VIETNAM NATIONAL UNIVERSITY, HANOI

INTERNATIONAL SCHOOL

******□******

REPORT OF FINAL EXAM PROJECT

Subject: Enterprise Analytics for Decision Support

Class: INS3063.02 - Group: 06

Topic: World Population

Member: Vũ Thị Quế Anh – 20070900

Nguyễn Thu Huyền - 200709

Nguyễn Thị Yến Nhi – 20070968

Lecturer: TS.Tran Thi Oanh

HANOI 2023

Table of Contents

I. DISCOVERY

1. What is it?

The world's population faces many challenges and problems. These include population growth, population aging, inequality, health services and education, and climate change effects. For a better understanding of the world's population and to capture changing trends, we performed data visualization research and developed models that are used to predict future population trends.

2. How important is it to organizations?

Analyzing the world population situation is an extremely important process for leading organizations in the world. It is the procedure of researching and analyzing data regarding a population, including its size, age composition, geographic distribution, and other aspects pertaining to the world population. The analysis of the global population enables us to identify the patterns of sex change, age structure, and geographic distribution, as well as the trajectory of population growth. Resource management, social and economic policy, and development planning are all based on this data. For example, if a region is predicted to experience significant population growth, a company may invest in expanding or developing the market in that area. Additionally, if predictions indicate a region or country facing population decline, companies need to plan for addressing potential labor shortages in the future. Analyzing the world population can help with resource and infrastructure management. Planning for the requirements of productive capacity, land, water, food, and infrastructure is essential to ensure sustainable development for nations with rapidly expanding populations. We can better understand and respond to global concerns like climate change, migration, disease outbreaks, and anti-infection rates by analyzing the world's population. It will be easier to develop more effective policies and remedies to lessen the effects of these problems if demographic data, particularly on vulnerable populations, is collected.

Furthermore, the demographic analysis process can influence the investment decisions of business organizations. If a region or country has an anticipated population growth and a favorable business environment, companies may decide to invest in expansion or to establish new production facilities. Conversely, if projections suggest that an area is experiencing population decline and limited growth potential, businesses may consider divesting or adjusting their business strategies.

3. What are potential benefits?

By predicting future population growth, governments, organizations, and policymakers can better plan and allocate resources in areas such as healthcare, education, infrastructure, and food production. It allows for more effective long-term planning and helps ensure that adequate resources are available to meet the needs of the growing population. Moreover, predicting world population offers valuable insights for decision-making, planning, and

addressing various social, economic, and environmental challenges. It helps stakeholders make informed choices to promote sustainable development, improve quality of life, and ensure the well-being of current and future generations.

4. Descriptive method

In this project, we use Python and Tableau as 2 main methods to solve the problem.
We first use Tableau to visualize the data in order to acquire a general understanding of it. To better understand the global population, interactive graphs and charts in the form of dashboards and sheets are helpful. Tableau offers a simple approach to explore and comprehend trends, outliers, and patterns in data by utilizing visual elements like charts, graphs, and maps.
Then, we analyze the performance of models like Linear Regression, Decision Tree, and Random Forest using Python to determine which model is more accurate at forecasting future global population changes.

II. DATA PREPARATION

We gather data for this project on Kaggle; this dataset was produced from the World Population Review, making it a trustworthy source.

About dataset:

1. Title: World Population
2. Sources: World Population Dataset | Kaggle
3. Relevant Information: This Dataset is created from World Population Review
4. Number of Instances: 234
5. Number of Attributes: 17
6. Attribute information:
   Rank: Rank by Population
   CCA3: 3 Digit country/Territories Code
   Country/Territories: Name of the Country/Territories
   Capital: Name of the Capital
   Continent: Name of the Continent
   2022 population: Population of the Country/Territories in the year 2022.
   2020 Population: Population of the Country/Territories in the year 2020.
   2015 Population: Population of the Country/Territories in the year 2015.
   2010 Population: Population of the Country/Territories in the year 2010.
   2000 Population: Population of the Country/Territories in the year 2000.
   1990 Population: Population of the Country/Territories in the year 1990.
   1980 Population: Population of the Country/Territories in the year 1980.
   1970 Population: Population of the Country/Territories in the year 1970.
   Area (km²): Area size of the Country/Territories in square kilometers.
   Density (per km²): Population Density per square kilometer.
   Growth Rate: Population Growth Rate by Country/Territories.

World Population Percentage: The population percentage by each Country/Territories.

III. MODEL PLANNING

Data pre-processing is a technique for gathering unprocessed data and turning it into information that is relevant and valuable.

a) Tableau

Using SQL to process data: By creating a new data file, use union all to combine data from the population tables of each year into one column.

b) Python

The duplicate or missing data was then sorted, screened for, and processed, and the data was then transformed into a format suited for analysis and straightforward result prediction.

IV. MODEL BUILDING

1. Tableau

In this section, we have raised some issues around world population and related issues to help analysts better understand dataset. By using the Tableau tool, we will visualize the data for a better view.

- World population trends: 1970 to 2025

Currently, the world's population is three times greater than it was in the mid-20th century. It is estimated that the global population will reach 8.0 billion by 2022, which is an increase of 1 billion since 2010 and 3 billion since 1995.

During the 56 years from 1970 to 2025, the world's population grew fastest in the period 1970-2010, when it increased by an average of 2.1% per year. As a result of declining fertility, population growth has slowed by more than half since then. In 2020, for the first time since 1950, the population growth rate will fall below 1% per year and is forecast to continue to slow down over the next few decades. This is forecast to continue until the end of the century, the global population could grow to 8.3 billion by 2025.

