**Team Research and Development Project**

**Final report title:** Is there a difference in the mean resale price of cars between different fuel types?

**Group ID:** A035

S

**Dataset number:** DS261

**Prepared by:** Emma Harrison, ID: 23040079,

Rutvik Kishorbhai Vadher, ID: 23034973

Muhammad Waleed Ashraf, ID: 23023507

Marshal Ashokkumar Pachani, ID: 23030538

Saiyad Mahamad Alfaiz Javidbhai, ID: 23022436

University of Hertfordshire

Hatfield, 2024

Table of Contents

1. Introduction
   1. Problem statement and research motivation
   2. The data set
   3. Research question
   4. Null hypothesis and alternative hypothesis (H0/H1)
2. Background research
   1. Research papers (at least 3 relevant to your topic / DS)
   2. Why RQ is of interest (research gap and future directions according to the literature)
3. Visualisation
   1. Appropriate plot for the RQ*output of an R script (NOT a screenshot)*
   2. Additional information relating to understanding the data (optional)
   3. Useful information for the data understanding
4. Analysis
   1. Statistical test used to test the hypotheses and output
   2. The null hypothesis is rejected /not rejected based on the p-value
5. Evaluation – group’s experience at 7COM1079
   1. What went well
   2. Pointsfor improvement
   3. Group’s time management
   4. Project’s overall judgement
   5. Comment on GitHub log output
6. Conclusions
   1. Results explained.
   2. Interpretation of the results
   3. Reasons and/or implications for future work, limitations of your study
7. Reference list

Harvard (author, date) format.

1. Appendices
2. R code used for analysis and visualisation.
3. Git Hub log output.
   * + 1. **Introduction**
   1. **Problem statement and research motivation**

As new technological and ecological advancements are discovered, specifically in the automotive industry, this paper addresses if there is a difference in mean resale price of cars between different fuel types within India.The purpose is to evaluate all fuel types in the given dataset DS261 Car resale data – 2023;and reflect on the impact‘real-life’resale value hason purchasingvehiclescompared totheir respective fuel types. It is important to fully understand and stay informed following the effects the fuel types have in this industry.Ewing and Sarigöllü (1998) suggests “…dramatically higher gasoline prices might prompt them, as one-car households, to purchase a second non-gasoline car.”

**1.2. The data set**

The data set used to explore this research question is DS261Car resale data – 2023. The dataset contains information surrounding car resale prices across different cities in India. There are fifteen columns in this dataset and seventeen thousand,four hundred and forty-seven rows. The dataset captures key information about each vehicle, including their resale value, fuel type and other features like milage and engine capacity. Below (Fig 1) is a snippet of the data set.

A screenshot of a computer

Description automatically generated

Fig. 1. Snippet of the data set

**1.3. Research question**

This paper will explore how to effectively answer the research question while incorporating relevant data. Firstly, investigating how nominal independent and interval dependent variables impact this data. Formulating a relevant hypothesis and then leading on to an evaluation of Pair wise Wilcox on visual statistical analysis, while comparing the normal bell-curve histogram and boxplot.

**1.4. Null hypothesis and alternative hypothesis (H0/H1)**

As this research is based on if resale price has an impact on fuel type, based on the dependent and independent variables respectively, and thus the data types being investigated are interval and nominal data respectively. Understanding this, allows individuals to formulate a comparison of means, and thus speculate if there is a null hypothesis or an alternative hypothesis.There is no difference in the mean resale price of cars between different fuel types or there is a difference in the mean resale price of cars between different fuel types, respectively.

* + - 1. **Background research**

**2.1. Research papers**

When analysing similar data and similar research, according to Gilmore, E.A. and Lave, L.B. (2013) “…These alternative vehicles, however, are more expensive to manufacture and sell at a higher initial retail price than the conventional gasoline vehicles.” Suggests that there is already a link between resale price of vehicles and alternative fuel types, specifically how electric vehicles tend to be more expensive than the traditional gasoline tank. Gilmore and Lave were able to calculate the cost of ownership for paired vehicles and the resale value through firstly evaluating how vehicles retain value through a function of mileage, then used the estimates of resale ratios to find the difference in resale prices in alternative and conventional vehicles.

