

SHETH L.U.J. AND SIR M.V. COLLEGE

SUBJECT:Data Analysis with R

Aim: Performing text manipulation using `str_sub()`, `str_split()` (R). import dataset.

The screenshot shows two RStudio sessions side-by-side. Both sessions have tabs for 'S103 7th Practical.R', 'S103 8th practical.R', and 'S103 10th practical.R'. The left session, titled 'df', displays a data frame with columns: Category, Price, Rating, Stock, Discount. The right session, titled 'df_calc', displays a similar data frame with an additional column: Discount_Amount.

Session df (Left):

```
R > # Data frame creation
> df <- data.frame(Category = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10"),
+                   Price = c(5548, 3045, 4004, 4808, 1817, 3522, 667, 7125, 2777, 443),
+                   Rating = c(1.870322, 4.757798, 0.000000, 1.492085, 0.000000, 0.000000, 3.668341, 4.983998, 2.678384, 4.49187),
+                   Stock = c("<NA>", "<NA>", "In Stock", "<NA>", "Out of Stock", "<NA>", "In Stock", "Out of Stock", "In Stock", "<NA>"),
+                   Discount = c(0, 38, 0, 33, 23, 0, 41, 7, 6, 0))
> df
#> # Data manipulation
> df_clean <- df %>%
+   mutate(
+     # pasted connects strings with no separator by default
+     # paste connects strings with a space by default
+     Product_Summary = paste0(category, ". item is", stock, " at $", price)
+   )
> print("---- Method C: Text Transformation ---")
[1] "---- Method C: Text Transformation ---"
> print(head(df_clean$Product_Summary))
[1] "1 NA item is NA at $ 5548"      "NA item is NA at $ 3045"      "NA item is In Stock at $ 4004"
[4] "NA item is NA at $ 1817"        "NA item is Out of Stock at $ 1817" "NA item is NA at $ 3522"
> final_dataset <- df_clean %>%
+   mutate(
+     Final_Price = price - (price * discount / 100),
+     Is_High_Value = ifelse(final_price > 2000, TRUE, FALSE),
+     Status_Report = paste0("Rating: ", round(rating, 1), "% Dis: ", discount, "%")
+   )
> print("---- Final Combined Dataset ---")
[1] "---- Final Combined Dataset ---"
> print(head(final_dataset))
Category Price Rating Stock Discount Final_Price Is_High_Value Status_Report
1 NA 5548 1.870322 <NA> 0 5548.00
2 NA 3045 4.757798 <NA> 38 1157.10 1887.90
3 NA 4004 0.000000 In Stock 0 4004.00
4 NA 4808 1.492085 <NA> 33 3221.36
5 NA 1817 0.000000 Out of Stock 23 417.91 1399.09
6 NA 3522 0.000000 <NA> 0 3522.00
7 C 667 3.668341 In Stock 41 273.47 393.53
8 A 7125 4.983998 Out of Stock 7 498.75 6626.25
9 A 2777 2.678384 In Stock 6 166.62 2610.38
10 NA 443 4.49187 <NA> 0 13.89 449.11
```

Session df_calc (Right):

```
R > # Data frame creation
> df_calc <- data.frame(Category = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10"),
+                         Price = c(5548, 3045, 4004, 4808, 1817, 3522, 667, 7125, 2777, 443),
+                         Rating = c(1.870322, 4.757798, 0.000000, 1.492085, 0.000000, 0.000000, 3.668341, 4.983998, 2.678384, 4.49187),
+                         Stock = c("<NA>", "<NA>", "In Stock", "<NA>", "Out of Stock", "<NA>", "In Stock", "Out of Stock", "In Stock", "<NA>"),
+                         Discount = c(0, 38, 0, 33, 23, 0, 41, 7, 6, 0),
+                         Discount_Amount = c(0.00, 1157.10, 0.00, 3221.36, 417.91, 3522.00, 273.47, 498.75, 166.62, 13.89),
+                         Final_Price = c(5548.00, 1887.90, 4004.00, 3221.36, 1399.09, 3522.00, 393.53, 6626.25, 2610.38, 449.11))
> df_calc
#> # Data manipulation
> df_clean <- df_calc %>%
+   mutate(
+     Discount_Amount = price * (discount / 100), # Step 1: Calc amount off
+     Final_Price = price - discount_amount # Step 2: Subtract from total
+   )
> print("---- Method A: Arithmetic Results (Final Price) ---")
[1] "---- Method A: Arithmetic Results (Final Price) ---"
> 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
```

