

PRACTICAL NO : 10

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

```

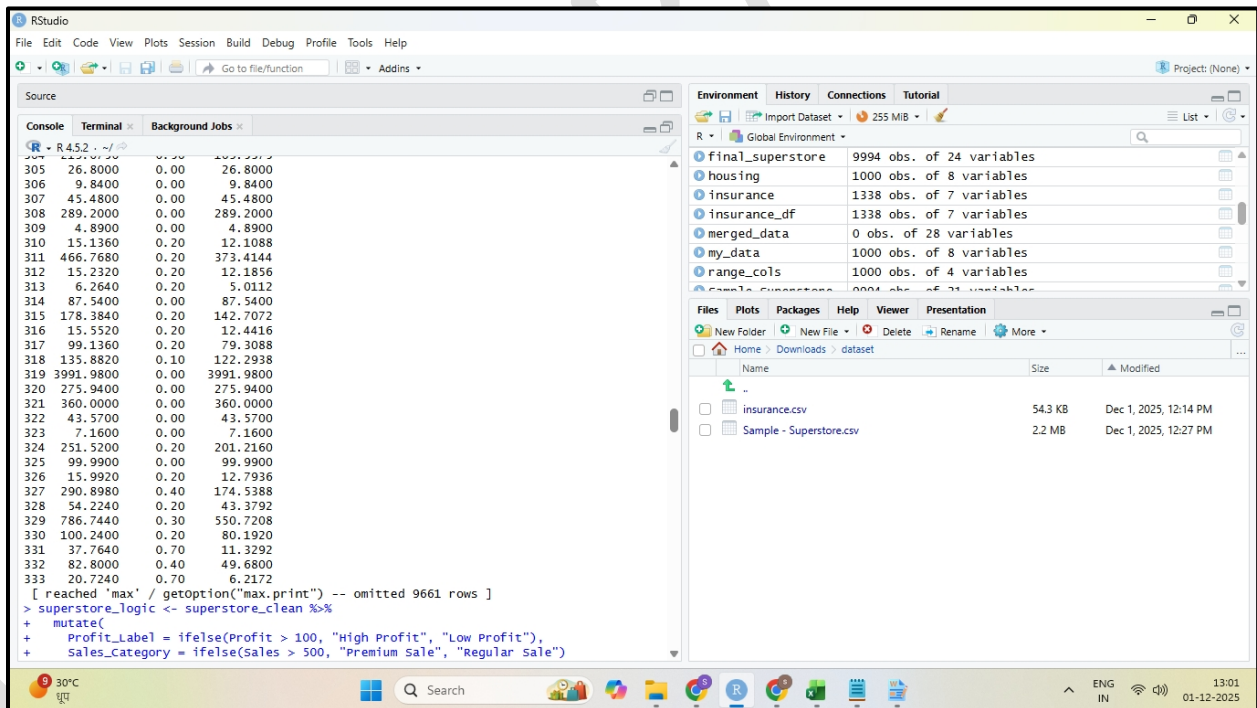
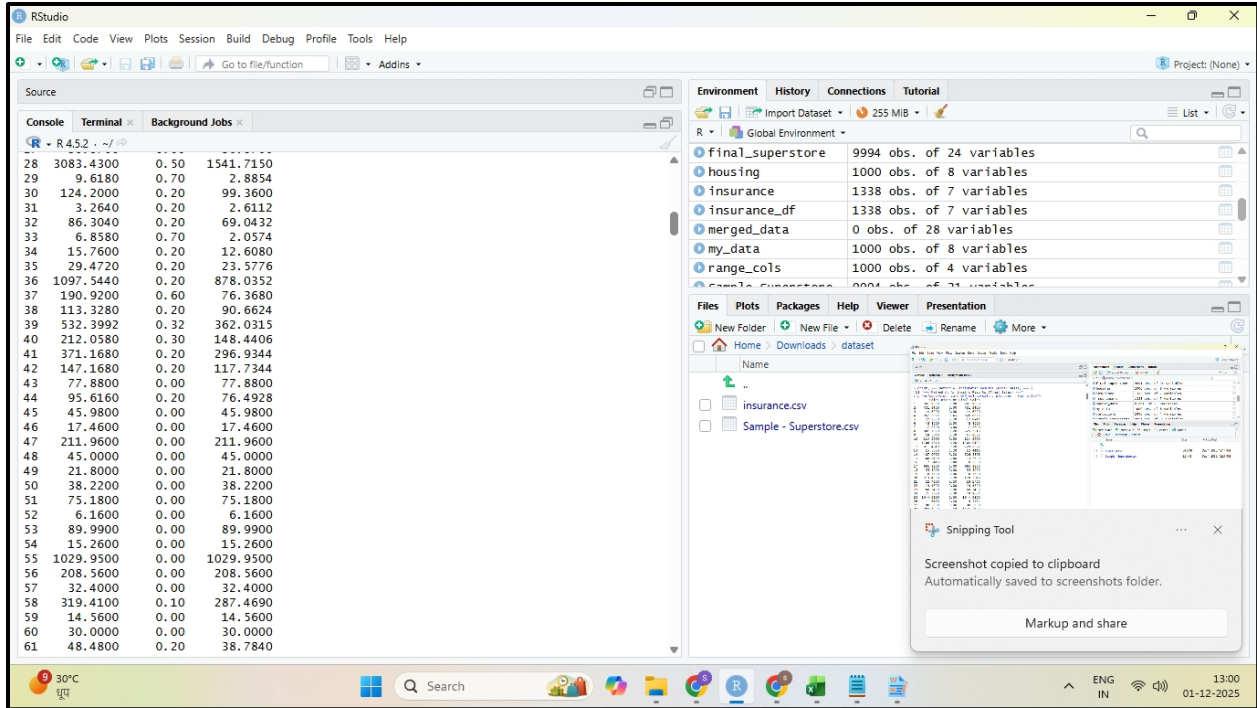
R - R4.5.2 - ~/
> library(dplyr)
> library(tidyverse) # Used to clean the data first
> # Import Superstore dataset
> superstore_df <- read.csv("Downloads/dataset/sample - Superstore.csv", na.strings = c(
+   "", "NA"))
> # PRE-CLEANING:
> # Transformations fail if numbers are missing (NA).
> # Fill missing Sales/Discount/Profit with 0 for calculations demo
> superstore_clean <- superstore_df %>%
+   mutate(
+     Sales = replace_na(Sales, 0),
+     Discount = replace_na(Discount, 0),
+     Profit = replace_na(Profit, 0)
+   )
> print("---- Cleaned Baseline data ----")
> print(head(superstore_clean[, c("Product.ID", "Category", "Sales", "Discount", "Profit")]))
  Product.ID Category Sales Discount Profit
1 FUR-BO-10001798 Furniture 261.9600 0.00 41.9136
2 FUR-CH-10000454 Furniture 731.9400 0.00 219.5820
3 OFF-LA-10000240 Office Supplies 14.6200 0.00 6.8714
4 FUR-TA-10000577 Furniture 957.5775 0.45 -383.0310
5 OFF-ST-10000760 Office Supplies 22.3680 0.20 2.5164
