1. Introduction

Problem Definition:

The project aims to predict car prices based on various features provided in the dataset, such as make, body style,

wheel base, engine size, horsepower, city mileage, highway mileage, and others.

Objective:

Develop a predictive model that can accurately estimate a car's price to assist stakeholders in making informed

decisions. This could be useful for buyers, sellers, and manufacturers to understand the price determinants and set

pricing strategies.

2. Data Exploration and Visualization

The dataset consists of various attributes related to cars, both numerical and categorical. Initial visualizations

provided insights into the relationships between different features and the target variable (price). Notable

observations included the impact of the make of the car and engine size on price, as well as the general correlation

trends between features like horsepower, engine size, and price.

3. Data Preprocessing

The preprocessing involved handling categorical variables through Label Encoding and scaling the numerical features

using Robust Scaling (chosen for its robustness to outliers), Min-Max Scaling, and Standardization to ensure that

the variance within features does not skew the model.

4. Feature Engineering

PCA was implemented to reduce dimensionality and create principal components that retain the majority of the

information from the original features. 14 principal components were selected as they explained approximately 95%

of the variance in the dataset.

5. Model Development

Model Selection: Linear Regression was chosen for its interpretability and simplicity as a starting point.

Model Training: The model was trained using the 14 principal components derived from PCA.

Model Evaluation: The model achieved an R2 Score of 0.839, indicating a reasonable fit and predictive capability

with the given principal components.

6. Conclusion

The model demonstrates a decent ability to predict car prices, providing a foundation upon which further

optimizations and explorations can be built. The use of PCA allowed for a reduction in dimensionality while

retaining the majority of the variance, which is crucial for predictive accuracy.

7. Ethical Considerations and Compliance

Ensuring ethical use of the model and compliance with relevant data protection and user privacy laws is crucial.

The model should be tested for biases and fairness, and usage should adhere to ethical and legal guidelines.

8. Next Steps

Further steps could involve trying different models, optimizing hyperparameters, and exploring alternative feature

engineering and selection methods to enhance predictive performance.