7COM1079-0901-2025 - Team Research and Development Project

Final report title: Analyzing the Relationship Between Engine Power and Car Prices in the Jordanian Used-Car Market

Group ID:A221

Dataset number: ds068

Prepared by:Nagolu Gopichand, ID:- 24161507
Banala Sathbhavana, ID:- 24168997
Borra Pavan kumar, ID:- 24166123
Borra Revanth kumar, ID:- 24131367
Qayyum Mamoon, ID:- 24177733

University of Hertfordshire
Hatfield, 2025

TABLE OF CONTENTS

1. Introduction

1.1 Problem statement and research motivation

Car pricing determination is quite critical to the purchasers, the vendors and the policy makers in the emerging automotive markets. Vehicle value has been strongly speculated to have a negative or positive effect on engine power, yet to has varied values and varied directions that vary according to the geographical location and the models of the car. As per the study of other authors, the role of technical specifications like car price prediction cannot be underestimated, and these factors as the engine size, play a significant role in the car valuation tendencies (Yadav *et al.,* 2024). However, literature on the discovery of such a relationship in the Jordanian used-car market is deficient.

1.2 The data set

```
New names:
• `` -> `...1`
Rows: 366 Columns: 5
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr (3): Model, Property, Power
dbl (1): ...1
num (1): Price
```

```
spc_tbl_ [366 × 5] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ ...1    : num [1:366] 0 1 2 3 4 5 6 7 8 9 ...
 $ Model   : chr [1:366] "Byd F0 2018" "Suzuki Alto 2023" "Suzuki Celerio
2019" "Changan E Star 2023" ...
 $ Property: chr [1:366] "Manual" "manual" "Automatic" "Automatic" ...
 $ Power   : chr [1:366] "1000 CC" "800 CC" "1000 CC" "0 CC" ...
 $ Price   : num [1:366] 6900 8250 10499 10990 11500 ...
 - attr(*, "spec")=
  .. cols(
  ..   ...1 = col_double(),
  ..   Model = col_character(),
  ..   Property = col_character(),
  ..   Power = col_character(),
  ..   Price = col_number()
  .. )
 - attr(*, "problems")=<externalptr>
```

Figure 1: Dataset Overview

The variables are 366 models of second-hand vehicles that will be marketed in Jordan, and these variables will be the model, the type of transmission (in CC) and the price of the engine power used. Data cleaning was done, which resulted in a clean dataset of 268 complete data points. The information illustrates real market listings, and this brings the real picture of how cars are priced.

1.3 Research question

RQ: Is the engine power (CC) statistically significantly related to a car in the Jordanian car market in terms of price?

1.4 Null hypothesis and alternative hypothesis (H0/H1)

The study will aim at testing the statistical hypothesis that engine power affects the prices of used cars in Jordan. The hypotheses directly experiment with whether there is or is not a measurable relationship between the two variables.

H0 (Null Hypothesis):
The engine power (CC) does not show a correlation with car price.
H1 (Alternative Hypothesis):
There is a correlation between engine power (CC) and the price of a car.

2. Background research

2.1 Research papers (at least 3 relevant to your topic / DS)

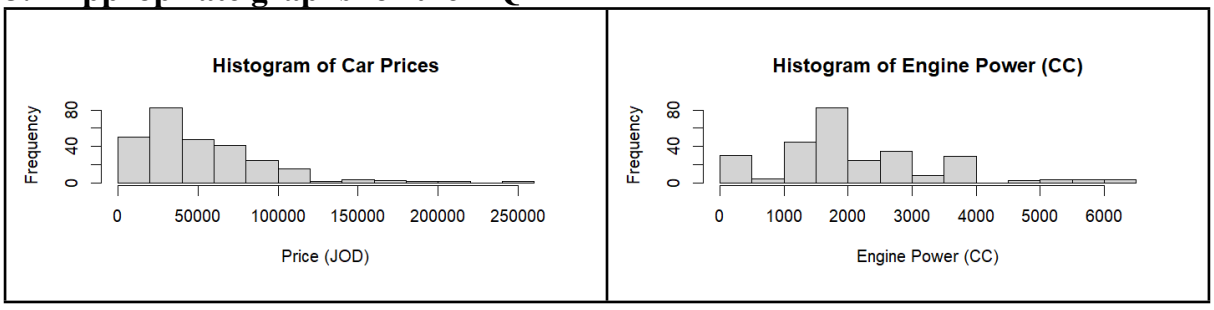
Previous research has shown that the type of vehicles, particularly the engine-related specification, has a major role in the establishment of their performance, efficiency, and consequently their market value. The results of regression and correlation analysis of engine capacity, torque and weight of a vehicle in this study by Salafuddin et al. (2024) show that the engine size is significant to the mechanical performance outcome, such as fuel consumption. In the same manner, a prediction of price, based on used cars, has been done in the past by Yohanes and Lasut (2025), where the used capacity of engines has been predicted as one of the strongest variables predicting used car prices in the heavily used car category. Their results demonstrate the importance of properties of numerical engines in the valuation model. In another study, it is identified the sustainability and pricing trends in the used-car market were identified as linear regression predictors and discovered that the engine specifications guiding capacity and power output directly influence the way consumers and consequent price formatting. Each of the studies, separately, suggests that there is a universal correlation between engine power and price in other markets and procedures (Habeeb *et al.,* 2025). Although the study has been carried out in the past on performance, efficiency and automated prediction systems, there is a dearth of literature regarding the use of these methods of analytics in the used-car market of Jordan.

