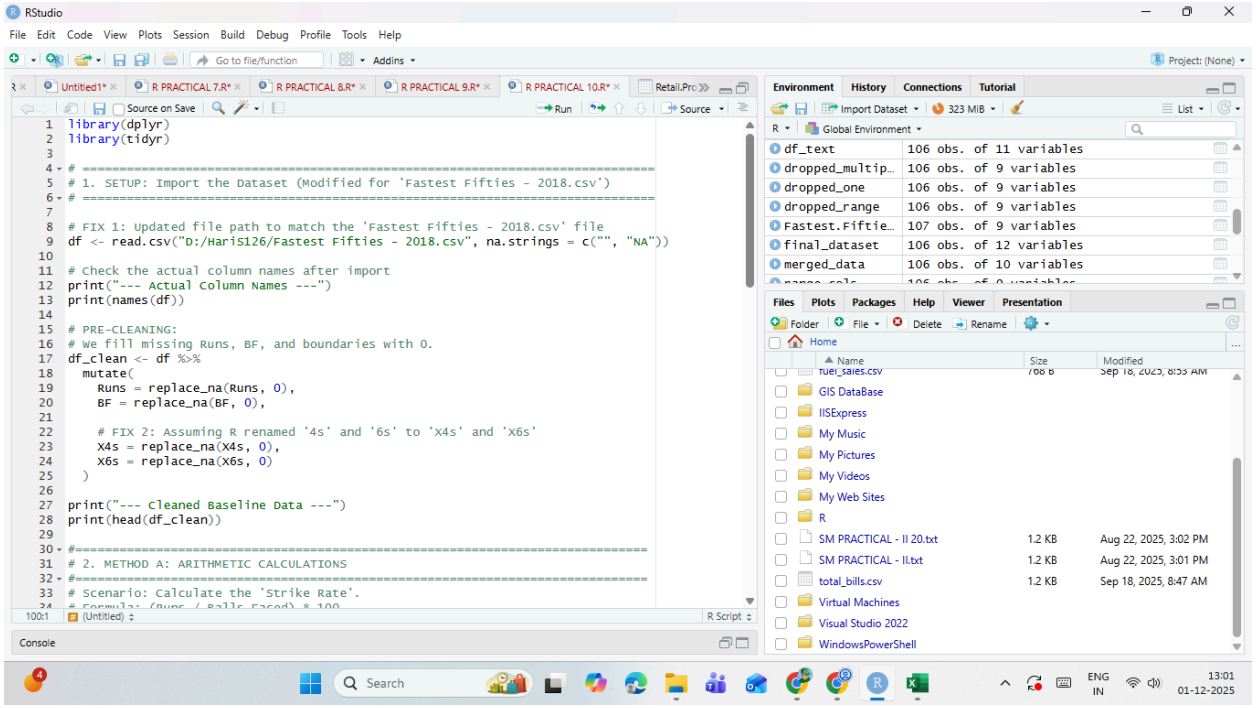
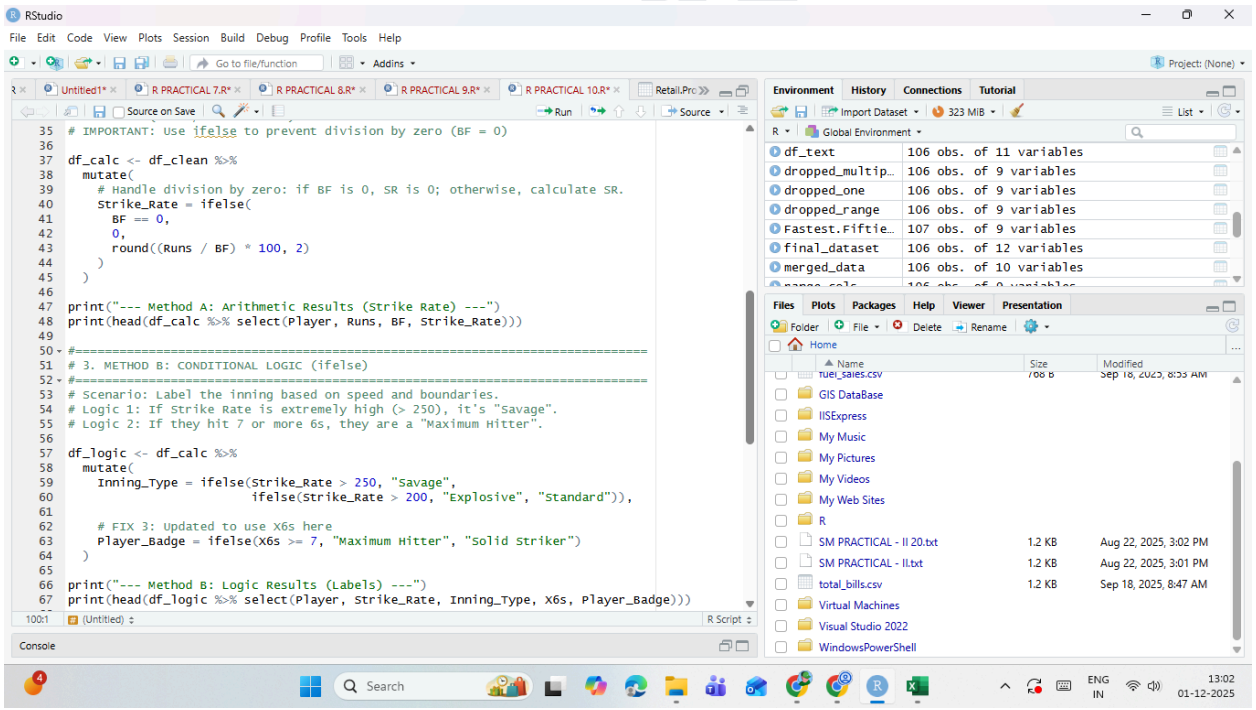
MVLU COLLEGE
PRACTICAL NO.10

