```
#https://github.com/docaotien/Lab-MKTG5883.N22.CTTT
```

```
from google.colab import drive
drive.mount('/content/gdrive')
```

```
Mounted at /content/gdrive
```

```
import pandas as pd
df = pd.read_csv('gdrive/My Drive/PastHires.csv')
```

```
%matplotlib inline
import numpy as np
```

```
df.head()
```

	Years Experience	Employed?	Previous employers	Level of Education	Top-tier school	Interned	Hired
0	10	Y	4	BS	N	N	Y
1	0	N	0	BS	Y	Y	Y
2	7	N	6	BS	N	N	N
3	2	Y	1	MS	Y	N	Y
4	20	N	2	PhD	Y	N	N

```
df.head(10)
```

	Years Experience	Employed?	Previous employers	Level of Education	Top-tier school	Interned	Hired
0	10	Y	4	BS	N	N	Y
1	0	N	0	BS	Y	Y	Y
2	7	N	6	BS	N	N	N
3	2	Y	1	MS	Y	N	Y
4	20	N	2	PhD	Y	N	N
5	0	N	0	PhD	Y	Y	Y
6	5	Y	2	MS	N	Y	Y
7	3	N	1	BS	N	Y	Y
8	15	Y	5	BS	N	N	Y
9	0	N	0	BS	N	N	N

```
df.tail(4)
```

	Years Experience	Employed?	Previous employers	Level of Education	Top-tier school	Interned	Hired
9	0	N	0	BS	N	N	N
10	1	N	1	PhD	Y	N	N
11	4	Y	1	BS	N	Y	Y
12	0	N	0	PhD	Y	N	Y

```
df.shape
```

```
(13, 7)
```

```
df.size
```

```
91
```

```
len(df)
```

```
13
```

```
df.columns
```

```
Index(['Years Experience', 'Employed?', 'Previous employers',
       'Level of Education', 'Top-tier school', 'Interned', 'Hired'],
      dtype='object')
```

```
df["Hired"]
```

```
0     Y
1     Y
2     N
3     Y
4     N
5     Y
6     Y
7     Y
8     Y
9     N
10    N
11    Y
12    Y
Name: Hired, dtype: object
```

```
df["Hired"][:5]
```

```
0    Y
1    Y
2    N
3    Y
4    N
Name: Hired, dtype: object
```

```
df["Hired"][5]
```

```
'Y'
```

```
df[["Years Experience", "Hired"]]
```

	Years Experience	Hired
0	10	Y
1	0	Y
2	7	N
3	2	Y
4	20	N
5	0	Y
6	5	Y
7	3	Y
8	15	Y
9	0	N
10	1	N
11	4	Y
12	0	Y

```
df[["Years Experience", "Hired"]][:5]
```

	Years Experience	Hired
0	10	Y
1	0	Y
2	7	N
3	2	Y
4	20	N

```
df.sort_values(['Years Experience'])
```

