Profit_Analysis

REGRESSION AND VISUALIZATION BASED ANALYSIS

Name :Vinay Kumar Jaiswal
Team ID : PTID-CDA-MAY-25-491
Project ID : PRDA-01

CONTENTS

1. Objective
2. Data Collection
3. Data Preperation
4. Data Visualization Dashboard
5. Correlation Analysis
6. Regression Model
7. Prediction
8. Conclusion & Recommendation

OBJECTIVE

- To analyze how different type of company expenditures RD Spend, Administration and Marketing Spend impact the Profit of the company.

- Perform a regression analysis to quantify the relationship between expenditures and profit.

- Use the model to predict profit for new scenarios based on input values.(**RD_Spend   Administration   Marketing   Profit**)

RD_Spend	Administration	Marketing	Profit
21892.92	81910.77	164270.7	?
23940.93	96489.63	137001.1	?

- Visualize the data using Power BI to uncover trends and insights across different states.

- Provide data-driven recommendations to guide future spending strategies and maximize profitability.

DATA COLLECTION

You've uploaded a file named (project_profit_analy - sis.csv) , which appears to contain the dataset. I'll begin by examining its contents.

Let me now inspect the file .

The dataset contains the following features:

- RD_Spend: Money spent on R&D

- Administration: Money spent on administration

- Marketing_Spend: Money spent on marketingState: Location of the business unit

- Profit: Target variable

DATA PREPARATION

- Import the CSV file using Pandas.

```
[3]: import pandas as pd

[21]: data = pd.read_csv("C:/Users/ramvi/Desktop/project_Profit_analysis.csv")
      print(data)
```

	RD_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
5	131876.90	99814.71	362861.36	New York	156991.12
6	134615.46	147198.87	127716.82	California	156122.51
7	130298.13	145530.06	323876.68	Florida	155752.60
8	120542.52	148718.95	311613.29	New York	152211.77
9	123334.88	108679.17	304981.62	California	149759.96
10	101913.08	110594.11	229160.95	Florida	146121.95
11	100671.96	91790.61	249744.55	California	144259.40
12	93863.75	127320.38	249839.44	Florida	141585.52
13	91992.39	135495.07	252664.93	California	134307.35
14	119943.24	156547.42	256512.92	Florida	132602.65
15	114523.61	122616.84	261776.23	New York	129917.04
16	78013.11	121597.55	264346.06	California	126992.93
17	94657.16	145077.58	282574.31	New York	125370.37
18	91749.16	114175.79	294919.57	Florida	124266.90
19	86419.70	153514.11	0.00	New York	122776.86
20	76253.86	113867.30	298664.47	California	118474.03
21	78389.47	153773.43	299737.29	New York	111313.02
22	73994.56	122782.75	303319.26	Florida	110352.25
23	67532.53	105751.03	304768.73	Florida	108733.99
24	77044.01	99281.34	140574.81	New York	108552.04
25	64664.71	139553.16	137962.62	California	107404.34
26	75328.87	144135.98	134050.07	Florida	105733.54
27	72107.60	127864.55	353183.81	New York	105008.31
28	66051.52	182645.56	118148.20	Florida	103282.38
29	65605.48	153032.06	107138.38	New York	101004.64
30	61994.48	115641.28	91131.24	Florida	99937.59
31	61136.38	152701.92	88218.23	New York	97483.56
32	63408.86	129219.61	46085.25	California	97427.84
33	55493.95	103057.49	214634.81	Florida	96778.92
34	46426.07	157693.92	210797.67	California	96712.80
35	46014.02	85047.44	205517.64	New York	96479.51
36	28663.76	127056.21	201126.82	Florida	90708.19
37	44069.95	51283.14	197029.42	California	89949.14
38	20229.59	65947.93	185265.10	New York	81229.06
39	38558.51	82982.09	174999.30	California	81005.76
40	28754.33	118546.05	172795.67	California	78239.91
41	27892.92	84710.77	164470.71	Florida	77798.83
42	23640.93	96189.63	148001.11	California	71498.49
43	15505.73	127382.30	35534.17	New York	69758.98
44	22177.74	154806.14	28334.72	California	65200.33
45	1000.23	124153.04	1903.93	New York	64926.08
46	1315.46	115816.21	297114.46	Florida	49490.75
47	0.00	135426.92	0.00	California	42559.73
48	542.05	51743.15	0.00	New York	35673.41
49	0.00	116983.80	45173.06	California	14681.40

