project-1-622

February 21, 2024

[ ]:
```
## LINEAR REGRESSION

## PROJECT - 1

## Importing pandas
# loading the file "Linear_regr_Salary_dataset.csv"

import pandas as pd
data = pd.read_csv("/content/Linear_regr_Salary_dataset.csv")
data
```

[ ]:

	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
10	10	4.0	63219.0
11	11	4.1	55795.0
12	12	4.1	56958.0
13	13	4.2	57082.0
14	14	4.6	61112.0
15	15	5.0	67939.0
16	16	5.2	66030.0
17	17	5.4	83089.0
18	18	6.0	81364.0
19	19	6.1	93941.0
20	20	6.9	91739.0
21	21	7.2	98274.0
22	22	8.0	101303.0
23	23	8.3	113813.0
24	24	8.8	109432.0

```
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
```

[ ]:
```
## Printing the size of the csv file

data.shape
```

[ ]:
```
(30, 3)
```

[ ]:
```
## checking  there are any null values in the dataframe Or not

data.isnull().sum()
```

[ ]:
```
Unnamed: 0         0
YearsExperience    0
Salary             0
dtype: int64
```

[ ]:
```
## storing values in the x and y variables

x = data[['YearsExperience']]
y = data[['Salary']]
```

[ ]:
```
# Training the content

from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size = 0.4,␣
 ↪random_state=58)
from sklearn.linear_model import LinearRegression

model = LinearRegression()
model
```

[ ]:
```
LinearRegression()
```

[ ]:
```
## Fitting the model that we created

model.fit(x_train,y_train)
```

[ ]:
```
LinearRegression()
```

[ ]:
```
# Trainded x variable values

x_train
```

[ ]:

	YearsExperience
12	4.1
26	9.6
27	9.7
19	6.1
2	1.6
10	4.0
6	3.1
28	10.4
5	3.0
1	1.4
20	6.9
29	10.6
21	7.2
14	4.6
25	9.1
23	8.3
0	1.2
3	2.1

[ ]:
```
# Predicting the x testing value and stroring

y_pred = model.predict(x_test)
y_pred
```

[ ]:
```
array([[ 65202.97736453],
       [ 77154.89276419],
       [ 75316.13654886],
       [ 57847.95250321],
       [ 66122.3554722 ],
       [ 73477.38033353],
       [101058.7235635 ],
       [ 82671.16141018],
       [ 62444.84304154],
       [108413.74842482],
       [ 57847.95250321],
       [ 48654.17142655]])
```

[ ]:
```
## Printing the size of the y_pred variable
y_pred.shape
```

[ ]:
```
(12, 1)
```

[ ]:
```
## Printing the column values of the y_test
y_test
```

[ ]:
```
      Salary
11   55795.0
17   83089.0
16   66030.0
7    54446.0
13   57082.0
15   67939.0
22  101303.0
18   81364.0
9    57190.0
24  109432.0
8    64446.0
4    39892.0
```

[ ]:
```
## Finding the accuarcy of the trained data

# importing numpy module

import numpy as np
from sklearn.metrics import accuracy_score
accuracy = accuracy_score(y_test,np.round(y_pred))
accuracy
```