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

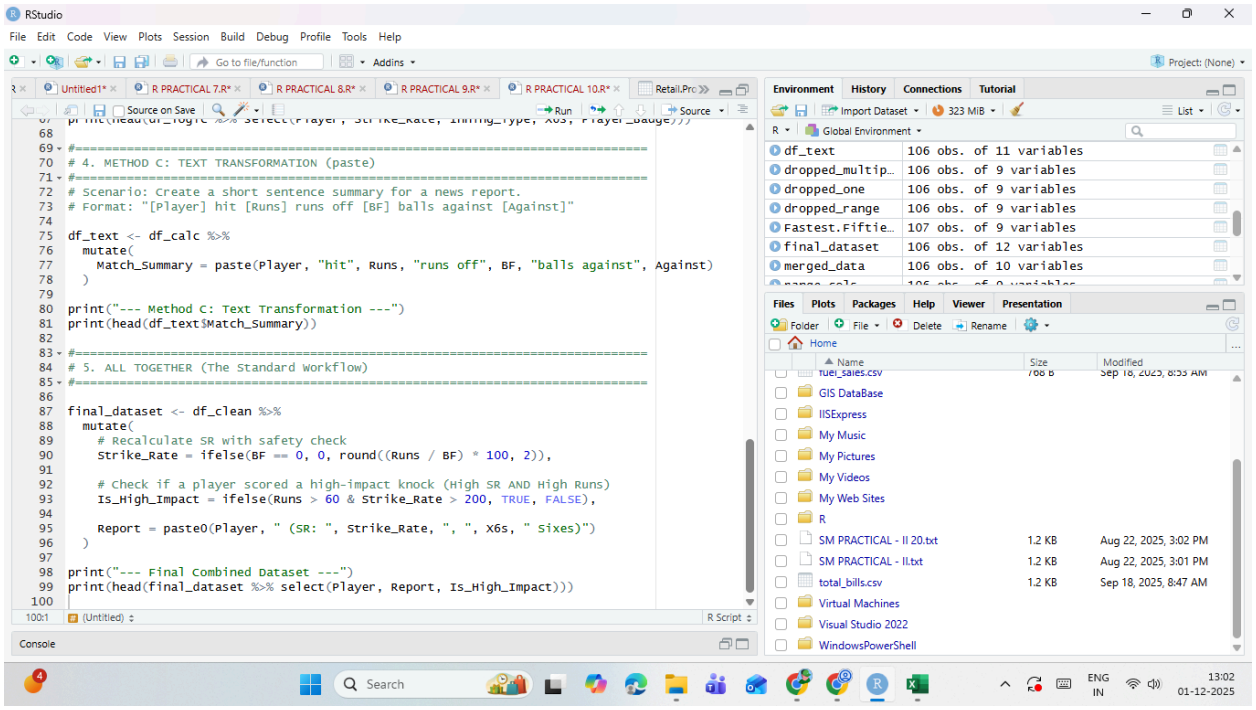
Aim:



