



## **PRICE DETERMINANTS FOR A USED CAR**

(STATISTICAL ANALYSIS OF THE SECOND- HAND CAR MARKET)

Candidate No: 220298193  
BNM854 – Descriptive Analytics

## Table of Contents

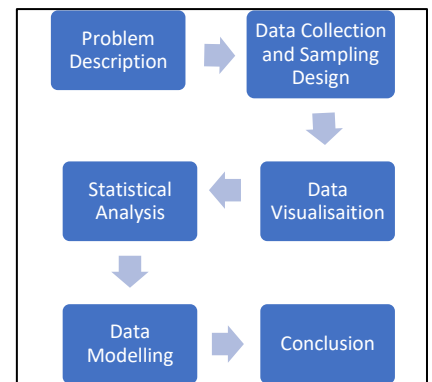
EXECUTIVE SUMMARY .....	3
STATISTICAL ANALYSIS OF THE USED CAR MARKET .....	4
STAGE 1: PROBLEM DESCRIPTION .....	4
STAGE 2: DATA COLLECTION AND SAMPLING DESIGN .....	4
i) SOURCE OF DATA - .....	4
ii) CHOOSING A SAMPLING TECHNIQUE .....	4
iv) JUSTIFICATIONS.....	5
v) LIMITATIONS .....	5
STAGE 3 : DATA VISUALISATION .....	5
PRINCIPLES FOLLOWED IN ALL GRAPHS .....	7
STAGE 4 : STATISTICAL ANALYSIS.....	7
i) DESCRIPTIVE STATISTICS .....	7
a) DESCRIPTIVE STATISTICS FOR ALL THE VARIABLES .....	7
OBSERVATIONS: .....	7
b) DESCRIPTIVE STATISTICS FOR THE PRICE VARIABLE CATEGORISED ACCORDING TO YEAR OF REGISTRATION....	8
OBSERVATIONS .....	8
ii) CONFIDENCE INTERVAL OF THE AVERAGE SECOND-HAND CAR PRICE .....	9
iii) HYPOTHESIS TESTING.....	9
iv) CORRELATION ANALYSIS.....	10
STAGE 5 : DATA MODELLING .....	11
i) REGRESSION ANALYSIS.....	11
a) REASON FOR CHOOSING THE MODEL: .....	12
ii) RESIDUAL ANALYSIS - CHECK FOR ADEQUACY.....	12
a) LINEARITY .....	12
b) HOMOSCEDASTICITY .....	12
c) INDEPENDENCE OF ERRORS.....	12
d) NORMALITY .....	12
e) MULTICOLLINEARITY .....	13
iii) GOODNESS OF FIT .....	13
iv) STATISTICAL MODEL.....	13
CONCLUSION :.....	14
REFERENCES .....	14

## EXECUTIVE SUMMARY

Automotive sector business managers and owners require information on which to make their decisions to maximise company profit. PRICE is one among these components, and it is a key driver in controlling supply and demand. Even though the UK is experiencing unprecedented levels of inflation, used automobile prices have remained strong in 2022.

The aim of the research was to find the most essential characteristics influencing the price of a used automobile, as well as a statistical model that can be used to determine the market worth of a car. The research was conducted using the " **ALL TIME FAVORITE CAR -FORD FIESTA** " as the point of study. The study was conducted for Birmingham's second-hand car market in the B1 2LJ postcode region. Due to the shortage of used Ford Fiesta cars in that area, **the area of interest was expanded by 5 miles**, and the used car market within 5 miles of the B1 2LJ region was investigated. The automobiles' ages were restricted to 5 years.

The research was carried out in six stages with the help of SPSS and Ms-Excel. There were approximately 465 used FORD FIESTA cars on the market, from which data for approximately 100 cars was gathered about their year of registration (age), mile count (number of miles driven), price, fuel type, number of doors, engine capacity, and engine power. The data was cleaned and wrangled in accordance with the specifications.



The graphs asserted that the FORD FIESTA is the most popular used car in the small cars category .

The average price of the automobile was **£ 11152.50**, with the largest number of cars costing **£ 8999**. Most automobiles are driven for **30,000 miles**, and the majority of cars on the market are four years old. It is also worth highlighting that the average price of the used Ford Fiesta car in B1 2LJ region is in line with the average price of car available in UK.

It was affirmed that the main price determinants for used Ford Fiesta cars in the B1 2LJ market are Age Of The Car, Mileage, Engine Power, and Engine Capacity. The price estimator (linear model) can be used to estimate the price of a used Ford Fiesta car as it is explaining 82% of the variation.

# STATISTICAL ANALYSIS OF THE USED CAR MARKET

## STAGE 1: PROBLEM DESCRIPTION

### AIM

The study's primary objective was to analyse the second-hand car market of an area with the post code B1 2LJ to identify

- a) The key characteristics of the car that influence its price.
- b) Develop a statistical model to project a car's market worth in relation to those same key variables.

