**SID: 510012005**

**Data dictionary:**

* **Country:** Country Name
* **Year:** Year of observation
* **Both sexes:** Adult mortality rate (probability of dying between 15 and 60 years per 1000 population) for both sexes
* **Male:** Adult mortality rate (probability of dying between 15 and 60 years per 1000 population) for Males
* **Female:** Adult mortality rate (probability of dying between 15 and 60 years per 1000 population) for Females

**Data Structure:**

Column: 3113

Row: 5 (variables)

**Metadata:**

**source:** WHO data catalog

**Link:** World Mortality Rates (2000-2017)

[**https://www.kaggle.com/datasets/navinmundhra/world-mortality/metadata?resource=download&select=Adult+mortality.csv**](https://www.kaggle.com/datasets/navinmundhra/world-mortality/metadata?resource=download&select=Adult+mortality.csv%20)

**Collaborators:** Navin Mundhra

**date of collection: 06/09/2020**

**strengths:**

This dataset contains adult mortality rates for a large number of countries around the world, including rates categorized into male and female genders. This data set is useful for people to study the effect of male and female gender on mortality, or to study and compare mortality between countries.

On this basis, the data set still has a large amount of data, which will not cause some problems due to the lack of data.

**limitations:**

The dataset only included mortality rates for people aged 15-60, so it was not possible to study the population as a whole. (For example, there is no way to study mortality in infants or people over 60.)

The number of countries included in this dataset is too large, and if researchers want to study mortality rates in specific countries, they need to be screened.

The data in this dataset points to the effect of gender on mortality, which is inaccurate in a single study. (For example, because mortality is also influenced by other factors.)

**Quality of data:**

I went through the code to check for null values in the dataset:

文本

描述已自动生成

The resulting output is as follows:



Next, I sorted and integrated the data in the dataset. And get 3 dictionaries with the lists about data at the end. During this process, I also checked the dataset for outliers (such as -1 for the year, etc.):

文本

描述已自动生成

In the output of the program, the information "error data" does not appear, indicating that there are no outliers in the data set.

**Data analysis:**

First, I classified the data into "bothsex" "male" and "female". And it is divided into 3 dictionaries, where the key is the country name, and the value is the average mortality rate.

文本

描述已自动生成

Next, on the basis of the above, I found the maximum/minimum average mortality rate for each country by gender (from 2000-2016) (ages 15-60)

文本

描述已自动生成

电脑萤幕画面

描述已自动生成

The resulting output is as follows:

图片包含 文本

描述已自动生成

Next, I analysed the data to Find the relationship between mortality (from 2000-2016) (ages 15-60) and gender in each country

电脑萤幕的截图

描述已自动生成

The resulting output is as follows:



Finally, I analyzed the data to Find which country has the highest/smallest difference in mortality between male and female

文本

描述已自动生成

The resulting output is as follows:



**Summary:**

From the results obtained from the above analysis, we can see that the overall average mortality rate and the average mortality rate of male and female in the country of Zimbabwe are the highest among all the countries in the data set, which can be inferred that there are some lagging aspects for Zimbabwe, but the specific aspects have to be analyzed according to other data.

The countries with the lowest overall, male, and female mortality rates were Iceland, Qatar, and Japan, respectively. This data is of great help to researchers. For example, if you want to study how to reduce the mortality rate of a certain gender or the whole, you can conduct research and analysis from these three countries.

The second result is that of the 183 countries in the dataset, only one country has a higher average mortality rate for women than for men, and that country is Bhutan. Here, we can speculate that the physiological structure of the female gender may have an impact on mortality and life expectancy, and the life expectancy of men may generally be lower than that of women.

The final result is that the difference in mortality between men and women is the largest in the Russian Federation. The smallest difference is in Gabon. This could potentially be used to study whether the benefits, policies, treatments, etc. received by women and men in a country lead to an increase in mortality. (For example, in countries with higher status of women, women may receive better medical conditions, so the average death rate will also be lower)