Approach :

Multiple Linear Regression is used when you want to predict a continuous dependent variable from a number of independent variables. If the dependent variable is dichotomous, then logistic regression should be used. CLV is a continuous dependent variable. So, You

Choose Multiple Linear Regression.

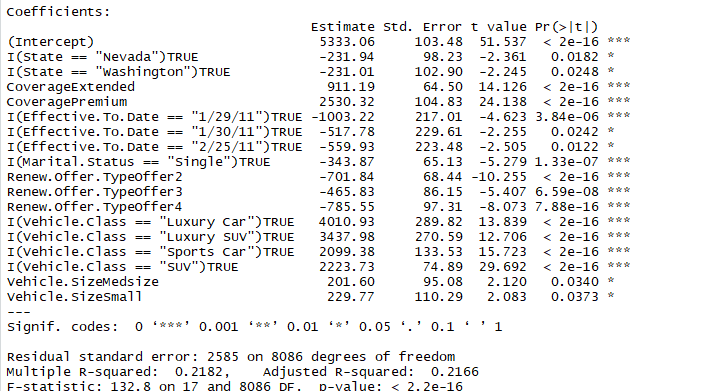
Steps :

1. Load Data in R
2. Check the structure and summary of data
3. Remove the outliers from dependent variable ( CLV )
4. Check the missing value
5. Create a model and check using summary function
6. Check the presence of multicollinearity
7. Calculate the mean absolute percentage error (MAPE)
8. Check the assumption

Result :

1. The residual standard deviation is simply the standard deviation of the residualvalues, or the difference between a set of observed and predicted values. The standard deviation of the residuals calculates how much the data points spread around the regression line.
2. the higher the R-squared, the better the model fits your data.

Interpretation and Significance of variable :



:

As the p-value is much less than 0.05, we reject the null hypothesis that *β*= 0. Hence there is a significant relationship between the variables in the linear regression model of the data set faithful

Business meaning :

Company is focus on independent variables instead of focusing on insignificant variables because only significant variables have any positive or negative effect on CLV which is our dependent variable