

**TASK**

**Exploratory Data Analysis on the UN Socioeconomic Demographic Dataset**

[](http://www.hyperiondev.com/portal/)

**Introduction**

This dataset was taken from the UNESCO Institution for Statistics website. It contains over 40 thousand records for information on (but not limited to) birth rates per woman, infant mortality rates per 1000 live births, GDP, population growth, rural population as percentage of total, and so on. The date range of this dataset is 2016 – 2022, though in some instances, the latest available data is 2020.

The column “Indicator” shows the feature being measured, such as GDP or total government expenditure. The first column is a unique identifier for each entry, which can be useful when considering if the same record keeps appearing in analysis. The other columns are self-explanatory.

The analysis in this EDA looks at birth rates and population demographics.

**DATA CLEANING**

First, columns were removed which were not deemed useful. These included a redundant column of country codes, which was not needed as the full country name was also included in the column “Country”, as well as the column “TIME”, which repeated the data in “Time”.

**MISSING DATA**

Missing data was handled by removing the columns “Flag Codes” and “Flags” as they were entirely NaN values.

**DATA STORIES AND VISUALISATIONS**

The first data extracted was on Fertility rates, total (births per woman). For this information, the latest available data was 2020.

This table shows that Niger had the highest fertility rate by this metric for all years measured.  
Graphical user interface, application

Description automatically generated

In order to get information on more countries, the top 20 results were returned (below).

Graphical user interface

Description automatically generated with low confidence

For comparison, the same extraction was made for the lowest fertility rates.

Top 5:

**Graphical user interface, application

Description automatically generated**

Top 20:

**Graphical user interface, text

Description automatically generated with medium confidence**

The pie charts below illustrate how much each country is represented in the data for top 20 highest birth rates, and top 20 lowest birth rates.

In the highest birth rates, we can see all countries are in Africa.

**Chart, pie chart

Description automatically generated**

For the top 20 lowest birth rates, 4 of the 6 countries in the list are in East Asia, and the remaining 2 are in North America and Europe.

**Chart, pie chart

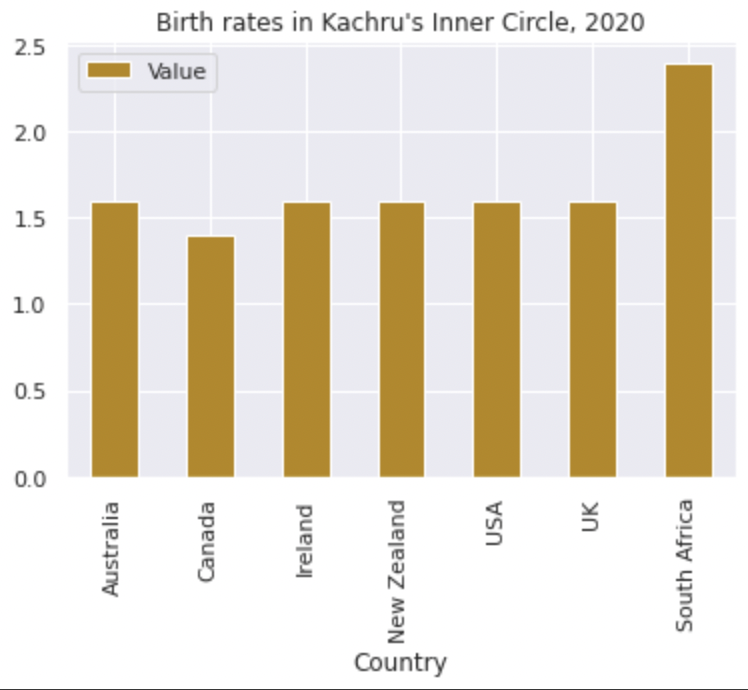
Description automatically generated**

The table below shows the birth rates in 2020 (the latest year available) for countries within Kachru’s Inner Circle of World Englishes. This was chosen as a subset demographic as these countries share a common language and culture but span multiple continents, and none of them featured in either of the above results.

