

```
In [4]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [5]: df=pd.read_csv('C://Users//ALWAYS RAMESH//Downloads//data.csv')
```

```
In [6]: df
```

```
Out[6]:
```

	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0
...
395	15691863	Female	46	41000	1
396	15706071	Male	51	23000	1
397	15654296	Female	50	20000	1
398	15755018	Male	36	33000	0
399	15594041	Female	49	36000	1

400 rows × 5 columns

```
In [7]: from sklearn.preprocessing import LabelEncoder
```

```
In [8]: le=LabelEncoder()
df['Gender']=le.fit_transform(df['Gender'])
```

```
In [9]: df.head(10)
```

Out[9]:

	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	1	19	19000	0
1	15810944	1	35	20000	0
2	15668575	0	26	43000	0
3	15603246	0	27	57000	0
4	15804002	1	19	76000	0
5	15728773	1	27	58000	0
6	15598044	0	27	84000	0
7	15694829	0	32	150000	1
8	15600575	1	25	33000	0
9	15727311	0	35	65000	0

```
In [10]: x=df.drop('EstimatedSalary',axis=1)
y=df['EstimatedSalary']
```

In [11]: x

Out[11]:

	User ID	Gender	Age	Purchased
0	15624510	1	19	0
1	15810944	1	35	0
2	15668575	0	26	0
3	15603246	0	27	0
4	15804002	1	19	0
...
395	15691863	0	46	1
396	15706071	1	51	1
397	15654296	0	50	1
398	15755018	1	36	0
399	15594041	0	49	1

400 rows × 4 columns

In [12]: y

```
Out[12]: 0      19000
         1      20000
         2      43000
         3      57000
         4      76000
         ...
        395     41000
        396     23000
        397     20000
        398     33000
        399     36000
        Name: EstimatedSalary, Length: 400, dtype: int64
```

```
In [18]: from sklearn.model_selection import train_test_split
```

```
In [19]: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=10)
```

```
In [21]: from sklearn.linear_model import LinearRegression
         model=LinearRegression()
         model.fit(x_train,y_train)
```

```
Out[21]: ▼ LinearRegression ⓘ ?
         LinearRegression()
```

```
In [22]: y_pred=model.predict(x_test)
```

```
In [23]: y_pred
```

```
Out[23]: array([58880.92644655, 57972.04066751, 84297.14962066, 59057.06888871,
68287.10713707, 85661.94885965, 55439.72173099, 50441.41954492,
54616.78060781, 57469.78980942, 56207.17631319, 68038.5060285 ,
81313.88566981, 83497.58515484, 85931.71665072, 58759.11916563,
61599.29322097, 63182.39602768, 64285.20801069, 86745.09426005,
66850.75962289, 68028.18950753, 52723.4991889 , 79039.23997931,
77826.89793056, 60879.0543755 , 58844.89736892, 85997.82954195,
89794.84212181, 60421.88285554, 63293.54134908, 58868.19346423,
54734.05924768, 88754.56617635, 82192.63578836, 55559.23423372,
84415.64510775, 94589.89001089, 52248.76775422, 59053.61462198,
50577.73549028, 90408.98437851, 61467.23488657, 59692.62688938,
61276.52988622, 63595.18044884, 91396.46687793, 56606.77089772,
63957.26404865, 65910.52030482, 87063.95919055, 95807.7627693 ,
90545.45879515, 52761.53596038, 62401.01026714, 60302.55859063,
80507.06837504, 64718.45196381, 81764.18216287, 82397.1264876 ,
60824.42182672, 88228.49338032, 56921.42412973, 88784.5030566 ,
87203.19458322, 53563.97962042, 51758.06740575, 81367.91889137,
61317.31126862, 59296.30625917, 59428.6147981 , 84812.17467384,
65608.61837542, 56112.13074886, 64707.93347269, 54086.86457182,
55452.78632784, 68058.03579758, 47407.31932143, 84522.13663999,
86219.84222473, 58200.53114763, 66883.4492787 , 64237.43557761,
61072.87371584, 56424.43075679, 58418.65795958, 60442.98985741,
64677.53841171, 59764.46803578, 85396.67397287, 57274.88437346,
58695.00589528, 52000.48244516, 58014.99168 , 66673.06055432,
80280.01744619, 62382.59385494, 65306.65212158, 55506.32822886,
65713.69334056, 60016.56001584, 57001.33986129, 76365.23960561,
52766.30800355, 84096.15677869, 61116.78914081, 66554.3114536 ,
88512.54261546, 60915.97191155, 58439.69621237, 58469.26713937,
57250.35534085, 64554.29627921, 80751.97778502, 82376.41192392,
54106.51602405, 75794.1161508 , 65582.64181027, 57504.62502398])
```

```
In [24]: from sklearn.metrics import r2_score,mean_absolute_error,mean_squared_error
```

```
In [25]: MAE=mean_absolute_error(y_test,y_pred)
```

```
In [26]: MAE
```

```
Out[26]: 24977.353706441412
```

```
In [27]: MSE=mean_squared_error(y_test,y_pred)
```

```
In [28]: MSE
```

```
Out[28]: 931888758.9677299
```

```
In [29]: RMSE=np.sqrt(MSE)
```

```
In [30]: RMSE
```

```
Out[30]: 30526.853079997123
```

```
In [31]: y_pred2=model.predict([[15624510,1,19,0]])
```

```
C:\ProgramData\anaconda3\Lib\site-packages\sklearn\base.py:493: UserWarning: X does not have valid feature names, but LinearRegression was fitted with feature names
warnings.warn(
```

```
In [32]: y_pred2
```

```
Out[32]: array([60833.14493315])
```

```
In [ ]:
```

```
In [ ]: # multicollinearity
#
```

```
In [ ]:
```