

SHETH L.U.J. AND SIR M.V. COLLEGE

Applying basic data cleaning functions: handling missing values using `na.omit()/replace_na()` in R. import dataset.

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins

Source
Console Terminal Background Jobs

R 4.5.2 - / ->

```
> install.packages("tidyverse")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:
https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/CHETAN/AppData/Local/R/win-library/4.5'
(as "lib" is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/tidyverse_1.3.1.zip'
Content type 'application/zip' length 1276404 bytes (1.2 MB)
downloaded 1.2 MB

package 'tidyverse' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:/Users/CHETAN/AppData/Local/Temp/Rtmpsg8l7p/downloaded_packages
> library(tidyverse)
> sleep_df <- read.csv("dataset_2191_sleep.csv", na.strings = c("", "NA"))
> sleep_df[1,]
#> #> 1. Original Data (First 6 Rows) ---"
> print(head(sleep_df))
#> #> #> body_weight brain_weight max_life_span gestation_time predation_index sleep_exposure_index
#> #> 1 6654.000 573.000 38.6 645 3 5
#> #> 2 3.000 6.6 45 42 3 1
#> #> 3 3.385 44.5 14 60 1 1
#> #> 4 0.920 5.7 ? 25 5 2
#> #> 5 2547.000 4603.0 69 624 3 5
#> #> 6 10.550 179.5 27 180 4 4
#> danger_index total_sleep
#> 1 3 3.3
#> 2 3 8.3
#> 3 1 12.5
#> 4 3 10.5
#> 5 4 3.9
#> 6 4 9.8
#> print("---- Count of Missing Values per Column ----")
```

Environment History Connections Tutorial
Import Dataset 1.25 GiB

merged_data 99 obs. of 27 variables
my_data 3000 obs. of 18 variables
new_rows 2 obs. of 14 variables
product_sales_dataset_200000 obs. of 14 variables
range_cols 20000 obs. of 10 variables
region_filter 3278 obs. of 6 variables
robot_related 1 obs. of 3 variables
selected_cols 20000 obs. of 4 variables
sleep_df 62 obs. of 8 variables

Files Plots Packages Help Viewer Presentation
New Folder New File Delete Rename More

Name Size Modified
PTROLICAL_N0_4B_JKX 573 B Nov 24, 2025, 8:42 PM
Practical_N0_5_R.R 728 B Nov 24, 2025, 7:53 PM
Practical_N0_5B_R.R 672 B Nov 24, 2025, 8:53 PM
Processed_Student_Mental_Health.csv 6.1 KB Nov 18, 2025, 10:58 AM
S093_CHEATAN_MANDAVKAR FEATURE ENGINEERING SUBMISSION - Colab.ipynb 2.3 MB Jul 27, 2025, 4:18 PM
S093_Chetan_Mandavkar_FP_Soft_Copy_Documentation_T1.docx 45.1 KB Oct 15, 2025, 8:30 PM
SHRUTI_MANDAVKAR_EVS_Report.pdf 357.3 KB Nov 21, 2025, 7:34 PM
Structure and functions of pancreas Shruti Mandavkar.pptx 639.7 KB Aug 24, 2025, 11:35 AM
Virtual Machines
WhatsApp Image 2025-11-25 at 19.51.57_2b9447da.jpg 17.9 KB Nov 25, 2025, 7:52 PM
Attendance_Prediction.csv 1.6 MB Dec 1, 2025, 10:48 AM
product_sales_dataset_final.csv 25.3 MB Dec 1, 2025, 10:52 AM
Practical_N0_6_R.R 696 B Dec 1, 2025, 11:07 AM
synthetic_fraud_dataset.csv 847.7 KB Dec 1, 2025, 11:09 AM
Practical_N0_7_R.R 1.1 KB Dec 1, 2025, 11:49 AM
dataset_2191.sleep.csv 1.7 KB Dec 1, 2025, 11:39 AM
Practical_N0_8_R.R 1.2 KB Dec 1, 2025, 11:51 AM

27°C Sunny
Search ENG IN 11:55 01-12-2025

The screenshot shows an RStudio interface with the following details:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Source Tab:** Contains the R code for the current session.
- Console Tab:** Displays the R session output, including the count of missing values per column and the final dataset structure.
- Terminal Tab:** Not visible in the screenshot.
- Background Jobs:** Shows a single job named "R 4.5.2" with ID 4.
- Environment Tab:** Lists global variables and their characteristics.
- History Tab:** Shows a history of previous sessions.
- Connections Tab:** Shows connections to databases.
- Tutorial Tab:** Shows available tutorials.
- File Browser:** Shows the project structure with files like "Practical_CW_48.Rnw", "Practical_No_5.R", etc.

