DATA 603 Project Report

DEC. 13, 2022

IMAD AHMAD, IBTASSAM RASHEED, ANDRII VOITKIV AND YIP CHI MAN

Table of Contents

1. Introduction

1.1 Motivation

1.1.1 Context

For over two decades now, the used car market has dominated most car sales worldwide, with used car sales more than doubling new car sales (Ellencweig 2019). With the rise of technology and more people utilizing websites like Facebook Marketplace and Kijiji, used car sales are accessible to almost anyone from anywhere. Below is a graphic from Ben Ellencweig's article on the used car market showcasing how consistently strong used car sales have been.

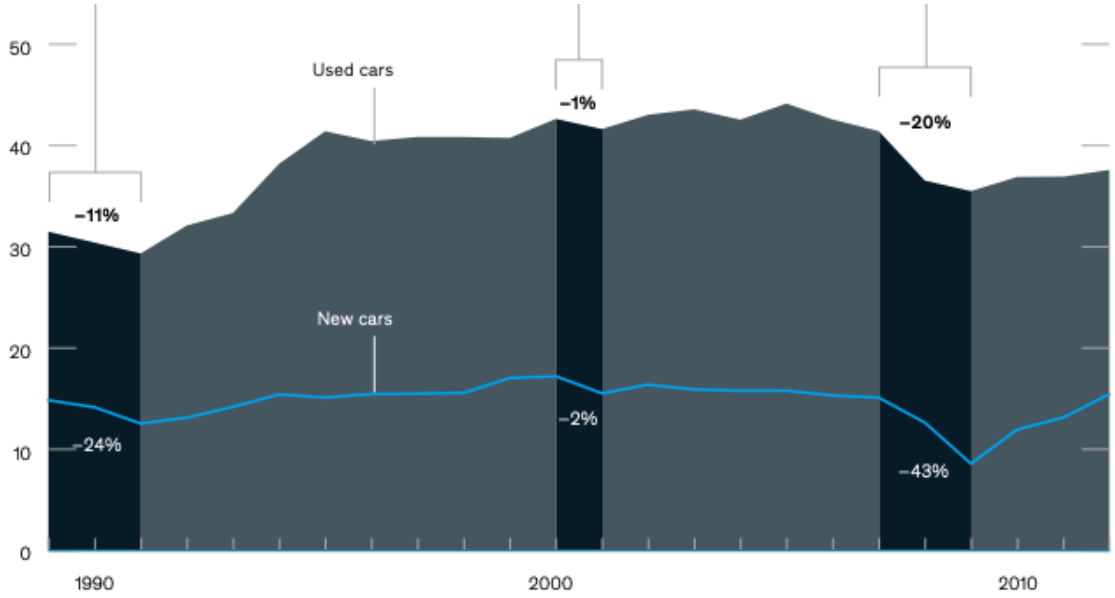


Figure 1A: Line graphs showing used vs. new car sales over twenty years

India has one of the fastest growing used car markets, being fifth in worldwide used car sales last year. Most of India's used car sales happen online, through websites like cardekho.com (Chadha 2021). Thus, it provides a large enough sample size to be used to study the global used car market.

1.1.2 Problem

COVID-19 impacted many industries in a large way. The car industry was not immune, suffering losses of close to 40% (Chadha 2021). This paired with fluctuating and inflated prices impacted almost all of us, either directly, or through a domino effect touching other industries. Used car prices specifically have soared in many countries since the pandemic, bringing sales to a decline. We are just now reaching a stage of recovery in their sales (Hoeft 2021).

Research has shown that the key to getting ahead of the next world emergency is understanding the deep ins and outs of how industries function (Sun 2020). Thus, gaining a deeper understanding of what predictors are involved in the listing price of a used car would aid us greatly in mitigating another set of extreme circumstances. We are going to address this issue

and attempt to gain a strong understanding of the most important variables involved in a car price.

1.2 Objectives

1.2.1 Overview

As previously stated, the overall intent of this project is to gain a deeper understanding of the predictors in a used car's listing price. Previous research has shown that mileage, kilometers driven, and year are the most important predictors (Gegic 2019).

We also hope to come up with some ideas on how the used industry can prepare for the next global emergency, focusing on what and what not to prioritize.

1.2.2 Goals and Research Questions

Our main research questions can be summarized as follows:

1. What are the most important predictors in the prices of a used car?
2. What can be done for the car industry to prepare for the next global emergency?

Our primary goal will be to use regression analysis to construct a model with car listing price as the dependent variable, and a set of both quantitative and qualitative predictors as independent variables. Along the way, we will use visuals to both test our model for assumptions, as well as show how different facets of the data are related.

Over 300 million people drive cars in North America alone (Carlier 2021). This is an industry that effects all of us in one of or another. Even so, understanding how to aid an industry during the next global emergency is paramount. Thus, this project is important and necessary as a small step in stabilizing industries during collapse.

2. Methodology

2.1 Data

The dataset was sourced from kaggle.com (Birla 2020). It was originally pulled from the API of cardekho.com, one of India's most popular used car websites (Car Dekho 2020). The dataset is part of the open data commons, available through a Database Contents License (DbCL) (License).

The dataset contains 13 columns in total, one of which will be our dependent variable, and 7 of which will be our predictor variables. Our explanatory variable is selling price, measured in INR (Indian Rupee). Below are the 12 other columns, our 12 predictor variables:

1. Name (String): The car name in the format "make model". Included because some car makes and models are favored over others and may influence price.
2. Year (Quantitative): The year the car was *initially* bought. Included because age is one of the biggest indicators of the price of a car.

3. Km_driven (Quantitative): The number of kilometers the car has been driven at the time of the listing. Included because km on a used car is one of the biggest indicators of price.
4. Fuel (Qualitative): The vehicles fuel type, one of **petrol**, **diesel**, **CNG**, **LPG**, or **electric**. Included because fuel prices vary by type, and so fuel may influence selling price.
5. Seller type (Qualitative): The type of seller, one of **individual**, **dealer**, or **other**. Included because people may be more trusting of a dealer, and so they may be charging more.
6. Transmission (Qualitative): The type of transmission, one of **automatic** or **manual**. Included as transmission affects mileage, which can impact price.
7. Owner: (Qualitative) The number of previous owners of the car, one of **first owner**, **second owner**, **third owner**, **fourth and above owner**, **test drive car**. Included because the number of owners the car has had will probably impact price.
8. Mileage (Quantitative): The mileage of the car, in a variety of units (ex. kmpl, km/kg, etc). Included as it is one of the most important things people look at when buying a car.
9. Engine (Quantitative): The engine capacity, measured in cubic centimeters. Included as car enthusiasts may value this, making high-capacity engines priced higher.
10. Max_power (Quantitative): The maximum power of the car, measured in brake-horsepower. Included for the same reason as predictor number 9.
11. Torque (Quantitative): The torque of the car, measured in a variety of units (ex. Nm @ rpm, etc). Included for the same reason as predictor number 9.
12. Seats (Quantitative): The number of seats in the car. Because of need, people may charge more for cars with more seats, impacting price.

As there aren't any dates of listing in the dataset, it is hard to fully know the sampling method. However, it appears that all the cars are common cars in India, and with 2000 unique cars, I think it's safe to assume the data was taken through random sampling from the website (either by randomly sampling, or just taking a sample of around 8000 listings on a given day). Thus, it should not contain any biases. Below is the head of our dataset:

Name	year	selling_price	km_driven	fuel	seller_type	transmission	owner	mileage	engine	max_power	torque	seats
Maruti Swift Dzire VDI	2014	450000	145500	Diesel	Individual	Manual	First Owner	23.4 kmpl	1248 CC	74 bhp	190Nm@ 2000rpm	5
Skoda Rapid 1.5 TDI Ambition	2014	370000	120000	Diesel	Individual	Manual	Second Owner	21.14 kmpl	1498 CC	103.52 bhp	250Nm@ 1500-2500rpm	5
Honda City 2017-2020 EXi	2006	158000	140000	Petrol	Individual	Manual	Third Owner	17.7 kmpl	1497 CC	78 bhp	12.7@ 2,700(kgm@ rpm)	5
Maruti Wagon R LXI DUO BSIII	2007	96000	175000	LPG	Individual	Manual	First Owner	17.3 km/kg	1061 CC	57.5 bhp	7.8@ 4,500(kgm@ rpm)	5

Figure 2A: The head of the raw data table

Show 10 entries Search:

	year	selling_price	km_driven	mileage	engine	max_power	seats	torque_rpm
Mean	2013.97	635708.25	69407.6	19.47	1456.17	91.16	5.42	2698.88
Standard Dev	3.86	786503.63	56860.5	3.94	503.26	35.2	0.96	1138.44
Median	2015	450000	60000	19.33	1248	82	5	2000
Min	1994	29999	1	9	624	32.8	4	175
Max	2020	10000000	2360457	42	3604	400	14	21800
Range	26	9970001	2360456	33	2980	367.2	10	21625
Percentile 25	2012	270000	35000	16.78	1197	68.05	5	1750
Percentile 50	2015	450000	60000	19.33	1248	82	5	2000
Percentile 75	2017	680000	96000	22.32	1582	102	5	4000

Showing 1 to 9 of 9 entries Previous 1 Next

Figure 2B: Some statistics for quantitative variables

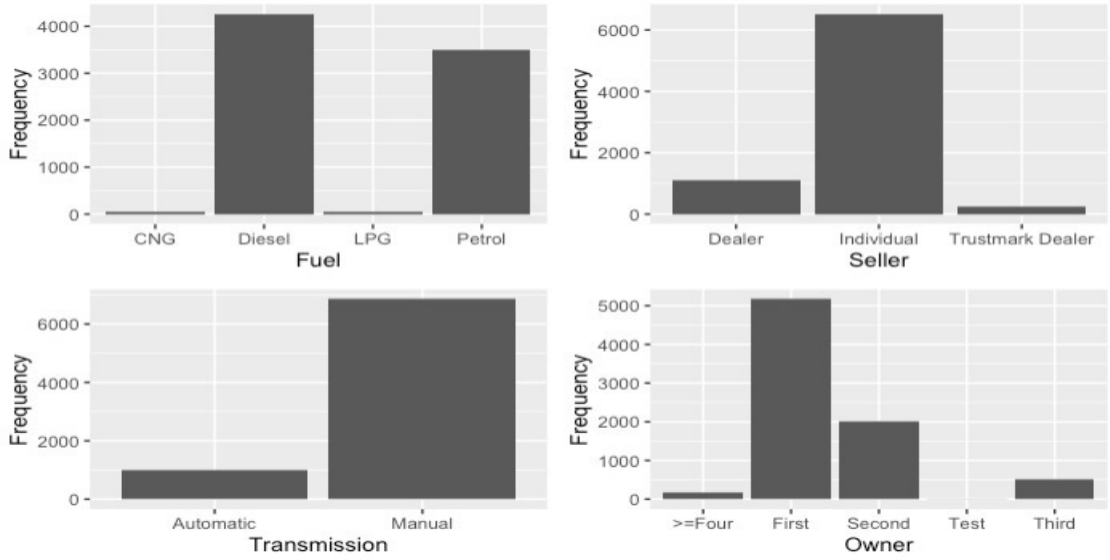


Figure 2C: Distribution of our categorical variables

2.2 Approach

We will be using multiple regression linear modelling. Our approach will be to first clean our data set, and deal with any null values as needed. We will also have to wrangle our dataset to deal with converting to consistent units. We will then construct our model through the typical workflow: first order model, interactions model, then higher order model. Appropriate assumptions will be checked before and after construction of the model. I think this will work well as previous research studying the same topic has used a very similar approach and been

successful (Monburinon 2018). All regression models, calculations, and analysis will be done using R in RStudio.

2.3 Workflow

Our first step before starting will be to test our predictor variables for multicollinearity, through calculating variance inflation factors for each variable. Dropping appropriate variables, we will perform a Global F-test on our model. After this, we will perform individual t-tests on all variables in our full model, as well as stepwise, forward selection, backward elimination, and all-possible-regressions selection procedures on the model. Along the way, we will perform partial F-tests on any variables we are thinking of dropping.

The second step will be to build our full interactions model. Performing individual t-tests on this model's beta coefficients, we will have our reduced model. We will again perform individual t-tests on this model, and if we drop any more terms we will check them with a partial F-test.

We will then move on to checking for higher order terms in our model.

2.4 Workload Distribution

The initial data cleaning was done by Chi Man. We then chose a day to meet up and build the model together. Ibtassam organized the plan for us, and all of us came together to build the model on Imad's laptop. While the model was being build, Imad began writing the report and finished it later. Everyone worked together to check conditions for the model as well

3. Results

Below is a general roadmap of our plan for performing the regression. Another similar map will be shown in section 4.1, showing what we actually ended up doing:

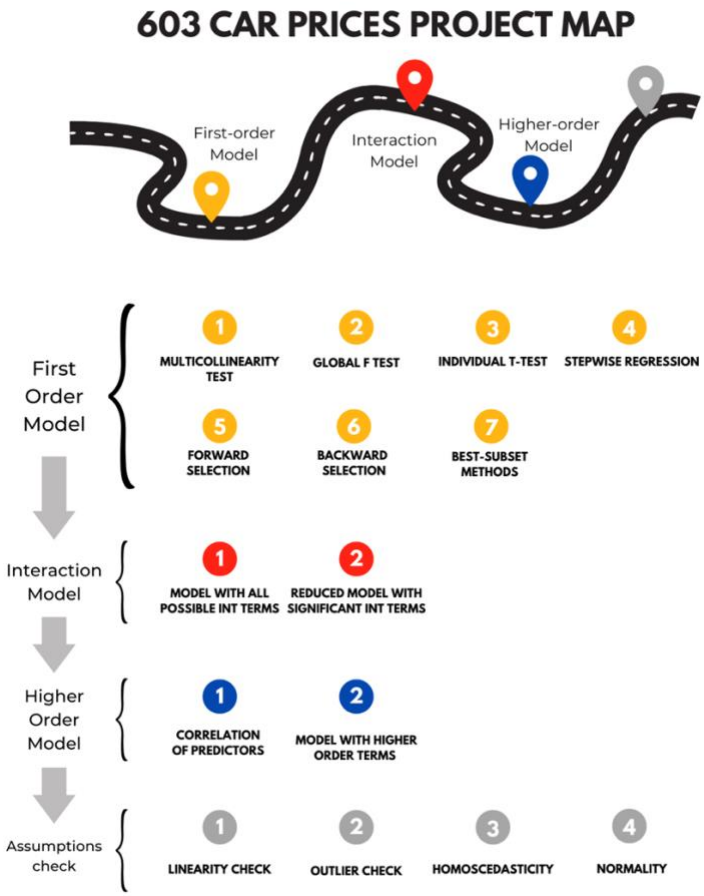


Figure 3A: General roadmap of our plan

3.1 Data Cleaning

Our first cleaning step was removing any rows containing null values. This involved dropping 221 rows from our total 8128 rows. We did this as it wasn't many rows that were removed, and we did not want any unintended bias introduced. We then converted all our mileage values to kmpl, where previously we had kmpl *and* km/kg values. This column also had 17 rows containing 0, which is not an appropriate mileage, so we dropped these rows. We then converted the torque column. Initially, it is in the format Nm @ rpm (see head of data above). We split this into two columns, one containing the Nm (torque_nm), and one containing the RPM (torque_rpm). We suspect that these columns will show strong collinearity and decided to only include the torque_rpm column. Some torque values did not have rpm values attached and resulted in null rows in the torque_rpm column, and so we removed these 35 null rows. Finally, we removed all the unit labels from the rows containing numbers, so they can be used as numerical columns.

