

How can I use SAS to filter for rows that contain a specific string?

Authored by
stats writer

July 1, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I use SAS to filter for rows that contain a specific string?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=164954>

SAS (Statistical Analysis System) is a powerful software used for data analysis and management. One of its useful functions is the ability to filter data based on specific criteria. This allows users to narrow down their data to a specific subset that is relevant to their analysis. In order to filter for rows that contain a specific string, users can utilize the "WHERE" statement in the DATA step or the "WHERE" clause in PROC SQL. These commands allow users to specify the string they want to filter for, and SAS will only display the rows that contain that string. This feature can be particularly helpful for finding specific data points or patterns within a large dataset. Overall, SAS offers a convenient and efficient way to filter data and extract relevant information for further analysis.

SAS: Filter for Rows that Contain String

You can use the following methods to filter SAS datasets for rows that contain certain strings:

Method 1: Filter Rows that Contain Specific String

```
/*filter rows where var1 contains "string1"*/  
data specific_data;  
set original_data;  
where var1 contains 'string1';  
run;
```

Method 2: Filter Row that Contain One of Several Strings

```
/*filter rows where var1 contains "string1", "string2", or  
"string3"*/
```

```
data specific_data;  
set original_data;  
where var1 in ('string1', 'string2', 'string3');  
run;
```

The following examples show how to use each method with the following dataset in SAS:

```
/*create dataset*/  
data nba_data;  
input team $ points;  
datalines;  
Mavs 95  
Spurs 99  
Warriors 104  
Rockets 98  
Heat 95  
Nets 90  
Magic 99  
Cavs 106  
;  
run;
```

```
/*view dataset*/
```

```
proc printdata=nba_data;
```

| Obs | team | points |
|-----|----------|--------|
| 1 | Mavs | 95 |
| 2 | Spurs | 99 |
| 3 | Warriors | 104 |
| 4 | Rockets | 98 |
| 5 | Heat | 95 |
| 6 | Nets | 90 |
| 7 | Magic | 99 |
| 8 | Cavs | 106 |

Method 1: Filter Rows that Contain Specific String

The following code shows how to filter the dataset for rows that contain the string "avs" in the team column:

```
/*filter rows where team contains the string 'avs'*/  
data specific_data;  
set nba_data;  
where team contains 'avs';  
run;  
  
/*view resulting rows*/proc printdata=specific_data;
```

| Obs | team | points |
|-----|------|--------|
| 1 | Mavs | 95 |
| 2 | Cavs | 106 |

The only two rows shown are the ones where the team column contains 'avs' in the name.

Method 2: Filter Rows that Contain One of Several Strings

The following code shows how to filter the dataset for rows that contain the strings "Mavs", "Nets", or "Rockets" in the team column:

```
/*filter rows where team contains the string 'Mavs',  
'Nets', or 'Rockets'*/  
data specific_data;  
set nba_data;  
where team in ('Mavs', 'Nets', 'Rockets');  
run;
```

```
/*view resulting rows*/proc printdata=specific_data;
```

| Obs | team | points |
|-----|---------|--------|
| 1 | Mavs | 95 |
| 2 | Rockets | 98 |
| 3 | Nets | 90 |

Additional Resources

The following tutorials explain how to perform other common tasks in SAS:

ARABPSYCHOLOGY.COM