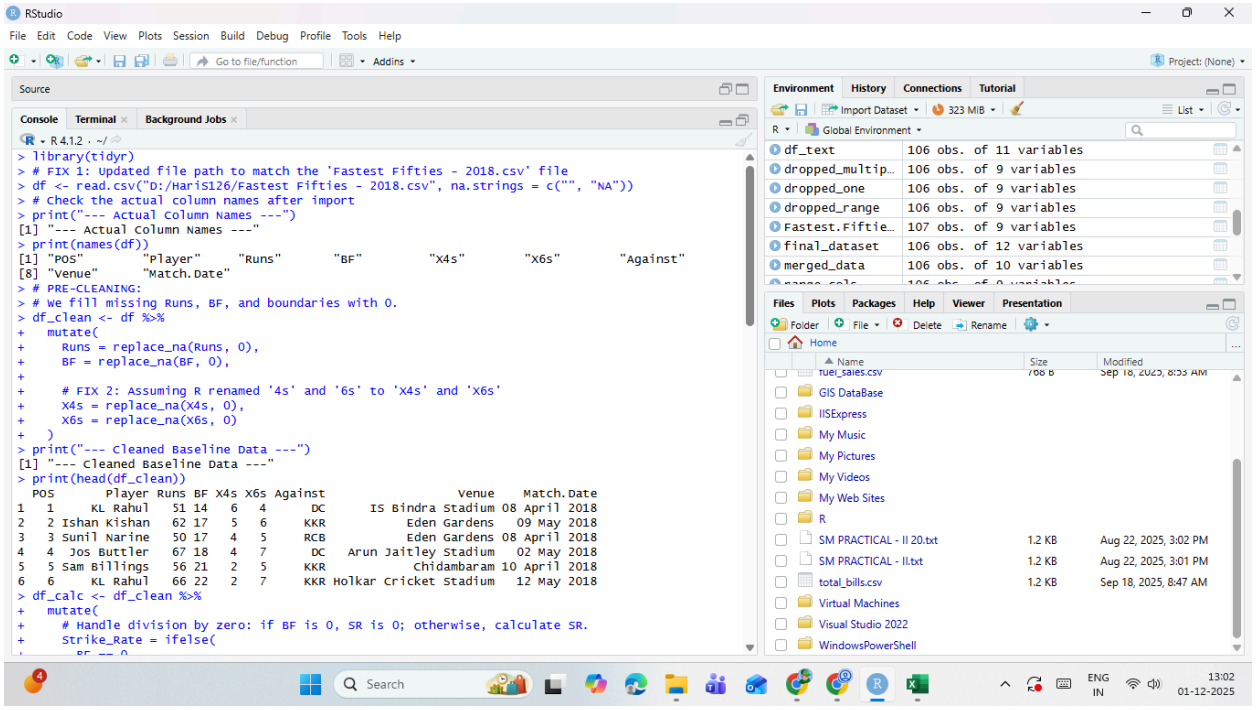

Hariprasad Vishwakarma
S126
SYCS

MVLU COLLEGE
PRACTICAL NO.10



```
#==========================================================================
# 4. METHOD C: TEXT TRANSFORMATION (paste)
#==========================================================================
# Scenario: Create a short sentence summary for a news report.
# Format: "[Player] hit [Runs] runs off [BF] balls against [Against]"

df_text <- df_calc %>%
  mutate(
    Match_Summary = paste(Player, "hit", Runs, "runs off", BF, "balls against", Against)
  )

print("--- Method C: Text Transformation ---")
print(head(df_text$Match_Summary))

#==========================================================================
# 5. ALL TOGETHER (The Standard Workflow)
#==========================================================================

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 ---")
print(head(final_dataset %>% select(Player, Report, Is_High_Impact)))
```

