

Background

The report conducted on behalf of Car4all, provides a comprehensive analysis of the pre-owned automobile market in the United Kingdom. The main goal of this analysis is to determine the crucial factors that substantially impact the pricing of used cars. Four characteristics encompassed are mileage, engine size, color, and the count of prior owners. This report aims to identify the key variables that significantly influence the value of a used automobile and develop a model to accurately predict the price of a used car based on these four qualities above.

Methodology

This analysis uses a dataset consisting of 100 used BMW 1 Series cars from the population of about 600 cars obtained from the website <https://www.autotrader.co.uk/>. The cars were selected from a 50-mile radius around the B25RS postcode. For many reasons, the dataset could be considered representative of the population. First, 100 vehicles are adequate to effectively detect existing patterns and create reliable population estimates. In addition, having a variety of automobile ages, conditions, and prices enables a complete market study. The geographical area is limited to a 50-mile radius around a single postcode, which could include urban and suburban areas to fully represent the market.

Limitation

The data collected might be affected by the partnership or advertising made by the sellers on the website which might reduce the randomness of the sample. In addition, the data might be influenced by a short-term trend since it is collected at only one particular point in time.

Graphical Visualisation

1. Histogram: Distribution of Price

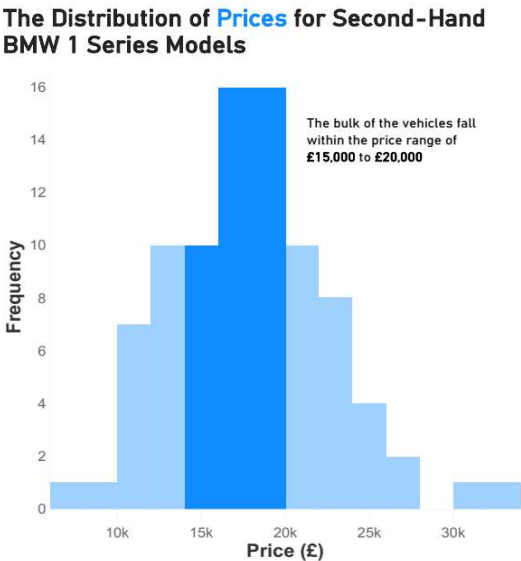


Figure 1: Distribution of Prices (Output: Power BI)

From the visualization, we can see the distribution of Prices for second-hand BMW 1 Series Models.

The graph makes use of the proximity principle by clustering cars with comparable prices into bars that show how often each range appears. In addition, by using a descriptive title and providing an immediate comment on the most important data item, the histogram complies with the "Say" principle of IBCS Standard

2. Bar Chart: Most Searched

Second-hand Car Makes

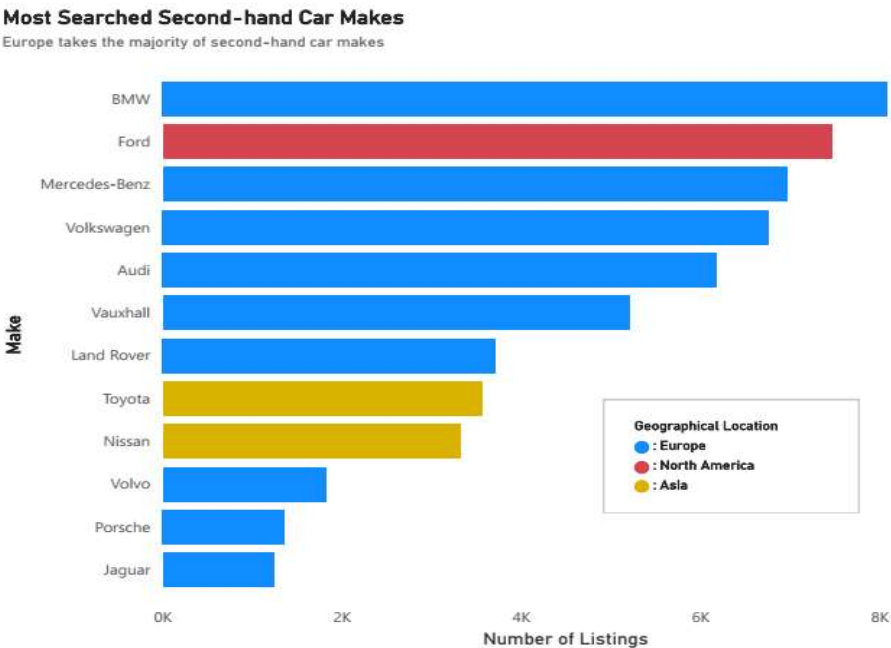


Figure 2: Most Searched Second-Hand Car Makes (Output: Power BI)

The graph illustrates that the majority of second-hand automobile listings are of cars manufactured in Europe, with BMW being the most popular among the various car brands.

