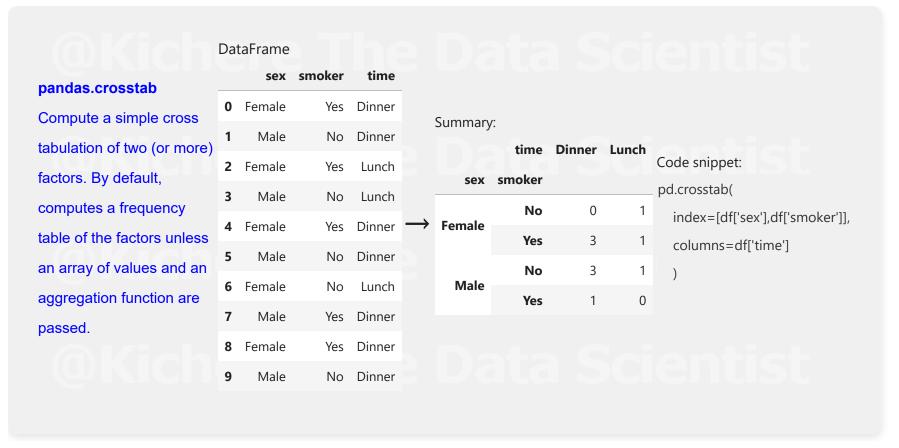
pandas.crosstab Overview

Count of smokers and non-smokers for each sex



Preface

What Makes This Ebook Unique

Contacts and Social Media

Sources & References

Contents

The "index" Parameter
The "columns" Parameter
The "values" Parameter
The "aggfunc" Parameter
The "rownames" Parameter
The "rownames" Parameter
The "margins" Parameter
The "margins Parameter

1st Edition

Project(Real Life application)

Why This E-book?

"The aim of this ebook is to give you the 'aha' moment right away at the start of learning a new concept."

• Practical step By Step Guide With Simple Examples

- Visual Illustrations and Interactive
- Simple Datasets
- Comprehensive Coverage(pandas Documentation used as reference)

Introduction

pandas.crosstab is a powerful function in the Pandas library used for creating contingency tables, which display the frequency distribution of variables. It allows you to cross-tabulate two or more factors, showing the counts or proportions of each combination of categories. You can easily add margins (totals) and normalize the data to show percentages or fractions, making it a versatile tool for data analysis and summarization.

The "pandas.crosstab" syntax

pandas.crosstab(index, columns, values=None, rownames=None,
colnames=None, aggfunc=None, margins=False, margins_name='All'
dropna=True, normalize=False)

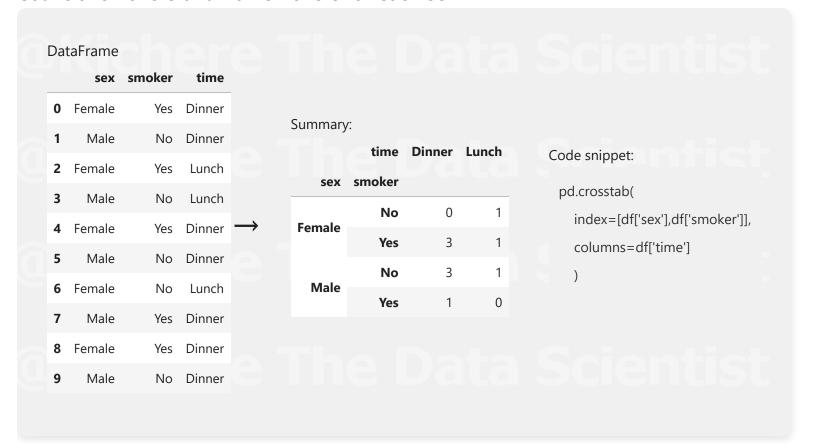
The index parameter

Specificies values to group by in the rows.

