

# SHETH L.U.J. AND SIR M.V. COLLEGE

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

The screenshot shows the RStudio interface. The console on the left displays the following commands and output:

```
> install.packages("stringr")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and
d install the appropriate version of Rtools before proceeding:
https://cran.rstudio.com/bin/windows/Rtools/
Warning: package 'stringr' is in use and will not be installed

> install.packages("tidyr")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and
d install the appropriate version of Rtools before proceeding:
https://cran.rstudio.com/bin/windows/Rtools/
Warning: package 'tidyr' is in use and will not be installed

> library(stringr)
> library(tidyr)
> library(dplyr)
> transport_df <- read_csv("Synthetic_Transportation_Dataset_Expanded_v2.csv")
Rows: 5500 Columns: 17
#> Column specification
Delimiter: ","
chr (9): Timestamp, Road_ID, Vehicle_Type, Traffic_Density, Weather, Road_Condition, Accide...
dbl (8): Record_ID, Latitude, Longitude, Vehicle_Count, Avg_Speed(km/h), Visibility(m), tem...
```

The Environment pane on the right shows the loaded objects:

- `region_filter`: 3278 obs. of 6 variables
- `robot_related`: 1 obs. of 3 variables
- `selected_cols`: 20000 obs. of 4 variables
- `sleep_df`: 62 obs. of 8 variables
- `special_buildings`: 3007 obs. of 6 variables
- `split_matrix`: chr [1:5500, 1] "Fog" "Rain" "Storm" "Clear" "Storm" "S..."
- `starts_with_s`: 20000 obs. of 3 variables
- `tidy_transport`: 5500 obs. of 24 variables
- `transport_df`: 5500 obs. of 23 variables

The screenshot shows the RStudio interface with the following commands and output in the console:

```
> transport_df$Day <- str_sub(transport_df$Timestamp, 1, 2)
> transport_df$Month <- str_sub(transport_df$Timestamp, 4, 5)
> transport_df$Year <- str_sub(transport_df$Timestamp, 7, 10)
> transport_df$Hour <- str_sub(transport_df$Timestamp, 12, 13)
> print("After str_sub() Extraction (Corrected) ----")
[1] "After str_sub() Extraction (Corrected) ----"
> print(transport_df %>% select(Timestamp, Day, Month, Year, Hour) %>% head())
#> A tibble: 6 x 5
  Timestamp      Day Month Year Hour
  <chr>         <chr> <chr> <chr> <chr>
1 03-10-2023 09:50 03 10 2023 09
2 09-06-2019 08:01 09 06 2019 08
3 08-01-2020 01:38 08 01 2020 01
4 18-10-2022 05:43 18 10 2022 05
5 29-06-2023 23:11 29 06 2023 23
6 02-03-2020 15:55 02 03 2020 15

# 19 more variables: Traffic_Density <chr>, Weather <chr>, Visibility(m) <dbl>,
# Road_Condition <chr>, Accident_Occurred <chr>, Accident_Severity <chr>,
# Alert_Generated <chr>, temperature <dbl>, humidity <dbl>
> transport_df$Weather_Detail <- str_split(transport_df$Weather, " ", simplify = TRUE)
> transport_df$Weather_Main <- split_matrix[, 1]
> if (ncol(split_matrix) > 2) {
+   transport_df$Weather_Detail <- split_matrix[, 2]
+ } else {
+   transport_df$Weather_Detail <- NA
+ }
> print("After Safe Weather Split ----")
[1] "After Safe Weather Split ----"
> print(transport_df %>% select(Weather, Weather_Main, Weather_Detail) %>% head())
#> A tibble: 6 x 3
  Weather Weather_Main Weather_Detail
  <chr>      <chr>      <list>
1 Fog      Fog      NA
2 Rain     Rain     NA
```

The Environment pane on the right shows the loaded objects:

