KLE Technological University Belagavi

A Course Project Report on

"GLOBAL POP ANALYZER"

A Course Project Report Submitted in Partial Fulfillment of the Requirement for the Course of
Exploratory Data Analysis
in
4th Semester of Computer Science and Engineering

By

Lubna Shaikh	02FE21BCS114
Pushkarani Pujari	02FE21BCS066
Sachin Kangralkar	02FE21BCS072
Sankalp Shamanewadi	02FE21BCS075

Under the guidance of

Prof. Priyanka Gavade
Assistant Professor,
Department of Computer Science and Engineering,
KLE Technological University's Dr. MSSCET, Belagavi.

KLE Technological University's
Dr. M. S. Sheshgiri College of Engineering and Technology, Belagavi – 590008.

July 2023

KLE
TECHNOLOGICAL UNIVERSITY
Creating Value, Leveraging Knowledge
— Belagavi Campus —

Dr.M.S.Sheshgiri College of Engineering & Technology

Department of Computer Science & Engineering

DECLARATION

We hereby declare that the matter embodied in this report entitled "**GLOBAL POP ANALYZER**" submitted to KLE Technological University for the course completion of Exploratory Data Analysis (21ECSC210) in the 4th Semester of Computer Science and Engineering is the result of the work done by us in the Department of Computer Science and Engineering, KLE Dr. M. S. Sheshgiri College of Engineering, Belagavi under the guidance of Prof. Priyanka Gavade, Assistant Professor, Department of Computer Science and Engineering. We further declare that to the best of our knowledge and belief, the work reported here in doesn't form part of any other project on the basis of which a course or award was conferred on an earlier occasion on this by any other student(s), also the results of the work are not submitted for the award of any course, degree or diploma within this or in any other University or Institute. We hereby also confirm that all of the experimental work in this report has been done by us.

Belagavi – 590 008

Date :24-07-2023

Lubna Shaikh
(02FE21BCS114)

Pushkarani Pujari
(02FE21BCS066)

Sachin Kangralkar
(02FE21BCS072)

Sankalp Shamanewadi
(02FE21BCS075)

KLE
TECHNOLOGICAL UNIVERSITY
Creating Value, Leveraging Knowledge
—— Belagavi Campus ——

Dr.M.S.Sheshgiri College of Engineering & Technology

Department of Computer Science & Engineering

CERTIFICATE

This is to certify that the project entitled **"GLOBAL POP ANALYZER"** submitted to KLE Technological University's Dr. MSSCET, Belagavi for the partial fulfillment of the requirement for the course - Exploratory Data Analysis (21ECSC210) by **Lubna**, **Pushkarani**, **Sachin** and **Sankalp**, students in the Department of Computer Science and Engineering, KLE Technological University's Dr. MSSCET, Belagavi, is a bonafide record of the work carried out by them under my supervision. The contents of this report, in full or in parts, have not been submitted to any other Institute or University for the award of any other course completion.

Belagavi – 590008

Prof. Priyanka Gavade   Dr. Rajashri Khanai

(Course Coordinator)   (Head of the Department)

Abstract

The world's population has been a subject of paramount importance due to its significant growth and its profound impact on various aspects of society and the environment. In this report, we explore a comprehensive dataset sourced from Kaggle, encompassing population data from the years 1970 to 2022. The dataset includes valuable information such as population density, growth rate, and world population percentage for countries and continents across the globe.

The analysis of the world population dataset from 1970 to 2022 provides valuable insights into global population trends. The dataset's wealth of information offers a comprehensive understanding of population growth, density, and distribution across countries and continents. Armed with these insights, policymakers and stakeholders can make informed decisions to address the challenges and opportunities presented by the world's evolving population dynamics. As the global population continues to grow, strategic planning and concerted efforts are essential to ensure a prosperous and sustainable future for humanity.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Since 1970, the world's population has increased substantially. Since 1970, the world's population has increased substantially. The world's population was 3.7 billion people in 1970. The population of the globe reached 8 billion people in 2022. Several factors, such as the following, have contributed to this growth:

1.Healthcare advancements: Since 1970, there have been tremendous advancements in healthcare, which have decreased infant death rates and increased life expectancy. Because of this, more people have reached reproductive age, which has aided in population expansion.