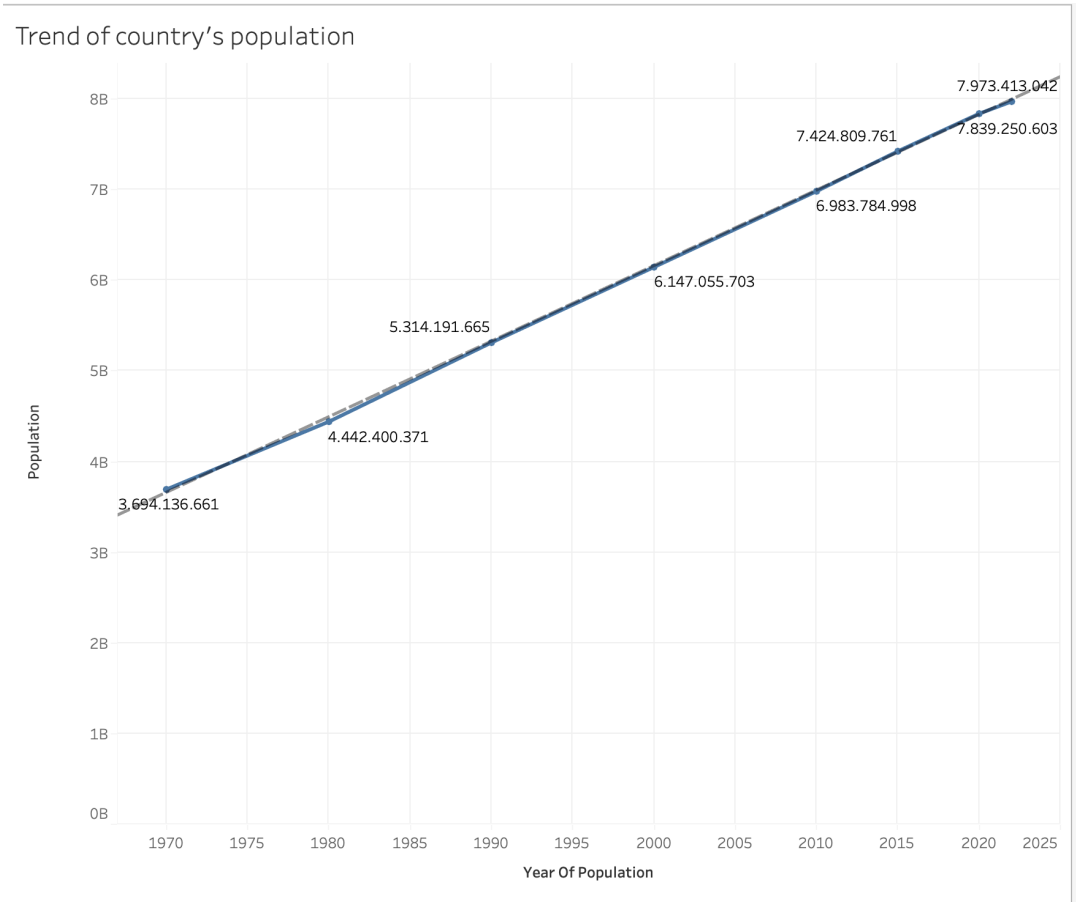


Figure 1: Trend of country's population

The malaria epidemic that hit Europe in the fourteenth century killed one-third to two-thirds of the continent's population. The Covid-19 pandemic did not cause such a serious death rate. However, the demographic impact of Covid-19 could be significantly larger as the nearly 3 million Covid-19 deaths have been accompanied by a worldwide decline in births. As a result, the pandemic may have pushed back the projected global population peak by a decade to the 2050s. Population decline is a welcome issue given the many challenges the world faces about the environment. However, according to the British magazine The Economist, a smaller population also means fewer new ideas. This leads to a future that could be very different from what optimists imagined.

Additionally, more and more young couples are delaying or deciding not to have children amid the economic downturn and high house prices. In 2022, the average age of pregnant women in Korea is 33.5 years old, an increase of 0.2 years compared to 2021. In addition, a part of young people has a concept of prioritizing individual freedom, so they tend to pursue "celibacy", avoiding marriage and having children. According to a survey result, up to 53.2% of respondents who are women aged 20-34 in Korea think that getting married and having children is not important to women. The low birth rate and declining population will inevitably have consequences for the economy, including the risk of a labor shortage, a

decrease in the number of taxpayers and the pressure on the pension system. The increasing cost of social security will put great pressure on the shoulders of the younger generation.

It may be claimed that the global population is on the decline and will likely continue to do so for many years to come. Each nation and territory's economy, culture, society, ecology, etc. may be impacted by the population decline trend.

