American International University-Bangladesh (AIUB)

Department of Computer Science

Faculty of Science &Technology (FST)

FALL 23-24

Section: C

INTRODUCTION TO DATA SCIENCE

REPORT SUBMITTED BY

SN	Student Name	Student ID
1	RIAZ SARKER	20-42362-1
2	MD. RAFID HASAN	20-42421-1

SUBMITTED TO

TOHEDUL ISLAM

Assistant Professor, Faculty, Department of Computer Science

Description:

The heart_attack_data dataset is widely recognized for its information on patient. According to the provided information, the medical dataset classifies either heart attack or none. It is frequently employed as a benchmark in data analysis and machine learning tasks. This dataset offers valuable insights into the demographics and characteristics of the heart attack patient, as well as their heart condition. By studying this dataset, researchers and data scientists can explore various factors that potentially influenced heart Disease rates on patient, including Age, Sex, ChestPain Type, RestingBp, Cholesterol, FastingBS, RestingECG, MaxHr etc.

Attributes:

Age: Age of the patient

Sex: Sex of the patient

ChestPain Type: There is four type of ChestPain Type of patient (Typical Angina, Atypical Angina, Non-Anginal Pain, Asymptomatic)

RestingBP: Resting blood pressure (in mm Hg) of Patient.

Cholesterol: Cholesterol in mg/dl fetched via BMI sensor

FastingBS: Fasting blood sugar > (120 mg/dl)

RestingECG: Resting Electrocardiographic results of Patient

MaxHR: Maximum heart rate achieved of Patient.

ExerciseAngina: Exercise induced angina on heart attack Patient.

Oldpeak: ST depression induced by exercise relative to rest.

ST_Slope: The slope of the peak exercise ST segment

HeartDisease: Diagnosis of heart disease (angiographic disease status).

Import Dataset:

heart_attack_data <- read.csv("D:/Dataset_MIdterm_Sectoin(C).csv",header=TRUE, sep=",")

View(heart_attack_data)

```
Console   Terminal   Background Jobs
R 4.3.2 · D:/Data Science/
> heart_attack_data <- read.csv("D:/Dataset_MIdterm_Sectoin(C).csv",header=TRUE, sep=",")
> View(heart_attack_data)
>
```

Explanation: Imports the heart_attack_data dataset from a designated file path, storing it in the variable "heart_attack_data". The dataset is read as a CSV file, with headers included and values separated by patient informetion. To facilitate exploration and analysis, executing "View(heart_attack_data)" opens a separate viewer window, displaying the imported dataset.

DISPLAYING THE DATA_SET (heart_attack_data):

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
1	40	M	ATA	140	289	0	Normal	172	N	0.0	Up	0
2	49	F	NAP	160	180	0	Normal	156	N	1.0	Flat	1
3	37	M	ATA	130	283	0	ST	98	N	0.0	Up	0
4	NA	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
5	54	M	NAP	-150	195	0	Normal	122	N	0.0	Up	0
6	39	M	NAP	120	339	0	Normal	170	N	0.0	Up	0
7	45	F	ATA	130	237	0	Normal	170		0.0	Up	0
8	54	M	ATA	110	208	0	Normal	142	N	0.0	Up	0
9	37		ASY	140	207	0	Normal	130	Y	1.5	Flat	1
10	48	F	ATA	120	284	0	Normal	120	N	0.0	Up	0
11	37	F	NAP	130	1000	0	Normal	142	N	0.0	Up	0
12	58	M	ATA	136	1005	0	ST	99	Y	2.0	Flat	1
13	39	M	ATA	120	204	0	Normal	145	N	0.0	Up	0
14	49	M	ASY	140	234	0	Normal	140	Y	1.0	Flat	1
15	42	M	NAP	115	211	0	ST	137	N	0.0	Up	0
16	54	M	ATA	120	273	0	Normal	150	N	1.5	Flat	0
17	38	M	ASY	110	196	0	Normal	166	N	0.0	Flat	1
18	43	F	ATA	120	201	0	Normal	165	N	0.0	Up	0
19	60	M	ASY	100	248	0	Normal	125	N	1.0	Flat	1
20	36	M	ATA	120	267	0	Normal	160	N	3.0	Flat	1

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
21	43	M	TA	100	223	0	Normal	142	N	0.0	Up	0
22	44	M	ATA	120	184	0	Normal	142	N	1.0	Flat	0
23	49	F	ATA	124	201	0	Normal	164	N	0.0	Up	0
24	*NA*	M	ATA	150	288	0	Normal	150	Y	3.0	Flat	1
25	40	M	NAP	130	215	0	Normal	138	N	0.0	Up	0
26	36		NAP	130	209	0	Normal	178	N	0.0	Up	0
27	53	M	ASY	124	260	0	ST	112		3.0	Flat	0
28	52	M	ATA	120	284	0	Normal	118	N	0.0	Up	0
29	53	F	ATA	113	468	0	Normal	127	N	0.0	Up	0
30	51	M	ATA	125	188	0	Normal	145	N	0.0	Up	0
31	53	M	NAP	145	518	0	Normal	130	N	0.0	Flat	1
32	56	M	NAP	130	167	0	Normal	114	N	0.0	Up	0
33	*NA*	M	ASY	125	224	0	Normal	122	N	2.0	Flat	1
34	41	M	ASY	130	172	0	ST	130	N	2.0	Flat	1
35	43	F	ATA	150	186	0	Normal	154	N	0.0	Up	0
36	54	M	ATA	125	254	0	Normal	155	N	0.0	Up	0
37	65	M	ASY	140	306	1	Normal	87	Y	1.5	Flat	1
38	41	F	ATA	110	250	0	ST	142	N	0.0	Up	0
39	48	M	ATA	120	177	1	ST	148	N	0.0	Up	0
40	48		ASY	150	227	0	Normal	130	Y	1.0	Flat	0

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
130	42	M	NAP	120	228	0	Normal	152	Y	1.5	Flat	0
131	54	M	NAP	145	292	0	Normal	130	N	0.0	Up	0
132	46	M	ASY	110	202	0	Normal	150	Y	0.0	Flat	1
133	56	M	ASY	170	388	0	ST	122	Y	2.0	Flat	1
134	56	M	ASY	150	230	0	ST	124	Y	1.5	Flat	1
135	61	F	ASY	130	294	0	ST	120	Y	1.0	Flat	0
136	54	M	NAP	115	265	0	Normal	175	N	0.0	Flat	1
137	43	F	ATA	120	215	0	ST	175	N	0.0	Up	0
138	39	M	ATA	120	241	0	ST	146	N	2.0	Up	0
139	54	M	ASY	140	166	0	Normal	118	Y	0.0	Flat	1
140	54	M	ASY	150	247	0	Normal	130	Y	2.0	Flat	1
141	52	M	ASY	160	331	0	Normal	94	Y	2.5	Flat	1
142	50	M	ASY	140	341	0	ST	125	Y	2.5	Flat	1
143	54	M	ASY	160	291	0	ST	158	Y	3.0	Flat	1
144	53	M	ASY	140	243	0	Normal	155	N	0.0	Up	0
145	54	F	ATA	120	279	0	Normal	150	N	1.0	Flat	1
146	39	M	ASY	110	273	0	Normal	132	N	0.0	Up	0
147	42	M	ATA	120	198	0	Normal	155	N	0.0	Up	0
148	43	F	ATA	120	249	0	ST	176	N	0.0	Up	0
149	50	M	ATA	120	168	0	Normal	160	N	0.0	Up	0
150	54	M	ASY	130	603	1	Normal	125	Y	1.0	Flat	1

Visualizing the Dataset:

names(heart_attack_data)

str(heart_attack_data)

dim(heart_attack_data)

summary(heart_attack_data)

```
Console   Terminal   Background Jobs
R 4.3.2 · D:/Data Science/
> names(heart_attack_data)
 [1] "Age"            "Sex"            "ChestPainType"  "RestingBP"      "Cholesterol"
 [6] "FastingBS"      "RestingECG"     "MaxHR"          "ExerciseAngina" "Oldpeak"
[11] "ST_Slope"       "HeartDisease"
> str(heart_attack_data)
'data.frame':	150 obs. of  12 variables:
 $ Age           : int  40 49 37 NA 54 39 45 54 37 48 ...
 $ Sex           : chr  "M" "F" "M" "F" ...
 $ ChestPainType : chr  "ATA" "NAP" "ATA" "ASY" ...
 $ RestingBP     : int  140 160 130 138 -150 120 130 110 140 120 ...
 $ Cholesterol   : int  289 180 283 214 195 339 237 208 207 284 ...
 $ FastingBS     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ RestingECG    : chr  "Normal" "Normal" "ST" "Normal" ...
 $ MaxHR         : int  172 156 98 108 122 170 170 142 130 120 ...
 $ ExerciseAngina: chr  "N" "N" "N" "Y" ...
 $ Oldpeak       : num  0 1 0 1.5 0 0 0 0 1.5 0 ...
 $ ST_Slope      : chr  "Up" "Flat" "Up" "Flat" ...
 $ HeartDisease  : int  0 1 0 1 0 0 0 0 1 0 ...
> dim(heart_attack_data)
[1] 150  12
```