In this graph, the principle of similarity is implemented by consistently utilizing color to symbolize distinct geographical areas of vehicle manufacturers. Viewers can rapidly categorize car makes by continent of origin as a result of this color labeling.

3. Scatter Plot: Prices vs Mileage by Number of Previous Owners

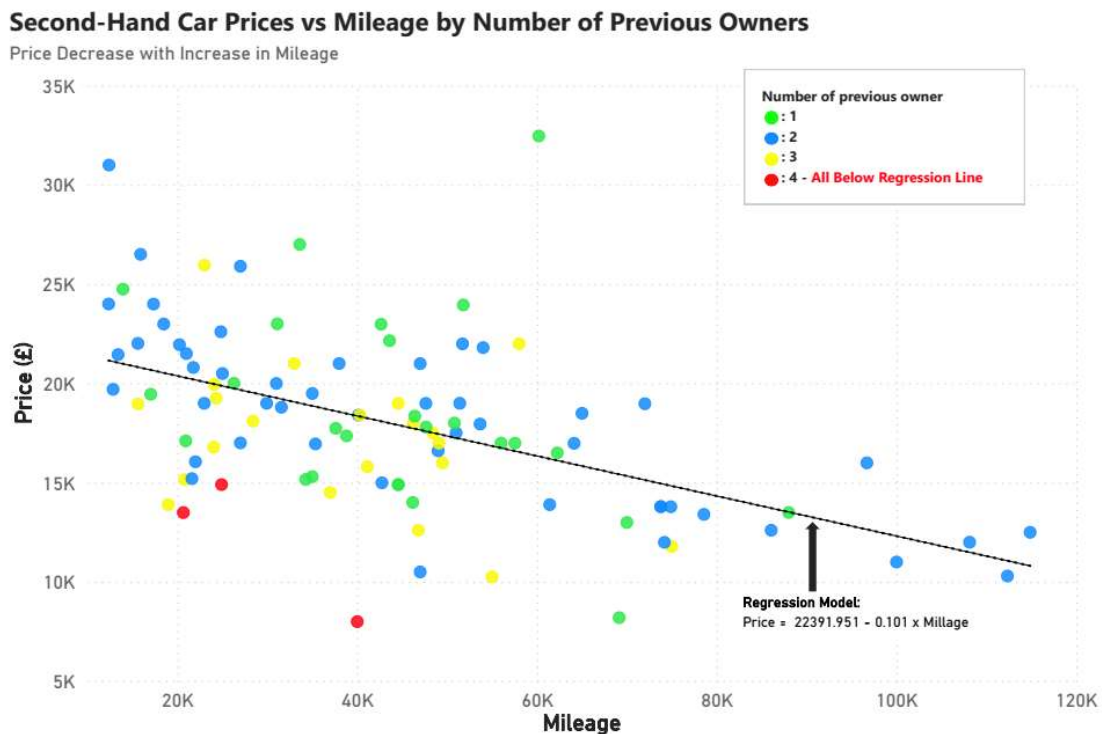


Figure 3: Relationship between Prices, Millage and Number of Previous Owners

It is evident from the graph that cars with greater costs in the dataset typically have lesser mileage, irrespective of the number of previous owners.

The graph adhered to the Gestalt principle of likeness, which states that objects are regarded as belonging to a group based on their shared characteristics. According to Tufte's Graphical Excellence, the chart also employs a precise regression line to portray quantitative information effectively.

Statistical Analysis

In this section, we will focus on the statistics of the dataset, and from that find out the characteristics of our sample, as well as come up with a model to predict the price of a used car. The tool that will be used for this task is a statistical software called SPSS

1. Summary Statistics Table

First, let's look at the information regarding the statistics of our data:

Descriptive Statistics							
	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Price	100	24468	7995	32463	17952.21	4513.467	20371388.774
Mileage	100	102519	12341	114860	44020.57	23400.600	547588090.13
Engine size	100	1.5	1.5	3.0	1.685	.3230	.104
Number of owners	100	3	1	4	2.01	.772	.596
Valid N (listwise)	100						

The following table provides an overview of essential details regarding our existing stock of pre-owned vehicles. The average price of these cars is approximately £17,952. The vehicles have accumulated mileage with an average of 44,021 miles. The engine capacities vary between 1.5 and 3.0 liters. Finally, the range of previous owners varies from 1 to 4, with an average of 2 owners per automobile.

Price Statistics by Vehicle Colour

Descriptive Statistics ^a							
	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Price	35	23005	7995	31000	17872.31	4917.633	24183109.928
Valid N (listwise)	35						

a. Colour = Black

Descriptive Statistics ^a							
	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Price	19	15605	10295	25900	18276.47	4079.494	16642271.152
Valid N (listwise)	19						

a. Colour = Blue

Descriptive Statistics ^a							
	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Price	18	17563	14900	32463	20079.56	4429.318	19618861.556
Valid N (listwise)	18						

a. Colour = Grey

Descriptive Statistics ^a							
	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Price	1	0	12495	12495	12495.00	.	.
Valid N (listwise)	1						

a. Colour = Red

Descriptive Statistics ^a							
	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Price	27	17755	8195	25950	16611.48	3919.690	15363967.721
Valid N (listwise)	27						

a. Colour = White

The analysis reveals that grey automobiles have the highest average price, roughly £20,080, followed by Blue vehicles with an average price of around £18,276. The highest degree of price variation is found among Black automobiles, with a price variability of £4,917.63. Grey vehicles have the highest maximum price, reaching £32,463, while the lowest price is observed in white vehicles, which cost £8,195.

2. Confidence Interval

This part will be used to determine the statistical characteristics of the population of second-hand cars based on the sample that we have extracted from that particular population. This can tell us how well our sample is used to describe our population as a whole.

95% Confidence Interval Of The Average Second-hand Car Price

			Statistic	Std. Error
Price	Mean		17952.21	451.347
	95% Confidence Interval for Mean	Lower Bound	17056.64	
		Upper Bound	18847.78	
	5% Trimmed Mean		17834.16	
	Median		17874.50	
	Variance		20371388.774	
	Std. Deviation		4513.467	
	Minimum		7995	
	Maximum		32463	
	Range		24468	
	Interquartile Range		6046	
	Skewness		.446	.241
	Kurtosis		.692	.478

According to our sample data, we are 95% sure that the average price of second-hand in the market falls within the range of £17,056 and £18,848.

99% Confidence Interval of The Average Second-hand Car Price

			Statistic	Std. Error
Price	Mean		17952.21	451.347
	99% Confidence Interval for Mean	Lower Bound	16766.79	
		Upper Bound	19137.63	
	5% Trimmed Mean		17834.16	
	Median		17874.50	
	Variance		20371388.774	
	Std. Deviation		4513.467	
	Minimum		7995	
	Maximum		32463	
	Range		24468	
	Interquartile Range		6046	
	Skewness		.446	.241
	Kurtosis		.692	.478

This range widens to around £16,767 and £19,138 when we want to have a chance of 99% that our average price of used car is within the above range.

3. One Sample T-test (Hypothesis Test)

The primary objective of this analysis is to ascertain whether a substantial disparity exists between the mean price of pre-owned vehicles as recorded in our dataset and the real mean price of a used automobile. The average cost of a used

automobile is derived from: Recent used car price information is available at Auto Trader Insight (autotraderinsight-blog.co.uk).

One-Sample Statistics						
	N	Mean	Std. Deviation	Std. Error Mean		
Price	100	17952.21	4513.467	451.347		

One-Sample Test						
Test Value = 17641						
	t	df	Significance		Mean Difference	95% Confidence Interval of the Difference
			One-Sided p	Two-Sided p		Lower Upper
Price	.690	99	.246	.492	311.210	-584.36 1206.78

One-Sample Effect Sizes					
	Standardizer ^a	Point Estimate	95% Confidence Interval		
			Lower	Upper	
Price	Cohen's d	4513.467	.069	-.127	.265
	Hedges' correction	4548.024	.068	-.126	.263

a. The denominator used in estimating the effect sizes.
 Cohen's d uses the sample standard deviation.
 Hedges' correction uses the sample standard deviation, plus a correction factor.

It is evident from this result that there is no statistically significant difference between the mean price of used cars in the sample and the aggregate mean price of used cars in the UK market.

4. Correlation Analysis

In this section, we will determine which variables are highly related to the price of a used car.

