

## PRACTICAL NO 7

**AIM:- Selecting and dropping variables using select() in R.  
import dataset.**

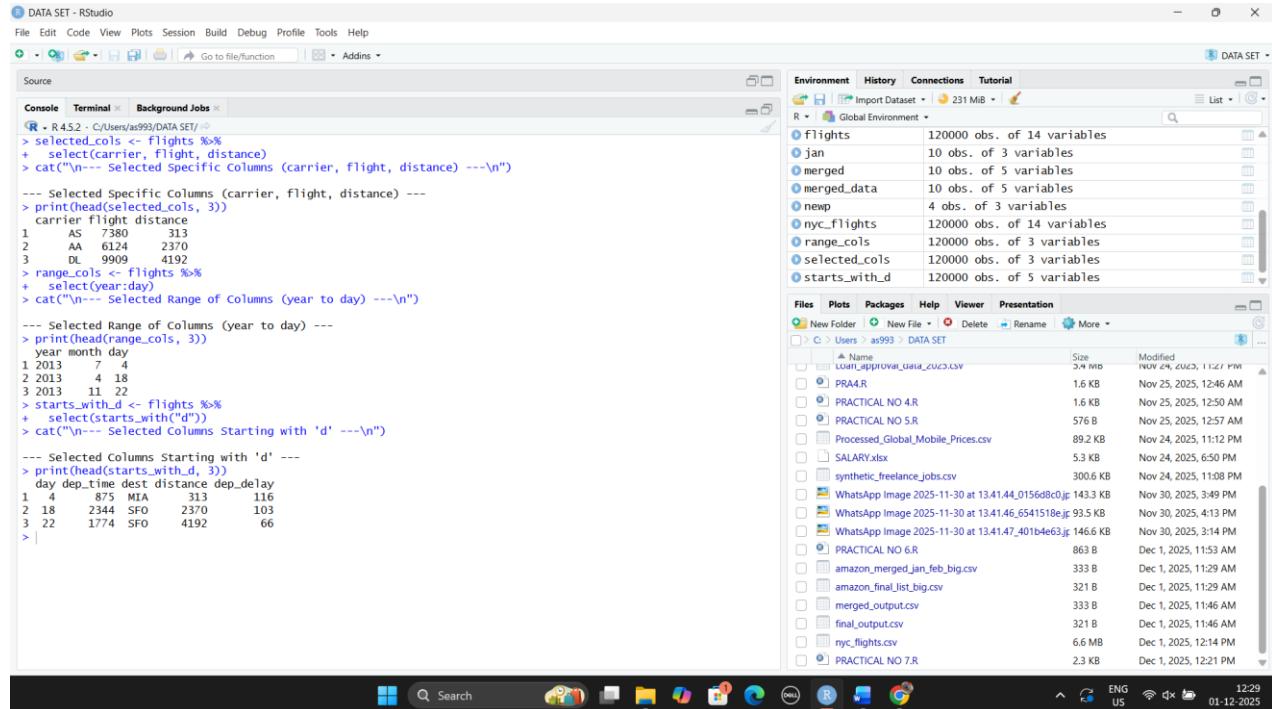
**OUTPUT:-**

### 1. IMPORT DATASET:

The screenshot shows the RStudio interface with the following details:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Code Editor:** A script named "PRACTICAL NO 7.R" containing R code to import the "nyc\_flights" dataset and select specific columns.
- Environment View:** Shows objects in the Global Environment, including "flights" (120000 obs. of 14 variables), "jan" (10 obs. of 3 variables), "merged" (10 obs. of 5 variables), "merged\_data" (10 obs. of 5 variables), "newp" (4 obs. of 3 variables), "nyc\_flights" (120000 obs. of 14 variables), "range\_cols" (120000 obs. of 3 variables), "selected\_cols" (120000 obs. of 3 variables), and "starts\_with\_d" (120000 obs. of 5 variables).
- File Explorer:** Shows files in the current directory, including "loan\_approval\_data\_clean.csv", "PRA4.R", "PRACTICAL NO 4.R", "Processed\_Global\_Mobile\_Prices.csv", "SALARY.xlsx", "synthetic\_freelance\_jobs.csv", "WhatsApp Image 2025-11-30 at 13:41:44\_0156d8c0.jpg", "WhatsApp Image 2025-11-30 at 13:41:46\_6541510e.jpg", "WhatsApp Image 2025-11-30 at 13:41:47\_401b4e63.jpg", "PRACTICAL NO 6.R", "amazon\_merged\_jan\_feb\_big.csv", "amazon\_final\_list\_big.csv", "merged\_output.csv", "final\_output.csv", "nyc\_flights.csv", and "PRACTICAL NO 7.R".
- Console:** Displays the R code execution results, including the head of the "flights" dataset.
- Terminal:** Displays the R version and working directory information.

