

Analysis of World Population Growth Using Python

Yash Tyagi^{1,*}, Neha Yadav²

Abstract

In this research work, we studied data analysis using the Python programming language. The fundamental steps in data analysis, such as cleansing, converting, and modelling of data is briefly explained in this study. In order to come up with good results, data analysis is required. Python has been used by us for data analysis. This language is interactive, interpreted, and follows an object-oriented programming paradigm. It is open source and comes with a variety of libraries, including MATplotlib, Seaborn, and Pandas. In this study, we focus primarily on the insights that might be discovered through exploratory data analysis of an existing dataset. The data from the dataset will be analysed graphically using various Python libraries and functions. The "World Population" dataset is used in this analysis to examine and extract different data in both numerical and visual form.

Keywords: Data analysis, exploratory data analysis, Pandas, Seaborn, matplotlib

INTRODUCTION

Population growth analysis is a vital field of study because data analysis can provide crucial understanding of the trends and patterns of population expansion across time. Python offers a wide range of tools and modules that make it easy to handle, manipulate, and visualise data. Analysis and display of historical population data for different countries are required for the project. Population data for many countries are contained in the dataset used for the project from 1990 through 2021. By using a variety of plots and charts to visualise the data, it is hoped to provide a better understanding of population growth in different countries.

BACKGROUND

Understanding and addressing a wide range of global concerns require a thorough understanding of world population analysis. The world population has changed significantly over time. Due to high death rates and poor technical development in ancient times, population growth was very modest. However, population growth rates quickened with the start of the industrial revolution and advancements in agriculture, sanitation, and healthcare. As societies modernized, this time period also saw the demographic transition, which was characterized by decreased fertility and mortality rates. Regional population growth rates vary; some see tremendous expansion while others struggle with an ageing population and declining birth rates. This study attempts to offer light on the present trends, issues, and implications that shape our global society by undertaking an extensive investigation of the world population.

*Author for Correspondence

Yash Tyagi
E-mail: yashtyagi181@gmail.com

^{1,2}Research Scholar, MCA, Thakur Institute of Management Studies, Career Development & Research (TIMSCDR), Mumbai, Maharashtra, India

Received Date: February 29, 2024
Accepted Date: April 03, 2024
Published Date: April 4, 2024

Citation: Yash Tyagi, Neha Yadav. Analysis of World Population Growth Using Python. Recent Trends in Programming Languages. 2024; 11(1): 15–20p.

LITERATURE REVIEW

The American Economic Review article "An econometric analysis of population growth" by Adelman, offers a thorough examination of the connection between population growth and economic variables. The author uses econometric models to analyse how factors like wealth, education, fertility rates, and mortality rates affect population dynamics. The study discovers a negative association between fertility rates and

population growth and a positive correlation between income levels and population growth. The study offers important new perspectives on population dynamics and lays the groundwork for future studies in the area of population economics [1].

The application of Python for exploratory data analysis (EDA) is the main topic of the research paper "Exploratory data analysis using Python" by Sahoo *et al.* [2]. The writers go over numerous Python techniques and modules that are used for statistical analysis, data cleansing, and visualisation. The paper provides a thorough review of the topic and is an invaluable resource for researchers and data analysts interested in using Python for EDA activities [2].

The significance of Python as a potent tool for data analysis is highlighted in the literature review for the research article "Python for data analysis: Data wrangling with Pandas, NumPy, and IPython" by McKinney, published by O'Reilly Media, Inc. in 2012. The study focuses on the use of well-known Python libraries for data wrangling activities, such as data cleaning, transformation, and manipulation, such as Pandas, NumPy, and IPython. For data analysts and researchers looking to take advantage of Python's capabilities for effective data analysis and exploration, it offers thorough examples and useful strategies for managing real-world datasets using Python [3]. The significance of Python as a flexible programming language for data analysis tasks is highlighted by the books "Data wrangling with Pandas, NumPy, and IPython" by McKinney [3] and "Python data analysis" by Idris [4], I. Both studies emphasise the use of well-liked Python modules like Pandas and NumPy for effective data exploration, manipulation, and visualisation. The articles offer concrete illustrations and methods for Python-based data cleansing, aggregation, statistical analysis, and other data processing activities [4].