6 FUR-FU-10001487 Furniture 48.8600 0.00 14.1694
> superstore_calc <- superstore_clean %>%
+   mutate(
+     Discount_Amount = Sales * Discount, # Discount column is already fractional
+     Final_Sales = Sales - Discount_Amount
+   )
> print("---- Method A: Arithmetic Results (Final Sales) ----")
> print(superstore_calc %>% select(Sales, Discount, Final_Sales))
  Sales Discount Final_Sales
1 261.9600 0.00 261.9600
2 731.9400 0.00 731.9400
3 14.6200 0.00 14.6200
4 957.5775 0.45 526.6675
5 22.3680 0.20 17.8944
6 48.8600 0.00 48.8600
7 7.2800 0.00 7.2800
8 907.1520 0.20 725.7216
9 18.5040 0.20 14.8032
10 114.9000 0.00 114.9000
11 1706.1840 0.20 1364.9472
12 911.4240 0.20 729.1392
13 15.5520 0.20 12.4416
14 407.9760 0.20 326.3808
15 68.8100 0.80 13.7620
16 2.5440 0.80 0.5088
17 665.8800 0.00 665.8800
18 55.5000 0.00 55.5000
19 8.5600 0.00 8.5600
20 213.4800 0.20 170.7840
21 22.7200 0.20 18.1760
22 19.4600 0.00 19.4600
23 60.3400 0.00 60.3400
24 71.3720 0.30 49.9604
25 1044.6300 0.00 1044.6300
26 11.6480 0.20 9.3184
27 90.5700 0.00 90.5700
28 3083.4300 0.50 1541.7150
29 9.6180 0.70 2.8854
30 174.2000 0.20 99.3600

