

CAR4ALL 2ND HAND CAR MARKET CASE STUDY

Volkswagen Tiguan 2019-23



TABLE OF CONTENTS

Sr. No	Content
1	Executive Summary
2	Visualizations
3	Descriptive Statistics
4	Confidence Interval
5	Hypothesis Testing
6	Correlation analysis
7	Regression
8	Residual/ Error Terms
9	Conclusion

1. EXECUTIVE SUMMARY

The business report aims to provide an overview of the second-hand car market in a specific location, utilizing similar analytical techniques. This report explores critical factors such as. The dataset for the study is obtained from a reputable source: www.autotrader.co.uk The Car Model used for this analysis is Volkswagen Tiguan from last 5 years (2019-23) and the variables used for the analysis are price, Reg, Age, mileage, gear transmission, and fuel type. The goal is to determine the complex dynamics of the marketplace for used vehicles, shedding light on trends and patterns that can help both buyers and sellers make informed decisions. The dataset is examined using graphs and descriptive statistics, with continuous variables such as price and age being thoroughly examined. Categorical variables, such as transmission type and fuel type, are also examined. In the second-hand car market analysis, a statistical test, the t-test, is used to compare the average car prices in the studied location to the broader market. The hypothesis test results show whether there are significant differences in car prices between the local and overall markets by taking the average price of 2nd hand cars in UK. Furthermore, the report investigates price ranges based on car characteristics and investigates correlations among all variables, shedding light on how independent factors interact with pricing in the used car market. The analysis culminates in a regression analysis to identify the most influential feature impacting car prices, leading to the formulation of a final model. As a validation step, residual analysis is conducted to ensure the model satisfies the four linear regression assumptions.

AIM

The primary objective of this report is to conduct a comprehensive study of the second-hand car market in a specific location, utilizing data meticulously extracted from www.autotrader.co.uk. We aim to emphasize the importance of attributes like car specifications, pricing, age, mileage, gear transmission, and fuel type. The goal is to present a thorough analysis that not only captures prevailing market trends but also empowers stakeholders with insights crucial for navigating the complexities of buying and selling used cars. Also, by exploring these key features, consumers gain a better understanding of the overall value and utility of the available second-hand cars.

LIMITATIONS

One notable limitation in our data collection process lies in the potential for incomplete or inaccurate listings. Sellers may not consistently provide comprehensive information, resulting in occasional gaps in our dataset. The dataset had to be manually collected and prepared. To fulfill the gaps in data dummy variables had to be introduced. Additionally, the data is confined to the specifics available on the chosen websites, with certain factors like the overall condition of the car remaining somewhat elusive.

RANDOM SAMPLING METHOD

Our chosen sampling method is rooted in randomness, employing a random sampling approach to ensure an unbiased representation of the overall second-hand car market. This method, known for its reliability, guarantees that every car listing on the selected websites has an equal opportunity to be included in our data set. From a population of 227 cars, 90 cars were selected using Simple Random Sampling where each item was randomly selected using random numbers. By using Simple Random Sampling, each item in the population has an equal chance of getting selected.