[ ]:
```
0.0
```

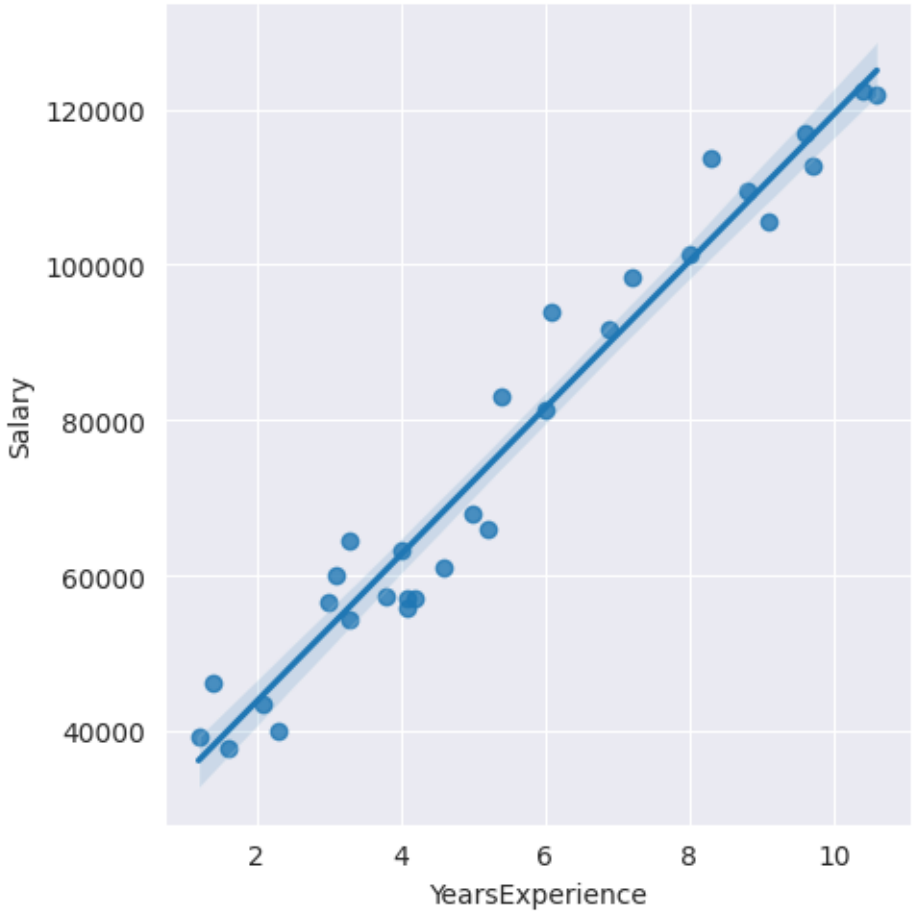
[ ]:
```
## Finding the prediction value of the column of the dataframe
inputdata = [[4.5]]
prediction = model.predict(inputdata)
np.round(y_pred)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning: X does
not have valid feature names, but LinearRegression was fitted with feature names
  warnings.warn(
```

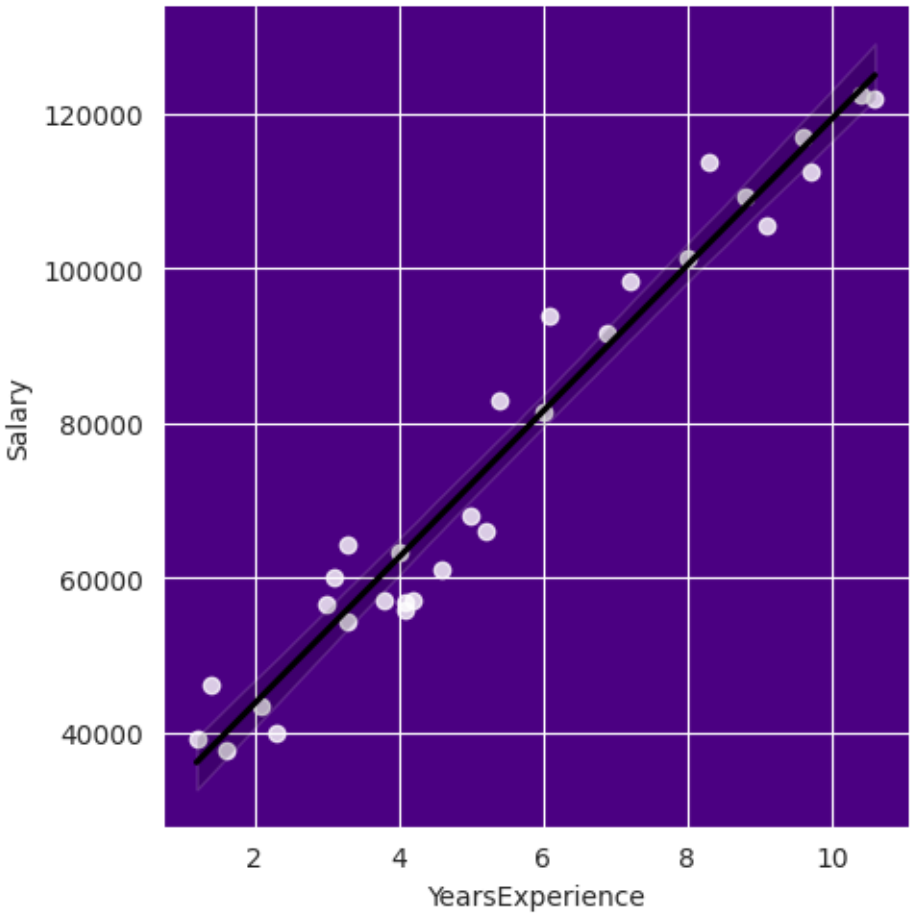
[ ]:
```
array([[ 65203.],
       [ 77155.],
       [ 75316.],
       [ 57848.],
       [ 66122.],
       [ 73477.],
       [101059.],
       [ 82671.],
       [ 62445.],
       [108414.],
       [ 57848.],
       [ 48654.]])
```

