

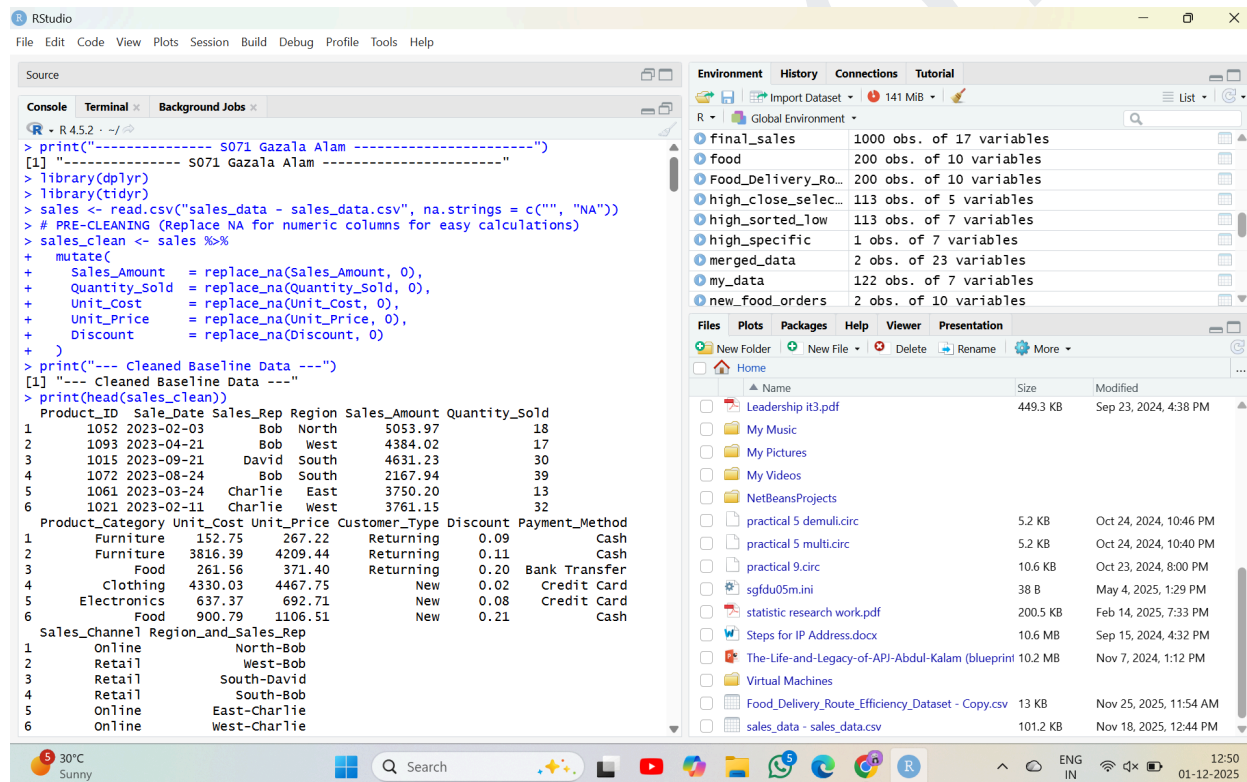
# SHETH L.U.J AND SIR M.V COLLEGE

Subject: Data Analysis with SAS / SPSS /R

## Practical no. 10

**Aim:** Creating new variables using transformations and calculations in R. import dataset.

**Outputs**→



The screenshot shows the RStudio interface. The Console pane on the left displays the following R code and its output:

```
> print("----- S071 Gazala Alam -----")
[1] "----- S071 Gazala Alam -----"
> library(dplyr)
> library(tidy)
> sales <- read.csv("sales_data - sales_data.csv", na.strings = c("", "NA"))
> # PRE-CLEANING (Replace NA for numeric columns for easy calculations)
> sales_clean <- sales %>%
+   mutate(
+     Sales_Amount = replace_na(Sales_Amount, 0),
+     Quantity_Sold = replace_na(Quantity_Sold, 0),
+     Unit_Cost = replace_na(Unit_Cost, 0),
+     Unit_Price = replace_na(Unit_Price, 0),
+     Discount = replace_na(Discount, 0)
+   )
> print("--- Cleaned Baseline Data ---")
[1] "--- Cleaned Baseline Data ---"
> print(head(sales_clean))
```

The output shows the first 6 rows of the cleaned data frame:

