Predicting Avocado Prices: Machine Learning with SVR, Linear Regression & Decision Trees

Unlocking market insights through predictive analytics.

Aradhya Mishra 23BIT0261

Parva Kothari 23BIT0053

Ansari Mohd. Umair 23BIT0292

Thiran Devesh 23BIT0340

Harsh Singh 23BIT0242

Why Predict Avocado Prices?

Market Volatility

Avocado market volatility significantly impacts growers, retailers, and consumers. Prices can fluctuate wildly based on supply, demand, and external factors.

Smarter Decisions

Accurate price forecasts enable smarter supply chain management, optimized pricing strategies, and better inventory planning.

Rich Dataset

A comprehensive Kaggle dataset offers rich historical sales, volume, and regional data, perfect for developing robust predictive models.

Understanding the Dataset

Key Data Points

- Source: Avocado Prices Dataset on Kaggle

   - Features: Date, AveragePrice, Type
(Conventional/Organic), Region, Volumes, Bags, PLU codes

   - Target Variable: AveragePrice

   - Records: ~18,000 (2015–2018)
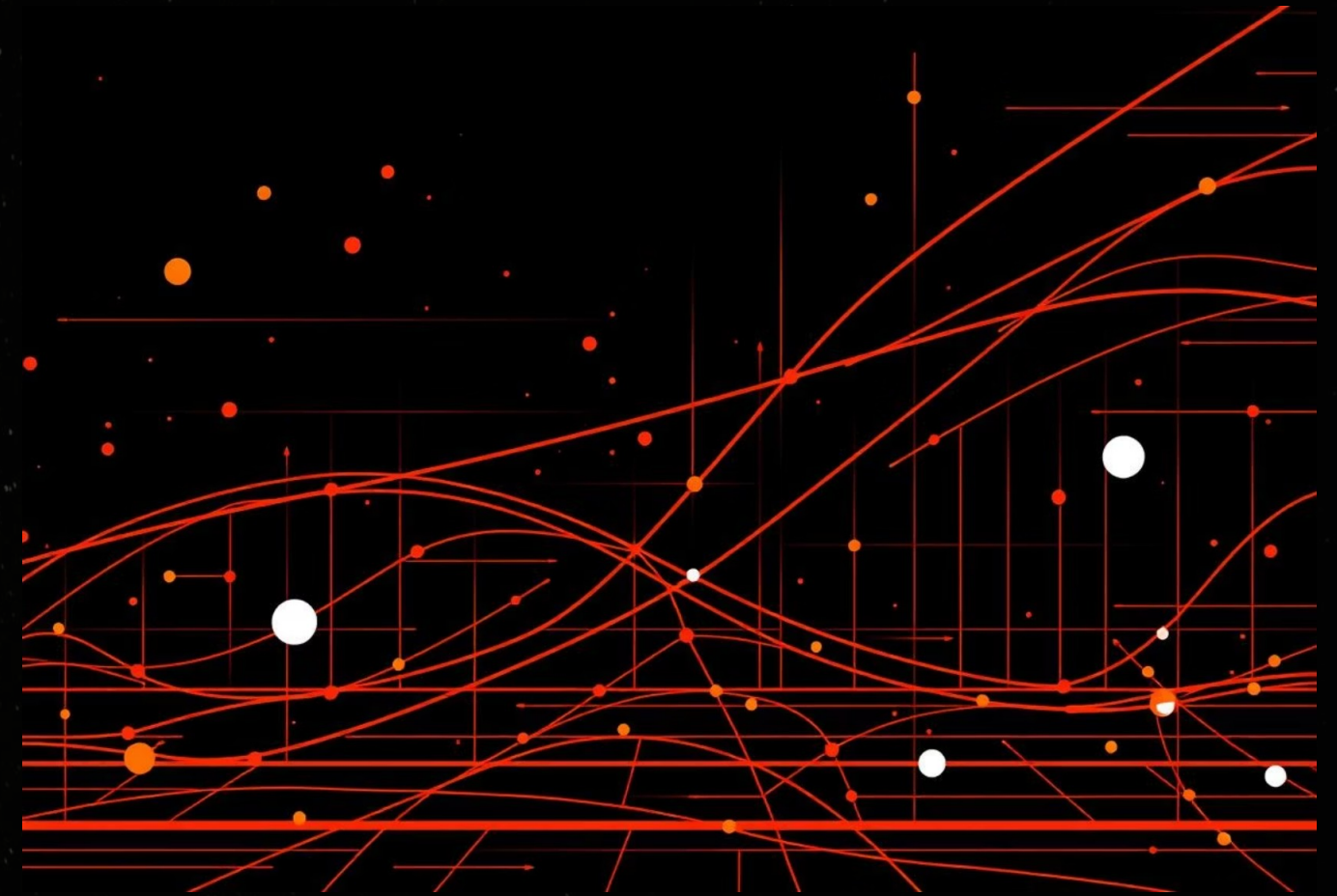
Challenges Ahead

- Missing values need imputation.

