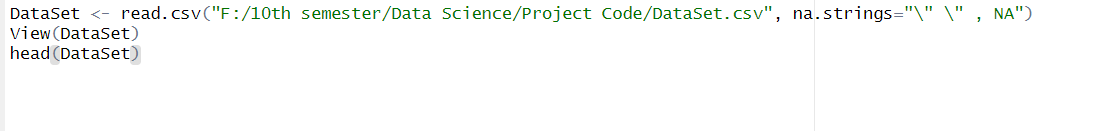
**Data Loading and Initial Overview**

**Code**

****

**Output**

**A computer screen shot of a computer code

AI-generated content may be incorrect.**

**Summarize Dataset**

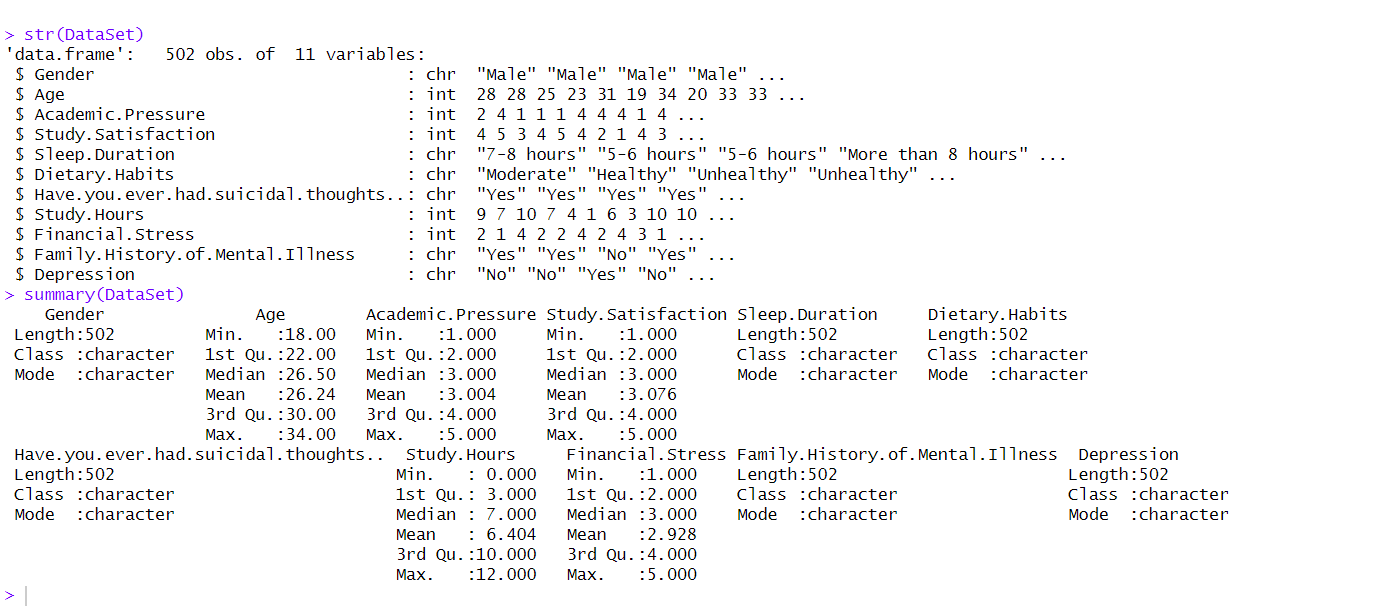
Provides a detailed summary of the dataset using skimr.

**Code**

**A black text on a white background

AI-generated content may be incorrect.**

**Output**

****

**Define Column Types**

**Code**

**A screenshot of a computer code

AI-generated content may be incorrect.**

**Output:**

**A screenshot of a computer code

AI-generated content may be incorrect.**

**Summary Statistics for Columns**

**Description**

Prints the summary of numeric columns and counts unique values in categorical columns.

