**Shilpa Kolekar**

**DSC 530 - Final Project**

**Date - 11/22/2020**

Final Project - Auto Insurance Rate Prediction Model

Auto Insurance Rate Predication Model is a model which represents how insurance companies predict the risk of each individual and offer them appropriate auto rate, or in other words, how actuaries use to predict customers’ risk factors and come up with the auto rate.

To start my data exploratory process, I picked five variables which I think were the best fit to research problem statement. Five variables are KIDSDRIV, AGE, MSTATUS, CLAM\_FREQ and REVOKED. I plotted each variable on histogram. Mean, median, and mode were calculated for non-binary variables and outliers were viewed.

Variables that I picked for split were revoked license and marital status of the driver. I compared them with age variable. After plotting them on graph, I don’t think I picked the correct variables for split. Marital status of driver and age graph is not relevant for my research question. Same way, revoked license and age graph is non-relevant for my research. However, individually REVOKED, MSTATUS, and AGE variables were significant in auto rate making.

After analyzing graphs, coefficient correlation, and logistic model, I concluded that auto rates vary for person to person and there are many factors that are involved in rate making. Factors that I chose to predict auto rate model gave me good idea on how insurance company predicts the rate. According to my model, married couples are more safe drivers and therefore rate is lower for this category. Surprisingly, age does not impact on auto rate much, but if you have revoked license, more kid drivers in the household, and more claim frequency in past, that might cost you more.