Multiple Linear Regression for Predicting Profit

Objective

The major objective of this project is to predict the profit of an organization based on the money spent on Research and Development, Administration, Marketing and the State factors.

Solution Architecture

Data Importing and Preprocessing

Necessary packages are imported in Jupyter Notebook.

Data is loaded and a copy of it is prepared.

Null values are examined and corrected.

Exploratory Data Analysis(EDA)

Bar plot: It interprets that more amount is spent on Marketing and less on Research and Development.

Scatter plot: Analysed greater the spend on Research and Development greater the profit.

Box plot: Identified the outliers in ’State’ of NewYork. Imputing outliers in the data.

Heatmap: Examining the heatmap gives the correlation is high between Profit and R&D Spend.

Model Building

Models were built on the basis of correlation between Profit and other columns. The R2\_Score of the models are as follows:

|  |  |
| --- | --- |
| Model | R2\_Score of the Model |
| Profit vs R&D,Marketing, Administration | 95.43% |
| Profit vs R&D | 95% |
| Profit vs Marketing | -4% |
| Profit vs Administration | 79% |

Result

Models 1 and 2 are more accurate with R2\_Scores 95.43% and 95%.

Conclusion

The project concludes that Profit can be yielded high with the two models of high R2\_Scores. The data driven approach can make a huge difference in the profit of an organization.