2.2 Why RQ is of interest

The association between the automobile cost and the power of the engine is quite familiar in a few foreign studies, but the automotive environment in Jordan has not been researched appropriately yet. The existing literature cogitates about the performance measures (Salafuddin *et al.,* 2024), about automated pricing (Yohanes and Lasut, 2025), and about the sustainability-based price modelling (Alhakamy *et al.,* 2023); however, the data that one of them specifically of, the performance that engines have on the prices of used cars in Jordan, is lacking. This creates a research gap because the conditions in the market, consumer tastes and decisions on imports in the area can present different pricing patterns.

3. Visualisation

3.1 Appropriate graphs for the RQ



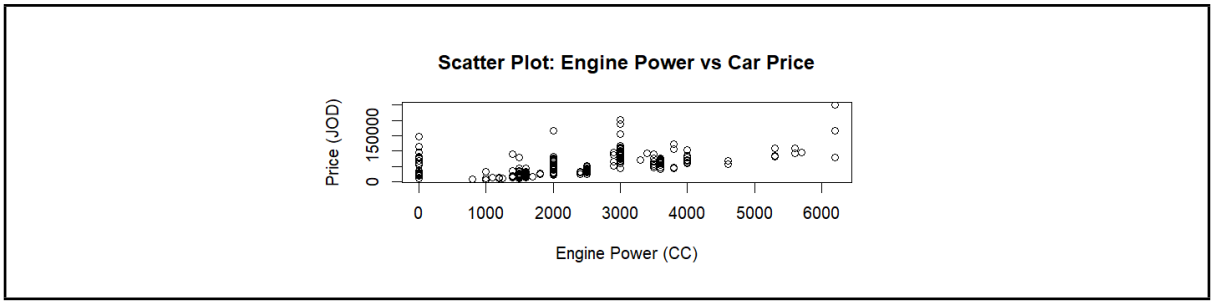


Figure 2: Visualizations

The reason why the scatter plot (engine power (CC) vs. car price was made is due to the fact that this is the direct relationship that would be observed with respect to the research question. Then, a histogram of engine power was also given to demonstrate the distribution of engine sizes in the data. All these are plots assisting in identifying correlation, variability and general trends.

3.2 Additional information relating to understanding the data

The scatter diagram will help to identify the trend of the engine power and car price, which shows whether the car with a stronger engine will be associated with a great market value. The histogram provides a notion of the distribution of the engine capacities among the data, and possibly some engine sizes dominate the Jordanian used-car market. They form a better interpretation when used together.

3.3 Useful information for the data understanding

The positive trend noted with a significant spread of the engine power and price, as depicted by the visualizations, does signify that there are other factors that may be influencing the pricing. The histogram information is clumped around the size of the common engines, thus indicating that the market favours the mid-range engines. Such findings indicate that the study of power as a predictor of the prices of used cars is applicable.