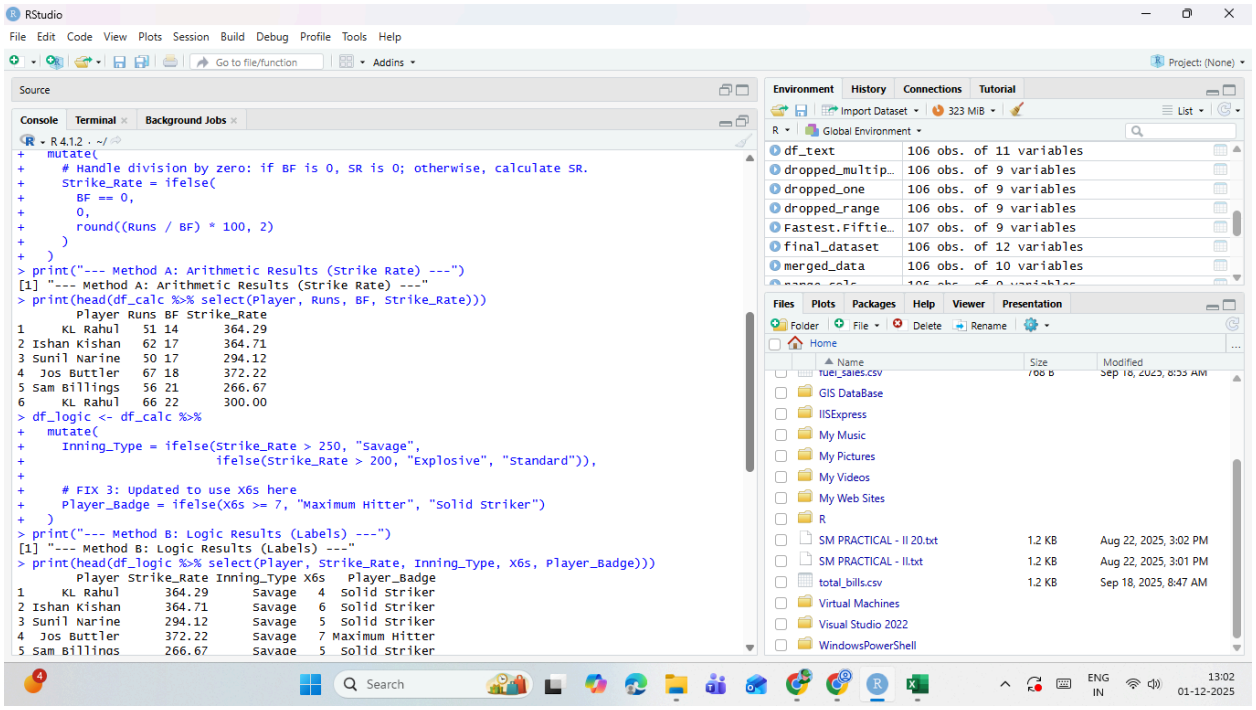
Output:



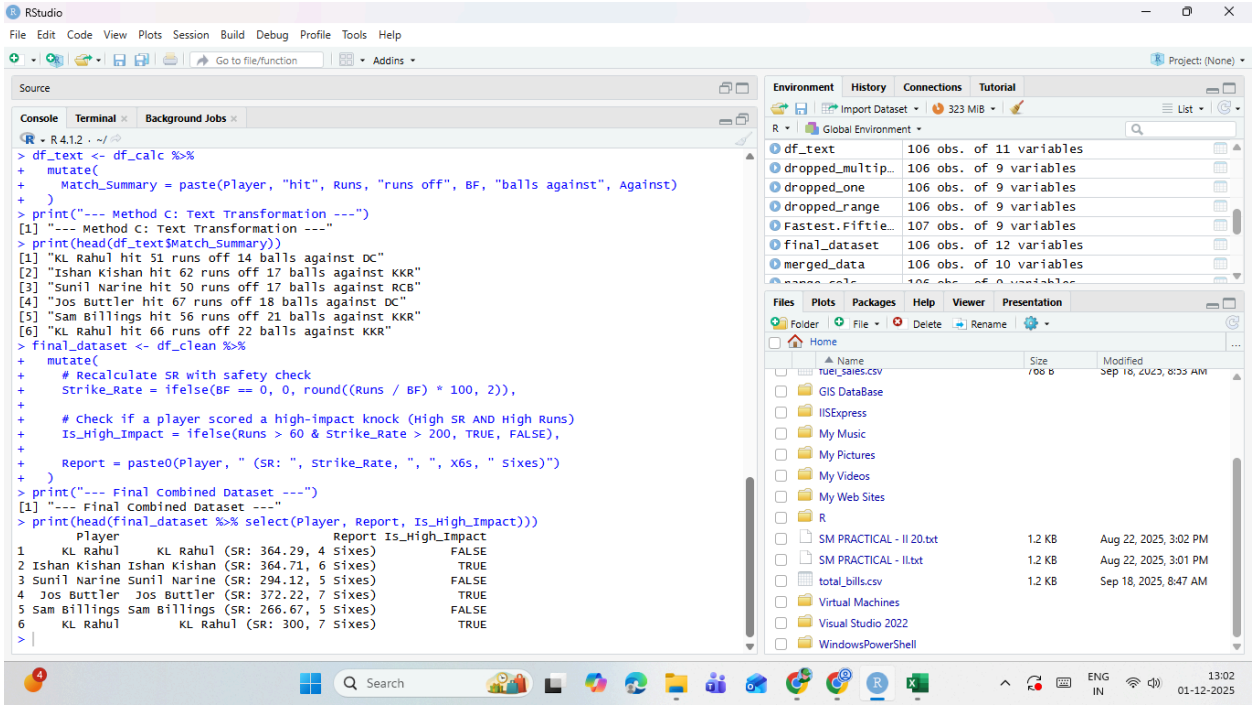
```
> library(tidyr)
> # FIX 1: Updated file path to match the 'Fastest Fifties - 2018.csv' file
> df <- read.csv("D:/HariS126/Fastest Fifties - 2018.csv", na.strings = c("", "NA"))
> # Check the actual column names after import
> print("--- Actual Column Names ---")
[1] "--- Actual Column Names ---"
> print(names(df))
[1] "POS"        "Player"     "Runs"       "BF"         "X4s"        "X6s"        "Against"
[8] "Venue"      "Match.Date"
> # PRE-CLEANING:
> # We fill missing Runs, BF, and boundaries with 0.
> df_clean <- df %>%
+   mutate(
+     Runs = replace_na(Runs, 0),
+     BF = replace_na(BF, 0),
+
+     # FIX 2: Assuming R renamed '4s' and '6s' to 'X4s' and 'X6s'
+     X4s = replace_na(X4s, 0),
+     X6s = replace_na(X6s, 0)
+   )
> print("--- Cleaned Baseline Data ---")
[1] "--- Cleaned Baseline Data ---"
> print(head(df_clean))
  POS       Player Runs BF X4s X6s Against                 Venue    Match.Date
1   1     KL Rahul   51 14   6   4      DC IS Bindra Stadium 08 April 2018
2   2 Ishan Kishan   62 17   5   6     KKR          Eden Gardens    09 May 2018
3   3 Sunil Narine   50 17   4   5     RCB          Eden Gardens 08 April 2018
4   4  Jos Buttler   67 18   4   7      DC  Arun Jaitley Stadium    02 May 2018
5   5 Sam Billings   56 21   2   5     KKR           Chidambaram 10 April 2018
6   6     KL Rahul   66 22   2   7     KKR Holkar Cricket Stadium   12 May 2018
> df_calc <- df_clean %>%
+   mutate(
+     # Handle division by zero: if BF is 0, SR is 0; otherwise, calculate SR.
+     Strike_Rate = ifelse(
```

Hariprasad Vishwakarma
S126
SYCS

MVLU COLLEGE
PRACTICAL NO.10



```
+   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
```



```
> 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     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, 7 Sixes)           TRUE
>
```

Hariprasad Vishwakarma
S126
SYCS