Used Car Price Prediction: End-to-End Report

# 1. Introduction

This report documents the entire workflow of a used car price prediction project. The objective is to build a regression model to predict car prices based on various features, optimize the model using Optuna, evaluate its performance, and visualize the results in Power BI.

# 2. Data Preprocessing

The dataset included several numerical and categorical features. We dropped irrelevant columns such as 'DealerName', 'Variant', and 'ManufactureDate'. Categorical columns were encoded using OneHotEncoder and numerical columns were scaled using StandardScaler.

# 3. Feature Correlation

We computed correlation values with the target variable 'Price'. Important features included 'ModelYear' (positively correlated), 'CarAge' (negatively correlated), and 'QualityScore'.

# 4. Model Building with XGBoost and Optuna

We used the XGBoost Regressor wrapped inside a pipeline and optimized its hyperparameters using Optuna. The objective was to minimize the RMSE on the validation set.

# 5. Model Evaluation

After training with the best parameters, we evaluated the model:  
✅ Final RMSE: ~181,765  
✅ Final R² Score: ~0.8619  
These results indicate a strong model with high predictive performance.

# 6. Feature Importance

We visualized feature importance using XGBoost's `plot\_importance`. Top contributing features (by gain) included 'ModelYear', 'QualityScore', and 'Kilometer'.

# 7. Power BI Integration

We exported the processed dataset with predictions to a CSV file and loaded it into Power BI. In Power BI, we created visualizations to display predicted vs actual prices, distributions, and feature-level breakdowns. An 'Index' column was added to help with merging tables.

# 8. Conclusion

This project demonstrated a complete machine learning pipeline from preprocessing to deployment visualization. The model achieved strong accuracy and was integrated with Power BI for business-friendly insights.