

```

import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

saldf=pd.read_csv("Salary_Data.csv")

saldf.head()

```

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0

```

saldf.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 2 columns):
 #   Column          Non-Null Count  Dtype  
---  -
 0   YearsExperience  30 non-null    float64
 1   Salary          30 non-null    float64
dtypes: float64(2)
memory usage: 612.0 bytes

saldf.head()

```

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0

```

saldf.duplicated().sum()

0

saldf.isnull().sum()

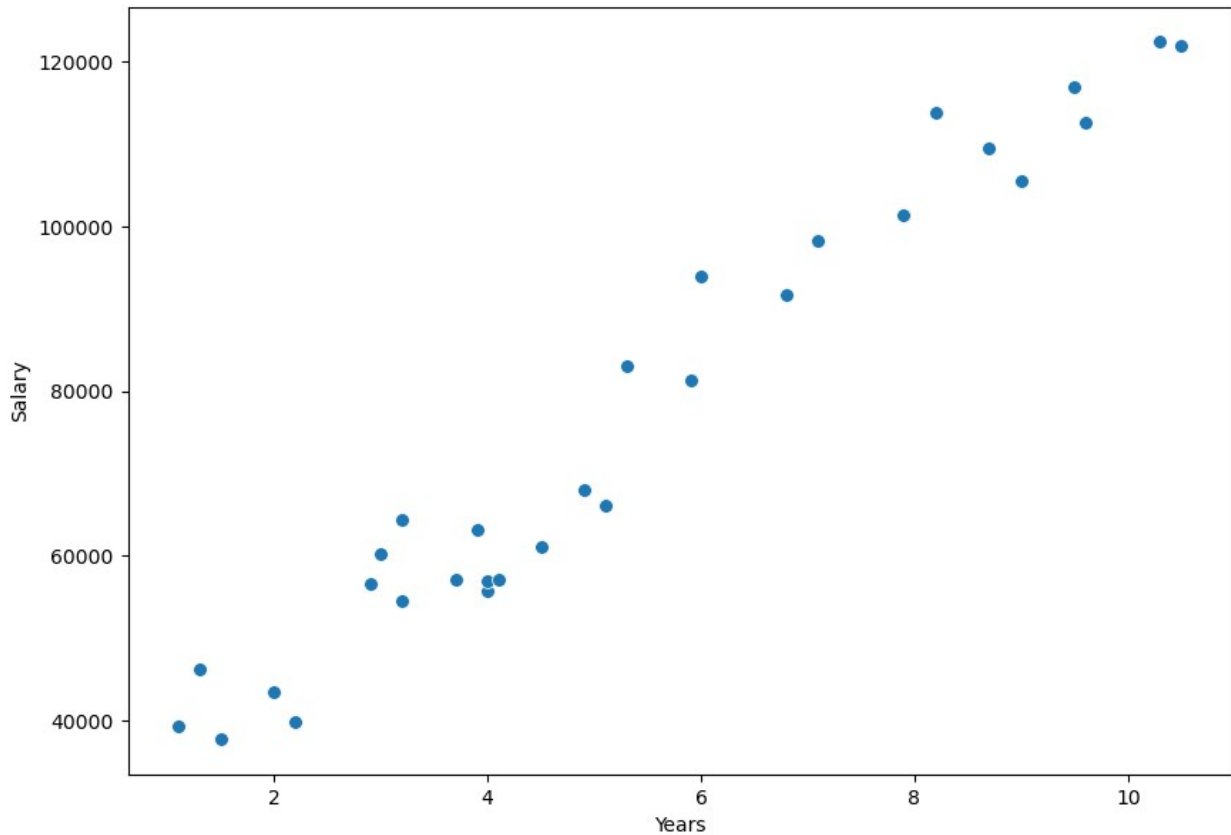
YearsExperience    0
Salary            0
dtype: int64

saldf.columns

Index(['YearsExperience', 'Salary'], dtype='object')

```

```
plt.figure(figsize=(10,7))
sns.scatterplot(data=saldf,x='YearsExperience',y='Salary',s=50)
plt.xlabel("Years")
plt.ylabel("Salary")
plt.show()
```



```
x=saldf.drop('Salary',axis=1)
y=saldf["Salary"]

from sklearn.model_selection import train_test_split
xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.3,random_state=42)

from sklearn.linear_model import LinearRegression
model=LinearRegression()

model.fit(xtrain,ytrain)

LinearRegression()

ypred=model.predict(xtest)

from sklearn.metrics import r2_score
```

```
rscore=r2_score(ytest,ypred)
```

```
print(rscore*100,"%")
```

```
94.14466227178214 %
```

```
model.score(xtest,ytest)
```

```
0.9414466227178214
```