# DS250 Project Report

*Akash Likhar (11840110), Rhythm Gupta (11840920), Shubham Arora (11841080)*

---

## **Aim**:

Analyse the data of different states and rank the districts (state wise) according to the points mentioned below, which will help in making observations regarding the need of organising health and child welfare programmes in various regions:

1. Household characteristics
2. Wealth index
3. Sex ratio
4. Literacy rate
5. School & work status
6. Age at time of marriage
7. Injury, disability, illness, personal habits
8. Fertility, abortion
9. Family planning practices
10. Child health after birth
11. Awareness on HIV/AIDS and some other diseases
12. Mortality

Also, given the outcome of pregnancy, predict the outcome of pregnancy for future cases (live birth, still birth, abortion).

## **Progress**:

1. **Data Collection:** (States selected - Uttarakhand, Chhattisgarh, Rajasthan)

   Source - Health Management Information System - Ministry of Health & Family Welfare

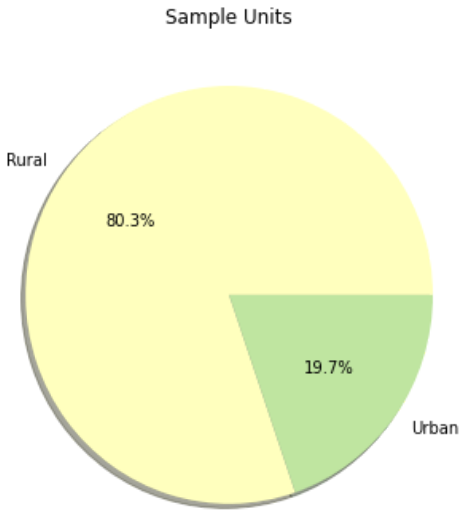
2. **Data Cleaning:**

   In surveys, a lot of questions are optional or depend on the answer of other questions and thus remain unfilled. This causes a lot of NaN values in the dataset. Also, there are 202 columns in the dataset and not all are required to analyse the data and predict the objective. Hence, the columns have been reduced to 101 after thoroughly analysing the data. The NaN values have also been taken care of.

3. **Visualisation:**

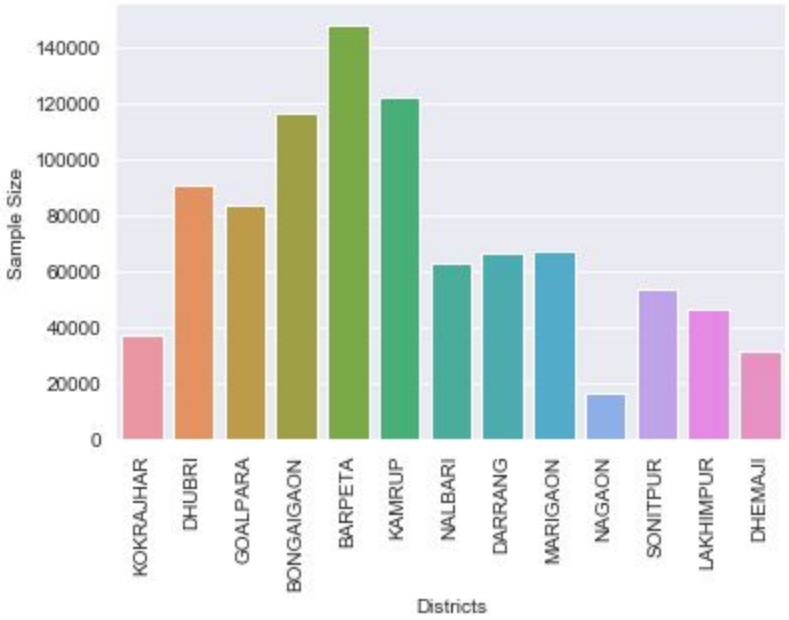
Some plots have been made to understand the dataset well, some of them are pasted below: (for **Uttarakhand**)

1.



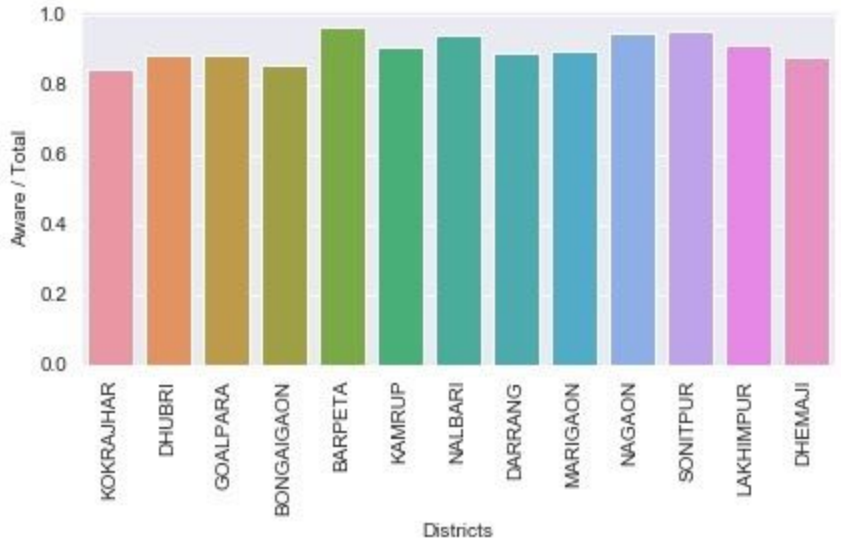

This pie chart shows the samples collected from rural and urban areas. As visible in the pie chart shown, most of the population which was interviewed belonged to rural areas.

2.




This bar graph shows the size of samples collected from each district in Uttarakhand.

3.




This plot shows the ratio of

(*person aware of HIV*)/(*total sample from that district*)

to show relative awareness of HIV/AIDS among districts of Uttarakhand.

## **Plans**:

1. **Assessing the economic profile of the households** using some of the characteristics. To calculate this value, a technique named Principal Component Analysis (PCA) will be used to convert all the characteristics into a single value, which will act as a wealth index.

2. **Visualizing relevant features of the dataset** to know about the conditions in various districts of that particular state. This will help in understanding the facilities that should be provided in a particular region. Moreover, after comparing with a relevant threshold value, the rank of a district regarding a particular observation can be found.

3. **Prediction of outcome of pregnancy** using neural networks.