Summary of Findings: Analysis of Used Car Prices

Objective

The objective of this analysis was to identify the key factors influencing used car prices and to build predictive models to assist dealerships in refining inventory and pricing strategies. This report provides insights derived from data preprocessing, modeling, and interpretation of results.

Key Findings

1. Factors Influencing Car Prices

- **Mileage (Odometer Reading):**
    - A strong inverse relationship exists between mileage and price, with higher mileage resulting in lower prices.
    - **Recommendation:** Dealers should prioritize low-mileage vehicles as they command higher prices and appeal to buyers looking for well-maintained cars.
- **Manufacturer and Model:**
    - Certain manufacturers and models, especially luxury brands (e.g., BMW, Mercedes) and popular models (e.g., Toyota Corolla, Honda Civic), have significantly higher price ranges.
    - **Recommendation:** Focus inventory acquisition on high-demand brands and models for better profitability.
- **Condition:**
    - Cars in "excellent" condition sell for substantially higher prices.
    - **Recommendation:** Highlight vehicle condition in marketing materials to justify premium pricing.

2. Model Performance and Insights

Three regression models were implemented: **Linear Regression**, **Ridge Regression**, and **Gradient Boosting Regression**.

Linear Regression

- **Performance Metrics:**
    - MAE: $3,098.48
    - RMSE: $4,742.28

    - R^2: 0.86
- **Insights:**
    - Captures 86% of the variance in car prices, making it an effective baseline model.
    - Offers strong interpretability and computational efficiency.

Ridge Regression

- **Best Parameters:** Alpha = 0.1
- **Performance Metrics:**
    - MAE: $3,102.53
    - RMSE: $4,730.04
    - R^2: 0.86
- **Insights:**
    - Slightly reduced overfitting compared to Linear Regression, though improvements were minimal.

Gradient Boosting Regression

- **Best Parameters:** Learning Rate = 0.2, Max Depth = 3, N Estimators = 300
- **Performance Metrics:**
    - MAE: $3,969.35
    - RMSE: $5,581.38
    - R^2: 0.80
- **Insights:**
    - Captured some nonlinear relationships but underperformed compared to simpler models due to limited hyperparameter tuning and computational constraints.

---

3. Evaluation Metrics

- **MAE (Mean Absolute Error):** Measures average error magnitude; used to assess how close predictions are to actual values.
- **RMSE (Root Mean Squared Error):** Penalizes larger deviations, emphasizing accuracy in predictions.
- **R^2 (Coefficient of Determination):** Indicates the proportion of variance explained by the model; higher values represent better fit.

4. Feature Importance

Coefficients from Linear and Ridge Regression models revealed that:

- Mileage and year of manufacture are significant predictors of price.
- Condition, manufacturer, and model type also strongly influence pricing.

Recommendations

1. **Inventory Strategy:**
    - Focus on acquiring low-mileage vehicles from high-demand brands and models.
    - Highlight vehicle condition in sales strategies to justify premium pricing.
2. **Model Deployment:**
    - Use Linear or Ridge Regression for predictive pricing due to strong performance and interpretability.
3. **Future Enhancements:**
    - Incorporate additional features such as fuel efficiency, safety ratings, and maintenance history to improve predictive accuracy.
    - Allocate more computational resources for advanced models like Gradient Boosting to explore complex nonlinear relationships.

Conclusion

This analysis successfully identifies key drivers of used car prices and provides actionable recommendations for dealerships. The models developed offer robust predictive capabilities, enabling better alignment of inventory and pricing strategies with customer preferences and market trends.