Regression Analysis on Car Prices in Serbia 2024

STAC67 - Group 25

Armaan Rehman Shah[1] - 1009641309,
Darren Guerina[2] - 1005511741,
Rachel Takacs[3] - 1001406294

April 4, 2025

[1] Quantitative Variable Filtering, Introduction (Background and significance), Model (Final model breakdown, Adding interaction terms on variables, Transformation of model variables for improvement).
[2] Discussion/Conclusion, Model (Model Diagnostics, LINE Assumptions, Model Validation).
[3] Qualitative Variable Filtering, Exploratory Data Analysis (Analysis of overall data pre-filtering, Analysis of overall data post-filtering, Correlation of each variable with each other).

Background and Significance

This study aims to predict used car prices in Serbia (2024) by analyzing the relationship between pricing and key vehicle features, including quantitative factors (mileage, horsepower, listing duration) and qualitative attributes (car type, fuel, A/C). The research question investigates how these variables collectively influence pricing, with the hypothesis that nonlinear relationships and interactions (e.g., mileage-age, engine size-fuel type) significantly improve predictive accuracy. To prepare the data, outliers were removed (e.g., cars priced above €20,000, extreme mileage values), and variables were transformed (log/square root) and centered to meet regression assumptions. The analysis employs a three-stage modeling approach: (1) a Main Effects Model to establish baselines, (2) an Interaction Model to test combined variable effects, and (3) a Power Model with polynomial terms and interactions for optimal fit. Variables like emission class and color were excluded due to redundancy or excessive categories, while car type and A/C were retained for their interpretable impact. By validating models through train-test splits and diagnostic checks, this study provides actionable insights for sellers pricing vehicles and identifies critical valuation drivers in Serbia's used car market.

How much should we price a car, based on a number of defining features? If we go too low, we get high demand and therefore a limited supply. If we go too high, we get limited demand and oversupply.

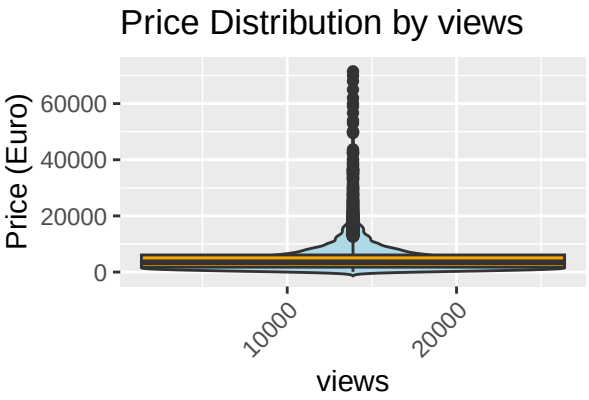
Exploratory Data Analysis

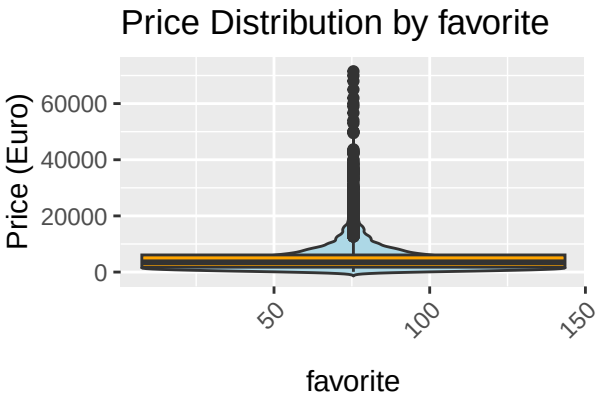
The Data

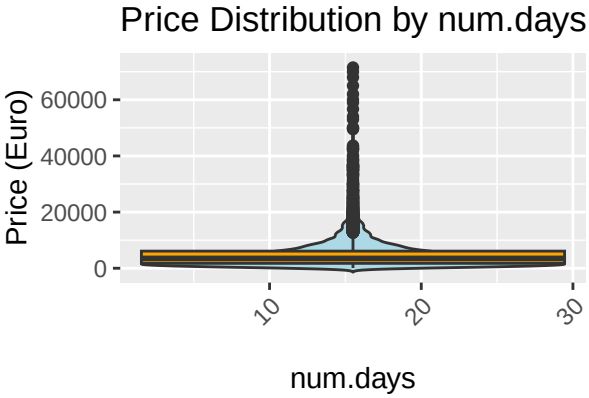
Y: price - The price of the car in Euro. X_1: views = The total number of views the car listing has received X_2: favorite = The number of users who have added the car to their favorite list X_3: num.days = The number of days since the post was made X_4: `car_mileage, km` = The car's mileage in kilometers X_5: `engine_capacity, cc` = The engine capacity in cubic centimeters (cc) X_6: year_to_date = The number of years since the car was manufactured X_7: hp = The engine power measured in horsepower. X_8: car_name = Manufacturer and model. X_9: seats_amount = The number of seats in the car. X_{10}: `A/C` = Indicates whether the car is equipped with air conditioning. X_{11}: emission_class = The car's emission standard classification. X_{12}: color = The exterior color of the car. X_{13}: type_of_drive = Indicates the type of drive, such as front-wheel drive or all-wheel drive. X_{14}: doors = The number of doors on the car. X_{15}: fuel = "The type of fuel the car uses, such as gasoline or diesel. X_{16}: car_type = The category or body style of the car (e.g., sedan, SUV, hatchback). X_{17}: gearbox = The type of transmission, such as manual or automatic.

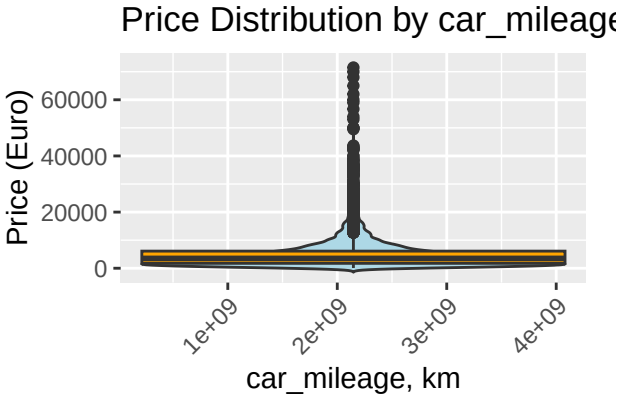
Analysis of Data Pre-filtering

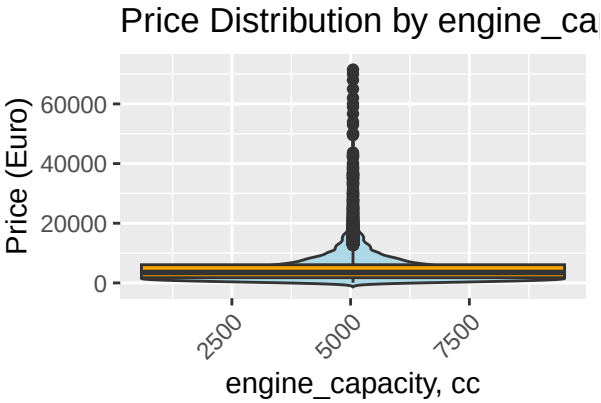
The data we receive includes seven quantitative variables and 10 qualitative variables, of which we transform two variables into quantitative. These were post info and horsepower, since these values contained numerical inputs. The relevance of the numerical values of post-info could relay information on recent popularity of a car whereas horsepower could relay information on the technical caliber of a car; higher quality cars have greater horsepower. Another variable transformed was the year the car was manufactured, transforming it to the number of years it has existed instead, for ease of interpretance.

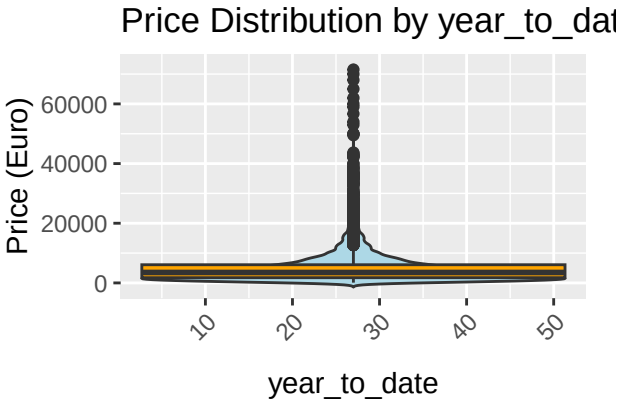

Price Distribution by views
Price (Euro)
60000
40000
20000
0
10000
20000
views

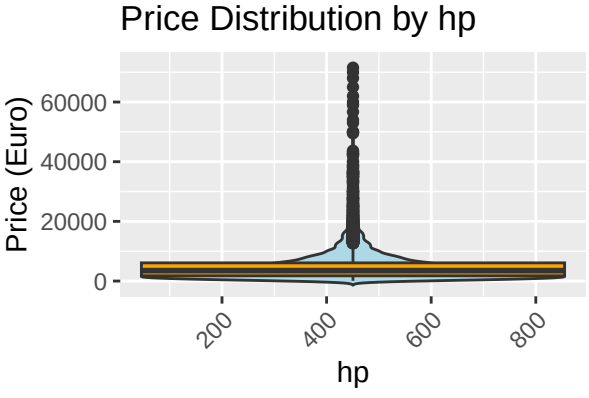

Price Distribution by favorite
Price (Euro)
60000
40000
20000
0
50
100
150
favorite

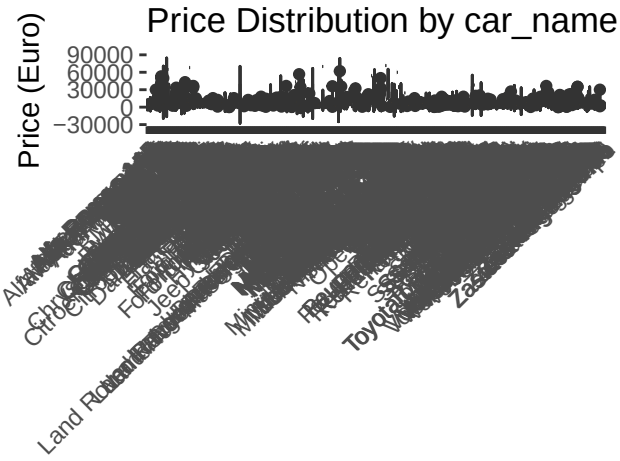

Price Distribution by num.days
Price (Euro)
60000
40000
20000
0
10
20
30
num.days

