**Carlos Arturo Ramirez**

**CIND820 – Literature Review**

**Student Number: 500425774**

**Date of Submission: June 8, 2023**

Contents

[**Documentation Collected** 5](#_Toc137138552)

[Document Decline 5](#_Toc137138553)

[Summary: 5](#_Toc137138554)

[Replacement Rate 5](#_Toc137138555)

[Summary: 6](#_Toc137138556)

[Low Birth Rate 6](#_Toc137138557)

[Summary: 6](#_Toc137138558)

[Population Aging: 6](#_Toc137138559)

[Summary: 6](#_Toc137138560)

[Low immigration rates: 6](#_Toc137138561)

[Summary: 6](#_Toc137138562)

[Disease and Pandemics: 6](#_Toc137138563)

[Summary: 7](#_Toc137138564)

[War and conflict 7](#_Toc137138565)

[Summary 7](#_Toc137138566)

[Economic factors 7](#_Toc137138567)

[Summary: 7](#_Toc137138568)

[Environmental factors 7](#_Toc137138569)

[Summary 7](#_Toc137138570)

[Terminology 7](#_Toc137138571)

[**Data Description:** 12](#_Toc137138572)

[General overview 12](#_Toc137138573)

[**Dataset Statistics** 12](#_Toc137138574)

[Variable types 13](#_Toc137138575)

[**Important Variables** 13](#_Toc137138576)

[HeatMap of data 43](#_Toc137138577)

[**Approach of the project** 44](#_Toc137138578)

[Flow diagram 44](#_Toc137138579)

[Problem Statement 44](#_Toc137138580)

[Data Collection 44](#_Toc137138581)

[Data Cleaning 44](#_Toc137138582)

[ Remove duplicate values 45](#_Toc137138583)

[ Remove irrelevant observations 45](#_Toc137138584)

[ Address missing values 45](#_Toc137138585)

[ Reformat data types 45](#_Toc137138586)

[ Filter unwanted outliers 45](#_Toc137138587)

[ Reformat strings 45](#_Toc137138588)

[ Validate 45](#_Toc137138589)

[Exploratory Data Analysis (EDA) 45](#_Toc137138590)

[ Univariate non-graphical: make observations of the population and understand sample distributions of a single variable. (e.g. the measure of spread, the measure of central tendency, outlier detection) 45](#_Toc137138591)

[ Univariate graphical: graphical analysis on a single variable. (e.g. Histograms, Boxplots, Stem and leaf) 45](#_Toc137138592)

[ Multivariate non-graphical: techniques which show the relationship between two or more variables. (e.g. covariance, correlations) 45](#_Toc137138593)

[ Multivariate graphical: graphically show the relationship between two or more variables. (e.g. bar plots, scatterplots) 45](#_Toc137138594)

[Feature Engineering 45](#_Toc137138595)

[ Feature Extraction: select and/or combine variables into features to reduce the dimensionality of your dataset. (e.g. Principle Component Analysis, Nonlinear dimensionality reduction, unsupervised clustering methods) 46](#_Toc137138596)

[ Feature Selection: select the features which contribute most to the problem you are solving. (e.g. Variance thresholding, Pearson correlation, LASSO) 46](#_Toc137138597)

[ Feature Construction: the process of manually building more efficient features from raw data. (e.g. Dynamic aggregation of relational attributes) 46](#_Toc137138598)

[ Feature Learning: the automatic identification and use of features. (e.g. Restricted Boltzmann machine, K-means clustering) 46](#_Toc137138599)

[Modelling 46](#_Toc137138600)

[ Preprocessing 46](#_Toc137138601)

[ Machine learning models 46](#_Toc137138602)

[Communication 46](#_Toc137138603)

[ Don’t overcrowd your slides (6 items max) 46](#_Toc137138604)

[ Use relevant visualisations 46](#_Toc137138605)

[ Know your audience 47](#_Toc137138606)

[ Make sure it flows 47](#_Toc137138607)

[**Presumptions and clarification after exploration of the data:** 47](#_Toc137138608)

[Assess why stratify clustering based on country is required 47](#_Toc137138609)

[Why wrappers and multicollinearity is required in this study having that some analysis will be useful with country, and year is also important 47](#_Toc137138610)

[Groups of information by column: 47](#_Toc137138611)

[Dependant variable: 47](#_Toc137138612)

[Independent variables: 47](#_Toc137138613)

# **Documentation Collected**

## Document Decline

### Summary:

A population decline (also sometimes called underpopulation, depopulation, or population collapse) in humans is a reduction in a human population size, either caused by higher mortality rates, emigration or reduced birth rates

Until the beginning of the Industrial Revolution, the global population grew very slowly, at about 0.04% per year. After about 1800, the growth rate accelerated to a peak of 2.1% annually during the 1962–1968 period, but since then, due to the worldwide collapse of the total fertility rate, it has slowed to 0.9% as of 2023.[2] The global growth rate in absolute numbers accelerated to a peak of 92.8 million in 1990, but slowed to 80.0 million in 2019.

Current trend: But after 1968 the global population growth rate started a long decline, and in the period 2022–2027 the UN estimates it to be about 0.9%,[2] less than half of its peak during the period 1962 - 1968. Although still growing, the UN predicts that global population will level out around 2086,[2] and some sources predict the start of a decline before then.[1][12] The principal cause of this phenomenon is the abrupt decline in the global total fertility rate, from 5.3 in 1963 to 2.4 in 2019, as the world continues to move through the stages of the Demographic Transition .[4] The decline in the total fertility rate has occurred in every region

Relevant data available: In this article the attributes to check the natural change for quantifying the natural change of the population are natural change over a period of time in years.

