**PROJECT ON “HEALTHCARE (CHD-Heart Disease)”**

**For**

**AI-ML**



**Name:** Arshpreet Singh

**Course Name :** AI-ML

**Roll No. :** 2210991370

**Table of Contents**

1. **Introduction**
2. **Dataset**
   1. **Sources**
   2. **Attributes**
   3. **Factors**
3. **Data-Shape**
4. **Describing The Dataset**
5. **Problem Statements**
   1. **What is the percentage of patient with or without heart problems in the given dataset ?**
   2. **What is the sex Wise ratio of Coronary Heart Disease Patients?**
   3. **Show the Heart Disease Frequency with respect to age-groups.**
   4. **Show Heart Disease Frequency with respect to Sex.**
   5. **What is the Heart diseases Data according to Fasting Blood sugar?**
   6. **Analyses of the chest pain ?**
   7. **Analyses of the Resting electrocardiographic measurement in patients**
   8. **Analysing number of major vessels (0-3) colored by flourosopy**
   9. **Analysing A blood disorder called thalassemia**

**CHD (Coronary Heart Disease) Case Study and Analysis.**

1. **INTRODUCTION:**

Heart disease refers to various types of conditions that can affect heart function. These types include:

* **Coronary artery (atherosclerotic) disease**: Disease that affects the blood supply to the heart.
* **Valvular heart disease**: Disease that affects how the valves function to regulate blood flow in and out of the heart.
* **Cardiomyopathies**: Disease that affect heart muscles, heart rhythm disturbances
* **Arrhythmias**: Disease that affect the electrical conduction and congenital heart diseases where the heart has structural problems that develop before birth.

Heart disease is the major cause of morbidity and mortality globally: it accounts for more deaths annually than any other cause. For example:

an estimated 17.9 million people died from heart diseases in 2016, representing 31% of all global deaths. Over three quarters of these deaths took place in low- and middle-income countries.

Most heart diseases are highly preventable and simple lifestyle modifications(such as reducing tobacco use, eating healthily, obesity and exercising) coupled with early treatment greately improve their prognoses.

It is, however, difficult to identify high risk patients because of the mulfactorial nature of several contributory risk factors such as diabetes, high blood pressure, high cholesterol etc. Due to such constraints, scientists have turned towards modern approaches like Data Mining and Machine Learning for predicting the disease.

Machine learning (ML), due to its superiority in pattern detection and classification, proves to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry on heart disease.

In this notebook, I will be exploring different Machine Learning approaches for predicting wheather a patient has 10-year risk of developing coronary heart disease (CHD) using the Framingham dataset that is publicly availabe on Kaggle.

1. **Dataset:**

**2.1 Source-**

The dataset is publically available on kaggle and it is from an ongoing cardiovascular study.The dataset provides the patients’ information.The dataset provides the patients’ information.

The classification goal is to predict whether the patient has 10-year risk of future coronary heart disease (CHD).

* 1. **Attributes:**

1. **age:** The person's age in years
2. **sex:** The person's sex (1 = male, 0 = female)
3. **cp:** The chest pain experienced (

Value 1: typical angina,

Value 2: atypical angina,

Value 3: non-anginal pain,

Value 4: asymptomatic)

1. **trestbps**: The person's resting blood pressure (mm Hg on admission to the hospital)
2. **chol:** The person's cholesterol measurement in mg/dl
3. **fbs:** The person's fasting blood sugar (> 120 mg/dl, 1 = true; 0 = false)
4. **restecg**: Resting electrocardiographic measurement (

0 = normal,

1 = having ST-T wave abnormality,

2 = showing probable or definite left ventricular hypertrophy by Estes' criteria)

1. **thalach:** The person's maximum heart rate achieved
2. **exang:** Exercise induced angina (1 = yes; 0 = no)
3. **oldpeak:** ST depression induced by exercise relative to rest ('ST' relates to positions on the ECG plot)
4. **slope:** the slope of the peak exercise ST segment (Value 1: upsloping, Value 2: flat, Value 3: downsloping)
5. **ca:** The number of major vessels (0-3)
6. **thal:** A blood disorder called thalassemia (3 = normal; 6 = fixed defect; 7 = reversable defect)
7. **target:** Heart disease (0 = no, 1 = yes)
   1. FACTORS

Heart disease risk factors to the following:

* High cholesterol
* High blood pressure
* Diabetes
* Weight
* Family history
* Smoking

According to another source , the major factors that can't be changed are:

* Increasing age
* Male gender
* Heredity.
* thalassemia, one of the variables in this dataset, is Heredity.

