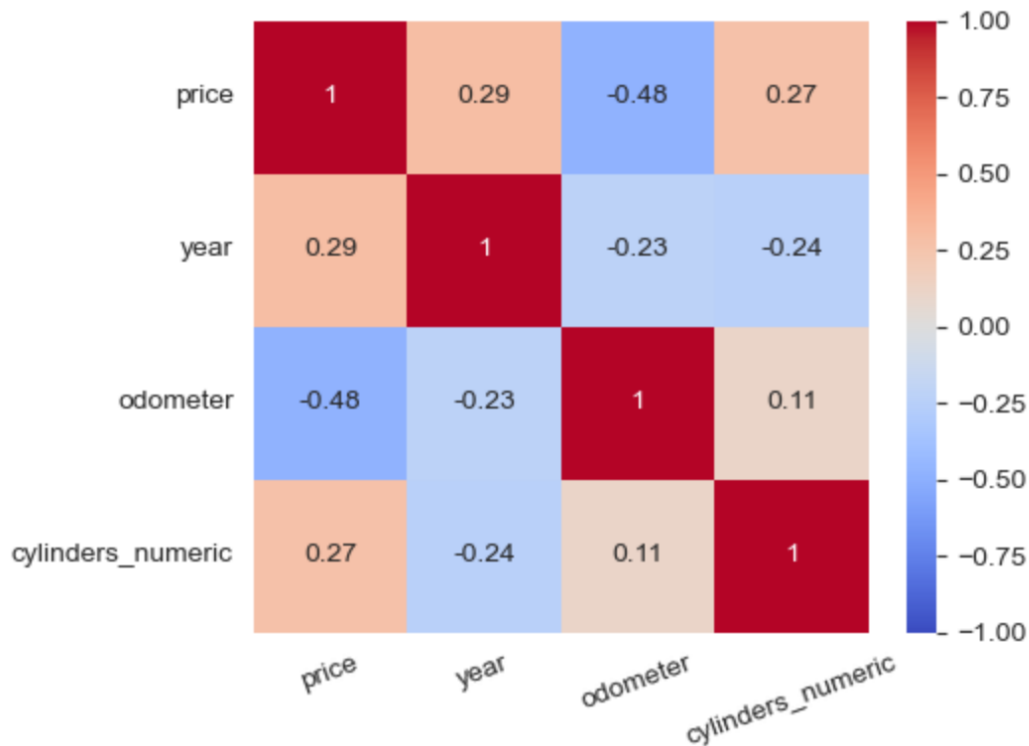


## Practical-Application-2

### What drives the price of a Car?

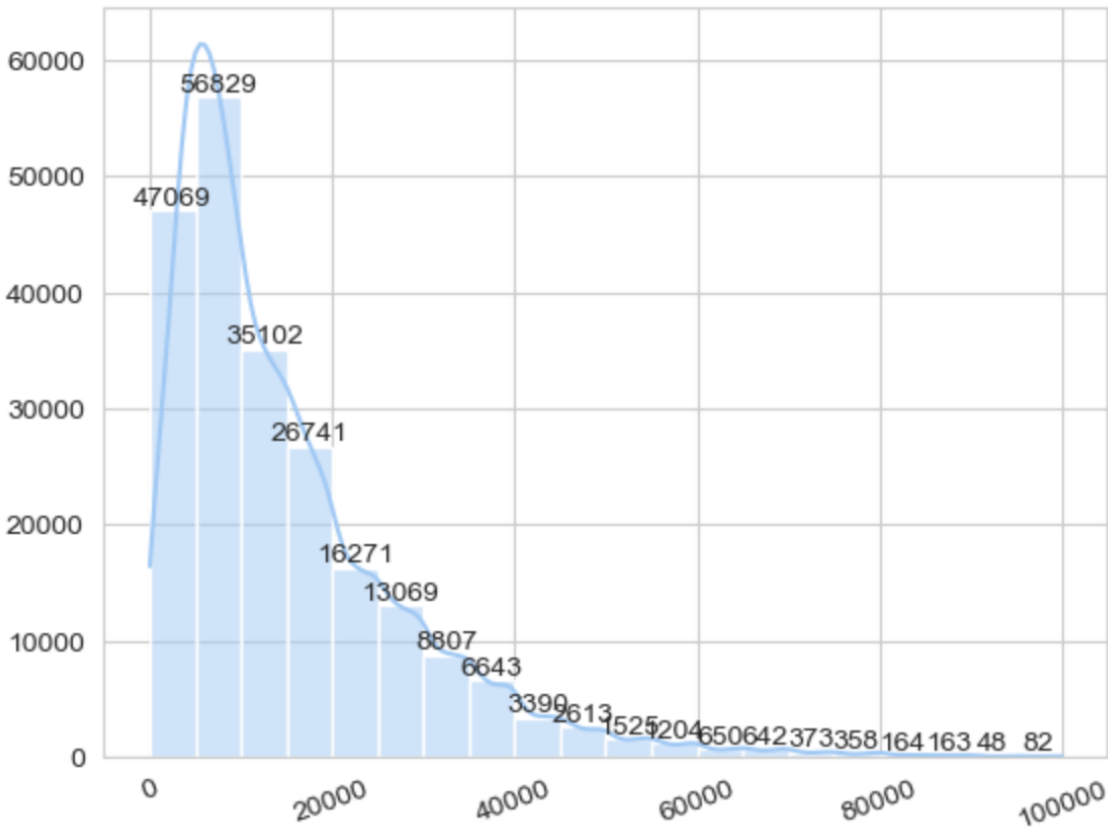
Harish Laxmi Narasimha Venugopal

#### Correlation of numerical features with price



- Price and Odometer are negatively correlated (-0.41). Higher the odometer reading, lower the price
- Price and years are positively correlated (0.29). Current or recent year cars are priced higher than old cars

## Distribution of price



Used car prices typically follow a right-skewed distribution, meaning most