[ ]:
```
## Printing the "linear mean plot"

import seaborn as sns
sns.lmplot(x='YearsExperience',y='Salary',data = data)
```

[ ]:
```
<seaborn.axisgrid.FacetGrid at 0x7955384c0460>
```

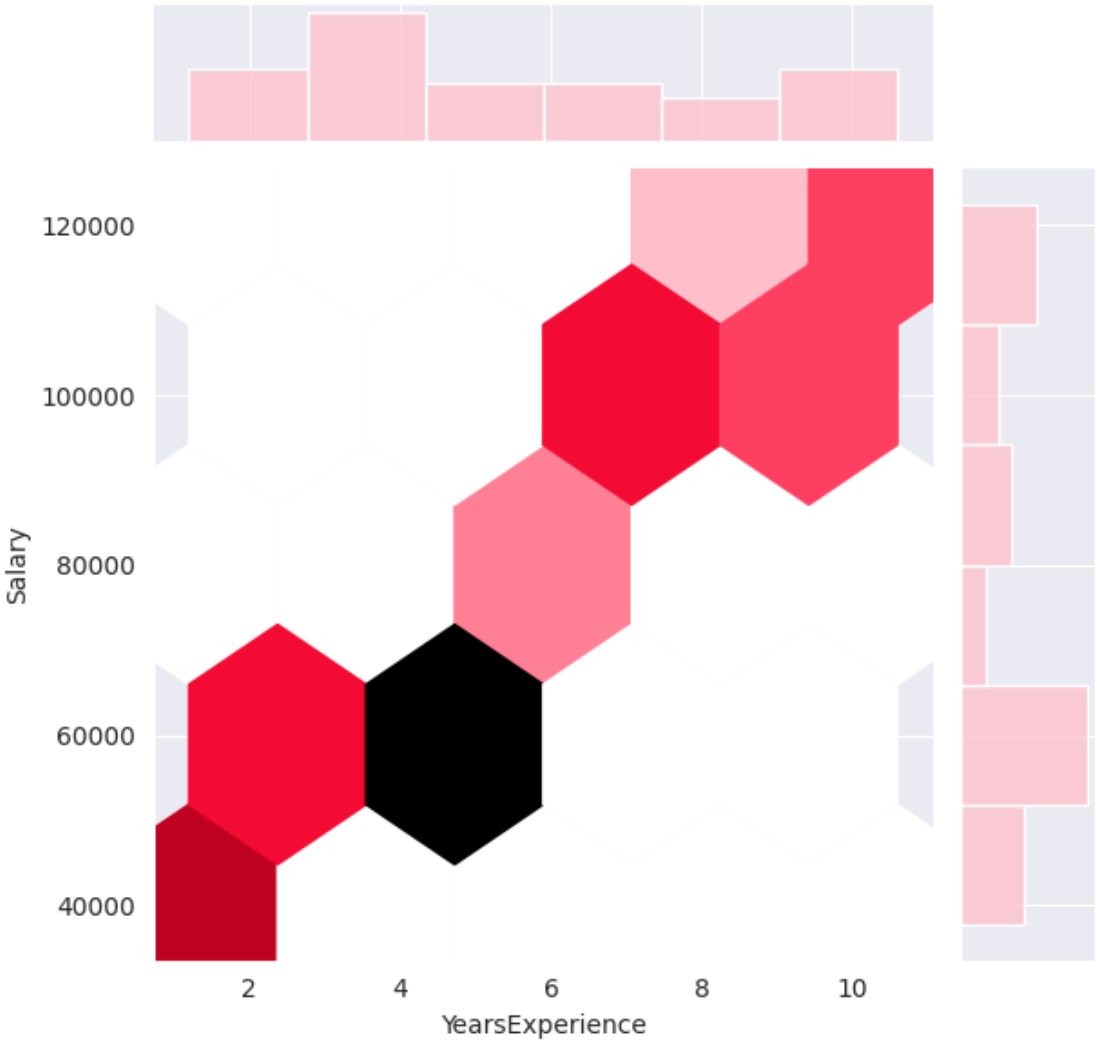


[ ]:
```
# Adding grid and facecolour to the lm plot

sns.lmplot(x='YearsExperience',y='Salary',data = data,scatter_kws={'color':
 ↪'whi'},line_kws={'color':'blue'})
sns.set_style('darkgrid')
ax = plt.gca()

plt.gca().set_facecolor('indigo')