- ❖ Top 10 most populous countries in 2022

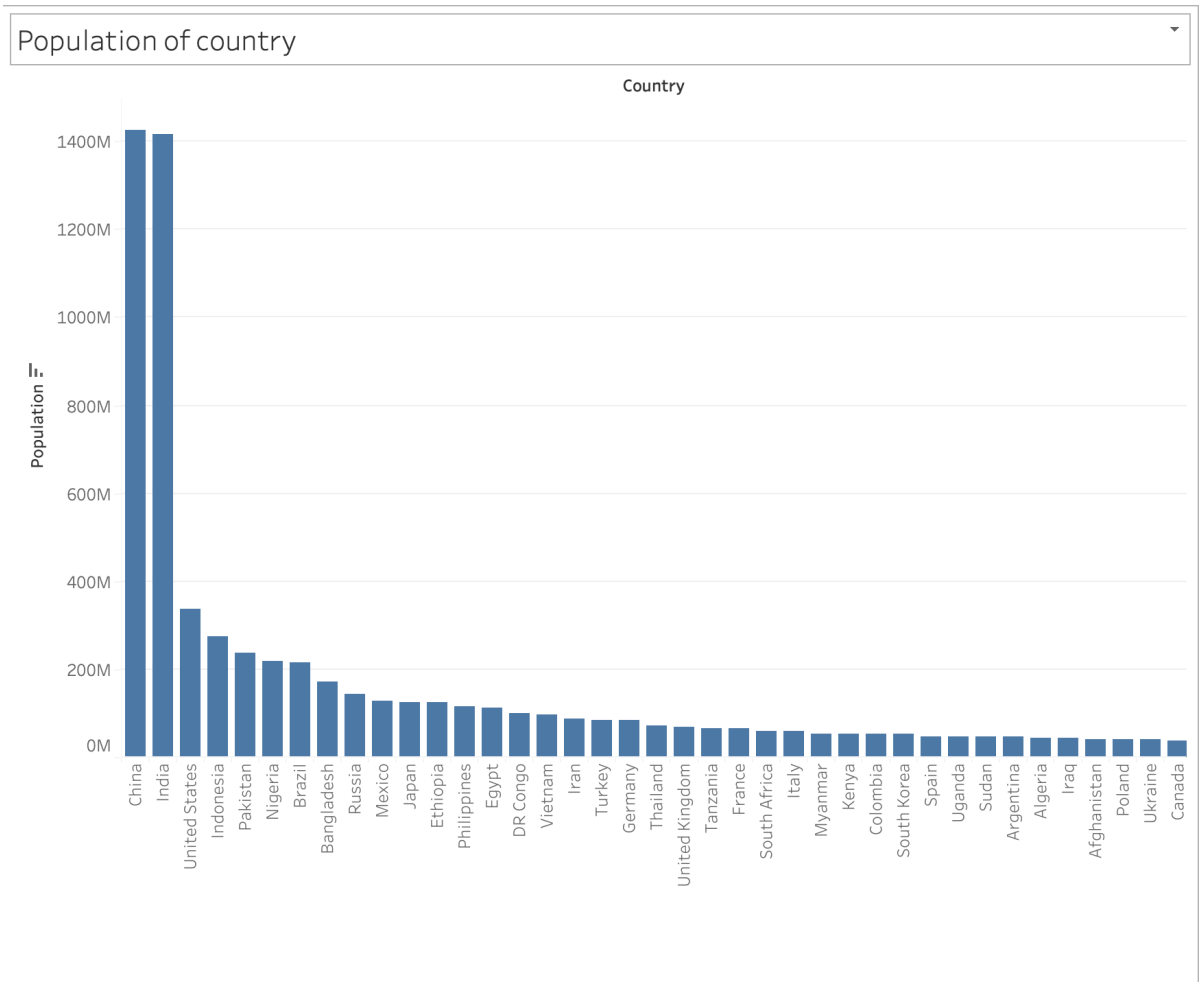


Figure 2: Top 10 most populous countries in 2022

From the graph, we can see that China is the most populous country in the world with 1.4 billion people followed by India, United States and others.

So, why are China and India the two most populous countries in the world?

Two of the world's most populous nations, China and India, are distinguished by distinctive cultural, historical, and educational elements that support their vast populations. These countries have ancient civilizations and have had a considerable increase in their population

through time. Their large population can be linked to historical conditions, educational institutions, and cultural customs.

To understand why China and India have such big populations depends heavily on cultural considerations. Family and community are deeply ingrained cultural values in both nations. Great fertility rates are a result of traditional traditions that frequently place a great priority on having children and carrying down the family name. In addition, early family formation and higher birth rates are influenced by cultural norms around marriage, where early and planned weddings are common.

Population growth is significantly influenced by education as well. Over the years, China and India have both made significant efforts to increase access to education. Despite obstacles, literacy rates and educational possibilities for both sexes have steadily increased in India. Similar to this, China has made significant investments in its educational system to provide universal access to education. Education gives people information and possibilities, but it may also put off getting married and having kids. As a result, the relationship between education and population increase is complicated and impacted by a number of variables, including socioeconomic status, cultural norms, and governmental regulations.

The dynamics of China's and India's populations have been influenced by historical events as well. Periods of political stability, economic expansion, and technical innovation in these nations have all fueled population rise. For instance, both countries had substantial advancements in public health, sanitation, and healthcare over the 20th century, which decreased death rates. Population growth has been influenced by the decrease in newborn and child mortality as well as rising life expectancy throughout time.

It is important to highlight that China and India have both acknowledged the problems brought on by fast population expansion and have taken action to overcome them. In an effort to slow population growth, China notably enforced the one-child policy from 1979 to 2015. The demographics of China were significantly impacted by this program, which caused the fertility rate to drop. Similar to other countries, India has launched a number of efforts to support family planning, reproductive health care, and population control awareness.

So, a mix of historical conditions, educational institutions, and cultural traditions led to the massive populations of China and India. High birth rates have been a problem in both countries because of cultural aspects that place a strong emphasis on family and community, as well as bettering educational possibilities. Additionally, historical factors including advancements in healthcare and economic growth have boosted life expectancy and lowered death rates. It's crucial to remember that both nations have acknowledged the problems caused by high population expansion and have taken action to solve them.

