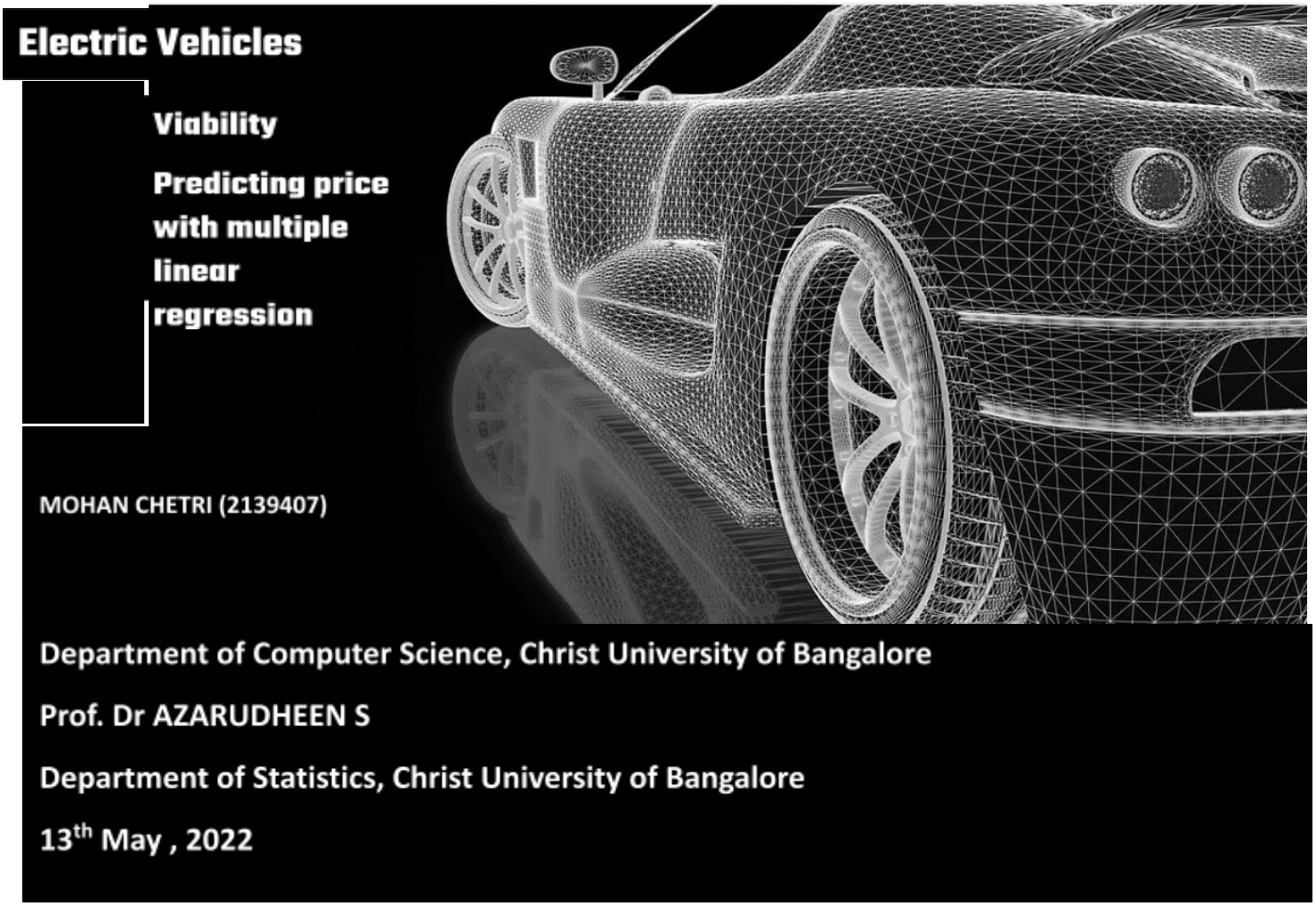


ABSTRACT

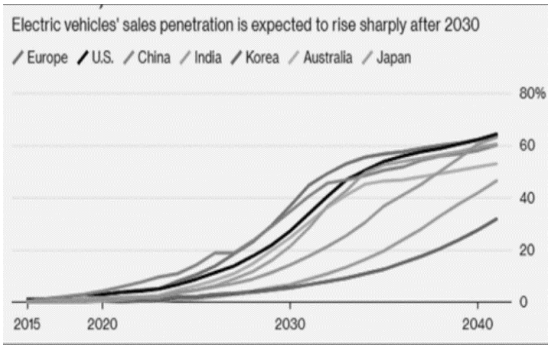
Environmental pollution is currently a worldwide problem. Toxic emissions from internal combustion engines are a major source of air pollution. Electric vehicles (EVs) are being aggressively promoted throughout the world to mitigate the effects of fossil fuel emissions and to address environmental concerns (ECS). Financial incentives are being offered by several governments to encourage people to switch to electric automobiles. According to prior studies, customer acceptability is hampered by the high cost of an electric car, the absence of charging infrastructure, and time and range concern. The Indian government hopes to have "entirely Electric Vehicles" on the roads by 2030. This article is current, and it looks at the various aspects that influence the price of an electric vehicle and a multiple linear regression. regression model to estimate the price of the vehicle and confirming it with the help of residual analysis, as well as its viability based on cost.

Keywords: EVs , ECS, viability, multiple linear regression, Residual Analysis .

INTRODUCTION

The automobile industry, which has existed for over a century, is experiencing a transformation. Individual mobility patterns have shifted as the price of fossil fuels has risen, as has the environmental effect of their emissions. Internal combustion engines have propelled the industry forward, but it is gradually turning to electric vehicles (EVs). Electric motors propel EVs, which are powered by a rechargeable battery or other portable energy storage device. These vehicles utilise less energy, generate fewer greenhouse gases (GHGs), and produce less noise.

Transportation emits over a quarter of all greenhouse gas emissions. Automobiles are the world's greatest source of greenhouse gas emissions, accounting for 25.9% in China, 13.87 percent in the US, and 13.87 percent in India. 7.45% of the population In India, the transition to electric vehicles is necessary but not urgent. Unplanned growth and heavy pollution are plaguing some cities. They are severely degraded, with traffic emissions being the principal cause.



But the real question is whether it would be viable all over the world ,Can we predict the price with good accuracy. Graph that shows the EV projections for a few of the major countries.

NEED OF STUDY

Although electric vehicles are healthier for the environment than gasoline-powered automobiles, many individuals are still hesitant to purchase them. One factor is the higher initial investment. As a result, most people found it impractical to purchase one. This study will look for factors that influence the price, as well as recommend viable options based on certain vehicle characteristics. As we all know, despite their higher sticker price, electric cars can save tons of money in the long run.

LITERATURE REVIEW

Many research have been carried out all around the world in order to better understand customer acceptability of electric cars. The price of the electric car, the driving range per charge, the time it takes to recharge the battery, the number of charging stations available, and the cost of the battery were all taken into account. Previous research has looked at range anxiety and the financial and non-financial incentives offered. Technology refers to the technological features of a vehicle. Previous study has identified range anxiety, or the little distance travelled per full charge of the battery, as a major hurdle to EV adoption. Infrastructure characteristics are concerned with the available charging infrastructure. In research, this ingredient has been found to have a positive influence. Appropriate charging facilities will save time and money during the search process, reducing buyers' concerns about range.

METHODOLOGY

There are 180 automobiles in the database, and some values are missing. To better comprehend the data and information, we do numerous data analytics visualisations, statistical tests, and hypothesis tests. This offers us a sense of the market as a whole and the distribution of data in general. This sort of information frequently provides market insight and enables companies to undertake market research. With sufficient investment and research, the electric car market may be enhanced and a better degree of efficiency obtained. Hypothesis testing and linear regression fitting on various numerical attributes are other variables influencing the pricing of electronic autos. Finally, all of the insights and essential points will be summarised to assure feasibility. Additionally, Hypothesis testing and fitting linear regression on some numerical attributes being factors affecting the price of Electronic cars. Eventually, we will be concluding with all the insights and key points to ensure feasibility. Residual analysis will be a major factor for confirming the model.