## 2. SELECTING VARIABLES (Keeping Columns)



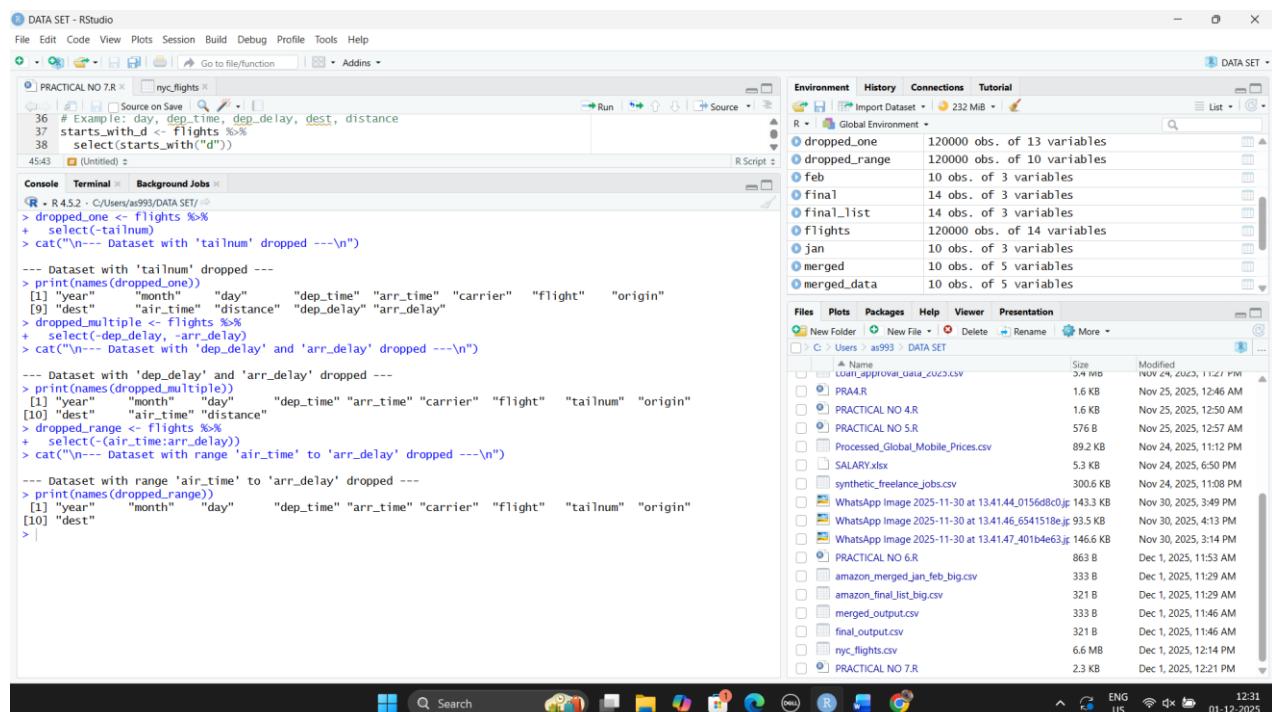
The screenshot shows the RStudio interface. The left pane displays R code for selecting specific columns from a dataset named 'flights'. The right pane shows the 'Environment' tab with several objects listed, including 'flights', 'jan', 'merged', 'merged\_data', 'newp', 'nyc\_flights', 'range\_cols', 'selected\_cols', and 'starts\_with\_d'. The bottom status bar indicates the date as 01-12-2025 and the time as 12:29.

```

DATA SET - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source Terminal Background Jobs
R - R 4.5.2 - C:/Users/as993/DATA SET
> selected_cols <- flights %>%
+   select(carrier, flight, distance)
> cat("\n--- Selected Specific Columns (carrier, flight, distance) ---\n")
--- Selected Specific Columns (carrier, flight, distance) ---
> print(head(selected_cols, 3))
  carrier flight distance
1       AS      7380     313
2       AA      6124    2370
3       DL      9909    4192
> range_cols <- flights %>%
+   select(year:day)
> cat("\n--- Selected Range of Columns (year to day) ---\n")
--- Selected Range of Columns (year to day) ---
> print(head(range_cols, 3))
  year month day
1 2013      7    4
2 2013      4   18
3 2013     11   22
> starts_with_d <- flights %>%
+   select(starts_with("d"))
> cat("\n--- Selected Columns Starting with 'd' ---\n")
--- Selected Columns Starting with 'd' ---
> print(head(starts_with_d, 3))
  day dep_time dest distance dep_delay
1   4      875  MIA      313     116
2  18     2344  SFO      2370     103
3  22     1774  SFO      4192      66
>

