**SCHOOL OF INFORMATION SCIENCES**

**Manipal University, Manipal.**


# Data Visualization and Analysis of Child Mortality based Geo-Spatial Data

**Submitted by**

**Shankramma Oursang**
**171046027**

**Shweta Aursang**
**171046029**

**ME "Big Data and Data Analytics"**

**Under the guidance of**

**Sandeep Patil,**
**Applied Cognition Systems Pvt. Ltd.**
**Big Data and Data Analytics,**
**School of Information Sciences,**
**Manipal University.**

# INDEX

# 1. Abstract

Child mortality is a core indicator for child health and well-being. The main objective of the project is to visualize and analyse the child mortal data so as to make people understand through visualization. We are trying to make an attempt to reach people around the globe and by letting people know about facts and causes for the child mortality. However, traditional methods of analysing child mortality operations data are costly, rigid, and complex. Clinical and operations data must first be logged in applications and then cleansed and fit into a standard data model. When a change is required, database experts can require weeks or months to adjust the data model and prepare new reports.

This report represents the latest estimates of Under-five, Infant mortality up to the year 2015 and assesses progress at the country, regional and global levels. This report is also provides an overview on the estimation methods used for child mortality indicators.

# 2. Objective

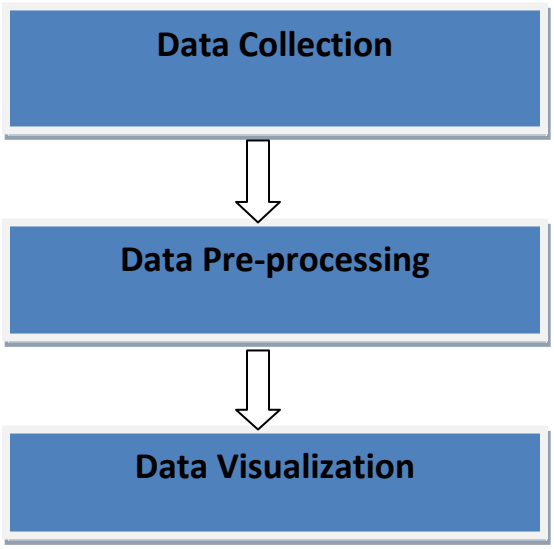
The objective of the project is to visualize and analyses the child mortality data so as to make people understand through visualization. We are trying to make an attempt to reach the people around the globe and by letting them to known facts, figures and causes for the child mortality that has happened.

# 3. Introduction

Child mortality is a core indicator for child health and well-being. Every child movement boosted global momentum in improving newborn and child survival. Evidence -based estimation of child mortality is a cornerstone for tracking progress towards child survival goals and for planning national and global health and well-being. The report represents the latest estimates of under-five, infant mortality up to the year 2015 and assesses progress at the country, regional and global levels. The report also provides an overview on the estimation methods used for child mortality indicators.

Main objective of the project is to analyse and visualize the child mortality data, to make people understand through visualization of graphs. We are trying to make an attempt to reach the people around the globe and to known facts, graphs and causes for the child mortality that has happened.

# 4. Operational Flow




The above figure shows the operational flow of our project. First we collected the data sets from the UNICEF .Org. In the second step, we have modified the datasets as per our requirements. In third step, we stored it in database, and wrote queries. From queries we visualised the data.

## 4.1 Data Collection

In this phase we have collected the datasets from the UNICEF Org. The datasets were in csv files. We have three datasets as “Infant mortality”, “Under-five mortality”, “Disease table”. https://data.unicef.org/topic/child-survival/under-five-mortality/.This is the link which we used to collect our datasets.

## 4.2 Data Pre-processing

In this phase we have rearranged the data as per the requirements. After that we have created database and wrote queries. Below are the screenshots for the structure of the tables.

## 4.3 Data Visualization

Visualization or visualisation is any technique for creating images, diagrams, or animations to communicate a message. Visualization through visual imagery has been an effective way to communicate both abstract and concrete ideas since the dawn of humanity. In this phase we have plotted different graphs using matplotlib and maps using high charts.

