How do I sum values in a column that match a given condition using pandas?

Ask Question

40

16

Suppose I have a column like so:

```
a   b
1   5
1   7
2   3
1   3
2   5
```

I want to sum up the values for `b` where `a = 1`, for example. This would give me `5 + 7 + 3 = 15`.

How do I do this in pandas?

python pandas dataframe data-analysis

edited Dec 11 '16 at 17:21

Alex Riley
83.3k 26 166 168

asked Jan 30 '15 at 12:48

adijo
537 1 8 16

2 use group by function in pandas to do this task – sundar nataraj Jan 30 '15 at 12:52

1 Answer

54

It's quite common to use boolean indexing for this kind of task. With this method, you find out where column `a` is equal to `1` and then sum the corresponding rows of column `b`. You can use `loc` to handle the indexing:

```
>>> df.loc[df['a'] == 1, 'b'].sum
15
```

part and pull out the value that the 1's added up to:

```
>>> df.groupby('a')['b'].sum()[1]
15
```

The `groupby` approach is much slower than using boolean indexing, but is useful if you want check the sums for other values in column `a`.

edited Oct 22 '15 at 21:50

answered Jan 30 '15 at 12:53

Alex Riley
83.3k 26 166 168

What is the `.b` part? How are you indexing the columns like that? – adijo Jan 30 '15 at 13:09

You can access column `a` of the DataFrame by writing `df['a']` or `df.a`. The second one is convenient but doesn't work well if you have a column name with multiple words like "person id" or if the column name coincides with a DataFrame method like "where" or "sum". – Alex Riley Jan 30 '15 at 13:11

Ok thanks, also how do I convert all the values in the column to an `int` thereby enabling me to sum them? I was using `map` to convert all of them to `ints` but I think there might be an inbuilt function in pandas to do this more efficiently. – adijo Jan 30 '15 at 13:16

1 You could write `df['a'] = df['a'].astype(int)` or `df['a'] = df['a'].convert_objects(convert_numeric=True)` to do that. – Alex Riley Jan 30 '15 at 13:18

3 @LucSpan: absolutely, you can write `df.loc[(df['a'] == 1) & (df['c'] == 2), 'b'].sum()` to sum the values. – Alex Riley May 8 '17 at 11:53