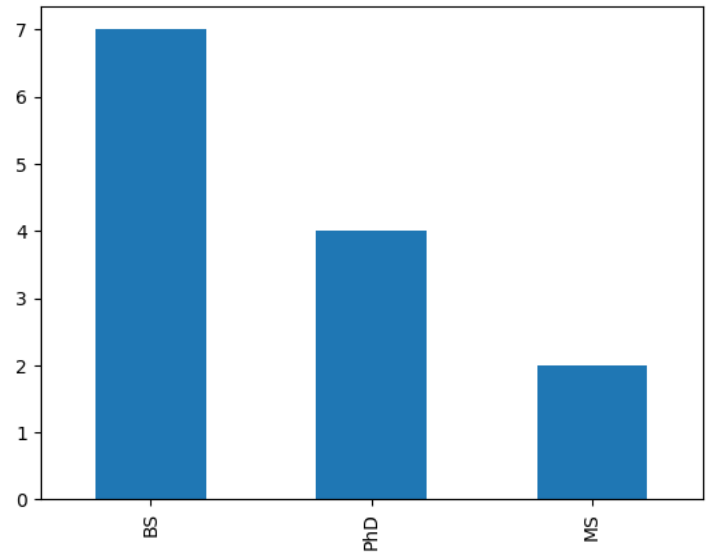
	Years Experience	Employed?	Previous employers	Level of Education	Top-tier school	Interned	Hired
1	0	N	0	BS	Y	Y	Y
5	0	N	0	PhD	Y	Y	Y
9	0	N	0	BS	N	N	N
12	0	N	0	PhD	Y	N	Y
10	1	N	1	PhD	Y	N	N
3	2	Y	1	MS	Y	N	Y
7	3	N	1	BS	N	Y	Y
11	4	Y	1	BS	N	Y	Y
6	5	Y	2	MS	N	Y	Y
2	7	N	6	BS	N	N	N
0	10	Y	4	BS	N	N	Y
8	15	Y	5	BS	N	N	Y
4	20	N	2	PhD	Y	N	N

```
degree_counts = df['Level of Education'].value_counts()
degree_counts
```

```
BS     7
PhD    4
MS     2
Name: Level of Education, dtype: int64
```

```
degree_counts.plot(kind='bar')
```

```
<Axes: >
```




```
import numpy as np
```

```
import pandas as pd
```

```
labels = ['a','b','c']
my_list = [10,20,30]
arr = np.array([10,20,30])
d = {'a':10,'b':20,'c':30}
```

```
pd.Series(data=my_list)
```

```
0    10
1    20
```

```
2    30
dtype: int64
```

```
pd.Series(data=my_list,index=labels)
```

```
a    10
b    20
c    30
dtype: int64
```

```
pd.Series(arr)
```

```
0    10
1    20
2    30
dtype: int64
```

```
pd.Series(arr,labels)
```

```
a    10
b    20
c    30
dtype: int64
```

```
pd.Series(d)
```

```
a    10
b    20
c    30
dtype: int64
```

```
pd.Series(data=labels)
```

```
0    a
1    b
2    c
dtype: object
```

```
pd.Series([sum,print,len])
```

```
0      <built-in function sum>
1    <built-in function print>
2      <built-in function len>
dtype: object
```

```
seri = pd.Series([1,2,3,4],index = ['USA','Germany','USSR','Japan'])
```

```
seri
```

```
USA        1
Germany    2
USSR       3
Japan      4
dtype: int64
```

```
serii = pd.Series([1,2,5,4],index = ['USA','Germany','Italy','Japan'])
```

```
serii
```

```
USA        1
Germany    2
Italy      5
Japan      4
dtype: int64
```

```
seri['USA']
```

```
1
```

```
seri + serii
```

```
Germany    4.0
Italy      NaN
```

```
Japan      8.0
USA        2.0
USSR       NaN
dtype: float64
```

```
import pandas as pd
import numpy as np
```

```
from numpy.random import randn
np.random.seed(101)
```

```
df = pd.DataFrame(randn(5,4),index = 'A B C D E'.split(),columns='W X Y Z'.split())
```

```
df
```

	W	X	Y	Z
A	2.706850	0.628133	0.907969	0.503826
B	0.651118	-0.319318	-0.848077	0.605965
C	-2.018168	0.740122	0.528813	-0.589001
D	0.188695	-0.758872	-0.933237	0.955057
E	0.190794	1.978757	2.605967	0.683509

```
df['W']
```

```
A    2.706850
B    0.651118
C   -2.018168
D    0.188695
E    0.190794
Name: W, dtype: float64
```

```
df[['W','Z']]
```

	W	Z
A	2.706850	0.503826
B	0.651118	0.605965
C	-2.018168	-0.589001
D	0.188695	0.955057
E	0.190794	0.683509

```
df.W
```

```
A    2.706850
B    0.651118
C   -2.018168
D    0.188695
E    0.190794
Name: W, dtype: float64
```

```
type(df['W'])
```

```
pandas.core.series.Series
```

```
df['new'] = df['W'] + df['Y']
```

```
df
```

	W	X	Y	Z	new
A	2.706850	0.628133	0.907969	0.503826	3.614819

```
df.drop('new',axis=1)
```

	W	X	Y	Z
A	2.706850	0.628133	0.907969	0.503826
B	0.651118	-0.319318	-0.848077	0.605965
C	-2.018168	0.740122	0.528813	-0.589001
D	0.188695	-0.758872	-0.933237	0.955057
E	0.190794	1.978757	2.605967	0.683509

```
df
```

	W	X	Y	Z	new
A	2.706850	0.628133	0.907969	0.503826	3.614819
B	0.651118	-0.319318	-0.848077	0.605965	-0.196959
C	-2.018168	0.740122	0.528813	-0.589001	-1.489355
D	0.188695	-0.758872	-0.933237	0.955057	-0.744542
E	0.190794	1.978757	2.605967	0.683509	2.796762

```
df.drop('E',axis = 0)
```

	W	X	Y	Z	new
A	2.706850	0.628133	0.907969	0.503826	3.614819
B	0.651118	-0.319318	-0.848077	0.605965	-0.196959
C	-2.018168	0.740122	0.528813	-0.589001	-1.489355
D	0.188695	-0.758872	-0.933237	0.955057	-0.744542

```
df.drop('E',axis=0)
```

