A STATISTICAL PERSPECTIVE ON IMPORTANCE OF UNEMPLOYMENT IN GLOBAL ECONOMIC MODELING

A Project Submitted to Bharathiar University

in Partial Fulfilment of the Requirements

for the Award of the Degree of

MASTER OF SCIENCE

IN

STATISTICS

Submitted By

MAHABHARATHI S

REG.NO.: 23STAC14

Under the Guidance of

Dr.S.JAYALAKSHMI

ASSISTANT PROFESSOR

DEPARTMENT OF STATISTICS

BHARATHIAR UNIVERSITY

COIMBATORE - 641 046

APRIL 2025

CERTIFICATE

Certificate

This is to certify that the project work entitled **"A STATISTICAL PERSPECTIVE ON IMPORTANCE OF UNEMPLOYMENT IN GLOBAL ECONOMIC MODELING"** submitted to Bharathiar University,Coimbatore,in partial fulfillment of the requirements for the award of the Degree of Master of Science in Statistics is a record of original research work done by **MAHA BHARATHI S (23STAC14)** during the period of her study in the Department of Statistics, Bharathiar University, under my supervision and guidance and the project work has not formed the basis for the award of any Degree / Diploma / Associateship / Fellowship or other similar title to any candidate of any University.

Signature of the Guide

(Dr.S.JAYALAKSHMI)

Head of the Department

(Dr.R.JAISANKAR)

Coimbatore
April 2025

DECLARATION

Declaration

I, **MAHABHARATHI S (23STAC14),** hereby declare that the project work entitled **" A STATISTICAL PERSPECTIVE ON IMPORTANCE OF UNEMPLOYMENT IN GLOBAL ECONOMIC MODELING"** submitted to the Bharathiar University, Coimbatore, in partial fulfillment of the requirements for the award of the Degree of Master of Science in Statistics is a record of original research work done by me under the supervision and guidance of **Dr.S. JAYALAKSHMI**, Assistant Professor, Department of Statistics, Bharathiar University, Coimbatore. This project work has not formed the basis for the award of any Degree, Diploma, Associateship, Fellowship or other similar title to any candidate of any University.

Signature of the Candidate

Signature of the Guide

(Dr.S.JAYALAKSHMI)

ACKNOWLEDGEMENT

Acknowledgement

I thank God almighty for giving me all the mental strength and the will power to pursue this task successfully against all odds. I take immense pleasure to convey my gratitude of thanks to all who helped me to make this project a successful one.

I would like to express my sincere gratitude and appreciation to my guide, **Dr.S.Jayalakshmi**, Assistant Professor in the Department of Statistics at Bharathiar University in Coimbatore, for initiating me to this area of my project topic and for her motivated leadership, helpful advice, and unwavering support in seeing this project through to completion.

I am very grateful to **Dr.R.Jaisankar**, Professor and Head, Department of Statistics, Bharathiar University, Coimbatore for providing me the necessary facilities to complete this work.

I extend my genuine regards and with sincere gratitude, heartfelt thanks to **Dr. R. Muthukrishnan**, Professor, **Dr. K. M. Sakthivel**, Professor, **Dr. V. Kaviyarasu**, Associate Professor, **Dr. K. Pradeepa Veerakumari**, Associate Professor, and **Dr. S. Gandhiya Vendhan**, Assistant Professor, for their continuous support throughout.

I sincerely thank our Former Head of the Department and Senior Professor **Dr.R.Vijayaraghavan** for giving guest lectures and sharing his knowledge and experience.

I render my sincere thanks to my beloved parents and my brother who have supported me for the whole time in completing the course and project work.

Last but not least, I convey my acknowledgement to my fellow friends who helped me out with their abilities.

-MAHABHARATHI S

CONTENTS

Contents

CHAPTER 1

Chapter 1

INTRODUCTION

Statistics is the study of the collection, analysis, interpretation, presentation, and organization of data. In other words, it is a mathematical discipline to collect, summarize data. Also, we can say that statistics is a branch of applied mathematics. However, there are two important and basic ideas involved in statistics; they are uncertainty and variation. The uncertainty and variation in different fields can be determined only through statistical analysis. The word of statistics has been derived from the 'status', which is Latin word or 'Statista' which is Italian word. In the18th century, Prof. G. F. Achenwall has been used it first time. These words were used for political state of the region during early period. The Italian word 'Statista' was used to maintain the records of census or data related to wealth of a state or nation. Successively, the meaning and usage of statistics extended and there onwards its nature also changed. For a common man, 'Statistics' means numerical information expressed in quantitative terms, which may relate to objects, subjects, activities, information, phenomena, or regions of space.

1.1 DEFINITION

The definition of statistics has been given by the different statistician in different ways. Some important definitions of statistics are given below; A.L. Bowley defined that "Statistics may be called the science of counting". He also said that "Statistics may rightly be called the science of average". According to Boddington "Statistics is the science of estimates and probabilities". According to Selligman "Statistics is the science which deals with the methods of collecting,

classifying, tabulation, comparing and interpreting numerical data collected to throw some light on any sphere of enquiry". Croxton and Cowden defined that "statistics as the collection, tabulation, presentation, analysis and interpretation of numerical data".

According to Merriam-Webster dictionary, statistics is defined as "classified facts representing the conditions of a people in a state – especially the facts that can be stated in numbers or any other tabular or classified arrangement". According to statistician Sir Arthur Lyon Bowley, statistics is defined as "Numerical statements of facts in any department of inquiry placed in relation to each other".

1.2 MATHEMATICAL STATISTICS

Mathematical statistics is the application of Mathematics to Statistics, which was initially conceived as the science of the state — the collection and analysis of facts about a country: its economy, and, military, population, and so forth. Mathematical techniques used for different analytics include mathematical analysis, linear algebra, stochastic analysis, differential equation and measure-theoretic probability theory.

1.3 HISTORY OF STATISTICS

The word 'Statistics' seems to have been 'derived from the Latin word' status' or the Italian word' statista' or the German word' statistik' each of which means a 'political state.In India, an efficient system of collecting official and administrative statistics existed even more than 2,000 years ago, in particular, during the reign of Chandra

- In India ,an efficient system of collecting official and administrative statistics existed even more than 2000 years ago, in particular, Chandra Gupta Maurya (324 -300 B.C.). From Kautilya's Arthshastra it is known that even before 300 B.C. a very good system of collecting 'Vital Statistics' and registration of births and deaths was in vogue. During Akbar's reign (1556 - 1605A.D.), Raja Todarmal , then land and revenue minister, maintained good records of land and agricultural statistics

- In Germany, the systematic collection of official statistics originated towards the end of the 18th century when, in order to have an idea of the relative strength of different German States, information regarding population and-output – industrial and agricultural - was collected.

- In England, statistics were the outcome of Napoleonic Wars. The Wars necessitated the systematic collection of numerical data to enable the government to assess the revenues and expenditure with greater precision and then to levy new taxes in order to meet the cost of war.

- Seventeenth century saw the origin of the vital statistics captain john grant of London (1960-1674), known as the father of vital statistics , was the first man to study the statistics of birth and deaths. computation of mortality tables and the calculation of expectation of life at different ages by a number of persons,viz casper newman, sir William petty(1623-1687),james Dodson, Dr.price,tomention only a few ,led to the idea of life Insurance and the first life insurance institution was founded in London in 1698

- The theoretical Development of the so-called modern statistics came during the mid seventeeth century with the introduction of 'theory of probability' and Theory of Games and Chance' the chief contributors being mathematicians and gamblers of france,Germany and England .The France mathematician Pascal (1623-1662),after lenthy correspondence withanother france mathematician p.fermat(1601-1665) solved the famous 'problem of points' posed by gambler chevalier de-mere. His study of the problem laid the foundation of the theory of probability which is the backbone of the modern theory of statistics. Pascal also investigated the properties of the coefficient of binomial expension and also invented mechanical computation machine. Other notable contributor in this field area ;James Bernoulli (1654-1705),who wrote the first treatise on the 'theory of probability';De-Moivre(1667-1754) who also worked on probabilities and annuities and published his important work "The Doctrine of Chances" in 1718,Laplace (1949-1827)who published in the 1782 his monumental work on the theory of probability ,and Gauss (1777-1855),perhaps the most original of all writers on statistical subjects ,who gave the 'principle of least square' and the 'normal law errors' later on ,most of the prominent mathematicians of 18th,19th,and 20th centuries ,viz., Euler, lagrange, Bayes, A.Markoff, Khintchin, Kolmogorov, to mention

only a few ,have made very outstanding contributions to the modern theory of probability. Modern Stalwarts in the development of the subject of the statistics to different disciplines. Francies Galton(1822-1921) pioneered the Study of 'Regression Analysis'in Biometry; Karl pearson(1857-1936) who founded the greatest statistical laboratory in England pioneered the study of 'correlation Analysis' His Chi-square Test of goodness of fit is the first and most important of the tests of significance in statistics .

- W.S Gosset with his t-test ushered in an era of exact sample test. perhaps most of the work in the statistical theory during the past few decades can be attributed to a single person sir Ronald Fisher (1890-1962) who applied statistics to variety of diversified fields such as genetics , biometry, psychology and education ,agriculture ,etc and who is rightly termed as the father of statistics . In Addition to enhancing the existing statistical theory he is the pioneer is estimation theory; exact sampling distributions ,analysis of variance and design of experiments. His contributions to the subject of statistics are described in the following word

  'R.A Fisher is the real giant in the development of the theory of statistics'

- The varied and outstanding contributions of R.A Fisher put the subject of statistics on a very firm footing and earned for it the status of full-fledged science. Indian Statisticians have also made notable contributions to the development of statistics in various diversified fields. The valuable contributions of P.C Mahalanobis and P.V Sukhatme (sample survys);R.C Bose,Panse,J.N Srivastava(Design of experiments in agriculture);S.N.Roy(Multivariate Analysis);C.R Rao(Statistical inference);parthasarathy(Theory of probability),to mention only a few ,have earned for india a high position in the world map of Statistics.

1.4 SCOPE OF STATISTICS

Statistics is used in many sectors such as psychology, geology, sociology, weather forecasting, probability and much more. The goal of statistics is to gain understanding from the data, it focuses on applications, and hence, it is distinctively considered as a mathematical science.

1.4.1 Statistics and Planning

In today's age of planning, statistics has become integral to governmental and organizational strategies. Governments world wide rely on statistical data to shape economic development plans and policies. This practice, commonly known as "the age of planning," showcases the vital role statistics plays in informed decision-making for the betterment of societies.

1.4.2 Statistics and Economics

The influence of statistical data and analysis in economics cannot be overstated. It aids in resolving complex economic issues, such as wage-price trends, inflation, and economic forecasting. The application of statistical techniques, including time series analysis and regression analysis, empowers economists to make accurate predictions and informed policy recommendations.

1.4.3 Statistics and Business

For businesses, statistics is an invaluable tool for understanding customer preferences and behavior. Through statistical analysis, business executives gain insights into market trends, customer demands, and purchasing patterns. This knowledge guides product development, marketing strategies, and customer relationship management, ensuring sustained growth and competitiveness.

1.4.4 Statistics and Industry

In the industrial sector, statistics plays a crucial role in quality control. Production engineering relies on statistical tools like inspection plans and control charts to assess whether products conform to specifications. This quality control ensures consistency and reliability in manufacturing processes, leading to improved product outcomes.

1.4.5 Statistics and Modern Science

In the realm of modern science, statistics aids in the collection and analysis of data related to medical advancements. In the field of medical science, statistical

tools are used to study disease incidence, drug efficacy, and treatment outcomes. This information is essential for evidence-based medical practices and improving patient care.

1.5 CHARACTERISTICS OF STATISTICS

- Aggregate of Facts: Single and non-connected facts or figures are not statistics, rather when the facts are aggregates, they are said to be statistics, as they can be compared.

- Affected to a substantial extent by a variety of reasons: This means that statistics are influenced to a substantial extent by a number of factors that operate together. For example, The statistics of rice production is based on various factors like a method of cultivation, climatic conditions, seeds, fertilizers and manures, etc.

- Numerical expression: Statistics are expressed in terms of numbers. Therefore, qualitative expressions such as happy, sad, right, wrong, good, or bad do not amount to statistics. For example: 'Production of ABC ltd. has risen' is not statistics, but 'Production of ABC ltd. has risen from 92000 units in 2020 to 110000 units in 2021' is statistics.

- Enumerated and Estimated as per reasonable standard of accuracy: Reasonable accuracy needs to be there in the statistical data, as it acts as a basis for the field of statistical enquiry. This is because, if the scope of the inquiry is narrow, then by using the method of actual counting, the data can be collected, whereas if the scope of inquiry is wide then the data collection will be based on estimate and estimates can be inaccurate.

- Data collection is carried out in a systematic manner: The collection of statistics should be performed in a systematic as well as planned manner, because in the absence of any system, the data collected can be unreliable and inaccurate, which may also lead to misleading conclusions. Further, the purpose for its collection needs to be stated beforehand to keep its usefulness intact.

- Data must be placed in relation to one another: Data collection is performed for the purpose of comparison and so the basis must be homogeneous. Be-

cause when the basis of two units is heterogeneous, the comparison is not possible.

1.6 TYPES OF STATISTICS

There are the two broad ways of classifying statistics, one is the on the basis of function and other one is the on the basis of distribution of data.

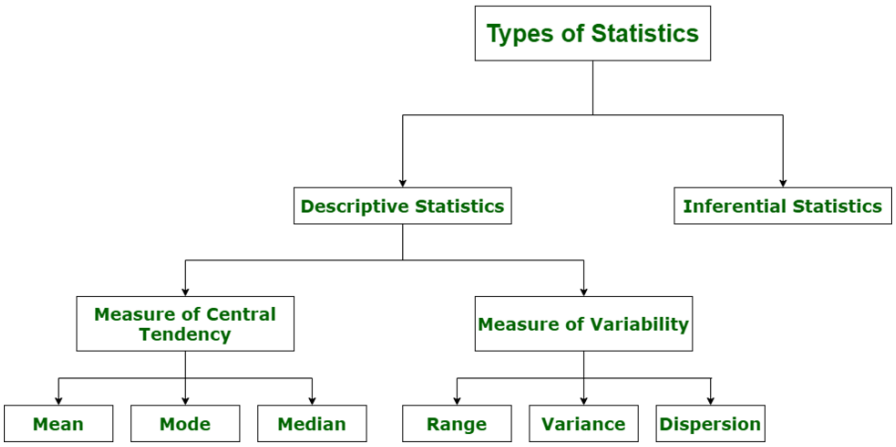

Figure 1.1: Types of Statistics

1.6.1 Statistics Based on Function

There are two types of statistics on the basis of subject matter function.

- Descriptive Statistics
- Inferential Statistics

In the case of descriptive statistics, the data or collection of data is described in summary. But in the case of inferential stats, it is used to explain the descriptive one. Both these types have been used on large scale.

1.6.2 Descriptive Statistics

The data is summarised and explained in descriptive statistics. The summarization is done from a population sample utilising several factors such as mean

and standard deviation. Descriptive statistics is a way of organising, representing, and explaining a set of data using charts, graphs, and summary measures. Histograms, pie charts, bars, and scatter plots are common ways to summarise data and present it in tables or graphs. Descriptive statistics are just that: descriptive. They don't need to be normalised beyond the data they collect.

Measures of Central Tendency

The central tendency is stated as the statistical measure that represents the single value of the entire distribution or a dataset. It aims to provide an accurate description of the entire data in the distribution. The central tendency of the dataset can be found out using the three important measures namely mean, median and mode.

Measures of Variability

Variability refers to how spread scores are in a distribution out; that is, it refers to the amount of spread of the scores around the mean. For example, distributions with the same mean can have different amounts of variability or dispersion. There are four frequently used measures of the variability of a distribution:

- Range
- Interquartile range
- Variance
- Standard deviation

1.6.3 Inferential Statistics

We attempt to interpret the meaning of descriptive statistics using inferential statistics. We utilise inferential statistics to convey the meaning of the collected data after it has been collected, evaluated, and summarised. The probability principle is used in inferential statistics to determine if patterns found in a study sample may be extrapolated to the wider population from which the sample was drawn. Inferential statistics are used to test hypotheses and study correlations between variables, and they can also be used to predict population sizes. Inferential

statistics are used to derive conclusions and inferences from samples, i.e. to create accurate generalisation. Inferential statistics have two main uses:

- making estimates about populations (for example, the mean SAT score of all 11th graders in the US).
- testing hypotheses to draw conclusions about populations (for example, the relationship between SAT scores and family income).

Inferential statistics encompasses two primary categories

- Hypothesis testing
- Regression analysis.

Hypothesis Testing

Hypothesis testing involves using a sample of data to test a hypothesis about a population parameter. A hypothesis is a statement about a population parameter, such as the population mean or proportion that the researcher wants to test. Hypothesis testing involves comparing the observed sample statistic to what would be expected if the null hypothesis (the statement being tested) were true. This comparison is used to determine whether there is enough evidence to reject the null hypothesis in favor of an alternative hypothesis.

