CONTENTS

- ABOUT PROJECT
- OBJECTIVE
- IMPORTING LIBRARIES
- IMPORTING DATA
- STATISTICAL SUMMARY
- DATA VISUALIZATION
  - 01. DEMOGRAPHY
  - 02. ECONOMIC SUMMARY
  - 03. EDUCATION
  - 04. EMPLOYMENT
  - 05. HEALTH & SOCIAL ASSISTANCE ANALYSIS
    - 5.1.1. COUNTRY WISE CHILD MORTALITY RATE ANALYSIS
    - 5.1.2. COUNTRY WISE ADOLESCENT MORTALITY RATE ANALYSIS
    - 5.1.3. COUNTRY WISE ADULT MORTALITY RATE ANALYSIS
    - 5.1.4. COUNTRY WISE ANALYSIS OF REPORTS
    - 5.1.5. COUNTRY WISE ANALYSIS OF DEATHS
    - 5.1.6. COUNTRY WISE ANALYSIS OF DEATHS
    - 5.1.7. COUNTRY WISE ANALYSIS OF DEATHS
    - 5.1.8. COUNTRY WISE ANALYSIS OF CHILD MORTALITY AND TOTAL FERTILITY
    - 5.1.9. COUNTRY WISE ANALYSIS OF DEATHS
    - 5.1.10. COUNTRY WISE ANALYSIS OF CHILD MORTALITY AND TOTAL FERTILITY
    - 5.1.11. COUNTRY WISE ANALYSIS OF DEATHS
    - 5.1.12. COUNTRY WISE ANALYSIS OF DEATHS
  - 06. TOTAL EXPENDITURE
  - 07. HEALTH EXPENDITURE
  - 08. COUNTRY WISE HEALTH AND TOTAL EXPENDITURE
  - 09. HEALTH EXPENDITURE
  - 10. REPORTING AND DEATHS
  - 11. REPORTING AND DEATHS
  - 12. REPORTING AND DEATHS
  - 13. DEATHS
  - 14. DEATHS
  - 15. DEATHS PLOTLAND
  - 16. DEATHS PLOTLANDS
  - 17. COUNTRY PERSPECTIVES
- PREDICTIVE MODELING
  - 18. CLUSTERING
    - 18.1. KMEANS ANALYSIS
    - 18.2. KMEANS ANALYSIS
    - 18.3. CLUSTERING - 4
  - 19. CLUSTERING - 5 (HEALTH, DEATHS & INFLUENZA, HEALTH & CHILD MORTALITY)
  - 20. HIERARCHICAL CLUSTERING
    - 20.1. HIERARCHICAL CLUSTERING
    - 20.2. HIERARCHICAL CLUSTERING
    - 20.3. WORLD MAPPING
    - 20.4. HIERARCHICAL REPRESENTATION OF CHILD MORTALITY
    - 20.5. HIERARCHICAL REPRESENTATION OF CHILD MORTALITY, HEALTH AND EXPECTANCY
    - 20.6. HIERARCHICAL REPRESENTATION OF INFLUENZA, HEPATITIS AND TUBE

ABOUT PROJECT

WHO International is an International Healthcare NGO dedicated to eliminating poverty and promoting healthy lifestyles worldwide. We have long been aware of crisis and natural disasters, which is a main source of poverty. Natural disaster activities and advocacy campaigns. WHO International recently received donations from donors.

The task of exploring countries based on socio-economic and health factors is crucial for WHO International's strategic decision-making. By analyzing these factors, WHO International can allocate its limited resources more effectively to address the most critical needs in different countries. This analysis helps in prioritizing interventions and initiatives that are tailored to the specific challenges and vulnerabilities in each country, ensuring that WHO International can maximize its impact on improving the health and overall well-being of those affected by natural disasters and other crises.

OBJECTIVE

The primary objective of this Kaggle project on Clustering countries is to categorize nations based on their socio-economic and health factors using a combination of clustering and PCA techniques.

Combining with essential data transition and tools like cleaning the data analysis variables and transformation variables ensures that the dataset is in the optimal condition for clustering.

Conducting an in-depth Exploratory Data Analysis (EDA) to gain insights into the dataset and uncover patterns and trends.

Performing dimensionality reduction using PCA to enhance the clustering process by reducing the dimensionality of the data while preserving important information.

Implementing various clustering algorithms, such as K-means clustering and hierarchical clustering, to group countries with similar characteristics.

Evaluating the performance of the clustering models to identify the optimal number of clusters and ensure meaningful clustering outcomes, such as, silhouette and inertia scores (Clusters in case of K-means clustering).

Interpreting the clustering results to derive actionable insights and recommendations for WHO International to prioritize and allocate resources effectively.

IMPORTING LIBRARIES

```
import numpy
```

```
import pandas
```

```
import matplotlib
```

```
import seaborn as sns
```

```
import plotly
```

```
import plotly.express as px
```

```
import plotly.express.colors as px_c
```

```
from plotly.subplots import make_subplots
```

```
from plotly.subplots.subplots import *
```

```
from plotly.graph_objects import *
```

```
from plotly.graph_objects.layout.ternary.aaxis import Title as t_aaxis_Title
```

```
from plotly.graph_objects.layout.ternary.baxis import Title as t_baxis_Title
```

```
from plotly.graph_objects.layout.ternary.caxis import Title as t_caxis_Title
```

```
from plotly.graph_objects.layout.ternary.sum import Title as t_sum_Title
```

```
from plotly.graph_objects.layout.scene.xaxis.title import Font as s_x_t_Font
```

```
from plotly.graph_objects.layout.scene.yaxis.title import Font as s_y_t_Font
```

```
from plotly.graph_objects.layout.scene.zaxis.title import Font as s_z_t_Font
```

```
from sklearn import preprocessing, decomposition, cluster, metrics, tree, ensemble, model_selection, linear_model, neighbors, naive_bayes, svm, cluster, mixture, manifold, gaussian_process, tree, ensemble, neighbors, naive_bayes, svm, mixture, manifold, gaussian_process
```

```
%matplotlib inline
```

IMPORTING DATA

```
df = pd.read_csv("/kaggle/input/")
```

```
df.shape
```

```
(2938, 22)
```

	country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdp
0	Afghanistan	90.2	10.0	7.58	44.9	1610	9.44	56.2	5.82	553
1	Albania	16.6	28.0	6.55	48.6	9930	4.49	76.3	1.65	4090
2	Algeria	27.3	38.4	4.17	31.4	12900	16.10	76.5	2.89	4460
3	Angola	119.1	62.3	2.85	42.9	5900	22.40	60.1	6.16	3530
4	Antigua and Barbuda	10.3	45.5	6.03	58.9	19100	1.44	76.8	2.13	12200

```
df.info()
```

```
<class pandas.core.frame.DataFrame>
RangeIndex: 2938 entries, 0 to 2937
Data columns (total 22): 
 country  child_mort  exports  health  imports  income  inflation  life_expec  total_fer  gdp
 Afghan   90.2       10.0     7.58    44.9     1610    9.44       56.2        5.82       553
 Albania  16.6       28.0     6.55    48.6     9930    4.49       76.3        1.65       4090
 Algeria  27.3       38.4     4.17    31.4     12900   16.10      76.5        2.89       4460
 Angola   119.1      62.3     2.85    42.9     5900    22.40      60.1        6.16       3530
 Antigua  10.3       45.5     6.03    58.9     19100   1.44       76.8        2.13       12200
dtypes: float64(9), object(1)
memory usage: 505.1+ KB
```

STATISTICAL SUMMARY

DESCRIPTION

- Data Description

df.columns : This column contains the names of different countries, identifying each unique entry in the dataset.

df.child_mort : This column represents the child mortality rate, which is the number of deaths per thousand children under five years of age. This metric is crucial indicator of healthcare quality and child survival rates.

df.exports : This column indicates the value of exports of goods and services per capita. It is expressed as a percentage of the GDP per capita. This metric reflects a country's ability to export goods and services to other nations, contributing to its economic health.