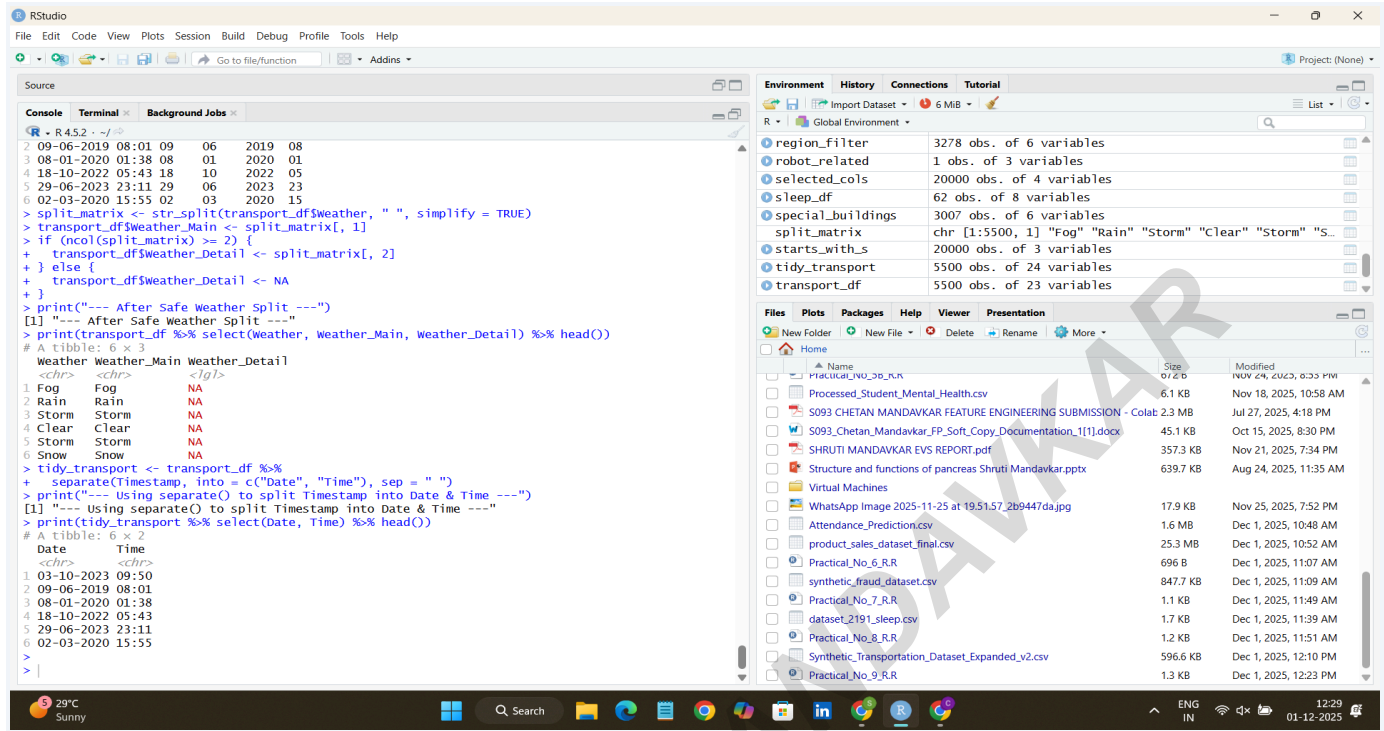
- `region_filter`: 3278 obs. of 6 variables
- `robot_related`: 1 obs. of 3 variables
- `selected_cols`: 20000 obs. of 4 variables
- `sleep_df`: 62 obs. of 8 variables
- `special_buildings`: 3007 obs. of 6 variables
- `split_matrix`: chr [1:5500, 1] "Fog" "Rain" "Storm" "Clear" "Storm" "S..."
- `starts_with_s`: 20000 obs. of 3 variables
- `tidy_transport`: 5500 obs. of 24 variables
- `transport_df`: 5500 obs. of 23 variables

NAME:- CHETAN MANDAVKAR

ROLL NO. S093

SUBJECT:- Data Analysis with SAS / SPSS / R

# SHETH L.U.J. AND SIR M.V. COLLEGE



The screenshot displays the RStudio environment with the following components:

- Source:** Contains the R script being executed.
- Console:** Shows the output of the R code, including a table of weather data and the results of the `tidy_transport` function.
- Environment:** Lists the objects in the environment, such as `region_filter`, `robot_related`, `selected_cols`, `sleep_df`, `special_buildings`, `split_matrix`, `starts_with_s`, `tidy_transport`, and `transport_df`.
- Files:** Shows the file explorer with various files and folders, including `Processed_Student_Mental_Health.csv`, `S093_CHETAN_MANDAVKAR_FEATURE_ENGINEERING_SUBMISSION - Colab`, `S093_Chetan_Mandavkar_FP_Soft_Copy_Documentation_1[1].docx`, `SHRUTI MANDAVKAR EVS REPORT.pdf`, `Structure and functions of pancreas Shruti Mandavkar.pptx`, `Virtual Machines`, `WhatsApp Image 2025-11-25 at 19:51:57_2b9447da.jpg`, `Attendance_Prediction.csv`, `product_sales_dataset_final.csv`, `Practical_No_6_RR`, `synthetic_fraud_dataset.csv`, `Practical_No_7_RR`, `dataset_2191_sleep.csv`, `Practical_No_8_RR`, `Synthetic_Transportation_Dataset_Expanded_v2.csv`, and `Practical_No_9_RR`.

```
R - R4.5.2 -- /  
2 09-06-2019 08:01 09 06 2019 08  
3 08-01-2020 01:38 08 01 2020 01  
4 18-10-2022 05:43 18 10 2022 05  
5 29-06-2023 23:11 29 06 2023 23  
6 02-03-2020 15:55 02 03 2020 15  
> split_matrix <- str_split(transport_df$Weather, " ", simplify = TRUE)  
> transport_df$Weather_Main <- split_matrix[, 1]  
> if (ncol(split_matrix) >= 2) {  
+   transport_df$Weather_Detail <- split_matrix[, 2]  
+ } else {  
+   transport_df$Weather_Detail <- NA  
+ }  
> print("---- After Safe Weather Split ----")  
[1] "---- After Safe Weather Split ----"  
> print(transport_df %>% select(Weather, Weather_Main, Weather_Detail) %>% head())  
# A tibble: 6 x 3  
  Weather Weather_Main Weather_Detail  
  <chr>    <chr>      <lg>  
1 Fog     Fog         NA  
2 Rain    Rain        NA  
3 Storm   Storm       NA  
4 Clear   Clear       NA  
5 Storm   Storm       NA  
6 Snow    Snow        NA  
> tidy_transport <- transport_df %>%  
+   separate(timestamp, into = c("Date", "Time"), sep = " ")  
> print("---- Using separate() to split Timestamp into Date & Time ----")  
[1] "---- Using separate() to split Timestamp into Date & Time ----"  
> print(tidy_transport %>% select(Date, Time) %>% head())  
# A tibble: 6 x 2  
  Date      Time  
  <chr>    <chr>  
1 03-10-2023 09:50  
2 09-06-2019 08:01  
3 08-01-2020 01:38  
4 18-10-2022 05:43  
5 29-06-2023 23:11  
6 02-03-2020 15:55  
>
```

NAME:- CHETAN MANDAVKAR

ROLL NO. S093

SUBJECT:- Data Analysis with SAS / SPSS / R