**PRDA – 01 Profit Analysis**

Task:1

1. Get data from the database with the given credentials.
2. Perform Regression Analysis for the given data to identify how the money spent on Marketing, R&D, and Administration is affecting the company’s Profit. Predict the Profit for the below-given input features.

|  |  |  |  |
| --- | --- | --- | --- |
| **R&D Spend** | **Administration** | **Marketing Spend** | **Profit** |
| 21892.92 | 81910.77 | 164270.7 |  |
| 23940.93 | 96489.63 | 137001.1 |  |

Aim: To Perform Regression Analysis for the given data to identify how the money spent on Marketing, R&D, and Administration is affecting the company’s Profit. Predict the Profit for the below-given input features.

Procedure:

1. **Understand the Objective**:
   * Identify the impact of Marketing, R&D, and Administration expenses on Profit.
   * Predict Profit based on given inputs.
2. **Data Collection**:
   * Use the dataset provided (from Data mites Internship Team).
   * Extract data from SQL to Excel for analysis.
3. **Data Preparation**:
   * Clean the data (handle missing values, outliers, and inconsistencies).
   * Perform Exploratory Data Analysis (EDA) to understand the relationships between variables.
   * Standardize or normalize the data if needed.
4. **Choose a Regression Model**:
   * Use **Linear Regression** as the primary model to identify the relationship.
5. **Model Building**:
   * Split the data into **Training** and **Testing** datasets (e.g., 80:20 split).
   * Fit the Linear Regression model using Marketing, R&D, and Administration as independent variables and Profit as the dependent variable.
6. **Model Evaluation**:
   * Assess the model using metrics like **R²**, **Adjusted R²**, and **Mean Squared Error (MSE)**.
   * Check for multicollinearity using **VIF (Variance Inflation Factor)**.
7. **Prediction**:
   * Use the trained model to predict Profit for the given input features.
   * Validate predictions by comparing them with actual values (if available).
8. **Visualization and Reporting**:
   * Visualize the regression results (scatter plots, regression line, residual plots).
   * Create dashboards in Power BI to present findings effectively.
9. **Interpretation**:
   * Analyse the regression coefficients to determine the impact of each expense type on Profit.
   * Provide actionable insights based on the results.
10. **Documentation and Presentation**:

* Document the process, findings, and insights in a structured 30-slide PowerPoint presentation.
* Include visualizations, model outputs, and prediction results.

Summary Output:

|  |  |
| --- | --- |
| **Column1** | **Column2** |
| Regression Statistics | |
| Multiple R | 0.972787387 |
| R Square | 0.9463153 |
| Adjusted R Square | 0.94273632 |
| Standard Error | 9334.261142 |
| Observations | 49 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |
| **Column1** | **Column2** | **Column3** | **Column4** | **Column5** | **Column6** |
|  | df | SS | MS | F | Significance F |
| Regression | 3 | 6.911E+10 | 2.304E+10 | 264.40922 | 1.40E-28 |
| Residual | 45 | 3.921E+09 | 87128431 |  |  |
| Total | 48 | 7.303E+10 |  |  |  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Column1** | **Coefficients** | **Standard Error** | **t Stat** | **P-value** | **Lower 95%** | **Upper 95%** | **Lower 95.0%** | **Upper 95.0%** |
| Intercept | 50092.75 | 6718.40841 | 7.4560442 | 2.16E-09 | 36561.181 | 63624.3 | 36561.181 | 63624.319 |
| 165349.2 | 0.805808 | 0.04575324 | 17.612048 | 8.19E-22 | 0.7136564 | 0.89796 | 0.71365644 | 0.8979599 |
| 136897.8 | -0.02672 | 0.05168968 | -0.516964 | 0.60771 | -0.13083 | 0.07739 | -0.13083003 | 0.0773867 |
| 471784.1 | 0.027309 | 0.01685407 | 1.6203162 | 0.11215 | -0.0066369 | 0.06125 | -0.00663692 | 0.0612548 |

OUTPUT:

Formula for Linear Regression:

The regression equation for this scenario is:

{Profit} = b\_0 + b\_1({Marketing}) + b\_2({R&D}) + b\_3({Administration})

Where:

* b0b\_0b0​: Intercept (Profit when all inputs are 0).
* b1,b2,b3b\_1, b\_2, b\_3b1​,b2​,b3​: Coefficients for Marketing, R&D, and Administration respectively.
* Input features (Marketing, R&D, Administration): Values provided for prediction.
* Output: Predicted Profit.

**Prediction**

For a specific set of input values (x1,x2,x3x\_1, x\_2, x\_3x1​,x2​,x3​), substitute them into the formula to calculate Profit.

Predicted Profit=b0+b1x1+b2x2+b3x3{Predicted Profit} = b\_0 + b\_1x\_1 + b\_2x\_2 + b\_3x\_3Predicted Profit=b0​+b1​x1​+b2​x2​+b3​x3​

=intercept +(R&D spend\*coefficient)+(Administration\*coefficient)+(Marketing Spend\*coefficient)

|  |  |  |  |
| --- | --- | --- | --- |
| **R&D Spend** | **Administration** | **Marketing Spend** | **Profit** |
| 21892.92 | 81910.77 | 164270.7 | 70031.507 |
| 23940.93 | 96489.63 | 137001.1 | 70547.536 |

Result: The regression analysis aims to predict the profit of a company based on its spending on Research & Development (R&D), Administration, and Marketing. The model is trained on a dataset with several data points, and then predictions are made for two new input sets.

**Predictions**

1. **Input:**
   * R&D Spend: 21892.92
   * Administration: 81910.77
   * Marketing Spend: 164270.7

**Predicted Profit:** 70031.507

1. **Input:**
   * R&D Spend: 23940.93
   * Administration: 96489.63
   * Marketing Spend: 137001.1

**Predicted Profit:** 70547.536

**Conclusion:**

The regression model appears to have a reasonable fit to the data, as evidenced by the predicted profits being within a reasonable range based on the input features. However, without further information about the model's performance metrics (e.g., R-squared, Mean Squared Error), it's difficult to assess the model's overall accuracy and reliability.