**Code**

**A screenshot of a computer program

AI-generated content may be incorrect.**

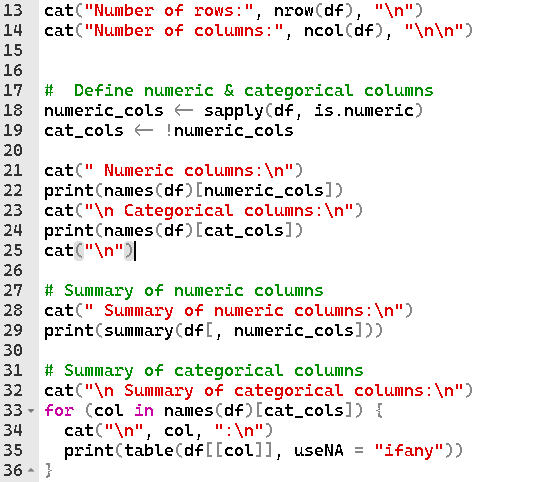
**Output**

**A screen shot of a computer code

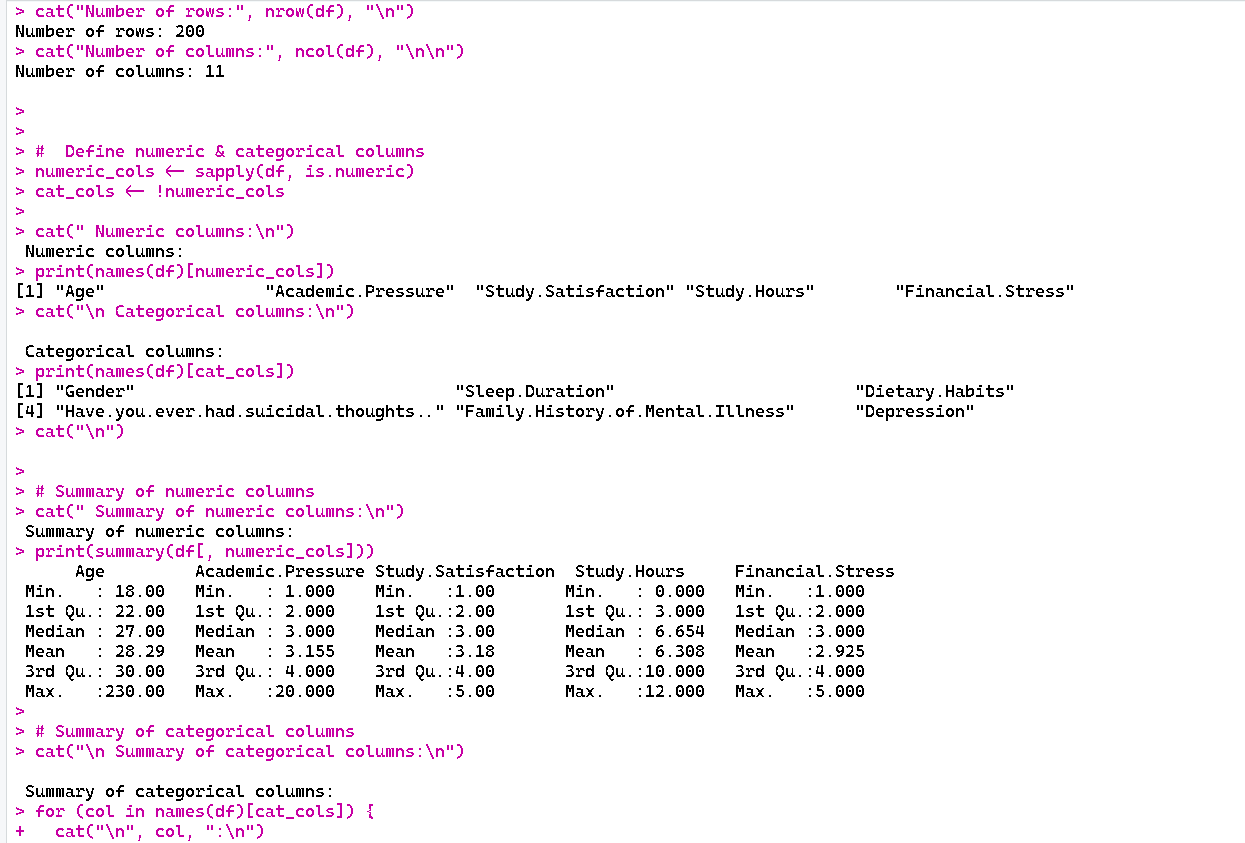
AI-generated content may be incorrect.**