# 5. Snapshots and Results

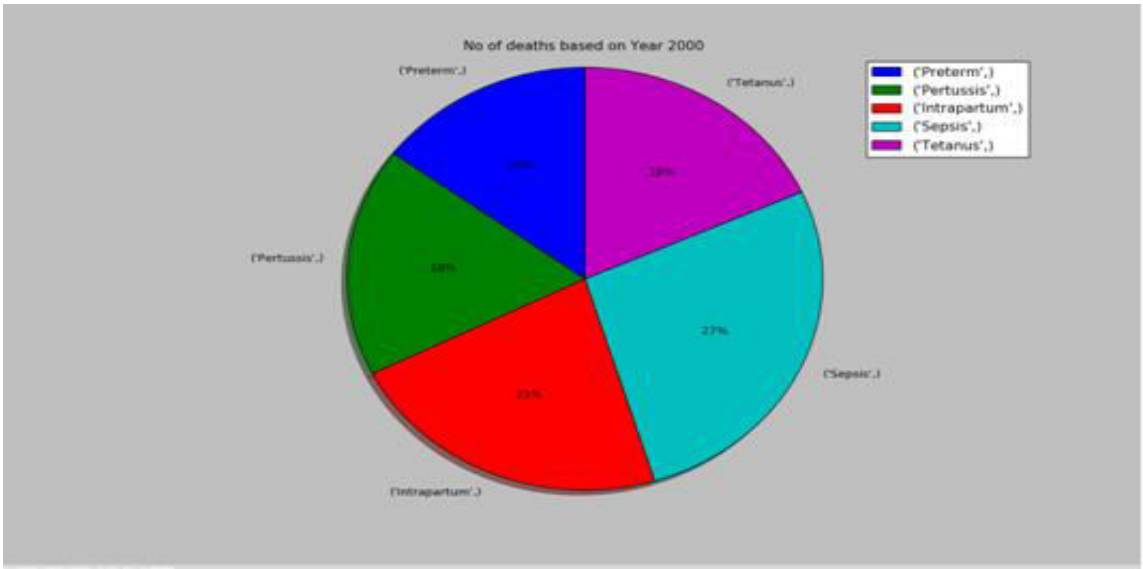



**Fig 1:** The above graph shows about number of death rates for 5 diseases in Year 2000

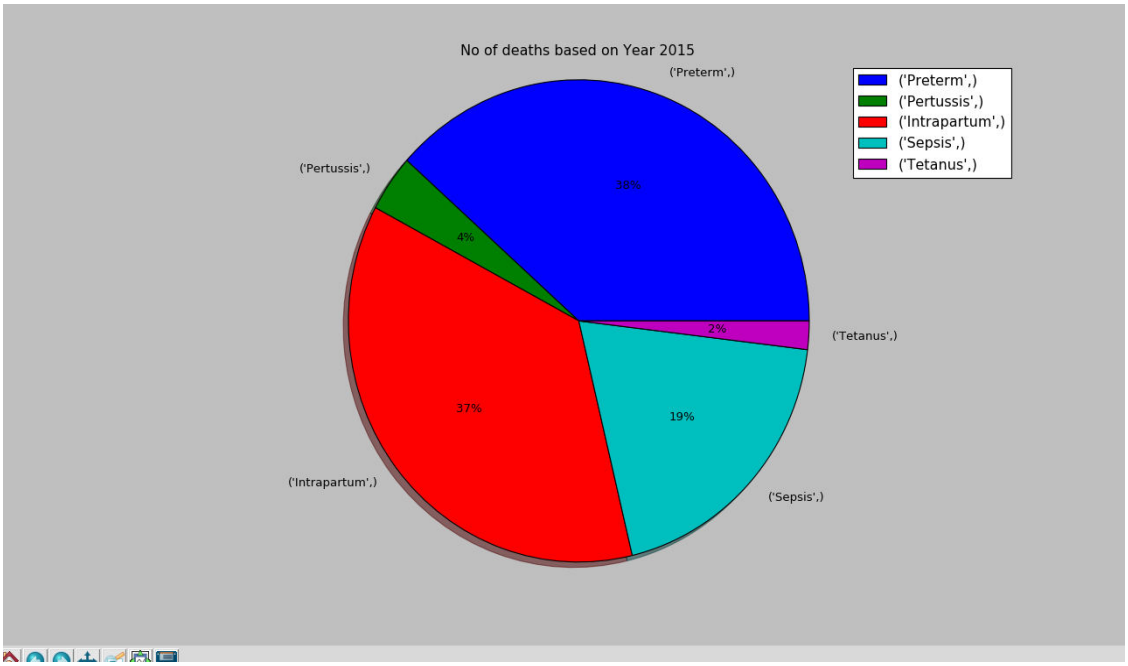



**Fig 2**: The above graph shows about number of death rates for 5 diseases in Year 2015

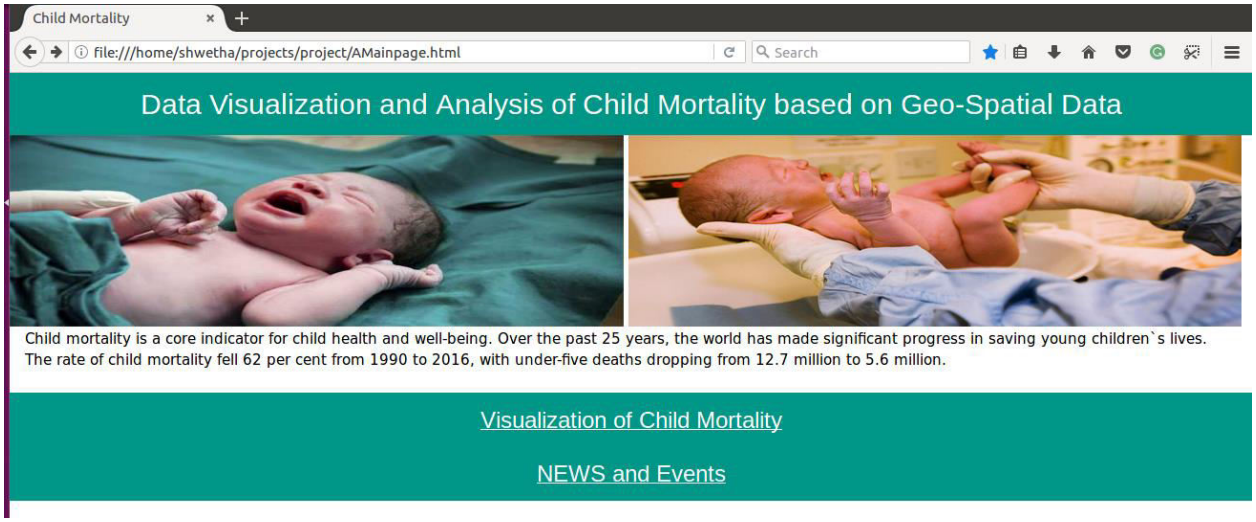



**Fig 3:** The first web page

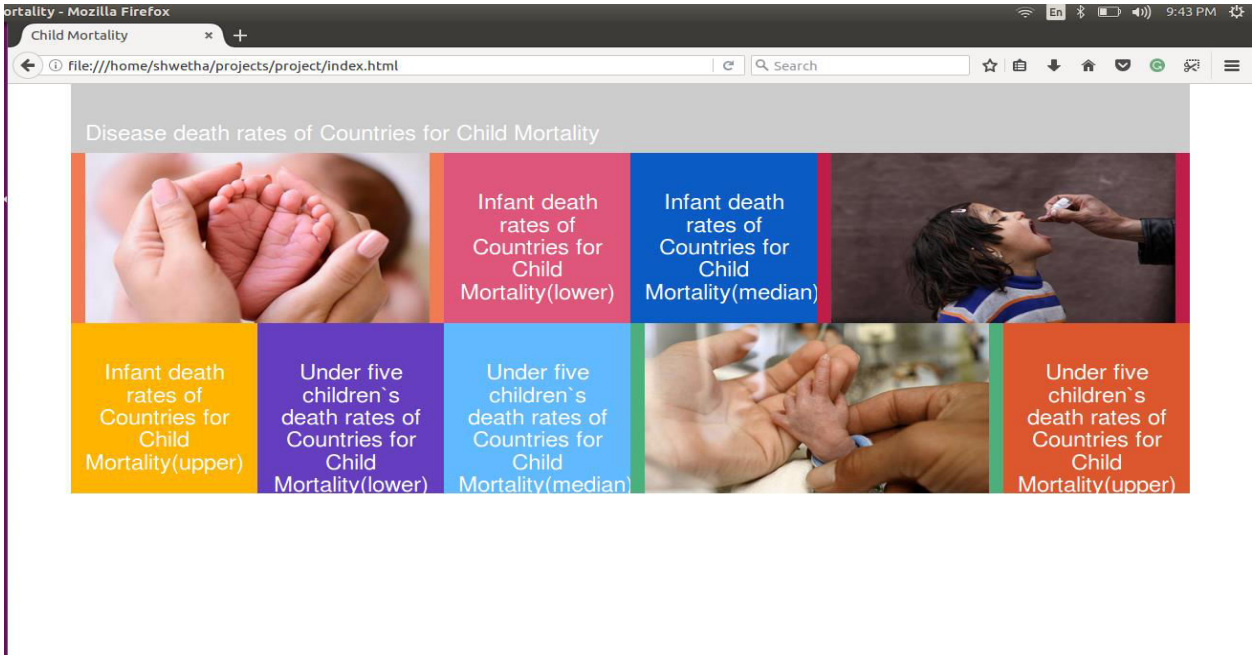



**Fig 4.** The Visualization page

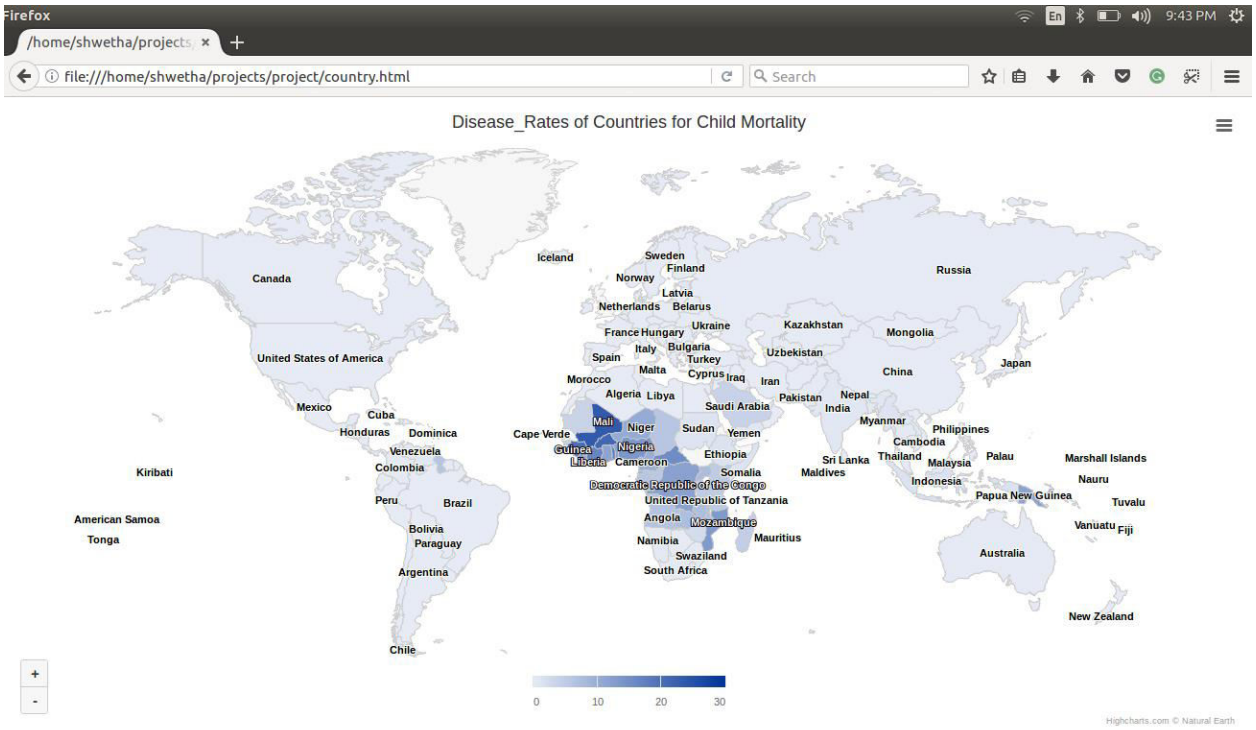



**Fig 5.** World map for Disease rates of countries for child mortality

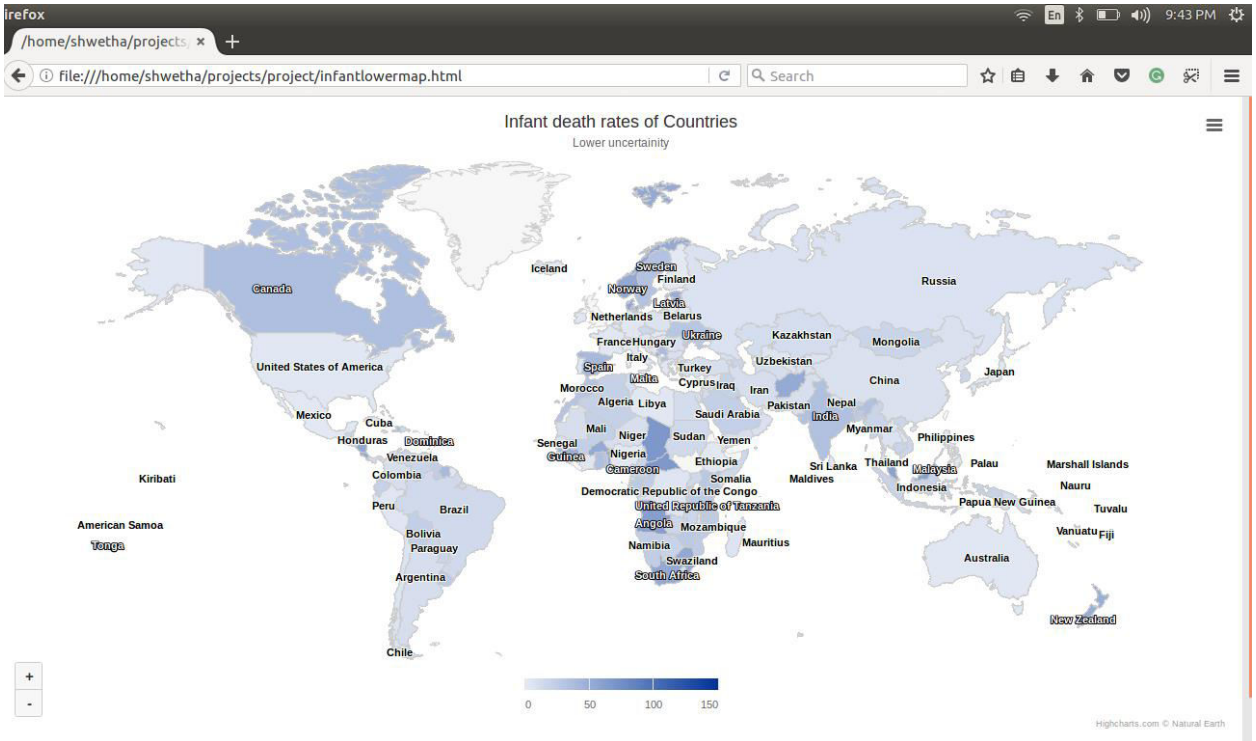



**Fig 6:** World map for Infant death rates of countries


**Fig 7:** Time series graph for Infant death rates at lower uncertainty

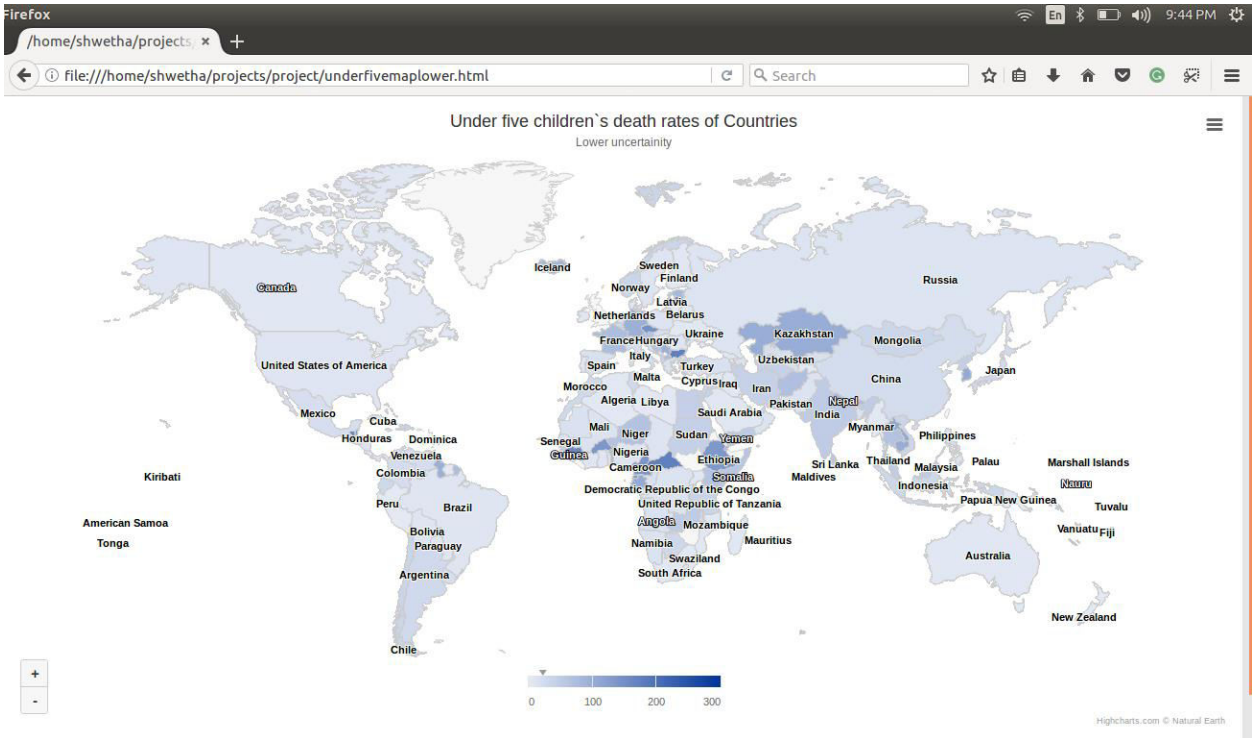



