

# World Demographic Analysis 2021

<sup>1</sup>Shankara Parameshwari Navya Sri, <sup>2</sup>Sukruthi Yakkala,

<sup>1</sup>sdameru1@umbc.edu , <sup>2</sup>syakkal1@umbc.edu

<sup>1,2</sup> Master in Professional Studies, Data Science

<sup>1,2</sup> University of Maryland, Baltimore County

## Abstract

As The world is evolving we can see drastic changes in our day to day life. Data has become the most important aspect in everyone's life. The changes that can be due to natural signs. We can predict the future data based on what we have . This might help us to concern more on what social domain is effecting the change among the people. As the Data and Technology is advancing every day and the drastic changes are occurring in every sector these days and so does in the Demographic sector. This industry will bring about many revolutions since it has integrated technology that enables scientists, researchers, physicians, and other professionals to utilize computers to accomplish their objectives. The main objective behind is to predict the facts which have already been there and to test them via live data .This paper will discuss several technological developments that might result in what we refer to as an informed user

experience, allowing us to see the drastic changes and comparing it with the facts.

## Introduction

The world population continues to grow but the pace of the growth is slowing down. Earth has a human population of over 8 billion, with an overall population density of 50 people per km<sup>2</sup> (130 per sq. mile). Most of the population lives in Asia, with about 2.8 billion people living in the 2 countries that are China and India. Accordingly, we are finding the population country, Area, Density, Number of births in 2021, Number of deaths\_2021, suicide rate, Fertility rate, and life expectancy, from all different countries. This would help us to analyze how the world's entire structure has changed in the year 2021. And we would also get the whole details about the number of birth rates and deaths that took place in different countries and the maximum area occupied by the countries.

Planning for national development and putting the 2030 Agenda for Sustainable Development into practice requires an understanding of population patterns and an ability to predict demographic change. The 2030 Agenda places a strong emphasis on the fact that people are the key to sustainable development. Understanding the population of the world's various nations is crucial. The majority of people on earth live in cities and suburbs, and there has been a substantial movement into metropolitan areas. Knowing the population increase rate and the number of deaths that occurred in a given nation will be useful in this case. The Recent demographic developments will be helpful in overcoming future SDG implementation issues. For instance, nations with a high rate of population increase, the majority of which are in sub-Saharan Africa.

Countries with high birth rates should get ready to satisfy the requirements of a growing population of children and teenagers. Countries that are experiencing a demographic dividend due to a fall in childbearing ought to invest in their human capital by providing possibilities for

productive work, access to health care, and excellent education for all ages. Adapting public services to the increasing share of older people in aging populations must be a priority for nations with aging populations. Depending on the issue at hand, the timeline being studied, the technology available, and the demographic, social, and economic context, population increase may not be the primary cause of environmental harm, but it may nevertheless worsen the issue or hasten the date of its appearance.

### **Literature Survey**

As we think it is easy to predict or analyze data, but it is not, it is certainly not so easy to analyze big-data within it and there are too many challenges to predict something to the end user in the right way. The problem of big data analysis while web scraping into SQLite or SQL and converting columns to its data types. This required a great deal of time. There really are two ways to fix it: construct a table with the correct data types for each column, or utilize change datatype to meet our needs. The alternative method is to load the csv using pandas, update the

datatype using pandas, and then use `pd.read_sql` to use the tables for analysis.

explains how migration, changes in birth and death rates, and changes in population size may all affect these factors. Current low birth rates below population replacement levels and increased life expectancy are indicators of demographic change in the Western industrialized countries.

World's population grew by 0.87% from 2020 to 2021, reaching 7,909,295,151 people. A 0.98% growth from 2019 brought the population of the world to 7,840,952,880. In 2022, the population of the globe is expected to increase at a pace of around 0.84% annually, which is a decrease from the 1.05% in 2020, 1.08% in 2019, 1.10% in 2018, and 1.12% in 2017. An estimated 67 million individuals are added to the population each year as of right now.

### **Technology used:**

**Web scraping :** The practice of deploying bots to gather information and material from a website is known as web scraping. Web scraping collects the underlying HTML code and, with it, data kept in a database, in contrast to screen scraping, which just

scrapes pixels seen on screen. After that, the scraper can duplicate a whole website's content elsewhere. Here, we extract content and data from a Website using selenium in Python and Pandas.