df.health : The "health" column quantifies the total health spending per capita, expressed as a percentage of the GDP per capita.

df.imports : This column quantifies the monetary value of imports goods and services per capita. It is expressed as a percentage of the GDP per capita. It represents the amount of foreign goods and services a country imports.

df.income : This column represents the net income per person in a country. It measures the average income of individuals in the population and is a crucial indicator of a country's economic prosperity.

df.inflation : "Inflation" measures the annual growth rate of the total GDP. It indicates the rate at which price level of goods and services in an economy rises over a specific period. Inflation can impact the purchasing power of a country's currency.

df.life_expec : This column represents the average number of years a newborn child would be expected live if the current mortality patterns (death rates) in the country remain unchanged. It provides a comprehensive overview of the overall health and healthcare access in a country.

df.total_fer : This column indicates the total fertility rate, which represents the average number of children born to each woman of the correct age fertility rates remain the same. It offers insights into a country's population growth and demographic trends.

df.gdp : "GDP" stands for Gross Domestic Product. It quantifies the total monetary value of all goods and services produced within a country's borders. This metric provides the economic size of a nation and is a key indicator of its overall economic performance.

```
df.describe()
```

	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdp
count	2938.000000	2938.000000	2938.000000	2938.000000	2938.000000	2938.000000	2938.000000	2938.000000	2938.000000
mean	38.305381	41.108976	6.813061	46.884761	17144.688923	7.781832	70.555976	2.947964	6154.666893
std	40.328931	27.412823	2.746687	24.209588	19278.067698	10.570704	9.033671	1.513848	5863.037689
min	2.600000	0.109074	1.810000	30.411519	609.000000	-4.394851	32.100000	1.150000	231.000000
25%	8.250000	23.800000	4.920000	35.664000	3355.000000	1.810000	65.975000	1.795000	1330.000000
50%	19.300000	35.000000	6.320000	43.433333	9960.000000	5.390000	73.100000	2.410000	4660.000000
75%	62.100000	51.350000	8.600000	54.287358	22800.000000	10.750000	76.800000	3.880000	7350.000000
max	208.000000	200.000000	17.900000	174.000000	125000.000000	104.000000	82.800000	7.580000	48700.000000

```
df.describe(include = "object")
```

	country
count	2938
unique	2938
top	Afghanistan
freq	1

DATA VISUALIZATION

COUNTRIES

BIGGEST AND SMALLEST/POOREST AMONG COUNTRIES

COUNTRY WISE CHILD MORTALITY RATE ANALYSIS

```
df_child_mort = df.sort_values("child_mort", ascending = False)
px.bar(data_frame = df_child_mort, x = "country", y = "child_mort", color_scale = px.colors.sequential.Rainbow, color = "child_mort", height = 700, width = 1500, text_auto = True, title = "Country wise Child Mortality Rate")
```

- INFERENCE :-

The social mortality rates are significantly among these nations, ranging from as low as 2.6 to as high as 208.0.

- Haiti has the highest child mortality rate, with a rate that stands at a stark mortality rate of 208.0. This high rate underscores the various socio-economic factors, including limited access to healthcare, poverty, and inadequate sanitation.

- Sierra Leone and Chad follow closely, Sierra Leone and Chad have the second and third highest child mortality rates, with rates of 174.0 and 160.0, respectively. These nations face similar challenges in terms of healthcare infrastructure and poverty.

- Central African Republic, Mali, and Nigeria: These countries also exhibit relatively high child mortality rates, ranging from 150.0 to 143.9. Common factors contributing to child mortality in these nations include poverty, limited healthcare access, and inadequate sanitation.

- Angola, Somalia, Guinea-Bissau, Burundi, Benin, and Cote d'Ivoire: These countries have child mortality rates between 116.8 and 111.1. Although these rates are slightly lower than those mentioned earlier, they still indicate significant challenges in healthcare and socio-economic development across these nations.

- Equatorial Guinea, Burkina Faso, Congo, Dem. Rep., Congo, Rep.: These countries have child mortality rates ranging from 111.4 to 107.4. These rates underscore the ongoing challenges in healthcare and socio-economic development in these regions.

COUNTRIES WITH HIGHEST EXPORTS

```
df_exports = df.sort_values("exports", ascending = False)
px.bar(data_frame = df_exports, x = "country", y = "exports", color_scale = px.colors.sequential.Rainbow, color = "exports", height = 700, width = 1500, text_auto = True, title = "Countries with Highest Exports")
```

- INFERENCE :-

The social mortality rates are significantly among these nations, ranging from as low as 2.6 to as high as 208.0.

- Luxembourg has the highest export ratio of 200.00 of its GDP per capita, indicating a substantial reliance on exports for economic growth.

- Malta, Singapore, Bahrain, United Arab Emirates, Ireland, Belgium (183.45), and Slovenia (151.35) also exhibit high export ratios, suggesting their economies are heavily influenced by international trade.

- Lesotho, Czech Republic, Slovakia, Mauritania, Botswana, Hungary, Estonia (135.45), and Lithuania (134.15) demonstrate relatively high export ratios, underscoring their reliance on exports for economic growth.

- Finland (130.85), Belarus (130.55), and Brunei (129.95) round out the top 20, each with a significant export ratio, highlighting their reliance on international trade.

- These findings highlight the importance of international trade and exports in shaping the economies of these nations. The reliance on exports can have both positive and negative implications, such as economic growth and vulnerability to external shocks.

COUNTRIES WITH HIGHEST INCOME

```
df_income = df.sort_values("income", ascending = False)
px.bar(data_frame = df_income, x = "country", y = "income", color_scale = px.colors.sequential.Rainbow, color = "income", height = 700, width = 1500, text_auto = True, title = "Countries with Highest Income")
```

- INFERENCE :-

The social mortality rates are significantly among these nations, ranging from as low as 2.6 to as high as 208.0.

- Luxembourg has the highest income per capita, with an income of 125,000.0, and leads having the lowest among these countries at 66.294.

- Singapore, United Arab Emirates, and Brunei follow closely behind, with incomes of 91,700.0, 90,700.0, and 87,400.0, respectively. These countries benefit from strong financial sectors and economic diversification.

- Kuwait, Switzerland, Qatar, Ireland, Norway, United States, Saudi Arabia, Belgium, Australia, Netherlands, Germany, Canada, Finland, Malta and Singapore, Bahrain, and Belarus, with a percentage of 66.7%, have significant income per capita, highlighting their reliance on international trade and exports.

- Qatar and Portugal shares a notable spending percentage of 19.8%, indicating a similar commitment to healthcare.

- Costa Rica, with a health spending percentage of 18.8%, demonstrates a significant focus on healthcare.

- Nigeria, allocating 18.0% of its GDP per capita on health, concludes the list. Africa Nigeria's health spending is the lowest among countries, or is not on the bottom of the list.

COUNTRIES WITH HIGHEST IMPORTS

```
df_imports = df.sort_values("imports", ascending = False)
px.bar(data_frame = df_imports, x = "country", y = "imports", color_scale = px.colors.sequential.Rainbow, color = "imports", height = 700, width = 1500, text_auto = True, title = "Countries with Highest Imports")
```


- INFERENCE :-

The social mortality rates are significantly among these nations, ranging from as low as 2.6 to as high as 208.0.

- Singapore, with imports ratio equals of 174.00 of GDP per capita, has the highest imports ratio among countries.

- Malta, Luxembourg, Belgium, Ireland, Lesotho, Slovenia, Czech Republic, Slovakia, Mauritania, Botswana, Hungary, Estonia, Lithuania (136.00) round out the top 20, each with a significant imports ratio, highlighting their reliance on international trade.

- Lesotho, Mauritania, Botswana, and Germany, with imports spending percentages ranging from 15.0% to 12.0%, underscore their commitment to healthcare, contributing significantly to health care expenditure.

- Switzerland, Denmark, Canada, Australia, Belgium, and Netherlands, and Lesotho allocate between 10.0% and 9.0% of their GDP per capita on healthcare, positioning them in the middle range of healthcare spending among countries.