2.Economic growth: Population expansion has been aided by economic growth. Countries tend to have lower birth rates as they develop. However, not all developing nations have followed this trend, and some have had tremendous population expansion

3.Modifications to fertility patterns: Since 1970, fertility trends have also shifted. In many countries, the number of children per woman has declined. However, this trend has not been universal, and some countries have experienced increases in fertility rates.

1.2 Problem Statement

Between 1970 and 2022, the population of the world has increased exponentially, with significant ramifications for the economy, society, and environment. To address the challenges posed by this rapid growth and to prepare for the future, this study aims to analyze the world population dataset and understand the key elements that have influenced population growth over this period.

1.2.1   Objectives

The primary objective of this report is to provide a comprehensive understanding of global population trends from 1970 to 2022, using the World Population dataset from World Bank. The specific objectives include:

1. To analyze the historical data and identify long-term trends in global population growth, including potential periods of accelerated or decelerated growth and to explore the variations in population density across different continents and regions and understand the distribution of the world's population.
2. To investigate the factors influencing population growth, such as birth rates, death rates, and migration patterns, and assess their impact on population dynamics.
3. To assess the environmental implications of population growth, including resource utilization, pollution, and climate change effects.
4. To examine the economic consequences of population growth, including its effects on labor markets, consumption patterns, and economic development.
5. To project future global population trends based on historical growth rates and demographic indicators, providing insights for future planning and policy development.
6. To propose sustainable strategies and recommendations for population management, considering the environmental, economic, and social dimensions.

Chapter 2

Knowing the Dataset

This chapter describes the details of the dataset used in this course project. It first details on the dataset properties and then presents a few of the statistical analysis that we performed on the dataset.

2.1   Dataset

- The dataset consists of 234 sample spaces and 17 features.

Data set: Worldbank (https://data.worldbank.org/)

- Below is the snapshot of the data set:

	Rank	CCA3	Country	Capital	Continent	2022 Popu	2020 Popu	2015 Popu	2010 Popu	2000 Popu	1990 Popu	1980 Popu	1970 Popu	Area (km/	Density (p	Growth Ra	World Population Percentage
1																	
2	36	AFG	Afghanist	Kabul	Asia	4.1E+07	3.9E+07	3.4E+07	2.8E+07	2E+07	1.1E+07	1.2E+07	1.1E+07	652230	63.0587	1.0257	0.52
3	138	ALB	Albania	Tirana	Europe	2842321	2866849	2882481	2913399	3182021	3295066	2941651	2324731	28748	98.8702	0.9957	0.04
4	34	DZA	Algeria	Algiers	Africa	4.5E+07	4.3E+07	4E+07	3.6E+07	3.1E+07	2.6E+07	1.9E+07	1.4E+07	2381741	18.8531	1.0164	0.56
5	213	ASM	American	Pago Pago	Oceania	44273	46189	51368	54849	58230	47818	32886	27075	199	222.477	0.9831	NaN
6	203	AND	Andorra	Andorra l	Europe	79824	77700	71746	71519	66097	53569	35611	19860	468	170.564	1.01	NaN
7	42	AGO	Angola	Luanda	Africa	3.6E+07	3.3E+07	2.8E+07	2.3E+07	1.6E+07	1.2E+07	8330047	6029700	1246700	28.5466	1.0315	0.45
8	224	AIA	Anguilla	The Valley	North Am	15857	15585	14525	13172	11047	8316	6560	6283	91	174.253	1.0066	NaN
9	201	ATG	Antigua a	Saint John	North Am	93763	92664	89941	85695	75055	63328	64888	64516	442	212.134	1.0058	NaN
10	33	ARG	Argentina	Buenos Ai	South Am	4.6E+07	4.5E+07	4.3E+07	4.1E+07	3.7E+07	3.3E+07	2.8E+07	2.4E+07	2780400	16.3683	1.0052	0.57
11	140	ARM	Armenia	Yerevan	Asia	2780469	2805608	2878595	2946293	3168523	3556539	3135123	2534377	29743	93.4831	0.9962	0.03
12	198	ABW	Aruba	Oranjesta	North Am	106445	106585	104257	100341	89101	65712	62267	59106	180	591.361	0.9991	NaN
13	55	AUS	Australia	Canberra	Oceania	2.6E+07	2.6E+07	2.4E+07	2.2E+07	1.9E+07	1.7E+07	1.5E+07	1.3E+07	7692024	3.4032	1.0099	0.33
14	99	AUT	Austria	Vienna	Europe	8939617	8907777	8642421	8362829	8010428	7678729	7547561	7465301	83871	106.588	1.002	0.11
15	91	AZE	Azerbaijar	Baku	Asia	1E+07	1E+07	9863480	9237202	8190337	7427836	6383060	5425317	86600	119.608	1.0044	0.13
16	176	BHS	Bahamas	Nassau	North Am	409984	406471	392697	373272	325014	270679	223752	179129	13943	29.4043	1.0051	0.01
17	154	BHR	Bahrain	Manama	Asia	1472233	1477469	1362142	1213645	711442	517418	362595	222555	765	1924.49	1.0061	0.02
18	8	BGD	Banglades	Dhaka	Asia	1.7E+08	1.7E+08	1.6E+08	1.5E+08	1.3E+08	1.1E+08	8.4E+07	6.8E+07	147570	1160.04	1.0108	2.15
19	186	BRB	Barbados	Bridgetow	North Am	281635	280693	278083	274711	264657	258868	253575	241397	430	654.965	1.0015	NaN
20	96	BLR	Belarus	Minsk	Europe	9534954	9633740	9700609	9731427	1E+07	1E+07	9817257	9170786	207600	45.9295	0.9955	0.12
21	81	BEL	Belgium	Brussels	Europe	1.2E+07	1.2E+07	1.1E+07	1.1E+07	1E+07	9959560	9828986	9629376	30528	381.811	1.0038	0.15
22	177	BLZ	Belize	Belmopan	North Am	405272	394921	359871	322106	240406	182589	145133	120905	22966	17.6466	1.0131	0.01
23	77	BEN	Benin	Porto-Nov	Africa	1.3E+07	1.3E+07	1.1E+07	9445710	6998023	5133419	3833939	3023443	112622	118.564	1.0274	0.17