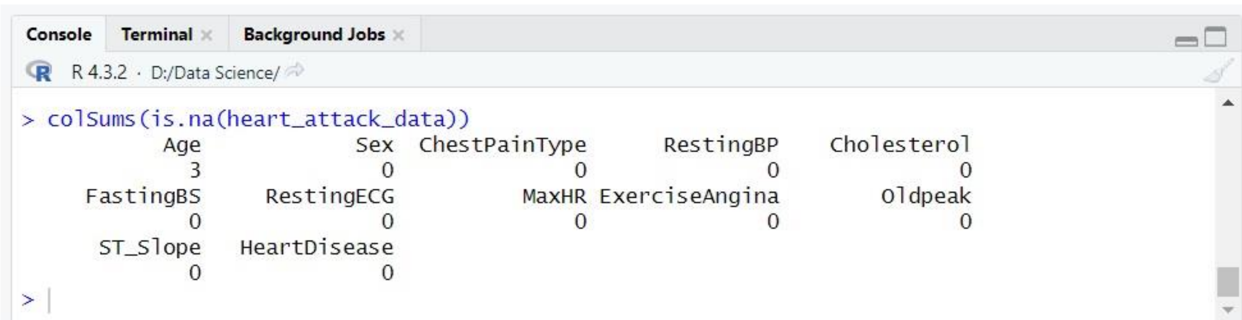
```
Console   Terminal   Background Jobs
R 4.3.2 · D:/Data Science/
> summary(heart_attack_data)
      Age              Sex            ChestPainType        RestingBP
 Min.   : 32.00   Length:150         Length:150         Min.   :-150.0
 1st Qu.: 42.00   Class :character   Class :character   1st Qu.: 120.0
 Median : 49.00   Mode  :character   Mode  :character   Median : 130.0
 Mean   : 49.81                                         Mean   : 129.3
 3rd Qu.: 54.00                                         3rd Qu.: 140.0
 Max.   :172.00                                         Max.   : 190.0
 NA's   :3
  Cholesterol       FastingBS        RestingECG            MaxHR
 Min.   :  85.0   Min.   :0.00000   Length:150         Min.   : 82.0
 1st Qu.: 205.2   1st Qu.:0.00000   Class :character   1st Qu.:124.0
 Median : 239.0   Median :0.00000   Mode  :character   Median :140.0
 Mean   : 258.3   Mean   :0.08667                      Mean   :140.1
 3rd Qu.: 277.0   3rd Qu.:0.00000                      3rd Qu.:155.8
 Max.   :1005.0   Max.   :1.00000                      Max.   :190.0

 ExerciseAngina        Oldpeak          ST_Slope          HeartDisease
 Length:150         Min.   :0.0000   Length:150         Min.   :0.00
 Class :character   1st Qu.:0.0000   Class :character   1st Qu.:0.00
 Mode  :character   Median :0.0000   Mode  :character   Median :0.00
                    Mean   :0.5933                      Mean   :0.38
                    3rd Qu.:1.0000                      3rd Qu.:1.00
                    Max.   :4.0000                      Max.   :1.00
```

Explanation: Visualizing the dataset involves exploring its variables and structure. The command "names(heart_attack_data)" retrieves the column names, providing an overview of the available data. "str(heart_attack_data)" displays the structure, data types, and summaries of the variables. Additionally, "dim(heart_attack_data)" reveals the dataset dimensions, giving insights into its size. Lastly, "summary(heart_attack_data)" provides statistical summaries, aiding in understanding the numerical aspects of the dataset. These visualization techniques contribute to comprehending the dataset and supporting data analysis and decision-making processes.

Find Missing Values:

colSums(is.na(heart_attack_data))



```
Console  Terminal  Background Jobs
R 4.3.2 · D:/Data Science/
> colSums(is.na(heart_attack_data))
           Age            Sex  ChestPainType      RestingBP    Cholesterol
             3              0              0              0              0
     FastingBS     RestingECG          MaxHR ExerciseAngina        Oldpeak
             0              0              0              0              0
      ST_Slope   HeartDisease
             0              0
>
```

Explanation: Identifying missing values in the heart_attack_data dataset involves assessing the presence of null or missing data points. This process allows researchers to determine the extent of missingness across different variables. By analyzing the column-wise sums of missing values, insights can be gained into the distribution and quantity of missing data, aiding in subsequent data cleaning and imputation strategies.

Remove Missing Values (Age):

missing_Age <- is.na(heart_attack_data$Age)

heart_attack_data_1 <- subset(heart_attack_data, !missing_Age)

colSums(is.na(heart_attack_data_1))

```
Console  Terminal ×  Background Jobs ×
R 4.3.2 · D:/Data Science/
> missing_Age <- is.na(heart_attack_data$Age)
> heart_attack_data_1 <- subset(heart_attack_data, !missing_Age)
> colSums(is.na(heart_attack_data_1))
           Age            Sex  ChestPainType      RestingBP    Cholesterol 
             0              0              0              0              0 
     FastingBS     RestingECG          MaxHR ExerciseAngina        Oldpeak 
             0              0              0              0              0 
      ST_Slope   HeartDisease 
             0              0 
> 
```

Explanation: To address missing values (Age) in the heart_attack_data dataset, a process of removal can be performed. By creating a logical variable, "missing_Age," to identify the missing values, the dataset can be subsetted to exclude these cases. The resulting dataset, "heart_attack_data_1," will no longer contain missing Age values. By subsequently checking for missing values using "colSums(is.na(heart_attack_data_1))," researchers can confirm that the Age variable is now free of missing data.

Find Empty Data:

colSums(heart_attack_data_1 == '')

```
Console  Terminal ×  Background Jobs ×
R 4.3.2 · D:/Data Science/
> colSums(heart_attack_data_1 == '')
           Age            Sex  ChestPainType      RestingBP    Cholesterol 
             0              3              0              0              0 
     FastingBS     RestingECG          MaxHR ExerciseAngina        Oldpeak 
             0              0              0              2              0 
      ST_Slope   HeartDisease 
             0              0 
> 
```

Remove Empty Data (Sex):

heart_attack_data_2 <- heart_attack_data_1[!heart_attack_data_1$Sex == "",]

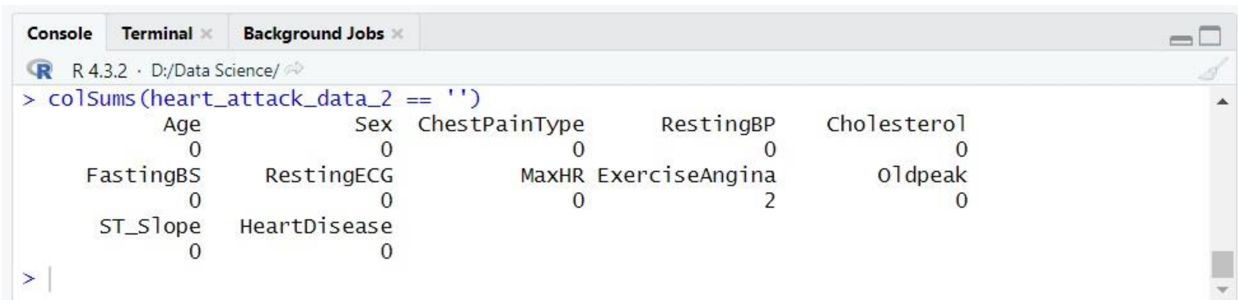
colSums(heart_attack_data_2 == '')

```
Console  Terminal ×  Background Jobs ×
R 4.3.2 · D:/Data Science/
> colSums(heart_attack_data_2 == '')
           Age            Sex  ChestPainType      RestingBP    Cholesterol 
             0              0              0              0              0 
     FastingBS     RestingECG          MaxHR ExerciseAngina        Oldpeak 
             0              0              0              2              0 
      ST_Slope   HeartDisease 
             0              0 
> 
```

Explanation: To remove empty or blank values from the heart_attack_data dataset, an assessment is conducted to determine the presence of empty data using column-wise calculations. After identifying the empty values, the dataset is filtered to exclude rows where specific variables have empty entries. By reevaluating the dataset, it can be confirmed that the empty data has been successfully removed.

Again Find Empty Data:

colSums(heart_attack_data_2 == '')

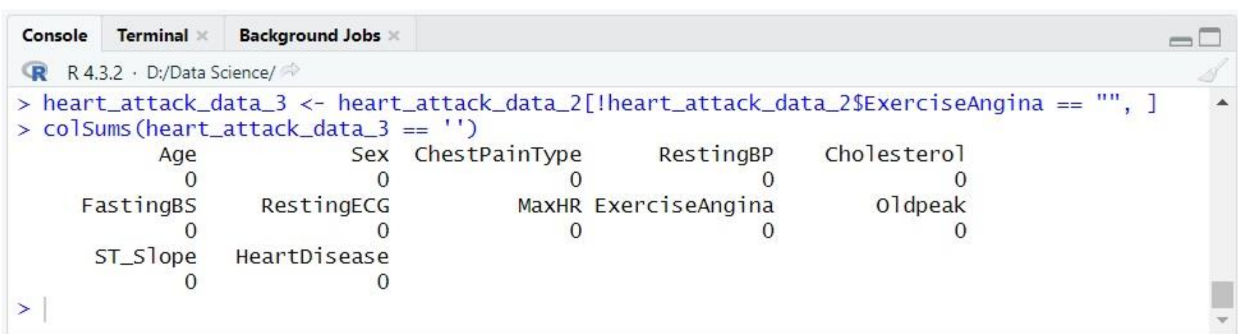


```
> colSums(heart_attack_data_2 == '')
          Age           Sex  ChestPainType      RestingBP    Cholesterol
            0             0              0              0              0
    FastingBS    RestingECG          MaxHR ExerciseAngina        Oldpeak
            0             0              0              2              0
     ST_Slope  HeartDisease
            0             0
>
```

Remove Empty Data (ExerciseAngina):

heart_attack_data_3 <- heart_attack_data_2[!heart_attack_data_2$ExerciseAngina == "",]

colSums(heart_attack_data_3 == '')



```
> heart_attack_data_3 <- heart_attack_data_2[!heart_attack_data_2$ExerciseAngina == "", ]
> colSums(heart_attack_data_3 == '')
          Age           Sex  ChestPainType      RestingBP    Cholesterol
            0             0              0              0              0
    FastingBS    RestingECG          MaxHR ExerciseAngina        Oldpeak
            0             0              0              0              0
     ST_Slope  HeartDisease
            0             0
>
```

Explanation: To remove empty or blank values from the heart_attack_data dataset, an assessment is conducted to determine the presence of empty data using column-wise calculations. After identifying the empty values, the dataset is filtered to exclude rows where specific variables have empty entries. By reevaluating the dataset, it can be confirmed that the empty data has been successfully removed.