### CHOICE OF CAR

The Ford Fiesta is the most popular used compact vehicle(Ford Fiesta is the UK's most popular used car , Haynes Publishing ,2022)in the UK. Additionally, it has received the Parkers Best Used Car of the Year title for the past two years(K.Adams,2021). Also, Ford fiesta is the frontrunner in offering a number of choices to its customers. Therefore, it was beneficial to investigate the factors that affect this car's price.

The study was conducted for Birmingham's second-hand car market in the B1 2LJ postcode region. Due to the shortage of used Ford Fiesta cars in that area, **the area of interest was expanded by 5 miles**, and the used car market within 5 miles of the B1 2LJ region was investigated. The automobiles' ages were restricted to 5 years.

## STAGE 2: DATA COLLECTION AND SAMPLING DESIGN.

### i) SOURCE OF DATA -

The data was collected from [www.autotrader.co.uk](http://www.autotrader.co.uk) , the website of one of the UK's largest digital and reliable automotive marketplaces for new and used cars.

### ii) CHOOSING A SAMPLING TECHNIQUE

There were around 465 cars with age range not exceeding 5 years in the market and were taken as the population (a group of related objects or occurrences that are relevant to a certain experiment or topic) out of which a sample of 100 cars (a subset of individuals from within the population of 400 cars) was selected using **SIMPLE RANDOM SAMPLING** (SRS) technique. SRS is a type of **Probability Sampling** where each unit of the population has an equal likelihood of being selected.

SRS not only minimizes bias and simplifies analysis of results but also allows to draw conclusions regarding the population from the results of a sample. However, Simple random sampling tends to have larger sampling errors and may not produce enough elements from small subgroups. (Thomas,2018)

### iii) SAMPLE - REPRESENTATIVE OF THE POPULATION

The 100-car sample was chosen at random in this case using the random number generator approach. Since the SRS approach ensures that the sample is an accurate representation of the population by **removing bias from the selection process**, the resulting random sample is a " **MINI VERSION OF THE**

**POPULATION"**. Furthermore, the **sample size is substantial enough** in this case to allow the sample to possibly reflect the characteristics of the population.

#### iv) JUSTIFICATIONS

The cars should not be more than five years old. This is justified as given that an automobile loses more than half of its original value in the first five years of ownership, this restriction seems sensible. In addition, the older a car is, the more likely it is that it does not have some of the wonderful technological and safety features that are now standard in modern vehicles.

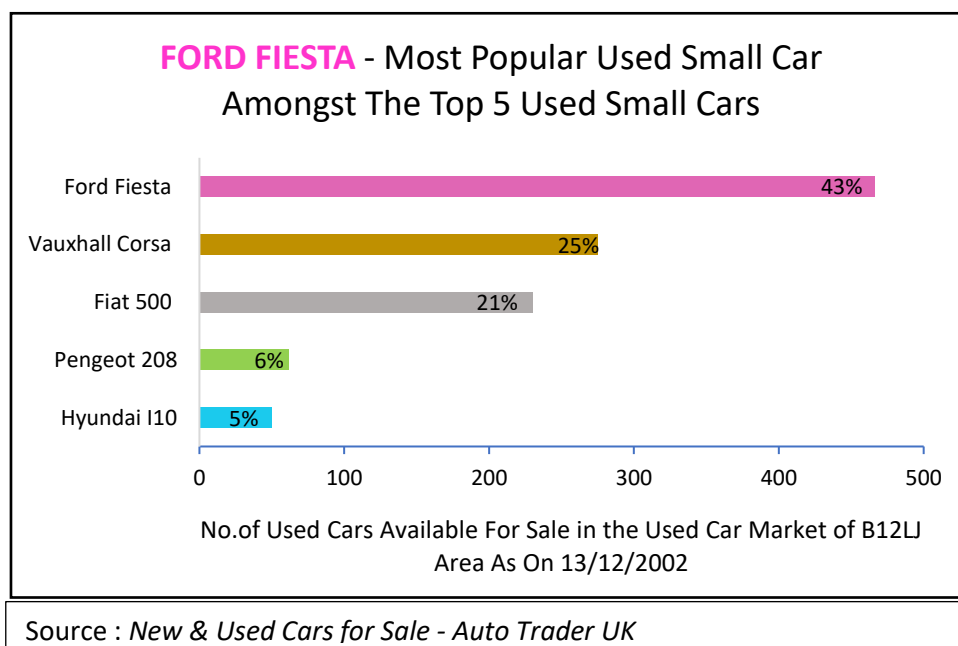
#### v) LIMITATIONS

- a) The sample had outliers i.e. it had some unusual values which could have distorted the results of the analysis. Thus, the outliers were removed from the data set.
- b) There were only 7 cars with age of 0 years in the population of 465 cars. As the number was too small, no car with age of 0 years appeared in the sample

### STAGE 3 : DATA VISUALISATION

The graphical display is an easy method to observe and comprehend trends and helps in comparing the data sets.

#### i) GRAPH 1

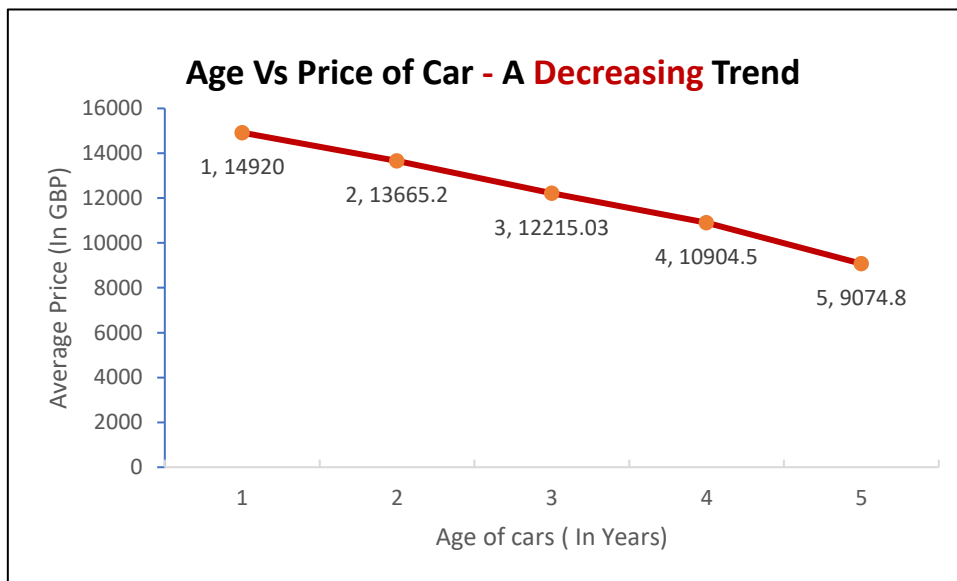


#### JUSTIFICATION

Here, a clustered bar graph has been considered a good option to display the breakdown of the car types in the region as there are a small number of categories.

ANALYSIS The above stacked bar graph demonstrates the **FORD FIESTA's** popularity in the B1 2LJ area's used car market. The number of vehicles available for sale in the used car market can be used to support the popularity. About 43% of the small vehicles for sale amongst the top 5 used small cars are Ford Fiestas. Whereas I10 doesn't appear to be a danger to **FORD FIESTA's** appeal, Vauxhall is just slightly behind **FORD FIESTA** in terms of popularity. Additionally, it is noted that there are around 450 **FORD FIESTA** used cars for sale, providing the customers with a wide range of purchasing options.

## ii) GRAPH 2

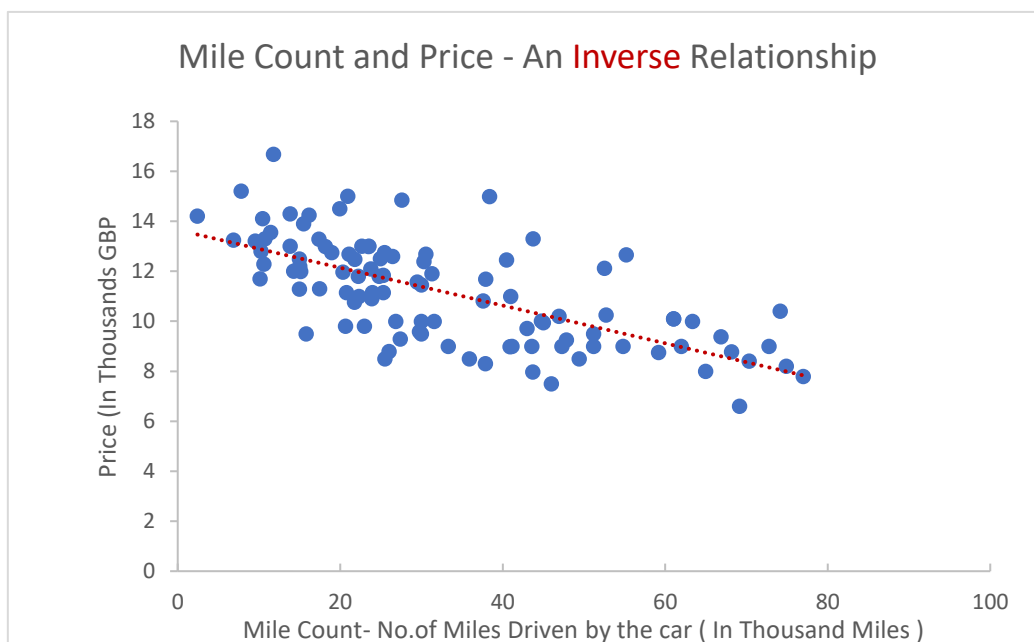


**JUSTIFICATION** : The research's goal is to identify the factors that influence pricing. Age of a used car is one of the variables that might impact price. In order to see the link between the age of the automobile and its average price during that Time Period, a line graph is created. Since we have to analyze the link

