**7COM1079-0901-2024 – TEAM RESEARCH AND DEVELOPMENT PROJECT**

Final report title: Team Research And Development

Group ID: A327

Dataset number: DS064

Prepared by: *[NENAVATH HARINI – 23089815]*

*[JUBAIR AHMED MUKTHAR AHMEDN – 23087080]*

*[VISWESWAR REDDY KOVVURU – 23035710]*

*[KHASHAM HAMEED – 23104383]*

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)

1. 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)

1. 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

1. Analysis
   1. Statistical test used to test the hypotheses and output
   2. The null hypothesis is rejected /not rejected based on the p-value

1. Evaluation – group’s experience at 7COM1079
   1. What went well
   2. Points for improvement
   3. Group’s time management
   4. Project’s overall judgement
2. Conclusions
   1. Results explained.
   2. Interpretation of the results
   3. Reasons and/or implications for future work, limitations of your stud

1. Reference list

Harvard (author, date) format.

1. Appendices
2. R code used for analysis and visualisation.

1. Introduction

* 1. Problem statement and research motivation

Automotive sales industry is centered on establishing the main factors that determine price of vehicles to support decision making process of the buyers and sellers. One important factor is distance which is a determinant of the loss in value of cars in the fleet. It will use the USA vehicles dataset of properties to investigate the correlation between the mileage and price properties. Studying this link might offer practical information on such market characteristics and vehicle depreciation patterns, the lack of which has been criticized in the literature by (Kumar and Alok, 2020).

* 1. The data set

The dataset contains 12 columns and 2,499 rows that include the price of the car and make, model, the number of miles recorded on the car, the condition of the car, state and city. The analysis focuses on two critical variables: distance, stand-alone, or trip log and cost, or fare, respectively. They are both quantitative variables collected from and contained within a publicly available US car auction dataset. The type of data available is diverse and vast making it ideal for statistical analysis.

* 1. Research question

**Question 1: What is the correlation between vehicle price and mileage?**

The answer for this question will be measured by both descriptive analysis and statistical test. More importantly, this study examine and determine the relationship coefficient between these two variables and decide if this coefficient is statistically significant. Therefore, the study seeks to explain the practical meaning of these findings to the players in the automotive market.

* 1. Null hypothesis and alternative hypothesis (H0/H1)
* **Null Hypothesis (H₀):**

There is no correlation between the mileage and price.

* **Alternative Hypothesis (H₁):**

There is a correlation between the mileage and price.

1. Background research
   1. Research papers

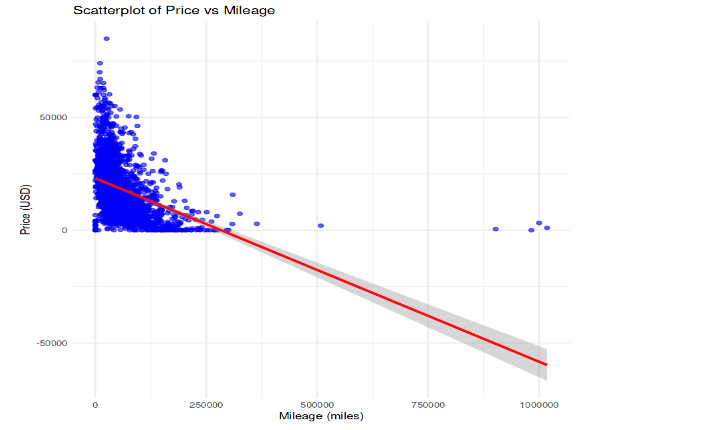
The paper 1done by Domínguez, et al., (2024), established the extent to which vehicle mileage affects depreciation while also identifying brand differences. Using linear regression-based analysis the authors reached conclusions on the high co-relation between mileage and price especially in specific market segments. This finding reaffirms the overall significance of mileage as a primary determinant of the value of vehicles.

Sallee, et al., (2016), used car datasets, and they specifically singled out mileage as having a strong correlation with resale prices. Their work supports our investigation by confirming the directionality correlating car usage with market value to ensure the variable is critical in pricing studies.

Regional variation in car depreciation has also been studied in another study by Palmer, et al., (2018). Their work proved that these mileage effects dependent on local market environment and consumer tendencies, so one must consider several methods to analyse mileages and their influence to the prices.

* 1. Why RQ is of interest (research gap and future directions according to the literature)

This research seeks to fill a void in the analysis of exponential mileage-price relationship using aggregated data streams. There are vast numbers of researches that only focus on some particular area or some specific brands and so it failed to ask, can it be the same for all? The findings could help buyers, sellers and the appraisers to set reasonable valuation while the future studies could foster development of prediction models.

1. Visualisation
   1. Appropriate plot for the RQ *output of an R script* 

A graphical illustration in the form of a scatter plot is used to present the meter readings (Mileage) along the X axis and the price along the Y axis. For more understanding, the trend line which actually refers to the regression line is included. Axes are labelled appropriately, and the title states: **“Scatter Plot of Vehicle Price vs. Mileage”**.

