

MVLU COLLEGE
PRACTICAL NO.9

Aim: Performing text manipulation using `str_sub()`, `str_split()` (R). import dataset

Code:

The screenshot shows the RStudio interface with the following details:

- Header:** RStudio, File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Go to file/function, Addins.
- Code Editor:** An R script titled "R PRACTICAL 6.R" containing code for data manipulation and analysis. The code includes loading libraries (dplyr, tidyverse), setting up data, renaming columns, and performing data manipulation.
- Environment Tab:** Shows the global environment with objects like Global Environment, dropped_range, Fastest.Fifties, merged_data, range_cols, Retail.Product, runs_by_month, and runs_by_opponent.
- Files Tab:** Displays the file structure of the current project, including files like tuel_sales.csv, GIS Database, IIS Express, 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:** A small window at the bottom left showing the command "R Script".
- Bottom Bar:** Icons for Search, Home, Recent, and other application switches.

The screenshot shows the RStudio interface with the following details:

- Top Bar:** R Studio, File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Go to file/function, Addins.
- Code Editor:** R PRACTICAL 6.R, Untitled*, R PRACTICAL 7.R*, R PRACTICAL 8.R*, R PRACTICAL 9.R*. The code is for data manipulation, specifically splitting the 'Match_Date' column into 'Day_of_Month', 'Month', and 'Year_Trunc'. It also calculates average runs per inning by opponent.
- Environment Tab:** Shows the Global Environment with objects like dropped_one, dropped_range, Fastest.Fifties, merged_data, range_cols, Retail.Product, runs_by_month, and runs_by_opponent.
- Files Tab:** Shows the project structure with files like tuel_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_bill.csv, Virtual Machines, Visual Studio 2022, and WindowsPowerShell.
- Console:** Shows the command "R Script" and the current file "Untitled".
- Bottom Bar:** Search, Home, Recent, Project (None), ENG IN, 12:55 PM, 01-12-2025.

MVLU COLLEGE

PRACTICAL NO.9

```

# Group the data by the opposition column
group_by(opposition) %>%
  summarise(
    Total_Innings = n(),
    Total_Runs = sum(Runs, na.rm = TRUE),
    Average_Runs = mean(Runs, na.rm = TRUE),
    Median_BF = median(BF, na.rm = TRUE)
  ) %>%
  # Arrange the results to see the best averages first
  arrange(desc(Average_Runs))

print("--- 3. Summarized Runs by Opposition ---")
print(runs_by_opponent)

# -----
# 4. BONUS: Highest Scoring Month
# Goal: Find the total runs scored in each month.
# -----
runs_by_month <- cricket_clean %>%
  # Group by the extracted Month column
  group_by(Month) %>%
  summarise(
    Total_Runs_in_Month = sum(Runs, na.rm = TRUE),
    Total_Fifties = n()
  ) %>%
  arrange(desc(Total_Runs_in_Month))

print("--- 4. Total Runs Scored by Month ---")
print(runs_by_month)

```

Output:

```

> library(stringr)
> # IMPORTANT: Using your corrected file path and the confirmed file name.
> # Note: R automatically converts '4's and '6's to 'x4s' and 'x6s'.
> cricket_df <- read.csv("D:/HariS126/Fastest Fifties - 2018.csv")
> # 1A. Rename columns for clarity (e.g., changing R's automatic x4s to Fours)
> cricket_df <- cricket_df %>%
+   rename(
+     Fours = x4s,
+     Sixes = x6s,
+     Opposition = against,
+     Match_Date = Match.Date
+   )
> print("--- 1. Data structure after Loading and Renaming ---")
[1] "--- 1. Data structure after Loading and Renaming ---"
> print(names(cricket_df))
[1] "POS"          "Player"        "Runs"          "BF"            "Fours"         "Sixes"         "Opposition"
[8] "Venue"        "Match.Date"
> print(head(cricket_df %>% select(Player, Fours, Sixes, Match.Date), 3))
  Player Fours Sixes Match.Date
1  KL Rahul     6    4 08 April 2018
2 Ishan Kishan   5    6 09 May 2018
3 Sunil Narine   4    5 08 April 2018
> cricket_clean <- cricket_df %>%
+   # use separate() to split the Match.Date column
+   separate(
+     col = Match.Date,
+     into = c("Day_of_Month", "Month", "Year_Trunc"),
+     sep = "-",
+     remove = FALSE # keep the original Match.Date column
+   )

Warning message:
Expected 3 pieces. Missing pieces filled with `NA` in 106 rows [1, 2, 3, 4, 5, 6, 7, 8, 9, 10,
11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].