## Replacement Rate

### Summary:

**PIP:**Replacement level fertility is the level of fertility at which a population exactly replaces itself from one generation to the next. In developed countries, replacement level fertility can be taken as requiring an average of 2.1 children per woman. In countries with high infant and child mortality rates, however, the average number of births may need to be much higher. Replacement level fertility is not associated with an unique set of age-specific birth rates.

## Low Birth Rate

### Summary:

When birth rates fall below the replacement level (generally considered to be around 2.1 children per woman), it can lead to a declining population over time.

## Population Aging:

### Summary:

An increase in life expectancy and declining birth rates can result in an aging population. As a larger proportion of the population grows older, it can lead to a decline in overall population size.

## Low immigration rates:

### Summary:

Insufficient levels of immigration to offset low birth rates or emigration can also result in population decline.

## Disease and Pandemics:

### Summary:

Severe disease outbreaks or pandemics can lead to a significant reduction in population, as seen in historical events like the Black Death or more recent examples like the COVID-19 pandemic.

## War and conflict

### Summary

Prolonged conflicts or wars can cause population decline through casualties, displacement, and migration.

## Economic factors

### Summary:

Economic instability, lack of job opportunities, and poor living conditions can contribute to outmigration and population decline.

## Environmental factors

### Summary

Natural disasters, environmental degradation, or inadequate access to resources can impact population size, especially in vulnerable regions.