Major factors that can be modified are:

* Smoking
* High cholesterol
* High blood pressure
* Physical inactivity
* Being overweight
* Having Diabetes.

Other factors include stress, alcohol and poor diet/nutrition.

1. **DATA-SHAPE**

“**data.shape**” is used to get the dimensions of the DataFrame or Series. It returns a tuple representing the dimensionality of the DataFrame or Series. The returned tuple consists of two values:

* The first value is the number of rows.
* The second value is the number of columns.

For example, if you have a DataFrame data with 5 rows and 3 columns, data.shape will return (5, 3)

1. **Describing the Dataset**

**“data.describe()**” is used to generate descriptive statistics of a DataFrame or Series. It provides a summary of the central tendency, dispersion, and shape of the distribution of a dataset, excluding NaN values.

what it returns for a DataFrame:

* **count:** Number of non-null observations.
* **mean:** Mean of the values.
* **std:** Standard deviation of the observations.
* **min:** Minimum of the values in the object.
* **25%:** The first quartile or 25th percentile.
* **50%:** The second quartile or median or 50th percentile.
* **75%:** The third quartile or 75th percentile.
* **max:** Maximum of the values in the object.

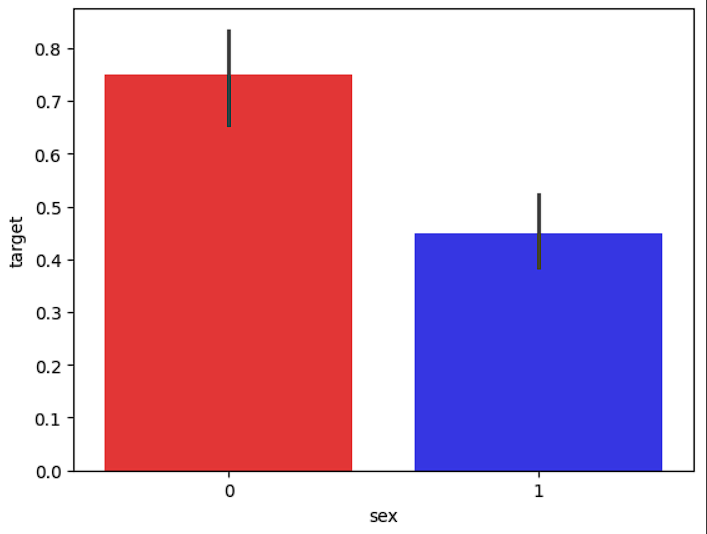
1. **PROBLEM STATEMENTS**

5.1 **What is the percentage of patient with or without heart problems in the given dataset ?**

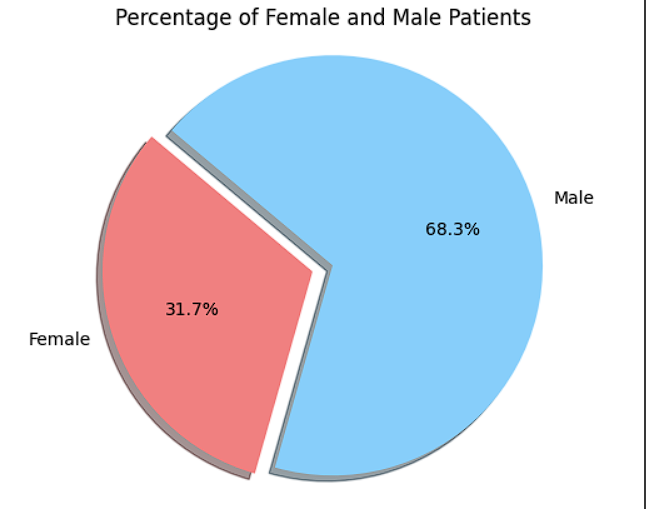
According to our Analysis,

From the total dataset of 303 patients, 165 (54%) have a heart disease (target=1)