This parameter specifies the column(s) whose unique values will be used to form the rows of the resulting cross-tabulation.

import numpy as np
import pandas as pd

Count of smokers and non-smokers for each sex



Code Snippet:

data = {
 'sex': ['Female', 'Male', 'Female', 'Male', 'Mal

	sex	smoker	time	
0	Female	Yes	Dinner	
1	Male	No	Dinner	
2	Female	Yes	Lunch	

```
male', 'Male'],
                                                                                               sex smoker
                                                                                                                 time
  'smoker': ['Yes', 'No', 'Yes', 'No', 'Yes', 'No', 'No', 'Yes', 'Yes', 'No'],
                                                                                              Male
                                                                                                               Lunch
                                                                                        3
                                                                                                         No
  'time': ['Dinner', 'Dinner', 'Lunch', 'Lunch', 'Dinner', 'Dinner', 'Lunch', 'Dinner',
                                                                                                         Yes Dinner
                                                                                        4 Female
'Dinner', 'Dinner']
                                                                                              Male
                                                                                                         No Dinner
df = pd.DataFrame(data)
                                                                                        6 Female
                                                                                                              Lunch
                                                                                                         No
df
                                                                                                         Yes Dinner
                                                                                             Male
                                                                                                         Yes Dinner
                                                                                        8 Female
                                                                                              Male
                                                                                                         No Dinner
```

The columns parameter

Specifies Values to group by in the columns.

The "columns" parameter accepts array-like, Series, or list of arrays/Series

Counts of combinations of smoker and time for each sex category



6 Female No Lunch time Dinner Lu 7 Male Yes Dinner sex 8 Female Yes Dinner Male 3 9 Male No Dinner	S	ex	smoker	time		smoker		No		Yes	columns=[df['smoker'], df['time']
8 Female Yes Dinner Male 3	ìā	ale	No	Lunch		time	Dinner	Lunch	Dinner	Lunch)
	1a	ale	Yes	Dinner		sex					
9 Male No Dinner	ìa	ale	Yes	Dinner	-0	Male	3	1	1	0	
J Male 140 Billier	1a	ale	No	Dinner							

Code Snippet:

```
data = {
    'sex': ['Female', 'Male', 'Female', 'Male', 'Female', 'Male', 'Female', 'Male', 'Fe
male', 'Male'],
    'smoker': ['Yes', 'No', 'Yes', 'No', 'Yes', 'No', 'Yes', 'Yes', 'No'],
    'time': ['Dinner', 'Dinner', 'Lunch', 'Lunch', 'Dinner', 'Dinner', 'Lunch', 'Dinner',
'Dinner', 'Dinner']
}
df = pd.DataFrame(data)
df
```

df:

	sex	smoker	time
0	Female	Yes	Dinner
1	Male	No	Dinner
2	Female	Yes	Lunch
3	Male	No	Lunch
4	Female	Yes	Dinner
5	Male	No	Dinner
6	Female	No	Lunch
7	Male	Yes	Dinner
8	Female	Yes	Dinner
9	Male	No	Dinner

The values parameter

Maximum tip amount for each combination of 'sex' and 'smoker' status

	sex	smoker	tip					
0	Female	Yes	5.5	_				Code snippet:
1	Female	Yes	3.0		Summary			pd.crosstab(
2	Male	No	4.5		smoker	No	Yes	index=df['sex'],
3	Male	No	3.5	\rightarrow	sex			columns=df['smoker'],
4	Female	No	4.0		Female	5.0	5.5	values=df['tip'],
5	Female		5.0		Male	4.5	6.0	aggfunc='max'
6	Male	Yes)
7	Male	Yes						

	df:			
		sex	smoker	tip
Code Snippet:	0	Female	Yes	5.5
data = {	1	Female	Yes	3.0
'sex': ['Female', 'Female', 'Male', 'Female', 'Female', 'Male', 'Male'], 'smoker': ['Yes', 'Yes', 'No', 'No', 'No', 'Yes', 'Yes'],	2	Male	No	4.5
'tip': [5.5, 3.0, 4.5, 3.5, 4.0, 5.0, 6.0, 2.5]	3	Male	No	3.5
}	4	Female	No	4.0
	5	Female	No	5.0

df = pd.DataFrame(data)
df

	sex	smoker	tip
6	Male	Yes	6.0
7	Male	Yes	2.5

The rownames parameter

"rownames" is used to label the rows in the resulting table.

The "rownames" parameter accepts sequence and it is None by default.

If passed, must match number of row arrays passed.

