

Assignment 2 - Data Cleaning and Transformation

Author : Gayi Komi Selassi

ID : RA2422021010003

Program : MSc Epidemiology & Biostatistics

Portfolio: [Click Here](#)Access the GitHub Repository: [Click Here](#)**1. First 20 Observations of the new created variables:**

<i>Obs</i>	<i>Age</i>	<i>Age Group</i>	<i>Target</i>	<i>Heart Disease (Yes/No)</i>
1	52	Middle-Aged	0	No
2	53	Middle-Aged	0	No
3	70	Senior	0	No
4	61	Senior	0	No
5	62	Senior	0	No
6	58	Middle-Aged	1	Yes
7	58	Middle-Aged	0	No
8	55	Middle-Aged	0	No
9	46	Middle-Aged	0	No
10	54	Middle-Aged	0	No
11	71	Senior	1	Yes
12	43	Middle-Aged	0	No
13	34	Young	1	Yes
14	51	Middle-Aged	0	No
15	52	Middle-Aged	0	No
16	34	Young	1	Yes
17	51	Middle-Aged	1	Yes
18	54	Middle-Aged	0	No
19	50	Middle-Aged	1	Yes
20	58	Middle-Aged	1	Yes

COMMENT:

Age Group and Heart Disease character variables were successfully created

- < 40 = Young ; 40-60 = Middle-Aged ; >60 = Senior

- target (0) = No ; target (1) = Yes

Labels were added to make variables more interpretable.

Exploratory Analysis of the HEART Dataset: Assignment 2

2. Check for missing values in the new variables using missin option in te frequency procedure :

The FREQ Procedure

<i>Age_Group</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Frequency</i>	<i>Cumulative Percent</i>
<i>Middle-Aged</i>	696	67.90	696	67.90
<i>Senior</i>	272	26.54	968	94.44
<i>Young</i>	57	5.56	1025	100.00

<i>Heart_Disease</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Frequency</i>	<i>Cumulative Percent</i>
<i>No</i>	499	48.68	499	48.68
<i>Yes</i>	526	51.32	1025	100.00

COMMENT:

No missing value were identify in te new created variables

GitHub Repository: [Click Here](#)

Portfolio: [Click Here](#)