

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 shows the RStudio interface with the following details:

- Console Tab:** Displays R code and its output. The code includes installing packages like tidyverse, handling missing values, and importing a CSV file named "Food_Delivery_Route_Efficiency_Dataset.csv". The output shows the dataset has 200 rows and 10 variables.
- File Explorer (File View):** Shows a list of files and folders in the current directory:
 - practical 5 demulti.circ (5.2 KB)
 - practical 5 multi.circ (5.2 KB)
 - practical 9.circ (10.6 KB)
 - sgfdlu05m.ini (38 B)
 - statistic research work.pdf (200.5 KB)
 - Steps for IP Address.docx (10.6 MB)
 - The-Life-and-Legacy-of-API-Abdul-Kalam (blueprint) (10.2 MB)
 - Virtual Machines
 - Food_Delivery_Route_Efficiency_Dataset - Copy.csv (13 KB)
 - sales_data - sales_data.csv (101.2 KB)
- Environment View:** Shows the global environment with various objects and their characteristics.
- Bottom Status Bar:** Displays system information including weather (28°C, Sunny), search bar, and system icons.

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

Subject: Data Analysis with SAS / SPSS /R

```

R + R 4.5.2 · ~/R
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
> # =====
> # 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
#> Bicycle Clear 2025-01-01 15:29 South North
#> Car Cloudy 2025-01-03 00:47 West North
#> Bike Rainy 2025-01-04 17:32 South Central
#> Scooter Rainy 2025-01-01 14:12 Central Central
#> Car Rainy 2025-01-02 16:50 West North
#> 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
route_length_km delivery_mode weather order_time
0 0 0 0
restaurant_zone customer_zone
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

```

```

R + R 4.5.2 · ~/R
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
> 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
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
#> Bicycle Clear 2025-01-01 15:29 South North
#> Car Cloudy 2025-01-03 00:47 West North
#> Bike Rainy 2025-01-04 17:32 South Central
#> Scooter Rainy 2025-01-01 14:12 Central Central
#> Car Rainy 2025-01-02 16:50 West North
#> Car Windy 2025-01-02 09:56 West North
#> # =====
#> # 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
#> deliveryv mode weather order time restaurant zone customer zone

```

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

Subject: Data Analysis with SAS / SPSS /R

gazala S071