

# 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 →

R Studio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source

Console Terminal Background Jobs

```
R > R 4.5.2 -->
> print("----- S071 Gazala Alam -----")
[1] "----- S071 Gazala Alam -----"
> library(dplyr)
> library(tidyverse)
> 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))
  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
1        Online           North-Bob
2       Retail            West-Bob
3       Retail           South-David
4       Retail           South-Bob
5      Online        East-Charlie
6      Online        West-Charlie
```

Environment History Connections Tutorial

Import Dataset 141 MiB Global Environment

files sales\_data - sales\_data.csv

list

Source

Environment History Connections Tutorial

Import Dataset 141 MiB Global Environment

files sales\_data - sales\_data.csv

list

Files Plots Packages Help Viewer Presentation

New Folder New File Delete Rename More

Home

Name	Size	Modified
Leadership i3.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
sgfldu05m.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-API-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

30°C Sunny

Search

12:50 01-12-2025

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

Subject: Data Analysis with SAS / SPSS /R

RStudio interface showing R code in the console and a data frame in the environment pane.

```
> 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

RStudio interface showing R code in the console and a data frame in the environment pane.

```
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
```

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

Subject: Data Analysis with SAS / SPSS /R

The screenshot shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The left sidebar has tabs for Source, Console, Terminal, and Background Jobs. The main area displays a large dataset from the console:

```
219 4784.28      30   0.10 143384.8716
220 3862.26      11   0.22 42391.3933
221 1576.34      28   0.20 44049.2450
222 1844.07      44   0.29 80903.7767
223 3955.75      12   0.28 47336.0868
224 1389.35      39   0.24 54054.6068
225 557.69       26   0.22 14468.0401
226 2913.78      13   0.08 37848.8367
227 4177.06       8   0.25 33332.9388
228 569.63       28   0.13 15928.9055
229 4281.79      30   0.08 128350.9370
230 4548.09       9   0.12 40883.6906
231 2889.67      39   0.25 112415.3872
232 685.10       31   0.04 21229.6048
233 971.57       21   0.14 20374.4058
234 3696.10      47   0.15 173456.1249
235 3740.13      34   0.01 127151.7036
236 4880.80      31   0.06 151214.0171
237 3337.40      25   0.25 83226.4125
238 863.03       12   0.25 10330.4691
239 865.76       1    0.09 864.9808
240 3651.42      22   0.26 80122.3788
241 1111.41      40   0.23 44354.1503
242 1906.64      37   0.13 70453.9706
243 2529.02      47   0.22 118602.4393
244 404.69        1   0.29 403.5164
245 1709.71        1   0.14 1707.3164
246 2363.90      42   0.07 99214.3013
247 1482.88      42   0.28 62106.5733
248 828.95        21   0.18 17376.6157
249 3896.19      15   0.03 58425.3171
250 4262.21      17   0.17 72334.3921
[ reached 'max' / getopt("max.print") -- omitted 750 rows ]
> sales_logic <- sales_Clean %>
+  mutate(
+   Sale_Type = ifelse(Sales_Amount > 10000, "High Value Sale", "Normal sale"),
+   Discount_Level = ifelse(Discount > 20, "High Discount", "Low Discount")
+ )
```

The right sidebar shows the Environment, History, Connections, and Tutorial panes. The Environment pane lists various objects:

- final\_sales: 1000 obs. of 17 variables
- food: 200 obs. of 10 variables
- Food\_Delivery\_Ro...: 200 obs. of 10 variables
- high\_close\_select...: 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

The Files pane shows a directory structure under 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 demuli.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

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source

Console Terminal Background Jobs

