

MVLU COLLEGE

PRACTICAL NO.10

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

Aim:

RStudio Environment pane showing:

- df_text: 106 obs. of 11 variables
- dropped_multip...: 106 obs. of 9 variables
- dropped_one: 106 obs. of 9 variables
- dropped_range: 106 obs. of 9 variables
- Fastest.Fifties: 107 obs. of 9 variables
- final_dataset: 106 obs. of 12 variables
- merged_data: 106 obs. of 10 variables
- names_col: 106 obs. of 0 variables

RStudio File browser pane showing project files:

- tuel_sales.csv
- GIS Database
- ISExpress
- My Music
- My Pictures
- My Videos
- My Web Sites
- R
- SM PRACTICAL - II 20.txt
- SM PRACTICAL - II.txt
- total_bills.csv
- Virtual Machines
- Visual Studio 2022
- WindowsPowerShell

```

1 library(dplyr)
2 library(tidyverse)
3 
4 # =====
5 # 1. SETUP: Import the Dataset (Modified for "Fastest Fifties - 2018.csv")
6 # =====
7 
8 # FIX 1: Updated file path to match the 'Fastest Fifties - 2018.csv' file
9 df <- read.csv("D:/Haris126/Fastest Fifties - 2018.csv", na.strings = c("", "NA"))
10 
11 # Check the actual column names after import
12 print(" --- Actual Column Names ---")
13 print(names(df))
14 
15 # PRE-CLEANING:
16 # We'll fill missing Runs, BF, and boundaries with 0.
17 df_clean <- df %>%
18   mutate(
19     Runs = replace_na(Runs, 0),
20     BF = replace_na(BF, 0),
21 
22     # FIX 2: Assuming R renamed '4s' and '6s' to 'x4s' and 'x6s'
23     x4s = replace_na(x4s, 0),
24     x6s = replace_na(x6s, 0),
25   )
26 
27 print(" --- Cleaned Baseline Data ---")
28 print(head(df_clean))
29 
30 # =====
31 # 2. METHOD A: ARITHMETIC CALCULATIONS
32 # =====
33 # Scenario: calculate the 'Strike Rate'.
34 # Formula: (Runs / Ball_Faced) * 100
35 # IMPORTANT: use ifelse to prevent division by zero (BF = 0)
36 df_calc <- df_clean %>%
37   mutate(
38     # Handle division by zero: if BF is 0, SR is 0; otherwise, calculate SR.
39     Strike_Rate = ifelse(
40       BF == 0,
41       0,
42       round((Runs / BF) * 100, 2)
43     )
44   )
45 
46 print(" --- Method A: Arithmetic Results (Strike Rate) ---")
47 print(head(df_calc %>% select(Player, Runs, BF, Strike_Rate)))
48 
49 # =====
50 # 3. METHOD B: CONDITIONAL LOGIC (ifelse)
51 # =====
52 # Scenario: Label the inning based on speed and boundaries.
53 # Logic 1: If Strike Rate is extremely high (> 250), it's "Savage".
54 # Logic 2: If they hit 7 or more 6s, they are a "Maximum Hitter".
55 
56 df_logic <- df_calc %>%
57   mutate(
58     Inning_Type = ifelse(Strike_Rate > 250, "Savage",
59                           ifelse(strike_rate > 200, "Explosive", "Standard")),
60 
61     # FIX 3: updated to use x6s here
62     Player_Badge = ifelse(X6s >= 7, "Maximum Hitter", "Solid Striker")
63   )
64 
65 print(" --- Method B: Logic Results (Labels) ---")
66 print(head(df_logic %>% select(Player, Strike_Rate, Inning_Type, X6s, Player_Badge)))
  
```

RStudio Environment pane showing:

- df_text: 106 obs. of 11 variables
- dropped_multip...: 106 obs. of 9 variables
- dropped_one: 106 obs. of 9 variables
- dropped_range: 106 obs. of 9 variables
- Fastest.Fifties: 107 obs. of 9 variables
- final_dataset: 106 obs. of 12 variables
- merged_data: 106 obs. of 10 variables
- names_col: 106 obs. of 0 variables

