

MVLU COLLEGE

PRACTICAL NO. 10

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

```

1 library(dplyr)
2 library(tidyverse)
3 df <- read.csv("C:/users/itlab/onedrive/documents/SI05/Retail Product - Retail Product.csv", na.strings = c("", "N/A"))
4 df_clean <- df %>%
5   mutate(
6     Price = replace_na(Price, 0),
7     Discount = replace_na(Discount, 0),
8     Rating = replace_na(Rating, 0)
9   )
10 print("--- Cleaned Baseline Data ---")
11 print(head(df_clean))
12 df_calc <- df_clean %>%
13   mutate(
14     Discount_Amount = Price * (Discount / 100), # Step 1: Calc amount off
15     Final_Price = Price - Discount_Amount # Step 2: Subtract from total
16   )
17 print("--- Method A: Arithmetic Results (Final Price) ---")
18 print(df_calc %>% select(Price, Discount, Final_Price))
19 df_logic <- df_clean %>%
20   mutate(
21     Quality_Label = ifelse(Rating > 4.0, "Top Rated", "Average"),
22     # Let's add a second logic: Is it expensive?
23     Price_Category = ifelse(Price > 4000, "Premium", "Budget")
24   )
25
26 print("--- Method B: Logic Results (Labels) ---")
27 print(df_logic %>% select(Rating, Quality_Label, Price, Price_Category))
28
29 df_text <- df_clean %>%
30   mutate(
31     # paste0 connects strings with no separator by default
32     # paste connects strings with a space by default
33     Product_Summary = paste0(Category, " item is ", Stock, " at $ ", Price)
34   )
35 print("--- Method C: Text Transformation ---")
36 print(head(df_text$Product_Summary))
37
38 final_dataset <- df_clean %>%
39   mutate(
40     Final_Price = Price - (Price * Discount / 100),
41     Is_High_Value = ifelse(Final_Price > 2000, TRUE, FALSE),
42     Status_Report = paste0("Rating: ", round(Rating, 1), " / Dis: ", Discount, "%")
43   )
44
45 print("--- Final Combined Dataset ---")
46 print(head(final_dataset))
47

```

The screenshot shows two instances of the RStudio interface. The top instance displays the R code for data cleaning and transformation. The bottom instance shows the results of the execution, where the code has been run and the resulting dataset is displayed. The environment pane shows various objects created during the session, such as appended, clean.omit, clean.replace, etc.

MVLU COLLEGE

PRACTICAL NO. 10

RStudio

```

> library(dplyr)
> library(tidyr)
> df <- read.csv("Retail Product.csv", na.strings = c("", "NA"))

Error in file(file, "rt") : cannot open the connection
In addition: Warning message:
in cannot open file 'Retail Product.csv': no such file or directory

# PRE-CLEANING:
# Transformations fail if numbers are missing (NA).
# we will fill missing Price/Discount with 0 for this calculation demo.
> df_clean <- df %>%
+   mutate(
+     Price = replace_na(Price, 0),
+     Discount = replace_na(Discount, 0),
+     Rating = replace_na(Rating, 0)
+   )

Error in useMethod("mutate") :
no applicable method for 'mutate' applied to an object of class "function"

> df <- read.csv("c:/Users/itlab/OneDrive/Documents/S105/Retail Product - Retail Product.csv", na.strings = c("", "N"))
> df_clean <- df %>%
+   mutate(
+     Price = replace_na(Price, 0),
+     Discount = replace_na(Discount, 0),
+     Rating = replace_na(Rating, 0)
+   )
> print("--- Cleaned Baseline Data ---")
[1] "--- Cleaned Baseline Data ---"
> print(head(df_clean))
  Category Price Rating Stock Discount
1      <NA> 5548 1.870322    <NA>     0
2      <NA> 3045 4.757798    <NA>    38
3      <NA> 4004 0.000000 In Stock     0
4      <NA> 4808 1.492085    <NA>    33
5      <NA> 1817 0.000000 Out of Stock  23
6      <NA> 3522 0.000000    <NA>     0