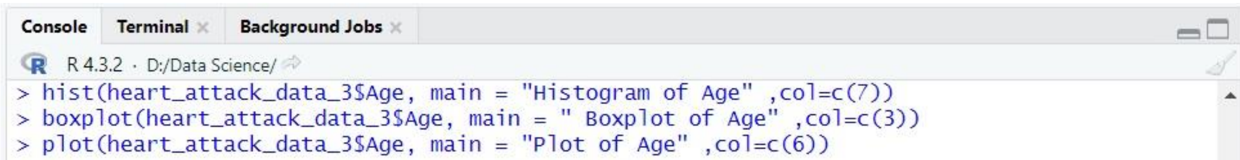
Find Outlier in Age attribute:

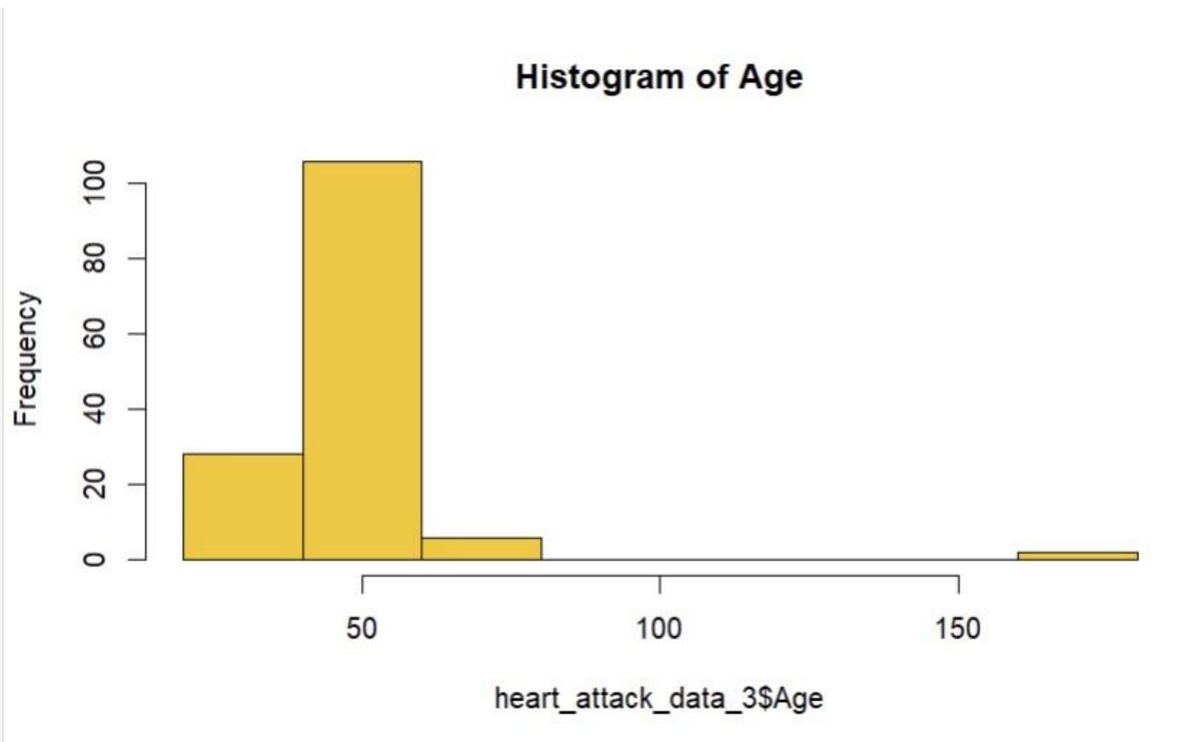
hist(heart_attack_data_3$Age, main = "Histogram of Age" ,col=c(7))

boxplot(heart_attack_data_3$Age, main = " Boxplot of Age" ,col=c(3))

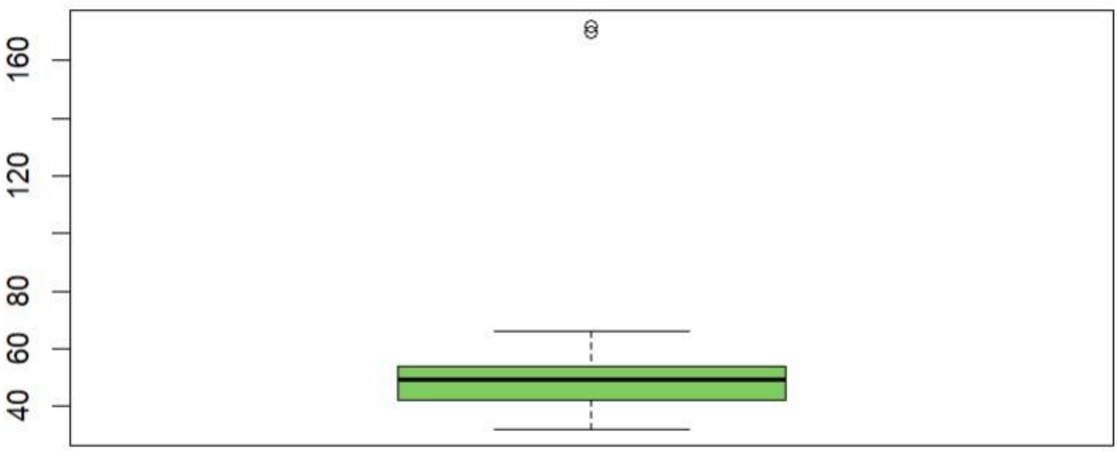
plot(heart_attack_data_3$Age, main = "Plot of Age" ,col=c(6))

barplot(heart_attack_data_3$Age, main = "Bar of Age" ,col=c(6))

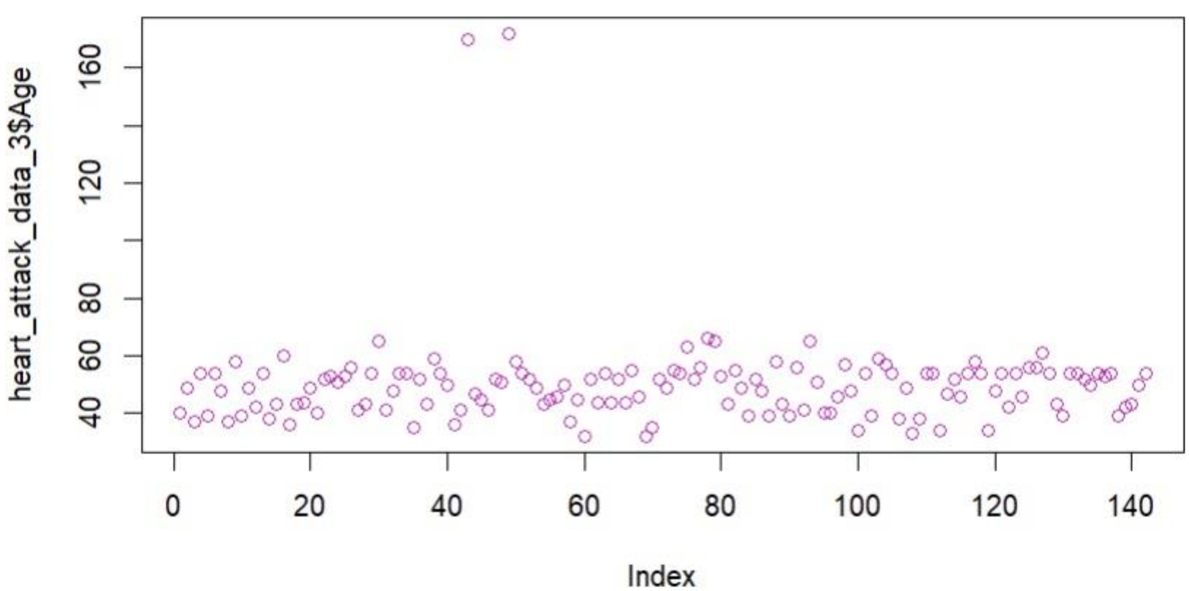


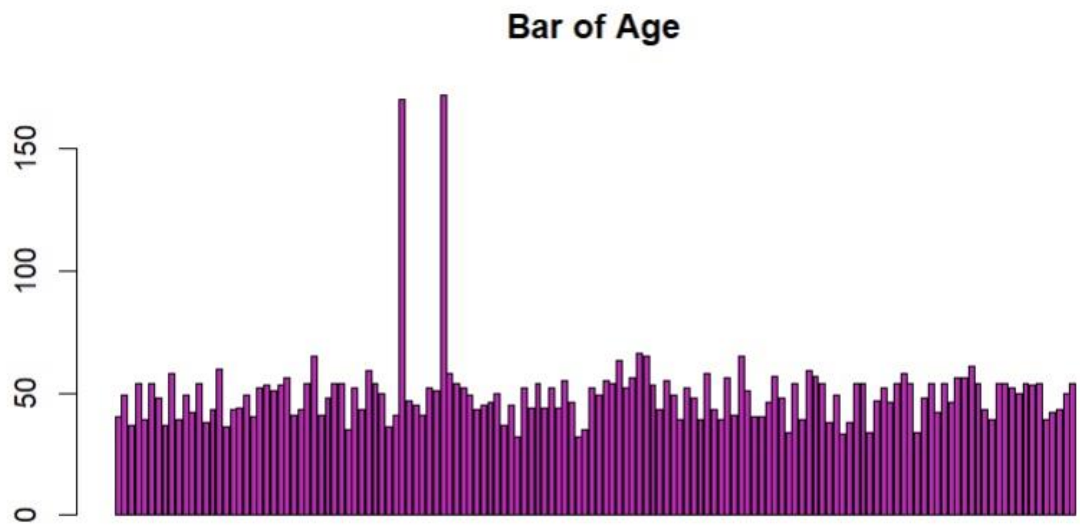


Boxplot of Age



Plot of Age





Explanation: To identify outliers in the Age attribute of the heart_attack_data dataset, various visualization techniques can be employed. A histogram provides an overview of the Age distribution, highlighting any unusual values. A box plot visually represents the range and distribution of Ages, making it easier to identify outliers. Scatter plots can also be utilized to examine individual data points and detect any extreme values that deviate significantly from the overall pattern. These visualization approaches help in identifying and understanding outliers within the Age attribute.