## Terminology

**Glossary of Demographic Terms**

| **Indicator** | **Topic** | **Definition** |
| --- | --- | --- |
| **Age-specific fertility rates** | **Fertility** | Number of births to women in a particular age group, divided by the number of women in that age group. The age groups used are: 15-19, 20-24, ..., 45-49. The data refer to annual civil calendar years from 1 January to 31 December. |
| **Births** | **Fertility** | Number of births over a given period. Refers to live births for annual civil calendar years from 1 January to 31 December. Data are presented in thousands. |
| **Births by age group of mother** | **Fertility** | Number of births over a given period classified by age group of mother (15-19, 20-24, 25-29, 30-34, ...45-49). Refers to live births for annual civil calendar years from 1 January to 31 December. Data are presented in thousands. |
| **Child dependency ratio** | **Population** | The ratio of the population under age x1 (e.g., 0-24) to the number of persons of working age x1-x2 (e.g., 25-64). It is expressed as number of dependants per 100 persons of working age (e.g., 25-64). Data are presented for different age categories. |
| **Crude birth rate** | **Fertility** | Number of births over a given period divided by the person-years lived by the population over that period. It is expressed as number of births per 1,000 population. |
| **Crude death rate** | **Mortality** | Number of deaths over a given period divided by the person-years lived by the population over that period. It is expressed as number of deaths per 1,000 population. |
| **Deaths by sex** | **Mortality** | Number of deaths over a given period. Refers to annual civil calendar years from 1 January to 31 December. Data are presented in thousands. |
| **Deaths under age 5** | **Mortality** | Number of deaths under age 5 over a given period. Refers to annual civil calendar years from 1 January to 31 December. Data are presented in thousands. |
| **Feminity Ratio by broad age groups** | **Population** | Number of Males per 100 Females by various functional combination of age groups (0-14, 0-17, primary and secondary school ages, 15-24, 15-49, ..., 18+, 50+, etc.). De facto population as of 1 July of the year indicated. Figures are expressed per 100 female population. |
| **Infant Deaths, under age 1** | **Mortality** | Number of deaths under age 1 over a given period. Refers to annual civil calendar years from 1 January to 31 December. Data are presented in thousands. |
| **Infant mortality Rate** | **Mortality** | Probability of dying between birth and exact age 1. It is expressed as deaths per 1,000 live births. |
| **Life expectancy at birth** | **Mortality** | The average number of years of life expected by a hypothetical cohort of individuals who would be subject during all their lives to the mortality rates of a given period. It is expressed as years. |
| **Life expectancy at age x** | **Mortality** | The average number of remaining years of life expected by a hypothetical cohort of individuals who already reached age x and would be subject during the remainder of their lives to the mortality rates of a given period. It is expressed as years. |
| **Live births Surviving to Age 1** | **Population** | Number of newborns in a given year who survive to the next calendar year, and reach their first birthday. This cohort survival measure is commonly used to plan for immunization programs. It is calculated as (B \* (1-1q0)) + (B \* (1q0 \* 1a0)) where B is the number of live births, 1q0 is the probability of dying between birth and age 1, and 1a0 is the average time lived by those who died before reaching age 1. Data are presented in thousands. |
| **Mean age at childbearing** | **Fertility** | The mean age at childbearing is the mean age of mothers at the birth of their children if women were subject throughout their lives to the age-specific fertility rates observed in a given year. |
| **Median age** | **Population** | Age that divides the population in two parts of equal size, that is, there are as many persons with ages above the median as there are with ages below the median. It is expressed as years. |
| **Mortality before age x** | **Mortality** | The probability of dying between birth and exact age x measures the risk of premature death and is expressed as deaths under age x per 1,000 births. |
| **Mortality between age x and x+n** | **Mortality** | The probability that an individual alive at exact age x would die before exact age x+n, given the mortality conditions in a given year. For example, the probability of dying between exact age 15 and exact age 60 reflects mortality risks among adults and is expressed as deaths per 1,000 persons reaching age 15. |
| **Natural Change** | **Population** | Births minus Deaths. It is expressed as thousands. |
| **Net migration** | **Migration** | Net number of migrants, that is, the number of immigrants minus the number of emigrants. It is expressed as thousands. |
| **Net migration rate** | **Migration** | The number of immigrants minus the number of emigrants over a period, divided by the person-years lived by the population of the receiving country over that period. It is expressed as net number of migrants per 1,000 population. |
| **Net reproduction rate** | **Fertility** | The average number of daughters that female members of a birth cohort would bear during their reproductive life span if they were subject throughout their lives to the observed age-specific fertility and mortality rates of the given time period. It is expressed as the number of daughters per woman. |
| **Old-age dependency ratio** | **Population** | The ratio of the population aged x2 and over (e.g., 65+) to the number of persons of working age x1-x2 (e.g., 25-64). It is expressed as number of dependants per 100 persons of working age (e.g., 25-64). Data are presented for different age categories. |
| **Old-age potential support ratio** | **Population** | The ratio of the number of persons of working age x1-x2 (e.g., 25-64) to the population aged x2 and over (e.g., 65+). It is expressed as number of persons of working age x1-x2 (e.g., 25-64 years) per person aged x2 years or over (e.g., 65+). Data are presented for different age categories. |
| **Percentage by broad age groups** | **Population** | Percentage of Total Population by various functional combination of age groups (0-14, 0-17, primary and secondary school ages, 15-24, 15-49, ..., 18+, 50+, etc.). De facto population as of 1 July of the year indicated. Figures are expressed per 100 population. |
| **Percentage rural** | **Population** | Rural population as a percentage of the total population. |
| **Percentage urban** | **Population** | Urban population as a percentage of the total population. |
| **Population** | **Population** | De facto population in a country, area or region as of 1 January or 1 July of the year indicated. Figures are presented in thousands. |
| **Population annual doubling time** | **Population** | The population doubling time corresponds to the number of years required for the total population to double in size if the annual rate of population change would remain constant. It is calculated as ln(2)/r where r is the annual population growth rate. Doubling time is computed only for fast growing populations with growth rates exceeding 0.5 per cent. |
| **Population by broad age groups** | **Population** | De facto population as of 1 July of the year indicated classified by sex (male, female, both sexes combined) and by various functional combination of age groups (0-14, 0-17, primary and secondary school ages, 15-24, 15-49, ..., 18+, 50+, etc.). Data are presented in thousands. |
| **Population by five-year age group and sex** | **Population** | De facto population as of 1 July of the year indicated classified by sex (male, female, both sexes combined) and by five-year age groups (0-4, 5-9, 10-14, ..., 95-99, 100+). Data are presented in thousands. |
| **Population by sex** | **Population** | De facto population as of 1 July of the year indicated classified by sex (male, female, both sexes combined). Data are presented in thousands. |
| **Population change** | **Population** | Population increment over a period, that is, the difference between the population at the end of the period and that at the beginning of the period. Refers to annual civil calendar years from 1 January to 31 December. Data are presented in thousands. |
| **Population density** | **Population** | Population per square Kilometer. |
| **Population growth rate** | **Population** | Average exponential rate of growth of the population over a given period. It is calculated as ln(P2/P1)/n where P1 and P2 are the populations on 1 January of subsequent years, and n is the length of the period between t1 and t2 (n=1 for annual data). It is expressed as a percentage. |
| **Population sex ratio** | **Population** | Number of males per 100 females in the population. |
| **Rate of Natural Change** | **Population** | Crude birth rate minus the crude death rate. Represents the portion of population growth (or decline) determined exclusively by births and deaths. It is expressed as number of births per 1,000 population. |
| **Rural population** | **Population** | De facto population living in areas classified as rural (that is, it is the difference between the total population of a country and its urban population). Data refer to 1 July of the year indicated and are presented in thousands. |
| **Sex ratio at birth** | **Fertility** | Number of male births per 100 female births. |
| **Sex Ratio by broad age groups** | **Population** | Number of Females per 100 Males by various functional combination of age groups (0-14, 0-17, primary and secondary school ages, 15-24, 15-49, ..., 18+, 50+, etc.). De facto population as of 1 July of the year indicated. Figures are expressed per 100 male population. |
| **Total dependency ratio** | **Population** | The ratio of the sum of the population under age x1 (e.g., 0-24) and that aged x2 and over (e.g., 65+) to the number of persons of working age x1-x2 (e.g., 25-64). It is expressed as number of dependants per 100 persons of working age (e.g., 25-64). Data are presented for different age categories. |
| **Total fertility rate** | **Fertility** | The average number of live births a hypothetical cohort of women would have at the end of their reproductive period if they were subject during their whole lives to the fertility rates of a given period and if they were not subject to mortality. It is expressed as live births per woman. |
| **Under-five Mortality Rate** | **Mortality** | Probability of dying between birth and exact age 5. It is expressed as deaths per 1,000 births. |
| **Urban population** | **Population** | De facto population living in areas classified as urban according to the criteria used by each area or country. Data refer to 1 July of the year indicated and are presented in thousands. |
| **Women aged 15-49** | **Fertility** | Number of women aged 15-49 as of 1 July of the year indicated, and that number as a percentage of the total female population as of 1 July of the year indicated. The number of women is presented in thousands. |

# **Data Description:**

## General overview

### **Dataset Statistics**

Dataset statistics

|  |  |
| --- | --- |
| **Number of variables** | 67 |
| **Number of observations** | 17538 |
| **Missing cells** | 16496 |
| **Missing cells (%)** | 1.4% |
| **Duplicate rows** | 0 |
| **Duplicate rows (%)** | 0.0% |
| **Total size in memory** | 9.1 MiB |
| **Average record size in memory** | 544.0 B |
|  |  |

### Variable types

|  |  |
| --- | --- |
| **Numeric** | 60 |
| **Text** | 3 |
| **Categorical** | 4 |

### **Important Variables**

[LocID](file:///C:\Users\Public\Documents\cind820_capstone\proyect_data_analysis\population.html#pp_var_7303704326026779423)  
Real number (ℝ)

|  |  |
| --- | --- |
| **Distinct** | 237 |
| **Distinct (%)** | 1.4% |
| **Missing** | 0 |
| **Missing (%)** | 0.0% |
| **Infinite** | 0 |
| **Infinite (%)** | 0.0% |
| **Mean** | 440.2827 |
| **Minimum** | 4 |
| **Maximum** | 894 |
| **Zeros** | 0 |
| **Zeros (%)** | 0.0% |
| **Negative** | 0 |
| **Negative (%)** | 0.0% |
| **Memory size** | 274.0 KiB |

[Location](file:///C:\Users\Public\Documents\cind820_capstone\proyect_data_analysis\population.html#pp_var_-8632165206330318936)  
Text

|  |  |
| --- | --- |
| **Distinct** | 237 |
| **Distinct (%)** | 1.4% |
| **Missing** | 0 |
| **Missing (%)** | 0.0% |
| **Memory size** | 274.0 KiB |

[Time](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-5427304558351154836)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 74 | |
| **Distinct (%)** | 0.4% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 1986.5 | |
| **Minimum** | | 1950 | |
| **Maximum** | | 2023 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[TPopulation1Jan](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-909445285330829770)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17426 | |
| **Distinct (%)** | 99.4% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 21315.413 | |
| **Minimum** | | 0.508 | |
| **Maximum** | | 1425925.4 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[TPopulationMale1July](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-2886227159196142758)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17306 | |
| **Distinct (%)** | 98.7% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 10778.91 | |
| **Minimum** | | 0.246 | |
| **Maximum** | | 736851.53 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[TPopulationMale1July](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-2886227159196142758)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17306 | |
| **Distinct (%)** | 98.7% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 10778.91 | |
| **Minimum** | | 0.246 | |
| **Maximum** | | 736851.53 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[TPopulationFemale1July](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_4629935417273986776)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17295 | |
| **Distinct (%)** | 98.6% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 10696.282 | |
| **Minimum** | | 0.265 | |
| **Maximum** | | 698498.54 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[PopDensity](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_3383767214632776164)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17419 | |
| **Distinct (%)** | 99.3% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 311.0506 | |
| **Minimum** | | 0.0488 | |
| **Maximum** | | 24879.53 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[PopSexRatio](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_2233476939868149624)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 16784 | |
| **Distinct (%)** | 95.7% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 100.18233 | |
| **Minimum** | | 63.6007 | |
| **Maximum** | | 327.4562 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[MedianAgePop](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-5503812170237705621)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 16778 | |
| **Distinct (%)** | 95.7% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 23.855083 | |
| **Minimum** | | 13.0783 | |
| **Maximum** | | 57.7188 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[NatChange](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_3725541731862260066)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 14788 | |
| **Distinct (%)** | 84.3% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 319.55694 | |
| **Minimum** | | -1061.083 | |
| **Maximum** | | 22509.645 | |
| **Zeros** | | 9 | |
| **Zeros (%)** | | 0.1% | |
| **Negative** | | 974 | |
| **Negative (%)** | | 5.6% | |
| **Memory size** | | 790.1 KiB | |

[NatChangeRT](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_3769896282980781694)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 14108 | |
| **Distinct (%)** | 80.4% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 17.46044 | |
| **Minimum** | | -62.082 | |
| **Maximum** | | 43.707 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 978 | |
| **Negative (%)** | | 5.6% | |
| **Memory size** | | 790.1 KiB | |

[PopChange](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-3228181019784159319)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 14978 | |
| **Distinct (%)** | 85.4% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 319.55731 | |
| **Minimum** | | -7264.93 | |
| **Maximum** | | 22289.974 | |
| **Zeros** | | 11 | |
| **Zeros (%)** | | 0.1% | |
| **Negative** | | 1786 | |
| **Negative (%)** | | 10.2% | |
| **Memory size** | | 790.1 KiB | |

[PopGrowthRate](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-7218686760665210257)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 5662 | |
| **Distinct (%)** | 32.3% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 1.7276559 | |
| **Minimum** | | -71.689 | |
| **Maximum** | | 36.299 | |
| **Zeros** | | 13 | |
| **Zeros (%)** | | 0.1% | |
| **Negative** | | 1785 | |
| **Negative (%)** | | 10.2% | |
| **Memory size** | | 790.1 KiB | |

[DoublingTime](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_5371853805289180076)  
Real number (ℝ)

HIGH CORRELATION  MISSING

|  |  |  |
| --- | --- | --- |
| **Distinct** | 4040 | |
| **Distinct (%)** | 28.9% | |
| **Missing** | 3546 | |
| **Missing (%)** | 20.2% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 43.187633 | |
| **Minimum** | | 1.9095 | |
| **Maximum** | | 138.3527 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[Births](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_4913512207752860088)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 15340 | |
| **Distinct (%)** | 87.5% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 535.28656 | |
| **Minimum** | | 0.002 | |
| **Maximum** | | 33567.34 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[Births1519](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_3979748025263545655)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 11617 | |
| **Distinct (%)** | 66.2% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 64.294392 | |
| **Minimum** | | 0 | |
| **Maximum** | | 5714.733 | |
| **Zeros** | | 83 | |
| **Zeros (%)** | | 0.5% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[CBR](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-5450891459210526096)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 14466 | |
| **Distinct (%)** | 82.5% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 28.60834 | |
| **Minimum** | | 4.089 | |
| **Maximum** | | 59.416 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[TFR](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_4795294077001089400)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 15146 | |
| **Distinct (%)** | 86.4% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 3.9794014 | |
| **Minimum** | | 0.7455 | |
| **Maximum** | | 8.8637 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[NRR](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_6321416072467647781)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 12334 | |
| **Distinct (%)** | 70.3% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 1.6119997 | |
| **Minimum** | | 0.2823 | |
| **Maximum** | | 3.5099 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[MAC](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-8500028186232076670)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 5639 | |
| **Distinct (%)** | 32.2% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 28.781747 | |
| **Minimum** | | 23.685 | |
| **Maximum** | | 33.023 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[SRB](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_3171343233238376261)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 185 | |
| **Distinct (%)** | 1.1% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 105.0565 | |
| **Minimum** | | 96.7 | |
| **Maximum** | | 128.2 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[Deaths](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-3392614148408164425)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 13895 | |
| **Distinct (%)** | 79.2% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 215.72961 | |
| **Minimum** | | 0.005 | |
| **Maximum** | | 19613.435 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[DeathsMale](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-8707993978025899997)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 13065 | |
| **Distinct (%)** | 74.5% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 114.57483 | |
| **Minimum** | | 0.002 | |
| **Maximum** | | 10352.999 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[DeathsFemale](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-7345344894006607540)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 12744 | |
| **Distinct (%)** | 72.7% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 101.15479 | |
| **Minimum** | | 0.002 | |
| **Maximum** | | 9260.436 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[CDR](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_8465301678791395416)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 11001 | |
| **Distinct (%)** | 62.7% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 11.1479 | |
| **Minimum** | | 0.795 | |
| **Maximum** | | 103.534 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[LEx](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_4453562486566386749)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17167 | |
| **Distinct (%)** | 97.9% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 63.853736 | |
| **Minimum** | | 11.9951 | |
| **Maximum** | | 87.0143 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[LExMale](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_348088296953211666)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17041 | |
| **Distinct (%)** | 97.2% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 61.418253 | |
| **Minimum** | | 10.0537 | |
| **Maximum** | | 85.1679 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[LExFemale](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-3280894856559892950)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17074 | |
| **Distinct (%)** | 97.4% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 66.39228 | |
| **Minimum** | | 12.721 | |
| **Maximum** | | 88.9943 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[LE15](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-4259741207164923020)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 16922 | |
| **Distinct (%)** | 96.5% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 54.916893 | |
| **Minimum** | | 8.6848 | |
| **Maximum** | | 72.4198 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[LE15Male](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_8126270463495702934)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 16830 | |
| **Distinct (%)** | 96.0% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 52.631341 | |
| **Minimum** | | 7.0621 | |
| **Maximum** | | 70.5694 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[LE15Female](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_3637102682413603294)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 16849 | |
| **Distinct (%)** | 96.1% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 57.280779 | |
| **Minimum** | | 11.2629 | |
| **Maximum** | | 74.4041 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[LE65](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_2908808122009942566)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 15817 | |
| **Distinct (%)** | 90.2% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 13.967675 | |
| **Minimum** | | 3.7412 | |
| **Maximum** | | 23.7298 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[LE65Male](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-9197943631248905017)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 15379 | |
| **Distinct (%)** | 87.7% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 12.799532 | |
| **Minimum** | | 3.2006 | |
| **Maximum** | | 22.1186 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[LE65Female](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_7289554825634059638)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 16025 | |
| **Distinct (%)** | 91.4% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 15.005129 | |
| **Minimum** | | 4.3896 | |
| **Maximum** | | 25.4487 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[LE80](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_2674290446437140966)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 14301 | |
| **Distinct (%)** | 81.5% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 6.1898123 | |
| **Minimum** | | 2.1042 | |
| **Maximum** | | 11.6846 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[LE80Male](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-3375158285567377448)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 13558 | |
| **Distinct (%)** | 77.3% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 5.622227 | |
| **Minimum** | | 1.8464 | |
| **Maximum** | | 10.4995 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[LE80Female](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-8702206000190493626)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 14516 | |
| **Distinct (%)** | 82.8% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 6.5769712 | |
| **Minimum** | | 2.2776 | |
| **Maximum** | | 12.9128 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[InfantDeaths](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-8700892949239081367)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 9716 | |
| **Distinct (%)** | 55.4% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 38.276884 | |
| **Minimum** | | 0 | |
| **Maximum** | | 3954.196 | |
| **Zeros** | | 145 | |
| **Zeros (%)** | | 0.8% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[IMR](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_2986246233300750764)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17404 | |
| **Distinct (%)** | 99.2% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 56.554086 | |
| **Minimum** | | 0.8442 | |
| **Maximum** | | 400.6434 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[LBsurvivingAge1](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_6720445502544253827)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 15283 | |
| **Distinct (%)** | 87.1% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 507.34287 | |
| **Minimum** | | 0.002 | |
| **Maximum** | | 31266.974 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[Under5Deaths](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_3487413736306473616)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 10201 | |
| **Distinct (%)** | 58.2% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 55.792883 | |
| **Minimum** | | 0 | |
| **Maximum** | | 7744.457 | |
| **Zeros** | | 126 | |
| **Zeros (%)** | | 0.7% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[Q5](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_4510601570379652160)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17447 | |
| **Distinct (%)** | 99.5% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 83.937328 | |
| **Minimum** | | 1.4623 | |
| **Maximum** | | 659.3338 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[Q0040](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-2383629977619701546)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17482 | |
| **Distinct (%)** | 99.7% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 162.53711 | |
| **Minimum** | | 6.4528 | |
| **Maximum** | | 986.7973 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[Q0040Male](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_7255577245359500551)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17423 | |
| **Distinct (%)** | 99.3% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 177.38317 | |
| **Minimum** | | 6.4917 | |
| **Maximum** | | 995.6609 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[Q0040Female](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_6020492526236682112)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17385 | |
| **Distinct (%)** | 99.1% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 147.08334 | |
| **Minimum** | | 5.0168 | |
| **Maximum** | | 962.566 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[Q0060](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-1159474100123543299)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17510 | |
| **Distinct (%)** | 99.8% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 297.60763 | |
| **Minimum** | | 27.9222 | |
| **Maximum** | | 999.7985 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[Q0060Male](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_970786674092516079)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17451 | |
| **Distinct (%)** | 99.5% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 335.77947 | |
| **Minimum** | | 32.12 | |
| **Maximum** | | 999.9713 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[Q0060Female](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_48173781692494409)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17428 | |
| **Distinct (%)** | 99.4% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 257.7853 | |
| **Minimum** | | 22.7805 | |
| **Maximum** | | 998.7521 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[Q1550](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-3212800158288429640)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17473 | |
| **Distinct (%)** | 99.6% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 134.63075 | |
| **Minimum** | | 8.9296 | |
| **Maximum** | | 996.1659 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[Q1550Male](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_6317881477587753621)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17303 | |
| **Distinct (%)** | 98.7% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 155.74932 | |
| **Minimum** | | 9.3056 | |
| **Maximum** | | 999.1017 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[Q1550Female](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-3352727934005961828)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17174 | |
| **Distinct (%)** | 97.9% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 112.81804 | |
| **Minimum** | | 6.5188 | |
| **Maximum** | | 985.1255 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[Q1560](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-1445768557488040042)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17498 | |
| **Distinct (%)** | 99.8% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 233.17885 | |
| **Minimum** | | 23.3059 | |
| **Maximum** | | 999.6025 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[Q1560Male](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_2897660173210578698)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17426 | |
| **Distinct (%)** | 99.4% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 271.52965 | |
| **Minimum** | | 27.6102 | |
| **Maximum** | | 999.9418 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[Q1560Female](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-5053048964849027440)  
Real number (ℝ)