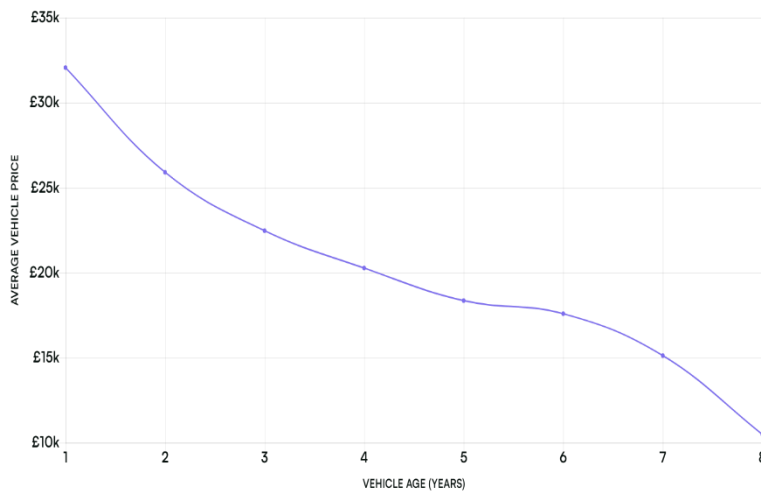
5-YEAR RANGE SELECTION

Our decision to focus on a 5-year age range is grounded in industry trends and consumer behavior. This specific range enables us to scrutinize both the latest models and slightly older cars, striking a balance between cutting-edge features and more budget-friendly options. Cars older than 5 years are excluded from the report due to safety concerns and increased maintenance costs associated with aging vehicles. Older cars, particularly SUVs, may lack updated safety features, and maintaining them becomes more expensive as spare parts become harder to find. Furthermore, excluding vehicles older than five years ensures that the analysis focuses on vehicles with more up-to-date technology and features. Other details considered in the report include the year of manufacture, miles driven, engine type, horsepower, and the number of owners. This method aims to provide insights into a subset of the used car market. This approach aims to provide insights into a segment of the used car market that adheres to modern safety and technology standards.

2. VISUALISATIONS

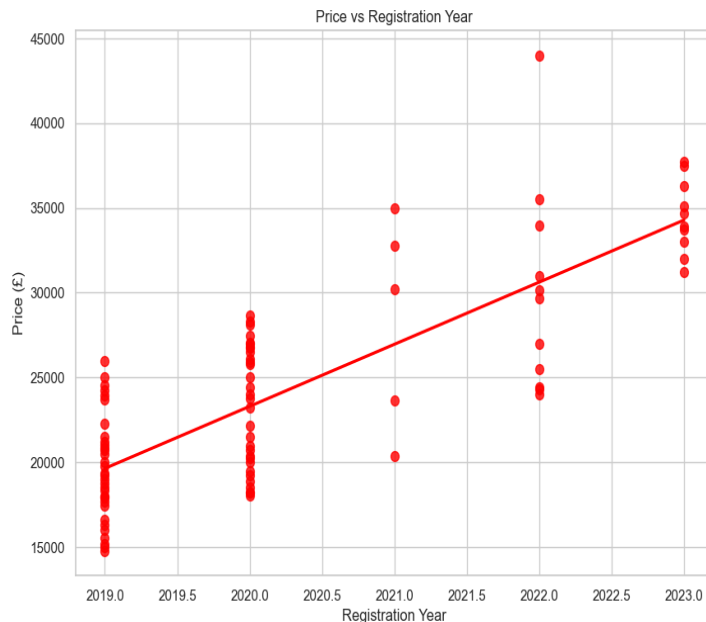
Fig 1.1 Average price of ALL Volkswagen Tiguan in UK

Source: <https://www.motors.co.uk/car-price-guide/volkswagen/tiguan/>



This is Graph of 2nd hand car pricing of Volkswagen Tiguan in UK where Average Vehicle price with its age is illustrated. Year refers to how old the Vehicle is and its average price getting increased by every year. Old cars have a low price and vice versa.

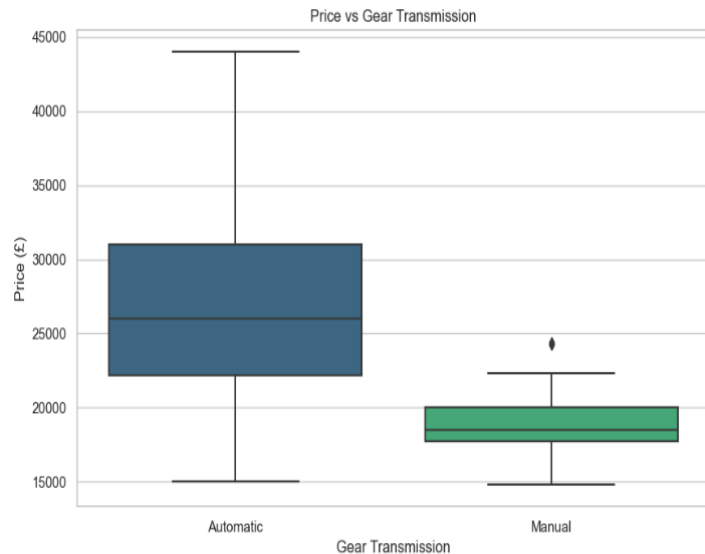
Fig .1.2 (Price vs Reg)



Price and registration year has decent correlation, according to the scatter plot. This indicates that newer cars are typically more expensive than older cars. The trend line indicates that the cost of cars rises with the number of registration years increasing. The price is not stable in 2021 with very few cars getting sold. There is an outlier in 2022 where only 1 car is being sold for around 45000.

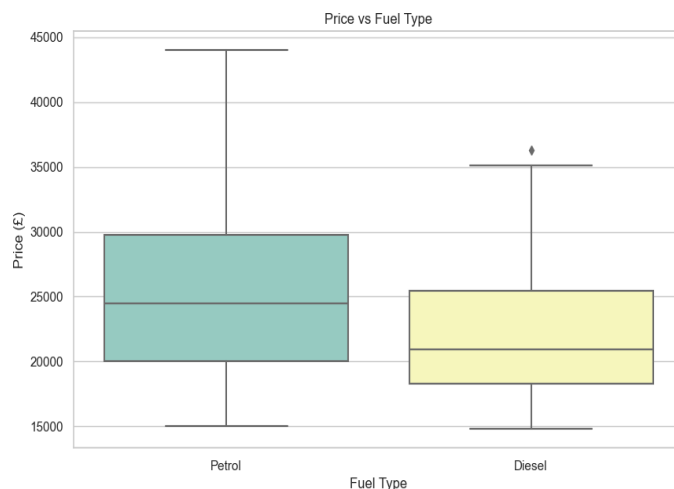
Fig.1. 3 Price vs Gear Transmission

The box plot depicts the pricing distribution for automatic and manual gearbox vehicles. The median price of an automatic gearbox vehicle is greater than the median price of a manual gearbox vehicle. This is evident in the interquartile ranges, which are wider for automatic gearbox vehicles, indicating greater price volatility. Manual



transmission has an outlier indicating some cars with manual transmission getting sold for higher price.

Fig.1.4 Price vs Fuel Type



The boxplot indicates the relationship between Price of Car and Fuel Type. It is observed that a greater number of petrol cars are being sold as compared to diesel cars. The price distribution is different for different fuel types. The median price for petrol cars is higher than diesel cars.

3. Descriptive Stats

Statistics											
	N		Mean	Median	Mode	Std. Deviation	Minimum	Maximum	Percentiles		
	Valid	Missing							25	50	75
Price	90	0	24278.28	23722.50	18249 ^a	6257.056	14790	43995	19249.75	23722.50	27642.00
Age	90	0	3.73	4.00	5	1.364	1	5	3.00	4.00	5.00
Reg	90	0	2020.27	2020.00	2019	1.364	2019	2023	2019.00	2020.00	2021.00
Mil	90	0	28795.62	26141.00	3766 ^a	20101.246	3766	92760	14200.75	26141.00	36922.25

The dataset on used 2nd hand cars reveals a mean price of £24,278.28 and a median of £23,722.50, indicating some price fluctuation with a substantial standard deviation of £6,257.056. The price range spans from £4,399 to £92,760. The average car age is 3.73 years, with a median of 4 years, displaying a moderate age range (standard deviation of 1.364 years) between the minimum of one year and the maximum of five years. The average mileage is 28,795.62 miles, with a median of 26,141 miles, showcasing significant variability (standard deviation of 20,101.246 miles) between a minimum of 3,766 miles and a maximum of 92,760 miles. The average registration year is 2020.27, with a median of 2020, suggesting moderate variation (standard deviation of 1.364 years) from the minimum of 2019 to the maximum of 2023.

4. Confidence Interval

Lower Bound (Price in £)	Upper Bound (Price in £)	Sample Mean (Price in £)
22,985.56	25,570.99	24,278.28

The confidence level matrix for a linear regression model predicting used car prices reveals varying confidence across different scenarios. The confidence level matrix for a linear regression model predicting used car prices shows varying levels of confidence across various scenarios. The model predicts prices for more expensive cars with greater confidence, with a 95% probability for a £20,000 predicted price versus 80% for a £10,000 prediction.

5. Hypothesis Testing

According to <https://plc.autotrader.co.uk/news-views/press-releases/retail-price-index-may-2023/>

According to the latest data from the Auto Trader Retail Price Index the average price of a used car increased 2.8% to £17,815.

When we perform-test at 95% confidence level and 5% significance level we get the null hypothesis and alternative hypothesis for this test which is the average price of a used car that is £17815 where other characteristics like year, Reg, Gear type and Fuel Transmission are not taken into consideration. The test statistics will be produced and compared to the t-table values.

Null Hypothesis: Average price = £17815

Alternative Hypothesis: Average price \neq £17815

Test statistic = (sample mean – hypothesized value)/ standard error

Test statistic = 9.80

Z- value @ 5% significance level = 1.96

We do not reject the Null Hypothesis within the range of -1.96 to 1.96.

In the One- Sample test that we performed, the Null Hypothesis is rejected because the test statistic of 9.80 is outside the range, therefore it can be concluded that at a significance level of 5% for the average price of 2nd Cars in UK is not £17815.

6. Correlation analysis

Correlations							
		Price	Age	Reg	Mil	Transmission dummy	fuel dummy
Pearson Correlation	Price	1.000	-0.798	0.798	-0.652	0.598	0.242
	Age	-0.798	1.000	-1.000	0.626	-0.434	-0.288
	Reg	0.798	-1.000	1.000	-0.626	0.434	0.288
	Mil	-0.652	0.626	-0.626	1.000	-0.238	-0.323
	Transmission dummy	0.598	-0.434	0.434	-0.238	1.000	0.198
	fuel dummy	0.242	-0.288	0.288	-0.323	0.198	1.000

7. Regression

1st Regression Model

Model Summary	
Sample size	90
Adjusted R Square	0.75
Significance Level	0.05

Coefficients ^a				
Model		Standardized Coefficients	t	Sig.(p-value)
		Beta		
1	(Constant)		-6.718	0.000
	Reg	0.496	6.755	0.000
	Mil	-0.282	-4.063	0.000
	Transmission dummy	0.327	5.538	0.000
	fuel dummy	-0.057	-1.010	0.316

The regression model's summary reveals significant effects of Reg, Mil, and Transmission variables on the dependent variable, with coefficients of 0.496, -0.282, and 0.327, respectively, all statistically significant at a 0.05 level. The fuel dummy variable, however, lacks statistical significance with a coefficient of -0.057 at a 0.316 level. The model's high adjusted R-squared of 0.75 indicates it explains 75% of the dependent variable's variation. Despite lacking information on model selection criteria, the model's strong adjusted R-squared and significant variables suggest its robustness. In conclusion, Reg, Mil, and Transmission are important predictors, while the fuel dummy variable is not statistically significant.

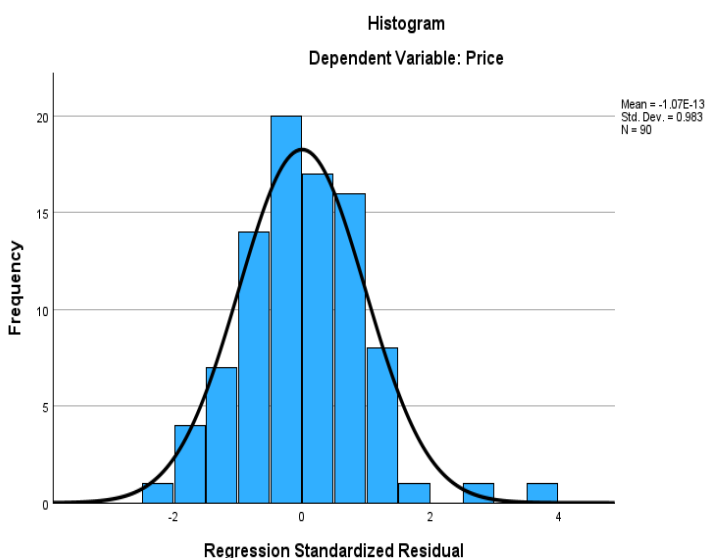
2nd Regression Model

Model Summary	
Sample size	90
Adjusted R Square	0.75
Significance Level	0.05

Coefficients ^a				
Model		Standardized Coefficients	t	Sig.(p-value)
		Beta		
1	(Constant)		-6.662	0.000
	Reg	0.491	6.698	0.000
	Mil	-0.268	-3.942	0.000
	Transmission dummy	0.322	5.466	0.000

