“Multiple Linear Regression Applications to Predict the Used Car Pricing”

Descriptive Statistics (all variables)

	N	Minimum	Maximum	Mean	Std. Deviation
Price	1436	4350.0	32500.0	10730.825	3626.9646
Age	1436	1.0	80.0	55.947	18.6000
KM	1436	1.0	243000.0	68533.260	37506.4489
HP	1436	69.0	192.0	101.502	14.9811
cc	1436	1300.0	16000.0	1576.856	424.3868
Doors	1436	2.0	5.0	4.033	.9527
Quarterly_Tax	1436	19.0	283.0	87.123	41.1286
Weight	1436	1000.0	1615.0	1072.460	52.6411
Valid N (listwise)	1436				

Descriptive Statistics :

Variable

Price in $	
Mean	10730.82451
Standard Error	95.71193469
Median	9900
Mode	8950
Standard Deviation	3626.964585
Sample Variance	13154872.1
Kurtosis	3.73778119
Skewness	1.70388498
Range	28150
Minimum	4350
Maximum	32500
Sum	15409464
Count	1436

Basic data about the response variable (the price of these cars).

Specifically, Mean=10730.82451, Median=9900, Standard deviation=3626.964585.

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Weight, KM, Doors, HP, cc, Age, Quarterly_Tax[b]	.	Enter

a. Dependent Variable: Price

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.929[a]	.863	.862	1346.2723

a. Predictors: (Constant), Weight, KM, Doors, HP, cc, Age, Quarterly_Tax

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	16289064300.951	7	2327009185.850	1283.903	.000[b]
	Residual	2588177162.826	1428	1812448.994		
	Total	18877241463.777	1435			

a. Dependent Variable: Price

b. Predictors: (Constant), Weight, KM, Doors, HP, cc, Age, Quarterly_Tax

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2750.555	1059.491		-2.596	.010
	Age	-121.793	2.623	-.625	-46.439	.000
	KM	-.021	.001	-.212	-16.395	.000
	HP	33.731	2.743	.139	12.297	.000
	cc	-.125	.090	-.015	-1.385	.166
	Doors	-24.534	39.389	-.006	-.623	.533
	Quarterly_Tax	4.123	1.312	.047	3.141	.002
	Weight	16.985	1.071	.247	15.860	.000

a. Dependent Variable: Price

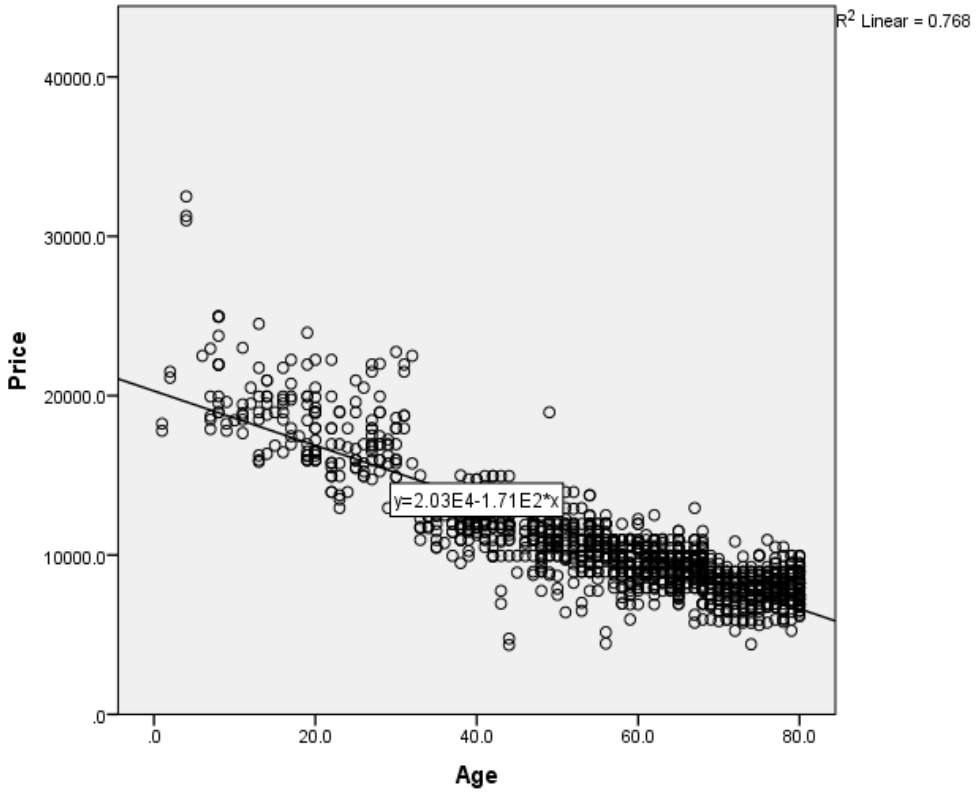
Based on all the information and outputs above, my initial model is:

Price= β 0AGE +β 1KM +β 2HP +β 3CC +β 4DOORS +β 5Quaterly_Tax +β 6WEIGHT+ constant .

