



EGYPT UNIVERSITY OF INFORMATICS  
FACULTY OF COMPUTING  
& INFORMATION SCIENCES

# Analysis of the determinants of used car prices

Submitted by:

Maryam Ewida 22-101218

Jana Elbakry 22-101159

Shahd Ayman 22-101045

Malak Shaker 22-101008

Supervised by:

Dr. Mohamed Taher ElRefaei

TA Nadine ELsaeed

## Introduction:

We're all aware of how inflation affects almost everything around us, including the prices we pay for goods and services. However, beyond inflation, there are many other factors at play when it comes to determining the prices of used cars.

In this analysis, we're interested in uncovering these additional factors that influence the prices of pre-owned vehicles. While inflation certainly plays a big role, we want to dig deeper and explore the various other variables that impact car prices.

## Research question:

What are the determinants of used car prices based on registration year, inflation, and other factors?"

## Hypothesis:

### Mileage vs. Price:

Null Hypothesis ( $H_0$ ): There is no significant relationship between the mileage of a used car and its price.

Alternative Hypothesis ( $H_1$ ): There is a significant relationship between the mileage of a used car and its price.

### Number of Owners vs. Price:

Null Hypothesis ( $H_0$ ): There is no significant relationship between the number of owners a used car has had and its price.

Alternative Hypothesis ( $H_1$ ): There is a significant relationship between the number of owners a used car has had and its price.

### Transmission Type vs. Price:

Null Hypothesis ( $H_0$ ): There is no significant difference in the prices of used cars based on their transmission type (automatic or manual).

Alternative Hypothesis ( $H_1$ ): There is a significant difference in the prices of used cars based on their transmission type (automatic or manual).

## Dataset:

The dataset used for this analysis comprises information on used cars from India. It encompasses a diverse range of car models, registration years, and relevant attributes such as mileage, number of owners, transmission type, and prices in Indian Rupees (INR).

(1388,18)

- `car_name`: The name or model of the car.
- `registration_year`: The year in which the car was registered.
- `insurance_validity`: The validity of insurance coverage for the car.
- `fuel_type`: The type of fuel used by the car (e.g., petrol, diesel).
- `seats`: The number of seats in the car.
- `kms_driven`: The total kilometers driven by the car.
- `ownership`: The number of owners the car has had.
- `transmission`: The type of transmission (e.g., automatic, manual).
- `manufacturing_year`: The year in which the car was manufactured.
- `mileage(kmpl)`: The mileage of the car in kilometers per liter.
- `engine(cc)`: The engine capacity of the car in cubic centimeters.
- `max_power(bhp)`: The maximum power output of the car's engine in brake horsepower.
- `torque(Nm)`: The torque generated by the car's engine in Newton-meters.
- `brand`: The brand or manufacturer of the car.
- `Rupees`: The price of the car in Indian Rupees (INR).

## Analysis:

In our analysis of used car prices in India, we considered various factors like mileage, ownership history, and transmission type. We also factored in inflation rates, using a dataset that had missing values for the year 2011. To handle this, we filled in the missing values with the average inflation rate for that year.

It's important to note that this method might have affected our results, especially if the inflation rates for 2011 differed significantly from the average. While we aimed to ensure our analysis was as accurate as possible, it's crucial to be aware of this potential impact on our findings.

After cleaning the dataset, the next step is both visualizing and analyzing the dataset to find patterns to support the hypothesis we came up with at the beginning of the report.

This section will cover all visualizations used in the Jupyter Notebook, platform used to analyze the dataset, that will help us answer the research question and conduct the hypothesis testing to either reject or fail to reject  $H_0$ .

## Hypothesis testing steps:

### Mileage vs. Price:

Step 1: Define null and alternative hypothesis:

- Null Hypothesis ( $H_0$ ): There is no significant correlation between mileage and price of used cars.
- Alternative Hypothesis ( $H_1$ ): There is a significant correlation between mileage and price of used cars.

