# ASSIGNMENT - 7

**import pandas as pd**

**import numpy as np**

**import matplotlib.pyplot as plt**

**dataset1=pd.read_csv('general_data.csv')**

**dataset1.head()**
Output:
```
   Age Attrition  ... YearsSinceLastPromotion YearsWithCurrManager
0  51     No ...                 0                 0
1  31     Yes ...                1                 4
2  32     No ...                 0                 3
3  38     No ...                 7                 5
4  32     No ...                 0                 4

[5 rows x 24 columns]
```

**dataset1.columns**
Output:
```
Index(['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',
    'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',
    'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',
    'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',
    'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',
    'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager'],
    dtype='object')
```

**dataset1.isnull()**
Output:
```
      Age  Attrition ... YearsSinceLastPromotion  YearsWithCurrManager
0    False    False ...          False                False
1    False    False ...          False                False
2    False    False ...          False                False
3    False    False ...          False                False
4    False    False ...          False                False

  ...     ... ...           ...            ...
4405  False    False ...          False                False
4406  False    False ...          False                False
```

| 4407 | False | False | ... | False | False |
| 4408 | False | False | ... | False | False |
| 4409 | False | False | ... | False | False |

[4410 rows x 24 columns]

**dataset1.duplicated()**
Output:
| 0 | False |
| 1 | False |
| 2 | False |
| 3 | False |
| 4 | False |

| 4405 | False |
| 4406 | False |
| 4407 | False |
| 4408 | False |
| 4409 | False |

Length: 4410, dtype: bool

**dataset1.drop_duplicates()**
Output:
| | Age | Attrition | ... | YearsSinceLastPromotion | YearsWithCurrManager |
|---|---|---|---|---|---|
| 0 | 51 | No | ... | 0 | 0 |
| 1 | 31 | Yes | ... | 1 | 4 |
| 2 | 32 | No | ... | 0 | 3 |
| 3 | 38 | No | ... | 7 | 5 |
| 4 | 32 | No | ... | 0 | 4 |
| ... | ... | ... | ... | ... | ... |
| 4405 | 42 | No | ... | 0 | 2 |
| 4406 | 29 | No | ... | 0 | 2 |
| 4407 | 25 | No | ... | 1 | 2 |
| 4408 | 42 | No | ... | 7 | 8 |
| 4409 | 40 | No | ... | 3 | 9 |

[4410 rows x 24 columns]

**dataset3=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].describe()**
Output:

**dataset3=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].median()**

**dataset3**
Output:
Age                      36.0
DistanceFromHome         7.0
Education                3.0
MonthlyIncome            49190.0
NumCompaniesWorked       2.0
PercentSalaryHike        14.0
TotalWorkingYears        10.0
TrainingTimesLastYear    3.0
YearsAtCompany           5.0
YearsSinceLastPromotion  1.0
YearsWithCurrManager     3.0
dtype: float64

**dataset3=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].mode()**

**dataset3**
Output:
  Age  DistanceFromHome  ...  YearsSinceLastPromotion  YearsWithCurrManager
0   35                 2 ...                        0                     2

[1 rows x 11 columns]

**dataset3=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].mean()**

**dataset3**
Output:
Age                       36.923810
DistanceFromHome           9.192517
Education                  2.912925
MonthlyIncome          65029.312925
NumCompaniesWorked         2.694830
PercentSalaryHike         15.209524
TotalWorkingYears         11.279936
TrainingTimesLastYear      2.799320
YearsAtCompany             7.008163
YearsSinceLastPromotion    2.187755
YearsWithCurrManager       4.123129
dtype: float64

**dataset3=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].var()**

**dataset3**
Output:
Age                       8.341719e+01
DistanceFromHome          6.569144e+01
Education                 1.048438e+00
MonthlyIncome             2.215480e+09
NumCompaniesWorked        6.244436e+00
PercentSalaryHike         1.338907e+01
TotalWorkingYears         6.056298e+01
TrainingTimesLastYear     1.661465e+00
YearsAtCompany            3.751728e+01
YearsSinceLastPromotion   1.037935e+01
YearsWithCurrManager      1.272582e+01
dtype: float64

**dataset3=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].skew()**

**dataset3**

Output:

| | |
|---|---|
| Age | 0.413005 |
| DistanceFromHome | 0.957466 |
| Education | -0.289484 |
| MonthlyIncome | 1.368884 |
| NumCompaniesWorked | 1.026767 |
| PercentSalaryHike | 0.820569 |
| TotalWorkingYears | 1.116832 |
| TrainingTimesLastYear | 0.552748 |
| YearsAtCompany | 1.763328 |
| YearsSinceLastPromotion | 1.982939 |
| YearsWithCurrManager | 0.832884 |

dtype: float64

**dataset3=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].kurt()**

**dataset3**

Output:

| | |
|---|---|
| Age | -0.405951 |
| DistanceFromHome | -0.227045 |
| Education | -0.560569 |
| MonthlyIncome | 1.000232 |
| NumCompaniesWorked | 0.007287 |
| PercentSalaryHike | -0.302638 |
| TotalWorkingYears | 0.912936 |
| TrainingTimesLastYear | 0.491149 |
| YearsAtCompany | 3.923864 |
| YearsSinceLastPromotion | 3.601761 |
| YearsWithCurrManager | 0.167949 |

dtype: float64

**dataset3=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome', 'NumCompaniesWorked', 'PercentSalaryHike','TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany','YearsSinceLastPromotion', 'YearsWithCurrManager']].info()**

