

# Project Final Report: A Study on Child Mortality Rates

**Sreekar Guggilam**

**Ashish Sanghvi**

**Rohith Kumar Poshala**

## **Introduction:**

According to UNICEF, “An estimated 5.3 million children under age five died in 2018—roughly half of those deaths occurred in sub-Saharan Africa”. Even though there has been a remarkable progress in reducing the mortality rates globally, the major causes of death vary widely across countries. Despite the challenges in understanding and addressing infant mortality, it is crucial to study this to improve the collective health of a nation.

Thus, the objective of the project is to analyze the factors affecting child mortality rate across the world by analyzing the data from all the countries under UNICEF surveillance. The following report is a brief overview of the proposed analysis and milestones for the project.

## **Data pre-processing in SQL:**

The aggregated data warehouse on child mortalities provided by UNICEF is chosen for the analysis. It provides data from 202 countries and areas monitored by UNICEF organization. Statistics on child survival, development and protection for the world’s countries, areas and regions collected since 1950s have been provided by UNICEF and can be used to study the trends across nations.

Datasets pertaining to mortality, mortality rates, and sex ratios, for under 5yrs as well as 5-14 yrs age groups have been chosen to collectively identify and analyze the different causal factors across various geographical locations in the world.

For analysis, the data was initially uploaded into a SQL database. For this, data from the excel sheets has been modified and made compatible to store into tables before saving into a csv file. For example :

1. Entire excel sheet has been formatted to text to avoid comma separated issues while inserting a number into the tables.

2. Replaced country 'Cote d'Ivoire' with 'Ivory Coast' to avoid issues while inserting data into the tables created.

Tables were merged and modified in SQL for improved analysis and the subsequent tables in the SQL database have been dumped into a .sql and stored into a data base created. Refer appendix for list of tables used from UNICEF and created/modified in SQL.

Few more tables have been created from the existing tables and data with median data has been populated into it for further visualization.

## Analysis and Visualization:

Mortality rates across various nations have been compared using visualizations. Pie charts were used to identify most significant causes of deaths in various countries. In particular, neonatal, infant, under 5yr, 5-14 yrs age ranges have been used to compare the mortality rates.

The trends in child mortality over past decades is chosen as a secondary objective. A slider graph of the world mortality rates have been used to study the relative change in the trends across the planet.

Trends in world mortality rates across different age groups have been visualized using line plots. A significant decreasing trend was observed across all age groups suggesting positive impact. We could observe a spike in death rates during 1970's which was due to increase in number of countries which joined UNICEF during that time and more sophisticated data collection. And the weightage of percentage contributed by each country towards mortality rate over around 60 years has been visualized using a choropleth world map with a slider.

More analysis has been done and data of number of deaths from all countries across the time period has been visualized but for further analysis it has observed that comparison of mortality rates between countries is more sensible than number of deaths.

Further analysis has been done on a particular country for given year has been visualized to understand the contribution of 4 different classes (neonatal, infant, under 5yr, 5-14 yrs age)

Later in order to understand how a specific country has been tackling (or need to focus) mortality (basically which class has a country being focusing on) has been visualized by observing the trends of different classes for a particular country over the years.

Data frames have been created in pandas by merging multiple tables. And Required data frames have been created by merging multiple data frames according to the visualization we need to do.

Line graphs to visualize mortality rates for all classes of a country over years along with the range (Upper and Lower) has been plotted. It had been observed that the data collection has become more precise over time for all countries. And mortality rate for almost all countries has been observed to be decreasing over time.

## Technical components and User interface:

- All functions defined have a **default input** (as latest year) in case year is not given by user.
- User search for the countries has been made **case insensitive**.
- A function for **Dynamic suggestions** for country name has been defined in such a way that a country is name is suggested from the list of countries by matching maximum streak of correct letters in the input given by the user. If no letter is matched, an error message is displayed to the user.

- **Try and except blocks** have been implemented during the user interface to prevent the errors caused by incorrect user inputs and in turn instructing the user with appropriate message to go further. (while passing country name or year intended)
- **Interactive slider** has been included to the interactive figures and charts to help the visualization of data across the years.

## Conclusion:

Data has been categorized and visualization has been done for all countries across the years. And we have observed that there are many factors influencing mortality rate for a country. As almost all countries have a decline in mortality rates, we could observe that a stable government and economy contributes the most for the decline in mortality rates for a country (observed through the countries which have high mortality rates are still either under developed or developing). Through the visualizations done, we can see where a country should focus to better its child mortality rate.

## Appendix:

Tables created using sqlite3:

```
'A_5_14_DEATHS',
'A_5_14_MORTALITY_RATE',
'INFANT_DEATHS',
'INFANT_MORTALITY_RATE',
'NEONATAL_DEATHS',
'NEONATAL_MORTALITY_RATE',
'U5MR_DEATHS',
'U5MR_MORTALITY_RATE',
'IMR_MALE_M',
'IMR_FEMALE_M',
'IMR_SEXRATIO_M',
'U5MR_MALE_M',
'U5MR_FEMALE_M',
'U5MR_SEXRATIO_M'
```

Tables created from existing tables using sqlite3 through python:

```
'A_5_14_DEATHS_B_M',
'A_5_14_MORTALITY_RATE_B_M',
'INFANT_DEATHS_B_M',
'INFANT_MORTALITY_RATE_B_M',
'NEONATAL_DEATHS_B_M',
'NEONATAL_MORTALITY_RATE_B_M',
'U5MR_DEATHS_B_M',
'U5MR_MORTALITY_RATE_B_M'
```

Data frames created using pandas:

```
'a_5_14_deaths', 'a_5_14_mortality_rate', 'infant_deaths', 'infant_mortality_rate',
'neonatal_deaths', 'neonatal_mortality_rate', 'u5mr_deaths', 'u5mr_mortality_rate'
```

## Data Source:

“Child Mortality Data.” *UNICEF*, Retrieved from <https://data.unicef.org/resources/dataset/child-mortality/>