Product_ID	Sale_Date	Sales_Rep	Region	Sales_Amount	Quantity_Sold	
1	1052	2023-02-03	Bob	North	5053.97	18
2	1093	2023-04-21	Bob	West	4384.02	17
3	1015	2023-09-21	David	South	4631.23	30
4	1072	2023-08-24	Bob	South	2167.94	39
5	1061	2023-03-24	Charlie	East	3750.20	13
6	1021	2023-02-11	Charlie	West	3761.15	32

The Environment pane on the right shows the following data frames:

Object	Size	Variables
final_sales	1000 obs.	17 variables
food	200 obs.	10 variables
Food_Delivery_Ro...	200 obs.	10 variables
high_close_selec...	113 obs.	5 variables
high_sorted_low	113 obs.	7 variables
high_specific	1 obs.	7 variables
merged_data	2 obs.	23 variables
my_data	122 obs.	7 variables
new_food_orders	2 obs.	10 variables

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

Subject: Data Analysis with SAS / SPSS /R

RStudio interface showing R code execution and environment details.

**Source**

```
R - R 4.5.2 - / -  
> sales_calc <- sales_clean %>%  
+ mutate(  
+   Gross_Amount = Unit_Price * Quantity_Sold,  
+   Discount_Amount = Gross_Amount * (Discount / 100),  
+   Net_Revenue = Gross_Amount - Discount_Amount  
+ )  
> print("--- Method A: Arithmetic Results (Net Revenue) ---")  
[1] "--- Method A: Arithmetic Results (Net Revenue) ---"  
> print(sales_calc %>% select(Unit_Price, Quantity_Sold, Discount, Net_Revenue))  
Unit_Price Quantity_Sold Discount Net_Revenue  
1 267.22 18 0.09 4805.6310  
2 4209.44 17 0.11 71481.7635  
3 371.40 30 0.20 11119.7160  
4 4467.75 39 0.02 174207.4016  
5 692.71 13 0.08 8998.0258  
6 1106.51 32 0.21 35333.9625  
7 2624.09 29 0.14 75992.0719  
8 3964.65 46 0.12 182155.0513  
9 1095.45 30 0.05 32847.0682  
10 2682.34 18 0.13 48219.3532  
11 2517.60 13 0.23 32653.5238  
12 1137.44 43 0.08 48870.7921  
13 641.09 21 0.00 13462.8900  
14 4270.65 30 0.20 127863.2610  
15 2869.60 21 0.29 60086.8414  
16 487.65 48 0.18 23365.0670  
17 4420.15 17 0.04 75112.4930  
18 5034.35 40 0.10 201172.6260  
19 3209.22 19 0.26 60816.6445  
20 2790.10 15 0.07 41822.2040  
21 4439.12 9 0.18 39880.1663  
22 5074.42 10 0.12 50683.3070  
23 4276.99 44 0.06 188074.6475  
24 2441.79 31 0.24 75513.8208  
25 3360.40 23 0.22 77119.1638  
26 3774.65 30 0.20 113013.0210  
27 1084.28 33 0.07 35756.1931  
28 2558.95 9 0.03 23023.6408
```

**Environment**

Object	Variables
final_sales	1000 obs. of 17 variables
food	200 obs. of 10 variables
Food_Delivery_Ro...	200 obs. of 10 variables
high_close_selec...	113 obs. of 5 variables
high_sorted_low	113 obs. of 7 variables
high_specific	1 obs. of 7 variables
merged_data	2 obs. of 23 variables
my_data	122 obs. of 7 variables
new_food_orders	2 obs. of 10 variables

**Files**

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 interface showing R code execution and environment details.