Fig. (1) Dataset

2.2　　　Features of the Dataset

VARIABLES	DATA TYPES	DESCRIPTION
RANK	INT	RANK OF THE COUNTRIES (ACCORDING POPULATION).
CCA3	OBJECT	CODE OF EACH COUNTRY.
COUNTRY	OBJECT	NAMES OF ALL THE COUNTRIES.
CONTINENT	OBJECT	NAMES OF THE CONTINET TO WHICH THE COUNTRIES BELONG.
POPULATION(1970-2022)	INT	POPULATION OF THE COUNTRIES FROM 1970 TO 2022.
AREA	INT	AREA OF EACH COUNTRY IN (KM)^2
DENSITY	FLOAT	DENSITY OF EACH COUNTRY IN (PER KM^2).
GROWTH RATE	FLOAT	TOTAL GROWTH RATE OF EACH COUNTRY.
WORLD POPULATION PERCENTAGE	FLOAT	EACH COUNTRIES CONTRIBUTION TO THE WORLD'S POPULATION (IN %)

Table (1) Features of dataset

2.3　　Observations

- How are the features? All categorical? Mix?

  - We have mixed features in our dataset.

- Are there any missing values? If yes, are they large or small?

  - There are small missing values in our data set.

- Are there any outliers?

  - Yes, there are outliers in our data set.

- Overall what are the characteristics of your dataset?

  - In our dataset, there is more categorical data

2.4   Statistical Data Analysis

The dataset encompasses population-related attributes for diverse countries and continents over five decades.

Statistical Findings:

1. Population Growth Trend: The world's population has steadily increased from approximately 3.7 billion in 1970 to over 7.9 billion in 2022. We observed variations in growth rates over the years, with periods of accelerated and decelerated growth.
2. Continental Analysis: Asia emerged as the most populous continent, consistently contributing a significant percentage to the world's total population. Africa exhibited the highest population growth rate, leading to a considerable increase in its population share.
3. Population Density: Some countries, particularly those with limited land area and large populations, demonstrated remarkably high population densities. Others, with extensive territories and lower populations, displayed lower population densities.
4. World Population Percentage Shifts: The dataset illustrated fluctuations in each country's share of the world's total population, reflecting changes in global demographics.