**SQLite :** SQLite is an Relational Database Management System which is written in ANSI-C. The `sqlite3` module provides a straightforward interface for interacting with SQLite databases. A connection object is created using `sqlite3.connect()`; the connection must be closed at the end of the session with the `.close()` command. While the connection is open, any interactions with the database require you to make a cursor object with the `.cursor()` command. The cursor is then ready to perform all kinds of operations with `.execute()`. Here, Joining two different tables with a unique primary key and Implementing Problem statements.

**Pandas :** Pandas is a software library written for the Python programming language for data manipulation and analysis. We can also use pandas to create new tables within an SQLite database. Here, As we have web scrapped the data into CSV files then use them as SQLite database.

## Result:

Top 10 countries with the highest Number of Birth.

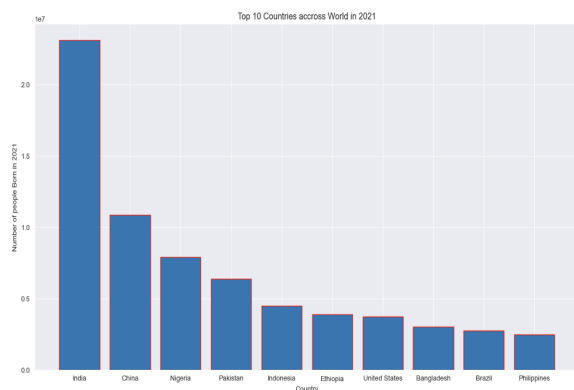
Today, the world's population is more than three times larger than it was in the mid-twentieth century. The global human population will reach 8.0 billion in mid-November 2022 from an estimated 2.5 billion people.

SQL query:

```
Problem_1= pd.read_sql_query('SELECT Country,Number_of_births_2021 from PROJECT order by Number_of_births_2021 DESC LIMIT 10;',conn)
```

]:

|   | Country       | Number_of_births_2021 |
|---|---------------|-----------------------|
| 0 | India         | 23114000              |
| 1 | China         | 10881567              |
| 2 | Nigeria       | 7923294               |
| 3 | Pakistan      | 6374741               |
| 4 | Indonesia     | 4496383               |
| 5 | Ethiopia      | 3895734               |
| 6 | United States | 3722822               |
| 7 | Bangladesh    | 3019672               |
| 8 | Brazil        | 2760958               |
| 9 | Philippines   | 2485008               |

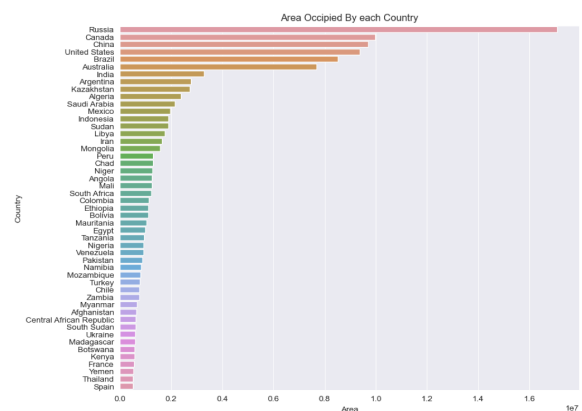


2. The countries with an area greater than 5,00,000 km.

SQL query:

```
Problem_2 = pd.read_sql_query('Select Country,Area from Project where area > 500000 order by Area DESC;',conn)
```

|    | Country       | Area     |
|----|---------------|----------|
| 0  | Russia        | 17098242 |
| 1  | Canada        | 9984670  |
| 2  | China         | 9706961  |
| 3  | United States | 9372610  |
| 4  | Brazil        | 8515767  |
| 5  | Australia     | 7692024  |
| 6  | India         | 3287590  |
| 7  | Argentina     | 2780400  |
| 8  | Kazakhstan    | 2724900  |
| 9  | Algeria       | 2381741  |
| 10 | Saudi Arabia  | 2149690  |
| 11 | Mexico        | 1964375  |
| 12 | Indonesia     | 1904569  |
| 13 | Sudan         | 1886068  |
| 14 | Libya         | 1759540  |
| 15 | Iran          | 1648195  |
| 16 | Mongolia      | 1564110  |
| 17 | Peru          | 1285216  |
| 18 | Chad          | 1284000  |
| 19 | Niger         | 1267000  |
| 20 | Angola        | 1246700  |
| 21 | Mali          | 1240192  |
| 22 | South Africa  | 1221037  |