DEMOGRAPHICS AND DATA UNDERSTANDING

Column and its description

Column Nam	Description
Brands	Here it gives the different kinds of EV manufacture there are and the name of the different brands.
Model	It gives the exact model of the EV cars provided by each brands.
AccelSec	The minimum time it takes to reach 100 km in seconds.
TopSpeed_K	The top speed of the different models of the EV Cars.
Range_Km	Total range provided in 100% charge of the battery in optimal condition.
Efficiency_W	The maximum productivity it can give in minimum wasted charge.
FastCharge_	The charging speed of the car in Km/h.
RapidCharge	The availability of being able to fully charge the car in minimum number of hours.
PowerTrain	The type of storing and supplying charge of the car.
PlugType	The type of Plugin used for charging The car.
BodyStyle	The style of the car.
Segment	The classification of the car based on its body style.
Seats	The total number of seats inside the car.
PriceEuro	The Price of buying the model of the car in Euros.

Numerical attributes overview

	AccelSec	TopSpeed	Range_Km	Effeciency	FastCharge	PriceEuro
Min.	2.1	123	95	104	170	20129
Max.	22.4	410	970	273	940	215000

Brand wise overview on features

	AccelSec		TopSpeed		Range_Km		Effeciency		FastCharge		PriceEuro	
	Brand	Model	Brand	Model	Brand	Model	Brand	Model	Brand	Model	Brand	Model
Min	Tesla	Roadstar	Nissan	e-NV200 E	Smart	EQ forfour	Lightyear	One	Volkswagen	e-Up!	SEAT	Mii Electric
Max	Renault	Kangoo Maxi Z	Tesla	Roadstar	Tesla	Roadster	Mercedes	EQV 300 Lo	Tesla	Model 3 Long Range D	Tesla	Roadster

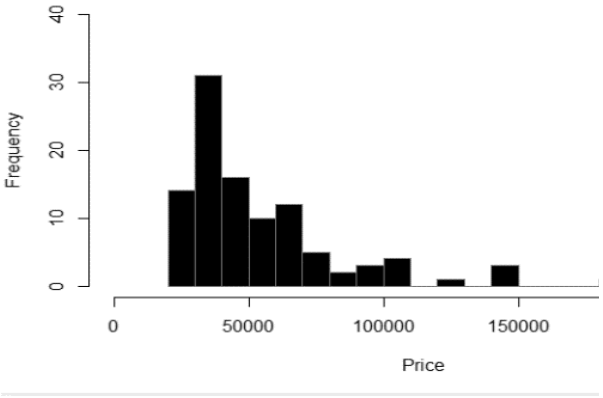
The data collection includes 33 manufacturers worldwide, over 100 models, nine various body designs, over six different types of segments and seat distribution, four different plug types, and three different powertrains. Data sample is distributed over different segments and is not biased so it’s a good fit for our study.

Glimpse of the dataset

	Brand	Model	AccelSec	TopSpeed_KmH	Range_Km	Efficiency_WhKm	FastCharge_KmH	RapidCharge	PowerTrain	PlugType	BodyStyle	Segment	Seats	PriceEuro
1	Tesla	Model 3 Long Range Dual Motor	4.6	233	450	161	940	Yes	AWD	Type 2 CCS	Sedan	D	5	55480
2	Volkswagen	ID.3 Pure	10.0	160	270	167	250	Yes	RWD	Type 2 CCS	Hatchback	C	5	30000
3	Polestar	2	4.7	210	400	181	620	Yes	AWD	Type 2 CCS	Liftback	D	5	56440
4	BMW	iX3	6.8	180	360	206	560	Yes	RWD	Type 2 CCS	SUV	D	5	68040

QUICK INSIGHTS.

