KIT718 Big Data Analytics

DATA ANALYTICS PROJECT

Phase 1

PREPARED BY: GROUP NO:20

Jayani Madusha Edirisinghe-(707202)

Ritham Chhetri (653362)

Kanchi Mahendra Dungarani (681889)

TABLE OF CONTENTS

1.0 Background and Problem Statement

Ensuring healthy lives and promoting well-being for all is a core focus of the United Nations' Sustainable Development Goal 3 (SDG 3). This goal is especially challenging for countries struggling with food insecurity, where access to nutritious diets is limited, and health outcomes are often poor. Food insecurity not only leads to malnutrition but is also closely tied to increased child and maternal mortality, shortened life expectancy, and higher incidences of preventable diseases. The complexity of this problem makes it a significant concern for policymakers, international organizations, and governments across the globe. A better understanding of how food security impacts health outcomes is crucial to designing effective interventions that can alleviate these issues.

The significance of this problem extends beyond national borders, affecting low and middle-income countries worldwide. Food insecurity limits access to the basic nutrition needed for good health, which in turn exacerbates health disparities. If solutions are found to make nutritious food more affordable and accessible, it could lead to better health outcomes, reduced mortality rates, and a longer life expectancy for millions of people. Solving this issue is not only important for improving individual well-being but also for driving global health equity and sustainable development.

To explore this challenge, we will use three datasets: Food Insecurity and Nourishment Indicators**,** Food Prices for Nutrition**,** and Health-Related Indicators.

1. Food Insecurity and Nourishment Indicators Dataset: This dataset tracks various indicators of food insecurity and nourishment in different countries. It provides key data points on undernourishment rates, food insecurity levels, and health-related expenditures.

- Key Characteristics:
    - Food Insecurity Level: Measures the percentage of the population that lacks regular access to sufficient food, indicating the severity of food insecurity.
    - Undernourishment Rates: The proportion of the population that does not meet the minimum dietary energy requirements.
    - Health Expenditures: The amount spent by countries on health services per capita, providing insight into how much investment is made in healthcare systems in relation to food insecurity.
- Significance: This dataset is crucial for understanding how food scarcity impacts nourishment and health outcomes. It helps to analyze whether countries with higher

food insecurity are also experiencing worse health indicators such as higher child and maternal mortality rates.

2. Food Prices for Nutrition Dataset:This dataset captures the cost of maintaining a nutritious diet in various countries and tracks the percentage of populations that are unable to afford such diets.

- Key Characteristics:
    - Cost of a Healthy Diet: The average cost required for an individual to maintain a diet that meets nutritional guidelines.
    - Percentage Unable to Afford a Healthy Diet: The portion of the population that cannot afford a healthy diet based on their income and economic conditions.
- Significance: The cost of nutrition is a key factor in assessing food security. Countries with higher food prices may have a larger portion of the population unable to afford a healthy diet, leading to poor health outcomes. This dataset helps identify economic barriers to achieving proper nutrition and highlights regions where affordability is a critical issue.

3. Health-Related Indicators Dataset: This dataset provides a comprehensive view of public health outcomes across different countries. It includes metrics such as maternal and child mortality rates, disease prevalence, vaccination coverage, and life expectancy.

- Key Characteristics:
    - Maternal and Child Mortality: Tracks the number of deaths of mothers during childbirth and children under five, indicating the quality of healthcare and nutrition.
    - Life Expectancy: The average number of years a person is expected to live based on current health conditions in the country.
    - Disease Prevalence: Prevalence rates of communicable diseases such as tuberculosis and HIV, which are often higher in food-insecure regions.
    - Vaccination Coverage: The percentage of the population that is vaccinated against major preventable diseases.
- Significance: These health-related indicators provide a direct link between food security, healthcare access, and public health outcomes. This dataset allows for the examination of how food insecurity and the affordability of nutrition impact critical health metrics such as mortality and life expectancy.

Together, these datasets allow us to explore the connections between food security, economic affordability, and public health outcomes, helping to identify where the most pressing issues lie. Understanding this data is essential for informing policy decisions on health services, nutrition programs, and food subsidies. By analyzing the relationship between food security and health outcomes, this study aims to provide insights into how countries can better allocate resources and create targeted interventions that will improve the health and well-being of vulnerable populations. A better solution to this problem could not only reduce preventable deaths and diseases but also contribute significantly to global efforts to meet SDG 3, ensuring a healthier future for all.

