

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

SUBJECT : Data Analysis with SAS / SPSS / R

AIM : Identifying and handling duplicates using distinct() (R).

OUTPUT

The first screenshot shows the initial dataset 'nirf' with 100 observations. The code identifies duplicates based on Institute.ID, Institute.Name, State, Score, and Rank. The output shows 2 duplicate rows (Exact Matches Found).

```
> library(dplyr)
> nirf <- read.csv("NIRF Ranking 2016.csv", na.strings = c("", "NA"))
> print("--- 1. Original Dataset (Preview) ---")
[1] "--- 1. Original Dataset (Preview) ---"
> print(head(nirf))
  Institute.ID Institute.Name TLR RPC GO OI PERCEPTION City
1 NIRF-ENG-77 Indian Institute Of Technology, Madras 88.26 94.02 81.81 86.11 98 Chennai
2 NIRF-ENG-312 Indian Institute Of Technology, Bombay 85.93 94.14 84.97 74.84 99 Bombay
3 NIRF-ENG-300 Indian Institute Of Technology, Kharagpur 76.23 92.68 83.95 78.05 97 Kharagpur
4 NIRF-ENG-79 Indian Institute Of Technology, Delhi 80.27 91.62 74.72 66.17 98 New Delhi
5 NIRF-ENG-228 Indian Institute Of Technology, Kanpur 66.08 93.52 85.62 70.59 98 Kanpur
6 NIRF-ENG-78 Indian Institute Of Technology, Roorkee 67.83 83.50 81.61 76.59 96 Roorkee

  State Score Rank
1 Tamil Nadu 89.41 1
2 Maharashtra 87.66 2
3 West Bengal 83.91 3
4 Delhi 82.02 4
5 Uttar Pradesh 81.07 5
6 Uttarakhand 78.68 6

> duplicates_report <- nirf %>%
+ group_by(Institute.ID, Institute.Name, State, Score, Rank) %>%
+ count() %>%
+ filter(n > 1)
> print("--- 2. Duplicate Rows (Exact Matches Found) ---")
[1] "--- 2. Duplicate Rows (Exact Matches Found) ---"
> print(duplicates_report)
# A tibble: 0 x 6
# Groups:   Institute.ID, Institute.Name, State, Score, Rank [0]
#   Institute.ID <chr>, Institute.Name <chr>, State <chr>, Score <dbl>, Rank <chr>, n <int>
1
> nirf_clean_exact <- nirf %>%
+ distinct()
> print("--- 3. Dataset After Removing Exact Duplicates ---")
```

The second screenshot shows the dataset after removing exact duplicates. The code prints the first 27 rows of the cleaned dataset.

```
> nirf_clean_exact <- nirf %>%
+ distinct()
> print("--- 3. Dataset After Removing Exact Duplicates ---")
[1] "--- 3. Dataset After Removing Exact Duplicates ---"
> print(nirf_clean_exact)
  Institute.ID
1 NIRF-ENG-77
2 NIRF-ENG-312
3 NIRF-ENG-300
4 NIRF-ENG-79
5 NIRF-ENG-228
6 NIRF-ENG-78
7 NIRF-ENG-345
8 NIRF-ENG-298
9 NIRF-ENG-340
10 NIRF-ENG-344
11 NIRF-ENG-299
12 NIRF-ENG-370
13 NIRF-ENG-768
14 NIRF-ENG-423
15 NIRF-ENG-373
16 NIRF-ENG-346
17 NIRF-ENG-1-2451694143
18 NIRF-ENG-225
19 NIRF-ENG-140
20 NIRF-ENG-1-NEW11
21 NIRF-ENG-338
22 NIRF-ENG-1-2453839794
23 NIRF-ENG-1-245363
24 NIRF-ENG-1-245357
25 NIRF-ENG-1-2451585489
26 NIRF-ENG-301
27 NIRF-ENG-296
```

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

SUBJECT : Data Analysis with SAS / SPSS / R

The image displays two screenshots of the RStudio environment, showing the execution of R code and the resulting data analysis output.