The phenomenological theory of world population expansion is discussed in the research article titled "The phenomenological theory of world population growth" by Kapitza, S. P. In order to explain the dynamics of population expansion using actual evidence, the paper proposes a theoretical framework. Kapitza suggests a mathematical model that takes into account a number of variables impacting population increase, including migration, the birth and death rates, and technological improvements. In order to shed light on the long-term trends and patterns of global population expansion, the paper analyses historical population data. It contributes to the subject of demography and population studies and offers a distinctive viewpoint on population dynamics [5].

IMPLEMENTATION

- *Handling Categorical Variables:* Since categorical variables cannot be used directly in mathematical models, they need specific handling as shown in Figures 1 and 2.
- *Splitting the Dataset:* Using the train test split function of scikit-learn, the dataset was divided into training and testing sets in order to assess the performance of machine learning models and prevent overfitting.

Data Collection

To conduct our analysis on population distribution, we used three datasets. The world population dataset, the population of India, and the population of cities in China were among these datasets.

1. Population statistics from 1990 through 2021 are available for a number of nations and regions in the dataset taken from the World Bank website. The population totals for each nation or region are shown in the year columns (1990–2021).
2. Population statistics of year 2021 are available for a number of states and cities of India in the dataset taken from the Kaggle website.
3. China Population dataset has been gathered via Kaggle for a number of Chinese cities, including Shanghai, Beijing, Tianjin, Guangzhou, and Shenzhen, among others [6].

Data Preprocessing

It entails converting unstructured raw data into a clean, organized state suitable for additional analysis or modelling.

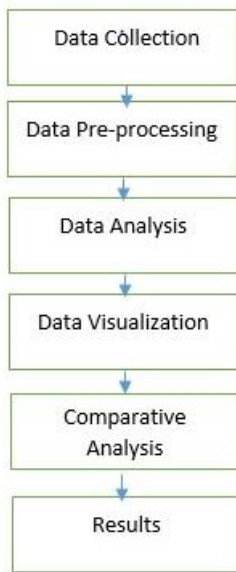


Figure 1. Flow chart of the proposed model.

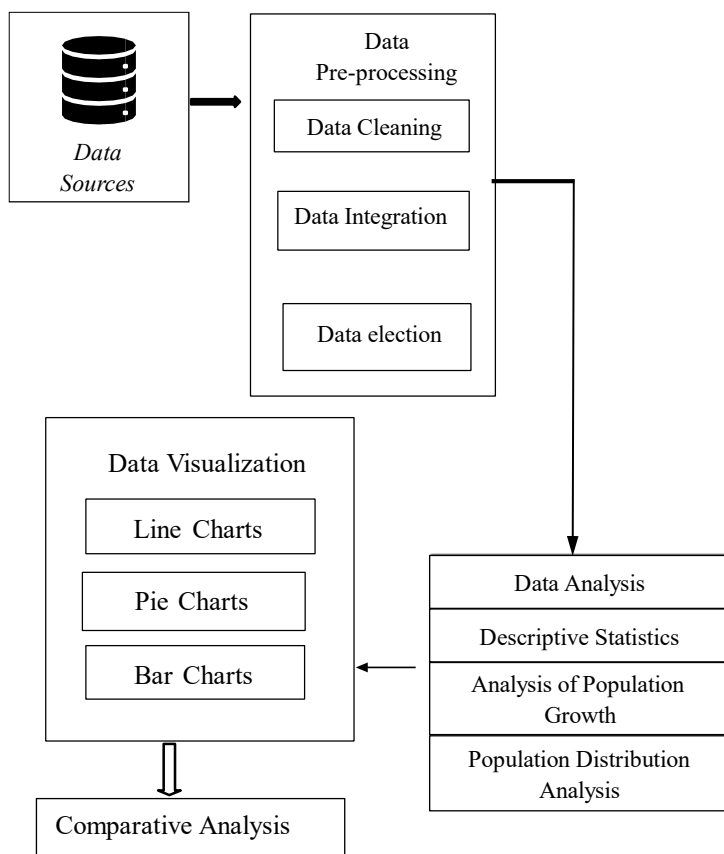


Figure 2. Framework of dataset visualization.

Handling Missing Values: We used the `isnull().sum()` function to look for missing numbers in order to solve this problem.

