

```
import pandas as pd
import matplotlib.pyplot as plt
pd.read_csv("general_data.csv")
```

Out[7]:

	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]

```
dataset=pd.read_csv("general_data.csv")
dataset.head()
```

Out[8]:

	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]

```
dataset=pd.read_csv("general_data.csv")
```

```
dataset.head()
```

```
Out[9]:
```

	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]
```

```
dataset.duplicated()
```

```
Out[10]:
```

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

```
dataset1
```

Out[11]:

```
Age DistanceFromHome ... YearsSinceLastPromotion YearsWithCurrManager
0 35          2 ...              0              2
```

[1 rows x 11 columns]

```
dataset2=dataset1[['Age','DistanceFromHome','Education','MonthlyIncome',
'NumCompaniesWorked','PercentSalaryHike','TotalWorkingYears','TrainingTimesLastYear',
'YearsAtCompany','YearsSinceLastPromotion','YearsWithCurrManager']].describe()
```

dataset2

Out[12]:

```
Age DistanceFromHome ... YearsSinceLastPromotion YearsWithCurrManager
count  1.0          1.0 ...              1.0              1.0
mean  35.0          2.0 ...              0.0              2.0
std   NaN          NaN ...              NaN              NaN
min   35.0          2.0 ...              0.0              2.0
25%   35.0          2.0 ...              0.0              2.0
50%   35.0          2.0 ...              0.0              2.0
75%   35.0          2.0 ...              0.0              2.0
max   35.0          2.0 ...              0.0              2.0
```

[8 rows x 11 columns]

```
plt.boxplot(dataset["MonthlyIncome"])
```

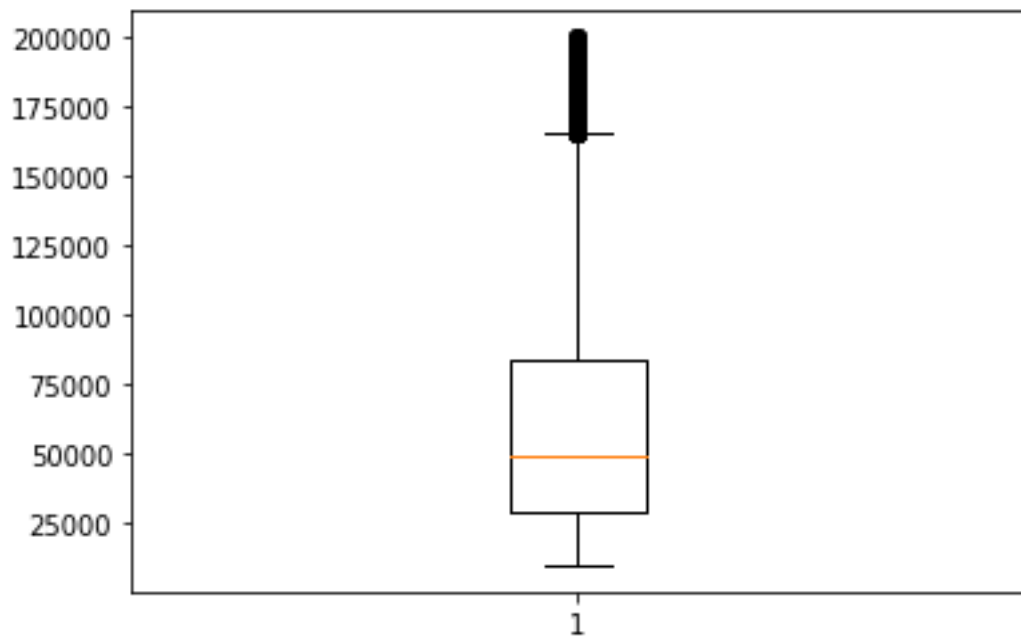
```
plt.boxplot(dataset["MonthlyIncome"])
```

Out[16]:

```
{'whiskers': [<matplotlib.lines.Line2D at 0x1e67de11e88>,
<matplotlib.lines.Line2D at 0x1e67de16d08>],
'caps': [<matplotlib.lines.Line2D at 0x1e67de16e08>,
<matplotlib.lines.Line2D at 0x1e67de1ac88>],
'boxes': [<matplotlib.lines.Line2D at 0x1e67de11d08>],
'medians': [<matplotlib.lines.Line2D at 0x1e67de1ad88>],
```

'fliers': [<matplotlib.lines.Line2D at 0x1e67de1ec08>],

'means': []}



In [17]: plt.scatter(dataset["MonthlyIncome"],dataset["PercentSalaryHike"])

Out[17]: <matplotlib.collections.PathCollection at 0x1e67e8d91c8>