3.2 Construction of First Order Model

3.2.1 Testing for Collinearity

Our first step in constructing our linear model will be to test it for collinearity, or strong correlations between two of our predictor variables. A strong linear correlation between two of our predictor variables can cause problems as the coefficients can becomes very sensitive to small changes, and the statistical power is weakened. We will test for multicollinearity by calculating the variance inflation factors (VIF) for each variable. This measures correlation between independent variables. The equation for calculating VIFs can be found in appendix B equation 1.

In calculating the VIFs, we saw that the only two variables that show strong collinearity are the variables for diesel fuel and petrol fuel (both are separate variables for fuel as it is a qualitative variable). Because of this, we did not drop these variables. All our raw R output of the VIF's can be seen in appendix A figure 1. The graph below also shows that there were no extremely high correlations in our model:

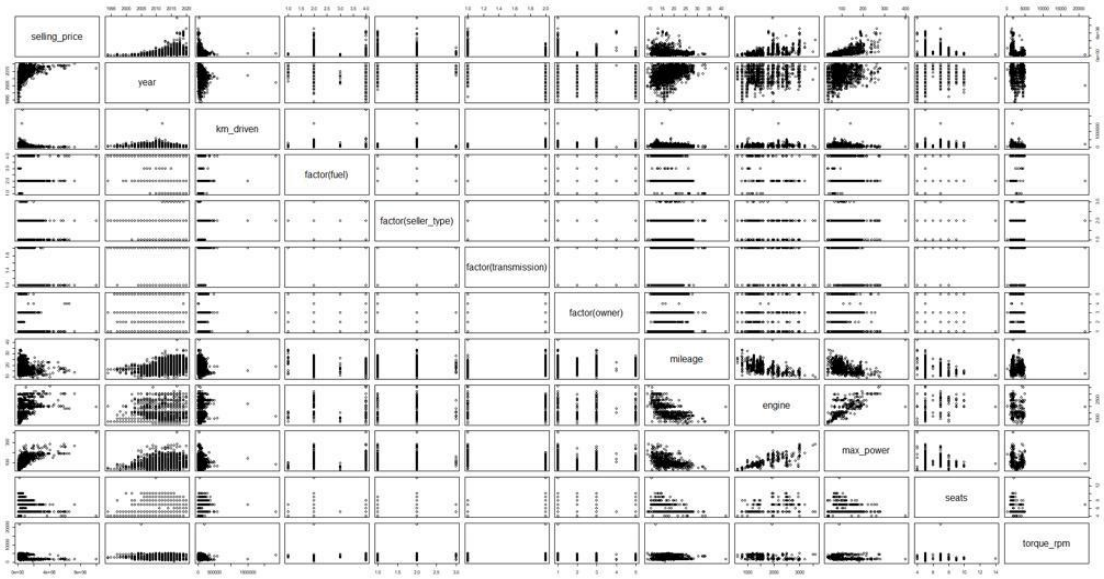


Figure 3B: Facet grid of correlation scatterplots between each of our predictor variables

3.2.2 Full Model F-test

We then performed a full model F-test on our current model. Our null hypothesis for this test was as follow:

$$H_0: \beta_i = 0$$
$$H_A: At\ least\ one\ \beta_i \neq 0$$
$$Where\ i\ is\ one\ of\ our\ predictor\ variables$$

We began constructing an ANOVA table to calculate the F-statistic to be used in our test. The first step was calculating the sum of squares for error (SSE) and sum of squares for regression (SSR) (Appendix B Equations 2 and 3). Adding these together will give us the total corrected

sum of squared of the Y's (SST). Inputting the degrees of freedom, we get the following ANOVA table, where the F-statistic was calculated through appendix B equation 4:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F-Statistic
Regression	17	3.3139E15	1.949353E14	989.17
Residual	7837	1.5445E15	1.97078E11	
Total	7854	4.8584E15		

Table 3A: ANOVA table for global F-test on model

The raw ANOVA table R output can be found in appendix A figure 2 From the calculated F-statistic, we get a p-value extremely close to zero (R is outputting 2.2e-16). **This means that we can reject the null hypothesis and conclude that at least one of these predictors has an influence on selling price.** Note that our full model has an adjusted R^2 value of 0.6814, and a residual standard error of 443900.

3.2.3 Individual T-tests

We next performed individual t-tests on each of our predictor variables. Our hypotheses for these tests were as follows:

$$H_0: \beta_i = 0$$
$$H_A: \beta_i \neq 0$$
$$Where\ i\ is\ one\ of\ our\ predictor\ variables$$

The raw R output of the individual t-tests can be seen in appendix A figure 3. No variables had a p-value below our chosen alpha of 0.05, except for two qualitative variables which had other levels that were significant and so they were kept. **For all variables, we reject the null hypothesis and say that they have an influence on listing price.**

3.2.4 Stepwise, Forward, and Backward Regression

We next performed stepwise, forward, and backward regression. In stepwise regression, variables with the highest t-value are added one at a time (i.e., The best one-variable model, then the best two-variable model, etc.). Variables are removed if they become nonsignificant. In forward regression, the steps are the same, but there is no rechecking for if variables need to be removed. Finally, in backward elimination regression, the software begins with the full model, and drops them if the t-value is less than a critical value, one at a time.

Upon running all three of these tests, we achieved the same results as from our individual t-tests. No predictor variables were removed, and the full model was chosen as the best model. The full model outputted from these procedures can be seen in appendix A figure 4.

3.2.5 All-Possible-Regression Selection

The specific criteria we are looking at is the R^2, adjusted R^2, RMSE, Mallows' Cp Criterion, AIC (Akaike's information criterion), and BIC (Bayesian information criteria). We are going to see if the addition of more predictors impacts these 6 criteria (in the right direction). If the addition

does not cause these values to change, there may be a variable worth taking out. The formulae for these six variables can be found in appendix B equations 5 – 10. Below are plots of these 6 criteria's values as variables were added:

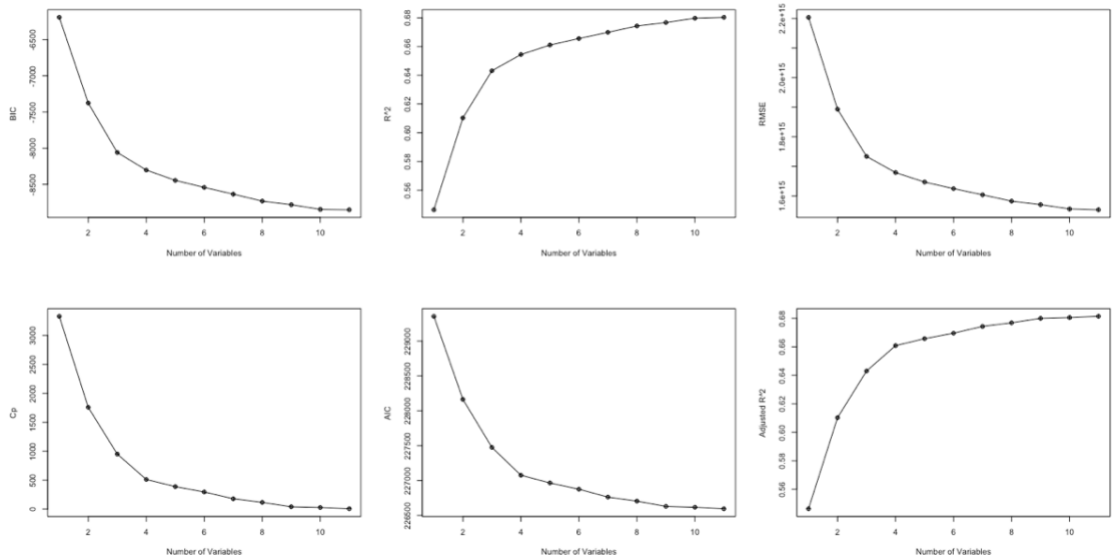


Figure 3C: Changes in chosen criteria as variables are added to the model

The raw R output for the changes in these criteria can be seen in appendix A figure 5. As can be seen from the graph, none of the lines become flat as more variables are being added. After about the fourth variable is added, the slope of the graph does become flatter, but not enough that we can say that the addition of more variables does not impact these values.

Looking at the raw R output, we see that seats does not significantly impact these metrics after it is added. Thus, we performed a partial F-test on seats with the following hypotheses:

$$H_0: \beta_{seats} = 0$$
$$H_A: \beta_{seats} \neq 0$$

The raw R output of the F-test can be seen in appendix A figure 6. We see that we get a p-value much less than 0.05, indicating that we keep the seats term. Thus, we still have not removed any terms from our full model. Our model with 11 predictors is the strongest model so far.

3.2.5 Final First Order Model

All our selection methods produced the same first order model, our initial model without any predictors removed. Thus, our model so far is as follows:

$$\widehat{sellingprice} = -64299291.9128 + (32093.5683)year - (0.9418)km\ driven$$
$$- (215195.8914)factor(fuel)diesel + (239869.4936)factor(fuel)LPG$$
$$- (1570.8221)factor(fuel)petrol$$
$$- (273160.0637)factor(seller\ type)individual$$
$$- (334283.3670)factor(seller\ type)dealer$$
$$- (427785.7085)factor(transmission)manual$$
$$+ (5248.2837)factor(owner)fourthOrAbove$$
$$- 43216.2798factors(owner)second$$
$$+ (1983617.3870)factor(owner)testDriveCar$$
$$- (13627.3172)factor(owner)third + (16920.7029)mileage$$
$$+ (129.3882)engine + (12247.3144)maxPower - (38131.3418)seats$$
$$- (120.6913)torque$$

With $R^2 = 0.6814$ and RMSE = 443900

3.3 Construction of Interactions Model

Our first step in constructing an interactions model was to create the full interactions model, containing all interaction terms. We seemed to be on the right track as our adjusted R^2 value had gone up to 0.9108, and RMSE had dropped to 234900. Performing individual t-test on our model, we get hypotheses of:

$$H_0: \beta_i = 0$$
$$H_A: \beta_i \neq 0$$
$$Where\ i\ is\ one\ of\ our\ predictor\ variables$$

Upon performing this test, we found that several interactions were insignificant, and removed these to produce a reduced interactions model with 32 terms. Retesting this reduced interactions model, we found that the interaction between owner and torque had become insignificant. We removed this term and performed a partial F-test on it, finding it not to be significant again. The raw R-output for both the F-test and final interaction model can be seen in appendix A figures 7 and 8. Note that this model has an adjusted **R^2 of 0.9085** and an **RMSE of 237900**.

3.4 Higher Order Models

Now, it was time to check for higher order terms. The first step in this process was to see which terms were most strongly correlated with our dependent variable. The raw R output of the correlations can be seen in appendix A figure 9.

Max power has the highest correlation with listing price, so this is the term we tested for higher order first. When making this term squared, we achieve constructing a model with a very slightly higher adjusted R^2 of 0.9086, and a slightly lower RMSE of 237900 (appendix A figure 10). We then made this term cubed, but this did not impact the adjusted R^2 or RMSE at all, and so we excluded this to ensure our model did not get too complicated (appendix A figure 11).

We then tested higher order terms for the engine variable. We added a squared term and saw that there was no change in adjusted R^2 or RMSE, and so we chose to leave it out.

We also decided to take out the max power squared term. It marginally changed R^2 and RMSE but added complications to an otherwise linear model. Because of this, we left it out.

Thus, we did not change our model at all since our interactions model.

3.5 Checking Assumptions

3.5.1 Linearity

The first assumption we checked was for linearity. There should be a straight-line relationship between the predictors and explanatory variable. To test this, we utilized a residual plot. We expect to see no pattern or relationship between the residuals and fitted values. Below is our residual plot:

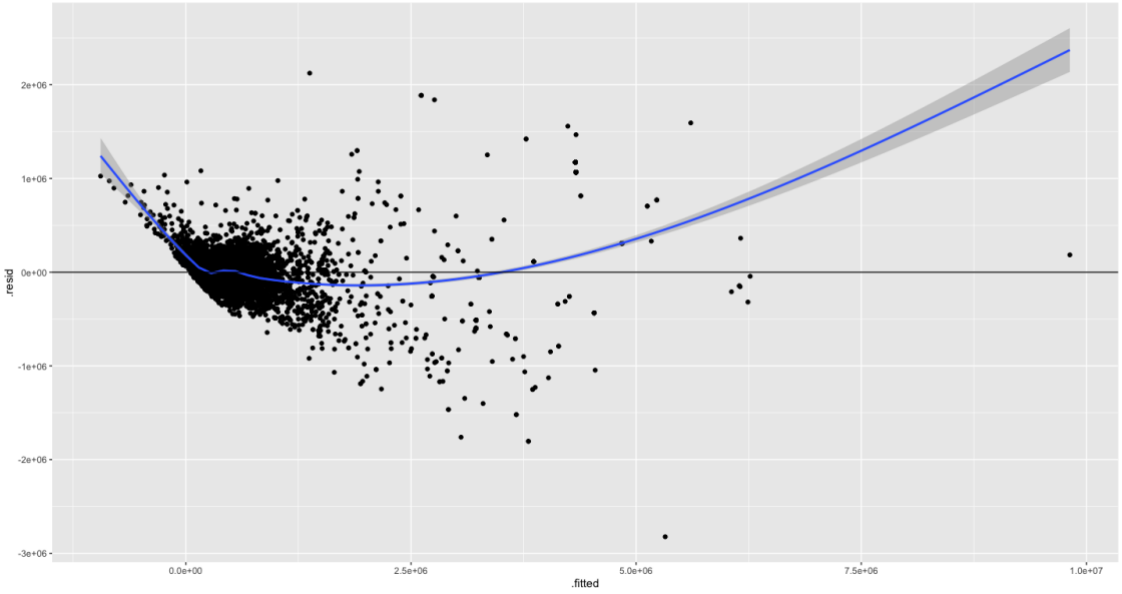


Figure 3D: Plot of residuals Vs. fitted values for data

As can be clearly seen from the graph, the data appears to be non-linear, as a clear parabola-like pattern is shown. This is a problem and indicates our model may not be statistically sound.

