

Frequency tables

pandas.crosstab()

- To compute a simple cross-tabulation of one, two (or more) factors
- By default computes a frequency table of the factors

```
pd.crosstab(index=cars_data2['FuelType'], columns='count',
            dropna=True)
```

Size of data

1436– Original data

1336 – after dropping
nan values

Out[3]:

col_0	count
FuelType	
CNG	15
Diesel	144
Petrol	1177

Most of the cars have petrol as fuel type

Two-way tables

pandas.crosstab()

- To look at the frequency distribution of gearbox types with respect to different fuel types of the cars

```
pd.crosstab(index = cars_data2['Automatic'],
             columns = cars_data2['FuelType'],
             dropna = True)
```

```
Out[5]:
FuelType  CNG  Diesel  Petrol
Automatic
0         15    144    1104
1         0     0     73
```

Automatic

0- Manual gear box

1- Automatic gearbox

Two-way table - joint probability

`pandas.crosstab()`

- Joint probability is the likelihood of two independent events happening at the same time

```
pd.crosstab(index      = cars_data2['Automatic'],  
            columns    = cars_data2['FuelType'],  
            normalize  = True,  
            dropna     = True)
```

Out[16]:

FuelType	CNG	Diesel	Petrol
Automatic			
0	0.010801	0.108011	0.828083
1	0.000000	0.000000	0.053105

Two-way table - marginal probability

pandas.crosstab()

- Marginal probability is the probability of the occurrence of the single event

```
pd.crosstab(index = cars_data2['Automatic'],
            columns = cars_data2['FuelType'],
            margins = True,
            dropna = True,
            normalize = True)
```

probability of cars having manual gear box when the fuel type are CNG or Diesel or Petrol is 0.95

```
Out[17]:
FuelType      CNG      Diesel      Petrol      All
Automatic
0      0.010801  0.108011  0.828083  0.946895
1      0.000000  0.000000  0.053105  0.053105
All      0.010801  0.108011  0.881188  1.000000
```

Two-way table - conditional probability

pandas.crosstab()

- Conditional probability is the probability of an event (A), given that another event (B) has already occurred
- Given the type of gear box, probability of different fuel type

```
pd.crosstab(index      = cars_data2['Automatic'],
            columns    = cars_data2['FuelType'],
            margins    = True,
            dropna     = True,
            normalize  = 'index')
```

Out[19]:

FuelType	CNG	Diesel	Petrol
Automatic			
0	0.011407	0.114068	0.874525
1	0.000000	0.000000	1.000000
All	0.010801	0.108011	0.881188

→ Row sum = 1

Two-way table - conditional probability

pandas.crosstab()

- Conditional probability is the probability of an event (A), given that another event (B) has already occurred

```
pd.crosstab(index      = cars_data2['Automatic'],
             columns    = cars_data2['FuelType'],
             margins    = True, dropna = True,
             normalize  = 'columns')
```

Out[20]:

FuelType	CNG	Diesel	Petrol	All
Automatic				
0	1.0	1.0	0.939734	0.946895
1	0.0	0.0	0.060266	0.053105



Column sum = 1