- ❖ Top 10 countries with the largest area

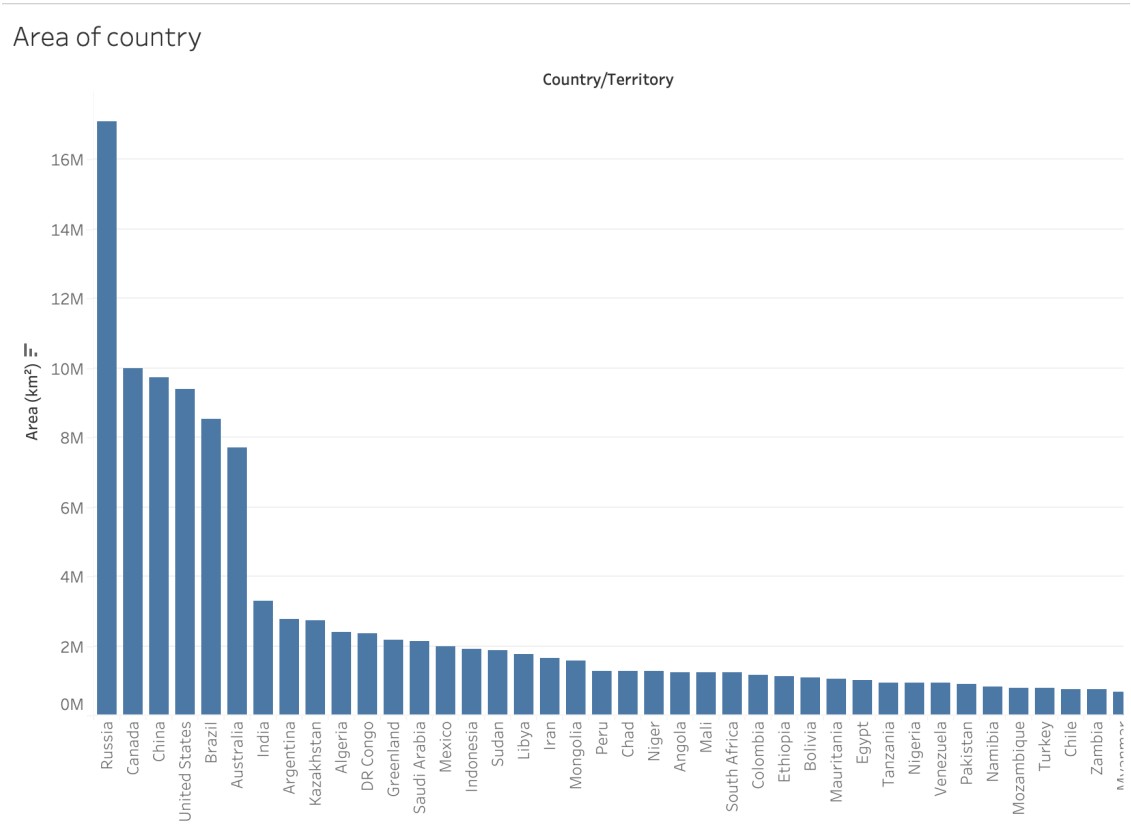


Figure 3: Top 10 countries with the largest area

Russia is the country with the largest area, followed by Canada, China and others. We can see that although Russia and Canada are the two largest countries by area, they are not the most populous. This raises a question: With such a large area, why don't people choose Russia or Canada to live? There are many factors that people consider when choosing where to live, and area is not the only determining factor. The vast expanses of Russia and Canada mean a wide variety of climates. Russia has a cold climate and the cold lasts for a large part of its territory. Meanwhile, Canada has a cold and harsh climate in many areas, especially the North. This extreme climate may reduce the attractiveness of the two countries to some.

The vast areas of Russia and Canada are unevenly dispersed. From the chart below, we can see that Russia and Canada have much lower population densities than China and India in 2022. In a square kilometer, more than 400 Indians live, while in Russia only about 9 people live and in Canada only 4 people live. Therefore, people will prioritize places with dense populations to develop the economy, education and people.

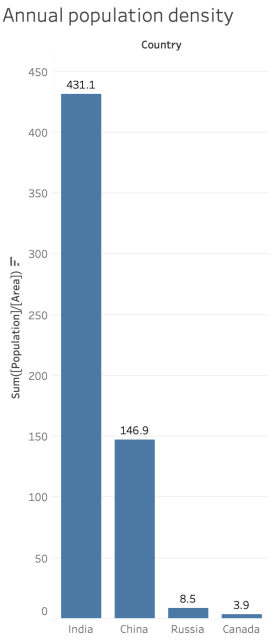


Figure 4: Density of India, China, Russia, Canada

- Population growth rate from 1970 to 2010

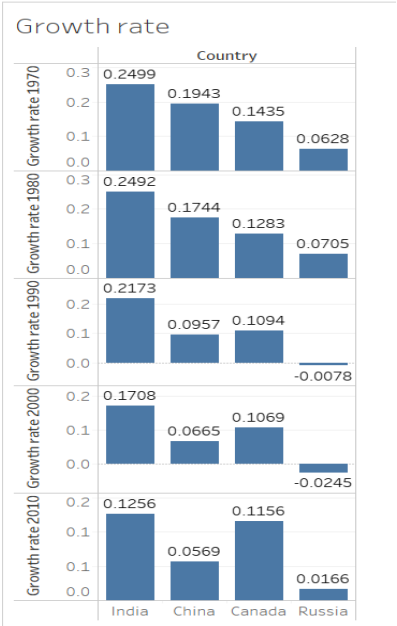


Figure 5: Population growth rate from 1970 to 2010

Because India, China, Russia, and Canada have the highest populations and largest land areas in the globe, we examine their 10-year population growth rates. As we can see, India has had the highest population growth worldwide throughout the years, but its growth rate is rapidly slowing down. Next, China currently has the highest population, but it is growing at a slower rate than India, and during the past ten years, that rate has been dramatically declining. In

contrast to the previous two nations, Russia and Canada have growth rates that are rather variable. Russia's growth rate fell between 1970 and 2000, but it rose (0.0166) from 2010 to 2011. The growth rate in Canada fell precipitously in 1990, even turning negative, indicating a general decline in population over the preceding ten years from 1980 to 1990. But by 2000, the population of Canada had increased once more and was maintaining a positive growth rate. The study above shows that China and India have population reduction strategies, presumably as a result of dropping birth rates, rising death rates, and effective family planning to concentrate on development. growth of societal issues including the economics, education, and so forth... With regulations encouraging individuals to have children and lowering the rate of migration to other nations as a result of the severe climate and a lack of infrastructure, Russia and Canada are seeing population growth.

2. Python

a. Data analysis and Visualization

About the overall population change, we see that the world population from 1970 to 2022 in general has an increasing trend. However, whereas the global population increased quickly between 1970 and 2015, it is expected to grow more slowly between 2020 and 2022. This is because the world has been affected by the Covid-19 pandemic for the past two years, which has killed billions of people and is partially the result of countries implementing population planning.

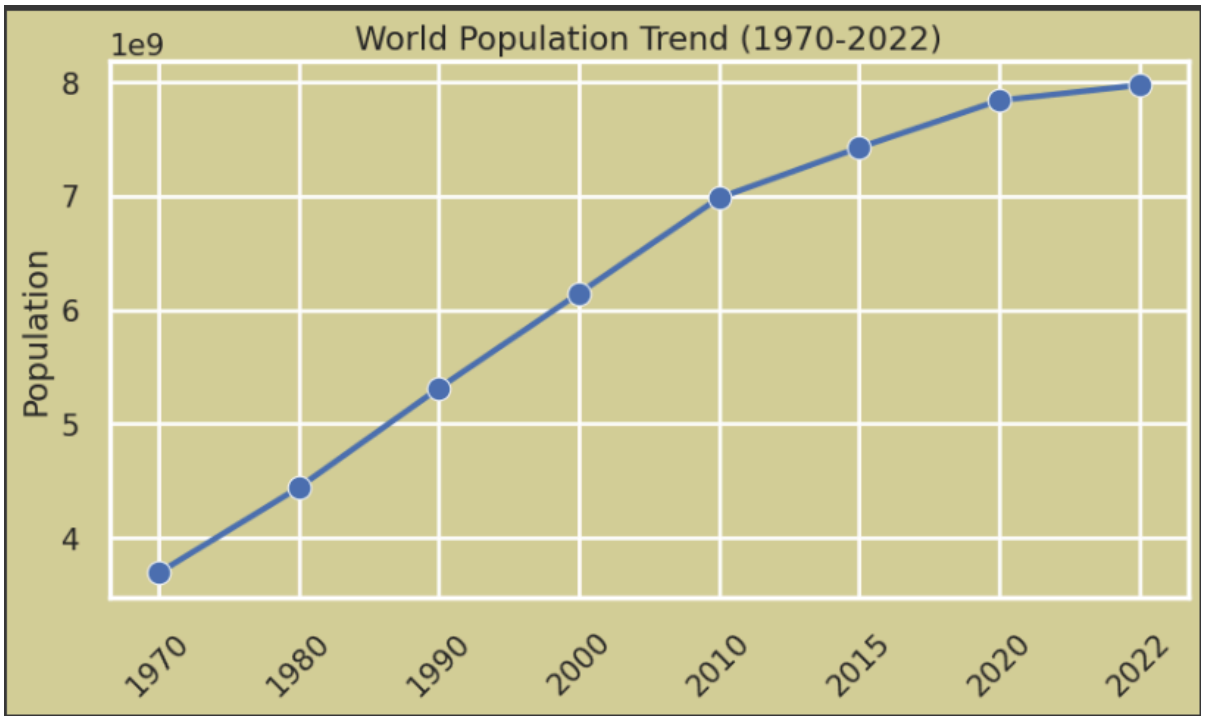


Figure 6: World population Trend (1970 - 2022)

Regarding the distribution of the world's population by continent, the graph shows that nearly all Asian nations have large populations and high population density per square kilometer.

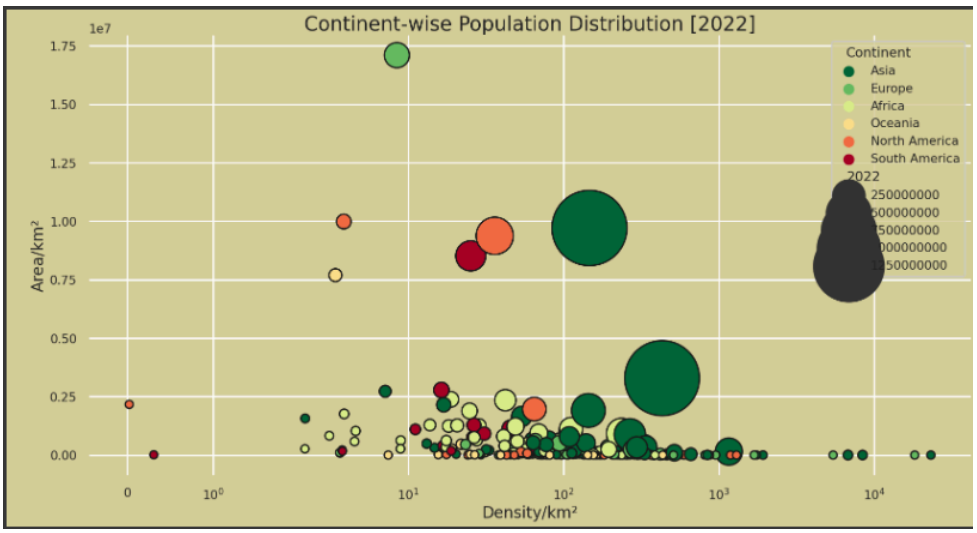


Figure 7: Continent-wise Population Distribution (2022)

Asia has first and foremost a lengthy history of colonization and civilization. Over the centuries and for thousands of years, those parts of Asia that were colonized early have thrived. Large cities and crowded urban regions have resulted as a result of this. After that, the continent's abundance of natural resources drew settlements and boosted population. Due to increased birth rates and declining mortality rates, Asia is likewise seeing a sizable increase in its population. High population densities place strain on some nations' infrastructure, like India, Pakistan, and Indonesia, and on vital services like healthcare and education. However, the large population and high population density create a large and diverse consumer market. This creates favorable conditions for economic development, investment and business expansion.