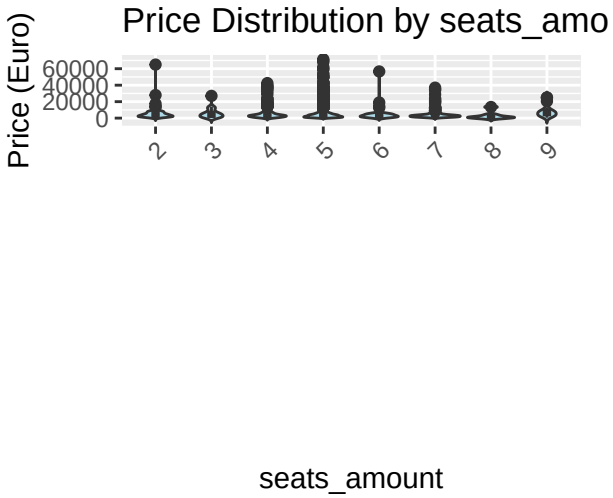

Price Distribution by car_mileage
Price (Euro)
60000
40000
20000
0
1e+09
2e+09
3e+09
4e+09
car_mileage, km

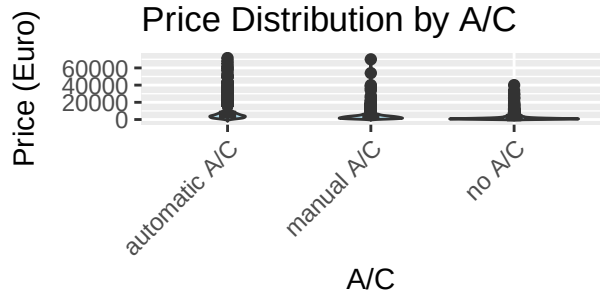

Price Distribution by engine_ca
Price (Euro)
60000
40000
20000
0
2500
5000
7500
engine_capacity, cc

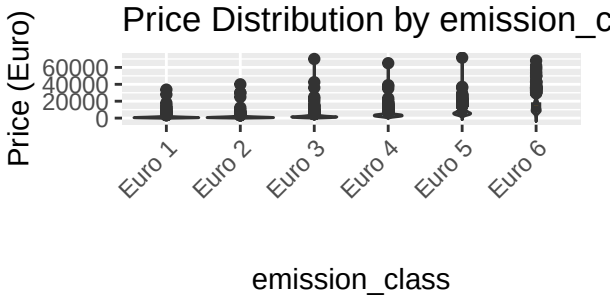

Price Distribution by year_to_dat
Price (Euro)
60000
40000
20000
0
10
20
30
40
50
year_to_date

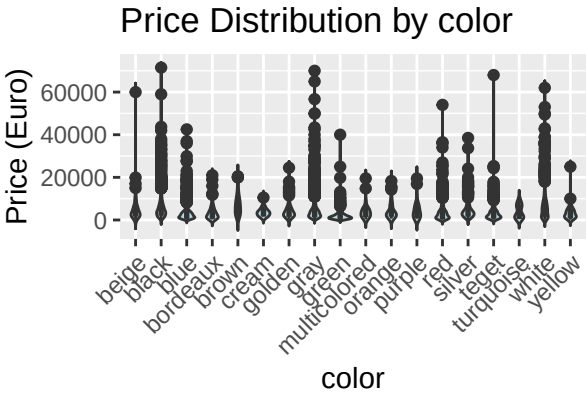

Price Distribution by hp
Price (Euro)
60000
40000
20000
0
200
400
600
800
hp

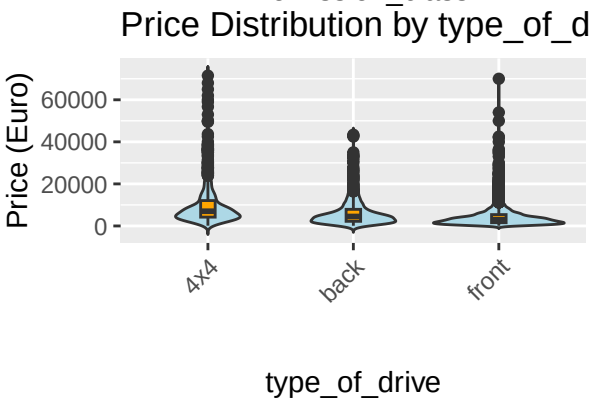

Price Distribution by car_name
Price (Euro)
90000
60000
30000
0
−30000
Toyota
car_name

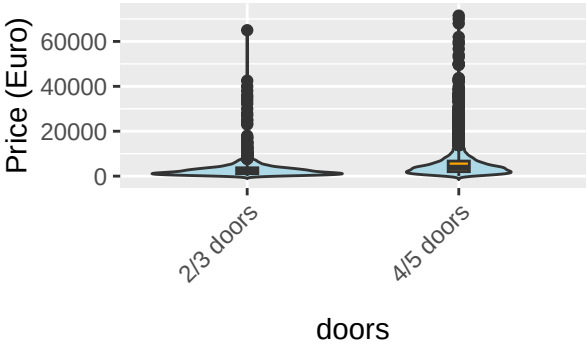

Price Distribution by seats_amo
Price (Euro)
60000
40000
20000
0
2
3
4
5
6
7
8
9
seats_amount

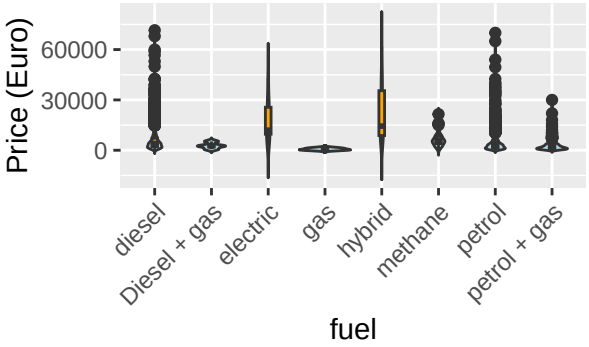

Price Distribution by A/C
Price (Euro)
60000
40000
20000
0
automatic A/C
manual A/C
no A/C
A/C

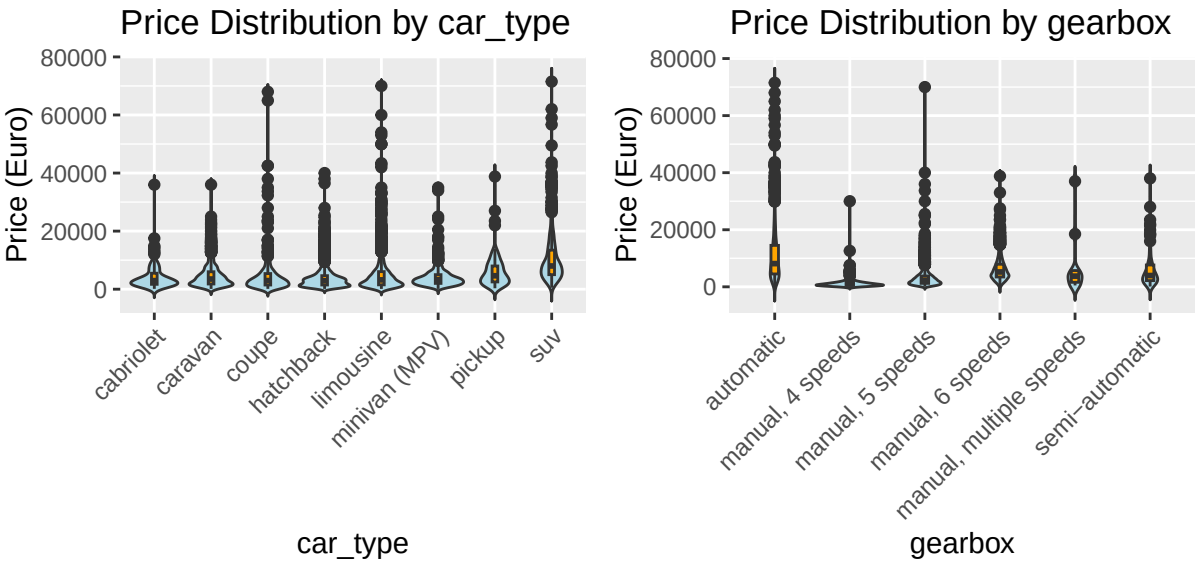

Price Distribution by emission_c
Price (Euro)
60000
40000
20000
0
Euro 1
Euro 2
Euro 3
Euro 4
Euro 5
Euro 6
emission_class


Price Distribution by color
Price (Euro)
60000
40000
20000
0
beige
black
blue
bordeaux
brown
cream
golden
gray
green
multicolored
orange
purple
red
silver
teget
turquoise
white
yellow
color


Price Distribution by type_of_dri
Price (Euro)
60000
40000
20000
0
4x4
back
front
type_of_drive


Price Distribution by doors
Price (Euro)
60000
40000
20000
0
2/3 doors
4/5 doors
doors


Price Distribution by fuel
Price (Euro)
60000
30000
0
diesel
Diesel + gas
electric
gas
hybrid
methane
petrol
petrol + gas
fuel



Quantitative Variables

From our initial results, we can see transformation of some variables and filtering will be necessary.

For views, in testing for homogeneity of variance using the Brown-Forsythe (BF) test, we get the p-value: 1.3865218 and thus violates equal variance. We have an $R^2 = 0.0003$ indicating a non-existent linear relationship. The residual plot indicates this violation of linearity. The concentration of the plot indicates we should transform the variable. After trial and error, $\sqrt{x}$ produces better variability.

For favorite, the BF test gives the p-value: 0.002715321 and thus violates equal variance. We have an $R^2 = -0.0764$ indicating a non-existent linear relationship. The residual plot indicates this violation of linearity. The concentration of the plot indicates we should transform the variable. After trial and error, $\sqrt{x}$ produces better variability.

For num.days, the BF test gives the p-value: 0.0006138374 and thus violates equal variance. We have an $R^2 = -0.0257$ indicating a non-existent linear relationship.The residual plot indicates this violation of linearity. The concentration of the plot indicates we should transform the variable. After trial and error, $\sqrt{x}$ produces better variability.

For car_mileage, km, the BF test gives the p-value: $2.171742e-33$ and thus violates equal variance. We have an $R^2 = -0.0159$ indicating a non-existent linear relationship. The residual plot indicates this violation of linearity. The concentration of the plot indicates we should transform the variable. After trial and error, $\log x$ produces better variability.

For engine_capacity, cc, the BF test gives the p-value: $2.332314e-24$ and thus violates equal variance. We have an $R^2 = 0.2453$ indicating a weak, positively correlated linear relationship. The residual plot indicates more random scatter, but is still quite condensed in relation to this weak linearity. The concentration of the plot indicates we should transform the variable. After trial and error, $\log x$ produces better variability.

For year_to_date, the BF test gives the p-value: $2.745272e-63$ and thus violates equal variance. We have an $R^2 = -0.5694$ indicating a strong, negatively correlated linear relationship. The residual plot has strong scatter, but not completely random looking. The concentration of the plot indicates we should transform the variable. After trial and error, $\log x$ produces better variability.

For hp, the BF test gives the p-value: $1.034599e-97$ and thus violates equal variance. We have an $R^2 = 0.5129$ indicating a strong, positively correlated linear relationship. The residual plot has weak randomization. Transformation does not improve the relationship, as it is clear that outliers skew the linear relationship.

