

Pandas crosstab explained

Source dataframe

make	body_style	drive_wheels	num_doors	curb_weight
mazda	hatchback	fwd	two	2385
volvo	sedan	rwd	four	2912
toyota	sedan	fwd	four	2140
mitsubishi	hatchback	fwd	two	1944
volkswagen	sedan	fwd	two	2261
mazda	sedan	fwd	four	2410
mazda	hatchback	rwd	two	2380
peugot	sedan	rwd	four	3197
mazda	sedan	fwd	four	1945
nissan	sedan	fwd	two	1918

Basic Usage

```
pd.crosstab(df.make, df.body_style)
```

body_style	convertible	hardtop	hatchback	sedan	wagon
make					
honda	0	0	7	5	1
mazda	0	0	10	7	0
mitsubishi	0	0	9	4	0
nissan	0	1	5	9	3
subaru	0	0	3	5	4
toyota	1	3	14	10	4
volkswagen	1	0	1	9	1
volvo	0	0	0	8	3

Margin Totals

```
pd.crosstab(df.make, df.num_doors,
            margins=True,
            margins_name="Total")
```

num_doors	four	two	Total
make			
honda	5	8	13
mazda	7	9	16
mitsubishi	4	9	13
nissan	9	9	18
subaru	9	3	12
toyota	18	14	32
volkswagen	8	4	12
volvo	11	0	11
Total	71	56	127

Label Rows and Columns

```
pd.crosstab(df.make, df.body_style,
            rownames=['Auto Manufacturer'],
            colnames=['Body Style'])
```

Body Style	convertible	hardtop	hatchback	sedan	wagon
Auto Manufacturer					
honda	0	0	7	5	1
mazda	0	0	10	7	0
mitsubishi	0	0	9	4	0
nissan	0	1	5	9	3
subaru	0	0	3	5	4
toyota	1	3	14	10	4
volkswagen	1	0	1	9	1
volvo	0	0	0	8	3

Grouping and Aggregating Values

```
pd.crosstab(df.make, [df.body_style, df.drive_wheels], values=df.curb_weight, aggfunc='mean').fillna('-')
```

body_style	convertible		hardtop		hatchback		sedan			wagon			
drive_wheels	fwd	rwd	fwd	rwd	4wd	fwd	rwd	4wd	fwd	rwd	4wd	fwd	rwd
make													
honda	-	-	-	-	-	1970	-	-	2288.8	-	-	2024	-
mazda	-	-	-	-	-	2148.33	2411.25	-	2231.6	2685	-	-	-
mitsubishi	-	-	-	-	-	2376.56	-	-	2394	-	-	-	-
nissan	-	-	2008	-	-	2176	3116.33	-	2237.89	-	-	2452.33	-
subaru	-	-	-	-	2240	2085	-	2447.5	2225	-	2535	2372.5	-
toyota	-	2975	-	2585	-	2177.25	2626.83	-	2258.57	2521.67	2700	2280	2151
volkswagen	2254	-	-	-	-	2221	-	-	2342.22	-	-	2563	-
volvo	-	-	-	-	-	-	-	-	-	3023	-	-	3077.67

Average curb weight for all fwd toyota wagons

Normalize All Values

```
pd.crosstab(df.make,
            df.body_style,
            normalize=True)
```

body_style	convertible	hardtop	hatchback	sedan	wagon
make					
honda	0.000000	0.000000	0.054688	0.039062	0.007812
mazda	0.000000	0.000000	0.078125	0.054688	0.000000
mitsubishi	0.000000	0.000000	0.070312	0.031250	0.000000
nissan	0.000000	0.007812	0.039062	0.070312	0.023438
subaru	0.000000	0.000000	0.023438	0.039062	0.031250
toyota	0.007812	0.023438	0.109375	0.078125	0.031250
volkswagen	0.007812	0.000000	0.007812	0.070312	0.007812
volvo	0.000000	0.000000	0.000000	0.062500	0.023438

Normalize Rows

```
pd.crosstab(df.make,
            df.body_style,
            normalize='index')
```

body_style	convertible	hardtop	hatchback	sedan	wagon
make					
honda	0.000000	0.000000	0.538462	0.384615	0.076923
mazda	0.000000	0.000000	0.588235	0.411765	0.000000
mitsubishi	0.000000	0.000000	0.692308	0.307692	0.000000
nissan	0.000000	0.055556	0.277778	0.500000	0.166667
subaru	0.000000	0.000000	0.250000	0.416667	0.333333
toyota	0.031250	0.093750	0.437500	0.312500	0.125000
volkswagen	0.083333	0.000000	0.083333	0.750000	0.083333
volvo	0.000000	0.000000	0.000000	0.727273	0.272727

Normalize Columns

```
pd.crosstab(df.make,
            df.body_style,
            normalize='columns')
```

body_style	convertible	hardtop	hatchback	sedan	wagon
make					
honda	0.0	0.00	0.142857	0.087719	0.0625
mazda	0.0	0.00	0.204082	0.122807	0.0000
mitsubishi	0.0	0.00	0.183673	0.070175	0.0000
nissan	0.0	0.25	0.102041	0.157895	0.1875
subaru	0.0	0.00	0.061224	0.087719	0.2500
toyota	0.5	0.75	0.285714	0.175439	0.2500
volkswagen	0.5	0.00	0.020408	0.157895	0.0625
volvo	0.0	0.00	0.000000	0.140351	0.1875