Questions

1. How does food insecurity impact key health indicators such as child mortality and maternal mortality across different countries? (This question links food insecurity to vital health indicators, helping to identify how it affects vulnerable populations, particularly children and mothers.)
2. What is the relationship between the cost of a healthy diet and health outcomes, such as life expectancy and disease incidence? (It explores economic barriers to nutrition, revealing how diet affordability influences overall health, guiding policies for better access to nutritious food.)
3. How do countries with higher health expenditures fare in reducing mortality rates and increasing life expectancy compared to those with lower health investments? (This assesses the effectiveness of healthcare spending, showing whether increased investments lead to improved health outcomes, which can inform resource allocation.)
4. 4.What is the relationship between fertility rate and income level of countries?

2.0 Data Selection

In the data selection process, we analyzed all the datasets, starting with Food_Insecurity_Nourishment_Indicators, Food_Prices_for_Nutrition, Health-Related Indicators and Overall health acheivement. Our analysis identified that the data in Food_Prices_for_Nutrition spans from 2017 to 2022. However, upon analyzing the Food_Insecurity_Nourishment_Indicators dataset, we found that only 37 values exist for 2022 (which is 97.7%), with the rest of the data missing for that year. Therefore, we decided to select data from 2017 to 2022 from all datasets.

Our goal is to obtain insights from different continents and countries with varying income levels. To achieve this, we researched the United Nations Sustainable Development Report, which provided us with information on the different income levels and categorizations of countries. Based on this report and the overall health achievement levels file, we identified four income categories: Low-Income Countries, Lower-Middle-Income Countries, Upper-Middle-Income Countries, and High-Income Countries.

Through our research, we found the World Bank's income group categorization for each country for each year, and we also discovered continent-wise data mappings for the countries. We choose Asia, Europe, Africa, Oceania, North America, Antarctica, and South America as the seven continents. noting that there are no countries in Antarctica., we selected the other six continents.

After that, we combined these datasets for our analysis. During the data selection process, we took a sample of 15 countries based on data availability, ultimately selecting one country from each income level across the six continents.

	Continent_y	Country	Income_Group
35	Europe	Albania	Upper-middle-income countries
25	Asia	Armenia	Upper-middle-income countries
55	Oceania	Australia	High-income countries
30	Europe	Austria	High-income countries
20	Asia	Bangladesh	Lower-middle-income countries
10	Africa	Botswana	Upper-middle-income countries
70	South America	Brazil	Upper-middle-income countries
0	Africa	Burkina Faso	Low-income countries
5	Africa	Cameroon	Lower-middle-income countries
40	North America	Canada	High-income countries
65	South America	Chile	High-income countries
50	North America	Costa Rica	Upper-middle-income countries
60	Oceania	Fiji	Upper-middle-income countries
45	North America	Honduras	Lower-middle-income countries
15	Asia	Israel	High-income countries

3.0 Data Preparation

In the dataset preparation process, the main issue we encountered was missing values across the datasets. To address this, we first replaced the missing values to accurately identify and manage them. We then analyzed the percentage of missing values in each column by calculating the proportion of missing data.. The percentage of missing values in each column was computed using the following command:

missing_pct = round(dataframe_name.isnull().sum() / len(dataframe_name) * 100, 1)

In the transformation phase,Redundant columns, such as 'Time Code', were dropped to remove unnecessary information Afterward, we renamed the columns to more intuitive labels, such as changing 'Affordability of a healthy diet: ratio of cost to the food poverty line [CoHD_pov]' to 'CoHD_pov' and 'Cost of a healthy diet [CoHD]' to 'CoHD', among others, to simplify our analysis.

we ensured that the data types were correctly formatted, such as converting key columns like 'CoHD_pov', 'CoHD', and 'CoHD_headcount' to float. We also reshaped the Food_Insecurity_Nourishment_Indicators dataset by melting and pivoting it to create a 'Year' column and reorganized other columns accordingly, which helped structure the data set for combining. The datasets were then merged using left and inner joins, where necessary, to combine relevant data from different sources. For data integration, we used left and inner joins to merge relevant data from different sources. Left joins ensured that no data was lost during merging, while inner joins were used when both datasets had complete data for the relevant columns, ensuring meaningful data integration

In the Food_Prices_for_Nutrition and health-related indicators datasets, rows with missing data percentages lower than 10% were dropped. For columns with missing values between 10% and 70%, we used group-wise forward fill with mean imputation, particularly grouped by country. Finally, we ensured that all columns were renamed appropriately, making the datasets ready for further analysis.

