

CH8_Q5

Fasih Atif

12/20/2020

Our task is to analyze Price vs Age of cars to find a good car deal. For this purpose, I web scrapped data from a cars website named truecars.com which sells used cars and is based in the United States. The brand and type of car we will use for analysis is the Ford Fusion.



Price vs Age Analysis for Ford Fusion

The top 5 rows of the cars database that we scrapped is shown below:

X1	model_name	make_year	price	car_description	location	car_colour	age
1	Ford Fusion	2019	17485	No accidents, 1 Owner, Fleet use	Boise, ID	Gray exterior, Black interior	1
2	Ford Fusion	2019	16500	No accidents, 1 Owner, Personal use	Wilson, NC	Gray exterior, Black interior	1
3	Ford Fusion	2017	11990	No accidents, 1 Owner, Fleet use	Hempstead, NY	Black exterior, Unknown interior	3
4	Ford Fusion	2013	9494	No accidents, 2 Owners, Fleet use	Mechanicsville, VA	White exterior, Black interior	7
5	Ford Fusion	2019	16000	No accidents, 2 Owners, Fleet use	Manning, SC	Black exterior, Black interior	1
6	Ford Fusion	2013	6500	No accidents, 2 Owners, Personal use	Chesapeake, VA	Black exterior, Black interior	7

Variable Transformation

Some non linear patterns can be better approximated by linear regressions if the X or Y or both are transformed by taking logs. This is often a good choice when the data is skewed. It is important to think beforehand what kind of transformations are feasible or not. We would like a transformation that explains the association between the dependent and independent variable clearly and interpretations that make the most sense.

Level-Level Model: $\text{price} = \alpha + \beta \cdot \text{age}$ (Figure 1)

Log-Log Model: $\ln_price = \alpha + \beta \cdot \ln_age$ (Figure 2)

Log-Level Model: $\ln_price = \alpha + \beta \cdot \text{age}$ (Figure 3)

Based on model comparison, our chosen model is **Log-Level Model**.

- Substantive Reasoning: Price decrease is measured in percentage changes so it can be easily related with age. Our interpretations will tell us the percentage change with which prices decrease as age increases in absolute terms.
- Statistical Reasoning: It is a linear model which is easy to interpret and captures the variation well.

Log Transformation of variables for Regression

Based on our investigation, we found out that log transformations would work best on both dependent variable. So, we created a new log variables for car_price named \ln_price .

Estimating different Regression models

We ran different regression models in order to try to capture the non linearity in the best we could. We used simple linear regression, quadratic regression, and piecewise linear regression to estimate. All models gave approximately the same result:

	\ln_price - linear	\ln_price - quadratic	\ln_spline - PLS
(Intercept)	9.81*** (0.00)	9.79*** (0.01)	9.80*** (0.00)
age	-0.11*** (0.00)	-0.10*** (0.00)	
age_sq		-0.00** (0.00)	
lspline(age, cutoff)1			-0.11*** (0.00)
lspline(age, cutoff)2			-0.14*** (0.01)
R ²	0.79	0.79	0.79
Adj. R ²	0.79	0.79	0.79
Num. obs.	3003	3003	3003
RMSE	0.16	0.16	0.16

*** p < 0.001; ** p < 0.01; * p < 0.05

Modelling Price vs Age of Cars

The model that i think is the best is the simple linear regression. It has the same result as other models and is simple to use.

Interpretation

Beta: Car price decreases by 11% as age increases by 1.