	Rank	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	1990 Population	1980 Population	1970 Population	Area (km²)	Density (p kr
count	234.000000	2.340000e+02	2.340000e+02	2.340000e+02	2.340000e+02	2.340000e+02	2.340000e+02	2.340000e+02	2.340000e+02	2.340000e+02	234.0000
mean	117.500000	3.407441e+07	3.350107e+07	3.172996e+07	2.984524e+07	2.626947e+07	2.271022e+07	1.898462e+07	1.578691e+07	5.814494e+05	452.1270
std	67.694165	1.367664e+08	1.355899e+08	1.304050e+08	1.242185e+08	1.116982e+08	9.783217e+07	8.178519e+07	6.779509e+07	1.761841e+06	2066.1219
min	1.000000	5.100000e+02	5.200000e+02	5.640000e+02	5.960000e+02	6.510000e+02	7.000000e+02	7.330000e+02	7.520000e+02	1.000000e+00	0.0261
25%	59.250000	4.197385e+05	4.152845e+05	4.046760e+05	3.931490e+05	3.272420e+05	2.641158e+05	2.296142e+05	1.559970e+05	2.650000e+03	38.4178
50%	117.500000	5.559944e+06	5.493074e+06	5.307400e+06	4.942770e+06	4.292907e+06	3.825410e+06	3.141146e+06	2.604830e+06	8.119950e+04	95.3467
75%	175.750000	2.247650e+07	2.144798e+07	1.973085e+07	1.915957e+07	1.576230e+07	1.186923e+07	9.826054e+06	8.817329e+06	4.304258e+05	238.9332
max	234.000000	1.425887e+09	1.424930e+09	1.393715e+09	1.348191e+09	1.264099e+09	1.153704e+09	9.823725e+08	8.225344e+08	1.709824e+07	23172.2667

Fig. (2) Central tendency

Chapter 3

Implement Framework

The implementation framework for EDA on world population data can be divided into the following steps:

- Data collection: The first step is to collect data on the world's population. This data has come from sources such as government reports and surveys.
- Data cleaning: Once the data has been collected, it is important to clean it to remove any errors or inconsistencies. This includes checking for missing values, outliers, and duplicate records. The data should be cleaned in a way that preserves the integrity of the data and that allows for accurate analysis.
- Data exploration: The next step is to explore the data to understand its distribution and to identify patterns and trends. This can be done using a variety of statistical and visualization techniques. Some of the statistical techniques that can be used include:
    - Univariate analysis: This involves analyzing the distribution of each variable in the data set. This can be done using descriptive statistics, such as the mean, median, and standard deviation.
    - Bivariate analysis: This involves analyzing the relationship between two variables. This can be done using scatter plots, correlation coefficients, and regression models. Some of the visualization techniques that can be used include:
        1) Histograms: These can be used to visualize the distribution of a single variable.
        2) Box plots: These can be used to visualize the distribution of a single variable and identify outliers.
        3) Heatmaps: These can be used to visualize the relationship between multiple variables.
- Hypothesis generation: The final step is to generate hypothesis about how different factors are associated with global population.

Chapter 4

Data Pre-processing

Data pre-processing is a process of preparing data for further analysis. We have performed the following data preprocessing steps on the dataset:

1. Data cleaning
2. Data visualization

- Cleaning of data:
  We checked the dataset for missing values and discovered that there were 57 missing values in the column World Population Percentage. We have handled these missing values by filling the cells with the most probable value i.e., 0.
- Visualization of data:
  Utilized a range of graphical representations, including line plots, bar charts, boxplots, heatmap and scatter plots, to visualize population trends and patterns.

Chapter 5

Exploratory Data Analysis

5.1 Hypothesis on the Problem Statement

- How population of Asian countries has shown diverse growth patterns over time, including varying rates of growth?

- Investigate how population density and growth rates correlate with environmental factors like area and climate.

- Utilize growth rates to project future population sizes for countries or continents?

- Visualize the correlation between different variables in the given dataset by implementing it through a heatmap.

- Implement the association between population density and area for the top 10 overpopulated countries using a scatter plot.

- How did the population and growth rates of the various nations over the span of 50 years change (from 1970 to 2020)?

- Visualize the percentage of the world's population for the top most populous countries.

- Calculate the total area of each continent in square kilometers, shedding light on the geographical sizes of different regions and their significance in the global landmass distribution.

- Visualize the correlation between different variables in the given dataset by implementing it through a heatmap.

5.2   Analysis

Our hypothesis is as mentioned below and we have analyzed them using various techniques.

