

How can I calculate the cumulative percentage in Pandas?

Authored by
stats writer

July 1, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I calculate the cumulative percentage in Pandas?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=164843>

Pandas is a popular data analysis library in Python that offers a wide range of functions for manipulating and analyzing data. One of the frequently used operations in data analysis is calculating the cumulative percentage of a particular column or series. This can be done in Pandas by using the "cumsum" function, which calculates the cumulative sum of a series. To obtain the cumulative percentage, the "cumsum" function can be divided by the total sum of the series and multiplied by 100. This process can be easily implemented using Pandas' powerful and intuitive syntax, making it an efficient tool for calculating the cumulative percentage in data analysis tasks.

Calculate Cumulative Percentage in Pandas

You can use the following basic syntax to calculate the cumulative percentage of values in a column of a pandas DataFrame:

```
#calculate cumulative sum of column
```

```
df = df.cumsum()
```

```
#calculate cumulative percentage of column (rounded to 2 decimal places)
```

```
df = round(100*df.cum_sum/df.sum(),2)
```

The following example shows how to use this syntax in practice.

Example: Calculate Cumulative Percentage in Pandas

Suppose we have the following pandas DataFrame that shows the number of units a company sells during

consecutive years:

```
import pandas as pd
```

```
#create DataFrame
```

```
df = pd.DataFrame({'year': ,  
'units_sold': })
```

```
#view DataFrame
```

```
print(df)
```

```
year units_sold
```

```
0 1 60
```

```
1 2 75
```

```
2 3 77
```

```
3 4 87
```

```
4 5 104
```

```
5 6 134
```

```
6 7 120
```

```
7 8 125
```

```
8 9 140
```

```
9 10 150
```

Next, we can use the following code to add a column that shows the cumulative number of units sold and

cumulative percentage of units sold:

```
#calculate cumulative sum of units sold
```

```
df = df.cumsum()
```

```
#calculate cumulative percentage of units sold
```

```
df = round(100*df.cum_sum/df.sum(),2)
```

```
#view updated DataFrame
```

```
print(df)
```

```
year units_sold cum_sum cum_percent
```

```
0 1 60 60 5.60
```

```
1 2 75 135 12.59
```

```
2 3 77 212 19.78
```

```
3 4 87 299 27.89
```

```
4 5 104 403 37.59
```

```
5 6 134 537 50.09
```

```
6 7 120 657 61.29
```

```
7 8 125 782 72.95
```

```
8 9 140 922 86.01
```

```
9 10 150 1072 100.00
```

We interpret the cumulative percentages as follows:

5.60% of all sales were made in year 1.12.59 of all sales were made in years 1 and 2 combined.19.78% of all sales were made in years 1, 2, and 3 combined.

And so on.

Note that you can simply change the value in the round() function to change the number of decimal points shown as well.

For example, we could round the cumulative percentage to zero decimal places instead:

```
#calculate cumulative sum of units sold
```

```
df = df.cumsum()
```

```
#calculate cumulative percentage of units sold
```

```
df = round(100*df.cum_sum/df.sum(),0)
```

```
#view updated DataFrame
```

```
print(df)
```

```
year units_sold cum_sum cum_percent
```

```
0 1 60 60 6.0
```

```
1 2 75 135 13.0
```

```
2 3 77 212 20.0
```

3 4 87 299 28.0

4 5 104 403 38.0

5 6 134 537 50.0

6 7 120 657 61.0

7 8 125 782 73.0

8 9 140 922 86.0

9 10 150 1072 100.0

The cumulative percentages are now rounded to zero decimal places.

Additional Resources

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