The adjusted R2 is .862, which indicates this model fits the data well. The F-test (ANOVA) tests whether β0 =β1 =β2 = β3 = β4 = β5 = β6 =0. The p-Value is < 0.0001, indicating that overall the explanatory variables are significant to the response variable.

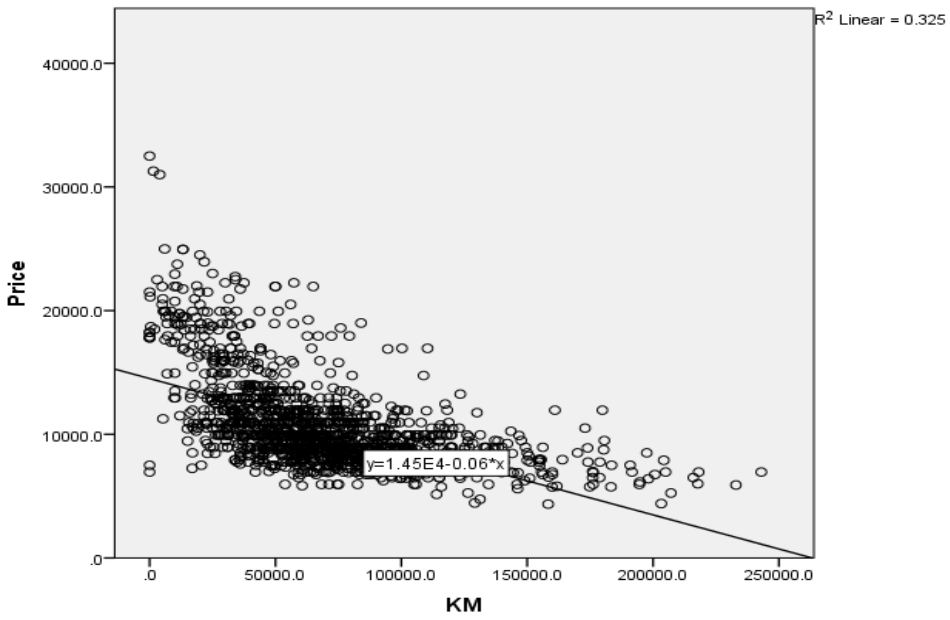
Coefficients and standard errors, along with t-tests and Sig are given in the above **Coefficients table**. The t-test tests whether each explanatory variable is zero. Hence, only Age, KM, HP, Quarterly-Tax and Weight are significant because of Sig<0.01. So we remove CC and DOORS from our model.

Graph : Scatterplot between Price of car and Age of car



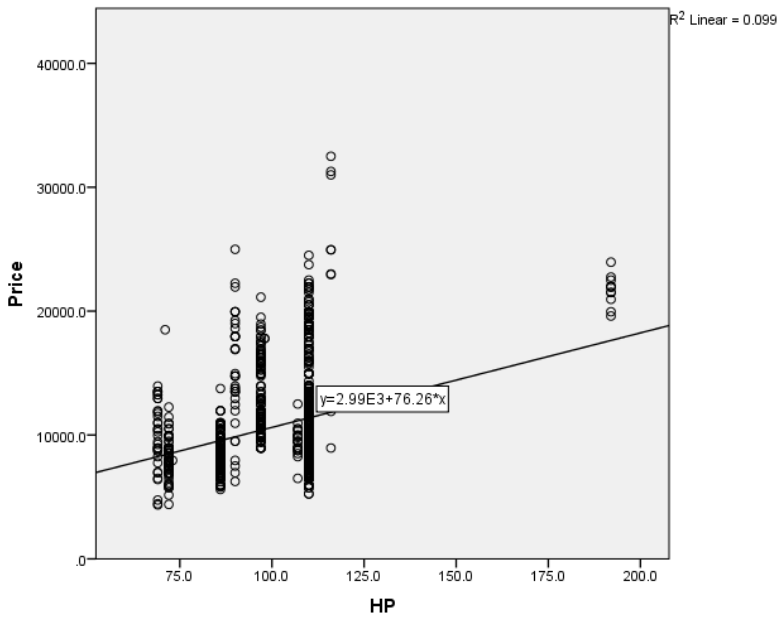
From the scatterplots, we can see that the linear relation is negative between price and age. The normal thinking is that ”the price will go down as the age goes up”. That is to say, the coefficient between price and age is negative.

