

How can the SAS language's SAS: Use CONTAINS be applied in PROC SQL?

Authored by
stats writer

June 25, 2024

RECOMMENDED CITATION

stats writer (2024). *How can the SAS language's SAS: Use CONTAINS be applied in PROC SQL?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=152813>

The SAS language's SAS: Use CONTAINS function is a powerful tool that can be utilized in PROC SQL to efficiently search and filter data within a specified column. This function allows for the identification of specific patterns or strings of characters within a larger string, making it ideal for tasks such as data cleansing and text mining. By incorporating the CONTAINS function into PROC SQL, users can easily perform complex data manipulations and retrieve desired results with greater precision and speed. This feature enhances the functionality and versatility of the SAS language, making it a valuable tool for data analysis and management.

SAS: Use CONTAINS in PROC SQL

You can use the CONTAINS operator in the PROC SQL statement in SAS to only return rows where a variable in a dataset contains some string pattern.

The following examples show how to use the CONTAINS operator in practice with the following dataset in SAS that contains information about various basketball players:

```
/*create dataset*/  
data my_data;  
input team $ points;  
datalines;  
Cavs 12  
Cavs 14  
Warriors 15  
Hawks 18
```

Mavs 31

Mavs 32

Mavs 35

Celtics 36

Celtics 40

;

run;

/*view dataset*/

proc printdata=my_data;

Obs	team	points
1	Cavs	12
2	Cavs	14
3	Warriors	15
4	Hawks	18
5	Mavs	31
6	Mavs	32
7	Mavs	35
8	Celtics	36
9	Celtics	40

Example 1: Select Rows where Variable Contains One Pattern

We can use the CONTAINS operator in PROC SQL to select only the rows where the team contains the pattern 'avs' somewhere in the name:

```
/*select all rows where team contains 'avs'*/  
proc sql;  
select *  
from my_data  
where team contains 'avs';  
quit;
```

team	points
Cavs	12
Cavs	14
Mavs	31
Mavs	32
Mavs	35

Notice that only the rows where the team contains 'avs' somewhere in the name are returned.

Example 2: Select Rows where Variable Contains One of Several Patterns

We can use the CONTAINS operator in PROC SQL to select only the rows where the team contains the pattern 'avs' or the pattern 'ics' somewhere in the name:

```
/*select all rows where team contains 'avs' or 'ics'*/  
proc sql;
```

```
select *  
from my_data  
where team contains 'avs' or team contains 'ics';  
quit;
```

team	points
Cavs	12
Cavs	14
Mavs	31
Mavs	32
Mavs	35
Celtics	36
Celtics	40

Only the rows where the team contains 'avs' or 'ics' somewhere in the name are returned.

Example 3: Select Rows where Variable Does Not Contain Pattern

The opposite of the CONTAINS operator in PROC SQL is NOT CONTAINS, which selects rows where some variable in a dataset does *not* contain a certain string pattern.

The following code shows how to use the NOT CONTAINS operator to select all rows where the team does not contain 'avs' in the name:

```
/*select all rows where team does not contain 'avs'*/  
proc sql;  
select *  
from my_data  
where team not contains 'avs';  
quit;
```

team	points
Warriors	15
Hawks	18
Celtics	36
Celtics	40

Notice that only the rows where the team does not contain 'avs' somewhere in the name are returned.

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