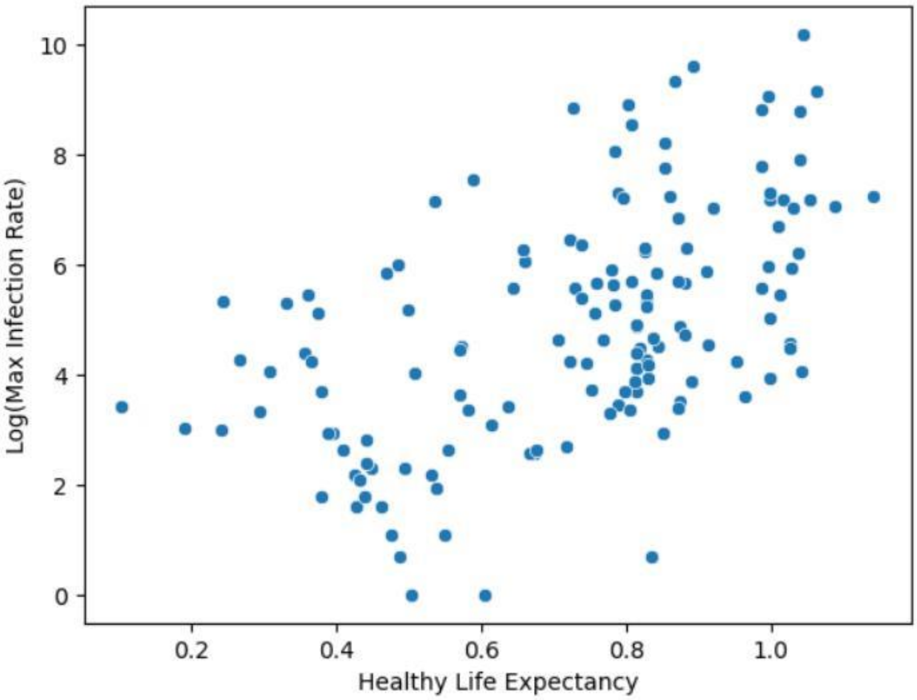


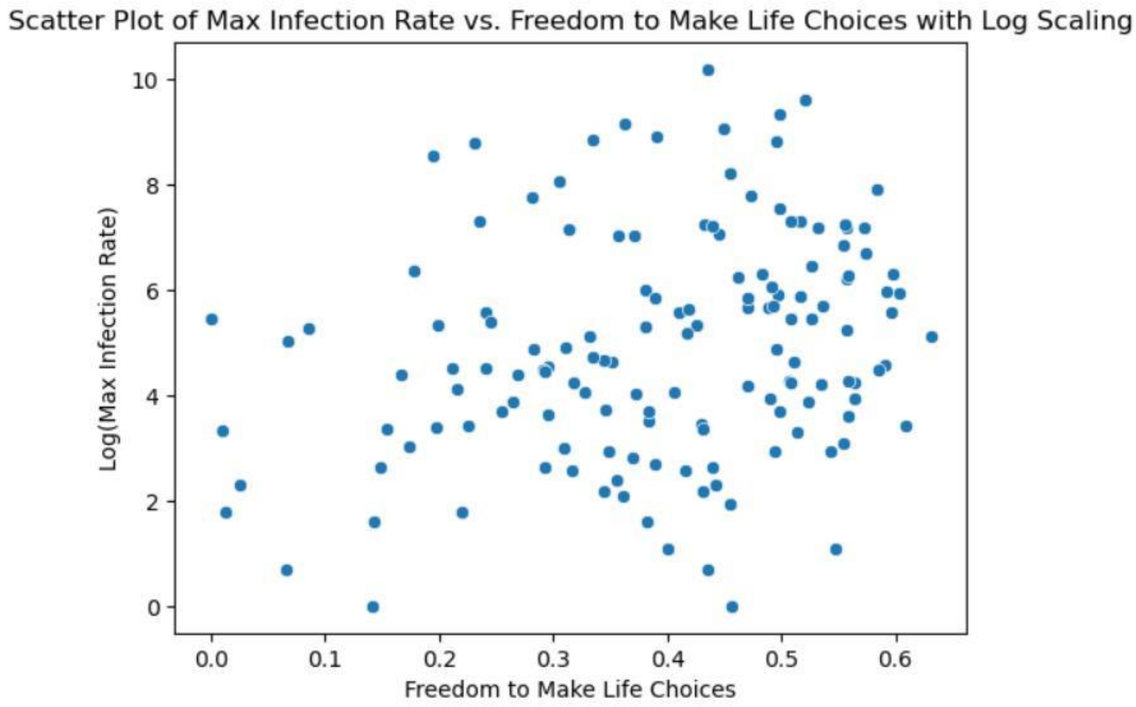



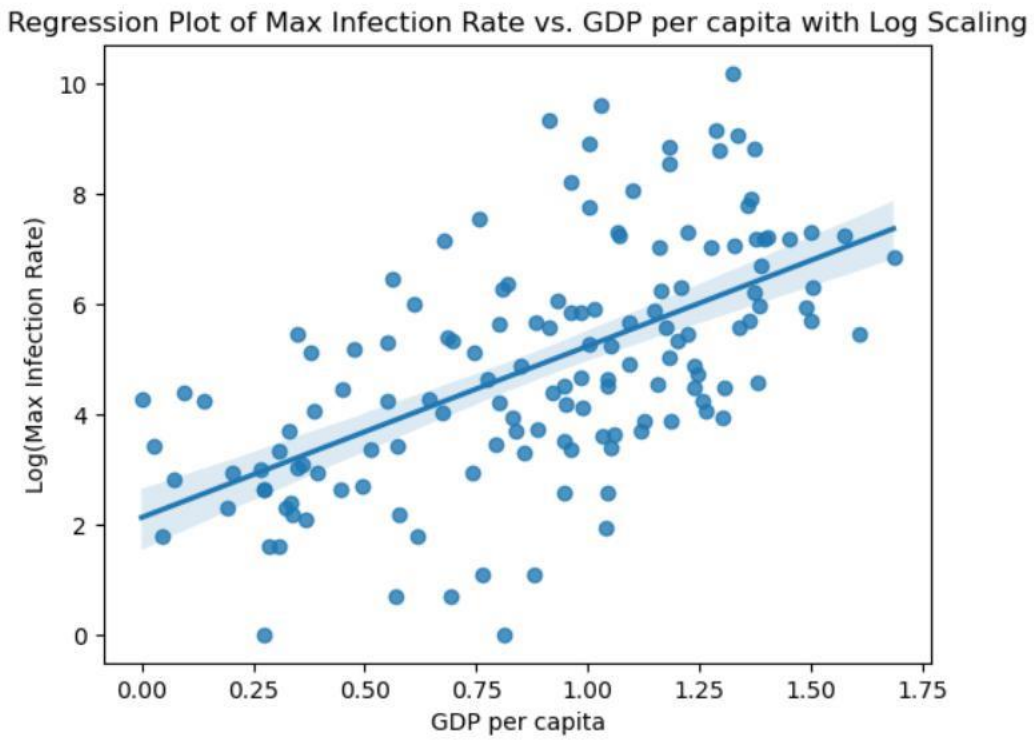
Regression Plot of Max Infection Rate vs. Freedom to Make Life Choices with Log Scaling

