

MVLU COLLEGE

PRACTICAL NO. 10

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

The screenshot shows the RStudio interface with a script editor on the left and an environment pane on the right. The script editor contains the following R code:

```
1 library(dplyr)
2 library(tidy)
3 df <- read.csv("C:/Users/itlab/OneDrive/Documents/S105/Retail Product - Retail Product.csv", na.strings = c("", "N"))
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 print("--- Method B: Logic Results (Labels) ---")
26 print(df_logic %>% select(rating, quality_label, price, price_category))
27 df_text <- df_clean %>%
28   mutate(
29     # paste0 connects strings with no separator by default
30     # paste connects strings with a space by default
31     product_summary = paste(category, "item is", stock, "at $", price)
32   )
33 print("--- Method C: Text Transformation ---")
34 print(head(df_text$product_summary))
35 final_dataset <- df_clean %>%
36   mutate(
37     final_price = price - (price * discount / 100),
38     is_high_value = ifelse(final_price > 2000, TRUE, FALSE),
39     status_report = paste0("Rating: ", round(rating, 1), " / Dis: ", discount, "%")
40   )
41 print("--- Final Combined Dataset ---")
42 print(head(final_dataset))
```

The environment pane on the right shows the following objects:

Object	Size
df	4362 obs. of 5 variables
df_clean	4362 obs. of 5 variables
df_calc	4362 obs. of 7 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_one	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 9 variables

The screenshot shows the RStudio interface with a script editor on the left and an environment pane on the right. The script editor contains the following R code:

```
1 library(dplyr)
2 library(tidy)
3 df <- read.csv("C:/Users/itlab/OneDrive/Documents/S105/Retail Product - Retail Product.csv", na.strings = c("", "N"))
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 print("--- Method B: Logic Results (Labels) ---")
26 print(df_logic %>% select(rating, quality_label, price, price_category))
27 df_text <- df_clean %>%
28   mutate(
29     # paste0 connects strings with no separator by default
30     # paste connects strings with a space by default
31     product_summary = paste(category, "item is", stock, "at $", price)
32   )
33 print("--- Method C: Text Transformation ---")
34 print(head(df_text$product_summary))
35 final_dataset <- df_clean %>%
36   mutate(
37     final_price = price - (price * discount / 100),
38     is_high_value = ifelse(final_price > 2000, TRUE, FALSE),
39     status_report = paste0("Rating: ", round(rating, 1), " / Dis: ", discount, "%")
40   )
41 print("--- Final Combined Dataset ---")
42 print(head(final_dataset))
```

The environment pane on the right shows the following objects:

Object	Size
df	4362 obs. of 5 variables
df_clean	4362 obs. of 5 variables
df_calc	4362 obs. of 7 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_one	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 9 variables

MVLU COLLEGE

PRACTICAL NO. 10

The screenshot displays the RStudio interface with the following components:

- Source Panel:** Contains R code for reading a CSV file, cleaning data, and calculating final prices.
- Console Panel:** Shows the execution of the code, including error messages and the output of the `print` statements.
- Environment Panel:** Lists the objects created in the R session, such as `df`, `df_clean`, and `df_calc`.

Code Snippets:

```
R - R 4.5.2 - ~/R
> library(dplyr)
> library(tidyverse)
> df <- read.csv("Retail Product.csv", na.strings = c("", "NA"))

Error in file(file, "rt") : cannot open the connection

In addition: Warning message:
In file(file, "rt") :
  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/itslab/OneDrive/Documents/S105/Retail Product - Retail Product.csv", na.strings = c("", "NA"))
> 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

> 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
+   )

Error: unexpected ')' in ")"

> print(df_calc %>% select(Price, Discount, Final_Price))
  Price Discount Final_Price
1 5548 0 5548.00
2 3045 38 1887.90
3 4004 0 4004.00
4 4808 33 3221.36
5 1817 23 1399.09
6 3522 0 3522.00
7 667 41 393.53
8 7125 7 6626.25
9 2777 6 2610.38
10 463 3 449.11
11 1151 0 1151.00
12 3772 45 2074.60
13 7719 4 7410.24
14 8416 29 5975.36
15 8530 10 7677.00
16 7936 44 4444.16
17 9319 28 6709.68
18 0 40 0.00
19 2066 30 1446.20
20 1764 12 1569.92
21 5589 39 3409.29
22 0 25 0.00
23 4176 42 2422.08
24 407 44 227.92
```

MVLU COLLEGE

PRACTICAL NO. 10

The screenshot shows the RStudio interface with the following components:

- Source:** A script file named "R 4.5.2" containing R code for data cleaning and labeling. The code includes comments and functions for handling missing values, creating quality labels, and selecting specific columns.
- Console:** Displays the output of the R code, showing a data frame with 4029 rows and 17 variables. The output is a table with columns for row number, rating, and price.
- Environment:** Shows the global environment with various data frames and objects, including "df", "df_clean", "df_logic", "df_text", "df1", "df2", "dropped_multiple", "dropped_one", "dropped_range", "final_dataset", "housing", "merged_full", "merged_inner", "merged_left", "range_cols", "retail_data", "retail_df", "retail_product_re", "selected_cols", "split_list", "split_matrix", "spotify", "spotify_data_clean", "starts_with_track", and "tidy_data".
- Plots:** Empty.
- Packages:** Empty.
- Help:** Empty.
- Viewer:** Empty.
- Presentation:** Empty.

MVLU COLLEGE

PRACTICAL NO. 10

The image displays two screenshots of the RStudio interface, showing the execution of R code and the resulting environment data.

Top Screenshot:

- Console:** Shows the execution of R code. The code defines a function `df_logic` that filters data based on `Rating` and `Price`. It then prints the results of `df_logic` using `select` and `print`.
- Environment:** Lists the objects created in the environment, including `df`, `df_clean`, `df_text`, `df1`, `df2`, `dropped_multiple`, `dropped_one`, `dropped_range`, `final_dataset`, `housing`, `merged_full`, `merged_inner`, `merged_left`, `range_cols`, `retail_data`, `retail_df`, `retail_product_re`, `selected_cols`, `split_list`, `split_matrix`, `spotify`, `spotify_data_clean`, `starts_with_track`, and `tidy_data`.

Bottom Screenshot:

- Console:** Shows the execution of R code. The code defines a function `df_text` that filters data based on `Rating` and `Price`. It then prints the results of `df_text` using `select` and `print`.
- Environment:** Lists the objects created in the environment, including `df`, `df_clean`, `df_text`, `df1`, `df2`, `dropped_multiple`, `dropped_one`, `dropped_range`, `final_dataset`, `housing`, `merged_full`, `merged_inner`, `merged_left`, `range_cols`, `retail_data`, `retail_df`, `retail_product_re`, `selected_cols`, `split_list`, `split_matrix`, `spotify`, `spotify_data_clean`, `starts_with_track`, and `tidy_data`.