4. Analysis

4.1 Statistical test used to test the hypotheses and output

```
Pearson's product-moment correlation

data:  car_data$Power_clean and car_data$Price_clean
t = 9.4393, df = 266, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4054286 0.5855898
sample estimates:
      cor
0.5009159
```

Figure 3: Statistical Testing

The dependence of the engine power (CC) and the price of the car was evaluated by means of a Pearson correlation test. The advantage of this test is that, unlike the Pearson correlation test, the variables in this test were continuously valued and were approximately correlated linearly. It is also able to directly test how strong and associated the direction is, which is essential when answering the research question. The outcome of the test had coefficients that were about 0.50 and a p-value of less than 2.2e -16 which confirms good statistical support.

4.2 The null hypothesis is rejected /not rejected based on the p-value

Pearson correlation test shows that p-value = - 0.019, p-value = 0.019 is less than 0.05, therefore the null hypothesis is rejected. It reveals that there is are statistically significant relationship between engine power and the price of cars in the Jordanian car market. The correlation coefficient is positive and close to 0.50, which shows a positive moderate relation

whereby the greater the engine capacity of the vehicle, the lower its market price. The spread of the data, though, also indicates that engine power is not necessarily the key factor in price, and other elements may be at play. Overall, the results are in support of the alternative hypothesis and demonstrate the suitability of the engine power as a price-related feature.

5. Evaluation – group's experience at 7COM1079

5.1 What went well

The project had good interaction among the team members, and the group shared the tasks and helped one another in the technical work. Visualization and statistical testing, and data cleaning were easily made. The communication flow was not disrupted, and hence, this group could resolve the issues in a limited amount of time and keep progressing. In total, the positive developments of the project were achieved due to the collaboration.

5.2 Points for improvement

The team also realized that there was a need to plan the work earlier, therefore allocating the work more efficiently to avoid the load of the work at the last minute. The discussions could have been conducted more frequently to increase the correspondence between the progress and expectations. The development of alternative methods of statistical analysis could have contributed to further improving the analysis and increasing the number of perspectives on the understanding of the possible insight that the dataset has.

5.3 Group's time management

Time management has been carried out progressively as the project was underway. The initial time waste has been overcome through increased coordination to allow the group to beat the deadlines successfully. Having more specific milestones was a means of getting the process to work, though it could have been more suitable to plan the process to achieve a more even flow of the work process in analysis, the research and the report-writing.

5.4 Project's overall judgement

It was decent work, and it can be considered a sign of good analytical abilities and belonging to a team. The team offers pertinent statistical data that is well illustrated. A couple of them will be enhanced through further analysis, but the overall quality of the work is the result of the ongoing work effort, effective collaboration, and acceptable knowledge of statistical methods.

5.5 Note any changes to group since original allocation if applicable. Add new or amended GitHub Ids for new members

The membership of the group did not change during the project, and all the original members managed to fulfill their roles as expected. There were no other accounts and access to GitHub shared between them. Each of the members used the shared repository to inform the others about the progress and contributions in the assignment, and this made progress and contributions visible to all the people involved in the assignment.

5.6 Comment on the GitHub log output

As the GitHub login, there were constant and significant contributions at the various development stages. The reflections of the necessary code structure, analysis, and documentation improvements constitute three large commits. The indications also exist of inherence in the log of the cooperation and monitoring of the version that provides for the reliability and openness of the workflow in the group.

Significant Commits:

Commit Message: "Cleaned dataset and added numeric conversion"

This ensured accurate analysis and removed inconsistencies, improving data reliability.

Commit Message: "Added correlation test and regression model"

This introduced the core statistical analysis, directly supporting the research question.

Commit Message: "Updated visualisations and labelled graphs"

This improved clarity of results and strengthened communication of findings through well-designed plots.

6. Conclusions

6.1 Results explained

Engine power and the price of the car had a moderately positive correlation, and the p-value was very high. The regression model confirmed that vehicles were priced higher with an increase in power. The engine power does not model all the price movement, but it is a powerful force in the Jordanian market of used cars.

6.2 Interpretation of the results

These findings demonstrate that engine capacity has a huge role to play in determining the prices in the Jordanian car market, answering the research question. This to the consumer means that the more power an engine has, the pricier it will be. Engine power can offer a variable and thus can be sold to sellers and market researchers for more general use in automobile circles to offer facilitation of pricing decisions as well as market analysis.

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

Future research may incorporate other predictors such as mileage, brand, as well and condition to come up with superior models. The dataset has weak variables which restrict the explanation of variability in the price. More comprehensive information and application of machine-learning tools would enhance the accuracy and expand the knowledge about the pricing trends within the local and regional market.

