The Mission: How not to get ripped off on Craig's List

Detective: Vicky Clayton


The Why:
Trying to buy a
car

CL fort smith > for sale > cars+trucks 2 hidden post account

cars & trucks

search cars & trucks

save search

all owner dealer

- search titles only
- has image
- posted today
- bundle duplicates
- include nearby areas

MILES FROM ZIP
miles from zip

PRICE
min max

MAKE AND MODEL
make / model

MODEL YEAR
min max

ODOMETER
min max

- cryptocurrency ok

- condition
- cylinders
- drive
- fuel
- paint color

$15500

Feb 2 Dodge Ram 1500 ST $15500 (Eufaula ok)

$8000

Feb 2 07 f-150 Lariat 4x4 $8000 (Fort smith

$850

$2555

No Min
No Max

Year Range

No Min No Max

Mileage Under

No Max

Special Offers

- CERTIFIED PRE-OWNED (146)
- RECENTLY REDUCED (286)
- UNDER $35K SPECIALS (102)

Make

- CHEVROLET (17)
- CHRYSLER (242)

New 2016 Honda Odyssey EX-L

ABOUT 9 MILES FROM 94103

Recently Reduced

Stock	GB093073
Mileage	322
Color	White
Conv Make	VMI
Conversion	Side Entry Automatic In Floor
Location	Oakland CA

price: $71,900
dealer discount: $8,000
sale price: **$63,900**

MORE DETAILS
REQUEST MORE INFO COMPARE

New 2016 Honda Odyssey EX-L

ABOUT 9 MILES FROM 94103

Stock	GB113417
Mileage	316
Color	Black
Conv Make	VMI
Conversion	Side Entry Automatic In Floor
Location	Oakland CA

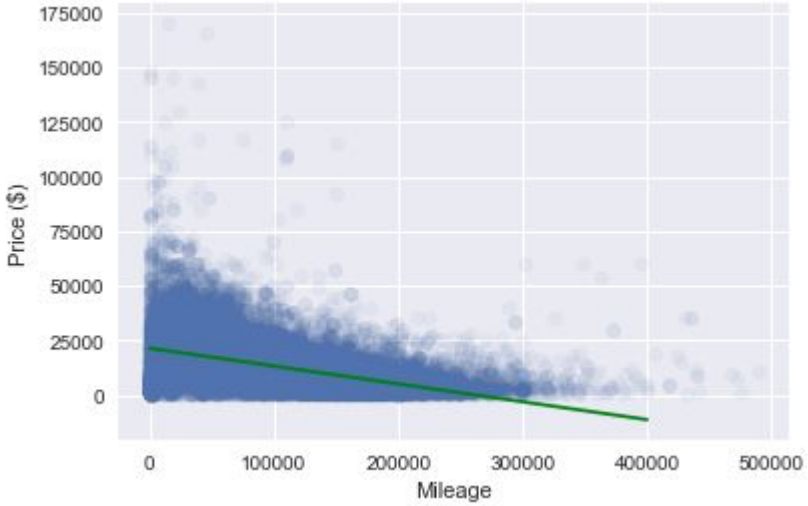
Request More Info for Price

MORE DETAILS
REQUEST MORE INFO COMPARE

N = 39000

Mileage
Age
Brand
“Condition”
Paint color

Start simple...