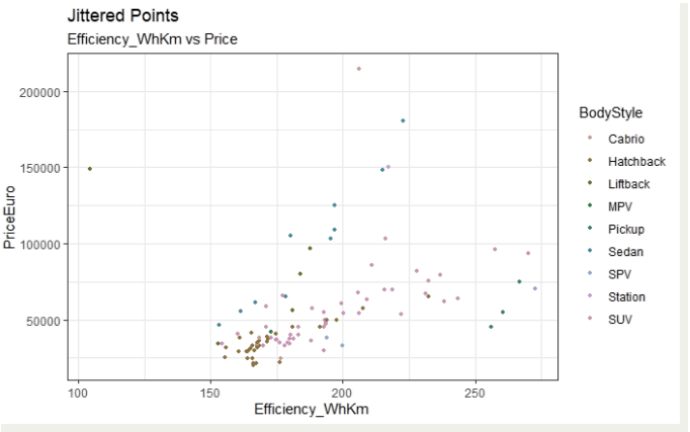
PRICING OF EVs



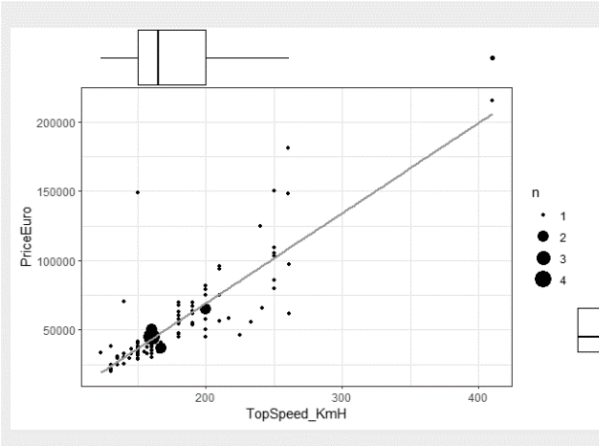
- Most electric cars are in the 45k euro range. Additionally, there are some cars that are very expensive, with prices that reach over 1l euro.

EFFICIENCY | PRICING WITH DIFFERENT BODY STYLES

- While the prices of SUV and SPV e-cars are evenly spread, others tend to cluster, which says SUV and SPV are available in different price range while other are available around the same range along with its body style.

