

# How to Replace Multiple Values in a PySpark Column

Authored by  
**stats writer**

February 4, 2026

## RECOMMENDED CITATION

stats writer (2026). *How to Replace Multiple Values in a PySpark Column*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=129365>

Replacing multiple values in one column in PySpark can be achieved by using the "replace" function. This function allows you to specify a list of values to be replaced and their corresponding replacement values. It can be used on a single column or multiple columns at once, making it a convenient solution for data cleaning or transformation tasks. Additionally, the "replace" function is supported in both PySpark DataFrame and Spark SQL, providing flexibility for users to choose their preferred method. By using this function, you can efficiently and effectively replace multiple values in a column in PySpark.

## PySpark: Replace Multiple Values in One Column

You can use the following syntax to replace multiple values in one column of a PySpark DataFrame:

```
from pyspark.sql.functions import when#replace
multiple values in 'team' column
df_new = df.withColumn('team', when(df.team=='A',
'Atlanta')
.when(df.team=='B', 'Boston')
.when(df.team=='C', 'Chicago'))
.otherwise(df.team))
```

This particular example makes the following replacements in the team column of the DataFrame:

Replace 'A' with 'Atlanta'  
Replace 'B' with 'Boston'  
Replace 'C' with 'Chicago'

The following examples show how to use this syntax in

**practice.**

**Example: How to Replace Multiple Values in Column of PySpark DataFrame**

**Suppose we have the following PySpark DataFrame that contains information about various basketball players:**

```
from pyspark.sql import SparkSession  
spark = SparkSession.builder.getOrCreate()
```

```
#define data
```

```
data = ,
```

```
,
```

```
,
```

```
,
```

```
,
```

```
,
```

```
,
```

```
]
```

```
#define column names
```

```
columns =
```

```
#create dataframe using data and column names
```

```
df = spark.createDataFrame(data, columns)
```

```
#view dataframe
```

```
df.show()
```

```
+----+-----+-----+
|team|conference|points|
+----+-----+-----+
| A| East| 11|
| A| East| 8|
| A| East| 31|
| B| West| 16|
| B| West| 6|
| C| East| 5|
| D| West| 12|
| D| West| 24|
+----+-----+-----+
```

We can use the following syntax to replace several values in the team column of the DataFrame:

```
from pyspark.sql.functions import when#replace
multiple values in 'team' column
df_new = df.withColumn('team', when(df.team=='A',
'Atlanta')
.when(df.team=='B', 'Boston'))
```

```
.when(df.team=='C', 'Chicago'))  
.otherwise(df.team))
```

```
#view new DataFrame  
df_new.show()
```

```
+-----+-----+-----+  
| team|conference|points|  
+-----+-----+-----+  
|Atlanta| East| 11|  
|Atlanta| East| 8|  
|Atlanta| East| 31|  
| Boston| West| 16|  
| Boston| West| 6|  
|Chicago| East| 5|  
| D| West| 12|  
| D| West| 24|  
+-----+-----+-----+
```

**Notice that several of the values in the team column have been replaced with specific new values.**

**Note that we did not specify a value to replace 'D' with in the team column so it simply remained the same.**

**Note: You can find the complete documentation for the PySpark when function .**

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

ARABPSYCHOLOGY.COM