

PRACTICAL NO 10

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

OUTPUT:-

The screenshot shows the RStudio interface with the following details:

- Top Bar:** DATA SET - RStudio, File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Go to file/function, Addins.
- Source Editor:** Displays an R script for data cleaning. The script includes imports for dplyr and tidyr, reads a CSV file, and performs various data transformations and calculations. It also includes sections for quick look and pre-cleaning numeric columns.
- Environment Tab:** Shows the global environment with objects like final_dataset, final_list, flights, flipkart, etc., and their respective sizes and types.
- File Browser:** Shows a list of files in the directory C:/Users/as993/DATA SET, including practical files, merged datasets, and output files.
- Bottom Bar:** Search, Home, Project, Files, Databases, Data, Plots, Help, ENG US, and a timestamp of 13:48 01-12-2025.

Sheth L.U.J. College of Arts And Sir M.V. College of Science and Commerce

Data Analysis with SAS / SPSS / R

R DATA SET - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Console Terminal Background Jobs

```
R - R 4.5.2 - C:/Users/as993/DATA SET/
```

```
+ Rating = as.numeric(Rating),
+ StockQty = as.integer(StockQty),
+
# replace NA with sensible defaults for calculations
+ Price = replace_na(Price, 0),
+ DiscountPercent = replace_na(DiscountPercent, 0),
+ Rating = replace_na(Rating, 0),
+ StockQty = replace_na(StockQty, 0)
+ )
>
> # -----
> # 3. METHOD A: ARITHMETIC CALCULATIONS (Final price)
> #
> # Recompute Discount_Amount and Final_Price to be sure (will overwrite if existed)
> df_calc <- df_clean %>%
+ mutate(
+   Discount_Amount = round(Price * (DiscountPercent / 100), 2),
+   Final_Price = round(Price - Discount_Amount, 2)
+ )
>
> # Show sample
> cat("---- Sample (Price, DiscountPercent, Discount_Amount, Final_Price) ----\n")
--- Sample (Price, DiscountPercent, Discount_Amount, Final_Price) ---
> print(head(df_calc %>% select(Price, DiscountPercent, Discount_Amount, Final_Price), 6))
  Price DiscountPercent Discount_Amount Final_Price
1 3591.65      36.8    1321.73    2269.92
2 1939.68      20.0     387.94    1551.74
3 883.67      13.8    121.95     761.72
4 3218.71      32.3    1039.64    2179.07
5 10032.03      4.8     481.54    9550.49
6 5163.22      4.5     232.34    4930.88

> # -----
> # 4. METHOD B: CONDITIONAL LOGIC (Labels)
> #
> # Quality_Label based on Rating, Price_Cat from Final_Price threshold
> df_logic <- df_calc %>%
+ mutate(
+   Quality_Label = ifelse(Rating > 4.0, "Top Rated", "Average"),
+   # Adjust threshold if you want; using 4000 as in your earlier code
+   # Price_Cat = ifelse(Final_Price < 2000, "Inexpensive", "Premium")
+   Price_Cat = ifelse(Final_Price < 2000, "Inexpensive", "Premium")
+ )
>
> cat("---- Sample (Rating, Quality_Label, Final_Price, Price_Cat) ----\n")
--- Sample (Rating, Quality_Label, Final_Price, Price_Cat) ---
> print(head(df_logic %>% select(Rating, Quality_Label, Final_Price, Price_Cat), 6))
  Rating Quality_Label Final_Price Price_Cat
1  4.95   Top Rated    2269.92   Budget
2  4.46   Top Rated    1551.74   Budget
3  3.64   Average     761.72   Budget
4  4.57   Top Rated    2179.07   Budget
5  4.44   Top Rated    9550.49  Premium
6  4.10   Top Rated    4930.88  Premium

> # -----
> # 5. METHOD C: TEXT TRANSFORMATION (Product_Summary)
> #
> # Using Category, StockQty and Price to create a short summary string
> df_text <- df_logic %>%
+   mutate(
+     Product_Summary = paste0(Category, " - ", Subcategory, ":", StockQty, " pcs, Rs.", Price)
+   )
>
> cat("---- Sample Product_Summary ----\n")
--- Sample Product_Summary ---
> print(head(df_text$product_summary, 6))
[1] "Toys & Baby - Toy Car: 42 pcs, Rs. 3591.65"           "Home & Kitchen - LED Bulb: 48 pcs, Rs. 1939.68"
[3] "Home & Kitchen - Air Purifier: 42 pcs, Rs. 883.67"    "Toys & Baby - Stroller: 61 pcs, Rs. 3218.71"
[5] "Electronics - Laptop: 64 pcs, Rs. 10032.03"           "Home & Kitchen - Cooker: 52 pcs, Rs. 5163.22"
>
> # -----
> # 6. ALL TOGETHER (final combined dataset)
> #
> final_dataset <- df_text %>%
+   mutate(
+     Is_High_Value = ifelse(Final_Price > 2000, TRUE, FALSE),
+     Status_Report = paste0("Rating: ", round(Rating, 1), " / Disc: ", DiscountPercent, "%")
+   )
`
```