Remove Outliers in Age attribute:

```
Age_Q1 <- quantile(heart_attack_data_3$Age, 0.25, na.rm = TRUE)

Age_Q3 <- quantile(heart_attack_data_3$Age, 0.75, na.rm = TRUE)

Age_IQR <- Age_Q3 - Age_Q1

lower_Age <- Age_Q1 - 1.5 * Age_IQR

upper_Age <- Age_Q3 + 1.5 * Age_IQR

Age_outliers <- heart_attack_data_3$Age < lower_Age | heart_attack_data_3$Age > upper_Age

heart_attack_data_4 <- heart_attack_data_3[!Age_outliers, ]
```

```
Console   Terminal   Background Jobs
R 4.3.2 · D:/Data Science/
> Age_Q1 <- quantile(heart_attack_data_3$Age, 0.25, na.rm = TRUE)
> Age_Q3 <- quantile(heart_attack_data_3$Age, 0.75, na.rm = TRUE)
> Age_IQR <- Age_Q3 - Age_Q1
> lower_Age <- Age_Q1 - 1.5 * Age_IQR
> upper_Age <- Age_Q3 + 1.5 * Age_IQR
> Age_outliers <- heart_attack_data_3$Age < lower_Age | heart_attack_data_3$Age > upper_Age
> heart_attack_data_4 <- heart_attack_data_3[!Age_outliers, ]
```

For Graph (Age):

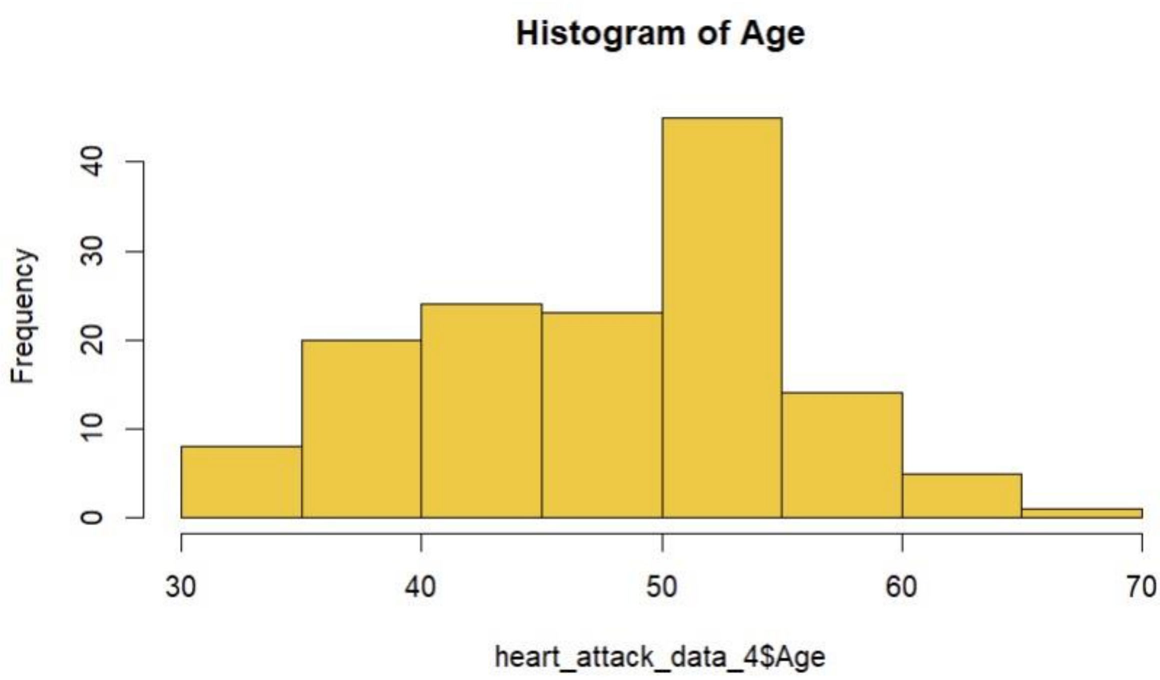
hist(heart_attack_data_4$Age, main = "Histogram of Age" ,col=c(7))

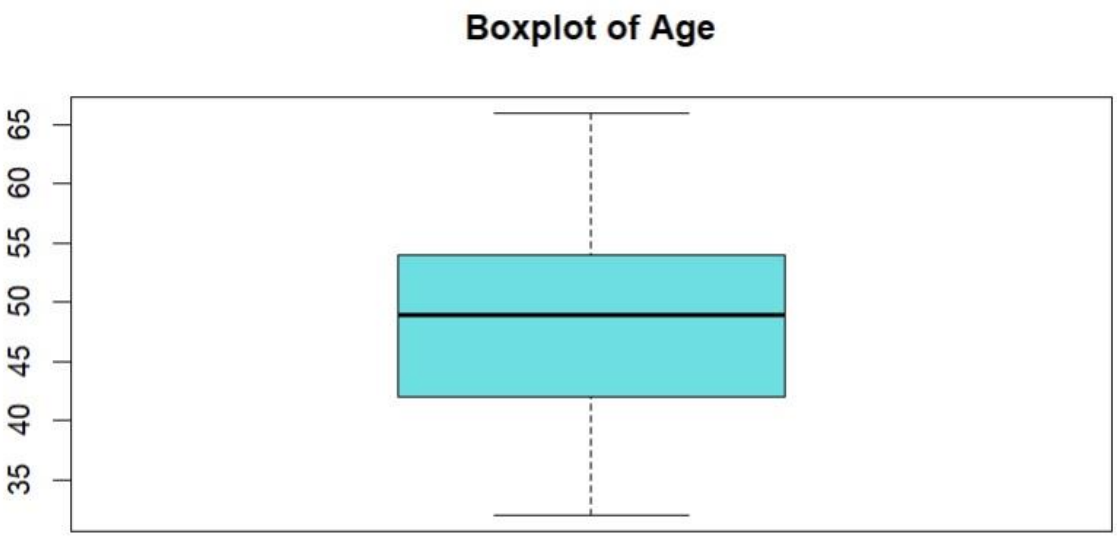
boxplot(heart_attack_data_4$Age, main = "Boxplot of Age" ,col=c(5))

plot(heart_attack_data_4$Age, main = "Plot of Age" ,col=c(3))

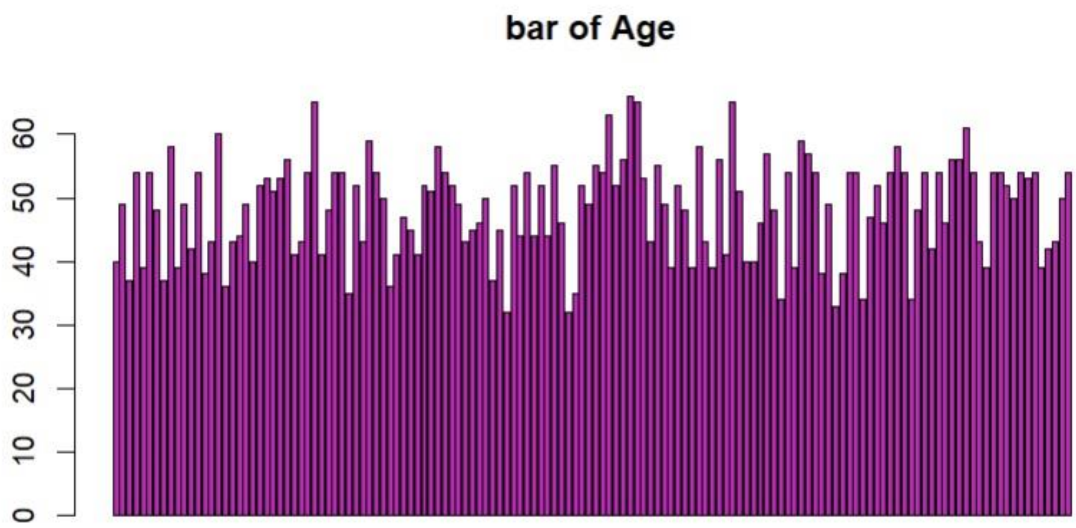
barplot(heart_attack_data_4$Age, main = "bar of Age" ,col=c(6))

```
Console   Terminal   Background Jobs
R 4.3.2 · D:/Data Science/
> hist(heart_attack_data_4$Age, main = "Histogram of Age" ,col=c(7))
> boxplot(heart_attack_data_4$Age, main = "Boxplot of Age" ,col=c(5))
> plot(heart_attack_data_4$Age, main = "Plot of Age" ,col=c(2))
> |
```




Boxplot of Age
35
40
45
50
55
60
65


Plot of Age
heart_attack_data_4$Age
35
40
45
50
55
60
65
0
20
40
60
80
100
120
140
Index



Explanation: To remove outliers in Age attribute from the heart_attack_data dataset using Tukey's fences method. The lower quartile (Q1) and upper quartile (Q3) are calculated, and the interquartile range (IQR) is determined. The lower and upper fences are established by subtracting and adding 1.5 times the IQR, respectively. Age values falling outside these fences are identified as outliers and filtered out, resulting in a new dataset, heart_attack_data_4, without outliers in Age. The resulting Age distribution can be visualized using a Histogram, Boxplot, Plot and Barplot.

