

**SHETH L.U.J AND SIR MV COLLEGE**  
**PRACTICAL NO:- 10**

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

**CODE:-**

```
library(dplyr)
library(tidyr)
retail_df <- read.csv("C:/Users/IT-29/Downloads/retail_sales_dataset.csv",
                     na.strings = c("", "NA"))
print("--- Original Data (First 6 Rows) ---")
print(head(retail_df))

retail_clean <- retail_df %>%
  mutate(
    Price.per.Unit = replace_na(Price.per.Unit, 0),
    Quantity = replace_na(Quantity, 0),
    Total.Amount = replace_na(Total.Amount, 0)
  )

retail_calc <- retail_clean %>%
  mutate(

    Final_Price = Quantity * Price.per.Unit
  )

print("--- Method A: Calculated Final Price ---")
print(retail_calc %>% select(Quantity, Price.per.Unit, Final_Price))

retail_logic <- retail_calc %>%
  mutate(
    Price_Category = ifelse(Final_Price > 500, "Expensive", "Affordable")
  )

print("--- Method B: Conditional Logic ---")
print(retail_logic %>% select(Final_Price, Price_Category))

retail_text <- retail_logic %>%
  mutate(
    Transaction_Summary = paste(Customer.ID, "bought", Quantity, Product.Category,
                                "items for $", Final_Price)
  )
```

## SHETH L.U.J AND SIR MV COLLEGE

### PRACTICAL NO:- 10

```
print("--- Method C: Text Transformation ---")
print(head(retail_text$Transaction_Summary))
```

```
final_dataset <- retail_clean %>%
  mutate(
    Final_Price = Quantity * Price.per.Unit,
    Price_Category = ifelse(Final_Price > 500, "Expensive", "Affordable"),
    Transaction_Summary = paste(Customer.ID, "bought", Quantity, Product.Category,
                                "items for $", Final_Price)
  )

print("--- Final Combined Dataset ---")
print(head(final_dataset))
```

OUTPUT:-

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains the R code for data transformation, including loading libraries, reading the CSV file, cleaning the data, and creating the final dataset.
- Environment:** Lists the objects created in the environment, such as `retail_text`, `selected_cols`, `sepal_data`, `split_matrix`, `starts_with_avg`, and `tidy_data`.
- Files:** Shows the file explorer with various files and folders, including `.RData`, `.Rhistory`, `amazon_selenium.py`, `Desktop - ShortcutLink`, `desktop.ini`, `facebook_scrape.py.py`, `GIS DataBase`, `housing_csv`, `housing_temp.csv`, `IISExpress`, `Medical_Cost_Personal_Datasets`, `My Music`, and `My Pictures`.
- Console:** Displays the output of the code, showing the original data (first 6 rows) and the final combined dataset.

The console output shows the original data (first 6 rows) and the final combined dataset:

```
[1] "--- Original Data (First 6 Rows) ---"
Transaction.ID Date Customer.ID Gender Age Product.Category Quantity
1 2023-11-24 CUST001 Male 34 Beauty 3
2 2023-02-27 CUST002 Female 26 Clothing 2
3 2023-01-13 CUST003 Male 50 Electronics 1
4 2023-05-21 CUST004 Male 37 Clothing 1
5 2023-05-06 CUST005 Male 30 Beauty 2
6 2023-04-23 CUST006 Female 45 Beauty 1

Price.per.Unit Total.Amount
1 50 150
2 500 1000
3 30 30
4 500 500
```

NAME:- RANI CHAVRE  
ROLL NO:- S128

# SHETH L.U.J AND SIR MV COLLEGE

## PRACTICAL NO:- 10

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Source

Console Terminal Background Jobs

```
R - R 4.5.2 ~ /  
326 3 25 75  
327 3 50 150  
328 2 50 100  
329 4 25 100  
330 4 50 200  
331 3 30 90  
332 4 300 1200  
333 4 300 1200  
[ reached 'max' / getoption("max.print") -- omitted 667 rows ]  
[1] "---- Method B: Conditional Logic ----"  
Final_Price Price_Category  
1 150 Affordable  
2 1000 Expensive  
3 30 Affordable  
4 500 Affordable  
5 100 Affordable  
6 30 Affordable  
7 50 Affordable  
8 100 Affordable  
9 600 Expensive  
10 200 Affordable  
11 100 Affordable  
12 75 Affordable  
13 1500 Expensive  
14 120 Affordable  
15 2000 Expensive  
16 1500 Expensive  
17 100 Affordable  
18 50 Affordable  
19 50 Affordable  
20 900 Expensive  
21 500 Affordable  
22 100 Affordable  
23 120 Affordable
```