Though dataset DS261 was not used for any other official related papers, another report on fuel economy and sale price, indirectly affects this paper. Turrentine, T.S and Kurani, K.S (2007) states that “HEVs can cost more to buy than conventional vehicles,”.“…On the other hand, resale value of HEVs (as a percentage of purchase price) is among the highest of any vehicle” which also endorses Gilmore and Laves research.

Finally, comparing fuel efficiency and prices, Alberini, A et al, “…the more fuel-efficient a car is, the higher the price.” Which indirectly influences fuel type, research around this paper, and supports the above-mentioned authors.

**2.2. Why Research Question is of interest**

Similarly to our research motivation, understanding differences in mean resale prices across fuel types addresses a critical research gap, previous studies have often focus on fuel efficiency or environmental impacts. It is important to have a better understanding of any market before committing to buying or selling. When investing huge amounts of money, it is of interest to stay well informed and understand what assets or liabilities may be influencing any decision making. This topic contributes to the broader understanding of how market changes influences purchasing behaviours. For consumers, resale value represents a significant consideration when investing in cars, particularly as fuel type preferences evolve with ecological and technical concerns.

* + - 1. **Visualisation**

**3.1. Appropriate plot for the Research Question**

As mentioned previously, with the data types being interval and nominal for our dependent variable and independent variable respectively, the only reasonable visualisation would be a histogram and box plot for a comparison of means, onto the Pair wise Wilcox on statistical test. Below is the R code for the histogram and box plot

A screenshot of a computer code

Description automatically generated

Fig. 1.Full R code to create a histogram

**A screenshot of a computer code

Description automatically generated**

Fig. 2.Full R code to create a box plot

**3.2. Additional information relating to understanding the data**

Below, represented by figure 2 is the histogram of resale price with normal curve. And figure 3 is the correspondent box plot.