Checking of Error Values:

```
attribute_names <- names(heart_attack_data_4)

for (attribute in attribute_names)

{

  unique_values <- unique(heart_attack_data_4[[attribute]])

  print(paste("Unique values in", attribute, ":"))

  print(unique_values)

}
```

Console Terminal Background Jobs

R 4.3.2 · D:/Data Science/

```
> attribute_names <- names(heart_attack_data_4)
> for (attribute in attribute_names)
+ {
+   unique_values <- unique(heart_attack_data_4[[attribute]])
+   print(paste("Unique values in", attribute, ":"))
+   print(unique_values)
+ }
```

Console Terminal Background Jobs

R 4.3.2 · D:/Data Science/

```
[1] "Unique values in Age :"
 [1] 40 49 37 54 39 48 58 42 38 43 60 36 44 52 53 51 56 41 65 35 59 50 47 45 46 32 55 63
[29] 66 57 34 33 61
[1] "Unique values in Sex :"
[1] "M" "F"
[1] "Unique values in ChestPainType :"
[1] "ATA" "NAP" "ASY" "TA"
[1] "Unique values in RestingBP :"
 [1]  140  160  130 -150  120  110  136  115  100  124  113  125  145  150  112  132  118
[18]  170  142  190  135  180  108
[1] "Unique values in Cholesterol :"
  [1]  289  180  283  195  339  208  284 1000 1005  204  234  211  273  196  201  248
 [17]  267  223  184  215  468  188  518  167  172  186  254  306  250  177  230  294
 [33]  264  259  175  318  216  340  205  224  245  194  213  365  342  253  277  202
 [49]  260  297  225  246  412  265  182  218  268  163  529  100  206  238  139  263
 [65]  291  229  307  210  329  207  147   85  269  275  179  392  466  214  129  241
 [81]  255  276  282  338  160  156  272  240  393  161  228  292  388  166  247  331
 [97]  341  243  279  198  249  168  603
[1] "Unique values in FastingBS :"
[1] 0 1
[1] "Unique values in RestingECG :"
[1] "Normal" "ST"
[1] "Unique values in MaxHR :"
 [1] 172 156  98 122 170 142 120  99 145 140 137 150 166 165 125 160 164 138 118 127 130
[22] 114 154 155  87 148 100 168 184 134  96 174 175 144  82 135 115 128 116  94 112 110
[43]  92 180 152 124 106 185 139 190 146 158 132 176
[1] "Unique values in ExerciseAngina :"
[1] "N" "Y"
[1] "Unique values in Oldpeak :"
[1] 0.0 1.0 2.0 1.5 3.0 4.0 0.5 2.5
[1] "Unique values in ST_Slope :"
[1] "Up"   "Flat"
[1] "Unique values in HeartDisease :"
[1] 0 1
> 
```

Explanation: During the analysis of the heart_attack_data dataset, a rigorous assessment was performed to identify and rectify error values. This involved iterating through each attribute and extracting unique values to detect anomalies. Among the attributes examined, special attention was given to the "Sex" attribute, where valid values were sought after verifying data accuracy. By meticulously addressing the short from of Male and Female and confirming the validity of attribute values, the dataset's integrity was ensured, laying the foundation for reliable analysis and interpretations.

Fixing the Value of Sex:

unique(heart_attack_data_4$Sex)

heart_attack_data_4$Sex <- ifelse(heart_attack_data_4$Sex == "M","Male", heart_attack_data_4$Sex)

heart_attack_data_4$Sex <- ifelse(heart_attack_data_4$Sex == "F","Female", heart_attack_data_4$Sex)

unique(heart_attack_data_4$Sex)

Previous Data (Sex):

View(heart_attack_data_3)

DS Project Test.R* | heart_attack_data_3 | heart_attack_data_4 | heart_attack_data

Filter

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_S
1	40	M	ATA	140	289	0	Normal	172	N	0.0	U
2	49	F	NAP	160	180	0	Normal	156	N	1.0	Fl
3	37	M	ATA	130	283	0	ST	98	N	0.0	U
5	54	M	NAP	-150	195	0	Normal	122	N	0.0	U
6	39	M	NAP	120	339	0	Normal	170	N	0.0	U
8	54	M	ATA	110	208	0	Normal	142	N	0.0	U
10	48	F	ATA	120	284	0	Normal	120	N	0.0	U
11	37	F	NAP	130	1000	0	Normal	142	N	0.0	U
12	58	M	ATA	136	1005	0	ST	99	Y	2.0	Fl
13	39	M	ATA	120	204	0	Normal	145	N	0.0	U
14	49	M	ASY	140	234	0	Normal	140	Y	1.0	Fl
15	42	M	NAP	115	211	0	ST	137	N	0.0	U
16	54	M	ATA	120	273	0	Normal	150	N	1.5	Fl
17	38	M	ASY	110	196	0	Normal	166	N	0.0	Fl
18	43	F	ATA	120	201	0	Normal	165	N	0.0	U
19	60	M	ASY	100	248	0	Normal	125	N	1.0	Fl
20	36	M	ATA	120	267	0	Normal	160	N	3.0	Fl
21	43	M	TA	100	223	0	Normal	142	N	0.0	U
22	44	M	ATA	120	184	0	Normal	142	N	1.0	Fl

Updated Data (Sex):

View(heart_attack_data_4)

Console | Terminal | Background Jobs

R 4.3.2 · D:/Data Science/

```
> View(heart_attack_data_4)
>
```

DS Project Test.R* | heart_attack_data_3 | heart_attack_data_4 | heart_attack_data

Filter

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_S
1	40	Male	ATA	140	289	0	Normal	172	N	0.0	U
2	49	Female	NAP	160	180	0	Normal	156	N	1.0	Fl
3	37	Male	ATA	130	283	0	ST	98	N	0.0	U
5	54	Male	NAP	-150	195	0	Normal	122	N	0.0	U
6	39	Male	NAP	120	339	0	Normal	170	N	0.0	U
8	54	Male	ATA	110	208	0	Normal	142	N	0.0	U
10	48	Female	ATA	120	284	0	Normal	120	N	0.0	U
11	37	Female	NAP	130	1000	0	Normal	142	N	0.0	U
12	58	Male	ATA	136	1005	0	ST	99	Y	2.0	Fl
13	39	Male	ATA	120	204	0	Normal	145	N	0.0	U
14	49	Male	ASY	140	234	0	Normal	140	Y	1.0	Fl
15	42	Male	NAP	115	211	0	ST	137	N	0.0	U
16	54	Male	ATA	120	273	0	Normal	150	N	1.5	Fl
17	38	Male	ASY	110	196	0	Normal	166	N	0.0	Fl
18	43	Female	ATA	120	201	0	Normal	165	N	0.0	U
19	60	Male	ASY	100	248	0	Normal	125	N	1.0	Fl
20	36	Male	ATA	120	267	0	Normal	160	N	3.0	Fl
21	43	Male	TA	100	223	0	Normal	142	N	0.0	U
22	44	Male	ATA	120	184	0	Normal	142	N	1.0	Fl

Explanation: To correct or write the full from of values in the "Sex" attribute of the heart_attack_data dataset, the unique values in this attribute were examined. Subsequently, erroneous entries such as "M" were replaced with the correct label "Male" using conditional statements. Similarly, "F" were corrected to "Female". By fixing these incorrect values, the dataset's integrity was restored. A final examination of the unique values confirmed that the corrections were successfully implemented.

Detect Noisy Value (Cholesterol):

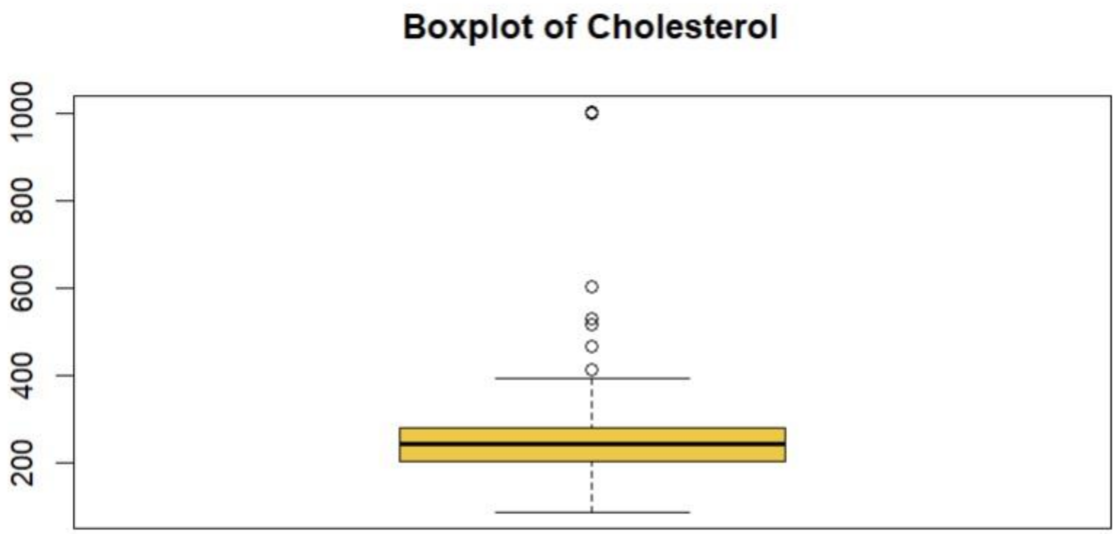
summary(heart_attack_data_4$Cholesterol)

boxplot(heart_attack_data_4$Cholesterol, main = "Boxplot of Cholesterol", col=c(7))

Console | Terminal | Background Jobs

R 4.3.2 · D:/Data Science/

```
> summary(heart_attack_data_4$Cholesterol)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   85.0   202.0   241.0   259.8   279.8  1005.0 
> boxplot(heart_attack_data_4$Cholesterol, main = "Boxplot of Cholesterol", col=c(7))
> 
```



After resolve the noisy value (Cholesterol):

Cholesterol_Q1 <- quantile(heart_attack_data_4$Cholesterol, 0.25, na.rm = TRUE)

Cholesterol_Q3 <- quantile(heart_attack_data_4$Cholesterol, 0.75, na.rm = TRUE)