Residual Analysis

We can use our residual results to find the best deals. The most lowest and negative residuals will be the best car deals for us. The best deals are shown below:

model_name	make_year	car_description	car_colour	location	car_price	linear_reg_res
Ford	2012	3 accidents, 3 Owners,	Unknown exterior,	Greensboro,	3800	-
Fusion		Fleet use	Unknown interior	NC		0.6592519
Ford	2011	3 accidents, 5 Owners,	Silver exterior, Black	Garland,	3495	-
Fusion		Fleet use	interior	TX		0.6289747
Ford	2012	No accidents, 2	White exterior,	Raleigh,	3450	-
Fusion		Owners, Personal use	Unknown interior	NC		0.7558787
Ford	2008	No accidents, 4	Unknown exterior,	Anniston,	1840	-
Fusion		Owners, Fleet use	Unknown interior	AL		0.9287079
Ford	2013	No accidents, 1 Owner,	Black exterior, Gray	Hicksville,	4195	-
Fusion		Personal use	interior	OH		0.6743044

Appendix

Figure 1

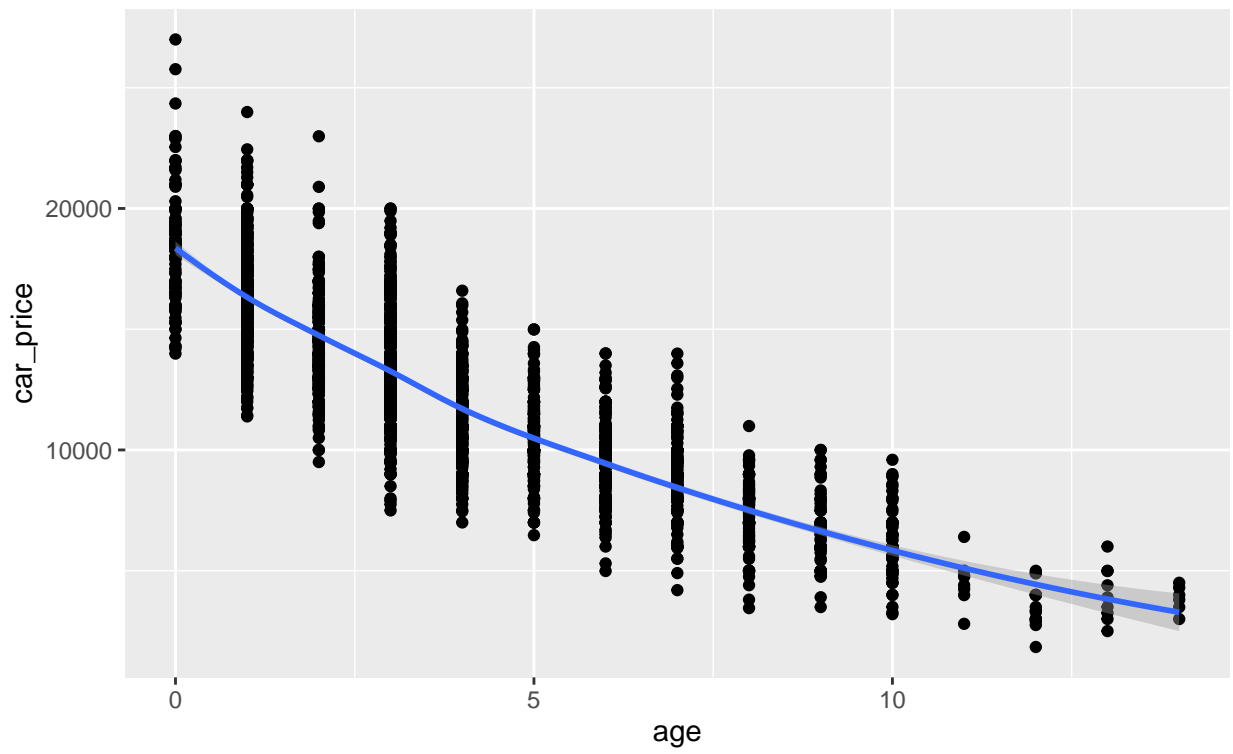


Figure 2

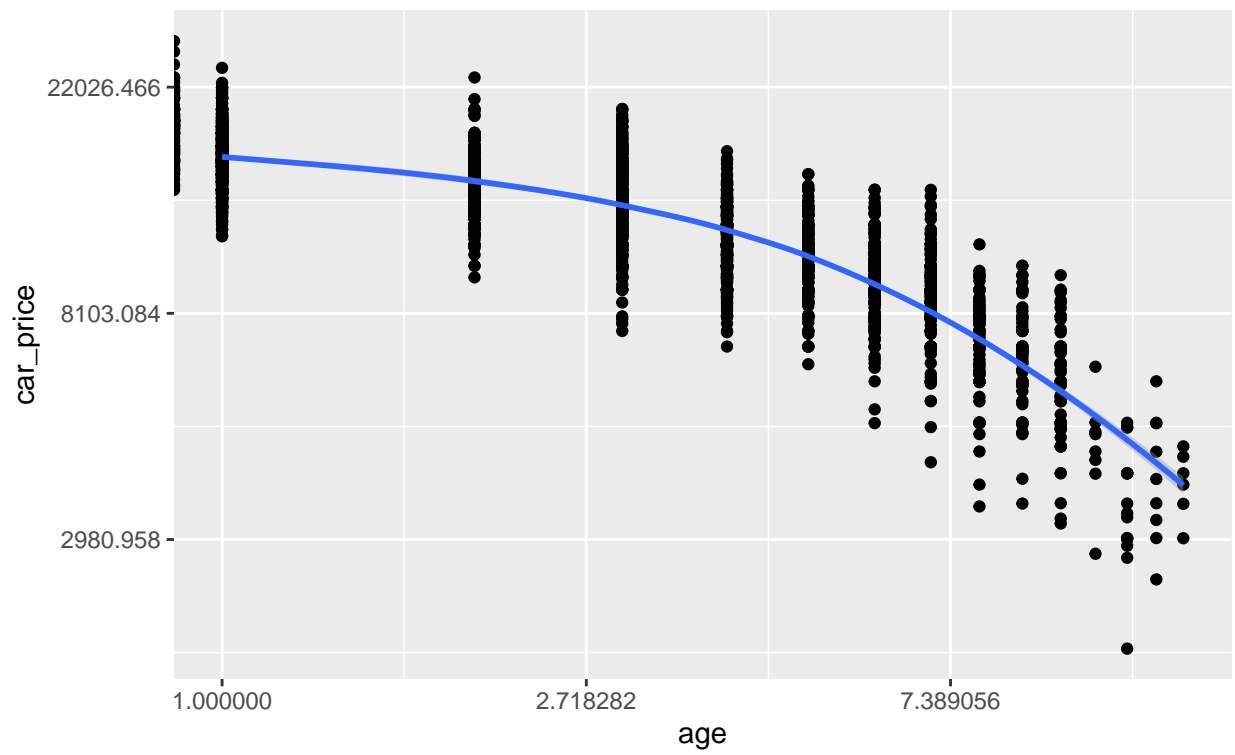


Figure 3

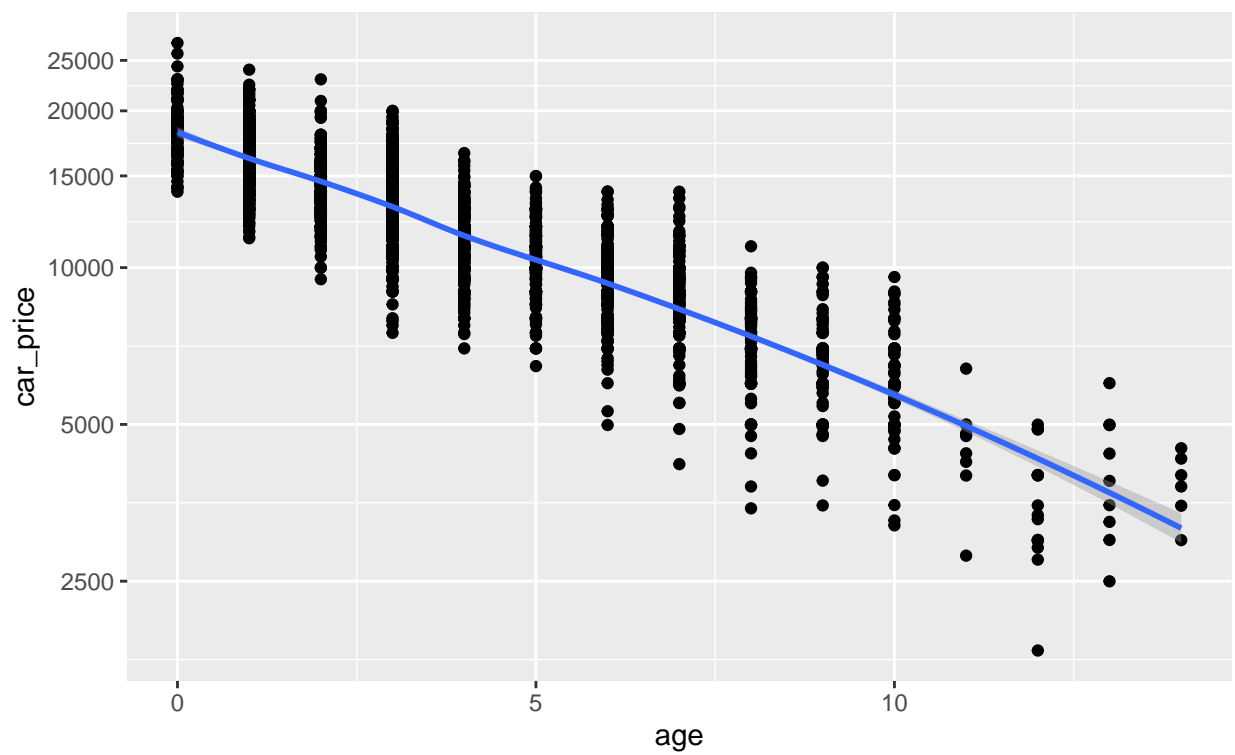


Figure 4

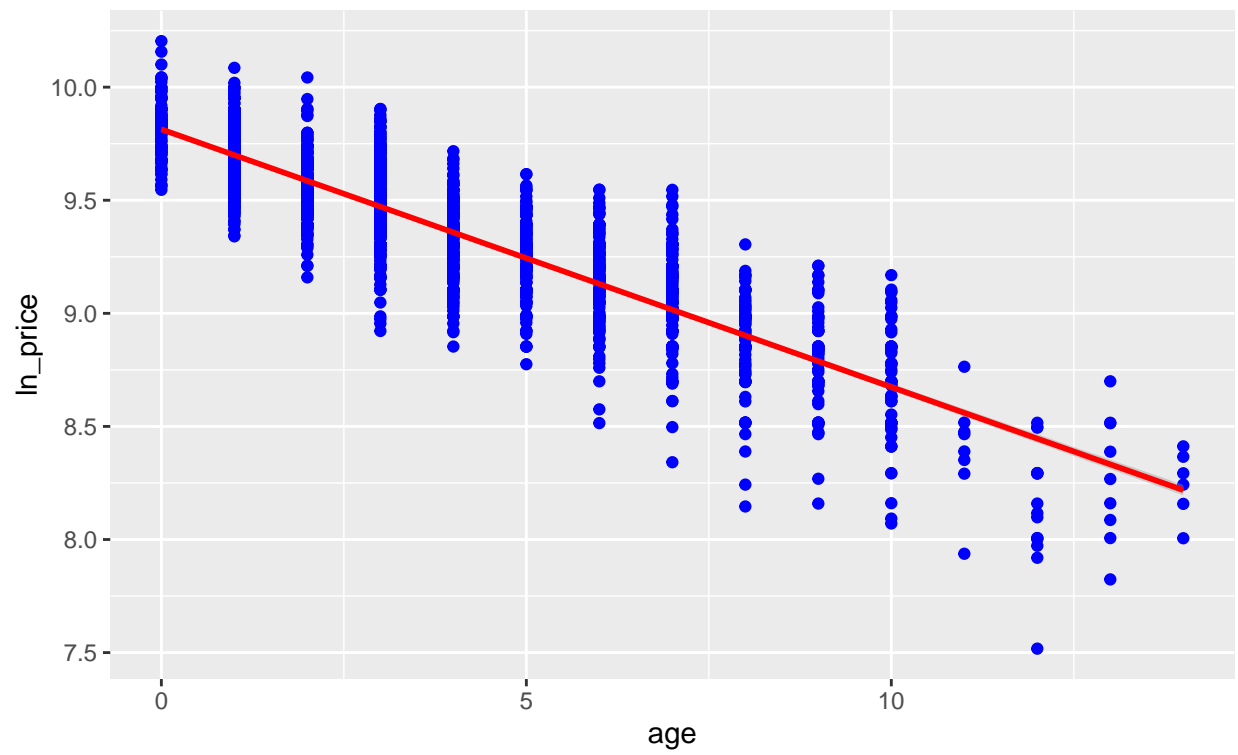


Figure 5

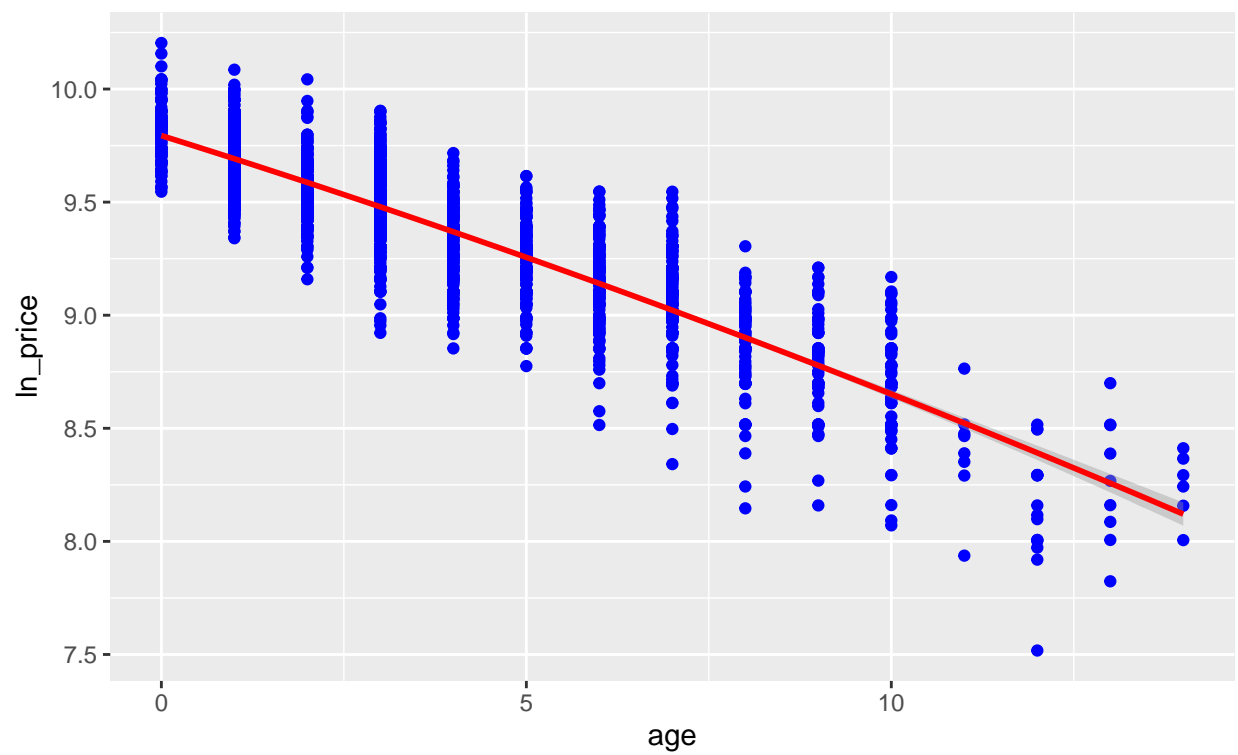


Figure 6