Top Screenshot: The console shows the execution of R code for data analysis. The code includes a list of 10 Indian Institute of Technology (IIT) locations and their corresponding NIRF rankings. The output shows the NIRF ranking for each IIT location.

```

R 4.5.1 ~ /
80 NIRF-ENGG-1-2451417043
81 NIRF-ENGG-INF-410
82 NIRF-ENGG-1-2452796944
83 NIRF-ENGG-1-2455130573
84 NIRF-ENGG-1-2453523111
85 NIRF-ENGG-1-2450743734
86 NIRF-ENGG-1-2454009090
87 NIRF-ENGG-INF-367
88 NIRF-ENGG-773
89 NIRF-ENGG-1-2453802770
90 NIRF-ENGG-1-2454313873

e. Name
1 Indian Institute Of Technology, M
2 Indian Institute Of Technology, B
3 Indian Institute Of Technology, Khar
4 Indian Institute Of Technology, K
5 Indian Institute Of Technology, K
6 Indian Institute Of Technology, Ro
7 Indian Institute Of Technology, Hyde
8 Indian Institute Of Technology, Gandhi
9 Indian Institute Of Technology, Ropar-Rup
10 Indian Institute Of Technology, Patna
  
```

Bottom Screenshot: The console shows the execution of R code for data analysis. The code includes a list of 10 Indian Institute of Technology (IIT) locations and their corresponding NIRF rankings. The output shows the NIRF ranking for each IIT location.

```

R 4.5.1 ~ /
anded
90
atore
TLR RPC GO OI PERCEPTION
1 88.26 94.02 81.81 86.11 98
2 85.93 94.14 84.97 74.84 99
3 76.23 92.68 83.95 78.05 97
4 80.27 91.62 74.72 66.17 98
5 66.08 93.52 85.62 70.59 98
6 67.83 83.50 81.61 76.59 96
7 76.76 76.83 80.89 63.76 95
8 80.36 72.09 64.91 71.94 96
9 89.96 73.56 86.42 60.67 38
10 79.80 68.29 74.13 68.78 93
11 71.92 78.66 71.82 65.62 94
12 56.99 80.37 81.50 79.50 92
13 66.95 74.72 77.38 77.11 91
14 59.08 80.63 96.80 56.73 96
15 59.02 85.24 78.43 68.81 83
16 86.75 75.78 89.85 39.23 46
17 55.39 81.06 82.47 65.50 90
18 59.98 73.64 69.05 82.70 85
19 60.78 75.69 81.49 73.21 74
20 57.76 83.07 76.07 59.50 87
21 85.83 70.35 81.05 47.29 43
22 58.61 66.40 82.92 74.67 95
23 60.26 67.86 77.29 70.22 88
24 54.63 72.07 74.08 75.04 82
25 56.72 64.67 77.30 72.10 94
26 73.02 66.41 60.71 60.11 86
27 67.40 75.73 61.06 51.57 85
28 49.43 78.70 77.69 64.58 83
29 58.20 71.82 75.41 61.20 82
  
```

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:

- Source Panel:** Contains R code for data manipulation, including a data frame with columns for Institute.ID, Institute.Name, and various numerical variables.
- Console Panel:** Shows the execution of the code, including a message indicating that 10 rows were omitted due to reaching the 'max' option.
- Environment Panel:** Lists the objects in the global environment, including 'nirf', 'nirf_cle', and 'nirf_uni', each with 100 observations and 11 variables.
- Files Panel:** Displays the file explorer, showing the current directory and its contents, including a file named 'NIRF Ranking 2016.csv'.

