**Executive Summary**

When predicting the price of a used car in the current Indian market, we discovered that the year, car brand, engine size, mileage, and power associated with a given used car were the most telling of price. The model we are proposed gives us approximately 5 percent information gain with a high amount of accuracy in our estimate. Additionally, for future implementation, non-linear models would be better performing than linear models. Cost associated with a given car and the new price equivalent of a given car brand would be helpful for future analysis as well.

When looking over the next steps associated with this company, I believe that creating or importing a repository of costs associated with purchasing and maintaining a used car would be highly beneficial. Additionally, creating a webservice that shows prices to our users and grades the value of the product being sold would be a great way to maximize transparency with our users. Finally, I would continue to explore other methods to produce a better profit model while maximizing customer satisfaction.

**Introduction**

The purpose of this analysis is to develop a model that can provide insight on how to properly price used vehicles in the market and determine the independent variables that drive the price using machine learning techniques. There appears to be a surge in customers wanting to purchase used cars just as much as new cars with the given economic circumstances that the world has faced in the past couple of years. With that said, if our company, Cars4U, does not provide a service that our customers are looking to be filled, we potentially miss out on a profitable strategy that we can leverage compared to our competitors.

The intended goal is to determine a pricing model that can effectively predict the price of a used vehicle within the Indian Market that our business partners can leverage and develop a strategy that makes us competitive in the marketplace. A couple of questions that come to mind before we begin this analysis is as follows.

1. Are there particular attributes of a vehicle that can lead to a particular price point?
2. Do certain car models or brands sell at different price points when they are used or are they all priced relatively similar?
3. What is the price that our sellers can offer to the customer that will maximize our profits but also maximize the utility of the customer?
4. Are there ways that we group or segment certain vehicles that can help us differentiate them from other vehicles?

Ultimately, we are looking to provide a solution to our business stakeholders by creating a robust regression model that gives us a competitive advantage in the used car sales sector with a strategy of how to properly price the used car market. This is a hard question to answer as it has been a problem that was first discussed by George Akerlof in 1970 cited as “*The Market for Lemons: Quality Uncertainty and the Market Mechanism”*. In which he concluded that the quality of goods traded in a market can be degrade in the presence of information asymmetry between buyers and sellers. This asymmetry causes some buyers to purchase cars that are found to be defective after it has been bought, also known as a “lemon”.[[1]](#footnote-1)

**Model Selection**

For this problem, we have decided to choose our tuned Random Forest and tuned Decision Tree model to be the best models to predict price. While our Random Forest model has the smallest difference in Root Mean Square Error, there is only a five percent information gain. While our Decision Tree has a slightly higher Root Mean Square Error and a 10 percent information gain.

Table

Description automatically generated

Additionally, for situations in which time is short or resources are thin, I would propose that we use the Decision Tree model as it is less computationally intensive and gives us estimates that are relatively similar but may be considered less accurate. For other times where we have a new lot of cars coming in and we have enough time to lead our model building process, I would consider utilizing the Random Forest Model. We can observe that all three of the linear models are not considered to be good models to choose from a bias-variance tradeoff issue. All suffer from severe over or underfitting, which means that our estimates will not be reliable and some of our features could be considered significant when, they are not. With that in mind, we would not recommend putting these models as a plausible solution for us to predict the price of a used car for our end users.

**Recommendations**

After looking through the data, we can really come down to four recommendations for the enterprise to capitalize on in order to best build up their brand and maximize their profits. These four recommendations are Focus on brand, be transparent, look for new technology, and ease of use.

For the first recommendation we tagline this as “Its More Than Just a Brand”. It appears that the brand of a car can impact the price sold of a used car. We can think of this with two examples. If we were to sell a Toyota and a Lexus, we would assume that the Lexus would sell for higher than the Toyota, right? The interesting part is that most Toyota and Lexus vehicles are made from the exact same parts as Lexus is owned by Toyota but considered to be a ‘luxury’ vehicle. Point being that brand matters and people are willing to pay for it even if the cars are made from the same parent company or even made in the same factory.

For the second example, let’s take the exact same Lexus, that we expect to make a substantial profit off that vehicle, all things equal to a Toyota. However, would we expect to make significant more profits on a Lamborghini as we would a Lexus? I would think not. My proposed recommendation is when we are looking to restock our inventory, we want to get brands that will sell for higher margins in order to increase our cashflow but couple that with an assortment of brands that we know have a significant impact on price. Identifying these brands that do sell compared to brands that do not is paramount in not keeping inventory that will sit and we will later sell for less profits or even a loss.

