**Customer Lifetime Value Prediction**

Problem Statement

An Auto Insurance company in the USA is facing issues in retaining its customers and wants to advertise promotional offers for its loyal customers. They are considering Customer Lifetime Value CLV as a parameter for this purpose. Customer Lifetime Value represents a customer’s value to a company over a period of time. It’s a competitive market for insurance companies, and the insurance premium isn’t the only determining factor in a customer’s decisions. CLV is a customer-centric metric, and a powerful base to build upon to retain valuable customers, increase revenue from less valuable customers, and improve the customer experience overall. Using CLV effectively can improve customer acquisition and customer retention, prevent churn, help the company to plan its marketing budget, measure the performance of their ads in more detail, and much more.

Project Overview

* The objective of the problem is to accurately predict the Customer Lifetime Value(CLV) of the customer for an Auto Insurance Company
* Performed EDA to understand the relation of target variable CLV with the other features.
* Statistical Analysis techniques like Mann–Whitney U and also Kruskal Wallis test for the categorical variables were performed to find the significance of the features with respect to the target.
* Supervised Regression Models like Linear Regression, DecisionTree Regression and Random Forest Regression.
* Using GridSearchCV with Random Forest Regression gave the best RMSE and R^2 score values.

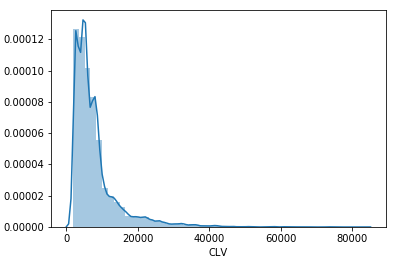
Dataset Description

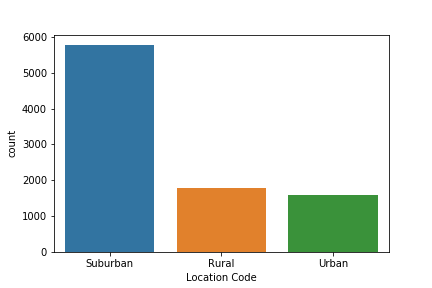
The dataset represents Customer lifetime value of an Auto Insurance Company in the United States, it includes over 24 features and 9134 records to analyze the lifetime value of Customer.

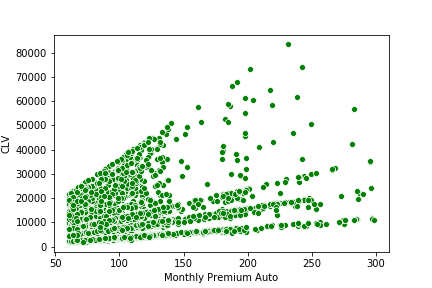
Exploratory Data Analysis

Univariate, Bivariate and Multivariate Analysis were performed to bring out important aspects of data into focus for further analysis. Some of the highlights from EDA are listed below.

**Univariate Analysis**

****

* CLV is heavily right skewed in the data

**** **Bivariate Analysis**

## CLV and Monthly premium auto have a positive correlation and there is a linear relationship between them.

## Multivariate Analysis

## There is a positive correlation between CLV and the monthly premium auto

## There is a slight positive correlation between the total claim amount and CLV.

## Income has a lesser positive correlation with CLV

## Evaluation Metric

## RMSE and R^2 score were chosen as the metric for the models.

## Supervised Models Used

|  |  |  |
| --- | --- | --- |
| Model | R^2 Score | RMSE |
| Linear Regression | 0.25 | 0.577 |
| DecisionTree Regression | 0.844 | 0.263 |
| RandomForest Regression | 0.910 | 0.198 |
| RandomForest with GridSearchCV | 0.914 | 0.195 |

## Final Model

## By comparing RMSE and R^2 score results of all models and then we choose the best model as the Random Forest with GridSearchCV, having the best evaluation scores.

## Conclusion

## Overall we can see that No of policies, Monthly Premium auto, Total Claim amount, Months Since Policy Inception, Income , Months Since Last Claim, Number of Open Complaints, Coverage\_Extended,vEmploymentStatus\_Employed and Renew Offer Type\_Offer2 are the important features in predicting the Customer Lifetime Value.

## The customers having more number of policies with high monthly premium will add more value to company.

## Ironically being an auto insurance company, the type of vehicle or size does not have an impact on the CLV prediction.

## The insurance agents should start increasing their policy advertisement for the customers who have more no. of policies, which is the major feature in predicting the CLV