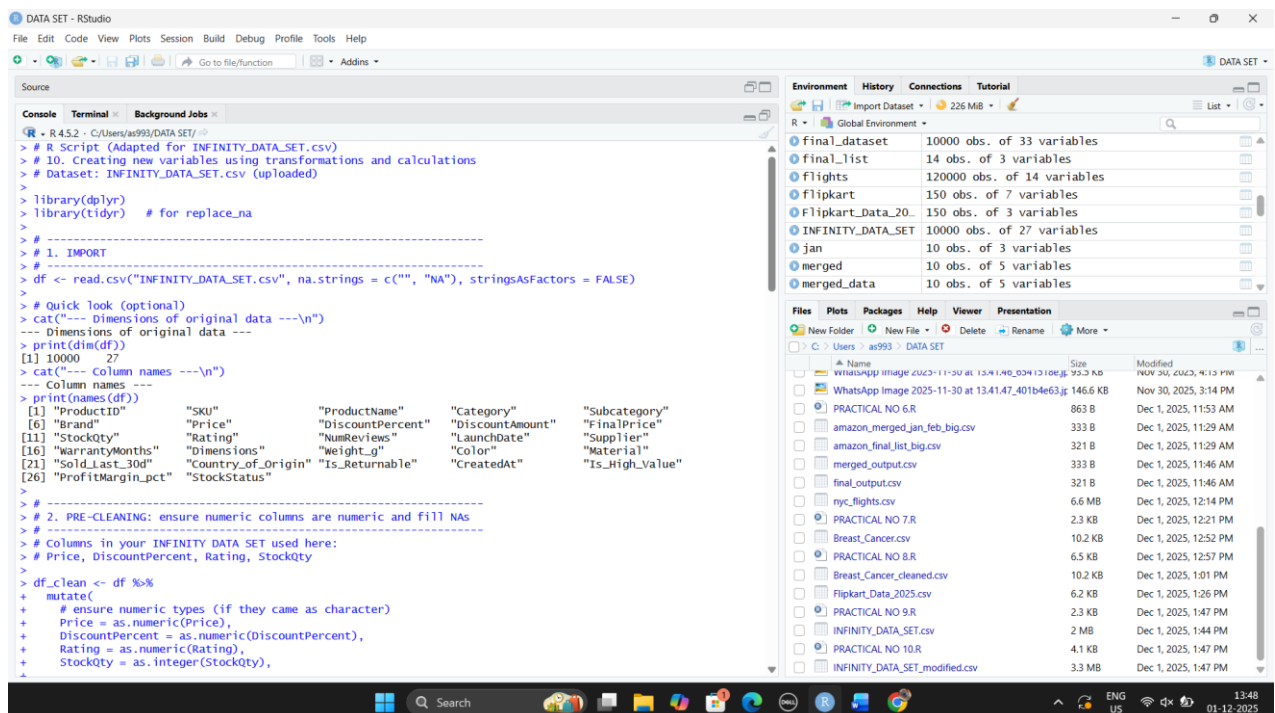


PRACTICAL NO 10

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

OUTPUT:-



```
> R Script (Adapted for INFINITY_DATA_SET.csv)
> # 10. Creating new variables using transformations and calculations
> # Dataset: INFINITY_DATA_SET.csv (uploaded)
>
> library(dplyr)
> library(tidyr) # for replace_na
>
> # -----
> # 1. IMPORT
> # -----
> df <- read.csv("INFINITY_DATA_SET.csv", na.strings = c("", "NA"), stringsAsFactors = FALSE)
>
> # Quick look (optional)
> cat("--- Dimensions of original data ---\n")
--- Dimensions of original data ---
> print(dim(df))
[1] 10000 27
> cat("--- Column names ---\n")
--- Column names ---
> print(names(df))
[1] "ProductID" "SKU" "ProductName" "Category" "Subcategory"
[6] "Brand" "Price" "DiscountPercent" "DiscountAmount" "FinalPrice"
[11] "StockQty" "Rating" "NumReviews" "LaunchDate" "Supplier"
[16] "WarrantyMonths" "Dimensions" "Weight_g" "Color" "Material"
[21] "Sold_Last_30d" "Country_of_Origin" "Is_Returnable" "CreatedAt" "Is_High_Value"
[26] "ProfitMargin_pct" "StockStatus"
>
> # -----
> # 2. PRE-CLEANING: ensure numeric columns are numeric and fill NAs
> # -----
> # Columns in your INFINITY DATA SET used here:
> # Price, DiscountPercent, Rating, StockQty
>
> df_clean <- df %>%
+ mutate(
+   # ensure numeric types (if they came as character)
+   Price = as.numeric(Price),
+   DiscountPercent = as.numeric(DiscountPercent),
+   Rating = as.numeric(Rating),
+   StockQty = as.integer(StockQty),
+ )
```