For the stronger correlated relationships, the variables have histograms that closely resemble normal plots and Q-Q plots that lie closer to $Y = X$ line, but still stray, indicating normality does not hold well for even the strongly correlated variables.

What we take away from the violin plots overall as well as all other plots, is that there are a significant amount of outliers under each variable, so by filtering, we can likely increase the strength each variable plays in model prediction.

Qualitative Variables

The quantitative variable car_name has too many levels. While a car brand name could mean more to specific buyers attracted to status, using this variable in our model would hinder the stronger predictive outcomes.

The variable seats_amount is overpowered by the number cars that have 5 seats. It is in our best interest for filter for cars with 5 seats.

The variable A/C can tell us a lot about the quality of a car in terms of comfort. A/C is considered a luxury which could correlate with a higher price.

The variable emission_class is based on the year it existed for certification under a specific class. Therefore we choose to ignore this variable since we already have number of years to indicate this already.

The variable color contains a wide variety but a majority of cars belong to the following colors: black and gray. Thus we choose to filter for gray and black cars.

The variable type_of_drive has few levels but is heavily carried by cars under front-wheel drive. Therefore we choose to filter for this type.

The variable doors has few levels but is heavily carried by cars with 4/5 doors. Therefore we choose to filter for this type.

The variable fuel has few levels but majority falls under diesel and petrol. We choose to filter for this type.

The variable gearbox has few levels but majority falls under manual. We choose to filter for this type, including all variations of manual transmission. The variable car type has 8 levels, but we choose to leave this in the regression as it is a variable that will make for some more insightful analysis, as the distinction between car types is significant.

Lastly we perform Box-Cox tests. Our result on the full model as of this point without filtering but including only the necessary variables, is 0.2222222. It is at this point we also center the predictive and response variables.

We can see these results respect normality and BF tests indicate higher p-values for views, favorite and year_to_date. Overall the Q-Q plot and histograms of residuals indicate stronger normality and the residuals are more randomized for all variables, even though we still have outliers.

Variable	t-statistic	p-value
Views	11.14438	2.00000
Favorite	2.119924	1.965953
Num.Days	-2.31241969	0.02078316
Car Mileage	-9.515485e+00	2.426601e-21
Engine Capacity	-3.932050e+00	8.502948e-05
Year to Date	0.7570804	1.5509731
Horsepower	-8.460689e+00	3.198377e-17

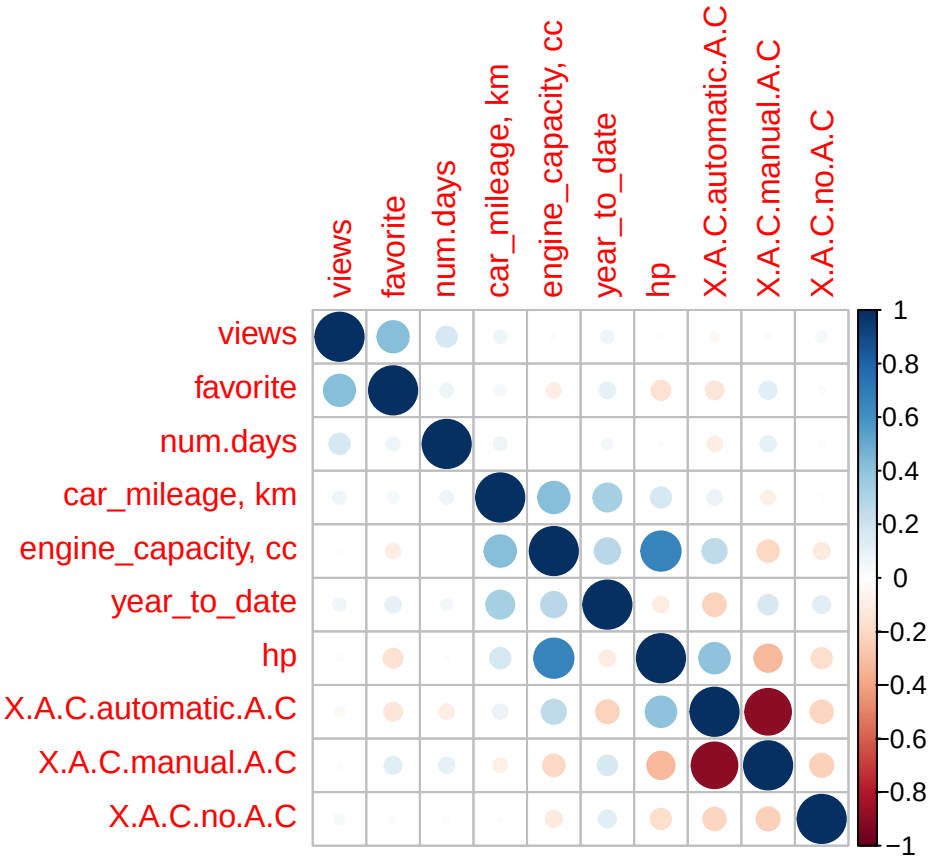
We filter under the following criteria in order to isolate from outliers:

```
filtering_function <- function(data) {
  filter.v1 <- filter(serbia_car_sales_price_2024, doors == "4/5 doors",
                      seats_amount == 5, type_of_drive == "front",
                      favorite > 0, `car_mileage, km` > 10000,
                      hp < 201, `car_mileage, km` < 500000,
                      year_to_date < 30, favorite > 1,
                      price < 20000, views < 500,
                      favorite < 20)
  filter.v1 <- filter.v1 %>% filter(str_detect(gearbox ,"manual"))
  filter.v1 <- dplyr::filter(filter.v1, grepl('gray|black', color))
  filter.v1 <- dplyr::filter(filter.v1, grepl('diesel|petrol', fuel))
  filter.v1 <- dplyr::filter(filter.v1, !grepl('+ gas', fuel))
  return(filter.v1)
}
```

Analysis of Data Post-Filtering

As a result of our filtering, we only include the qualitative variables A/C and Car_Type as our qualitative variables. In referring to our appendix, we notice a remarkable improvement in the correlation between the factors of each these variables when we use the data that is transformed, centered and filtered. Overall for the quantitative variables, we notice an increase in the correlation coefficients, better scatter in the residual plots, increasingly normalized histograms, Q-Q plots closer to $Y = X$ line, and an overall lack of outliers.

Correlation Plot



Model

Main Effect Model

Table 1: Model 1 Performance Metrics

Metric	Value
Number of Predictors	55.0000
R-squared	0.5829
Adjusted R-squared	0.5796
F-statistic	178.3804
MSE	13305526.5422

The interaction model considers relationship between Mileage-Age as mileage might depend on car age, Age-AC as newer cars are more likely to be equipped with AC, and Engine Size-Fuel Type as Engine size might depend on fuel type.

Interaction Model:

Using Backward Elimination & AIC, we obtain our final model:

Final Power Model:

Variable	Coefficient	p-value
(Intercept)	16197.9097	2.27e-06
views	-19.9544	0.0012918
favorite	-1.0477	0.99608
num.days	131.9635	0.062722
`car_mileage, km`	0.0094	0.2604
`engine_capacity, cc`	-8.5512	0.01269
year_to_date	-1252.1465	1.36e-05
hp	111.9306	0.13869
factor(`A/C`)manual A/C	-2501.2330	2.76e-07
factor(`A/C`)no A/C	-548.3344	0.60968
factor(car_type)coupe	530.9379	0.34361
factor(car_type)hatchback	592.5350	1e-04
factor(car_type)limousine	629.3826	1.28e-04
factor(car_type)minivan (MPV)	393.3757	0.069537
factor(car_type)pickup	1211.7106	0.0342
factor(car_type)suv	2243.8919	2.1e-11
factor(fuel)petrol	2976.9832	1.96e-04
I(views^2)	0.0809	0.0037196
I(views^3)	-0.0001	0.0075272
I(favorite^2)	-16.8928	0.61282
I(favorite^3)	0.9816	0.50266
I(num.days^2)	-9.0985	0.10131
I(num.days^3)	0.1989	0.12461
I(`car_mileage, km`^2)	0.0000	0.17935
I(`car_mileage, km`^3)	0.0000	0.31587
I(`engine_capacity, cc`^2)	0.0078	0.0015847
I(`engine_capacity, cc`^3)	0.0000	7.33e-04
I(year_to_date^2)	16.5047	0.3873

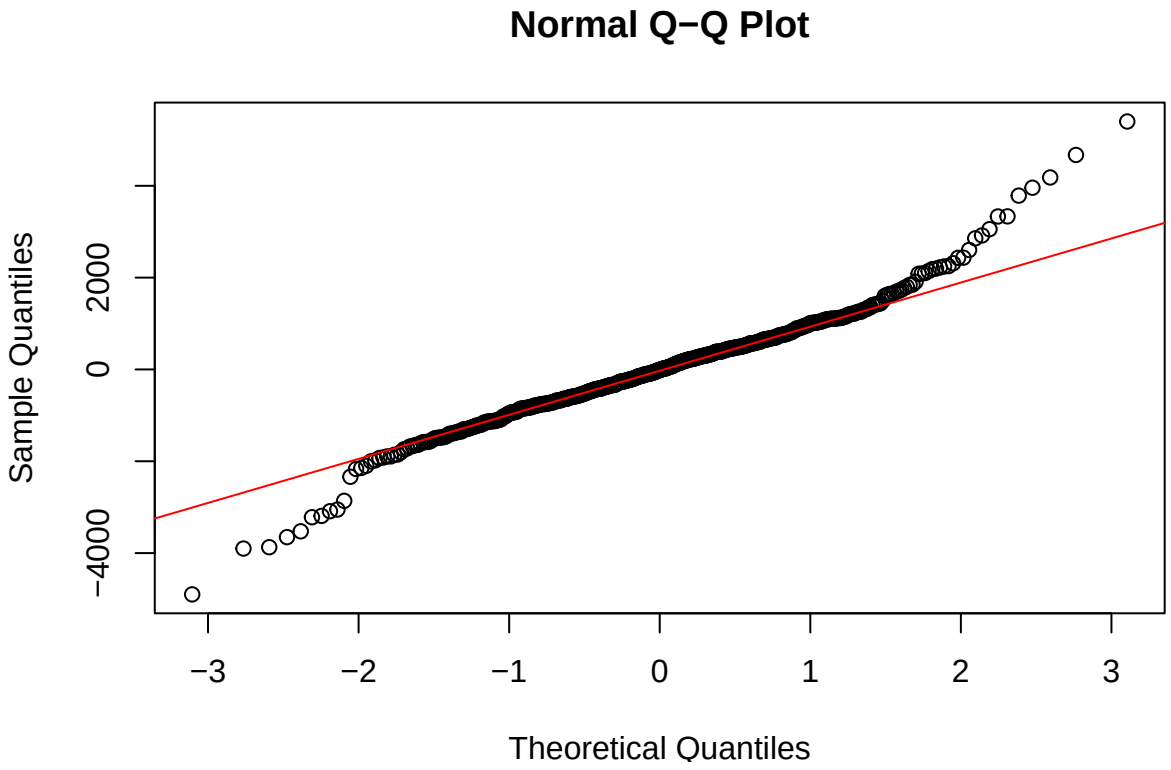
Variable	Coefficient	p-value
I(year_to_dateˆ3)	0.1910	0.63247
I(hpˆ2)	-1.1572	0.091976
I(hpˆ3)	0.0042	0.036802
`car_mileage, km`:year_to_date	0.0001	0.61569
year_to_date:factor(`A/C`)manual A/C	123.8130	7.22e-06
year_to_date:factor(`A/C`)no A/C	26.9193	0.62814
`engine_capacity, cc`:factor(fuel)petrol	-1.9108	1.48e-04

