```
import pandas as pd
from sklearn.linear_model import LinearRegression


sal=pd.read_csv("/content/Salary Data.csv")


sal.head()
```

	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

Next steps: Generate code with sal | View recommended plots | New interactive sheet

```
LR=LinearRegression()


sal.isnull().sum()
```

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

dtype: int64

```
sal.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
```

```
sal.dropna(inplace=True)


sal.isnull().sum()
```

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

dtype: int64

```
sal.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 373 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: 20.4+ KB
```

```
a=sal[['Years of Experience']]
b=sal['Salary']
```

```
LR.fit(a,b)
```

```
LinearRegression()
```

```
LR.predict([[3]])
```

```
/usr/local/lib/python3.11/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not have valid feature names, but LinearRe
  warnings.warn(
array([52454.7484605])
```

```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
```

```
sal['gen']=le.fit_transform(sal['Gender'])
```

```
sal.head()
```

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

Next steps: Generate code with sal | View recommended plots | New interactive sheet

```
le1=LabelEncoder()
sal['Edu']=le1.fit_transform(sal['Education Level'])
sal.head()
```

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

Next steps: Generate code with sal | View recommended plots | New interactive sheet

```
le2=LabelEncoder()
sal['Job']=le2.fit_transform(sal['Job Title'])
sal.head()
```

	Age	Gender	Education Level	Job Title	Years of Experience	Salary	gen	Edu	Job
0	32.0	Male	Bachelor's	Software Engineer	5.0	90000.0	1	0	159
1	28.0	Female	Master's	Data Analyst	3.0	65000.0	0	1	17
2	45.0	Male	PhD	Senior Manager	15.0	150000.0	1	2	130
3	36.0	Female	Bachelor's	Sales Associate	7.0	60000.0	0	0	101
4	52.0	Male	Master's	Director	20.0	200000.0	1	1	22

Next steps: Generate code with sal | View recommended plots | New interactive sheet

```
c=sal[['Age','gen','Edu','Job','Years of Experience']]
d=sal['Salary']


LR1=LinearRegression()


LR1.fit(c,d)
```

```
LinearRegression
LinearRegression()
```

```
LR1.predict([[25,1,1,1,3]])
```

```
array([53136.31554386])
```