- Categorical variables require encoding.

- Time series nature demands specific modeling approaches.

Preprocess and clean the avocado dataset

Column Management

Convert categorical features (type, region) into numeric form.

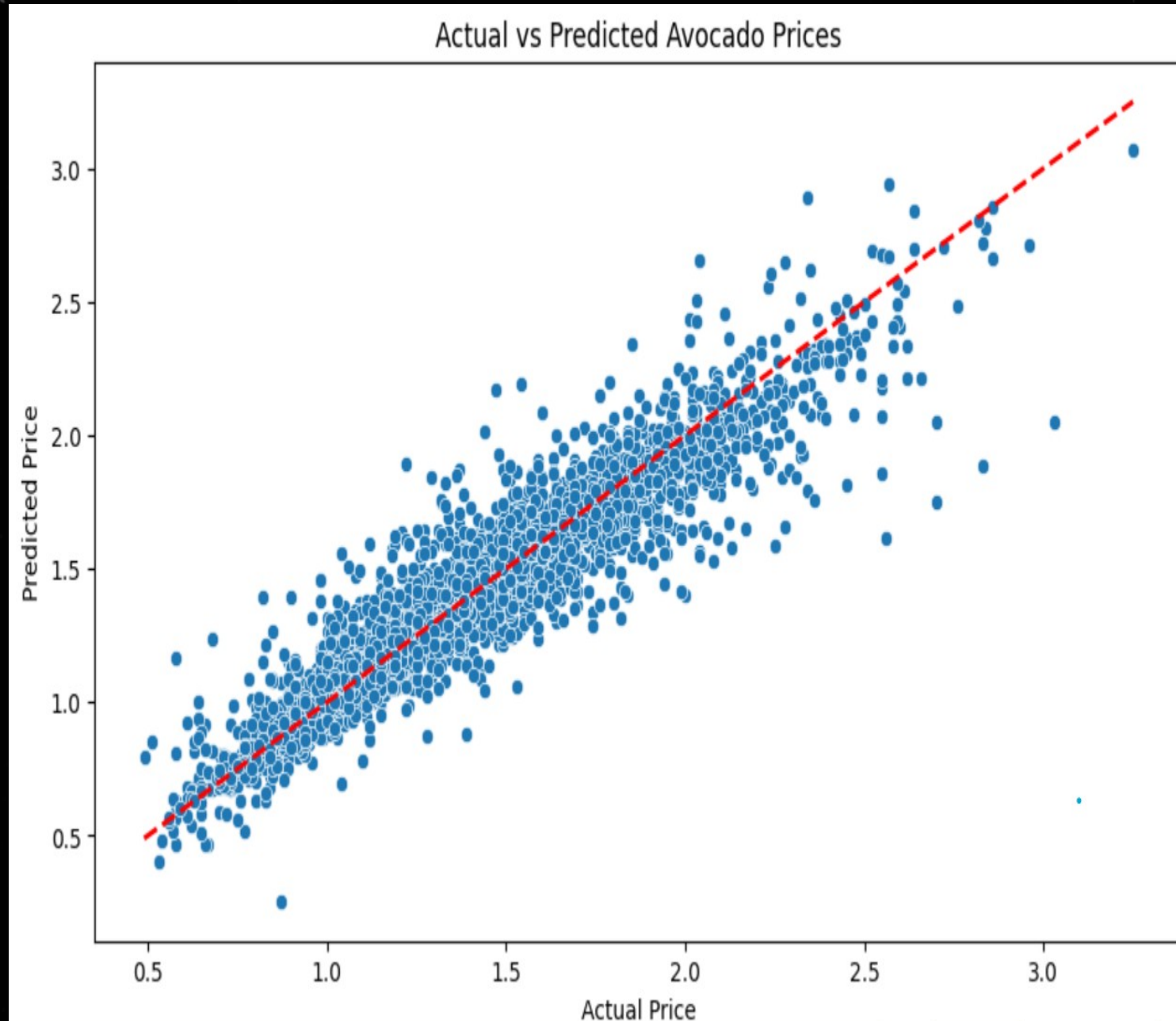## Time-Based Features

month, seasonal sine/cosine encoding

Categorical Encoding

Ratios, PLU volumes, price momentum

Normalization & Outliers

Handled outliers and normalized volume features to ensure model stability and improve performance.

Model 1: Linear Regression (LR)

Actual vs Predicted Avocado Prices

The Foundation

Our baseline model, Linear Regression, helps us understand the fundamental linear relationships within the data.

- **Strength:** Simplicity and high interpretability, making it easy to understand initial trends.