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4410 entries, 0 to 4409
Data columns (total 11 columns):
 #  Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0  Age                 4410 non-null   int64
```

| 1 | DistanceFromHome | 4410 non-null | int64 |
| 2 | Education | 4410 non-null | int64 |
| 3 | MonthlyIncome | 4410 non-null | int64 |
| 4 | NumCompaniesWorked | 4391 non-null | float64 |
| 5 | PercentSalaryHike | 4410 non-null | int64 |
| 6 | TotalWorkingYears | 4401 non-null | float64 |
| 7 | TrainingTimesLastYear | 4410 non-null | int64 |
| 8 | YearsAtCompany | 4410 non-null | int64 |
| 9 | YearsSinceLastPromotion | 4410 non-null | int64 |
| 10 | YearsWithCurrManager | 4410 non-null | int64 |

dtypes: float64(2), int64(9)
memory usage: 379.1 KB

**box_plot=dataset1.Age**

**plt.boxplot(box_plot)**
Output:
{'whiskers': [<matplotlib.lines.Line2D at 0x20ae69f6708>,
 <matplotlib.lines.Line2D at 0x20ae85fba88>],
 'caps': [<matplotlib.lines.Line2D at 0x20ae5df2048>,
 <matplotlib.lines.Line2D at 0x20ae69f9c88>],
 'boxes': [<matplotlib.lines.Line2D at 0x20ae85ff308>],
 'medians': [<matplotlib.lines.Line2D at 0x20ae68ff1c8>],
 'fliers': [<matplotlib.lines.Line2D at 0x20ae85f57c8>],
 'means': []}
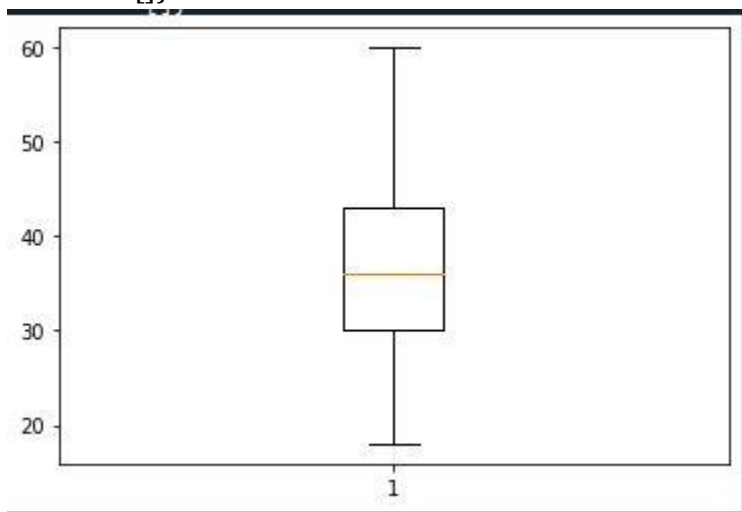


**box_plot=dataset1.MonthlyIncome**

**plt.boxplot(box_plot)**
Output:
{'whiskers': [<matplotlib.lines.Line2D at 0x20ae8e6c588>,
 <matplotlib.lines.Line2D at 0x20ae8f1ba08>],

```
'caps': [<matplotlib.lines.Line2D at 0x20ae8f1bb08>,
 <matplotlib.lines.Line2D at 0x20ae8f20988>],
'boxes': [<matplotlib.lines.Line2D at 0x20ae8f19908>],
'medians': [<matplotlib.lines.Line2D at 0x20ae8f20a88>],
'fliers': [<matplotlib.lines.Line2D at 0x20ae8f26908>],
'means': []}
```
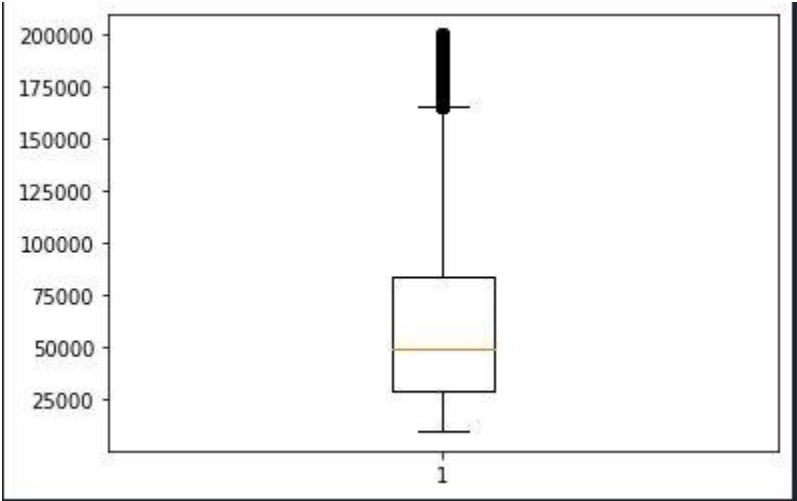


**box_plot=dataset1.YearsAtCompany**

**plt.boxplot(box_plot)**
Output:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x20ae8f5b788>,
 <matplotlib.lines.Line2D at 0x20ae8f8cd88>],
'caps': [<matplotlib.lines.Line2D at 0x20ae8f8ce88>,
 <matplotlib.lines.Line2D at 0x20ae8f91dc8>],
'boxes': [<matplotlib.lines.Line2D at 0x20ae8f87c88>],
'medians': [<matplotlib.lines.Line2D at 0x20ae8f91d48>],
'fliers': [<matplotlib.lines.Line2D at 0x20ae8f91f88>],
'means': []}
```