

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

```
In [2]: import os
```

```
In [3]: os.getcwd()
```

```
Out[3]: 'C:\\Users\\cmahl\\project notebook'
```

```
In [4]: os.chdir("C:\\Users\\cmahl\\Desktop")
```

```
In [5]: df=pd.read_csv("Salary.csv")
```

```
In [6]: df.head(10)
```

```
Out[6]:
```

	Unnamed: 0	YearsExperience	Salary
0	0	1.2	39344.0
1	1	1.4	46206.0
2	2	1.6	37732.0
3	3	2.1	43526.0
4	4	2.3	39892.0
5	5	3.0	56643.0
6	6	3.1	60151.0
7	7	3.3	54446.0
8	8	3.3	64446.0
9	9	3.8	57190.0

```
In [7]: df.tail()
```

```
Out[7]:
```

	Unnamed: 0	YearsExperience	Salary
25	25	9.1	105583.0
26	26	9.6	116970.0
27	27	9.7	112636.0
28	28	10.4	122392.0
29	29	10.6	121873.0

```
In [8]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 3 columns):
 #   Column          Non-Null Count  Dtype  
---  -
 0   Unnamed: 0      30 non-null    int64  
 1   YearsExperience  30 non-null    float64
 2   Salary          30 non-null    float64
dtypes: float64(2), int64(1)
memory usage: 848.0 bytes
```

```
In [9]: df.columns
```

```
Out[9]: Index(['Unnamed: 0', 'YearsExperience', 'Salary'], dtype='object')
```

```
In [10]: df.describe()
```

Out[10]:

	Unnamed: 0	YearsExperience	Salary
count	30.000000	30.000000	30.000000
mean	14.500000	5.413333	76004.000000
std	8.803408	2.837888	27414.429785
min	0.000000	1.200000	37732.000000
25%	7.250000	3.300000	56721.750000
50%	14.500000	4.800000	65238.000000
75%	21.750000	7.800000	100545.750000
max	29.000000	10.600000	122392.000000

```
In [11]: df.shape
```

Out[11]: (30, 3)

```
In [12]: df.size
```

Out[12]: 90

```
In [13]: df.ndim
```

Out[13]: 2

```
In [14]: df.isnull().sum()
```

Out[14]: Unnamed: 0 0
YearsExperience 0
Salary 0
dtype: int64

```
In [19]: #Assigning values in X & Y  
x = df.iloc[:, :-1].values  
y = df.iloc[:, -1].values  
#X = df['YearsExperience']  
#y = df['Salary']
```

```
In [20]: print(x)
```

```
[[ 0.  1.2]  
 [ 1.  1.4]  
 [ 2.  1.6]  
 [ 3.  2.1]  
 [ 4.  2.3]  
 [ 5.  3. ]  
 [ 6.  3.1]  
 [ 7.  3.3]  
 [ 8.  3.3]  
 [ 9.  3.8]  
[10.  4. ]  
[11.  4.1]  
[12.  4.1]  
[13.  4.2]  
[14.  4.6]  
[15.  5. ]  
[16.  5.2]  
[17.  5.4]  
[18.  6. ]  
[19.  6.1]  
[20.  6.9]  
[21.  7.2]  
[22.  8. ]  
[23.  8.3]  
[24.  8.8]  
[25.  9.1]
```

```
[26.  9.6]
[27.  9.7]
[28. 10.4]
[29. 10.6]]
```

```
In [21]: print(y)
```

```
[ 39344.  46206.  37732.  43526.  39892.  56643.  60151.  54446.  64446.
  57190.  63219.  55795.  56958.  57082.  61112.  67939.  66030.  83089.
  81364.  93941.  91739.  98274. 101303. 113813. 109432. 105583. 116970.
112636. 122392. 121873.]
```

```
In [35]: from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=.3,random_state=42)
```

```
In [36]: print(x_train)
```

```
[[ 0.  1.2]
 [ 4.  2.3]
[16.  5.2]
 [ 5.  3. ]
[13.  4.2]
[11.  4.1]
[22.  8. ]
 [ 1.  1.4]
 [ 2.  1.6]
[25.  9.1]
 [ 3.  2.1]
[21.  7.2]
[26.  9.6]
[18.  6. ]
[29. 10.6]
[20.  6.9]
 [ 7.  3.3]
[10.  4. ]
[14.  4.6]
[19.  6.1]
 [ 6.  3.1]]
```

```
In [32]: print(X_test)
```

```
[[27.  9.7]
[15.  5. ]
[23.  8.3]
[17.  5.4]
 [ 8.  3.3]
 [ 9.  3.8]
[28. 10.4]
[24.  8.8]
[12.  4.1]]
```

```
In [37]: print(y_train)
```

```
[ 39344.  39892.  66030.  56643.  57082.  55795. 101303.  46206.  37732.
 105583.  43526.  98274. 116970.  81364. 121873.  91739.  54446.  63219.
  61112.  93941.  60151.]
```

```
In [38]: print(y_test)
```

```
[112636.  67939. 113813.  83089.  64446.  57190. 122392. 109432.  56958.]
```