In [3]:

```
import pandas as pd
df=pd.read_csv("haircolor.csv")
df
```

Out[3]:

		name	country	gender	age	hair_color
0	NaN	Ram	India	M	23	black
1	NaN	Mathew	UK	M	27	brown
2	NaN	Gillian	UK	F	43	brown
3	NaN	Tom	USA	M	33	brown
4	NaN	Anna	USA	F	25	blonde
5	NaN	Sophia	USA	F	27	blonde
6	NaN	Emma	UK	F	52	blonde
7	NaN	Sweta	India	F	23	black
8	NaN	Mohan	India	M	44	black
9	NaN	Amelia	UK	F	24	blonde

In [4]:

```
df_drop = pd.crosstab([df.gender, df.country], df.hair_color)
```

In [5]:

```
df_drop=pd.crosstab([df.gender,df.country],df.hair_color)
df
```

Out[5]:

		name	country	gender	age	hair_color
0	NaN	Ram	India	M	23	black
1	NaN	Mathew	UK	M	27	brown
2	NaN	Gillian	UK	F	43	brown
3	NaN	Tom	USA	M	33	brown
4	NaN	Anna	USA	F	25	blonde
5	NaN	Sophia	USA	F	27	blonde
6	NaN	Emma	UK	F	52	blonde
7	NaN	Sweta	India	F	23	black
8	NaN	Mohan	India	M	44	black
9	NaN	Amelia	UK	F	24	blonde

In [6]:

```
df_drop.drop('M')
```

Out[6]:

		hair_color	black	blonde	brown
gender	country				
F	India		1	0	0
	UK		0	2	1
	USA		0	2	0

In [8]:

```
df_drop.drop(index='M')
```

Out[8]:

		hair_color	black	blonde	brown
gender	country				
F	India		1	0	0
	UK		0	2	1
	USA		0	2	0

In [10]:

```
df_drop.drop(index = 'India', level=1)
```

Out[10]:

		hair_color	black	blonde	brown
gender	country				
F	UK		0	2	1
	USA		0	2	0
M	UK		0	0	1
	USA		0	0	1

In [11]:

```
df_drop.drop(index = ['India', 'UK'], level=1)
```

Out[11]:

		hair_color	black	blonde	brown
gender	country				
F	USA		0	2	0
M	USA		0	0	1

In [12]:

```
df_drop.drop(columns = 'brown')
```

Out[12]:

	hair_color	black	blonde
gender	country		
F	India	1	0
	UK	0	2
	USA	0	2
M	India	2	0
	UK	0	0
	USA	0	0

In [13]:

```
df_drop.drop(columns = ['brown', 'black'])
```

Out[13]:

	hair_color	blonde
gender	country	
F	India	0
	UK	2
	USA	2
M	India	0
	UK	0
	USA	0

In [14]:

```
df_drop = pd.crosstab(df.country, [df.gender, df.hair_color])
```

In [15]:

```
df_drop
```

Out[15]:

gender	F			M	
hair_color	black	blonde	brown	black	brown
country					
India	1	0	0	2	0
UK	0	2	1	0	1
USA	0	2	0	0	1

In [16]:

```
df_drop.drop(columns = ['black', 'blonde'], level=1)
```

Out[16]:

gender	F	M
hair_color	brown	brown
country		
India	0	0
UK	1	1
USA	0	1

In [17]:

```
df_drop.drop(index='UK', columns='M')
```

Out[17]:

gender			F
hair_color	black	blonde	brown
country			
India	1	0	0
USA	0	2	0

In [19]:

```
import pandas as pd
```

In [21]:

```
df = pd.DataFrame({
"name":['arun', 'varun', 'neha', 'varun', 'varun', 'arun'],
'instruments':['violin', 'drum', 'flute', 'guitar', 'bongo','tabla'],
'start_date': ['Jan 10, 2020', 'Mar 3, 2003', 'Feb 6, 2005', 'Dec 8, 2008',
'Nov 5, 2011', 'Mar 10, 2011']
})
df.start_date = pd.to_datetime(df.start_date)
df
```

Out[21]:

	name	instruments	start_date
0	arun	violin	2020-01-10
1	varun	drum	2003-03-03
2	neha	flute	2005-02-06
3	varun	guitar	2008-12-08
4	varun	bongo	2011-11-05
5	arun	tabla	2011-03-10

In [22]:

```
df.drop_duplicates('name')
```

Out[22]:

	name	instruments	start_date
0	arun	violin	2020-01-10
1	varun	drum	2003-03-03
2	neha	flute	2005-02-06

In [24]:

```
df.drop_duplicates('name', keep='last')
```

Out[24]:

	name	instruments	start_date
2	neha	flute	2005-02-06
4	varun	bongo	2011-11-05
5	arun	tabla	2011-03-10

In [25]:

```
df.drop_duplicates('name', keep=False)
```

Out[25]:

	name	instruments	start_date
2	neha	flute	2005-02-06

In [26]:

```
df.drop_duplicates('name', keep='last', ignore_index=True)
```

Out[26]:

	name	instruments	start_date
0	neha	flute	2005-02-06
1	varun	bongo	2011-11-05
2	arun	tabla	2011-03-10

In [27]:

```
import pandas as nd
import numpy as np
```

In [29]:

```
df = pd.DataFrame({
'alphabet':list('dpbtbkc'),
'num1':[1,2,np.nan,4,3,7,2],
'num2':[3,4,3,4,2,5,4]
})
df
```

Out[29]:

	alphabet	num1	num2
0	d	1.0	3
1	p	2.0	4
2	b	NaN	3
3	t	4.0	4
4	b	3.0	2
5	k	7.0	5
6	c	2.0	4

In [30]:

```
df.sort_values('alphabet')
```

Out[30]:

	alphabet	num1	num2
2	b	NaN	3
4	b	3.0	2
6	c	2.0	4
0	d	1.0	3
5	k	7.0	5
1	p	2.0	4
3	t	4.0	4

In [31]:

```
df.sort_values(by=['alphabet', 'num2'])
```

Out[31]:

	alphabet	num1	num2
4	b	3.0	2
2	b	NaN	3
6	c	2.0	4
0	d	1.0	3
5	k	7.0	5
1	p	2.0	4
3	t	4.0	4

In [32]:

```
df.sort_values(by='alphabet',ascending=False)
```

Out[32]:

	alphabet	num1	num2
3	t	4.0	4
1	p	2.0	4
5	k	7.0	5
0	d	1.0	3
6	c	2.0	4
2	b	NaN	3
4	b	3.0	2

In [33]:

```
df.sort_values(by='num1')
```

Out[33]:

	alphabet	num1	num2
0	d	1.0	3
1	p	2.0	4
6	c	2.0	4
4	b	3.0	2
3	t	4.0	4
5	k	7.0	5
2	b	NaN	3

In [34]:

```
df.sort_values(by='num1', na_position='first')
```

Out[34]:

	alphabet	num1	num2
2	b	NaN	3
0	d	1.0	3
1	p	2.0	4
6	c	2.0	4
4	b	3.0	2
3	t	4.0	4
5	k	7.0	5

In [ ]:

```

```