1. Demographic data parameters, distribution patterns of world population percentage, growth rate, population density, land area, and the projected population for the year 2022.
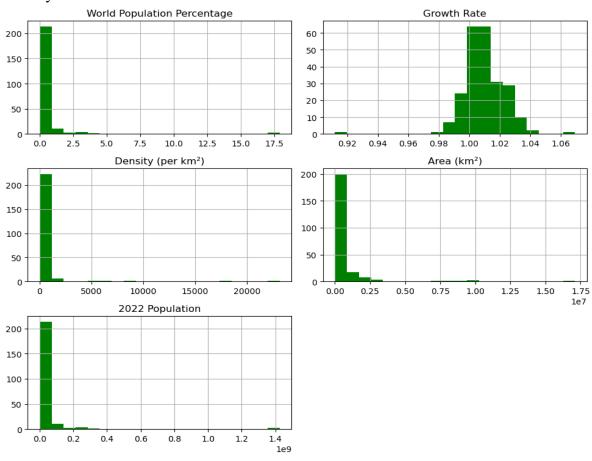

Fig. (3) Univariate Graph

These insights can be valuable for researchers, policymakers, and anyone interested in understanding global or regional population dynamics and demographics. Additionally, it can help identify areas that may require specific attention in terms of population management, urban planning, or resource allocation.

2. The population of different countries in the year 2020 and their corresponding growth percentages from 1970 to 2020.
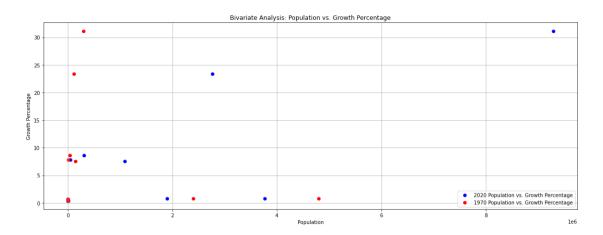
Fig. (4) Bivariate graph

The plot visually represents how the population of each country has grown from 1970 to 2020, with the x-axis showing the population size and the y-axis showing the growth percentage. The two separate scatter plots with different colors allow us to compare the growth patterns between the two time points

3. How population trend over time for multiple Asian countries ?
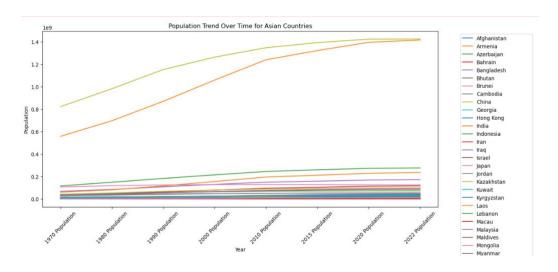
Fig. (5) Multivariate graph

The above graph allows us to observe how the populations of various countries have evolved and whether there are any significant trends or patterns across the region and allowing for easy comparison of population changes for different countries. give more explanation.

4. Gain insights into how the world's most populated countries are compared to the geographically largest countries.
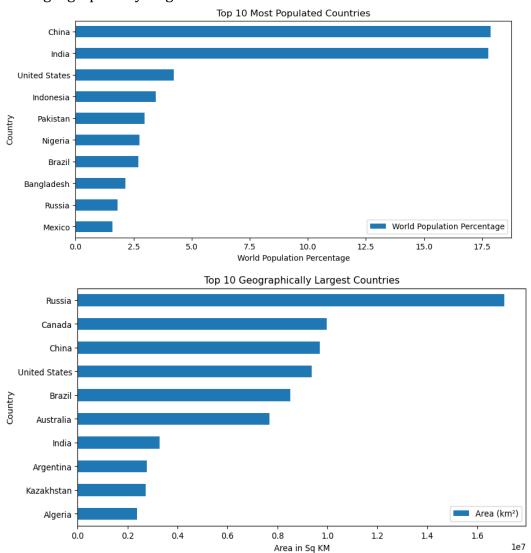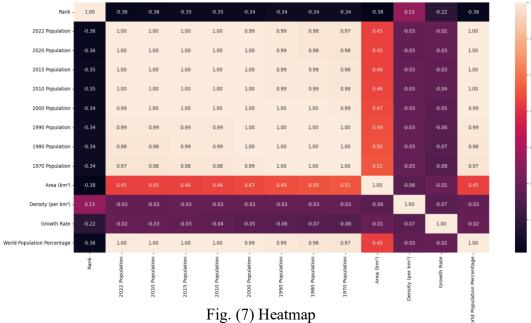

Fig. (6) Most populated countries compared to geographically largest countries.