Cholesterol_IQR <- Cholesterol_Q3 - Cholesterol_Q1

lower_Cholesterol <- Cholesterol_Q1 - 1.5 * Cholesterol_IQR

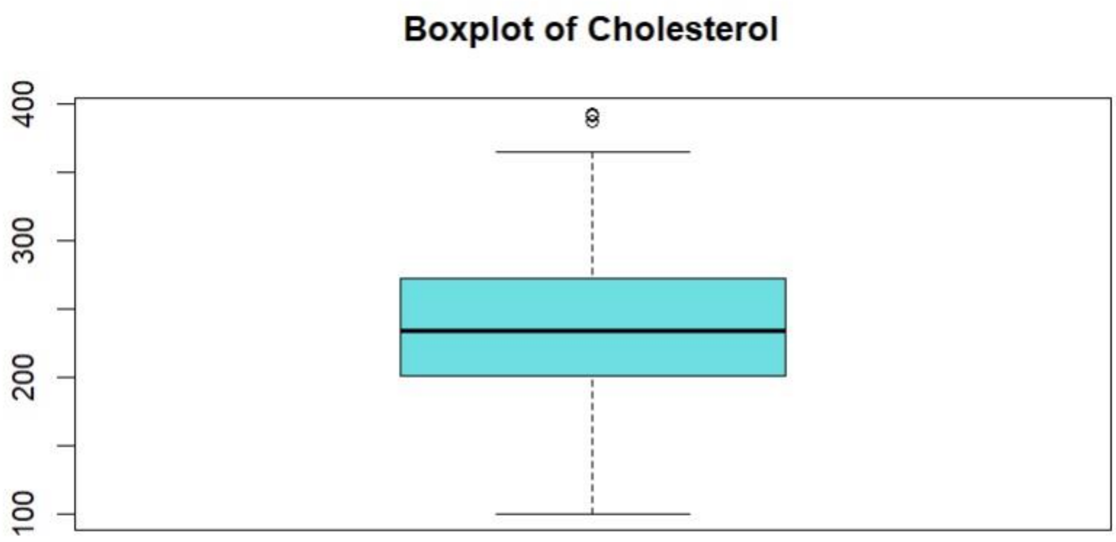
upper_Cholesterol <- Cholesterol_Q3 + 1.5 * Cholesterol_IQR

Cholesterol_noisyvalue <- heart_attack_data_4$Cholesterol < lower_Cholesterol | heart_attack_data_4$Cholesterol > upper_Cholesterol

heart_attack_data_5 <- heart_attack_data_4[!Cholesterol_noisyvalue,]

boxplot(heart_attack_data_5$Cholesterol, main = "Boxplot of Cholesterol", col=c(5))

```
> Cholesterol_Q1 <- quantile(heart_attack_data_4$Cholesterol, 0.25, na.rm = TRUE)
> Cholesterol_Q3 <- quantile(heart_attack_data_4$Cholesterol, 0.75, na.rm = TRUE)
> Cholesterol_IQR <- Cholesterol_Q3 - Cholesterol_Q1
> lower_Cholesterol <- Cholesterol_Q1 - 1.5 * Cholesterol_IQR
> upper_Cholesterol <- Cholesterol_Q3 + 1.5 * Cholesterol_IQR
> Cholesterol_noisyvalue <- heart_attack_data_4$Cholesterol < lower_Cholesterol | heart_attack_data_4$Ch
olesterol > upper_Cholesterol
> heart_attack_data_5 <- heart_attack_data_4[!Cholesterol_noisyvalue, ]
> boxplot(heart_attack_data_5$Cholesterol, main = "Boxplot of Cholesterol", col=c(5))
>
```



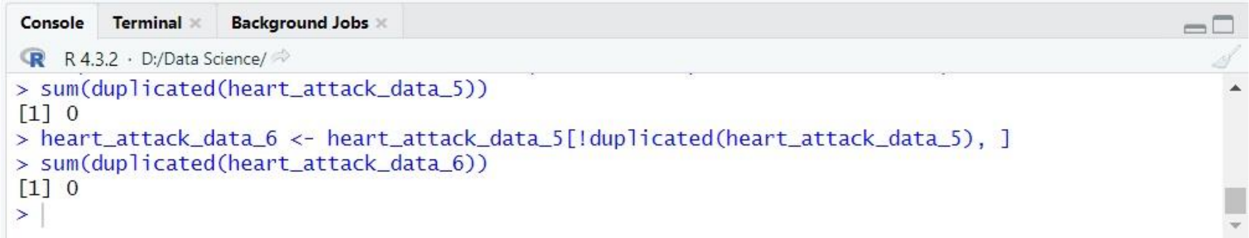
Explanation: To identify noisy values in the "Cholesterol" attribute of the heart_attack_data dataset, the summary statistics of Cholesterol were examined. This provided insights into the distribution and range of values. Additionally, a box plot was created to visualize the distribution, enabling the detection of potential outliers or extreme values. By analyzing these statistical measures and visual representations, noisy values within the "Cholesterol" attribute could be detected for further investigation and data quality assurance.

Remove Duplicate Data:

sum(duplicated(heart_attack_data_5))

heart_attack_data_6 <- heart_attack_data_5[!duplicated(heart_attack_data_5),]

sum(duplicated(heart_attack_data_6))



Explanation: To eliminate duplicate data from the heart_attack_data dataset, a process of removing identical observations based on all variables can be performed. This helps ensure that each row in the dataset is unique and prevents any redundant information from skewing the analysis or results. By eliminating duplicate data, researchers can work with a more accurate and reliable dataset for further exploration and analysis. But fortunately there is no Duplicate data on the heart_attack_data dataset.

Conversion Data Type of Attributes:

heart_attack_data_6$RestingECG <- factor(heart_attack_data_6$RestingECG,levels=c('Normal','ST'),labels=c("NORMAL","ABNORMALITY"))

heart_attack_data_6$Sex <- factor(heart_attack_data_6$Sex,levels=c('Male','Female'),labels=c("MALE","FEMALE"))

heart_attack_data_6$ExerciseAngina <- factor(heart_attack_data_6$ExerciseAngina,levels=c('Y','N'),labels=c("YES","NO"))

unique(heart_attack_data_5$ChestPainType)

heart_attack_data_6$ChestPainType <- factor(heart_attack_data_6$ChestPainType,levels=c('ATA','NAP','ASY','TA'),labels=c("ATYPICAL ANGINA","NON-ANGINAL PAIN","ASYMPTONIC

","TYPICAL ANGINA"))

heart_attack_data_6$HeartDisease <- factor(heart_attack_data_6$HeartDisease,levels=c(1,0),labels=c("MORE CHANCE","LESS CHANCE"))

unique(heart_attack_data_5$ST_Slope)

heart_attack_data_6$ST_Slope <- factor(heart_attack_data_6$ST_Slope,levels=c('Up','Flat'),labels=c("UP","FLAT")) unique(heart_attack_data_5$FastingBS)

heart_attack_data_6$FastingBS <- factor(heart_attack_data_6$FastingBS,levels=c(0,1),labels=c("FALSE","TRUE"))

head(heart_attack_data_6)

```
Console   Terminal   Background Jobs
R 4.3.2 · D:/Data Science/
> heart_attack_data_6$RestingECG <- factor(heart_attack_data_6$RestingECG,levels=c('Normal','ST'),labels=c("NORMAL","ABNORMALI
TY"))
> heart_attack_data_6$Sex <- factor(heart_attack_data_6$Sex,levels=c('Male','Female'),labels=c("MALE","FEMALE"))
> heart_attack_data_6$ExerciseAngina <- factor(heart_attack_data_6$ExerciseAngina,levels=c('Y','N'),labels=c("YES","NO"))
> unique(heart_attack_data_5$ChestPainType)
[1] "ATA" "NAP" "ASY" "TA"
> heart_attack_data_6$ChestPainType <- factor(heart_attack_data_6$ChestPainType,levels=c('ATA','NAP','ASY','TA'),labels=c("ATY
PICAL ANGINA","NON-ANGINAL PAIN","ASYMPTONIC
+              ","TYPICAL ANGINA"))
> heart_attack_data_6$HeartDisease <- factor(heart_attack_data_6$HeartDisease,levels=c(1,0),labels=c("MORE CHANCE","LESS CHANC
E"))
> unique(heart_attack_data_5$ST_Slope)
[1] "Up"   "Flat"
> heart_attack_data_6$ST_Slope <- factor(heart_attack_data_6$ST_Slope,levels=c('Up','Flat'),labels=c("UP","FLAT"))
> unique(heart_attack_data_5$FastingBS)
[1] 0 1
> heart_attack_data_6$FastingBS <- factor(heart_attack_data_6$FastingBS,levels=c(0,1),labels=c("FALSE","TRUE"))
> head(heart_attack_data_6)
  Age    Sex    ChestPainType RestingBP Cholesterol FastingBS  RestingECG MaxHR ExerciseAngina Oldpeak
1  40   MALE  ATYPICAL ANGINA       140         289     FALSE      NORMAL   172             NO       0
2  49 FEMALE NON-ANGINAL PAIN       160         180     FALSE      NORMAL   156             NO       1
3  37   MALE  ATYPICAL ANGINA       130         283     FALSE ABNORMALITY    98             NO       0
5  54   MALE NON-ANGINAL PAIN      -150         195     FALSE      NORMAL   122             NO       0
6  39   MALE NON-ANGINAL PAIN       120         339     FALSE      NORMAL   170             NO       0
8  54   MALE  ATYPICAL ANGINA       110         208     FALSE      NORMAL   142             NO       0
  ST_Slope HeartDisease
1       UP  LESS CHANCE
2     FLAT  MORE CHANCE
3       UP  LESS CHANCE
5       UP  LESS CHANCE
6       UP  LESS CHANCE
8       UP  LESS CHANCE
```

Explanation: To ensure appropriate data representation, several attribute data types were converted in the heart_attack_data dataset. The "Sex" attribute was transformed into a factor variable with labels "MALE" and "FEMALE". The "ChestPainType" attribute was converted there full form after rounding. Similarly, the "RestingECG" attribute was changed to a factor variable with labels "NORMAL", "ABNORMALITY". The "ExerciseAngina" attribute was converted to a factor variable with labels "YES" and "NO". Lastly, the "HeartDisease" attribute was transformed into a factor variable with labels "HIGH CHANCE" and "LOW CHANCE". A preview of the updated dataset can be observed through the head of "heart_attack_data_6".