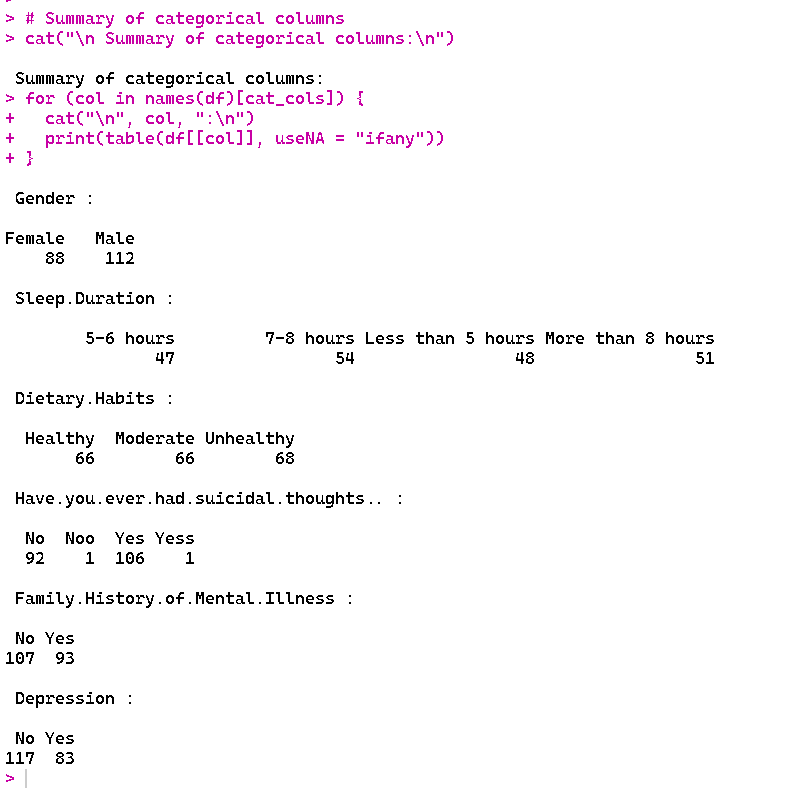
**Numeric And Categorical Column Define**

**Code**

****

**Output**

****

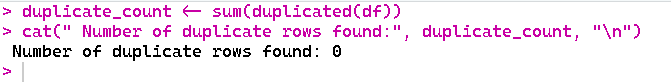
****

**Check for duplicate**

**Code**

****

**Output:**

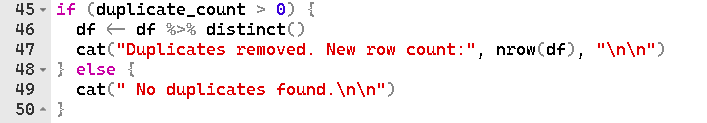
****

**Handle duplicate**

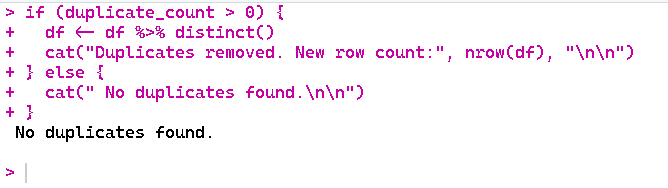
**Description**

remove duplicate row if there has any duplicate row.

**Code**

****

**Output:**

****

**Missing values**

**Description**

Check the null value and there has any missing value then Convert missing values to NA