cars are priced on the lower end, while a few high-end models drive up the maximum price range. The median price is often much lower than the mean due to expensive outliers.

**Outliers** - Some listings have extremely low prices (e.g., \$0, \$1, or a few hundred dollars), which may be incorrect or refer to damaged/salvage vehicles or placeholder prices.

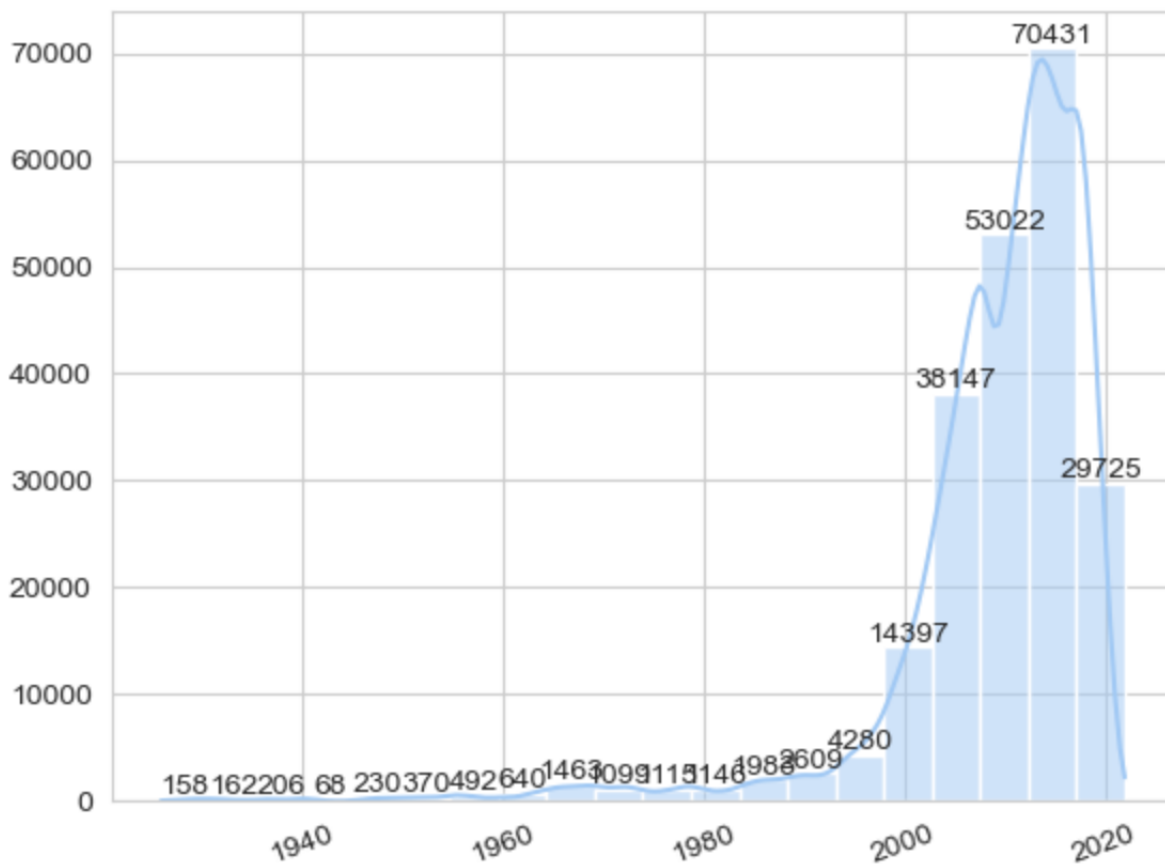
**Most competitive price range** - The majority of used cars fall between \$5,000 and \$25,000, suggesting that this is the most competitive price range.