Rename the row index to 'Sexoo'



ichere The Data

Code Snippet:

df

```
data = {
    'sex': ['Female', 'Male', 'Female', 'Male', 'Female', 'Male', 'Female', 'Male', 'Female',
'Male'],
    'smoker': ['Yes', 'Yes', 'Yes', 'No', 'Yes', 'No', 'Yes', 'Yes', 'Yes', 'No']
}
df = pd.DataFrame(data)
```

	sex	smoker
0	Female	Yes
1	Male	Yes
2	Female	Yes
3	Male	No
4	Female	Yes
5	Male	No
6	Female	No



	sex	smoker
7	Male	Yes
8	Female	Yes
9	Male	No

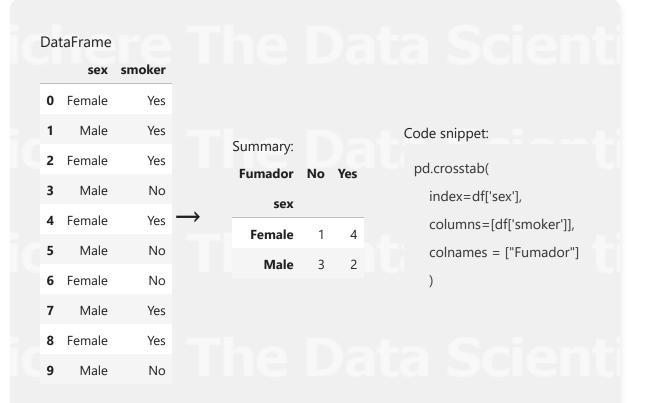
The colnames parameter

"colnames" is used to label the cols in the resulting table.

The "colnames " parameter accepts sequence and it is None by default.

If passed, must match number of col arrays passed.

Rename the column index(smoker) to 'Fumador'



ichere The Data

Code Snippet:

```
data = {
    'sex': ['Female', 'Male', 'Female', 'Male', 'Female', 'Male', 'Female', 'Male', 'Female',
'Male'],
    'smoker': ['Yes', 'Yes', 'Yes', 'No', 'Yes', 'No', 'Yes', 'Yes', 'No']
}
df = pd.DataFrame(data)
df
```

	sex	smoker
0	Female	Yes
1	Male	Yes
2	Female	Yes
3	Male	No
4	Female	Yes
5	Male	No
6	Female	No

	sex	smoker
7	Male	Yes
8	Female	Yes
9	Male	No

The aggfunc parameter

Accepts a function and it is optional..

If specified, requires values be specified as well.

Sum and the mean of tips for each combination of 'sex' and 'smoker'

	sex	smoker	tip							Code snippet:
)	Female	Yes	5.5		Summary	:				pd.crosstab(
l	Female	Yes	3.0				sum	a	mean	index=df['sex'],
2	Male	No	4.5		smoker	No	Yes	No	Yes	columns=df['smoker'],
3	Male	No	3.5	\rightarrow	sex					values=df['tip'],
ļ	Female	No	4.0		Female	9.0	8.5	4.5	4.25	aggfunc=['sum', 'mean']
5	Female	No	5.0		Male	8.0	8.5	4.0	4.25)
5	Male	Yes	6.0							
7	Male	Yes	2.5							

The margins parameter

The "margins" parameter Adds row/column margins (subtotals).

It is boolean and defaults to False.

Sum and the mean of tips for each combination of 'sex' and 'smoker', and also include row and column totals

Fer	male	Ves			Summary							
		103	5.5		carminary	•						pd.crosstab(
Fer	male	Yes	3.0		e.T	'n	e	sum	a		mean	index=df['sex'],
N	Male	No	4.5		smoker	No	Yes	All	No	Yes	All	columns=df['smoker'],
N	Male	No	3.5	\rightarrow	sex							values=df['tip'],
Fer	male	No	4.0		Female	9.0		17.5	4.50	4.25	4.375	aggfunc=['sum', 'mean'],
Fer	male	No	5.0		Male	8.0					4.125	margins = True
N	Male	Yes	6.0		All	17.0	17.0	34.0	4.25	4.25	4.250)
	Male	Yes	2.5									

The margins_name parameter

The "margins_name" parameter specifies the name of the row or column that will contain the totals when margins=True."

It is string and By default, the name is set to 'All'.