By comparing the two histograms, we can gain insights into how the world's most populated countries compare to the geographically largest countries. It allows us to see if there is any correlation between a country's population size and its land area. Additionally, we can identify countries that have a relatively small population but cover a large land area or vice versa.

5. Visualize the relationships and dependencies between various numerical in the datasets.

Fig. (7) Heatmap

1.Population Correlation: The heatmap shows that various population-related variables, such as "2022 Population," "2020 Population," "2015 Population," "2010 Population," and so on, are highly positively correlated with each other.

2. World Population Percentage: The "World Population Percentage" variable has a high positive correlation with various population-related variables. This indicates that the population percentage of a country in the world population tends to increase or decrease along with its overall population.

3. Density Correlation: The heatmap reveals a positive correlation between "Density (per km²)" and "Area (km²)" variables, though it is not very strong. This suggests that countries with larger areas tend to have slightly lower population densities, and vice versa.

4.Rate Correlation: The "Growth Rate" variable exhibits a very low correlation with other numerical variables. Population CorrelationThe heatmap shows that various population-related variables.

2. World Population Percentage: The "World Population Percentage" variable has a high positive correlation with various population-related variables. This indicates that the population percentage of a country in the world population tends to increase or decrease along with its overall population.

6. Shows the distribution of population growth rates across different continents, the variability and central tendency of growth rates within each continent.
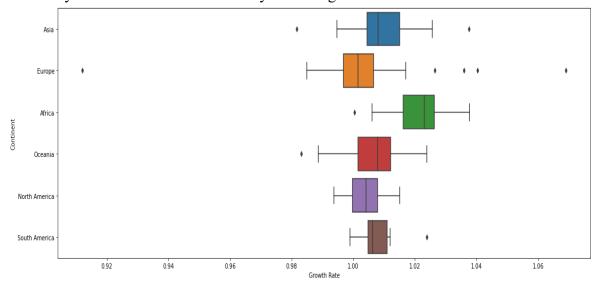

Fig. (8) Box Plot

Asia has the highest median growth rate: The horizontal line inside the box represents the median growth rate for Asian countries, and it is positioned higher compared to other continents, indicating a higher median growth rate for countries in Asia.

Africa and South America have relatively high variability:i This suggests that there is a higher variation in population growth rates across different African and South American countries.

Europe has the smallest variability: The box for "Europe" is relatively shorter, suggesting that the growth rates for European countries are less varied compared to other continents. The distribution of growth rates is more concentrated around the median in Europe.

Oceania has a large number of outliers:These data points are considered outliers, which means that certain Oceania countries have significantly higher or lower growth rates compared to the majority of other countries in the continent.

North America has a relatively lower median growth rate: The box for "North America" is positioned at a lower growth rate compared to other continents.

Chapter 6

Results and Outcomes

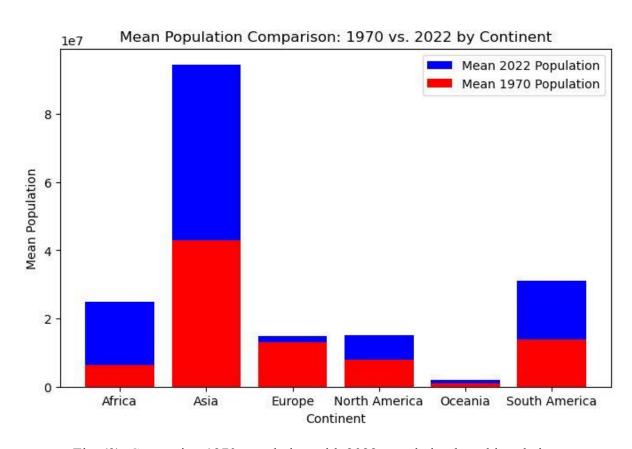1. Population changes over time for each continent

Fig. (9) Comparing 1970 population with 2022 population by taking their means

It helps to identify changes in population over time for each continent and to observe how the mean population values have evolved from 1970 to 2022. In conclusion, analyzing changes in population and mean population values over time can provide valuable insights into the social, economic, and environmental dynamics of different continents. Governments and policymakers can utilize this information to design and implement appropriate policies and strategies to address population-related issues, such as family planning, immigration, and regional development.