```



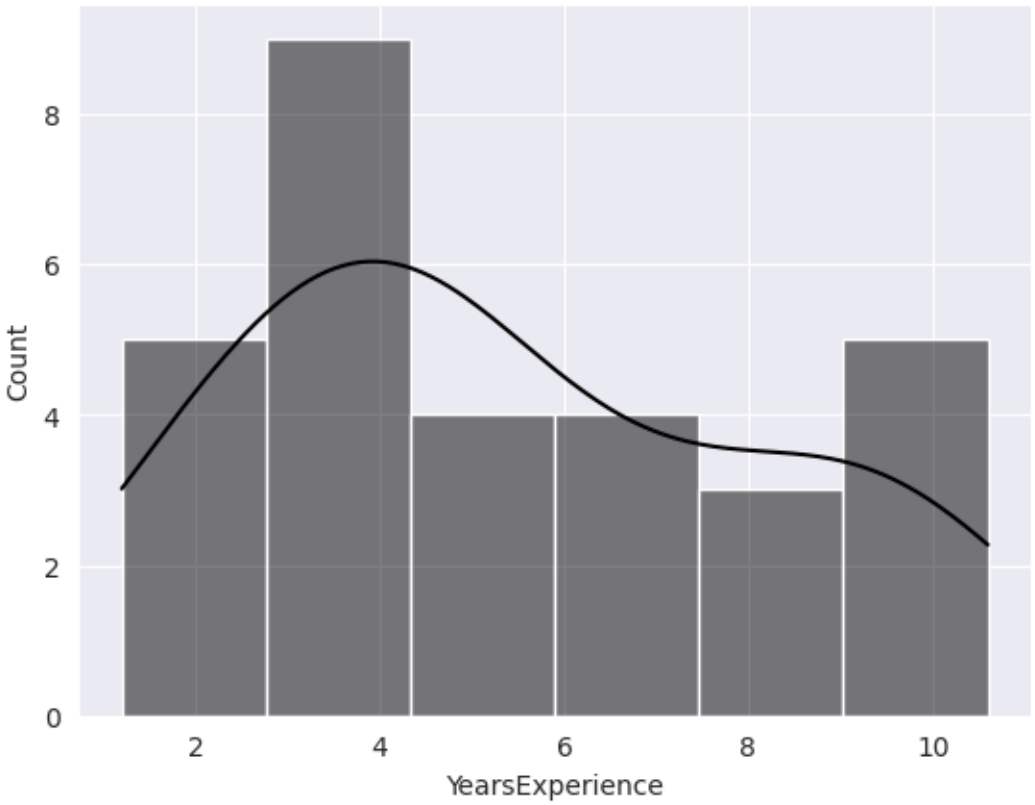
[ ]:
```
## plotting the jointplot graph

sns.jointplot(x='YearsExperience',y='Salary',data =␣
↪data,kind='hex',color='pink')
```

[ ]:
```
<seaborn.axisgrid.JointGrid at 0x79553adef670>
```



[ ]:
```
import seaborn as sns
sns.histplot(x= 'YearsExperience',y = 'Salary',color = 'orange',kde=␣
↪True,data=data)
```

[ ]: <Axes: xlabel='YearsExperience', ylabel='Count'>


Count
0
2
4
6
8
2
4
6
8
10
YearsExperience

[ ]: