

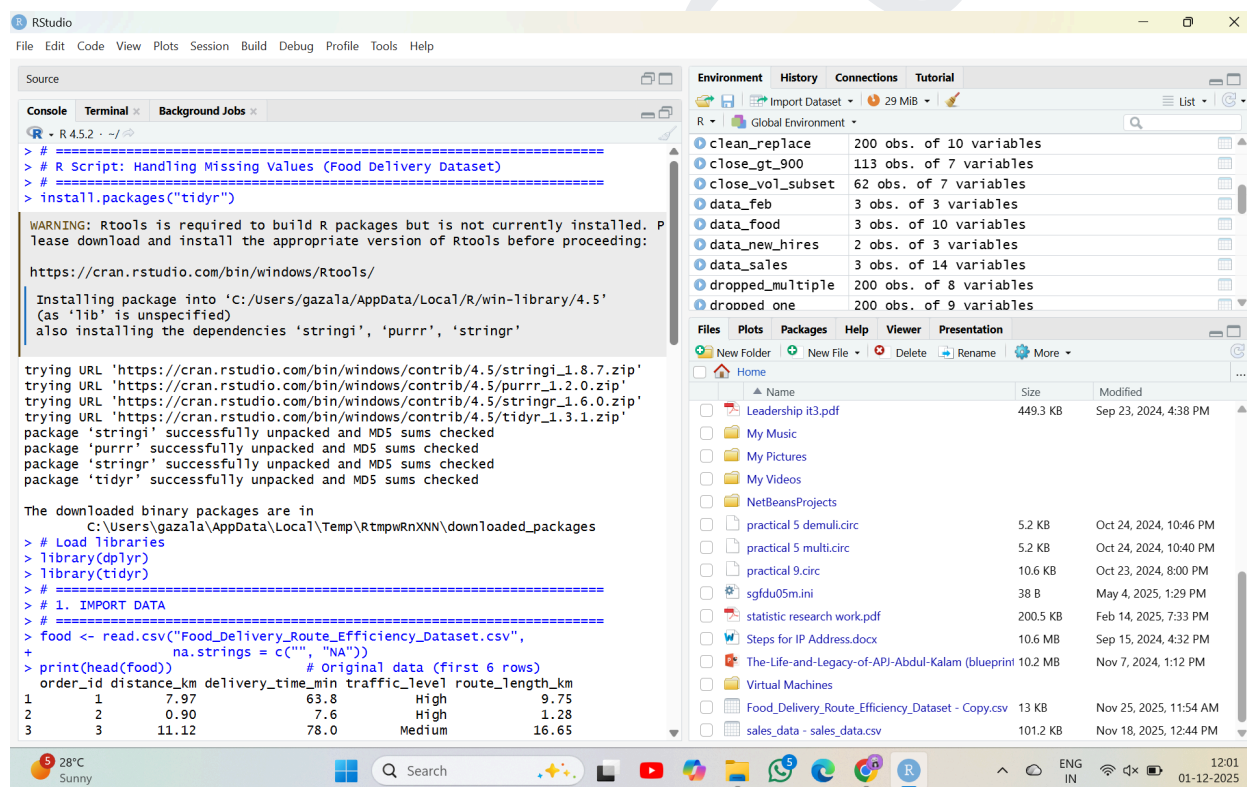
SHETH L.U.J AND SIR M.V COLLEGE

Subject: Data Analysis with SAS / SPSS /R

Practical no. 8

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

Outputs→



The screenshot displays the RStudio environment with the following components:

- Source:** Contains the R script being executed.
- Console:** Shows the output of the R code, including package installation messages and the head of the imported dataset.
- Environment:** Lists the objects in the global environment, including datasets like 'clean_replace', 'close_gt_900', 'close_vol_subset', 'data_feb', 'data_food', 'data_new_hires', 'data_sales', 'dropped_multiple', and 'drooped_one'.
- Files:** Shows the file explorer with various files and folders.

Console Output:

```
R - R 4.5.2 - ~/R
> # =====
> # R Script: Handling Missing Values (Food Delivery Dataset)
> # =====
> install.packages("tidyr")

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/gazala/AppData/Local/R/win-library/4.5'
(as 'lib' is unspecified)
also installing the dependencies 'stringi', 'purrr', 'stringr'

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/stringi_1.8.7.zip'
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/purrr_1.2.0.zip'
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/stringr_1.6.0.zip'
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/tidyr_1.3.1.zip'
package 'stringi' successfully unpacked and MD5 sums checked
package 'purrr' successfully unpacked and MD5 sums checked
package 'stringr' successfully unpacked and MD5 sums checked
package 'tidyr' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:/Users/gazala/AppData/Local/Temp/RtmpwRnXNN/downloaded_packages
> # Load libraries
> library(dplyr)
> library(tidyr)
> # =====
> # 1. IMPORT DATA
> # =====
> food <- read.csv("Food_Delivery_Route_Efficiency_Dataset.csv",
+               na.strings = c("", "NA"))
> print(head(food)) # Original data (first 6 rows)
  order_id distance_km delivery_time_min traffic_level route_length_km
1         1         7.97          63.8           High           9.75
2         2         0.90           7.6           High           1.28
3         3        11.12          78.0        Medium          16.65
```

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

Subject: Data Analysis with SAS / SPSS /R

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help

Source
Console Terminal Background Jobs
R - R4.5.2 - /
> # =====
> # 1. IMPORT DATA
> # =====
> food <- read.csv("Food_Delivery_Route_Efficiency_Dataset.csv",
+                 na.strings = c("", "NA"))
> print(head(food)) # Original data (first 6 rows)
  order_id distance_km delivery_time_min traffic_level route_length_km
1         1       7.97           63.8           High           9.75
2         2       0.90           7.6            High           1.28
3         3      11.12          78.0           Medium          16.65
4         4       4.90          24.8            Low            5.25
5         5      10.04          56.0           High           11.34
6         6      10.96          76.8           High           13.62
  delivery_mode weather order_time restaurant_zone customer_zone
1     Bicycle   Clear 2025-01-01 15:29             South      North
2         Car   Cloudy 2025-01-03 00:47             West      North
3         Bike   Rainy 2025-01-04 17:32             South      Central
4     Scooter   Rainy 2025-01-01 14:12             Central    Central
5         Car   Rainy 2025-01-02 16:50             West      North
6         Car   Windy 2025-01-02 09:56             West      North
> print(colSums(is.na(food))) # Count missing values per column
  order_id distance_km delivery_time_min traffic_level
0         0         0         0         0
  route_length_km delivery_mode weather order_time
0         0         0         0         0
  restaurant_zone customer_zone
0         0         0
> # =====
> # 2. METHOD A: REMOVE MISSING VALUES
> # =====
> clean_omit <- na.omit(food)
> print(paste("Original rows:", nrow(food)))
[1] "Original rows: 200"
> print(paste("Rows after na.omit:", nrow(clean_omit)))
[1] "Rows after na.omit: 200"
> print(head(clean_omit))
  order_id distance_km delivery_time_min traffic_level route_length_km
1         1       7.97           63.8           High           9.75
  delivery_mode weather order_time restaurant_zone customer_zone
1     Bicycle   Clear 2025-01-01 15:29             South      North
2         Car   Cloudy 2025-01-03 00:47             West      North
3         Bike   Rainy 2025-01-04 17:32             South      Central
4     Scooter   Rainy 2025-01-01 14:12             Central    Central
5         Car   Rainy 2025-01-02 16:50             West      North
6         Car   Windy 2025-01-02 09:56             West      North