Interpretation of regression coefficients:

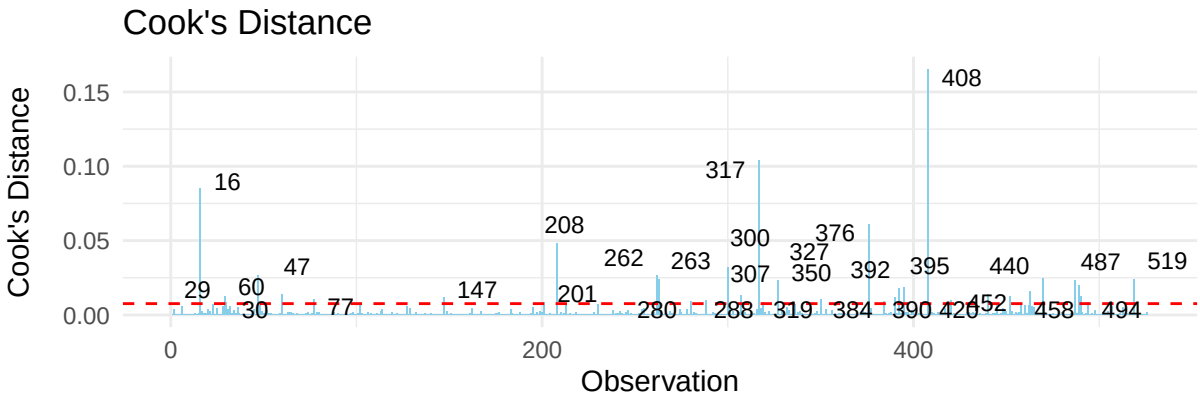
- $\hat{\beta_0}$ (Intercept): The expected mean car price is €16,197.91 when all numeric predictors are zero, for an automatic A/C sedan with baseline fuel type.
- $\hat{\beta_1}$ (views): For each additional view, the expected mean car price decreases by €19.95 while holding all other variables constant.
- $\hat{\beta_2}$ (favorite): For each additional favorite, the expected mean car price decreases by €1.05 while holding all other variables constant.
- $\hat{\beta_3}$ (num.days): For each additional day a listing is active, the expected mean car price increases by €131.96 while holding all other variables constant.
- $\hat{\beta_4}$ (car_mileage, km): For each additional kilometer, the expected mean car price increases by €0.009 while holding all other variables constant.
- $\hat{\beta_5}$ (engine_capacity, cc): For each additional cc of engine size, the expected mean car price decreases by €8.55 while holding all other variables constant.
- $\hat{\beta_6}$ (factor(A/C)manual A/C): The expected mean price difference between manual A/C and automatic A/C (baseline) cars is -€2,501.23 while holding all other variables constant.
- $\hat{\beta_7}$ (factor(A/C)no A/C): The expected mean price difference between no A/C and automatic A/C cars is -€548.33 while holding all other variables constant.
- $\hat{\beta_8}$ (factor(car_type)coupe): The expected mean price difference between coupes and sedans (baseline) is €530.94 while holding all other variables constant.
- $\hat{\beta_9}$ (factor(car_type)hatchback): The expected mean price difference between hatchbacks and sedans is €592.54 while holding all other variables constant.
- $\hat{\beta_{10}}$ (I(views^2)): The expected change in slope for the views-price relationship is +€0.08 per squared view while holding all other variables constant.
- $\hat{\beta_{11}}$ (I(views^3)): The expected change in slope for the views-price relationship is -€0.0001 per cubed view while holding all other variables constant.
- $\hat{\beta_{12}}$ (car_mileage, km:year_to_date): The expected slope difference in the mileage-price relationship for each additional year of car age is +€0.0001 while holding all other variables constant.
- $\hat{\beta_{13}}$ (year_to_date:factor(A/C)manual A/C): The expected slope difference in the age-price relationship for manual A/C versus automatic A/C cars is +€123.81 per year while holding all other variables constant.
- $\hat{\beta_{14}}$ (engine_capacity, cc:factor(fuel)petrol): The expected slope difference in the engine size-price relationship for petrol versus baseline fuel type cars is -€1.91 per cc while holding all other variables constant.

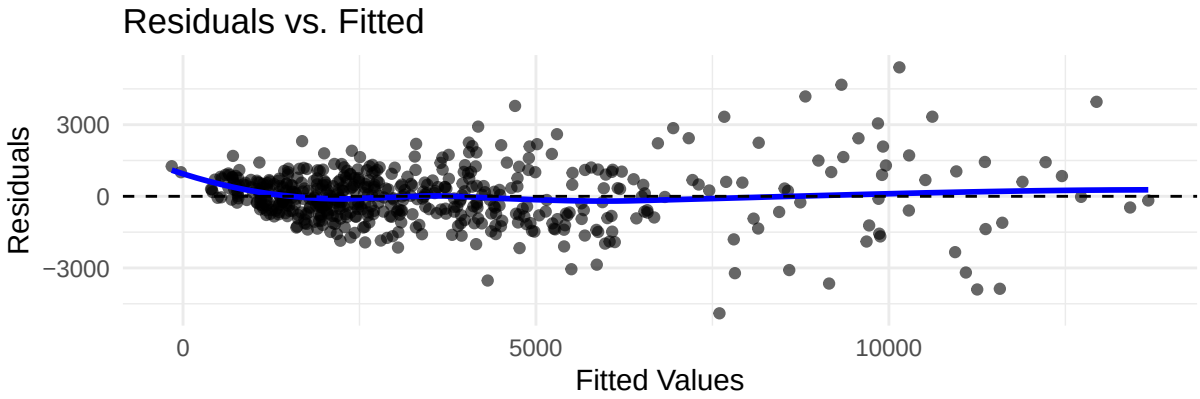
Model Name	R^2	Adjusted R^2	AIC Score	PRESS Statistic
Main Effect Model	0.583	0.580	136267.329	97454235514
Interaction Model	0.788	0.779	9097.306	1039104954
Power Model	0.838	0.826	8984.886	828871922

Model Diagnostics

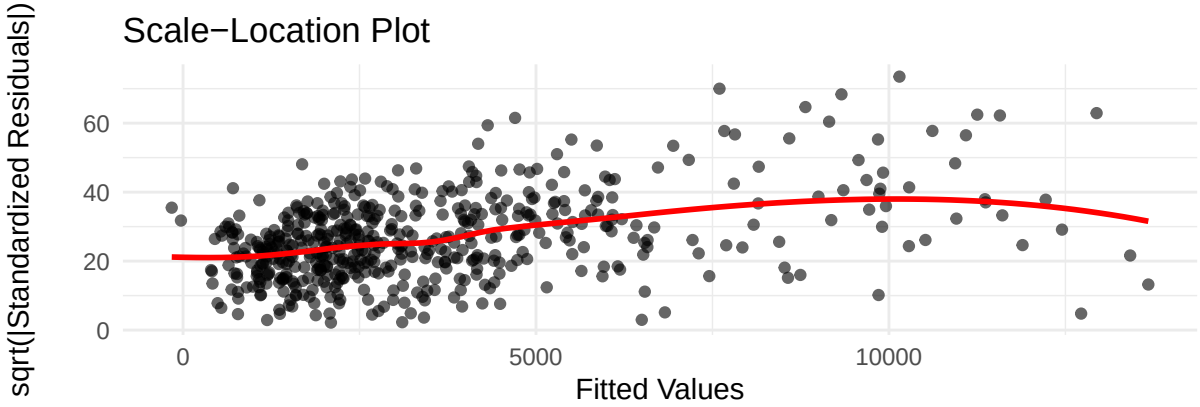


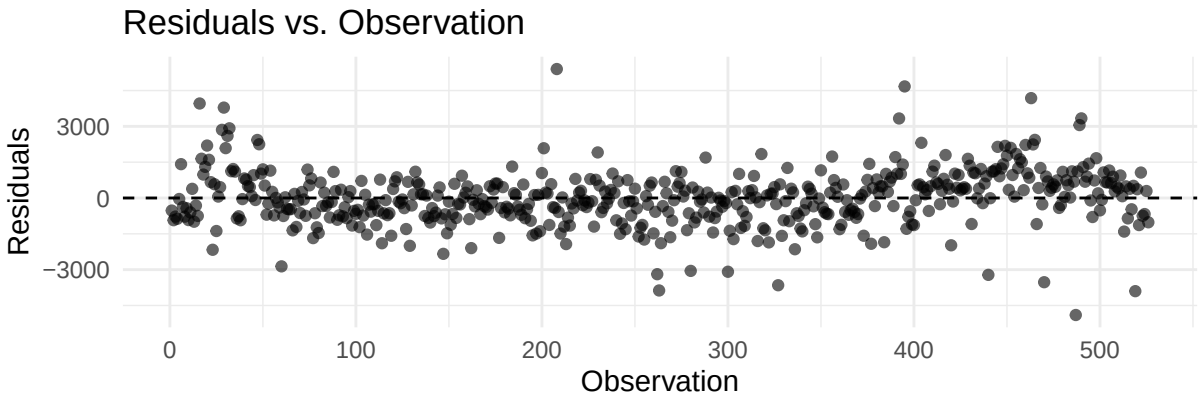

```
## `geom_smooth()` using formula = 'y ~ x'
```






```
## `geom_smooth()` using formula = 'y ~ x'
```





Model Validation

```
## [1] 1431999
```

	MSPE	MSE
Validation Statistics	1815866	1376649

	sqrt(MSPE)	sqrt(MSE)
Validation Statistics	1348	1173

The validation results for the Power Model shows that it achieves a Mean Squared Prediction Error (MSPE) of 1,815,866 and a Mean Squared Error (MSE) of 1,376,649 on the validation and training data, respectively. When converted to the root scale, these translate to average prediction errors of ± 1,348 MSPE and ± 1,173 MSE for the training data. This indicates the model has moderate overfitting, performing slightly worse on unseen data, but still maintains reasonable predictive accuracy within a ~6.5% error margin for an average-priced car. The results suggest the model generalizes decently but could benefit from further refinement to reduce the gap between training and validation performance.

Residuals vs Fitted: We get a trumpet shaped pattern with most points spreading out further to the right. This suggests heteroscedasticity, which means that our assumption that we have equal variance may have be violated.