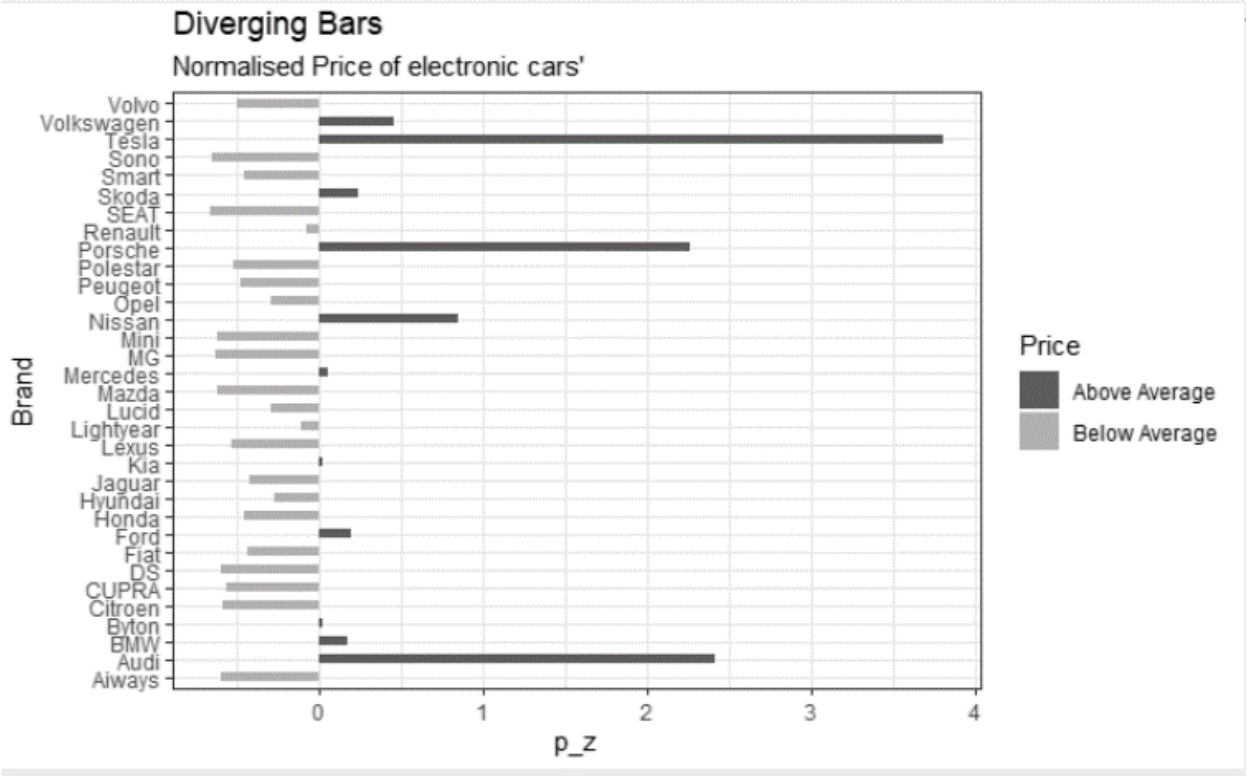


TOP SPEED | PRICE



- Topspeed has a positive relationship with price, Also indicating that our pricing contains some outliers.

BRAND WISE PRICING..



- After normalising the price with its mean and standard deviation, Divided it into two categories: below average and above average pricing.
- Tesla, Audi, and Porsche all have prices that are significantly higher than the norm.
- Many brands have price that is comparably inexpensive and affordable, and others are in the moderate level as we can interpret from the graph.

APPROACH TO MODELLING

Multiple linear regression (MLR), often known as multiple regression, is a statistical approach that predicts the result of a response variable by combining numerous explanatory factors. Multiple regression is a variant of linear (OLS) regression in which just one explanatory variable is used.

$$y = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n + \varepsilon$$

y = the dependent variable's anticipated value

B0 represents the y-intercept (value of y when all other parameters are set to 0)

B1X1= the first independent variable's (X1) regression coefficient (B1) (a.k.a. the effect that increasing the value of the independent variable has on the predicted y value)

... = repeat for as many independent variables as you want to test.

BnXn = the final independent variable's regression coefficient

e = model error (a.k.a. how much variation there is in our estimate of y)

We will develop a model to predict the price of an EV given its features and narrow down the significant factors. As there are so many variables, we will use a step-by-step technique to narrow down the key attributes with the lowest AIC. We'll build a model to estimate the price of an EV given its variables and narrow down the significant variables. Because there are so many variables, we'll take a step-forward approach, narrowing down the significant attributes with the lowest AIC and checking the assumptions of the linear model. The model should at least explain 75 percent of variability.

Initial model

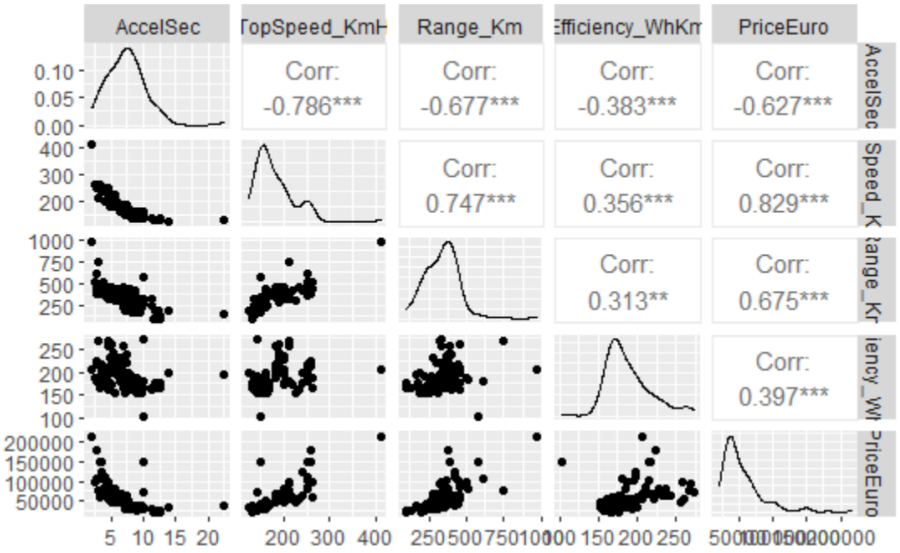
```
{r}
step(I_model, direction='forward', scope=formula(m))
```

```
Start:  AIC=2151.24
PriceEuro ~ 1

                  Df  Sum of Sq        RSS    AIC
+ TopSpeed_KmH     1 8.1688e+10 3.7160e+10 2033.5
+ Range_Km         1 5.4125e+10 6.4723e+10 2090.6
+ AccelSec         1 4.6748e+10 7.2099e+10 2101.8
+ Efficiency_WhKm  1 1.8704e+10 1.0014e+11 2135.6
<none>                          1.1885e+11 2151.2

Step:  AIC=2033.49
```

