**Project Report: Predicting Car Price**

**Introduction**

The objective of this project is to predict the price of a car model using a dataset containing various columns, including "Price," "Age\_08\_04," "KM," "HP," "cc," "Doors," "Gears," "Quarterly Tax," and "Weight." This report outlines the methodology and steps taken to achieve this objective.

**Data Preprocessing**

**Data Collection:**

The initial step involved data collection, which included copying the dataset.

**Feature Selection:**

To focus on the analysis objective, we selected the columns relevant to the price prediction: "Price," "Age\_08\_04," "KM," "HP," "cc," "Doors," "Gears," "Quarterly Tax," and "Weight."

**Data Quality Check:**

A thorough inspection of the data revealed no missing values in the selected columns, ensuring that the dataset was complete.

**Exploratory Data Analysis (EDA)**

**Visual Representation:** The first step in EDA was to visualize the relationships between variables. the correlation between the variables can be visualized with the help of scatterplot , barplot, lineplot & regplot.

**Correlation Analysis:**

A heatmap was used to identify correlations between the selected columns and the "Price" column. The heatmap showed that "Price" had a strong positive correlation with "Age," "Weight," "Hp," and "Km," while being less correlated with "Quarterly Tax," "Gears," "Doors," and "cc."

**Checking for outliers:**

The analysis identified a significant number of outliers in the"Age", "Km", "Hp", "cc", "Gears", "Quarterly\_Tax", "Weight", "Price" column. The price column A total of 110 outliers were detected, accounting for approximately 7.66% of the data. The outliers were subsequently imputed using an upper threshold value to prepare the data for modeling.

**Model Planning and Execution**

**Independent and Dependent Variables:**

For the predictive modeling, "Age\_08\_04," "Km," "Hp," and "Weight" were selected as independent variables, and "Price" was chosen as the dependent variable.

**Data Splitting:**

The data was divided into training and testing sets using the "train\_test\_split" method with an 80:20 split ratio.

**Cross validation: The training dataset putting 30% data to cross validation is making accurate predication on unseen data.**

**Model Training and Evaluation:**

A prediction model was built to estimate car prices. The model achieved an R-squared (R2) score of **87.15** indicating the proportion of variance in the dependent variable explained by the independent variables. The Adjusted R2 score of **86.7**8 The Root Mean Squared Error (RMSE) value was calculated to be **1073.72.** The mean squared error(MSE) value was calculated to be1152890. The mean absolute error(MAE) value was calculated to be **824.68.**

**Conclusion**

This project successfully achieved the objective of predicting car prices. Key steps included data preprocessing, exploratory data analysis, and predictive modeling. The chosen independent variables ("Age\_08\_04," "Km," "Hp,", "cc", "Gears", "Quarterly\_Tax" and "Weight") were found to have a significant impact on car prices. Furthermore, the analysis addressed the presence of outliers, ensuring a more accurate predictive model.

The results indicate that the model can predict car prices with an R2 score of **87.15**. The findings from this analysis can be utilized to make informed decisions in the context of the automotive industry.