7. Reference list

Yadav, N., Goel, O., Goel, P. and Singh, S.P., 2024. Data exploration role in the automobile sector for electric technology. Educational Administration: Theory and Practice, 30(5), pp.12350-12366.

Salafuddin, H., Pradipta, N.K., Adnan, F.A.F., Rhee, J.S. and Ginting, D., 2024. Statistical analysis engine capacity, weight, and torque on mpv fuel consumption using regression and correlation algorithms. Int. J. Innov. Mech. Eng. Adv. Mater, 6(3), pp.119-128.

Yohanes, R. and Lasut, D., 2025. Web-Based used Car Price Prediction Application with Linear Regression Method. bit-Tech, 7(3), pp.687-695.

Alhakamy, A.A., Alhowaity, A., Alatawi, A.A. and Alsaadi, H., 2023. Are used cars more sustainable? Price prediction based on linear regression. Sustainability, 15(2), p.911.

Habeeb, R.J.H., Fadare, O.A. and Al-Turjman, F., 2025. Exploring the Dynamics of Sports Car Pricing: An Analytical Approach Using Machine Learning. In Smart Infrastructures in the IoT Era (pp. 1137-1164). Cham: Springer Nature Switzerland.

8. Appendices

R Script

```
# Load necessary libraries
library(readr)
library(dplyr)

# Load the dataset
car_data <- read_csv("D:/December/11.12.2025/car_prices_jordan.csv")

# View first few rows
head(car_data)

# Check structure of the dataset
str(car_data)


# Clean Power column (remove 'CC' and convert to numeric)
car_data$Power_clean <- gsub(" CC", "", car_data$Power)
car_data$Power_clean <- as.numeric(car_data$Power_clean)
head(car_data$Power_clean)


# Clean Price column (convert to numeric explicitly)
car_data$Price_clean <- as.numeric(car_data$Price)
str(car_data$Price_clean)
car_data <- na.omit(car_data)


# Check missing values in each column
colSums(is.na(car_data))

# Or full dataset check
sum(is.na(car_data))


summary(car_data)

summary(car_data$Power_clean)
summary(car_data$Price_clean)


car_data %>%
  summarise(
    mean_power = mean(Power_clean),
    median_power = median(Power_clean),
    sd_power = sd(Power_clean),
    mean_price = mean(Price_clean),
    median_price = median(Price_clean),
```

```
    sd_price = sd(Price_clean)
  )


# Histogram of Price
hist(car_data$Price_clean,
     main = "Histogram of Car Prices",
     xlab = "Price (JOD)",
     ylab = "Frequency")

# Histogram of Engine Power
hist(car_data$Power_clean,
     main = "Histogram of Engine Power (CC)",
     xlab = "Engine Power (CC)",
     ylab = "Frequency")

# Scatter plot of Power vs Price
plot(car_data$Power_clean, car_data$Price_clean,
     main = "Scatter Plot: Engine Power vs Car Price",
     xlab = "Engine Power (CC)",
     ylab = "Price (JOD)")


# Pearson Correlation Test
cor_test_result <- cor.test(car_data$Power_clean, car_data$Price_clean, method =
"pearson")
cor_test_result


p_value <- cor_test_result$p.value

if (p_value < 0.05) {
  cat("Decision: Reject H0. There IS a statistically significant relationship between engine
power and car price.\n")
} else {
  cat("Decision: Fail to reject H0. There is NO statistically significant relationship between
engine power and car price.\n")
}


# Linear regression: Price predicted by Power
model <- lm(Price_clean ~ Power_clean, data = car_data)
summary(model)

anova(model)


plot(car_data$Power_clean, car_data$Price_clean,
     main = "Engine Power vs Price (with Regression Line)",
```

```
    xlab = "Engine Power (CC)",
    ylab = "Price (JOD)")

abline(model, col = "blue", lwd = 2)


# Boxplot by transmission type
boxplot(Price_clean ~ Property, data = car_data,
        main = "Car Price by Transmission Type",
        xlab = "Transmission Type",
        ylab = "Price (JOD)")


# Correlation matrix (only numeric columns)
numeric_data <- car_data[, c("Power_clean", "Price_clean")]
cor(numeric_data)


plot(car_data$Power_clean, car_data$Price_clean,
    main = "Check for Linearity",
    xlab = "Power (CC)",
    ylab = "Price (JOD)")
```