|  |  |  |
| --- | --- | --- |
| **Distinct** | 17353 | |
| **Distinct (%)** | 98.9% | |
| **Missing** | 0 | |
| **Missing (%)** | 0.0% | |
| **Infinite** | 0 | |
| **Infinite (%)** | 0.0% | |
| **Mean** | 193.54883 | |
| **Minimum** | | 18.1986 | |
| **Maximum** | | 997.5971 | |
| **Zeros** | | 0 | |
| **Zeros (%)** | | 0.0% | |
| **Negative** | | 0 | |
| **Negative (%)** | | 0.0% | |
| **Memory size** | | 790.1 KiB | |

[NetMigrations](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-4023493991908055322)  
Real number (ℝ)

HIGH CORRELATION  ZEROS

|  |  |  |  |
| --- | --- | --- | --- |
| **Distinct** | 12802 | | |
| **Distinct (%)** | 73.0% | | |
| **Missing** | 0 | | |
| **Missing (%)** | 0.0% | | |
| **Infinite** | 0 | | |
| **Infinite (%)** | 0.0% | | |
| **Mean** | 4.5786293 × 10-5 | | |
| **Minimum** | | -6673.58 |
| **Maximum** | | 3366.387 |
| **Zeros** | | 623 |
| **Zeros (%)** | | 3.6% |
| **Negative** | | 9672 |
| **Negative (%)** | | 55.1% |
| **Memory size** | | 790.1 KiB |

