

How can I use the describe() function in Pandas to summarize descriptive statistics for categorical variables?

Authored by
stats writer

June 24, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I use the describe() function in Pandas to summarize descriptive statistics for categorical variables?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=151245>

The describe() function in Pandas is a useful tool for summarizing descriptive statistics for categorical variables. It provides a concise and informative overview of the distribution of categorical data, including the count, unique values, top value, and frequency of the top value. Additionally, it can also calculate the mean, standard deviation, and quartile values for numerical variables within each category. This function allows for a quick and efficient way to gain insights into the categorical data, making it a valuable tool for data analysis and decision making.

Pandas: Use describe() for Categorical Variables

By default, the describe() function in pandas calculates descriptive statistics for all numeric variables in a DataFrame.

However, you can use the following methods to calculate descriptive statistics for as well:

Method 1: Calculate Descriptive Statistics for Categorical Variables

```
df.describe(include='object')
```

This method will calculate count, unique, top and freq for each categorical variable in a DataFrame.

Method 2: Calculate Categorical Descriptive Statistics for All Variables

```
df.astype('object').describe()
```

This method will calculate count, unique, top and freq for every variable in a DataFrame.

The following examples show how to use each method with the following pandas DataFrame that contains information about various basketball players:

```
import pandas as pd

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

#view DataFrame
print(df)

team points assists rebounds
0 A 18 5 11
1 B 22 7 8
2 C 19 7 10
3 D 14 9 6
4 E 14 12 6
5 F 11 9 5
```

6 G 20 9 9

7 H 28 4 12

Example 1: Calculate Descriptive Statistics for Categorical Variables

We can use the following syntax to calculate descriptive statistics for each categorical variable in the DataFrame:

#calculate descriptive statistics for categorical variables only

```
df.describe(include='object')
```

team

count 8

unique 8

top A

freq 1

The output shows various descriptive statistics for the only categorical variable (team) in the DataFrame.

Here's how to interpret the output:

count: There are 8 values in the team column.unique:

There are 8 unique values in the team column.top: The "top" value (i.e. highest in the alphabet) is A.freq: This top value occurs 1 time.

Example 2: Calculate Categorical Descriptive Statistics for All Variables

#calculate categorical descriptive statistics for all variables

df.astype('object').describe()

team points assists rebounds

count 8 8 8 8

unique 8 7 5 7

top A 14 9 6

freq 1 2 3 2

The output shows count, unique, top and freq for every variable in the DataFrame, including the numeric variables.

The following tutorials explain how to perform other common operations in pandas: