Linear Regression Project MVP

1- Using sklearn to fit the training data we got R2 of 0.67

In [115]:
```
#seperating the target variable in a different datafram and spliting the dataset to 60% train, 20% validation, 20% testing
X = cars2.drop('price', axis = 1)
y = cars2['price']

X, X_test, y, y_test = train_test_split(X, y, test_size = .2, random_state = 10)

X_train, X_val, y_train, y_val = train_test_split(X, y, test_size = .25, random_state = 10 )
```

In [135]:
```
lm = LinearRegression()
lm.fit(X_train, y_train)
lm.score(X_train, y_train)
```

Out[135]: 0.6715071486659262

2- Using statsmodel (OLS) to fit the training data we got R2 of 0.91

In [134]:
```
cars_model = sm.OLS(y_train, X_train, data = cars2 )

results = cars_model.fit()

results.summary()
```

Out[134]:

OLS Regression Results

Dep. Variable:	price	**R-squared (uncentered):**	0.915
Model:	OLS	**Adj. R-squared (uncentered):**	0.915
Method:	Least Squares	**F-statistic:**	7.298e+04
Date:	Fri, 24 Sep 2021	**Prob (F-statistic):**	0.00
Time:	18:20:39	**Log-Likelihood:**	-3.4561e+06
No. Observations:	332889	**AIC:**	6.912e+06
Df Residuals:	332840	**BIC:**	6.913e+06
Df Model:	49		
Covariance Type:	nonrobust		