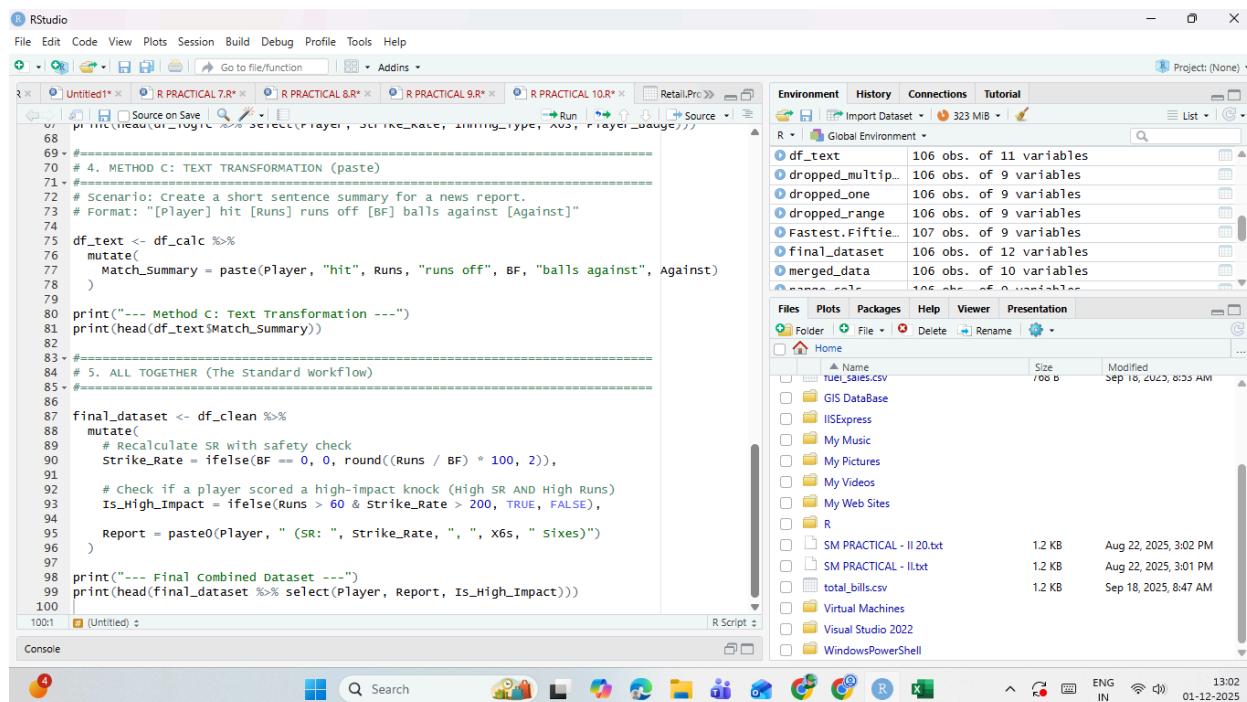
RStudio File browser pane showing project files:

- tuel_sales.csv
- GIS Database
- ISExpress
- My Music
- My Pictures
- My Videos
- My Web Sites
- R
- SM PRACTICAL - II 20.txt
- SM PRACTICAL - II.txt
- total_bills.csv
- Virtual Machines
- Visual Studio 2022
- WindowsPowerShell