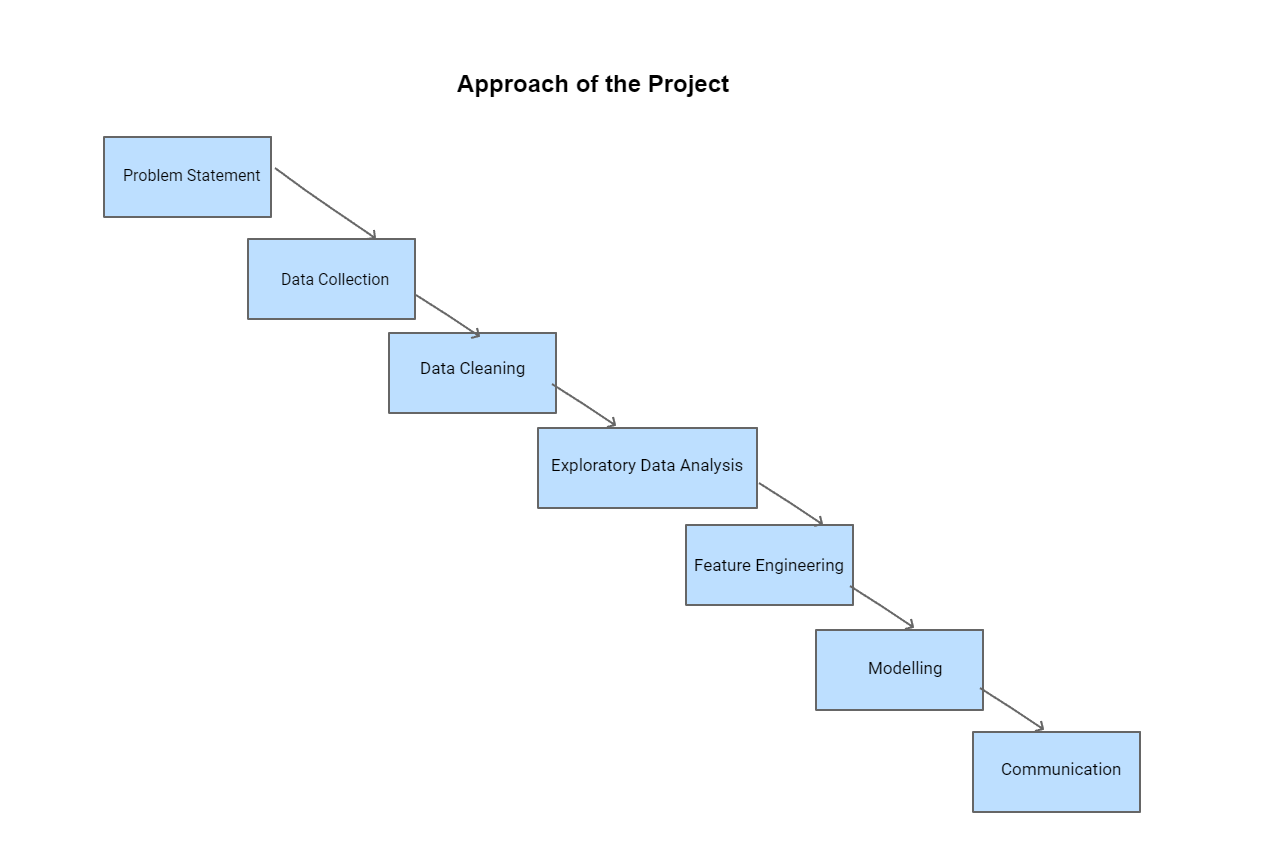
[CNMR](http://localhost:8889/notebooks/pandas_profiling_population-improved.ipynb#pp_var_-5841574420743773068)  
Real number (ℝ)

HIGH CORRELATION  ZEROS

|  |  |  |  |
| --- | --- | --- | --- |
| **Distinct** | 12070 | | |
| **Distinct (%)** | 68.8% | | |
| **Missing** | 0 | | |
| **Missing (%)** | 0.0% | | |
| **Infinite** | 0 | | |
| **Infinite (%)** | 0.0% | | |
| **Mean** | -0.057138613 | | |
| **Minimum** | | -526.323 |
| **Maximum** | | 415.239 |
| **Zeros** | | 623 |
| **Zeros (%)** | | 3.6% |
| **Negative** | | 9672 |
| **Negative (%)** | | 55.1% |
| **Memory size** | | 790.1 KiB |
| HeatMap of data | |  |

# **Approach of the project**

## Flow diagram



## Problem Statement

## The UN projects that the global population will increase from a population of around 8 billion in 2022 to 10.4 billion by the end of the century. By that time, the UN projects, fast global population growth will come to an end. According to [worldometers](https://www.worldometers.info/world-population/#:~:text=World%20Population%20Clock%3A%208%20Billion,(LIVE%2C%202023)%20%2D%20Worldometer) 8 billion people in the world at present, and this number is likely to keep rising as we add 315,000 people to the planet every day(1)

## Global population growth is determined by the number of births and deaths. Improving health is increasing the size of the population as it is decreasing mortality. The countervailing trend is falling fertility rates – the trend of couples having fewer children is what brought rapid population growth to an end in many countries already, and what will bring an end to rapid population growth globally.

## The human population could outweigh resources available to sustain it. This is why it is important to study population data and devise a means to predict future population. Population projection can help governments to make decisions and plans about the future.

## Data Collection

There are many statistics agencies for every country. Our data collection will only use the united nations data collection, extracted from the department of economic and social affairs

https://population.un.org/wpp/Download/

* Demographic Indicator
* Percentage of population on 01 July, by 5-year age groups.

## Data Cleaning

Cleaning your data is a process of ensuring your data is in the correct format; consistent and errors are identified and dealt with appropriately.

### The study is intended to be conducted as **Country** based not by regions, we are removing **regions** included in the UN datasets:

#The study needs only countries, we ommit regions by deleting them:

withNulls =df[ (df['ISO3\_code'].isnull() ) ].index

df.drop(withNulls , inplace=True)

* There are missing values in the fields.

