

# Used car pricing

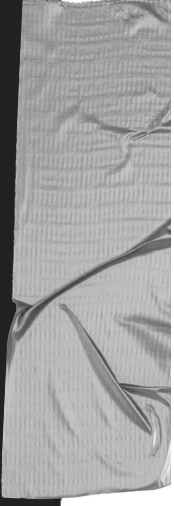
Anh Luong

# Problem

This project analyses the 2020 used car data to identify what features have a strong relationship with resale price of used car.

The analysis also looks to determine if the model can be used to accurately predict the resale price.

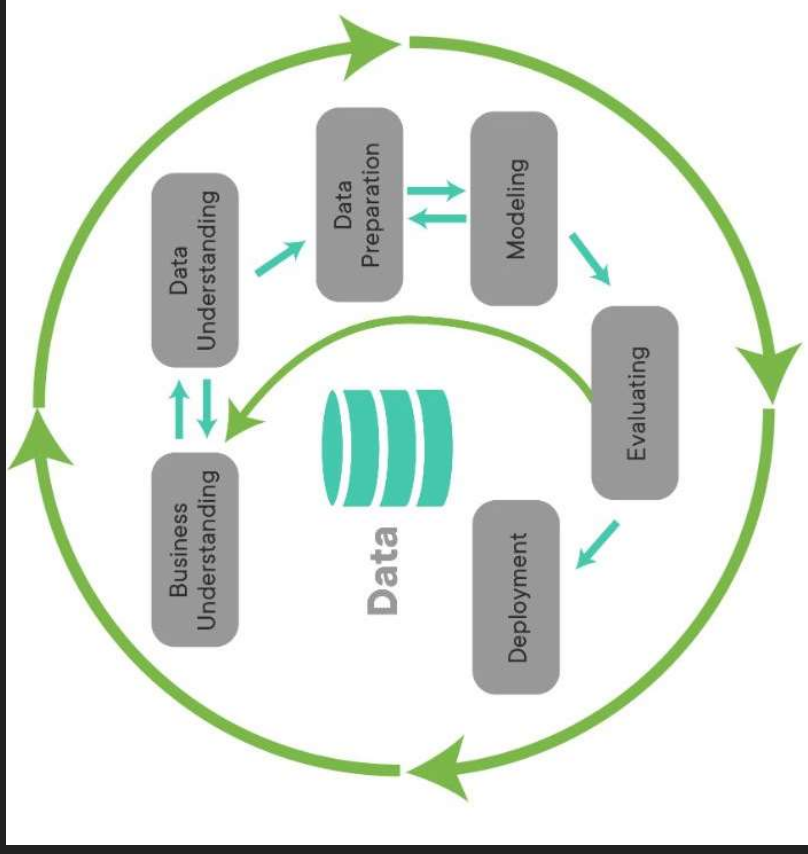




## What did we look at?

- 2020 used car sales data
- Multiple features
  - Age
  - Current price in the show room
  - Kms driven
  - Fuel type
  - Seller type
  - Transmission
  - No. of owners

# How did we do it?



# Results:

Significant features:

- Age
- Kms driven
- Seller that is individual and
- Number of owners
- Current price for the same car in the show room

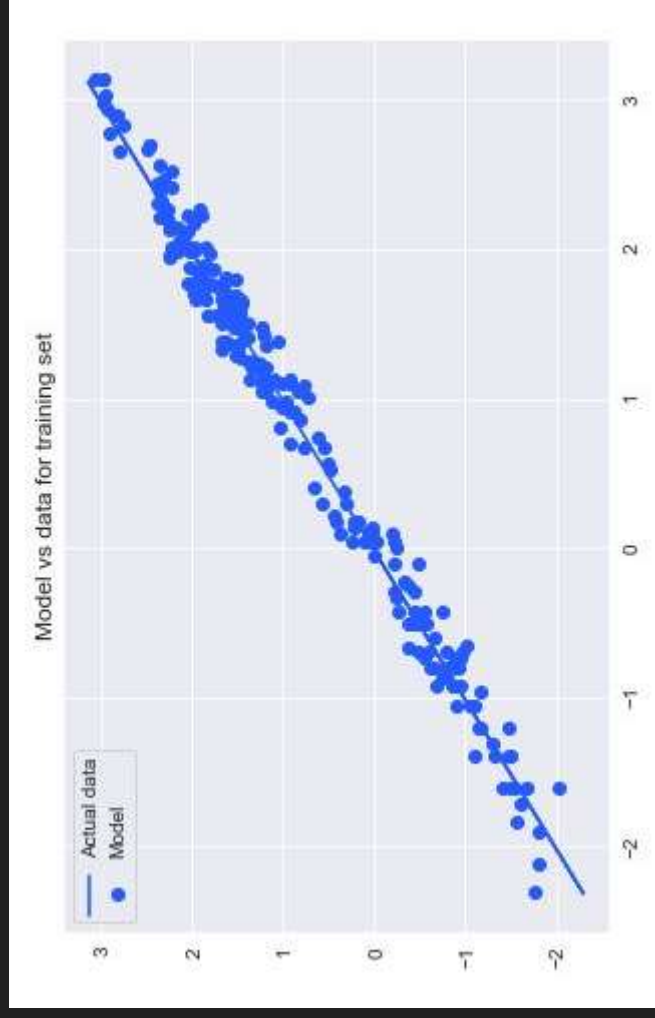
Model accuracy:

The model R square is 0.973 - this mean that 97.3% of the resale price of an used car can be explained by the dependent variables in the model.

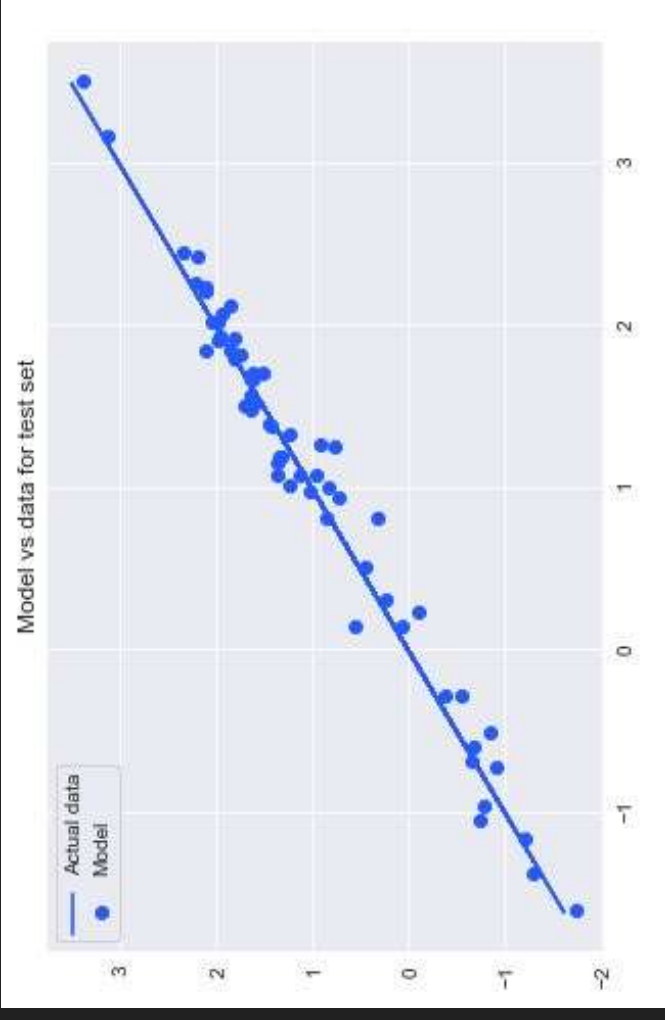
## OLS Regression Results

Dep. Variable:	Selling_Price_log	R-squared:	0.974			
Model:	OLS	Adj. R-squared:	0.973			
Method:	Least Squares	F-statistic:	1824.			
Date:	Sun, 26 Feb 2023	Prob (F-statistic):	3.73e-228			
Time:	16:45:29	Log-Likelihood:	53.515			
No. Observations:	299	AIC:	-93.03			
Df Residuals:	292	BIC:	-67.13			
Df Model:	6					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	1.2679	0.141	8.990	0.000	0.990	1.545
Age_log	-0.7096	0.037	-19.286	0.000	-0.782	-0.637
Kms_Driven_log	-0.0369	0.017	-2.220	0.027	-0.070	-0.004
Present_Price_log	0.9365	0.021	43.620	0.000	0.894	0.979
Seller_Type_Individual	-0.1941	0.053	-3.649	0.000	-0.299	-0.089
Owner_1	-0.1697	0.067	-2.540	0.012	-0.301	-0.038
Owner_3	-0.8130	0.218	-3.730	0.000	-1.242	-0.384

# Model evaluation:



Train Mean Squared Error: 0.025801024134270403  
Test Mean Squared Error: 0.03341965250967328



# Conclusion:

Good resale price is driven by

- Age
- Kms driven
- Seller that is individual and
- Number of owners
- Present Price

The model can be used to predict the price highly accurately.



**Thank you**