In our analysis we did not focus on normalizing the data since all the data are is in the range for our analysis to get correct idea regarding dataset and different indicators related to health and food we did not do the normalization for this analysis at this stage.

Finally, for our analysis, we selected 15 countries representing different income levels that had complete and clean datasets. These countries were chosen based on the availability of all relevant values, ensuring that our analysis would be accurate and robust. By focusing on these countries, we were able to streamline the dataset, eliminate inconsistencies, and ensure that all data was properly cleaned and ready for further in-depth analysis.

4.0 Exploratory Data Analysis

In this exploratory data analysis (EDA), the use of bar charts is a deliberate and well-suited choice for the dataset. Here's the reasoning behind the selection of this visualization type and the approach taken for exploring the data:

In this exploratory data analysis (EDA), bar charts are chosen because they are very effective for the type of data we are working with. Here's why we picked this type of chart and how we used it to analyze the data:

Bar charts are perfect for comparing different categories of data, which is exactly what we need for this analysis. Our dataset includes variables such as health spending, the percentage of people not getting enough food, and rates of death for mothers and children, as well as vaccination rates. we need to compare them across different countries, years, and income levels.

Since we are looking at data grouped by continent, country, and income level, bar charts help us easily see how each country is doing in terms of different health measures (like how many people are undernourished, the fertility rate, or vaccination rates) when compared to other countries in the same income group or region.

In the first group of charts, where we organize by 'Year', bar charts show how countries are doing in various health areas over time. This helps to see if there are improvements or if some countries are not making progress.

In the second group of bar charts, the data is organized by 'Continent_Country', with 'Continent' and 'Country' sorted in a logical order yearly. This way, we can see the geographic location and also consider the income levels of the countries yearly.

Using 'Income_Group' as a color code, the charts allow us to compare countries within the same economic groups. This is helpful for understanding how countries with similar locations, but different income levels perform in health areas.

We use bar charts as the, Bar charts are very useful when you want to compare specific health metrics, like spending on health or birth rates, across different groups, such as continents or income groups.

The EDA shows that rich countries (shown in pink in the bar charts) usually have better health results (like lower deaths during childbirth and higher vaccination rates), while poor countries (orange bars) often have problems with not having enough food and high death rates.

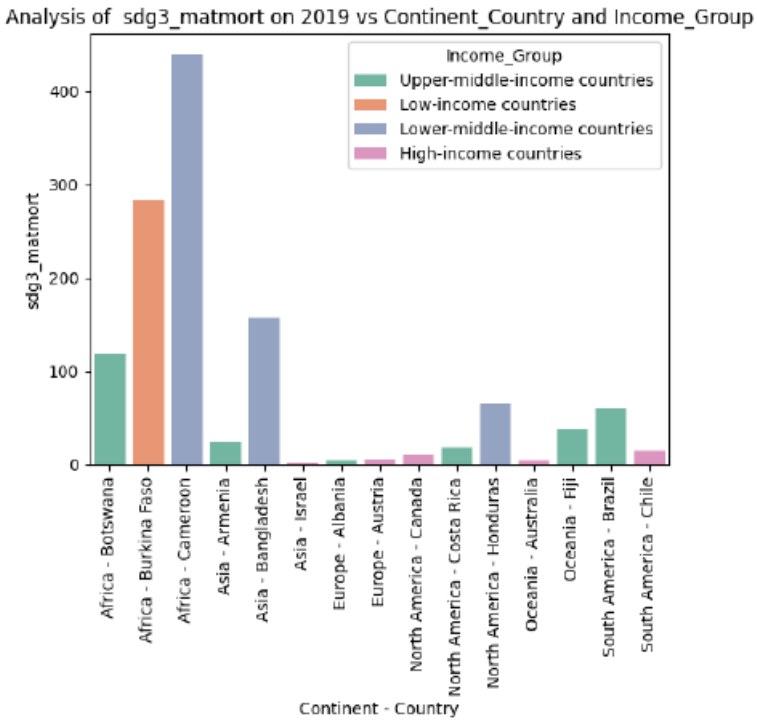
In the group of bar charts, we can see how things change from year to year, like more people getting vaccinated or fewer children dying in some countries over time. These changes help us understand how well health programs work in different places.