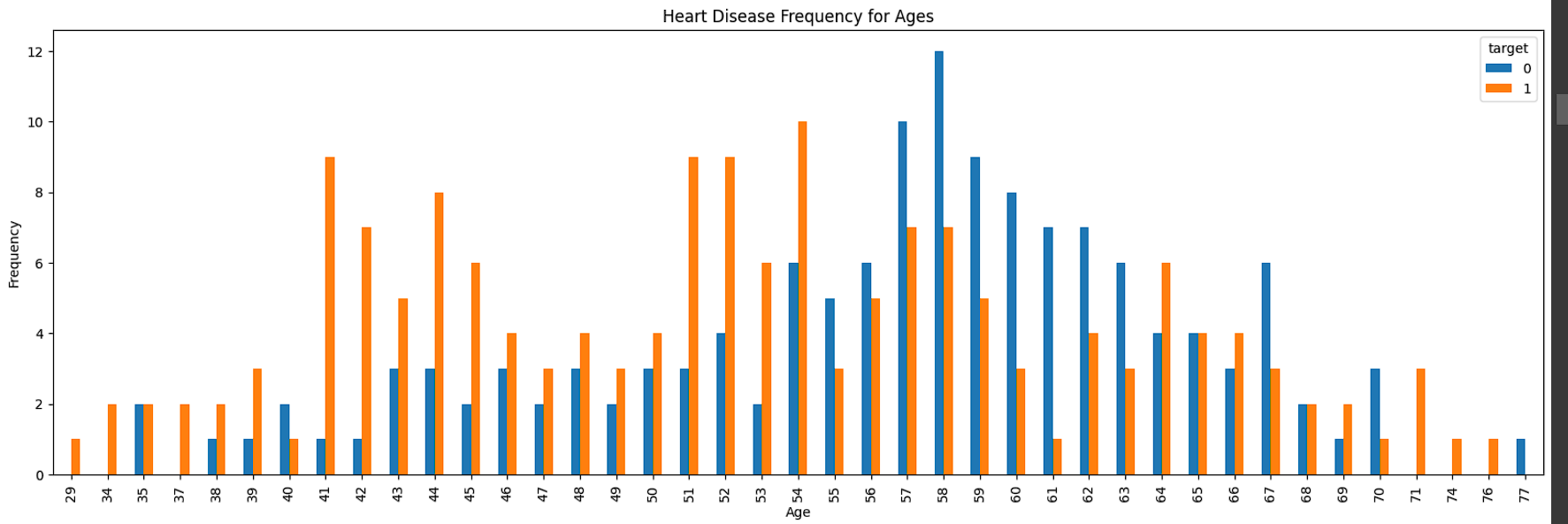
* Percentage of patients without heart problems: 45.54%
* Percentage of patients with heart problems: 54.46%



* 1. **What is the sex Wise ratio of Coronary Heart Disease Patients?**
* Percentage of Female Patients:31.68%
* Percentage of Male Patients:68.32%