Vehicles below \$5,000 are often older cars or high mileage, while those above \$25,000 are usually newer models or premium brands.

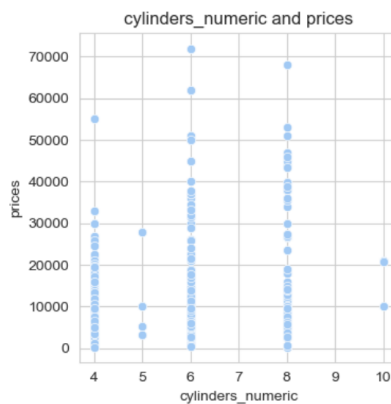
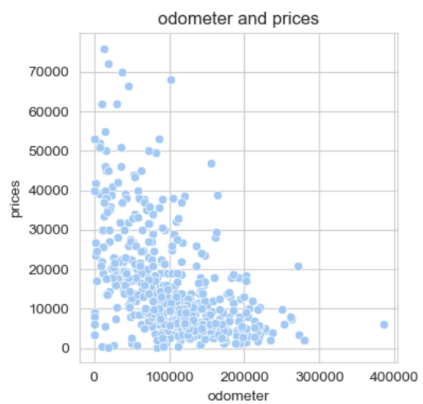
Newer vehicles (e.g., 2018 and later) tend to be priced significantly higher. Vehicles with low mileage (<50K miles) are priced higher, while those exceeding 150K miles see a steep decline in value.

On the higher end, luxury models or rare vehicles can significantly raise the maximum price to \$100,000+.

