EXECUTIVE SUMMARY

1. Data was collected for BMW 3 series cars within a 40-mile radius of the NW1 4NP postcode, with a focus on cars sold between 2019 to 2023. The data collected includes price, registration year, segment, mileage(miles), Engine-size, BHP, Gearbox, Fuel-type, colour, number of owners and number of doors.

2. Mean price of BMW 3 series is £ 25338.35.

3. All of the linear and multi-linear regression criteria are satisfied.

4. The tools utilized in this study include **SPSS**, **MS Excel**, and **Python**.

5. A regression model has been constructed for the purpose of forecasting the price of BMW 3 series based on Age, miles, BHP, Fuel-type, Segment and Transmission.

$$\text{Price (£)} = 28155.040 - 3204.526(\text{age}) - 0.107(\text{miles}) + 45.642(\text{BHP}) + 1963.6(\text{diesel}) - 1319.785(\text{sedan}) + 3976.913(\text{automatic})$$

CHAPTER 1 - INTRODUCTION

Understanding the complex factors that affect the pricing of used cars, creating a scenario where buyers face uncertainty due to inadequate information on pricing and vehicle conditions. The absence of clear information obstructs buyers from making informed decisions and prevents sellers from establishing fair pricing practices. Statistical methods analyse market data to reveal pricing trends, brand quality insights, and factors influencing used car values, aiming to increase transparency for market participants.

- The BMW 3 series has been chosen for the study within a 40-mile radius of the NW1 4NP postcode.
- Source: https://www.autotrader.co.uk/car-search?advertising-location=at_cars&make=BMW&model=3%20Series&postcode=NW1%204NP&radius=40&sort=relevance&year-from=2019&year-to=2023
- After drawing out significant data around 422 population data from the Autotrader website, it is critical to recognize and deal with null values in order to properly structure the data for analysis and interpretation which ensures accuracy of the predictive model.

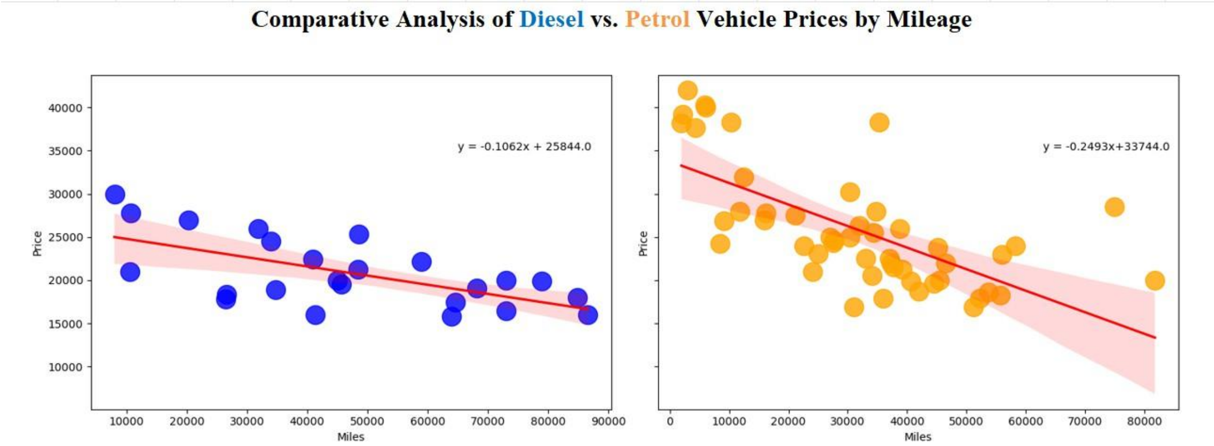
Limitations:

- The collected population data is cleaned to remove null values without altering the data's context, resulting in 315 data.
- Random sampling is then employed to select 100 data for statistical analysis, ensuring each data point has an equal chance of being included in the sample.
- The study focuses on cars within the age range of 2019-2023 to maintain a specific timeframe for analysis and consistency in the study parameters.

CHAPTER 2 - DATA VISUALIZATION

Data visualization is a powerful tool that transforms data into visual representations to improve understanding, communication, and decision-making across various domains, from business to scientific research.

i. Comparative Analysis of Diesel vs Petrol BMW 3 series



A **scatter plot** in simple terms is a graphical representation that illustrates the relationship between two sets of data by plotting points on a graph to show how they are interconnected.

In the above plot, we can observe the correlation between car's price and its mileage. Notably, there is a negative correlation for both petrol and diesel cars, indicating that as the mileage increases, the car's price decreases. The term 'Mileage' refers to the total distance in miles that a car has been driven.

