

## PRACTICAL NO 4

AIM:- Applying conditional filters subset() or filter() in R.

OUTPUT:-

The screenshot shows the RStudio interface. The console displays the following output:

```
R - R 4.5.2 - C:/Users/as993/DATA SET />
The following objects are masked from 'package:stats':
  filter, lag
The following objects are masked from 'package:base':
  intersect, setdiff, setequal, union

Warning message:
In file.copy(savedcopy, lib, recursive = TRUE) :
  problem copying C:/Users/as993/AppData/Local/R/win-library/4.5/00LOCK/dplyr/lib/x64/dplyr.dll to C:/Users/as993/AppData/Local/R/win-library/4.5/dplyr/lib/x64/dplyr.dll: Permission denied

> library(readr)
> loans <- read_csv("Loan_approval_data_2025.csv")
Rows: 50000 Columns: 20
Column specification
Delimiter: ","
chr (4): customer_id, occupation_status, product_type, loan_intent
dbl (16): age, years_employed, annual_income, credit_score, credit_history_years, savings_assets, curren...
```

The Environment pane on the right shows the loaded dataset:

Object	Size	Modified
home_edu_loans_f...	17587 obs. of 20 variables	
ILPD	10000 obs. of 7 variables	
in_set	10000 obs. of 7 variables	
large_approved_l...	6265 obs. of 20 variables	
Loan_approval_da...	50000 obs. of 20 variables	
loans	50000 obs. of 20 variables	
low_economic_loss	50000 obs. of 13 variables	
low_interest_fil...	4819 obs. of 20 variables	
male_or_low_albu...	7128 obs. of 7 variables	

The screenshot shows the RStudio interface with the following code in the script editor:

```
loans <- read_csv("Loan_approval_data_2025.csv")
head(loans)
high_income_subset <- subset(loans, annual_income > 80000)
cat("Number of high-income customers (annual income > 80000):", nrow(high_income_subset), "\n")
summary(high_income_subset$annual_income)
good_credit_low_dti_subset <- subset(loans,
  credit_score > 700 &
  debt_to_income_ratio < 0.3)
cat("Number of customers with good credit and low DTI:", nrow(good_credit_low_dti_subset), "\n")
head(good_credit_low_dti_subset)
```

The console output shows the results of these operations:

```
R - R 4.5.2 - C:/Users/as993/DATA SET />
> high_income_subset <- subset(loans, annual_income > 80000)
> cat("Number of high-income customers (annual income > 80000):", nrow(high_income_subset), "\n")
Number of high-income customers (annual income > 80000): 7181
> summary(high_income_subset$annual_income)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
80010. 89006. 102004. 112487. 125151. 250000
> good_credit_low_dti_subset <- subset(loans,
+   credit_score > 700 &
+   debt_to_income_ratio < 0.3)
> cat("Number of customers with good credit and low DTI:", nrow(good_credit_low_dti_subset), "\n")
Number of customers with good credit and low DTI: 5518
> head(good_credit_low_dti_subset)
# A tibble: 6 x 20
  customer_id age occupation_status years_employed annual_income credit_score credit_history_years
  <chr>      <dbl> <chr>              <dbl>          <dbl>      <dbl>      <dbl>
1 CUST100000 40 Employed             17.2        25379      692        5.3
2 CUST100001 33 Employed             7.3        43087      627        3.5
3 CUST100002 42 Student              1.1        20840      689        8.4
4 CUST100003 53 Student              0.5        29147      692        9.8
5 CUST100004 32 Employed            12.5        63657      630        7.2
6 CUST100005 32 Employed            13.4        32015      570        7.3
```

The Environment pane on the right shows the updated dataset list:

Object	Size	Modified
good_credit_low...	5518 obs. of 20 variables	
high_casualties	21933 obs. of 13 variables	
high_casualty_an...	25036 obs. of 13 variables	
high_income_sub...	7181 obs. of 20 variables	
home_edu_loans_f...	17587 obs. of 20 variables	
ILPD	10000 obs. of 7 variables	
in_set	10000 obs. of 7 variables	
large_approved_l...	6265 obs. of 20 variables	
Loan_approval_da...	50000 obs. of 20 variables	

# 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 the RStudio interface with the following components:

- Script Editor:** Contains R code for filtering loans. The code defines a subset of loans with a credit score greater than 700 and a debt-to-income ratio less than 0.3. It then filters this subset for customers with defaults on file or delinquencies in the last two years.
- Console:** Displays the output of the R code, showing a tibble with 6 rows and 7 columns: customer\_id, age, occupation\_status, years\_employed, annual\_income, credit\_score, and credit\_history\_years.
- Environment:** Lists the objects in the R environment, including 'Loan\_approval\_data\_2025', 'loans', 'low\_economic\_loss', 'low\_interest\_filter', 'male\_or\_low\_albu', 'older\_high\_bili', 'patients\_subset', 'patients\_tbl', and 'risky\_customers\_subset'.
- Files:** Shows a list of files in the project directory, including 'Rhistory', 'BC.txt', 'Breast\_Cancer.csv', 'DATA SET.Rproj', 'eeg-headset.csv', 'ILPD.csv', 'Loan\_approval\_data\_2025.csv', 'PRAAR', 'PRACTICAL NO 4.R', 'PRACTICAL NO 5.R', 'Processed\_Global\_Mobile\_Prices.csv', 'SALARY.xlsx', and 'synthetic\_freelance\_jobs.csv'.