	W	X	Y	Z	new
A	2.706850	0.628133	0.907969	0.503826	3.614819
B	0.651118	-0.319318	-0.848077	0.605965	-0.196959
C	-2.018168	0.740122	0.528813	-0.589001	-1.489355
D	0.188695	-0.758872	-0.933237	0.955057	-0.744542

```
df.loc['A']
```

```
W      2.706850
X      0.628133
Y      0.907969
Z      0.503826
new    3.614819
Name: A, dtype: float64
```

```
df.iloc[2]
```

```
W     -2.018168
X      0.740122
Y      0.528813
Z     -0.589001
new   -1.489355
Name: C, dtype: float64
```

```
df.loc['B','Y']
```

```
-0.8480769834036315
```

```
df.loc[['A','B'],['W','Y']]
```

	W	Y
A	2.706850	0.907969
B	0.651118	-0.848077

```
df
```

	W	X	Y	Z	new
A	2.706850	0.628133	0.907969	0.503826	3.614819
B	0.651118	-0.319318	-0.848077	0.605965	-0.196959
C	-2.018168	0.740122	0.528813	-0.589001	-1.489355
D	0.188695	-0.758872	-0.933237	0.955057	-0.744542
E	0.190794	1.978757	2.605967	0.683509	2.796762

```
df>0
```

	W	X	Y	Z	new
A	True	True	True	True	True
B	True	False	False	True	False
C	False	True	True	False	False
D	True	False	False	True	False
E	True	True	True	True	True

```
df[df>0]
```

	W	X	Y	Z	new
A	2.706850	0.628133	0.907969	0.503826	3.614819
B	0.651118	NaN	NaN	0.605965	NaN
C	NaN	0.740122	0.528813	NaN	NaN
D	0.188695	NaN	NaN	0.955057	NaN
E	0.190794	1.978757	2.605967	0.683509	2.796762

```
df[df['W']>0]
```

	W	X	Y	Z	new
A	2.706850	0.628133	0.907969	0.503826	3.614819
B	0.651118	-0.319318	-0.848077	0.605965	-0.196959
D	0.188695	-0.758872	-0.933237	0.955057	-0.744542
E	0.190794	1.978757	2.605967	0.683509	2.796762

```
df[df['W']>0]['Y']
```

```
A    0.907969
B   -0.848077
D   -0.933237
E    2.605967
Name: Y, dtype: float64
```

```
df[df['W']>0][['Y','X']]
```

	Y	X
A	0.907969	0.628133

```
df[(df['W']>0) & (df['Y']>1)]
```

	W	X	Y	Z	new
E	0.190794	1.978757	2.605967	0.683509	2.796762

```
df
```

	W	X	Y	Z	new
A	2.706850	0.628133	0.907969	0.503826	3.614819
B	0.651118	-0.319318	-0.848077	0.605965	-0.196959
C	-2.018168	0.740122	0.528813	-0.589001	-1.489355
D	0.188695	-0.758872	-0.933237	0.955057	-0.744542
E	0.190794	1.978757	2.605967	0.683509	2.796762

```
df1 = df.drop('new',axis=1)
```

```
df1
```

	W	X	Y	Z
A	2.706850	0.628133	0.907969	0.503826
B	0.651118	-0.319318	-0.848077	0.605965
C	-2.018168	0.740122	0.528813	-0.589001
D	0.188695	-0.758872	-0.933237	0.955057
E	0.190794	1.978757	2.605967	0.683509

```
df1.reset_index()
```

	index	W	X	Y	Z
0	A	2.706850	0.628133	0.907969	0.503826
1	B	0.651118	-0.319318	-0.848077	0.605965
2	C	-2.018168	0.740122	0.528813	-0.589001
3	D	0.188695	-0.758872	-0.933237	0.955057
4	E	0.190794	1.978757	2.605967	0.683509

```
newind = 'CA NY WY OR CO'.split()
```

```
df1['States'] = newind
```

```
df1
```

	W	X	Y	Z	States
A	2.706850	0.628133	0.907969	0.503826	CA
B	0.651118	-0.319318	-0.848077	0.605965	NY
C	-2.018168	0.740122	0.528813	-0.589001	WY
D	0.188695	-0.758872	-0.933237	0.955057	OR
E	0.190794	1.978757	2.605967	0.683509	CO

```
df1.set_index('States')
```

States	W	X	Y	Z
CA	2.706850	0.628133	0.907969	0.503826
NY	0.651118	-0.319318	-0.848077	0.605965
WY	-2.018168	0.740122	0.528813	-0.589001
OR	0.188695	-0.758872	-0.933237	0.955057
CO	0.190794	1.978757	2.605967	0.683509

```
df1
```

	W	X	Y	Z	States
A	2.706850	0.628133	0.907969	0.503826	CA
B	0.651118	-0.319318	-0.848077	0.605965	NY
C	-2.018168	0.740122	0.528813	-0.589001	WY
D	0.188695	-0.758872	-0.933237	0.955057	OR
E	0.190794	1.978757	2.605967	0.683509	CO

✓ 0 giây hoàn thành lúc 23:24