Dataset Description

**Description**

This dataset contains medical and lifestyle information for 1500 patients, designed to predict the presence of cancer based on various features. The dataset is structured to provide a realistic challenge for predictive modeling in the medical domain.

**Dataset Structure**

**Features**

1. **Age**: Integer values representing the patient's age, ranging from 20 to 80.
2. **Gender**: Binary values representing gender, where 0 indicates Male and 1 indicates Female.
3. **BMI**: Continuous values representing Body Mass Index, ranging from 15 to 40.
4. **Smoking**: Binary values indicating smoking status, where 0 means No and 1 means Yes.
5. **GeneticRisk**: Categorical values representing genetic risk levels for cancer, with 0 indicating Low, 1 indicating Medium, and 2 indicating High.
6. **PhysicalActivity**: Continuous values representing the number of hours per week spent on physical activities, ranging from 0 to 10.
7. **AlcoholIntake**: Continuous values representing the number of alcohol units consumed per week, ranging from 0 to 5.
8. **CancerHistory**: Binary values indicating whether the patient has a personal history of cancer, where 0 means No and 1 means Yes.
9. **Diagnosis**: Binary values indicating the cancer diagnosis status, where 0 indicates No Cancer and 1 indicates Cancer.

**Target Variable**

**Diagnosis**: The main variable to predict, indicating if a patient has cancer.