**Fig 8:** World Map for Under-five child death rates for countries

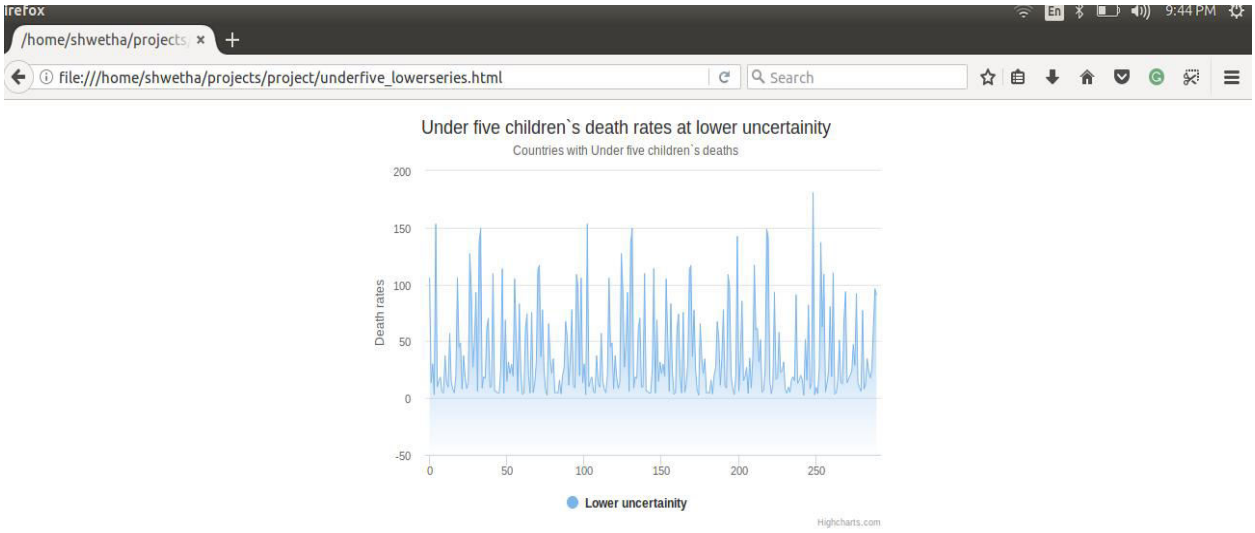



**Fig 9:** Time series graph for Under-five child death rates

# 6. Web Scraping

Web scraping (web harvesting or web data extraction) is data scraping used for extracting data from websites. Web scraping software may access the World Wide Web directly using the Hypertext Transfer Protocol, or through a web browser. While web scraping can be done manually by a software user, the term typically refers to automated processes implemented using a web crawler. It is a form of copying, in which specific data is gathered and copied from the web, typically into a central