```

```

R - R4.5.2 - ~/
> print("---- Method A: Arithmetic Results (Final Sales) ----")
> print(superstore_calc %>% select(Sales, Discount, Final_Sales))
  Sales Discount Final_Sales
1 261.9600 0.00 261.9600
2 731.9400 0.00 731.9400
3 14.6200 0.00 14.6200
4 957.5775 0.45 526.6675
5 22.3680 0.20 17.8944
6 48.8600 0.00 48.8600
7 7.2800 0.00 7.2800
8 907.1520 0.20 725.7216
9 18.5040 0.20 14.8032
10 114.9000 0.00 114.9000
11 1706.1840 0.20 1364.9472
12 911.4240 0.20 729.1392
13 15.5520 0.20 12.4416
14 407.9760 0.20 326.3808
15 68.8100 0.80 13.7620
16 2.5440 0.80 0.5088
17 665.8800 0.00 665.8800
18 55.5000 0.00 55.5000
19 8.5600 0.00 8.5600
20 213.4800 0.20 170.7840
21 22.7200 0.20 18.1760
22 19.4600 0.00 19.4600
23 60.3400 0.00 60.3400
24 71.3720 0.30 49.9604
25 1044.6300 0.00 1044.6300
26 11.6480 0.20 9.3184
27 90.5700 0.00 90.5700
28 3083.4300 0.50 1541.7150
29 9.6180 0.70 2.8854
30 174.2000 0.20 99.3600

```



```

R - R 4.5.2 ~ /
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
> print("--- Method B: Logic Results (Labels) ---")
[1] "--- Method B: Logic Results (Labels) ---"
> print(superstore_logic %>% select(Sales, Profit, Profit_Label, Sales_Category))
  Sales Profit Profit_Label Sales_Category
1 261.9600 41.9136 Low Profit Regular Sale
2 731.9400 219.5820 High Profit Premium Sale
3 14.6200 6.8714 Low Profit Regular Sale
4 957.5775 -383.0310 Low Profit Premium Sale
5 22.3680 2.5164 Low Profit Regular Sale
6 48.8600 14.1694 Low Profit Regular Sale
7 7.2800 1.9656 Low Profit Regular Sale
8 907.1520 90.7152 Low Profit Premium Sale
9 18.5040 5.7825 Low Profit Regular Sale
10 114.9000 34.4700 Low Profit Regular Sale
11 1706.1840 85.3092 Low Profit Premium Sale
12 911.4240 68.3568 Low Profit Premium Sale
13 15.5520 5.4432 Low Profit Regular Sale
14 407.9760 132.5922 High Profit Regular Sale
15 68.8100 -123.8580 Low Profit Regular Sale
16 2.5440 -3.8160 Low Profit Regular Sale
17 665.8800 13.3176 Low Profit Premium Sale
18 55.5000 9.9900 Low Profit Regular Sale
19 8.5600 2.4824 Low Profit Regular Sale
20 213.4800 16.0110 Low Profit Regular Sale
21 22.7200 7.3840 Low Profit Regular Sale
22 19.4600 5.0596 Low Profit Regular Sale
23 60.3400 15.6884 Low Profit Regular Sale
24 71.3720 -1.0196 Low Profit Regular Sale
25 1044.6300 240.2649 High Profit Premium Sale
26 11.6480 4.2224 Low Profit Regular Sale
27 90.5700 11.7741 Low Profit Regular Sale
28 3083.4300 -1665.0522 Low Profit Premium Sale
29 9.6180 -7.0532 Low Profit Regular Sale