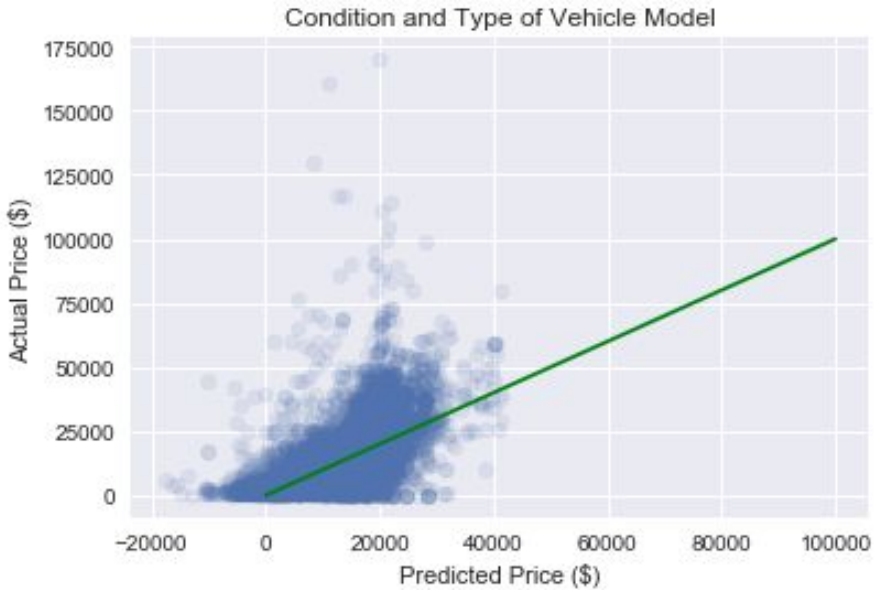


RMSE (test):
$10948

(Mean $: 12190)

When you don’t succeed, try, try again...

RMSE (test): $9223



RMSE (test): $8886

Mileage	Age	Condition	Type	Paint color	Brand

Let's look at some p-values...

const	2.345e+04	255.690	91.705	0.000	2.29e+04	2.39e+04
age	-229.1697	5.731	-39.986	0.000	-240.403	-217.936
mileage	-0.0622	0.001	-81.055	0.000	-0.064	-0.061
excellent	-1731.0408	120.371	-14.381	0.000	-1966.971	-1495.110
fair	-5632.5516	267.040	-21.093	0.000	-6155.957	-5109.146
good	-4492.2881	139.602	-32.179	0.000	-4765.911	-4218.665
like_new	1291.0620	179.208	7.204	0.000	939.810	1642.314
new	1.052e+04	558.725	18.821	0.000	9420.679	1.16e+04
SUV	1592.3041	138.560	11.492	0.000	1320.723	1863.885
bus	1.129e+04	1531.501	7.375	0.000	8292.799	1.43e+04
mini-van	26.3157	388.332	0.068	0.946	-734.824	787.456

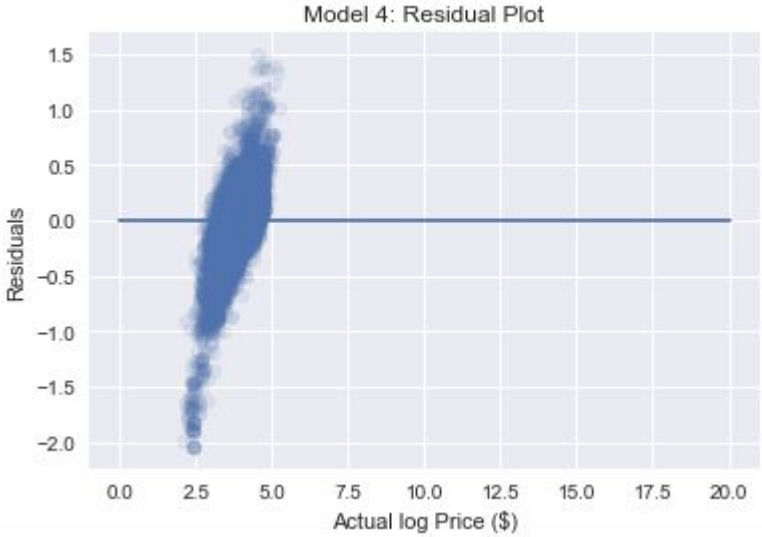
```
('age', -0.009816733299638369),
('mileage', -2.0708583557851332e-06),
```