```

Hariprasad Vishwakarma
S126
SYCS

MVLU COLLEGE

PRACTICAL NO.9

RStudio Environment pane:

- Global Environment: dropped_range, Fastest.Fifties, merged_data, range_cols, Retail.Product, runs_by_month, runs_by_opponent
- File list: tue_sales.csv, GIS DataBase, IISExpress, My Music, My Pictures, My Videos, My Web Sites, R, total_bill.csv, Virtual Machines, Visual Studio 2022, WindowsPowerShell

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source Terminal Background Jobs
R 4.1.2 ~/ ...
warning messages:
Expected 3 pieces. Missing pieces filled with 'NA' in 106 rows [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
> print(" --- 2. Date Splitting Results ---")
[1] " --- 2. Date Splitting Results ---"
> print(head(cricket_clean %>% select(Player, Match_Date, Month), 3))
  Player Match_Date Month
1 KL Rahul 08 April 2018 <NA>
2 Ishan Kishan 09 May 2018 <NA>
3 Sunil Narine 08 April 2018 <NA>
> runs_by_opponent <- cricket_clean %>%
+   # Group the data by the Opposition column
+   group_by(Opposition) %>%
+   # Summarise the key statistics for each group
+   summarise(
+     Total_Innings = n(),
+     Total_Runs = sum(Runs, na.rm = TRUE),
+     Average_Runs = mean(Runs, na.rm = TRUE),
+     Median_BF = median(BF, na.rm = TRUE)
+   ) %>%
+   # Arrange the results to see the best averages first
+   arrange(desc(Average_Runs))
> print(" --- 3. Summarized Runs by opposition ---")
[1] " --- 3. Summarized Runs by opposition ---"
> print(runs_by_opponent)
# A tibble: 8 x 5
  Opposition Total_Innings Total_Runs Average_Runs Median_BF
<chr>           <int>      <int>        <dbl>       <dbl>
1 SRH              13      1025       78.8        33
2 MI               10       734       73.4        35
3 RCB              15      1042       69.5        32
4 RR               14       949       67.8        32.5
5 CSK              14       934       66.7        31.5

```

RStudio Environment pane:

- Global Environment: dropped_range, Fastest.Fifties, merged_data, range_cols, Retail.Product, runs_by_month, runs_by_opponent
- File list: tue_sales.csv, GIS DataBase, IISExpress, My Music, My Pictures, My Videos, My Web Sites, R, total_bill.csv, Virtual Machines, Visual Studio 2022, WindowsPowerShell

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source Terminal Background Jobs
R 4.1.2 ~/ ...
+ ) %>%
+   # Arrange the results to see the best averages first
+   arrange(desc(Average_Runs))
> print(" --- 3. Summarized Runs by opposition ---")
[1] " --- 3. Summarized Runs by opposition ---"
> print(runs_by_opponent)
# A tibble: 8 x 5
  Opposition Total_Innings Total_Runs Average_Runs Median_BF
<chr>           <int>      <int>        <dbl>       <dbl>
1 SRH              13      1025       78.8        33
2 MI               10       734       73.4        35
3 RCB              15      1042       69.5        32
4 RR               14       949       67.8        32.5
5 CSK              14       934       66.7        31.5
6 DC                13       866       66.6        28
7 KKR              12       738       61.5        30
8 PBKS              15       913       60.9        36
> runs_by_month <- cricket_clean %>%
+   # Group by the extracted Month column
+   group_by(Month) %>%
+   summarise(
+     Total_Runs_in_Month = sum(Runs, na.rm = TRUE),
+     Total_Fifties = n()
+   ) %>%
+   arrange(desc(Total_Runs_in_Month))
> print(" --- 4. Total Runs Scored by Month ---")
[1] " --- 4. Total Runs Scored by Month ---"
> print(runs_by_month)
# A tibble: 1 x 3
  Month Total_Runs_in_Month Total_Fifties
<chr>           <int>          <int>
1 NA                  2201            106

```