**Code**

****

**Output:**

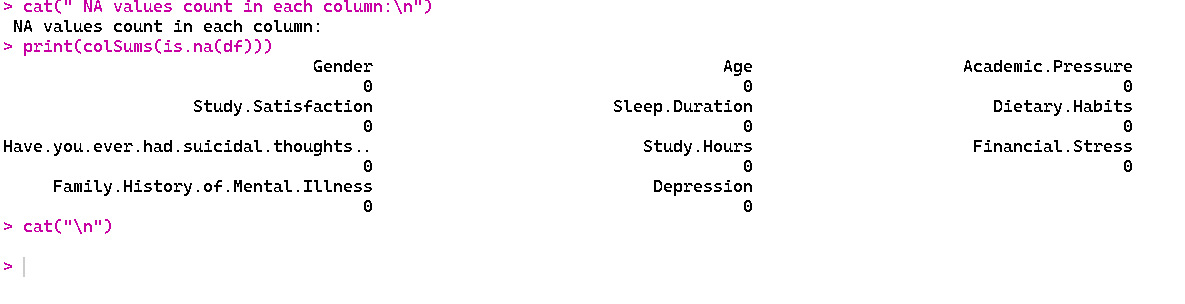
****

**Null Value**

**Code:**

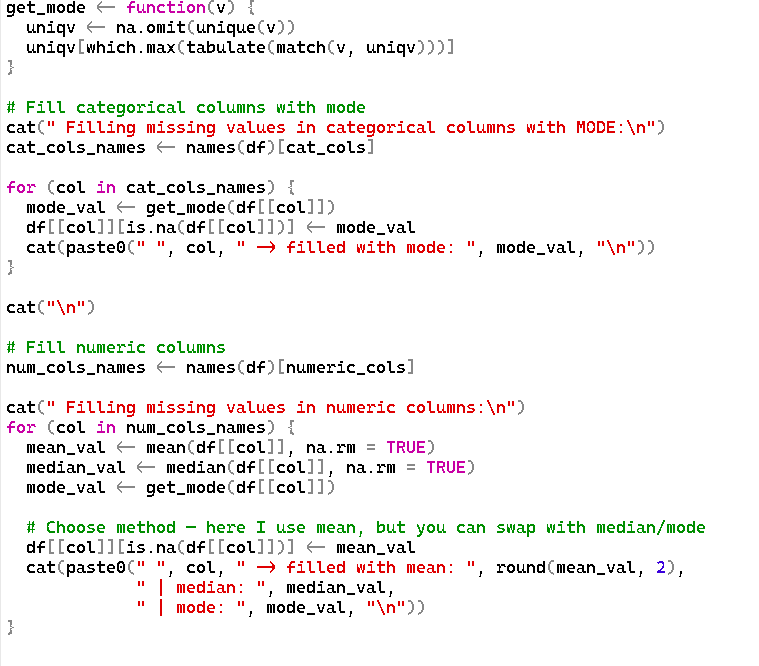
****

**Output**

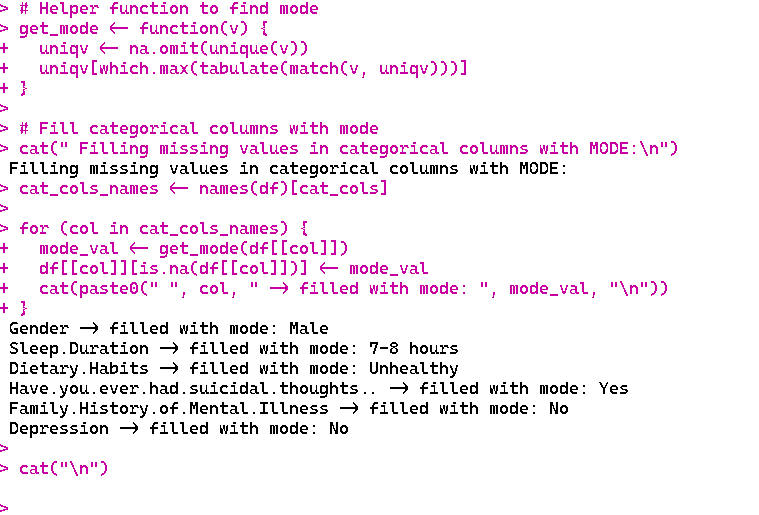
****

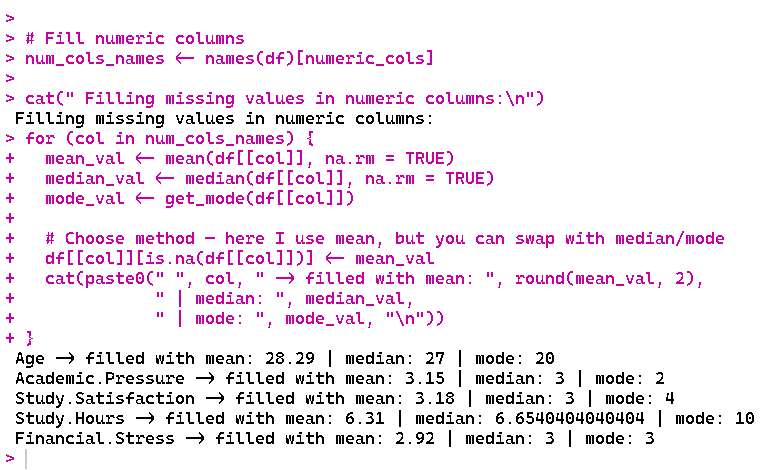
**Handle missing values**

**Code**

****

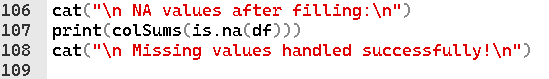
**Output**

****

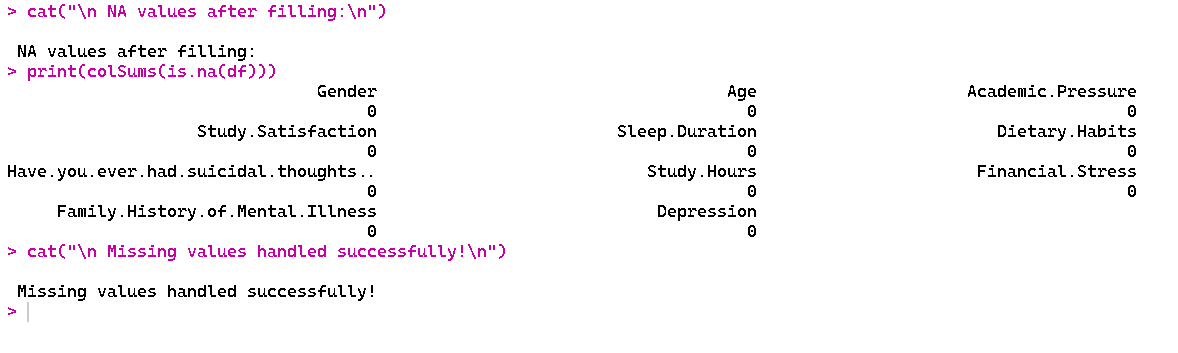
****

**Check Missing Value**

**Code**

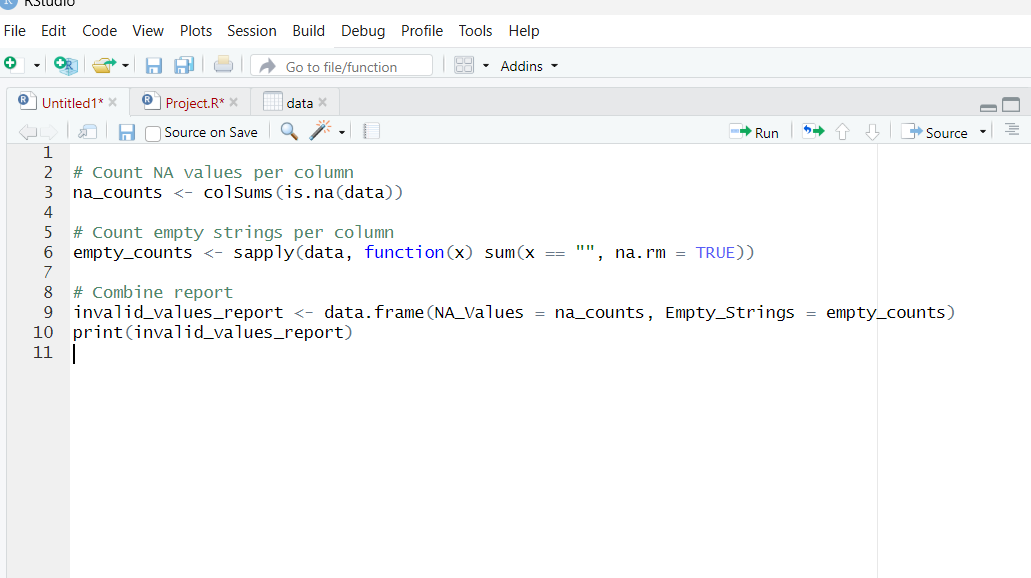
****

**Output**

****

**Invalid Value**

**Code**

****

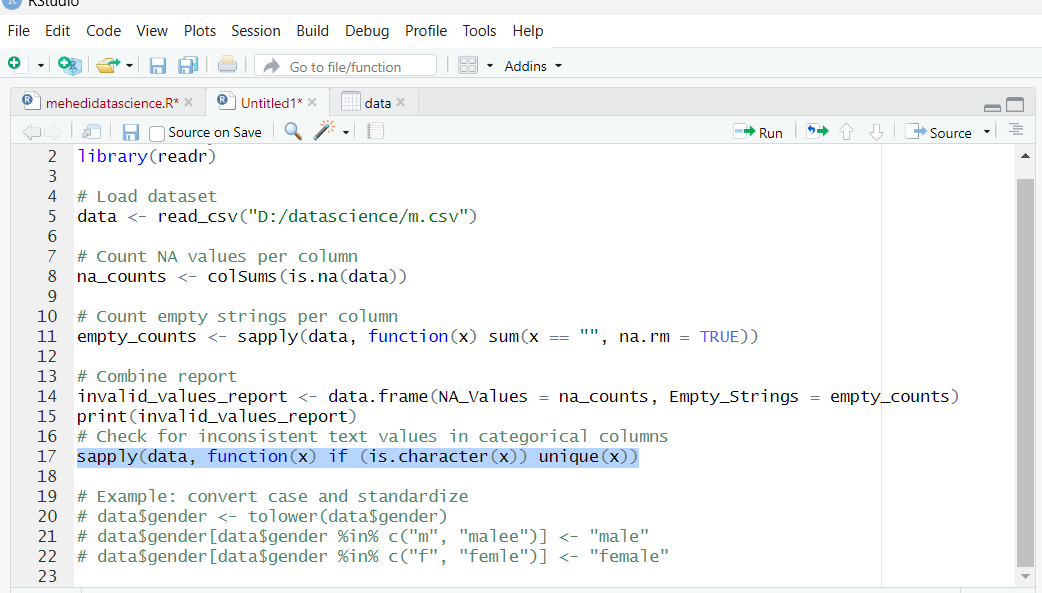
**Output**

**A screenshot of a computer

AI-generated content may be incorrect.**