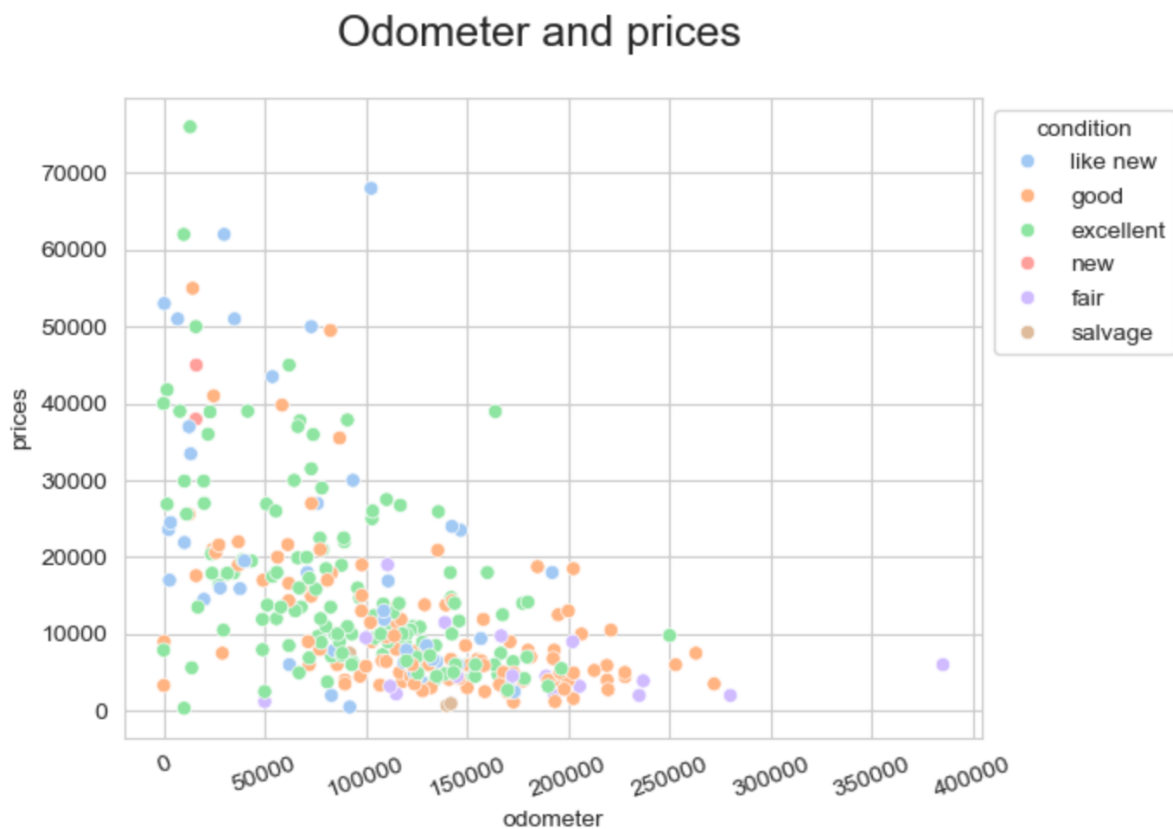
## Distribution of Year of the vehicle



- Year - Newer cars tend to be more expensive, with a steep price drop for cars over 10 years old.
- 