		Correlations							
		Price	Mileage	Engine size	Number of owners	Colour=Black	Colour=Blue	Colour=Grey	Colour=White
Price	Pearson Correlation	1	-.523**	.424**	-.188	-.013	.035	.222*	-.182
	Sig. (2-tailed)		<.001	<.001	.061	.897	.730	.026	.071
	N	100	100	100	100	100	100	100	100
Mileage	Pearson Correlation	-.523**	1	.144	-.134	-.194	.067	-.192	.246*
	Sig. (2-tailed)	<.001		.153	.184	.053	.506	.056	.014
	N	100	100	100	100	100	100	100	100
Engine size	Pearson Correlation	.424**	.144	1	-.230*	-.031	-.001	.095	-.035
	Sig. (2-tailed)	<.001	.153		.021	.760	.991	.348	.732
	N	100	100	100	100	100	100	100	100
Number of owners	Pearson Correlation	-.188	-.134	-.230*	1	.018	-.073	-.040	.080
	Sig. (2-tailed)	.061	.184	.021		.861	.472	.693	.428
	N	100	100	100	100	100	100	100	100
Colour=Black	Pearson Correlation	-.013	-.194	-.031	.018	1	-.355**	-.344**	-.446**
	Sig. (2-tailed)	.897	.053	.760	.861		<.001	<.001	<.001
	N	100	100	100	100	100	100	100	100
Colour=Blue	Pearson Correlation	.035	.067	-.001	-.073	-.355**	1	-.227*	-.295**
	Sig. (2-tailed)	.730	.506	.991	.472	<.001		.023	.003
	N	100	100	100	100	100	100	100	100
Colour=Grey	Pearson Correlation	.222*	-.192	.095	-.040	-.344**	-.227*	1	-.285**
	Sig. (2-tailed)	.026	.056	.348	.693	<.001	.023		.004
	N	100	100	100	100	100	100	100	100
Colour=White	Pearson Correlation	-.182	.246*	-.035	.080	-.446**	-.295**	-.285**	1
	Sig. (2-tailed)	.071	.014	.732	.428	<.001	.003	.004	
	N	100	100	100	100	100	100	100	100

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

According to the output the evaluation of the mileage and engine capacity should be prioritize when determining the price of a used vehicle. By having a stronger correlation with the car's price, it is understandable that their influence on the car's value will be more significant. This analysis also shows that there is no significant correlation observed between price and color. Furthermore, although there is a correlation between price and the number of previous owners, this is not substantial and may be analyzed after engine capacity and mileage.

5. Regression Analysis

In this part, we will do the analysis to find out the function that can predict the price of a used car using its attributes.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.759 ^a	.577	.545	3045.833

a. Predictors: (Constant), Colour=White, Engine size , Number of owners , Colour=Grey, Mileage , Colour=Blue, Colour=Black

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1163274232.4	7	166182033.20	17.913	<.001 ^b
	Residual	853493256.17	92	9277100.611		
	Total	2016767488.6	99			

a. Dependent Variable: Price

b. Predictors: (Constant), Colour=White, Engine size , Number of owners , Colour=Grey, Mileage , Colour=Blue, Colour=Black

Coefficients ^a								
		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
Model		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	18860.673	3889.965		4.849	<.001		
	Mileage	-.127	.015	-.661	-8.496	<.001	.760	1.315
	Engine size	6754.413	993.687	.483	6.797	<.001	.910	1.099
	Number of owners	-928.905	412.694	-.159	-2.251	.027	.923	1.083
	Colour=Black	-5566.835	3312.027	-.591	-1.681	.096	.037	26.900
	Colour=Blue	-4177.128	3301.560	-.365	-1.265	.209	.055	18.083
	Colour=Grey	-4398.469	3378.546	-.376	-1.302	.196	.055	18.161
	Colour=White	-4733.838	3247.049	-.468	-1.458	.148	.045	22.400

a. Dependent Variable: Price

The analysis reveals that as an automobile's selling price tends to decrease as it collects mileage or changes ownership. In addition, the car's pricing is positively affected by the size of its engine.

However, this is not the most parsimonious model and we still have to do further analysis to find the appropriate function to help predict the price of a used car.

Remove insignificant factor

We will eliminate the four colors factor as it has been stated from the above model that these characteristics do not have a major impact on the price value of a used car. We will perform a Regression Analysis that includes three factors that are left:

Model Summary					ANOVA ^a				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Model	Sum of Squares	df	Mean Square	F
1	.743 ^a	.552	.538	3066.822	1 Regression	1113849519.6	3	371283173.20	39.476
					2 Residual	902917969.00	96	9405395.510	
					3 Total	2016767488.6	99		

a. Predictors: (Constant), Number of owners, Mileage, Engine size

b. Predictors: (Constant), Number of owners, Mileage, Engine size

Coefficients ^a								
		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
Model		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	13851.272	2088.962		6.631	<.001		
	Mileage	-.118	.013	-.613	-8.833	<.001	.969	1.033
	Engine size	6642.864	987.506	.475	6.727	<.001	.934	1.071
	Number of owners	-939.484	412.600	-.161	-2.277	.025	.937	1.068

a. Dependent Variable: Price

The results indicate that all three factors are statistically relevant in determining the price of a second-hand car. Now, let's examine whether there is any correlation between these three parameters since this can affect the accuracy of our function of predicting the overall price of a used car:

Check for correlations (Multicollinearity)

		Correlations			
		Mileage	Engine size	Number of owners	Price
Mileage	Pearson Correlation	1	.144	-.134	-.523**
	Sig. (2-tailed)		.153	.184	<.001
	N	100	100	100	100
Engine size	Pearson Correlation	.144	1	-.230*	.424**
	Sig. (2-tailed)	.153		.021	<.001
	N	100	100	100	100
Number of owners	Pearson Correlation	-.134	-.230*	1	-.188
	Sig. (2-tailed)	.184	.021		.061
	N	100	100	100	100
Price	Pearson Correlation	-.523**	.424**	-.188	1
	Sig. (2-tailed)	<.001	<.001	.061	
	N	100	100	100	100

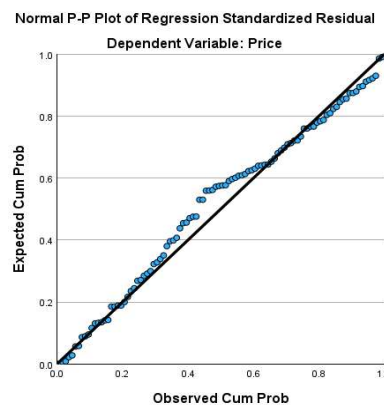
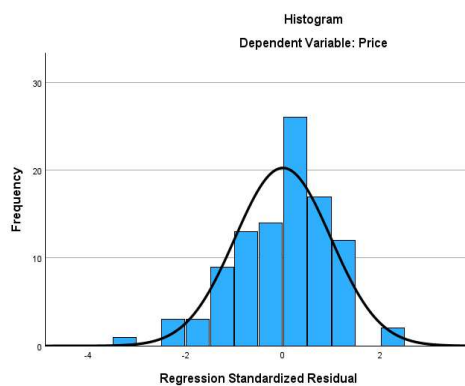
** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

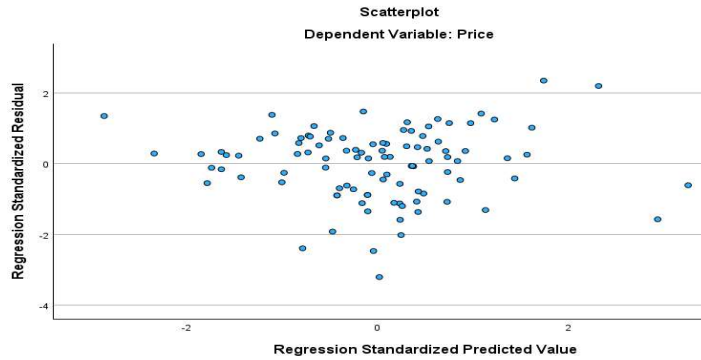
From the result, it can be seen that there is no link between these three elements. Hence, this is the most accurate model that we can derive from this sample to forecast the price of a used car.

6. Residual Analysis

Now the final step is to check the adequacy of the final model that we have concluded. This will determine if the model that we have is good enough to be used to predict the price of a used car.



Normality: Since the histogram displays an essentially normal distribution and the dots are extremely near to aligning with the diagonal line, we can say that the residuals follow a normal distribution.



Linearity: We can conclude that the scatterplot does not violate the linearity assumption as there is no evident pattern or systematic curve.

Independence: The independence assumptions is not violated because there is no discernible pattern or clustering of dots at specified ranges.

Homoscedasticity: There is no funnel shape in the scatter plot, hence the assumption of homoscedasticity is not violated.

Multicollinearity: Since the independent variables are unrelated to one another, multicollinearity is not violated.

7. Statistical Model

The final model to predict price of an second-hand car are:

$$Price = 13852.272 - 0.118Mileages + 6642.864EngineSize - 939.484Numberofowners + e$$

This model satisfies all the conditions to be considered a good equation to predict the price of a used car. However, one downside of this equation is that the value of Adjusted - R Square is 0.538. This proves that there are still factors that can affect the price of a used car, but these indicators are not covered by our data set.

An example of the usage of the model:

The price for a manual BMW 1 Series car, had ran 12000 mileages has an engine size of 2.0 and have 4 previous owners:

$$Price = 13852.272 - 0.118(12000) + 6642.864(2) - 939.484(4) = £21964.064$$