INDEPENDENT VARIABLE DISTRIBUTION



With least AIC score

```
Call:
lm(formula = PriceEuro ~ TopSpeed_KmH + Efficiency_WhKm, data = df1)

Coefficients:
    (Intercept)     TopSpeed_KmH  Efficiency_WhKm
       -80209.2            617.0            134.6
```

With the help of AIC , the explanatory variables are Efficiency and TopSpeed.

AIC = 290 (Least)

Price = x + x1*Efficiency + x2*Topspeed

Hypothesis:

Null Hypothesis : *x, x1, x2 is not significant .*

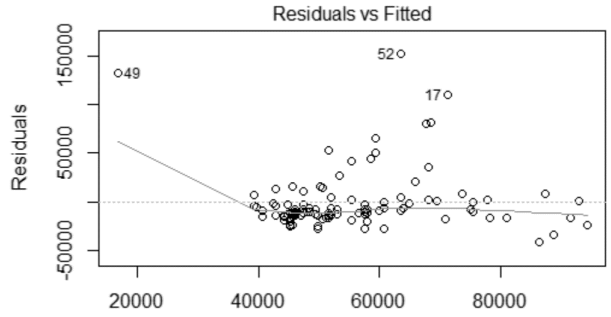
Alternate Hypothesis : *x, x1, x2 is significant*

We will check the level of significance at 5% i.e alpha = 0.05

ASSUMPTIONS:

Linear regression makes several assumptions about the data.

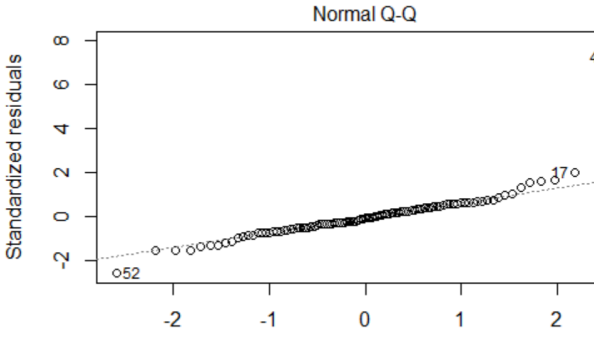
- **Linearity of the data.**



Observation - The line is nearly aligned to the horizontal line,

Inference - Data is almost linear, Hence assume data is satisfactory linear.

- **Normality of residuals.**



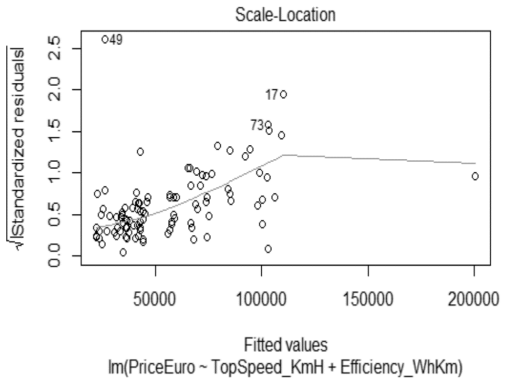

```
        Shapiro-Wilk normality test

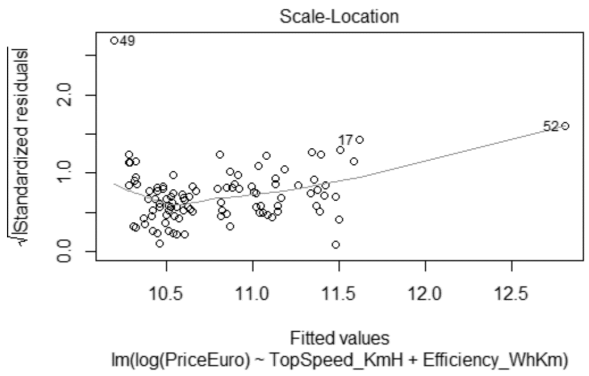
data:  df1$PriceEuro
W = 0.77308, p-value = 2.581e-11
```

Observation – points are aligned with the diagonal line, p-value < 0.05

Inference - data with a normal distribution are considered "normal."

