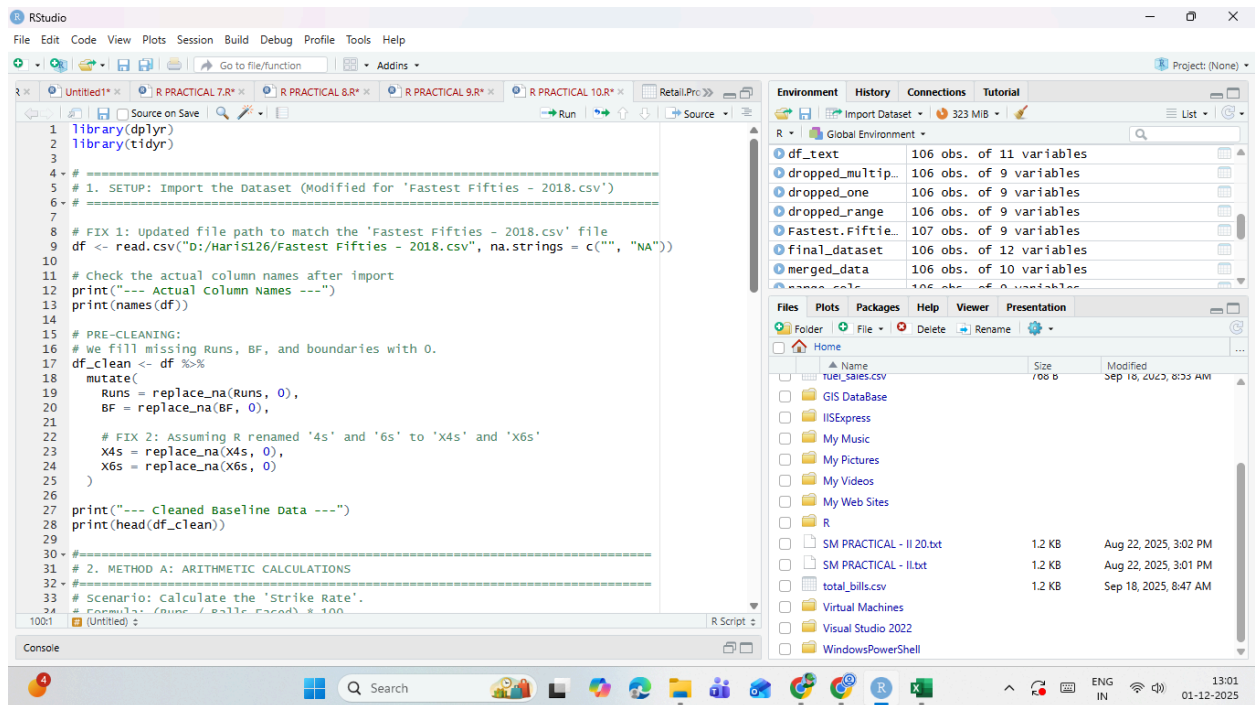


MVLU COLLEGE PRACTICAL NO.10

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

Aim:



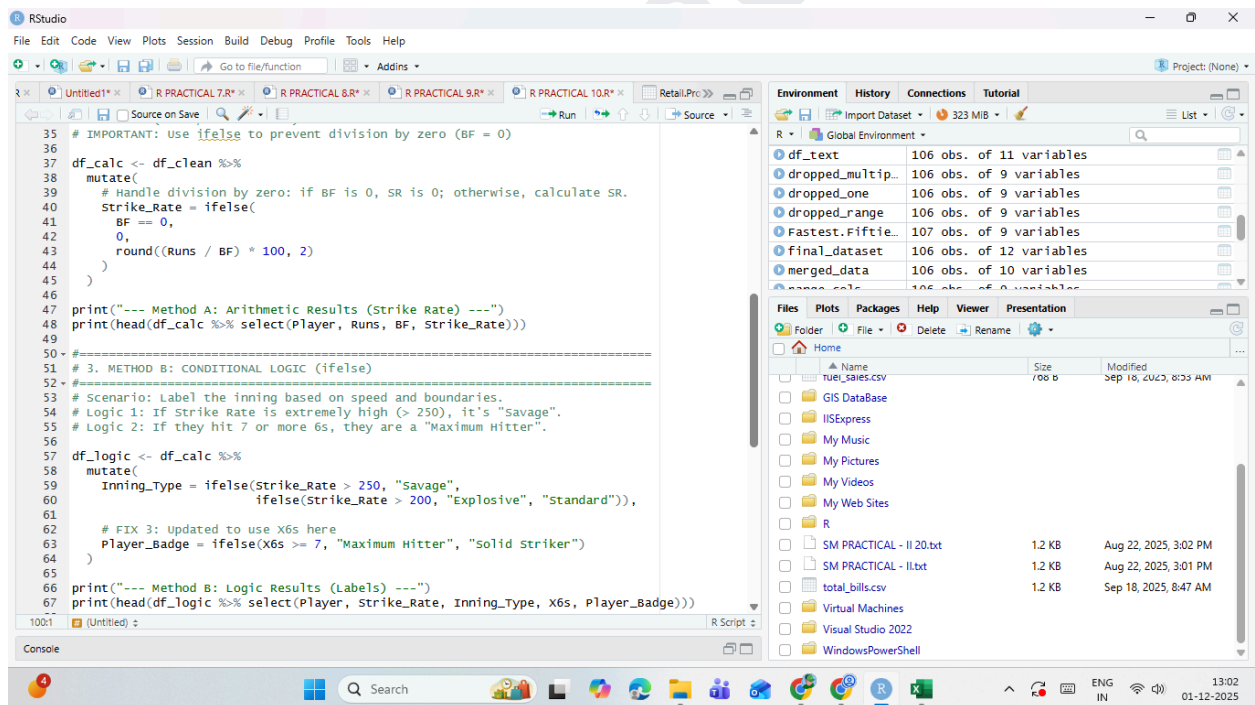
```
1 library(dplyr)
2 library(tidy)
3
4 #
5 # 1. SETUP: Import the Dataset (Modified for 'Fastest Fifties - 2018.csv')
6 #
7
8 # FIX 1: Updated file path to match the 'Fastest Fifties - 2018.csv' file
9 df <- read.csv("D:/Haris126/Fastest Fifties - 2018.csv", na.strings = c("", "NA"))
10
11 # Check the actual column names after import
12 print("--- Actual Column Names ---")
13 print(names(df))
14
15 # PRE-CLEANING:
16 # We 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 / Batsman's Balls Faced) * 100
```

Environment

Object	Obs	Vars
df_text	106	11
dropped_multip...	106	9
dropped_one	106	9
dropped_range	106	9
Fastest.Fiftie...	107	9
final_dataset	106	12
merged_data	106	10

Files

Name	Size	Modified
tue1_sales.csv	706 B	Sep 18, 2025, 8:35 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		



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

Environment

Object	Obs	Vars
df_text	106	11
dropped_multip...	106	9
dropped_one	106	9
dropped_range	106	9
Fastest.Fiftie...	107	9
final_dataset	106	12
merged_data	106	10

Files

Name	Size	Modified
tue1_sales.csv	706 B	Sep 18, 2025, 8:35 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		

MVLU COLLEGE PRACTICAL NO.10

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for data transformation. It includes comments for 'METHOD C: TEXT TRANSFORMATION (paste)' and '5. ALL TOGETHER (The standard workflow)'. The code uses `df_text` and `final_dataset` objects.
- Environment:** Lists objects in the Global Environment, including `df_text` (106 obs. of 11 variables), `dropped_multi...` (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), and `merged_data` (106 obs. of 10 variables).
- Files:** Shows a list of files in the Home directory, including `tue1_sales.csv`, `GIS DataBase`, `IISExpress`, `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`, and `WindowsPowerShell`.
- Console:** Shows the output of the R code, including the text 'Method C: Text Transformation' and the head of the `df_text` object.

Output:

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for data cleaning. It includes comments for 'FIX 1: Updated file path to match the 'Fastest Fifties - 2018.csv' file' and 'PRE-CLEANING'. The code uses `df` and `df_clean` objects.
- Environment:** Lists objects in the Global Environment, including `df_text` (106 obs. of 11 variables), `dropped_multi...` (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), and `merged_data` (106 obs. of 10 variables).
- Files:** Shows a list of files in the Home directory, including `tue1_sales.csv`, `GIS DataBase`, `IISExpress`, `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`, and `WindowsPowerShell`.
- Console:** Shows the output of the R code, including the text 'Cleaned Baseline Data' and the head of the `df_clean` object.

Hariprasad Vishwakarma
S126
SYCS

MVLU COLLEGE PRACTICAL NO.10

The image displays two screenshots of the RStudio interface, showing the execution of R code for data manipulation and analysis.

Top Screenshot:

- Source:** The code defines a function `mutate()` to handle division by zero and calculate the strike rate (SR) as $\text{SR} = \frac{\text{Runs}}{\text{BF}} \times 100$. It then prints the results for Method A: Arithmetic Results (Strike Rate).
- Console:** The output shows the results for Method A, listing players and their SR values.
- Environment:** The environment pane shows the objects created, including `df_text`, `dropped_multip...`, `dropped_one`, `dropped_range`, `Fastest.Fiftie...`, `final_dataset`, and `merged_data`.

Bottom Screenshot:

- Source:** The code continues with Method B: Logic Results (Labels), defining `Inning_Type` and `Player_Badge` based on SR and X6s. It then prints the results for Method B.
- Console:** The output shows the results for Method B, listing players and their SR values, along with the `Inning_Type` and `Player_Badge`.
- Environment:** The environment pane shows the objects created, including `df_text`, `dropped_multip...`, `dropped_one`, `dropped_range`, `Fastest.Fiftie...`, `final_dataset`, and `merged_data`.

Hariprasad Vishwakarma
S126
SYCS