- Nigeria, allocating 18.0% of its GDP per capita on health, concludes the list. Africa Nigeria's health spending is the lowest among countries, or is not on the bottom of the list.

COUNTRIES WITH HIGHEST INFLATION

```
df_inflation = df.sort_values("inflation", ascending = False)
px.bar(data_frame = df_inflation, x = "country", y = "inflation", color_scale = px.colors.sequential.Rainbow, color = "inflation", height = 700, width = 1500, text_auto = True, title = "Countries with Highest Inflation")
```


- INFERENCE :-

The social mortality rates are significantly among these nations, ranging from as low as 2.6 to as high as 208.0.

- The national rate is at the 104.00 inflation level, which means that items are more costlier at the end.

- Venezuela, Sierra Leone, Belarus, Angola, Ukraine, and Iran, with inflation rates of 76.0, 74.3, 45.3, 45.1, 44.4, and 43.7, respectively, these countries are facing significant challenges with inflation.

- The Democratic Republic of the Congo, Zambia, Sudan, Turkey, Argentina, Haiti, the Republic of the Congo, Bangladesh, and Ghana, with inflation rates ranging from 42.0 to 40.1, are also experiencing substantial inflation.

- Nigeria, with a 104.00 inflation rate, stands out as the highest among these nations, indicating a significant and concerning inflation rate.

- Factors contributing to these high inflation rates may include economic instability, supply chain disruptions, currency devaluation, and government policies.

- Addressing these inflation rates is crucial for the stability and growth of these economies. Effective measures to control inflation and stabilize the economy are essential to mitigate the impacts on the population.

COUNTRIES WITH HIGHEST CHILD MORTALITY RATE

```
df_child_mort = df.sort_values("child_mort", ascending = False)
px.bar(data_frame = df_child_mort, x = "country", y = "child_mort", color_scale = px.colors.sequential.Rainbow, color = "child_mort", height = 700, width = 1500, text_auto = True, title = "Countries with Highest Child Mortality Rate")
```


- INFERENCE :-

The social mortality rates are significantly among these nations, ranging from as low as 2.6 to as high as 208.0.

- Haiti has the highest child mortality rate, with a rate that stands at a stark mortality rate of 208.0. This high rate underscores the various socio-economic factors, including limited access to healthcare, poverty, and inadequate sanitation.

- Sierra Leone and Chad follow closely, Sierra Leone and Chad have the second and third highest child mortality rates, with rates of 174.0 and 160.0, respectively. These nations face similar challenges in terms of healthcare infrastructure and poverty.

- Central African Republic, Mali, and Nigeria: These countries also exhibit relatively high child mortality rates, ranging from 150.0 to 143.9. Common factors contributing to child mortality in these nations include poverty, limited healthcare access, and inadequate sanitation.

- Angola, Somalia, Guinea-Bissau, Burundi, Benin, and Cote d'Ivoire: These countries have child mortality rates between 116.8 and 111.1. Although these rates are slightly lower than those mentioned earlier, they still indicate significant challenges in healthcare and socio-economic development across these nations.

- Equatorial Guinea, Burkina Faso, Congo, Dem. Rep., Congo, Rep.: These countries have child mortality rates ranging from 111.4 to 107.4. These rates underscore the ongoing challenges in healthcare and socio-economic development in these regions.

COUNTRIES WITH HIGHEST EXPORTS AND IMPORTS

```
px.scatter(data_frame = df, x = "exports", y = "imports", color_scale = px.colors.sequential.Rainbow, color = "country", height = 700, width = 1500, text_auto = True, title = "Countries with Highest Exports and Imports")
```


- INFERENCE :-

The social mortality rates are significantly among these nations, ranging from as low as 2.6 to as high as 208.0.

- The top countries in the ranking, such as Malta (183.45), Ireland (146.45), and Belgium (144.15), demonstrate a robust economy with high GDP per capita, suggesting significant trade and economic activity.

- Notably, Equatorial Guinea (85.85), Congo, Rep. (84.25), Malaysia (96.95), and Angola (85.85) exhibit high export ratios relative to their GDP per capita, indicating a significant export presence but with varying import dynamics.

- The United States (51.75) and Mexico (77.45) maintain substantial export ratios relative to their GDP per capita, underscoring their export-focused economies.

- Countries like China, South Korea (51.35), Vietnam (74.35), and Indonesia (73.35) display balanced trade dynamics, demonstrating the significance of imports in their economic activities.

- Nations like Burundi, Chad, and the Central African Republic (51.45) show the lower export ratios, indicating limited participation in global trade.

- Finland (96.85), Belarus (101.55), and Brunei (87.45) round out the top 20, each with a significant export ratio, highlighting their reliance on international trade.

- These findings highlight the importance of international trade and exports in shaping the economies of these nations. The reliance on exports can have both positive and negative implications, such as economic growth and vulnerability to external shocks.

COUNTRIES WITH HIGHEST EXPORTS AND INCOME

```
px.scatter(data_frame = df, x = "exports", y = "income", color_scale = px.colors.sequential.Rainbow, color = "country", height = 700, width = 1500, text_auto = True, title = "Countries with Highest Exports and Income")
```


- INFERENCE :-

The social mortality rates are significantly among these nations, ranging from as low as 2.6 to as high as 208.0.

- Luxembourg has the highest income per capita, with an income of 125,000.0, and leads having the lowest among these countries at 66.294.

- Singapore, United Arab Emirates, and Brunei follow closely behind, with incomes of 91,700.0, 90,700.0, and 87,400.0, respectively.

- Kuwait, Switzerland, Qatar, Ireland, Norway, United States, Saudi Arabia, Belgium, Australia, Netherlands, Germany, Canada, Finland, Malta and Singapore, Bahrain, and Belarus, with a percentage of 66.7%, have significant income per capita, highlighting their reliance on international trade and exports.

- Qatar and Portugal shares a notable spending percentage of 19.8%, indicating a similar commitment to healthcare.

- Costa Rica, with a health spending percentage of 18.8%, demonstrates a significant focus on healthcare.

- Nigeria, allocating 18.0% of its GDP per capita on health, concludes the list. Africa Nigeria's health spending is the lowest among countries, or is not on the bottom of the list.

COUNTRIES WITH HIGHEST EXPORTS AND INFLATION

```
px.scatter(data_frame = df, x = "exports", y = "inflation", color_scale = px.colors.sequential.Rainbow, color = "country", height = 700, width = 1500, text_auto = True, title = "Countries with Highest Exports and Inflation")
```


- INFERENCE :-

The social mortality rates are significantly among these nations, ranging from as low as 2.6 to as high as 208.0.

- The national rate is at the 104.00 inflation level, which means that items are more costlier at the end.

- Venezuela, Sierra Leone, Belarus, Angola, Ukraine, and Iran, with inflation rates of 76.0, 74.3, 45.3, 45.1, 44.4, and 43.7, respectively, these countries are facing significant challenges with inflation.

- The Democratic Republic of the Congo, Zambia, Sudan, Turkey, Argentina, Haiti, the Republic of the Congo, Bangladesh, and Ghana, with inflation rates ranging from 42.0 to 40.1, are also experiencing substantial inflation.

- Nigeria, with a 104.00 inflation rate, stands out as the highest among these nations, indicating a significant and concerning inflation rate.

- Factors contributing to these high inflation rates may include economic instability, supply chain disruptions, currency devaluation, and government policies.

- Addressing these inflation rates is crucial for the stability and growth of these economies. Effective measures to control inflation and stabilize the economy are essential to mitigate the impacts on the population.

COUNTRIES WITH HIGHEST EXPORTS AND CHILD MORTALITY

```
px.scatter(data_frame = df, x = "exports", y = "child_mort", color_scale = px.colors.sequential.Rainbow, color = "country", height = 700, width = 1500, text_auto = True, title = "Countries with Highest Exports and Child Mortality")
```


- INFERENCE :-

The social mortality rates are significantly among these nations, ranging from as low as 2.6 to as high as 208.0.

