**Collect the Dataset**

There are many popular open sources for collecting the data. Eg: kaggle.com, UCI repository, etc.   
In this project we have used .csv data. This data is downloaded from kaggle.com. Please refer to the link given below to download the dataset.   
Link: <https://www.kaggle.com/datasets/andrewmvd/fetal-health-classification>As the dataset is downloaded. Let us read and understand the data properly with the help of some visualization techniques and some analyzing techniques.   
Note: There are a number of techniques for understanding the data. But here we have used some of it. In an additional way, you can use multiple techniques. 

**Importing the libraries**

Import the necessary libraries as shown in the image.



Our dataset format might be in .csv, excel files, .txt, .json, etc. We can read the dataset with the help of pandas.

In pandas we have a function called read\_csv() to read the dataset. As a parameter we have to give the directory of the csv file.



**Data Preparation**

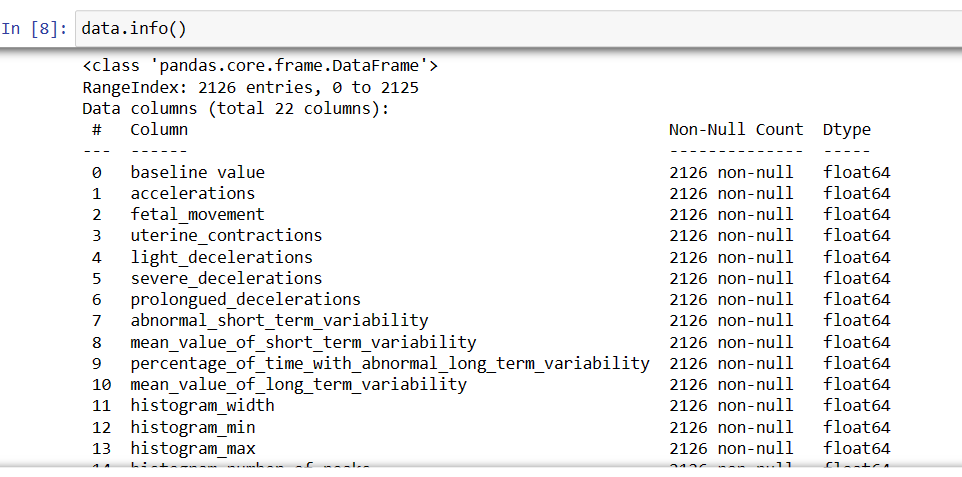
As we have understood how the data is, let us pre-process the collected data.

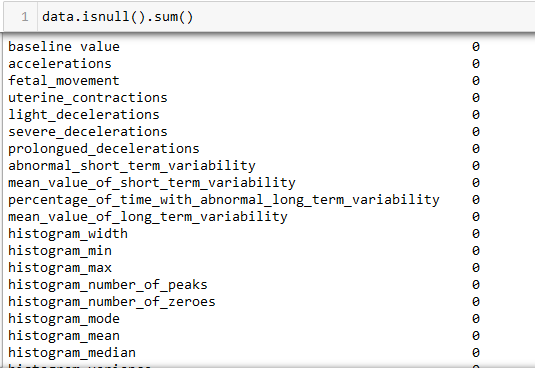
The Machine Learning model cannot be trained on the imported data directly. The dataset might have randomness, we might have to clean the dataset and bring it in the right form. This activity involves the following steps:

* Handling Missing Values
* Handling Categorical Data
* Handling Imbalance Data

Note: These are the general steps of pre-processing the data before using it for machine learning. Depending on the condition of your dataset, you may or may not have to go through all these steps.

**Handling Missing Values:**



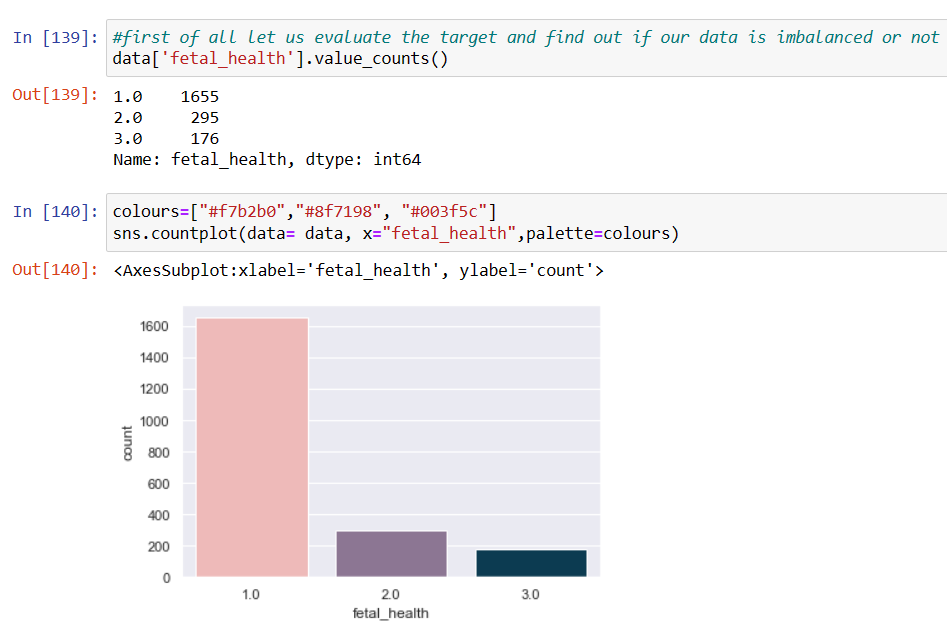


There are no missing values in the dataset. That is why we can skip this step.

**Handling Categorical Data.**

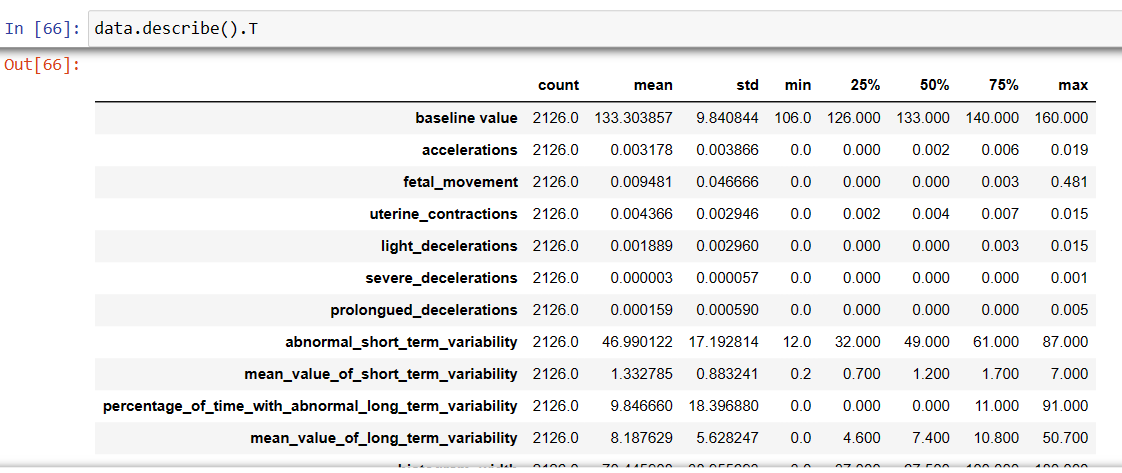
There are no categorical values in the dataset. That is why we can skip this step.

**Handling Imbalance Data**

****After checking, we get to know that the dataset is highly imbalanced. So in the later stages we have balanced the dataset before training the model.

**Descriptive Statistics**

Descriptive analysis is to study the basic features of data with the statistical process. Here pandas has a worthy function called describe. With this describe function we can understand the unique, top and frequent values of categorical features. And we can find mean, std, min, max and percentile values of continuous features.



### Visual Analysis

Visual analysis is the process of using visual representations, such as charts, plots, and graphs, to explore and understand data. It is a way to quickly identify patterns, trends, and outliers in the data, which can help to gain insights and make informed decisions. 