Normal Q-Q Plot: We can see on the plot that most of the points - with the exception of a few points near the ends - fall on the line, which suggests that the normality assumption has not been violated.

Scale-Location Plot: The plot shows that the red line is not exactly linear. There is a small hill near the right, going up then down. The points are also spread out more to the right This suggests that the homoscedasticity assumption may not hold, and variance may not be constant.

Residuals vs. Observation Plot: The points are completely scattered with no obvious pattern. It doesn't indicate that any assumptions were violated.

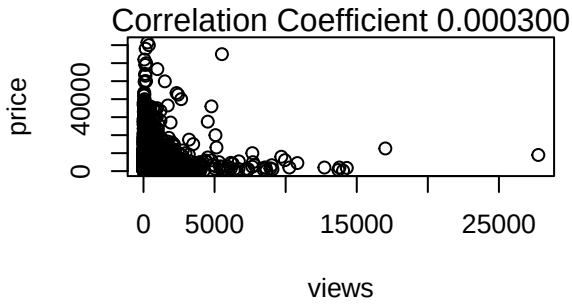
Discussion and Conclusion

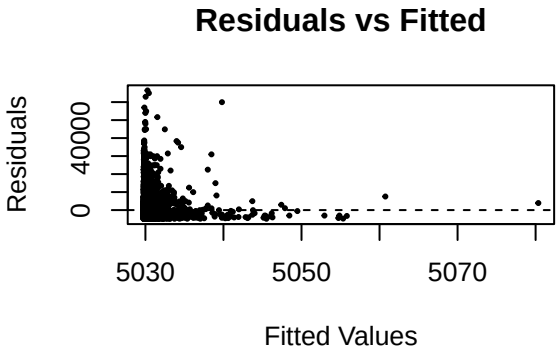
The goal of our study was to predict the price of a car based on different car features, using regression techniques. After looking through possible predictors, testing different interaction terms and polynomial transformations, we developed our final model, which includes predictors such as views, favorite, car mileage, engine capacity, horsepower. Qualitative factors such as A/C, car type, and fuel, were included as well. Polynomial terms up to the second and third order, as well as interaction terms like mileage-age and fuel-efficienct were also included after discovering they improved model performance. Part of our analysis focused on filtering the data based on qualitative variables and ensuring that we worked with a cleaner subset of the data free from outliers. After transforming, centering, and filtering the data, we observed improvement in the correlations between variables. The correlation plots demonstrated better relationships between features, and the residual plots, histograms, and Q-Q plots showed better normalization and fewer outliers. The use of filtering and transformations notably improved the quality of the analysis, resulting in more reliable models. Our final model appears to satisfy the assumption of normality, but there are clear violations of homoscedasticity as seen through the residual plots. Given these violations, remedial measures such as Weighted Least Squares should be applied. In conclusion, the analysis has provided valuable insights into the factors that drive car prices, and the models developed can serve to make future predictions.

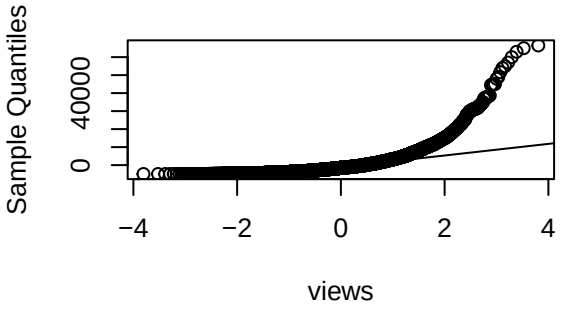
References

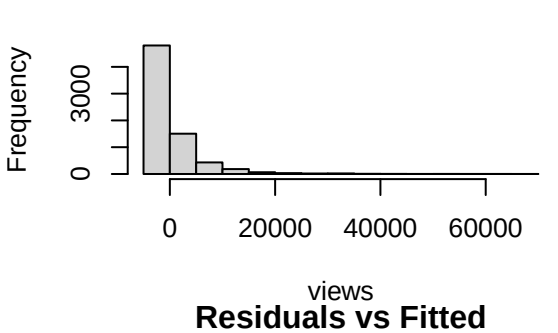
Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). Applied Linear Regression Models: Michael H. Kutner, Christopher J. Nachtsheim, John Neter. McGraw-Hill.