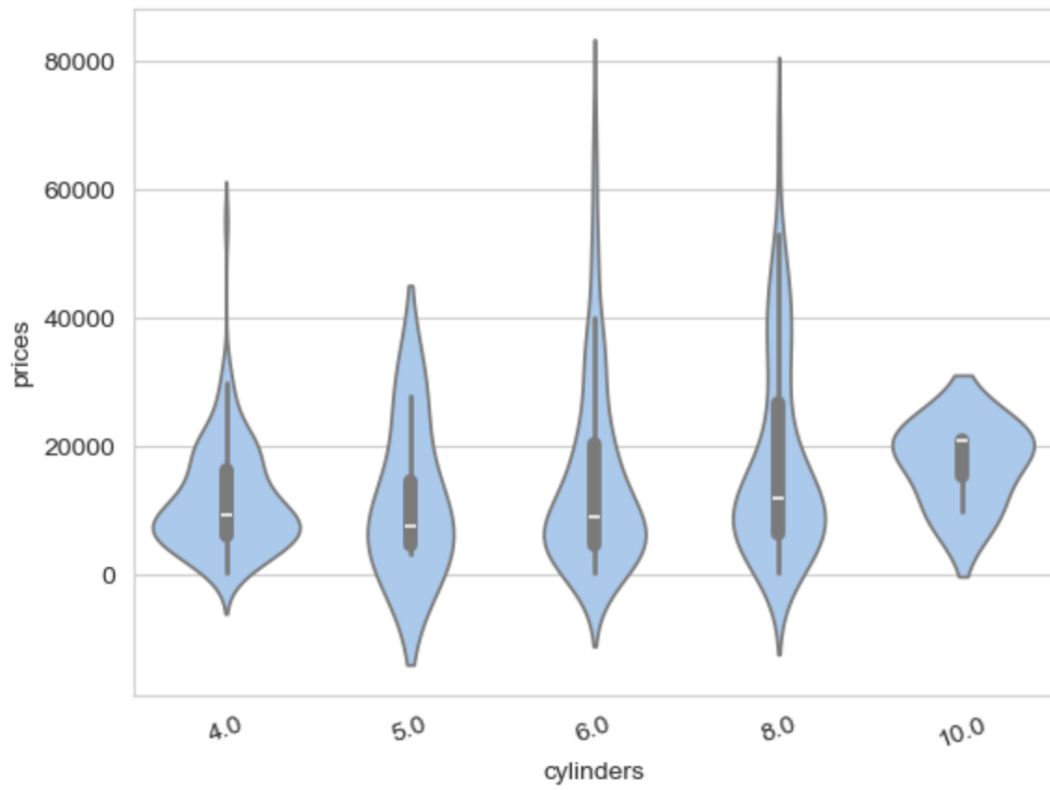


- Mileage (Odometer) - Higher mileage correlates with lower price, especially beyond 150K miles.
- Positive correlation between price and cylinders (0.25). Cars with higher number of cylinders attracts more price.
- Price and years are positively correlated (0.29). Current or recent year cars are priced higher than old cars



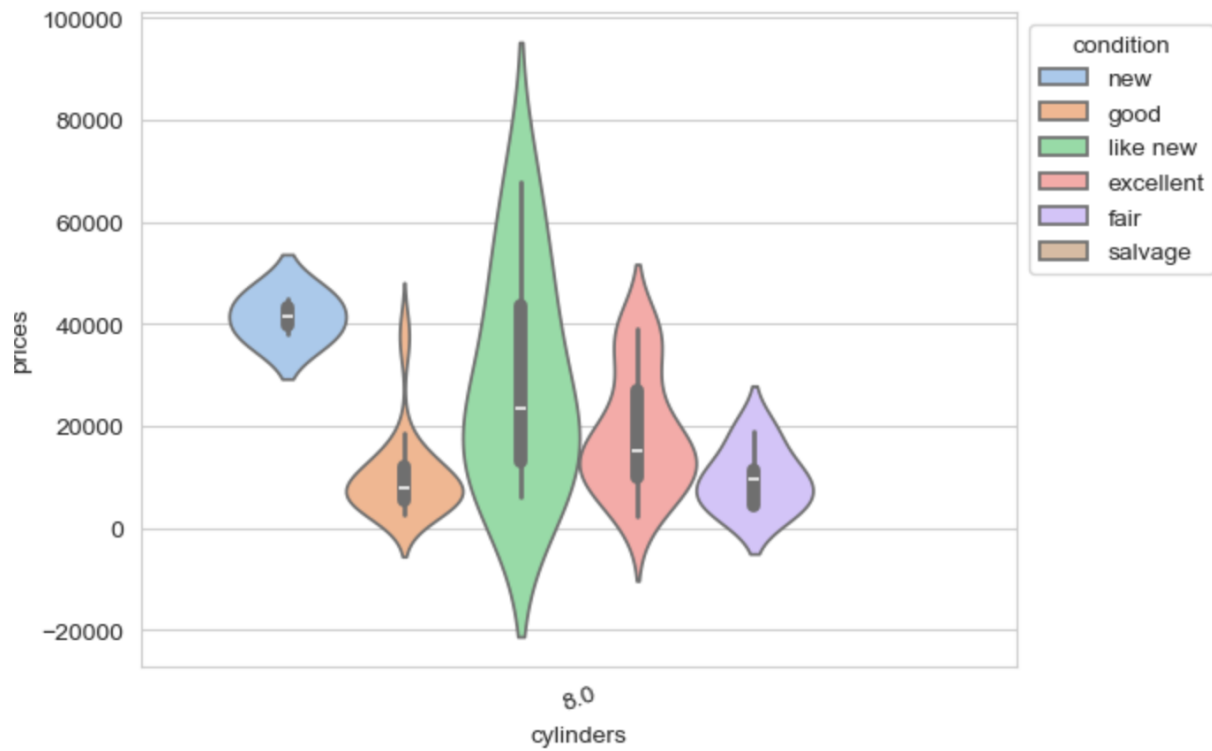
- Price and Odometer are negatively correlated (-0.41). Higher the odometer reading, lower the price

## Cylinders and prices



Cars with higher number of cylinders attracts more price.