```
R + R 4.5.2 - ~/
```

Index	Product_ID	Category	Region	Rep	Amount	Discount
69	3068.03	Normal	Sale	0.21	Low	Discount
70	6499.94	Normal	Sale	0.16	Low	Discount
71	9744.52	Normal	Sale	0.09	Low	Discount
72	8485.90	Normal	Sale	0.13	Low	Discount
73	333.59	Normal	Sale	0.09	Low	Discount
74	8995.75	Normal	Sale	0.08	Low	Discount
75	7853.66	Normal	Sale	0.05	Low	Discount
76	7825.72	Normal	Sale	0.11	Low	Discount
77	4634.16	Normal	Sale	0.08	Low	Discount
78	4040.25	Normal	Sale	0.21	Low	Discount
79	3098.87	Normal	Sale	0.21	Low	Discount
80	750.38	Normal	Sale	0.15	Low	Discount
81	2359.53	Normal	Sale	0.23	Low	Discount
82	2541.38	Normal	Sale	0.25	Low	Discount
83	4892.36	Normal	Sale	0.17	Low	Discount
84	7499.70	Normal	Sale	0.19	Low	Discount
85	4790.72	Normal	Sale	0.08	Low	Discount
86	672.66	Normal	Sale	0.27	Low	Discount
87	9582.12	Normal	Sale	0.01	Low	Discount
88	9432.97	Normal	Sale	0.10	Low	Discount
89	7873.38	Normal	Sale	0.22	Low	Discount
90	9914.15	Normal	Sale	0.02	Low	Discount
91	5490.38	Normal	Sale	0.05	Low	Discount
92	9631.41	Normal	Sale	0.18	Low	Discount
93	848.49	Normal	Sale	0.17	Low	Discount
94	3720.24	Normal	Sale	0.22	Low	Discount
95	2331.27	Normal	Sale	0.24	Low	Discount
96	2038.75	Normal	Sale	0.04	Low	Discount
97	1493.95	Normal	Sale	0.01	Low	Discount
98	6261.90	Normal	Sale	0.21	Low	Discount
99	7835.09	Normal	Sale	0.09	Low	Discount
100	5825.15	Normal	Sale	0.19	Low	Discount
101	1554.93	Normal	Sale	0.05	Low	Discount
102	8130.13	Normal	Sale	0.14	Low	Discount
103	6395.95	Normal	Sale	0.15	Low	Discount
104	3942.84	Normal	Sale	0.18	Low	Discount
105	6773.89	Normal	Sale	0.15	Low	Discount
106	2673.06	Normal	Sale	0.15	Low	Discount

Environment History Connections Tutorial

Import Dataset Global Environment

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 Plots Packages Help Viewer Presentation

New Folder New File Delete Rename More

Home Leadership it3.pdf My Music My Pictures My Videos NetBeansProjects practical 5 demulti.circ practical 5 multi.circ practical 9.circ sgfdlu05m.ini static research work.pdf Steps for IP Address.docx The-Life-and-Legacy-of-API-Abdul-Kalam (blueprint) Virtual Machines Food\_Delivery\_Route\_Efficiency\_Dataset - Copy.csv sales\_data - sales\_data.csv

Name Size Modified

Leadership it3.pdf 449.3 KB Sep 23, 2024, 4:38 PM

My Music

My Pictures

My Videos

NetBeansProjects

practical 5 demulti.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

sgfdlu05m.ini 38 B May 4, 2025, 1:29 PM

static 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-API-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

ENG IN 12:51 01-12-2025

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source

Console Terminal Background Jobs

```
R + R 4.5.2 - ~/
```

Index	Product_ID	Category	Region	Rep	Amount	Discount
223	3595.20	Normal	Sale	0.28	Low	Discount
224	8951.53	Normal	Sale	0.24	Low	Discount
225	9469.92	Normal	Sale	0.22	Low	Discount
226	8936.33	Normal	Sale	0.08	Low	Discount
227	4252.54	Normal	Sale	0.25	Low	Discount
228	7825.62	Normal	Sale	0.13	Low	Discount
229	4815.72	Normal	Sale	0.08	Low	Discount
230	5025.64	Normal	Sale	0.12	Low	Discount
231	2126.33	Normal	Sale	0.25	Low	Discount
232	5952.19	Normal	Sale	0.04	Low	Discount
233	1942.75	Normal	Sale	0.14	Low	Discount
234	3382.49	Normal	Sale	0.15	Low	Discount
235	8564.80	Normal	Sale	0.01	Low	Discount
236	2150.05	Normal	Sale	0.06	Low	Discount
237	804.47	Normal	Sale	0.25	Low	Discount
238	783.18	Normal	Sale	0.25	Low	Discount
239	9413.77	Normal	Sale	0.09	Low	Discount
240	5118.51	Normal	Sale	0.26	Low	Discount
241	4153.18	Normal	Sale	0.23	Low	Discount
242	8127.70	Normal	Sale	0.13	Low	Discount
243	8374.68	Normal	Sale	0.22	Low	Discount
244	3388.69	Normal	Sale	0.29	Low	Discount
245	6966.82	Normal	Sale	0.14	Low	Discount
246	7734.12	Normal	Sale	0.07	Low	Discount
247	6581.04	Normal	Sale	0.28	Low	Discount
248	1600.79	Normal	Sale	0.18	Low	Discount
249	8771.24	Normal	Sale	0.03	Low	Discount
250	5437.04	Normal	Sale	0.17	Low	Discount

```
[ 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
```

```
+   )
```

```
+ )
```

Environment History Connections Tutorial

Import Dataset Global Environment

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 Plots Packages Help Viewer Presentation

New Folder New File Delete Rename More

Home Leadership it3.pdf My Music My Pictures My Videos NetBeansProjects practical 5 demulti.circ practical 5 multi.circ practical 9.circ sgfdlu05m.ini static research work.pdf Steps for IP Address.docx The-Life-and-Legacy-of-API-Abdul-Kalam (blueprint) Virtual Machines Food\_Delivery\_Route\_Efficiency\_Dataset - Copy.csv sales\_data - sales\_data.csv

Name Size Modified

Leadership it3.pdf 449.3 KB Sep 23, 2024, 4:38 PM

My Music

My Pictures

My Videos

NetBeansProjects

practical 5 demulti.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

sgfdlu05m.ini 38 B May 4, 2025, 1:29 PM

static 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-API-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

ENG IN 12:51 01-12-2025

# 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 sections:

- File**: File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Source**: A tab bar with Console, Terminal, and Background Jobs.
- Console**: Displays R code and its output. The code performs text transformation, prints the head of a sales dataset, and creates a final sales dataset by combining multiple data frames. It also includes a mutate function to calculate Net\_Revenue and High\_Value based on other variables.
- Environment**: Shows the global environment with various objects and their details.
- Files**: Shows the file system structure under Home, including files like Leadership\_it3.pdf, My Pictures, My Videos, NetBeansProjects, practical 5 demul.circ, practical 5 multi.circ, practical 9.circ, sgfd05m.ini, statistic research work.pdf, Steps for IP Address.docx, The-Life-and-Legacy-of-APJ-Abdul-Kalam (blueprint), Virtual Machines, Food\_Delivery\_Route\_Efficiency\_Dataset - Copy.csv, and sales\_data - sales\_data.csv.

The screenshot shows the RStudio interface with the following details:

- Source** tab: Contains R code for creating a final combined dataset from multiple sources.
- Console** tab: Displays the output of the R code, including the final dataset structure and its contents.
- Environment** tab: Shows the global environment with various data frames and objects.
- Data Frames**:
  - `final_sales`: 1000 obs. of 17 variables.
  - `food`: 200 obs. of 10 variables.
  - `Food_Delivery_Ro...`: 200 obs. of 10 variables.
  - `high_close_select...`: 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** tab: Lists files in the current directory, including PDFs, images, and project files.
- Plots**, **Packages**, **Help**, **Viewer**, **Presentation** tabs: Standard RStudio navigation tabs.

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

Subject: Data Analysis with SAS / SPSS /R

gazala S071