The Box Cox transformation, though primarily used to solve the issue of heteroscedasticity and nonnormality, has shown that it can aid in linearizing data. The first step to using this transformation is ensuring that our response variable is always positive, which it is in the case of listing price. Below is a residual plot of our data after applying this transformation:

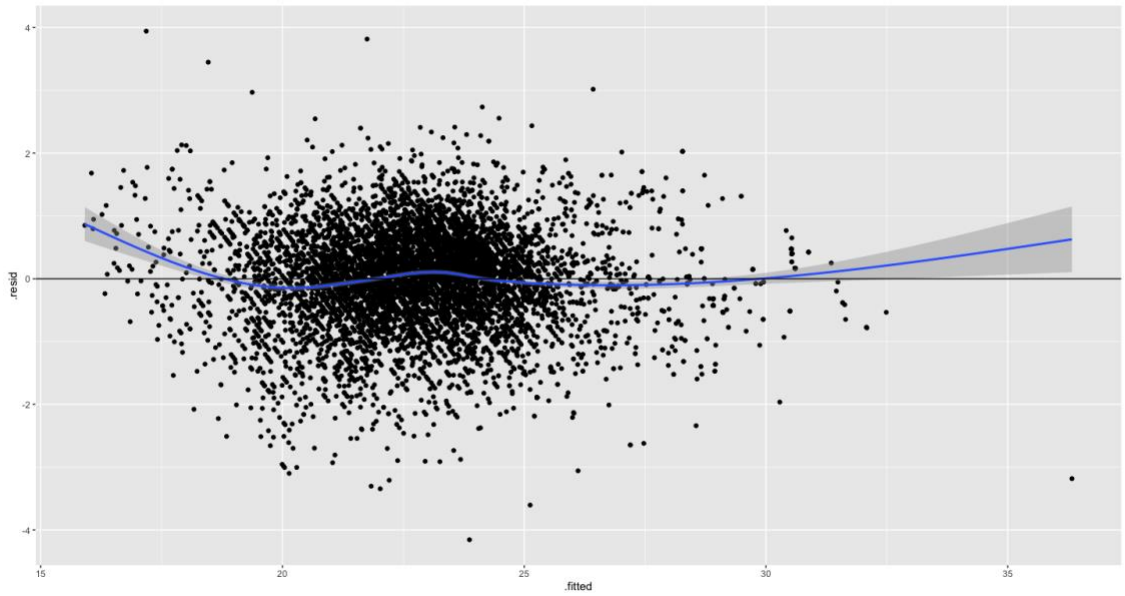


Figure 3E: Plot of residuals Vs. fitted values for data post Box-Cox transformation

As can be seen from the plot residual plot post transformation, the points are scattered a lot more uniformly, with a greatly less pronounced pattern. Thus, our final model is now the interactions model with the Box-Cox transformation applied, with an adjusted R^2 of 0.894. The raw R output of this model can be found in appendix A figure 12.

3.5.2 Outliers

To test for outliers, we utilized Cook's distance (D_i). D_i measures the effect of deleting a *single* given observation has on the model and coefficients. A high D_i means that the observed point has a strong influence on the coefficients, and thus may be an outlier. We check our Cook's distances with a residuals vs leverage plot. This would truly tell us if any points had high enough Cook's distances to be considered outliers. Below is our residuals vs. leverage plot:

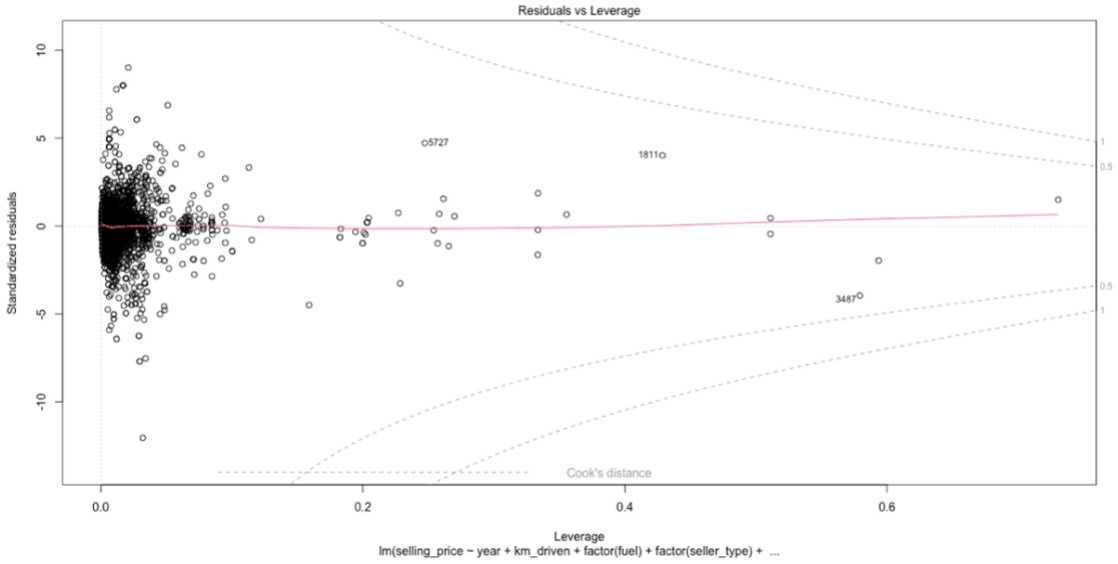


Figure 3F: Residuals Vs. Leverage plot of data

As can be seen, we have no points falling outside of the 0.5/1 lines. Thus, none of our points have reached the critical Cook's distance threshold to be considered outliers, and our model is not changed.

3.5.3 Homoscedasticity

We next aimed to check that our data was homoscedastic (i.e., the data points have similar variances). We would hope this is true, as otherwise the spread of the data can change depending on the measured values, which is an issue when trying to construct a linear model. We checked for homoscedasticity using the Breusch-Pagan test, which follows a chi-square distribution (appendix B equation 12). Below are our hypotheses:

$$H_0: Heteroscedasticity\ is\ not\ present$$
$$H_A: Heteroscedasticitiy\ is\ present$$

Upon performing the Breusch-Pagan test, we received a p-value well below 0.05. This indicates that we can reject the null hypothesis, suggesting that heteroscedastic data exists. This is an issue as our variance varies with the data points, again lowering the statistical power of our model. We will discuss this further in section 4.1. The raw R output of our test can be seen in appendix A figure 13.

3.5.4 Normality

Finally, we began testing for normality. The multiple linear regression analysis requires that the errors between observed and predicted values (i.e., the residuals of the regression) should be normally distributed We did this by performing a Shapiro-Wilk test, as well as a histogram of residuals and a standard Q-Q plot. Below are our hypotheses for our Shapiro-Wilk test:

$$H_0: The\ sample\ data\ is\ significantly\ normally\ distributed$$
$$H_A: The\ sample\ data\ is\ not\ significantly\ normally\ distributed$$

We received a p-value well below 0.05, indicating that we should reject the null hypothesis, and can say that the data is not significantly normally distributed. The raw R code output for the Shapiro-Wilk test can be seen in appendix A figure 14.

Below, is our Q-Q plot:

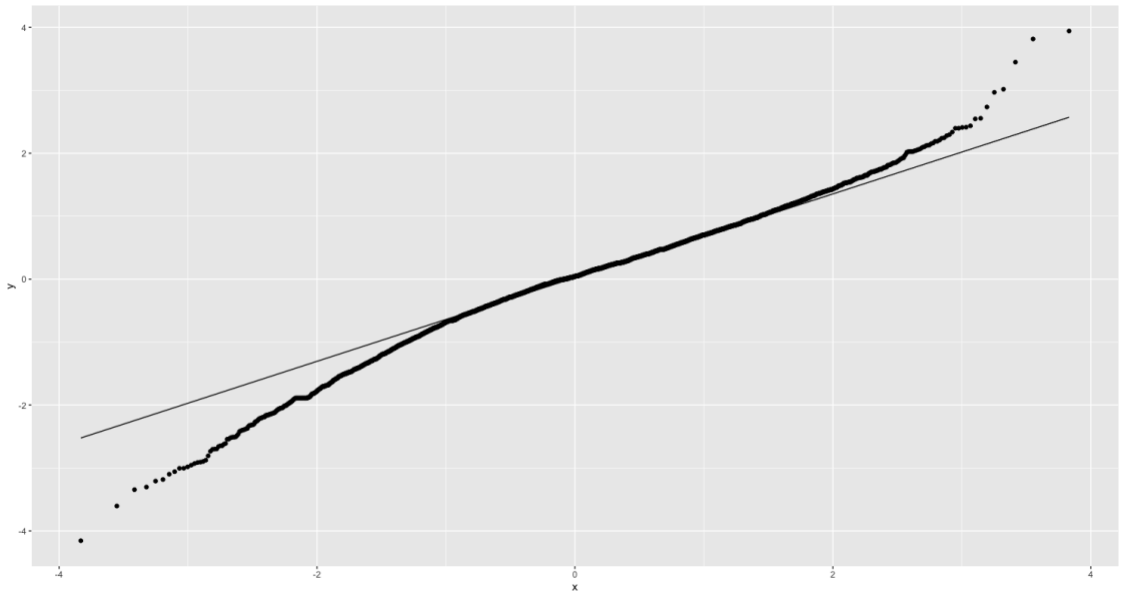


Figure 3G: Q-Q Normality Plot

From this plot, we can see many points not far from the center falling off the line. The data does not appear normal, further verifying our conclusion from our Shapiro-Wilk test.

Our histogram of residuals can be seen below:

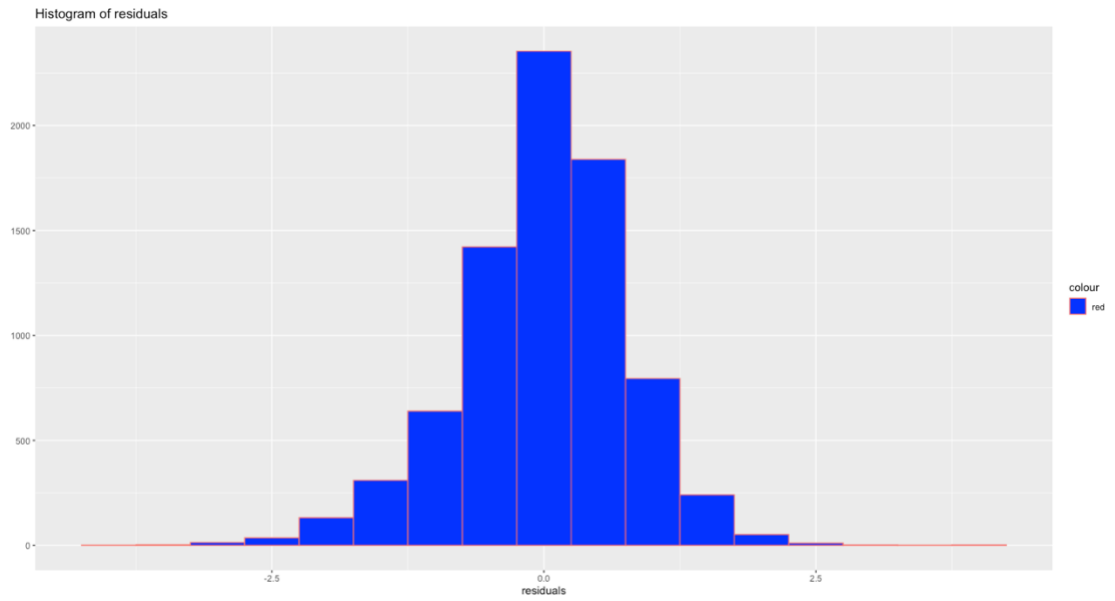


Figure 3H: Histogram of residuals of data

This histogram of residuals appears normal; however, our other tests have proven too strong to ignore. We can conclude that the data in not normally distributed, and the **normality assumption has not been met**. We cannot transform the model further, as a Box-Cox

transformation already been applied. Thus, it appears we are at a loss here. Limitations of this will be discussed in section 4.1.

4. Conclusion and Discussion

4.1 Approach

Below is the final roadmap we ended up following:

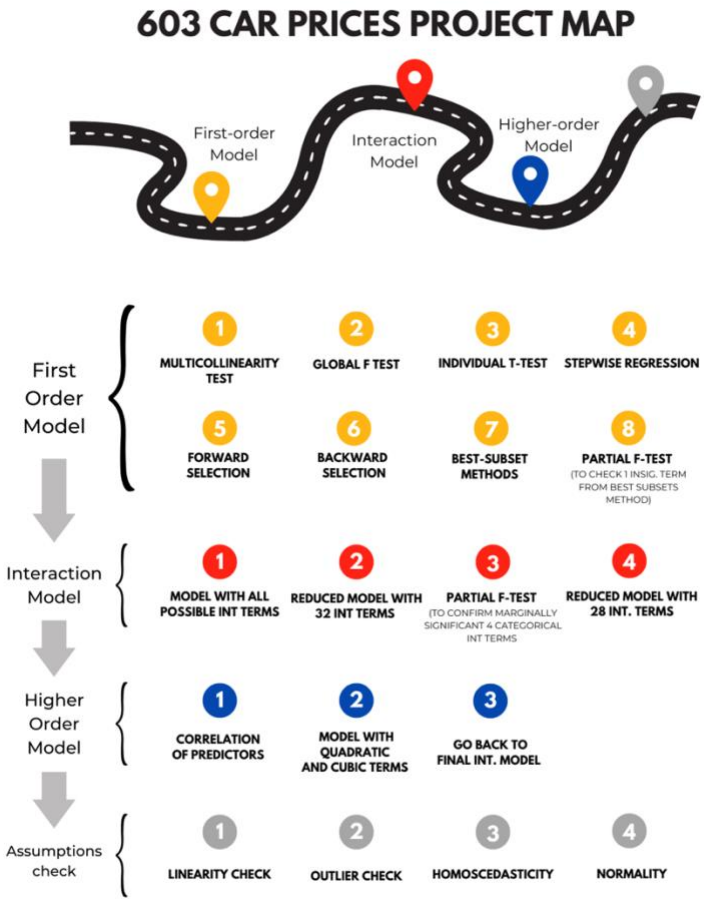


Figure 4A: Actual flow of work after following all steps

This differs slightly from our initial plan, in the cases of some partial f-tests done when the model was reduced, as well as some transformations. Overall, I think we had a strong approach, ensuring to double check variables being removed with partial f-tests. Additionally, the order that we constructed the model in was logically planned and ensured a strong model.

