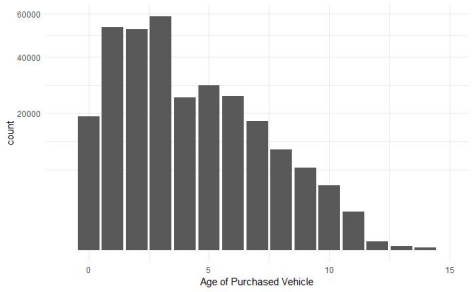
2020 Carmax Analytics Showcase Report

Introduction

In this project I analyzed a dataset provided by CarMax that was representative of U.S. used vehicle sales. I utilized a variety of statistical models and data visualizations to identify unique insights on the factors that influence vehicle sales, and the results of my analysis are explained in the sections below.

Vehicle Purchase Year

In terms of inventory strategy, one of the variables I focused my analysis on was the vehicle purchase year. I created an additional variable that I refer to as 'vehicle year lag' to represent the age of the vehicle (using vehicle year - 2013). In general, the count of vehicle sales peaked at with vehicles at the age of 3 years, with the featured bar plot showing a steep drop in total purchases following the third year. When analyzing the age of purchased vehicles I discovered that the trends were highly uniform across all variables besides the manufacturer, leading to the conclusion that it is highly important for the age of the purchased vehicle to be analyzed based on sales trends for each specific manufacturer when structuring a CarMax location's inventory.



Logistic Regression for Subsequent Purchase Prediction

I built a series of Logistic Regression models in order to identify which variables were the most important in terms of predicting subsequent purchases. Through this process I was able to identify the Vehicle Trade as a highly important variable towards predicting subsequent purchases.

$$logit(p) = \beta_0 + \beta_1 * trade\_in$$

The model above results in an intercept of **-.837812**, which represents the log odds ratio that a customer makes a subsequent purchase. The *trade_in* coefficient of **.230566** corresponds to the log odds ratio between the *trade_in* and non *trade_in* groups. By using the following formula, we can convert these log odds ratios into probabilities, making them much easier to interpret.

$$e^{logit} \div (1 + e^{logit})$$

By applying the formula above, we achieve results of **.3019958** and **.5573874** for the intercept and *trade_in* coefficient respectively. These results signify that we would predict a customer who trades in a vehicle to be roughly **55.74%** more likely to make a subsequent purchase.