**Source**

```
R - R 4.5.2 - / -  
77 2929.32 40 0.08 117079.0618  
78 3844.51 19 0.21 72892.2941  
79 3931.25 39 0.21 152996.7806  
80 1043.45 28 0.15 29172.7751  
81 2839.73 17 0.23 48164.3766  
82 2742.97 6 0.25 16416.6754  
83 2180.83 19 0.17 41365.3292  
84 2836.94 38 0.19 107598.8929  
85 2168.91 28 0.08 60680.8964  
86 1189.35 2 0.27 2372.2775  
87 2527.00 38 0.01 96016.3974  
88 860.95 37 0.10 31823.2949  
89 2910.51 4 0.22 11616.4275  
90 4576.72 22 0.02 100667.7024  
91 4002.63 26 0.05 104016.3458  
92 2147.14 49 0.18 105020.4823  
93 531.02 43 0.17 22795.0424  
94 1167.33 36 0.22 41931.4275  
95 2882.85 13 0.24 37387.1051  
96 1492.48 32 0.04 47740.2563  
97 2979.64 17 0.01 50648.8146  
98 1444.97 41 0.21 59119.3581  
99 4720.27 36 0.09 169776.7833  
100 2338.66 46 0.19 107373.9611  
101 1715.83 19 0.05 32584.4696  
102 4304.70 35 0.14 150453.5697  
103 1830.27 46 0.15 84066.1314  
104 1067.53 40 0.18 42624.3378  
105 3108.30 28 0.15 86901.8514  
106 3027.74 1 0.15 3023.1984  
107 3745.91 42 0.10 157170.8918  
108 1273.98 18 0.02 22927.0537  
109 2486.14 48 0.03 119298.9196  
110 1875.75 40 0.13 74932.4610  
111 918.17 33 0.11 30266.2804  
112 4856.22 12 0.15 58187.2280  
113 3729.27 8 0.09 29807.3093  
114 1739.45 24 0.21 41659.1317
```

**Environment**

Object	Variables
final_sales	1000 obs. of 17 variables
food	200 obs. of 10 variables
Food_Delivery_Ro...	200 obs. of 10 variables
high_close_selec...	113 obs. of 5 variables
high_sorted_low	113 obs. of 7 variables
high_specific	1 obs. of 7 variables
merged_data	2 obs. of 23 variables
my_data	122 obs. of 7 variables
new_food_orders	2 obs. of 10 variables

**Files**

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

The screenshot shows the RStudio interface. The console displays the output of a data cleaning process, including a list of 750 rows of data with columns for Sales\_Amount, Sale\_Type, Discount, and Discount\_Level. The environment pane on the right lists several datasets: final\_sales (1000 obs. of 17 variables), food (200 obs. of 10 variables), Food\_Delivery\_Ro... (200 obs. of 10 variables), high\_close\_selec... (113 obs. of 5 variables), high\_sorted\_low (113 obs. of 7 variables), high\_specific (1 obs. of 7 variables), merged\_data (2 obs. of 23 variables), my\_data (122 obs. of 7 variables), and new\_food\_orders (2 obs. of 10 variables).

The screenshot shows the RStudio interface. The console displays the output of a data cleaning process, including a list of 750 rows of data with columns for Sales\_Amount, Sale\_Type, Discount, and Discount\_Level. The environment pane on the right lists several datasets: final\_sales (1000 obs. of 17 variables), food (200 obs. of 10 variables), Food\_Delivery\_Ro... (200 obs. of 10 variables), high\_close\_selec... (113 obs. of 5 variables), high\_sorted\_low (113 obs. of 7 variables), high\_specific (1 obs. of 7 variables), merged\_data (2 obs. of 23 variables), my\_data (122 obs. of 7 variables), and new\_food\_orders (2 obs. of 10 variables).

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

Subject: Data Analysis with SAS / SPSS /R

The screenshot shows the RStudio environment with the following components:

- Source:** A data table with 106 rows and 7 columns. The columns are: Row Number, Sales Amount, Product Type, Sales Region, Sales Representative, Discount, and Status. The data is sorted by Sales Amount in descending order.
- Environment:** A list of objects in the Global Environment, including: final\_sales (1000 obs. of 17 variables), food (200 obs. of 10 variables), Food\_Delivery\_Ro... (200 obs. of 10 variables), high\_close\_selec... (113 obs. of 5 variables), high\_sorted\_low (113 obs. of 7 variables), high\_specific (1 obs. of 7 variables), merged\_data (2 obs. of 23 variables), my\_data (122 obs. of 7 variables), and new\_food\_orders (2 obs. of 10 variables).
- Files:** A list of files in the Home directory, including: Leadership it3.pdf (449.3 KB), My Music, My Pictures, My Videos, NetBeansProjects, practical 5 demul.circ (5.2 KB), practical 5 multi.circ (5.2 KB), practical 9.circ (10.6 KB), sgfd05m.ini (38 B), statistic research work.pdf (200.5 KB), Steps for IP Address.docx (10.6 MB), The-Life-and-Legacy-of-APJ-Abdul-Kalam (blueprint) (10.2 MB), Virtual Machines, Food\_Delivery\_Route\_Efficiency\_Dataset - Copy.csv (13 KB), and sales\_data - sales\_data.csv (101.2 KB).

The screenshot shows the RStudio environment with the following components:

- Source:** A data table with 250 rows and 7 columns. The columns are: Row Number, Sales Amount, Product Type, Sales Region, Sales Representative, Discount, and Status. The data is sorted by Sales Amount in descending order.
- Environment:** A list of objects in the Global Environment, including: final\_sales (1000 obs. of 17 variables), food (200 obs. of 10 variables), Food\_Delivery\_Ro... (200 obs. of 10 variables), high\_close\_selec... (113 obs. of 5 variables), high\_sorted\_low (113 obs. of 7 variables), high\_specific (1 obs. of 7 variables), merged\_data (2 obs. of 23 variables), my\_data (122 obs. of 7 variables), and new\_food\_orders (2 obs. of 10 variables).
- Files:** A list of files in the Home directory, including: Leadership it3.pdf (449.3 KB), My Music, My Pictures, My Videos, NetBeansProjects, practical 5 demul.circ (5.2 KB), practical 5 multi.circ (5.2 KB), practical 9.circ (10.6 KB), sgfd05m.ini (38 B), statistic research work.pdf (200.5 KB), Steps for IP Address.docx (10.6 MB), The-Life-and-Legacy-of-APJ-Abdul-Kalam (blueprint) (10.2 MB), Virtual Machines, Food\_Delivery\_Route\_Efficiency\_Dataset - Copy.csv (13 KB), and sales\_data - sales\_data.csv (101.2 KB).

The Console shows the following R code being executed:

```
[ reached 'max' / getoption("max.print") -- omitted 750 rows ]
> sales_text <- sales_clean %>%
+ mutate(
+   Sales_Summary = paste(
+     "Product", Product_ID,
+     "sold by", Sales_Rep,
+     "in", Region,
+     "for ₹", Sales_Amount
+   )
+ )
```



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

## Subject: Data Analysis with SAS / SPSS /R

The screenshot shows the RStudio interface with the following components:

- Source:** Contains R code for data transformation and printing.
- Console:** Displays the output of the R code, including a head of the final dataset and a detailed report.
- Environment:** Lists the objects in the global environment, including 'final\_sales', 'food', 'Food\_Delivery\_Ro...', 'high\_close\_selec...', 'high\_sorted\_low', 'high\_specific', 'merged\_data', 'my\_data', and 'new\_food\_orders'.
- Files:** Shows the file explorer with various files and folders.

```
> print("--- Method C: Text Transformation ---")
[1] "--- Method C: Text Transformation ---"
> print(head(sales_text$Sales_Summary))
[1] "Product 1052 sold by Bob in North for ₹ 5053.97"
[2] "Product 1093 sold by Bob in West for ₹ 4384.02"
[3] "Product 1015 sold by David in South for ₹ 4631.23"
[4] "Product 1072 sold by Bob in South for ₹ 2167.94"
[5] "Product 1061 sold by Charlie in East for ₹ 3750.2"
[6] "Product 1021 sold by Charlie in West for ₹ 3761.15"
> final_sales <- sales_clean %>%
+ mutate(
+   Net_Revenue = (Unit_Price * Quantity_Sold) -
+     ((Unit_Price * Quantity_Sold) * Discount / 100),
+   High_Value = ifelse(Net_Revenue > 15000, TRUE, FALSE),
+   Report = paste0("Rep: ", Sales_Rep, " | Dis: ", Discount, "%")
+ )
> print("--- Final Combined Dataset ---")
[1] "--- Final Combined Dataset ---"
> print(head(final_sales))
  Product_ID Sale_Date Sales_Rep Region Sales_Amount Quantity_Sold
1      1052 2023-02-03      Bob  North      5053.97           18
2      1093 2023-04-21      Bob  West      4384.02           17
3      1015 2023-09-21     David  South      4631.23           30
4      1072 2023-08-24      Bob  South      2167.94           39
5      1061 2023-03-24    Charlie  East      3750.20           13
6      1021 2023-02-11    Charlie  West      3761.15           32
  Product_Category Unit_Cost Unit_Price Customer_Type Discount Payment_Method
1      Furniture     152.75     267.22      Returning  0.09      Cash
2      Furniture     3816.39    4209.44      Returning  0.11      Cash
3           Food      261.56      371.40      Returning  0.20  Bank Transfer
4      Clothing     4330.03    4467.75        New     0.02    Credit Card
5      Electronics   637.37     692.71        New     0.08    Credit Card
6           Food      900.79    1106.51        New     0.21      Cash
  Sales_Channel Region_and_Sales_Rep Net_Revenue High_Value
1      Online      North-Bob      4805.631      FALSE
2      Retail      West-Bob      71481.763      TRUE
3      Retail      South-David  11119.716      FALSE
4      Retail      South-Bob      174207.402      TRUE
5      Online      East-Charlie  8998.026      FALSE
6      Online      West-Charlie  35333.963      TRUE
  Report
1 Rep: Bob | Dis: 0.09%
2 Rep: Bob | Dis: 0.11%
3 Rep: David | Dis: 0.2%
4 Rep: Bob | Dis: 0.02%
5 Rep: Charlie | Dis: 0.08%
6 Rep: Charlie | Dis: 0.21%
```