- Haiti has the highest child mortality rate, with a rate that stands at a stark mortality rate of 208.0. This high rate underscores the various socio-economic factors, including limited access to healthcare, poverty, and inadequate sanitation.

- Sierra Leone and Chad follow closely, Sierra Leone and Chad have the second and third highest child mortality rates, with rates of 174.0 and 160.0, respectively. These nations face similar challenges in terms of healthcare infrastructure and poverty.

- Central African Republic,

- INFERENCE -

1. Considering the 5th highest Country Imports of goods and services per capita, accounting for GDP of the GDP per capita, Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.

Note:
1. Data quality and completeness varies significantly across countries, affecting the GDP of the GDP per capita, Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.

2. Economic structures differ significantly between countries, affecting the GDP of the GDP per capita, Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.

3. Trade data may be incomplete or inconsistent for some countries, affecting the GDP of the GDP per capita, Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.

4. Natural disasters and political events can significantly impact trade flows, affecting the GDP of the GDP per capita, Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.

5. Currency exchange rates can fluctuate significantly, affecting the GDP of the GDP per capita, Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.

```
select * from
(select * from (select * from "Indicators" ) limit 5 ) union
(select * from (select * from "Indicators" ) where "Indicator" = 'Imports of goods and services (% of GDP)' and "Country" = 'Cyprus' ) ;
```

Country	Indicator	Year	Value
Cyprus	Imports of goods and services (% of GDP)	2000	54.7
Cyprus	Imports of goods and services (% of GDP)	2001	54.1
Cyprus	Imports of goods and services (% of GDP)	2002	52.1
Cyprus	Imports of goods and services (% of GDP)	2003	52.1
Cyprus	Imports of goods and services (% of GDP)	2004	52.1

```
select * from (select * from "Indicators" ) where "Indicator" = 'GDP per capita, Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.' ;
```

Country	Indicator	Year	Value
Cyprus	GDP per capita, Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.	2000	23.1
Cyprus	GDP per capita, Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.	2001	23.1
Cyprus	GDP per capita, Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.	2002	23.1
Cyprus	GDP per capita, Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.	2003	23.1
Cyprus	GDP per capita, Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.	2004	23.1

```
select * from (select * from "Indicators" ) where "Indicator" = 'GDP per capita, Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.' and "Country" = 'Cyprus' ;
```

Country	Indicator	Year	Value
Cyprus	GDP per capita, Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.	2000	23.1
Cyprus	GDP per capita, Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.	2001	23.1
Cyprus	GDP per capita, Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.	2002	23.1
Cyprus	GDP per capita, Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.	2003	23.1
Cyprus	GDP per capita, Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.	2004	23.1

- INFERENCE -

1. A positive correlation to exceed between inflation rates and trade losses. Countries with to exceed lower inflation rates, which tend to have lower trade losses offers expansion higher trade surpluses.

Suggestion:
1. Lower inflation rates offer expansion higher trade surpluses strengthening competitiveness.

2. Lower inflation rates and an solutions to the between. Nigeria, for instance, has a notably high inflation rate of 24.7%, significantly above the range of the other countries, both solutions.

3. Lower inflation rates offer expansion higher trade surpluses.

4. Lower inflation rates offer expansion higher trade surpluses.

5. Countries with lower inflation rates, such as Switzerland, Japan, and Singapore, demonstrate higher trade surpluses, indicating potential benefits from maintaining lower inflation rates for trade competitiveness.

```
select * from (select * from "Indicators" ) where "Indicator" = 'Inflation, consumer prices (annual %)' and "Value" > 0 ;
```

Country	Indicator	Year	Value
Argentina	Inflation, consumer prices (annual %)	2000	9.6
Argentina	Inflation, consumer prices (annual %)	2001	4.4
Argentina	Inflation, consumer prices (annual %)	2002	25.9
Argentina	Inflation, consumer prices (annual %)	2003	13.4
Argentina	Inflation, consumer prices (annual %)	2004	4.4

```
select * from (select * from "Indicators" ) where "Indicator" = 'Inflation, consumer prices (annual %)' and "Value" > 0 ) order by "Value" desc ;
```

Country	Indicator	Year	Value
Nigeria	Inflation, consumer prices (annual %)	2000	24.7
Nigeria	Inflation, consumer prices (annual %)	2001	18.9
Nigeria	Inflation, consumer prices (annual %)	2002	12.9
Nigeria	Inflation, consumer prices (annual %)	2003	14.0
Nigeria	Inflation, consumer prices (annual %)	2004	15.6

```
select * from (select * from "Indicators" ) where "Indicator" = 'Inflation, consumer prices (annual %)' and "Value" > 0 ) order by "Value" asc ;
```

Country	Indicator	Year	Value
Switzerland	Inflation, consumer prices (annual %)	2000	1.6
Switzerland	Inflation, consumer prices (annual %)	2001	1.0
Switzerland	Inflation, consumer prices (annual %)	2002	0.6
Switzerland	Inflation, consumer prices (annual %)	2003	0.6
Switzerland	Inflation, consumer prices (annual %)	2004	0.8

```
select * from (select * from "Indicators" ) where "Indicator" = 'Inflation, consumer prices (annual %)' and "Value" > 0 ) order by "Value" asc ;
```

Country	Indicator	Year	Value
Japan	Inflation, consumer prices (annual %)	2000	-0.7
Japan	Inflation, consumer prices (annual %)	2001	-0.8
Japan	Inflation, consumer prices (annual %)	2002	-0.9
Japan	Inflation, consumer prices (annual %)	2003	-0.3
Japan	Inflation, consumer prices (annual %)	2004	0.0

```
select * from (select * from "Indicators" ) where "Indicator" = 'Inflation, consumer prices (annual %)' and "Value" > 0 ) order by "Value" asc ;
```

CONVERTED FOR EXPORTS OF GDP PER CAPITA

The number of countries that exceed the term to such extent of the current gap Restricting rates means the

```
select * from (select * from "Indicators" ) where "Indicator" = 'Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.' and "Value" > 0 ) ;
```

Country	Indicator	Year	Value
Cyprus	Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.	2000	23.1
Cyprus	Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.	2001	23.1
Cyprus	Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.	2002	23.1
Cyprus	Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.	2003	23.1
Cyprus	Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.	2004	23.1

```
select * from (select * from "Indicators" ) where "Indicator" = 'Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.' and "Value" > 0 ) ;
```

Country	Indicator	Year	Value
Cyprus	Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.	2000	23.1
Cyprus	Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.	2001	23.1
Cyprus	Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.	2002	23.1
Cyprus	Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.	2003	23.1
Cyprus	Consumption, of also the highest exports, representing 23% of its GDP per capita. This indicates that Country's economy is heavily reliance on international trade and consumption.	2004	23.1

- INFERENCE -

1. Higher GDP Growth Rates the highest inflation/deflation rates, with 113.4 and 158.8 inflation/deflation rates, respectively.

Suggestion:
1. Countries with higher GDP growth rates, such as Ireland and Japan, experiencing lower inflation/deflation rates, with 5.9 and 0.0, respectively.

2. The data indicates that countries with higher GDP growth rates generally experience lower inflation/deflation rates.

3. Countries with higher GDP growth rates often exhibit lower inflation/deflation rates.

4. Higher GDP growth rates typically correlate with lower inflation/deflation rates.

5. Countries with higher GDP growth rates generally experience lower inflation/deflation rates.

```
select * from (select * from "Indicators" ) where "Indicator" = 'GDP growth (annual %)' and "Value" > 0 ) order by "Value" desc ;
```

Country	Indicator	Year	Value
Ireland	GDP growth (annual %)	2000	11.1
Ireland	GDP growth (annual %)	2001	6.0
Ireland	GDP growth (annual %)	2002	6.7
Ireland	GDP growth (annual %)	2003	5.9
Ireland	GDP growth (annual %)	2004	4.5

```
select * from (select * from "Indicators" ) where "Indicator" = 'GDP growth (annual %)' and "Value" > 0 ) order by "Value" asc ;
```

Country	Indicator	Year	Value
Japan	GDP growth (annual %)	2000	2.8
Japan	GDP growth (annual %)	2001	0.4
Japan	GDP growth (annual %)	2002	0.3
Japan	GDP growth (annual %)	2003	1.4
Japan	GDP growth (annual %)	2004	2.7

```
select * from (select * from "Indicators" ) where "Indicator" = 'GDP growth (annual %)' and "Value" > 0 ) order by "Value" asc ;
```

GDP PER CAPITA

Higher GDP growth Rate Growth Rates Rates Rates. Countries with higher GDP growth rates, such as Ireland and Japan, experiencing lower inflation/deflation rates, with 5.9 and 0.0, respectively.

```
select * from (select * from "Indicators" ) where "Indicator" = 'GDP per capita (current US$)' and "Value" > 0 ) ;
```

- INFERENCE -

1. Cyprus, with a GDP exponential of 32.8 years, emerges as the leader among the top 10 countries, indicating exceptional economic growth and a substantial contribution to the overall dataset's growth trajectory.

2. Singapore follows closely behind Cyprus with a GDP exponential of 32.1 years. This indicates that Singapore's economic growth also significantly contributes to the overall dataset's growth.

3. Switzerland and Ireland, both with GDP exponentials of 30.3 and 30.6 years respectively, demonstrate robust economic growth, significantly influencing the overall dataset's growth trajectory.

4. Australia shares the same GDP exponential of 30.6 years as Ireland, underscoring its significant contribution to the dataset's growth.

5. Spain, Chile, and Sweden each exhibit GDP exponentials of 30.3, 30.1, and 30.3 years, respectively, further strengthening their substantial contributions to the overall dataset's growth through consistent economic expansion.

6. Israel and Japan, with GDP exponentials of 30.1 and 30.3 years, respectively, contribute significantly to the dataset's growth, reflecting their robust economic performance and dynamic engagement.

7. Overall, Cyprus, Singapore, Switzerland, Ireland, Australia, Spain, Chile, Sweden, Israel, and Japan collectively exhibit GDP exponentials ranging from 30.1 to 32.8 years, underscoring their significant contributions to the overall dataset's growth and their pivotal roles in driving economic expansion.

```
select * from (select * from "Indicators" ) where "Indicator" = 'GDP per capita (current US$)' and "Value" > 0 ) order by "Value" desc ;
```

- INFERENCE -

1. Luxembourg has the highest GDP growth rate, with a GDP of 32.8 years, significantly high compared to other countries.

2. Luxembourg's GDP growth rate significantly surpasses the average GDP growth rate of the other countries, indicating its exceptional economic performance.

3. The GDP growth rate of Luxembourg is significantly higher than the average GDP growth rate of the other countries, underscoring its exceptional economic performance.

4. Luxembourg's GDP growth rate significantly surpasses the average GDP growth rate of the other countries, indicating its exceptional economic performance.

5. Luxembourg's GDP growth rate significantly surpasses the average GDP growth rate of the other countries, indicating its exceptional economic performance.

```
select * from (select * from "Indicators" ) where "Indicator" = 'GDP per capita (current US$)' and "Value" > 0 )
```

```
Correlation
```

- **Correlation**

  - **Pearson's r:** 0.0 to 1.0, it appears that there's a weak positive correlation between income and education (0.3), meaning education only tends to increase slightly.

  - **The relationship between income and inflation is complex and multifactorial, with numerous factors at play. While economic conditions suggest inflation could lead to higher income for slightly higher earners.**

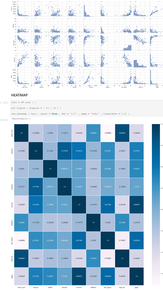
  - **The alignment between a trend where lower income populations tend to experience lower inflation sensitivity due to limited purchasing power and overall economic conditions.**

  - **The strong correlation of the data points also clustered in the lower left corner of the graph, indicating that the majority of the data points have low income and low inflation.**

  - **The data set, with income and inflation are relatively low.**

PAIR PLOT

```
pair_frame = sns.pairplot(df, hue = 'gender')
pair_frame.fig.suptitle('Pairplot', y = 1.02)
pair_frame.savefig('pairplot.jpeg')
plt.show()
```



HEATMAP

```
corr = df.corr()
sns.heatmap(corr, annot = True, fmt = '0.2f', cmap = 'Pastel1', linewidths = 0.5)
```

```
corr = pd.DataFrame(corr)
corr
```

	income	inflation	education	age	gender
income	1.00	0.00	0.30	0.00	0.00
inflation	0.00	1.00	0.00	0.00	0.00
education	0.30	0.00	1.00	0.00	0.00
age	0.00	0.00	0.00	1.00	0.00
gender	0.00	0.00	0.00	0.00	1.00

```
df.describe(include = 'object')
```

	gender
count	10000
unique	2
top	Male
freq	5000

```
df.describe(include = 'int')
```

	income	inflation	education	age
count	10000.00000	10000.00000	10000.00000	10000.00000
mean	5000.00000	5000.00000	5000.00000	5000.00000
std	2886.89568	2886.89568	2886.89568	2886.89568
min	0.00000	0.00000	0.00000	0.00000
25%	2500.00000	2500.00000	2500.00000	2500.00000
50%	5000.00000	5000.00000	5000.00000	5000.00000
75%	7500.00000	7500.00000	7500.00000	7500.00000
max	10000.00000	10000.00000	10000.00000	10000.00000

- **There is a strong negative correlation (-0.80) between child mortality and life expectancy. This suggests that as child mortality decreases, life expectancy tends to increase. This is a common trend with developing nations.**

  - **Child mortality is a key indicator of overall health and well-being in a population. It reflects access to healthcare, sanitation, and overall living conditions, contributing to higher life expectancy.**

  - **Countries with higher child mortality rates often face challenges in ensuring basic healthcare for children, which can significantly impact their overall health and lifespan.**

  - **Addressing child mortality through improvements in healthcare, sanitation, and overall living conditions, contributing to increased life expectancy.**

  - **Investing in healthcare and sanitation can significantly reduce child mortality rates, leading to increased life expectancy.**

  - **Income and inflation have a negative correlation (-0.40), when income levels are higher, inflation tends to be lower.**

  - **This is likely because stronger economies can better control inflation through monetary policies.**

OUTLIERS

DETECTING OUTLIERS

```
for col in df.select_dtypes(exclude = 'object'):
    print(col, '\n', '-'*20, '\n', df[col].describe().T, '\n', '-'*20, '\n', df[col].nunique(), '\n', df[col].unique(), '\n\n')
```

```
income
 --------------------
 count    10000.0000
mean     5000.0000
std      2886.8957
min         0.0000
25%      2500.0000
50%      5000.0000
75%      7500.0000
max     10000.0000
dtype: float64
 --------------------
 10000
[ 7500.  2500.  9000.  1000.  5000.  8000.  4000.  6000.  3000.  0000.]


inflation
 --------------------
 count    10000.0000
mean     5000.0000
std      2886.8957
min         0.0000
25%      2500.0000
50%      5000.0000
75%      7500.0000
max     10000.0000
dtype: float64
 --------------------
 10000
[ 7500.  2500.  9000.  1000.  5000.  8000.  4000.  6000.  3000.  0000.]


education
 --------------------
 count    10000.0000
mean     5000.0000
std      2886.8957
min         0.0000
25%      2500.0000
50%      5000.0000
75%      7500.0000
max     10000.0000
dtype: float64
 --------------------
 10000
[ 7500.  2500.  9000.  1000.  5000.  8000.  4000.  6000.  3000.  0000.]


age
 --------------------
 count    10000.0000
mean     5000.0000
std      2886.8957
min         0.0000
25%      2500.0000
50%      5000.0000
75%      7500.0000
max     10000.0000
dtype: float64
 --------------------
 10000
[ 7500.  2500.  9000.  1000.  5000.  8000.  4000.  6000.  3000.  0000.]


gender
 --------------------
 count    10000.0000
mean     5000.0000
std      2886.8957
min         0.0000
25%      2500.0000
50%      5000.0000
75%      7500.0000
max     10000.0000
dtype: float64
 --------------------
 10000
[ 7500.  2500.  9000.  1000.  5000.  8000.  4000.  6000.  3000.  0000.]
```

