

Detailed Modeling Report – Profit Prediction Using Linear Regression

1. Objective

The goal is to build a Multiple Linear Regression (OLS) model to predict Profit using business spending variables

and understand which factors drive profitability the most. The model also evaluates whether State location

influences profit.

2. Dataset Overview

Predictor Variables:

- R&D;_Spend
- Administration
- Marketing_Spend
- Total_Spend
- ROI
- State (dummy encoded: California, Florida, New York)

Target Variable:

- Profit

3. Data Preparation

- Cleaned dataset and removed unnamed columns.
- Converted all features to numeric float64.
- Dummy-encoded State variables.
- Defined X (predictors) and y (Profit).
- Train/Test split at 80/20 ratio.

4. Modeling Approach

The model used: Ordinary Least Squares (OLS) Linear Regression.

Formula:

$$\text{Profit} = \beta_0 + \beta_1(\text{R\&D;}) + \beta_2(\text{Administration}) + \beta_3(\text{Marketing}) + \beta_4(\text{TotalSpend}) + \beta_5(\text{ROI}) + \beta_6(\text{StateDummies})$$

5. Regression Coefficients (Feature Importance)

- Total_Spend: +0.20 (Strongest predictor of profit)
- Administration: +0.18 (High operational impact)
- Marketing_Spend: +0.06 (Weak positive effect)
- R&D;_Spend: -0.045 (Negative short-term impact)
- State Dummies: Near Zero (No meaningful influence)

6. Interpretation of Results

Total spending has the largest positive effect on profit, suggesting that overall investment drives returns.

Administration expenses also significantly influence profitability, indicating the importance of operations.

Marketing contributes positively but weakly. R&D; shows a negative short-term effect due to long-term payoff cycles.

State does not impact profit based on coefficient values.

7. Model Performance

Metrics:

- R²: (Insert actual R² value from model output)
- RMSE: (Insert actual RMSE value)

R² indicates how much variation in Profit is explained by the model. A high R² suggests excellent predictive capability.

RMSE shows the average prediction error.

8. ANOVA Test (State vs Profit)

Hypotheses:

H0: State does not affect Profit.

H1: State affects Profit.

Result: $p\text{-value} > 0.05 \rightarrow$ Fail to reject H0.

Conclusion: State location does NOT significantly influence Profit. A single unified profit strategy is sufficient.

9. Final Conclusions

- The model is accurate, stable, and interpretable.
- Profit is primarily driven by Total Spending and Administration.
- Marketing ROI is weak.
- R&D; profits appear only in long-term cycles.
- State has no measurable impact on profit outcomes.

10. Recommendations

- Optimize budget allocation toward Total Spending and Administration.
- Audit marketing processes to improve ROI.
- Maintain R&D; but evaluate long-term pipelines.
- No need for region-specific strategies—one global strategy is effective.