R Project

Predicting CLV (Customer Life Value) using R programming

The steps to approach to the data

1. Observe the data and then choose the model which we have to run by looking at the data. As our Independent variable was in numeric format It is more likely to go with the linear regression model.
2. Clean the data. Check whether there are any empty rows, as there were none then we convert the necessary independent variables into factor format from numeric format. (Such as Number.of.Complaints, Number.of.Policies)
3. Then check the outliers and omit them as there were some in the given data.
4. Then fit the model and one by one removing the unnecessary constrains which doesn’t have any correlation with the dependent variable.
5. After getting the cleaned model we run some statistical tests to valid the model.

* We calculate MAPE.
* Check the data with the multicollinearity test. (If there are any significant correlation between the dependent variables)
* Run the homoscedasticity test.
* After that Normality test.

1. Then save the Predicted value along with the data in csv format.

Significant of Variables

1. Coverage
2. Education
3. Effective to date (01/03/11, 01/08/11)
4. Employment status
5. Gender (M)
6. Income
7. Marital status
8. Monthly premium auto
9. Number of open Complaints
10. Number of policies
11. Vehicle class (Luxury SUV, SUV)

Statistics on the MODEL

* R-squared = 0.925 & Adjusted R-squared = 0.9247
* MAPE value = 0.1022
* Bptest (p-value is low)