- **Homogeneity of variance**





Observation –horizontal line is not with equally spread points.

Inference/Action - Reduced the heteroscedasticity problem by using a log transformation of the outcome variable (y). After transformation it tends to have equal spread points.

All the assumptions are satisfied we move forward with the final modelling

OBSERVATION

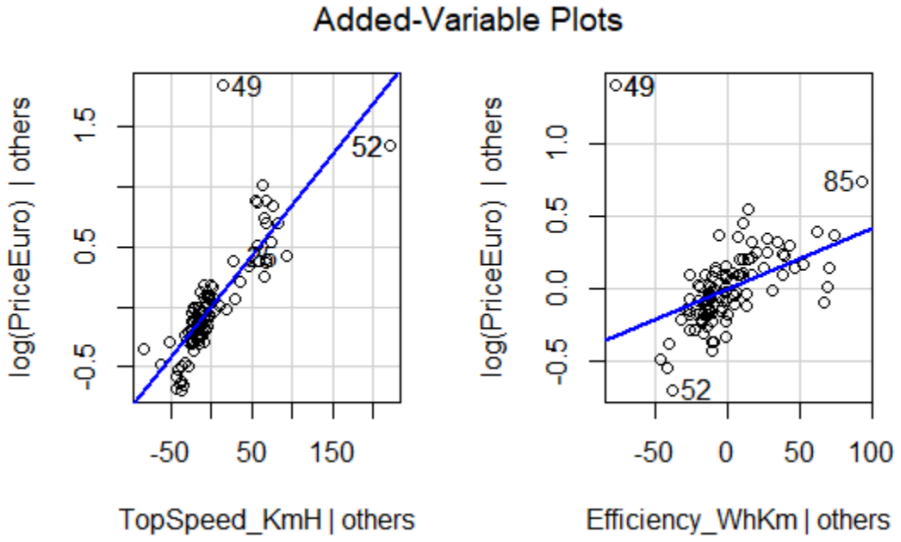
```
Call:
lm(formula = log(PriceEuro) ~ TopSpeed_KmH + Efficiency_WhKm,
    data = df1)

Residuals:
     Min       1Q   Median       3Q      Max
-0.53093 -0.12114 -0.01654  0.10252  1.71179

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     8.5089775  0.1662607   51.18  < 2e-16 ***
TopSpeed_KmH    0.0084246  0.0006034   13.96  < 2e-16 ***
Efficiency_WhKm 0.0041082  0.0008892    4.62 1.15e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2482 on 100 degrees of freedom
Multiple R-squared:  0.7501,    Adjusted R-squared:  0.7451
F-statistic: 150.1 on 2 and 100 DF,  p-value: < 2.2e-16
```

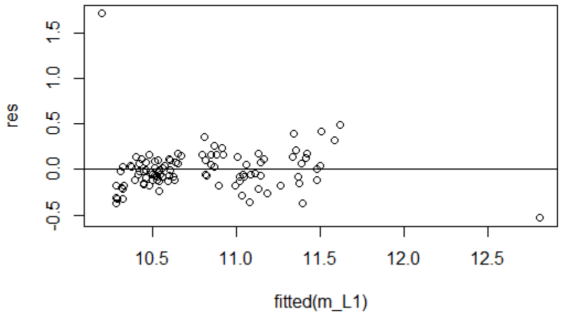
p-value < 0.05, and R and F-statistics are good for all parameters. As a result, the null hypothesis is rejected and the alternative hypothesis is accepted. The model explains 75.01% of variability.



Price = 8.509 + 0.0041*Efficiency_WhKm + 0.0084*Topspeed_KmH

Residual Analysis

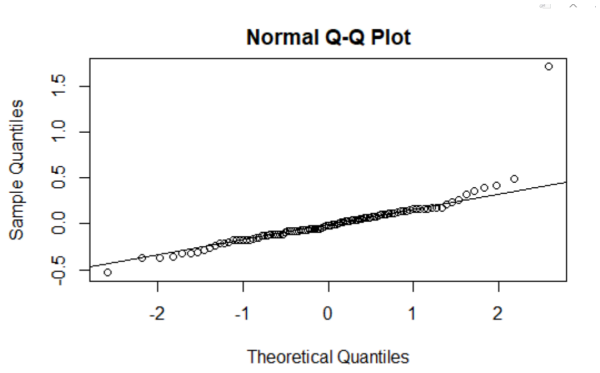
The residual analysis is crucial for confirming the regression model. The regression model is regarded legitimate if the error term in the regression model meets the conditions.