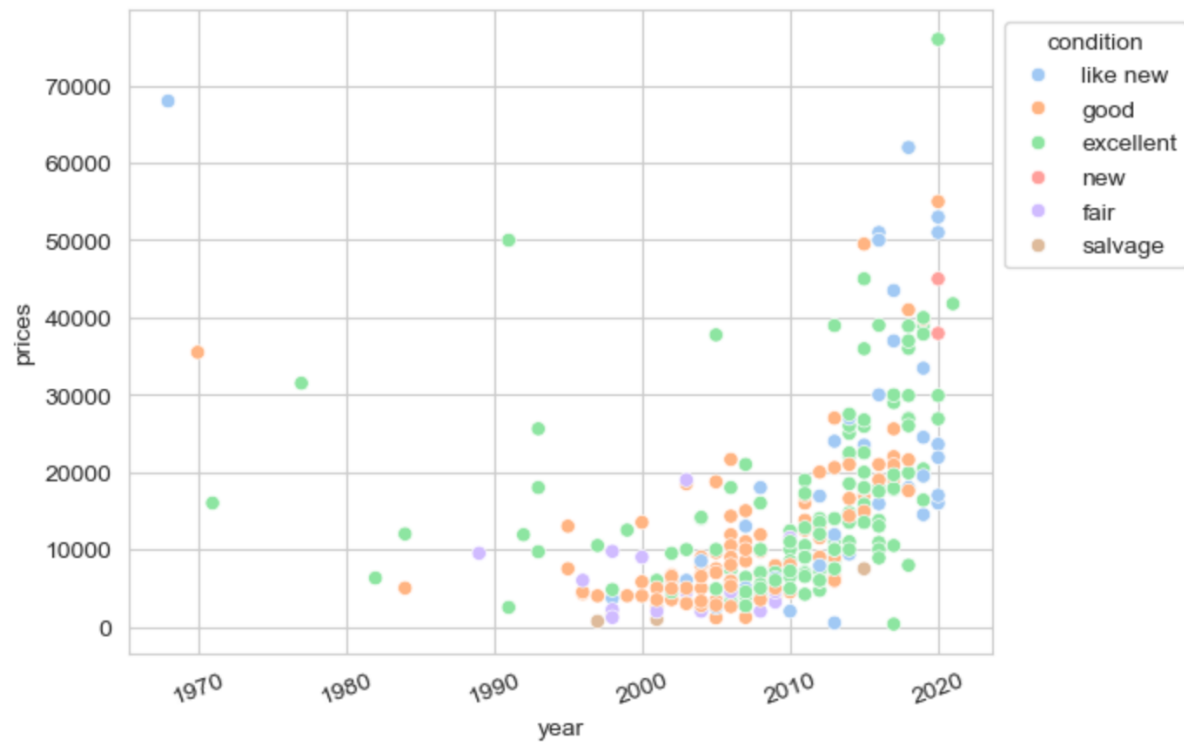
## Cylinders and prices



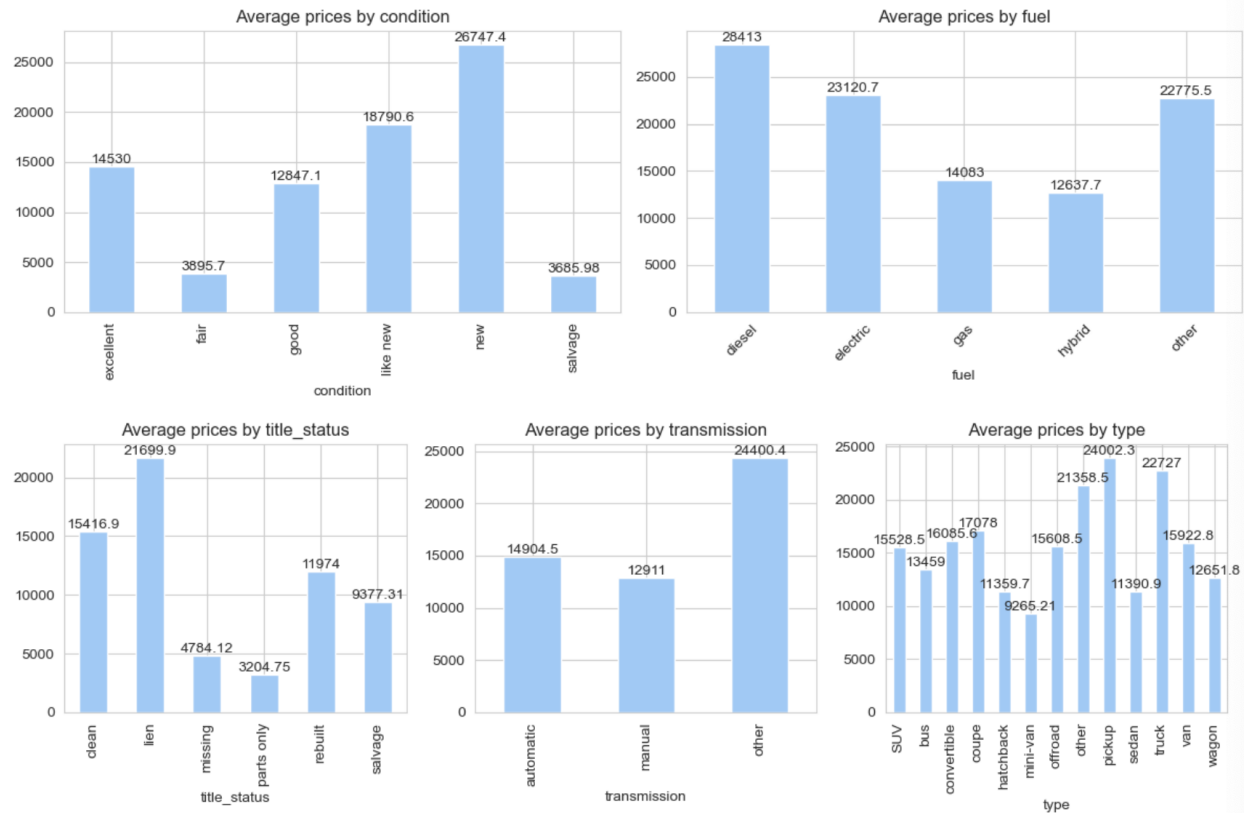
Cars with higher number of cylinders attracts more price.

Condition also plays a vital role in