```

Environment History Connections Tutorial Project: (None)

Data

- appended 298 obs. of 17 variables
- clean.omit 540 obs. of 5 variables
- clean.replace 4362 obs. of 5 variables
- Cleaned_BMW_Sales_ 99 obs. of 11 variables
- cleaned_car_Price_ 199 obs. of 10 variables
- df 4362 obs. of 5 variables
- df_calc 4362 obs. of 7 variables
- df_clean 4362 obs. of 5 variables
- df_logic 4362 obs. of 7 variables
- df_text 4362 obs. of 6 variables
- df1 99 obs. of 11 variables
- df2 199 obs. of 10 variables
- dropped_multiple 4573 obs. of 13 variables
- dropped_omni 4573 obs. of 14 variables
- dropped_range 4573 obs. of 11 variables
- final_dataset 4362 obs. of 8 variables
- housing 4573 obs. of 15 variables
- merged_full 294 obs. of 17 variables
- merged_inner 4 obs. of 17 variables
- merged_left 99 obs. of 17 variables
- range_cols 4573 obs. of 6 variables
- retail_data 5 obs. of 7 variables
- retail_df 4362 obs. of 5 variables
- Retail_Product_Re_ 4362 obs. of 5 variables
- selected_cols 4573 obs. of 3 variables
- split_list List of 5
- split_matrix chr [1:5, 1:2] "Electronics" "Home" "Clothing" ...
- spotify 4573 obs. of 15 variables
- spotify_data_clean 8573 obs. of 15 variables
- starts_with_track 4573 obs. of 5 variables
- tidy_data 5 obs. of 0 variables

Files Plots Packages Help Viewer Presentation

30°C Sunny Search ENG IN 13:03 01-12-2025

RStudio

```

> df <- read.csv("Retail Product.csv", na.strings = c("", "NA"))

1      <NA> 5548 1.870322    <NA>     0
2      <NA> 3045 4.757798    <NA>    38
3      <NA> 4004 0.000000 In Stock     0
4      <NA> 4808 1.492085    <NA>    33
5      <NA> 1817 0.000000 Out of Stock  23
6      <NA> 3522 0.000000    <NA>     0
> df_calc <- df_clean %>%
+   mutate(
+     Discount_Amount = Price * (Discount / 100), # step 1: Calc amount off
+     Final_Price = Price - Discount_Amount # step 2: Subtract from total
+   )
> print("")

Error: unexpected ')' in ""

> print(df_calc %>% select(Price, discount, Final_Price))
  Price Discount Final_Price
1 5548.00        0 5548.00
2 3045.38        38 1887.90
3 4004.00        0 4004.00
4 4808.33        33 3221.36
5 1817.23        23 1535.99
6 3522.00        0 3522.00
7 667.41        41 393.53
8 7125.7        7 6626.25
9 2777.6        6 2610.38
10 463.3        3 449.11
11 1151.0        0 1151.00
12 3772.45       45 2074.60
13 7719.4        4 7545.14
14 8416.29       29 5975.36
15 8530.10       10 7677.00
16 7936.44       44 4444.16
17 9319.28       28 6709.68
18 0.40          40 0.00
19 2066.30       30 1446.20
20 1784.12       12 1569.92
21 5589.39       39 3409.29
22 0.25          25 0.00
23 4176.42       42 2422.08
24 407.44        44 227.92

```

Environment History Connections Tutorial Project: (None)