Environment History Connections Tutorial

Import Dataset 226 MiB

Global Environment

final_dataset 10000 obs. of 33 variables

final_list 14 obs. of 3 variables

flights 120000 obs. of 14 variables

flipkart 150 obs. of 7 variables

flipkart_Data_20... 150 obs. of 3 variables

INFINITY_DATA_SET 10000 obs. of 27 variables

jan 10 obs. of 3 variables

merged 10 obs. of 5 variables

merged_data 10 obs. of 5 variables

Files Plots Packages Help Viewer Presentation

New Folder New File Delete Rename More

C:\Users\as993\DATA SET

Name Size Modified

viberapp image 2022-11-30 at 13:41:40_0294121065.jpg 95.3 kB Nov 30, 2025, 4:13 PM

WhatsApp Image 2025-11-30 at 13:41:47_401bfe63.j 146.6 kB Nov 30, 2025, 3:14 PM

PRACTICAL NO 6.R 663 B Dec 1, 2025, 11:53 AM

amazon_merged_jan_feb_big.csv 333 B Dec 1, 2025, 11:29 AM

amazon_final_list_big.csv 321 B Dec 1, 2025, 11:29 AM

merged_output.csv 333 B Dec 1, 2025, 11:46 AM

final_output.csv 321 B Dec 1, 2025, 11:46 AM

nyc_flights.csv 6.6 MB Dec 1, 2025, 12:14 PM

PRACTICAL NO 7.R 2.3 kB Dec 1, 2025, 12:21 PM

Breast.Cancer.csv 10.2 kB Dec 1, 2025, 12:52 PM

PRACTICAL NO 8.R 6.5 kB Dec 1, 2025, 12:57 PM

Breast_Cancer_cleaned.csv 10.2 kB Dec 1, 2025, 1:01 PM

Flipkart_Data_2025.csv 6.2 kB Dec 1, 2025, 1:26 PM

PRACTICAL NO 9.R 2.3 kB Dec 1, 2025, 1:47 PM

INFINITY_DATA_SET.csv 2 MB Dec 1, 2025, 1:44 PM

PRACTICAL NO 10.R 4.1 kB Dec 1, 2025, 1:47 PM

INFINITY_DATA_SET_modified.csv 3.3 MB Dec 1, 2025, 1:47 PM

ENG US 13:48 01-12-2025

R DATA SET - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Console Terminal Background Jobs

```
R - R 4.5.2 - C:/Users/as993/DATA SET/
```

```
+ Quality_Label = ifelse(Rating > 4.0, "Top Rated", "Average"),
+   # Adjust threshold if you want; using 4000 as in your earlier code
+   Price_Cat = ifelse(Final_Price > 4000, "Premium", "Budget")
+ )
>
> cat("---- Sample (Rating, Quality_Label, Final_Price, Price_Cat) ----\n")
--- Sample (Rating, Quality_Label, Final_Price, Price_Cat) ---
> print(head(df_logic %>% select(Rating, Quality_Label, Final_Price, Price_Cat), 6))
  Rating Quality_Label Final_Price Price_Cat
1  4.95   Top Rated    2269.92   Budget
2  4.46   Top Rated    1551.74   Budget
3  3.64   Average     761.72   Budget
4  4.57   Top Rated    2179.07   Budget
5  4.44   Top Rated    9550.49  Premium
6  4.10   Top Rated    4930.88  Premium

> # -----
> # 5. METHOD C: TEXT TRANSFORMATION (Product_Summary)
> #
> # Using Category, StockQty and Price to create a short summary string
> df_text <- df_logic %>%
+   mutate(
+     Product_Summary = paste0(Category, " - ", Subcategory, ":", StockQty, " pcs, Rs.", Price)
+   )
>
> cat("---- Sample Product_Summary ----\n")
--- Sample Product_Summary ---
> print(head(df_text$product_summary, 6))
[1] "Toys & Baby - Toy Car: 42 pcs, Rs. 3591.65"           "Home & Kitchen - LED Bulb: 48 pcs, Rs. 1939.68"
[3] "Home & Kitchen - Air Purifier: 42 pcs, Rs. 883.67"    "Toys & Baby - Stroller: 61 pcs, Rs. 3218.71"
[5] "Electronics - Laptop: 64 pcs, Rs. 10032.03"           "Home & Kitchen - Cooker: 52 pcs, Rs. 5163.22"
>
> # -----
> # 6. ALL TOGETHER (final combined dataset)
> #
> final_dataset <- df_text %>%
+   mutate(
+     Is_High_Value = ifelse(Final_Price > 2000, TRUE, FALSE),
+     Status_Report = paste0("Rating: ", round(Rating, 1), " / Disc: ", DiscountPercent, "%")
+   )
`
```