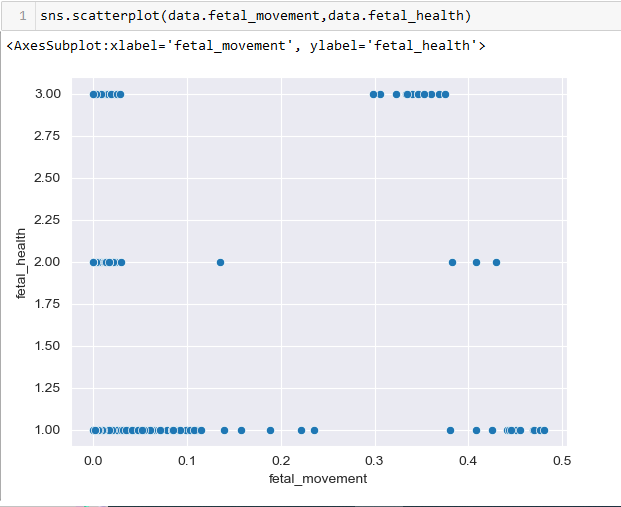
**Univariate analysis**

In simple words, univariate analysis is understanding the data with a single feature. Here we have displayed different graphs such as histogram and boxplot.

The Seaborn and matplotlib package provides a wonderful functions histogram and boxplot. With the help of histogram and boxplot, we can find the distribution of the feature. To make multiple graphs in a single plot, we use subplot.