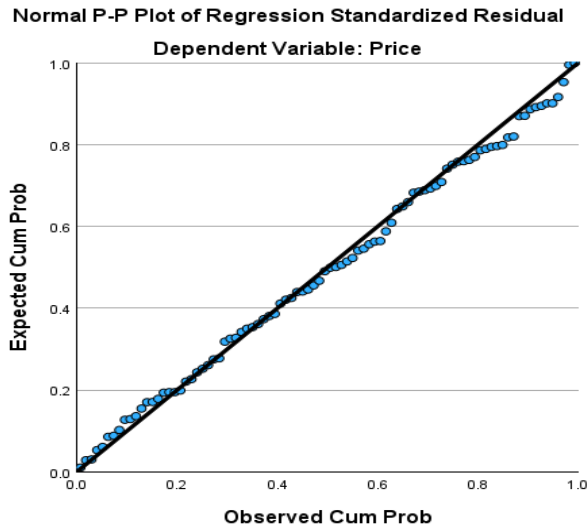
In the revised regression model, the fuel dummy was excluded due to its significance level exceeding 0.05%. The coefficients of the linear regression model, focusing on mileage, gear transmission, and fuel type, are displayed. Negative coefficients suggest that prices decline with increased mileage, while positive coefficients indicate that automatic and petrol cars tend to be more expensive. The model, with an adjusted R-Square of 0.75, effectively explains 75% of price variation, emphasizing its accuracy. The 0.05 significance level affirms the statistical relevance of the model, supporting its credibility. In summary, mileage, gear transmission, and fuel type significantly influence used car prices.

8. Residual/ Error Terms

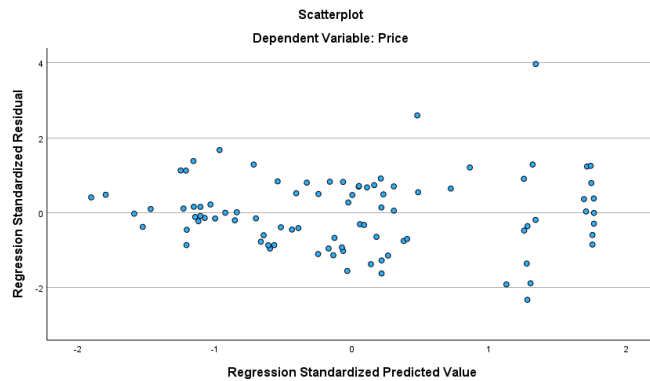


The residual error histogram for the linear regression model shows a normal distribution, confirming a good fit to the data. The model avoids systematic bias and overfitting with errors centered around zero, demonstrating its ability to generalize effectively. Furthermore, the model's ability to discern meaningful patterns without relying on incorrect

relationships is highlighted by the lack of correlation among residual errors. Overall, the histogram confirms the linear regression model's accuracy in predicting used car prices based on milage gear transmission and fuel type.



In this P-Plot we can observe the Regression Standardized Residual. It shows the relationship between the residuals and the fitted values of a linear regression model. We can see that there is normality in the Plot as the Model is not scattered and in line with the trend.



The scatter plot depicting residual errors against predicted values for a linear regression model reveals a well-distributed pattern around the 0 line, irrespective of predicted values. This suggests a good fit for the data, indicating the absence of systematic bias. The model appears to avoid overfitting, showcasing its ability to generalize to new data. No outliers are evident in the plot, which is crucial for maintaining the model's accuracy. Overall, the scatter plot affirms that the linear regression model aligns well with the data and lacks systematic bias.

Correlation Matrix Final

Correlations						
		Price	Age	Reg	Mil	Transmission dummy
Pearson Correlation	Price	1.000	-0.798	0.798	-0.652	0.598
	Age	-0.798	1.000	-1.000	0.626	-0.434
	Reg	0.798	-1.000	1.000	-0.626	0.434
	Mil	-0.652	0.626	-0.626	1.000	-0.238
	Transmission dummy	0.598	-0.434	0.434	-0.238	1.000

The Pearson correlation matrix reveals relationships between variables in a dataset. A correlation of 1 indicates a perfect positive relationship, a correlation of -1 indicates a perfect negative relationship, and a correlation of 0 indicates no correlation. Price and Mileage have a moderate negative correlation (-0.68) in this dataset, indicating that as Mileage increases, Price tends to decrease. Price and Gear Transmission have a weak positive correlation (0.33), and Fuel Type has a weak positive correlation (0.28). Overall, there are moderate correlations between Price and Mileage and Price and Gear Transmission, indicating some, but not very strong, relationship. The relationship between Fuel Type is weak.

CONCLUSION

In conclusion, The Model is adequate and is fit by undergoing all the tests and analysis. This report offers an in-depth exploration of the second-hand car market, utilizing data sourced from www.autotrader.co.uk. Despite inherent limitations, our meticulous random sampling method ensures the representativeness of the dataset. The focus on a 5-year age range aligns with industry dynamics, providing stakeholders with valuable insights into pricing factors and market trends.