Data

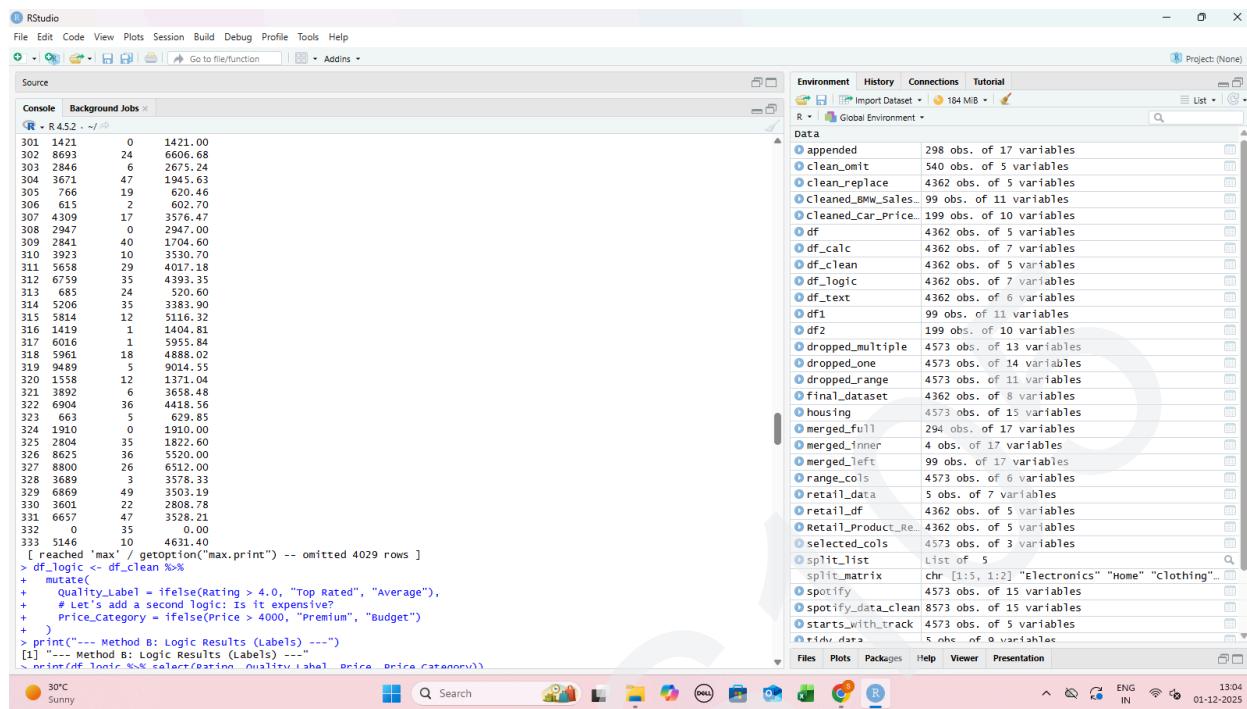
- appended 298 obs. of 17 variables
- clean.omit 540 obs. of 5 variables
- clean.replace 4362 obs. of 5 variables
- Cleaned_BMW_Sales_ 99 obs. of 11 variables
- cleaned_car_Price_ 199 obs. of 10 variables
- df 4362 obs. of 5 variables
- df_calc 4362 obs. of 7 variables
- df_clean 4362 obs. of 5 variables
- df_logic 4362 obs. of 7 variables
- df_text 4362 obs. of 6 variables
- df1 99 obs. of 11 variables
- df2 199 obs. of 10 variables
- dropped_multiple 4573 obs. of 13 variables
- dropped_omni 4573 obs. of 14 variables
- dropped_range 4573 obs. of 11 variables
- final_dataset 4362 obs. of 8 variables
- housing 4573 obs. of 15 variables
- merged_full 294 obs. of 17 variables
- merged_inner 4 obs. of 17 variables
- merged_left 99 obs. of 17 variables
- range_cols 4573 obs. of 6 variables
- retail_data 5 obs. of 7 variables
- retail_df 4362 obs. of 5 variables
- Retail_Product_Re_ 4362 obs. of 5 variables
- selected_cols 4573 obs. of 3 variables
- split_list List of 5
- split_matrix chr [1:5, 1:2] "Electronics" "Home" "Clothing" ...
- spotify 4573 obs. of 15 variables
- spotify_data_clean 8573 obs. of 15 variables
- starts_with_track 4573 obs. of 5 variables
- tidy_data 5 obs. of 0 variables

Files Plots Packages Help Viewer Presentation

30°C Sunny Search ENG IN 13:04 01-12-2025

MVLU COLLEGE

PRACTICAL NO. 10



MVLU COLLEGE

PRACTICAL NO. 10