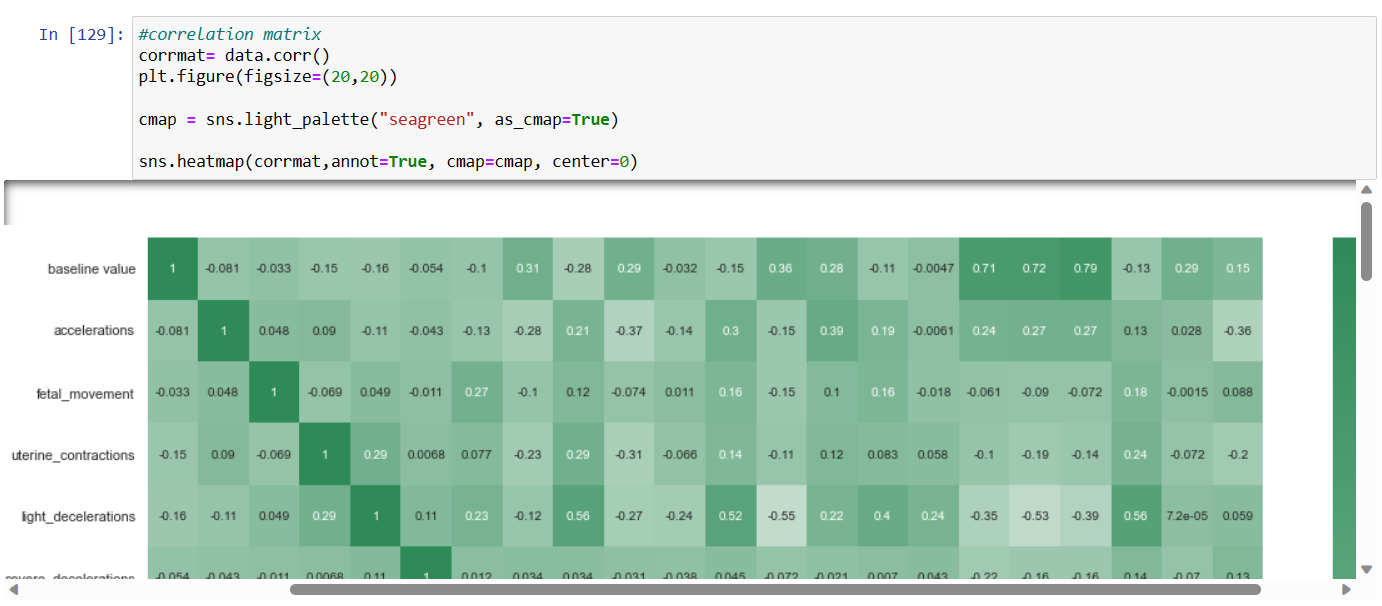


**Bivariate Analysis:**

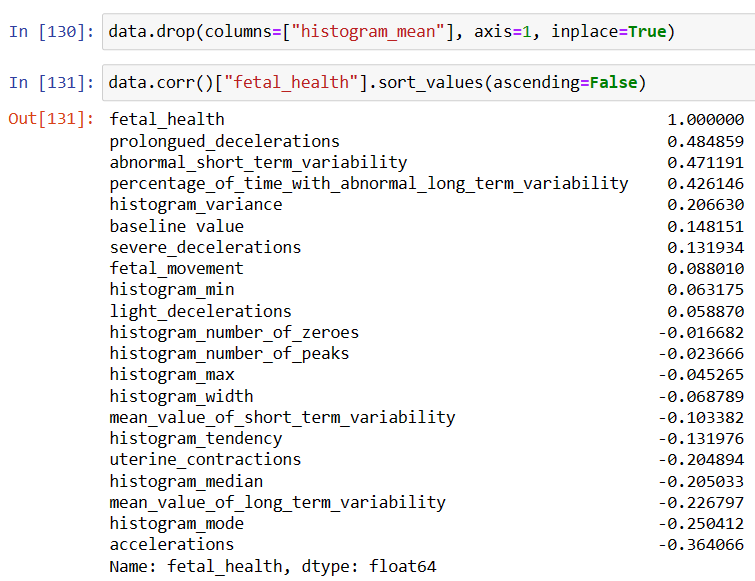


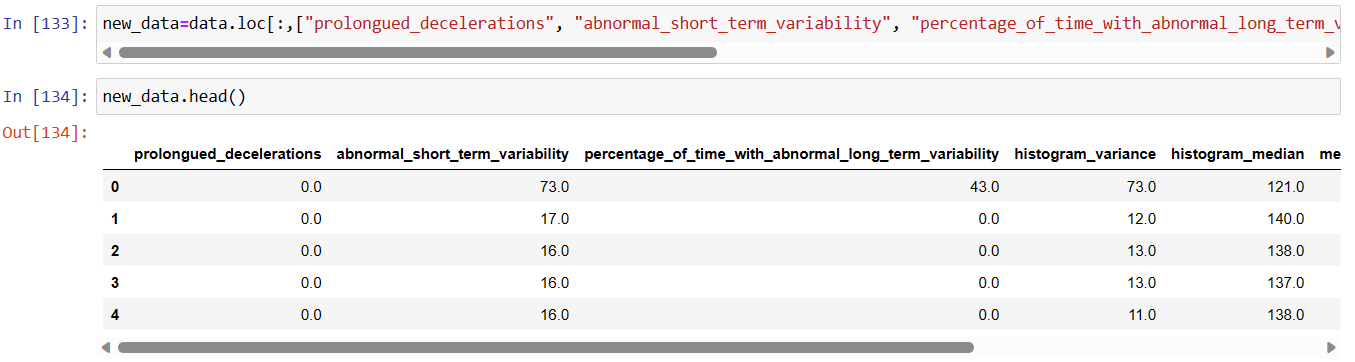
**Multivariate analysis**

In simple words, multivariate analysis is to find the relation between multiple features. Here we have used correlation matrix.

