In [ ]:
```
#Assignment 1
#Simple linear regress
# Amit pawar
# PRN 25070126501
```

In [13]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [3]:
```
df = pd.read_fwf 'data.csv'
, colspecs = 'infer' , header = None)
df.head()
```

Out[3]:

	0
0	YearsExperience,Salary
1	1.1,39343.00
2	1.3,46205.00
3	1.5,37731.00
4	2.0,43525.00

In [15]:
```
df = pd.read_csv('data.csv')
print(df.columns)
```

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

In [16]:
```
print(df.shape)
```

```
(30, 2)
```

In [17]:
```
df.columns = ['YearsExperience' , 'Salary']
```

In [6]:
```
print(df.head())
```

```
                         0
,0  YearsExperience,Salary
,1            1.1,39343.00
,2            1.3,46205.00
,3            1.5,37731.00
,4            2.0,43525.00
```

In [18]:
```
df.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
```

In [19]:
```
df.describe()
```

Out[19]:

	YearsExperience	Salary
count	30.000000	30.000000
mean	5.313333	76003.000000
std	2.837888	27414.429785
min	1.100000	37731.000000
25%	3.200000	56720.750000
50%	4.700000	65237.000000
75%	7.700000	100544.750000
max	10.500000	122391.000000

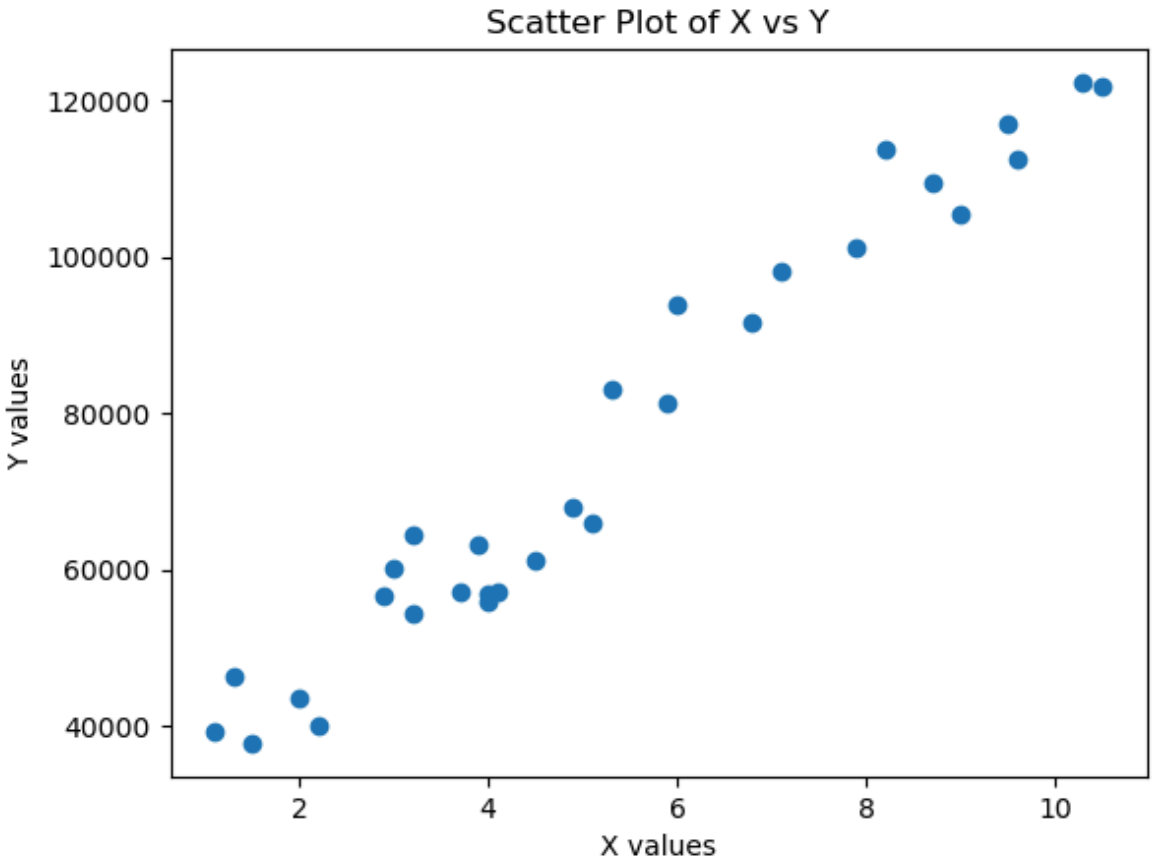
In [20]:
```
x = df['YearsExperience'].values
y = df['Salary'].values
```

In [21]:
```
x
```

Out[21]:
```
array([ 1.1,  1.3,  1.5,  2. ,  2.2,  2.9,  3. ,  3.2,  3.2,  3.7,  3.9,
        4. ,  4. ,  4.1,  4.5,  4.9,  5.1,  5.3,  5.9,  6. ,  6.8,  7.1,
        7.9,  8.2,  8.7,  9. ,  9.5,  9.6, 10.3, 10.5])
```

In [22]:
```
plt.scatter(x, y)
plt.xlabel('X values')
plt.ylabel('Y values')
plt.title('Scatter Plot of X vs Y')

plt.show()
```



In [39]:
```
print(x.shape)
print(y.shape)
```

```
(30, 1)
,(30, 1)
```

In [40]:
```
x=x.reshape(-1,1)
y=y.reshape(-1,1)
```

In [41]:
```
print(x.shape)
print(y.shape)
```

```
(30, 1)
,(30, 1)
```

In [42]:
```
from sklearn.model_selection import train_test_split
```

In [43]:
```
X_train,X_test,y_train,y_test = train_test_split(x, y, test_size=0.20,random_sta
```

In [44]:
```
print(X_train.shape)
print(X_train.shape)
print(y_train.shape)
print(y_train.shape)
```

```
(24, 1)
,(24, 1)
,(24, 1)
,(24, 1)
```

In [45]:
```
from sklearn.linear_model import LinearRegression
ln = LinearRegression()
```

In [46]:
```
ln.fit(X_train,y_train)
```

Out[46]:

▾ LinearRegression

```
LinearRegression()
```

In [47]:
```
y_pred=ln.predict(X_test)
```

In [48]:
```
a=ln.coef_
b=ln.intercept_
print("Estimated Model slope, a:", a)
print("Estimated model intercept, b:", b)
```

```
Estimated Model slope, a: [[9423.81532303]]
,Estimated model intercept, b: [25321.58301178]
```

In [49]:
```
from sklearn.metrics import mean_squared_error
```

In [51]:
```
from sklearn.metrics import r2_score
print("R2 Score value: {:.2f}".format(r2_score(y_test, y_pred)))
```

```
R2 Score value: 0.90
```

In [55]:
```
plt.scatter(x, y, color='blue', label="Scatter Plot.")
plt.plot(X_test, y_pred, color='black', linewidth=2, label='Regression Line')
plt.title("Relationship Between Years of Experience and Salary")
```

```
plt.xlabel('YearsExperience')
plt.ylabel('Salary')
plt.legend()
plt.show()
```