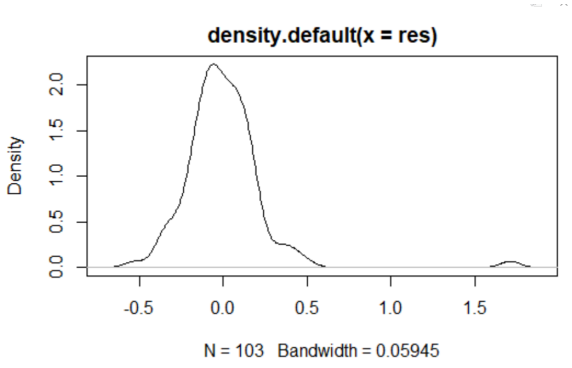


Heteroscedasticity – a systematic change in the spread of residuals over a range of values.

We can see that the spread of the residuals tends to around fitted values and very few are high , but it doesn't look serious enough that we would need to make any changes to the model.

QQPLOT AND DENSITY PLOT FOR THE RESIDUALS





OBSERVATIONS FOR RESIDUAL ANALYSIS:

- The residuals have a normal distribution.
- Homoscedasticity -We can see that the spread of the residuals tends to around fitted values
- The data set has some linear relationship
- Data has no auto - correlation.

CONCLUSION

Hence we confirm the model on the basis of residual analysis ,our initial hypothesis and the required accuracy, In this entire study there are strong evidences that it can be further studied and better accuracy can be attained with the right data and correct approach.

Model..

```
Price = 8.509 + 0.0041*Efficiency_WhKm + 0.0084*Topspeed_KmH
```

Viability and Pricing factors..

The price of an electronic vehicle is influenced by a number of factors, including brand, acceleration/sec, efficiency, powertrain type, and TopSpeed, and the majority of electronic vehicles fall into this category. We've gone through how many factors affect the cost of an automobile in this post. While the initial investment is high, it is a win-win situation in the long run. Returning to affordability, the average price of 45K Euros with all standard features is low, but it is unlikely to be affordable for certain impoverished countries. However, depending on the alternatives, some people may choose a car with a high acc/sec, while others may prefer a certain brand, therefore cost is decided by the choices.

References

- Bevans, R. (2022, May 6). *An introduction to multiple linear regression*. Scribbr. Retrieved May 14, 2022, from https://www.scribbr.com/statistics/multiple-linear-regression/
- Encyclopædia Britannica, inc. (n.d.). *Residual analysis*. Encyclopædia Britannica. Retrieved May 14, 2022, from https://www.britannica.com/science/statistics/Residual-analysis
- *F E A S I B I L I T Y O F e l e c T R I C V E H I C L E S I N I N D I A*. (n.d.). Retrieved May 13, 2022, from

https://www.cppr.in/wp-content/uploads/2019/10/Feasibility-of-Electric-Vehicles-in-India.pdf

- Prabhakaran, S. (n.d.). *If(typeof ez_ad_units != 'undefined'){ez_ad_units.Push([[728,90],'r_statistics_co-box-3','Ezslot_1',109,'0','0'])};if(typeof __ez_fad_position != 'undefined'){__ez_fad_position('div-GPT-ad-r_statistics_co-box-3-0')};top 50 GGPLOT2 visualizations - the master list (with full R code)*. Top 50 ggplot2 Visualizations - The Master List (With Full R Code). Retrieved May 14, 2022, from http://r-statistics.co/Top50-Ggplot2-Visualizations-MasterList-R-Code.html
- *A study on the adoption of electric vehicles in India ... - sage journals*. (n.d.). Retrieved May 13, 2022, from https://journals.sagepub.com/doi/full/10.1177/0972262919875548