Statement of Purpose

Data Description

Industry: Heart Diseases

Application: Statistical Analysis and Visualization

Tools/ Technologies required: Excel, Jupyter Notebook, GitHub, Statistical Data Analysis, Exploratory Data Analysis, Data Pre-processing

Dataset: heart.csv

No.	Column Names	Description/ Comments
1	Age	Age of the Patient
2	Sex	The gender of the Patient • 1=male • 0=female
3	Cp	Chest pain type • 0: typical angina • 1: atypical angina • 2: non angina pain • 3: asymptomatic
4	trestbps	Resting blood pressure (in mm Hg admission to the hospital)
5	Chol	serum cholesterol in mg/dl
6	Fbs	(Fast blood sugar>120 mg/dl) • 1=true • 0=false
7	Restecg	Resting electrocardiographic results • 0: normal • 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) • 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
8	Thalach	Maximum heart rate achieved
9	Exang	exercise induced angina • 1=yes • 0=no
10	Oldpeak	ST depression induced by exercise relative to rest
11	Slope	the slope of the peak exercise ST segment • 0: upsloping • 1: flat • 2: downsloping

12	Ca	number of major vessels (0-3) colored by flourosopy
13	thal	• 0 = error (in the original dataset 0 maps to NaN's) • 1 = fixed defect • 2 = normal • 3 = reversable defect
14	target	• 0 = no disease • 1 = disease

Problem Statement

"Heart disease is term used for diseases and conditions affecting the heart and circulatory system. Since heart is amongst the most vital organs of the body, its diseases affect other organs and part of the body as well. There are several different types and forms of heart diseases. The most common ones cause narrowing or blockage of the coronary arteries, malfunctioning in the valves of the heart, enlargement in the size of heart and several others leading to heart failure."

Task

1. Explore the heart disease dataset using exploratory data analysis (EDA)
2. Make use of classification algorithms such as decision tree, KNN, logistic regression etc., for prediction (modelling)
3. Perform visualization of data set using graphs and tableau.
4. Build a Regression model to predict the possibility of heart failure.

Now use this model to predict whether the first patient has a possibility of heart failure

No	Column Names
1	Age = 52
2	Sex = 1
3	Cp = 0
4	Trestbps =125
5	Chol = 212
6	Fbs = 0
7	Restecg =1
8	Thalach = 168
9	Exang = 0
10	Oldpeak = 1
11	Slope = 2
12	Ca = 2
13	Thal = 3
14	Target = 0