The main stand-out issue from our results came in the assumption checking. Both the assumptions of homoscedasticity and normality were not met. I will be discussing both issues below:

Previous research has shown that heteroscedasticity in data used for linear regression can lead to increased Type I error rates and decreased statistical power (Rosopa et al, 2016). It has also been said that heteroscedasticity is mainly bad when it is caused by other errors like nonlinearity,

which was not the case for our model after applying the Box-Cox transformation (Long and Laurie, 2000). Essentially, heteroscedasticity makes it so your regression estimator (the equation on the right side of the equal sign) is not the best linear unbiased estimator of the regression slopes, and thus the standard errors are incorrect (as well as their p-values when we perform t-tests on them). Thus, a possible solution to this would have been to utilize a robust standard error and carry on fitting the model (Noah, 2022). Overall, however, predictions made by our model should not be greatly impacted, with impact increasing the more extreme the value we are predicting for.

Normality is required for the very backbone of the assumptions of multiple linear regression, and its absence can lead to improper inferences (Schmidt and Finan, 2018). A possible solution to this (for future research) is to use the *glm* function in R, from the *stats* library, which can fit generalized linear models (ex. Gamma distributions with left skewness). Further, it has been shown that nonnormality is not a *huge* issue. Gauss-Markov theorem does not require normality, and the estimator is still the best linear unbiased estimator (John, 2020). However, it may be difficult to do inference, such as hypothesis testing and confidence intervals for finite sample sizes (John, 2020). These latter issues could be solved with bootstrap.

Additionally, we found that since the column "year" had so many repeating values, we may have been wrong in treating it is a numerical variable. Year is often treated as a categorical variable, something we did not do (Fnguyen, 2019). We decided to run individual t-tests again on the model keeping "year" as a factor (qualitative) variable. It was seen that the output showed that all the year terms were insignificant (the raw R output can be seen in appendix A figure 15). This did not make sense, as year seems like one of the most important things people look at when buying a car. Thus, we decided that, for future studies, it may be better to either convert the year column into ages (by doing 2022 – year), or to bin the ages into new, medium, and old.

Finally, we had a name column that had different sorts of formatting (make, model or model, make), which we dropped. In hindsight, we should have parsed this column and pulled out the car makes. This is another very important factor in used car listing price.

Overall, our assumptions not being met may not make our model the best possible model; however, it still holds statistical power, and can be improved by the suggestions made in this section.

4.2 Future Work

The first thing to consider with future work is what better data could be collected. I think that, rather than just having listings, it should be clearly indicated whether a car sold or not (possibly as a boolean data column "sold?"). People tend to inflate the value of their car when first listing it, and thus our data may have been biased as it did not indicate which of these cars sold at the given price (Pal et al., 2018). Our dependent variable is not so much used car selling price but used car *listing* price.

Our model showed many interaction terms. It may be insightful to study why certain terms interact with each other. Some have already been investigated, such as the parabolic relationship

between max power and torque (Seck et al., 1995); while others have not, such as the interaction between max power and number of seats.

Another possible future study would be to compare this to other used goods regression analyses. This could be compared to new goods to see if there is a difference in the way the industries should be dealt with.

4.3 Conclusion

In answering our first guiding question, we saw that all the predictors that we suspected ended up being important in determining a vehicle's listing price. Additionally, interactions between these variables make up most predictors.

In exploring our second guiding question, more investigation is needed before a concrete plan can come into play. Our regression model fell short of certain assumptions critical to make any concrete inferences.

Overall, both the listing price of used cars and construction of a multiple linear model are complex and have many moving parts. We are happy with our approach, and felt we performed the appropriate tests at the appropriate times.

5. References

Birla, Nehal, et al. “Vehicle Dataset” *Kaggle.com* <https://www.kaggle.com/datasets/nehalbirla/vehicle-dataset-from-cardekho?select=Car+details+v3.csv> (2020).

Car Dekho. “Car Dekho API (Public).” *Car Dekho* (2020).

Carlier, Mathilde. “Canada: Number of Licensed Drivers.” *Statista* (2021).

Chadha, Sunainaa. India's used-car market is booming and startups are Capitalising like never before. *The Times of India* (2021).

Ellencweig, Ben, et al. "Used cars, new platforms: Accelerating sales in a digitally disrupted market." *McKinsey & Company* (2019).

Fnguyen (https://datascience.stackexchange.com/users/79227/fnguyen), How to handle "year" variable for Machine Learning models, URL (version: 2019-26-11): https://datascience.stackexchange.com/questions/63785/how-to-handle-year-variable-for-machine-learning-models

Furcher, Thomas, et al. "How consumers’ behavior in car buying and mobility is changing amid COVID-19." (2020).

Gegic, Enis, et al. "Car price prediction using machine learning techniques." *TEM Journal* 8.1 (2019): 113.

Hoeft, Fabian. "The case of sales in the automotive industry during the COVID-19 pandemic." *Strategic Change* 30.2 (2021): 117-125.

JohnK (https://stats.stackexchange.com/users/31420/johnk), How incorrect is a regression model when assumptions are not met?, URL (version: 2020-06-11): https://stats.stackexchange.com/q/188666

Long, J. Scott, and Laurie H. Ervin. "Using heteroscedasticity consistent standard errors in the linear regression model." *The American Statistician* 54.3 (2000): 217-224.

Monburinon, Nitis, et al. "Prediction of prices for used car by using regression models." *2018 5th International Conference on Business and Industrial Research (ICBIR)*. IEEE, 2018.

Noah (https://stats.stackexchange.com/users/116195/noah), Linear regression's (OLS) coefficient interpretation with heteroscedasticity, URL (version: 2022-05-23): <https://stats.stackexchange.com/q/576218>

Pal, Nabarun, et al. "How much is my car worth? A methodology for predicting used cars' prices using random forest." *Future of Information and Communication Conference*. Springer, Cham, 2018.

Rosopa, Patrick J., Meline M. Schaffer, and Amber N. Schroeder. "Managing heteroscedasticity in general linear models." *Psychological Methods* 18.3 (2013): 335.

Schmidt, Amand F., and Chris Finan. "Linear regression and the normality assumption." *Journal of clinical epidemiology* 98 (2018): 146-151.

Seck, D., et al. "Maximal power and torque-velocity relationship on a cycle ergometer during the acceleration phase of a single all-out exercise." *European journal of applied physiology and occupational physiology* 70.2 (1995): 161-168.

Sun, Pengfei, et al. "Understanding of COVID-19 based on current evidence." *Journal of medical virology* 92.6 (2020): 548-551.

6. Appendix

A: Raw R Output

VIF Multicollinearity Diagnostics

	VIF	detection
year	2.0953	0
km_driven	1.4386	0
factor(fuel)Diesel	40.2806	1
factor(fuel)LPG	1.6951	0
factor(fuel)Petrol	39.1833	1
factor(seller_type)Individual	1.4686	0
factor(seller_type)Trustmark Dealer	1.2542	0
factor(transmission)Manual	1.6409	0
factor(owner)Fourth & Above Owner	1.1005	0
factor(owner)Second Owner	1.2754	0
factor(owner)Test Drive Car	1.0109	0
factor(owner)Third Owner	1.1996	0
mileage	3.0548	0
engine	5.5554	0
max_power	3.2619	0
seats	2.3255	0
torque_rpm	4.2029	0

Figure 1: Raw VIF outputs on full model

```
Analysis of Variance Table

Model 1: selling_price ~ 1
Model 2: selling_price ~ year + km_driven + factor(fuel) + factor(seller_type) +
    factor(transmission) + factor(owner) + mileage + engine +
    max_power + seats + torque_rpm
  Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
1   7854 4.8584e+15
2   7837 1.5445e+15 17 3.3139e+15 989.17 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 2: ANOVA table for Global F-test on full model

```
Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                     -6.430e+07  3.756e+06 -17.118  < 2e-16 ***
year                             3.209e+04  1.879e+03  17.077  < 2e-16 ***
km_driven                       -9.418e-01  1.057e-01  -8.913  < 2e-16 ***
factor(fuel)Diesel              -2.152e+05  6.381e+04  -3.373 0.000748 ***
factor(fuel)LPG                  2.399e+05  9.791e+04   2.450 0.014316 *
factor(fuel)Petrol              -1.571e+03  6.307e+04  -0.025 0.980130
factor(seller_type)Individual   -2.732e+05  1.615e+04 -16.912  < 2e-16 ***
factor(seller_type)Trustmark Dealer -3.343e+05  3.286e+04 -10.173  < 2e-16 ***
factor(transmission)Manual      -4.278e+05  1.924e+04 -22.233  < 2e-16 ***
factor(owner)Fourth & Above Owner  5.248e+03  3.731e+04   0.141 0.888142
factor(owner)Second Owner       -4.322e+04  1.296e+04  -3.335 0.000858 ***
factor(owner)Test Drive Car      1.984e+06  1.997e+05   9.934  < 2e-16 ***
factor(owner)Third Owner        -1.363e+04  2.231e+04  -0.611 0.541258
mileage                          1.692e+04  2.219e+03   7.624 2.75e-14 ***
engine                           1.294e+02  2.346e+01   5.515 3.59e-08 ***
max_power                        1.225e+04  2.570e+02  47.646  < 2e-16 ***
seats                           -3.813e+04  7.955e+03  -4.793 1.67e-06 ***
torque_rpm                      -1.207e+02  9.021e+00 -13.380  < 2e-16 ***
---
```

Figure 3: Results from individual t-tests on full model predictors

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	-64299291.913	3756136.979		-17.118	0.000	-71662322.272	-56936261.553
factor(seller_type)Individual	-273160.064	16151.789	-0.131	-16.912	0.000	-304821.879	-241498.249
factor(seller_type)Trustmark Dealer	-334283.367	32860.037	-0.073	-10.173	0.000	-398697.805	-269868.929
factor(transmission)Manual	-427785.709	19241.410	-0.181	-22.233	0.000	-465504.005	-390067.412
engine	129.388	23.460	0.083	5.515	0.000	83.400	175.376
max_power	12247.314	257.046	0.548	47.646	0.000	11743.436	12751.192
year	32093.568	1879.369	0.157	17.077	0.000	28409.504	35777.632
torque_rpm	-120.691	9.021	-0.175	-13.380	0.000	-138.374	-103.009
factor(owner)Fourth & Above Owner	5248.284	37311.756	0.001	0.141	0.888	-67892.710	78389.277
factor(owner)Second Owner	-43216.280	12959.253	-0.024	-3.335	0.001	-68619.873	-17812.687
factor(owner)Test Drive Car	1983617.387	199670.648	0.064	9.934	0.000	1592209.658	2375025.116
factor(owner)Third Owner	-13627.317	22305.549	-0.004	-0.611	0.541	-57352.143	30097.508
km_driven	-0.942	0.106	-0.068	-8.913	0.000	-1.149	-0.735
seats	-38131.342	7955.381	-0.047	-4.793	0.000	-53726.010	-22536.674
factor(fuel)Diesel	-215195.891	63807.264	-0.136	-3.373	0.001	-340275.148	-90116.635
factor(fuel)LPG	239869.494	97914.837	0.020	2.450	0.014	47930.296	431808.691
factor(fuel)Petrol	-1570.822	63066.708	-0.001	-0.025	0.980	-125198.392	122056.748
mileage	16920.703	2219.399	0.085	7.624	0.000	12570.089	21271.317

Figure 4: Output model from stepwise regression, backward elimination regression, and foreword regression

```
             cp       BIC         RMSE AdjustedR   rsquare      AIC
 [1,] 3331.20899 -6191.965 2.203705e+15 0.5463546 0.5464124 229355.4
 [2,] 1757.25232 -7376.248 1.893128e+15 0.6102392 0.6103384 228164.2
 [3,]  950.58254 -8058.025 1.733761e+15 0.6430044 0.6431408 227475.4
 [4,]  674.71311 -8301.157 1.679001e+15 0.6542360 0.6544121 227075.8
 [5,]  513.94464 -8443.710 1.646924e+15 0.6607986 0.6610145 226965.1
 [6,]  403.61864 -8541.036 1.624787e+15 0.6653151 0.6655708 226875.5
 [7,]  299.50357 -8633.824 1.603875e+15 0.6695807 0.6698752 226762.1
 [8,]  192.79794 -8730.481 1.582452e+15 0.6739526 0.6742847 226704.1
 [9,]  135.23129 -8779.999 1.570713e+15 0.6763300 0.6767009 226628.9
[10,]   63.05667 -8844.475 1.556095e+15 0.6793013 0.6797097 226616.2
[11,]   50.36254 -8850.137 1.553200e+15 0.6798573 0.6803057 226595.2
```

Figure 5: Raw R output showing changes in chosen criteria as variables are added

```
Analysis of Variance Table

Model 1: selling_price ~ year + km_driven + factor(fuel) + factor(seller_type) +
    factor(transmission) + factor(owner) + mileage + engine +
    max_power + torque_rpm
Model 2: selling_price ~ year + km_driven + factor(fuel) + factor(seller_type) +
    factor(transmission) + factor(owner) + mileage + engine +
    max_power + seats + torque_rpm
  Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
1   7838 1.5490e+15
2   7837 1.5445e+15  1 4.5276e+12 22.974 1.672e-06 ***
---
```

Figure 6: Raw R output of ANOVA table from partial F-test on seats

