# Problem Statement

---

In India, as per the latest National Family Health Survey (NFHS) 2015-16, 29.5% of women have experienced physical violence since age 15. Moreover, as per the latest National Crime Records Bureau (NCRB) report, an average of 87 rape cases were registered daily in India in 2019 (NCRB, 2020).
Violence against women is deep-rooted and has unequal power and gender relations (CARE International, 2018), it transcends social and economic boundaries, and affects women and girls of all socio-economic backgrounds. So, to create awareness about the enormity of the issue, the following dataset has been chosen.

**Link to Dataset:**
https://drive.google.com/file/d/1OEzc9pXfAcOZizJ8UtNWGP5tycEhrGhP/view?usp=sharing

The following tasks can be performed in any order

**Task 1:**

A dataset on crime against women is provided. Using this, analyze the data and find the state/union territory which is the safest (relative lower crime rate).

Analyse by what percentage the crimes have been reduced in a particular state or against the national average. Also, provide a suitable explanation for your analysis.

**Task 2:**

A new column must be added to the existing dataset named ‘Safety index’. This column will contain values based on how you interpret the safety of the district. For example, if the number of abductions in a district is greater than 10, then decrease the safety index by 2%. Ideally, a district with zero crime will have a safety index of 100%.

The participants can define the safety index with their own rule set. By coding (not manually) update the dataset with the safety index for each district every year.

**Task 3:**

This being the final task, you must develop a model to predict the safety of women in a particular location. Consider the district and the year as the input parameters. The safety index (from task 2) is expected as output. Your output will be tested against a range of results to measure the accuracy of your model.

Test the created predictive model using the following inputs and interpret the results.
District: 'BANGALORE COMMR.'
Year: '2021'

**Submission link:** https://cutt.ly/MahilAI-Datathon_Submission

**Judging Criteria:**
**Graded on a scale of 1-4 (1 being the lowest, 4 being the highest)**

**Impact:** Did the presenters articulate and interpret the problem statement correctly? Was the problem statement defined properly from the participant's point of view?

**Data visualization:** Did the presenters provide clear and aesthetically pleasing visuals such as graphs, tables, charts, maps, or infographics to help present their findings?

**Communication of findings:** Were the presenters able to articulate their findings and conclusions from their analysis?

**Innovation:** Did the presenters define their rule set in uniquely, was it an accurate definition?How innovative was your rule set definition?

**Vision:** Do the presenters have a clear vision of next steps: concepts for new technologies for development, or questions for further research?