Methodology:

Followed machine learning life cycle steps

Produced box plots and scatter plots, which determines outliers.

Found outliers in Profit column.

Removed those outlier values using z score method.

One hot encoding is done for State column.

Fitted multiple linear regression model.

In excel we got Rsquare value 0.95 which means 95% accuracy.

In python we got Rsquare value 0.93 which means 93% accuracy.

Regression summary interpretation:

Negative coefficients:

|  |
| --- |
| Administration |
| State\_New York |
| State\_California |