

SHETH L.U.J AND SIR M.V. COLLEGE
SUBJECT NAME: **Data Analysis with SAS / SPSS / R**
Practical No. 10

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

Output-

```
R - R4.1.2 - ~/...
> library(dplyr)
> library(tidyverse)
> # FIX 1: updated file name (removed extra spaces based on your folder screenshot)
> # Verify if your file is "Fastest_Fifties..." or "Fastest_Fifties..."
> # I am using the one that likely matches your windows folder structure.
> df <- read.csv("D:/dhairya/R/Fastest_Fifties_All_Seasons_Combine.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 'D:/dhairya/R/Fastest_Fifties_All_Seasons_Combine.csv': No such file or directory

> # check the actual column names to see if R renamed '6s' to 'X6s'
> print("---- Actual column Names ----")
> print(names(df))
[1] "X"      "Player"  "Runs"    "BF"      "X4s"     "X6s"     "Against"  "Venue"

> # PRE-CLEANING:
> # We fill missing Runs, BF, and 6s with 0.
> df_clean <- df %>%
+   mutate(
+     Runs = replace_na(Runs, 0),
+     BF = replace_na(BF, 0),
+     # FIX 2: changed '6s' to X6s
+     # (R usually adds 'X' to columns starting with numbers)
+     X6s = replace_na(X6s, 0)
+   )
> print("---- Cleaned Baseline Data ----")
[1] "---- Cleaned Baseline Data ----"
> print(head(df_clean))
  X Player Runs BF X4s X6s Against Venue Match.Date
1 0 Yusuf Pathan 61 21 4 6 DEC Rajiv Gandhi Intl. Cricket Stadium 24 April 2008
2 1 Kumar Sangakkara 94 23 13 1 MI IS Bindra Stadium 25 April 2008
3 2 Virender Sehwag 51 23 2 5 PBKS Arun Jaitley Stadium 17 May 2008
4 3 Kumar Sangakkara 50 23 7 2 DEC IS Bindra Stadium 23 May 2008
5 4 Virender Sehwag 71 24 7 4 CSK Chidambaram 02 May 2008
6 5 James Hopes 71 24 10 3 CSK IS Bindra Stadium 19 April 2008

> df_clean <- df_clean %>%
+   mutate(
```

```
R - R4.1.2 - ~/...
+   Inning_Type = ifelse(Strike_Rate > 200, "Explosive", "standard"),
+   # FIX 3: updated to use X6s here as well
+   Player_Badge = ifelse(X6s >= 5, "Power Hitter", "Technical Batter")
+ )
> print("---- Method B: Logic Results (Labels) ----")
[1] "---- Method B: Logic Results (Labels) ----"
> print(head(df_logic %>% select(Player, Strike_Rate, Inning_Type, Player_Badge)))
  Player Strike_Rate Inning_Type Player_Badge
1 Yusuf Pathan 290.48 Explosive Power Hitter
2 Kumar Sangakkara 408.70 Explosive Technical Batter
3 Virender Sehwag 221.74 Explosive Power Hitter
4 Kumar Sangakkara 217.39 Explosive Technical Batter
5 Virender Sehwag 295.83 Explosive Technical Batter
6 James Hopes 295.83 Explosive Technical Batter

> df_text <- df_clean %>%
+   mutate(
+     Match_Summary = paste(Player, "scored", Runs, "runs against", Against)
+   )
> print("---- Method C: Text Transformation ----")
[1] "---- Method C: Text Transformation ----"
> print(head(df_text %>% select(Player, Match_Summary)))
  Player Match_Summary
1 Yusuf Pathan scored 61 runs against DEC "Kumar Sangakkara scored 94 runs against MI"
3 "Virender Sehwag scored 51 runs against PBKS" "Kumar Sangakkara scored 50 runs against DEC"
5 "Virender Sehwag scored 71 runs against CSK" "James Hopes scored 71 runs against CSK"

> final_dataset <- df_clean %>%
+   mutate(
+     Strike_Rate = round((Runs / BF) * 100, 2),
+     Is_High_Performance = ifelse(Runs > 70, TRUE, FALSE),
+     Report = paste0(Player, " (SR: ", round((Runs/BF)*100, 1), ")")
+   )
> print("---- Final Combined Dataset ----")
[1] "---- Final Combined Dataset ----"
> print(head(final_dataset %>% select(Player, Report, Is_High_Performance)))
  Player Report Is_High_Performance
1 Yusuf Pathan Yusuf Pathan (SR: 290.5) FALSE
2 Kumar Sangakkara Kumar Sangakkara (SR: 408.7) TRUE
3 Virender Sehwag Virender Sehwag (SR: 221.7) FALSE
4 Kumar Sangakkara Kumar Sangakkara (SR: 217.4) FALSE
5 Virender Sehwag Virender Sehwag (SR: 295.8) TRUE
6 James Hopes James Hopes (SR: 295.8) TRUE
```