Environment History Connections Tutorial

R - Global Environment

- retail\_text 1000 obs. of 12 variables
- selected\_cols 5000 obs. of 3 variables
- sepal\_data 150 obs. of 3 variables
- split\_matrix chr [1:1000, 1] "Beauty" "Clothing" "Electronics" "Cl...
- starts\_with\_avg 5000 obs. of 4 variables
- tidy\_data 1000 obs. of 16 variables

values

avg\_price 179.89

Files Plots Packages Help Viewer Presentation

New Folder New File Delete Rename More

Home

Name	Size	Modified
.RData	298.5 KB	Nov 24, 2025, 1:01 PM
.Rhistory	1.3 KB	Nov 29, 2025, 3:01 PM
amazon_selenium.py	195 B	Sep 13, 2025, 11:50 AM
Desktop - ShortcutLink	720 B	Sep 12, 2025, 11:52 AM
desktop.ini	402 B	Jul 31, 2025, 3:58 PM
facebook_scrape.py.py	1 KB	Sep 13, 2025, 12:06 PM
GIS DataBase		
housing_csv	14.3 KB	Nov 24, 2025, 12:32 PM
housing_temp.csv	14.3 KB	Nov 24, 2025, 12:34 PM
IIExpress		
Medical_Cost_Personal_Datasets	54.3 KB	Sep 20, 2025, 10:05 AM
My Music		
My Pictures		

29°C Sunny 12:37 01-12-2025

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Source

Console Terminal Background Jobs

```
R - R 4.5.2 ~ /  
454 25 Affordable  
455 100 Affordable  
456 60 Affordable  
457 900 Expensive  
458 100 Affordable  
459 1200 Expensive  
460 50 Affordable  
461 1000 Expensive  
462 1200 Expensive  
463 1500 Expensive  
464 600 Expensive  
465 150 Affordable  
466 100 Affordable  
467 150 Affordable  
468 25 Affordable  
469 75 Affordable  
470 1000 Expensive  
471 150 Affordable  
472 900 Expensive  
473 50 Affordable  
474 1500 Expensive  
475 75 Affordable  
476 2000 Expensive  
477 120 Affordable  
478 60 Affordable  
479 1200 Expensive  
480 2000 Expensive  
481 1200 Expensive  
482 1200 Expensive  
483 30 Affordable  
484 1200 Expensive  
485 30 Affordable  
486 25 Affordable  
487 2000 Expensive  
488 900 Expensive
```

Environment History Connections Tutorial

R - Global Environment

- retail\_text 1000 obs. of 12 variables
- selected\_cols 5000 obs. of 3 variables
- sepal\_data 150 obs. of 3 variables
- split\_matrix chr [1:1000, 1] "Beauty" "Clothing" "Electronics" "Cl...
- starts\_with\_avg 5000 obs. of 4 variables
- tidy\_data 1000 obs. of 16 variables

values

avg\_price 179.89

Files Plots Packages Help Viewer Presentation

New Folder New File Delete Rename More

Home

Name	Size	Modified
.RData	298.5 KB	Nov 24, 2025, 1:01 PM
.Rhistory	1.3 KB	Nov 29, 2025, 3:01 PM
amazon_selenium.py	195 B	Sep 13, 2025, 11:50 AM
Desktop - ShortcutLink	720 B	Sep 12, 2025, 11:52 AM
desktop.ini	402 B	Jul 31, 2025, 3:58 PM
facebook_scrape.py.py	1 KB	Sep 13, 2025, 12:06 PM
GIS DataBase		
housing_csv	14.3 KB	Nov 24, 2025, 12:32 PM
housing_temp.csv	14.3 KB	Nov 24, 2025, 12:34 PM
IIExpress		
Medical_Cost_Personal_Datasets	54.3 KB	Sep 20, 2025, 10:05 AM
My Music		
My Pictures		

29°C Sunny 12:38 01-12-2025

NAME:- RANI CHAVRE  
ROLL NO:- S128

# SHETH L.U.J AND SIR MV COLLEGE

## PRACTICAL NO:- 10

The screenshot displays the RStudio environment with the following components:

- Console:** Shows R code execution output. It includes a loop for 500 iterations, a method for text transformation, and a final combined dataset. The output shows transaction details for 6 transactions, including Transaction.ID, Date, Customer.ID, Gender, Age, Product.Category, and Quantity.
- Environment:** Lists variables in the Global Environment:
  - retail\_text: 1000 obs. of 12 variables
  - selected\_cols: 5000 obs. of 3 variables
  - sepal\_data: 150 obs. of 3 variables
  - split\_matrix: chr [1:1000, 1] "Beauty" "Clothing" "Electronics" "Cl..."
  - starts\_with\_avg: 5000 obs. of 4 variables
  - tidy\_data: 1000 obs. of 16 variables
  - values: values
  - avg\_price: 179.89
- Files:** Shows a list of files in the Home directory, including .RData, .Rhistory, amazon\_selenium.py, Desktop - Shortcut.Lnk, desktop.ini, facebook\_scrape.py.py, GIS DataBase, housing\_csv, housing\_temp.csv, IISExpress, Medical\_Cost\_Personal\_Datasets, My Music, and My Pictures.

NAME:- RANI CHAVRE  
ROLL NO:- S128