

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

SUBJECT : Data Analysis with SAS / SPSS /R

AIM : Applying basic data cleaning functions: handling missing values using `na.omit()/replace_na()` in R. import dataset.

OUTPUT

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source Terminal Background Jobs
Console R 4.5.1 - ~/~>
> video_df <- read.csv("video_data.csv", na.strings = c("", "NA"))
> print("--- 1. Original Data ---")
[1] "--- 1. Original Data ---"
> print(head(video_df))
  video_id      title duration_sec hashtags_count   views likes comments shares upload_hour category
1 vid_1000 Short Video #0       43        9 198775 21933    3228   400        8     Tech
2 vid_1001 Short Video #1       56        2 290336 20063    3719   1942       16 Comedy
3 vid_1002 Short Video #2       33        6 264206 37032    3228   1817        7 Food
4 vid_1003 Short Video #3       19        9 85076 27269    2371   980        1 Lifestyle
5 vid_1004 Short Video #4       47        8 90780 8041    2891   1109       23 Tech
6 vid_1005 Short Video #5       12        3 153617 14488    3756   249        12 Travel
> print("--- Count of Missing Values per Column ---")
[1] "--- Count of Missing Values per Column ---"
> print(colsSums(is.na(video_df)))
  video_id      title duration_sec hashtags_count   views likes comments shares upload_hour category
0          0          0            0            0        0    0        0    0        0        0        0
> clean.omit <- na.omit(video_df)
> print("--- 2. Data after na.omit() ---")
[1] "--- 2. Data after na.omit() ---"
> print(paste("Original rows:", nrow(video_df)))
[1] "Original rows: 300"
> print(paste("Rows remaining:", nrow(clean.omit)))
[1] "Rows remaining: 300"
> print(head(clean.omit))
  video_id      title duration_sec hashtags_count   views likes comments shares upload_hour category
1 vid_1000 Short Video #0       43        9 198775 21933    3228   400        8     Tech
2 vid_1001 Short Video #1       56        2 290336 20063    3719   1942       16 Comedy
3 vid_1002 Short Video #2       33        6 264206 37032    3228   1817        7 Food
4 vid_1003 Short Video #3       19        9 85076 27269    2371   980        1 Lifestyle
5 vid_1004 Short Video #4       47        8 90780 8041    2891   1109       23 Tech
6 vid_1005 Short Video #5       12        3 153617 14488    3756   249        12 Travel
> avg.duration <- mean(video_df$duration_sec, na.rm = TRUE)
> avg.views <- mean(video_df$views, na.rm = TRUE)
> clean.replace <- video_df %>%
+   replace_na(list(
+     category = "Unknown",
+     likes = 0,
+     shares = 0,
+     duration_sec = avg.duration,
+     views = avg.views
+   ))
> print("--- 3. Data after replace_na() ---")
[1] "--- 3. Data after replace_na() ---"
> print(clean.replace[3,])
  video_id      title duration_sec hashtags_count   views likes comments shares upload_hour category
3 vid_1002 Short Video #2       33        6 264206 37032    3228   1817        7 Food
> print(head(clean.replace))
  video_id      title duration_sec hashtags_count   views likes comments shares upload_hour category
1 vid_1000 Short Video #0       43        9 198775 21933    3228   400        8     Tech
2 vid_1001 Short Video #1       56        2 290336 20063    3719   1942       16 Comedy
3 vid_1002 Short Video #2       33        6 264206 37032    3228   1817        7 Food
4 vid_1003 Short Video #3       19        9 85076 27269    2371   980        1 Lifestyle
5 vid_1004 Short Video #4       47        8 90780 8041    2891   1109       23 Tech
6 vid_1005 Short Video #5       12        3 153617 14488    3756   249        12 Travel
> print("--- Remaining NAs after replacement ---")
[1] "--- Remaining NAs after replacement ---"
> print(colsSums(is.na(clean.replace)))
  video_id      title duration_sec hashtags_count   views likes comments shares upload_hour category
0          0          0            0            0        0    0        0    0        0        0        0

```

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

SUBJECT : Data Analysis with SAS / SPSS /R

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

S108-R PROGRAM PRAC8.R x clean OMIT x

video_id title duration_sec hashtags_count views likes comments shares upload_hour category

1	vid_1000	Short Video #0	43	9	198775	21933	3228	400	8	Tech
2	vid_1001	Short Video #1	56	2	290336	20063	3719	1942	16	Comedy
3	vid_1002	Short Video #2	33	6	264206	37032	3228	1817	7	Food
4	vid_1003	Short Video #3	19	9	85076	27269	2371	980	1	Lifestyle
5	vid_1004	Short Video #4	47	8	90780	8041	2891	1109	23	Tech
6	vid_1005	Short Video #5	12	3	153617	14488	3756	249	12	Travel
7	vid_1006	Short Video #6	25	0	22689	5669	4225	390	2	Tech
8	vid_1007	Short Video #7	43	1	274318	4400	94	1805	15	Tech
9	vid_1008	Short Video #8	23	0	351605	39137	2674	334	5	Comedy
10	vid_1009	Short Video #9	27	4	389318	36686	1955	628	18	Food
11	vid_1010	Short Video #10	15	4	366177	12115	793	1542	6	Travel
12	vid_1011	Short Video #11	15	6	10348	45102	1834	136	9	Education
13	vid_1012	Short Video #12	28	8	417930	6354	2116	882	8	Education
14	vid_1013	Short Video #13	57	8	39102	17244	4106	687	1	Tech
15	vid_1014	Short Video #14	40	2	75460	24314	3325	1863	14	Comedy
16	vid_1015	Short Video #15	44	2	365778	40218	2919	1521	4	Comedy
17	vid_1016	Short Video #16	28	2	90930	15283	2597	826	21	Education
18	vid_1017	Short Video #17	7	3	7801	6338	3205	726	3	Travel

Showing 1 to 19 of 300 entries. 10 total columns

Console Terminal Background Jobs

