Assignment 1 - Import and Explore Data

Author   : Gayi Komi Selassi

ID   : RA2422021010003

Program : MSc Epidemiology & Biostatistics

Portfolio: Click Here

Access the GitHub Repository: Click Here

1. First 10 Observations of the HEART Dataset:

Age (years)	Sex (0 = Female, 1 = Male)	Chest Pain Type	Resting Blood Pressure (mm Hg)	Serum Cholesterol (mg/dL)	Fasting Blood Sugar > 120 mg/dL (1 = Yes, 0 = No)	Resting Electrocardiographic Results	Maximum Heart Rate Achieved
52	1	0	125	212	0	1	168
53	1	0	140	203	1	0	155
70	1	0	145	174	0	1	125
61	1	0	148	203	0	1	161
62	0	0	138	294	1	1	106
58	0	0	100	248	0	0	122
58	1	0	114	318	0	2	140
55	1	0	160	289	0	0	145
46	1	0	120	249	0	0	144
54	1	0	122	286	0	0	116

Exercise Induced Angina (1 = Yes, 0 = No)	ST Depression Induced by Exercise	Slope of Peak Exercise ST Segment	Number of Major Vessels Colored by Fluoroscopy	Thalassemia (3 = Normal, 6 = Fixed Defect, 7 = Reversible Defect)	Heart Disease Diagnosis (1 = Disease, 0 = No Disease)
0	1	2	2	3	0
1	3.1	0	0	3	0
1	2.6	0	0	3	0
0	0	2	1	3	0
0	1.9	1	3	2	0
0	1	1	0	2	1
0	4.4	0	3	1	0
1	0.8	1	1	3	0
0	0.8	2	0	3	0
1	3.2	1	2	2	0

Exploratory Analysis of the HEART Dataset: Assignment 1
2. Variable Attributes and Metadata

The CONTENTS Procedure

Data Set Name	DATA.HEART	Observations	1025
Member Type	DATA	Variables	14
Engine	V9	Indexes	0
Created	05/09/2025 15:29:50	Observation Length	112
Last Modified	05/09/2025 15:29:50	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information

Data Set Page Size	131072
Number of Data Set Pages	2
First Data Page	1
Max Obs per Page	1168
Obs in First Data Page	1025
Number of Data Set Repairs	0
Filename	/home/u64176007/sas_pratical_assignment/data/heart.sas7bdat
Release Created	9.0401M8
Host Created	Linux
Inode Number	6442451989
Access Permission	rw-r--r--
Owner Name	u64176007
File Size	384KB
File Size (bytes)	393216

Variables in Creation Order

#	Variable	Type	Len	Format	Informat	Label
1	age	Num	8	BEST12.	BEST32.	Age (years)
2	sex	Num	8	BEST12.	BEST32.	Sex (0 = Female, 1 = Male)
3	cp	Num	8	BEST12.	BEST32.	Chest Pain Type
4	trestbps	Num	8	BEST12.	BEST32.	Resting Blood Pressure (mm Hg)
5	chol	Num	8	BEST12.	BEST32.	Serum Cholesterol (mg/dL)
6	fbs	Num	8	BEST12.	BEST32.	Fasting Blood Sugar > 120 mg/dL (1 = Yes, 0 = No)
7	restecg	Num	8	BEST12.	BEST32.	Resting Electrocardiographic Results
8	thalach	Num	8	BEST12.	BEST32.	Maximum Heart Rate Achieved
9	exang	Num	8	BEST12.	BEST32.	Exercise Induced Angina (1 = Yes, 0 = No)
10	oldpeak	Num	8	BEST12.	BEST32.	ST Depression Induced by Exercise
11	slope	Num	8	BEST12.	BEST32.	Slope of Peak Exercise ST Segment
12	ca	Num	8	BEST12.	BEST32.	Number of Major Vessels Colored by Fluoroscopy
13	thal	Num	8	BEST12.	BEST32.	Thalassemia (3 = Normal, 6 = Fixed Defect, 7 = Reversible Defect)
14	target	Num	8	BEST12.	BEST32.	Heart Disease Diagnosis (1 = Disease, 0 = No Disease)

GAYI KOMI SELASSI, RA2422021010003, MSc Epidemiology & Biostatistics

COMMENT:

All variables are numeric (Type = Num).

Length (Len=8) indicates the memory allocated per variable.

Format/Informat = BEST12. means values are displayed with default numeric formatting.

Labels were added to make variables more interpretable.

Exploratory Analysis of the HEART Dataset: Assignment 1
3. Summary Statistics (Mean, Median, Minimum, Maximum)

The MEANS Procedure

Variable	Label	Mean	Median	Minimum	Maximum
age	Age (years)	54.43	56.00	29.00	77.00
sex	Sex (0 = Female, 1 = Male)	0.70	1.00	0.00	1.00
cp	Chest Pain Type	0.94	1.00	0.00	3.00
trestbps	Resting Blood Pressure (mm Hg)	131.61	130.00	94.00	200.00
chol	Serum Cholesterol (mg/dL)	246.00	240.00	126.00	564.00
fbs	Fasting Blood Sugar > 120 mg/dL (1 = Yes, 0 = No)	0.15	0.00	0.00	1.00
restecg	Resting Electrocardiographic Results	0.53	1.00	0.00	2.00
thalach	Maximum Heart Rate Achieved	149.11	152.00	71.00	202.00
exang	Exercise Induced Angina (1 = Yes, 0 = No)	0.34	0.00	0.00	1.00
oldpeak	ST Depression Induced by Exercise	1.07	0.80	0.00	6.20
slope	Slope of Peak Exercise ST Segment	1.39	1.00	0.00	2.00
ca	Number of Major Vessels Colored by Fluoroscopy	0.75	0.00	0.00	4.00
thal	Thalassemia (3 = Normal, 6 = Fixed Defect, 7 = Reversible Defect)	2.32	2.00	0.00	3.00
target	Heart Disease Diagnosis (1 = Disease, 0 = No Disease)	0.51	1.00	0.00	1.00

The MEANS Procedure

Variable	Label	N	N Miss
age	Age (years)	1025	0
sex	Sex (0 = Female, 1 = Male)	1025	0
cp	Chest Pain Type	1025	0
trestbps	Resting Blood Pressure (mm Hg)	1025	0
chol	Serum Cholesterol (mg/dL)	1025	0
fbs	Fasting Blood Sugar > 120 mg/dL (1 = Yes, 0 = No)	1025	0
restecg	Resting Electrocardiographic Results	1025	0
thalach	Maximum Heart Rate Achieved	1025	0
exang	Exercise Induced Angina (1 = Yes, 0 = No)	1025	0
oldpeak	ST Depression Induced by Exercise	1025	0
slope	Slope of Peak Exercise ST Segment	1025	0
ca	Number of Major Vessels Colored by Fluoroscopy	1025	0
thal	Thalassemia (3 = Normal, 6 = Fixed Defect, 7 = Reversible Defect)	1025	0
target	Heart Disease Diagnosis (1 = Disease, 0 = No Disease)	1025	0

COMMENT:

There is no missing value in the dataset.

GitHub Repository: Click Here

Portfolio: Click Here

GAYI KOMI SELASSI, RA2422021010003, MSc Epidemiology & Biostatistics