NAME:- CHETAN MANDAVKAR
ROLL NO. S093
SUBJECT:- Data Analysis with SAS / SPSS /R

SHETH L.U.J. AND SIR M.V. COLLEGE

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Console Terminal Background Jobs

```

> clean_replace <- sleep_df %>
+   replace_na(list(
+     body_weight = mean(sleep_df$body_weight, na.rm = TRUE),
+     brain_weight = mean(sleep_df$brain_weight, na.rm = TRUE),
+     max_life_span = mean(sleep_df$max_life_span, na.rm = TRUE),
+     gestation_time = mean(sleep_df$gestation_time, na.rm = TRUE),
+     predation_index = mean(sleep_df$predation_index, na.rm = TRUE),
+     sleep_exposure_index = mean(sleep_df$sleep_exposure_index, na.rm = TRUE),
+     danger_index = mean(sleep_df$danger_index, na.rm = TRUE),
+     total_sleep = mean(sleep_df$total_sleep, na.rm = TRUE)
+   ))
Warning messages:
1: In mean.default(sleep_df$max_life_span, na.rm = TRUE) :
  argument is not numeric or logical: returning NA
2: In mean.default(sleep_df$gestation_time, na.rm = TRUE) :
  argument is not numeric or logical: returning NA
3: In mean.default(sleep_df$total_sleep, na.rm = TRUE) :
  argument is not numeric or logical: returning NA
> print("--- 3. Data after replace_na() ---")
[1] "--- 3. Data after replace_na() ---"
> print(cad(clean_replace))
  body_weight brain_weight max_life_span gestation_time predation_index sleep_exposure_index
1    6654.000      5712.0       38.6        645            3             5
2     1.000         6.6        4.5        42            3             1
3     3.385        44.5       14          60            1             1
4     0.920        5.7         ?          25            5             2
5   2547.000      4603.0       69          624            3             5
6    10.550       179.5       27          180            4             4
  danger_index total_sleep
1            3           3.3
2            3           8.3
3            1          12.5
4            3          16.5
5            4           3.9
6            4           9.8
> print("--- Remaining NAs after replacement ---")
[1] "--- Remaining NAs after replacement ---"
> print(colSums(is.na(clean_replace)))

```

Files Plots Packages Help Viewer Presentation

New Folder New File Delete Rename More

Home

Name	Size	Modified
Practical_No_4_R.R	913 B	Nov 24, 2025, 0:42 PM
Practical_No_5_R.R	728 B	Nov 24, 2025, 7:53 PM
Practical_No_5B_R.R	672 B	Nov 24, 2025, 8:53 PM
Processed_Student_Mental_Health.csv	6.1 KB	Nov 18, 2025, 10:58 AM
S093 CHETAN MANDAVKAR FEATURE ENGINEERING SUBMISSION - Colab 2.3 MB		Jul 27, 2025, 4:18 PM
S093_Chetan_Mandavkar_FP_Soft_Copy_Documentation_1[1].docx	45.1 KB	Oct 15, 2025, 8:30 PM
SHRUTI MANDAVKAR EVS REPORT.pdf	357.3 KB	Nov 21, 2025, 7:34 PM
Structure and functions of pancreas Shruti Mandavkar.pptx	639.7 KB	Aug 24, 2025, 11:35 AM
Virtual Machines		
WhatsApp Image 2025-11-25 at 19.51.57_2b9447da.jpg	17.9 KB	Nov 25, 2025, 7:52 PM
Attendance_Prediction.csv	1.6 MB	Dec 1, 2025, 10:48 AM
product_sales_dataset_final.csv	25.3 MB	Dec 1, 2025, 10:52 AM
Practical_No_6_R.R	696 B	Dec 1, 2025, 11:07 AM
synthetic_fraud_dataset.csv	847.7 KB	Dec 1, 2025, 11:09 AM
Practical_No_7_R.R	1.1 KB	Dec 1, 2025, 11:49 AM
dataset_2191_sleep.csv	1.7 KB	Dec 1, 2025, 11:39 AM
Practical_No_8_R.R	1.2 KB	Dec 1, 2025, 11:51 AM