- Check Structure

```
[25]: print(data.info())

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 5 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   RD_Spend         50 non-null     float64
 1   Administration   50 non-null     float64
 2   Marketing_Spend  50 non-null     float64
 3   State            50 non-null     object
 4   Profit           50 non-null     float64
dtypes: float64(4), object(1)
memory usage: 2.1+ KB
None
```

```
[29]: print(data.describe())
```

	RD_Spend	Administration	Marketing_Spend	Profit
count	50.000000	50.000000	50.000000	50.000000
mean	73721.615600	121344.639600	211025.097800	112012.639200
std	45902.256482	28017.802755	122290.310726	40306.180338
min	0.000000	51283.140000	0.000000	14681.400000
25%	39936.370000	103730.875000	129300.132500	90138.902500
50%	73051.080000	122699.795000	212716.240000	107978.190000
75%	101602.800000	144842.180000	299469.085000	139765.977500
max	165349.200000	182645.560000	471784.100000	192261.830000

- Check For missing values

```
[33]: print(data.isnull().sum())

RD_Spend           0
Administration     0
Marketing_Spend    0
State              0
Profit             0
dtype: int64
```

DATA VISUALIZATION DASHBOARD
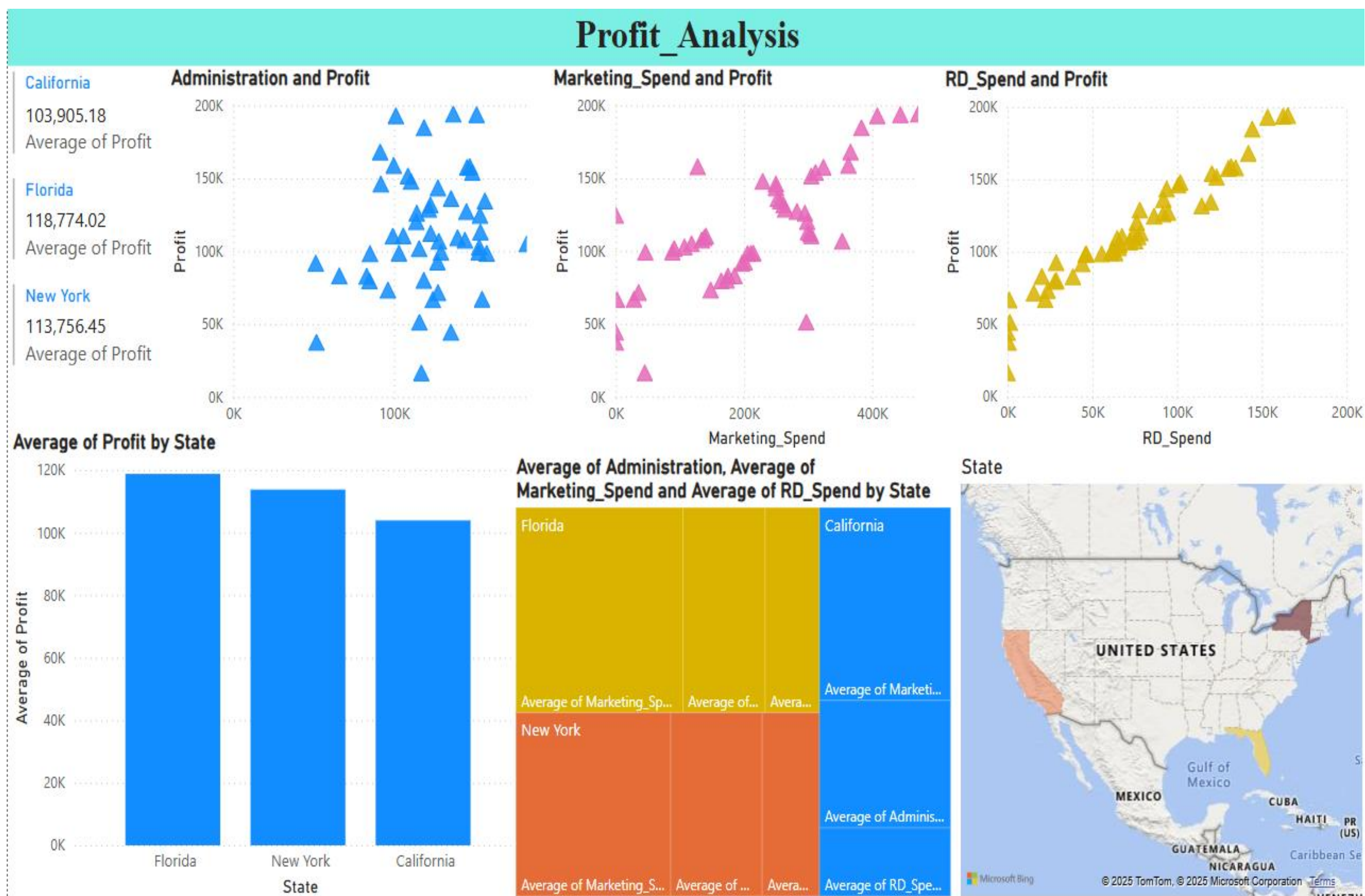
CORRELATION ANALYSIS

Correlation analysis is a statistical method used to measure and evaluate the strength and direction of the relationship between two or more variables.

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	50122.19	6572.353	7.626	0	36892.733	63351.65	36892.733	63351.65
RD_Spend	0.806	0.045	17.846	0	0.715	0.897	0.715	0.897
Administration	-0.027	0.051	-0.526	0.602	-0.13	0.076	-0.13	0.076
Marketing_Spend	0.027	0.016	1.655	0.105	-0.006	0.06	-0.006	0.06

- RD_Spend has strong correlation with Profit.

- Administration_Spend has weak/negative.

- Marketing_Spend is moderate/weak.

A **scatter plot** (also known as a **scatter chart**, **scatter graph**, or **scattergram**) is a type of graph that uses dots to represent the values for two different numerical variables.