Step 2: Choose the appropriate test

We chose the Pearson correlation test to analyze the relationship between mileage and price because both variables are continuous and the test assesses linear relationships. This choice aligns with our research question of whether there's a significant correlation between mileage and price. The test assumes linearity, which is plausible in the context of our variables. Overall, the Pearson correlation test is a suitable method for quantifying the strength and direction of the relationship between mileage and price.

Step 3: Calculate the p-value

P-value: 0.006885388032457903

Step 4: Determine the statistical significance

reject the null hypothesis, there is a significant correlation between mileage and price of used cars

### Number of Owners vs. Price:

Step 1: Define null and alternative hypothesis:

- Null Hypothesis ( $H_0$ ): There is no significant difference in prices of used cars between different numbers of owners.
- Alternative Hypothesis ( $H_1$ ): There is a significant difference in prices of used cars between at least two categories of owners.

Step 2: Choose the appropriate test

We chose the ANOVA test to compare the prices of used cars across different categories of ownership because it allows us to assess whether there are significant differences in mean prices among multiple groups simultaneously. This choice is suitable for our research question, which aims to determine if there are statistically significant differences in prices based on the number of owners a car has had. It is designed for comparing means across multiple groups, making it ideal for analyzing the impact of ownership categories on prices. Additionally, ANOVA assumes independence of observations within each group, which is reasonable for our dataset. By using ANOVA, we can effectively evaluate whether there are significant differences in mean prices between cars with different numbers of owners, providing valuable insights into how ownership history influences car prices in the used car market.

Step 3: Calculate the p-value

P-value: 0.001265353194125373

Step 4: Determine the statistical significance

reject the null hypothesis, there are significant differences in car prices across ownership categories

### Transmission Type vs. Price:

Step 1: Define null and alternative hypothesis:

- Null Hypothesis ( $H_0$ ): There is no significant difference in mean prices of used cars between automatic and manual transmission types.
- Alternative Hypothesis ( $H_1$ ): There is a significant difference in mean prices of used cars between automatic and manual transmission types.

Step 2: Choose the appropriate test

We chose the independent samples t-test to compare the prices of cars based on transmission type (automatic vs. manual) because it is specifically designed to assess whether there is a significant difference in means between two independent groups. This choice aligns with our research objective of determining if there is a statistically significant difference in mean prices between cars with automatic transmissions and those with manual transmissions. Additionally, the t-test assumes normality within each group, which is often reasonable for large sample sizes. By using this test, we can confidently evaluate whether there is a significant price difference between the two transmission types, providing valuable insights into the pricing dynamics of cars with different transmission systems.

Step 3 Calculate the p-value  
P-value: 0.25989525314241907

Step 4: Determine the statistical significance  
Even though the visualization included in the notebook suggested that automatic cars are higher in price, we still fail to reject the null hypothesis, there is no significant difference in mean prices of used cars between automatic and manual transmission types.

## Conclusion:

In conclusion, our analysis aimed to uncover the determinants of used car prices in India, focusing on factors such as mileage, ownership history, and transmission type. We began by formulating hypotheses related to these factors and conducted hypothesis testing to evaluate their significance.

For the hypothesis regarding mileage vs. price, we found a significant correlation between these variables, suggesting that mileage does indeed influence the price of used cars. Similarly, our analysis revealed significant differences in car prices across ownership categories, indicating that the number of owners a car has had impacted its price.

However, when comparing prices based on transmission type, we found no significant difference between cars with automatic and manual

transmissions. This suggests that transmission type may not be a significant determinant of used car prices in the Indian market

## Potential issues:

There may be other important variables not included in the analysis that could confound the relationships observed. For example, features like vehicle brand reputation, model popularity, maintenance history, and market trends could all influence used car prices but were not accounted for in our analysis. Furthermore, our analysis focused specifically on the Indian used car market, and the findings may not be applicable to other countries or regions with different market dynamics, regulations, and consumer preferences.