Regression Analysis

Regression analysis is used to model the relationship between one or more independent variables and a dependent variable. The goal of regression analysis is to create a mathematical equation that can be used to predict the value of the dependent variable based on the values of the independent variables. Regression analysis is commonly used in fields such as economics, psychology, and social sciences to understand how changes in one variable can affect another variable.

1.7 STAGES OF STATISTICS

Johnson and Jackson said that "Statistics is the process of the methods which are related to collection, classification, tabulation, analysis, interpretation and presentation of data."

- Collection of data
- Organisation of data
- Presentation of data
- Analysis and interpretation of data
- Forecasting of data

1.7.1 Collection of Data

It is the first step in a statistical enquiry. There are two ways of collecting data: primary and secondary. The data provide the basis of statistical enquiry. If data is biased, the entire study will become unreliable. Therefore, it has to be collected with great caution.

1.7.2 Organization of Data

The next step after collection of data is to organize this data in a proper way. Organizing involves classification, grouping and editing.

1.7.3 Presentation of Data

Once the collected data are organized these are presented in a systematic manner to make statistical analysis clear and easier. Data can be presented through tables, diagrams, graphs or pictures.

1.7.4 Analysis of Data

After presentation data are analyzed by using averages, dispersion, correlation, regression, time series analysis etc. It helps us to reach at some conclusions on the basis of data.

1.7.5 Interpretation of Data

It is the final stage of statistical study. In this stage data are used to derive some conclusions to take decisions.

1.7.6 Forecasting of Data

Forecasting is a technique that uses historical data as inputs to make informed estimates that are predictive in determining the direction of future trends.

1.8 LIMITATIONS OF STATISTICS

Statistics, with its wide applications in almost every sphere of human activity; is not without limitations. The following are some of its important limitations:

1.8.1 Statistics is not suited to the study of qualitative phenomenon

Statistics, being a science dealing with a set of numerical data, is applicable to the study of only those subjects of enquiry which are capable of quantitative measurement. As such; qualitative phenomena like honesty, poverty, culture, etc., which cannot be expressed numerically, are not capable of direct statistical analysis.

1.8.2 Statistics does not study individuals

Statistics deals with an aggregate of objects and does not give any specific recognition to the individual items of a series. Individual items, taken separately, do not constitute statistical data and are meaningless for any statistical enquiry

1.8.3 laws are not exact

Unlike the laws of physical and natural sciences, statistica1 laws are only approximations and not exact. On the basis of statistical analysis we can talk only in terms of probability and chance and not in terms of certainty. Statistical conclusions are not universally true - they are true only on an average.

1.9 ECONOMIC STATISTICS

Economic statistics encompass the numerical data used to understand and analyze the performance of an economy. They include various indicators such as GDP, inflation, unemployment, and more. These statistics are essential for policy-making, economic forecasting, and understanding economic trends.Economics can be broadly classified into two main areas:

- microeconomics
- macroeconomics

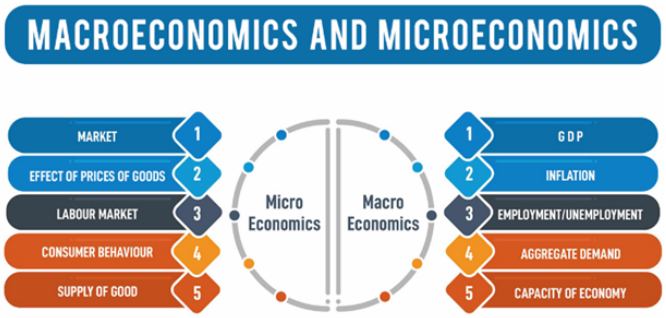

Figure 1.2: Types of Economics

1.9.1 Microeconomics

Microeconomics studies how individual consumers and firms make decisions to allocate resources. Whether a single person, a household, or a business, economists may analyze how these entities respond to changes in price and why they demand what they do at particular price levels. Microeconomics analyzes how and why goods are valued differently, how individuals make financial decisions, and how they trade, coordinate, and cooperate. Within the dynamics of supply and demand, the costs of producing goods and services, and how labor is divided and allocated, microeconomics studies how businesses are organized and how individuals approach uncertainty and risk in their decision-making.

Market

A market is a place where buyers and sellers come together to exchange goods and services. It can be physical, like a shop, or virtual, like an online platform. In microeconomics, markets determine the prices and quantities of goods based on demand and supply. Competitive markets lead to efficient resource allocation. Different types of markets include perfect competition, monopoly, oligopoly, and monopolistic competition

Effects of Prices Of Goods

When the price of a good changes, it affects both demand and supply. A price rise usually decreases demand but increases supply, while a price drop increases demand and reduces supply. This behavior is explained by the "law of demand" and "law of supply." Prices also influence related goods—substitutes and complements. For example, if coffee prices rise, people might buy more tea.

Labour Market

The labour market is where workers offer their skills and employers demand labor. Wages are determined by the interaction between labour supply and labour demand. Factors like education, skills, and experience affect a worker's wage. Unemployment, minimum wages, and labour unions also influence the market. The labour market can vary widely across different industries and regions

Consumer Behaviour

Consumer behaviour studies how individuals make decisions to spend their money on goods and services. It is influenced by preferences, income, prices, and social factors. Consumers aim to maximize their satisfaction (utility) given their budget. Psychological, cultural, and economic factors also play a role. The theory of consumer choice and indifference curves are important tools to analyze behaviour.

Supply of Good

Supply refers to the quantity of a good that producers are willing and able to sell at different prices. According to the law of supply, as the price increases, the

quantity supplied generally increases. Factors like production cost, technology, number of sellers, and government policies affect supply. A supply curve usually slopes upward from left to right. Shifts in supply occur when non-price factors change

1.9.2 Macroeconomics

Macroeconomics is a branch of economics that deals with how an economy functions on a large scale. It differs from microeconomics, which deals with how individual economic players, such as consumers and firms, make decisions. Its primary focus is recurrent economic cycles and broad economic growth and development. It focuses on foreign trade, government fiscal and monetary policy, unemployment rates, the level of inflation, interest rates, the growth of total production output, and business cycles that result in expansions, booms, recessions, and depressions. Using aggregate indicators, economists use macroeconomic models to help formulate economic policies and strategies.

Employment and Unemployment

When a person is actively engaged in a job, business, trade, or profession, they are considered to be employed. Unemployed people are those who are actively looking for work but are currently unemployed. Every government's macroeconomic goal is to keep the unemployment rate low and solve the problem of unemployment. Full employment occurs when every worker is employed. The unemployment rate is 0 percentage which means that everyone who is able and willing to work is working.

Inflation

In economics, inflation is an increase in the money supply. The effect of inflation is a general increase in the prices of goods and services in an economy. This is usually measured using a consumer price index. When the general price level rises, each unit of currency buys fewer goods and services; consequently, inflation corresponds to a reduction in the purchasing power of money. The opposite of CPI inflation is deflation, a decrease in the general price level of goods and services. The common measure of inflation is the inflation rate, the annualized percentage

change in a general price index. As prices faced by households do not all increase at the same rate, the consumer price index (CPI) is often used for this purpose.

Gross Domestic Product

Gross domestic product is a monetary measure of the total market value of all the final goods and services produced and rendered in a specific time period by a country or countries. GDP is often used to measure the economic performance of a country or region.

Aggregate demand

In macroeconomics, aggregate demand (AD) represents the total demand for all final goods and services in an economy at a given price level, encompassing consumer spending, investment, government spending, and net exports.

Capacity of Economy

In macroeconomics, the "capacity of the economy" refers to its maximum potential output of goods and services, also known as its potential GDP, which is determined by factors like available resources and technology

1.10 OBJECTIVE OF THE STUDY

This study examines the importance of unemployment as a key macroeconomic factor analyzed through econometric methods. Unemployment is one of the major indicators used to assess a country's economic growth. Economics plays a vital role in the development of every nation.

1. To describe our data using descriptive analysis and graphical representation of the data.

2. To find the correlate the relationship between the one variables to other variables

3. To find the average of relationship between the value of our data and accuracy predict of the data

4. To Visualize the top 20 countries with the highest unemployment rates, along with influencing factors such as population, GDP, life expectancy, and urban population

5. To explore the factors affecting a country's durability across different socio-economic aspects—such as environmental changes, CPI, health expenditure, GDP, gross primary education, gross tertiary education, and labor force participation—using dimension reduction techniques

6. To design and analyze data, recognize patterns, and predict unemployment rates using machine learning models such as linear regression, random forest, and gradient boosting, and to compare the models to determine which one provides the best fit.

7. To identify and analyze the significance of unemployment and its impact on society and the economy

CHAPTER 2

Chapter 2

UNEMPLOYMENT

Figure 2.1: Unemployment

Unemployment is a critical economic issue that affects both individuals and society at large. It occurs when people who are capable of working, are actively seeking work, but are unable to find any suitable employment opportunities. High levels of unemployment are often indicative of a struggling economy, while low unemployment can suggest a healthy economy with ample job opportunities

Unemployment can have far-reaching consequences, not only on the economic stability of a country but also on the social and psychological well-being of individuals. It is often used as an indicator of the health of an economy, and governments usually monitor and implement policies to control and reduce unemployment rates.

Unemployment brings, in its wake, a large number of socioeconomic problems in many different ways. The negative impact of unemployment is diverse and acute. In addition to lowering output and income, it can also lead to rising income

inequality, loss of human capital and skill, ill health and mortality, migration, loss of motivation for future works, loss of human relations, and social life. Therefore, it is very important to understand the nature and effects of unemployment, and to design appropriate policies and programmers to combat it

Unemployment is defined as the condition of being unemployed, or, it refers to the number or proportion of people in the working population who are unemployed (have no jobs). An unemployed person is one who is an active member of the labor force and is able to and seeks work, but is unable to find work during a specified reference period (a week or a month or a year). Let us divide the total population of a country into four categories as illustrated below.

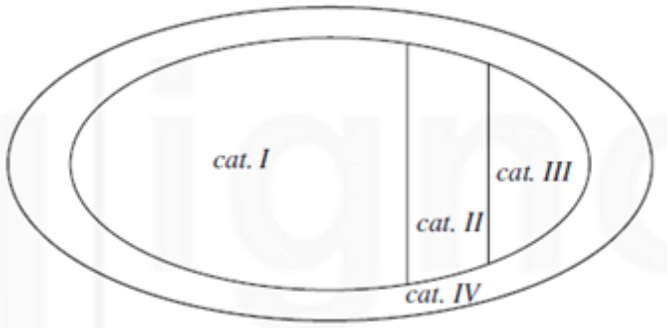

Figure 2.2: categories

Category I is comprised of all those economically active people, say between 15 and 65 years, who are currently employed during a specified reference period (a week, or a month, or a year). Category II represents all those economically active people, say between 15 and 65 years, who are currently unemployed during a specified reference period (a week, or a month, or a year), but are capable of, and are looking for employment Category III represents all those people say, between 15 and 65 years, who are neither employed nor are seeking any employment, in other words, those who are not in the labor force. Category IV represents all those people aged below 15 years and above 65 years, and other people excluded from the above mentioned three categories for any reason. The total population of a country is the sum of all these four categories: I + II +III + IV; and labor force is defined as I + II. The unemployment rate is defined as the percentage of the labor force that is Unemployed , which can be represented as:

Unemployment = (II/ I+II) X 100

2.1 DEFINITION

The state of being without any work yet looking for work is called unemployment. Economists distinguish between various overlapping types of and theories of unemployment, including cyclical or Keynesian unemployment, frictional unemployment, structural unemployment and classical unemployment definition. Some additional types of unemployment that are occasionally mentioned are seasonal unemployment, hardcore unemployment, and hidden unemployment.

Though there have been several definitions of "voluntary" and "involuntary unemployment" in the economics literature, a simple distinction is often applied. Voluntary unemployment is attributed to the individual's decisions, but involuntary unemployment exists because of the socio-economic environment (including the market structure, government intervention, and the level of aggregate demand) in which individuals operate. In these terms, much or most of frictional unemployment is voluntary since it reflects individual search behavior. Voluntary unemployment includes workers who reject low-wage jobs, but involuntary unemployment includes workers fired because of an economic crisis, industrial decline, company bankruptcy, or organizational restructuring.

2.2 HISTORY OF UNEMPLOYMENT

There are relatively limited historical records on unemployment because it has not always been acknowledged or measured systematically. Industrialization involves economies of scale,

which often prevent individuals from having the capital to create their own jobs to be self-employed. An individual who cannot join an enterprise or create a job is unemployed. As individual farmers, ranchers, spinners, doctors and merchants are organized into large enterprises, those who cannot join or compete become unemployed

The Depression of 1873–79: New York City police violently attacking unemployed workers in Tompkins Square Park, Manhattan, New York City, 1874

2.2.1 16th century

The closing of the monasteries in the 1530s increased poverty, as the Roman Catholic Church had helped the poor. In addition, there was a significant rise in enclosures during the Tudor period. Also, The population was rising. Those unable to find work had a stark choice: starve or break the law.In 1535, a bill was drawn up calling for the creation of a system of public works to deal with the problem of unemployment, which were to be funded by a tax on income and capital. A law that was passed a year later allowed vagabonds to be whipped and hanged

2.2.2 Industrial Revolution to late 19th century

Poverty was a highly visible problem in the nineteenth century, both in cities and in the countryside. In France and Britain by the end of the century, an estimated 10 percent of the people depended on charity or begging for their food. A description of the miserable living standards of the mill workers in England in 1844 was given by Fredrick Engels in The Condition of the Working Class in England in 1844. In the preface to the 1892 edition, Engels noted that the extreme poverty he had written about in 1844 had largely disappeared. David Ames Wells also noted that living conditions in England had improved near the end of the 19th century and that unemployment was low.

2.2.3 20th century

There were labor shortages during World War I Ford Motor Co. doubled wages to reduce turnover. After 1925, unemployment gradually began to rise. The 1930s saw the Great Depression impact unemployment across the globe. In Germany and the United States, the unemployment rate reached about 25% in 1932. In some towns and cities in the northeast of England, unemployment reached as high as 70%; the national unemployment level peaked at more than 22% in 1932. Unemployment in Canada reached 27% at the depth of the Depression in 1933. In1929, the U.S. unemployment rate

2.2.4 21st century

The official unemployment rate in the 16 European Union (EU) countries that use the euro rose to 10% in December 2009 as a result of another recession. Latvia had the highest unemployment rate in the EU, at 22.3% for November 2009. Europe's young workers have been especially hard hit. In November 2009, the unemployment rate in the EU27 for those aged 15–24 was 18.3%. For those under 25, the unemployment rate in Spain was 43.8%. Unemployment has risen in two thirds of European countries since 2010.

2.3 UNEMPLOYMENT RATE IN INDIA

Figure 2.3: Unemployment in India

According to the India Employment Report 2024, created jointly by the Institute for Human Development and the International Labour Organization (ILO), India's working population increased from 61 percent in 2011 to 64 percent in 2021, and it is projected to reach 65 percent in 2036. However, the percent of youth involved in economic activities declined to 37 percent in 2022. Continued vigilance and effective policy measures remain crucial to foster sustainable job growth and secure the nation's future prosperity. According to the latest data from the Centre for Monitoring Indian Economy (CMIE), an independent think tank, the unemployment rate in India stood at 7.8 percent in September 2024, a decline from 8.5 percent in August 2024. The labour participation rate fell from 41.6 percent to 41 percent, and the employment rate fell from 38 percent in August to 37.8 percent in September.

2.4 UNEMPLOYMENT RATE IN GLOBAL

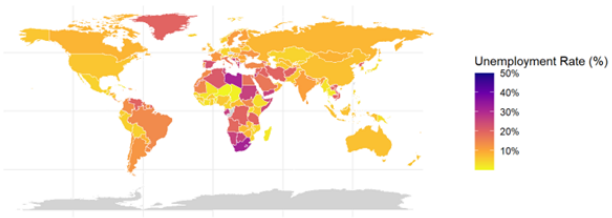

Figure 2.4: Unemployment in world

This is a list of countries by unemployment rate. Methods of calculation and presentation of unemployment rate vary from country to country. Some countries count insured unemployed only, some count those in receipt of welfare benefit only, some count the disabled and other permanently unemployable people, some countries count those who choose (and are financially able) not to work, supported by their spouses and caring for a family, some count students at college and so on. There may also be differences in the minimum requirements and some consider people employed even if only marginally associated with employment market (for example, working only one hour per week). There can be differences in the age limit

2.5 TYPES OF UNEMPLOYMENT

The various types of unemployment. Broadly, unemployment can be divided into two types: Voluntary, and involuntary, unemployment. Voluntary unemployment arises due to reasons that are specific to an individual, while involuntary unemployment is caused by a large number of socio-economic factors such as structure of the market, level and composition of aggregate demand, government intervention, and so on. Thus, there are different kinds of unemployment depending on the nature, causes, and duration of unemployment. Let us now discuss various types of unemployment. Unemployment is broadly classified in following categories

- Structural unemployment
- Cyclical unemployment
- Frictional unemployment

- Seasonal unemployment
- Natural rate of unemployment
- Open unemployment

2.5.1 Structural unemployment

This kind of unemployment occurs when there is any change in consumer demand and technology in the economy. For instance, when computers were introduced, many workers were dislodged because of a mismatch between the existing skills of the workers and the requirement of the job. Although jobs were available, there was a demand for a new kind of skill and qualification. So, persons with old skills did not get employment in the changed economic regime, and remain unemployed. This is called structural unemployment.