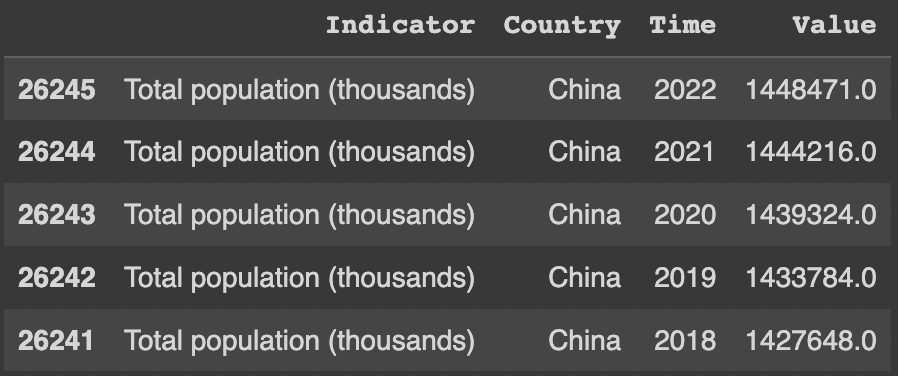
Graphical user interface, application

Description automatically generated

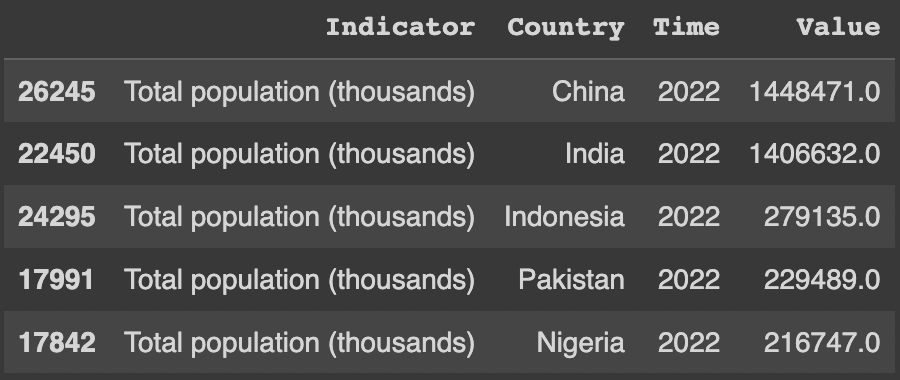
The bar chart below visualises the above data, showing South Africa with the highest birth rate amongst this demographic, which is consistent with African nations having higher birth rates overall. What is interesting to note in this graph is that there is no difference amongst most of these countries, except a slightly lower rate in Canada and a significantly higher rate in South Africa; the rest are the exact same.

****

Next other population data was extracted. Such as the highest population in the dataset. This produced the table below, which shows that for all the past 5 years, China has had the highest population.



For the latest year available for this data (2022), the top 5 highest populations are found in China, India, Indonesia, Pakistan, and Nigeria. Simply by looking at the number in the Value column, we can see that China and India have an exponentially larger population than the next three countries (approximately 1.2 billion). This suggests that these two countries will dominate further analysis on population sizes.

****

The below charts represent further subsets of population demographic. Specifically, by age. The categories can be seen in the titles for each graph and the boundaries for each category are <14, 15 – 24, 25 – 64, and 65+ respectively. These groups are determined by the dataset and cannot be adjusted.

**Chart, bar chart

Description automatically generated**

**Chart, bar chart

Description automatically generated**

**Chart, bar chart

Description automatically generated**

**Chart, bar chart

Description automatically generated**

As to be expected given the earlier analysis of overall population size, China and India feature, not only on every graph, but in the first two positions of every graph. What is interesting, however, is that there is a clear age demographic difference between these 2 countries. Namely, India has a much younger population, with exponentially more Children and Teens & Young Adults than China. China, on the other hand, has almost twice as many Retirees as India.

Also, in the table of 5 most populous countries, Nigeria is 5th, but it only appears on 2 of these age cohort tables – Children and Teens & Young Adults respectively. Equally, Nigeria is 3rd for most Children, yet 5th for most Teens & Young Adults. This suggests that the overall population in Nigeria is overwhelmingly young, with more Children than any other age demographic.

Further, on the Retirees graph, 2 of these countries do not feature on any other graph, nor the initial population table. These are Russian Federation and Germany. This indicates that they have the opposite age demographic composition from Nigeria, where they have a higher Retiree population than any other cohort.

When considering these age demographic charts and the earlier birth rates, we see that Asia is once again overrepresented. Four out of 5 countries for most Children and Teens & Young Adults are in Asia, all 5 countries with the highest Adult population are in Asia, and 3 (/4, as Russia spans 2 continents) out of the 5 highest Retirees population are also in Asia. However, these Asian countries which have the highest population do not have the highest birth rates. In fact, other Asian countries dominated the *lowest* birth rates. Conversely, although Africa was overrepresented in the highest birth rates, only Nigeria features on the highest population table and subsequent charts. To understand this more, further analysis of infant mortality rates and life expectancy would need to be carried out to account for why there is a disparity in the number of babies being born and overall population size.

**THIS REPORT WAS WRITTEN BY : Sara Lidguard**

