

SHETH L.U.J & SIR M.V COLLEGE OF SCIENCE

SUBJECT : Data Analysis with SAS / SPSS / R

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

OUTPUT

The screenshot displays the RStudio interface with the following components:

- Console:** Shows the execution of R code to load the 'dplyr' and 'readr' packages, read the 'spotify_data.csv' file, and apply filters to create subsets of the data.
- Environment:** Lists the loaded packages and data objects, including 'album_or_single_filt...', 'explicit_and_famous...', 'high_popularity_subs...', 'long_single_filter', 'low_popularity_filter', 'special_tracks_subset', and 'spotify'.
- Files:** Shows the file explorer with the 'spotify_data.csv' file highlighted.

Console Output:

```
R - R4.5.2 - ~/R
> library(dplyr)

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':
  filter, lag

The following objects are masked from 'package:base':
  intersect, setdiff, setequal, union

> library(readr)
> spotify <- read_csv("spotify_data.csv")
Rows: 8582 Columns: 15
Column specification
Delimiter: ","
chr (8): track_id, track_name, artist_name, artist_genres, album_id, album_name, album_rele...
dbl (6): track_number, track_popularity, artist_popularity, artist_followers, album_total_t...
lg1 (1): explicit

i Use 'spec()' to retrieve the full column specification for this data.
i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
> head(spotify)
# A tibble: 6 x 15
  track_id track_name track_number track_popularity explicit artist_name artist_popularity
  <chr>    <chr>          <dbl>         <dbl> <lg1>    <chr>          <dbl>
1 3EJ5SLyekDm... Trippy Ma...      4             0 TRUE    Diplo           77
2 loQW6GZziWu... OMGI          1             0 TRUE    Yelawolf        64
3 7ndkjzoYlF1... Hard 2 Fi...    1             1 TRUE    Riff Raff       48
4 67rw0Zl7oB3q... Still Get...    8             30 TRUE    Diplo           77
5 15xptTFRBrjs... ride me l...    2             0 TRUE    Rumelis         48
6 4ccpCcZyseq8... BLEED         1             2 FALSE   Minzie          46

# 8 more variables: artist_followers <dbl>, artist_genres <chr>, album_id <chr>,
# album_name <chr>, album_release_date <chr>, album_total_tracks <dbl>, album_type <chr>,
# track_duration_min <dbl>
> high_popularity_subset <- subset(spotify, track_popularity > 80)
> cat("Number of highly popular tracks (popularity > 80):", nrow(high_popularity_subset), "\n")
Number of highly popular tracks (popularity > 80): 592
> summary(high_popularity_subset$track_popularity)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   81.00  82.00   83.00  84.22  86.00   99.00
> explicit_and_famous_subset <- subset(spotify, explicit == TRUE & artist_followers > 1000000)
> cat("Number of explicit tracks by artists with >1M followers:", nrow(explicit_and_famous_subset),
"\n")
Number of explicit tracks by artists with >1M followers: 1798
> head(explicit_and_famous_subset)
# A tibble: 6 x 15
  track_id track_name track_number track_popularity explicit artist_name artist_popularity
  <chr>    <chr>          <dbl>         <dbl> <lg1>    <chr>          <dbl>
1 3EJ5SLyekDm... Trippy Ma...      4             0 TRUE    Diplo           77
2 loQW6GZziWu... OMGI          1             0 TRUE    Yelawolf        64
3 67rw0Zl7oB3q... Still Get...    8             30 TRUE    Diplo           77
4 2Zq3Trov36G0... superstar    2             59 TRUE    Artemas         74
5 1NXbNEACPvY5... Dracula      3             80 TRUE    Impala          84
6 6GgRwmK8b8BV... PSYCHWARD... 1             51 TRUE    Diplo           77

# 8 more variables: artist_followers <dbl>, artist_genres <chr>, album_id <chr>,
# album_name <chr>, album_release_date <chr>, album_total_tracks <dbl>, album_type <chr>,
# track_duration_min <dbl>
> special_tracks_subset <- subset(spotify, grepl("dark r&b", tolower(artist_genres)) | track_durati
on_min > 4)
> cat("Number of special tracks (dark r&b OR duration > 4 mins):", nrow(special_tracks_subset),
"\n")
Number of special tracks (dark r&b OR duration > 4 mins): 2205
> head(special_tracks_subset)
# A tibble: 6 x 15
  track_id track_name track_number track_popularity explicit artist_name artist_popularity
  <chr>    <chr>          <dbl>         <dbl> <lg1>    <chr>          <dbl>
1 15xptTFRBrjs... ride me l...    2             0 TRUE    Rumelis         48
2 4ccpCcZyseq8... BLEED         1             2 FALSE   Minzie          46
3 3QoQ3HqXTA7g... Te Procur...    1             20 FALSE   AZERDK         30
4 1YEZbT4T75f... come clos...    1             27 FALSE   Rumelis         49
5 0NZKfCua68mu... Root of a...    5             65 FALSE   Daniel Cae...   85
6 66F6YyfmY6C... MELODY - ... 11             27 FALSE   A2O            46
```