2.5.2 Cyclical unemployment

When there is an economy-wide decline in aggregate demand for goods and services, employment declines and unemployment correspondingly increases. Therefore, it is sometime referred to as 'demand deficient unemployment'. For instance, during the recent global slowdown, in late 2008, many workers around the globe lost their jobs

2.5.3 Frictional unemployment

This type of unemployment refers to a transition period of looking for a new job, for different reasons, such as seeking a better job, being fired from a current job, or having voluntarily quit a current job. The period of time between the current to a new job is referred to as frictional, or temporary unemployment

2.5.4 Seasonal unemployment

Seasonal unemployment a type of frictional unemployment, occurs in specific activities or occupations which are characterized by seasonal work. An example of seasonal unemployment is the joblessness during non-cultivation in rural areas.

2.5.5 Natural rate of unemployment

The sum total of frictional and structural unemployment is referred as the natural rate of unemployment

2.5.6 Open unemployment

Open unemployment arises when a person, voluntarily or involuntarily, keeps himself or herself out of consideration for certain jobs. It is important to note that the type and nature of unemployment differs significantly in developing and developed countries. Unemployment in developed countries arises due to the lack of effective demand and/or economic slowdown, recession, or depression. In developing countries, unemployment occurs largely due to a lower demand for labor and/or inadequate employment opportunities in the economy. Such a situation occurs due to the subsistence nature of agriculture, a low industrial base and the small size of the tertiary sector.

The various concepts of unemployment applicable to developing nations, including India. All developing countries, including India suffer from structural unemployment, which exists both in open and disguised forms. The problem in developing countries can better be summarized as under employment - a partial lack of work, low employment income, and under utilization of skills or low productivity rather than unemployment as discussed above. Thus, under employment describes the condition of those who work part time because full time jobs are unavailable or employed on a full time basis but the services they render may actually be much less than full time (disguised underemployment) or who are employed in occupations requiring lower levels of skills than they are qualified for (hidden underemployment). A related concept is that of working poor -those who actually work long hours but earn only a low income below the poverty line. In other words, working poor is defined as a situation when individuals or households, in spite of being employed, remain in relative poverty due to low levels of wages and earnings.

2.6 MEASUREMENT OF UNEMPLOYMENT

Measurement of unemployment is a difficult task. In India's statistical system, the most comprehensive and reliable data on employment and unemployment

are compiled by the National Sample Survey Organization (NSSO). Based on different reference period (a year, a week, and each day of a week), NSSO provides four different measures of employment and unemployment. The following are some methods of measuring unemployment.

2.6.1 Usual Principal Status Unemployment (UPS)

This is measured as the number of persons who remained unemployed for a major part of the year The persons covered by the survey may be classified into those working and/or available for work in their principal activity, and those working and/or available for work in a subsidiary activity, that is, a sector other than their principal activity. Hence, within the usual status concept, the estimates are now derived on the usual principal status as well as the usual principal and subsidiary status basis.

The usual status unemployment rate is a person rate and indicates chronic unemployment, because all those who are found usually unemployed in the reference year are counted as unemployed. This measure is more appropriate to those in search of regular employment, e.g., educated and skilled persons who may not accept casual work. This is also referred to as 'open unemployment'.

2.6.2 Usual Principal and Subsidiary Status Unemployment (UPSS)

Here person is considered unemployed, if besides UPS, those available but unable to find work on a subsidiary basis during a year

2.6.3 Current Weekly Status Unemployment (CWS)

This refers to the number of persons who did not find even an hour of work during the survey week.

2.6.4 Current Daily Status Unemployment (CDS)

This refers to the number of persons who did not find work on a day, or on some days, during the survey week.

Rates of unemployment differ based on different concepts and these concepts have their own advantages and limitations and, therefore, they have different Implications for analysis and policy formulation. The UPS and UPSS measure reflect only long term unemployment spells. The CWS measure captures shorter unemployment spells, but ignores unemployment for less than a week. The CDS measure is the most inclusive, capturing both open as well as partial unemployment.

2.7 CAUSES OF UNEMPLOYMENT

Some of the causes of unemployment, you are commonly aware, are over population and lack of work. Some of the important causes of unemployment are: the high rate of population growth and the consequent increase in labour force; low rate of economic growth, lack of adequate employment opportunities in non-agricultural activities; generation of seasonal employment, or lack of full time employment in agriculture; low labour absorption capacities in manufacturing and tertiary activities; shifting from labour intensive to capital- and skill-intensive production techniques; expansion in education system, and so on. Let us now briefly elaborate some of the above causes.

- Low and fluctuating levels of economic growth do not generate adequate
- Employment opportunities, as needed. Besides, rates the sectoral composition of growth is also an important determinant of unemployment
- Excessive dependence on agriculture and slow growth of non-farm activities limit employment generation
- Use of capital-intensive methods and production practices not only dislodges currently employed persons, but also slows down the generation of new employment opportunities.
- policy, may lead to a mismatch between the need and availability of relevant
- Lack of a clear and well developed human resource policy or manpower
- skills and training, which results in unemployment, especially of youth and educated.

- Rapid growth in population, lack of employability due to poor health and nutrition also lead to unemployment.
- Lack of investment and infrastructure development are factors that do not Generate adequate levels of employment in the economy, and consequently unemployment rises.
- Inadequate availability of unemployment insurance and public employment programmes also cause a rise in unemployment.

How these factors are operating to bring changes in the nature and magnitude of unemployment will be discussed in the following sections. It is important to mention here that one or more than one of the above factors may operate simultaneously to impact unemployment rates

2.7.1 Unemployment and Inflation Relationship

In economic literature there are many different accounts of labour market theory . All labour market outcomes have been explained by theoretical arguments with the help of three main forces: the market forces of demand and supply; institutional forces in the form of government, union, and; sociological factors such as cultures, class, customs, and family background. With this framework, different schools of thought - the classical, the neoclassical, and the institutional schools – emerged in labour economics literature. They were based on relative importance and the working of these forces to bring about specific labour market outcomes.

The neoclassical school focused primarily on the operation of market forces in determining the wages and allocation of labour, and considered other institutional and social factors as given. On the contrary, the institutional school emphasized the role of institutional forces, such as internal labour markets and unions, and sociological factors, such as class and discrimination, segmentation, and stratification in the labour market

2.8 POLICIES AND PROGRAMMES TO REDUCE UNEMPLOYMENT IN INDIA

Various government schemes have been undertaken in India to tackle the problem of unemployment. But these programmes have multiple objectives comprising the reduction of poverty, generation of employment, and the provisioning of crucial basic services. The employment programmes can broadly be categorized into two types: self employment, and wage employment programmes. Some of the important employment generation programmes are described next.

2.8.1 Integrated Rural Development Programme (October 1980)

IRDP is acentrally sponsored scheme providing self employment to the rural poor through the acquisition of productive assets or appropriate skills.Assistance, under this programme, is provided in the form of subsidy and bank credit. The pattern of subsidy was 25 per cent for small farmers, 33 per cent for marginal and agricultural farmers and rural artisans, 50 per cent for scheduled castes, scheduled tribes,and physically handicapped persons.

2.8.2 Training of Rural Youth for Self-Employment (August 1979)

Training of Rural Youth For Self-Employment is a centrally sponsored programme, supporting component of the IRDP. It aims at providing technical and entrepreneurial skills to rural unemployed youths in the age group,18-35 years from families below the poverty line, to enable them to take up income generating schemes. It has been merged with SGSY since April 1999.

2.8.3 Employment Assurance Scheme (October 1993)

initially it has been launched in selected blocks of drought prone, desert and hill areas, and later, extended to all rural blocks with effect from April 1997. The main objective of EAS is to provide about 100 days of assured casual manual employment during the lean agricultural season, at statutory minimum wages, to

all persons above the age of 18 years and below 60 years who need and seek employment on economically productive and labour intensive social and community works

Government Policies Updated

- The Indian government launched three Employment Linked Incentive (ELI) schemes to boost formal employment.
- Under ELI Scheme A, first-time employees receive a wage subsidy after financial literacy training.
- ELI Scheme B promotes job creation in the manufacturing sector with hiring incentives.
- ELI Scheme C rewards employers expanding their workforce with cash benefits.
- Tamil Nadu's Unemployment Assistance Scheme provides monthly payments to educated, unemployed youth.
- Unemployed Youth Employment Generation Programme (UYEGP) supports self-employment with loans and subsidies.
- Financial Assistance to Unemployed Graduates gives direct financial aid to jobless graduates and differently-abled.
- Overall, these efforts target 41 million new jobs nationally over five years with a 2 lakh crore budget.

2.9 SOME FACTORS THAT INFLUENCED UNEMPLOYMENT

Several factors influence unemployment, often interacting in complex ways. Below are key factors that contribute to or exacerbate unemployment

2.9.1 Economic Factors

- Economic Recession: During recessions, businesses face reduced demand for their products and services, leading to job cuts to manage costs. For example, the 2008 global financial crisis caused a spike in unemployment across many industries, particularly in real estate and banking.

- Inflation and Deflation: High inflation reduces purchasing power,squeezing business profits and triggering layoffs. On the other hand, deflation can discourage business investments as prices drop, leading to reduced hiring or job losses.

- Globalization: Companies often move jobs to countries with lower labor costs. While this benefits the global economy by improving efficiency, it creates unemployment in high-wage countries. For instance, manufacturing jobs have declined in developed nations due to outsourcing to developing countries.

2.9.2 Technological Advancements

- Automation and Artificial Intelligence: Many routine jobs, such as assembly line work, have been automated. While this boosts productivity, it displaces workers, especially in manufacturing and logistics. For instance, automated checkouts in retail stores reduce the need for cashiers.

- Digital Transformation: Rapid technological change creates new industries but requires workers with new skills. A lack of retraining opportunities leaves many unemployed, particularly those in traditional sectors.

2.9.3 Demographic Factors

- Population Growth: Rapid population growth, especially in countries with limited job creation, leads to a surplus of workers. This is often observed in developing nations where youth unemployment is high.

- Aging Population: In countries like Japan and parts of Europe, an aging workforce leads to shortages in certain sectors, while also creating opportu-

nities in healthcare. However, older workers may face challenges in adapting to new technologies, leading to underemployment.

2.9.4 Education and Skills

- Skill Mismatches: The job market may demand specialized skills, like coding or data analytics, while the workforce possesses outdated or unrelated qualifications. This is a major issue in rapidly changing economies.
- Lack of Access: Inequities in access to quality education and vocational training, especially in rural or marginalized areas, limit employment opportunities. For example, many countries face a disparity in urban versus rural job readiness

2.9.5 Structural Issues in the Economy

- Sectoral Decline: Certain industries, like coal mining or traditional manufacturing, decline due to environmental concerns or changing energy policies. Workers in these industries often lack the skills needed for emerging sectors.
- Geographic Mismatch: Jobs may exist in urban centers, while many unemployed individuals reside in rural or underdeveloped areas. The cost of relocation or lack of infrastructure exacerbates this issue.

- Rigid Labor Markets: Excessive regulations, such as high minimum wages or stringent hiring/firing rules, can deter companies from expanding their workforce. For example, in some European countries, strict labor laws contribute to high youth unemployment.

2.9.6 Government Policies

- Regulatory Environment: Excessive bureaucracy can delay business permits and stifle entrepreneurship. For example, lengthy procedures to start a business discourage job creation.

- Taxation: High taxes on corporations can reduce profits, limiting their capacity to hire. Conversely, tax incentives can stimulate employment by encouraging businesses to expand.
- Social Security Policies: While unemployment benefits provide essential support, overly generous schemes might reduce motivation to seek jobs if not designed properly

2.9.7 Social and Political Factors

- Conflict and Instability: Wars and political turmoil disrupt economies, forcing businesses to shut down. For example, the Syrian conflict led to massive unemployment and displacement.
- Discrimination: Prejudice based on gender, ethnicity, age, or disability can marginalize certain groups, leading to lower participation in the workforce.
- Union Activity: While unions protect workers rights, excessive demands can lead to business closures or reluctance to hire. For instance, strikes and high labor costs can reduce competitiveness.

2.9.8 Market Changes

- Consumer Behavior: Shifts in consumer preferences, such as the rise of e-commerce, reduce demand for traditional retail jobs. For instance, department stores have seen significant layoffs as online platforms dominate.
- Global Market Dynamics: Fluctuations in commodity prices, like oil or agricultural goods, impact sectors reliant on these industries, leading to cyclical unemployment.

2.9.9 Pandemics or Natural Disasters

- Health Crises: Pandemics like COVID-19 disrupt global economies, with lockdowns leading to job losses in hospitality, travel, and retail. Remote work trends have also reduced demand for jobs in urban infrastructure-related sectors.

- Environmental Factors: Natural disasters destroy infrastructure and livelihoods, particularly in agriculture and tourism. For example, hurricanes can devastate local economies in affected regions. Unemployment remains one of the most significant challenges for economies and societies worldwide. It results from a complex interplay of economic, technological, demographic, educational, and policy-related factors. Its impact extends beyond financial distress, affecting mental health, social stability, and overall economic growth.

CHAPTER 3

Chapter 3

METHODOLOGY

In this chapter Describe the statistical models and the methodology for carrying out the Descriptive statistics, Bar charts are presented. The Correlation, Regression, Principal component analysis.The concept of model Comparison

3.1 FUNCTIONS OF STATISTICS

Prof R.A.Fisher defined statistic as, "The science of statistics is essentially a branch of applied mathematics and may be regarded as a mathematics applied to observational data".

- Primary data: Primary data are those which are collected from the units or individuals directly and those data have never been used for any purpose earlier.
- Secondary data: The data, which had been collected by some individuals or agency and statistically treated to draw certain conclusions. Again the same data are used to draw certain conclusions and analyzed to extract some other information, are termed as secondary data.

3.2 PRESENTATION OF DATA

After the data has been systematically collected and edited, the next step is classification of data.Classification is a process of arranging data into different classes according to their resemblances and affinities. The objectives of classification of data are Simplifying the raw data

- Facilitating comparison
- Duplicating the silent features of data
- Making the data more intelligent
- Eliminating unnecessary details
- Facilitating statistical interpretation

3.3 CLASSIFICATION

Classification is the process of arranging things or item in group or class. According to their resemblance and affinities and give expression to the units of attributes that may subsist amongst the diversity of individuals .The classification mainly deals with

- Geographical classification, i.e., in relation to place
- Chronological classification, i.e., on the basis of time
- Qualitative classification, i.e., according to some attributes
- Quantitative classification, i.e., based on figures or characteristics.

3.4 TABULATION

It is the process of presenting data collected through survey, experiment or record in rows and columns. So that it can more easily understood and it can be used for further statistical analysis. There are five parts of tables,

- Title -this is a brief description of the contents and is shown at the top of the table.
- Stubs- The extreme left part of the table, were descriptions of rows are shown, is called stubs.
- Caption and Box-head-The upper part of table which the description of columns is called caption. Units of measurement and column-numbers, if any, are called Box-head.
- Body-It is a part of the table which shows the figure.
- Source note-This is the part below the Body, where the source of data and any explanations are shown.

3.5 VISUALIZATION

A graphical is a pictorial presentation of the relationship between variables. Many types of graphs are employed in statistic depending on the data on involved and the purpose of graphs is intended. Sometimes referred to as charts or diagrams.

3.5.1 Bar Chart

A bar chart is a graphical representation used to display and compare discrete categories of data through rectangular bars, where the length or height of each bar is proportional to the frequency or value of the corresponding category.

- Vertical Bar Chart: Bars are displayed vertically, with categories on the x-axis and values on the y-axis.
- Horizontal Bar Chart: Bars are displayed horizontally, with categories on the y-axis and values on the x-axis.

3.6 ANALYSIS OF DATA

Data analysis has multiple facts and approaches, encompassing diverse techniques under a variety of names, in different business, science, and social science

domains. Informative presentation of tables graph and statistics is very important in data analysis. It is process of inspecting, cleaning, transforming, and modelling data with the goal of discovering useful information, suggesting conclusions, and supporting decision-making. This study is carried out on statistical software like minitab,spss and python,R. The following statistical tool were used to carry out this study.

