

Matthew Witschorke
Yarely Vargas
Sam Slomowitz
Joseph Onwukeme

Used Car Price Prediction



ABSTRACT

Our project sets out to find the parameters to determine the price in used cars on the market.

TABLE OF CONTENT

Introduction.....pp. 3

Our Hypothesis.....pp. 4

Cleaning Data

Introduction

Previous research suggests the positive correlation between year and price and an inverse relationship between mileage and price (Pudaruth, 2014).¹ Other studies have employed regression tree model, multiple regression, and lasso regression models to used car datasets. Future work has been proposed for random forest methods and other machine learning models (Venkatasubbu and Ganesh, 2019)².

¹ Pudaruth, Sameerchand. "Predicting the Price of Used Cars using Machine Learning Techniques." *International Journal of Information & Computation Technology*, vol. 4, no. 7, 2014, pp. 753-764.

² Venkatasubbu, P., & Ganesh, M. (2019). Used cars price prediction using supervised learning techniques. *International Journal of Engineering and Advanced Technology*, 9(1S3), 216–223. <https://doi.org/10.35940/ijeat.a1042.1291s319>

Our Hypothesis

The context of our work is to use python to merge multiple data frames followed by a NULL value analysis and outlier analysis. Our ultimate goal is to tell a data story to evaluate if models within brands of Toyota and Ford have a higher price as well as note the variation in price between brands. We also were interested in the role of inflation in our dataset as well as the impact year on the price of a car in the used car market. Another point of interest is to see if we can expand the dataset with new car data gathered from APIs, once we have understood the basic trends and patterns of the used car data.