The R code in the Source panel is as follows:

```
R 4.5.1 ~ /  
87 46.73 52.37 66.90 55.69 37 Patna Bihar 51.08 87  
88 49.21 50.65 72.09 67.19 0 Guntur Andhra Pradesh 50.65 88  
89 48.74 47.61 67.42 51.75 41 Nanded Maharashtra 50.64 89  
90 46.62 56.15 67.80 65.62 6 Coimbatore Tamil Nadu 50.55 90  
[ reached 'max' / getOption("max.print") -- omitted 10 rows ]  
> nirf_unique_institutes <- nirf %>%  
+ distinct(Institute.Name, .keep_all = TRUE)  
> print("--- 4. Unique Institutes Only (1 entry per college) ---")  
[1] "--- 4. Unique Institutes Only (1 entry per college) ---"  
> print(nirf_unique_institutes)  
Institute.ID  
1 NIRF-ENGG-INF-77  
2 NIRF-ENGG-INF-312  
3 NIRF-ENGG-INF-300  
4 NIRF-ENGG-INF-79  
5 NIRF-ENGG-INF-228  
6 NIRF-ENGG-INF-78  
7 NIRF-ENGG-INF-345  
8 NIRF-ENGG-INF-298  
9 NIRF-ENGG-INF-340  
10 NIRF-ENGG-INF-344  
11 NIRF-ENGG-INF-299  
12 NIRF-ENGG-INF-370  
13 NIRF-ENGG-768  
14 NIRF-ENGG-INF-423  
15 NIRF-ENGG-INF-373  
16 NIRF-ENGG-INF-346  
17 NIRF-ENGG-1-2451694143  
18 NIRF-ENGG-INF-225  
19 NIRF-ENGG-INF-140  
20 NIRF-ENGG-INF-NEW11  
21 NIRF-ENGG-INF-338  
22 NIRF-ENGG-1-2453839794
```

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

SUBJECT : Data Analysis with SAS / SPSS / R

The top screenshot shows the following data:

City	State	Score	Rank
Chennai	Tamil Nadu	89.41	1
Bombay	Maharashtra	87.66	2
Kharagpur	West Bengal	83.91	3
New Delhi	Delhi	82.02	4
Kanpur	Uttar Pradesh	81.07	5
Roorkee	Uttarakhand	78.68	6
Hyderabad	Telangana	77.22	7
Ahmedabad	Gujarat	75.20	8
Rupnagar	Punjab	74.88	9
Patna	Bihar	74.62	10
Guwahati	Assam	74.62	11
Tiruchirappalli	Tamil Nadu	74.45	12
Vellore	Tamil Nadu	74.40	13
Varanasi	Uttar Pradesh	74.39	14
Surat	Gujarat	73.13	15
Indore	Madhya Pradesh	72.00	16
RANCHI	Jharkhand	71.80	17
Nagpur	Maharashtra	71.29	18
Rourkela	Odisha	70.80	19
Howrah	West Bengal	70.35	20A
Mandi	Himachal Pradesh	70.32	20
PUNE	Maharashtra	69.71	21
Mangalore	Karnataka	68.95	22
Allahabad	Uttar Pradesh	67.94	23
COIMBATORE	Tamil Nadu	67.80	24
Jodhpur	Rajasthan	67.68	25
Bhubaneswar	Odisha	67.58	26
Bhubaneswar	Odisha	67.51	27

The bottom screenshot shows the following data:

City	State	Score	Rank
Noida	Uttar Pradesh	56.79	60
PUNE	Maharashtra	56.56	61
SATHYAMANGALAM	Tamil Nadu	56.21	62
Raipur	Chhattisgarh	55.93	63
PUNE	Maharashtra	55.70	64
Silchar	Assam	55.49	65
Kanyakumari	Tamil Nadu	55.45	66
Kashmir	Jammu and Kashmir	55.30	67
JALGAON	Maharashtra	54.91	68
BERHAMPUR	Odisha	54.81	69
TUMKUR	Karnataka	54.67	70
HYDERABAD	Telangana	54.66	71
Chennai	Tamil Nadu	54.32	72
BHIMAVARAM	Andhra Pradesh	54.02	73
BHUBANESWAR	Odisha	53.91	74
ISLAMPUR	Maharashtra	53.69	75
Ponda	Goa	53.65	76
Jabalpur	Madhya Pradesh	53.37	77
Jamshedpur	Jharkhand	53.35	78
KOLKATA	West Bengal	53.26	79
Chennai	Tamil Nadu	52.64	80
JATNI	Odisha	52.58	81
UDAIPUR	Rajasthan	52.39	82
MUMBAI	Maharashtra	52.17	84
NASHIK	Maharashtra	51.92	85
NAGPUR	Maharashtra	51.50	86
Patna	Bihar	51.08	87
Guntur	Andhra Pradesh	50.65	88
NANDED	Maharashtra	50.64	89
COIMBATORE	Tamil Nadu	50.55	90

The bottom screenshot also shows a message: "reached 'max' / getoption("max.print") -- omitted 10 rows]".