```
Analysis of Variance Table

Model 1: selling_price ~ year + km_driven + factor(fuel) + factor(seller_type) +
    factor(transmission) + factor(owner) + mileage + engine +
    max_power + seats + torque_rpm + year * km_driven + year *
    factor(seller_type) + year * factor(transmission) + year *
    mileage + year * engine + year * max_power + year * torque_rpm +
    km_driven * factor(seller_type) + km_driven * factor(transmission) +
    km_driven * max_power + factor(fuel) * factor(transmission) +
    factor(fuel) * factor(owner) + factor(seller_type) * factor(transmission) +
    factor(seller_type) * factor(owner) + factor(seller_type) *
    mileage + factor(seller_type) * engine + factor(seller_type) *
    max_power + factor(transmission) * factor(owner) + factor(transmission) *
    mileage + factor(transmission) * engine + factor(transmission) *
    max_power + factor(transmission) * seats + factor(transmission) *
    torque_rpm + factor(owner) * mileage + factor(owner) * engine +
    mileage * engine + mileage * max_power + engine * max_power +
    engine * torque_rpm + max_power * seats + max_power * torque_rpm
Model 2: selling_price ~ year + km_driven + factor(fuel) + factor(seller_type) +
    factor(transmission) + factor(owner) + mileage + engine +
    max_power + seats + torque_rpm + year * km_driven + year *
    factor(seller_type) + year * factor(transmission) + year *
    mileage + year * engine + year * max_power + year * torque_rpm +
    km_driven * factor(seller_type) + km_driven * factor(transmission) +
    km_driven * max_power + factor(fuel) * factor(transmission) +
    factor(fuel) * factor(owner) + factor(seller_type) * factor(transmission) +
    factor(seller_type) * factor(owner) + factor(seller_type) *
    mileage + factor(seller_type) * engine + factor(seller_type) *
    max_power + factor(transmission) * factor(owner) + factor(transmission) *
    mileage + factor(transmission) * engine + factor(transmission) *
    max_power + factor(transmission) * seats + factor(transmission) *
    torque_rpm + factor(owner) * mileage + factor(owner) * engine +
    factor(owner) * torque_rpm + mileage * engine + mileage *
    max_power + engine * max_power + engine * torque_rpm + max_power *
    seats + max_power * torque_rpm
  Res.Df        RSS Df  Sum of Sq      F  Pr(>F)
1   7782 4.4059e+14
2   7779 4.4020e+14  3 3.8886e+11 2.2906 0.07619 .
---
```

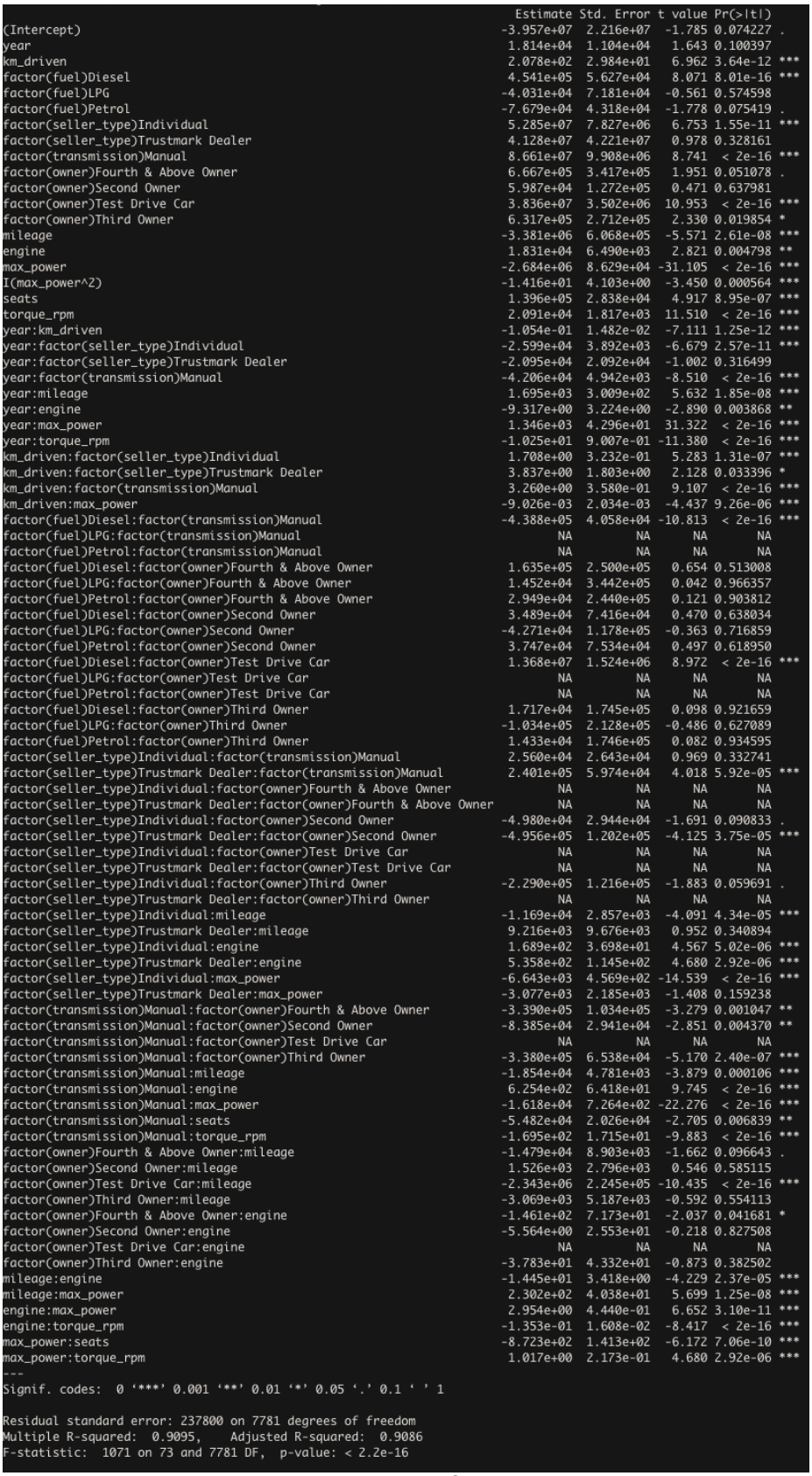
Figure 7: Raw R output of ANOVA table from partial F-test on interaction between owner and torque

```
Coefficients: (11 not defined because of singularities)
                                                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                                                     -4.410e+07  2.214e+07  -1.992 0.046412 *
year                                                             2.043e+04  1.103e+04   1.852 0.064011 .
km_driven                                                        2.099e+02  2.986e+01   7.031 2.22e-12 ***
factor(fuel)Diesel                                               4.621e+05  5.626e+04   8.213 2.50e-16 ***
factor(fuel)LPG                                                 -3.421e+04  7.184e+04  -0.476 0.633916
factor(fuel)Petrol                                              -6.874e+04  4.315e+04  -1.593 0.111183
factor(seller_type)Individual                                    5.169e+07  7.825e+06   6.606 4.21e-11 ***
factor(seller_type)Trustmark Dealer                              4.170e+07  4.224e+07   0.987 0.323559
factor(transmission)Manual                                       9.040e+07  9.854e+06   9.174  < 2e-16 ***
factor(owner)Fourth & Above Owner                                6.515e+05  3.419e+05   1.906 0.056738 .
factor(owner)Second Owner                                        4.965e+04  1.273e+05   0.390 0.696527
factor(owner)Test Drive Car                                      3.817e+07  3.504e+06  10.893  < 2e-16 ***
factor(owner)Third Owner                                         6.300e+05  2.714e+05   2.321 0.020287 *
mileage                                                         -3.408e+06  6.072e+05  -5.613 2.05e-08 ***
engine                                                           1.869e+04  6.493e+03   2.879 0.004005 **
max_power                                                       -2.664e+06  8.616e+04 -30.922  < 2e-16 ***
seats                                                            1.073e+05  2.682e+04   4.001 6.37e-05 ***
torque_rpm                                                       2.099e+04  1.818e+03  11.547  < 2e-16 ***
year:km_driven                                                  -1.065e-01  1.483e-02  -7.181 7.54e-13 ***
year:factor(seller_type)Individual                              -2.539e+04  3.890e+03  -6.526 7.18e-11 ***
year:factor(seller_type)Trustmark Dealer                        -2.114e+04  2.093e+04  -1.010 0.312619
year:factor(transmission)Manual                                 -4.400e+04  4.913e+03  -8.955  < 2e-16 ***
year:mileage                                                     1.709e+03  3.011e+02   5.677 1.42e-08 ***
year:engine                                                     -9.416e+00  3.227e+00  -2.918 0.003531 **
year:max_power                                                   1.335e+03  4.287e+01  31.130  < 2e-16 ***
year:torque_rpm                                                 -1.029e+01  9.012e-01 -11.418  < 2e-16 ***
km_driven:factor(seller_type)Individual                          1.755e+00  3.231e-01   5.432 5.75e-08 ***
km_driven:factor(seller_type)Trustmark Dealer                    3.866e+00  1.805e+00   2.142 0.032220 *
km_driven:factor(transmission)Manual                             3.210e+00  3.580e-01   8.968  < 2e-16 ***
km_driven:max_power                                             -8.784e-03  2.035e-03  -4.317 1.60e-05 ***
factor(fuel)Diesel:factor(transmission)Manual                   -4.418e+05  4.060e+04 -10.880  < 2e-16 ***
factor(fuel)LPG:factor(transmission)Manual                             NA         NA      NA       NA
factor(fuel)Petrol:factor(transmission)Manual                          NA         NA      NA       NA
factor(fuel)Diesel:factor(owner)Fourth & Above Owner             1.540e+05  2.501e+05   0.615 0.538269
factor(fuel)LPG:factor(owner)Fourth & Above Owner                6.880e+03  3.444e+05   0.020 0.984062
factor(fuel)Petrol:factor(owner)Fourth & Above Owner             2.126e+04  2.441e+05   0.087 0.930614
factor(fuel)Diesel:factor(owner)Second Owner                     3.694e+04  7.421e+04   0.498 0.618620
factor(fuel)LPG:factor(owner)Second Owner                       -3.817e+04  1.178e+05  -0.324 0.745973
factor(fuel)Petrol:factor(owner)Second Owner                     3.949e+04  7.539e+04   0.524 0.600422
factor(fuel)Diesel:factor(owner)Test Drive Car                   1.360e+07  1.525e+06   8.918  < 2e-16 ***
factor(fuel)LPG:factor(owner)Test Drive Car                            NA         NA      NA       NA
factor(fuel)Petrol:factor(owner)Test Drive Car                         NA         NA      NA       NA
factor(fuel)Diesel:factor(owner)Third Owner                      1.156e+04  1.747e+05   0.066 0.947250
factor(fuel)LPG:factor(owner)Third Owner                        -1.054e+05  2.129e+05  -0.495 0.620524
factor(fuel)Petrol:factor(owner)Third Owner                      9.407e+03  1.747e+05   0.054 0.957061
factor(seller_type)Individual:factor(transmission)Manual         2.746e+04  2.644e+04   1.038 0.299166
factor(seller_type)Trustmark Dealer:factor(transmission)Manual   2.399e+05  5.978e+04   4.013 6.04e-05 ***
factor(seller_type)Individual:factor(owner)Fourth & Above Owner        NA         NA      NA       NA
factor(seller_type)Trustmark Dealer:factor(owner)Fourth & Above Owner  NA         NA      NA       NA
factor(seller_type)Individual:factor(owner)Second Owner         -4.810e+04  2.946e+04  -1.633 0.102540
factor(seller_type)Trustmark Dealer:factor(owner)Second Owner   -4.833e+05  1.202e+05  -4.021 5.85e-05 ***
factor(seller_type)Individual:factor(owner)Test Drive Car              NA         NA      NA       NA
factor(seller_type)Trustmark Dealer:factor(owner)Test Drive Car        NA         NA      NA       NA
factor(seller_type)Individual:factor(owner)Third Owner          -2.314e+05  1.217e+05  -1.902 0.057189 .
factor(seller_type)Trustmark Dealer:factor(owner)Third Owner           NA         NA      NA       NA
factor(seller_type)Individual:mileage                           -1.343e+04  2.814e+03  -4.775 1.83e-06 ***
factor(seller_type)Trustmark Dealer:mileage                      7.107e+03  9.664e+03   0.735 0.462135
factor(seller_type)Individual:engine                             1.630e+02  3.696e+01   4.409 1.05e-05 ***
factor(seller_type)Trustmark Dealer:engine                       5.313e+02  1.146e+02   4.637 3.59e-06 ***
factor(seller_type)Individual:max_power                         -6.796e+03  4.551e+02 -14.934  < 2e-16 ***
factor(seller_type)Trustmark Dealer:max_power                   -3.118e+03  2.187e+03  -1.426 0.154007
factor(transmission)Manual:factor(owner)Fourth & Above Owner    -3.295e+05  1.034e+05  -3.187 0.001445 **
factor(transmission)Manual:factor(owner)Second Owner            -8.297e+04  2.943e+04  -2.819 0.004829 **
factor(transmission)Manual:factor(owner)Test Drive Car                 NA         NA      NA       NA
factor(transmission)Manual:factor(owner)Third Owner             -3.364e+05  6.542e+04  -5.141 2.80e-07 ***
factor(transmission)Manual:mileage                              -1.735e+04  4.772e+03  -3.635 0.000279 ***
factor(transmission)Manual:engine                                5.098e+02  5.478e+01   9.307  < 2e-16 ***
factor(transmission)Manual:max_power                            -1.450e+04  5.400e+02 -26.858  < 2e-16 ***
factor(transmission)Manual:seats                                -3.607e+04  1.954e+04  -1.846 0.064863 .
factor(transmission)Manual:torque_rpm                           -1.685e+02  1.716e+01  -9.814  < 2e-16 ***
factor(owner)Fourth & Above Owner:mileage                       -1.446e+04  8.909e+03  -1.623 0.104547
factor(owner)Second Owner:mileage                                1.648e+03  2.797e+03   0.589 0.555839
factor(owner)Test Drive Car:mileage                             -2.331e+06  2.247e+05 -10.376  < 2e-16 ***
factor(owner)Third Owner:mileage                                -2.785e+03  5.190e+03  -0.537 0.591576
factor(owner)Fourth & Above Owner:engine                        -1.385e+02  7.174e+01  -1.930 0.053590 .
factor(owner)Second Owner:engine                                -3.141e+00  2.554e+01  -0.123 0.902132
factor(owner)Test Drive Car:engine                                     NA         NA      NA       NA
factor(owner)Third Owner:engine                                 -3.611e+01  4.335e+01  -0.833 0.404819
mileage:engine                                                  -1.241e+01  3.368e+00  -3.684 0.000231 ***
mileage:max_power                                                1.789e+02  3.758e+01   4.761 1.96e-06 ***
engine:max_power                                                 1.768e+00  2.815e-01   6.281 3.54e-10 ***
engine:torque_rpm                                               -1.366e-01  1.609e-02  -8.492  < 2e-16 ***
max_power:seats                                                 -6.986e+02  1.321e+02  -5.287 1.28e-07 ***
max_power:torque_rpm                                             1.027e+00  2.174e-01   4.722 2.38e-06 ***
---
```

Figure 8: Summary of final interactions model

```
[1] "year 0.406275"
[1] "km_driven -0.216168"
[1] "Mileage -0.125685"
[1] "Engine 0.454658"
[1] "max_power 0.739197"
[1] "seats 0.051179"
[1] "torque -0.195415"
```

Figure 9: Correlations of independent variables with listing price to check for higher order terms



	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-3.957e+07	2.216e+07	-1.785	0.074227	.
year	1.814e+04	1.104e+04	1.643	0.100397	
km_driven	2.078e+02	2.984e+01	6.962	3.64e-12	***
factor(fuel)Diesel	4.541e+05	5.627e+04	8.071	8.01e-16	***
factor(fuel)LPG	-4.031e+04	7.181e+04	-0.561	0.574598	
factor(fuel)Petrol	-7.679e+04	4.318e+04	-1.778	0.075419	.
factor(seller_type)Individual	5.285e+07	7.827e+06	6.753	1.55e-11	***
factor(seller_type)Trustmark Dealer	4.128e+07	4.221e+07	0.978	0.328161	
factor(transmission)Manual	8.661e+07	9.908e+06	8.741	< 2e-16	***
factor(owner)Fourth & Above Owner	6.667e+05	3.417e+05	1.951	0.051078	.
factor(owner)Second Owner	5.987e+04	1.272e+05	0.471	0.637981	
factor(owner)Test Drive Car	3.836e+07	3.502e+06	10.953	< 2e-16	***
factor(owner)Third Owner	6.317e+05	2.712e+05	2.330	0.019854	*
mileage	-3.381e+06	6.068e+05	-5.571	2.61e-08	***
engine	1.831e+04	6.490e+03	2.821	0.004798	**
max_power	-2.684e+06	8.629e+04	-31.105	< 2e-16	***
I(max_power^2)	-1.416e+01	4.103e+00	-3.450	0.000564	***
seats	1.396e+05	2.838e+04	4.917	8.95e-07	***
torque_rpm	2.091e+04	1.817e+03	11.510	< 2e-16	***
year:km_driven	-1.054e-01	1.482e-02	-7.111	1.25e-12	***
year:factor(seller_type)Individual	-2.599e+04	3.892e+03	-6.679	2.57e-11	***
year:factor(seller_type)Trustmark Dealer	-2.095e+04	2.092e+04	-1.002	0.316499	
year:factor(transmission)Manual	-4.206e+04	4.942e+03	-8.510	< 2e-16	***
year:mileage	1.695e+03	3.009e+02	5.632	1.85e-08	***
year:engine	-9.317e+00	3.224e+00	-2.890	0.003868	**
year:max_power	1.346e+03	4.296e+01	31.322	< 2e-16	***
year:torque_rpm	-1.025e+01	9.007e-01	-11.380	< 2e-16	***
km_driven:factor(seller_type)Individual	1.708e+00	3.232e-01	5.283	1.31e-07	***
km_driven:factor(seller_type)Trustmark Dealer	3.837e+00	1.803e+00	2.128	0.033396	*
km_driven:factor(transmission)Manual	3.260e+00	3.580e-01	9.107	< 2e-16	***
km_driven:max_power	-9.026e-03	2.034e-03	-4.437	9.26e-06	***
factor(fuel)Diesel:factor(transmission)Manual	-4.388e+05	4.058e+04	-10.813	< 2e-16	***
factor(fuel)LPG:factor(transmission)Manual	NA	NA	NA	NA	
factor(fuel)Petrol:factor(transmission)Manual	NA	NA	NA	NA	
factor(fuel)Diesel:factor(owner)Fourth & Above Owner	1.635e+05	2.500e+05	0.654	0.513008	
factor(fuel)LPG:factor(owner)Fourth & Above Owner	1.452e+04	3.442e+05	0.042	0.966357	
factor(fuel)Petrol:factor(owner)Fourth & Above Owner	2.949e+04	2.440e+05	0.121	0.903812	
factor(fuel)Diesel:factor(owner)Second Owner	3.489e+04	7.416e+04	0.470	0.638034	
factor(fuel)LPG:factor(owner)Second Owner	-4.271e+04	1.178e+05	-0.363	0.716859	
factor(fuel)Petrol:factor(owner)Second Owner	3.747e+04	7.534e+04	0.497	0.618950	
factor(fuel)Diesel:factor(owner)Test Drive Car	1.368e+07	1.524e+06	8.972	< 2e-16	***
factor(fuel)LPG:factor(owner)Test Drive Car	NA	NA	NA	NA	
factor(fuel)Petrol:factor(owner)Test Drive Car	NA	NA	NA	NA	
factor(fuel)Diesel:factor(owner)Third Owner	1.717e+04	1.745e+05	0.098	0.921659	
factor(fuel)LPG:factor(owner)Third Owner	-1.034e+05	2.128e+05	-0.486	0.627089	
factor(fuel)Petrol:factor(owner)Third Owner	1.433e+04	1.746e+05	0.082	0.934595	
factor(seller_type)Individual:factor(transmission)Manual	2.560e+04	2.643e+04	0.969	0.332741	
factor(seller_type)Trustmark Dealer:factor(transmission)Manual	2.401e+05	5.974e+04	4.018	5.92e-05	***
factor(seller_type)Individual:factor(owner)Fourth & Above Owner	NA	NA	NA	NA	
factor(seller_type)Trustmark Dealer:factor(owner)Fourth & Above Owner	NA	NA	NA	NA	
factor(seller_type)Individual:factor(owner)Second Owner	-4.980e+04	2.944e+04	-1.691	0.090833	.
factor(seller_type)Trustmark Dealer:factor(owner)Second Owner	-4.956e+05	1.202e+05	-4.125	3.75e-05	***
factor(seller_type)Individual:factor(owner)Test Drive Car	NA	NA	NA	NA	
factor(seller_type)Trustmark Dealer:factor(owner)Test Drive Car	NA	NA	NA	NA	
factor(seller_type)Individual:factor(owner)Third Owner	-2.290e+05	1.216e+05	-1.883	0.059691	.
factor(seller_type)Trustmark Dealer:factor(owner)Third Owner	NA	NA	NA	NA	
factor(seller_type)Individual:mileage	-1.169e+04	2.857e+03	-4.091	4.34e-05	***
factor(seller_type)Trustmark Dealer:mileage	9.216e+03	9.676e+03	0.952	0.340894	
factor(seller_type)Individual:engine	1.689e+02	3.698e+01	4.567	5.02e-06	***
factor(seller_type)Trustmark Dealer:engine	5.358e+02	1.145e+02	4.680	2.92e-06	***
factor(seller_type)Individual:max_power	-6.643e+03	4.569e+02	-14.539	< 2e-16	***
factor(seller_type)Trustmark Dealer:max_power	-3.077e+03	2.185e+03	-1.408	0.159238	
factor(transmission)Manual:factor(owner)Fourth & Above Owner	-3.390e+05	1.034e+05	-3.279	0.001047	**
factor(transmission)Manual:factor(owner)Second Owner	-8.385e+04	2.941e+04	-2.851	0.004370	**
factor(transmission)Manual:factor(owner)Test Drive Car	NA	NA	NA	NA	
factor(transmission)Manual:factor(owner)Third Owner	-3.380e+05	6.538e+04	-5.170	2.40e-07	***
factor(transmission)Manual:mileage	-1.854e+04	4.781e+03	-3.879	0.000106	***
factor(transmission)Manual:engine	6.254e+02	6.418e+01	9.745	< 2e-16	***
factor(transmission)Manual:max_power	-1.618e+04	7.264e+02	-22.276	< 2e-16	***
factor(transmission)Manual:seats	-5.482e+04	2.026e+04	-2.705	0.006839	**
factor(transmission)Manual:torque_rpm	-1.695e+02	1.715e+01	-9.883	< 2e-16	***
factor(owner)Fourth & Above Owner:mileage	-1.479e+04	8.903e+03	-1.662	0.096643	.
factor(owner)Second Owner:mileage	1.526e+03	2.796e+03	0.546	0.585115	
factor(owner)Test Drive Car:mileage	-2.343e+06	2.245e+05	-10.435	< 2e-16	***
factor(owner)Third Owner:mileage	-3.069e+03	5.187e+03	-0.592	0.554113	
factor(owner)Fourth & Above Owner:engine	-1.461e+02	7.173e+01	-2.037	0.041681	*
factor(owner)Second Owner:engine	-5.564e+00	2.553e+01	-0.218	0.827508	
factor(owner)Test Drive Car:engine	NA	NA	NA	NA	
factor(owner)Third Owner:engine	-3.783e+01	4.332e+01	-0.873	0.382502	
mileage:engine	-1.445e+01	3.418e+00	-4.229	2.37e-05	***
mileage:max_power	2.302e+02	4.038e+01	5.699	1.25e-08	***
engine:max_power	2.954e+00	4.440e-01	6.652	3.10e-11	***
engine:torque_rpm	-1.353e-01	1.608e-02	-8.417	< 2e-16	***
max_power:seats	-8.723e+02	1.413e+02	-6.172	7.06e-10	***
max_power:torque_rpm	1.017e+00	2.173e-01	4.680	2.92e-06	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

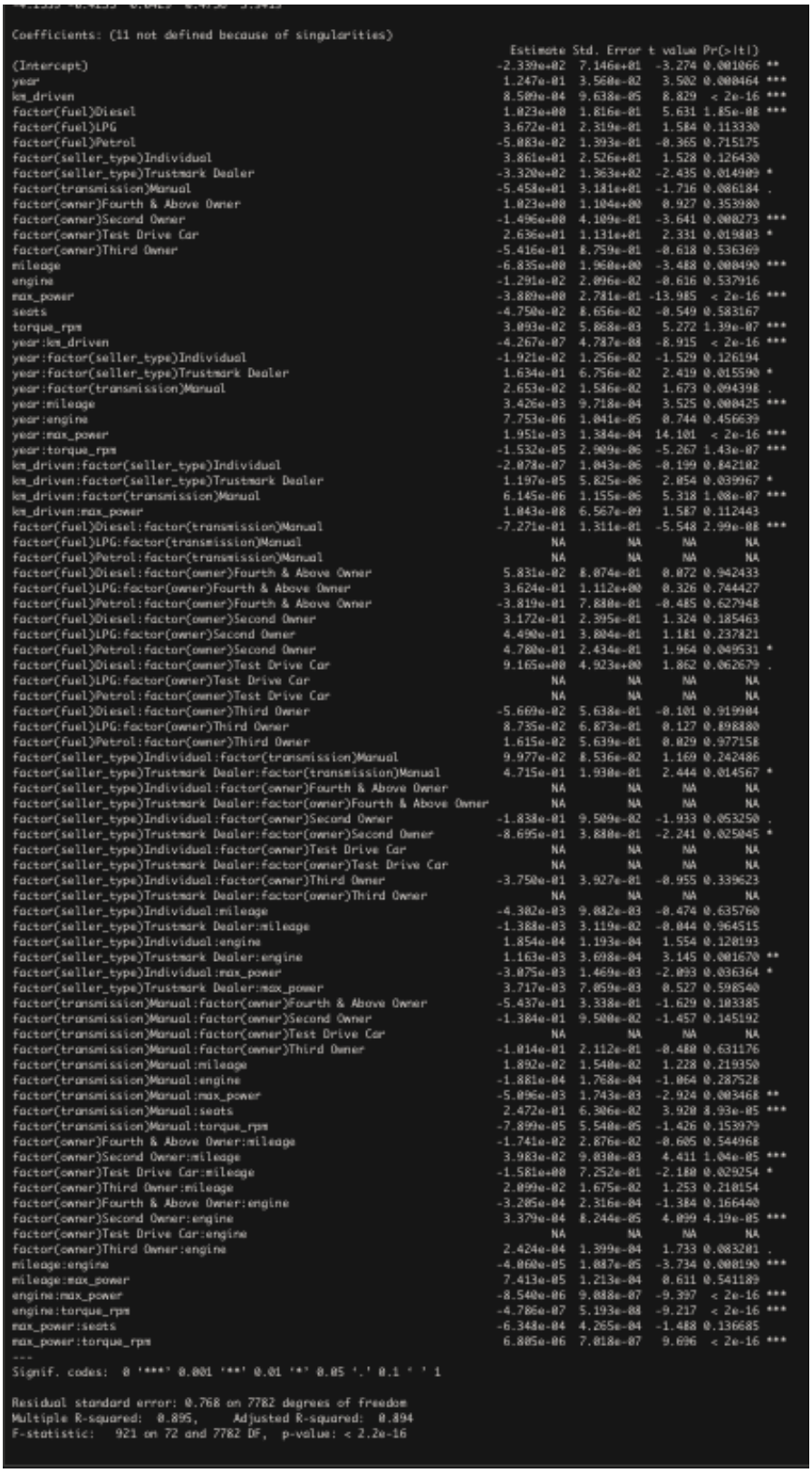
Residual standard error: 237800 on 7781 degrees of freedom
Multiple R-squared: 0.9095, Adjusted R-squared: 0.9086
F-statistic: 1071 on 73 and 7781 DF, p-value: < 2.2e-16

Figure 10: Higher order model with max_power^2