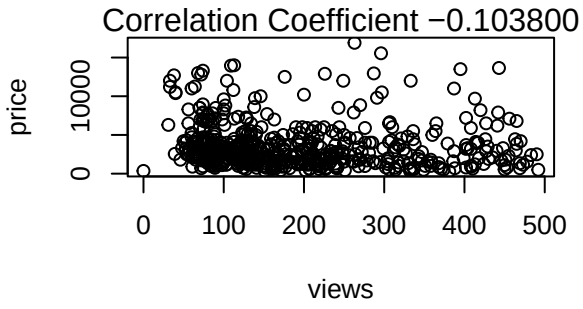
Appendix

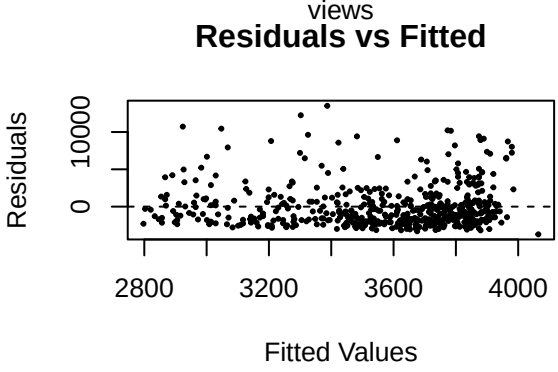

Correlation Coefficient 0.000300
price
0
40000
0
5000
15000
25000
views

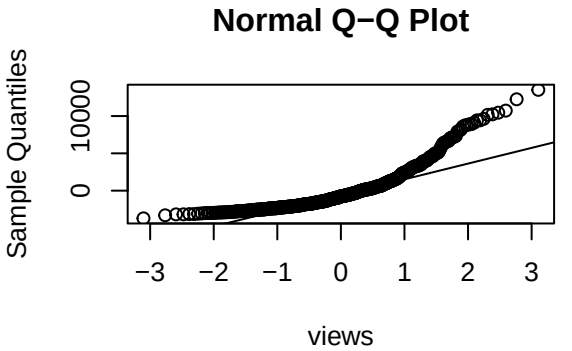

Residuals vs Fitted
Residuals
0
40000
5030
5050
5070
Fitted Values

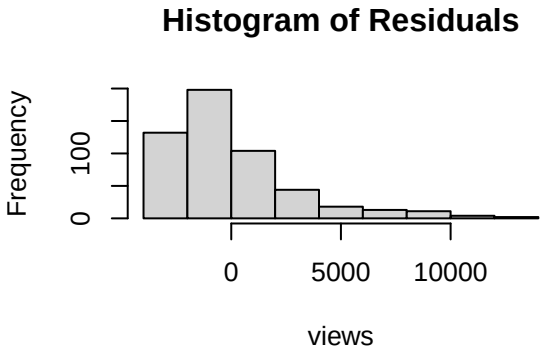

Normal Q−Q Plot
Sample Quantiles
0
40000
−4
−2
0
2
4
views

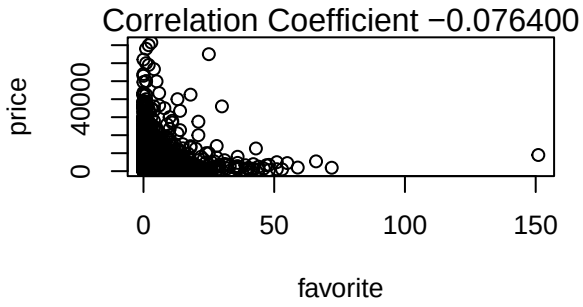

Histogram of Residuals
Frequency
0
3000
0
20000
40000
60000
views

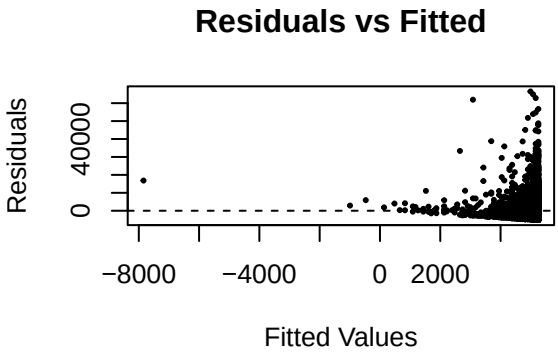

Correlation Coefficient −0.103800
price
0
10000
0
100
200
300
400
500
views

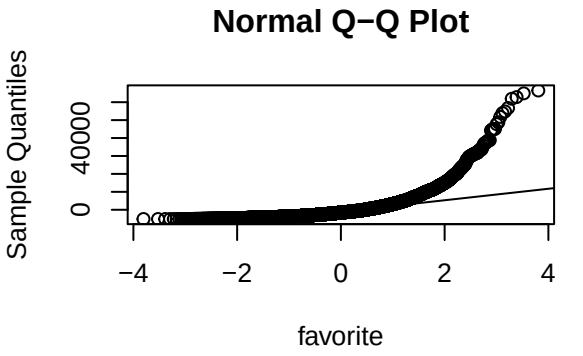

Residuals vs Fitted
Residuals
0
10000
2800
3200
3600
4000
Fitted Values

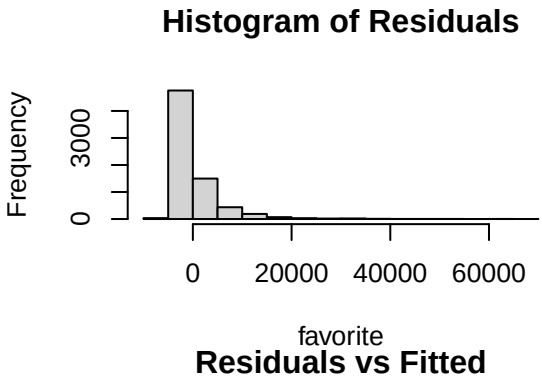

Normal Q−Q Plot
Sample Quantiles
0
10000
−3
−2
−1
0
1
2
3
views

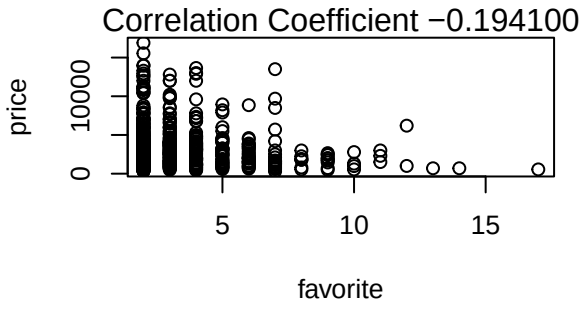

Histogram of Residuals
Frequency
0
100
0
5000
10000
views

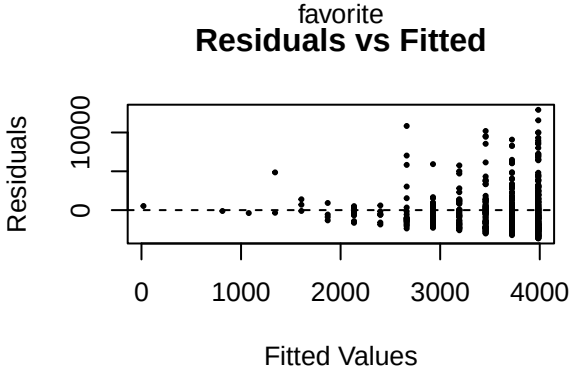

Correlation Coefficient −0.076400
price
0
40000
0
50
100
150
favorite

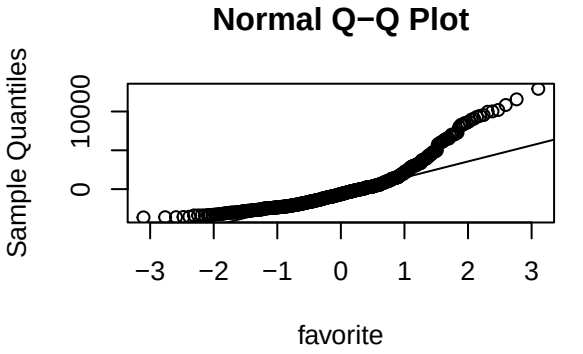

Residuals vs Fitted
Residuals
0
40000
−8000
−4000
0
2000
Fitted Values

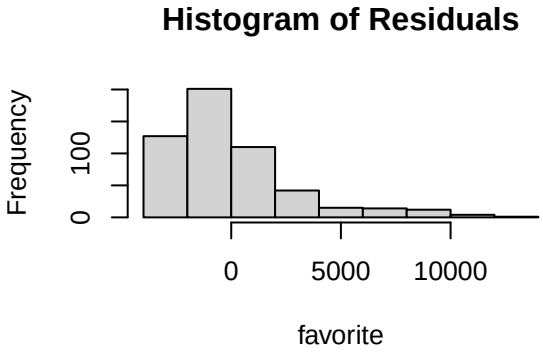

Normal Q−Q Plot
Sample Quantiles
0
40000
−4
−2
0
2
4
favorite

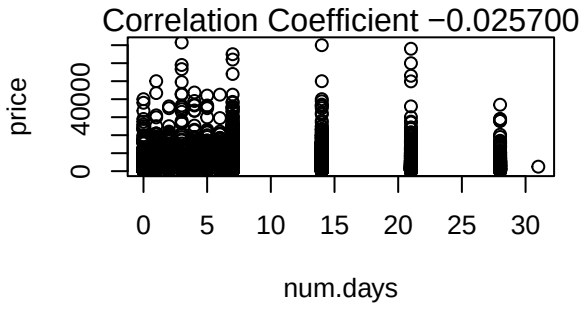

Histogram of Residuals
Frequency
0
3000
0
20000
40000
60000
favorite

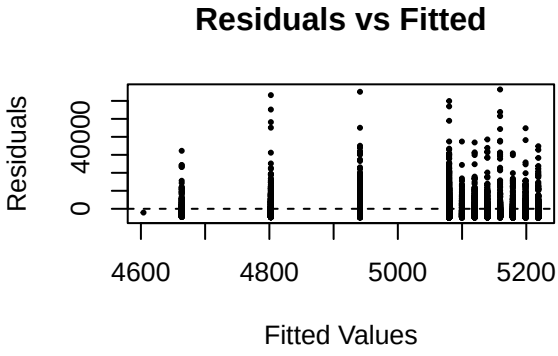

Correlation Coefficient −0.194100
price
0
10000
5
10
15
favorite

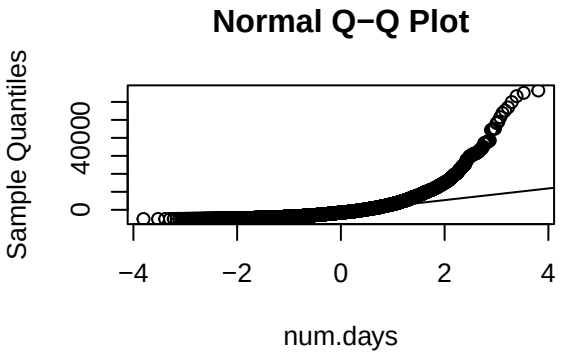

Residuals vs Fitted
Residuals
0
10000
0
1000
2000
3000
4000
Fitted Values

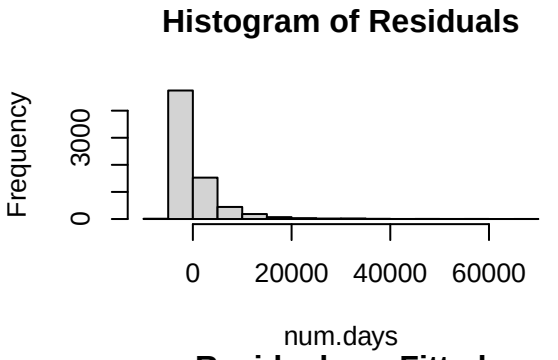

Normal Q−Q Plot
Sample Quantiles
0
10000
−3
−2
−1
0
1
2
3
favorite

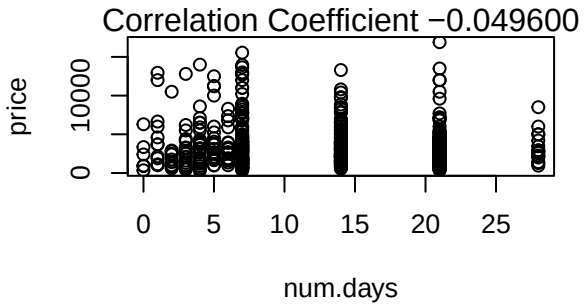

Histogram of Residuals
Frequency
0
100
0
5000
10000
favorite

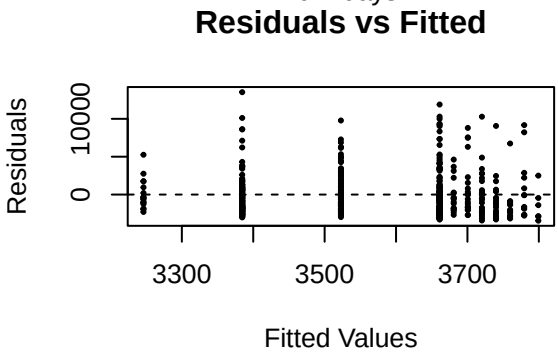

Correlation Coefficient −0.025700
price
0
40000
0
5
10
15
20
25
30
num.days

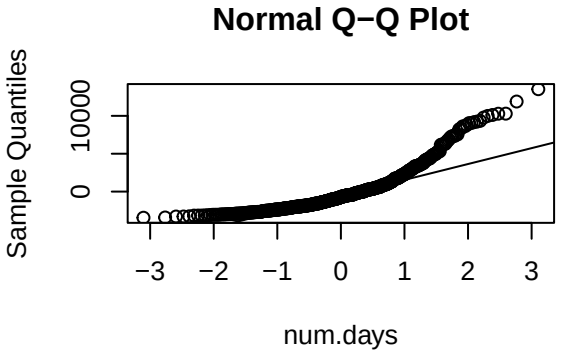

Residuals vs Fitted
Residuals
0
40000
4600
4800
5000
5200
Fitted Values

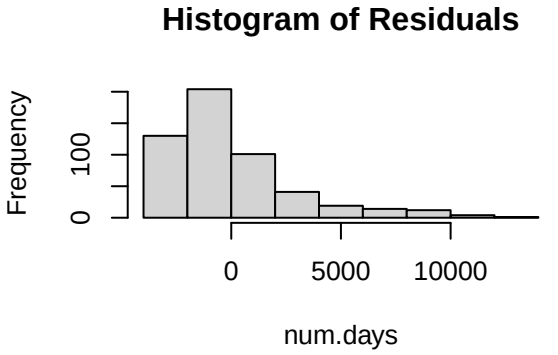

Normal Q−Q Plot
Sample Quantiles
0
40000
−4
−2
0
2
4
num.days