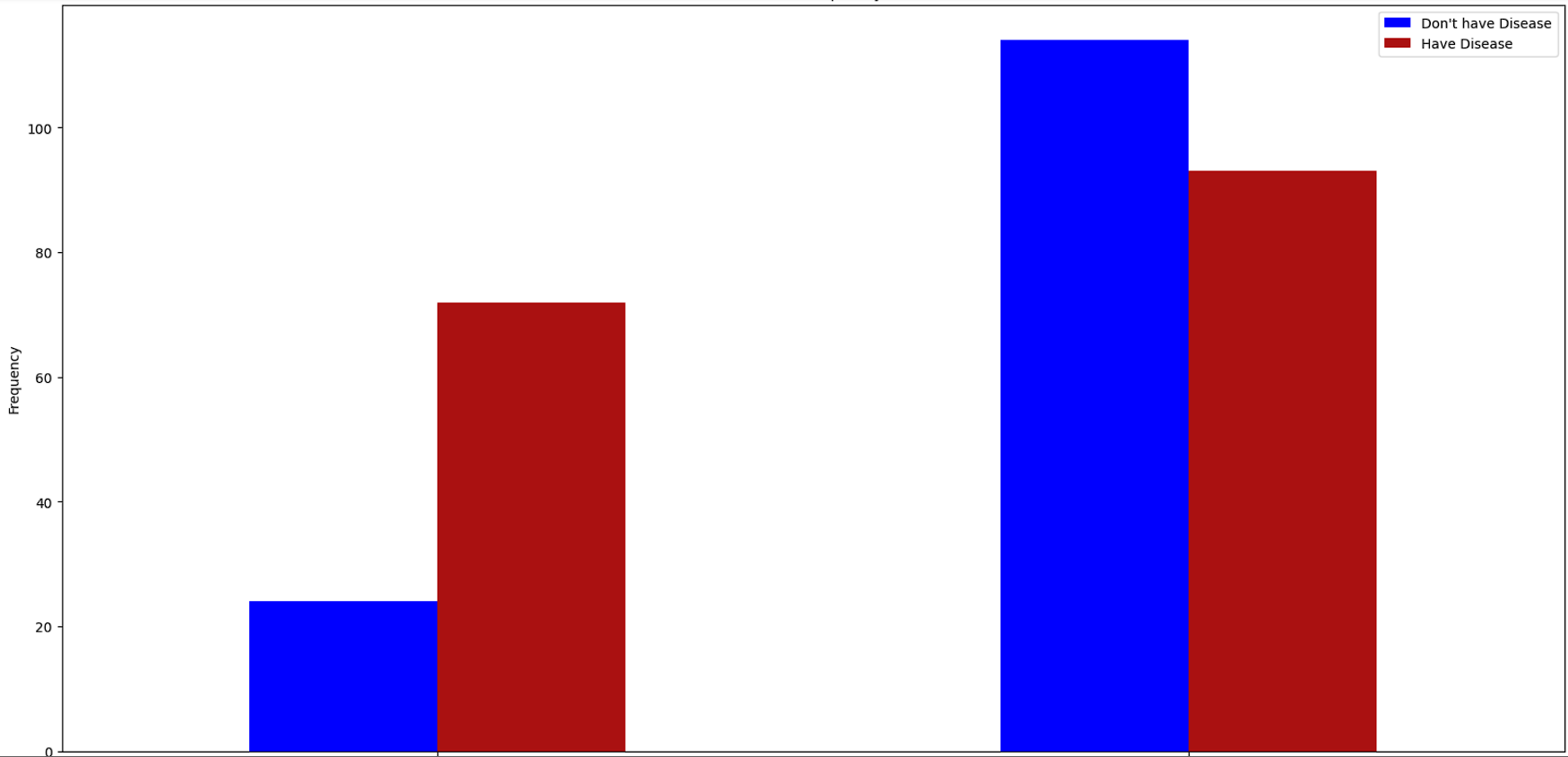


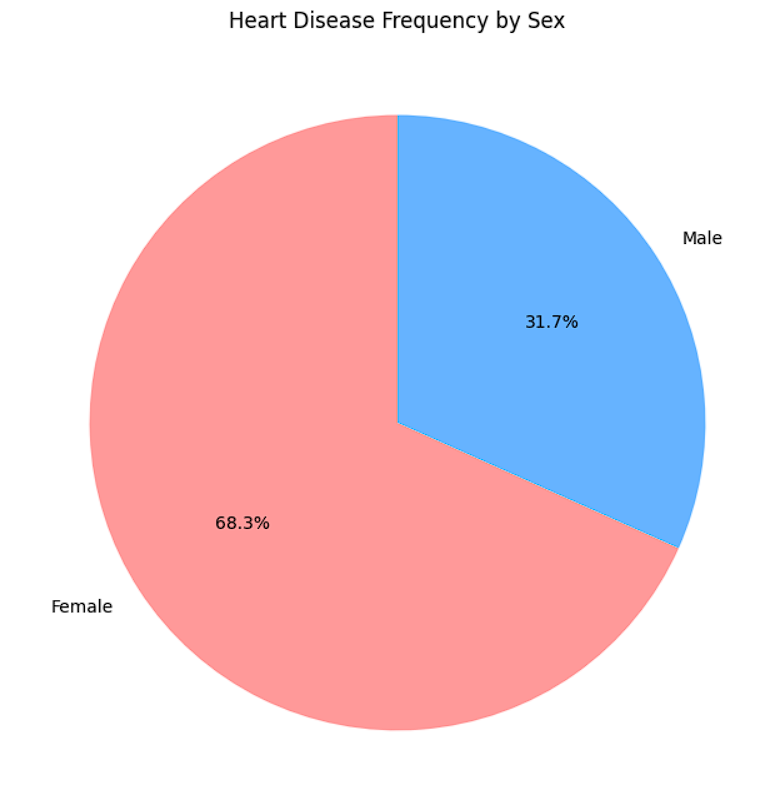
* 1. **Show the Heart Disease Frequency with respect to age-groups.**