Environment History Connections Tutorial

Import Dataset 226 MiB

Global Environment

final_dataset 10000 obs. of 33 variables

final_list 14 obs. of 3 variables

flights 120000 obs. of 14 variables

flipkart 150 obs. of 7 variables

flipkart_Data_20... 150 obs. of 3 variables

INFINITY_DATA_SET 10000 obs. of 27 variables

jan 10 obs. of 3 variables

merged 10 obs. of 5 variables

merged_data 10 obs. of 5 variables

Files Plots Packages Help Viewer Presentation

New Folder New File Delete Rename More

C:\Users\as993\DATA SET

Name Size Modified

viberapp image 2022-11-30 at 13:41:40_0294121065.jpg 95.3 kB Nov 30, 2025, 4:13 PM

WhatsApp Image 2025-11-30 at 13:41:47_401bfe63.j 146.6 kB Nov 30, 2025, 3:14 PM

PRACTICAL NO 6.R 663 B Dec 1, 2025, 11:53 AM

amazon_merged_jan_feb_big.csv 333 B Dec 1, 2025, 11:29 AM

amazon_final_list_big.csv 321 B Dec 1, 2025, 11:29 AM

merged_output.csv 333 B Dec 1, 2025, 11:46 AM

final_output.csv 321 B Dec 1, 2025, 11:46 AM

nyc_flights.csv 6.6 MB Dec 1, 2025, 12:14 PM

PRACTICAL NO 7.R 2.3 kB Dec 1, 2025, 12:21 PM

Breast.Cancer.csv 10.2 kB Dec 1, 2025, 12:52 PM

PRACTICAL NO 8.R 6.5 kB Dec 1, 2025, 12:57 PM

Breast_Cancer_cleaned.csv 10.2 kB Dec 1, 2025, 1:01 PM

Flipkart_Data_2025.csv 6.2 kB Dec 1, 2025, 1:26 PM

PRACTICAL NO 9.R 2.3 kB Dec 1, 2025, 1:47 PM

INFINITY_DATA_SET.csv 2 MB Dec 1, 2025, 1:44 PM

PRACTICAL NO 10.R 4.1 kB Dec 1, 2025, 1:47 PM

INFINITY_DATA_SET_modified.csv 3.3 MB Dec 1, 2025, 1:47 PM

ENG US 13:48 01-12-2025

ABHISHEK DINESH SINGH
S116

Sheth L.U.J. College of Arts And Sir M.V. College of Science and Commerce
Data Analysis with SAS / SPSS /R

