

I/O Operations with external files

Reading a CSV file & XLS file format

```
df = pd.read_csv('xyz.csv',header=None)
```

```
df.head() #This will display 1st 5 records
```

```
df.tail() #This will display last 5 records
```

The above dataset doesn't have header, we shall attach our own header.

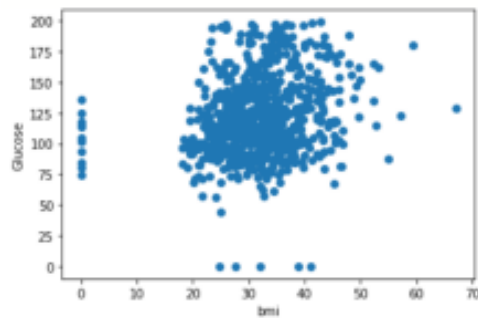
```
df.columns=['preg','glu','bp','sft','ins','bmi','dpf','age','class']
```

#Let us visualize the scatter plot of two continuous variable.

```
plt.scatter(df['bmi'],df['glu'])
```

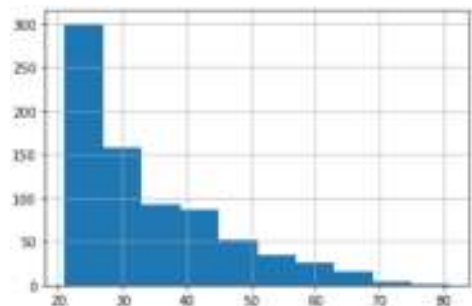
```
plt.xlabel('bmi')
```

```
plt.ylabel('Glucose')
```

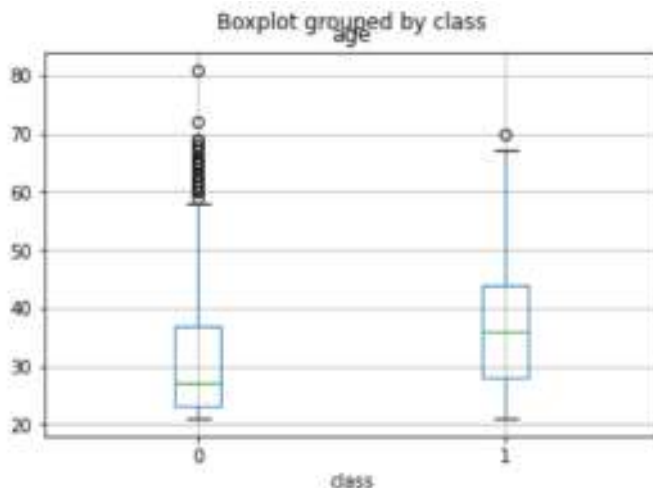


#Let us visualize the histogram of another continuous variable 'Age'

```
df['age'].hist()
```



#Let us visualize the distribution 'Age' with respect to Categories: Label-0(Healthy), Label-1 (Diabetes)



#We can observe the median age of diabetes is slightly greater than median age of healthy

```
W = pd.read_csv('xyz.xls',header=None)
```

```
W.head() #XLS file format also, we can read using pd.read_csv
```

```
D= np.loadtxt('xyz.data',delimiter=",")
```

```
D[:5,:] # this file is loaded in Numpy 2D array format
```

Reading a XLSX file format

```
G=pd.read_excel('xyz.xlsx',sheet_name='Sheet1')
```

```
G.head()
```

Here additionally we need to pass the sheet_name. If not specified, it will read the first page by default.

Reading a HTML file format

Pandas can read table tabs off of html. For example:

```
B = pd.read_html('http://www.fdic.gov/bank/individual/failed/banklist.html')
```

Reading a TXT file format

```
H = pd.read_table('HR.txt')
```

```
H.head()
```

```
f=H["Department"].value_counts()
```

```
f
```

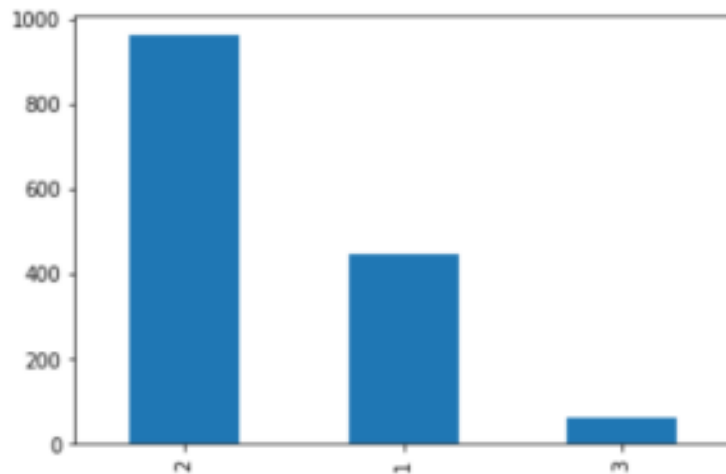
```
2 961
```

```
1 446
```

```
3 63
```

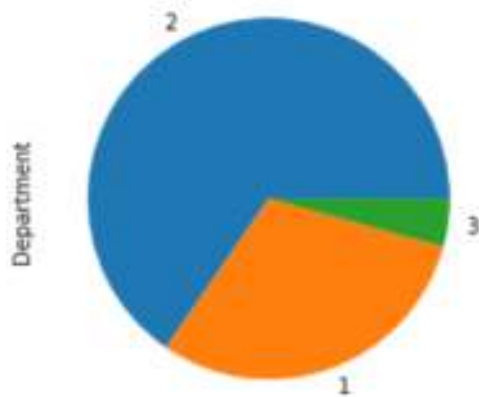
#We can visualize the distribution of categorical values using bar plot and pie chart

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



The above bar plot can be perceived in terms of Pie chart which would give % percentage information

```
f.plot(kind='pie')
```



#We can visualize two categorical variables at a time

```
fa=pd.crosstab(H['Gender'],H['Attrition'])
```

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