Here is a graph showing the correlation between the variables:

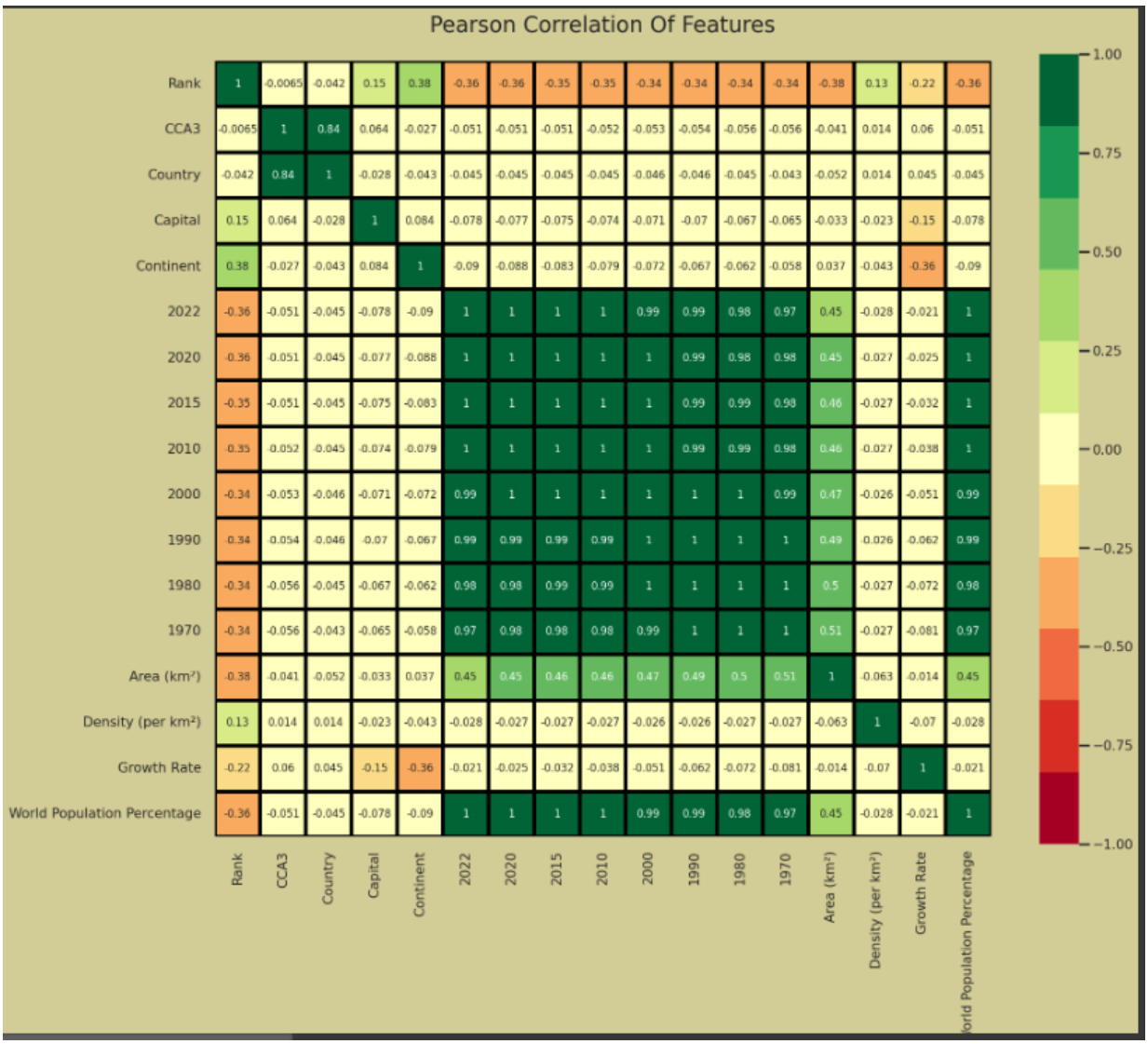


Figure 8: Pearson Correlation of Features

High correlation between population and world population percentage. Also, high correlation between country and CCA3 code as expected. Medium correlation between area and population, area and world population percentage. Medium inverse correlation between growth rate and continent, rank and world population percentage, rank and area, rank and population.

b. Prediction

Given the nature of the dataset, we may assume that the data are time series and that regression models have been fitted to train and test them. To estimate the population of the various nations in 2022, we employ three models: linear regression, decision trees, and random forests. As a result of statistical correlation between variables, we find that the population in 2022 has highly correlated with World Population Percentage, Growth Rate, Density, Area, and the population of the years 1970, 1980, 1990, 2000, 2010, 2015, 2020.

Therefore, we selected those variables as the input for the data model. The population in 2022 is the main output. We choose ***random_state*** to be equal to 1 and divide the train and test datasets by an 80:20 ratio.

There are three models we built so that we would be able to compare related evaluation metrics like R-square score, MAPE score, and RMSPE score.
R-square measures the explanatory power of the model for the dependent variable. It is also known as the Coefficient of Determination. It shows the percentage variation of the dependent variable explained by the model. The optimal model can be chosen by comparing the explainability of various models with R-square. R-square can be calculated using the equation:

$$1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum \left(y_i - \bar{y}_i \right)^2}$$

In which $\bar{y}_i$ is the mean value, and $\hat{y}_i$ is the predicted value. Or we can use import *r2_score* from *sklearn. metrics.*

Mean Absolute Percentage Error (MAPE) measures the mean error of the model when applied to predict the value of the dependent variable on a test set, however MAPE calculates the mean absolute value of the error divided by actual value, multiplied by 100. MAPE has the advantage that its unit of measure is a percentage, which makes it easier to understand the model's error. It also shows the average percentage error between the predicted value and the actual value. In this problem, we will import mean_absolute_percentage_error from sklearn.metrics.

Root Mean Square Percentage Error (RMSPE) also measures the mean error of the model when applied to predict the value of the dependent variable on the test set, however the RMSPE is calculated using the square root of the mean squared of the error divided by the actual value, multiplied by 100. RMSPE has the advantage that its unit of measurement is a percentage, which makes it easy to understand the error of the model. It also helps to evaluate the dispersion of errors, helping to detect observations that are noisy or have a large influence on the model.

Following each model being run, the R-square, MAPE, and RMSPE values are shown here:

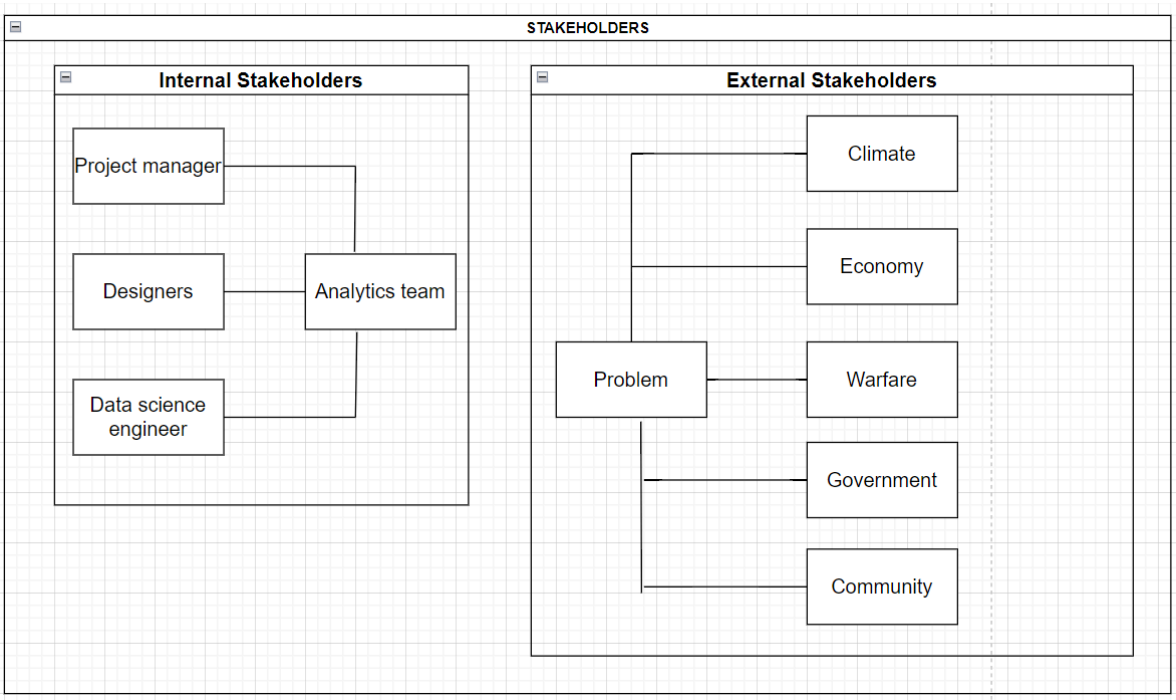
	R-square	MAPE	RMSPE
Linear Regression	0.99	0.60	255.05
Decision Tree	0.97	0.06	7.65
Random Forest	0.91	0.10	12.99

We can see that for both MAPE and RMSPE, the Decision Tree model yields the lowest result (0.06) and lowest result (7.65), respectively. The R-squared of the Decision Tree model also yields a value close to 1, which denotes that the predictor variable can accurately and error-free describe the responder variable. This demonstrates that the Decision Tree model outperforms the other two models as a predictor model.

There are other models that are currently used to predict the world population, however for this analysis, we only utilize Linear Regression, Decision Trees, and Random Forest. And among the three, we discovered that the Decision Tree model is the most effective. To assist organizations, small businesses, and other companies with initiatives including population forecasts, we highly recommend this model.

III. RESULTS COMMUNICATION

1. Identify stakeholders



Stakeholders are involved in data analysis, including data analysts, project managers, government officials, academics, NGOs, multinational corporations, businesses, and the general public.

To inform their judgments, research, policies, and strategic planning, these stakeholders use data from population analysis. Senior management and external stakeholders will receive the evaluated results from internal stakeholders, who are the ones that actually work with the data. The outcomes that internal stakeholders present will be used by external stakeholders to make the right judgments. Governments and regulators can comprehend and take advantage of population-related concerns and opportunities by using the vital data that population analysis gives about the population. Decisions about politics, society, the economy, and resource management frequently depend on data from demographic studies. Information from

demographic analysis is used by non-governmental and international organizations to design and carry out development initiatives and programs, including nonprofits and non-governmental organizations in the fields of health, education, development, and resource management. Additionally, population analysis offers crucial data on consumption markets and consumer trends that enterprises and business organizations may use to define markets, produce products and services that cater to the demands of the population, and build business strategy.

2. Key Findings

- Population Growth: Over the years, there has been a steady rise in global population. It is predicted that there will be 7.9 billion people on earth by 2022. Although it has been gradually slowing down, population growth is still increasing.