```
Coefficients: (11 not defined because of singularities)
                                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                                             -4.600e+07  2.249e+07  -2.045 0.040874 *
year                                                     2.139e+04  1.121e+04   1.908 0.056431 .
km_driven                                                2.071e+02  2.984e+01   6.938 4.28e-12 ***
factor(fuel)Diesel                                       4.532e+05  5.626e+04   8.056 9.08e-16 ***
factor(fuel)LPG                                         -3.960e+04  7.180e+04  -0.551 0.581324
factor(fuel)Petrol                                      -7.608e+04  4.318e+04  -1.762 0.078117 .
factor(seller_type)Individual                            5.354e+07  7.836e+06   6.832 9.02e-12 ***
factor(seller_type)Trustmark Dealer                      4.069e+07  4.221e+07   0.964 0.335092
factor(transmission)Manual                               8.809e+07  9.946e+06   8.856  < 2e-16 ***
factor(owner)Fourth & Above Owner                        6.730e+05  3.417e+05   1.970 0.048925 *
factor(owner)Second Owner                                6.075e+04  1.272e+05   0.477 0.633043
factor(owner)Test Drive Car                              3.838e+07  3.502e+06  10.961  < 2e-16 ***
factor(owner)Third Owner                                 6.452e+05  2.713e+05   2.378 0.017418 *
mileage                                                 -3.301e+06  6.086e+05  -5.423 6.03e-08 ***
engine                                                   1.835e+04  6.489e+03   2.828 0.004693 **
max_power                                               -2.656e+06  8.787e+04 -30.230  < 2e-16 ***
I(max_power^2)                                           9.908e-01  9.958e+00   0.099 0.920746
I(max_power^3)                                          -3.116e-02  1.866e-02  -1.669 0.095084 .
seats                                                    1.371e+05  2.842e+04   4.823 1.44e-06 ***
torque_rpm                                               2.094e+04  1.817e+03  11.527  < 2e-16 ***
year:km_driven                                          -1.050e-01  1.482e-02  -7.086 1.51e-12 ***
year:factor(seller_type)Individual                      -2.634e+04  3.897e+03  -6.760 1.48e-11 ***
year:factor(seller_type)Trustmark Dealer                -2.065e+04  2.091e+04  -0.987 0.323435
year:factor(transmission)Manual                         -4.280e+04  4.962e+03  -8.626  < 2e-16 ***
year:mileage                                             1.654e+03  3.018e+02   5.480 4.39e-08 ***
year:engine                                             -9.314e+00  3.224e+00  -2.889 0.003876 **
year:max_power                                           1.330e+03  4.390e+01  30.307  < 2e-16 ***
year:torque_rpm                                         -1.027e+01  9.006e-01 -11.399  < 2e-16 ***
km_driven:factor(seller_type)Individual                  1.680e+00  3.236e-01   5.193 2.12e-07 ***
km_driven:factor(seller_type)Trustmark Dealer            3.809e+00  1.803e+00   2.112 0.034686 *
km_driven:factor(transmission)Manual                     3.267e+00  3.580e-01   9.126  < 2e-16 ***
km_driven:max_power                                     -9.552e-03  2.058e-03  -4.641 3.53e-06 ***
factor(fuel)Diesel:factor(transmission)Manual           -4.350e+05  4.064e+04 -10.704  < 2e-16 ***
factor(fuel)LPG:factor(transmission)Manual                      NA         NA      NA       NA
factor(fuel)Petrol:factor(transmission)Manual                   NA         NA      NA       NA
factor(fuel)Diesel:factor(owner)Fourth & Above Owner     1.576e+05  2.500e+05   0.631 0.528342
factor(fuel)LPG:factor(owner)Fourth & Above Owner        9.369e+03  3.442e+05   0.027 0.978282
factor(fuel)Petrol:factor(owner)Fourth & Above Owner     2.516e+04  2.440e+05   0.103 0.917857
factor(fuel)Diesel:factor(owner)Second Owner             3.656e+04  7.416e+04   0.493 0.622002
factor(fuel)LPG:factor(owner)Second Owner                -4.448e+04  1.178e+05  -0.378 0.705603
factor(fuel)Petrol:factor(owner)Second Owner             3.856e+04  7.533e+04   0.512 0.608766
factor(fuel)Diesel:factor(owner)Test Drive Car           1.369e+07  1.524e+06   8.982  < 2e-16 ***
factor(fuel)LPG:factor(owner)Test Drive Car                     NA         NA      NA       NA
factor(fuel)Petrol:factor(owner)Test Drive Car                  NA         NA      NA       NA
factor(fuel)Diesel:factor(owner)Third Owner              1.492e+04  1.745e+05   0.085 0.931900
factor(fuel)LPG:factor(owner)Third Owner                 -1.089e+05  2.128e+05  -0.512 0.608979
factor(fuel)Petrol:factor(owner)Third Owner              1.111e+04  1.746e+05   0.064 0.949240
factor(seller_type)Individual:factor(transmission)Manual 2.312e+04  2.647e+04   0.873 0.382468
factor(seller_type)Trustmark Dealer:factor(transmission)Manual 2.392e+05  5.974e+04   4.004 6.29e-05 ***
factor(seller_type)Individual:factor(owner)Fourth & Above Owner       NA         NA      NA       NA
factor(seller_type)Trustmark Dealer:factor(owner)Fourth & Above Owner NA         NA      NA       NA
factor(seller_type)Individual:factor(owner)Second Owner -5.072e+04  2.945e+04  -1.723 0.085017 .
factor(seller_type)Trustmark Dealer:factor(owner)Second Owner -4.939e+05  1.202e+05  -4.111 3.98e-05 ***
factor(seller_type)Individual:factor(owner)Test Drive Car       NA         NA      NA       NA
factor(seller_type)Trustmark Dealer:factor(owner)Test Drive Car NA         NA      NA       NA
factor(seller_type)Individual:factor(owner)Third Owner  -2.261e+05  1.216e+05  -1.860 0.062929 .
factor(seller_type)Trustmark Dealer:factor(owner)Third Owner    NA         NA      NA       NA
factor(seller_type)Individual:mileage                   -1.104e+04  2.883e+03  -3.828 0.000130 ***
factor(seller_type)Trustmark Dealer:mileage              9.041e+03  9.676e+03   0.934 0.350119
factor(seller_type)Individual:engine                     1.643e+02  3.707e+01   4.432 9.48e-06 ***
factor(seller_type)Trustmark Dealer:engine               5.311e+02  1.145e+02   4.638 3.58e-06 ***
factor(seller_type)Individual:max_power                 -6.455e+03  4.705e+02 -13.721  < 2e-16 ***
factor(seller_type)Trustmark Dealer:max_power           -3.102e+03  2.185e+03  -1.419 0.155845
factor(transmission)Manual:factor(owner)Fourth & Above Owner -3.394e+05  1.034e+05  -3.284 0.001029 **
factor(transmission)Manual:factor(owner)Second Owner    -8.130e+04  2.945e+04  -2.761 0.005781 **
factor(transmission)Manual:factor(owner)Test Drive Car          NA         NA      NA       NA
factor(transmission)Manual:factor(owner)Third Owner     -3.397e+05  6.538e+04  -5.196 2.08e-07 ***
factor(transmission)Manual:mileage                      -1.794e+04  4.794e+03  -3.743 0.000183 ***
factor(transmission)Manual:engine                        6.181e+02  6.431e+01   9.611  < 2e-16 ***
factor(transmission)Manual:max_power                    -1.607e+04  7.291e+02 -22.047  < 2e-16 ***
factor(transmission)Manual:seats                        -5.337e+04  2.028e+04  -2.631 0.008519 **
factor(transmission)Manual:torque_rpm                   -1.686e+02  1.716e+01  -9.823  < 2e-16 ***
factor(owner)Fourth & Above Owner:mileage               -1.493e+04  8.902e+03  -1.678 0.093458 .
factor(owner)Second Owner:mileage                        1.406e+03  2.796e+03   0.503 0.615165
factor(owner)Test Drive Car:mileage                     -2.345e+06  2.245e+05 -10.443  < 2e-16 ***
factor(owner)Third Owner:mileage                        -3.489e+03  5.192e+03  -0.672 0.501672
factor(owner)Fourth & Above Owner:engine                -1.443e+02  7.173e+01  -2.012 0.044241 *
factor(owner)Second Owner:engine                        -6.806e+00  2.554e+01  -0.266 0.789885
factor(owner)Test Drive Car:engine                              NA         NA      NA       NA
factor(owner)Third Owner:engine                         -4.080e+01  4.335e+01  -0.941 0.346604
mileage:engine                                          -1.640e+01  3.611e+00  -4.542 5.65e-06 ***
mileage:max_power                                        2.669e+02  4.598e+01   5.804 6.74e-09 ***
engine:max_power                                         2.827e+00  4.504e-01   6.276 3.66e-10 ***
engine:torque_rpm                                       -1.330e-01  1.614e-02  -8.243  < 2e-16 ***
max_power:seats                                         -8.654e+02  1.414e+02  -6.122 9.70e-10 ***
max_power:torque_rpm                                     1.017e+00  2.173e-01   4.683 2.88e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 237700 on 7780 degrees of freedom
Multiple R-squared:  0.9095,    Adjusted R-squared:  0.9086
F-statistic:  1056 on 74 and 7780 DF,  p-value: < 2.2e-16
```

Figure 11: Higher order model with max_power[3]



```
Coefficients: (11 not defined because of singularities)
                                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                                             -2.339e+02  7.146e+01  -3.274 0.001066 **
year                                                     1.247e-01  3.560e-02   3.502 0.000464 ***
km_driven                                                8.509e-04  9.638e-05   8.829  < 2e-16 ***
factor(fuel)Diesel                                       1.023e+00  1.816e-01   5.631 1.85e-08 ***
factor(fuel)LPG                                          3.672e-01  2.319e-01   1.584 0.113330
factor(fuel)Petrol                                      -5.083e-02  1.393e-01  -0.365 0.715175
factor(seller_type)Individual                            3.861e+01  2.526e+01   1.528 0.126430
factor(seller_type)Trustmark Dealer                     -3.320e+02  1.363e+02  -2.435 0.014909 *
factor(transmission)Manual                              -5.458e+01  3.181e+01  -1.716 0.086184 .
factor(owner)Fourth & Above Owner                        1.023e+00  1.104e+00   0.927 0.353980
factor(owner)Second Owner                               -1.496e+00  4.109e-01  -3.641 0.000273 ***
factor(owner)Test Drive Car                              2.636e+01  1.131e+01   2.331 0.019803 *
factor(owner)Third Owner                                -5.416e-01  8.759e-01  -0.618 0.536369
mileage                                                 -6.835e+00  1.960e+00  -3.488 0.000490 ***
engine                                                  -1.291e-02  2.096e-02  -0.616 0.537916
max_power                                               -3.889e+00  2.781e-01 -13.985  < 2e-16 ***
seats                                                   -4.750e-02  8.656e-02  -0.549 0.583167
torque_rpm                                               3.093e-02  5.868e-03   5.272 1.39e-07 ***
year:km_driven                                          -4.267e-07  4.787e-08  -8.915  < 2e-16 ***
year:factor(seller_type)Individual                      -1.921e-02  1.256e-02  -1.529 0.126194
year:factor(seller_type)Trustmark Dealer                 1.634e-01  6.756e-02   2.419 0.015590 *
year:factor(transmission)Manual                          2.653e-02  1.586e-02   1.673 0.094398 .
year:mileage                                             3.426e-03  9.718e-04   3.525 0.000425 ***
year:engine                                              7.753e-06  1.041e-05   0.744 0.456639
year:max_power                                           1.951e-03  1.384e-04  14.101  < 2e-16 ***
year:torque_rpm                                         -1.532e-05  2.909e-06  -5.267 1.43e-07 ***
km_driven:factor(seller_type)Individual                 -2.078e-07  1.043e-06  -0.199 0.842102
km_driven:factor(seller_type)Trustmark Dealer            1.197e-05  5.825e-06   2.054 0.039967 *
km_driven:factor(transmission)Manual                     6.145e-06  1.155e-06   5.318 1.08e-07 ***
km_driven:max_power                                      1.043e-08  6.567e-09   1.587 0.112443
factor(fuel)Diesel:factor(transmission)Manual           -7.271e-01  1.311e-01  -5.548 2.99e-08 ***
factor(fuel)LPG:factor(transmission)Manual                      NA         NA      NA       NA
factor(fuel)Petrol:factor(transmission)Manual                   NA         NA      NA       NA
factor(fuel)Diesel:factor(owner)Fourth & Above Owner     5.831e-02  8.074e-01   0.072 0.942433
factor(fuel)LPG:factor(owner)Fourth & Above Owner        3.624e-01  1.112e+00   0.326 0.744427
factor(fuel)Petrol:factor(owner)Fourth & Above Owner     -3.819e-01 7.880e-01  -0.485 0.627948
factor(fuel)Diesel:factor(owner)Second Owner             3.172e-01  2.395e-01   1.324 0.185463
factor(fuel)LPG:factor(owner)Second Owner                4.490e-01  3.804e-01   1.181 0.237821
factor(fuel)Petrol:factor(owner)Second Owner             4.780e-01  2.434e-01   1.964 0.049531 *
factor(fuel)Diesel:factor(owner)Test Drive Car           9.165e+00  4.923e+00   1.862 0.062679 .
factor(fuel)LPG:factor(owner)Test Drive Car                     NA         NA      NA       NA
factor(fuel)Petrol:factor(owner)Test Drive Car                  NA         NA      NA       NA
factor(fuel)Diesel:factor(owner)Third Owner             -5.669e-02  5.638e-01  -0.101 0.919304
factor(fuel)LPG:factor(owner)Third Owner                 8.735e-02  6.873e-01   0.127 0.898880
factor(fuel)Petrol:factor(owner)Third Owner              1.615e-02  5.639e-01   0.029 0.977158
factor(seller_type)Individual:factor(transmission)Manual 9.977e-02  8.536e-02   1.169 0.242486
factor(seller_type)Trustmark Dealer:factor(transmission)Manual 4.715e-01 1.930e-01 2.444 0.014567 *
factor(seller_type)Individual:factor(owner)Fourth & Above Owner         NA         NA      NA       NA
factor(seller_type)Trustmark Dealer:factor(owner)Fourth & Above Owner   NA         NA      NA       NA
factor(seller_type)Individual:factor(owner)Second Owner -1.838e-01 9.509e-02 -1.933 0.053250 .
factor(seller_type)Trustmark Dealer:factor(owner)Second Owner -8.695e-01 3.880e-01 -2.241 0.025045 *
factor(seller_type)Individual:factor(owner)Test Drive Car       NA         NA      NA       NA
factor(seller_type)Trustmark Dealer:factor(owner)Test Drive Car NA         NA      NA       NA
factor(seller_type)Individual:factor(owner)Third Owner  -3.750e-01  3.927e-01  -0.955 0.339623
factor(seller_type)Trustmark Dealer:factor(owner)Third Owner    NA         NA      NA       NA
factor(seller_type)Individual:mileage                   -4.382e-03  9.082e-03  -0.474 0.635760
factor(seller_type)Trustmark Dealer:mileage             -1.388e-03  3.119e-02  -0.044 0.964515
factor(seller_type)Individual:engine                     1.854e-04  1.193e-04   1.554 0.120193
factor(seller_type)Trustmark Dealer:engine               1.163e-03  3.698e-04   3.145 0.001670 **
factor(seller_type)Individual:max_power                 -3.075e-03  1.469e-03  -2.093 0.036364 *
factor(seller_type)Trustmark Dealer:max_power            3.717e-03  7.059e-03   0.527 0.598540
factor(transmission)Manual:factor(owner)Fourth & Above Owner -5.437e-01 3.338e-01 -1.629 0.103385
factor(transmission)Manual:factor(owner)Second Owner    -1.384e-01  9.500e-02  -1.457 0.145192
factor(transmission)Manual:factor(owner)Test Drive Car          NA         NA      NA       NA
factor(transmission)Manual:factor(owner)Third Owner     -1.014e-01  2.112e-01  -0.480 0.631176
factor(transmission)Manual:mileage                       1.892e-02  1.540e-02   1.228 0.219350
factor(transmission)Manual:engine                       -1.881e-04  1.768e-04  -1.064 0.287528
factor(transmission)Manual:max_power                    -5.096e-03  1.743e-03  -2.924 0.003468 **
factor(transmission)Manual:seats                         2.472e-01  6.306e-02   3.920 8.93e-05 ***
factor(transmission)Manual:torque_rpm                   -7.899e-05  5.540e-05  -1.426 0.153979
factor(owner)Fourth & Above Owner:mileage               -1.741e-02  2.876e-02  -0.605 0.544968
factor(owner)Second Owner:mileage                        3.983e-02  9.030e-03   4.411 1.04e-05 ***
factor(owner)Test Drive Car:mileage                     -1.581e+00  7.252e-01  -2.180 0.029254 *
factor(owner)Third Owner:mileage                         2.099e-02  1.675e-02   1.253 0.210154
factor(owner)Fourth & Above Owner:engine                -3.205e-04  2.316e-04  -1.384 0.166440
factor(owner)Second Owner:engine                         3.379e-04  8.244e-05   4.099 4.19e-05 ***
factor(owner)Test Drive Car:engine                              NA         NA      NA       NA
factor(owner)Third Owner:engine                          2.424e-04  1.399e-04   1.733 0.083201 .
mileage:engine                                          -4.060e-05  1.087e-05  -3.734 0.000190 ***
mileage:max_power                                        7.413e-05  1.213e-04   0.611 0.541189
engine:max_power                                        -8.540e-06  9.888e-07  -9.397  < 2e-16 ***
engine:torque_rpm                                       -4.786e-07  5.193e-08  -9.217  < 2e-16 ***
max_power:seats                                         -6.348e-04  4.265e-04  -1.488 0.136685
max_power:torque_rpm                                     6.805e-06  7.018e-07   9.696  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.768 on 7782 degrees of freedom
Multiple R-squared:  0.895,     Adjusted R-squared:  0.894
F-statistic:   921 on 72 and 7782 DF,  p-value: < 2.2e-16
```

Figure 12: Model post Box-Cox transformation

```
        studentized Breusch-Pagan test

data:  bcmodel
BP = 1014.3, df = 72, p-value < 2.2e-16
```

Figure 13: Results from Breusch-Pagan test on model

```
        Shapiro-Wilk normality test

data:  residuals(bcmodel)[0:5000]
W = 0.98689, p-value < 2.2e-16
```

Figure 14: Results from Shapiro-Wilk normality test on model

```
Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                      2.218e+05  3.133e+05   0.708 0.479065
factor(year)1995                 2.920e+04  5.183e+05   0.056 0.955079
factor(year)1996                 2.360e+04  4.233e+05   0.056 0.955540
factor(year)1997                -1.590e+04  3.310e+05  -0.048 0.961693
factor(year)1998                -5.782e+03  3.310e+05  -0.017 0.986061
factor(year)1999                -3.741e+04  3.201e+05  -0.117 0.906961
factor(year)2000                 5.127e+04  3.190e+05   0.161 0.872312
factor(year)2001                -8.808e+03  3.456e+05  -0.025 0.979670
factor(year)2002                -7.507e+04  3.148e+05  -0.238 0.811535
factor(year)2003                -2.181e+05  3.075e+05  -0.709 0.478214
factor(year)2004                -2.540e+05  3.053e+05  -0.832 0.405403
factor(year)2005                -2.588e+05  3.035e+05  -0.853 0.393864
factor(year)2006                -3.120e+05  3.025e+05  -1.031 0.302482
factor(year)2007                -2.956e+05  3.013e+05  -0.981 0.326537
factor(year)2008                -3.050e+05  3.011e+05  -1.013 0.311044
factor(year)2009                -2.979e+05  3.009e+05  -0.990 0.322219
factor(year)2010                -2.934e+05  3.005e+05  -0.976 0.328909
factor(year)2011                -3.135e+05  3.002e+05  -1.044 0.296342
factor(year)2012                -2.806e+05  3.002e+05  -0.935 0.349947
factor(year)2013                -2.268e+05  3.003e+05  -0.755 0.450150
factor(year)2014                -1.928e+05  3.003e+05  -0.642 0.520997
factor(year)2015                -1.670e+05  3.003e+05  -0.556 0.578180
factor(year)2016                -9.375e+04  3.004e+05  -0.312 0.754972
factor(year)2017                -2.123e+04  3.004e+05  -0.071 0.943652
factor(year)2018                 2.516e+04  3.005e+05   0.084 0.933268
factor(year)2019                 4.883e+05  3.008e+05   1.623 0.104579
factor(year)2020                 3.084e+04  3.042e+05   0.101 0.919250
km_driven                       -5.343e-01  1.023e-01  -5.224 1.80e-07 ***
factor(fuel)Diesel              -1.564e+05  6.094e+04  -2.566 0.010318 *
factor(fuel)LPG                  3.110e+05  9.353e+04   3.325 0.000887 ***
factor(fuel)Petrol               1.389e+04  6.020e+04   0.231 0.817495
factor(seller_type)Individual   -2.663e+05  1.544e+04 -17.252  < 2e-16 ***
factor(seller_type)Trustmark Dealer -3.709e+05  3.182e+04 -11.659  < 2e-16 ***
factor(transmission)Manual      -3.472e+05  1.858e+04 -18.682  < 2e-16 ***
factor(owner)Fourth & Above Owner -6.110e+04  3.590e+04  -1.702 0.088742 .
factor(owner)Second Owner       -2.003e+04  1.249e+04  -1.604 0.108806
factor(owner)Test Drive Car      1.648e+06  1.909e+05   8.633  < 2e-16 ***
factor(owner)Third Owner        -1.709e+04  2.143e+04  -0.798 0.425121
mileage                          1.839e+04  2.130e+03   8.633  < 2e-16 ***
engine                           1.438e+02  2.242e+01   6.416 1.48e-10 ***
max_power                        1.222e+04  2.457e+02  49.745  < 2e-16 ***
seats                           -3.539e+04  7.631e+03  -4.638 3.58e-06 ***
torque_rpm                      -1.121e+02  8.677e+00 -12.919  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 15: Individual t-tests on model with year as a qualitative variable

```
Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                      1.155e+06  1.684e+05   6.857 7.61e-12 ***
factor(brand)Ashok              -2.358e+05  3.310e+05  -0.712  0.47618
factor(brand)Audi                7.958e+05  1.579e+05   5.039 4.78e-07 ***
factor(brand)BMW                 2.418e+06  1.530e+05  15.804  < 2e-16 ***
factor(brand)Chevrolet          -3.545e+05  1.499e+05  -2.365  0.01805 *
factor(brand)Daewoo              2.357e+05  2.266e+05   1.040  0.29828
factor(brand)Datsun             -3.879e+05  1.538e+05  -2.521  0.01172 *
factor(brand)Fiat               -3.644e+05  1.564e+05  -2.329  0.01987 *
factor(brand)Force              -3.257e+05  1.913e+05  -1.702  0.08872 .
factor(brand)Ford               -2.978e+05  1.493e+05  -1.995  0.04612 *
factor(brand)Honda              -2.649e+05  1.496e+05  -1.771  0.07654 .
factor(brand)Hyundai            -2.759e+05  1.490e+05  -1.852  0.06405 .
factor(brand)Isuzu               2.825e+05  1.991e+05   1.419  0.15591
factor(brand)Jaguar              1.176e+06  1.542e+05   7.624 2.76e-14 ***
factor(brand)Jeep                5.257e+05  1.598e+05   3.289  0.00101 **
factor(brand)Kia                 8.852e+04  2.098e+05   0.422  0.67307
factor(brand)Land                2.233e+06  2.102e+05  10.622  < 2e-16 ***
factor(brand)Lexus               3.286e+06  1.610e+05  20.416  < 2e-16 ***
factor(brand)Mahindra           -3.094e+05  1.490e+05  -2.077  0.03787 *
factor(brand)Maruti             -1.919e+05  1.490e+05  -1.288  0.19764
factor(brand)Mercedes-Benz       8.432e+05  1.558e+05   5.410 6.49e-08 ***
factor(brand)MG                  2.041e+05  2.576e+05   0.792  0.42833
factor(brand)Mitsubishi          1.554e+04  1.693e+05   0.092  0.92688
factor(brand)Nissan             -3.203e+05  1.522e+05  -2.104  0.03539 *
factor(brand)Opel                1.257e+05  3.304e+05   0.380  0.70369
factor(brand)Renault            -2.941e+05  1.502e+05  -1.958  0.05032 .
factor(brand)Skoda              -3.455e+05  1.515e+05  -2.281  0.02260 *
factor(brand)Tata               -4.018e+05  1.489e+05  -2.699  0.00698 **
factor(brand)Toyota             -6.336e+03  1.494e+05  -0.042  0.96616
factor(brand)Volkswagen         -3.924e+05  1.504e+05  -2.610  0.00908 **
factor(brand)Volvo               7.713e+05  1.589e+05   4.853 1.24e-06 ***
age                             -4.809e+04  1.343e+03 -35.813  < 2e-16 ***
km_driven                       -6.118e-01  7.168e-02  -8.535  < 2e-16 ***
factor(fuel)Diesel               1.079e+05  4.373e+04   2.467  0.01365 *
factor(fuel)LPG                  1.282e+05  6.685e+04   1.918  0.05516 .
factor(fuel)Petrol               3.478e+04  4.293e+04   0.810  0.41793
factor(seller_type)Individual   -8.368e+04  1.138e+04  -7.351 2.18e-13 ***
factor(seller_type)Trustmark Dealer -8.756e+04  2.369e+04  -3.696  0.00022 ***
factor(transmission)Manual      -1.097e+05  1.413e+04  -7.762 9.48e-15 ***
factor(owner)Fourth & Above Owner -2.779e+04  2.518e+04  -1.103  0.26995
factor(owner)Second Owner       -5.871e+04  8.844e+03  -6.639 3.39e-11 ***
factor(owner)Test Drive Car      2.412e+06  1.358e+05  17.762  < 2e-16 ***
factor(owner)Third Owner        -3.344e+04  1.510e+04  -2.214  0.02685 *
mileage                         -1.129e+04  1.711e+03  -6.597 4.48e-11 ***
engine                           7.987e+01  1.799e+01   4.440 9.14e-06 ***
max_power                        4.983e+03  2.052e+02  24.284  < 2e-16 ***
seats                           -1.680e+04  5.989e+03  -2.806  0.00503 **
torque_rpm                      -2.797e+01  6.595e+00  -4.241 2.25e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 294900 on 7467 degrees of freedom
Multiple R-squared:  0.8577,    Adjusted R-squared:  0.8568
F-statistic: 957.6 on 47 and 7467 DF,  p-value: < 2.2e-16
```

Figure 16: Individual t-tests on model with brand as a qualitative variable

B: Relevant Equations

$$VIF(\hat{\beta}_j) = \frac{1}{1 - R^2_{X_j|X_j}}$$

Equation 1: Equation used in calculating Variance Inflation Factors

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \left(y_i - \left(\widehat{\beta_0} + \widehat{\beta_1} X_{1i} + \widehat{\beta_2} X_{2i} + \cdots + \widehat{\beta_p} X_{pi} \right) \right)^2$$

Equation 2: Equation used in calculating the sum of squares for error (SSE)

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

Equation 3: Equation used in calculating the sum of squares for regression (SSR)

$$F_{cal} = \frac{MSR}{MSE} = \frac{\frac{SSR}{p}}{\frac{SSE}{(n-p-1)}}$$

Equation 4: Equation used in calculating the F statistic

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

Equation 5: Equation used in calculating R^2

$$R^2_{adj} = 1 - \frac{\frac{SSE}{n-p-1}}{\frac{SST}{n-1}}$$

$$R^2_{adj} = 1 - (n-1)\frac{MSE}{SST}$$

Equation 6: Equation used in calculating adjusted R^2

$$s = RMSE = \sqrt{\frac{1}{n-p-1} SSE}$$

Equation 7: Equation used in calculating adjusted root mean square error

$$AIC = n \ln\left(\frac{SSE}{n}\right) + 2p$$

Equation 8: Equation used in calculating Akaike's information criterion (AIC)

$$C_p = \frac{SS(Res)_p}{s^2} + 2p' - n$$

Equation 9: Equation used in calculating Mallows' Cp Criterion

$$BIC = n \ln\left(\frac{SSE}{n}\right) + (p) \ln(n)$$

Equation 10: Equation used in calculating Bayesian information criteria (BIC)

$$Y_i^{(\lambda)} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \epsilon_i$$

$$where$$

$$Y_i^{(\lambda)} = \begin{cases} \frac{Y^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log_e Y, & \lambda = 0 \end{cases}$$

Equation 11: Transformation applied to regression equation for Box-Cox method. Note: Lambda is an estimated value

$$\chi^2 = nR^2 \sim \chi^2_{p-1}$$

Equation 12: Equation used in calculating chi values for the Breusch-Pagan test