A graph with a red line

Description automatically generated

Fig.3.Histogram of Resale Price with Normal Curve

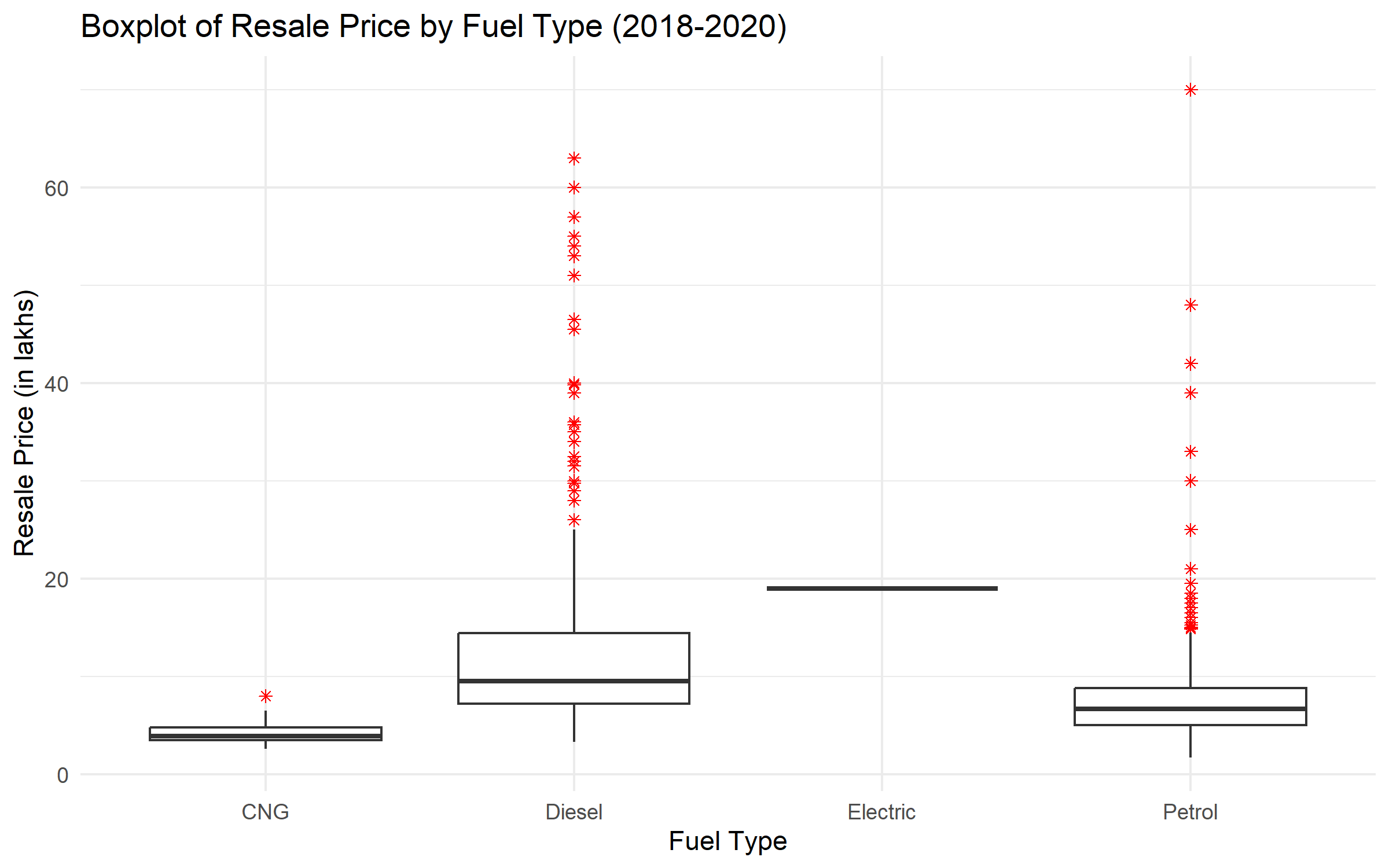


Fig. 4. Boxplot of Resale Price by Fuel Type (2018-2020)

**3.3. Useful information for the data understanding**

The normal curve overlay does not follow the shape of the underlying histogram, indicating that the data is not normally distributed. The histogram shows a right-skewed distribution, which is typical of car resale prices, as most cars are priced in the lower range, with fewer high-value outliers.

* + - 1. **Analysis**

**4.1. Statistical test used to test the hypotheses and output**

The dataset has four levels for the independent variable (fueltype): Petrol, Diesel, Electric, and CNG. Categories. The normal curve overlay does not follow the shape of the underlying data, so we use the non-parametric test that does not assume normality, Pairwise Wilcoxon statistical test. This test allows us to compare each pair of the four fuel types to determine if there are significant differences in resale prices between these categories.

**A white screen with black text

Description automatically generated**

Fig. 5. Full R code used to analyse the Pairwise Wilcoxon Statistical Test

**4.2. The null hypothesis is rejected based on the p-value**

The results are significant for all pair wise comparisons.

Based on the p-values from the Pair wise Wilcox on test, we reject the null hypothesis for most fuel type pairs, indicating that the resale price differs between many of the fuel types.

Since all p-values are less than 0.05, we have differences between these fuel types in terms of resale price.

The fuel type impacts the resale value of cars. Like before, these results are useful for buyers and dealers because, buyers might prefer certain fuel types over others to maintain higher resale value. Dealerships can use this analysis to set competitive pricing for cars based on their fuel types and expected depreciation.

**A screenshot of a computer

Description automatically generated**

Fig. 6. Pair wise Wilcox on as created in R Studios.

* + - 1. **Evaluation**

**5.1 What went well**

The team had a good and clear understanding of each task, we were able to research what consists of a good research question and the analysis respectively. We presented our research question and analysis and were able to work around the feedback that was given. We were able to utilise programmes such as R Studio to be able to carry out potentially lengthy calculations and difficult analysis.

**5.2 Points for improvement**

Something to consider when looking at ways to improve this task could be that we should have worked better as a team. Better communication and task allocation was required to share the weight of this project more evenly between peers.

**5.3 Group’s time management**

When evaluating the groups time management from the start of the project to the end, at first, we had a few time management issues. Due to unforeseen circumstances, we found we were late for presenting our research question; however, we were able to work hard and easily got back on track.

**5.4 Projects overall judgement**

Overall, the project was successful as we were able to determine an appropriate hypothesis for our research question and analysis proved that we had estimated correctly from the start. We were able to meet all our objectives in time and to a high standard. Essentially, the project fulfilled its intended purpose was concise and accurate.

**5.5 Changes to group since submission of Assignment 1**

N/A

**5.6 Git Hub log output(as seen in appendix B)**

1. Commit Message: [Added PowerPoint presentations, newest R script and histogram images] This commit included our research question.

2. Commit Message: [Added PowerPoint presentations, newest R script and histogram images] This commit included visualisation analysis.

1. Commit Message: [please use your folder]This allowed us to separate our individual work. Who worked towards this project.
   * + 1. **Conclusion**

**6.1 Results explained**

Our analysis revealed significant differences in the resale prices of cars based on each fuel type. Using the Pairwise Wilcoxon Rank Sum Test, we observed that resale prices varied across petrol, diesel, electric, and CNG vehicles. This supports the hypothesis that fuel type impacts the resale price, with all p-values being less than 0.05. The results confirm that fuel type is a considered factor for car resale value, including buyers and dealerships strategizing pricing.

**6.2 Interpretation of the results**

The findings indicate that fuel type significantly influences the resale value of vehicles, likely due to devaluation rates, market demand, and fuel efficiency. For our research question, the results show that buyers may benefit from choosing fuel types with historically higher resale values, such as electric or petrol. Overall, the results show the economic impact of fuel type preferences and may persuade market trends and pricing strategies in the automotive industry.

**6.3 Reasons and/or implications for future work, limitations of your study**

Future work could include additional variables, such as mileage or car brand, to further provide a better understanding of resale price determinants. Limitations could include usage of a specific dataset or exclusion of outliers, which may have constrained the results in automotive market.

* + - 1. **References**

Ewing, G.O. and Sarigöllü, E. (1998) ‘Car fuel-type choice under travel demand management and economic incentives’, *Transportation Research Part D: Transport and Environment*, 3(6), pp. 429–444. DOI: 10.1016/S1361-9209(98)00019-4.

Gilmore, E.A. and Lave, L.B. (2013) ‘Comparing resale prices and total cost of ownership for gasoline, hybrid and diesel passenger cars and trucks’, *Transport Policy*, 27, pp. 200–208. Available at: https://www.sciencedirect.com/science/article/pii/S0967070X13000036 (Accessed: 22 December 2024).

Turrentine, T.S. and Kurani, K.S. (2007) ‘Car buyers and fuel economy?’, *Energy Policy*, 35, pp. 1213–1223. Available at: https://www.sciencedirect.com/science/article/pii/S0301421506001200 (Accessed:22 December 2024)

Alberini, A., Di Cosmo, V. and Bigano, A. (2019) ‘How are fuel efficient cars priced? Evidence from eight EU countries’, *Energy Policy*, 134, 110978. DOI: 10.1016/j.enpol.2019.110978.

* + - 1. **Appendices**

**Appendix A**

Below is the R code used to create ahistogram. See Figure 4. For plain text and Figure 2. For code as seen in R Studios

A screenshot of a computer code

Description automatically generated

Fig. 1. Full R code to create a histogram

A graph with a red line

Description automatically generated

Fig. 3.Histogram as created in R studios

Below is the R code used to create a boxplotSee Figure 5. For plain text and Figure 3. The histogram created in R Studios using Fig. 5. Code.

**A screenshot of a computer code

Description automatically generated**

Fig. 2. Full R code to create a boxplot

A screen shot of a graph

Description automatically generated

Fig. 4.Boxplot as created in R Studios

Below is the R code used to analyse the Pairwise Wilcoxon Statistical Test**.** See Figure 6. For plain text and Figure 7. For code as seen in R Studios

**A white screen with black text

Description automatically generated**

Fig. 5. Full R code used to analyse the Pairwise Wilcoxon Statistical Test

**A screenshot of a computer

Description automatically generated**

Fig.6. Pairwise Wilcoxon as created in R Studios

**Appendix B**

Below is Figure. 7. In which shows the GitHub log output:

A screenshot of a computer

Description automatically generated

Fig. 7. GitHub log output