- Descriptive Statistics
- Graphical Representation
- Correlation
- Regression
- Principal Component Analysis
- Model Comparison

3.7 DESCRIPTIVE STATISTICS

Descriptive statistics are a set of techniques used to summarize and describe the main features of a dataset. These statistics provide a concise and meaningful overview of the essential characteristics of a dataset, helping researchers and analysts to understand the data and draw insights. Descriptive statistics can be broadly categorized into measures of central tendency and measures of dispersion.

3.7.1 Measures of Central Tendency

- Mean (Average): The sum of all values in a dataset divided by the number of observations.
- Median: The middle value in a dataset when it is ordered from least to greatest. It is less affected by extreme values than the mean.
- Mode: The most frequently occurring value in a dataset.

3.7.2 Measures of Dispersion (Variability)

- Range: The difference between the maximum and minimum values in a dataset.
- Variance: A measure of the average squared deviation of each data point from the mean.
- Standard Deviation: The square root of the variance. It indicates the average amount of deviation of individual data points from the mean.

3.7.3 Measures of Shape and Distribution

- Skewness: A measure of the asymmetry of the distribution. Positive skewness indicates a tail on the right, while negative skewness indicates a tail on the left.
- Kurtosis: A measure of the "tailedness" of the distribution. It describes the sharpness of the peak or the flatness of the tails compared to a normal distribution.

3.8 CORRELATION ANALYSIS

Correlation analysis is a statistical method that identifies the strength of a relationship between two or more variables, the process reveals patterns within a dataset's many variables. It's all about identifying relationships between variables–specifically in research.Correlation between two variables can be either a positive correlation, a negative correlation, or no correlation. Let's look at examples of each of these three types.

- Positive correlation: A positive correlation between two variables means both the variables move in the same direction. An increase in one variable leads to an increase in the other variable and vice versa.
- Negative correlation: A negative correlation between two variables means that the variables move in opposite directions. An increase in one variable leads to a decrease in the other variable and vice versa.

- Weak/Zero correlation: No correlation exists when one variable does not affect the other.

If x & y are the two variables of discussion, then the correlation coefficient can be calculated using the formula

$$r = \frac{(N\sum XY - \sum X.\sum Y)}{(\sqrt{(N\sum X^2 - (\sum X)^2)}\sqrt{(N\sum Y^2 - (\sum Y)^2)})}$$

where,

- rxy = strength of the correlation between variables x and y
- n= sample size
- $\sum$ = sum of what follows
- X = every x-variable value
- Y = every y-variable value
- XY = the product of each x-variable score and the corresponding y-variable score

Correlation in the Unemployment refers to the statistical relationship between the returns of different unemployment or between a Unemployment returns and another variable, such as an macro economic indicator. It helps investors understand how the predict of assets move relative to each other.

3.9 REGRESSION ANALYSIS

In statistics, simple linear regression (SLR) is a linear regression model with a single explanatory variable That is, it concerns two-dimensional sample points with one independent variable and one dependent variable (conventionally, the x and y coordinates in a Cartesian coordinate system) and finds a linear function (a non-vertical straight line) that, as accurately as possible, predicts the dependent variable values as a function of the independent variable.The adjective simple refers to the fact that the outcome variable is related to a single predictor. In this case, the slope of the fitted line is equal to the correlation between y and x corrected by the ratio of standard deviations of these variables. The intercept of the fitted line is such that the line passes through the center of mass (x, y) of the data points.

Formulation and computation

Consider the model function

$$Y = \alpha + \beta X \tag{3.1}$$

which describe a line with slope β and Y-intercept α In general, such a relationship may not hold exactly for the largely unobserved population of values of the independent and dependent variables; we call the unobserved deviations from the above equation the errors. Suppose we observe n data pairs and call them (x_i, y_i), i = 1,...,n.

$$Y_i = \alpha + \beta xi + \varepsilon_i$$

This relationship between the true (but unobserved) underlying parameters α and β and the data points is called a linear regression model. The goal is to find estimated values $\hat{\beta}$ and $\hat{\alpha}$ for the parameters α and βwhich would provide the "best" fit in some sense for the data points. As mentioned in the introduction, in this article the "best" fit will be understood as in the least-squares approach: a line that minimizes the sum of squared residuals (see also Errors and residuals) (differences between actual and predicted values of the dependent variable y),each of which is given by, for any candidate parameter values α and β

$$\hat{\varepsilon}_i = y_i - \alpha - \beta x_i.$$

In other words $\hat{\alpha}$ and $\hat{\beta}$ solve the following minimization problem

$$Q(\alpha, \beta) = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

By expanding to get a quadratic expression in α and β we can derive minimizing values of the function arguments, denoted $\hat{\beta}$ and $\hat{\alpha}$

$$\widehat{\alpha} = \bar{y} - (\widehat{\beta}\bar{x}),$$
$$\widehat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} \Delta x_i \Delta y_i}{\sum_{i=1}^{n} \Delta x_i^2}$$

Here we have introduced

- $\bar{x}_i$ and $\bar{y}_i$ as the average of the x_i and y_i,respectively.
- Δx_i and Δy_i as the deviations in $\bar{x}_i$ and $\bar{y}_i$ with respect to their

3.10 MULTIPLE LINEAR REGRESSION ANALYSIS

Multiple linear regression (MLR), also known simply as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of MLR is to model the linear relationship between the explanatory (independent) variables and response (dependent) variables. In essence, multiple regression is the extension of ordinary least-squares (OLS) regression because it involves more than one explanatory variable.

Formula and Calculation of Multiple Linear Regression (MLR)

$$y_i = \beta 0 + \beta 1 \times i1 + \beta 2 \times i2 + \ldots + \beta_{\mathrm{p}} x_{\mathrm{ip}} + \epsilon$$

where, for i=n observations:
y_i=dependent variable
x_i=explanatory variables
β_0=y-intercept(constant term)
β_p=slope coefficients for each explanatory variable
ε=the model's error term (also known as the residuals)

3.11 PRINCIPAL COMPONENT ANALYSIS

Principal component analysis ,or PCA, is a dimensionality-reduction method that is often used to reduce the dimensionality of large data set, by transforming a large set of variables into a smaller one that still contains most of the information in the large set. PCA is extremely useful when working with data sets that have a lot of features . common applications such as image processing ,genome research always have to deal with thousant's if not tens of thousants of columns. PCA is defined as an orthogonal linear transformation that transform the data to a new coordinate system such that the greatest variance by some scalar projection of the data comes to lie on the coordinate (called the first principal component),the second greatest variance on the second coordinate ,and so on.

3.12 MODEL COMPARISON

3.12.1 Machine Learning

Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalize to unseen data, and thus perform tasks without explicit instructions. Within a sub discipline in machine learning, advances in the field of deep learning have allowed neural networks, a class of statistical algorithms, to surpass many previous machine learning approaches in performance

ML finds application in many fields, including natural language processing, computer vision, speech recognition, email filtering, agriculture, and medicine. The application of ML to business problems is known as predictive analytics. Statistics and mathematical optimization (mathematical programming) methods comprise the foundations of machine learning.

Random Forest Algorithm

Random forest is a supervised learning algorithm. The "forest" it builds is an ensemble of decision trees, usually trained with the bagging method. The general idea of the bagging method is that a combination of learning models increases the overall result Random forests or random decision forests is an ensemble learning method for classification, regression and other tasks that works by creating a multitude of decision trees during training. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the output is the average of the predictions of the trees. Random forests correct for decision trees' habit of over fitting to their training set

Random Forest Regression

Random Forest Regression works by creating multiple of decision trees each trained on a random subset of the data. The process begins with Bootstrap sampling where random rows of data are selected with replacement to form different training datasets for each tree. After this we do feature sampling where only a random subset of features is used to build each tree ensuring diversity in the models. After the trees are trained each tree make a prediction and the final prediction for

regression tasks is the average of all the individual tree predictions and this process is called as Aggregation. This approach is beneficial because individual decision

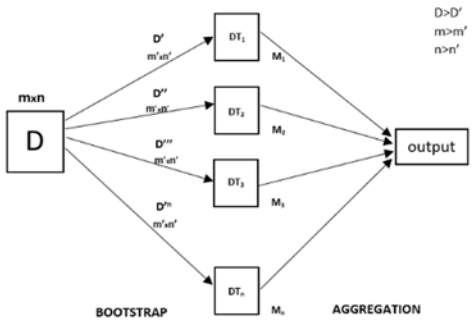

Figure 3.1: Random Forest Regression

trees may have high variance and are prone to over fitting especially with complex data. However by averaging the predictions from multiple decision trees Random Forest minimizes this variance leading to more accurate and stable predictions and hence improving generalization of model

Random Forest Classification

In a random forest classification, multiple decision trees are created using different random subsets of the data and features. Each decision tree is like an expert, providing its opinion on how to classify the data. Predictions are made by calculating the prediction for each decision tree and then taking the most popular result.

Gradient Boosting

Gradient Boosting is a popular boosting algorithm in machine learning used for classification and regression tasks. Boosting is one kind of ensemble Learning method which trains the model sequentially and each new model tries to correct the previous model. Gradient Boosting is a powerful boosting algorithm that combines several weak learners into strong learners, in which each new model is trained to minimize the loss function such as mean squared error or cross-entropy of the previous model using gradient descent. In each iteration, the algorithm computes the gradient of the loss function with respect to the predictions of the current ensemble and then trains a new weak model to minimize this gradient.

The predictions of the new model are then added to the ensemble, and the process is repeated until a stopping criterion is met.

Algorithm

Many supervised learning problems involve an output variable y and a vector of input variables x, related to each other with some probabilistic distribution. The goal is to find some function f(x) that best approximates the output variable from the values of input variables. This is formalized by introducing some loss function L(Y,f(x)) and minimizing it in expectation:

$$\hat{F} = \arg\min_{F} \mathbb{E}_{x,y}[L(y, F(x))]$$

The gradient boosting method assumes a real-valued y. It seeks an approximation F(x) in the form of a weighted sum of M functions hm(x) from some class Hcalled base (or weak) learners:

$$\hat{F}(x) = \sum_{m=1}^{M} \gamma_m h_m(x) + const$$

We are usually given a training set $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ of known values of x and corresponding values of y .In accordance with the empirical risk minimization principle, the method tries to find an approximation F(x) that minimizes the average value of the loss function on the training set, i.e., minimizes the empirical risk. It does so by starting with a model, consisting of a constant function $F_{0(x)}$, and incrementally expands it in a greedy fashion:

$$F_0(x) = \arg\min_{h_m \in \mathcal{H}} \sum_{i=1}^{n} L(y_i, h_m(x_i)),$$

$$F_m(x) = F_{m-1}(x) + \left(\arg\min_{h_m \in \mathcal{H}} \left[\sum_{i=1}^{n} L\left(y_i, F_{m-1}(x_i) + h_m(x_i)\right) \right] \right)(x)$$

for $m \geq 1$, where $h_m \in \mathcal{H}$ is a base learner function.

Unfortunately, choosing the best function hm at each step for an arbitrary loss function L is a computationally infeasible optimization problem in general. Therefore, we restrict our approach to a simplified version of the problem. The idea is to apply a steepest descent step to this minimization problem (functional

gradient descent). The basic idea is to find a local minimum of the loss function by iterating on $F_{m-1}(x)$. In fact, the local maximum-descent direction of the loss function is the negative gradient. Hence, moving a small amount such that the linear approximation remains valid:

$$F_m(x) = F_{m-1}(x) - \gamma \sum_{i=1}^{n} \nabla_{F_{m-1}} L(y_i, F_{m-1}(x_i))$$

where $\gamma > 0$. For small γ, this implies that $L(y_i, F_m(x_i)) \leq L(y_i, F_{m-1}(x_i))$.

Gradient Boosting Grid

Gradient boosting grid search is a hyperparameter tuning technique used to optimize the performance of gradient boosting algorithms. It involves systematically exploring a range of parameter values (like learning rate, number of estimators, max depth) for a gradient boosting model and evaluating the model's performance for each combination. This helps identify the parameter combination that yields the best results on a validation or test set.

CHAPTER 4

Chapter 4

ANALYSIS AND ITS INTERPRETATION

In this chapter presents the Analysis and Interpretation of its our data using analysis are as given below:

- Descriptive Statistics
- Graphical Representation
- Correlation Analysis
- Regression Analysis
- Principal Component Analysis
- Model Comparison

4.1 DESCRIPTIVE STATISTICS

Descriptive statistics were conducted on the following variables: density, agricultural land, land area, birth rate, consumer price index (CPI) and its change, fertility rate, GDP, gross enrollment in primary and tertiary education, infant mortality, life expectancy, maternal mortality, minimum wage, health expenditure, total population, labour wage, tax rate, tax revenue, unemployment rate, and urban population.

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation
Density	123	1378	2	1380	144.67	206.076
Agricultural_land	123	.80	.03	.83	.4154	.19572
Land_area	123	9984354	316	9984670	834073.36	1845696.202
Birth_rate	123	39.68	6.40	46.08	20.8951	10.09515
CPI	123	2638.40	101.87	2740.27	178.7538	263.06041
cpichange	123	2.580	-.030	2.550	.06691	.240693
Fertility_rate	123	5.93	.98	6.91	2.7643	1.31448
GDP	123	2.1427E+13	917059000.0	2.1428E+13	6.35006E+11	2.66371E+12
Grossprimaryeducation	123	.81	.62	1.43	1.0324	.11964
Grosstertitaryeducation	123	1.36	.01	1.37	.3921	.29427
infandmoratlity	123	76.80	1.70	78.50	21.7919	19.37286
lifeexpectancy	123	29.30	54.00	83.30	72.2057	7.33504
maternalmortality	122	1138.00	2.00	1140.00	154.1885	223.64946
minimumwage	122	13.59	.00	13.59	2.2241	3.05285
healthexpenditure	123	.79	.03	.82	.3566	.19166
population	123	1397617375	97625.0000	1397715000	54685784.60	179374031.1
labourwage	123	.86	.00	.86	.6120	.11298
tax_revenue	118	.37	.00	.37	.1539	.07088
taxrate	123	.97	.09	1.06	.4080	.14992
unemployment	123	.20	.00	.20	.0669	.04443
urban_population	123	842878200	55762	842933962	30066540.22	91734578.65
Valid N (listwise)	117					

Figure 4.1: Descriptive Statistics

INTERPRETATION

From the above table, in obtained by minimum, maximum, mean, standard deviation, and range values for each variable. The highest maximum value is for Population (1,397,715,000), and the lowest minimum value is for Minimum Wage (0.00). This indicates a wide variation across countries in terms of these variables.

4.2 GRAPHICAL REPRESENTATION

4.2.1 Bar Charts

Top 20 Countries Populated

These variables were used to create a bar chart for the following countries: China,India,US, Indonesia,Pakistan,Brazil,Nigeria, Bangladesh,Maxico,Philippines,Vietnam,The Republic of the Congo,Turkey,Germany,Iran,Thailand,France ,UK,Tanzania,Myanmar

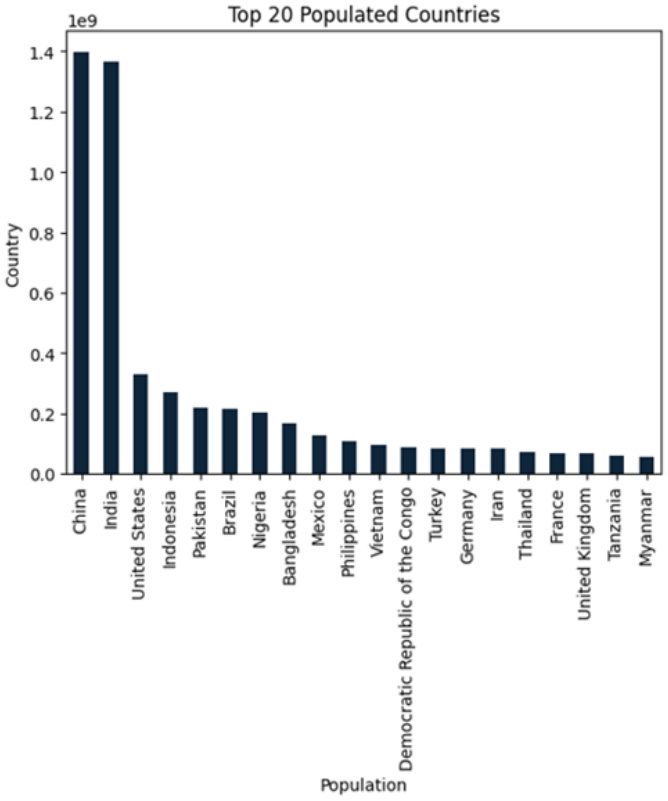

Figure 4.2: Top 20 countries populated

INTERPRETATION

The above bar Chart shows the world data for the top 20 most populated countries. China has the highest population, followed by India. The country with the lowest population among them is Myanmar

Bottom 20 Countries Populated

These variables were used to create a bar chart for the following countries:Seychelles,Barbados, Vanuatu,Belize,Cape Verde,Malta, Montenegro,Luxembourg,Guyana ,fiji,Mauritius,Estonia, Equatorial Guinea, Trinida and Tobago, Latvia,Guinea Bissau, Slovenia,Gabon, Botswana,Gambia