Sum and the mean of tips for each combination of 'sex' and 'smoker', and also include row and column totals

DataFrame

Code snippet:

	sex	smoker	tip	-	Summary:							pd.crosstab(
0	Female	Yes	5.5					sum			mean	index=df['sex'],
1	Female	Yes	3.0		smoker	No	Yes	Subtotal	No	Yes	Subtotal	columns=df['smoker'],
2	Male	No	4.5	\rightarrow	sex							values=df['tip'],
3	Male	No	3.5		Female	9.0	8.5	17.5	4.50	4.25	4.375	aggfunc=['sum', 'mean'],
4	Female	No	4.0		Male	8.0	8.5	16.5	4.00	4.25	4.125	margins = True,
5	Female	No	5.0		Subtotal	17.0	17.0	34.0	4.25	4.25	4.250	margins_name = 'Subtotal'
6	Male	Yes	6.0)
7	Male	Yes	2.5									

The dropna parameter

The "dropna" parameter does not include columns whose entries are all NaN. If True, rows with a NaN value in any column will be omitted before computing margins.

It is a boolean and defaults to True.

Show the total sales (sum of 'Price') for each manager across different products.

When dropna=False:											- /	1.6.10			
							Product	CPU	RAM	nan	All	When dropn			
							Manager					Product	CPU	RAM	All
								600.0	100.0	NI-NI	700.0	Manager			
df:							Debra	600.0	100.0	NaN	700.0	Debra	600.0	100.0	700.0
	Manager	Product	Quantity	Price	Status		Fred	300.0	0.0	NaN	300.0	Fred	300.0	0.0	300.0
0	Debra	CPU	2.0	600.0	None		NaN	NaN	NaN	0.0	NaN	All		100.0	1000.0
1	Debra	RAM	NaN	100.0	None		All	900.0	100.0	NaN	1000.0				
		CPU				\rightarrow						nd crosst	ah(
2	Fred	CPU	1.0	300.0	None	pd.crosstab(<pre>pd.crosstab(index=df['Manager'],</pre>					
3	Fred	RAM	3.0	NaN	None		index=df['Manager'],					columns=df['Product'],			
							columns=df['Product'],					. ,			

anager	Product	Quantity	Price	Status	values=df['Price'],	values=df['Price'],		
None	None	NaN	NaN	None	aggfunc='sum', margins=True, dropna = False	aggfunc='sum', margins=True, dropna =True)		

```
Code Snippet:
                                                             df:
 data = {
                                                                 Manager Product Quantity Price Status
   "Manager": ["Debra", "Debra", "Fred", "Fred", None],
                                                                                CPU
                                                              0
                                                                    Debra
                                                                                            2.0 600.0
                                                                                                        None
   "Product": ["CPU", "RAM", "CPU", "RAM", None],
   "Quantity": [2.0, None, 1.0, 3.0, None],
                                                                                          NaN 100.0
                                                                    Debra
                                                                               RAM
                                                                                                        None
   "Price": [600.0, 100.0, 300.0, None, None],
                                                              2
                                                                      Fred
                                                                                CPU
                                                                                            1.0 300.0
                                                                                                        None
   "Status": [None, None, None, None, None]
                                                                               RAM
                                                                                            3.0
                                                                                                 NaN
                                                                                                        None
                                                              3
                                                                      Fred
                                                                              None
                                                                     None
                                                                                          NaN
                                                                                                 NaN
                                                                                                        None
 df = pd.DataFrame(data)
 df
```

The normalize parameter

The "normalize" parameter is boolean. It accepts $\{'all', 'index', 'columns'\}$, or $\{0,1\}$ and defaults to False. Its purpose is to normalize(Proportion) by dividing all values by the sum of values.

Code Snippet:

```
sex smoker
data = {
            'sex': ['Female', 'Male', 'Female', 'Male', 'Male'
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   0 Female
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        Yes
male', 'Male'],
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    Male
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        Yes
            'smoker': ['Yes', 'No', 'Yes', 'No', 'Yes', 'No', 'No', 'Yes', 'Yes', 'No'],
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   2 Female
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        Yes
            'time': ['Dinner', 'Dinner', 'Lunch', 'Lunch', 'Dinner', 'Dinner', 'Lunch', 'Dinner',
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    Male
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         No
 'Dinner', 'Dinner']
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   4 Female
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        Yes
df = pd.DataFrame(data)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    Male
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          No
df
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     6 Female
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          No
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    Male
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        Yes
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   8 Female
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       Yes
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             Male
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         No
```

If passed 'all' or True, will normalize over all values.