SHETH L.U.J. AND SIR M.V. COLLEGE

SUBJECT:Data Analysis with R

The screenshot shows the RStudio interface with the following components:

- Data View:** Displays a table of 4362 entries across 5 columns: Category, Price, Rating, Stock, and Discount.
- Console View:** Shows R code and its output. The code involves data cleaning, logic selection, and printing results. The output includes a summary of the data and a subset of rows.
- File Explorer:** Shows the file structure of the current workspace, including files like Rhistory, Database1.accdb, and various CSV and PDF files.
- Plots, Packages, Help, Viewer, Presentation:** Standard RStudio navigation tabs.
- System Status Bar:** Shows the date (01-12-2025), time (12:18), and system status (ENG IN).

Data View Output (Subset):

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
7 C	667	3.668341	In Stock	41
8 A	7125	4.983998	Out of Stock	7
9 A	2777	2.678384	In Stock	6
10 MA	463	4.626187	NA	2

Console Output (Subset):

```

[R] - R 4.5.2 - C:/Users/Abu/Desktop/
32/ 8800 2b 6112.00
328 3689 3 3578.33
329 6869 49 3503.19
330 3601 22 2808.78
331 6657 47 3528.21
332 5146 0 35 0.00
333 5146 10 4631.40
[ reached 'max' / getoption("max.print") -- omitted 4029 rows ]
> df_logic <- df_clean %>
+ mutate(
+   quality_Label = ifelse(Rating > 4.0, "Top Rated", "Average"),
+   # Let's add a second logic: Is it expensive?
+   Price_Category = ifelse(Price > 4000, "Premium", "Budget")
+ )
> print(" --- Method B: Logic Results (Labels) --- ")
[1] " --- Method B: Logic Results (Labels) --- "
> print(df_logic %>% select(Rating, quality_Label, Price, Price_Category))
#> #> #> Rating Quality_Label Price Price_Category
#> 1 1.870322 Average 5548 Premium
#> 2 4.757798 Top Rated 3045 Budget
#> 3 0.000000 Average 4004 Premium
#> 4 1.492085 Average 4808 Premium
#> 5 0.000000 Average 1817 Budget
#> 6 0.000000 Average 3522 Budget
#> 7 C 667 3.668341 In Stock 41 Average Budget
#> 8 A 7125 4.983998 Out of Stock 7 Top Rated Premium
#> 9 A 2777 2.678384 In Stock 6 Average Budget
#> 10 MA 463 4.626187 NA 2 Top Rated Premium

```

SHETH L.U.J. AND SIR M.V. COLLEGE

SUBJECT:Data Analysis with R

Showing 1 to 13 of 4,362 entries, 6 total columns

Category	Price	Rating	Stock	Discount	Product_Summary
1 NA	5548	1.870322	NA	0	NA item is NA at \$ 5548
2 NA	3045	4.757798	NA	38	NA item is NA at \$ 3045
3 NA	4004	0.000000	In Stock	0	NA item is In Stock at \$ 4004
4 NA	4808	1.492085	NA	33	NA item is NA at \$ 4808
5 NA	1817	0.000000	Out of Stock	23	NA item is Out of Stock at \$ 1817
6 NA	3522	0.000000	NA	0	NA item is NA at \$ 3522
7 C	667	3.668341	In Stock	41	C item is In Stock at \$ 667
8 A	7125	4.983998	Out of Stock	7	A item is Out of Stock at \$ 7125
9 A	2777	2.678384	In Stock	6	A item is In Stock at \$ 2777
10 NA	463	4.626187	NA	3	NA item is NA at \$ 463
11 A	1151	2.947838	NA	0	A item is NA at \$ 1151
12 A	3772	4.890750	In Stock	45	A item is In Stock at \$ 3772

```
R > 4.5.2 . C:\Users\Itali\Desktop\ 12:18 01-12-2025
+ <NA>_is_High_Value = ifelse(is.na(Price) > 2000, TRUE, FALSE);
+ <NA>_Status_Report = paste0("Rating: ", round(Rating, 1), " / Dis: ", Discount, "%")
+ )
> print("--- Final Combined Dataset ---")
[1] "--- Final Combined Dataset ---"
> print(head(final_dataset))
Category Price Rating Stock Discount Final_Price Is_High_Value Status_Report
1 <NA> 5548 1.870322 NA 0 5548.00 TRUE Rating: 1.9 / Dis: 0%
2 <NA> 3045 4.757798 <NA> 38 1887.90 FALSE Rating: 4.8 / Dis: 38%
3 <NA> 4004 0.000000 In Stock 0 4004.00 TRUE Rating: 0 / Dis: 0%
4 <NA> 4808 1.492085 <NA> 33 3221.36 TRUE Rating: 1.5 / Dis: 33%
5 <NA> 1817 0.000000 Out of Stock 23 1399.09 FALSE Rating: 0 / Dis: 23%
6 <NA> 3522 0.000000 <NA> 0 3522.00 TRUE Rating: 0 / Dis: 0%
>
> View(df)
> View(df_calc)
> View(df_clean)
> View(df_logic)
> View(df_text)
> |
```