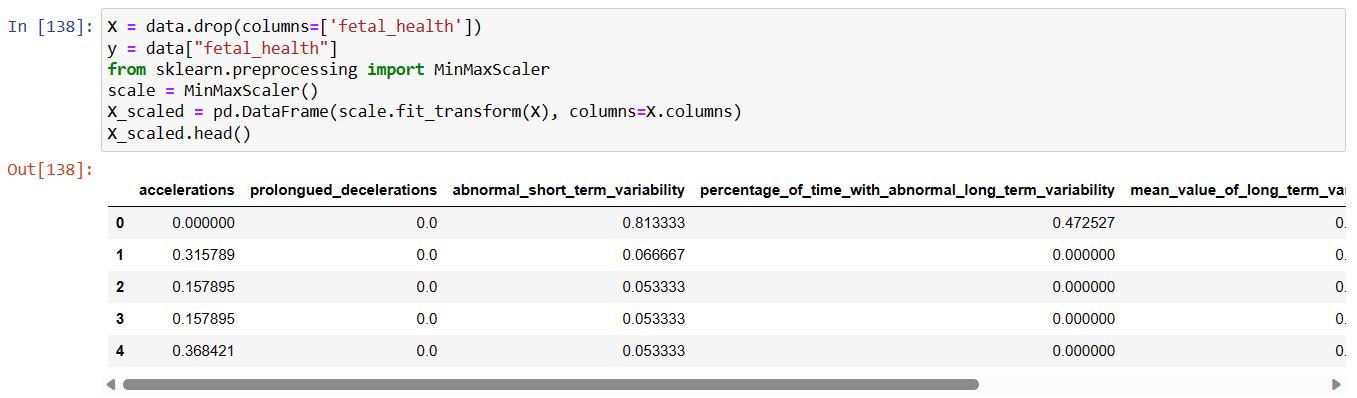


**Feature Selection**

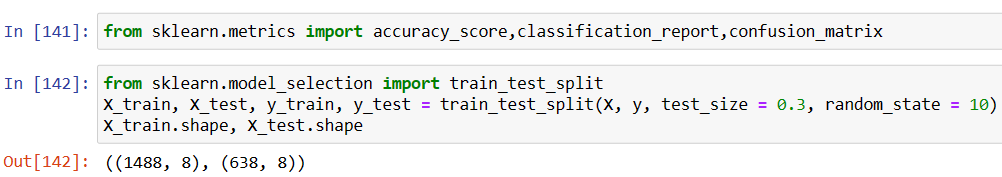
****

****

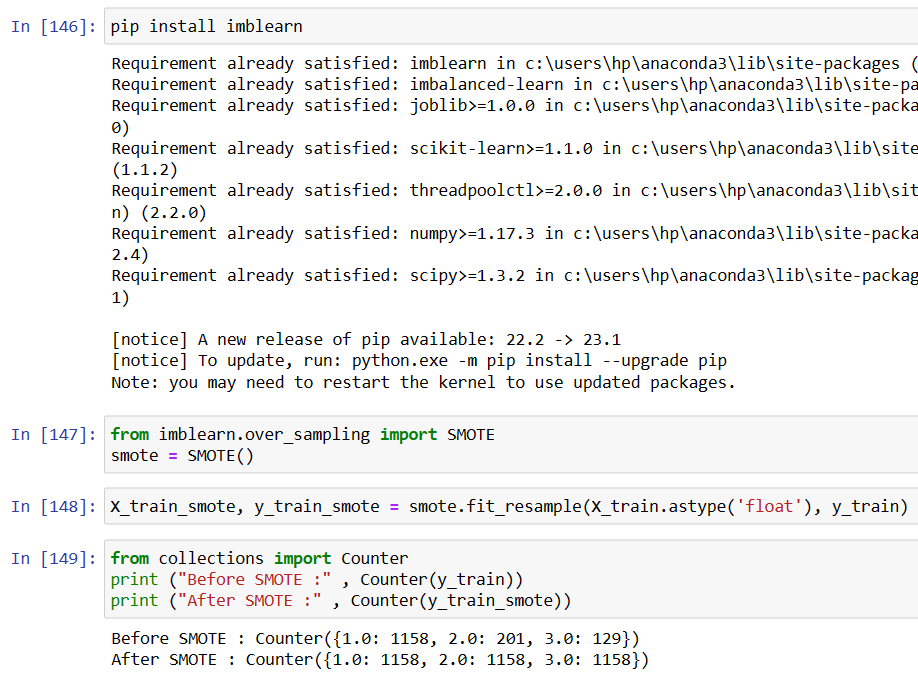
**Scaling Data:**

****

**Splitting data into train and test**



**Applying SMOTE for balancing the data**



After applying SMOTE, the dataset is balanced. And now we will train the model after balancing the dataset to check the accuracy.