By organizing the data and using colors for different income levels in the bar charts, we can easily see where richer countries do better than poorer ones in health areas. For example, poorer countries might have more people not getting enough food and less access to medical care.

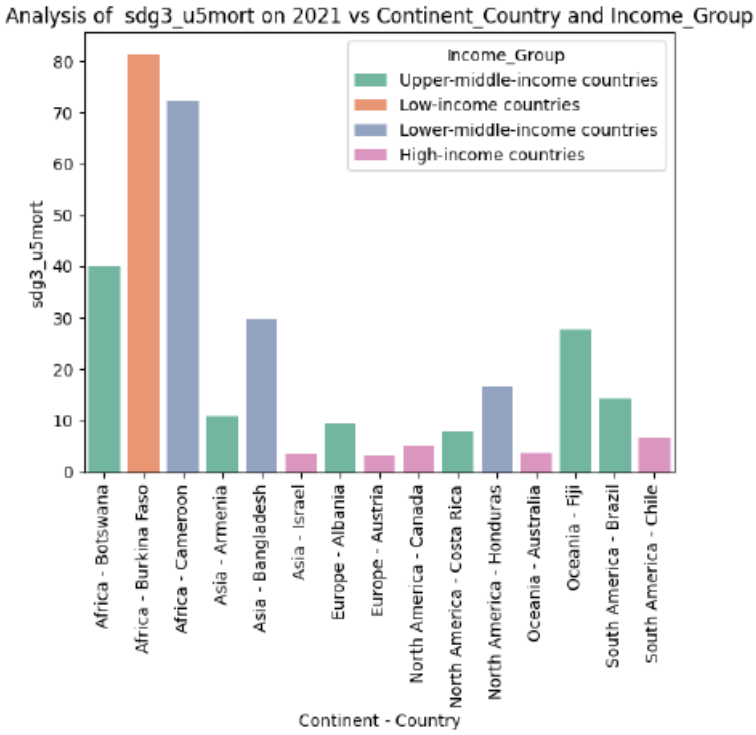
Questions

1.How does food insecurity impact key health indicators such as child mortality and maternal mortality across different countries?

Most of the Countries with lower income levels (Low and Lower-middle income groups) such as Burkina Faso Cameroon and Bangladesh have very importantly higher maternal mortality rates compared to higher-income countries. For example Cameroon and Burkina Faso (low-income and lower-middle-income countries respectively) have maternal mortality rates exceeding 100–400 deaths per 100000 live births. In contrast countries from the high-income group (e.g. Israel Austria Canada) show extremely low maternal mortality rates often close to zero highlighting the disparity in maternal healthcare between wealthier and poorer countries. As shown in following figure example year 2019 similar distribution shown in other years..



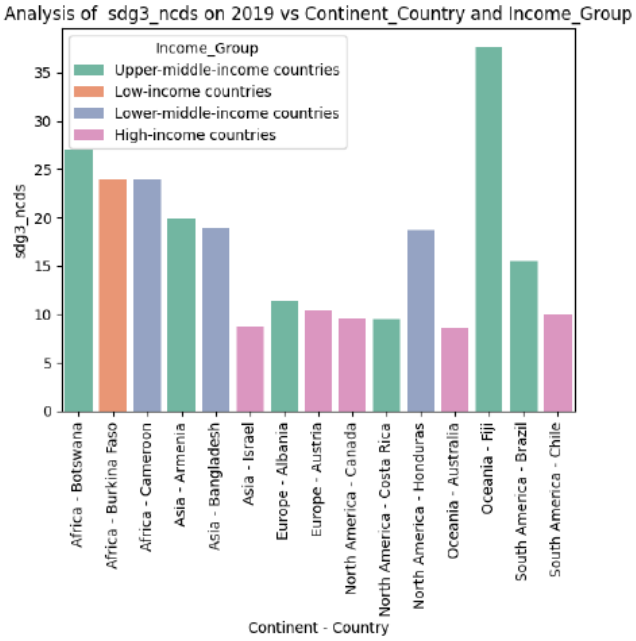
Similarly under-five mortality rates are highest in Burkina Faso and Cameroon with rates above 70 per 1000 live births again showing that low-income countries face very importantly higher child mortality. In contrast high-income countries such as Israel Austria and Canada have extremely low under-five mortality rates typically below 5 deaths per 1000 live births suggesting that child healthcare in these regions is much better as shown in following figure considering year 2021 similar distribution shown in other years.