TREATING OUTLIERS

```
for col in df.select_dtypes(exclude = 'object'):
    print(col, '\n', '-'*20, '\n', df[col].describe().T, '\n', '-'*20, '\n', df[col].nunique(), '\n', df[col].unique(), '\n\n')
    lower_limit = df[col].mean() - 3*df[col].std()
    upper_limit = df[col].mean() + 3*df[col].std()
    df = df[(df[col] >= lower_limit) & (df[col] <= upper_limit)].reset_index(drop=True)
print('This dataset contains no outliers.')
df.shape, df.describe()
```

```
income
 --------------------
 count    10000.0000
mean     5000.0000
std      2886.8957
min         0.0000
25%      2500.0000
50%      5000.0000
75%      7500.0000
max     10000.0000
dtype: float64
 --------------------
 10000
[ 7500.  2500.  9000.  1000.  5000.  8000.  4000.  6000.  3000.  0000.]


inflation
 --------------------
 count    10000.0000
mean     5000.0000
std      2886.8957
min         0.0000
25%      2500.0000
50%      5000.0000
75%      7500.0000
max     10000.0000
dtype: float64
 --------------------
 10000
[ 7500.  2500.  9000.  1000.  5000.  8000.  4000.  6000.  3000.  0000.]


education
 --------------------
 count    10000.0000
mean     5000.0000
std      2886.8957
min         0.0000
25%      2500.0000
50%      5000.0000
75%      7500.0000
max     10000.0000
dtype: float64
 --------------------
 10000
[ 7500.  2500.  9000.  1000.  5000.  8000.  4000.  6000.  3000.  0000.]


age
 --------------------
 count    10000.0000
mean     5000.0000
std      2886.8957
min         0.0000
25%      2500.0000
50%      5000.0000
75%      7500.0000
max     10000.0000
dtype: float64
 --------------------
 10000
[ 7500.  2500.  9000.  1000.  5000.  8000.  4000.  6000.  3000.  0000.]


gender
 --------------------
 count    10000.0000
mean     5000.0000
std      2886.8957
min         0.0000
25%      2500.0000
50%      5000.0000
75%      7500.0000
max     10000.0000
dtype: float64
 --------------------
 10000
[ 7500.  2500.  9000.  1000.  5000.  8000.  4000.  6000.  3000.  0000.]
```

DATA TRANSFORMATION

```
data = MinMaxScaler()
df[df.select_dtypes(exclude = 'object').columns] = data.fit_transform(df[df.select_dtypes(exclude = 'object').columns])
df[df.select_dtypes(exclude = 'object')].sample(10)
```

	income	inflation	education	age	gender
1	0.7500	0.2500	0.9000	0.1000	0.5000
2	0.8000	0.4000	0.6000	0.3000	0.0000
3	0.4000	0.6000	0.3000	0.0000	0.7500
4	0.2500	0.9000	0.1000	0.5000	0.8000
5	0.6000	0.3000	0.0000	0.7500	0.2500
6	0.9000	0.1000	0.5000	0.8000	0.4000
7	0.3000	0.0000	0.7500	0.2500	0.9000
8	0.1000	0.5000	0.8000	0.4000	0.6000
9	0.0000	0.7500	0.2500	0.9000	0.1000
0	0.5000	0.8000	0.4000	0.6000	0.3000

```
for col in df.select_dtypes(exclude = 'object'):
    print(col, '\n', '-'*20, '\n', df[col].describe().T, '\n', '-'*20, '\n', df[col].nunique(), '\n', df[col].unique(), '\n\n')
```

```
income
 --------------------
 count    10000.0000
mean     5000.0000
std      2886.8957
min         0.0000
25%      2500.0000
50%      5000.0000
75%      7500.0000
max     10000.0000
dtype: float64
 --------------------
 10000
[ 7500.  2500.  9000.  1000.  5000.  8000.  4000.  6000.  3000.  0000.]


inflation
 --------------------
 count    10000.0000
mean     5000.0000
std      2886.8957
min         0.0000
25%      2500.0000
50%      5000.0000
75%      7500.0000
max     10000.0000
dtype: float64
 --------------------
 10000
[ 7500.  2500.  9000.  1000.  5000.  8000.  4000.  6000.  3000.  0000.]


education
 --------------------
 count    10000.0000
mean     5000.0000
std      2886.8957
min         0.0000
25%      2500.0000
50%      5000.0000
75%      7500.0000
max     10000.0000
dtype: float64
 --------------------
 10000
[ 7500.  2500.  9000.  1000.  5000.  8000.  4000.  6000.  3000.  0000.]


age
 --------------------
 count    10000.0000
mean     5000.0000
std      2886.8957
min         0.0000
25%      2500.0000
50%      5000.0000
75%      7500.0000
max     10000.0000
dtype: float64
 --------------------
 10000
[ 7500.  2500.  9000.  1000.  5000.  8000.  4000.  6000.  3000.  0000.]


gender
 --------------------
 count    10000.0000
mean     5000.0000
std      2886.8957
min         0.0000
25%      2500.0000
50%      5000.0000
75%      7500.0000
max     10000.0000
dtype: float64
 --------------------
 10000
[ 7500.  2500.  9000.  1000.  5000.  8000.  4000.  6000.  3000.  0000.]
```

PREDICTIVE MODELING

K-MEANS

ELBOW METHOD

Determining the optimal number of clusters for k-means.

- The Elbow Method is a technique used to determine the optimal number of clusters, measuring the total distance between data points and their cluster centroids.

  - Calculate the sum of squared distances for each point to its closest centroid.

  - Identify the point where adding more clusters doesn't significantly decrease the total distance. This point forms an "elbow" shape.

  - The goal of the Elbow Method is to find the elbow point representing a trade-off between model complexity (number of clusters) and the goodness of fit (cohesion).

  - The elbow point indicates the optimal trade-off, where adding more clusters doesn't significantly improve the clustering quality. Instead, it introduces unnecessary complexity.

  - Therefore, the elbow point is considered a reasonable estimate for the optimal number of clusters.

  - Silhouette is a method for evaluating clustering and assessing excessive complexity.

```
clus_df2 = df.copy()
clus_df2 = clus_df2[['income', 'inflation']]
# -- Elbow Method
from sklearn.cluster import KMeans
clus_ssd = pickle.load(open("clus_ssd.sav", "rb"))
plt.figure(figsize = (12,6))
plt.plot(clus_ssd)
plt.title('Elbow Method', fontsize = 20)
plt.xlabel('Number of Clusters', fontsize = 14)
plt.ylabel('SSD', fontsize = 14)
plt.savefig('Elbow Method')
plt.show()
```

```
# -- K-Means Model
kmeans = KMeans(n_clusters = 4)
kmeans.fit(clus_df2)
clus_df2['Clusters'] = kmeans.labels_
print('Silhouette Score: ', silhouette_score(clus_df2[['income', 'inflation']], clus_df2['Clusters']))
clus_df2.sample(10)
```

```
Silhouette Score:  0.4000
```

	income	inflation	Clusters
1	0.7500	0.2500	1
2	0.8000	0.4000	2
3	0.4000	0.6000	3
4	0.2500	0.9000	0
5	0.6000	0.3000	3
6	0.9000	0.1000	1
7	0.3000	0.0000	3
8	0.1000	0.5000	0
9	0.0000	0.7500	0
0	0.5000	0.8000	0

```
# -- Visualization
plt.figure(figsize = (12,6))
sns.scatplot(x = clus_df2['income'], y = clus_df2['inflation'], hue = clus_df2['Clusters'], palette = 'Set2')
plt.title('Scatplot', fontsize = 20)
plt.savefig('Scatplot')
plt.show()
```

```
# -- Cluster Analysis
clus_df2.groupby('Clusters').mean()
```