Data Analysis

To better understand demographic trends, the data analysis phase involved examining and interpreting the dataset.

1. *Descriptive Statistics*: This gave a broad summary of the population information [7].
2. *Analysis of Population Growth*: The population statistics were collected over a number of years. We could identify nations with notable population increases or declines by examining the population growth over time.
3. *Population Distribution Analysis*: Examining the population distribution among various nations gave us information on the world's population trends.
4. *Gender Analysis*: Pie charts were used to show the distribution of men and women, allowing us to see any imbalances or differences in gender ratio.
5. *Literacy Rate Analysis*: Male and female literacy rates were compared in the analysis. Examining the population's gender differences in literacy was the goal [8].

Data Visualization

The presentation and analysis of population-related data involved the use of data visualization tools.

1. Line Chart,
2. Pie Chart, and
3. Bar Chart.

Python Libraries

Due to its extensive ecosystem of libraries and tools, Python is an effective programming language that is frequently used for data analysis and visualization.

1. NumPy,
2. Pandas,
3. SciPy,
4. Matplotlib,
5. Seaborn, and
6. Tkinter.

Comparative Analysis

The comparison of population dynamics among the chosen countries is made possible by the side-by-side visualisation of population increase. By comparing and contrasting growth rates, trends, and population sizes, we can make inferences about the population dynamics of various parts of the world [9].

1. *Relative Growth Rates*: The comparison enables us to track the relative population growth rates over time for each nation.
2. *Trend Analysis*: Examining the population growth curves' shapes allows us to spot trends in the population growth of each nation.
3. *Population Size Comparison*: Comparing the population growth plots allows us to gauge the relative size of the populations in each nation.
4. *Long-Term trends*: We can detect long-term population growth trends by looking at the plots over an extended period of time [10].

RESULT

Following are the findings of the analysis of population growth in several nations:

1. *Population Growth over Time*: The population growth over time for the chosen nation. Line Chart provides a visual picture of the growth pattern by showing the shift in population size from 1990 to 2020 in Figure 3.
2. *Population Distribution*: The population distribution among several nations, India is shown in the pie chart in Figure 4.
3. *Population Distribution Comparison*: A comparison of the population distribution in the chosen nations. They used graphically is given by the bar plot. Vertical bars are used to graphically depict each country's population size, allowing for a fast comparison of population sizes.
4. *Most Populated Countries*: It has a feature that allows the names of the most populous nations to be displayed.

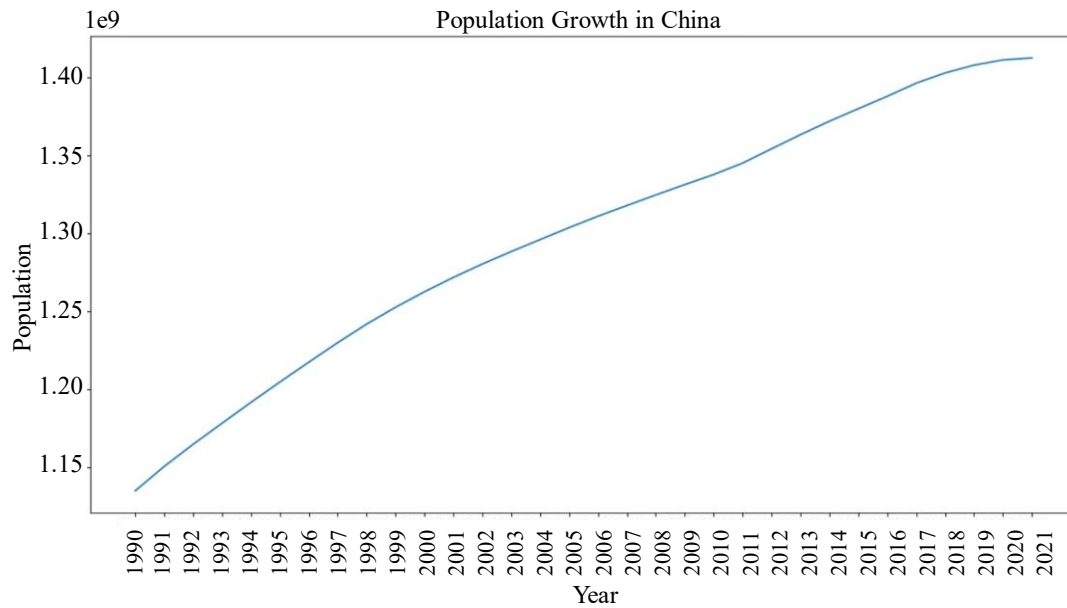


Figure 3. Population growth in China.