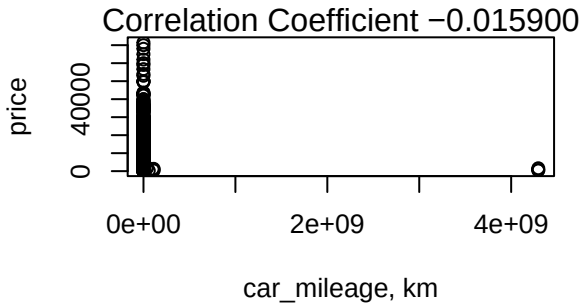

Histogram of Residuals
Frequency
0
3000
0
20000
40000
60000
num.days

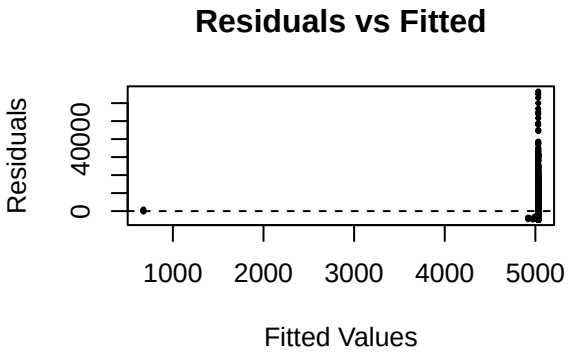

Correlation Coefficient −0.049600
price
0
10000
0
5
10
15
20
25
num.days

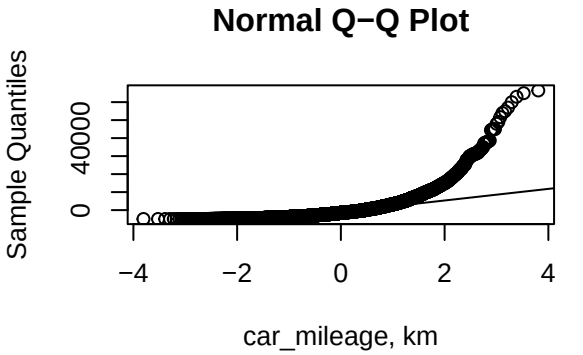

Residuals vs Fitted
Residuals
0
10000
3300
3500
3700
Fitted Values

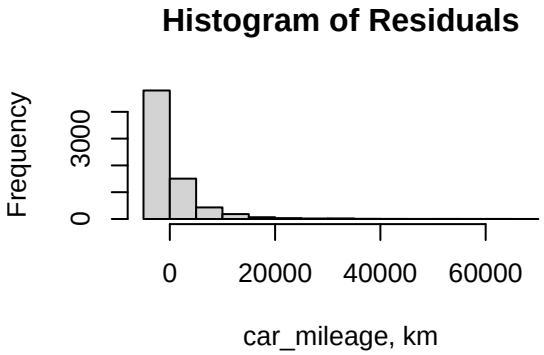

Normal Q−Q Plot
Sample Quantiles
0
10000
−3
−2
−1
0
1
2
3
num.days

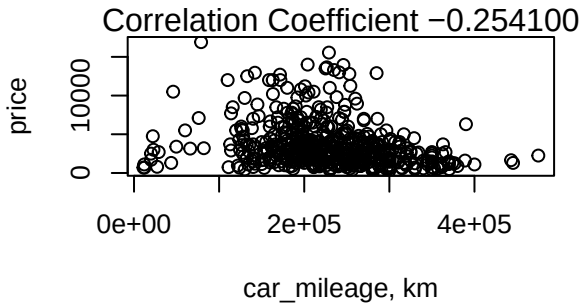

Histogram of Residuals
Frequency
0
100
0
5000
10000
num.days

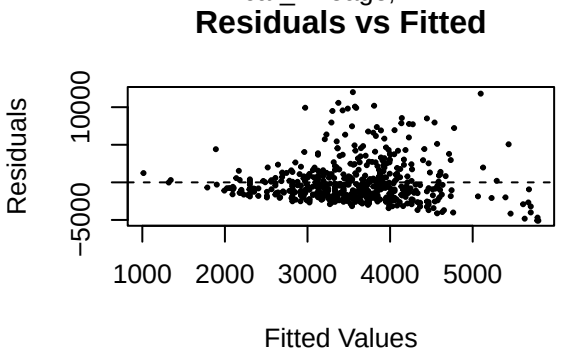

Correlation Coefficient −0.015900
price
0
40000
0e+00
2e+09
4e+09
car_mileage, km

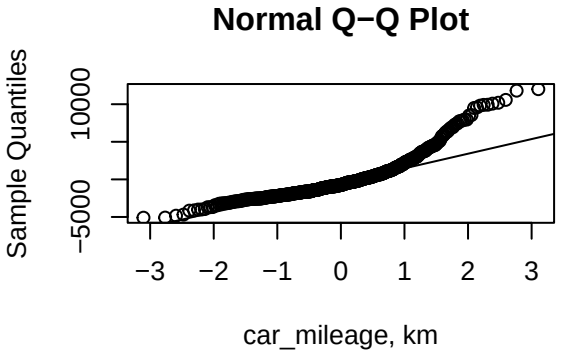

Residuals vs Fitted
Residuals
0
40000
1000
2000
3000
4000
5000
Fitted Values

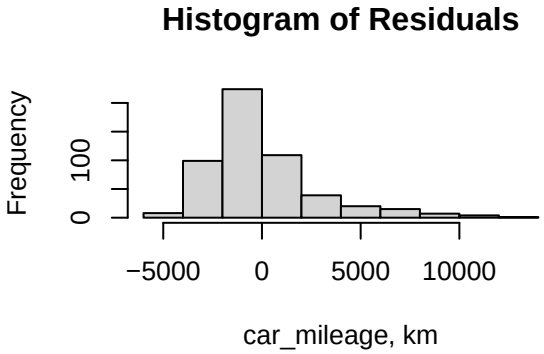

Normal Q−Q Plot
Sample Quantiles
0
40000
−4
−2
0
2
4
car_mileage, km

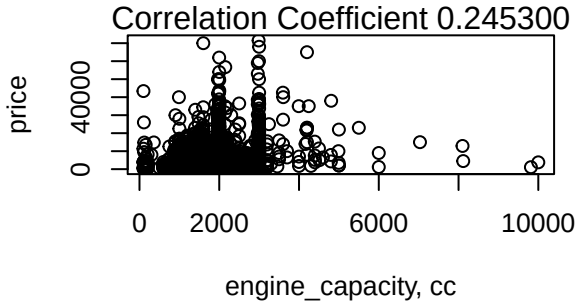

Histogram of Residuals
Frequency
0
3000
0
20000
40000
60000
car_mileage, km

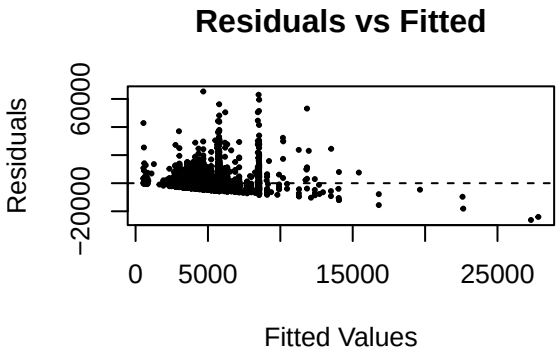

Correlation Coefficient −0.254100
price
0
10000
0e+00
2e+05
4e+05
car_mileage, km

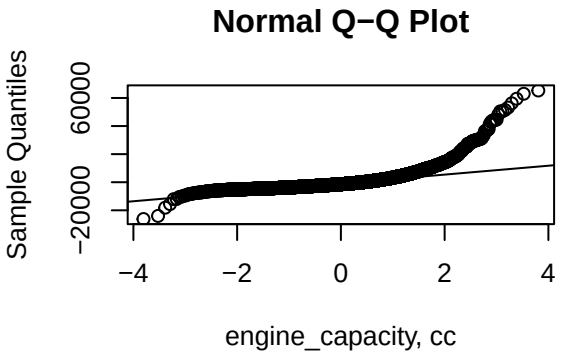

Residuals vs Fitted
Residuals
−5000
10000
1000
2000
3000
4000
5000
Fitted Values

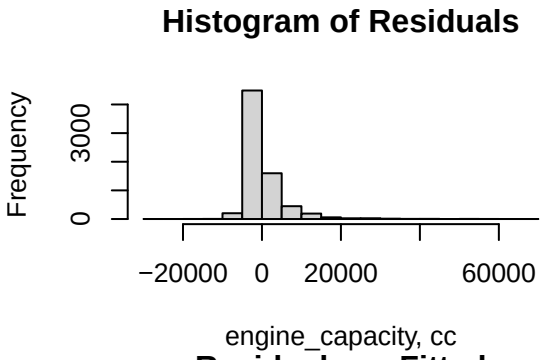

Normal Q−Q Plot
Sample Quantiles
−5000
10000
−3
−2
−1
0
1
2
3
car_mileage, km

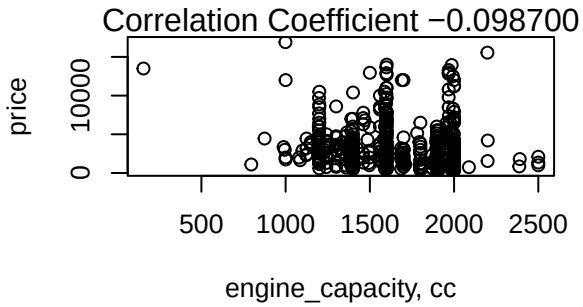

Histogram of Residuals
Frequency
0
100
−5000
0
5000
10000
car_mileage, km

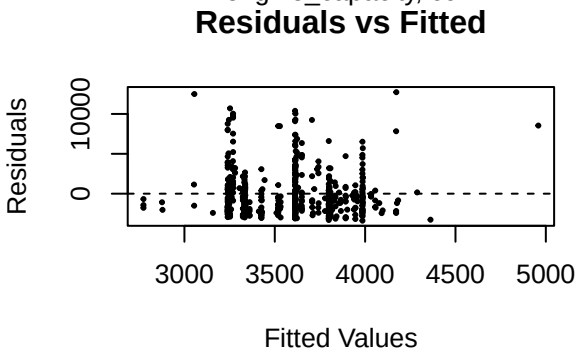

Correlation Coefficient 0.245300
price
0
40000
0
2000
6000
10000
engine_capacity, cc

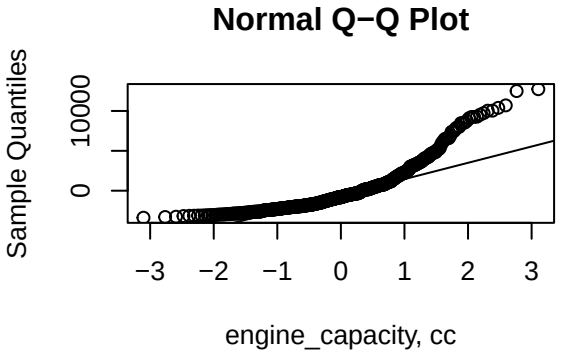

Residuals vs Fitted
Residuals
−20000
0
60000
0
5000
15000
25000
Fitted Values

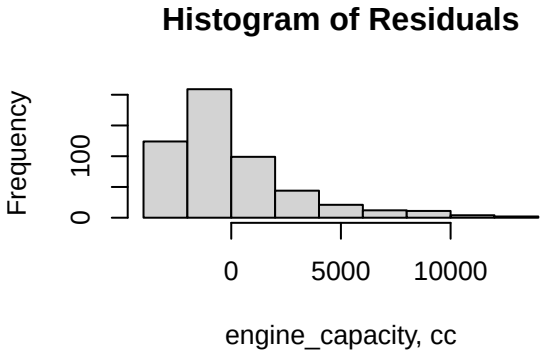

Normal Q−Q Plot
Sample Quantiles
−20000
0
60000
−4
−2
0
2
4
engine_capacity, cc

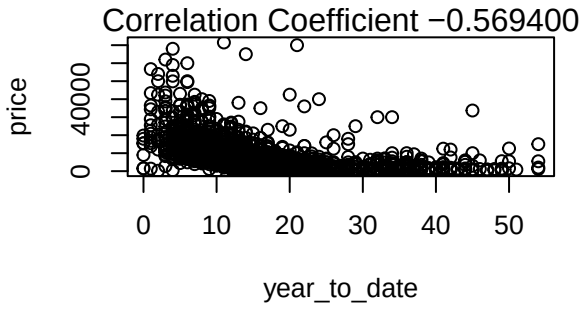

Histogram of Residuals
Frequency
0
3000
−20000
0
20000
60000
engine_capacity, cc

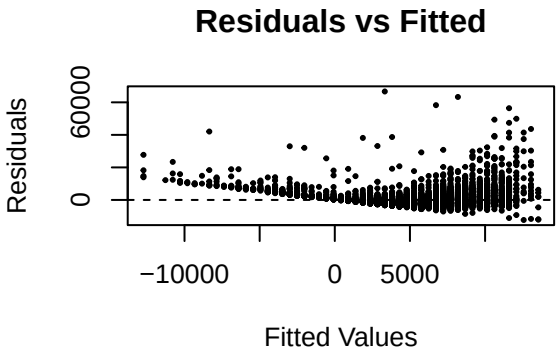

Correlation Coefficient −0.098700
price
0
10000
500
1000
1500
2000
2500
engine_capacity, cc

