

*The work We submitted as part of this assignment is original, and due credit is given to others where appropriate. We accept and acknowledge that We are solely responsible for if the assignment found to be plagiarized in any way, and we will be subject to School's Academic Integrity policy*

Group 10: Jiadai Yu(U63347232), Yuesen Zhang(U25148433), Xin Su(U07869307)

Date: 22<sup>rd</sup> Aug 2022

We use data of cars sold by Ford company that were available on <https://www.kaggle.com/datasets/adhurimquku/ford-car-price-prediction>. This data included model, transmission, mileage, fuel type, annual tax, miles per gallon, engine size, and price information for 17965 vehicles produced from 1996-2020. We concentrate on price, which is of most interest to consumers, and summarize the contribution of different variables to price.

First, the average price of cars from the different Ford car brands ranged widely, among the 24 models, the most economical model was Streetka, which had an average price of \$1924.50, while the most advanced brand was Mustang, with an average price of \$34631.26. But through analysis, the most popular models were the mid-priced Fiesta, Focus and Kuga, which accounted for three quarters of the total sales.

The secondary factor was the year of production. The average selling price of the vehicles declined slightly in the first a few years, hitting the lowest in 2004, and had risen since then, reaching \$20819.87 in 2020.

The transmission and fuel type of the car also had a significant impact on the price. Manual cars require greater technical proficiency from the drivers, and thus had the lowest average sales price. While the most convenient automatic vehicles had the highest average sales price. In addition, vehicles powered by traditional fuels, such as petrol and diesel, were less costly, while new energy models were more expensive and sold less than the former kinds.

Last but not least, as the most important summarizing indicator to define the capability of a vehicle, engine size generally had a positive correlation with the selling price, just like the old saying, you get what you pay for.

As for the attribution of mile age, annual tax and miles per gallon to the selling price of the vehicle, we conducted more detailed grouping analysis and regressions in the subsequent discussion, thus we do not dwell on them here.

We use z-score to determine outlier. if the data points has a z-score more than 3 or less than -3, it indicates that it is distributed quite different from other data point, outside the 3rd standard deviations. Thus we can conclude that the data point is an outlier.

We have the first outlier with a price of 42489 compare to the overall price mean of 12279. And compare to the price outlier mean 31887. The main reason for that is the car is new and the mileage is low so the sales price is relatively high. Also, we can see that the mileage of 3500 is pretty low compared to the overall mileage mean of 23362. The reason is that the car is new so the mileage is low. We have an outlier mpg of 22.1 compared to the overall mpg mean of 57. The reason is that the engine size is 5.0 which is big so the mpg is low. The tax is 145 compared to the overall tax mean of 57 which is high but it is reasonable that if the price is high then the tax is high. We also realized that the tax for the 364 row is 570 which compared to the tax mean of 57. Our guess would be this might be a typo since the data is too far from the mean.

Since the clients were concerned about the price the most, our group plotted the price with the variables mileage, tax, and mpg. From the plots, we notice the price has

a negative correlation with mileage and the correlation coefficient is around -0.53. In other words, the higher mileage has the lower price, because the higher mileage means the car is used more frequently, and it may have a higher risk of repair than the car which is not used a lot. In addition, both tax and mpg have relatively weak correlations with the price and the correlation coefficients are respectively 0.41 and -0.35. The tax rates are varied based on many factors. For example, each state has a different tax rate, or people can gift their cars to others without tax. Hence, the tax rate depends on various situations and does not correlate significantly with the car's price. For the variable mpg, it can also depend on other factors. The high vehicle emission causes high mpg but the price of this model can be cheap. Instead, the expensive model can also have high mpg. Therefore, mpg does not have a direct relationship with the price. In conclusion, only the mileage has a negative correlation with the price. The tax and mpg have a weak correlation with the price.