# **README: Predictive Modeling to identify Used Car Prices**

Github Repo- <https://github.com/ankitphophalia-code/UsedCarPricePrediction>

## **Business Problem**

Used car dealerships need to understand the key factors that drive the prices of used vehicles. With a competitive market and diverse inventory, having data-driven insights can help dealers optimize their inventory and pricing strategies. Our project aims to:

* Identify the primary drivers of used car prices.
* Build robust regression models to predict used car prices.
* Provide actionable recommendations for inventory management and pricing based on model insights.

## **Approach**

We followed the CRISP-DM framework to structure our project, with the following phases:

1. **Business Understanding**
   * Defined the objective: To determine which features most significantly influence used car prices.
   * Framed the task as a supervised regression problem with the target variable being the log-transformed price.
2. **Data Understanding**
   * Explored a dataset of used cars (Kaggle subset with 426K records) to inspect data quality, missing values, outliers, and the distribution of key features.
   * Identified critical attributes such as car age, odometer, manufacturer, fuel type, and regional information.
3. **Data Preparation**
   * Cleaned the data by dropping irrelevant columns (e.g., VIN, raw IDs) and imputing missing values.
   * Engineered new features (e.g., calculated car\_age from the manufacturing year).
   * Applied encoding (one-hot, frequency) to categorical variables.
   * Transformed the target variable (price) using a log transformation to reduce skewness.
   * Scaled numerical features for models sensitive to feature magnitude.
4. **Modeling & Evaluation**
   * Evaluated several baseline regression models (Linear, Ridge, Lasso, Random Forest, Gradient Boosting) using 5-fold cross-validation.
   * Selected RandomForest as it performed well and Performed hyperparameter tuning with GridSearchCV and RandomizedSearchCV on key models
   * Compared models using performance metrics (RMSE, R²) on both cross-validation and test sets.
   * Extracted feature importances from the best-performing models to identify which attributes are most influential in driving used car prices.
5. **Deployment & Reporting**
   * Prepared a comprehensive report summarizing our findings and actionable insights.
   * The final insights are tailored to help dealerships refine their inventory selection and pricing strategies.

## **Summary of Findings**

* **Model Performance:** Tree-based models, particularly Random Forest and Gradient Boosting, outperformed linear models, suggesting that non-linear relationships are key in predicting used car prices. Out of this we chose Random Forest as it has the lowest MSE
* **Key Drivers:**
  + **Car Age:** A major factor affecting price—newer cars tend to command higher prices.
  + **Mileage (Odometer):** Lower mileage is associated with higher vehicle value.
* **Other top 5 Key factors :**
  + Drive type (when not known) , Fuel Type(Diesel) ,Cylinder(4), Transmission(when not known), Condition(Excellent)
* **Actionable Insights:**
  + **Inventory Management:** Focus on vehicles with favorable attributes such as optimal age and low mileage.
  + **Pricing Strategy:** Leverage the insights on influential features to adjust pricing dynamically, aligning inventory with market demand.
  + **Marketing:** Target promotions based on popular manufacturers and fuel types that drive higher prices.

## **Recommendations & Next Steps**

* **Implement the Best Model:** Deploy the tuned Random Forest model for live price predictions to support dynamic pricing decisions.
* **Continuous Improvement:** Regularly update the model with new data to keep predictions accurate and relevant as market conditions evolve.
* **Further Exploration:** Consider incorporating additional features (e.g., economic indicators, seasonal trends) for even finer-grained insights.

By adopting this data-driven approach, used car dealerships can optimize their pricing and inventory strategies, ultimately driving profitability and gaining a competitive edge in the market.