**Noisy Value**

**Code**



**Output**

**A screenshot of a computer

AI-generated content may be incorrect.**

**Measure of central tendency**

**Code:**

A close-up of a computer code

Description automatically generated

**Output**

A white background with black and white clouds

Description automatically generated

**Measure of spread**

**Code**

A screen shot of a computer code

Description automatically generated

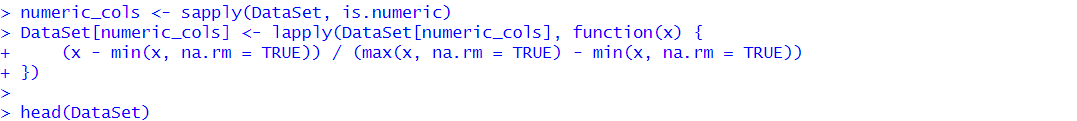
**Output**

A screenshot of a computer

Description automatically generated

**Normallization**

**Code**



**Output**

A white background with colorful text

Description automatically generated with medium confidence

**Outliers**

**Code**

A screenshot of a computer program

AI-generated content may be incorrect.

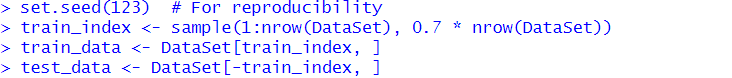
**Output**

**A white rectangular object with a black border

AI-generated content may be incorrect.**

**Train Test Split**.

**Code**



**Output**

A close-up of a computer code

Description automatically generated

**Sampling**

**Description**

**sampling** refers to randomly selecting a portion of the dataset’s rows to create subsets for model training and testing. It ensures that each row has an equal chance of being chosen, helping maintain the dataset’s overall distribution while preventing bias.

**Code**



**Output**

A screenshot of a computer

Description automatically generated