between the time component and the dependent variable, a line graph is the most appropriate choice. Following the conventions for making a line graph, the age of the car has been taken on the x-axis and the dependent variable on the y-axis.

**ANALYSIS** : The relationship between price and car age is clearly demonstrated by the line graph, indicating that it is a determining factor. Additionally, it is shown that age and average price have an inversely proportionate relationship, meaning that as a car gets older, its price goes down. Observing the trend line, it can be asserted that between the ages of 4 and 5, there is the greatest price depreciation.

## iii) GRAPH 3



**JUSTIFICATION** : The other factor significantly affecting the price can be the no. of miles driven by the car. Given that we must compare a huge number of data points without respect to time in this situation, a scatter plot will be the

most appropriate for analysing the connection between these two variables.

**ANALYSIS:** Here, the scatter plot gives a fair idea of the relationship between no. of miles driven by the car and price. It is clearly evident that there is a negative linear relationship between the no. of miles driven and the price of the car. It implies that as the no. of miles driven by the car will increase, the price will decrease. Also, it can be observed that the maximum no of cars available in the used car market have covered less than 50,000 miles.

### **PRINCIPLES FOLLOWED IN ALL GRAPHS**

Tufte's Principles of Graphical Excellence is being followed in all the above graphs by showing most ideas in least amount of ink , by a clear labelling of the axes and by mentioning the source of data. Moreover, the graphs are data driven, scales are standardised and there is context in all the graphs. The axes have been labelled and gridlines have been removed. However, ticks are used to give more precision to the graphs. Gestalt's principles have been followed through similarity of colour between the title and data it refers to.

### **STAGE 4 : STATISTICAL ANALYSIS**

#### **i) DESCRIPTIVE STATISTICS**

The key features of a data set can easily be described using descriptive statistics.

#### **a) DESCRIPTIVE STATISTICS FOR ALL THE VARIABLES**

<b>TABLE 1-DESCRIPTIVE STATISTICS</b>					
<b>VARIABLES</b>	<b>Price( £)</b>	<b>Miles Driven ( Miles)</b>	<b>Age of Car (Years)</b>	<b>Engine Capacity (Litres)</b>	<b>Engine Power (BHP)</b>
Mean	11152.50	33088.49	3.72		100.07
Quartile 1	9439.00	20150.00	3		84
Median	11150.00	27400.00	4	1	99
Quartile 3	12688.00	44886.00	5		99
Mode	8999.00	30000.00	4	1	99
Standard Deviation	2067.20	18448.39	1.03		17.18
Sample Variance	4273304.92	340342936.53	1.05		295.16
Range	10083.00	74574	4.00		57.00
Minimum	6595.00	2426	1	1	81
Maximum	16678.00	77000	5	1.5	138
Count	103	103	103	103	103.00

#### **OBSERVATIONS:**

#### **i) PRICE**

- The price of a used Ford Fiesta car with age range of 5 years is around £11152(**Mean**) where most cars are available for £8999(**Mode**). The minimum and maximum price of the car available in the market is £6595 and £16678 respectively.
- Almost 25 % of the Ford Fiesta Used Cars are having prices below the £9439(**Quartile 1**) price mark with 50% of the cars being priced below the price mark of £11150(**Median**)and the rest 50% will be priced higher than this price mark.

- Anyone with a budget between £9085 and £13215 (approximately) will be able to buy a Ford Fiesta used car.(11152  $\pm$  2067).

## ii) MILES DRIVEN

- A used Ford Fiesta car available in the market has been driven for around 33088 miles (Mean) where most cars are driven for 30,000 miles (Mode). The miles covered by cars are having an extensive range of 74574 miles with minimum of 2426 miles with maximum touching 77,000 miles.
- Almost three-fourth of the Ford Fiesta used cars which are available in the market have been driven for less than 50,000 miles.(Quartile 3)
- There is a **good spread in the average number of miles driven** as shown by the **standard deviation**.

## iii) OTHERS

- The maximum no. of cars available in the market has an age of 4 years .
- Similarly, the cars with an engine capacity of 1 litre are highly available in the market.