Environment History Connections Tutorial
R - Global Environment
clean_replace 200 obs. of 10 variables
close_gt_900 113 obs. of 7 variables
close_vol_subset 62 obs. of 7 variables
data_feb 3 obs. of 3 variables
data_food 3 obs. of 10 variables
data_new_hires 2 obs. of 3 variables
data_sales 3 obs. of 14 variables
dropped_multiple 200 obs. of 8 variables
dropped_one 200 obs. of 9 variables
Files Plots Packages Help Viewer Presentation
New Folder New File Delete Rename More
Home
Name Size Modified
Leadership it3.pdf 449.3 KB Sep 23, 2024, 4:38 PM
My Music
My Pictures
My Videos
NetBeansProjects
practical 5 demul.circ 5.2 KB Oct 24, 2024, 10:46 PM
practical 5 multi.circ 5.2 KB Oct 24, 2024, 10:40 PM
practical 9.circ 10.6 KB Oct 23, 2024, 8:00 PM
sgfdu05m.ini 38 B May 4, 2025, 1:29 PM
statistic research work.pdf 200.5 KB Feb 14, 2025, 7:33 PM
Steps for IP Address.docx 10.6 MB Sep 15, 2024, 4:32 PM
The-Life-and-Legacy-of-APJ-Abdul-Kalam (blueprint) 10.2 MB Nov 7, 2024, 1:12 PM
Virtual Machines
Food_Delivery_Route_Efficiency_Dataset - Copy.csv 13 KB Nov 25, 2025, 11:54 AM
sales_data - sales_data.csv 101.2 KB Nov 18, 2025, 12:44 PM
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help

Source
Console Terminal Background Jobs
R - R4.5.2 - /
> # =====
> # 3. METHOD B: REPLACE MISSING VALUES
> # =====
> avg_time <- mean(food$delivery_time_min, na.rm = TRUE)
> avg_dist <- mean(food$distance_km, na.rm = TRUE)
> clean_replace <- food %>%
+   replace_na(list(
+     traffic_level = "Unknown",
+     delivery_mode = "Not specified",
+     weather = "Clear",
+     delivery_time_min = avg_time,
+     distance_km = avg_dist
+   ))
> print(head(clean_replace)) # Data after replacing missing values
  order_id distance_km delivery_time_min traffic_level route_length_km
1         1       7.97           63.8           High           9.75
2         2       0.90           7.6            High           1.28
3         3      11.12          78.0           Medium          16.65
4         4       4.90          24.8            Low            5.25
5         5      10.04          56.0           High           11.34
6         6      10.96          76.8           High           13.62
  delivery_mode weather order_time restaurant_zone customer_zone
1     Bicycle   Clear 2025-01-01 15:29             South      North
2         Car   Cloudy 2025-01-03 00:47             West      North
3         Bike   Rainy 2025-01-04 17:32             South      Central
4     Scooter   Rainy 2025-01-01 14:12             Central    Central
5         Car   Rainy 2025-01-02 16:50             West      North
6         Car   Windy 2025-01-02 09:56             West      North

Environment History Connections Tutorial
R - Global Environment
clean_replace 200 obs. of 10 variables
close_gt_900 113 obs. of 7 variables
close_vol_subset 62 obs. of 7 variables
data_feb 3 obs. of 3 variables
data_food 3 obs. of 10 variables
data_new_hires 2 obs. of 3 variables
data_sales 3 obs. of 14 variables
dropped_multiple 200 obs. of 8 variables
dropped_one 200 obs. of 9 variables
Files Plots Packages Help Viewer Presentation
New Folder New File Delete Rename More
Home
Name Size Modified
Leadership it3.pdf 449.3 KB Sep 23, 2024, 4:38 PM
My Music
My Pictures
My Videos
NetBeansProjects
practical 5 demul.circ 5.2 KB Oct 24, 2024, 10:46 PM
practical 5 multi.circ 5.2 KB Oct 24, 2024, 10:40 PM
practical 9.circ 10.6 KB Oct 23, 2024, 8:00 PM
sgfdu05m.ini 38 B May 4, 2025, 1:29 PM
statistic research work.pdf 200.5 KB Feb 14, 2025, 7:33 PM
Steps for IP Address.docx 10.6 MB Sep 15, 2024, 4:32 PM
The-Life-and-Legacy-of-APJ-Abdul-Kalam (blueprint) 10.2 MB Nov 7, 2024, 1:12 PM
Virtual Machines
Food_Delivery_Route_Efficiency_Dataset - Copy.csv 13 KB Nov 25, 2025, 11:54 AM
sales_data - sales_data.csv 101.2 KB Nov 18, 2025, 12:44 PM
```

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

Subject: Data Analysis with SAS / SPSS /R

gazala S071