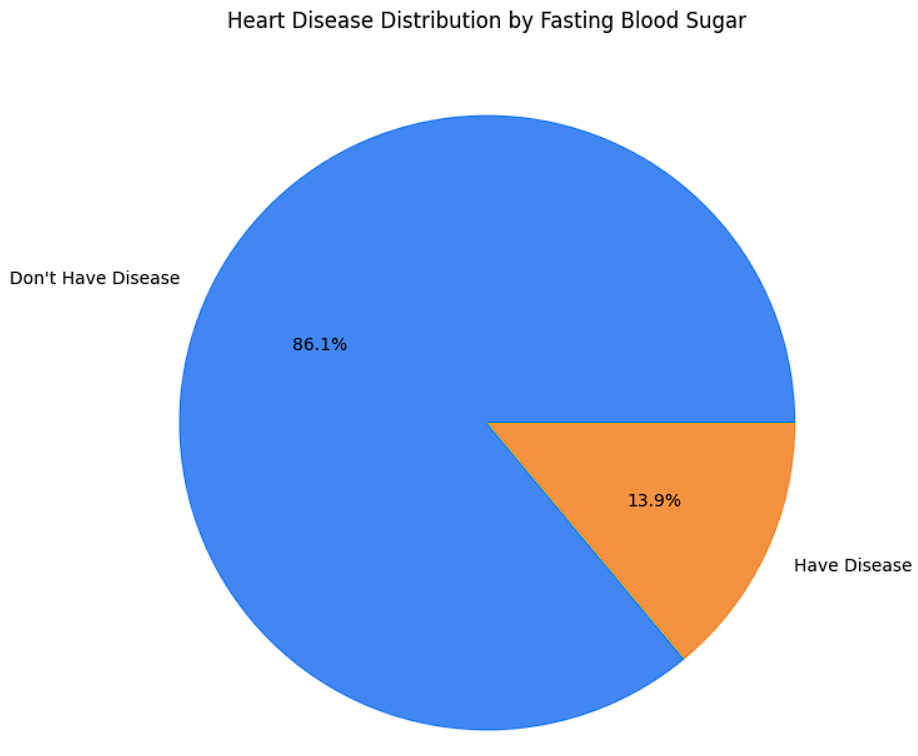
Heart Disease frequency for sex (where 0 is female and 1 is male and "red" is have heart disease and "blue" is don't have heart disease)