The screenshot shows the RStudio interface with the following components:

- Script Editor:** Contains R code for filtering loans based on interest rate and loan amount. The code filters loans with an interest rate less than 10% and then filters the resulting subset for loans with a loan amount greater than 50,000.
- Console:** Displays the output of the R code, showing a tibble with 6 rows and 7 columns: customer\_id, age, occupation\_status, years\_employed, annual\_income, credit\_score, and credit\_history\_years.
- Environment:** Lists the objects in the R environment, including 'Loan\_approval\_data\_2025', 'loans', 'low\_economic\_loss', 'low\_interest\_filter', 'male\_or\_low\_albu', 'older\_high\_bili', 'patients\_subset', 'patients\_tbl', and 'risky\_customers\_subset'.
- Files:** Shows a list of files in the project directory, including 'Rhistory', 'BC.txt', 'Breast\_Cancer.csv', 'DATA SET.Rproj', 'eeg-headset.csv', 'ILPD.csv', 'Loan\_approval\_data\_2025.csv', 'PRAAR', 'PRACTICAL NO 4.R', 'PRACTICAL NO 5.R', 'Processed\_Global\_Mobile\_Prices.csv', 'SALARY.xlsx', and 'synthetic\_freelance\_jobs.csv'.

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

## Data Analysis with SAS / SPSS / R

The screenshot displays the RStudio environment with the following components:

- Source Editor:** Contains R code for filtering data based on interest rate and loan intent.
- Console:** Shows the execution of the code, including the creation of a filtered dataset and a summary table.
- Environment Pane:** Lists the objects in the global environment, including the filtered dataset.
- Files Pane:** Shows the file explorer for the current project.

**R Code (Source Editor):**

```
24  
25 cat("Number of risky customers (defaults or many delinquencies):", nrow(risky_customers_subset), "\n")  
26 head(risky_customers_subset)  
27  
28 low_interest_filter <- loans |>  
29   filter(interest_rate < 10)  
30  
31 cat("Number of low-interest loans (interest_rate < 10):", nrow(low_interest_filter), "\n")  
32 summary(low_interest_filter$interest_rate)  
33
```

**Console Output:**

```
R - R 4.5.2 - C:/Users/as993/DATA SET/ >  
> home_edu_loans_filter <- loans |>  
+   filter(loan_intent %in% c("Home Improvement", "Education"))  
> cat("Number of loans for Home Improvement or Education:", nrow(home_edu_loans_filter), "\n")  
Number of loans for Home Improvement or Education: 17587  
> table(home_edu_loans_filter$loan_intent)  
  
      Education Home Improvement  
          10134           7453  
> |
```

**Environment Pane:**

Object	Variables	Observations
home_edu_loans_f...	20	17587
ILPD	7	10000
in_set	7	10000
large_approved_l...	20	6265
Loan_approval_da...	20	50000
loans	20	50000
low_economic_loss	13	50000
low_interest_fil...	20	4819
male_or_low_albu...	7	7128

**Files Pane:**

Name	Size	Modified
.Rhistory	721 B	Nov 24, 2025, 7:53 PM
BC.txt	1.7 KB	Nov 24, 2025, 6:39 PM
Breast_Cancer.csv	386.8 KB	Nov 24, 2025, 4:25 PM
DATA SET.Rproj	218 B	Nov 24, 2025, 7:41 PM
eeg-headset.csv		
ILPD.csv	296.6 KB	Nov 24, 2025, 9:35 PM
Loan_approval_data_2025.csv	5.4 MB	Nov 24, 2025, 11:27 PM
PRA4.R	0 B	Nov 25, 2025, 12:03 AM
PRACTICAL NO 4.R	2.7 KB	Nov 24, 2025, 11:40 PM
PRACTICAL NO 5.R	0 B	Nov 24, 2025, 11:56 PM
Processed_Global_Mobile_Prices.csv	89.2 KB	Nov 24, 2025, 11:12 PM
SALARY.xlsx	5.3 KB	Nov 24, 2025, 6:50 PM
synthetic_freelance_jobs.csv	300.6 KB	Nov 24, 2025, 11:08 PM