When purchasing or selling a used car, finding the perfect price both to avoid overpaying and to attract buyers is crucial. It is a commonly accepted fact that the price of a used car is associated with the milage of the car. Other accepted factors are the age and the type of car. However, although this information is widely acknowledged as the primary marker for a used car’s price, there is still a heavy variance between used cars that have highly similar factors. Therefore, the crucial factors still need to be identified and extracted. Currently, one of the primary obstacles to identify these factors is a lack of correlating data.

However, we have found a complete dataset of 1.2 million vehicles to overcome this obstacle. Therefore, through a thorough test of several different data models, we will discover the crucial factors that can identify a used car’s price. Specifically, our hypothesis is that there is a set of factors that can solely determine the price of a used car. Using various regression methods, we will determine which factors contribute the most, isolate those factors, and create a model that can predict an accurate price given any combination of those factors. Through the completion of this project, we will be able to effectively determine a proper purchasing or selling price for a used car, thereby allowing buyers to avoid overpaying and allowing sellers to attract willing buyers.

Aim 1 will identify the specific factors that most determine the price of a used car. These will be found by first preprocessing the data and pruning any information irrelevant to the car’s price. Then the resulting data will be standardized. This will keep only the most important factors from the original dataset.

Aim 2 will take the preprocessed data from Aim 1 and form multiple models from the data. Each model will apply a different Machine Learning methodology, thereby increasing the chances for a viable, highly accurate model to appear.

The proposed study will implement several new models for an accurate estimate of the price of a used car based on the car’s factors. Both accuracy and training time will be measured for each model, therefore the most accurate model with the shortest training time will be selected as the best model. This model will provide better financial opportunities for both used car sellers and used car purchasers, as more buyers can be attracted to purchase due to more accurate prices, and buyers will not have to fear about overpaying for a used car.