Problem Statement:
Anova Insurance, a global health insurance company, seeks to optimize its insurance policy premium pricing based on the health status of applicants. Understanding an applicant's health condition is crucial for two key decisions:
- Determining eligibility for health insurance coverage.
- Deciding on premium rates, particularly if the applicant's health indicates higher risks.

Your objective is to Develop a predictive model that utilizes health data to classify individuals as 'healthy' or 'unhealthy'. This classification will assist in making informed decisions about insurance policy premium pricing.

Overview
The dataset contains 9549 rows and 20 columns (original data without preprocessing), the no. of columns becomes 23 post preprocessing because of encoding, the 23 columns includes both numerical and categorical variables. Here is the data dictionary.

- Age: Represents the age of the individual. Negative values seem to be present, which might indicate data entry errors or a specific encoding used for certain age groups.

- BMI (Body Mass Index): A measure of body fat based on height and weight. Typically, a BMI between 18.5 and 24.9 is considered normal.

- Blood_Pressure: Represents systolic blood pressure. Normal blood pressure is usually around 120/80 mmHg.

- Cholesterol: This is the cholesterol level in mg/dL. Desirable levels are usually below 200 mg/dL.

- Glucose_Level: Indicates blood glucose levels. It might be fasting glucose levels, with normal levels usually ranging from 70 to 99 mg/dL.

- Heart_Rate: The number of heartbeats per minute. Normal resting heart rate for adults ranges from 60 to 100 beats per minute.

- Sleep_Hours: The average number of hours the individual sleeps per day.

- Exercise_Hours: The average number of hours the individual exercises per day.

- Water_Intake: The average daily water intake in liters.

- Stress_Level: A numerical representation of stress level.

- Target: This is a binary outcome variable, with '1' indicating 'Unhealthy' and '0' indicating 'Healthy'.

- Smoking: A categorical variable indicating smoking status. Contains values - (0,1,2) which specify the regularity of smoking with 0 being no smoking and 2 being regular smmoking.

- Alcohol: A categorical variable indicating alcohol consumption status. Contains values - (0,1,2) which specify the regularity of alcohol consumption with 0 being no consumption quality and 2 being regular consumption.

- Diet: A categorical variable indcating the quality of dietary habits. Contains values - (0,1,2) which specify the quality of the habit with 0 being poor diet quality and 2 being good quality.

- MentalHealth: Possibly a measure of mental health status. Contains values - (0,1,2) which specify the severity of the mental health with 0 being fine and 2 being highly severe

- PhysicalActivity: A categorical variable indicating levels of physical activity. Contains values - (0,1,2) which specify the instensity of the medical history with 0 being no Physical Activity and 2 being regularly active.

- MedicalHistory: Indicates the presence of medical conditions or history. Contains values - (0,1,2) which specify the severity of the medical history with 0 being nothing and 2 being highly severe.

- Allergies: A categorical variable indicating allergy status. Contains values - (0,1,2) which specify the severity of the allergies with 0 being nothing and 2 being highly severe.

- Diet_Type: Categorical variable indicating the type of diet an individual follows. Contains values(Vegetarian, Non-Vegetarian, Vegan).
- (this column has been encoded into three different columns during the preprocessing stage)
 - Diet_Type_Vegan,Diet_Type_Vegetarian

- Blood_Group: Indicates the blood group of the individual Contains values (A, B, AB, O), this column values are encoded too .