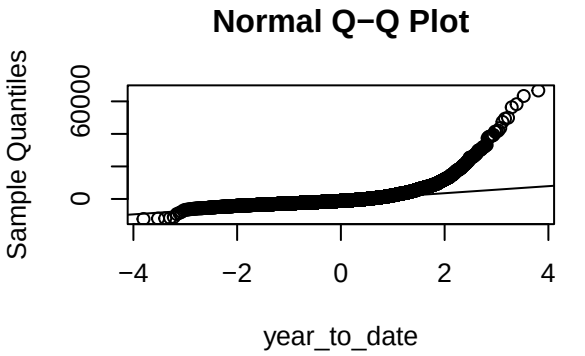

Residuals vs Fitted
Residuals
0
10000
3000
3500
4000
4500
5000
Fitted Values

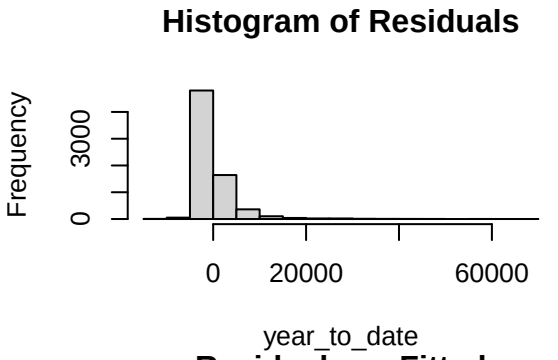

Normal Q−Q Plot
Sample Quantiles
0
10000
−3
−2
−1
0
1
2
3
engine_capacity, cc

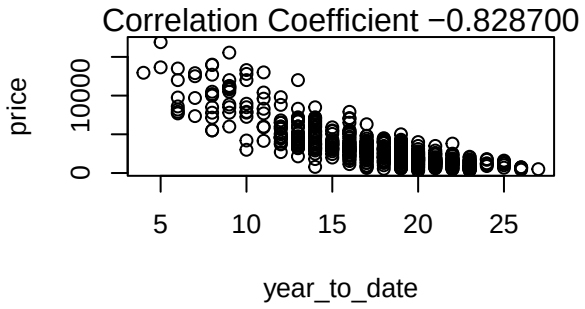

Histogram of Residuals
Frequency
0
100
0
5000
10000
engine_capacity, cc

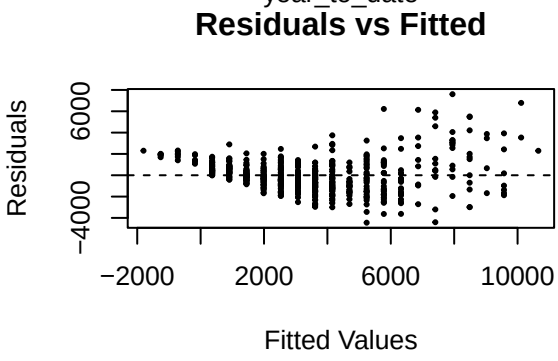

Correlation Coefficient −0.569400
price
0
40000
0
10
20
30
40
50
year_to_date

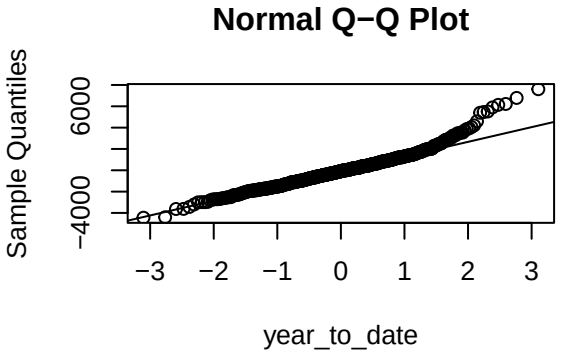

Residuals vs Fitted
Residuals
0
60000
−10000
0
5000
Fitted Values

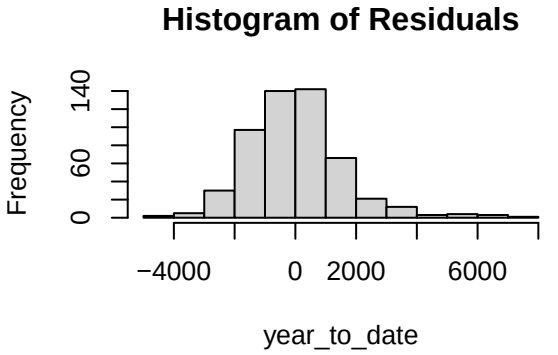

Normal Q−Q Plot
Sample Quantiles
0
60000
−4
−2
0
2
4
year_to_date

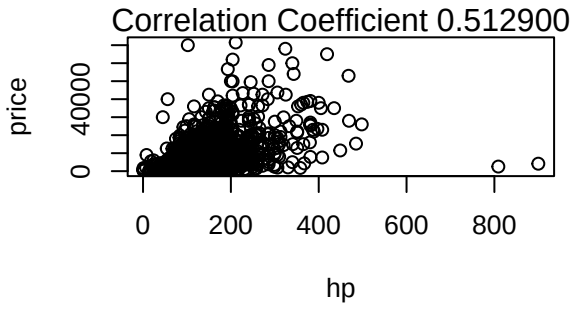

Histogram of Residuals
Frequency
0
3000
0
20000
60000
year_to_date

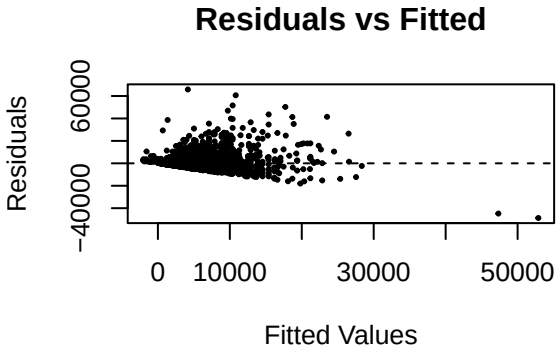

Correlation Coefficient −0.828700
price
0
10000
5
10
15
20
25
year_to_date

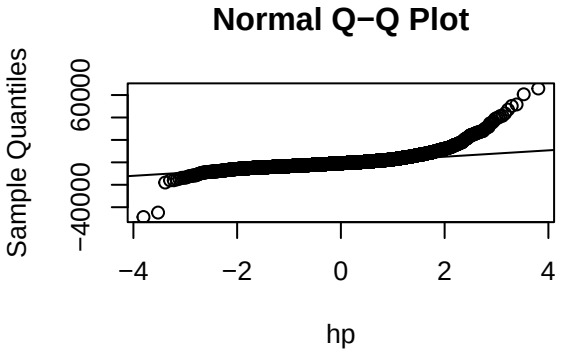

Residuals vs Fitted
Residuals
−4000
6000
−2000
2000
6000
10000
Fitted Values

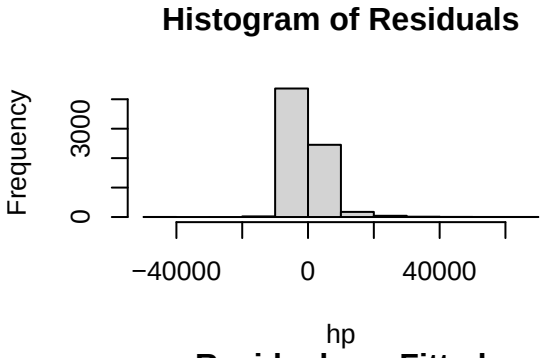

Normal Q−Q Plot
Sample Quantiles
−4000
6000
−3
−2
−1
0
1
2
3
year_to_date

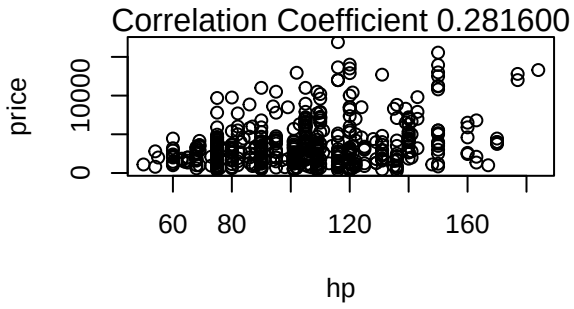

Histogram of Residuals
Frequency
0
60
140
−4000
0
2000
6000
year_to_date

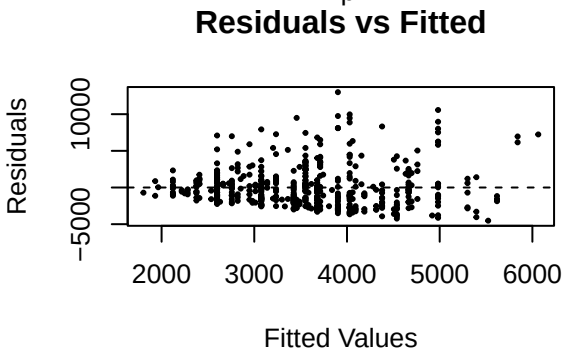

Correlation Coefficient 0.512900
price
0
40000
0
200
400
600
800
hp

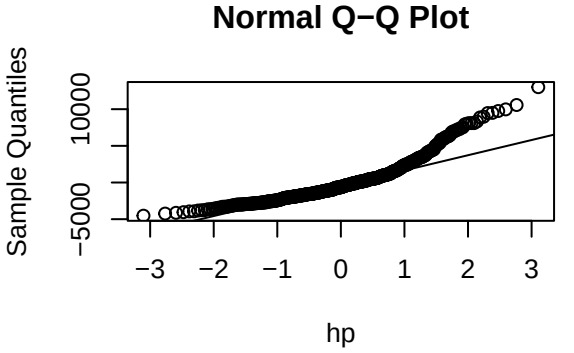

Residuals vs Fitted
Residuals
−40000
0
60000
0
10000
30000
50000
Fitted Values

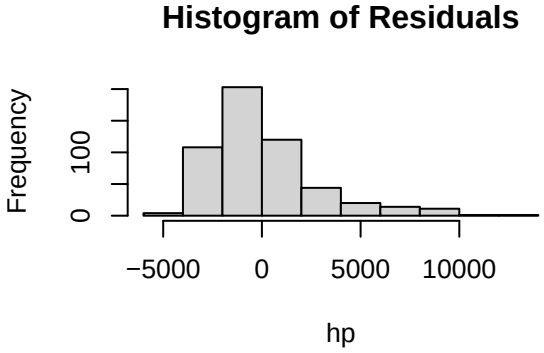

Normal Q−Q Plot
Sample Quantiles
−40000
0
60000
−4
−2
0
2
4
hp


Histogram of Residuals
Frequency
0
3000
−40000
0
40000
hp


Correlation Coefficient 0.281600
price
0
10000
60
80
120
160
hp


Residuals vs Fitted
Residuals
−5000
0
10000
2000
3000
4000
5000
6000
Fitted Values


Normal Q−Q Plot
Sample Quantiles
−5000
0
10000
−3
−2
−1
0
1
2
3
hp


Histogram of Residuals
Frequency
0
100
−5000
0
5000
10000
hp