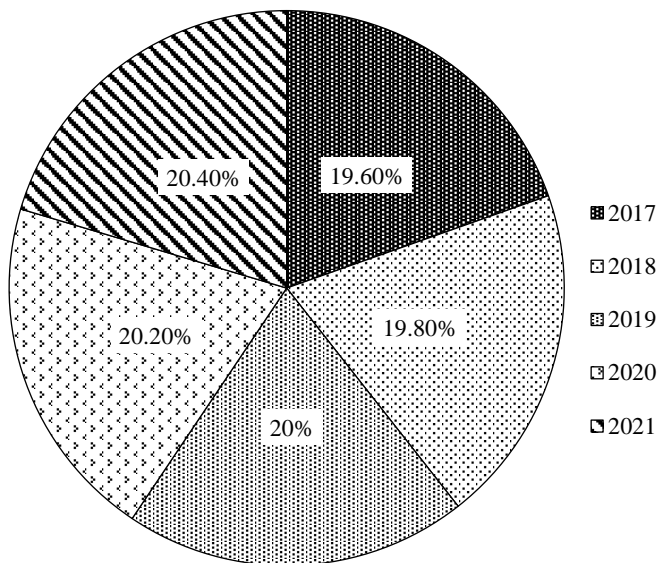


Figure 4. Population distribution in India.

- a. *Five most Populated cities:* There is a feature that allows you to graph the five most populous cities in the most populous nations.
5. *India's Population Distributions:* The population of India can be visualised according to several distributions. These distributions shed light on a number of population dynamics-related topics in India.
 - Gender Distribution.
 - Literacy Distribution.
 - Graduates Distribution.

CONCLUSION AND FUTURE WORK

Predictive modelling methods could be added to the project to help with policy planning and forecasting population growth in the future. Lastly, including real-time data sources and making sure updates are regular would improve the accuracy and applicability of the project's conclusions.

REFERENCES

1. Adelman Irwin. An Econometric Analysis of Population Growth. *Am Econ Rev.* 1963 Jun; 53(3): 314–339.
2. Sahoo K, Samal AK, Pramanik J, Pani SK. Exploratory Data Analysis Using Python. *International Journal of Innovative Technology and Exploring Engineering (IJITEE).* 2019 Oct; 8(12): 4727–4735.
3. McKinney Wes. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media, Inc. in 2012. <https://wesmckinney.com/book/>
4. Idris Ivan. Python Data Analysis. Packt Publishing Ltd. United Kingdom; 2014 Oct 28.
5. Kapitza SP. The phenomenological theory of world population growth. *Physics-Uspekhi.* 1996;39:57–71. doi: 10.1070/PU1996v039n01ABEH000127.
6. Jackson L. Scripting Solutions: Python at Your Service. [Online]. Hire Python Developer. 2023. Available from: <https://hirepythondeveloper.com/python-scripting-solutions/>
7. American Institute of CPAs. (2024). Justin Beals. AICPA conferences. [Online]. Available from: <https://aicpaconferences.com/b/sp/justin-beals-58305>
8. Arif Z. (2023 Aug 23). Building the Future with Supervised Learning: Unleashing the Potential of Labeled Data. [Online]. Medium. Available from: <https://medium.com/@mr-zeeshan-arif/building-the-future-with-supervised-learning-unleashing-the-potential-of-labeled-data-3f10624ca48f>
9. Bologna M, Aquino G. Deforestation and world population sustainability: a quantitative analysis. *Sci Rep.* 2020 May 6; 10(1): 7631.
10. Gerland P, Raftery AE, Ševčíková H, Li N, Gu D, Spoorenberg T, Alkema L, Fosdick BK, Chunn J, Lalic N, Bay G. World population stabilization unlikely this century. *Science.* 2014 Oct 10; 346(6206): 234–7.