Used Car Price Analysis Final Report

What Drives the Price of a Used Car?

Audience: Used Car Dealerships **Objective:** Identify key drivers of used car prices and build predictive models to fine-tune inventory decisions.

---

1. Business Objective

Our goal was to determine **what factors most influence used car prices** and how to use this information to:

- Accurately estimate market value
- Optimize vehicle purchasing decisions
- Identify vehicles with strong resale potential

---

2. Data Overview

We analyzed a dataset with **over 400,000 used car listings** from across the U.S., containing:

- Price, Year, Manufacturer, Model
- Odometer, Condition, Transmission, Cylinders, Fuel type, and more

After cleaning:

- Focused on cars from the last **30 years**
- Removed price and odometer outliers
- Retained records with realistic and complete data

---

3. Feature Engineering

We created powerful new features to better explain price variation:

- `car_age` = 2025 - year
- `miles_per_year` = odometer / car_age
- `price_per_mile` and `price_by_age` = value indicators
- `manufacturer_tier` = luxury vs. standard
- `usage_intensity` = very low to very high usage
- `is_high_power` = 6 or 8 cylinders
- `price_segment` = bucketed price category

---

4. Modeling Approach

We tested 3 regression models:

- **Linear Regression** – easy to interpret

- **Ridge Regression** – avoids overfitting with regularization
- **Random Forest Regression** – non-linear, flexible, accurate

Evaluation Metrics:

- **R^2**: How well the model explains price variance
- **Adjusted R^2**: Corrected for number of features
- **MSE**: Mean squared error of predictions

---

5. Model Performance Summary

Model	Train R^2	Test R^2	Test Adj. R^2	Test MSE	Notes
Linear Regression	0.917	0.917	0.914	0.061	Reliable, interpretable
Ridge Regression	0.917	0.917	0.914	0.061	Adds regularization
Random Forest	1.000	0.999	0.999	0.0001	May be overfitting, but highly accurate

---

6. Key Insights for Dealers

Factors that drive higher used car prices:

- Newer vehicles (low age)
- Low mileage per year
- Luxury brand (e.g., BMW, Lexus)
- 6-8 cylinders (high performance)
- Good condition, clean title, automatic transmission

Red flags that lower price:

- High odometer for age
- Older cars from low-demand manufacturers
- Manual transmission or salvage title

---

7. Actionable Recommendations

- Focus purchasing on **late-model, low-mileage** inventory
- Prioritize **luxury or high-demand brands** with clean titles
- Avoid overpaying for cars with **high usage intensity**
- Use the model to **predict fair price** before acquiring vehicles

---

8. Next Steps

- Deploy Ridge or Random Forest model in pricing system
- Build an internal dashboard for predictions
- Enhance future models with location or seasonal factors