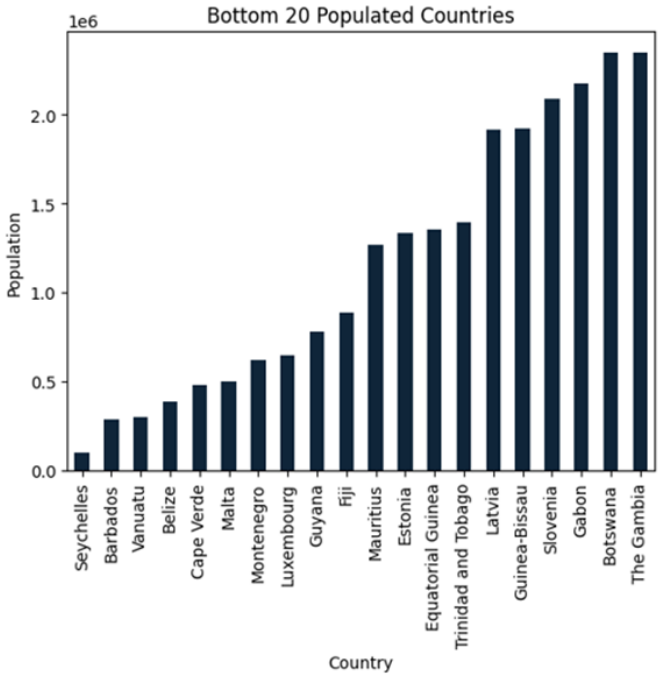

Figure 4.3: Bottom 20 countries populated

INTERPRETATION

The above Bar Chart shows the world data for the 20 Bottom populated countries. Botswana and The Gambia have equally high populations among the bottom 20 countries in the world . Seychelles has the lowest population among these countries

Top 20 Countries Highest Domestic Product

These variables were used to create a bar chart for the following countries: Venezuela,Belgium, Thailand,Poland,Turkey,Saudi Arabia,Netherland, Indonesia,Mexico,Australia,Spain,Canada ,Brazil,South Korea, India,France,UK,Germany,China,US

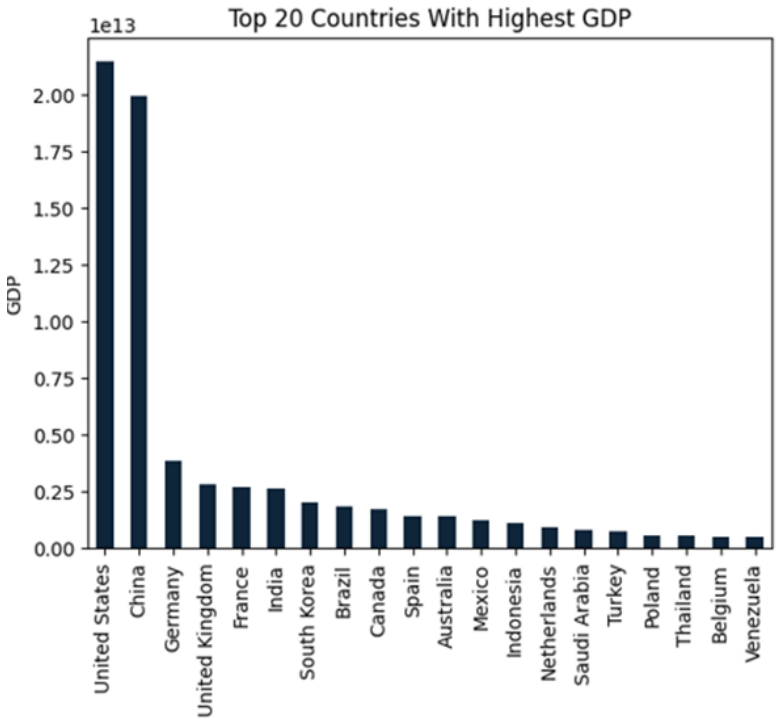

Figure 4.4: Top 20 countries with highest gross domestic product

INTERPRETATION

The above Bar chart shows that the United States holds the highest GDP(Gross Domestic Product), surpassing $20 trillion, making it the leading global economy. China follows as the second-largest economy, with a GDP slightly lower than that of the U.S. but still significantly ahead of other countries. Germany, the United Kingdom, France, and India rank next, showing strong but relatively lower GDP figures. South Korea, Brazil, Canada, Spain, Australia, and Mexico fall into the mid-range GDP category. Indonesia, the Netherlands, Saudi Arabia, Turkey, Poland, Thailand, Belgium, and Venezuela have the lowest GDP among the top 20 economies, with Venezuela at the bottom of the list.

Top 20 Countries Lowest Gross Domestic Product

These variables were used to create a bar chart for the following countries:US,China,Germany, UK,France,India,South Korea,Brazil,Canada,Spain,Australia,Mexico,Indonesia,Netherlands, Saudi Arabia,Turkey,Poland,Thailand,Belgium,Venezuela

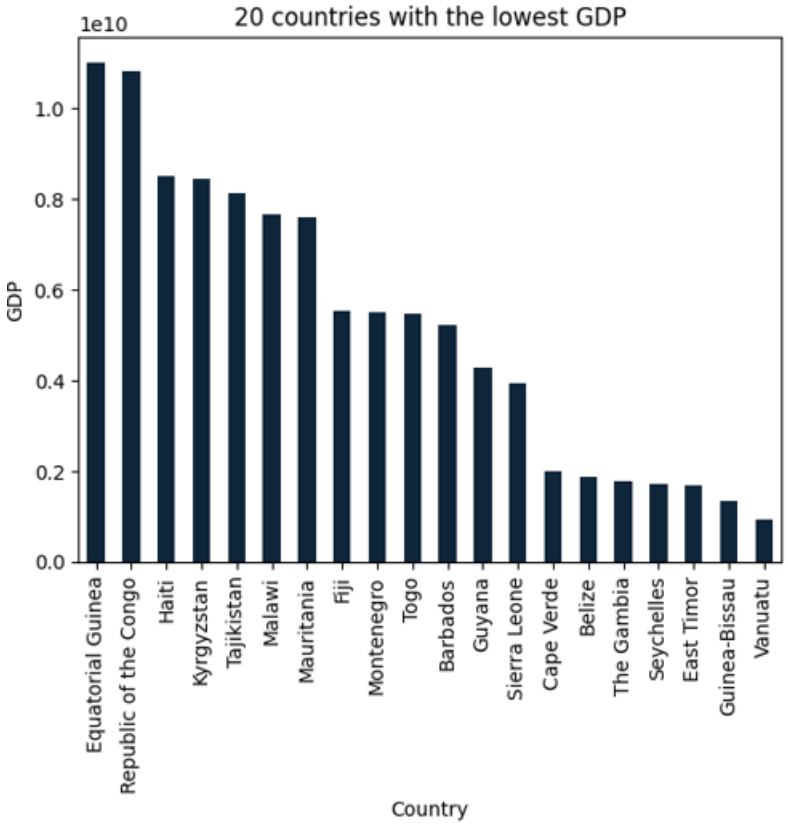

Figure 4.5: Top 20 Countries Lowest Gross Domestic Product

INTERPRETATION

The above bar chart shows the top 20 countries with the lowest GDP (Gross Domestic Product) in the world for the census . The country with the highest GDP among them is Equatorial Guinea, followed by the Republic of the Congo. The GDP of the remaining countries gradually decreases, with Vanuatu having the lowest GDP among them.

Gross Domestic Product Per Capita of 20 Countries with Highest Gross Domestic Product

These variables were used to create a bar chart for the following countries:US,China,Germany, UK,France,India,South Korea,Brazil,Canada,Spain,Australia,Mexico,Indonesia, Netherlands,Saudi Arabia,Turkey,Poland,Thailand,Belgium,Venezuela

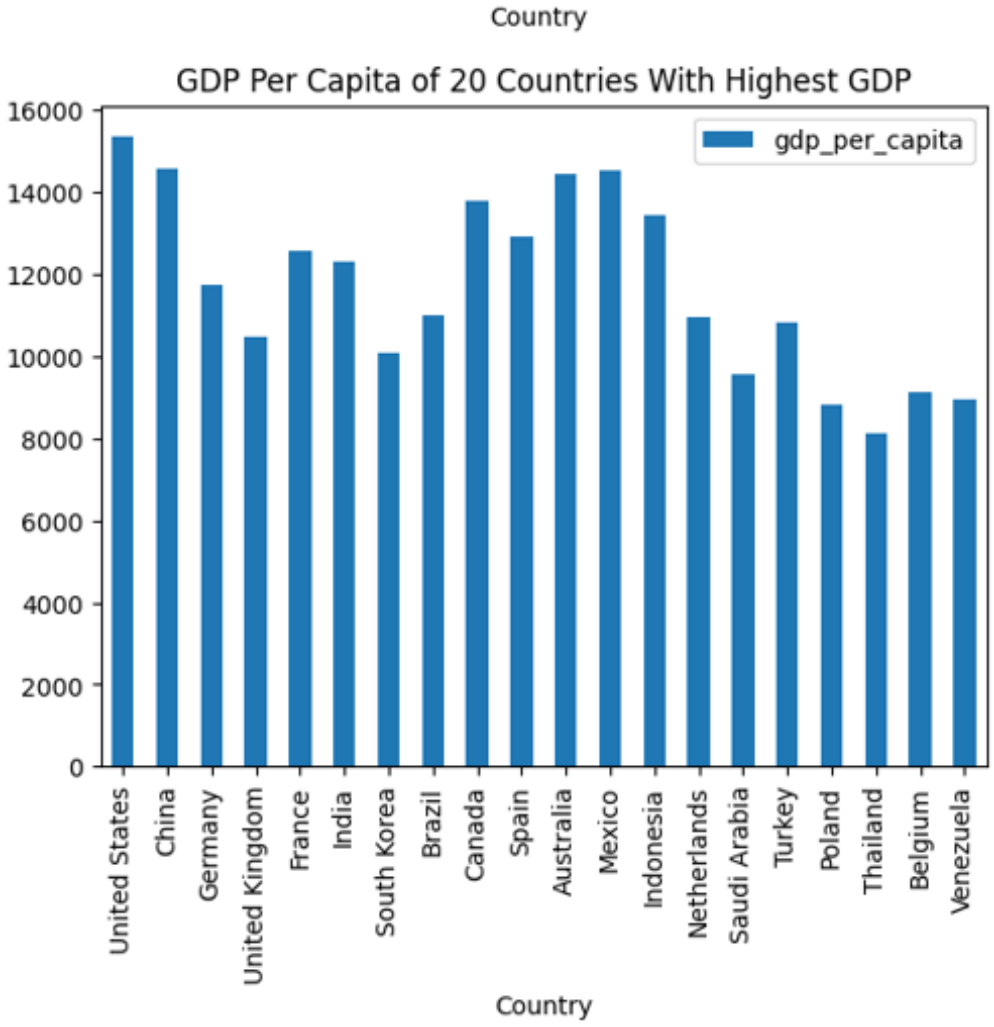

Figure 4.6: Gross Domestic Product Per Capita of 20 Countries with Highest Gross Domestic Product

INTERPRETATION

The above bar chart shows the GDP per capita of the 20 countries with the highest GDP in the world for the census . The United States leads with the highest GDP per capita, followed by China and other developed nations like Germany, the United Kingdom, and France. Developing economies such as India, Indonesia, and Brazil have lower GDP per capita despite having high total GDP. Countries like Thailand, Poland, and Venezuela have the lowest GDP per capita among the top 20 highest GDP nations. The data highlights economic disparities, showing that a high overall GDP does not necessarily equate to high individual income

Gross Domestic Product Per Capita of 20 Countries with Lowest Gross Domestic Product

These variables were used to create a bar chart for the following countries:Equatorial Guinea, Republic of the Congo Congo,Haiti,Kyrgyzstan,Tajikistan,Malawi, Mauritania,Fiji,Montenergro,Togo Barbados,Guyana,Sierra, Leone, Cape verde, Belize, The Gambia, Seychelles,East Timor,Guinea Bissau,Vanuatu

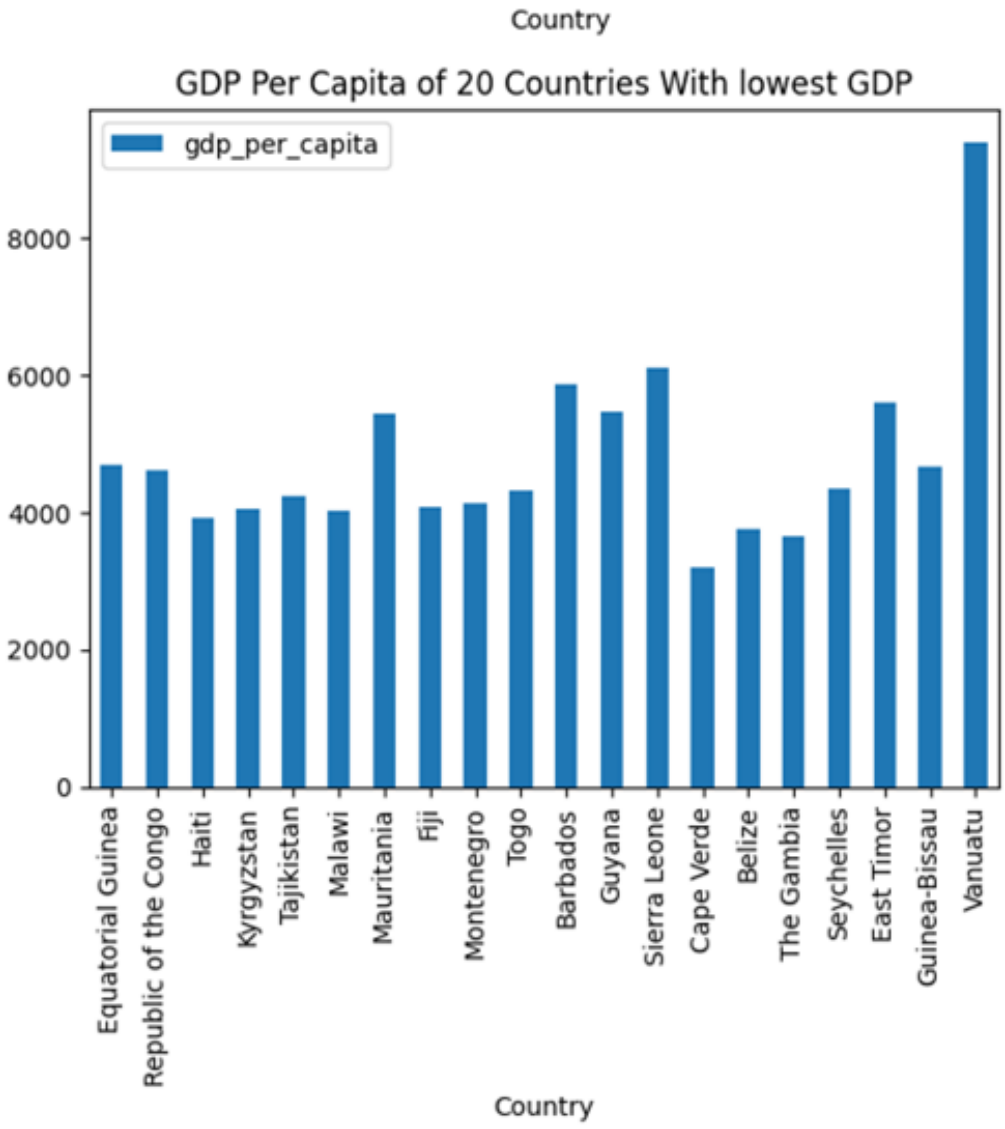

Figure 4.7: Gross Domestic Product Per Capita of 20 Countries with Lowest Gross Domestic Product

INTERPRETATION

The Bar chart shows the GDP per capita of the 20 countries with the lowest GDP. Vanuatu has the highest GDP per capita among them, followed by Sierra Leone, Barbados, and Mauritania, with Guyana and East Timor also ranking relatively high. Cape Verde has one of the lowest GDP per capita values in this group. Despite having low total GDP, some countries in this group show moderate GDP per capita, indicating that their wealth is distributed among a smaller population or that they have specific economic strengths. There are also economic disparities within low-GDP nations, as GDP per capita varies notably due to differences in development levels, population size, and economic activities

Top 20 Countries with Highest Life Expectancy

These variables were used to create a bar chart for the following countries:Barbados, Chile,Costa Rica,Germany,Slovenia,Greece,UK,Portugal,Belgium,Netharlands,New Zealand,Canada,Luxembourg,Republic of Ireland,Malta,France,South Korea,Australia, Israel,Spain

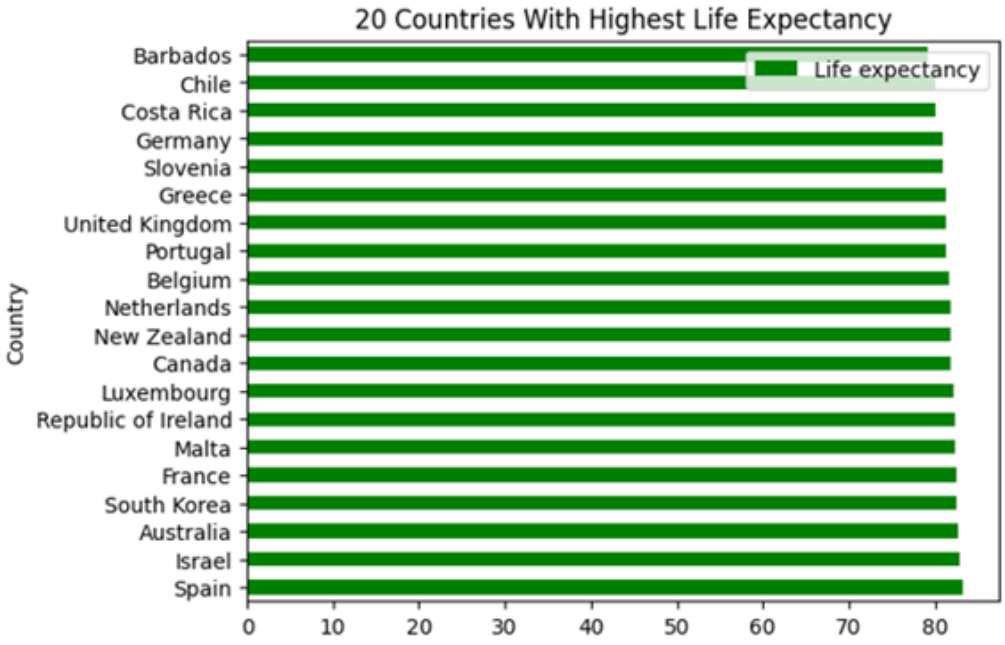

Figure 4.8: Top 20 countries with highest life expectancy

INTERPRETATION

The Bar chart shows the 20 countries with the highest life expectancy, all exceeding 80 years. European nations dominate the list, while non-European countries like South Korea, Australia, Israel, Canada, Chile, and Costa Rica also feature. High life expectancy in these countries is likely due to strong healthcare systems, healthy lifestyles, low infant mortality, and overall economic stability. Additionally, access to quality healthcare, balanced diets, public health initiatives, and economic stability play a crucial role in achieving longer lifespans.