4 27°C Sunny ENG IN 11:56 01-12-2025

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Console Terminal Background Jobs

```

> clean_replace <- sleep_df %>
+   replace_na(list(
+     gestation_time = mean(sleep_df$gestation_time, na.rm = TRUE),
+     predation_index = mean(sleep_df$predation_index, na.rm = TRUE),
+     sleep_exposure_index = mean(sleep_df$sleep_exposure_index, na.rm = TRUE),
+     danger_index = mean(sleep_df$danger_index, na.rm = TRUE),
+     total_sleep = mean(sleep_df$total_sleep, na.rm = TRUE)
+   ))
Warning messages:
1: In mean.default(sleep_df$max_life_span, na.rm = TRUE) :
  argument is not numeric or logical: returning NA
2: In mean.default(sleep_df$gestation_time, na.rm = TRUE) :
  argument is not numeric or logical: returning NA
3: In mean.default(sleep_df$total_sleep, na.rm = TRUE) :
  argument is not numeric or logical: returning NA
> print("--- 3. Data after replace_na() ---")
[1] "--- 3. Data after replace_na() ---"
> print(cad(clean_replace))
  body_weight brain_weight max_life_span gestation_time predation_index sleep_exposure_index
1    6654.000      5712.0       38.6        645            3             5
2     1.000         6.6        4.5        42            3             1
3     3.385        44.5       14          60            1             1
4     0.920        5.7         ?          25            5             2
5   2547.000      4603.0       69          624            3             5
6    10.550       179.5       27          180            4             4
  danger_index total_sleep
1            3           3.3
2            3           8.3
3            1          12.5
4            3          16.5
5            4           3.9
6            4           9.8
> print("--- Remaining NAs after replacement ---")
[1] "--- Remaining NAs after replacement ---"
> print(colSums(is.na(clean_replace)))
  body_weight brain_weight max_life_span gestation_time predation_index sleep_exposure_index
0            0           0           0           0           0           0
  danger_index total_sleep
0            0           0
>

```

Files Plots Packages Help Viewer Presentation

New Folder New File Delete Rename More

Home

Name	Size	Modified
Practical_No_4_R.R	913 B	Nov 24, 2025, 0:42 PM
Practical_No_5_R.R	728 B	Nov 24, 2025, 7:53 PM
Practical_No_5B_R.R	672 B	Nov 24, 2025, 8:53 PM
Processed_Student_Mental_Health.csv	6.1 KB	Nov 18, 2025, 10:58 AM
S093 CHETAN MANDAVKAR FEATURE ENGINEERING SUBMISSION - Colab 2.3 MB		Jul 27, 2025, 4:18 PM
S093_Chetan_Mandavkar_FP_Soft_Copy_Documentation_1[1].docx	45.1 KB	Oct 15, 2025, 8:30 PM
SHRUTI MANDAVKAR EVS REPORT.pdf	357.3 KB	Nov 21, 2025, 7:34 PM
Structure and functions of pancreas Shruti Mandavkar.pptx	639.7 KB	Aug 24, 2025, 11:35 AM
Virtual Machines		
WhatsApp Image 2025-11-25 at 19.51.57_2b9447da.jpg	17.9 KB	Nov 25, 2025, 7:52 PM
Attendance_Prediction.csv	1.6 MB	Dec 1, 2025, 10:48 AM
product_sales_dataset_final.csv	25.3 MB	Dec 1, 2025, 10:52 AM
Practical_No_6_R.R	696 B	Dec 1, 2025, 11:07 AM
synthetic_fraud_dataset.csv	847.7 KB	Dec 1, 2025, 11:09 AM
Practical_No_7_R.R	1.1 KB	Dec 1, 2025, 11:49 AM
dataset_2191_sleep.csv	1.7 KB	Dec 1, 2025, 11:39 AM
Practical_No_8_R.R	1.2 KB	Dec 1, 2025, 11:51 AM

4 27°C Sunny ENG IN 11:57 01-12-2025

NAME:- CHETAN MANDAVKAR
ROLL NO. S093
SUBJECT:- Data Analysis with SAS / SPSS /R