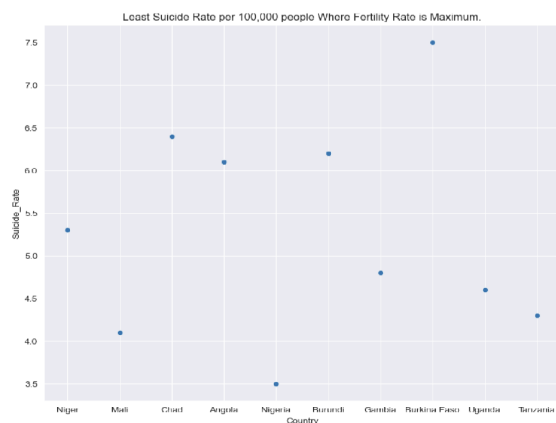


3. The 10 least suicide rates in countries with a max fertility rate Select TOP 10 .

SQL query:

```
Problem_3 = pd.read_sql_query('SELECT country,Suicide_Rate,Fertility_Rate from PROJECT order by Fertility_Rate DESC,Suicide_Rate ASC LIMIT 10;',conn)
```

|   | Country      | Suicide_Rate | Fertility_Rate |
|---|--------------|--------------|----------------|
| 0 | Niger        | 5.3          | 6.9            |
| 1 | Mali         | 4.1          | 5.9            |
| 2 | Chad         | 6.4          | 5.7            |
| 3 | Angola       | 6.1          | 5.5            |
| 4 | Nigeria      | 3.5          | 5.4            |
| 5 | Burundi      | 6.2          | 5.4            |
| 6 | Gambia       | 4.8          | 5.2            |
| 7 | Burkina Faso | 7.5          | 5.2            |
| 8 | Uganda       | 4.6          | 5.0            |
| 9 | Tanzania     | 4.3          | 4.9            |



4. The Number of births count and Country names with minimum Life expectancy.

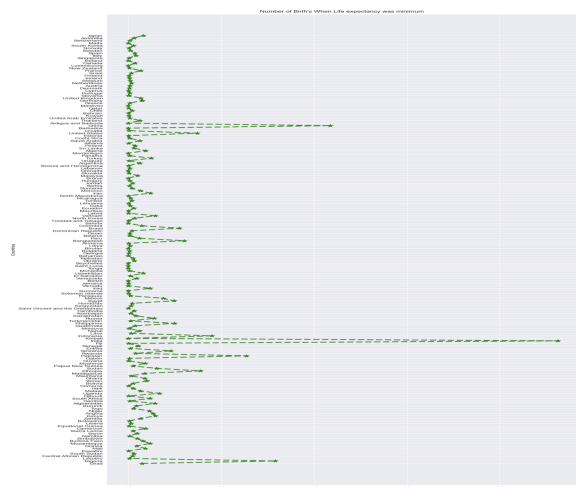
SQL query:

```
Problem_4 = pd.read_sql_query("Select Country,Number_of_births_2021,life_expec
```

tancy from PROJECT order by life\_expectancy;",conn)

|     | Country                  | Number_of_births_2021 | life_expectancy |
|-----|--------------------------|-----------------------|-----------------|
| 0   | Chad                     | 744834                | 52.5            |
| 1   | Nigeria                  | 7923294               | 52.7            |
| 2   | Lesotho                  | 59581                 | 53.1            |
| 3   | Central African Republic | 234821                | 53.9            |
| 4   | South Sudan              | 313347                | 55.0            |
| ... | ...                      | ...                   | ...             |
| 167 | South Korea              | 289006                | 83.7            |
| 168 | Malta                    | 4664                  | 83.8            |
| 169 | Switzerland              | 87086                 | 84.0            |
| 170 | Australia                | 298496                | 84.5            |
| 171 | Japan                    | 818509                | 84.8            |

172 rows x 3 columns

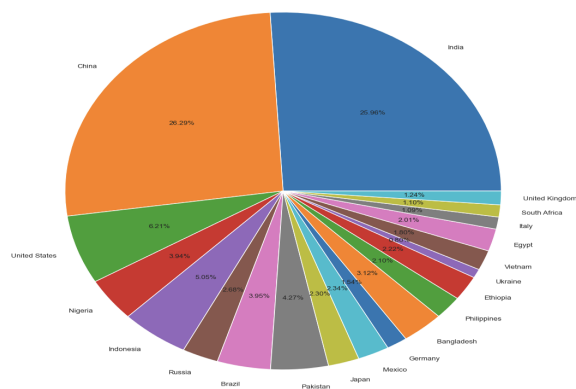


5. The Density and Population across countries where Max number of deaths took place.

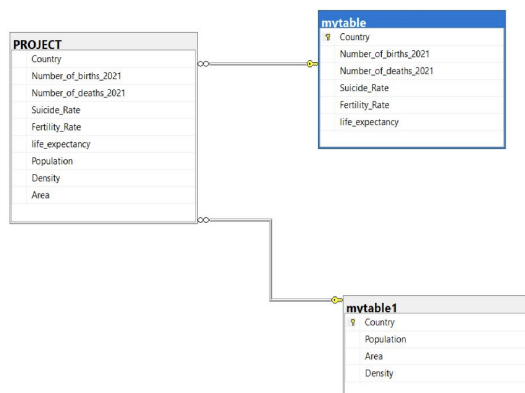
```
Problem_5 = pd.read_sql_query("Select country,Density,Population,Number_of_deaths_2021 from PROJECT order by Number_of_deaths_2021 DESC LIMIT 20;",conn")
```

|    | Country        | Density | Population | Number_of_deaths_2021 |
|----|----------------|---------|------------|-----------------------|
| 0  | India          | 477.0   | 1407563842 | 13299618              |
| 1  | China          | 151.0   | 1425893465 | 10617528              |
| 2  | United States  | 37.0    | 336997624  | 3280682               |
| 3  | Nigeria        | 240.0   | 213401323  | 2792704               |
| 4  | Indonesia      | 147.0   | 273753191  | 2755190               |
| 5  | Russia         | 9.0     | 145102755  | 2458539               |
| 6  | Brazil         | 26.0    | 214326223  | 1784239               |
| 7  | Pakistan       | 306.0   | 231402117  | 1660400               |
| 8  | Japan          | 340.0   | 124612530  | 1574533               |
| 9  | Mexico         | 66.0    | 126705138  | 1191718               |
| 10 | Germany        | 239.0   | 83408554   | 1039440               |
| 11 | Bangladesh     | 1315.0  | 169356251  | 962479                |
| 12 | Philippines    | 388.0   | 113880328  | 832888                |
| 13 | Ethiopia       | 109.0   | 120283026  | 815212                |
| 14 | Ukraine        | 69.0    | 43531422   | 731364                |
| 15 | Vietnam        | 313.0   | 97468029   | 713131                |
| 16 | Egypt          | 111.0   | 109262178  | 691381                |
| 17 | Italy          | 200.0   | 59240329   | 680373                |
| 18 | South Africa   | 49.0    | 59392255   | 678798                |
| 19 | United Kingdom | 279.0   | 67281039   | 652723                |

Top 20 Country Population When Number of Death's were high



## ER - Diagram



## Conclusion:

From this analysis We can jot down few predictions such as ,

1. India was ranking highest in terms of Number of Births.

2. Russia has the largest Area next comes Canada and so on.

3. Nigeria has the lowest suicide rate , but has maximum fertility rate.(The higher the fertility rate the lower the suicide rate.) This is a fact we can see in the graph.

4. ( If long life is valued and death control is established, then either the birth rate must come down to match the new death rate, or the population must rise.)The Number of births that took place when the life expectancy was minimum () has the highest number of births. That's the case happening in those rising countries.

5. (The Mortality Rate(Death's) Increases, the population growth decreases), where india and china could make up to withstand the population.

I learnt that SQL using real world datasets with all math and calculations are predicting the fact's with all those number's , The relations between the countries based on the demographic calculations.

## Reference :

- [1] Elhaik, E., Tatarinova, T., Chebotarev, D. *et al.* Geographic population structure analysis of worldwide human populations infers their biogeographical origins. *Nat Commun* 5, 3513 (2014). <https://doi.org/10.1038/ncomms4513>
- [2] Bruno Lunenfeld, An Aging World – demographics and challenges, Jul 2009 <https://doi.org/10.1080/09513590701718364>.
- [3] Rónán O’Caoimh, Prevalence of frailty in 62 countries across the world: a systematic review and meta-analysis of population-level studies, <https://doi.org/10.1093/ageing/afaa219>.
- [4] Madeleine Short Fabic, YoonJoung Choi & Sandra Bird. A systematic review of Demographic and Health Surveys: data availability and utilization for research [https://www.scielosp.org/article/ssm/content/raw/?resource\\_ssm\\_path=/media/assets/bwh/v90n8/a12v90n8.pdf](https://www.scielosp.org/article/ssm/content/raw/?resource_ssm_path=/media/assets/bwh/v90n8/a12v90n8.pdf).
- [5] P. P. Singh, M. Singh & Dr S. S. Mastana, APOE distribution in world populations with new data from India and the UK, Jul 2009, <https://doi.org/10.1080/03014460600594513>