FORECASTING THE VALUE OF PRE-OWNED VECHICLES

TEAM MEMBERS:

- Nikhith Krishna Vinduru
- Chaitanya Gopinadh Madala
- Divya Mamuru
- SaiKumar Dacharla

EXECUTIVE SUMMARY:

This project addresses the dynamic pricing challenge in the pre-owned vehicle market. The aim is to model the cost of cars using various independent variables, thereby understanding the relationship between these variables and price variations. Key technologies employed include Python 3.7, Jupyter Notebook, Flask Framework, HTML, CSS, JS, and AWS Cloud. The project leverages advanced data engineering and machine learning techniques, with a focus on the Random Forest Model for its accuracy in predicting vehicle prices.

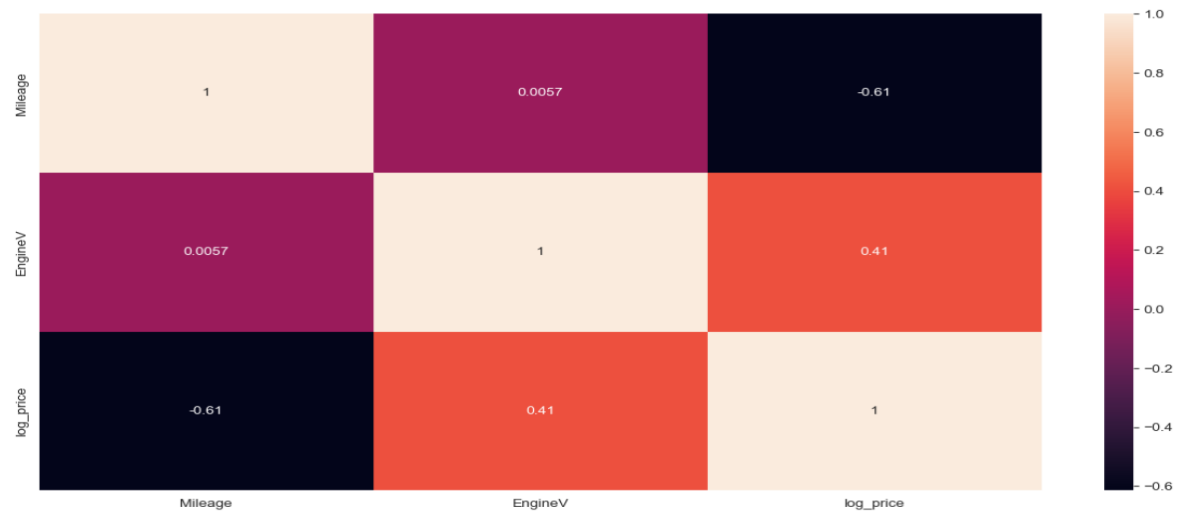
INTRODUCTION:

The pre-owned vehicle market faces the challenge of accurately determining vehicle values, influenced by factors like make, model, age, mileage, and condition. This project aims to develop a scalable model that provides real-time, accurate valuation of used vehicles. This model is crucial for both buyers and sellers, assisting them in making informed decisions.

METHODOLOGY:

The methodology encompasses gathering data from various sources including historical sales records, online listings, and manufacturer specifications. This data is then processed into a cohesive format for training the forecasting model. The project focuses on selecting relevant features that significantly impact vehicle valuation. The Random Forest Model is emphasized for its high accuracy in dealing with complex datasets and multiple variables.

RESULTS:



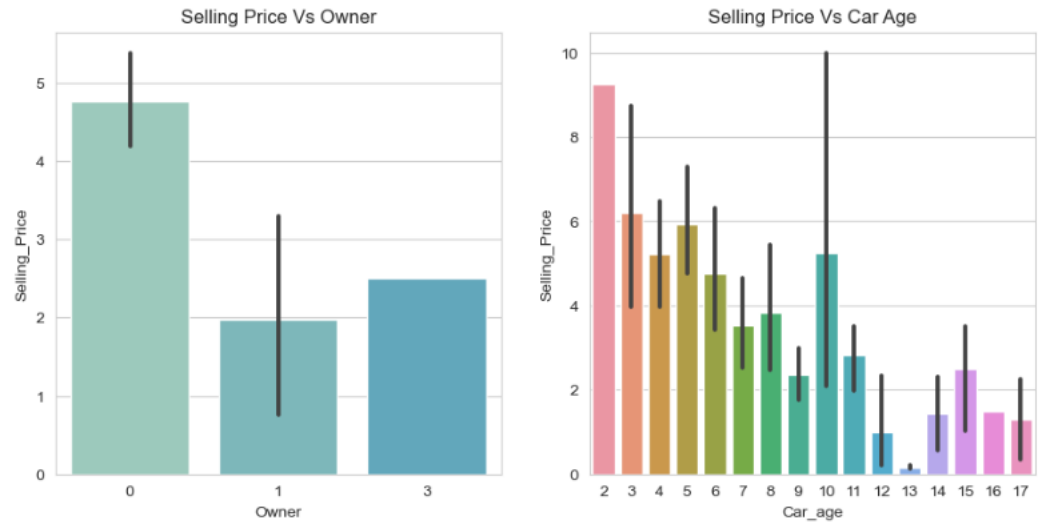
There are three variables: Mileage, EngineV (presumably Engine Volume), and log_price. The color intensity reflects the strength of the correlation between these variables. The correlation coefficient is a measure that determines the degree to which two variables move in relation to each other, with values ranging from -1 to 1. A value of 1 implies a perfect positive correlation, -1 a perfect negative correlation, and 0 no correlation.