#There are missing in some numeric datatype attributes, we are filling them with zeros since zeros does #it is the right value for not showing population growth in certain countries such the Vatican City

df['DoublingTime'] = df['DoublingTime'].fillna(0)

* The dataset includes projections for population up 2100

#The country only needs proyections and data up to 2022

estimationsGreaterCurrent = df[ (df['Time'] > 2023 ) ].index

df.drop(estimationsGreaterCurrent , inplace=True)estimationsGreaterCurrent = df[ (df['Time'] > 2023 ) ].index

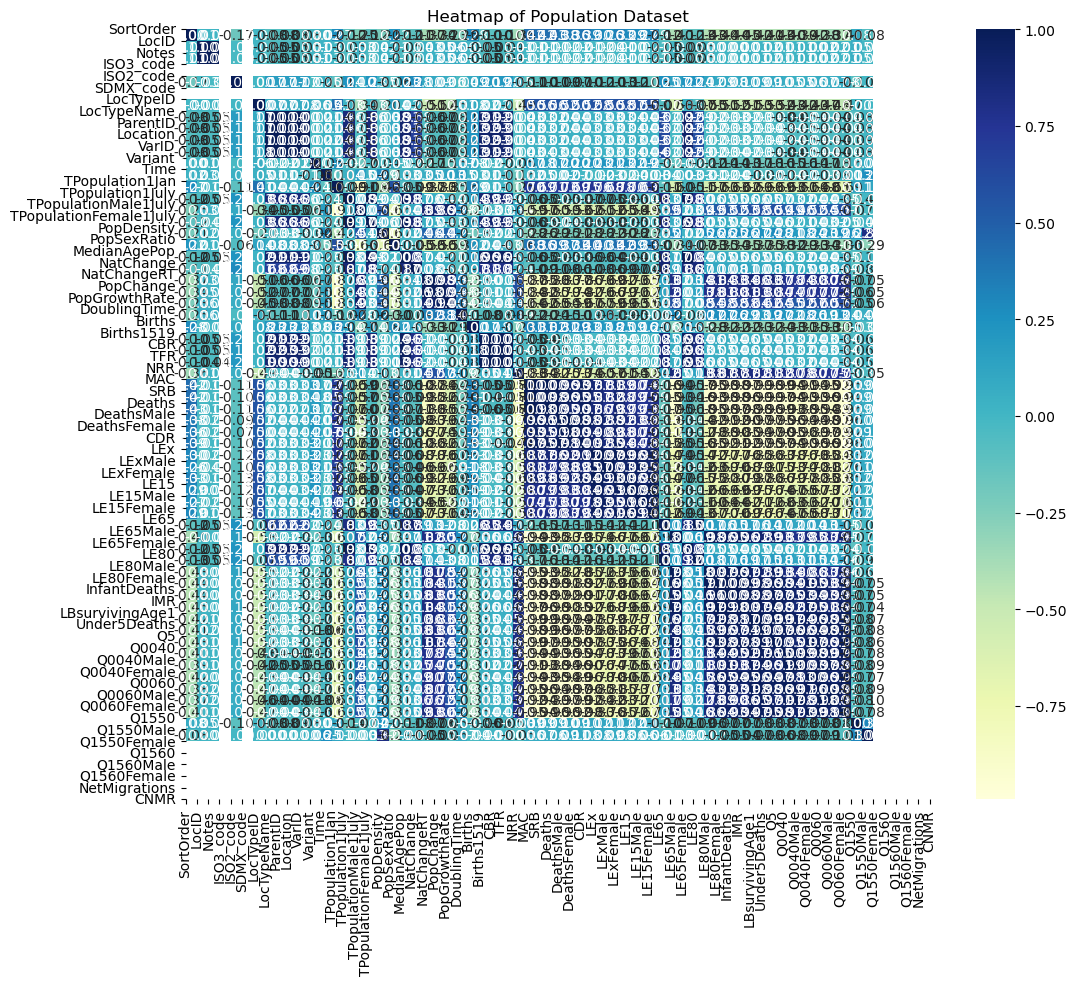
df.drop(estimationsGreaterCurrent , inplace=True)

## Exploratory Data Analysis (EDA)

Process for underlying patterns within the data, detect outliers and test assumptions with the final aim of finding a model that fits the data well.

### Univariate non-graphical: Already included in the document, since

* Heat map of the current dataset



## Feature Engineering

Is defined as the process of transforming raw data into features that better represent the underlying problem to the predictive models, resulting in improved model accuracy on unseen data.

### There are 63 columns and 43472 records for this study. I am applying the following methods for reduction:

* + Correlation, and correlation matrix
  + PCA: Principal component Analysis: in which I have plotted the scree plot for applying the rule of thumb for eigen values and vectors vs components.
  + Backwards elimination
  + Plotting the most important collinear duplas
  + remove collinear features based on the correlation coefficient between features

## Modelling

Application of All machine learning models are classified either a**Supervised**or**Unsupervised**learning problem

### Supervised modeling applying multiple linear regression, where the target variable is population size

* Stratify clustering based on Country

## Communication

It is essential to communicate your results. This can be done through a **presentation, formal report** or even a **blog post**. The point is the world has to see the amazing work you have done. A few key points to remember:

* Don’t overcrowd your slides (6 items max)
* Use relevant visualisations
* Know your audience
* Make sure it flows

# **Presumptions and clarification after exploration of the data:**

## Assess why stratify clustering based on country is required.

## Why wrappers and multicollinearity are required in this study having that some analysis will be useful with country, and year is also important.

## Groups of information by column:

* Time in years
* Population
* Fertility
* Mortality
* Migration

## Dependant variable:

* Population size

## Independent variables:

* Fields within the groups Fertility, Mortality, Migration
* Time: in years
* Demographics: such total population, male and female population, median age, natural change births minus deaths
* Policy and governance
* Education and health care
* Migration( already included )