Next, we want to establish a way in which we are giving our customers the feeling that we are being honest and open with them. Just providing a good product is not enough, we need to convey to our customers that the quality of the product is good. Much like companies such as Carfax shows the quality of the car, we need to make sure that we do our diligence when selling a car. The car is the product, and you are the company associated with selling that car. For us to be competitive in a saturated market that has countless numbers of used cars selling companies all over the world, we need to build our brand on the foundation that we are ethical and looking to provide the best product possible, at a price that is fair for both sides.

My third recommendation is more future facing and really is just an observation that I saw from a limited amount of data points within our data set. For us to be successful in the current economic landscape, we must sell cars that people want and make profit margins to keep us in business. However, I did observe that there are not many vehicles in our data set that are automatic or electric vehicles. Automatic vehicles are either not resold often or are not as common in India but do, on average, sell for a higher price point compared to their manual counterparts. This could be a great way for us to differentiate our products and provide a unique product to our users that is not as prevalent for this area of the world.

Chart, box and whisker chart

Description automatically generated

Additionally, electric vehicles are becoming much more prevalent in the automobile sector and while they are not going to be up for resale anytime soon, this is a long-term strategy that we can utilize or even investigate Hybrid vehicles that are primary fueled by petrol and use electric power from the batteries to decrease overall carbon emissions. Based off current global concerns, we could sell a used electric or hybrid vehicle for a higher premium while helping reduce our carbon footprint while delivering superior products.

Chart, box and whisker chart

Description automatically generated

Lastly, none of these are relevant if access to our product is not reliable and convenient. In an ever-changing world, we see numerous companies meeting their clients where they are with mobile applications and new payment options such as buy-now, pay-later. We must build an interface that allows our customers to see our product where they are and be able to browse options that fit their needs and inform them of the offers that are competitive in our place of business. I would also want to consider establishing some sort of cost accounting for us to utilize cost as a measure of economic forecasting and future machine learning algorithms.

**Benefits & Costs**

With a product like this we can assume there are countless number of costs and benefits. I believe that if we implement this pricing model, we will have a simple, quick, and reliable pricing model that will make use competitive in the market. This model will give us insights on the cars to buy and the cars to pass on as well has how to price them in order to maximize profits, all things equal. With expected profit margins, we can expect to grow our risk portfolio and grow as a company in the used car space which means hiring more labor and having more capital to create an even better experience for our customers. Lastly, we will be able to reduce our overall carbon footprint by using less computationally intensive methods that we are currently using and be able to contribute more hybrid or electrical vehicles to our customers at a reasonable price.

When it comes to cost, I think this model can hurt us in some areas. First, we may get too fixated on what makes us the most profit and pass up on cars that will give us a slim marginal gain which could potentially lead to having a harder time replenishing our inventory. There may also be circumstances in which the market has dried up and there are no cars to buy but nonprofitable cars which makes forecasting very hard. Another issue with this model is that we will inevitably have to grow in order to compete and that takes a lot of labor and data, with labor and infrastructure costs increasing, we will need to be vigilant on these hurting our profit margins long-term. Lastly, with the introduction of electrical vehicles, as they become more affordable, they have a longer lifespan than the average petrol vehicle, which leads to a slower selling cycle. With 1.4 billion people living in India, it is hard to believe that this is a major concern at this current time, but something we want to consider for future ventures.

**Risks & Challenges**

With Cars4U being in the automobile industry we must be concerned with how the economy is always performing in order to hedge ourselves when the economy is bad. One major risk is that as the economy gets worse, people are less willing to spend money on frivolous things such as vehicles. Creating some sort of contingency plan or diversifying our ventures will help us mitigate the volatility of the business we have chosen to be in. The other risk is that we are in a saturated market. Meaning that there are many players with little barriers to entry. We will have to compete at the margins and be intentional about the way we interact with our customers. We must go the extra mile in order to gain market share in this industry and make sure that Cars4U is the first name that comes to their mind when wanting to buy a used car. The challenges are like the risks, but another area would be the quality of data that we are feeding our models. Being able to change pricing structures quickly based on external economic factors will make use more competitive in the marketplace and give us an edge on what price to offer our users.

**Future Analysis Considerations**

While I think this model is a great model in predicting the price of a used car, I think there are alternatives to come up with a reliable answer. I think using different methods such as neural networks, KNN methods, or even deep learning models would be a great start to see how we can predict price for our customers. Additionally, I would consider doing a time series analysis to see how the over used car market is performing compared to the economy to make sure that we are prepared for a sort of recession/ depression that could be around the corner. I also would want to gather more features such as cost of purchasing the used car, price of maintaining the car, the color of the car, and the type of car (Sedan, SUV, Truck, etc.) of the car. These can help simplify our models and take less computational power while hopefully provide just as accurate or even more accurate models.

1. (Wikipedia n.d.) [↑](#footnote-ref-1)