Looking at the heatmap:

There is a strong negative correlation between Mileage and log_price, indicated by a darker color and a value of approximately -0.61. This suggests that as the mileage on a vehicle increases, the logarithm of its price tends to decrease.

The correlation between EngineV and log_price is moderately positive, around 0.41, suggesting that vehicles with larger engine volumes tend to have higher prices.

There is a negligible correlation between Mileage and EngineV, shown by the neutral color and the correlation coefficient being close to 0. This indicates that there is no significant linear relationship between a vehicle's mileage and its engine volume.



In the first chart, "Selling Price Vs Owner," the selling price is plotted against the number of previous owners a vehicle has had. The categories of owners are labeled as 0, 1, and 3. It can be observed that vehicles with 0 previous owners, which are likely new or like-new, fetch the highest selling prices, whereas the selling price seems to decrease with the number of owners, with vehicles that have had 3 previous owners exhibiting the lowest selling prices. This trend indicates that the perceived value of a vehicle diminishes with an increasing number of owners, which could be due to factors like increased wear and tear or the perception of reduced reliability.

The second chart, "Selling Price Vs Car Age," shows the selling price in relation to the car age, with the ages ranging from 2 to 17 years. The chart indicates a general trend of decreasing selling price with increasing car age. This is consistent with the common understanding that the value of a car depreciates over time. Interestingly, there is notable variance in selling prices among cars of the same age, which suggests that other factors, in addition to age, also significantly influence the selling price. The error bars indicate the variability of the selling price within each age group, highlighting that while age is a significant factor, individual vehicle conditions and other attributes can cause substantial differences in pricing.

Both charts together suggest that newer, less-used vehicles tend to have a higher selling price, reflecting the common consumer preference for newer and less-owned vehicles in the pre-owned car market. These visualizations are useful for buyers and sellers to understand how the number of previous owners and the age of a car could affect its market value.

The first plot, "Selling Price Vs Present Price," is a scatter plot with a trend line, indicating a positive correlation between a vehicle's present price and its selling price. This suggests that vehicles with a higher present value tend to be sold for more, which aligns with market expectations where a vehicle's current market value influences its resale value.

The middle plot, "Distribution of Selling Price," is a kernel density estimation (KDE) that shows the distribution of vehicle selling prices. The KDE plot reveals a distribution with multiple peaks, suggesting

that the selling prices of vehicles might have several common values or clusters, possibly indicating different groups or segments in the vehicle market, such as economy, mid-range, and luxury vehicles.

The third plot, "Distribution of KMS Driven," is another KDE plot illustrating the distribution of the kilometers driven by the vehicles. The distribution appears to be right skewed, with most of the data concentrated on the left, indicating that a larger number of vehicles have lower mileage, with fewer vehicles having been driven for longer distances. This skewness could reflect that vehicles with lower mileage are more common in the market or that they are more likely to be sold.

The model reveals several key insights:

- The current value of a car significantly influences its selling price, with a direct correlation observed.
- The age of a car negatively impacts its selling price, with older vehicles commanding lower prices.
- Diesel-fueled cars have higher selling prices compared to other fuel types.
- Automatic cars tend to have higher prices than manual cars.
- Vehicles sold by dealers fetch higher selling prices than those sold by individuals.

CONCLUSION:

The project successfully demonstrates the importance of various factors such as a vehicle's age, fuel type, transmission type, and seller type in determining its selling price. These findings provide valuable insights for market participants and contribute to more informed decision-making in the pre-owned vehicle market.