Top 20 countries with lowest life expectancy

These variables were used to create a bar chart for the following countries:Malawi,Haiti, Zambia,Uganda,Niger,Gambia,Benin,Burkina Faso,Togo,Angola,Democratic Republic of the Congo,Mozambique,Mali,Cameroon,Equatorial Guinea,Guinea Bissau,Ivory Coast,Nigeria, Sierra Leone,Chad

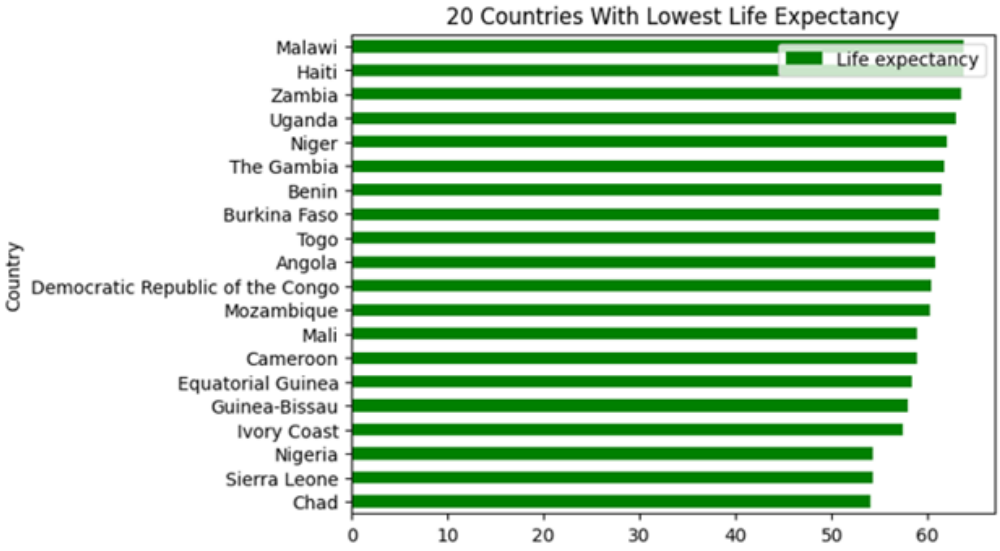

Figure 4.9: Top 20 countries with lowest life expectancy

INTERPRETATION

The Bar chart displays the 20 countries with the lowest life expectancy, showing a gradual decrease from Malawi, which has the highest life expectancy among them, to Chad, which has the lowest. This indicates a significant disparity in health outcomes among these nations. Factors contributing to lower life expectancy may include limited access to quality healthcare, high infant mortality rates, malnutrition, infectious diseases, and economic instability. Addressing these issues through healthcare improvements, better public health policies, and economic development could help increase life expectancy in these countries.

Top 20 Countries highest unemployment rate

These variables were used to create a bar chart for the following countries:South africa, Lesotho,Saint Lucia,Namibia,Gabon,Saint Vincent and the Grenadines,Libya, Bosnia and Herzegovina,Botswana,Greece,Armenia,Sudan,Tunisia,Montenegro, Jordan,US,Georgia,Spain,Haiti,Turkey

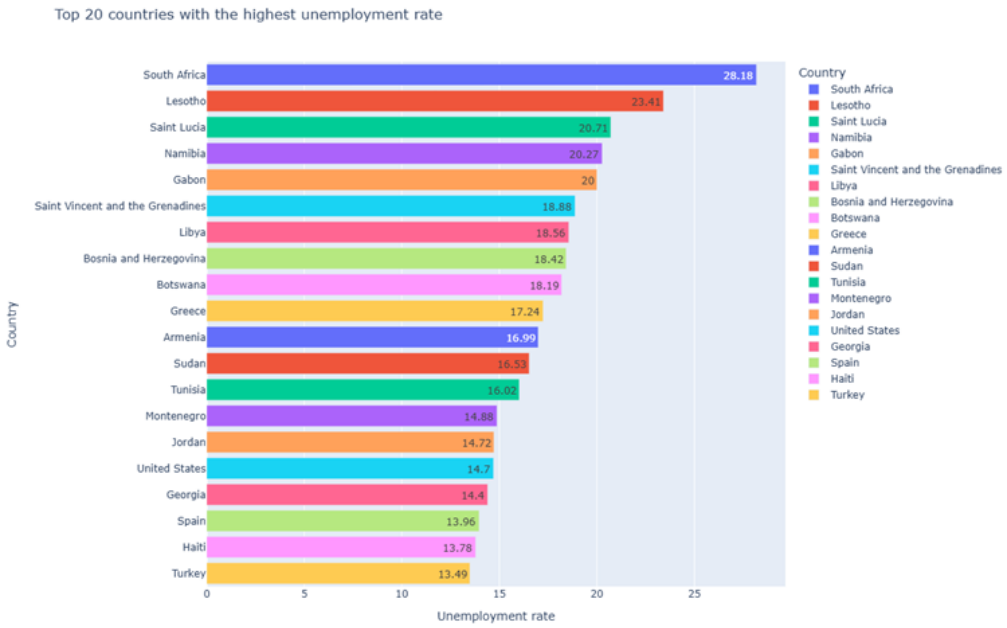

Figure 4.10: Top 20 Countries highest unemployment rate

INTERPRETATION

The Bar chart displays the top 20 countries with the highest unemployment rates. South Africa has the highest unemployment rate at (28.18%), followed by Lesotho (23.41%) and Saint Lucia (20.71%). The presence of multiple African countries, including Namibia, Gabon, and Botswana, indicates that unemployment is a significant challenge in the region. Additionally, European countries like Greece and Spain, as well as nations from the Americas and the Middle East, are also represented, showing that high unemployment is a global issue. At the lower end of the list, Turkey has the lowest unemployment rate among these 20 countries at (13.49%), followed by Haiti (13.78%) and Spain (13.96%). While these

rates are still relatively high, they are lower compared to the top-ranked nations. Factors contributing to high unemployment may include economic instability, lack of job opportunities, political issues, and structural weaknesses in labor markets. Addressing these challenges requires targeted policies to promote job creation, economic growth, and workforce development.

Largest city unemployment rate

These variables were used to create a bar chart for the following countries: Johannesburg,Maseru,Castries,Windhoek,Libreville,Calliaqua,Unknown, Tuzla canton,Gaborone,Macedonia

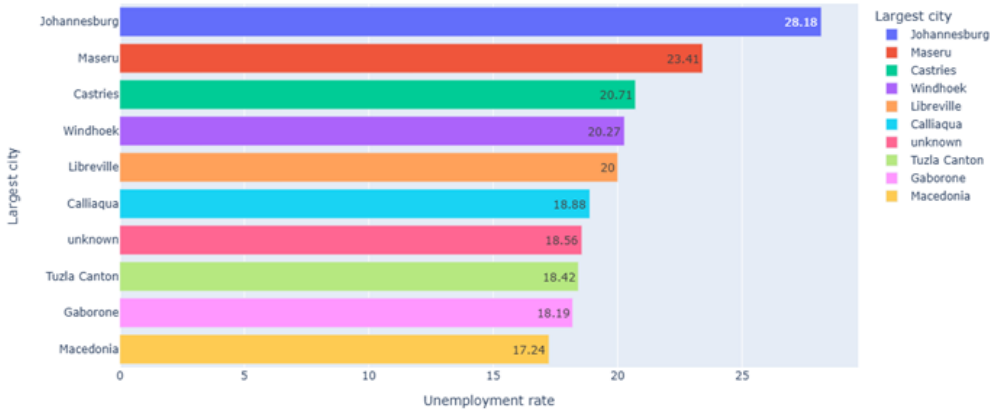

Figure 4.11: Largest city unemployment rate

INTERPRETATION

The Bar chart compares the unemployment rates of the largest cities. The city with the highest unemployment rate is Johannesburg (28.18%), followed by Maseru (23.41%). The unemployment rates of other cities decrease as follows: Castries (20.71%), Windhoek (20.27%), Libreville (20%), Calliaqua (18.88%), Libya (18.56%), Tuzla Canton (18.42%), Gaborone (18.19%), and Macedonia (17.24%). The city with the lowest unemployment rate is Macedonia (17.24%).

Top 20 countries with lowest unemployment rate

These variables were used to create a bar chart for the following countries: Oman,Kenya,Guinea Bissau,Guatemala,Papua New Guinea, Benin,Kuwait,Phillippines, Togo,Vietnam,Tanzania,Czech Republic,Chad,Uganda,Madagascar,Myanmar,Nepal, Thailand,Laos,Niger

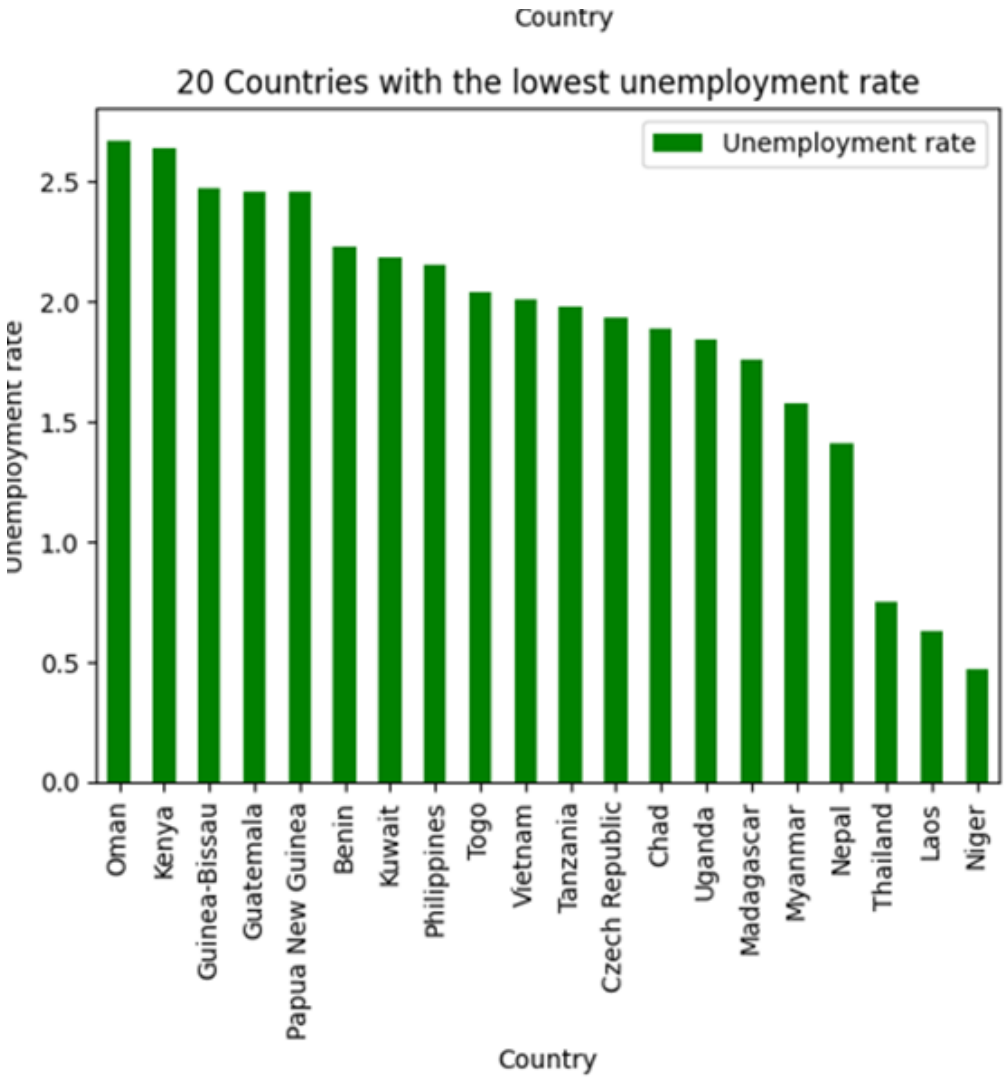

Figure 4.12: Top 20 countries with lowest unemployment rate

INTERPRETATION

The above bar chart presents a comparison of the 20 countries with the lowest unemployment rates. Oman has the highest unemployment rate among these 20 countries, while Niger has the lowest. The trend indicates that these countries have relatively stable labor markets with low unemployment. Countries such as Kenya, Guinea-Bissau, Guatemala, and Papua New Guinea also exhibit low unemployment rates, while nations like Myanmar, Nepal, Thailand, Laos, and Niger have the lowest rates overall structural economic factors contributing to minimal unemployment

Top 20 countries with the highest urban population percentage

These variables were used to create a bar chart for the following countries: Kuwait, Belgium,Uruguay,Malta,Israel,Argentina,Netherlands,Jordan,Gabon, Lebanon, Venezuela,New Zealand,Chile,Luxembourg, Brazil,Australia,Saudi Arabia,UK, Canada,US

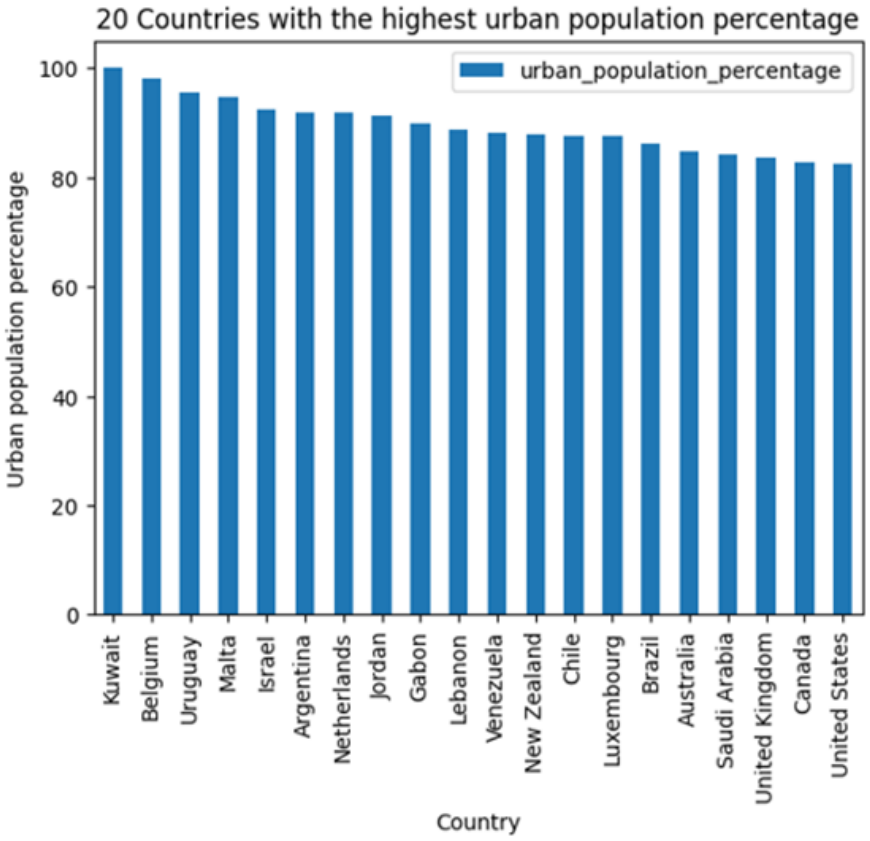

Figure 4.13: Top 20 countries with the highest urban population percentage

INTERPRETATION

The above bar chart presents the top 20 countries with the highest urban population percentage . Kuwait has the highest urban population percentage, followed closely by Belgium, Uruguay, and Malta, lowest urban population country is US. The trend indicates that these countries have a highly urbanized population, with the majority of residents living in cities rather than rural areas. Other countries on the list, such as Argentina, the Netherlands, Jordan, and Brazil, also exhibit high

levels of urbanization. The presence of developed nations like the United States, Canada, the United Kingdom, and Australia suggests that economic development and infrastructure play a significant role in urban population concentration