price. Like new condition attracts high price.

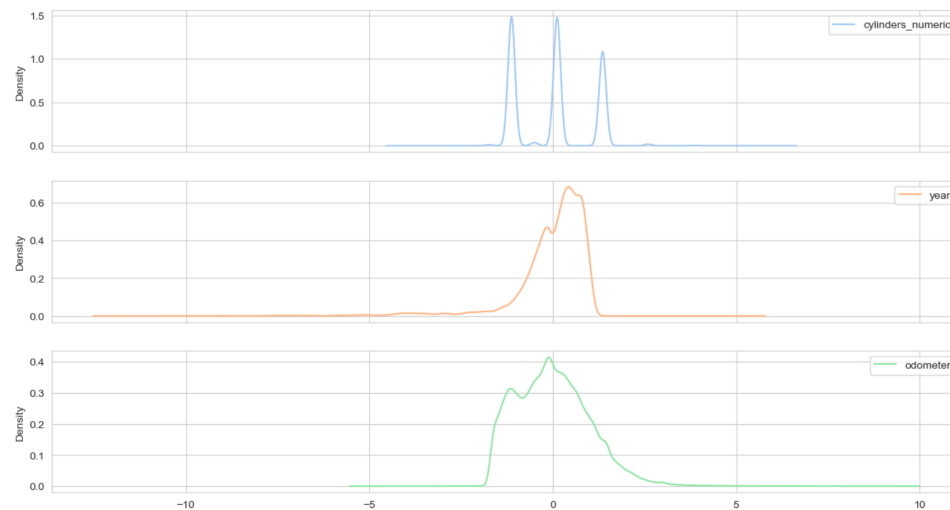
## Year and prices



- Year - Newer cars tend to be more expensive, with a steep price drop for cars over 10 years old.
-