BAJFINANCE -1.13%

Showing 1 to 18 of 4,362 entries, 8 total columns

Category	Price	Rating	Stock	Discount	Final_Price	Is_High_Value	Status_Report
1 NA	5548	1.870322	NA	0	5548.00	TRUE	Rating: 1.9 / Dis: 0%
2 NA	3045	4.757798	NA	38	1887.90	FALSE	Rating: 4.8 / Dis: 38%
3 NA	4004	0.000000	In Stock	0	4004.00	TRUE	Rating: 0 / Dis: 0%
4 NA	4808	1.492085	NA	33	3221.36	TRUE	Rating: 1.5 / Dis: 33%
5 NA	1817	0.000000	Out of Stock	23	1399.09	FALSE	Rating: 0 / Dis: 23%
6 NA	3522	0.000000	<NA>	0	3522.00	TRUE	Rating: 0 / Dis: 0%
7 C	667	3.668341	In Stock	41	393.53	FALSE	Rating: 3.7 / Dis: 41%
8 A	7125	4.983998	Out of Stock	7	6626.25	TRUE	Rating: 5 / Dis: 7%
9 A	2777	2.678384	In Stock	6	2610.38	TRUE	Rating: 2.7 / Dis: 6%
10 NA	463	4.626187	NA	3	449.11	FALSE	Rating: 4.6 / Dis: 3%
11 A	1151	2.947838	NA	0	1151.00	FALSE	Rating: 2.9 / Dis: 0%
12 A	3772	4.890750	In Stock	45	2074.60	TRUE	Rating: 4.9 / Dis: 45%
13 NA	7719	2.982242	In Stock	4	7410.24	TRUE	Rating: 3 / Dis: 4%
14 C	8416	1.270943	NA	29	5975.36	TRUE	Rating: 1.3 / Dis: 29%
15 B	8530	0.000000	NA	10	7677.00	TRUE	Rating: 0 / Dis: 10%
16 A	7936	3.032832	In Stock	44	4444.16	TRUE	Rating: 3 / Dis: 44%
17 B	9319	3.479064	In Stock	28	6709.68	TRUE	Rating: 3.5 / Dis: 28%

```
R > 4.5.2 . C:\Users\Itali\Desktop\ 12:19 01-12-2025
+ )
> print("--- Final Combined Dataset ---")
[1] "--- Final Combined Dataset ---"
> print(head(final_dataset))
Category Price Rating Stock Discount Final_Price Is_High_Value Status_Report
1 <NA> 5548 1.870322 <NA> 0 5548.00 TRUE Rating: 1.9 / Dis: 0%
2 <NA> 3045 4.757798 <NA> 38 1887.90 FALSE Rating: 4.8 / Dis: 38%
3 <NA> 4004 0.000000 In Stock 0 4004.00 TRUE Rating: 0 / Dis: 0%
4 <NA> 4808 1.492085 <NA> 33 3221.36 TRUE Rating: 1.5 / Dis: 33%
5 <NA> 1817 0.000000 Out of Stock 23 1399.09 FALSE Rating: 0 / Dis: 23%
6 <NA> 3522 0.000000 <NA> 0 3522.00 TRUE Rating: 0 / Dis: 0%
>
```

BAJFINANCE -1.13%