```
R - R 4.5.1 . ~/ 
> view(clean OMIT)
>
```

27°C Sunny

Search

Files Plots Packages Help Viewer Presentation

Environment History Connections Tutorial

R Global Environment

clean OMIT 300 obs. of 10 variables

clean REPL 300 obs. of 10 variables

video DF 300 obs. of 10 variables

Values avg_duri... 31.73

avg views 245058.166666667

Files Plots Packages Help Viewer Presentation

Home Name Size Modified

- .RData 49 B Nov 2
- .Rhistory 131 B Dec 1
- Custom Office Templates
- GIS DataBase
- IISExpress
- My Web Sites
- NetBeansProjects
- Power BI Desktop
- Virtual Machines
- Visual Studio 2022
- final_data.csv 4.5 MB Dec 1
- video_data.csv 18.6 KB Dec 1

11:45 01-12-2025 ENG IN

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

S108-R PROGRAM PRAC8.R x clean OMIT x clean_replace x

video_id title duration_sec hashtags_count views likes comments shares upload_hour category

1	vid_1000	Short Video #0	43	9	198775	21933	3228	400	8	Tech
2	vid_1001	Short Video #1	56	2	290336	20063	3719	1942	16	Comedy
3	vid_1002	Short Video #2	33	6	264206	37032	3228	1817	7	Food
4	vid_1003	Short Video #3	19	9	85076	27269	2371	980	1	Lifestyle
5	vid_1004	Short Video #4	47	8	90780	8041	2891	1109	23	Tech
6	vid_1005	Short Video #5	12	3	153617	14488	3756	249	12	Travel
7	vid_1006	Short Video #6	25	0	22689	5669	4225	390	2	Tech
8	vid_1007	Short Video #7	43	1	274318	4400	94	1805	15	Tech
9	vid_1008	Short Video #8	23	0	351605	39137	2674	334	5	Comedy
10	vid_1009	Short Video #9	27	4	389318	36686	1955	628	18	Food
11	vid_1010	Short Video #10	15	4	366177	12115	793	1542	6	Travel
12	vid_1011	Short Video #11	15	6	10348	45102	1834	136	9	Education
13	vid_1012	Short Video #12	28	8	417930	6354	2116	882	8	Education
14	vid_1013	Short Video #13	57	8	39102	17244	4106	687	1	Tech
15	vid_1014	Short Video #14	40	2	75460	24314	3325	1863	14	Comedy
16	vid_1015	Short Video #15	44	2	365778	40218	2919	1521	4	Comedy
17	vid_1016	Short Video #16	28	2	90930	15283	2597	826	21	Education
18	vid_1017	Short Video #17	7	3	7801	6338	3205	726	3	Travel

Showing 1 to 19 of 300 entries. 10 total columns

Console Terminal Background Jobs

```
R - R 4.5.1 . ~/ 
> view(clean_replace)
>
```

27°C Sunny

Search

Files Plots Packages Help Viewer Presentation

Environment History Connections Tutorial

R Global Environment

clean OMIT 300 obs. of 10 variables

clean REPL 300 obs. of 10 variables

video DF 300 obs. of 10 variables

Values avg_duri... 31.73

avg views 245058.166666667

Snipping Tool

Screenshot copied to clipboard
Automatically saved to screenshots folder.

Mark-up and share

Files Plots Packages Help Viewer Presentation

Home Name Size Modified

- .RData 49 B Nov 2
- .Rhistory 131 B Dec 1
- Custom Office Templates
- GIS DataBase
- IISExpress
- My Web Sites
- NetBeansProjects
- Power BI Desktop
- Virtual Machines
- Visual Studio 2022
- final_data.csv 4.5 MB Dec 1
- video_data.csv 18.6 KB Dec 1

11:45 01-12-2025 ENG IN

**SHETH L.U.J & SIR M.V COLLEGE OF SCIENCE
SUBJECT : Data Analysis with SAS / SPSS /R**

The screenshot shows an RStudio interface with the following components:

- Top Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Project Bar:** Go to file/function, Addins.
- Data View:** A data grid titled "S108-R PROGRAM PRAC8.R" showing 18 rows of video data. The columns are: video_id, title, duration_sec, hashtags_count, views, likes, comments, shares, upload_hour, and category. The data includes various video titles like "Short Video #0" through "#17", their durations, view counts, and categories such as Tech, Comedy, Food, Lifestyle, Travel, Education, and more.
- Environment View:** Shows the global environment with objects: clean OMIT, clean REPLACE, and video DF. It also displays summary statistics: avg_duration_... (31.73) and avg_views (245058.166666667).
- Console View:** Displays R code and its output. The code includes `R > R-4.5.1 - ~/` and `> view(video_01)`.
- Bottom Status Bar:** Shows system icons for battery, signal, and time (01-12-2023), along with language (ENG IN) and font size (11.48).

SHUBHAM RAORANE S108