local database or spreadsheet, for later retrieval or analysis. Web scraping a web page involves fetching it and extracting from it. Fetching is the downloading of a page (which a browser does when you view the page). Therefore, web crawling is a main component of web scraping, to fetch pages for later processing. Once fetched, then extraction can take place. The content of a page may be parsed, searched, reformatted, its data copied into a spreadsheet, and so on. Web scrapers typically take something out of a page, to make use of it for another purpose somewhere else. An example would be to find and copy names and phone numbers, or companies and their URLs, to a list.

We have used Beautiful soup library to scrap the data in python. Beautiful Soup is a Python library for pulling data out of HTML and XML files. It works with your favourite parser to provide idiomatic ways of navigating, searching, and modifying the parse tree. It commonly saves programmers hours or days of work. We have done for “The Hindu” NEWS paper. Based on the key words like ‘mortality’, ’infant’, ’child’, ’under-five’, we scraped the front page and stored it a file.

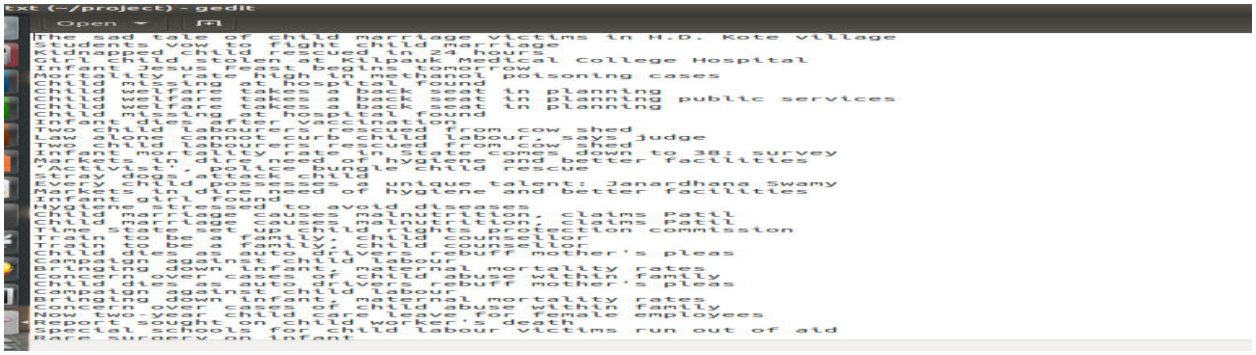


**Fig 10**: scrapped result

# 7. Plan of Action

| SL.No. | From Date | To Date | Work | Status |
|---|---|---|---|---|
| 1 | 28th August 2017 | 12th September 2017 | Requirement analysis related Dataset collection. | Done |
