How to Calculate Summary Statistics for a Pandas DataFrame?

Authored by

stats writer

November 29, 2025

To calculate summary statistics for a Pandas DataFrame, use the DataFrame's describe() method. This returns a summary of the data, including the count, mean, standard deviation, minimum, maximum, and quartiles. The describe() method can be used on both numerical and categorical data, and is useful for getting a quick overview of the data.

You can use the following methods to calculate summary statistics for variables in a pandas DataFrame:

Method 1: Calculate Summary Statistics for All Numeric Variables

df.describe()

Method 2: Calculate Summary Statistics for All String Variables

df.describe(include='object')

Method 3: Calculate Summary Statistics Grouped by a Variable

df.groupby('group_column').mean()

df.groupby('group_column').median()

df.groupby('group_column').max()

...

The following examples show how to use each method in practice with the following pandas DataFrame:

import pandas as pd
import numpy as np

#create DataFrame
df = pd.DataFrame({'team': ,
'points': ,
'assists': ,
'rebounds': })

#view DataFrame
print(df)

team points assists rebounds
0 A 18 5.0 11.0
1 A 22 NaN 8.0
2 A 19 7.0 10.0
3 A 14 9.0 6.0
4 B 14 12.0 6.0
5 B 11 9.0 5.0
6 B 20 9.0 9.0
7 B 28 4.0 NaN
8 B 30 5.0 6.0

Example 1: Calculate Summary Statistics for All Numeric Variables

The following code shows how to calculate the summary statistics for each numeric variable in the DataFrame:

df.describe()

points assists rebounds
count 9.000000 8.000000 8.000000
mean 19.555556 7.500000 7.625000
std 6.366143 2.725541 2.199838
min 11.000000 4.000000 5.000000
25% 14.000000 5.000000 6.000000
50% 19.000000 8.000000 7.000000
75% 22.000000 9.000000 9.250000
max 30.000000 12.000000 11.000000

We can see the following summary statistics for each of the three numeric variables:

count: The count of non-null values
mean: The mean value
std: The standard deviation
min: The minimum value
25%: The value at the 25th percentile
50%: The value at the 50th percentile (also the median)
75%: The value at the 75th percentile
max: The maximum value

Example 2: Calculate Summary Statistics for All String Variables

The following code shows how to calculate the summary statistics for each string variable in the DataFrame:

df.describe(include='object')

team
count 9
unique 2
top B
freq 5

count: The count of non-null values
unique: The number of unique values
top: The most frequently occurring value
freq: The count of the most frequently occurring value

Example 3: Calculate Summary Statistics Grouped by a Variable

The following code shows how to calculate the mean value for all numeric variables, grouped by the **team** variable:

df.groupby('team').mean()

points assists rebounds
team
A 18.25 7.0 8.75
B 20.60 7.8 6.50

The output displays the mean value for the **points**, **assists**, and **rebounds** variables, grouped by the **team** variable.

Note that we can use similar syntax to calculate a different summary statistic, such as the median:

df.groupby('team').median()

points assists rebounds
team
A 18.5 7.0 9.0
B 20.0 9.0 6.0

The output displays the median value for the **points**, **assists**, and **rebounds** variables, grouped by the **team** variable.

Note: You can find the complete documentation for the **describe** function in pandas .