



```
In [4]: import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt
```

```
In [5]: saldf=pd.read_csv('/content/drive/MyDrive/Colab Notebooks/panda/Salary Data.csv')
```

```
In [6]: saldf.head()
```

Out[6]:

	Age	Gender	Education Level	Job Title	Years of Experience	Salary
0	32.0	Male	Bachelor's	Software Engineer	5.0	90000.0
1	28.0	Female	Master's	Data Analyst	3.0	65000.0
2	45.0	Male	PhD	Senior Manager	15.0	150000.0
3	36.0	Female	Bachelor's	Sales Associate	7.0	60000.0
4	52.0	Male	Master's	Director	20.0	200000.0

```
In [7]: saldf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 375 entries, 0 to 374
Data columns (total 6 columns):
 #   Column           Non-Null Count  Dtype  
 --- 
 0   Age              373 non-null    float64
 1   Gender           373 non-null    object  
 2   Education Level  373 non-null    object  
 3   Job Title        373 non-null    object  
 4   Years of Experience  373 non-null    float64
 5   Salary            373 non-null    float64
dtypes: float64(3), object(3)
memory usage: 17.7+ KB
```

```
In [8]: saldf.isnull().sum()
```

```
Out[8]:
```

	0
Age	2
Gender	2
Education Level	2
Job Title	2
Years of Experience	2
Salary	2

dtype: int64

```
In [13]: saldf.dropna(inplace=True, how='all')
```

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

```
Out[14]:
```

	0
Age	0
Gender	0
Education Level	0
Job Title	0
Years of Experience	0
Salary	0

dtype: int64

```
In [ ]:
```

```
In [20]: inp=saldf[['Years of Experience']]
out=saldf['Salary']
```

```
In [21]: LR=LinearRegression()
```

```
In [22]: LR.fit(inp,out)
```

```
Out[22]:
```

▼ LinearRegression ⓘ ⓘ

LinearRegression()

```
In [23]: LR.predict([[5]])
```

```
/usr/local/lib/python3.12/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not have valid feature names, but LinearRegression was fitted with feature names
  warnings.warn(
```

```
Out[23]: array([66143.76948947])
```