DATA SET - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source

```
R > # R.4.5.2 - C:\Users\as993\DATA_SET.R
# > # MUTHU C: TEXT TRANSFORMATION (PRODUCT_SUMMARY)
# >
# Using Category, StockQty and Price to create a short summary string
df_text <- df_logic %>%
  mutate(
    +   Product_Summary = paste0(Category, " - ", Subcategory, "; ", StockQty, " pcs, Rs. ", Price)
  )
>
> cat("---- Sample Product_Summary ----\n")
--- Sample Product_Summary ---
> print(head(df_text$Product_Summary, 6))
[1] "Toys & Baby - Toy Car: 42 pcs, Rs. 3591.65"      "Home & Kitchen - LED Bulb: 48 pcs, Rs. 1939.68"
[2] "Home & Kitchen - Air Purifier: 42 pcs, Rs. 883.67" "Toys & Baby - Stroller: 61 pcs, Rs. 3218.71"
[3] "Electronics - Laptop: 64 pcs, Rs. 10032.03"       "Home & Kitchen - Cooker: 52 pcs, Rs. 5163.22"
>
> # > # 6. ALL TOGETHER (final combined dataset)
> #
> final_dataset <- df_text %>%
  +   mutate(
  +     Is_High_Value = ifelse(Final_Price > 2000, TRUE, FALSE),
  +     Status_Report = paste0("Rating: ", round(RATING,1), " / Disc: ", DiscountPercent, "%")
  +   )
>
> cat("---- Final combined sample ----\n")
--- Final combined sample ---
> print(head(final_dataset %>% select(ProductID, SKU, ProductName, Price, DiscountPercent, Final_Price,
  +                                         StockQty, Rating, Quality_Label, Price_Catogory, Product_Summary,
  +                                         Is_High_Value, Status_Report), 6))
  ProductID   SKU ProductName  Price DiscountPercent Final_Price StockQty Rating
1 P000001 SKU00001 Aurelia Toy Car Max-381 3591.65 36.8 2269.92 42 4.95
2 P000002 SKU00002 Urbanwear LED Bulb X-674 1939.68 20.0 1551.74 48 4.46
3 P000003 SKU00003 Sportify Air Purifier Plus-384 883.67 13.8 761.72 42 3.64
4 P000004 SKU00004 GreenHarvest Stroller X-487 3218.71 32.3 2179.07 61 4.57
5 P000005 SKU00005 Zenova Laptop Mini-384 10032.03 4.8 9550.49 64 4.44
6 P000006 SKU00006 KiddooFy Cooker Mini-376 5163.22 4.5 4930.88 52 4.10
  Quality_Label Price_Catogory Product_Summary Is_High_Value
1 Top Rated Budget Toys & Baby - Toy Car: 42 pcs, Rs. 3591.65 TRUE
2 Top Rated Budget Home & Kitchen - LED Bulb: 48 pcs, Rs. 1939.68 FALSE
3 Average Budget Home & Kitchen - Air Purifier: 42 pcs, Rs. 883.67 FALSE
4 Top Rated Budget Toys & Baby - Stroller: 61 pcs, Rs. 3218.71 TRUE
```

DATA SET - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Source

```
R > R452 - C:/Users/as993/DATA SET/
> final_dataset <- df_text %>%
+   mutate(
+     Is_High_Value = ifelse(Final_Price > 2000, TRUE, FALSE),
+     Status_Report = paste0("Rating: ", round(Rating,1), " / Disc: ", DiscountPercent, "%")
+   )
>
> cat("--- Final combined sample ---\n")
--- Final combined sample ---
> print(head(final_dataset %>% select(ProductID, SKU, ProductName, Price, DiscountPercent, Final_Price,
+                                         StockQty, Rating, Quality_Label, Price_Category, Product_Summary,
+                                         Is_High_Value, Status_Report), 6))
ProductID SKU ProductName Price DiscountPercent Final_Price StockQty Rating
1 P000001 SKU000001 Aurelia Toy Car Max-381 3951.65 36.8 2269.92 42 4.95
2 P000002 SKU000002 Urbanwear LED Bulb X-674 1939.68 20.0 1551.74 48 4.46
3 P000003 SKU000003 Sportify Air Purifier Plus-384 883.67 13.8 761.72 42 3.64
4 P000004 SKU000004 GreenHarvest Stroller X-487 3218.71 32.3 2179.07 61 4.57
5 P000005 SKU000005 Zenova Laptop Mini-384 10032.03 4.8 9550.49 64 4.44
6 P000006 SKU000006 Kiddoo Cooker Mini-376 5163.22 4.5 4930.88 52 4.10
Quality_Label Price_Category Product_Summary Is_High_Value
1 Top Rated Budget Toys & Baby - Toy Car: 42 pcs, Rs. 3951.65 TRUE
2 Top Rated Budget Home & Kitchen LED Bulb: 48 pcs, Rs. 1839.68 FALSE
3 Average Budget Home & Kitchen Air Purifier: 42 pcs, Rs. 883.67 FALSE
4 Top Rated Budget Toys & Baby - Stroller: 61 pcs, Rs. 3218.71 TRUE
5 Top Rated Premium Electronics - Laptop: 64 pcs, Rs. 10032.03 TRUE
6 Top Rated Premium Home & Kitchen Cooker: 52 pcs, Rs. 5163.22 TRUE
Status_Report
1 Rating: 5 / Disc: 36.8%
2 Rating: 4.5 / Disc: 20%
3 Rating: 3.6 / Disc: 13.8%
4 Rating: 4.6 / Disc: 32.3%
5 Rating: 4.4 / Disc: 4.8%
6 Rating: 4.1 / Disc: 4.5%
>
> # -----
> # 7. SAVE modified dataset
> #
> write.csv(final_dataset, "INFINITY_DATA_SET_modified.csv", row.names = FALSE)
> cat("Saved modified dataset as: INFINITY_DATA_SET_modified.csv\n")
Saved modified dataset as: INFINITY_DATA_SET_modified.csv
|
```

ABHISHEK DINESH SINGH
S116