Environment

Object	Class	Attributes
final_dataset	data.frame	10000 obs. of 33 variables
final_list	data.frame	14 obs. of 3 variables
flights	data.frame	120000 obs. of 14 variables
flipkart	data.frame	150 obs. of 7 variables
Flipkart_Data_20	data.frame	150 obs. of 3 variables
INFINITY_DATA_SET	data.frame	10000 obs. of 27 variables
jan	data.frame	10 obs. of 3 variables
merged	data.frame	10 obs. of 5 variables
merged_data	data.frame	10 obs. of 5 variables

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
Console Terminal Background Jobs
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_Category 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_Category = ifelse(Final_Price > 4000, "Premium", "Budget")
+   )
+
+ > # Show sample
+ > cat("---- Sample (Rating, Quality_Label, Final_Price, Price_Category) ----\n")
+ ---- Sample (Rating, Quality_Label, Final_Price, Price_Category) ----
+ > print(head(df_logic %>% select(Rating, Quality_Label, Final_Price, Price_Category), 6))
+
+ Rating Quality_Label Final_Price Price_Category
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 Average 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)
+   )
+
+ > # Show sample
+ > cat("---- Sample Product_Summary ----\n")
+ ---- Sample Product_Summary ----
+ > print(head(df_text %>% select(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
R 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
WhatsApp image 2023-11-30 at 13:41:40_0941310e.jpg 95.3 KB Nov 30, 2023, 4:13 PM
WhatsApp image 2025-11-30 at 13:41:47_401b4e63.jpg 146.6 KB Nov 30, 2025, 3:14 PM
PRACTICAL NO 6.R 863 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
```

```
DATA SET - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
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_Category = ifelse(Final_Price > 4000, "Premium", "Budget")
+
+
+ > cat("---- Sample (Rating, Quality_Label, Final_Price, Price_Category) ----\n")
+ ---- Sample (Rating, Quality_Label, Final_Price, Price_Category) ----
+ > print(head(df_logic %>% select(Rating, Quality_Label, Final_Price, Price_Category), 6))
+
+ Rating Quality_Label Final_Price Price_Category
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 Average 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)
+   )
+
+ > # Show sample
+ > cat("---- Sample Product_Summary ----\n")
+ ---- Sample Product_Summary ----
+ > print(head(df_text %>% select(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
R 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
WhatsApp image 2023-11-30 at 13:41:40_0941310e.jpg 95.3 KB Nov 30, 2023, 4:13 PM
WhatsApp image 2025-11-30 at 13:41:47_401b4e63.jpg 146.6 KB Nov 30, 2025, 3:14 PM
PRACTICAL NO 6.R 863 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
```

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

The screenshot shows the RStudio interface with the following components:

- Console:** Displays R code for data manipulation and a sample of the resulting data frame.
- Environment:** Lists loaded objects including 'final_dataset', 'final_list', 'flights', 'flipkart', 'INFINITY_DATA_SET', 'jan', 'merged', and 'merged_data'.

```
R > R 4.5.2 - C:/Users/as993/DATA SET/
> # 5. ME I HOU 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_Category, Product_Summary,
+   Is_High_Value, Status_Report), 6))
  ProductID  SKU ProductName Price DiscountPercent Final_Price StockQty Rating
1 P000001 SKU100001 Aurelia Toy Car Max-381 3591.65 36.8 2269.92 42 4.95
2 P000002 SKU100002 Urbanwear LED Bulb X-674 1939.68 20.0 1551.74 48 4.46
3 P000003 SKU100003 Sportify Air Purifier Plus-384 883.67 13.8 761.72 42 3.64
4 P000004 SKU100004 GreenHarvest Stroller X-487 3218.71 32.3 2179.07 61 4.57
5 P000005 SKU100005 Zenova Laptop Mini-384 10032.03 4.8 9550.49 64 4.44
6 P000006 SKU100006 Kiddofun 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. 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
5 Top Rated Premium Electronics - Laptop: 64 pcs, Rs. 10032.03 TRUE
6 Top Rated Budget 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
>
```

The screenshot shows the RStudio interface with the following components:

- Console:** Displays R code for data manipulation and a sample of the resulting data frame.
- Environment:** Lists loaded objects including 'final_dataset', 'final_list', 'flights', 'flipkart', 'INFINITY_DATA_SET', 'jan', 'merged', and 'merged_data'.

```
R > R 4.5.2 - 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 SKU100001 Aurelia Toy Car Max-381 3591.65 36.8 2269.92 42 4.95
2 P000002 SKU100002 Urbanwear LED Bulb X-674 1939.68 20.0 1551.74 48 4.46
3 P000003 SKU100003 Sportify Air Purifier Plus-384 883.67 13.8 761.72 42 3.64
4 P000004 SKU100004 GreenHarvest Stroller X-487 3218.71 32.3 2179.07 61 4.57
5 P000005 SKU100005 Zenova Laptop Mini-384 10032.03 4.8 9550.49 64 4.44
6 P000006 SKU100006 Kiddofun 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. 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
5 Top Rated Premium Electronics - Laptop: 64 pcs, Rs. 10032.03 TRUE
6 Top Rated Budget 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