# **Executive summary of Fetal Health Dataset**

This project is relevant for the decision makers who wants to analyze the life expectancy of people in a country such as world health organization to know about the factors or variables affecting the life expectancy of people.

The two questions that might be relevant to the decision maker are i) Does the data represent the information of various countries equally without any bias in data collection? ii) Does this dataset represents the pre-covid era and post-covid era records to know the present status of the life expectancy rates of different countries?

The key finding that might be relevant to the decision maker are i) There are no zero/ near zero variance producing variables in the data set, so one can include all the data. But for now, the Status and country were not included in the analysis and the predictors had too many NAs which were removed from the dataset. ii) The Root Mean Square Error of the models such as Random Forests, Boosted Trees, Generalised Additive Models is lower than other models with values 2.0670089, 2.0710561, 2.6664033 respectively. iii) Inclusion of Principal components as predictors might further reduce the interpretability of the sophisticated models such as Random Forest, Boosted Tree, Support Vector Machines, etc. Also, the multiple linear regression had high interpretability than all the other models if one must know about the predictor’s affect on the outcome variable.

Since the data is produced by the World Health Organization it has enough information of various countries however, the age bucket of the records for the data set is in the range of 33-89 years, so this data might not be accurate 100 percent for the prediction of life expectancy. The data collection might not be biased as it is taken from the repository of world health organization(WHO). This Data does just represent information related to the years between 2000-2015. So, it does not contain pre-pandemic i.e., data between 2018- 2020 and post-pandemic era i.e., 2021-2022. Thus, the dataset is not useful for predicting the current life expectancy rates, however it might be helpful for getting to know the process of prediction and perform them on the current data.