```

```

R - R 4.5.2 ~ /
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
> # 4. METHOD C: TEXT TRANSFORMATION (paste)
> # Scenario: Create 'Product_Summary' combining Category and Product Name
> superstore_text <- superstore_clean %>%
+   mutate(
+     Product_Summary = paste(Category, "product:", Product.Name, "with Sales $", Sale
+   )
+ )
> print("--- Method C: Text Transformation ---")
[1] "--- Method C: Text Transformation ---"
> print(head(superstore_text$Product_Summary))
[1] "Furniture product: Bush Somerset collection Bookcase with Sales $ 261.96"
[2] "Furniture product: Hon Deluxe Fabric upholstered Stacking Chairs, Rounded Back wit
h Sales $ 731.94"
[3] "Office Supplies product: Self-Adhesive Address Labels for Typewriters by universal
with Sales $ 14.62"
[4] "Furniture product: Bretford CR4500 Series Slim Rectangular Table with Sales $ 957.
5775"
[5] "Office Supplies product: Eldon Fold 'N Roll Cart System with Sales $ 22.368"
[6] "Furniture product: Eldon Expressions Wood and Plastic Desk Accessories, Cherry woo
d with Sales $ 48.86"
> final_superstore <- superstore_clean %>%
+   mutate(
+     Final_Sales = Sales - (Sales * Discount),
+     Is_High_Value = ifelse(Final_Sales > 500, TRUE, FALSE),
+     Status_Report = paste0("Profit: $", round(Profit, 2), " / Discount: ", Discount *
100, "%")
+   )

```

The screenshot displays the RStudio interface. The console shows the execution of R code that calculates final sales, determines if a product is high-value, and generates a status report. The environment pane on the right lists various data objects, including 'final_superstore' with 9994 observations and 24 variables. The file explorer shows the location of the dataset files.

```

R - R4.5.2 - ~/RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
R - R4.5.2 - ~/RStudio
h Sales $ 731.94"
[3] "Office Supplies product: Self-Adhesive Address Labels for Typewriters by universal
with Sales $ 14.62"
[4] "Furniture product: Bretford CR4500 Series slim Rectangular Table with Sales $ 957.
5775"
[5] "Office Supplies product: Eldon Fold 'N Roll Cart System with Sales $ 22.368"
[6] "Furniture product: Eldon Expressions Wood and Plastic Desk Accessories, cherry woo
d with Sales $ 48.86"
> final_superstore <- superstore_clean %>%
+   mutate(
+     Final_Sales = Sales - (Sales * Discount),
+     Is_High_Value = ifelse(Final_Sales > 500, TRUE, FALSE),
+     Status_Report = paste0("Profit: $", round(Profit, 2), " / Discount: ", Discount *
100, "%")
+   )
> print("--- Final Combined Dataset ---")
[1] "--- Final Combined Dataset ---"
> print(head(final_superstore[, c("Product.ID", "Category", "Final_Sales", "Is_High_Val
ue", "Status_Report")]))
  Product.ID Category Final_Sales Is_High_Value
1 FUR-B0-10001798 Furniture 261.9600 FALSE
2 FUR-CH-10000454 Furniture 731.9400 TRUE
3 OFF-LA-10000240 Office Supplies 14.6200 FALSE
4 FUR-TA-10000577 Furniture 526.6676 TRUE
5 OFF-ST-10000760 Office Supplies 17.8944 FALSE
6 FUR-FU-10001487 Furniture 48.8600 FALSE
  Status_Report
1 Profit: $41.91 / Discount: 0%
2 Profit: $219.58 / Discount: 0%
3 Profit: $6.87 / Discount: 0%
4 Profit: $-383.03 / Discount: 45%
5 Profit: $2.52 / Discount: 20%
6 Profit: $14.17 / Discount: 0%
>
Environment History Connections Tutorial
R - Global Environment
final_superstore 9994 obs. of 24 variables
housing 1000 obs. of 8 variables
insurance 1338 obs. of 7 variables
insurance_df 1338 obs. of 7 variables
merged_data 0 obs. of 28 variables
my_data 1000 obs. of 8 variables
range_cols 1000 obs. of 4 variables
Sample_Superstore 1000 obs. of 31 variables
Files Plots Packages Help Viewer Presentation
New Folder New File Delete Rename More
Home Downloads dataset
Name Size Modified
insurance.csv 54.3 KB Dec 1, 2025, 12:14 PM
Sample - Superstore.csv 2.2 MB Dec 1, 2025, 12:27 PM
30°C
Search
ENG IN
13:01 01-12-2025

```