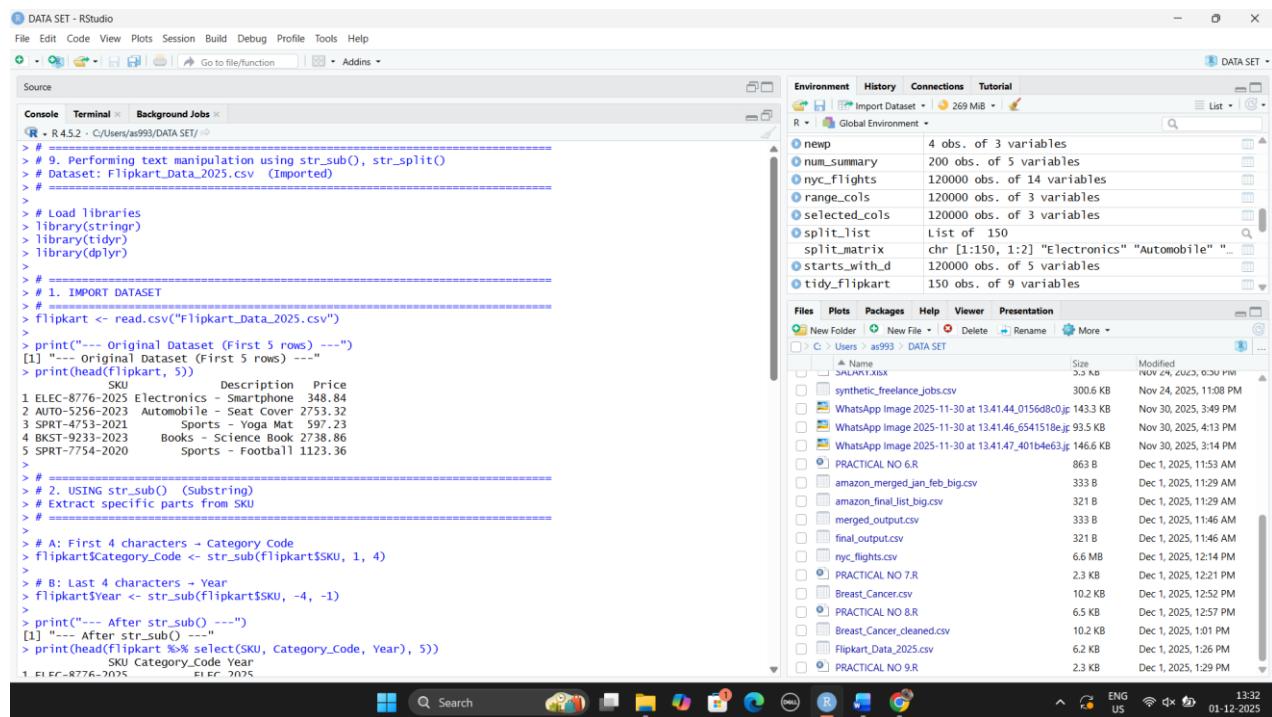


PRACTICAL NO 9

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

OUTPUT:-



The screenshot shows the RStudio interface with the following details:

- Environment pane:** Shows various objects created in the session, including `newp`, `num_summary`, `nyc_flights`, `range_cols`, `selected_cols`, `split_list`, `split_matrix`, `starts_with_d`, and `tidy_Flipkart`.
- Files pane:** Shows a list of files in the current directory, including `DATA SET.RData`, `synthetic_freelance.jobs.csv`, and several WhatsApp Image files.
- Code pane:** Displays the R code for the practical, which includes importing the dataset, performing str_sub() operations, and printing the results.

```

# =====#
# 9. Performing text manipulation using str_sub(), str_split()
# Dataset: Flipkart_Data_2025.csv (Imported)
# =====#
# Load libraries
library(stringr)
library(tidyverse)
library(dplyr)
# =====#
# 1. IMPORT DATASET
# Import
flipkart <- read.csv("Flipkart_Data_2025.csv")
# print("--- Original Dataset (First 5 rows) ---")
[1] "--- original dataset (First 5 rows) ---"
print(head(flipkart, 5))
  SKU Description Price
1 ELEC-8776-2025 Electronics - Smartphone 348.84
2 AUTO-5256-2023 Automobile - Seat Cover 2753.32
3 SPRT-4753-2021 Sports - Yoga Mat 597.23
4 BKST-9233-2023 Books - Science Book 2738.86
5 SPRT-7754-2020 Sports - Football 1123.36
# =====#
# 2. USING str_sub() (Substring)
# Extract specific parts from sku
# =====#
# A: First 4 characters - Category Code
flipkart$Category_Code <- str_sub(flipkart$SKU, 1, 4)
# 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

```

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

Data Analysis with SAS / SPSS / R

The screenshot shows an RStudio interface with the following details:

- Console:** Displays R code for manipulating a dataset named `flipkart`. The code includes:
 - Extracting the last 4 characters from the `Year` column.
 - Using `str_sub` to remove the last 4 characters from the `SKU` column.
 - Printing the result of the `str_sub` operation.
 - Selecting specific rows from the dataset based on the modified `SKU` and `Category_Code`.
 - Displaying the first few rows of the resulting dataset.
 - Using `str_split` to split the `Description` column into `Main_Cat` and `Sub_Cat`.
 - Comparing manual splitting with the `simplify = TRUE` option of `str_split`.
 - Handling missing values in the `Sub_Cat` column.
 - Using `separate` to split the `SKU` column into `Dept`, `Prod_ID`, and `Mfg_Year`.
 - Final output showing the first few rows of the tidy dataset.
- Environment:** Shows the global environment with objects like `newp`, `num_summary`, `nyc_flights`, etc.
- File Explorer:** Shows files in the current directory, including CSV files like `synthetic_freelance_jobs.csv` and `WhatsApp Image 2025-11-30 at 13:41:44_0156d8c0.jpg`.
- File Bar:** Includes tabs for `DATA SET`, `Source`, `Terminal`, and `Background Jobs`.
- System Bar:** Shows the date (01-12-2025), time (13:32), and system status (ENG US).

This screenshot shows an RStudio session with similar content to the first one, but with additional code and results:

- Console:** Continues the data manipulation process, including:
 - Method A: Basic split using `str_split`.
 - Method B: Simplifying the matrix for easy column creation using `str_split` with `simplify = TRUE`.
 - Splitting the `flipkart` dataset into `Main_Cat` and `Sub_Cat`.
 - Comparing manual splitting with the `simplify = TRUE` option of `str_split`.
 - Handling missing values in the `Sub_Cat` column.
 - Using `separate` to split the `SKU` column into `Dept`, `Prod_ID`, and `Mfg_Year`.
 - Final output showing the first few rows of the tidy dataset.
- Environment:** Shows the global environment with objects like `newp`, `num_summary`, `nyc_flights`, etc.
- File Explorer:** Shows files in the current directory, including CSV files like `synthetic_freelance_jobs.csv` and `WhatsApp Image 2025-11-30 at 13:41:44_0156d8c0.jpg`.
- File Bar:** Includes tabs for `DATA SET`, `Source`, `Terminal`, and `Background Jobs`.
- System Bar:** Shows the date (01-12-2025), time (13:33), and system status (ENG US).