Generate a table of sex and smoker status in a DataFrame, showing proportions relative to the entire dataset df:

	sex	smoker	Without nnormalize:									
				smoker	No	Yes	Total	When norn	When normalize = True or "all":			
0	Female	Yes		sex				smoker	No	Yes		
1	Male	Yes	-	Female	1	4	5	sex				
2	Female	Yes		remale	'	4	3					
_	N 4 - 1 -	NI -		Male	3	2	5	Female	0.1	0.4		
3	Male	No		Total	4	6	10	Male	0.3	0.2		
4	Female	Yes	\longrightarrow	Total	·	Ü	10					
5	Male	No		pd.crosstab(pd.crosstab(
6	Female No			index colun			, noker']],	index=[df['sex']], columns=[df['smoker']],				

```
sexsmokermargins = True,<br/>margins_name='Total'<br/>)normalize = True<br/>)7MaleYes9MaleNo
```

If passed 'index' will normalize over each row.

df:

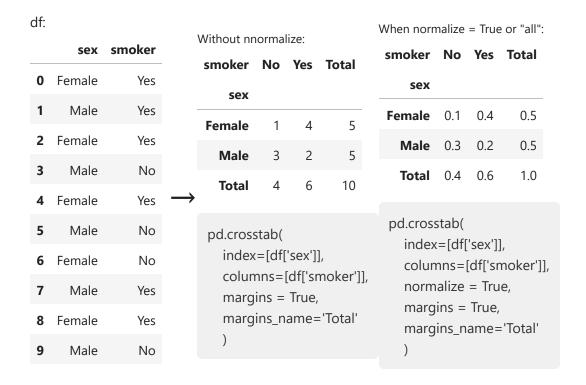
	sex	smoker	,	Without nr										
				smoker	No	Yes	Total	When r	orn	nalize	= inde	ex or "0":		
0	Female	Yes		sex				smok	er	No	Yes			
1	Male	Yes							S	ex				
2	Female	Yes		Female	1	4	5							
_		N.		Male	3	2	5	Fema	le	0.2	0.8			
3	Male	No		Total	4	6	10	Ma	le	0.6	0.4			
4	Female	Yes	\rightarrow											
5	Male	No		pd.cross	stab(pd.cr	pd.crosstab(
6	Female	No		index					<pre>index=[df['sex']], columns=[df['smoker']],</pre>					
7	Male	Yes		colun			noker']]	1	normalize = 'index'					
8	Female	Yes		marg)							
9	Male	No)										

If passed 'columns' will normalize over each column.

	sex	smoker	Without no	rmalı	ze:		
			smoker	No	Yes	Total	When normalize = 'index' or 1:
0	Female	Yes	sex				
1	Male	Yes		1			
			Female	I	4	5	

2 F				smoker	INO	Yes	Total		smoker	No	Yes			
	Female	Yes	\rightarrow	sex					sex					
3	Male	No		Male	3	2	5		Female	0.25	0.666667			
4 F	Female	Yes		Total	4	6	10		Male	0.75	0.333333			
5	Male	No												
6 F	Female	No		pd.cross index	'sex'l		pd.crosstab(index=[df['sex']],							
7	Male	Yes		colun		columns=[df['smoker']]								
8 F	Female	Yes		margins = True, margins_name='Total'						normalize = 'index'				
9	Male	No)	airie-	,								

If margins is True, will also normalize margin values.



Project(Real Life application)

pandas.crosstab (https://interactivechaos.com/en/python/function/pandascrosstab)

Sources & References

pandas.crosstab Documentation (https://pandas.pydata.org/docs/reference/api/pandas.crosstab.html#pandas.crosstab)
pandas.crosstab (https://interactivechaos.com/en/python/function/pandascrosstab)

Contacts and Social Media

Kichere Magubu

- A Dar es salaam, Tanzania
- in Kichere Magubu
- ▶ Kichere The Data Scientist
- KichereTheDataScientist
- K KichereTheDataScientist
- kicherethedatascientist@gmail.com
- © kicherethedatascientist
- **J** +255 654 729 851