Top 20 countries with least urban population Percentage

These variables were used to create a bar chart for the following countries:Laos,Sudan, Tanzania,India,Barbados,Myanmar,Burkina Faso,Kenya,Tajikistan,Guyana,Afghanistan, Vanuatu,Uganda,Chad,Nepal,Sri lanka,Malawi,Niger,Papua new Guinea,East Timor

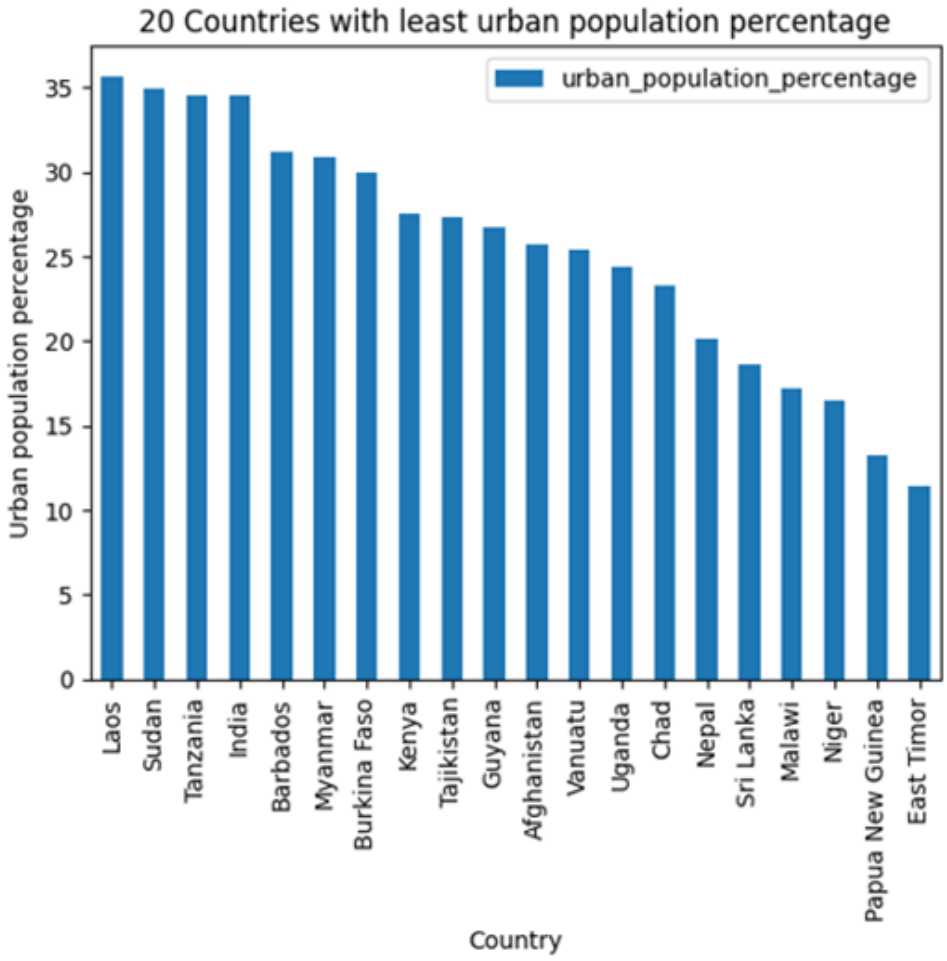

Figure 4.14: Top 20 countries with least urban population Percentage

INTERPRETATION

The above bar chart presents the 20 countries with the lowest urban population percentages. Laos has the high least urban population percentage among these 20 countries, followed closely by Sudan, Tanzania, and India. The trend in-

dicates that these countries have a predominantly rural population, with a smaller proportion of residents living in urban areas. Other countries on the list, such as Burkina Faso, Kenya, Afghanistan, and Chad, also exhibit low urbanization rates. East Timor has the lowest urban population percentage among the countries in the chart. The presence of developing nations suggests that factors such as limited infrastructure, lower levels of industrialization, and agricultural dependence contribute to the lower urbanization rates in these regions. This highlights the contrast between urban and rural development across different countries.

4.3 CORRELATION ANALYSIS

The variables used for correlation analysis are: GDP, Gross Primary Education, Gross Tertiary Education, Life Expectancy, Minimum Wage, Population, Labour Wage, Tax Rate, Unemployment, and Urban Population.

	CPI	GDP	Gross primary ed
GDP	-0.029		
Gross primary ed	-0.067	-0.019	
Gross tertiary e	0.047	0.216	-0.105
Life expectancy	-0.090	0.181	-0.003
Minimum wage	-0.137	0.207	-0.071
Out of pocket he	0.112	-0.140	-0.165
Population	0.000	0.628	0.039
Population: Labo	-0.044	0.031	0.266
Total tax rate	0.201	0.082	0.043
Unemployment rat	0.147	0.068	-0.050
Urban_population	0.000	0.780	0.023

	Gross tertiary e	Life expectancy	Minimum wage
Life expectancy	0.780		
Minimum wage	0.567	0.629	
Out of pocket he	-0.273	-0.379	-0.468
Population	0.012	0.011	-0.057
Population: Labo	-0.105	-0.182	-0.049
Total tax rate	0.003	-0.135	-0.076
Unemployment rat	0.159	0.111	-0.073
Urban_population	0.084	0.075	-0.001

	Out of pocket he	Population	Population: Labo
Population	0.116		
Population: Labo	-0.101	-0.025	
Total tax rate	0.154	0.136	-0.066
Unemployment rat	0.110	-0.055	-0.343
Urban_population	0.040	0.955	-0.005

	Total tax rate	Unemployment rat
Unemployment rat	0.115	
Urban_population	0.162	-0.015

Figure 4.15: correlation analysis

INTERPRETATION

The above correlation shows to reveals key relationships among socioeconomic indicators. Notably, life expectancy shows a strong positive correlation with gross tertiary education (0.780) indicating Life Expectancy Increasing the relationship of GDP is positively increasing. The higher education levels are linked to better health outcomes. Minimum wage is negatively associated with out-of-pocket health expenditure (-0.468) Minimum Wage increase ,there is out-of-pocket health expenditure decrease, suggesting better healthcare access in countries with higher wages. Urban population and labor force participation are highly correlated (0.955), reflecting urban areas as economic. GDP shows moderate positive correlations with life expectancy and minimum wage, Improved social conditions in wealthier economies.

4.4 MULTIPLE LINEAR REGRESSION ANALYSIS

"Multiple linear regression was performed using the following variables: CPI, GDP, gross primary education, gross tertiary education, life expectancy, health expenditure, population, labour wage, tax revenue, tax rate, and urban population."

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.483[a]	.233	.145	.04061

a. Predictors: (Constant), urban_population, CPI, labourwage, minimumwage, Grosstertitaryeducation, taxrate, Grossprimaryeducation, healthexpenditure, tax_revenue, GDP, lifeexpectancy, population

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	.0176	.1316	.0656	.02120	117
Residual	-.11995	.14930	.00000	.03846	117
Std. Predicted Value	-2.264	3.115	.000	1.000	117
Std. Residual	-2.953	3.676	.000	.947	117

a. Dependent Variable: unemployment

b. Dependent Variable: unemployment

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	.076	.077		.995	.322
	CPI	1.445E-5	.000	.089	.959	.340
	GDP	2.753E-15	.000	.171	.941	.349
	Grossprimaryeducation	.036	.034	.100	1.074	.286
	Grosstertitaryeducation	.026	.022	.176	1.201	.232
	lifeexpectancy	8.347E-5	.001	.014	.087	.931
	minimumwage	.001	.001	.072	.756	.451
	healthexpenditure	.031	.024	.133	1.263	.209
	population	-7.470E-11	.000	-.313	-.812	.419
	labourwage	-.139	.037	-.357	-3.767	.000
	tax_revenue	-.046	.067	-.075	-.691	.491
	taxrate	.030	.028	.102	1.077	.284
	urban_population	5.158E-11	.000	.110	.232	.817

a. Dependent Variable: unemployment

Figure 4.16: Multiple Linear Regression Analysis

INTERPRETATION

The multiple regression analysis was conducted to predict unemployment based on several independent variables like GDP, CPI, tax revenue, and others. The model shows a moderate correlation ($R = 0.483$) and explains about 23.3% ($R^2 = 0.233$) of the variance in unemployment. The Standard Error of the Estimate is 0.40601, suggesting moderate prediction accuracy. From the coefficients table, minimum wage have significant negative impacts on unemployment (p-values < 0.05), while urban population shows a significant positive impact. Other predictors like GDP, taxrate, and population are not statistically significant. Residual statistics suggest that errors are reasonably centered around zero with small variation.

The regression equation for predicting unemployment is:

Unemployment $= 0.076 + 0.00001445 \times$ CPI $+ 2.753E - 15 \times$ GDP $+ 0.036\times$ GrossPrimaryEducation $+0.026\times$ GrossTertiaryEducation $+0.00008347\times$ LifeExpectancy $+0.001\times$ MinimumWage $+0.031\times$ HealthExpenditure $-7.470E - 11\times$ Population $-0.139\times$ LabourWage $-0.046\times$ TaxRevenue $+0.030\times$ TaxRate $+5.158E-11\times$ UrbanPopulation

4.5 PRINCIPAL COMPONENT ANALYSIS

"Principal component analysis was carried out using the following variables: density, agricultural land, land area, birth rate, CPI and its change, fertility rate, GDP, gross primary and tertiary education, infant mortality, life expectancy, maternal mortality, minimum wage, health expenditure, population, labour wage, tax rate, tax revenue, unemployment, and urban population."

Communalities

	Initial	Extraction
Density	1.000	.762
Agricultural_land	1.000	.496
Land_area	1.000	.734
Birth_rate	1.000	.933
CPI	1.000	.924
cplchange	1.000	.949
Fertility_rate	1.000	.905
GDP	1.000	.788
Grossprimaryeducation	1.000	.653
Grosstertitaryeducation	1.000	.787
infandmoratlity	1.000	.905
lifeexpectancy	1.000	.919
maternalmortality	1.000	.714
minimumwage	1.000	.603
healthexpenditure	1.000	.639
population	1.000	.860
labourwage	1.000	.772
tax_revenue	1.000	.518
taxrate	1.000	.535
unemployment	1.000	.696
urban_population	1.000	.940

Extraction Method: Principal Component Analysis.

Figure 4.17: communalities

INTERPRETATION

This table shows the communalities from a Principal Component Analysis (PCA), representing the proportion of each variable's variance explained by the extracted components. Initial communalities are all 1.000, indicating total variance before extraction. The "Extraction" values show how much variance is retained for each variable after dimensionality reduction. Variables like Birth rate (0.933), urban population (0.940), and lifeexpectancy (0.919) are well-explained by the components, while Agricultural land (0.496) and taxrate (0.535) are less so. Generally, higher communalities (>0.6) suggest the variable is well-represented in the factor solution

Total Variance Explained

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.013	23.872	23.872	5.013	23.872	23.872
2	3.235	15.407	39.279	3.235	15.407	39.279
3	2.285	10.883	50.162	2.285	10.883	50.162
4	1.692	8.058	58.221	1.692	8.058	58.221
5	1.417	6.748	64.969	1.417	6.748	64.969
6	1.305	6.214	71.183	1.305	6.214	71.183
7	1.084	5.161	76.344	1.084	5.161	76.344
8	.897	4.274	80.617			
9	.749	3.565	84.182			
10	.707	3.369	87.551			
11	.602	2.865	90.416			
12	.524	2.494	92.910			
13	.434	2.067	94.978			
14	.325	1.547	96.525			
15	.260	1.236	97.760			
16	.204	.970	98.731			
17	.143	.679	99.409			
18	.050	.239	99.649			
19	.046	.221	99.869			
20	.018	.086	99.955			
21	.009	.045	100.000			

Extraction Method: Principal Component Analysis.

Figure 4.18: Total Variance Explained

INTERPRETATION

This table shows how much variance is explained by each principal component in the PCA. The first six components have eigenvalues greater than 1 and together explain approximately 71.18% of the total variance, indicating they capture most of the important information. The first component alone accounts for 23.87% of the variance. Components with lower eigenvalues contribute minimally and can typically be discarded. This helps in reducing the dimensionality while retaining key data patterns.

Component Matrix

Component Matrix[a]

	Component						
	1	2	3	4	5	6	7
Density	-.202	-.063	-.169	.325	-.081	.755	.081
Agricultural_land	.038	.161	-.065	.401	.161	.014	.527
Land_area	-.197	.756	-.093	-.243	.001	-.227	-.073
Birth_rate	.953	.045	-.076	-.080	-.039	-.088	.039
CPI	.134	.190	.849	-.132	.247	.250	-.091
cpichange	.073	.208	.862	-.143	.268	.251	-.049
Fertility_rate	.933	.052	-.080	-.084	-.067	-.118	-.012
GDP	-.291	.803	-.175	-.016	.005	-.125	-.109
Grossprimaryeducation	.046	.006	-.242	-.100	.495	.071	.576
Grosstertitaryeducation	-.823	.031	.250	.047	.105	-.182	.000
infandmoratlity	.939	.112	-.055	.061	-.055	.004	.002
lifeexpectancy	-.948	-.087	.048	-.089	.021	.043	.031
maternalmortality	-.002	-.072	-.224	-.689	-.281	.315	.076
minimumwage	-.125	-.159	.059	.698	.044	-.192	-.178
healthexpenditure	.383	.111	.211	.533	-.322	.217	-.016
population	-.106	.856	-.211	.142	-.087	.207	-.011
labourwage	.215	.049	-.262	.079	.780	-.130	-.151
tax_revenue	-.525	-.386	-.213	-.006	-.114	.003	.185
taxrate	.185	.326	.348	-.062	-.100	-.068	.505
unemployment	-.155	.077	.438	.073	-.398	-.473	.294
urban_population	-.178	.919	-.200	.097	-.058	.100	-.034

Extraction Method: Principal Component Analysis.

a. 7 components extracted.

Figure 4.19: Component Matrix

INTERPRETATION

The component matrix shows the loading of each variable on the extracted components, indicating how strongly a variable correlates with each component.The Birthrate (0.953), Fertility rate (0.933), and infant mortality (0.939) load highly on Component it represents health-related factors. urban population (0.919) and GDP (0.803) load strongly on Component likely reflecting urbanization and economic development. High loadings (absolute values 0.7) are key for interpreting the nature of each component. This matrix helps identify patterns and reduce data complexity.

Scree plot

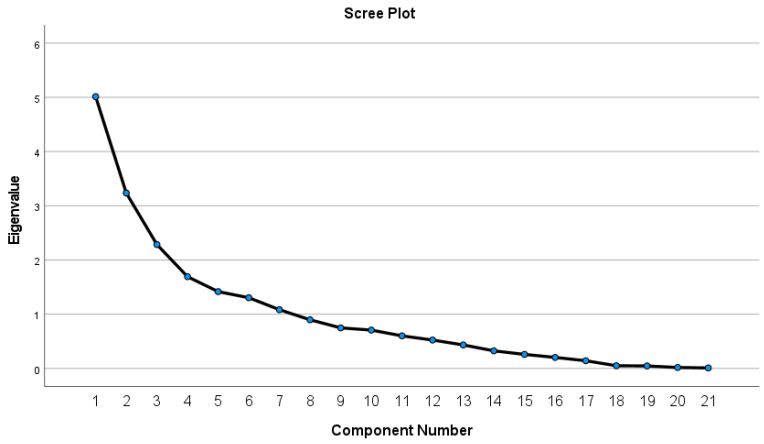

Figure 4.20: Scree plot

INTERPRETATION

The scree plot displays the eigenvalues of each principal component in descending order. A clear "elbow" appears at the six component, indicating that the first six components explain most of the variance in the data and are worth retaining. After this point, the eigenvalues level off, suggesting that additional components contribute minimally. This visual supports the selection of six components for meaningful dimensionality reduction, consistent with the Total Variance Explained table.