## b) DESCRIPTIVE STATISTICS FOR THE PRICE VARIABLE CATEGORISED ACCORDING TO YEAR OF REGISTRATION

TABLE 2 -DESCRIPTIVE STATISTICS OF PRICE CATEGORISED ACCORDING TO YEAR OF REGISTRATION					
MEASURES \ YEAR	2017	2018	2019	2020	2021
Mean	9074.48	10904.50	12215.03	13665.20	14920
Median	8999.00	10249.50	12240.00	13727.00	14920
Mode	8999.00	10100.00	12990.00	#N/A	#N/A
Standard Deviation	1071.90	1875.94	1156.47	882.02	98.9949494
Range	5204	9188.00	5211.00	3000.00	140
Minimum	6595	7490.00	9989.00	11999.00	14850
Maximum	11799	16678.00	15200.00	14999.00	14990
Count	27	34	30	10	2

(Note: Here, the continuous variable price is distributed according to the categorical variable Year of Make)

## OBSERVATIONS:

- The maximum no of cars available in the market are 2018 registered (34) with the least availability of comparatively newer cars (2).
- The average price of the cars are increasing as their age is decreasing i.e. a 2017 registered is relatively cheaper than a 2018 car.
- There is **not much spread in the average prices (Standard Deviation)** in the used car prices of 2020 registered cars whereas the used cars available in the market which were registered in 2018 are having the maximum variation in their prices.



## ii) CONFIDENCE INTERVAL OF THE AVERAGE SECOND-HAND CAR PRICE

Here, confidence interval will give us an interval estimate in which the average price of a second-hand Ford Fiesta will lie in the second hand car market of B1 2LJ.

**95 % CONFIDENCE INTERVAL:** Average Price  $\pm 1.96 * \text{Standard Deviation}$

$$= 11152.50 \pm (1.96 * 2067.20) \text{ (Refer to Table 1 for values)}$$

$$= [10753.280, 11551.720]$$

$$= 10753.280 \leq \text{Average Price} \leq 11551.720$$

**CONCLUSION :** It means that it can be said with **95% confidence** that the average price of a second-hand Ford Fiesta car in the B1 2LJ second hand car market will be between **£10,753.280 and £11,551.720**. There is only **5% chance that the average price will lie outside the interval**. In other words , it is 95 % sure that any person who has an average **budget** between **£10,753.280 and £11,551.720 will be able to buy a second-hand Ford Fiesta car**.

**99 % CONFIDENCE INTERVAL:** Average Price  $\pm 2.58 * \text{Standard Deviation}$

$$= 11152.50 \pm (2.58 * 2067.20) \text{ (Refer to Table 1 for values)}$$

$$= [10626.98, 11678.02]$$

$$= 10626.980 \leq \text{Average Price} \leq 11678.020$$

**CONCLUSION :** It means that it can be said with **99% confidence** that the average price of a second-hand Ford Fiesta car in the B1 2LJ second hand car market will be between **£10,626.980 and £11,678.020**. There is only **1% chance that the average price lies outside the interval**. In other words , it is 99 % sure that any person who has an average **budget** between **£10,626.980 and £11,678.020 will be able to buy a second-hand Ford Fiesta car**.

## iii) HYPOTHESIS TESTING

Hypothesis Testing is about putting forward an idea in the form of a testable proposition that captures the properties of the problem that we are interested in. To test whether the average price of the Second-hand **Ford Fiesta used petrol cars with 1.1 litre engine capacity** car in B1 2LJ car market is not different from the average price of the selected Ford Fiesta car in the UK market.

**Average Price in UK = £11,900** ( <https://priceanycar.com/Ford/Fiesta/2018/?q=1.0+Petrol>)

**Average Price in B1 2LJ = £11,445.50 ( n = 22)**

Here, Null Hypothesis = The average price of the selected model of the car in B12LJ is in line with the average price of the car available in the market(  $H_0 : \mu = 11900$ )

Alternative Hypothesis = The average price of the selected model of the car in B12LJ is in line with the average price of the car available in the market( $H_a: \mu \neq 11900$ )

One sample t-test will be conducted where the test is used to determine the one-sample t-test is used to examine if an unknown population mean differs from a given value. It will be conducted at conducted at 5

% level of significance. If the calculated p value will be greater than 0.05 , we will accept the null hypothesis and if it is smaller than 0.05, we will reject the null hypothesis.

Here, p-value > 0.05. So, we accept the null hypothesis and can say with 95% confidence that the average price of the car in B1 2LJ is in line with the average price of the car available in the market.