| 2 | 18th September 2017 | 24th September 2017 | Data pre-processing and required software installation. | Done |
| 3 | 25th September 2017 | 3rd October 2017 | Setting-up of project, Database creation. | Done |
| 4 | 4th October 2017 | 9th October 2017 | Code work and visualization of data by plotting graphs in python. | Done |
| 5 | 9th October 2017 | 10th October 2017 | Submission of 1st part of project. | Done |
| 6 | 19th October 2017 | 21st October 2017 | Up gradation in the obtained in terms of visualization for large dataset using time series. | Done |
| 7 | 22nd October 2017 | 28th October 2017 | Plotting of maps using high charts. | Done |
| 8 | 29th October 2017 | 5th November 2017 | Web Scrapping to infer the data. | In Progress |
| 9 | 6th November 2017 | | Final Submission of project. | |

# 8. Future Work

Web scrapping has to be done to infer the data that is to provide the probable reasons for an increase or decrease in the child mortality rate for a particular mortality point. The whole idea is to say why the death of a child has occurred in a country for a particular year. This would help the people to know the reasons for death as soon as they see our graphs or visualizations.

# References

[1]. https://data.unicef.org/topic/child-survival/under-five-mortality/

[2]. http://www.who.int/gho/child_health/en/

[3]. https://www.crummy.com/software/BeautifulSoup/bs4/doc/

[4]. https://www.highcharts.com/products/highcharts/

[5]. http://code.highcharts.com/mapdata/