4.6 MODEL COMPARISON

Model comparison is used to find the best algorithm for a dataset by evaluating their performance.it helps understand how different models behave with the same data tuned models grid search cam improve accuracy over default setting

Model comparison was conducted using the following variables: density, agricultural land, land area, birth rate, CPI and its change, fertility rate, GDP, gross primary and tertiary education, infant mortality, life expectancy, maternal mortality, minimum wage, health expenditure, population, labour wage, tax rate, tax revenue, and urban population. In this model consider unemployment is dependent variable

Modeling

Train	156	28
Test	39	28

Linear Regression

MAE score Train	3.46
MAE score Test	4.96
RMSE score Train	4.46
RMSE score Test	7.09
R^2 score Train	0.29
R^2 score Test	-1.12

INTERPRETATION

- Training: Relatively poor fit (MAE=3.46, RMSE=4.46, R^2=0.29).
- Testing: Very poor generalization (MAE=4.96, RMSE=7.09, R^2=-1.12).

The negative R-squared indicates the model performs worse than a simple average predictor. Linear Regression is not a suitable model for this data, likely due to non-linear relationships or other complexities it cannot capture.

Random Forest Regressor

Random Forest Regressor is an ensemble learning method that builds multiple decision trees on different subsets of data and averages their predictions for regression tasks. It handles non-linear relationships well, reduces overfitting, and is robust to outliers. To improve performance, Grid Search can be used to tune hyperparameters like the number of trees (n estimators) or tree depth (max depth). This process tests various parameter combinations using cross-validation to find the best model. Together, Random Forest and Grid Search provide a powerful and accurate regression approach.

MAE score Train	1.24
MAE score Test	3.06
RMSE score Train	1.65
RMSE score Test	3.88
R^2 score Train	0.9
R^2 score Test	0.37

INTERPRETATION

Training: Good fit (MAE=1.24, RMSE=1.65, R^2=0.90), suggesting it learns the training data well. Testing: Decent generalization (MAE=3.06, RMSE=3.88, R^2=0.37), a significant improvement over Linear Regression. Some overfitting is present (performance drop from train to test).

Random forest grid

MAE score Train	1.51
MAE score Test	3.44
RMSE score Train	2.13
RMSE score Test	4.14
R^2 score Train	0.84
R^2 score Test	0.28

INTERPRETATION

Training: Slightly worse fit than default (MAE=1.51, RMSE=2.13, R^2=0.84). Testing: Slightly worse generalization than default (MAE=3.44, RMSE=4.14, R^2=0.28). In this case, the default hyperparameters of Random Forest seem to generalize slightly better than the tuned ones found by Grid Search. The Grid Search might have over emphasized training performance.

Gradient boosting

Gradient Boosting is a machine learning technique used for regression and classification that builds models sequentially, where each new tree corrects the errors of the previous one. Its purpose is to create a strong model by combining many weak learners , leading to high accuracy. It's especially useful for handling complex, nonlinear relationships in data. However, it can overfit if not properly tuned. That's where Grid Search comes in—it helps find the best combination of hyperparameters (like learning rate, number of trees, and tree depth) to improve performance and avoid overfitting. Together, they aim to build an accurate and well-generalized predictive model.

MAE score Train	0.62
MAE score Test	3.24
RMSE score Train	0.79
RMSE score Test	4.13
R^2 score Train	0.98
R^2 score Test	0.28

INTERPRETATION

Training: Very good fit (MAE=0.62, RMSE=0.79, R^2=0.98), indicating a strong ability to learn complex patterns. Testing: Moderate generalization (MAE=3.24, RMSE=4.13, R^2=0.28), a significant drop from training performance, indicating substantial overfitting.

Gradient boosting grid

MAE score Train	0.0
MAE score Test	2.79
RMSE score Train	0.0
RMSE score Test	3.85
R^2 score Train	1.0
R^2 score Test	0.38

INTERPRETATION

Training: Perfect fit (MAE=0.00, RMSE=0.00, R^2=1.00), indicating the model has essentially memorized the training data. Testing: Best performance among all models (MAE=2.79, RMSE=3.85, R^2=0.38). While the training performance shows extreme overfitting, the tuning process has yielded the best generalization on the test set.

Model Comparisons

	Metric	Linear regression	Random forest	Random forest grid	Gradient boosting	Gradient boosting grid
0	MAE score Train	3.46	1.24	1.51	0.62	0.00
1	MAE score Test	4.96	3.06	3.44	3.24	2.79
2	RMSE score Train	4.46	1.65	2.13	0.79	0.00
3	RMSE score Test	7.09	3.88	4.14	4.13	3.85
4	R^2 score Train	0.29	0.90	0.84	0.98	1.00
5	R^2 score Test	-1.12	0.37	0.28	0.28	0.38

Figure 4.21: Comparison of different Models

INTERPRETATION

Best Generalizing Model: Gradient Boosting achieves the best performance on the unseen test data across all metrics .Over fitting: Over fitting is a concern for both Random Forest and especially Gradient Boosting, as evidenced by the large discrepancy between training and testing performance.

The "Gradient Boosting grid" model exhibits extreme over fitting on the training data but still generalizes best. Testing Performance: The testing metrics are crucial for evaluating how well each model generalizes to unseen data. Lower MAE and RMSE, and higher R-squared on the test set indicate better generalization.

CONCLUSION

Chapter 5

CONCLUSION

In this project study deals with A Statistical Perspective on Importance of Unemployment in Global Economic Modeling. Unemploymentis is macroeconomic.In macroeconomic like country,birth rate,consumer price index,fertility rate, GDP,primary education,tertitary education,life expectancy,mortality rate, population,labour force, unemployment, urban population etc., Unemployment is a key indicator of a country's economic health and labor market efficiency. High unemployment can lead to poverty, reduced spending, and social issues. It helps policy makers design effective job creation and economic policies. Addressing unemployment is essential for sustainable development and social stability

This study aims to understand the factors influencing unemployment and its implications for economic development. The dataset used for this analysis was collected from the Kaggle website and includes 23 observations spanning various socio-economic and demographic indicators. In unemployment is to reduce poverty, promote economic stability, and ensure social inclusion. In our study used the various statistical methodologies are Descriptive statistics.Graphical Representation,Correlation,Regression,Principal Component Analysis,Model Comparison

In Descriptive statistics,gives the information of our data. It reveal the Economic and socio economic factor,based on 23 observations. Exhibit comparable central Tendencies. In Minimum Value abd Maximum vaule obtained our data is highest value is Population of the country lowest value is minimu wages, Standard deviation , Range also obtained In visualization of our data using Bar chart.

In correlation Analysis , there is relationships among socioeconomic in-

dicators. Notably, life expectancy shows a strong positive correlation with gross tertiary education (0.780) increase life expectancy also increase Gross tertiary education, indicating that higher education levels are linked to better health outcomes. Minimum wage is negatively associated with out-of-pocket health expenditure (-0.468) minimum wage increase then health expenditure decrease, suggesting better health care access in countries with higher wages. Urban population and labor force participation are highly correlated (0.955), reflecting urban areas as economic hubs. GDP shows moderate positive correlations with life expectancy and minimum wage, hinting at improved social conditions in wealthier economies.

In Multiple Regression Analysis was conducted to predict unemployment based on several independent variables like GDP, CPI, tax revenue, and others. The model shows a moderate correlation ($R = 0.483$) and explains about 23.3% ($R^2 = 0.233$) of the variance in unemployment. The Standard Error of the Estimate is 0.40601, suggesting moderate prediction accuracy. From the coefficients table, minimum wage have significant negative impacts on unemployment (p-values < 0.05), while urban population shows a significant positive impact. Other predictors like GDP, taxrate, and population are not statistically significant. Residual statistics suggest that errors are reasonably centered around zero with small variation.

In Principal Component Analysis (PCA) reveals that six components with eigenvalues greater than 1 explain approximately 71.18% of the total variance, indicating effective dimensionality reduction. Variables such as Birth rate, Urban population, and Life expectancy are well-represented with high communalities, while Agricultural land and Tax rate are less explained. The component matrix highlights interpretable patterns, with health and development indicators forming distinct components. The scree plot supports the retention of six components, as additional ones add little value. Overall, the PCA successfully simplifies the dataset while preserving the most important underlying structures.

In Unemployment Rates, three machine learning models were used: linear regression, random forest, and gradient boosting. These models were evaluated using training and testing performance metrics. As expected, models performed better on training data than on testing data, suggesting some level of overfitting. Testing performance was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared values.
Grid Search (hyperparameter tuning) was applied to improve the performance of

random forest and gradient boosting models.
While tuning helped improve the testing accuracy, it also slightly increased overfitting in the gradient boosting model.

In conclusion, this project provides valuable insights into global unemployment trends and the socio-economic factors that influence them. If unemployment continues to rise in the future, it will create serious challenges for individuals and nations. Therefore, it is important to increase awareness about the importance of employment, especially among the youth. Encouraging skill development, education, and economic participation will play a crucial role in reducing unemployment and supporting national development. This study highlights the need for data-driven policies and programs to address the issue of unemployment on a global scale.

BIBLIOGRAPHY

Bibliography

BOOKS

- Fundamentals Mathematical Statistics –S.C. Gupta ,V.K. Kapoor
- Montgomery, D.C., Peck ,E.A.,Vining,G.G(2012) Introduction to linear Regression Analysis. John Wiley sons

ARTICLES

- Akvile(Aleksandraviciene)-Assessment of the impact macroeconomic factors on unemployment .,vytautas magnus university

WEBSITES

- https://www.kaggle.com/
- https://github.com/sebalemus/world-data-2023/blob/main/world data2023.ipynb
- https://www.investopedia.com/terms/s/statistics.asp
- https://egyankosh.ac.in/bitstream/123456789/39247/1/Unit-4.pdf
- https://en.wikipedia.org/wiki/Unemployment History
- https://www.imf.org/en/Home
- https://www.worldbank.org/ext/en/home
- https://www.geeksforgeeks.org/machine-learning/
- https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/
- https://en.wikipedia.org/wiki/Random forest

- https://www.geeksforgeeks.org/ml-gradient-boosting/
- https://en.wikipedia.org/wiki/Gradient boosting
- https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS
- https://www.ilo.org/
- https://tradingeconomics.com/country-list/unemployment-rate
- https://www.bls.gov/

APPENDIX

Table 5.1: Demography and social Economic Data

Country	Density	Agricultural Land	Land Area	Birth Rate	CPI
Afghanistan	60	0.58	652,230	32.49	149.9
Albania	105	0.43	28,748	11.78	119.05
Algeria	18	0.17	2,381,741	24.28	151.36
Angola	26	0.48	1,246,700	40.73	261.73
Argentina	17	0.54	2,780,400	17.02	232.75
Armenia	104	0.59	29,743	13.99	129.18
Australia	3	0.48	7,741,220	12.6	119.8
Azerbaijan	123	0.58	86,600	14	156.32
Bangladesh	1,265	0.71	148,460	18.18	179.68
Barbados	668	0.23	430	10.65	134.09
Belgium	383	0.45	30,528	10.3	117.11
Belize	17	0.07	22,966	20.79	105.68
Benin	108	0.33	112,622	36.22	110.71
Botswana	4	0.46	581,730	24.82	149.75
Brazil	25	0.34	8,515,770	13.92	167.4
Bulgaria	64	0.46	110,879	8.9	114.42
Burkina Faso	76	0.44	274,200	37.93	106.58
Ivory Coast	83	0.65	322,463	35.74	111.61
Cape Verde	138	0.2	4,033	19.49	110.5
Cameroon	56	0.21	475,440	35.39	118.65
Canada	4	0.07	9,984,670	10.1	116.76
Chad	13	0.4	1,284,000	42.17	117.7
Chile	26	0.21	756,096	12.43	131.91
China	153	0.56	9,596,960	10.9	125.08
Colombia	46	0.4	1,138,910	14.88	140.95
Republic of the Congo	16	0.31	342,000	32.86	124.74
Costa Rica	100	0.35	51,100	13.97	128.85
Croatia	73	0.28	56,594	9	109.82
Czech Republic	139	0.45	78,867	10.7	116.48

CurrencyCode	F-Rate	GDP	Gross primary	Gross tertiary	Infantmortality
AFN	4.47	1.91E+10	1.04	0.1	47.9
ALL	1.62	1.53E+10	1.07	0.55	7.8
DZD	3.02	1.70E+11	1.1	0.51	20.1
AOA	5.52	9.46E+10	1.14	0.09	51.6
ARS	2.26	4.50E+11	1.1	0.9	8.8
AMD	1.76	1.37E+10	0.93	0.55	11
AUD	1.74	1.39E+12	1	1.13	3.1
AZN	1.73	3.92E+10	1	0.28	19.2
BDT	2.04	3.03E+11	1.17	0.21	25.1
BBD	1.62	5.21E+09	0.99	0.65	11.3
EUR	1.62	5.30E+11	1.04	0.8	2.9
BZD	2.31	1.88E+09	1.12	0.25	11.2
XOF	4.84	1.44E+10	1.22	0.12	60.5
BWP	2.87	1.83E+10	1.03	0.25	30
BRL	1.73	1.84E+12	1.15	0.51	12.8
BGN	1.56	8.6E+10	0.89	0.71	5.9
XOF	5.19	1.57E+10	0.96	0.07	49
XOF	4.65	5.88E+10	1	0.09	59.4
CVE	2.27	1.98E+09	1.04	0.24	16.7
XAF	4.57	3.88E+10	1.03	0.13	50.6
CAD	1.5	1.74E+12	1.01	0.69	4.3
XAF	5.75	1.13E+10	0.87	0.03	71.4
CLP	1.65	2.82E+11	1.01	0.89	6.2
CNY	1.69	1.99E+13	1	0.51	7.4
COP	1.81	3.24E+11	1.15	0.55	12.2
XAF	4.43	1.08E+10	1.07	0.13	36.2
CRC	1.75	6.18E+10	1.13	0.55	7.6
HRK	1.47	6.04E+10	0.97	0.68	4
CZK	1.69	2.46E+11	1.01	0.64	2.7

Maternal	wage	exp	Population	Labor	Tax.R	tax rate	UMP	Urban
638	0.43	0.78	38,041,754	0.49	0.09	0.71	0.11	9,797,273
15	1.12	0.57	2,854,191	0.56	0.19	0.37	0.12	1,747,593
112	0.95	0.28	43,053,054	0.41	0.37	0.66	0.12	31,510,100
241	0.71	0.33	31,825,295	0.78	0.09	0.49	0.07	21,061,025
39	3.35	0.18	44,938,712	0.61	0.1	1.06	0.1	41,339,571
26	0.66	0.82	2,957,731	0.56	0.21	0.23	0.17	1,869,848
6	13.59	0.2	25,766,605	0.66	0.23	0.47	0.05	21,844,756
26	0.47	0.79	10,023,318	0.67	0.13	0.41	0.06	5,616,165
173	0.51	0.72	167,310,838	0.59	0.09	0.33	0.04	60,987,417
27	3.13	0.45	287,025	0.65	0.28	0.36	0.1	89,431
5	10.31	0.18	11,484,055	0.54	0.24	0.55	0.06	11,259,082
36	1.65	0.23	390,353	0.65	0.26	0.31	0.06	179,039
397	0.39	0.41	11,801,151	0.71	0.11	0.49	0.02	5,648,149
144	0.29	0.05	2,346,179	0.71	0.2	0.25	0.18	1,616,550
60	1.53	0.28	212,559,417	0.64	0.14	0.65	0.12	183,241,641
10	1.57	0.48	6,975,761	0.55	0.2	0.28	0.04	5,256,027
320	0.34	0.36	20,321,378	0.66	0.15	0.41	0.06	6,092,349
617	0.36	0.36	25,716,544	0.57	0.12	0.5	0.03	13,176,900
58	0.68	0.23	483,628	0.61	0.2	0.38	0.12	364,029
529	0.35	0.7	25,876,380	0.76	0.13	0.58	0.03	14,741,256
10	9.51	0.15	36,991,981	0.65	0.13	0.25	0.06	30,628,482
1140	0.6	0.56	15,946,876	0.71		0.64	0.02	3,712,273
13	2	0.32	18,952,038	0.63	0.18	0.34	0.07	16,610,135
29	0.87	0.32	1,397,715,000	0.68	0.09	0.59	0.04	842,933,962
83	1.23	0.18	50,339,443	0.69	0.14	0.71	0.1	40,827,302
378	0.88	0.44	5,380,508	0.69	0.09	0.54	0.09	3,625,010
27	1.84	0.22	5,047,561	0.62	0.14	0.58	0.12	4,041,885
8	2.92	0.15	4,067,500	0.51	0.22	0.21	0.07	2,328,318
3	3	0.15	10,669,709	0.61	0.15	0.46	0.02	7,887,156