```

## 3. DROPPING VARIABLES (Removing Columns)



The screenshot shows the RStudio interface. The left pane displays R code for dropping specific columns from a dataset named 'nyc\_flights'. The right pane shows the 'Environment' tab with objects like 'dropped\_one', 'dropped\_range', 'Feb', 'final', 'final\_list', 'flights', 'jan', 'merged', and 'merged\_data'. The bottom status bar indicates the date as 01-12-2025 and the time as 12:31.

```

DATA SET - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source Terminal Background Jobs
R - R 4.5.2 - C:/Users/as993/DATA SET
36 # Example: day, dep_time, dep_delay, dest, distance
37 starts_with_d <- flights %>%
38   select(starts_with("d"))
4543 (Untitled).c

Console Terminal Background Jobs
R - R 4.5.2 - C:/Users/as993/DATA SET
> dropped_one <- flights %>%
+   select(-tailnum)
> cat("\n--- Dataset with 'tailnum' dropped ---\n")
--- Dataset with 'tailnum' dropped ---
> print(names(dropped_one))
[1] "year"      "month"     "day"       "dep_time"   "arr_time"   "carrier"    "flight"     "origin"
[9] "dest"      "air_time"   "distance"  "dep_delay"  "arr_delay"
> dropped_multiple <- flights %>%
+   select(-dep_delay, -arr_delay)
> cat("\n--- Dataset with 'dep_delay' and 'arr_delay' dropped ---\n")
--- Dataset with 'dep_delay' and 'arr_delay' dropped ---
> print(names(dropped_multiple))
[1] "year"      "month"     "day"       "dep_time"   "arr_time"   "carrier"    "flight"     "tailnum"   "origin"
[10] "dest"     "air_time"   "distance"
> dropped_range <- flights %>%
+   select(-(air_time:arr_delay))
> cat("\n--- Dataset with range 'air_time' to 'arr_delay' dropped ---\n")
--- Dataset with range 'air_time' to 'arr_delay' dropped ---
> print(names(dropped_range))
[1] "year"      "month"     "day"       "dep_time"   "arr_time"   "carrier"    "flight"     "tailnum"   "origin"
[10] "dest"
>

```