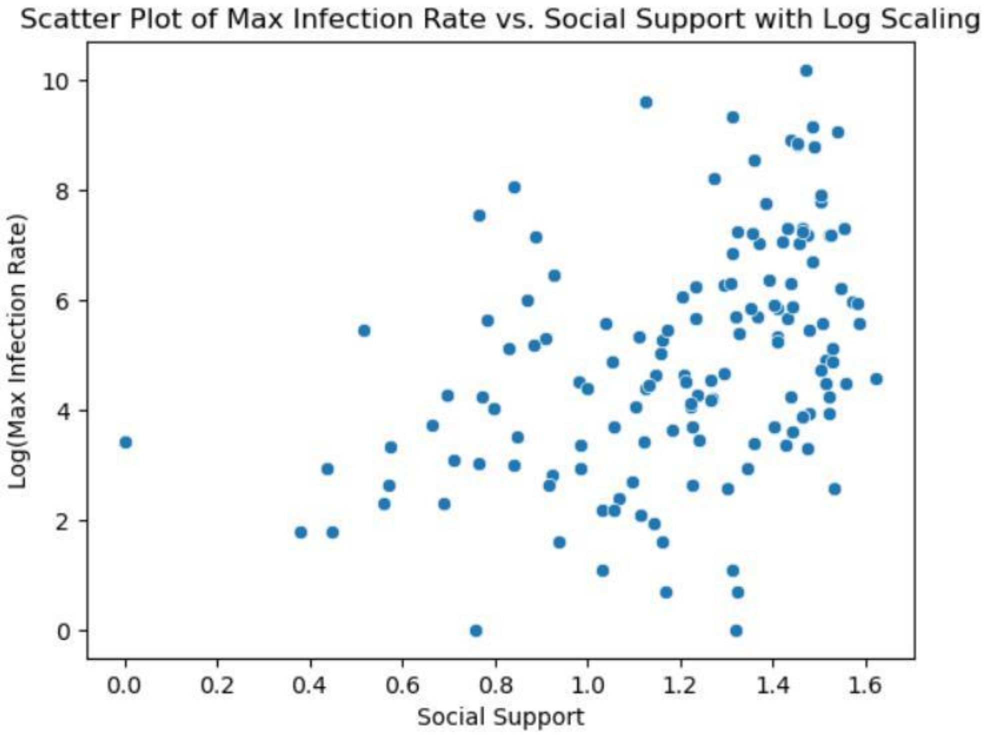


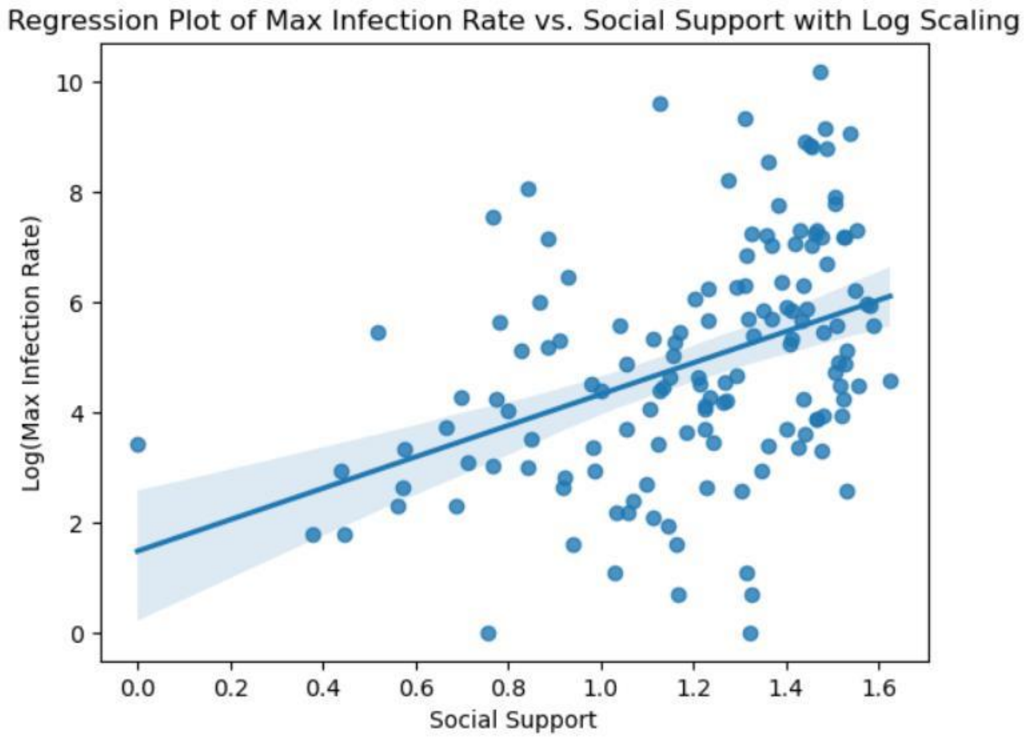

Scatter Plot of Max Infection Rate vs. Healthy Life Expectancy with Log Scaling




Scatter Plot of Max Infection Rate vs. Freedom to Make Life Choices with Log Scaling
Log(Max Infection Rate)
Freedom to Make Life Choices
0
2
4
6
8
10
0.0
0.1
0.2
0.3
0.4
0.5
0.6


Regression Plot of Max Infection Rate vs. GDP per capita with Log Scaling
Log(Max Infection Rate)
GDP per capita
0
2
4
6
8
10
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75


Scatter Plot of Max Infection Rate vs. Social Support with Log Scaling
Log(Max Infection Rate)
Social Support
0
2
4
6
8
10
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6


Regression Plot of Max Infection Rate vs. Social Support with Log Scaling
Log(Max Infection Rate)
Social Support
0
2
4
6
8
10
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6

# Chapter 6

# Conclusion

The analysis of 19 documents in the World Happiness Report on COVID19 and their relationship with socio-economic conditions was carried out because they contain some important information:

1. Correlation with economic society: Correlation analysis shows that the peak of the epidemicCOVID19 is positively associated with health indicators such as GDP per capita,health promotion, healthy life expectancy. This suggests that countries with greater health development may have more diseases.

2. NOTICE: When a correlation is observed, it is important to remember that correlation is not causation. More research is needed to understand the complex interactions between the economy and the spread of infectious diseases such as COVID19.

3. Importance of data collection: Integrating COVID19 data with socioeconomic indicators provides a better understanding of the impact of the epidemic. This collaboration can guide policymakers in developing effective international interventions and recovery strategies.

4. Data Transparency and Collaboration: The analysis highlights the importance of data transparency and international collaboration to respond to global health challenges. Access to accurate and timely information allows researchers and policymakers to make informed decisions and take important steps to reduce the spread of disease.

Overall, the analysis highlights the interaction between health outcomes and population health and emphasizes the need to join hands to address global health challenges such as COVID19.