The screenshot shows the RStudio interface with the following components:

- Source:** Contains R code for data transformation and printing.
- Console:** Displays the output of the R code, including a head of the final dataset and a detailed report.
- Environment:** Lists the objects in the global environment, including 'final\_sales', 'food', 'Food\_Delivery\_Ro...', 'high\_close\_selec...', 'high\_sorted\_low', 'high\_specific', 'merged\_data', 'my\_data', and 'new\_food\_orders'.
- Files:** Shows the file explorer with various files and folders.

```
> print("--- Method C: Text Transformation ---")
[1] "--- Method C: Text Transformation ---"
> print(head(sales_text$Sales_Summary))
[1] "Product 1052 sold by Bob in North for ₹ 5053.97"
[2] "Product 1093 sold by Bob in West for ₹ 4384.02"
[3] "Product 1015 sold by David in South for ₹ 4631.23"
[4] "Product 1072 sold by Bob in South for ₹ 2167.94"
[5] "Product 1061 sold by Charlie in East for ₹ 3750.2"
[6] "Product 1021 sold by Charlie in West for ₹ 3761.15"
> final_sales <- sales_clean %>%
+ mutate(
+   Net_Revenue = (Unit_Price * Quantity_Sold) -
+     ((Unit_Price * Quantity_Sold) * Discount / 100),
+   High_Value = ifelse(Net_Revenue > 15000, TRUE, FALSE),
+   Report = paste0("Rep: ", Sales_Rep, " | Dis: ", Discount, "%")
+ )
> print("--- Final Combined Dataset ---")
[1] "--- Final Combined Dataset ---"
> print(head(final_sales))
  Product_ID Sale_Date Sales_Rep Region Sales_Amount Quantity_Sold
1      1052 2023-02-03      Bob  North      5053.97           18
2      1093 2023-04-21      Bob  West      4384.02           17
3      1015 2023-09-21     David  South      4631.23           30
4      1072 2023-08-24      Bob  South      2167.94           39
5      1061 2023-03-24    Charlie  East      3750.20           13
6      1021 2023-02-11    Charlie  West      3761.15           32
  Product_Category Unit_Cost Unit_Price Customer_Type Discount Payment_Method
1      Furniture     152.75     267.22      Returning  0.09      Cash
2      Furniture     3816.39    4209.44      Returning  0.11      Cash
3           Food      261.56      371.40      Returning  0.20  Bank Transfer
4      Clothing     4330.03    4467.75        New     0.02    Credit Card
5      Electronics   637.37     692.71        New     0.08    Credit Card
6           Food      900.79    1106.51        New     0.21      Cash
  Sales_Channel Region_and_Sales_Rep Net_Revenue High_Value
1      Online      North-Bob      4805.631      FALSE
2      Retail      West-Bob      71481.763      TRUE
3      Retail      South-David  11119.716      FALSE
4      Retail      South-Bob      174207.402      TRUE
5      Online      East-Charlie  8998.026      FALSE
6      Online      West-Charlie  35333.963      TRUE
  Report
1 Rep: Bob | Dis: 0.09%
2 Rep: Bob | Dis: 0.11%
3 Rep: David | Dis: 0.2%
4 Rep: Bob | Dis: 0.02%
5 Rep: Charlie | Dis: 0.08%
6 Rep: Charlie | Dis: 0.21%
```

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

Subject: Data Analysis with SAS / SPSS /R

gazala S071