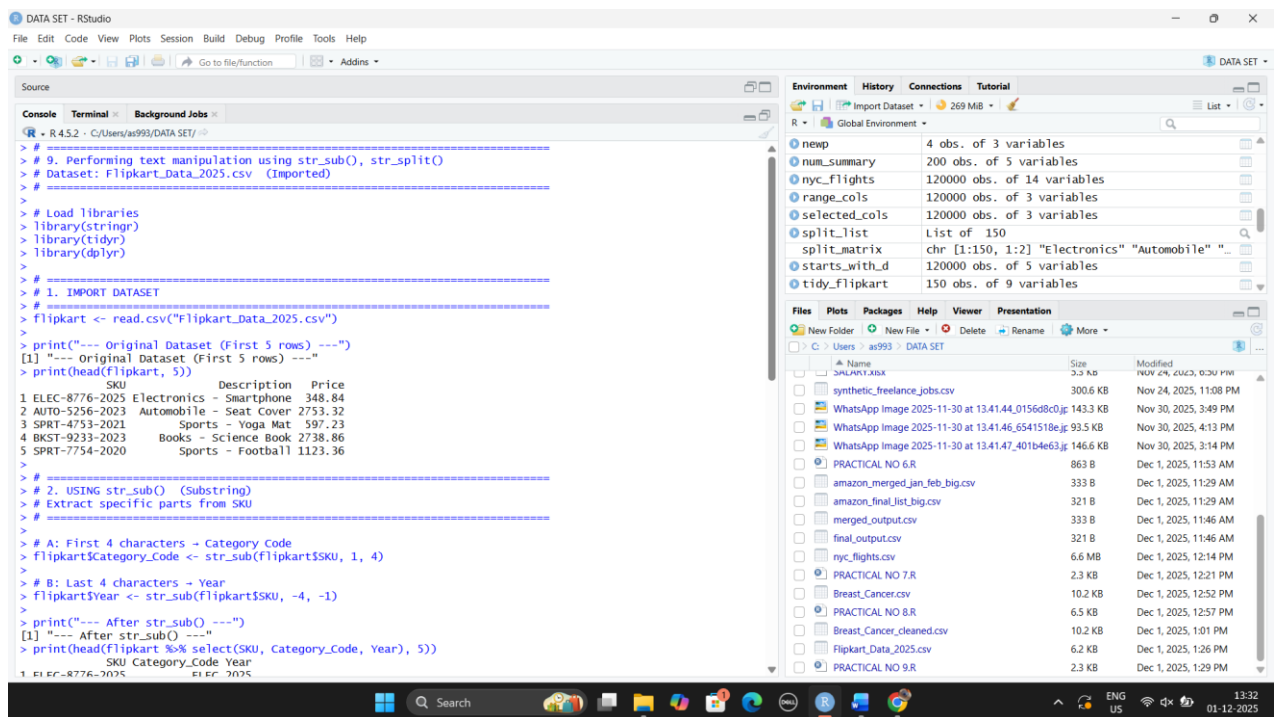


PRACTICAL NO 9

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

OUTPUT:-



ABHISHEK DINESH SINGH
S116

Sheth L.U.J. College of Arts And Sir M.V. College of Science and Commerce

Data Analysis with SAS / SPSS /R

```
DATA SET - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
R • R 4.5.2 • C:/Users/as993/DATA SET/
> # B: Last 4 characters = Year
> flipkart$Year <- str_sub(flipkart$SKU, -4, -1)
> print("---- After str_sub() ----")
[1] "---- After str_sub() ----"
> print(head(flipkart %>% select(SKU, Category_Code, Year), 5))
  SKU Category_Code Year
1 ELEC-8776-2025    ELEC 2025
2 AUTO-5256-2023    AUTO 2023
3 SPRT-4753-2021    SPRT 2021
4 BKST-9233-2023    BKST 2023
5 SPRT-7754-2020    SPRT 2020
> #
> # 3. USING str_split() (Split string)
> # Split 'Description' like 'Electronics - Smart TV'
> #
> # Method A: Basic split (list format)
> split_list <- str_split(flipkart$Description, " - ")
> print("---- Basic Split (First row) ----")
[1] "---- Basic Split (First row) ----"
> print(split_list[[1]])
[1] "Electronics" "Smartphone"
> #
> # Method B: Simplify to matrix for easy column creation
> split_matrix <- str_split(flipkart$Description, " - ", simplify = TRUE)
>
> flipkart$Main_Cat <- split_matrix[, 1]
> flipkart$Sub_Cat <- split_matrix[, 2]
>
> print("---- After str_split() (Manual) ----")
[1] "---- After str_split() (Manual) ----"
> print(head(flipkart %>% select(Description, Main_Cat, Sub_Cat), 5))
  Description Main_Cat Sub_Cat
1 Electronics - Smartphone Electronics Smartphone
2 Automobile - Seat Cover Automobile Seat Cover
3 Sports - Yoga Mat Sports Yoga Mat
4 Books - Science Book Books Science Book
5 Sports - Football Sports Football
> #
> # 4. BONUS: Using separate()
> # Automatically splits SKU into 3 columns
> tidy_flipkart <- flipkart %>%
+ separate(SKU, into = c("Dept", "Prod_ID", "Mfg_Year"), sep = "-")
>
> print("---- Using separate() ----")
[1] "---- Using separate() ----"
> print(head(tidy_flipkart %>% select(Dept, Prod_ID, Mfg_Year), 5))
  Dept Prod_ID Mfg_Year
1 ELEC 8776 2025
2 AUTO 5256 2023
3 SPRT 4753 2021
4 BKST 9233 2023
5 SPRT 7754 2020
> #
```

```
DATA SET - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
R • R 4.5.2 • C:/Users/as993/DATA SET/
> # Method A: Basic split (list format)
> split_list <- str_split(flipkart$Description, " - ")
> print("---- Basic Split (First row) ----")
[1] "---- Basic Split (First row) ----"
> print(split_list[[1]])
[1] "Electronics" "Smartphone"
> #
> # Method B: Simplify to matrix for easy column creation
> split_matrix <- str_split(flipkart$Description, " - ", simplify = TRUE)
>
> flipkart$Main_Cat <- split_matrix[, 1]
> flipkart$Sub_Cat <- split_matrix[, 2]
>
> print("---- After str_split() (Manual) ----")
[1] "---- After str_split() (Manual) ----"
> print(head(flipkart %>% select(Description, Main_Cat, Sub_Cat), 5))
  Description Main_Cat Sub_Cat
1 Electronics - Smartphone Electronics Smartphone
2 Automobile - Seat Cover Automobile Seat Cover
3 Sports - Yoga Mat Sports Yoga Mat
4 Books - Science Book Books Science Book
5 Sports - Football Sports Football
> #
> # 4. BONUS: Using separate()
> # Automatically splits SKU into 3 columns
> tidy_flipkart <- flipkart %>%
+ separate(SKU, into = c("Dept", "Prod_ID", "Mfg_Year"), sep = "-")
>
> print("---- Using separate() ----")
[1] "---- Using separate() ----"
> print(head(tidy_flipkart %>% select(Dept, Prod_ID, Mfg_Year), 5))
  Dept Prod_ID Mfg_Year
1 ELEC 8776 2025
2 AUTO 5256 2023
3 SPRT 4753 2021
4 BKST 9233 2023
5 SPRT 7754 2020
> #
```