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

SUBJECT : Data Analysis with SAS / SPSS / R

The screenshot displays the RStudio environment. The main editor shows a data frame with 17 rows and 8 columns. The columns are Institute.ID, Institute.Name, TLR, RPC, GO, OI, PERCEPTION, and City. The data lists various Indian Institutes of Technology (IITs) and their corresponding scores. The Environment pane on the right shows the Global Environment with variables like duplicat..., nirf, nirf_cle..., and nirf_uni... listed. The Console at the bottom shows the output of an R command, including the R version (4.5.1) and the results of a data frame operation.

Institute.ID	Institute.Name	TLR	RPC	GO	OI	PERCEPTION	City	
1	NIRF-ENGG-INF-77	Indian Institute Of Technology, Madras	88.26	94.02	81.81	86.11	98	Chennai
2	NIRF-ENGG-INF-312	Indian Institute Of Technology, Bombay	85.93	94.14	84.97	74.84	99	Bombay
3	NIRF-ENGG-INF-300	Indian Institute Of Technology, Kharagpur	76.23	92.68	83.95	78.05	97	Kharagpur
4	NIRF-ENGG-INF-79	Indian Institute Of Technology, Delhi	80.27	91.62	74.72	66.17	98	New Delhi
5	NIRF-ENGG-INF-228	Indian Institute Of Technology, Kanpur	66.08	93.52	85.62	70.59	98	Kanpur
6	NIRF-ENGG-INF-78	Indian Institute Of Technology, Roorkee	67.83	83.50	81.61	76.59	96	Roorkee
7	NIRF-ENGG-INF-345	Indian Institute Of Technology, Hyderabad	76.76	76.83	80.89	63.76	95	Hyderabad
8	NIRF-ENGG-INF-298	Indian Institute Of Technology, Gandhinagar	80.36	72.09	64.91	71.94	96	Ahmedabad
9	NIRF-ENGG-INF-340	Indian Institute Of Technology, Ropar-Rupnagar	89.96	73.56	86.42	60.67	38	Rupnagar
10	NIRF-ENGG-INF-344	Indian Institute Of Technology, Patna	79.80	68.29	74.13	68.78	93	Patna
11	NIRF-ENGG-INF-299	Indian Institute Of Technology, North Guwahati	71.92	78.66	71.82	65.62	94	Guwahati
12	NIRF-ENGG-INF-370	National Institute Of Technology, Tiruchirappalli	56.99	80.37	81.50	79.50	92	Tiruchirappalli
13	NIRF-ENGG-768	Vellore Institute Of Technology	66.95	74.72	77.38	77.11	91	Vellore
14	NIRF-ENGG-INF-423	Indian Institute Of Technology (Banaras Hindu University), V...	59.08	80.63	96.80	56.73	96	Varanasi
15	NIRF-ENGG-INF-373	Sardar Vallabhbhai National Institute Of Technology	59.02	85.24	78.43	68.81	83	Surat
16	NIRF-ENGG-INF-346	Indian Institute Of Technology, Indore	86.75	75.78	89.85	39.23	46	Indore
17	NIRF-ENGG-1-2451694143	Birla Institute Of Technology	55.39	81.06	82.47	65.50	90	RANCHI

Showing 1 to 17 of 100 entries, 11 total columns

Console Output:

```
R - R 4.5.1 - ~/...  
90 46.62 56.15 67.80 65.62 6  
[ reached 'max' /getOption("max.print") -- omitted 10 rows ]
```