Remove Negative Values from Dataset:

heart_attack_data_7 <- heart_attack_data_6

heart_attack_data_7[heart_attack_data_6 < 0] <- NA

```
Console   Terminal   Background Jobs
R 4.3.2 · D:/Data Science/
> heart_attack_data_7 <- heart_attack_data_6
> heart_attack_data_7[heart_attack_data_6 < 0] <- NA
```

Before:

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG
1	40	MALE	ATYPICAL ANGINA	140	289	0	NORMAL
2	49	FEMALE	NON-ANGINAL PAIN	160	180	0	NORMAL
3	37	MALE	ATYPICAL ANGINA	130	283	0	STANDARD
5	54	MALE	NON-ANGINAL PAIN	-150	195	0	NORMAL
6	39	MALE	NON-ANGINAL PAIN	120	339	0	NORMAL
8	54	MALE	ATYPICAL ANGINA	110	208	0	NORMAL

After:

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG
1	40	MALE	ATYPICAL ANGINA	140	289	0	NORMAL
2	49	FEMALE	NON-ANGINAL PAIN	160	180	0	NORMAL
3	37	MALE	ATYPICAL ANGINA	130	283	0	STANDARD
5	54	MALE	NON-ANGINAL PAIN	NA	195	0	NORMAL
6	39	MALE	NON-ANGINAL PAIN	120	339	0	NORMAL
8	54	MALE	ATYPICAL ANGINA	110	208	0	NORMAL
10	48	FEMALE	ATYPICAL ANGINA	120	284	0	NORMAL

colSums(is.na(heart_attack_data_7))

```
> colSums(is.na(heart_attack_data_7))
          Age           Sex  ChestPainType      RestingBP    Cholesterol      FastingBS     RestingECG
            0             0              0              1              0              0              0
        MaxHR ExerciseAngina        Oldpeak       ST_Slope   HeartDisease
            0             0              0              0              0
>
```

missing_RestingBP <- is.na(heart_attack_data_7$RestingBP)

heart_attack_data_8 <- subset(heart_attack_data_7, !missing_RestingBP)

colSums(is.na(heart_attack_data_8))

```
Console  Terminal ×  Background Jobs ×
R 4.3.2 · D:/Data Science/
> missing_RestingBP <- is.na(heart_attack_data_7$RestingBP)
> heart_attack_data_8 <- subset(heart_attack_data_7, !missing_RestingBP)
> colSums(is.na(heart_attack_data_8))
           Age            Sex  ChestPainType      RestingBP    Cholesterol      FastingBS     RestingECG 
             0              0              0              0              0              0              0 
         MaxHR ExerciseAngina        Oldpeak       ST_Slope   HeartDisease 
             0              0              0              0              0 
>
```

Explanation: To eliminate Negative data from the heart_attack_data dataset, a process of removing negative data based on all variables can be performed. This helps ensure that each row in the dataset is Positive avlues and prevents any redundant information from skewing the analysis or results. By eliminating Negative data, researchers can work with a more accurate and reliable dataset for further exploration and analysis.

Univariate Data Exploration:

summary(heart_attack_data_8$Age)

summary(heart_attack_data_8$MaxHR)

```
Console  Terminal ×  Background Jobs ×
R 4.3.2 · D:/Data Science/
> summary(heart_attack_data_8$Age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  32.00   42.25   49.00   48.31   54.00   66.00 
> summary(heart_attack_data_8$MaxHR)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   82.0   124.2   142.0   140.8   159.5   190.0 
>
```

For Age attribute:

mean(heart_attack_data_8$Age)

median(heart_attack_data_8$Age)

var(heart_attack_data_8$Age)

sd(heart_attack_data_8$Age)

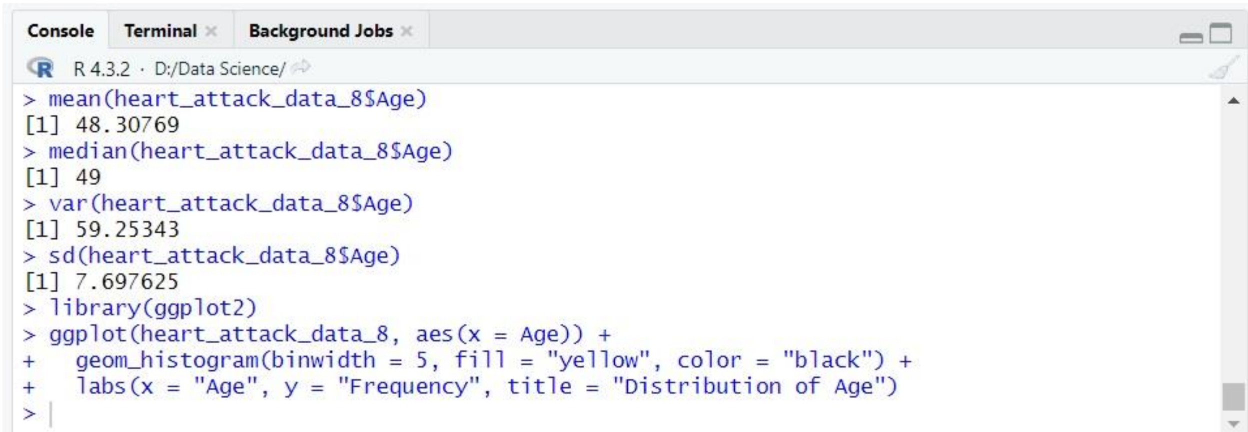
install.packages("ggplot2")

library(ggplot2)

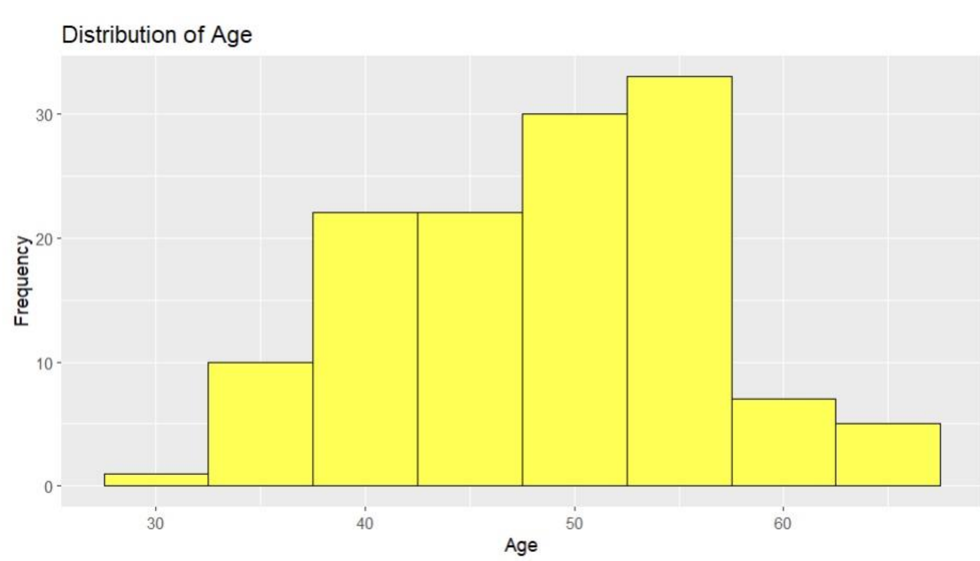
ggplot(heart_attack_data_8, aes(x = Age)) +

geom_histogram(binwidth = 5, fill = "yellow", color = "black") +

labs(x = "Age", y = "Frequency", title = "Distribution of Age")



```
Console   Terminal ×   Background Jobs ×
R 4.3.2 · D:/Data Science/
> mean(heart_attack_data_8$Age)
[1] 48.30769
> median(heart_attack_data_8$Age)
[1] 49
> var(heart_attack_data_8$Age)
[1] 59.25343
> sd(heart_attack_data_8$Age)
[1] 7.697625
> library(ggplot2)
> ggplot(heart_attack_data_8, aes(x = Age)) +
+   geom_histogram(binwidth = 5, fill = "yellow", color = "black") +
+   labs(x = "Age", y = "Frequency", title = "Distribution of Age")
>
```



For MaxHR attribute:

mean(heart_attack_data_8$MaxHR)

median(heart_attack_data_8$MaxHR)

var(heart_attack_data_8$MaxHR)

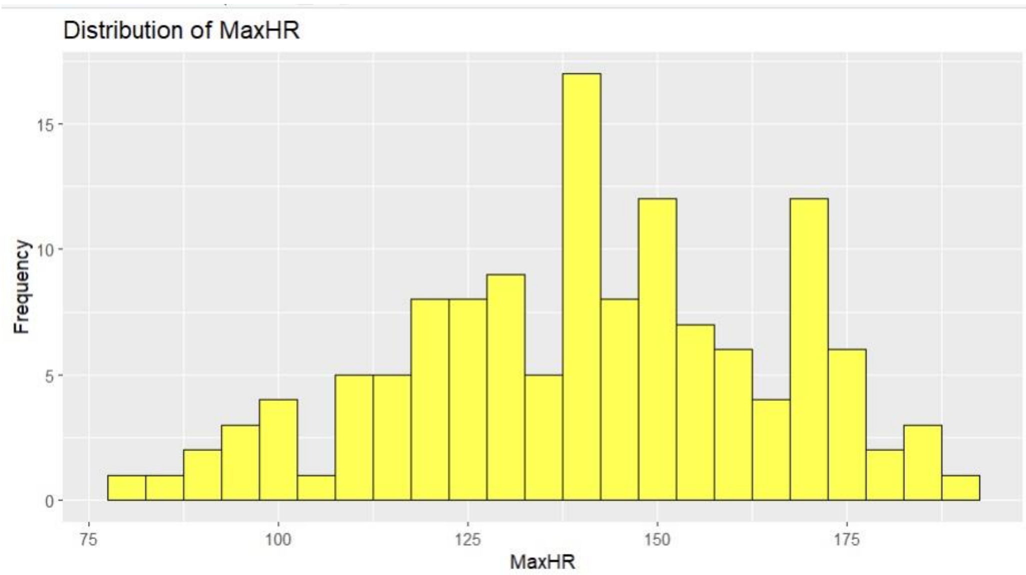
sd(heart_attack_data_8$MaxHR)

```
library(ggplot2)

ggplot(heart_attack_data_8, aes(x = MaxHR)) +

 geom_histogram(binwidth = 5, fill = "yellow", color = "black") +

labs(x = "MaxHR", y = "Frequency", title = "Distribution of MaxHR")
```



Explanation: Univariate data exploration involves examining individual variables in the dataset. Measures such as the mean, median, variance, and standard deviation provide insights into the central tendency, spread, and variability of the data. The "ggplot2" library is commonly used for creating visualizations, such as histograms, to visualize the distribution of variables. Additionally, labeling the axes and adding titles using "labs()" further enhances the interpretability of the visualizations, enabling a comprehensive exploration of the dataset.

