# Assigment0

Efstratios Ioannidis tem2505

8 October 2024

## 1 Introduction

This is the report for assigment 0 of HY-577 Machine Learning. All the calculation comes from the dataset "heart.csv" in my file.

### 1.1 Data Manipulation and Operations using Pandas

#### 1.1.2 Minimum Value of the whole dataframe

Mean value of:
- age = 29.0
- sex = 0.0
- cp = 0.0
- trestbps = 94.0
- chol = 126.0
- fbs = 0.0
- restecg = 0.0
- thalachh = 71.0
- exang = 0.0
- oldpeak = 0.0
- slope = 0.0
- ca = 0.0
- thal = 0.0
- hasHeartDisease = 0.0

#### 1.1.3 Feature with the highest mean

The feature with the highest mean is 'chol' with a mean value of 246.26402640264027.

#### 1.1.4 Standard deviation of each feature that has at least 10 unique values

The formula for the sample standard deviation is:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Where :
x is the sample mean.
n is the number of data points in the sample.

In this case the standard deviation of the features with at least 10 unique values is :
oldpeak=1.161075
age=9.082101
trestbps=17.538143
thalachh=22.905161
chol=51.830751

## 1.2 Visualizations using Matplotlib

### 1.2.2 Scatter plot

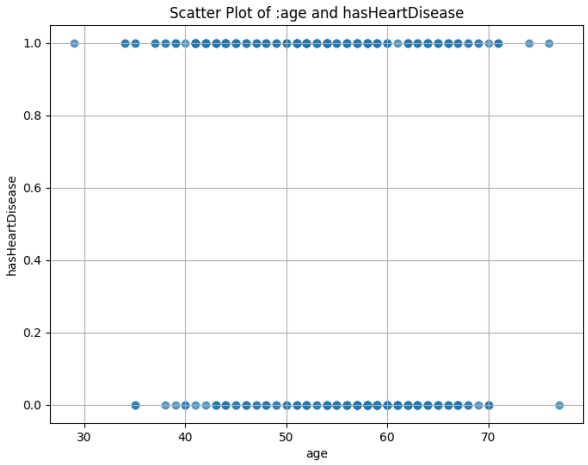


Figure 1: Scatter plot with the feature age on the x-axis and the feature hasHeart Disease on the y-axis

### 1.2.3 Scatter plot of two new features

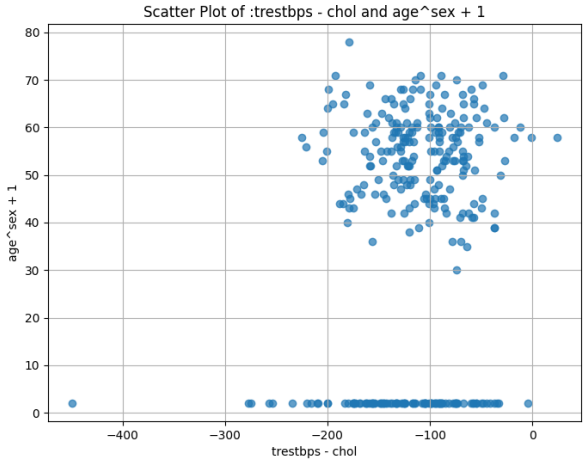


Figure 2: scatter plot of F1 = trestbps - chol on the x-axis and F2 = agesex +1 on the y-axis

### 1.2.4 Histogram

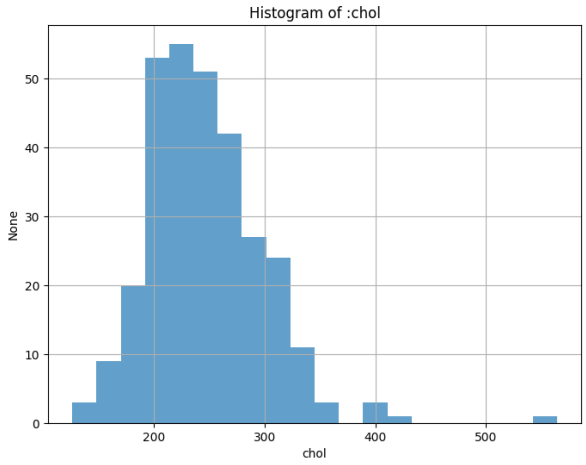


Figure 3: histogram of the feature with the highest mean, from the features that has at least 10 unique values

### 1.2.5 Line

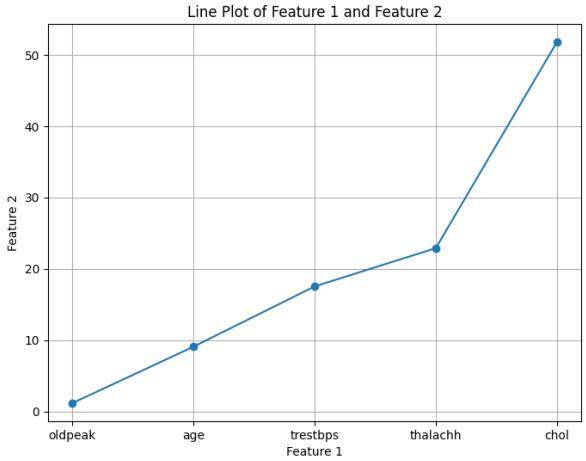


Figure 4: Line plot of the standard deviations of continuous features with at least 10 unique values