Graph : Scatterplot between Price of car and KM driven



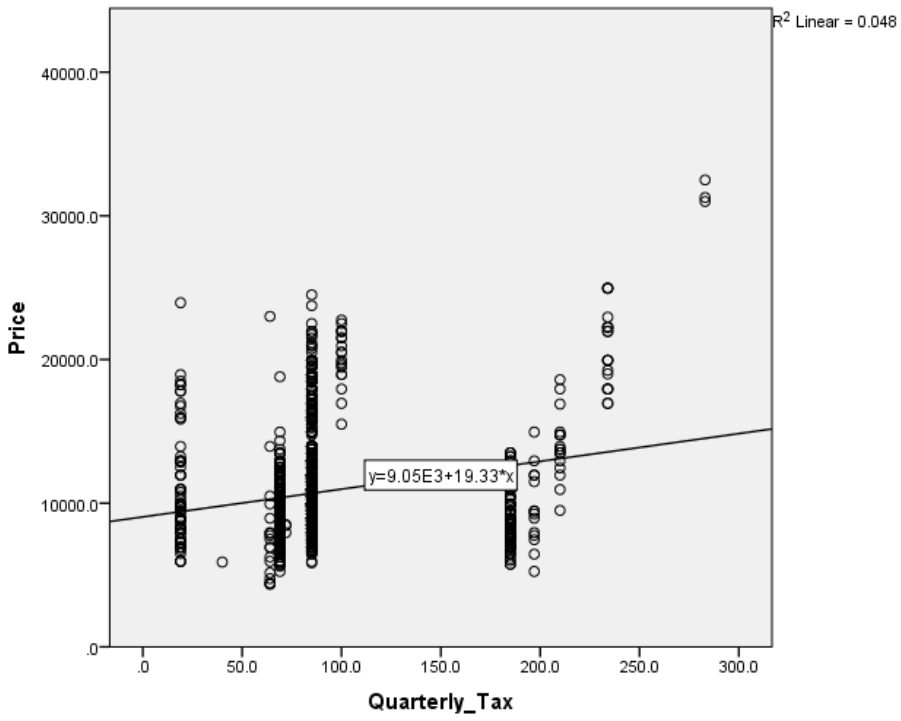
From the scatterplots, we can see that the linear relation is negative between price and km. The normal thinking is that “the price will go down as the km goes up”. That is to say, the coefficient between price and km is negative.

Graph : Scatterplot between Price of car and KM driven\



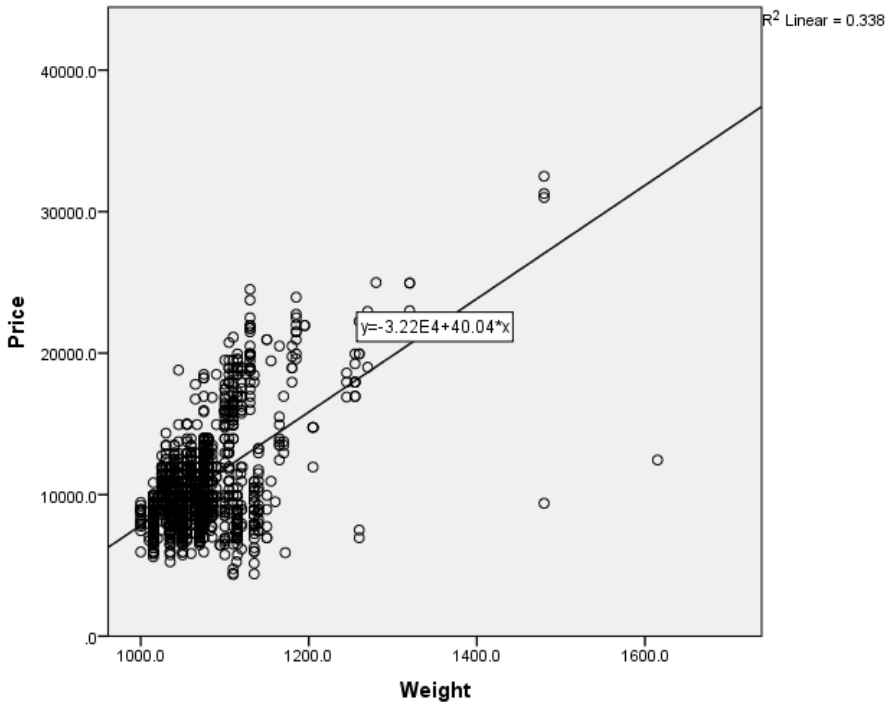
From the scatterplots, we can see that the linear relation is positive between price and HP. The normal thinking is that “the price will go up as the HP goes up”. That is to say, the coefficient between price and HP is positive.