- Population Distribution: The distribution of the world population is uneven, with some regions experiencing rapid population growth while others have slower growth or even declining populations. The most populous countries include China, India, the United States, Indonesia, and Pakistan.

- Regional Variations: Population growth rates vary significantly across regions. While some countries continue to experience high population growth, others are experiencing population stagnation or decline. Factors such as fertility rates, access to healthcare, and migration patterns contribute to these variations.

- Impact on Resources and Environment: The increasing world population places pressure on natural resources and the environment. Issues such as food security, water scarcity, deforestation, and climate change are closely linked to population growth and consumption patterns.

- Family Planning and Education: Access to family planning services and education, particularly for women, has been recognized as crucial for managing population growth. Empowering individuals with knowledge and resources to make informed choices about family planning can contribute to sustainable population management.

- Urbanization: The world is becoming increasingly urbanized. More people are moving to cities in search of better economic opportunities, leading to the growth of megacities and metropolitan areas. Urbanization presents both opportunities and challenges for sustainable development.

- Infectious Diseases and Pandemics: Infectious diseases remain a significant concern globally. The emergence and re-emergence of infectious diseases, as well as the occurrence of pandemics like COVID-19, highlight the ongoing challenges in disease prevention, surveillance, and response.

3. Business Value

A precise population prediction can help governments and organizations can better plan and allocate resources such as food, water, energy, and infrastructure. It helps in determining the demand for resources and developing strategies to meet those demands sustainably.

Population forecasts aid businesses in understanding consumer behavior and preferences. By studying demographic trends, businesses can identify target demographics, develop effective marketing strategies, and customize products and services to meet specific consumer needs and preferences. In addition, by anticipating changes in the labor market, businesses can align their hiring strategies, talent acquisition efforts, and skill development programs to ensure a skilled and diverse workforce that meets future demands.

Although the world population is not completely reliable, the more often governments and organizations do it, the better they will get. Business and government can benefit from accurately predicting the population of the world through making knowledgeable decisions, concentrating on specific markets, scheduling staffing, optimizing supply chains, and adapting to changing consumer needs. As a result, businesses can remain competitive, identify growth opportunities, and develop strategies that are in line with market trends.

4. Narrative to summarize and convey findings to stakeholders

Even if it is rising more slowly, the global population is nevertheless increasing. The world's population may top 8.3 billion by 2025. The ranking order by size will alter according to the various growth rates of the major nations in the world. India is currently the country with the second-highest population density in the world, behind China. India is anticipated to surpass China as the world's most populated nation by 2023.

If population growth continues at the current rate, the economies of nations with dense populations will not be able to keep up. All people's attempts to bring about peace, comfort, and welfare for everybody will fail if the population is not managed within appropriate bounds, and suffering will become increasingly obvious. Government pressure to provide essential services like health care, education, infrastructure, and social security may increase as the population grows. The infrastructure system and associated policies must be adjusted by the government to accommodate the needs of the expanding population. The health system is under pressure to offer healthcare to everyone as a result of population expansion. The demand for health and wellness services may be on the rise, and healthcare providers and industry experts may need to adapt. Additionally, population increase may put stress on the environment and its resources. Increased resource demand, environmental deterioration, and habitat loss are all effects of population growth. This increases the need for efficient resource management and the search for sustainable alternatives. Additionally, local communities may experience social and cultural effects of population expansion. Local communities may need to adjust and create multicultural and diverse living environments due to changes in population structure and greater population variety.

In conclusion, various stakeholders, including the government, businesses, investors, the health sector, the environment, and local communities, can be impacted by population expansion. To promote sustainable and harmonious development for all stakeholders, it is crucial to recognize the effects of population expansion and take appropriate action.

IV. CONCLUSION AND RECOMMENDATION

In conclusion, the prediction of world population is crucial in a variety of sectors, including government, business, and social development. It provides valuable insights into resource planning, economic forecasting, healthcare management, environmental impact assessment, and policy formulation. In addition to market analysis, investment planning, workforce management, supply chain optimization, and sustainable business practices, businesses can benefit from population projections.

Recommendation:

- Stay informed about population projections and demographic trends to anticipate market shifts and consumer behavior changes.
- Conduct thorough market research to understand the specific needs, preferences, and behaviors of target demographics in different regions.
- Utilize population projections to plan your workforce needs and skills development programs.
- Anticipate future market demands, infrastructure needs, and resource requirements to ensure sustainable growth and competitiveness.
- Decision Tree models should be used by businesses, especially small and medium-sized ones, to forecast population.

V. REFERENCE

1. Dân Số Giảm - Lợi Hay Hại đối với nền Kinh tế? 07/04/2021. Available at: https://mof.gov.vn/webcenter/portal/ttpltc/pages_r/l/chi-tiet-tin-ttpltc?dDocName=MOFUCM199098
2. One-child policy Encyclopædia Britannica. Available at: https://www.britannica.com/topic/one-child-policy/Consequences-of-Chinas-one-child-policy
3. Raj, R. (no date) Evaluation metrics to check performance of Regression Models. Available at: https://www.enjoyalgorithms.com/blog/evaluation-metrics-regression-models
4. Tình trạng Dân Số Giảm: Bài Toán Hóc Búa (2023) Báo Nhân Dân điện tử. Available at: https://nhandan.vn/tinh-trang-dan-so-giam-bai-toan-hoc-bua-post741081.html
5. World population prospects 2019: Highlights. 10 key findings (no date) MEDBOX. Available at: https://www.medbox.org/document/world-population-prospects-2019-highlights-10-key-findings#GO

6. World population prospects 2022: Summary of results | population division (no date) United Nations. Available at: https://www.un.org/development/desa/pd/content/World-Population-Prospects-2022