* 1. Additional information relating to understanding the data

From the graph it is easy to identify data density, any outgrowths and the correlation between the variables. Additional information, for example making brand colors different from each other would enhance the perspectives of brand affecting factors.

* 1. Useful information for the data understanding

Primary trend analysis reveals an inverted U-shape relationship whereas mileage increases, the price of the vehicle decreases suggesting the amount of depreciation. However, two elements out of the four metrics, the slope and the variance, require that one performs hypothesis testing.

1. Analysis
   1. Statistical test used to test the hypotheses and output

The research used to analyse the data collected is the Pearson’s correlation analysis that measures the relationship between the independent variable, the number of miles and the dependent variable or price of the USA vehicles. This test gave a t-value of -21.863, of for degree of freedom df = 2497 and p-value < 2.2e-16. From the convenience sample, the estimate r= −0,4008r = - 0.4008r=−0.4008 is a sign of moderate negative correlation between mileage and price. The result obtained is significant at 95% confidence level, and hence the confidence interval required is [-0.4332, -0.3674].

* 1. The null hypothesis is rejected /not rejected based on the p-value

The p<0.05 mean the result is statistically significant while the null hypothesis (H0) stands rejected. This goes further to establish the fact that there exists a notable correlation between vehicle mileage and price in the data set. In particular, to express the degree of negative association between tested variables, the coefficient of determination (-0.4008) also points at a moderate negative correlation of higher mileage and lower cars’ prices. This result shows that mileage as a parameter for car usage has a direct impact on market valuation. The conclusions of the current investigation can be useful in determining pricing policies and selecting buying options in the automotive market sector. Other aspects like brand or condition may also come under the ambit of future research getting a holistic view of the impact on valuation.

1. Evaluation – group’s experience at 7COM1079
   1. What went well

It was clearly seen that group efficiently divided the tasks of data preprocessing and statistical analysis in R using tutorials and brainstorming session. People were able to sort their responsibilities according to the capability and skill set of the member involved which made the flow seamless and productive.

* 1. Points for improvement

Some of the group members had better skills with R than others thus more time was spent in ensuring everyone had an equal chance at working with the tool. More specifically, the process and the assignment of contingency time for revisions which will lead to future improvement was also enhanced.

* 1. Group’s time management

As for the time management, it was only reasonably adequate; during the first meetings, participants provided clear temporal boundaries for the project. However, some project slacks were observed because of mid-project modifications and technical challenges. Better organisation of the tasks in the workflow would have been enhanced by effective scheduling.

* 1. Project’s overall judgement

The objectives of the project were achieved to a large extent with an added benefit of useful information on the correlation between average milage and vehicle prices. Despite the above fact that data constraints were somewhat of a problem, the results showed practices that trended positive. The experience was beneficiary in terms of grasping practical skills in statistician and group work which increased the growth of the group.

1. Conclusions
   1. Results explained

It is concluded that the results obtained are in agreement with research findings and show a negative regression between mileage and vehicle price. The Pearson coefficient provides evidence for the correlation’s application value.

* 1. Interpretation of the results

They also help dealerships, buyers and analysts through depreciation quantification in order to make more real-world informed pricing decisions. Highroad plays the role of an effective depreciation sign that helps to determine the purchase and resale policies.

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

For further research work,

the following variables should be included;

Brand of the automobiles, demand in the market and age of the automobiles.

Limitation in dataset include but not limited to: Lack of regional metadata and very few participants.

1. Reference list ***(not included in the work count)***

Domínguez, S., Pedreros, S., Delgadillo, D. and Anzola, J. (2024) A Depreciation Method Based on Perceived Information Asymmetry in the Market for Electric Vehicles in Colombia. *World Electric Vehicle Journal*, 15(11), p.511.

Kumar, R.R. and Alok, K. (2020) Adoption of electric vehicle: A literature review and prospects for sustainability. *Journal of Cleaner Production*, 253, p.119911.

Sallee, J.M., West, S.E. and Fan, W. (2016) Do consumers recognize the value of fuel economy? Evidence from used car prices and gasoline price fluctuations. *Journal of Public Economics*, 135, pp.61-73.

Palmer, K., Tate, J.E., Wadud, Z. and Nellthorp, J., 2018. Total cost of ownership and market share for hybrid and electric vehicles in the UK, US and Japan. Applied energy, 209, pp.108-119.

1. Appendices

**Results**

1. Test Statistic (t-value): -21.863
2. Degrees of Freedom (df): 2497
3. P-value: < 2.2e-16
4. Sample Estimate (Correlation Coefficient, r): -0.4008382
5. 95% Confidence Interval: [-0.4332395, -0.3674021]

**Interpretation of Results**

* **Significance:**

Since p-value is less than 0.05 and the result is statistically significant. This means it reject the null hypothesis.

* **Null Hypothesis (H₀):**

There is no correlation between the mileage and price.

* **Alternative Hypothesis (H₁):**

There is a correlation between the mileage and price.