Direction of Relationship
➢ **Positive Correlation**: If the dots generally trend upwards from left to right, it suggests that as one variable increases, the other tends to increase.
➢ **Negative Correlation**: If the dots generally trend downwards from left to right, it suggests that as one variable increases, the other tends to decrease.
➢ **No Correlation**: If the dots appear randomly scattered with no discernible pattern, it indicates little to no linear relationship between the variables.
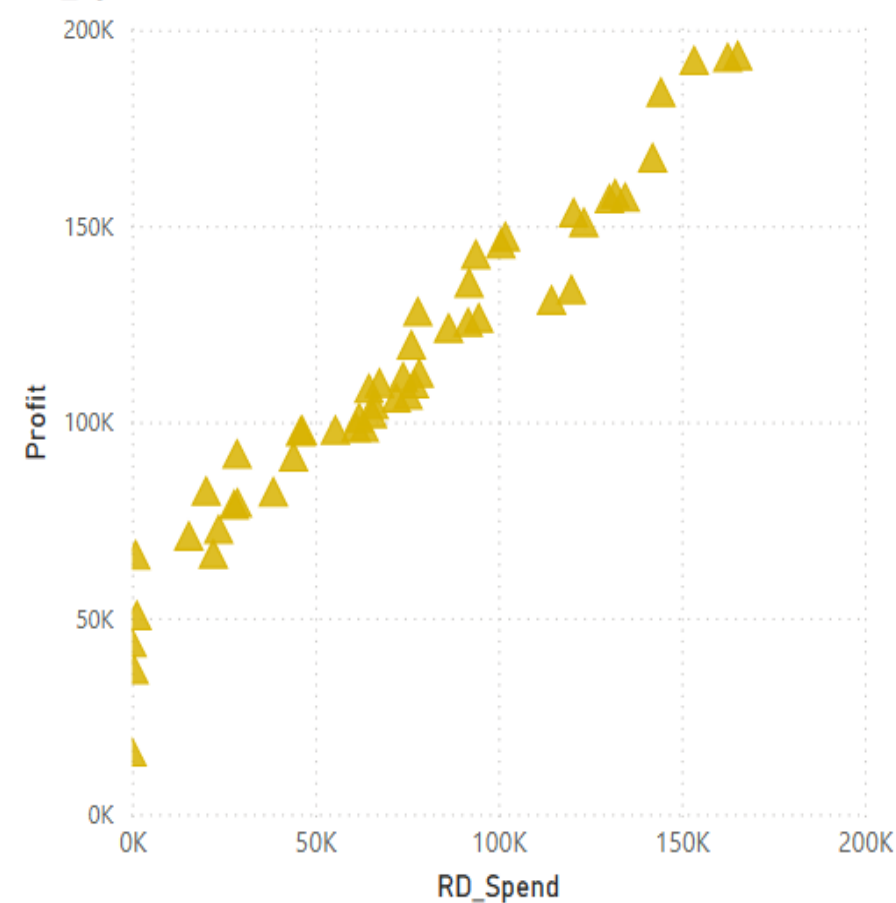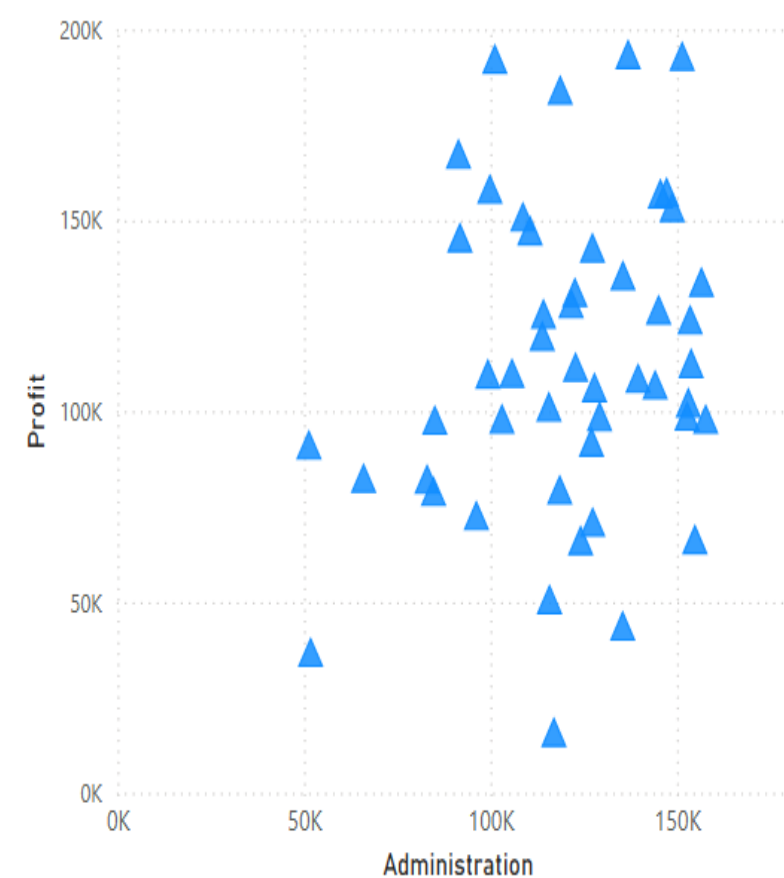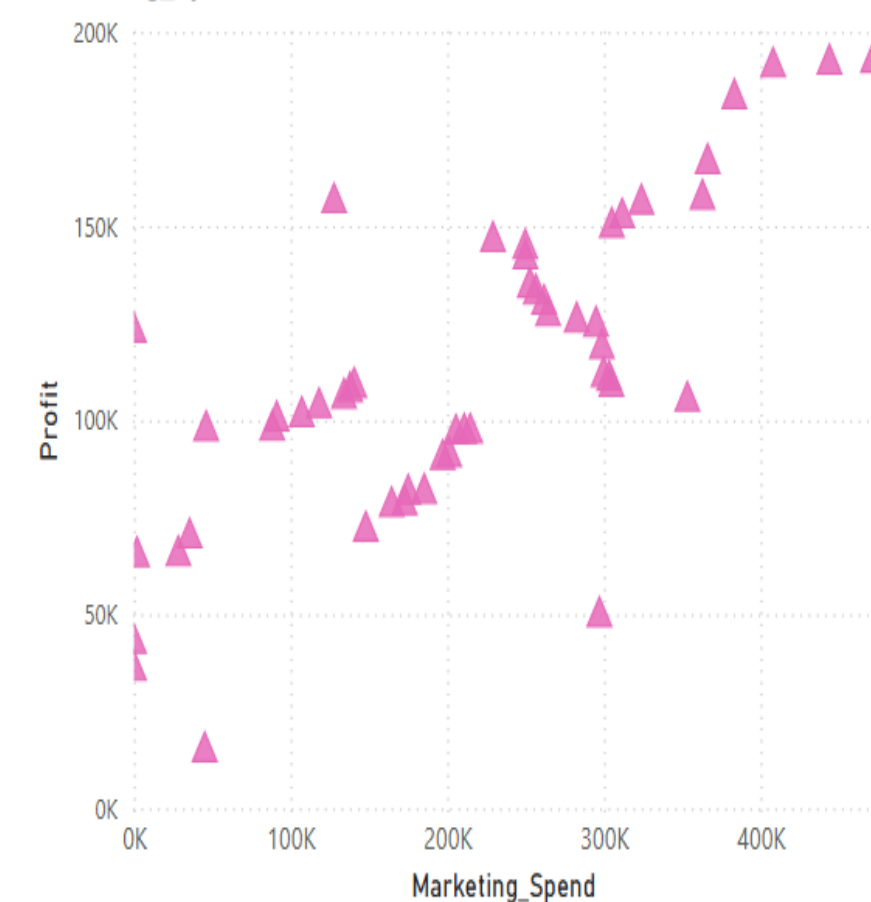
REGRESSION MODEL

