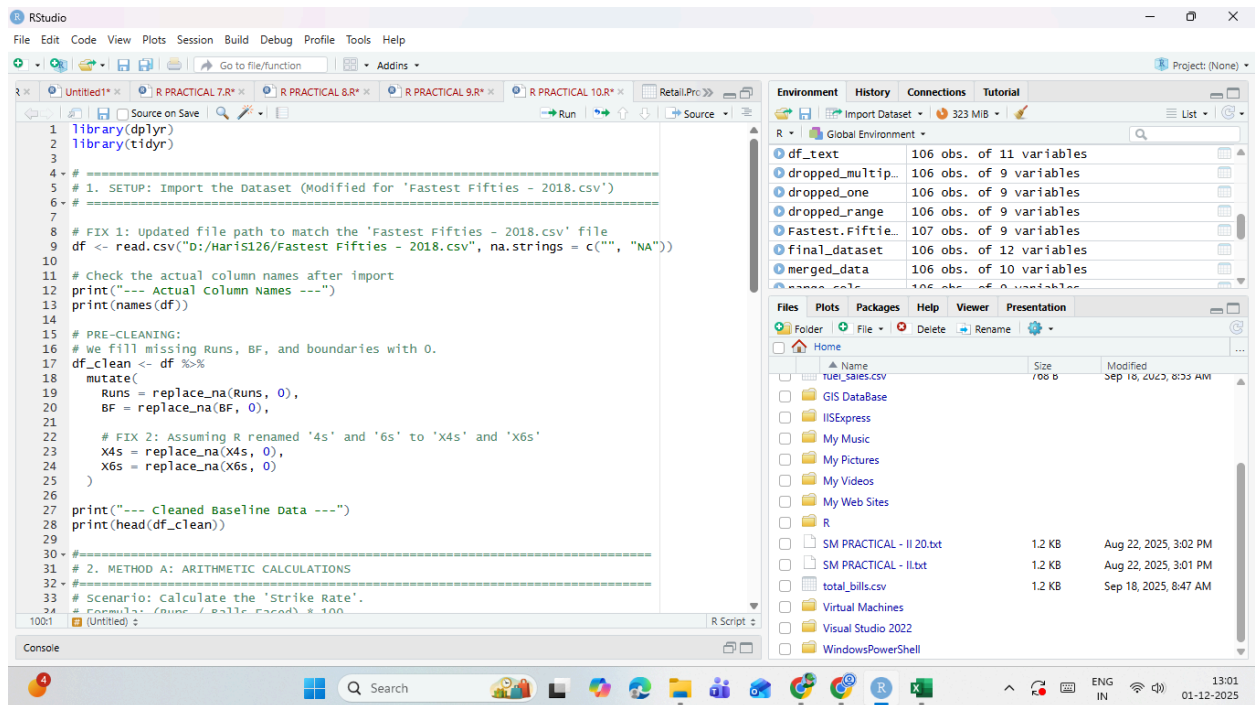


MVLU COLLEGE PRACTICAL NO.10

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

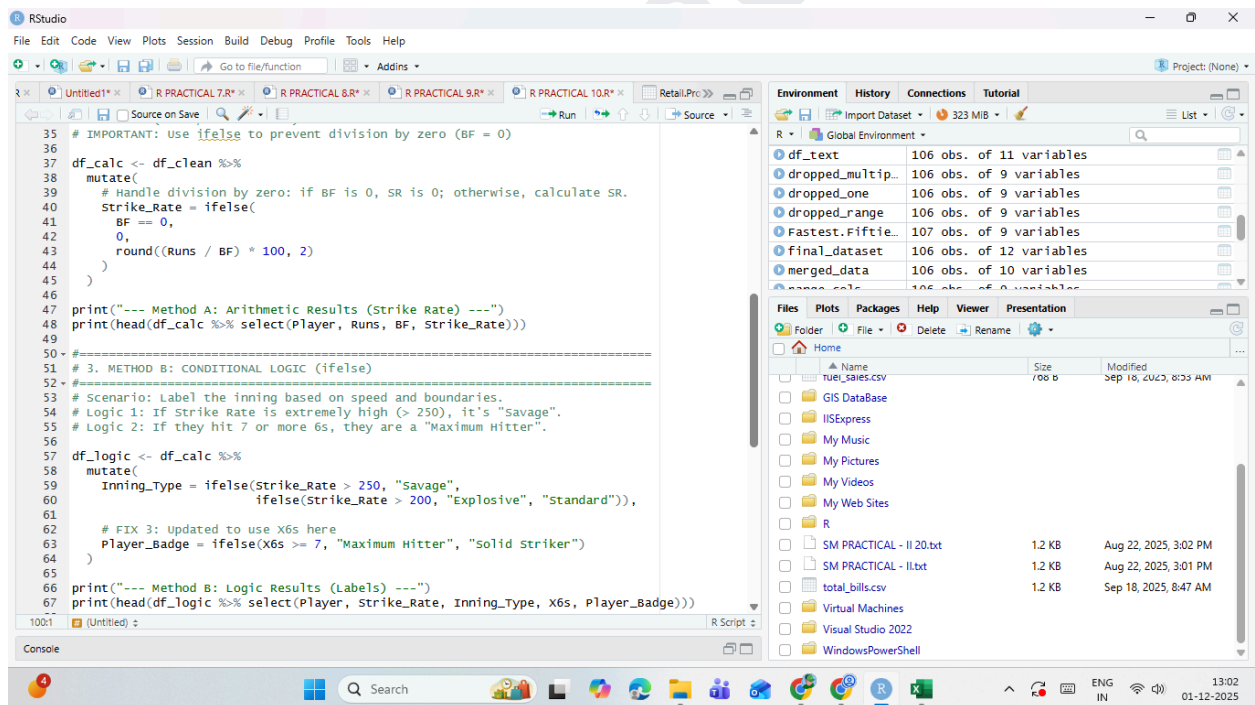
Aim:



```
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 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



```
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

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`.

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_clean` and `df_calc` objects.
- Console:** Displays the output of the R code, including the names of the columns in the `df` object and the cleaned baseline data.
- 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).

Hariprasad Vishwakarma
S126
SYCS

MVLU COLLEGE PRACTICAL NO.10

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Project: (None)

Source
Console Terminal Background Jobs
R - R4.1.2 ~\...
+ 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  51  14      364.29
2 Ishan Kishan  62  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
6  KL Rahul  300.00      Savage  5 Solid Striker
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Project: (None)

Source
Console Terminal Background Jobs
R - R4.1.2 ~\...
> df_text <- df_calc %>%
+   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  KL Rahul (SR: 364.29, 4 sixes) FALSE
2 Ishan Kishan Ishan Kishan (SR: 364.71, 6 sixes) TRUE
3 Sunil Narine Sunil Narine (SR: 294.12, 5 sixes) FALSE
4 Jos Buttler Jos Buttler (SR: 372.22, 7 sixes) TRUE
5 Sam Billings Sam Billings (SR: 266.67, 5 sixes) FALSE
6  KL Rahul  KL Rahul (SR: 300.00, 7 sixes) TRUE
```

Hariprasad Vishwakarma
S126
SYCS