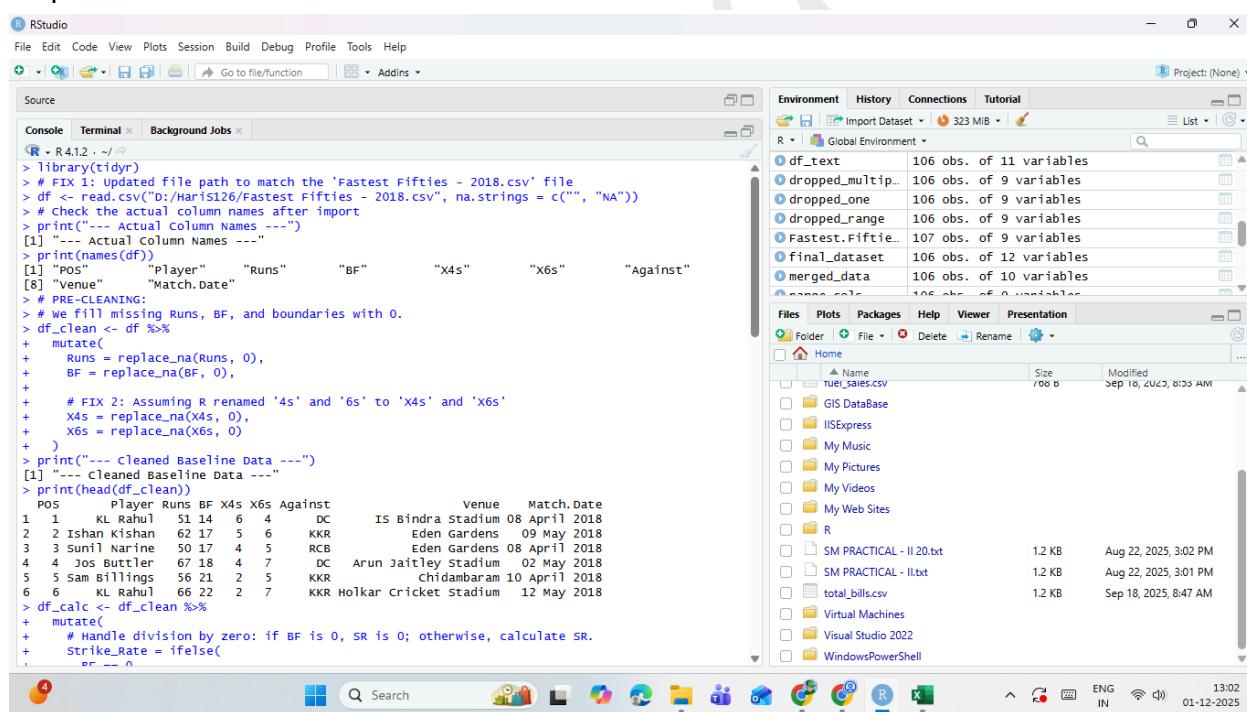
```

35 # IMPORTANT: use ifelse to prevent division by zero (BF = 0)
36 df_calc <- df_clean %>%
37   mutate(
38     # Handle division by zero: if BF is 0, SR is 0; otherwise, calculate SR.
39     Strike_Rate = ifelse(
40       BF == 0,
41       0,
42       round((Runs / BF) * 100, 2)
43     )
44   )
45 
46 print(" --- Method A: Arithmetic Results (Strike Rate) ---")
47 print(head(df_calc %>% select(Player, Runs, BF, Strike_Rate)))
48 
49 # =====
50 # 3. METHOD B: CONDITIONAL LOGIC (ifelse)
51 # =====
52 # Scenario: Label the inning based on speed and boundaries.
53 # Logic 1: If Strike Rate is extremely high (> 250), it's "Savage".
54 # Logic 2: If they hit 7 or more 6s, they are a "Maximum Hitter".
55 
56 df_logic <- df_calc %>%
57   mutate(
58     Inning_Type = ifelse(Strike_Rate > 250, "Savage",
59                           ifelse(strike_rate > 200, "Explosive", "Standard")),
60 
61     # FIX 3: updated to use x6s here
62     Player_Badge = ifelse(X6s >= 7, "Maximum Hitter", "Solid Striker")
63   )
64 
65 print(" --- Method B: Logic Results (Labels) ---")
66 print(head(df_logic %>% select(Player, Strike_Rate, Inning_Type, X6s, Player_Badge)))
  
```

MVLU COLLEGE
PRACTICAL NO.10



Output:



Hariprasad Vishwakarma
S126
SYCS

MVLU COLLEGE

PRACTICAL NO.10

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source

```
R - R 4.1.2 - ~/ ~
+ mutate(
+   # Handle division by zero: if BF is 0, SR is 0; otherwise, calculate SR.
+   Strike_Rate = ifelse(
+     BF == 0,
+     0,
+     round(runs / BF) * 100, 2)
+   )
> print("--- Method A: Arithmetic Results (Strike Rate) ---")
[1] "--- Method A: Arithmetic Results (Strike Rate) ---"
> print(head(df_calc %>% select(Player, Runs, BF, Strike_Rate)))
  Player Runs BF Strike_Rate
1  KL Rahul  31 14    364.29
2 Ishan Kishan  52 17    364.71
3 Sunil Narine  50 17    294.12
4  Jos Buttler  67 18    372.22
5 Sam Billings  56 21    266.67
6  KL Rahul  66 22    300.00
> df_logic <- df_calc %>%
+   mutate(
+     Inning_Type = ifelse(Strike_Rate > 250, "Savage",
+                           ifelse(Strike_Rate > 200, "Explosive", "standard")),
+     )
+   # FIX 3: Updated to use X6s here.
+   Player_Badge = ifelse(X6s >= 7, "Maximum Hitter", "solid striker")
+   )
> print("--- Method B: Logic Results (Labels) ---")
[1] "--- Method B: Logic Results (Labels) ---"
> print(head(df_logic %>% select(Player, strike_rate, Inning_Type, X6s, Player_Badge)))
  Player strike_rate Inning_Type X6s  Player_Badge
1  KL Rahul    364.29    Savage  4 Solid Striker
2 Ishan Kishan    364.71    Savage  6 Solid Striker
3 Sunil Narine    294.12    Savage  5 Solid Striker
4  Jos Buttler    372.22    Savage  7 Maximum Hitter
5 Sam Billings    266.67    Savage  5 solid Striker
```

Environment History Connections Tutorial

Project: (None)

df_text 106 obs. of 11 variables
dropped_multip... 106 obs. of 9 variables
dropped_one 106 obs. of 9 variables
dropped_range 106 obs. of 9 variables
Fastest.Fiftie... 107 obs. of 9 variables
final_dataset 106 obs. of 12 variables
merged_data 106 obs. of 10 variables
names_col 106 obs. of 0 variables

Files Plots Packages Help Viewer Presentation

Home

Name	Size	Modified
tue_sales.csv	708 b	Sep 18, 2025, 8:55 AM
GIS DataBase		
IISExpress		
My Music		
My Pictures		
My Videos		
My Web Sites		
R		
SM PRACTICAL - II 20.txt	1.2 KB	Aug 22, 2025, 3:02 PM
SM PRACTICAL - II.txt	1.2 KB	Aug 22, 2025, 3:01 PM
total_bills.csv	1.2 KB	Sep 18, 2025, 8:47 AM
Virtual Machines		
Visual Studio 2022		
WindowsPowerShell		

ENG IN 13:02 01-12-2025

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source

```
R - R 4.1.2 - ~/ ~
> df_text <- df_calc %>%
+   mutate(
+     Match_Summary = paste(Player, "hit", Runs, "runs off", BF, "balls against", Against)
+   )
> print("--- Method C: Text Transformation ---")
[1] "--- Method C: Text Transformation ---"
> print(head(df_text$Match_Summary))
[1] "KL Rahul hit 51 runs off 14 balls against DC"
[2] "Ishan Kishan hit 62 runs off 17 balls against KKR"
[3] "Sunil Narine hit 50 runs off 17 balls against RCB"
[4] "Jos Buttler hit 67 runs off 18 balls against DC"
[5] "Sam Billings hit 56 runs off 21 balls against KKR"
[6] "KL Rahul hit 66 runs off 22 balls against KKR"
> final_dataset <- df_clean %>%
+   mutate(
+     # Recalculate SR with safety check
+     Strike_Rate = ifelse(BF == 0, 0, round((Runs / BF) * 100, 2)),
+
+     # Check if a player scored a high-impact knock (High SR AND High Runs)
+     Is_High_Impact = ifelse(Runs > 60 & Strike_Rate > 200, TRUE, FALSE),
+
+     Report = paste0(Player, " (SR: ", strike_rate, ", ", X6s, " sixes)")
+   )
> print("--- Final Combined Dataset ---")
[1] "--- Final Combined Dataset ---"
> print(head(final_dataset %>% select(Player, Report, Is_High_Impact)))
  Player                               Report Is_High_Impact
1  KL Rahul (SR: 364.29, 6 Sixes)      FALSE
2 Ishan Kishan (SR: 364.71, 6 Sixes)    TRUE
3 Sunil Narine (SR: 294.12, 5 Sixes)   FALSE
4  Jos Buttler (SR: 372.22, 7 Sixes)    TRUE
5 Sam Billings (SR: 266.67, 5 Sixes)   FALSE
6  KL Rahul (SR: 300, 7 Sixes)         TRUE
```

Environment History Connections Tutorial

Project: (None)

df_text 106 obs. of 11 variables
dropped_multip... 106 obs. of 9 variables
dropped_one 106 obs. of 9 variables
dropped_range 106 obs. of 9 variables
Fastest.Fiftie... 107 obs. of 9 variables
final_dataset 106 obs. of 12 variables
merged_data 106 obs. of 10 variables
names_col 106 obs. of 0 variables

Files Plots Packages Help Viewer Presentation

Home

Name	Size	Modified
tue_sales.csv	708 b	Sep 18, 2025, 8:55 AM
GIS DataBase		
IISExpress		
My Music		
My Pictures		
My Videos		
My Web Sites		
R		
SM PRACTICAL - II 20.txt	1.2 KB	Aug 22, 2025, 3:02 PM
SM PRACTICAL - II.txt	1.2 KB	Aug 22, 2025, 3:01 PM
total_bills.csv	1.2 KB	Sep 18, 2025, 8:47 AM
Virtual Machines		
Visual Studio 2022		
WindowsPowerShell		

ENG IN 13:02 01-12-2025