- Condition - Excellent and Like New cars command high prices
- Pickups tend to attract higher prices, mini-vans on the other hand appear to sell for less
- Diesel vehicles tend to attract higher prices
- Vehicles with more cylinders are more expensive





## Model Evaluation

model index:	model	feature_type	number of features	degree	training rmse	testing rmse	r <sup>2</sup> test data	best_alpha	time to fit model (seconds):
1.11	linear	All Numeric	3	1	10663.441428	10686.773965	0.345645	n/a	0.055417
1.12	linear	All Numeric	9	2	9626.263470	9662.716134	0.465043	n/a	0.107707
1.13	linear	All Numeric	19	3	9330.447622	9364.728926	0.497530	n/a	0.221591
1.2	ridge	All Numeric	19	3	9330.447622	9364.728893	0.497530	{0.1}	4.576518
1.3	lasso	All Numeric	19	3	9330.449083	9364.741888	0.497528	{1}	17.845767
2.1	Linear	Numerical and Categorical	91	3	8111.820167	8145.970206	0.619806	n/a	2.244500
2.2	ridge	Numerical and Categorical	91	3	8111.881336	8145.243408	0.619873	{1}	25.411776
2.3	lasso	Numerical and Categorical	91	3	8111.870144	8145.716592	0.619829	{0.1}	59.240793
2.4	linear	Numerical and Categorical incl Ordinal	81	3	8126.398532	8160.704393	0.618429	n/a	2.009647

LinearRegression model was improved with additional of categorical variables. (Model performance 61%)

## Key features (Top 20) that drive price (+vely/-vely) of an used car

Top 20 features that have the most impact on increasing the price of the car

	Feature	Coefficient
42	categorical__manufacturer_ferrari	46358.079382
69	categorical__manufacturer_tesla	19872.511463
33	categorical__manufacturer_aston-martin	17214.310435
64	categorical__manufacturer_porsche	12665.923858
86	categorical__fuel_diesel	10951.901089
1	numerical__year	9320.512191
66	categorical__manufacturer_rover	5467.101258
26	categorical__title_status_lien	4757.751403
6	numerical__year^2	4337.250209
81	categorical__type_pickup	3902.236323
53	categorical__manufacturer_land rover	3603.552704
25	categorical__title_status_clean	3301.222495
79	categorical__type_offroad	3063.798483
83	categorical__type_truck	2729.844265
75	categorical__type_convertible	2450.978703
22	categorical__condition_like new	2218.290332
0	numerical__cylinders_numeric	1896.598739
4	numerical__cylinders_numeric year	1803.409778
19	categorical__condition_excellent	1797.318440
32	categorical__manufacturer_alfa-romeo	1621.263254

Top 20 features that have the most impact on lowering the price of the car

	Feature	Coefficient
43	categorical__manufacturer_fiat	-8741.033691
46	categorical__manufacturer_harley-davidson	-7605.084632
60	categorical__manufacturer_mitsubishi	-7130.076763
52	categorical__manufacturer_kia	-6801.465303
87	categorical__fuel_electric	-6195.529830
71	categorical__manufacturer_volkswagen	-6059.329126
48	categorical__manufacturer_hyundai	-5925.164431
62	categorical__manufacturer_nissan	-5653.252048
39	categorical__manufacturer_chrysler	-5421.488779
61	categorical__manufacturer_morgan	-5164.797878
41	categorical__manufacturer_dodge	-4913.065923
58	categorical__manufacturer_mercury	-4760.873854
28	categorical__title_status_parts only	-4374.862319
36	categorical__manufacturer_buick	-4190.583779
56	categorical__manufacturer_mazda	-4178.105706
74	categorical__type_bus	-4118.836605
67	categorical__manufacturer_saturn	-4104.480025
59	categorical__manufacturer_mini	-3836.510791
27	categorical__title_status_missing	-3538.711851
77	categorical__type_hatchback	-3299.327447

- Key features that drive price (+vely/-vely) of an used car are manufacturer, fuel type, title\_status, type, condition. These features make car more or less expensive in the used car market
- Positive price impact Manufacturer (Ferrari, Tesla, Ashton Martin, Porsche), fuel type (Diesel), title (Lien), type (pickup), condition (like new)