PRACTICAL 3

Aim:To build the model for prediction of profit

Imported libraries

In [21]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score,mean_squared_error
from sklearn.linear_model import LinearRegression
```

Reading Dataset

In [22]:
```
data = pd.read_csv(r"practical2.csv")
```

In [23]:
```
data.head()
```

Out[23]:

	R&D Spend	Administration	Marketing Spend	State	Profit
0	165349.20	136897.80	471784.10	New York	192261.83
1	162597.70	151377.59	443898.53	California	191792.06
2	153441.51	101145.55	407934.54	Florida	191050.39
3	144372.41	118671.85	383199.62	New York	182901.99
4	142107.34	91391.77	366168.42	Florida	166187.94

Data Preprocessing

In [24]:
```
data.isnull()
```

Out[24]:

	R&D Spend	Administration	Marketing Spend	State	Profit
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
5	False	False	False	False	False
6	False	False	False	False	False
7	False	False	False	False	False
8	False	False	False	False	False
9	False	False	False	False	False
10	False	False	False	False	False
11	False	False	False	False	False
12	False	False	False	False	False
13	False	False	False	False	False
14	False	False	False	False	False
15	False	False	False	False	False
16	False	False	False	False	False
17	False	False	False	False	False
18	False	False	False	False	False
19	False	False	False	False	False
20	False	False	False	False	False
21	False	False	False	False	False
22	False	False	False	False	False
23	False	False	False	False	False
24	False	False	False	False	False
25	False	False	False	False	False
26	False	False	False	False	False
27	False	False	False	False	False
28	False	False	False	False	False
29	False	False	False	False	False
30	False	False	False	False	False
31	False	False	False	False	False
32	False	False	False	False	False
33	False	False	False	False	False
34	False	False	False	False	False
35	False	False	False	False	False

	R&D Spend	Administration	Marketing Spend	State	Profit
36	False	False	False	False	False
37	False	False	False	False	False
38	False	False	False	False	False
39	False	False	False	False	False
40	False	False	False	False	False
41	False	False	False	False	False
42	False	False	False	False	False
43	False	False	False	False	False
44	False	False	False	False	False
45	False	False	False	False	False
46	False	False	False	False	False
47	False	False	False	False	False
48	False	False	False	False	False
49	False	False	False	False	False

In [30]:
```
column_to_remove = 'State'
df = data.drop(column_to_remove, axis=1)
```

Making data Ready for regression model

In [32]:
```
X = df.drop('Profit', axis=1)  # Exclude the target column
y = df['Profit']
```

Splitting the Dataset

In [33]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_sta
```

Training The Model

In [34]:
```
model = LinearRegression()
model.fit(X_train, y_train)
```

Out[34]:
```
▾ LinearRegression
LinearRegression()
```

Making predictions based on testing data

In [35]:
```
y_pred = model.predict(X_test)
```

In [39]:
```
print(y_pred-y_test)
diff=y_pred-y_test
```

```
13    -7604.322835
39     3888.990816
30    -1044.171840
45   -18424.371850
17     3758.027344
48    15319.284863
26     3283.013658
25    -6525.875855
32      272.756386
19    -9670.707078
Name: Profit, dtype: float64
```

In [40]:
```
print(diff**2)
```

```
13    5.782573e+07
39    1.512425e+07
30    1.090295e+06
45    3.394575e+08
17    1.412277e+07
48    2.346805e+08
26    1.077818e+07
25    4.258706e+07
32    7.439605e+04
19    9.352258e+07
Name: Profit, dtype: float64
```

Evaluting the model

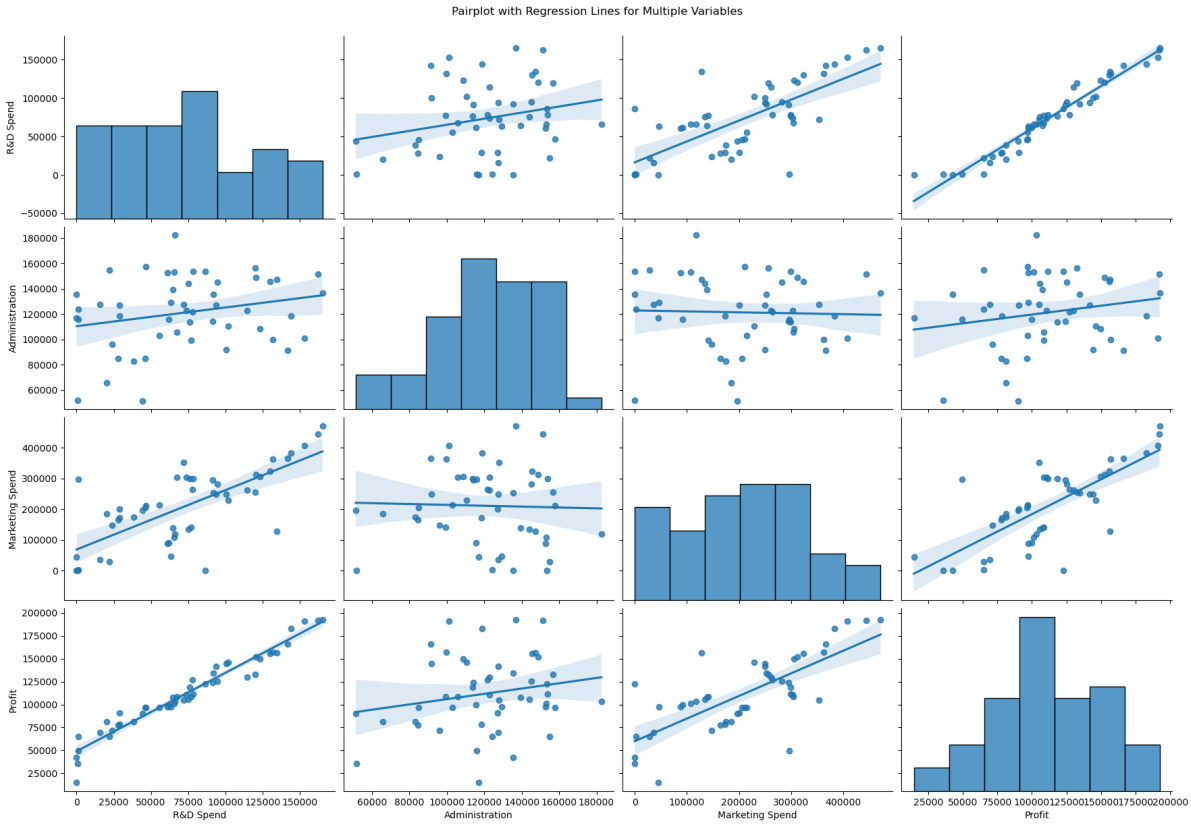
In [36]:
```
mse = mean_squared_error(y_test, y_pred)
print(f'Mean Squared Error: {mse}')
```

```
Mean Squared Error: 80926321.22295167
```

Visualizing the Conclusion

In [48]:
```
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt
```

In [49]:
```
sns.pairplot(df, kind="reg", height=3, aspect=1.5)
plt.suptitle('Pairplot with Regression Lines for Multiple Variables', y=1.02)
plt.show()
```


Pairplot with Regression Lines for Multiple Variables
R&D Spend
Administration
Marketing Spend
Profit
0 25000 50000 75000 100000 125000 150000
60000 80000 100000 120000 140000 160000 180000
0 100000 200000 300000 400000
25000 50000 75000 100000 125000 150000 175000 200000