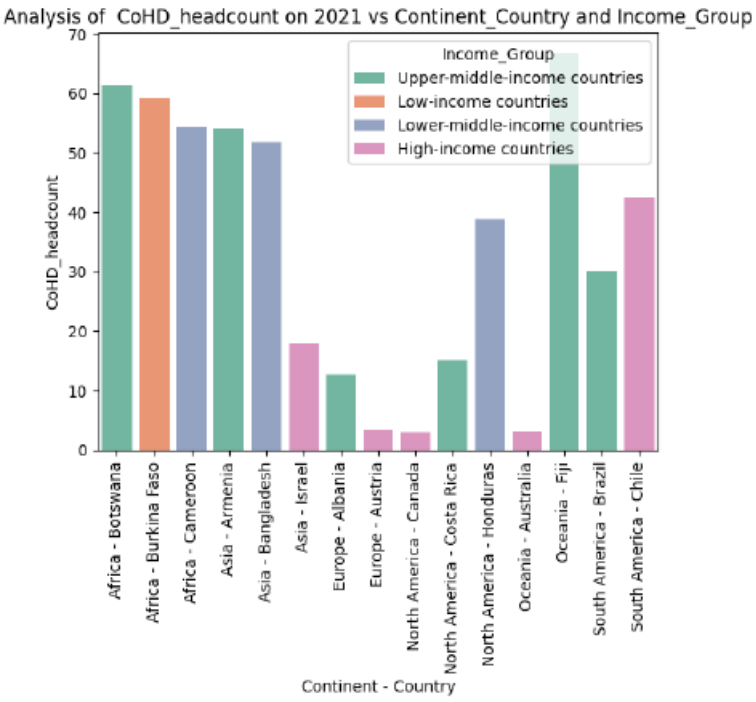


Maternal and child mortality have a substantial correlation with food insecurity. The potential mortality is increased by the average approach to nutrition in low-income countries. It's possible that these nations' restricted access to wholesome food contributes to the poor health of mothers and increased vulnerability to diseases common in children. Burkina Faso and Cameroon, two low- and lower-middle-income countries, probably suffer serious problems with food security, as seen by relatively high death rates. High-income nations are far better equipped to control and lower the rates of maternal and infant mortality because they have stronger healthcare systems, more food security, and both.

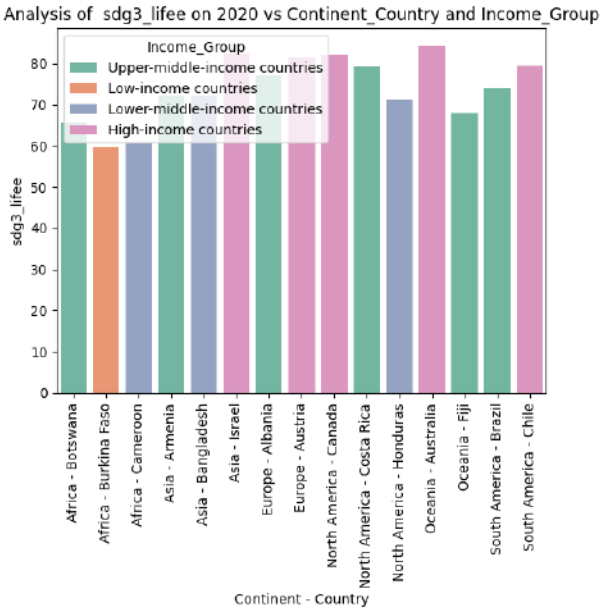
2.What is the relationship between the cost of a healthy diet and health outcomes, such as life expectancy and disease incidence?

Non-communicable diseases (NCDs) are more prevalent in lower-income (low- and lower-middle-income) countries. These areas tend to have higher nutrient costs relative to income as shown on, making it difficult for residents to provide consistent, healthful diets. This, in turn, raises the risk of diet-related diseases like diabetes, obesity, and arsenic eye damage. For and example illustrate in following figure. Similar distribution shown in other years





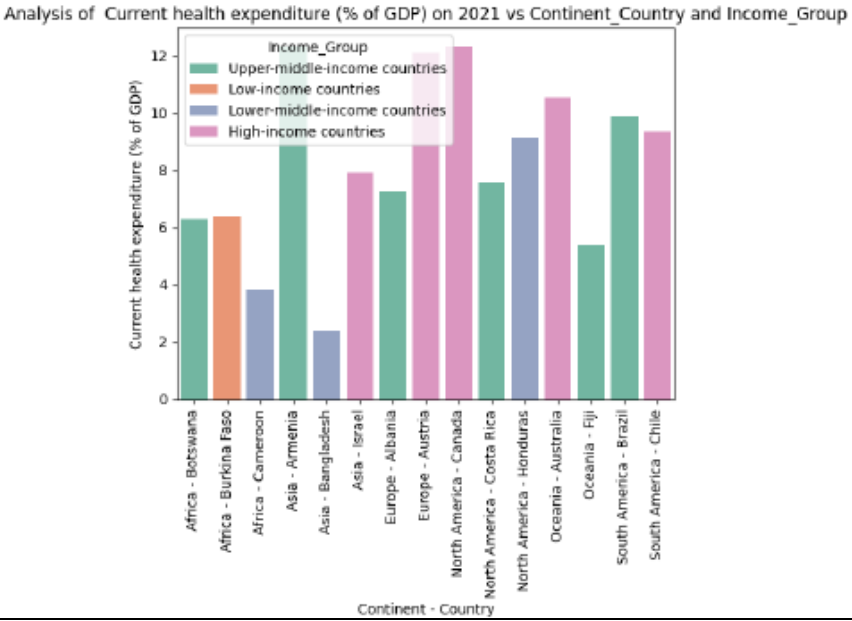
The high-income countries (shown by pink bars in the following figure) have consistently longer life expectancies, which may indicate superior health outcomes. These nations' potential for a free-for-all approach to healthcare and further affordable, nutritious nutrition up to long life expectancies



Low-income nations (orange bars), such as Burkina Faso and Cameroon, have shorter life expectancies, possibly because availability to reasonably priced, nutrient-dense food is limited or because maintaining a good diet is expensive. imply nourishment contributes to shortened life spans and higher illness exposure.

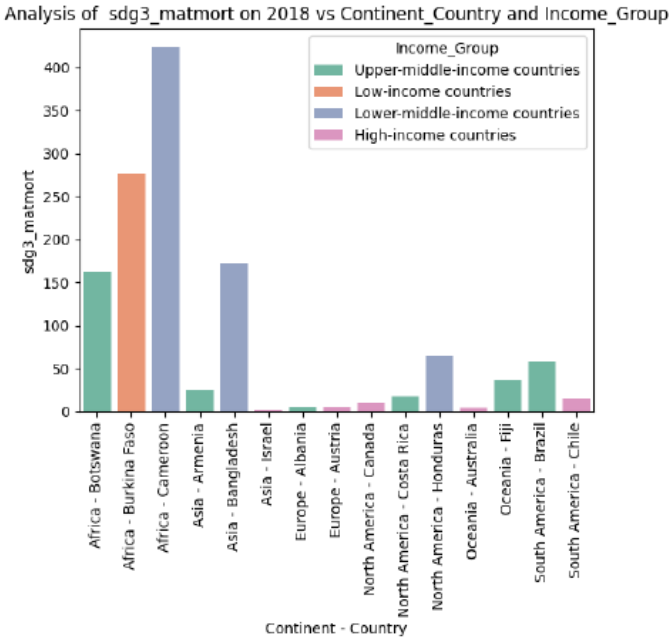
In higher-income countries, better access to affordable nutritious food contributes to longer life expectancy. Conversely, in low-income countries, the cost of a healthy diet and limited access to food contribute to shorter life expectancies.

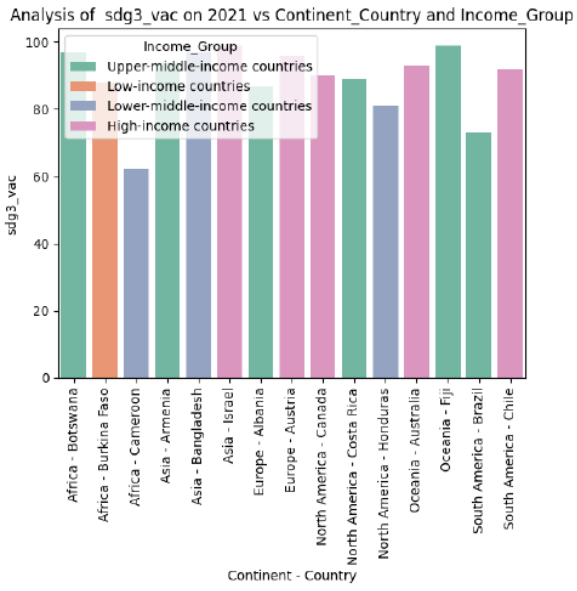
3.How do countries with higher health expenditures fare in reducing mortality rates and increasing life expectancy compared to those with lower health investments?



High-income countries invest heavily in healthcare systems, leading to higher life expectancies. Countries like Australia and Austria have high life expectancies for example illustrate as in above figure, with comprehensive services and preventive measures. Middle-income countries like Brazil and Costa Rica have slightly lower life expectancies, but still hold significant health programs and global aid. Low-income countries like Africa, Burkina Faso, and Cameroon have shorter life spans due to limited healthcare access and disease burden.

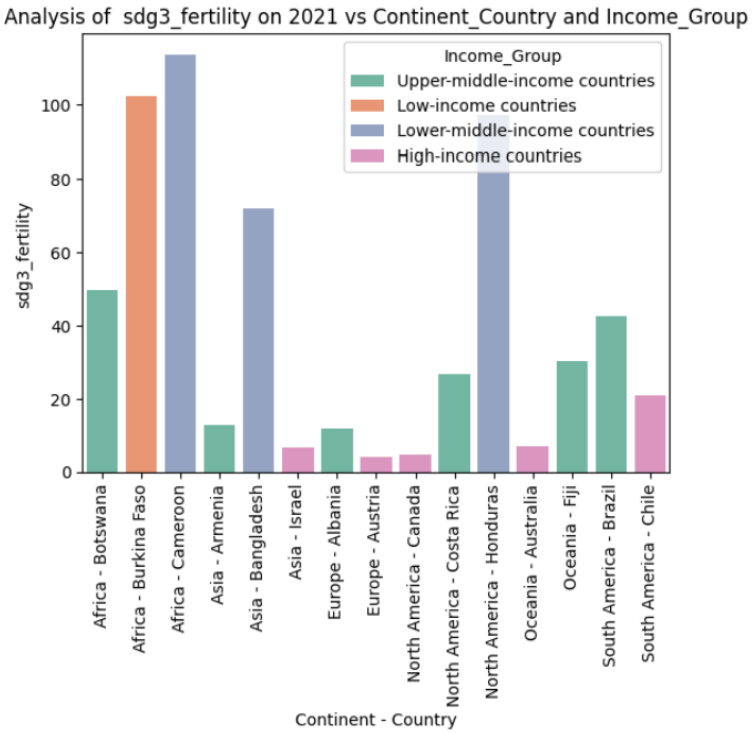
High healthcare investment leads to lower mortality rates as example illustrate in following chart, as countries with more investment in healthcare programs, vaccinations shown in below chart(sdg3_vac), disease bar measures, and advanced checkup treatments reduce morbidity rates. Conversely, low healthcare investment in countries like Honduras and Burkina Faso results in higher morbidity rates due to inadequate healthcare infrastructure and supplies. Countries with higher health expenditures tend to fare better in reducing mortality rates and increasing life expectancy.





4.What is the relationship between fertility rate and income level of countries?

There is a clear inverse relationship between a country's fertility rate and its income level. High-income countries such as Israel, Austria, Canada, and Australia generally exhibit low fertility rates, despite having better healthcare systems, food security, and overall living standards. On the other hand, low-income countries like Burkina Faso and Honduras tend to have higher fertility rates, even though they experience poorer healthcare and food insecurity



4.0 Conclusion

The Exploratory Data Analysis (EDA) highlighted significant disparities in health and socioeconomic indicators across various countries, particularly emphasizing the impact of food insecurity on maternal and child health. Low- and lower-middle-income countries, such as Burkina Faso and Cameroon, showed alarmingly high maternal mortality rates exceeding 300-400 deaths per 100,000 live births, in stark contrast to negligible rates in high-income countries like Israel and Canada. Additionally, under-five mortality rates in these nations were above 80 per 1,000 live births, while wealthier countries reported rates below five, illustrating how food insecurity exacerbates health challenges.

The analysis also revealed a connection between the cost of a healthy diet and overall health, with lower-income countries facing higher nutrient costs relative to income, leading to increased incidences of non-communicable diseases (NCDs). High-income countries enjoy longer life expectancies due to better access to nutritious food and comprehensive healthcare systems, while low-income nations grapple with high food prices and limited healthcare access, resulting in shorter life spans.

Moreover, countries that invest more in healthcare typically experience lower mortality rates. High-income nations like Australia and Austria demonstrate effective disease prevention strategies and better health outcomes, while low-income countries like Burkina Faso and Cameroon face higher morbidity rates due to inadequate healthcare investment. The findings reveal that 50-60% of the population in low-income countries cannot afford a healthy diet, contributing to malnutrition and diet-related diseases. This emphasizes the urgent need for targeted nutritional support programs. Additionally, vaccination coverage plays a crucial role in improving health outcomes, as high-income countries show better vaccination rates and lower incidences of preventable diseases.

Furthermore, exploring more questions like mentioned below can help us better understand how health, food security, and economic status are related to one another. This understanding can then inform policy recommendations meant to close the gap in health.

1. Which specific health indicators are most influenced by food insecurity in different countries?
2. Which countries have made the most significant improvements in health indicators related to nutrition over the past few years?

In summary, addressing food insecurity and improving healthcare access are essential for reducing mortality rates and enhancing health outcomes in vulnerable populations

5.0 References

Sustainable Development Solutions Network. (n.d.). SDG index & dashboards explorer. SDG Index. Retrieved September 27, 2024, from https://dashboards.sdgindex.org/explorer

Ritchie, H., Roser, M., Ortiz-Ospina, E., & Ortiz-Ospina, E. (n.d.). World Bank income groups. Our World in Data. Retrieved September 27, 2024, from https://ourworldindata.org/grapher/world-bank-income-groups?tab=table

Sustainable Development Solutions Network. (n.d.). The 2024 SDG index score and ranking. SDG Index. Retrieved September 27, 2024, from https://dashboards.sdgindex.org/chapters/part-2-the-sdg-index-and-dashboards#the-2024-sdg-index-score-and-ranking

Appendix

Individual Task Division and Contribution:

1. Data Preparation:

i. Jayani:

- ✓ Variables Prepared: Food Insecurity and Nourishment Indicators
- ✓ Tasks: Cleaning and transforming food insecurity-related variables (handling missing values, normalizing data if needed).

ii. Kanchi:

- ✓ Variables Prepared: Food Prices for Nutrition
- ✓ Tasks: Cleaning food prices data variables (handling missing values, normalizing data if needed).

iii. Ritham:

- ✓ Variables Prepared: Health-Related Indicators
- ✓ Tasks: Cleaning health-related variables variables (handling missing values, normalizing data if needed).

2. Exploratory Data Analysis (EDA):

i. Jayani:

- ✓ Countries Explored: 5 countries (Albania,Armenia,Australia,Austria,Bangladesh).
- ✓ Variables Explored: Relationship between food insecurity, nourishment and health indicators.

ii. Kanchi:

- ✓ Countries Explored: 5 countries (Israel,Chile,Honduras,Costa Rica,Botswana).
- ✓ Variables Explored: Relationship between food insecurity, nourishment and health indicators.

iii. Ritham:

- ✓ Countries Explored: 5 countries (Brazil,Cameroon,Canada,Burkina Faso,Fiji).
- ✓ Variables Explored Relationship between food insecurity, nourishment and health indicators.