Regression analysis is a powerful statistical method used to model and examine the relationship between a dependent variable (also called the outcome, response, or target variable) and one or more independent variables (also called predictor, explanatory, or regressor variables).

Regression Statistics					
Multiple R	0.975				
R Square	0.951				
Adjusted R Square	0.948				
Standard Error	9232.335				
Observations	50				
ANOVA					
	df	SS	MS	F	Significance F
Regression	3	75683964196	2.52E+10	295.978	0
Residual	46	3920856301	85236007		
Total	49	79604820497			

Regression Formula

Y = 0.806*RD_Spend +(- 0.027*Administration) + 0.027*Marketing_spend + 50122.193

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	50122.19	6572.353	7.626	0	36892.733	63351.65	36892.733	63351.65
RD_Spend	0.806	0.045	17.846	0	0.715	0.897	0.715	0.897
Administration	-0.027	0.051	-0.526	0.602	-0.13	0.076	-0.13	0.076
Marketing_Spend	0.027	0.016	1.655	0.105	-0.006	0.06	-0.006	0.06

PREDICTION

	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT								
2									
3	Regression Statistics								
4	Multiple R	0.975							
5	R Square	0.951							
6	Adjusted R Square	0.948							
7	Standard Error	9232.335							
8	Observations	50							
9									
10	ANOVA								
11		df	SS	MS	F	Significance F			
12	Regression	3	75683964196	2.52E+10	295.978	0			
13	Residual	46	3920856301	85236007					
14	Total	49	79604820497						
15	h								
16		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
17	Intercept	50122.19	6572.353	7.626	0	36892.733	63351.65	36892.733	63351.65
18	RD_Spend	0.806	0.045	17.846	0	0.715	0.897	0.715	0.897
19	Administration	-0.027	0.051	-0.526	0.602	-0.13	0.076	-0.13	0.076
20	Marketing_Spend	0.027	0.016	1.655	0.105	-0.006	0.06	-0.006	0.06

Formula

$Y = 0.806*RD\_Spend +(- 0.027*Administration) + 0.027*Marketing\_spend + 50122.193$

Predict the profit for the below-given input features.

RD_Spend	Administration	Marketing_Spend	PROFIT
21892.92	81910.77	164270.7	70037.91
23940.93	96489.63	137001.1	70554.57

Predict the profit.

RD_Spend	Administration	Marketing	Profit
21892.92	81910.77	164270.7	70037.91
23940.93	96489.63	137001.1	70554.57

CONCLUSION & RECOMMENDATION

1. RD Spend

❖ Strongest positive impact on profit

❖ Every additional dollar spent increases profit significantly

❖ Highly correlated with business success

2. Marketing Spend

❖ Slight positive impact

❖ Can boost profit, but less effective than R&D

❖ May need better targeting or analysis of ROI

3. Administration Spend

❖ No significant effect on profit

❖ Could be a candidate for cost reduction or optimization

4. State-wise Differences

❖ Minor differences in profit across states

❖ Could be due to operational or market differences