One-Sample Test							
Test Value = 11900							
	t	df	Significance		Mean Difference	95% Confidence Interval of the Difference	
			One-Sided p	Two-Sided p		Lower	Upper
V1	-1.171	21	.127	.255	-454.500	-1261.96	352.96

iv) CORRELATION ANALYSIS

Correlation Analysis is used to discover if there is a relationship between two variables/datasets, and how strong that relationship may be. Here, price is a dependent variable, and the independent variables are age of car, mile count, engine power, no of doors, fuel type and engine capacity. 4 dummy variables are created , **1.1 L , 1.2 L, 1.3 L and 1.5 L for engine capacity** with no dummy variable for 1 L as there are maximum no of cars for this engine capacity.

Table 3 -Correlation Matrix										
	Price (GBP)	Age of Car(Years)	Fuel Type	No Of Doors	Mile Count(Miles)	Engine power	1.1 L	1.2 L	1.3 L	1.5 L
Price	1	-0.745	0.128	0.242	-0.674	0.29	-0.146	-0.22	-0.31	-0.128
Age of Car	-0.75	1	-0.109	-0.255	0.48	-0.104	-0.097	0.125	0.34	0.109
Fuel Type	0.128	-0.109	1	-0.175	-0.28	0.079	0.126	0.025	0.067	-1
No Of Doors	0.242	-0.255	-0.175	1	-0.133	-0.158	0.048	-0.141	-0.057	0.175
Mile Count	-0.67	0.48	-0.28	-0.133	1	0.089	-0.23	0.195	-0.021	0.28
Engine power	0.29	-0.104	0.079	-0.158	0.089	1	-0.41	-0.093	-0.25	-0.079
1.1 L	-0.146	-0.097	0.126	0.048	-0.234	-0.405	1	-0.05	-0.137	-0.126
1.2 L	-0.22	0.125	0.025	-0.141	0.195	-0.093	-0.05	1	-0.027	-0.025
1.3 L	-0.31	0.34	0.067	-0.057	-0.021	-0.247	-0.137	-0.027	1	-0.067
1.5 L	-0.128	0.109	-1	0.175	0.28	-0.079	-0.126	-0.025	-0.067	1

It can be observed that all the variables are correlated to each other with varying levels of correlation. The independent variables ; **age of car and mile count** are **negatively corelated** with **price** meaning that with increase in any of these variables will decrease the price of the car and vice-versa. The **strongest negative** correlation is observed between Price and Age of Car.

Also, the variables **fuel type, no.of doors and engine power** are **positively corelated** with price i.e. meaning that increase in any of these variables will increase the price of the car and vice-versa. Among these 3 variables, the **strongest positive** correlation is between Engine Power and Price.

Fuel type and price are the variables with **least** correlation. Also, it can be observed that price and 1.1 L engine capacity are the most correlated. A strong negative correlation between **1.5 L and Fuel Type** is also highlighted here which gives a hint to multicollinearity. Except for these 2 variables, the other independent variables are not strongly correlated.

## **STAGE 5 : DATA MODELLING**

### **i) REGRESSION ANALYSIS**

Regression analysis is a process for estimating the relationships between a dependent variable (Variable which depends on the other variable) and one or more independent variables (Variables which are not dependent on any other variable, often called as Predictors).

After executing the regression function in SPSS, this is the first model is received.

<b>Coefficients<sup>a</sup></b>									
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	14659.548	793.754		18.469	0.000			
	Age of Car	-828.529	106.784	-0.410	-7.759	0.000	-0.745	-0.625	-0.319
	Mile Count	-0.061	0.006	-0.544	-10.523	0.000	-0.674	-0.735	-0.433
	Engine	17.495	6.031	0.145	2.901	0.005	0.290	0.287	0.119
	1.1 L	-1446.659	249.436	-0.283	-5.800	0.000	-0.146	-0.513	-0.239
	1.3 L	-1460.193	394.155	-0.179	-3.705	0.000	-0.308	-0.357	-0.153
	NoOfDoors	352.234	195.596	0.081	1.801	0.075	0.242	0.183	0.074
	1.2 L	-1176.028	903.790	-0.056	-1.301	0.196	-0.219	-0.133	-0.054
	1.5 L	155.346	394.327	0.018	0.394	0.695	-0.128	0.041	0.016

a. Dependent Variable: Price

This is not a **Parsimonious Model** as the sig values of **No of Doors, 1.2 L and 1.5 L** are greater than **0.05** which makes their coefficient insignificant. Thus, the model will be re estimated again after removing these variables from the model. Fuel type will also be included this time as it was earlier excluded from the model.

On re-estimation, it was found that the variable fuel, type is also insignificant. Thus, the model was re-estimated again by removing the fuel type and the **MOST PARSIMONIOUS** model was derived.