* 1. **Show Heart Disease Frequency with respect to Sex.**





* 1. **What is the Heart diseases Data according to Fasting Blood sugar?**



* 1. **Analyses of the chest pain ?**

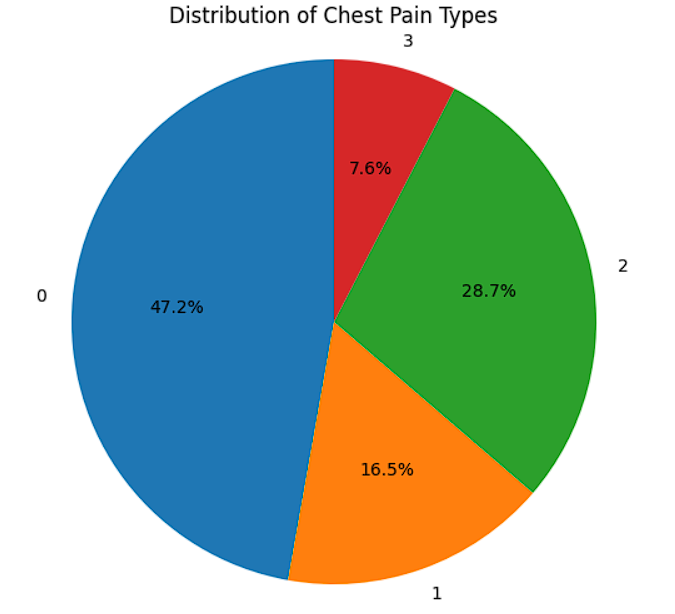
4 types of chest pain

1: Typical angina

2: Atypical angina

3: Non-anginal pain

4: Asymptomatic

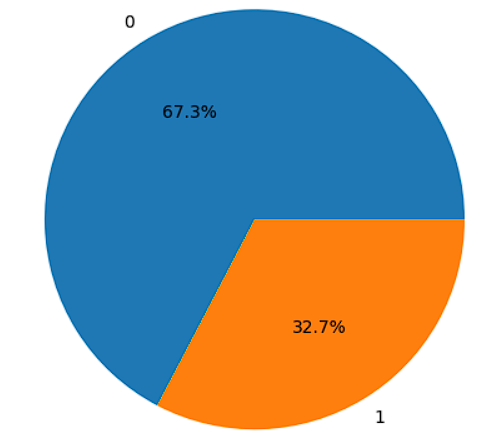
****

* 1. **Analyses of the Resting electrocardiographic measurement in patients**

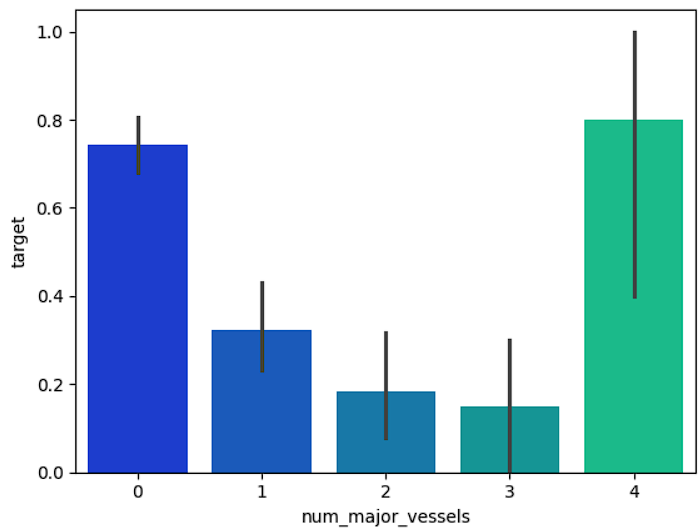
0 = normal

1 = having ST-T wave abnormality

2 = showing probable or definite left ventricular hypertrophy by Estes' criteria



* 1. **Analysing number of major vessels (0-3) colored by flourosopy**

****

* 1. **Analysing A blood disorder called thalassemia**

3 = normal

6 = fixed defect

7 = reversable defect