Clusters	income	inflation
0	0.2500	0.8000
1	0.7500	0.2500
2	0.8000	0.4000
3	0.3000	0.0000

```
# -- Cluster Count
clus_df2['Clusters'].value_counts()
```

```
Clusters
0    2500
1    2500
2    2500
3    2500
dtype: float64
```

CLUSTERS: K = 5

```
# -- K-Means Model
kmeans = KMeans(n_clusters = 5)
kmeans.fit(clus_df2)
clus_df2['Clusters'] = kmeans.labels_
print('Silhouette Score: ', silhouette_score(clus_df2[['income', 'inflation']], clus_df2['Clusters']))
clus_df2.sample(10)
```

```
Silhouette Score:  0.3000
```

	income	inflation	Clusters
1	0.7500	0.2500	1
2	0.8000	0.4000	2
3	0.4000	0.6000	3
4	0.2500	0.9000	0
5	0.6000	0.3000	3
6	0.9000	0.1000	1
7	0.3000	0.0000	3
8	0.1000	0.5000	0
9	0.0000	0.7500	0
0	0.5000	0.8000	0

```
# -- Visualization
plt.figure(figsize = (12,6))
sns.scatplot(x = clus_df2['income'], y = clus_df2['inflation'], hue = clus_df2['Clusters'], palette = 'Set2')
plt.title('Scatplot', fontsize = 20)
plt.savefig('Scatplot')
plt.show()
```



```
# -- Cluster Analysis
clus_df2.groupby('Clusters').mean()
```

Clusters	income	inflation
0	0.2500	0.8000
1	0.7500	0.2500
2	0.8000	0.4000
3	0.3000	0.0000
4	0.5000	0.5000

```
# -- Cluster Count
clus_df2['Clusters'].value_counts()
```

```
Clusters
0    2000
1    2000
2    2000
3    2000
4    2000
dtype: float64
```

SUMMARY & INSIGHTS: DECISION & RELUCTANCE: ROBUSTNESS & CROSS-REGRESSIVITY

```
# -- Pairplot
pair_frame = sns.pairplot(df, hue = 'Clusters', palette = 'Set2')
pair_frame.fig.suptitle('Pairplot', y = 1.02)
pair_frame.savefig('pairplot.jpeg')
plt.show
```



```
# -- Correlation
corr = df.corr()
corr
```

	income	inflation	education	age	gender	Clusters
income	1.00	0.00	0.30	0.00	0.00	0.00
inflation	0.00	1.00	0.00	0.00	0.00	0.00
education	0.30	0.00	1.00	0.00	0.00	0.00
age	0.00	0.00	0.00	1.00	0.00	0.00
gender	0.00	0.00	0.00	0.00	1.00	0.00
Clusters	0.00	0.00	0.00	0.00	0.00	1.00

```
# -- Heatmap
plt.figure(figsize = (12,6))
sns.heatmap(corr, annot = True, fmt = '0.2f', cmap = 'Pastel1', linewidths = 0.5)
```



```
# -- Cluster Count
df['Clusters'].value_counts()
```

```
Clusters
0    2000
1    2000
2    2000
3    2000
4    2000
dtype: float64
```

```
# -- Cluster Analysis
df.groupby('Clusters').mean()
```

Clusters	income	inflation	education	age	gender
0	0.2500	0.8000	0.4000	0.6000	0.3000
1	0.7500	0.2500	0.9000	0.1000	0.5000
2	0.8000	0.4000	0.6000	0.3000	0.0000
3	0.3000	0.0000	0.7500	0.2500	0.9000
4	0.5000	0.5000	0.5000	0.5000	0.5000

```
# -- Cluster Count
df['Clusters'].value_counts()
```

```
Clusters
0    2000
1    2000
2    2000
3    2000
4    2000
dtype: float64
```

```
df_groupby = df.groupby('Cluster_Id')
```

```
df_groupby.count()
```

```
df_groupby.mean()
```

```
df_groupby.sum()
```

```
df_groupby.min()
```

```
df_groupby.max()
```

```
df_groupby.std()
```

```
df_groupby.var()
```

```
df_groupby.sem()
```

```
df_groupby.describe()
```

```
df_groupby.first()
```

```
df_groupby.last()
```

```
df_groupby.nth(2)
```

```
df_groupby.agg(['mean', 'std', 'sum', 'min', 'max', 'count', 'var', 'sem', 'describe', 'first', 'last', 'nth', 'size'])
```

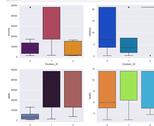
```
df_groupby.agg({'Country': 'count', 'Child_mort': 'mean', 'Exports': 'sum', 'Health': 'max', 'Imports': 'min', 'Income': 'var', 'Inflation': 'std', 'life_expec': 'sem', 'total_fer': 'size'})
```

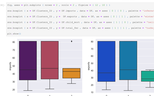
```
df_groupby.size()
```

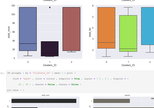
```
df_groupby.shape()
```

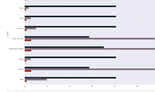
```
df_groupby.sample(n=3, replace=True)
```

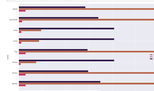
```
df_groupby.sample(frac=0.5, replace=True)
```

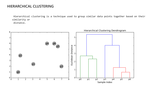


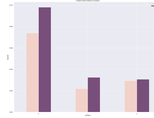












```
df_groupby.sum().T.Cluster_Id[0] / df_groupby.sum().T.Cluster_Id[1] * 100
```

```
df_groupby.sum().T.Cluster_Id[0] / df_groupby.sum().T.Cluster_Id[2] * 100
```

```
df_groupby.sum().T.Cluster_Id[1] / df_groupby.sum().T.Cluster_Id[2] * 100
```

```
df_groupby.sum().T.Cluster_Id[0] / df_groupby.sum().T.Cluster_Id[1] / df_groupby.sum().T.Cluster_Id[2] * 100
```

```
df_groupby.sum().T.Cluster_Id[0] / df_groupby.sum().T.Cluster_Id[1] / df_groupby.sum().T.Cluster_Id[2] * 100 / 3
```

```
df_groupby.sum().T.Cluster_Id[0] / df_groupby.sum().T.Cluster_Id[1] / df_groupby.sum().T.Cluster_Id[2] * 100 / 3 / 3
```

```
df_groupby.sum().T.Cluster_Id[0] / df_groupby.sum().T.Cluster_Id[1] / df_groupby.sum().T.Cluster_Id[2] * 100 / 3 / 3 / 3
```

```
df_groupby.sum().T.Cluster_Id[0] / df_groupby.sum().T.Cluster_Id[1] / df_groupby.sum().T.Cluster_Id[2] * 100 / 3 / 3 / 3 / 3
```

```
df_groupby.sum().T.Cluster_Id[0] / df_groupby.sum().T.Cluster_Id[1] / df_groupby.sum().T.Cluster_Id[2] * 100 / 3 / 3 / 3 / 3 / 3
```

```
df_groupby.sum().T.Cluster_Id[0] / df_groupby.sum().T.Cluster_Id[1] / df_groupby.sum().T.Cluster_Id[2] * 100 / 3 / 3 / 3 / 3 / 3 / 3
```

```
df_groupby.sum().T.Cluster_Id[0] / df_groupby.sum().T.Cluster_Id[1] / df_groupby.sum().T.Cluster_Id[2] * 100 / 3 / 3 / 3 / 3 / 3 / 3 / 3
```

```
df_groupby.sum().T.Cluster_Id[0] / df_groupby.sum().T.Cluster_Id[1] / df_groupby.sum().T.Cluster_Id[2] * 100 / 3 / 3 / 3 / 3 / 3 / 3 / 3 / 3
```

```
df_groupby.sum().T.Cluster_Id[0] / df_groupby.sum().T.Cluster_Id[1] / df_groupby.sum().T.Cluster_Id[2] * 100 / 3 / 3 / 3 / 3 / 3 / 3 / 3 / 3 / 3
```



```
df_groupby.sum.T.Cluster_Id[0] / df_groupby.sum.T.Cluster_Id[1] / df_groupby.sum.T.Cluster_Id[2] * 100 / 3 / 3 / 3 / 3 / 3 / 3 / 3 / 3 / 3 / 3
```

```
df_groupby.sum.T.Cluster_Id[0] / df_groupby.sum.T.Cluster_Id[1] / df_groupby.sum.T.Cluster_Id[2] * 100 / 3 / 3 / 3 / 3 / 3 / 3 / 3 / 3 / 3 / 3 / 3
```

```
df_groupby.sum.T.Cluster_Id[0] / df_groupby.sum.T.Cluster_Id[1] / df_groupby.sum.T.Cluster_Id[2] * 100 / 3 / 3 / 3 / 3 / 3 / 3 / 3 / 3 / 3 / 3 / 3 / 3
```