<b>Coefficients<sup>a</sup></b>									
Model		Unstandardized		Standardized	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-	Partial	Part
1	(Constant)	15255.947	721.937		21.132	0.000			
	Age of Car	-879.863	105.726	-0.436	-8.322	0.000	-0.745	-0.645	-0.349
	Mile Count	-0.061	0.006	-0.548	-10.835	0.000	-0.674	-0.740	-0.454
	Engine power	15.975	5.859	0.133	2.727	0.008	0.290	0.267	0.114
	1.1 L	-1467.444	249.673	-0.287	-5.877	0.000	-0.146	-0.512	-0.246
	1.5 L	-1455.061	396.341	-0.178	-3.671	0.000	-0.308	-0.349	-0.154

a. Dependent Variable: Price

a) REASON FOR CHOOSING THE MODEL:

Here, as all the sig. values are **less than 0.05**, which makes us reject the proposition that the co-efficient is equal to 0 and asserts that the variables Age of Car, Mile Count, Engine power, E1(1.1 L) and E3(1.3 L) exerts a significant impact on Price.

ii) RESIDUAL ANALYSIS - CHECK FOR ADEQUACY

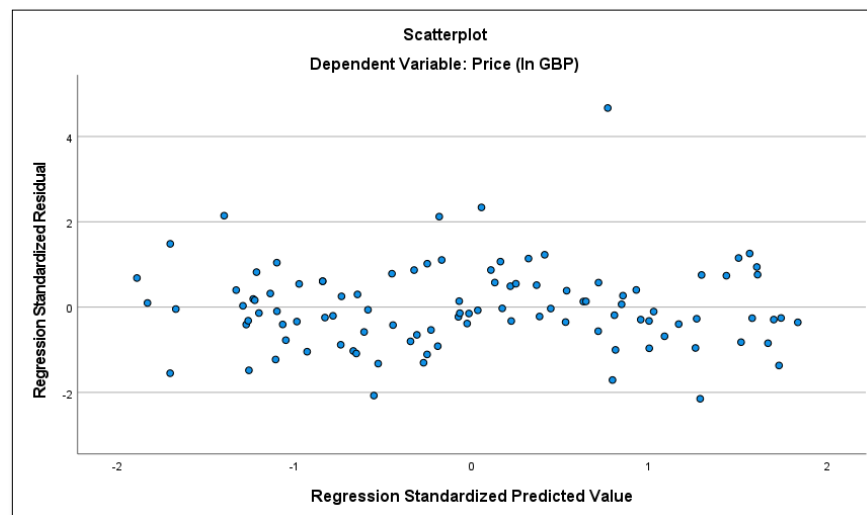
To check that whether the model is adequate i.e. can be used for **estimation process** or not we need to check the five regression assumptions. It can be inferred by studying the graphs.

a) LINEARITY

Here, the above scatter diagram has the same number of residuals above 0 and the same below 0, thus the assumption of **linearity holds true** as the residuals are scattered around zero making their average equal to 0.

b) HOMOSCEDASTICITY

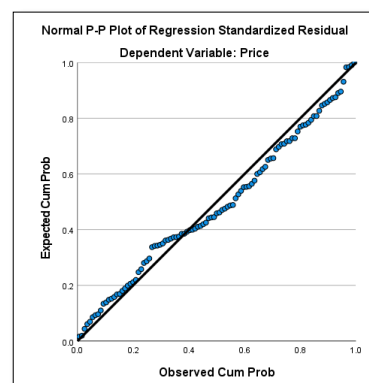
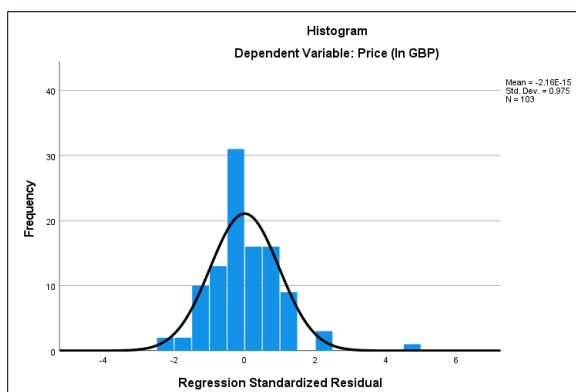
From the above scatter diagram, it can be observed that there is no clear pattern of the residuals increasing or decreasing with an increase of the estimated values of the dependent variable. Thus, the assumption of **homoscedasticity holds true** as the standard deviation of the residuals ( $\epsilon$ ) is the same for all values of the estimated dependent variable.



c) INDEPENDENCE OF ERRORS

From the above scatter diagram, it can be observed that it is a random one i.e. there is no clear pattern in the residuals in any specific form. Thus, the assumption of independence of errors holds true as the residuals are independent.

d) NORMALITY



Here, the histogram of residuals appears to be normally distributed and the residuals in the normal P-P plot are close to the line of the perfect normal distribution. Thus, it can be inferred that the residuals are normally distributed, and the assumption of linearity holds true.

As, all the assumptions hold true, it can be inferred that **model is adequate** and can be used for **price estimation** purposes.

