

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

## SUBJECT:Data Analysis with R

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

**Session 1 (Left): Original Dataset**

```
R > retail_df <- read.csv("student_admission_record_dirty.csv", na.strings = c("", "NA"))
> print(head(retail_df))
[1] "---- 1. Original Data (First 6 Rows) ----"
> print(head(retail_df))
Name Age Gender Admission.Test.Score High.School.Percentage City Admission.Status
1 Shehroz 24 Female 50 68.90 Quetta Rejected
2 Waqar 21 Female 99 60.73 Karachi NA
3 Bushra 17 Male 89 NA Islamabad Accepted
4 Aliya 17 Male 55 85.29 Karachi Rejected
5 Bilal 20 Male 65 61.13 Lahore NA
6 Murtaza 23 Female NA NA Islamabad Accepted
7 Asad 18 Male NA 97.31 Multan Accepted
8 Rabia 20 Female 82 55.67 Lahore Accepted
9 Rohail 17 Male 64 NA Karachi Accepted
10 Kamran 18 Male 53 98.98 Multan Rejected
11 Shafiq 17 Male 78 -10.00 Quetta Rejected
12 Nashit 18 Male 89 NA Lahore Accepted
13 NA 19 Male 66 88.17 NA NA
14 Ahmed 21 Male 62 79.03 Karachi Accepted
15 Fareeha 22 Male 86 50.77 Islamabad Rejected
16 Salm 19 Male 84 97.19 Peshawar Rejected
```

**Session 2 (Right): Cleaned Dataset**

```
R > clean_omit <- na.omit(retail_df)
> print(clean_omit)
[1] "---- 2. Data after na.omit() ----"
> print(clean_omit[3,]) # check row 3 for replacements
Name Age Gender Admission.Test.Score High.School.Percentage City Admission.Status
3 Bushra 17 Male 89 75.68473 Islamabad Accepted
> print(head(clean_replace))
Name Age Gender Admission.Test.Score High.School.Percentage City Admission.Status
1 Shehroz 24 Female 50 68.90000 Quetta Rejected
2 Waqar 21 Female 99 60.73000 Karachi Pending
3 Bushra 17 Male 89 75.68473 Islamabad Accepted
4 Aliya 17 Male 55 85.29000 Karachi Rejected
5 Bilal 20 Male 65 61.13000 Lahore Pending
6 Murtaza 23 Female 0 75.68473 Islamabad Accepted
```

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

## SUBJECT:Data Analysis with R

The screenshot shows the RStudio interface with the following components:

- Data View:** A grid showing student admission records. The columns are Name, Age, Gender, Admission.Test.Score, High.School.Percentage, City, and Admission.Status.
- Environment View:** Shows objects in the global environment: clean\_omit (100 obs. of 7 variables), clean\_replace (157 obs. of 7 variables), retail1\_df (157 obs. of 7 variables), and avg\_percen... (75.6847260273973).
- Console View:** Displays R code and its output. The code includes reading a CSV file, printing the head of the cleaned dataset, and verifying remaining NA values. The output shows the cleaned dataset with all NA values replaced by 0.
- File Explorer:** Shows files in the current directory, including various project files like S103\_Cn\_1st.pdf, T050\_Ajeet.R, and vgsales.csv, along with R scripts and CSV data files.
- Bottom Bar:** Includes system icons for weather (28°C, Sunny), search, taskbar, and system status (ENG IN, 11:51, 01-12-2025).