**CAR PRICE PREDICTION**

We are tasked with modeling the price of cars using the available independent variables. This model will enable management to understand precisely how prices fluctuate with these variables. Consequently, they can adjust car designs, business strategies, and other factors to achieve specific pricing levels. Additionally, the model will provide valuable insights for management to comprehend the pricing dynamics in a new market.

Our Data :  
Une image contenant texte, menu, capture d’écran, document

Description générée automatiquement

Check of Outlier:  
Une image contenant capture d’écran, diagramme, texte, Rectangle

Description générée automatiquement

To gain a better understanding of the car price distribution in our dataset, we use a box plot. The box plot provides a visual summary of the minimum, first quartile (Q1), median, third quartile (Q3), and maximum values, as well as any potential outliers.

From the box plot, we observe that the majority of car prices range from approximately 5,000 to 25,000, with a median price around 10,000. There are several outliers with prices above 30,000, indicating some cars are significantly more expensive than the rest.

This exploratory analysis helps in understanding the overall distribution and spread of car prices, which is essential for building an accurate prediction model.

To prepare the dataset for generating the machine learning model, I encoded the categorical variables into numerical format. This is necessary because machine learning algorithms require numerical inputs to process the data effectively.

To understand the relationship between the car prices and the independent variables, I calculated the correlation matrix. This matrix helps in identifying which variables have a significant linear relationship with the car prices.

After computing the correlation matrix, I plotted a heatmap to visualize the correlations among the numerical variables in the dataset. This visualization helps in identifying the strength and direction of relationships between different pairs of variables.

Une image contenant texte, capture d’écran, nombre

Description générée automatiquement

The analysis of the scatter plot and heatmap reveals that several independent variables significantly influence car prices. Variables such as drive wheel type, wheelbase, car length, car width, curb weight, cylinder number, engine size, fuel system, bore ratio, and horsepower exhibit a positive correlation with car prices, indicating that as these values increase, car prices tend to rise. Conversely, city MPG and highway MPG show a negative correlation, meaning higher fuel efficiency in city and highway driving is associated with lower car prices. These insights are essential for understanding the key factors that impact car pricing and will be crucial in developing an accurate predictive model.

I'll create a machine learning model using multiple variables for predicting car prices using Linear Regression. The independent variables influencing the car price include 'drivewheel', 'wheelbase', 'carlength', 'carwidth', 'curbweight', 'cylindernumber', 'enginesize', 'fuelsystem', 'boreratio', 'horsepower', 'citympg' (negatively correlated), and 'highwaympg' (negatively correlated). We'll employ Linear Regression for this task.

Une image contenant texte, capture d’écran, Police, ligne

Description générée automatiquement

In conclusion, the analysis of the results indicates that the machine learning model developed for car price prediction using multiple variables and linear regression performs admirably. With a score of 0.83, the model demonstrates a robust ability to predict car prices accurately. Moreover, the predicted values closely align with the actual values, affirming the model's efficacy. These findings suggest that the created machine learning model is not only effective but also reliable for car price prediction tasks.