Discussion and Conclusion:

The dataset that was given to us at the beginning of the study showed a significant level of disarray. It had a large number of outliers, missing values, and null values. A number of data preparation procedures were used to solve these problems and get the data ready for analysis. To learn more about each variable in the dataset, a univariate data exploration was also carried out on it.

Unprepared Dataset-

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
1	40	M	ATA	140	289	0	Normal	172	N	0.0	Up	0
2	49	F	NAP	160	180	0	Normal	156	N	1.0	Flat	1
3	37	M	ATA	130	283	0	ST	98	N	0.0	Up	0
4	NA	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
5	54	M	NAP	-150	195	0	Normal	122	N	0.0	Up	0
6	39	M	NAP	120	339	0	Normal	170	N	0.0	Up	0
7	45	F	ATA	130	237	0	Normal	170		0.0	Up	0
8	54	M	ATA	110	208	0	Normal	142	N	0.0	Up	0
9	37		ASY	140	207	0	Normal	130	Y	1.5	Flat	1
10	48	F	ATA	120	284	0	Normal	120	N	0.0	Up	0
11	37	F	NAP	130	1000	0	Normal	142	N	0.0	Up	0
12	58	M	ATA	136	1005	0	ST	99	Y	2.0	Flat	1
13	39	M	ATA	120	204	0	Normal	145	N	0.0	Up	0
14	49	M	ASY	140	234	0	Normal	140	Y	1.0	Flat	1
15	42	M	NAP	115	211	0	ST	137	N	0.0	Up	0
16	54	M	ATA	120	273	0	Normal	150	N	1.5	Flat	0
17	38	M	ASY	110	196	0	Normal	166	N	0.0	Flat	1
18	43	F	ATA	120	201	0	Normal	165	N	0.0	Up	0
19	60	M	ASY	100	248	0	Normal	125	N	1.0	Flat	1
20	36	M	ATA	120	267	0	Normal	160	N	3.0	Flat	1

Showing 1 to 20 of 150 entries, 12 total columns

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
131	54	M	NAP	145	292	0	Normal	130	N	0.0	Up	0
132	46	M	ASY	110	202	0	Normal	150	Y	0.0	Flat	1
133	56	M	ASY	170	388	0	ST	122	Y	2.0	Flat	1
134	56	M	ASY	150	230	0	ST	124	Y	1.5	Flat	1
135	61	F	ASY	130	294	0	ST	120	Y	1.0	Flat	0
136	54	M	NAP	115	265	0	Normal	175	N	0.0	Flat	1
137	43	F	ATA	120	215	0	ST	175	N	0.0	Up	0
138	39	M	ATA	120	241	0	ST	146	N	2.0	Up	0
139	54	M	ASY	140	166	0	Normal	118	Y	0.0	Flat	1
140	54	M	ASY	150	247	0	Normal	130	Y	2.0	Flat	1
141	52	M	ASY	160	331	0	Normal	94	Y	2.5	Flat	1
142	50	M	ASY	140	341	0	ST	125	Y	2.5	Flat	1
143	54	M	ASY	160	291	0	ST	158	Y	3.0	Flat	1
144	53	M	ASY	140	243	0	Normal	155	N	0.0	Up	0
145	54	F	ATA	120	279	0	Normal	150	N	1.0	Flat	1
146	39	M	ASY	110	273	0	Normal	132	N	0.0	Up	0
147	42	M	ATA	120	198	0	Normal	155	N	0.0	Up	0
148	43	F	ATA	120	249	0	ST	176	N	0.0	Up	0
149	50	M	ATA	120	168	0	Normal	160	N	0.0	Up	0
150	54	M	ASY	130	603	1	Normal	125	Y	1.0	Flat	1

Prepared Dataset-

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
1	40	MALE	ATYPICAL ANGINA	140	289	FALSE	NORMAL	172	NO	0.0	UP	LESS CHANCE
2	49	FEMALE	NON-ANGINAL PAIN	160	180	FALSE	NORMAL	156	NO	1.0	FLAT	MORE CHANCE
3	37	MALE	ATYPICAL ANGINA	130	283	FALSE	ABNORMALITY	98	NO	0.0	UP	LESS CHANCE
6	39	MALE	NON-ANGINAL PAIN	120	339	FALSE	NORMAL	170	NO	0.0	UP	LESS CHANCE
8	54	MALE	ATYPICAL ANGINA	110	208	FALSE	NORMAL	142	NO	0.0	UP	LESS CHANCE
10	48	FEMALE	ATYPICAL ANGINA	120	284	FALSE	NORMAL	120	NO	0.0	UP	LESS CHANCE
13	39	MALE	ATYPICAL ANGINA	120	204	FALSE	NORMAL	145	NO	0.0	UP	LESS CHANCE
14	49	MALE	ASYMPTONIC	140	234	FALSE	NORMAL	140	YES	1.0	FLAT	MORE CHANCE
15	42	MALE	NON-ANGINAL PAIN	115	211	FALSE	ABNORMALITY	137	NO	0.0	UP	LESS CHANCE
16	54	MALE	ATYPICAL ANGINA	120	273	FALSE	NORMAL	150	NO	1.5	FLAT	LESS CHANCE
17	38	MALE	ASYMPTONIC	110	196	FALSE	NORMAL	166	NO	0.0	FLAT	MORE CHANCE
18	43	FEMALE	ATYPICAL ANGINA	120	201	FALSE	NORMAL	165	NO	0.0	UP	LESS CHANCE
19	60	MALE	ASYMPTONIC	100	248	FALSE	NORMAL	125	NO	1.0	FLAT	MORE CHANCE
20	36	MALE	ATYPICAL ANGINA	120	267	FALSE	NORMAL	160	NO	3.0	FLAT	MORE CHANCE
21	43	MALE	TYPICAL ANGINA	100	223	FALSE	NORMAL	142	NO	0.0	UP	LESS CHANCE
22	44	MALE	ATYPICAL ANGINA	120	184	FALSE	NORMAL	142	NO	1.0	FLAT	LESS CHANCE
23	49	FEMALE	ATYPICAL ANGINA	124	201	FALSE	NORMAL	164	NO	0.0	UP	LESS CHANCE
25	40	MALE	NON-ANGINAL PAIN	130	215	FALSE	NORMAL	138	NO	0.0	UP	LESS CHANCE
28	52	MALE	ATYPICAL ANGINA	120	284	FALSE	NORMAL	118	NO	0.0	UP	LESS CHANCE
30	51	MALE	ATYPICAL ANGINA	125	188	FALSE	NORMAL	145	NO	0.0	UP	LESS CHANCE

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
118	54	FEMALE	ASYMPTONIC	130	338	TRUE	ABNORMALITY	130	YES	1.5	FLAT	MORE CHANCE
119	54	FEMALE	TYPICAL ANGINA	120	160	FALSE	ABNORMALITY	185	NO	0.0	UP	LESS CHANCE
120	34	MALE	TYPICAL ANGINA	140	156	FALSE	NORMAL	180	NO	0.0	FLAT	MORE CHANCE
121	47	FEMALE	NON-ANGINAL PAIN	135	248	TRUE	NORMAL	170	NO	0.0	FLAT	MORE CHANCE
122	52	FEMALE	NON-ANGINAL PAIN	125	272	FALSE	NORMAL	139	NO	0.0	UP	LESS CHANCE
123	46	MALE	ASYMPTONIC	110	240	FALSE	ABNORMALITY	140	NO	0.0	UP	LESS CHANCE
124	54	FEMALE	ATYPICAL ANGINA	180	393	FALSE	NORMAL	110	YES	1.0	FLAT	MORE CHANCE
125	58	MALE	ATYPICAL ANGINA	130	230	FALSE	NORMAL	150	NO	0.0	UP	LESS CHANCE
126	54	MALE	ATYPICAL ANGINA	120	246	FALSE	NORMAL	110	NO	0.0	UP	LESS CHANCE
127	34	FEMALE	ATYPICAL ANGINA	130	161	FALSE	NORMAL	190	NO	0.0	UP	LESS CHANCE
128	48	FEMALE	ASYMPTONIC	108	163	FALSE	NORMAL	175	NO	2.0	UP	LESS CHANCE
129	54	FEMALE	ATYPICAL ANGINA	120	230	TRUE	NORMAL	140	NO	0.0	UP	LESS CHANCE
130	42	MALE	NON-ANGINAL PAIN	120	228	FALSE	NORMAL	152	YES	1.5	FLAT	LESS CHANCE
131	54	MALE	NON-ANGINAL PAIN	145	292	FALSE	NORMAL	130	NO	0.0	UP	LESS CHANCE
132	46	MALE	ASYMPTONIC	110	202	FALSE	NORMAL	150	YES	0.0	FLAT	MORE CHANCE
133	56	MALE	ASYMPTONIC	170	388	FALSE	ABNORMALITY	122	YES	2.0	FLAT	MORE CHANCE
134	56	MALE	ASYMPTONIC	150	230	FALSE	ABNORMALITY	124	YES	1.5	FLAT	MORE CHANCE
135	61	FEMALE	ASYMPTONIC	130	294	FALSE	ABNORMALITY	120	YES	1.0	FLAT	LESS CHANCE
136	54	MALE	NON-ANGINAL PAIN	115	265	FALSE	NORMAL	175	NO	0.0	FLAT	MORE CHANCE
137	43	FEMALE	ATYPICAL ANGINA	120	215	FALSE	ABNORMALITY	175	NO	0.0	UP	LESS CHANCE
138	39	MALE	ATYPICAL ANGINA	120	241	FALSE	ABNORMALITY	146	NO	2.0	UP	LESS CHANCE
139	54	MALE	ASYMPTONIC	140	166	FALSE	NORMAL	118	YES	0.0	FLAT	MORE CHANCE
140	54	MALE	ASYMPTONIC	150	247	FALSE	NORMAL	130	YES	2.0	FLAT	MORE CHANCE