Graph : Scatterplot between Price of car and KM driven



From the scatterplots, we can see that the linear relation is positive between price and Quarterly tax. The normal thinking is that “the price will go up as the Quarterly tax of the car goes up”. That is to say, the coefficient between price and Quarterly tax is positive.

Graph : Scatterplot between Price of car and KM driven



From the scatterplots, we can see that the linear relation is positive between price and weight. The normal thinking is that “the price will go up as the weight of the car goes up”. That is to say, the coefficient between price and weight is positive.

Multiple Regression Analysis

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.929[a]	.863	.862	1346.4061

a. Predictors: (Constant), Weight, KM, HP, Age, Quarterly_Tax

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	16284924180.964	5	3256984836.193	1796.651	.000[b]
	Residual	2592317282.813	1430	1812809.289		
	Total	18877241463.777	1435			

a. Dependent Variable: Price

b. Predictors: (Constant), Weight, KM, HP, Age, Quarterly_Tax

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2554.436	1048.246		-2.437	.015
	Age	-121.837	2.623	-.625	-46.454	.000
	KM	-.021	.001	-.214	-16.543	.000
	HP	33.341	2.731	.138	12.209	.000
	Quarterly_Tax	3.972	1.298	.045	3.060	.002
	Weight	16.586	1.022	.241	16.224	.000

a. Dependent Variable: Price

the p-value<0.01, it means that these variables are significant to the response variable. At the same time, the Adjusted R-Sq=0.862, which means that these variables can explain 86.20% of the alterations in price of the cars. Finally, we can get the equation:

Price= - β 0AGE - β 1KM +β 2HP +β 5Quaterly_Tax +β 6WEIGHT- constant.

Price= - (121.837) AGE – (.021) KM +(33.341) HP +(3.972) Quarterly +(16.586)WEIGHT- 2554.436 .

Thus, the above equation is our final equation of the model. We can use this equation to find the actual price of the used cars.

CONCLUSION

The purpose of this research is to investigate what variables impacted automobile prices. We started with a preliminary investigation and selected seven factors to forecast the price. Then we utilised the multivariate linear regression approach to look at how these characteristics impact automobile prices. After the analysis, we chose 5variables (Age, KM, HP, Quarterly-Tax and Weight) to include in our model. Only the car's age, kilometres, horsepower, quarterly tax, and weight are important, according to our findings because of Sig<0.01. So, we removed CC and DOORS from our model. Finally, after again running the analysis we got the equation that is **Price= - (121.837) AGE – (.021) KM +(33.341) HP +(3.972) Quarterly +(16.586) WEIGHT- 2554.436 .**

FINDINGS

From the above analysis we can say that the coefficient of HP, CC and qtr_tax is positive and the coefficient of age and km is negative. At the same time, we can observe that the age coefficient has a large absolute value whereas the absolute value of the other coefficients is tiny, which means that age may have bigger influence on the price than others and the nature of the effect is negative. Once we have this question, I think age is the most useful index to estimate the price of vehicle. At the same time, these indexes have the most important influence on the price. When we are going to buy a car, we can put the information of Age, KM, HP, Quarterly-Tax and Weight of the car into this regression model, and then we can get an estimate price to compare with the real price. Then we can decide whether to buy or not

RECOMMENDATION

When we decide to buy a car, we should think about other aspects that impact the price, such as the season, if it has been in an accident, or the repair times, and so on. All of these factors will have a significant impact on the pricing. Many additional elements should be considered if we want a more exact model. It may take a long time to gather data and select the optimal regression model, but the end result will be excellent.