SPRINT-1

Date	18 November 2022
Team ID	PNT2022TMID52957
Project Name	Visualizing and Predicting Heart Diseases with an Interactive Dashboard
Maximum Marks	2 Marks

Dataset:

Data set is downloaded from the Kaggle in .CSV format.

Context: The leading cause of death in the developed world is heart disease. Therefore there needs to be work done to help prevent the risks of of having a heart attack or stroke.

Content: Use this dataset to predict which patients are most likely to suffer from a heart disease in the near future using the features given.

Acknowledgement: This data comes from the University of California Irvine's Machine Learning Repository at https://archive.ics.uci.edu/ml/datasets/Heart+Disease.

Data set contains 14 attributes each contains 270 rows of informations.

Dataset Description:

1. Age:
   Age of the patient in years.
2. Sex:
   Gender of the patient.

3. Chest Pain Type:
   Four types of chest pain type in our dataset,
   1.Typical Angina
   2.Atypical Angina
   3.Non-Anginal type
   4.Asymptomatic

4. BP:
   Level of blood pressure in mm/HG(Numerical)

5. Cholesterol:
   Serum Cholesterol in mg/dl (Numeric)

6. Fasting Blood Pressure:
   Blood sugar levels on fasting >120 mg/dl represents as 1 in case of true and 0 in case of false.

7. EKG Results:
   Result of electrocardiogram while at rest are represented in 3 distinct values:
   Value 0:Nominal
   Value 1:Having ST-T wave abnormality
   Value 2:Showing probable or definite left ventricular hypertrophy .

8. Max HR:
   maximum heart rate achieved

9. Exercise Angina:
   Exercise induced angina

10. Old peak:
    ST depression induced by exercise relative to rest

11. ST Slope:
    The slope of the peak exercise ST segment

12. Number of fluro:
    Number of major vessels (0-3) colored by flourosopy

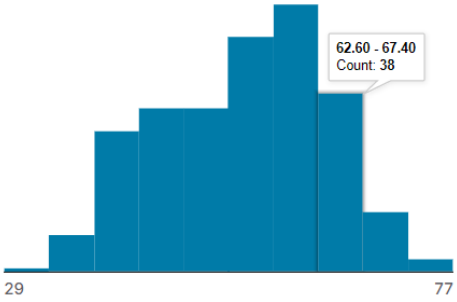
13. Thallium:
    thal: 0 = normal
    1 = fixed defect
    2 = reversable defect

The names and social security numbers of the patients were recently removed from the database, replaced with dummy values.

14. Heart Disease:
    0=Absence
    1=Present

Understanding the Dataset:

Age

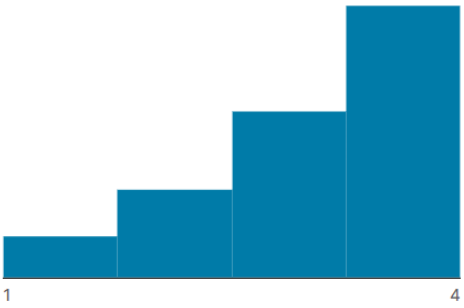


Valid	270	100%
Mismatched	0	0%
Missing	0	0%
Mean	54.4	
Std. Deviation	9.09	
Quantiles	29	Min
	48	25%
	55	50%
	61	75%
	77	Max

Sex

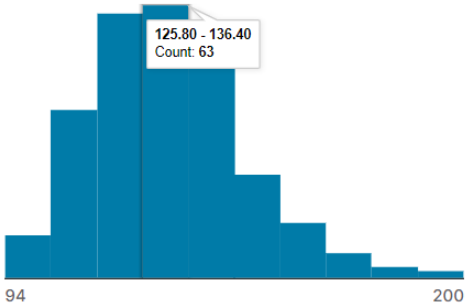
Valid	270	100%
Mismatched	0	0%
Missing	0	0%
Mean	0.68	
Std. Deviation	0.47	
Quantiles	0	Min
	0	25%
	1	50%
	1	75%
	1	Max

Chest pain type



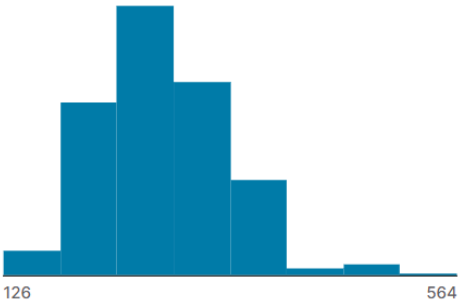
Valid	270	100%
Mismatched	0	0%
Missing	0	0%
Mean	3.17	
Std. Deviation	0.95	
Quantiles	1	Min
	3	25%
	3	50%
	4	75%
	4	Max

BP



Valid	270	100%
Mismatched	0	0%
Missing	0	0%
Mean	131	
Std. Deviation	17.8	
Quantiles	94	Min
	120	25%
	130	50%
	140	75%
	200	Max

Cholesterol



Valid	270	100%
Mismatched	0	0%
Missing	0	0%
Mean	250	
Std. Deviation	51.6	
Quantiles	126	Min
	213	25%
	245	50%
	281	75%
	564	Max

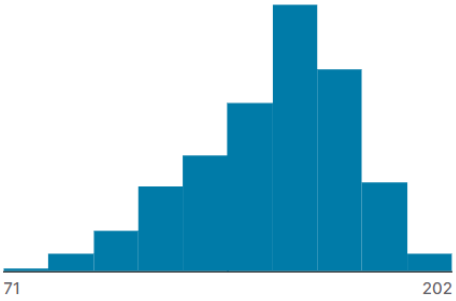
FBS over 120

Valid	270	100%
Mismatched	0	0%
Missing	0	0%
Mean	0.15	
Std. Deviation	0.36	
Quantiles	0	Min
	0	25%
	0	50%
	0	75%
	1	Max

EKG results

Valid	270	100%
Mismatched	0	0%
Missing	0	0%
Mean	1.02	
Std. Deviation	1	
Quantiles	0	Min
	0	25%
	2	50%
	2	75%
	2	Max

Max HR



Valid	270	100%
Mismatched	0	0%
Missing	0	0%
Mean	150	
Std. Deviation	23.1	
Quantiles	71	Min
	133	25%
	154	50%
	166	75%
	202	Max

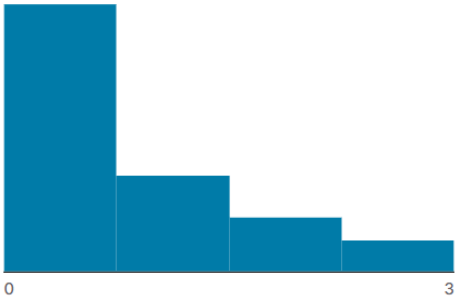
Exercise angina

Valid	270	100%
Mismatched	0	0%
Missing	0	0%
Mean	0.33	
Std. Deviation	0.47	
Quantiles	0	Min
	0	25%
	0	50%
	1	75%
	1	Max

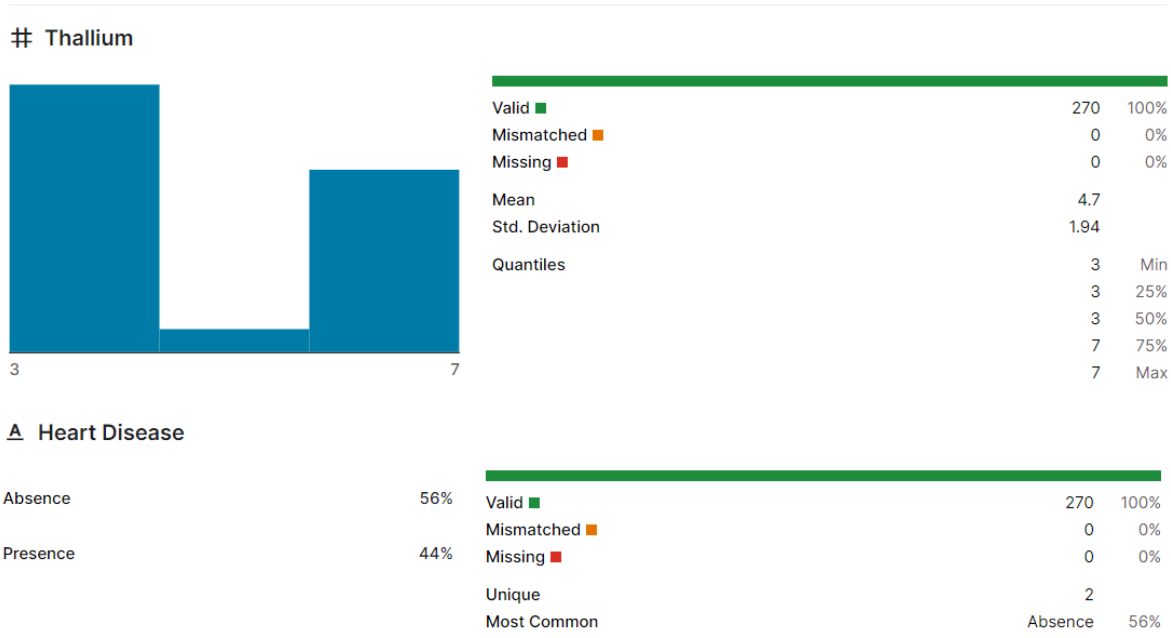
Slope of ST

Valid	270	100%
Mismatched	0	0%
Missing	0	0%
Mean	1.59	
Std. Deviation	0.61	
Quantiles	1	Min
	1	25%
	2	50%
	2	75%
	3	Max

Number of vessels fluro



Valid	270	100%
Mismatched	0	0%
Missing	0	0%
Mean	0.67	
Std. Deviation	0.94	
Quantiles	0	Min
	0	25%
	0	50%
	1	75%
	3	Max



After understanding the dataset. Dataset is loaded into IBM COGNOS.