The linear regression model for Diesel cars is represented as

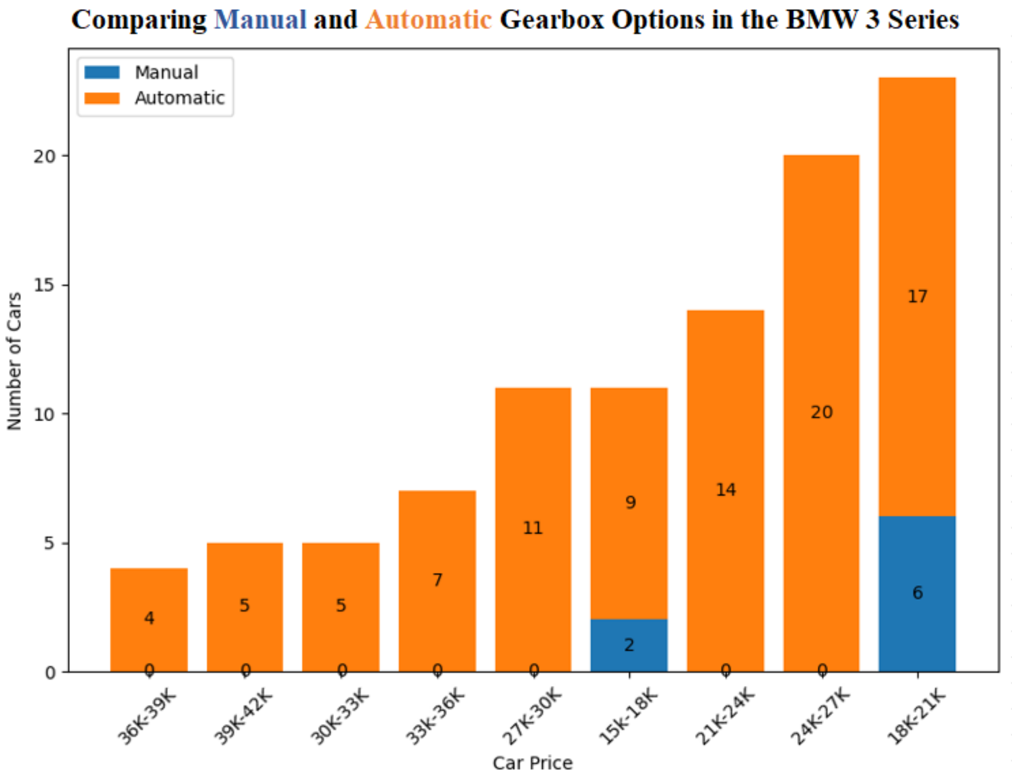
$$Y = -0.2493X + 33744$$

While for Petrol cars it is

$$Y = -0.1062X + 25844$$

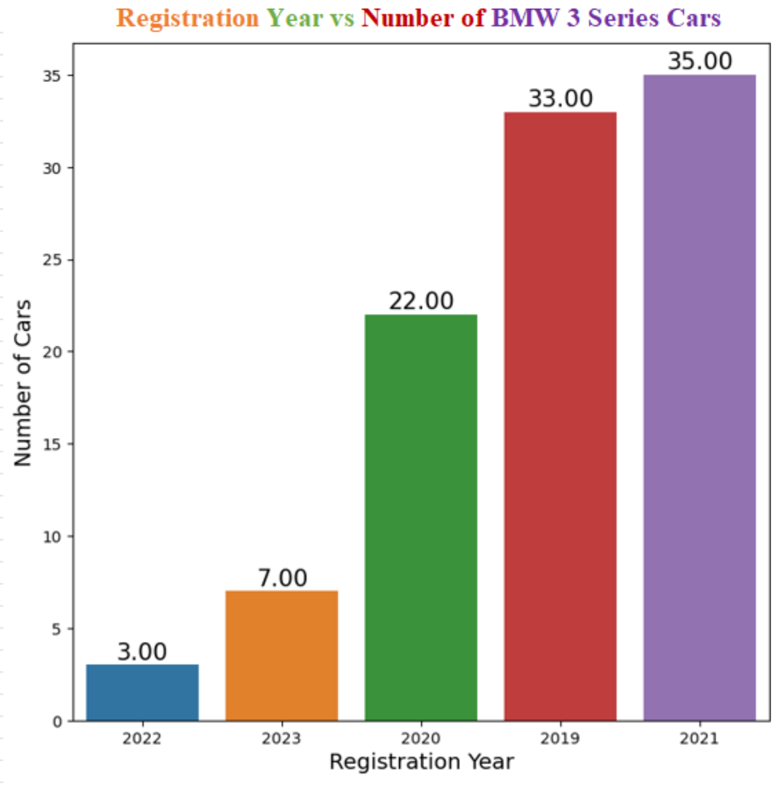
In these equations, X represents the mileage of the car.

ii. Comparison of BMW 3 Series Gearbox Options



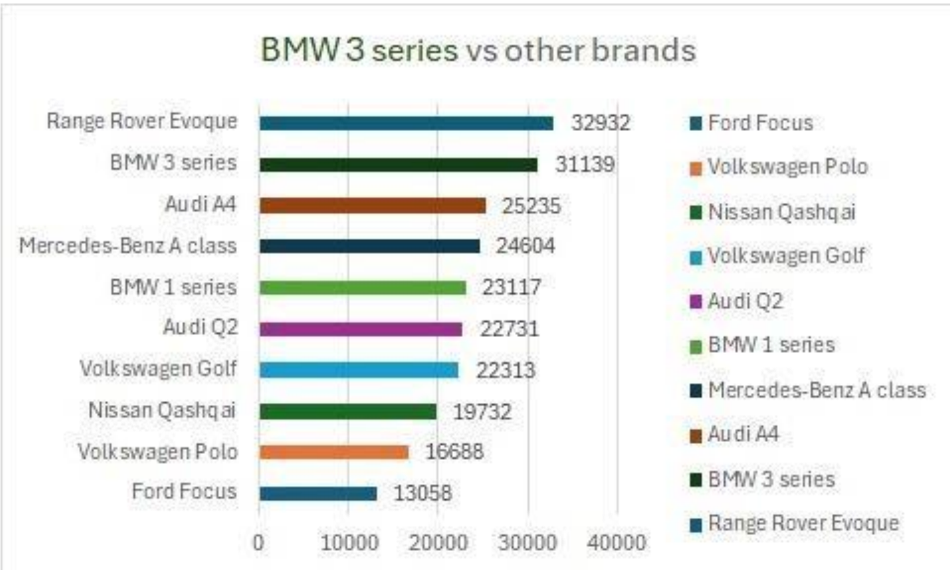
The distribution of manual and automatic cars in various price levels is seen in the stacked column chart above. The price ranges from £ 15000-42000 show up on the X-axis in £ 2000 increments. While the number of cars appears on the Y-axis. Every column denotes a range of prices, and the height of each column shows the total number of cars. Orange sections represent automatic cars, whereas blue section represents manual cars. All price points show a greater dominance of automated cars according to the data. There are more automatic cars than manual cars in my sample data for the BMW 3 series, and the majority of automatic cars found in the 24-27k price bracket. On the other hand, just 8 manual cars within the price 15-21k.

iii. Analysis of Registration Year vs. Number of BMW 3 Series



The bar graph illustrates the number of car registrations from 2019 to 2023, displaying varying counts: In 2019, 33 cars were registered. In 2020, 22 cars. A record high of 35 registrations was reached in 2021, yet there was an abrupt drop to 3 in 2022 and 7 by 2023. For clarity, each year's registration data is displayed in a different colour on the chart, demonstrating fluctuating trends in registrations over the time.

iv. Comparative Analysis of BMW 3 Series Pricing Against Competitors



The provided image depicts a horizontal bar chart that compares Price across various car models. The Range Rover Evoque has the highest value, followed by the BMW 3 Series and the Audi A4. The Mercedes-Benz A-Class and BMW 1 Series also have relatively high values. On the lower end of the chart, the Nissan Qashqai, Volkswagen Polo, and Ford Focus have the lowest values, with the Ford Focus being the lowest among the displayed models.

Each car model is represented by a color-coded bar, with a corresponding legend on the right side of the chart. This type of visualization allows for a quick comparison of the relative performance of the BMW 3 Series against its competitors in the automotive market.

The above all visualized graphs satisfy the Gestalt principles and also follows the Tufte’s principles.

Tufte’s principle focus on clear data visualization through accurate number representation, clear labelling, and prioritizing data variation over design elements to maintain focus on the data.
Gestalt principles are principles of human perception that describe how we group similar elements, recognize patterns, and simplify complex images when we perceive objects.

CHAPTER 3 - STATISTICAL ANALYSIS

The average value within a set of data can be derived through statistical measures of central tendency such as the mean, median, and mode in descriptive analysis.

- Mean: The mean is the average value of a dataset. It is calculated by adding up all values and dividing by the total number of values.
- Median: The median is the middle value in a dataset when values are ordered. If there is an even number of values, the median is the average of the two middle values.
- Mode: The mode is the value that appears most frequently in a dataset. A dataset can have one or more modes or no mode at all.
- Variance: Measures the spread of data points from the mean.
- Range: Difference between the highest and lowest values in a dataset.
- Standard Deviation: statistical measure that quantifies the amount of variation or dispersion in a set of values. It indicates how spread out the values are from the mean.

	Price	Year	Segment	Miles	Engine	BHP	Owners	No of Doors
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000
mean	25338.350000	2020.290000	0.770000	35084.270000	2.190000	211.620000	1.590000	4.230000
std	6702.353237	1.165974	0.422953	19046.965918	0.394277	63.791925	0.805223	0.422953
min	15800.000000	2019.000000	0.000000	1940.000000	2.000000	148.000000	1.000000	4.000000
25%	19948.750000	2019.000000	1.000000	23503.500000	2.000000	181.000000	1.000000	4.000000
50%	24275.000000	2020.000000	1.000000	33997.500000	2.000000	187.000000	1.000000	4.000000
75%	29219.000000	2021.000000	1.000000	44623.500000	2.000000	190.000000	2.000000	4.000000
max	41979.000000	2023.000000	1.000000	86614.000000	3.000000	374.000000	4.000000	5.000000

From the table, it is evident that our sample includes 100 data points for the BMW 3 series. The average price is £ 25,338.35 with a maximum price of £ 41,979 and a minimum of £ 15880. Within these 100 samples, there are approximately 77 sedan cars with an average price of £ 25,338 and around 23 estate cars with an average price of £ 25287. Based on this data, we can conclude that sedan cars have a slightly higher average price compared to estate cars, although the difference is not significant.

Descriptive Statistics of Price of Diesel v/s Petrol v/s Diesel-Hybrid v/s Petrol-Hybrid:

Diesel		*Petrol*		*Diesel-Hybrid*		*Petrol-Hybrid*	
Mean	20912	Mean	25849.83673	Mean	29100.55556	Mean	34233.75
Standard Error	761.956189	Standard Error	988.440752	Standard Error	1304.573454	Standard Error	631.6030366
Median	19855	Median	24250	Median	29443	Median	33997.5
Mode	19950	Mode	22490	Mode	26157	Mode	N/A
Standard Deviation	4103.259654	Standard Deviation	6919.085264	Standard Deviation	5534.836415	Standard Deviation	1263.206073
Sample Variance	16836739.79	Sample Variance	47873740.89	Sample Variance	30634414.14	Sample Variance	1595689.583
Kurtosis	-0.043357798	Kurtosis	0.023887261	Kurtosis	-0.962604589	Kurtosis	1.101040578
Skewness	0.940417643	Skewness	0.982265838	Skewness	-0.00748503	Skewness	0.984010308
Range	14195	Range	25080	Range	18705	Range	2960
Minimum	15800	Minimum	16899	Minimum	20295	Minimum	32990
Maximum	29995	Maximum	41979	Maximum	39000	Maximum	35950
Sum	606448	Sum	1266642	Sum	523810	Sum	136935
Count	29	Count	49	Count	18	Count	4

A comparison of diesel, petrol, diesel-hybrid, and petrol-hybrid cars can be seen in the table above. With 49, petrol cars have the greatest count, followed by diesel cars with 29, diesel-hybrid cars with 18 and petrol-hybrid cars with just 4. In general, diesel cars are more affordable, than the other three fuel types, with petrol vehicles coming in second. While petrol-hybrid cars are substantially more expensive than the other three fuel-types.

Descriptive Statistics of Price of Automatic v/s Manual Cars:

Automatic		*Manual*	
Mean	25932.55435	Mean	18505
Standard Error	692.7936665	Standard Error	640.0467
Median	24997	Median	18850
Mode	21300	Mode	N/A
Standard Deviation	6645.04341	Standard Deviation	1810.326
Sample Variance	44156601.92	Sample Variance	3277279
Kurtosis	-0.477822746	Kurtosis	-0.56804
Skewness	0.652820692	Skewness	-0.53795
Range	25989	Range	5190
Minimum	15990	Minimum	15800
Maximum	41979	Maximum	20990
Sum	2385795	Sum	148040
Count	92	Count	8

The comparison table above shows that there are more automatic cars (92) than manual cars (8), with an average price of £ 25932.55 for automatic cars and £ 18505 for manual cars. Automatic cars are regarded as superior to manual ones in terms of standard-deviation and range, indicating that most customers prefer to buy automatic cars when purchasing a new car.

CHAPTER 4 - CONFIDENCE INTERVALS

In statistics, a confidence interval is a probability that a population parameter will fall within a set of values a particular proportion of the time. A 95% confidence interval demonstrates that there is a 95% chance that the true population parameter is within the stated range of values around the mean.

The following presents the 95% confidence intervals for the BMW 3 series sample.

```
Confidence intervals (95%):
              Price         Year         Miles     Engine         BHP
Lower  24024.712904  2020.061473  31351.133279  2.112723  199.117012
Upper  26651.987096  2020.518527  38817.406721  2.267277  224.122988
```

24024.71<25338.35<26651.98, this range indicates that the average car price (£) is within these significant intervals.

One-sample T test

A one sample T-test will be used to examine whether the mean price of a BMW 3 series model in our data set lines up with the UK average price. The average price and mileage from 2019 to 2023 were estimated using https://www.carsite.co.uk/used-car-price-guide/bmw/3-series . After which, we compare our mean price and mean miles to the UK average price.

Within our BMW 3 series sample:

- The mean price of the BMW 3 series in our sample: £ 25338.35
- The average cost of the BMW 3 series in the UK: £ 31139
- The mean mileage of the BMW 3 series in our sample: £ 35084.27
- The typical mileage of the BMW 3 series in the UK: £ 32648

SPSS Output for one sample T-test for price:

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Price	100	25338.35	6702.353	670.235

One-Sample Test

	Test Value = 31139						
			Significance			95% Confidence Interval of the Difference	
	t	df	One-Sided p	Two-Sided p	Mean Difference	Lower	Upper
Price	-8.655	99	<.001	<.001	-5800.650	-7130.54	-4470.76

One-Sample Effect Sizes

				95% Confidence Interval	
		Standardizer[a]	Point Estimate	Lower	Upper
Price	Cohen's d	6702.353	-.865	-1.094	-.634
	Hedges' correction	6753.669	-.859	-1.086	-.629

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.

In this case, we reject the null hypothesis (HO) implies a significant difference between the sample mean and the test value. The negative t-value of -8.655 indicates that the sample value is lower than the test value. Furthermore, the very low P-value (much less than 0.001) indicates that the sample value is less than the 5% significance level, implying that it does not correspond to the mean price been evaluated.

SPSS Output for one sample T-test for miles:

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Miles	100	35084.27	19046.966	1904.697

One-Sample Test

	Test Value = 32648						
			Significance			95% Confidence Interval of the Difference	
	t	df	One-Sided p	Two-Sided p	Mean Difference	Lower	Upper
Miles	1.279	99	.102	.204	2436.270	-1343.06	6215.60

One-Sample Effect Sizes

				95% Confidence Interval	
		Standardizer[a]	Point Estimate	Lower	Upper
Miles	Cohen's d	19046.966	.128	-.069	.324
	Hedges' correction	19192.796	.127	-.069	.322

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.

In this context, we fail to reject the null hypothesis (HO) implies that there is no significant difference between the sample mean and the expected mean. The positive t-value of 1.279 indicates that the sample mean is slightly higher than the expected mean by 2436.27 units, furthermore, the p-value of 0.204 being above the significance level, indicates that our sample mileage is consistent with the true mean mileage.

Conducting Chi-square test for Fuel-type v/s Colours:

Fuel * Colour Crosstabulation

Count

		Colour							
		Black	Blue	Bronze	Grey	Orange	Silver	White	Total
Fuel	Diesel	8	5	1	7	0	0	8	29
	Diesel Hybrid	5	4	0	4	0	0	5	18
	Petrol	18	6	0	7	1	2	15	49
	Petrol Hybrid	1	0	0	1	0	0	2	4
Total		32	15	1	19	1	2	30	100

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	9.599[a]	18	.944
Likelihood Ratio	11.240	18	.884
N of Valid Cases	100		

a. 19 cells (67.9%) have expected count less than 5. The minimum expected count is .04.

The findings above show a p-value of 0.944, significantly surpasses the significance level of 0.05, making it statistically insignificant. As a result, we don't have enough evidence to reject the null hypothesis, indicating that based on our sample data, there is no casual connection between Fuel-type and Colours.

CHAPTER 5 - CORRELATION MATRIX

Correlations

		Price	Age	Miles	BHP	Fuel=Diesel	Segment=Sedan	Gearbox=Automatic	Engine	Owners	No of Doors
Price	Pearson Correlation	1	-.761	-.598	.514	-.424	-.169	.325	.521	-.118	.169
	Sig. (2-tailed)		<.001	<.001	<.001	<.001	.093	<.001	<.001	.241	.093
	N	100	100	100	100	100	100	100	100	100	100

		Colour=Black	Colour=Blue	Colour=Bronze	Colour=Grey	Colour=Orange	Colour=Silver	Colour=White	Fuel=Diesel Hybrid	Fuel=Petrol	Fuel=Petrol Hybrid	Segment=Estate	Gearbox=Manual
Price	Pearson Correlation	.071	.045	-.118	.009	-.016	-.045	-.072	.264	.075	.272	.169	-.325
	Sig. (2-tailed)	.483	.657	.242	.932	.871	.658	.478	.008	.457	.006	.093	<.001
	N	100	100	100	100	100	100	100	100	100	100	100	100

Analysing the correlation matrix above reveals key insights:

- Both age and mileage exhibit a strong negative correlation with price, at -0.761 and -0.598 respectively, with correlation coefficients less than 0.01. This indicates that older cars and those with higher mileage tend to have lower prices.
- There is a negative correlation of -0.424 between price and diesel cars, suggesting that diesel vehicles are generally cheaper in this dataset.
- BHP and engine-size show a moderate positive correlation with price, at 0.514 and 0.52 respectively, indicating that cars with higher BHP and larger engines tend to have higher prices.
- The correlation between price and automatic transmission is weakly positive, implying that cars with automatic gearboxes maybe slightly more expensive than those with manual transmission.

CHAPTER 6 - REGRESSION ANALYSIS

Regression analysis: Statistical method used to analyze the relationship betweenvariables, specifically estimating the impact of one or more independent variables on a dependent variable.

A **parsimonious model** in statistics is a simplified model that achieves good fit with minimal variables.

Regression

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Gearbox=Automatic, Colour=Grey, Miles, Fuel=Petrol Hybrid, Colour=Silver, Colour=Orange, Owners , Colour=Blue, Fuel=Diesel Hybrid, No of Doors , Colour=White, Fuel=Diesel, Engine, Age, BHP, Colour=Black[b]	.	Enter

a. Dependent Variable: Price

b. Tolerance = .000 limit reached.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.921[a]	.848	.819	2853.974

a. Predictors: (Constant), Gearbox=Automatic, Colour=Grey, Miles, Fuel=Petrol Hybrid, Colour=Silver, Colour=Orange, Owners , Colour=Blue, Fuel=Diesel Hybrid, No of Doors , Colour=White, Fuel=Diesel, Engine, Age, BHP, Colour=Black

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3771183516.2	16	235698969.76	28.937	<.001[b]
	Residual	676048836.55	83	8145166.705		
	Total	4447232352.7	99			

a. Dependent Variable: Price

b. Predictors: (Constant), Gearbox=Automatic, Colour=Grey, Miles, Fuel=Petrol Hybrid, Colour=Silver, Colour=Orange, Owners , Colour=Blue, Fuel=Diesel Hybrid, No of Doors , Colour=White, Fuel=Diesel, Engine, Age, BHP, Colour=Black

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	22105.969	4847.015		4.561	<.001
	Age	-3445.141	405.097	-.599	-8.504	<.001
	Miles	-.102	.021	-.290	-4.846	<.001
	Engine	-1290.858	2149.917	-.076	-.600	.550
	BHP	55.804	14.052	.531	3.971	<.001
	Owners	179.113	454.164	.022	.394	.694
	No of Doors	1198.914	730.854	.076	1.640	.105
	Colour=Black	1024.463	3006.991	.072	.341	.734
	Colour=Blue	262.469	3049.721	.014	.086	.932
	Colour=Grey	902.425	3042.733	.053	.297	.768
	Colour=Orange	1642.387	4272.907	.025	.384	.702
	Colour=Silver	3264.954	3774.838	.069	.865	.390
	Colour=White	1164.884	3037.788	.080	.383	.702
	Fuel=Diesel	2394.665	923.841	.163	2.592	.011
	Fuel=Diesel Hybrid	94.156	868.316	.005	.108	.914
	Fuel=Petrol Hybrid	-1260.062	1870.470	-.037	-.674	.502
	Gearbox=Automatic	3978.983	1216.512	.171	3.271	.002

a. Dependent Variable: Price

During regression analysis in SPSS, the car price is marked as the dependent variable, while all other variables are considered independent. With an Adjusted R-square of 0.819, indicating a good model fit, this signifies the model's predictive capability is 81.9%. Insignificant independent variables with a significance level less than 0.05 are preserved, while those with a significance level greater than 0.05 are eliminated one at a time subject to their significance values.

Regression

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Gearbox=Automatic, Miles, Segment=Sedan, BHP, Fuel=Diesel, Age[b]	.	Enter

a. Dependent Variable: Price

b. Tolerance = .000 limit reached.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.918[a]	.843	.833	2742.284

a. Predictors: (Constant), Gearbox=Automatic, Miles, Segment=Sedan, BHP, Fuel=Diesel, Age

b. Dependent Variable: Price

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3747860961.9	6	624643493.65	83.063	<.001[b]
	Residual	699371390.84	93	7520122.482		
	Total	4447232352.7	99			

a. Dependent Variable: Price

b. Predictors: (Constant), Gearbox=Automatic, Miles, Segment=Sedan, BHP, Fuel=Diesel, Age

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	28155.040	1859.741		15.139	<.001
	Age	-3204.526	319.806	-.557	-10.020	<.001
	Miles	-.107	.018	-.304	-5.969	<.001
	BHP	45.642	4.494	.434	10.157	<.001
	Fuel=Diesel	1963.600	765.330	.134	2.566	.012
	Segment=Sedan	-1319.785	664.566	-.083	-1.986	.050
	Gearbox=Automatic	3976.913	1065.586	.171	3.732	<.001

a. Dependent Variable: Price

Excluded Variables[a]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
1	No of Doors	.[b]	.	.	.	.000

a. Dependent Variable: Price

b. Predictors in the Model: (Constant), Gearbox=Automatic, Miles, Segment=Sedan, BHP, Fuel=Diesel, Age

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	15778.52	37059.02	25338.35	6152.819	100
Residual	-6424.576	6431.599	.000	2657.886	100
Std. Predicted Value	-1.554	1.905	.000	1.000	100
Std. Residual	-2.343	2.345	.000	.969	100

a. Dependent Variable: Price

After removing all insignificant variables, the Adjusted R-square improved from 0.819 to 0.833, indicating that our model is now considered parsimonious. The analysis reveals the following insights on the predictors and their impact on car prices:

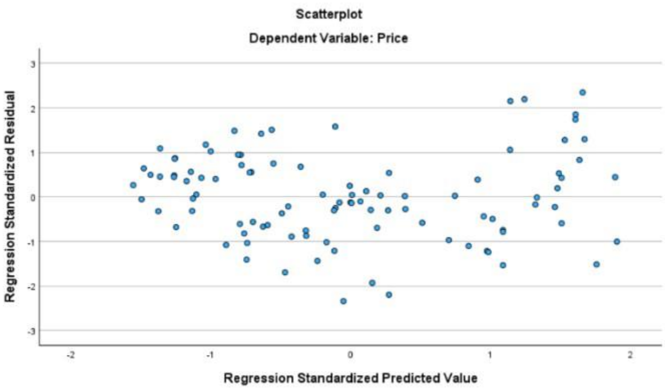
- Age is a significant predictor of car price, with each additional year resulting in a £ 3204.526 decrease (p-value<0.001)
- For every extra mile, the price drops by 0.107, which is statistically significant.
- Higher BHP results in higher prices, with each unit increase adding £ 45.642 to the price, indicating a strong and substantial predictor.
- Diesel cars are approximately £ 1963.600 more expensive than the petrol fuel-type indicating a significant predictor with a p-value of 0.012.
- Sedan cars are £ 1319.785 cheaper than the baseline segment, a finding on the verge of statistical significance (p-value = 0.050)
- Cars with automated gearbox cost £ 3976.913 more on average than cars with manual gearbox, making this a significant predictor in the analysis

CHAPTER 7 – RESIDUAL ANALYSIS

Having established our model as parsimonious in the previous chapter, this chapter will focus on testing its adequacy by ensuring it satisfies five essential assumptions for Linear regression.

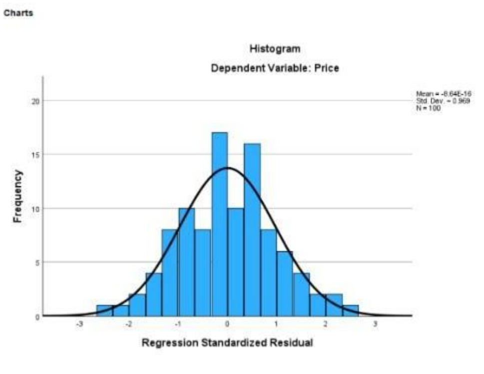
Assumption 1: Linear Relationship

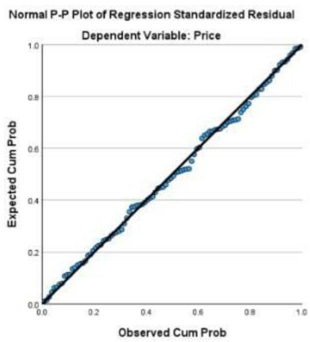
The first step in linear regression is to ensure that the correlation between the independent and dependent variables is linear. Checking for outliers is crucial because linear regression is sensitive to their effects, and scatter plots are the best way to test the linearity assumption.



Assumption 2: Multi-variate normality

Second, linear regression analysis requires that all variables be multivariate normal. This assumption can be examined most effectively using a histogram or Q-Q plot. The graphs below demonstrate that our model is linear and that the errors follow a normal distribution.





Assumption 3: No or little multicollinearity

Linear regression is based on the assumption that there is little or no multicollinearity in the data, which occurs when independent variables have strong correlations with one another. The table reveals that our matrix exhibits no significant correlations, implying an absence of multicollinearity. Therefore, the coefficients in our model are likely to remain steady.

Correlations

Correlations

		Price	Age	Miles	BHP	Fuel=Diesel	Segment=Sedan	Gearbox=Automatic
Price	Pearson Correlation	1	-.761	-.598	.514	-.424	-.169	.325
	Sig. (2-tailed)		<.001	<.001	<.001	<.001	.093	<.001
	N	100	100	100	100	100	100	100
Age	Pearson Correlation	-.761	1	.576	-.125	.502	.027	-.229
	Sig. (2-tailed)	<.001		<.001	.217	<.001	.788	.022
	N	100	100	100	100	100	100	100
Miles	Pearson Correlation	-.598	.576	1	-.012	.316	.025	-.039
	Sig. (2-tailed)	<.001	<.001		.907	.001	.802	.697
	N	100	100	100	100	100	100	100
BHP	Pearson Correlation	.514	-.125	-.012	1	-.229	-.152	.145
	Sig. (2-tailed)	<.001	.217	.907		.022	.131	.150
	N	100	100	100	100	100	100	100
Fuel=Diesel	Pearson Correlation	-.424	.502	.316	-.229	1	.140	-.415
	Sig. (2-tailed)	<.001	<.001	.001	.022		.165	<.001
	N	100	100	100	100	100	100	100
Segment=Sedan	Pearson Correlation	-.169	.027	.025	-.152	.140	1	-.089
	Sig. (2-tailed)	.093	.788	.802	.131	.165		.379
	N	100	100	100	100	100	100	100
Gearbox=Automatic	Pearson Correlation	.325	-.229	-.039	.145	-.415	-.089	1
	Sig. (2-tailed)	<.001	.022	.697	.150	<.001	.379	
	N	100	100	100	100	100	100	100

Assumption 4: Residuals are independent

The residuals in the regression analysis should be independent of one another, which means that the residuals from one observation should not predict those from the other. The above scatter plot shows that my model lacks a specific pattern, indicating that this assumption is satisfied.

Assumption 5: Homoscedasticity

The variances of the errors should be consistent, that signifies the pattern of distribution of residuals must be uniform across all levels of independent variables. The scatter plot above shows that the residuals remain close to zero, backing the validity of this assumption.

ADEQUACY TEST

After confirming that our model fits all assumptions and is parsimonious, the next step is to estimate the price using the equation y = mx + c.

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	28155.040	1859.741		15.139	<.001
	Age	-3204.526	319.806	-.557	-10.020	<.001
	Miles	-.107	.018	-.304	-5.969	<.001
	BHP	45.642	4.494	.434	10.157	<.001
	Fuel=Diesel	1963.600	765.330	.134	2.566	.012
	Segment=Sedan	-1319.785	664.566	-.083	-1.986	.050
	Gearbox=Automatic	3976.913	1065.586	.171	3.732	<.001

a. Dependent Variable: Price

Based on the filtered predictors we have incorporated into our dataset; we will now solve the equation to assess how well our predicted values match the actual market values.

Price(Y) in £ = 28155.040 + (–3204.526*Age) + (–0.107*Miles) + (45.642*BHP) + (1963.6*Fuel=Diesel) + (-1319.785*Segment=sedan) + (3976.913*Gearbox=Automatic)

	A	B	C	D	E	F	G	H	I	J	K	L
1	Price	Year	Segme	Miles	Engine	BHP	Gear-box	Fuel Type	Colour	Owners	No of Doors	Age
28	30,995	2021	Sedan	20,300	2	190	Automatic	Diesel	Blue	1	4	3

When we analyse the 28th row of population data alongside the unstandardized value and compare it with the statistical model, we notice that the car price closely aligns with the predicted value.

Price(Y) = 28155.040 – (3504.526*3) – (0.107*20300) + (45.642*190) + (1963.6) – (1319.785) + (3976.913)

Price(Y) = 29662.07 ~ £ 29662

Our predicted value was £ 29662, while the actual market value is £ 30995, showing a slight variance. This indicates that our prediction is close to the market value, suggesting accuracy. Factors such as Registration year, segment, mileage, BHP, gearbox, and fuel type contributes to predict the price of BMW 3 series cars.

CONCLUSION

The model has accurately captured the primary determinants of used BMW 3 series pricing, considering critical predictors such as the car's age, performance, and configuration. The model's ability to capture these linkages demonstrates its reliability in accurately calculating the market value of these used BMW 3 series cars. Furthermore, the model's alignment with expected and actual market prices demonstrates its effectiveness as a reliable tool for directing pricing strategies and decisions in the purchasing and selling of used BMW 3 series.

Overall, this model is a useful asset in the used car market since it identifies key pricing factors and provides precise price projections to help buyers make informed purchasing decisions.

REFERENCES

[1] https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/assumptions-of-multiple-linear-regression/

[2] https://www.statology.org/parsimonious-model/

[3] https://journals.ala.org/index.php/ltr/article/view/6289/8215

[4] https://www.dictionary.com/e/average-vs-mean-vs-median-vs-mode/

[5] https://www.techtarget.com/searchbusinessanalytics/definition/data-visualization