SHETH L.U.J & SIR M.V COLLEGE OF SCIENCE

SUBJECT : Data Analysis with SAS / SPSS /R

The screenshot displays the RStudio environment with the following components:

- Console:** Shows R code execution for filtering data based on popularity and album type. The code includes comments in Hindi and English, and the output shows the number of rows in the filtered datasets.
- Environment:** Lists the objects created in the environment, including 'album_or_single_filt...', 'explicit_and_famous...', 'high_popularity_subs...', 'long_single_filter', 'low_popularity_filter', 'special_tracks_subset', and 'spotify'.
- Files:** Shows the file explorer with various files and folders, including 'RData', 'Rhistory', 'Custom Office Templates', 'Dell', 'desktop.ini', 'My Music', 'My Videos', 'NetBeansProjects', 'Shortcut to Documents (OneDrive - Personal)', and 'spotify_data.csv'.

The code in the console is as follows:

```
R - R 4.5.2 - ~/R
# i 8 more variables: artist_followers <dbl>, artist_genres <chr>, album_id <chr>,
# album_name <chr>, album_release_date <chr>, album_total_tracks <dbl>, album_type <chr>,
# track_duration_min <dbl>
> low_popularity_filter <- spotify |>
+ filter(track_popularity < 30)
> cat("Number of low-popularity tracks (< 30):", nrow(low_popularity_filter), "\n")
Number of low-popularity tracks (< 30): 1481
> summary(low_popularity_filter$track_popularity)
Min. 1st Qu. Median Mean 3rd Qu. Max.
0.00 0.00 4.00 10.47 22.00 29.00
> long_single_filter <- spotify |>
+ filter(album_type == "single", track_duration_min > 3.5)
> cat("Number of singles longer than 3.5 minutes:", nrow(long_single_filter), "\n")
Number of singles longer than 3.5 minutes: 607
> head(long_single_filter)
# A tibble: 6 x 15
  track_id track_name track_number track_popularity explicit artist_name artist_popularity
  <chr> <chr> <dbl> <dbl> <lg1> <chr> <dbl>
1 3QOQ3HqXTA9g Te Procur... 1 20 FALSE AZERDK 30
2 2zzjpwG8wu3s Eggsshell ... 1 52 FALSE The All-Am... 67
3 2MbsC601xeyJ Eyes on Y... 1 22 FALSE Smith & Li... 17
4 0C42Yia3Qy90 Hold On C... 1 34 FALSE Suark 32
5 1yLkAST7ip9 Kay 2 FALSE Vince Bello 6
6 00riikd4ETp5 Talk To Y... 1 72 TRUE Sam Fender 73
# i 8 more variables: artist_followers <dbl>, artist_genres <chr>, album_id <chr>,
# album_name <chr>, album_release_date <chr>, album_total_tracks <dbl>, album_type <chr>,
# track_duration_min <dbl>
> album_or_single_filter <- spotify |>
+ filter(album_type %in% c("album", "single"))
> cat("Number of rows with album_type album or single:", nrow(album_or_single_filter), "\n")
Number of rows with album_type album or single: 8075
> table(album_or_single_filter$album_type)
album single
5856 2219
>
```