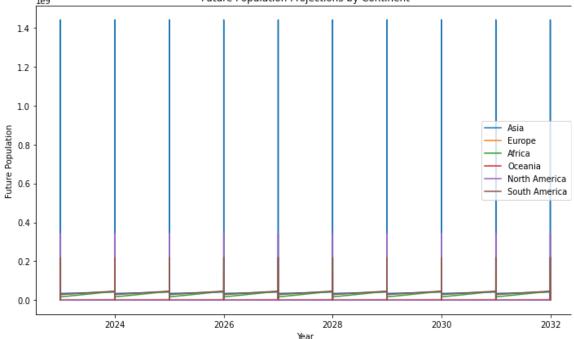2.Utilize growth rates to project future population sizes for countries or continents

Fig. (10) Predicting the future world population

For each year from 2023 to 2032 (i.e., '2022 + num_years'), calculate the future population for each country. The formula used is:

Future Population = Current Population * (1 + Growth Rate / 100)

This formula calculates the population for each year by adding the growth rate percentage to the current population. For example, if the growth rate is 2%, the future population for the next year would be 1.02 times the current population.

Overall, provides a useful tool for estimating and visualizing future population trends for specific areas based on the growth rates available in the dataset. However, users should interpret the results with caution, understanding that projections are only estimates and that real-world population changes may deviate from the projections due to various factors and uncertainties.

2. Investigate how population density and growth rates correlate with environmental factors like area and climate
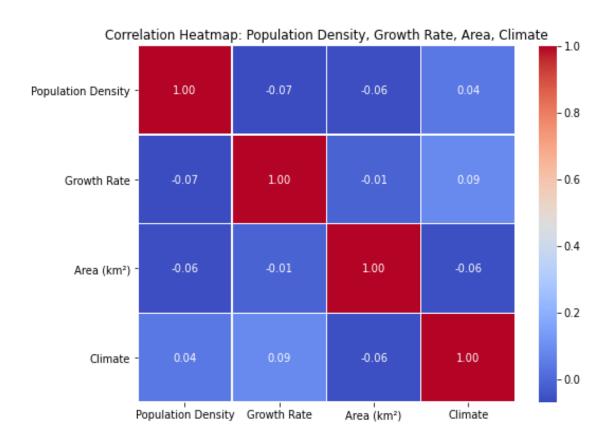

Fig. (11) How different variables correlate with climate.

Climate and Population: Investigate how population density and growth rates correlate with environmental factors like area and climate

In conclusion, this investigation can help uncover relationships between population variables and environmental factors, providing valuable information for sustainable development, urban planning, and policy formulation. By understanding these correlations, policymakers and researchers can make more informed decisions to address the challenges posed by population growth and environmental impacts.

Conclusions

Problem statement: Understanding the key elements influencing world population growth between 1970 and 2022 is crucial for addressing the challenges posed by this exponential increase. The insights gained from this study will serve as a valuable foundation for policymakers, researchers, and organizations to devise informed strategies and policies to promote sustainable development, manage resources effectively, and ensure a better future for global citizens.

Need of solving this problem is of utmost importance to ensure a harmonious balance between population growth and the well-being of humanity and the environment.   Climate Change and Environmental Impact: Population growth affects carbon emissions and environmental degradation. Analyzing the data will help comprehend the link between population growth and environmental pressures, enabling the formulation of climate change mitigation and adaptation strategies.

Sustainable Development: Population growth has direct implications for sustainable development. Studying the dataset will help identify regions experiencing rapid population growth, allowing for the allocation of resources and infrastructure development to accommodate the growing population sustainably.

Economic Planning: Population growth has a significant impact on economic development. This study will shed light on regions with high population growth rates, enabling policymakers to plan for increased labor force requirements, job creation, and economic diversification.

Results we got from this analysis : we were able to study the population analyze it and the analysis is expected to reveal the key elements influencing population growth, including birth rates, death rates, migration patterns, and socio-economic factors. It will provide insights into regions experiencing rapid population growth, potential demographic imbalances, and variations in population dynamics across different continents.

Conclusion: This investigation can help uncover relationships between population variables and environmental factors, providing valuable information for sustainable development, urban planning, and policy formulation. By understanding these correlations, policymakers and researchers can make more informed decisions to address the challenges posed by population growth and environmental impacts. The research will be valuable for governments, international organizations, and researchers working towards achieving a balanced and equitable global population growth trajectory and a sustainable future.

Bibliography

https://data.worldbank.org/

.