

Analyzing Employee Turnover

import the pandas module

In [1]:

```
import pandas as pd
```

Create path and file variables

In [2]:

```
path='C:/Users/HP/Downloads/'  
file='employee_turnover.csv'
```

Load the file into a DataFrame called 'data'

In [3]:

```
data=pd.read_csv(path+file)
```

Although there is a difference between Attrition and Turnover, these terms are used interchangeably

Attrition is a more popular choice of word than turnover(sounds fancier)

Attrition Rate = No. of Employees that Left Workforce / Average No. of Employees

For the purpose of this case study average no of employees is the same as the total number of employees

What is the attrition rate of this company?

In [4]:

```
l=list(data['Attrition'])  
c=0  
n=len(l)  
for i in l:  
    if i=='Yes':  
        c=c+1  
Attrition_rate=c/n  
print("Attrition Rate : ",Attrition_rate)
```

Attrition Rate : 0.16122448979591836

What is the average monthly income of all the employees in the company?

In [5]:

```
l=list(data['MonthlyIncome'])  
n=len(l)  
s=0  
for i in l:  
    s=s+i  
avg=s/n  
print("average salary : ",avg)
```

average salary : 6502.931292517007

Copy and Paste this line of code into your cell:

```
attrition_true=data[data['Attrition']=='Yes']
```

This will create a list of all people who are still with the company

In [6]:

```
attrition_true=data[data['Attrition']=='Yes']
```

What is the most popular rating in the attrition_true dataset?

In [7]:

```
l=list(attrition_true['DailyRate'])
res=l[0]
max=0
for i in l:
    freq=l.count(i)
    if freq>max:
        max=freq
        res=i
print("most popular rating : ",res)
```

```
most popular rating : 813
```

Copy and Paste this line of code into your cell:

```
attrition_false=data[data['Attrition']=='No']
```

This will create a list of all people who are still with the company

In [8]:

```
attrition_false=data[data['Attrition']=='No']
```

What is the most common Job Satisfaction rating for people still working at the company?

In [9]:

```
l=list(attrition_false['JobSatisfaction'])
res=l[0]
max=0
for i in l:
    freq=l.count(i)
    if freq>max:
        max=freq
        res=i
print("most common job satisfaction rating : ",res)
```

```
most common job satisfaction rating : 4
```

Are more females attriting compared to males?

In [10]:

```
l=list(attrition_true['Gender'])
m=l.count('Male')
f=l.count('Female')
if m>f:
    print("No")
else:
    print("Yes")
```

```
No
```

- a) What is the male to female ratio in the people who are still working at the company?
b) What inference can you draw from this?

In [11]:

```
l=list(attrition_false['Gender'])
m=l.count('Male')
f=l.count('Female')
r=m/f
print(r)
if r>1:
    print("More male")
elif r<1:
    print("More female")
else:
    print("Equal")
```

```
1.4610778443113772
More male
```

Is it true that unmarried people tend to attrite more often than married people?

Find out...

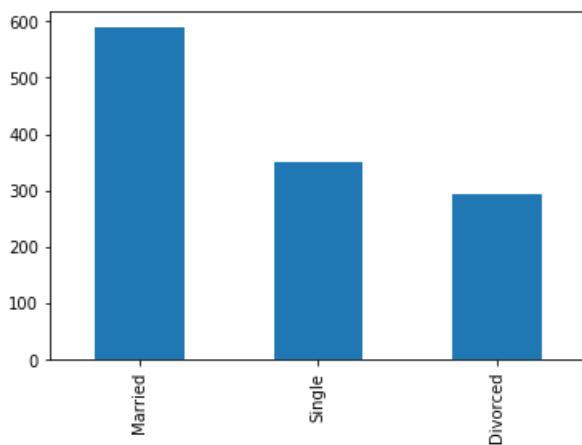
Hint: Use the 'MaritalStatus' column with 'value_counts'!

In [15]:

```
%matplotlib inline
attrition_false['MaritalStatus'].value_counts().plot.bar()
```

Out[15]:

<matplotlib.axes._subplots.AxesSubplot at 0x1871bcaf1c8>

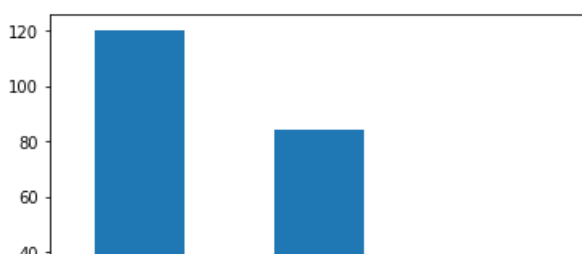


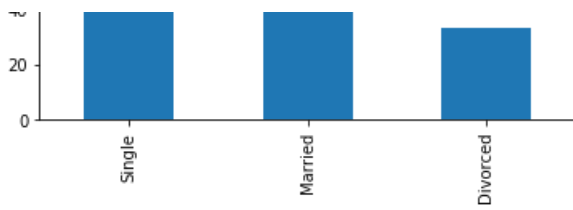
In [16]:

```
attrition_true['MaritalStatus'].value_counts().plot.bar()
```

Out[16]:

<matplotlib.axes._subplots.AxesSubplot at 0x1871cb631c8>





Is Overtime a common factor for attrition?

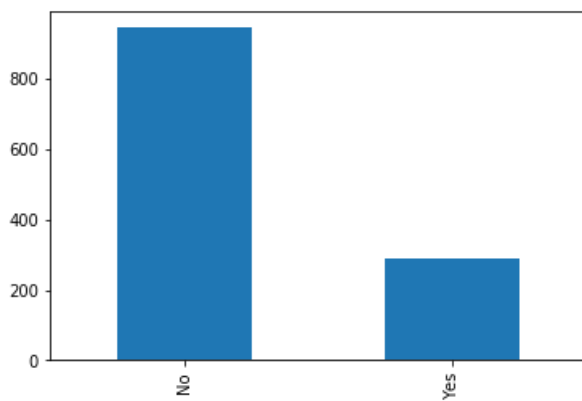
Use `value_counts()` function on 'OverTime' column for both datasets to find out!

In [17]:

```
attrition_false['OverTime'].value_counts().plot.bar()
```

Out[17]:

<matplotlib.axes._subplots.AxesSubplot at 0x1871cbc7808>

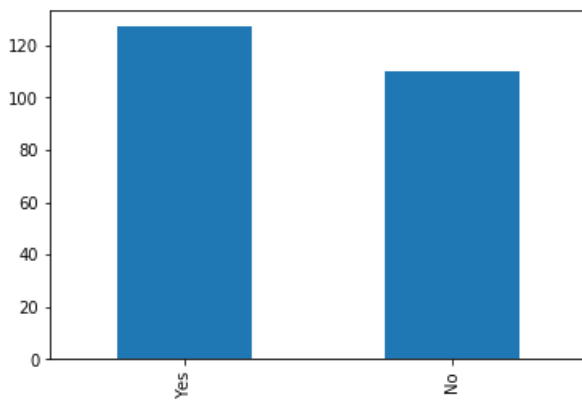


In [18]:

```
attrition_true['OverTime'].value_counts().plot.bar()
```

Out[18]:

<matplotlib.axes._subplots.AxesSubplot at 0x1871cc2d388>



In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []: