**Codes**: - Given in Note pad

**Approach**: - As we need to predict the CLV and on the basis of it we will decide whether we need to acquire a new customer or to retain the customer so for the mention scenario we are using a Linear Regression model.

**Result and interpretation**: - While running the Linear Regression on the given data we found the below result

R square -> .875

Mape -> .13

**Assumption**: -

**No serial correlation** as the p value is more than .05, the p value is .55

**Heteroscedasticity** **assumption violated**: - p value is less than .05

Data is **not normally distributed** because is p vale is less than .05

**VIF** (Variance Influence factor) is < 2, so there is no multicollinearity presence in the data

Which signifies it’s a reasonably good model and there is some of the variables like **Coverage, Income, Number of Policies, Total Claim Amount, Vehicle Class** are the important variables which help to decide whether to acquire a new customer or to retain an existing customer (brief explanation given in PPT) or to target the future customer on the basis of mention valuable variables.