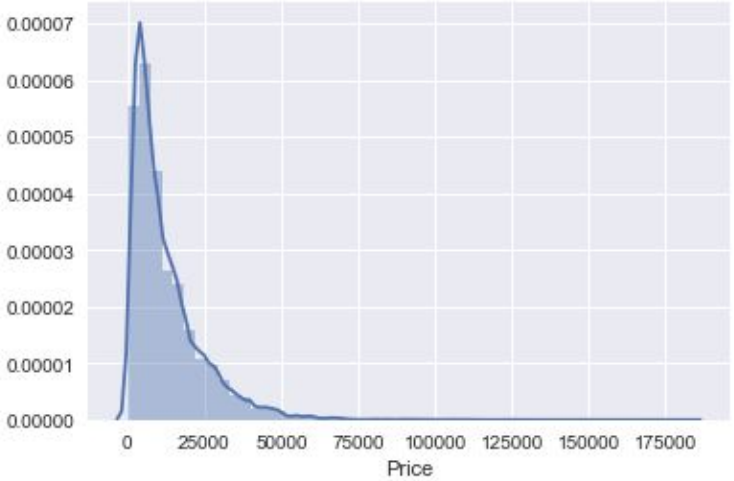
	age	mileage
age	1.000000	0.294187
mileage	0.294187	1.000000

Let's look at some p-values...

const	2.345e+04	255.690	91.705	0.000	2.29e+04	2.39e+04
age	-229.1697	5.731	-39.986	0.000	-240.403	-217.936
mileage	-0.0622	0.001	-81.055	0.000	-0.064	-0.061
excellent	-1731.0408	120.371	-14.381	0.000	-1966.971	-1495.110
fair	-5632.5516	267.040	-21.093	0.000	-6155.957	-5109.146
good	-4492.2881	139.602	-32.179	0.000	-4765.911	-4218.665
like_new	1291.0620	179.208	7.204	0.000	939.810	1642.314
new	1.052e+04	558.725	18.821	0.000	9420.679	1.16e+04
SUV	1592.3041	138.560	11.492	0.000	1320.723	1863.885
bus	1.129e+04	1531.501	7.375	0.000	8292.799	1.43e+04
mini-van	26.3157	388.332	0.068	0.946	-734.824	787.456

Omnibus:	29693.414	**Durbin-Watson:**	1.431
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1395603.745
Skew:	3.231	**Prob(JB):**	0.00
Kurtosis:	31.583	**Cond. No.**	1.18e+16

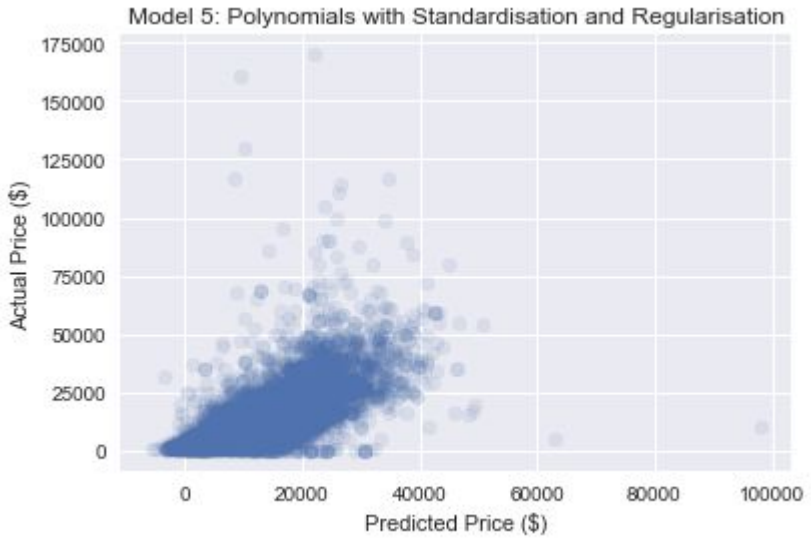




Still lots of heteroskedasticity even after transforming it

Missing Signal.

Ok, we need to add some complexity



RMSE: $8880

(Worse than the linear regression of order 2 with an RMSE of $8042)

Next Steps

- Focus on one dataset with more features before doing regularisation
- Then grid search across lambdas to find optimal regularisation
- Cross-validate models to make estimates of error more robust

Or, I could just go to Kelley's Blue Book