Linear Regression – Practice Assignment

Key takeaways:

- Building Multiple linear regression model using the function lm()
- Reading diagnostic plots to test the linear regression assumptions
- Identifying influential observations and handling them
- Data Pre-processing
    - Binning - Converting a numeric attribute to categorical
    - Dummies - Converting a categorical variable to numeric
    - Standardization - Effect of scaling the data
    - Imputation - Handling the missing values
    - Train-Test splitting of the data
    - Using Correlation plots, correlation values to identify appropriate variables for model building

Practice Assignment - CustomerData.csv

A large child education toy company (company name is confidential, and data is masked) which sells edutainment tablets and gaming systems both online and in retail stores in the US wanted to analyze the customer data. They are operating from last 15 years and maintaining all transactional information data. The given data 'CustomerData.csv' is a sample of customer level data extracted and processed for the analysis from various set of transactional files. Using this data, they want us to understand the lifetime value of each customer (LTV). This will enable them to design marketing strategies and customize the product offerings. The objective of activity is building a regression model to predict the customer revenue based on one/more factors that influences revenue.

- Read the given "CustomerData.csv" data into R (Rmd)
- Perform EDA
- Split the data into train (70%) and test (30%)
- Perform data pre-processing
- Are there any missing values in the data? If there are, then impute using central imputation method.
- The target for this problem is "Total Revenue generated"
- Model Building

- Select the most influencing variables as predictor for predicting the revenue generated. How would you do this?

- Build a Multiple linear regression model to predict the target with the selected variables

- Check the residual plots and report your observations

- Report performance metrics