### e) MULTICOLLINEARITY

Here, from the correlation matrix, it was observed that there is a strong correlation between 2 independent variables E4 and Fuel Type. But as both the independent variables are not a part of parsimonious model, it is not of a concern anymore. Also, there are **no 2 independent variables** which have a **correlation coefficient greater than 0.7 and less than 0.7**.Thus, there is **no multicollinearity** .

### iii) GOODNESS OF FIT

Here, the **Adjusted R Square = 0.821** , which makes the **model a good fit to the data**. It indicates that **82% of the variation in the price is dependent on Mile Count, Age of Car , Engine Power and on the Engine Capacity(1.1 L and 1.3 L)**

Model Summary <sup>b</sup>				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.911 <sup>a</sup>	0.830	0.821	874.854
a. Predictors: (Constant), 1.3L, Mile Count, Engine power, 1.1 L, Age of Car				
b. Dependent Variable: Price				

### iv) STATISTICAL MODEL

The price of a second-hand Ford Fiesta car can be estimated by

$$\text{Price} = 15255.947 - 879.863 * \text{Age of Car} - 0.061 * \text{Mile Count} + 15.975 * \text{Engine Power} - 1467.444 \text{ E1} - 1455.061 \text{ E3} + e$$

For a car with Engine Capacity 1.1 L (E1 =1 , E3 =0)

$$\text{Price} = 15255.947 - 879.863 * \text{Age of Car} - 0.061 * \text{Mile Count} + 15.975 * \text{Engine Power} - 1467.444 + e$$

For a car with Engine Capacity 1.3 L (E1 =0 , E3 =1)

$$\text{Price} = 15255.947 - 879.863 * \text{Age of Car} - 0.061 * \text{Mile Count} + 15.975 * \text{Engine Power} - 1455.061 + e$$

For a car with Engine Capacity 1 L , 1.2 L and 1.5 L (E1 = E3 = 0 )

$$\text{Price} = 15255.947 - 879.863 * \text{Age of Car} - 0.061 * \text{Mile Count} + 15.975 * \text{Engine Power} + e$$

#### a) MAGNITUDE OF THE EFFECT OF THE INDEPENDENT VARIABLES

It can be inferred that, with an increase in the age of a car by 1 year, the price will fall by 879.863 £ .  
 The price of the car will decrease by 0.061 £ ,with a unit increase of mile driven. If the engine power increases by 1 HP ,then price will increase by £15.975.

#### b) STANDARDISED BETA COEFFIECIENTS

A standardised beta coefficient measures how strongly each independent variable has an impact on the dependent variable. The effect is stronger the higher the beta coefficient's absolute value. Here, the variable,

age of car will have the maximum effect on price followed by mile count , 1.1 L , 1.3 L and at then engine power.

### c) PREDICTING THE PRICE OF THE CAR

Estimating the price of a used car with age = 4 years, miles driven = 45,000 miles , Engine power = 84 BHP and Engine Capacity = 1 L .

#### Using the equation

$$\begin{aligned}\text{Price} &= 15255.947 - 879.863 * \text{Age of Car} - 0.061 * \text{Mile Count} + 15.975 * \text{Engine Power} + e \\ &= 15255.947 - 879.863 * 4 - 0.061 * 45000 + 15.975 * 84 \\ &= \text{£ } 10333.40\end{aligned}$$

#### CONCLUSION :

Age of car, Mile Count , Engine Capacity and Engine power are the main factors that affect the price of a used car. The statistical model so obtained can be used to determine the price of a used car as it is an adequate model and a good fit explaining 82% of the variation in the price of a car. However, there are some other factors as well affecting the price of the car and needs to be determined.

#### REFERENCES

- Adams, K. (2021) *Used car of the year 2022: Ford Fiesta (2008-2017)*, Parkers. Parkers. Available at: <https://www.parkers.co.uk/car-awards/2022/used-car-of-the-year/#:~:text=The%20Ford%20Fiesta%20wins%20the,many%20good%20reasons%20for%20this.> (Accessed: December 14, 2022).
- *Ford Fiesta is the UK's most popular used car* (2022) Haynes Publishing. Available at: <https://haynes.com/en-gb/tips-tutorials/ford-fiesta-uk-s-most-popular-used-car> (Accessed: December 14, 2022).
- *1,372 cars for sale* (no date) *New & Used Cars for Sale - Auto Trader UK*. Available at: <https://www.autotrader.co.uk/car-search?postcode=b12lj&radius=5&make=Ford&include-delivery-option=on&year-from=2017&year-to=2022> (Accessed: December 14, 2022).
- Thomas, G. (2018) *LORECENTRAL*. Available at: <https://www.lorecentral.org/2017/11/advantages-disadvantages-random-sampling.html> (Accessed: December 4, 2022).