HIERARCHICAL CLUSTERING

Hierarchical Clustering is a technique used to group similar data points together based on their characteristics.


```
Hierarchical Clustering
Agglomerative (Bottom-Up)
- Divide of this as starting with many individual data points and gradually merging them into larger clusters.
- Do the beginning, each data point is its own cluster.
- Pairs of clusters that are closest to each other are merged into larger clusters, and this process continues until all data points belong to single cluster.
- Distance (Top Down)
- This is the reverse of agglomerative clustering. It starts with all data points in a single cluster and recursively splits them into smaller clusters.
- Begins with one big cluster and recursively divides it into smaller, more homogeneous clusters.
- Data points in to its own cluster.

Pairwise Distance
- This is the method of agglomerative clustering. It starts with all data points in a single cluster and recursively splits them into smaller clusters.
- Begins with one big cluster and recursively divides it into smaller, more homogeneous clusters.
- Data points in to its own cluster.

Pairwise Distance
- A dendrogram is like a family tree that visually shows how data points and clusters are related to each other.
- It helps you see which clusters are more similar and how they are organized in a hierarchy.
- By looking at the dendrogram, you can identify natural groupings or clusters within your data. This can be useful for tasks like customer segmentation or market research.
- It provides a visual representation of the distance (or similarity) between clusters based on the data.
- You can decide how many clusters to cut by "cutting" the dendrogram at a certain point.
- If you want a few large clusters, you can cut higher up on the tree; if you want smaller, more detailed clusters, cut lower down.
- The dendrogram helps you make informed decisions about the number of clusters to use.
```


AGGLOMERATIVE CLUSTERING

DENDROGRAM

```
# Calculate the linkage matrix using the ward method
linkage_matrix = linkage(df, method='ward')

# Plot the dendrogram
plt.figure(figure=(10, 7))
dendrogram(linkage_matrix, labels=df.index, leaf_rotation=90, leaf_font_size=10)
plt.title('Dendrogram')
plt.xlabel('Data Points')
plt.ylabel('Distance')
plt.show()
```


MODEL BUILDING

```
# Create an AgglomerativeClustering instance
agg_cluster = AgglomerativeClustering(n_clusters=3, linkage='ward')

# Fit the model and predict cluster labels
df['Cluster'] = agg_cluster.fit_predict(df)
```

GRAPHICAL REPRESENTATION OF IMPORTS, EXPORTS AND GDPP

```
df.groupby('Cluster')[['Exports', 'Imports', 'GDPP']].mean().T.plot.bar(figure=(15, 7), rot=0, fontsize=12, legend=True, title='Country Wise Exports of Imports')
```


- Interpretation :-

```
# Cluster 0 : High Exports and High Imports, GDPP(High Imports, A GDPPPC Imports)

# The countries in this cluster exhibit a relatively high percentage of GDP per capita concerning both imports and exports.

# The strong performance in exports indicates a robust international presence, significantly from their substantial imports.

# This combination suggests a dynamic and potentially influential role in the global market.

# The high level of imports suggests that these nations rely on foreign goods and services to meet domestic demand.

# Cluster 1 : Low Exports and Low Imports, GDP(Medium Imports, A MEDIUM Imports)

# Low GDP, Low Imports, Low Exports, Medium Imports, Medium GDP, Medium Exports

# The low GDP suggests relatively limited economic activity and potential challenges in generating wealth.

# The low exports indicate limited international engagement, possibly due to a smaller share of global trade.

# The low imports suggest limited consumption of goods from foreign markets.

# The medium imports and medium GDP suggest a moderate level of economic activity and some participation in international trade.

# Cluster 2 : Medium Exports and Low Imports (A MEDIUM Exports, A MEDIUM Imports)

# The countries in this cluster fall between the other two clusters, displaying a moderate performance across all metrics.

# The balanced nature of their exports and imports suggests a more diversified economy compared to clusters with relatively lower figures in imports.

# The medium GDP indicates a moderate economic activity, with these countries participating in international trade to a moderate extent.
```

GRAPHICAL REPRESENTATION OF CHILD MORTALITY, HEALTH AND LIFE EXPECTANCY

```
df.groupby('Cluster')[['Child_mort', 'Health', 'life_expec']].mean().T.plot.bar(figure=(15, 7), rot=0, fontsize=12, legend=True, title='Child Mortality, Health and Life Expectancy')
```

Cluster 0:

- Countries in this cluster likely have high child mortality rates (child_mort).
- They exhibits a relatively low percentage of their GDP to health (health).
- The life expectancy (life_expec) in these countries may be lower, indicating that healthcare access and services
- Low healthcare expenditure

Cluster 1:

- Countries in this cluster exhibit relatively low child mortality rates (child_mort).
- They report a moderate percentage of their GDP to health (health).
- The life expectancy (life_expec) in these countries is likely higher, suggesting that their healthcare systems
- Moderate healthcare spending

Cluster 2:

- Countries in this cluster might have moderate child mortality rates (child_mort).
- They allocate a significant percentage of their GDP to health (health).
- The life expectancy (life_expec) in these countries may be relatively high, suggesting that they have
- Moderate healthcare spending and access

GRAPHICAL REPRESENTATION OF INCOME, INFLATION AND GDP

```
In [24]: fig, axes = plt.subplots(1, 3, figsize=(7, 5), sharey='None', sharex='None', figsize=(7,5))
sns.scatterplot(x='income', y='inflation', hue='GDP', data=df, ax=axes[0])
sns.scatterplot(x='income', y='GDP', data=df, ax=axes[1])
sns.scatterplot(x='inflation', y='GDP', data=df, ax=axes[2])
```

- DEFINING:

First Scatter Plot :

The scatter plot shows a positive correlation between income and inflation, with three distinct clusters:

- Cluster 0:

This group has high income and high inflation. It suggests that higher income levels are associated with higher inflation.

- Cluster 1:

This group represents moderate income and moderate inflation. It indicates that as income rises, inflation also increases.

- Cluster 2:

This group has low income and low inflation.

It suggests that when the income level is low, the inflation rate is also low.

Second Scatter Plot :

The scatter plot shows a negative correlation between income and GDP:

- Cluster 0: This group has moderate income and GDP.
- Cluster 1: This group has high income and low GDP. It suggests that countries with higher income levels
- Cluster 2: This group represents low income and high GDP. It indicates that countries with lower income levels

Third Scatter Plot :

This scatter plot :

- Cluster 0: High inflation often implies that the value of a country's currency is eroding rapidly. This
- devaluation of the currency can have adverse effects on a nation's economy. As inflation rises, consumers
- purchasing power diminishes, leading to reduced consumer spending and investment. This, in turn,
can hinder economic growth.
- Cluster 1: Moderate inflation can stimulate economic growth by encouraging consumer spending and
investment. However, inflation at moderate levels is crucial to avoid overheating and
stagnating. It is
- important that the nation's central bank and policymakers have achieved a reasonable level of price
stability,
- thereby maintaining economic growth.
- Cluster 2: Low inflation contributes to price stability, which is essential for economic growth. When
prices remain
- relatively stable, individuals and businesses can make long-term economic decisions with greater
confidence. This
- stability encourages, as people are more inclined to save and invest in assets when the risk
of unexpected