- **Limitation:** Struggles to capture complex non-linear price fluctuations and the inherent seasonality of avocado prices.

# Model 2: Support Vector Regression (SVR)

Advanced Predictive Power

SVR leverages kernel functions to effectively capture intricate, non-linear patterns in the avocado price data.

- **Refined Tuning:** Hyperparameters (C, epsilon, kernel) were optimized using grid search CV for superior accuracy.

- **Robustness:** Offers better handling of outliers compared to Linear Regression, though it is more computationally intensive.

Captures non linear relationships.

Prediction Error Distribution
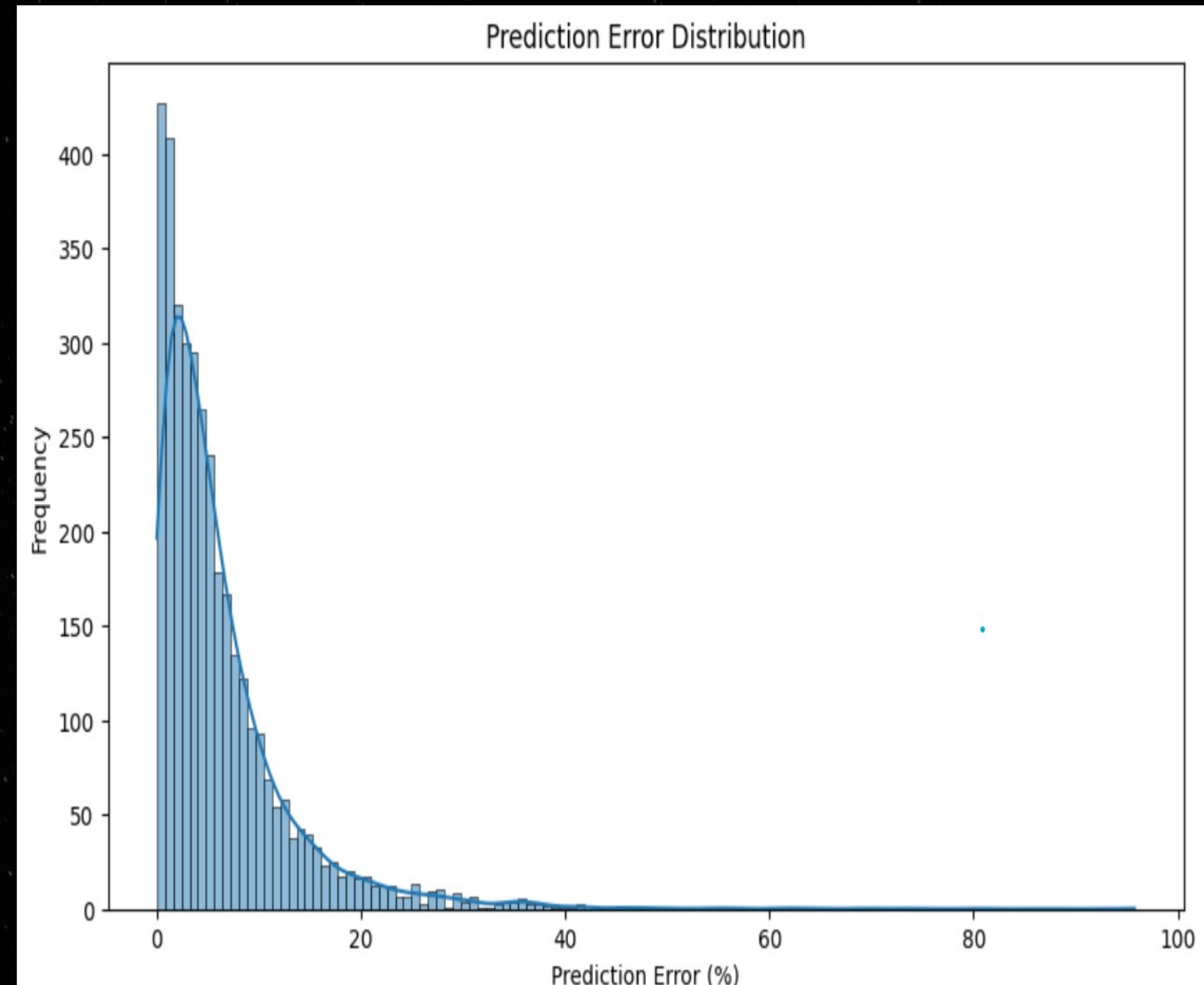
Model 3: Decision Tree Regressor
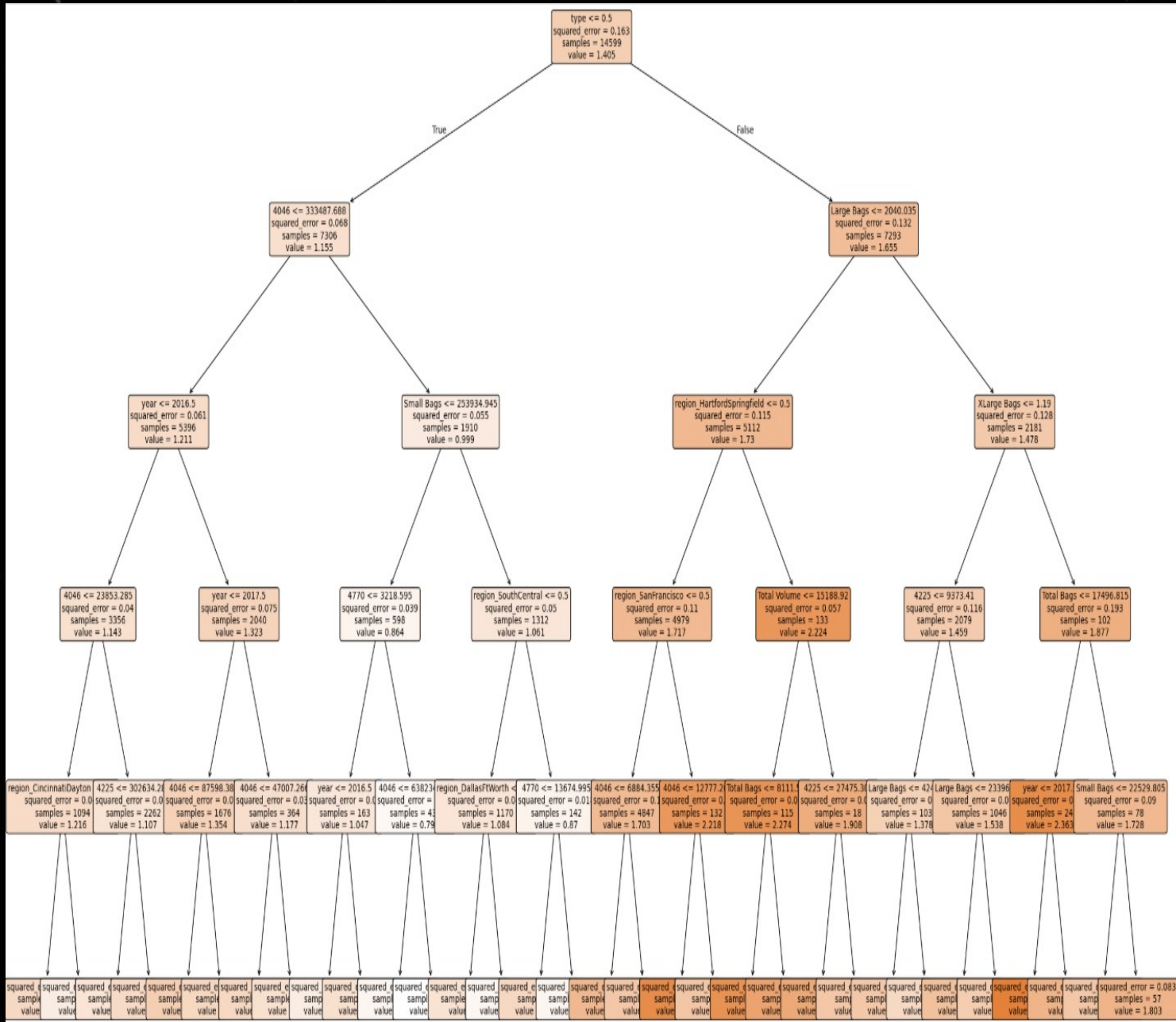
Insightful Segmentation

This tree-based model excels at capturing non-linearities and identifying important feature interactions within the dataset.

- **Interpretability:** Easy to understand with visualizable decision paths, offering clear insights into price drivers.

- **Consideration:** Carries a risk of overfitting if not properly pruned or if its maximum depth is not carefully controlled.

Model Performance Comparison

Model	RMSE	MAE	Adjusted R² Score (Real)
Linear Regression	0.138	0.0954	0.8786 (0.8807)
Support Vector Regression	0.1336	0.0874	0.8870 (0.8889)
Decision Tree	0.2035	0.1375	0.7375 (0.7421)

The Support Vector Regressor (SVR) delivered the best accuracy, successfully balancing bias and variance in its predictions. The Decision Tree followed closely, offering excellent interpretability alongside strong performance. Linear Regression provided a valuable and understandable baseline.

Visualizing Predictions vs Actual Prices

Our visualizations confirm that SVR is particularly effective at capturing seasonal price spikes. The Decision Tree, meanwhile, provides clear insights into how specific volume and regional factors influence avocado prices, showcasing its interpretability.

Actual vs Predicted Avocado Prices
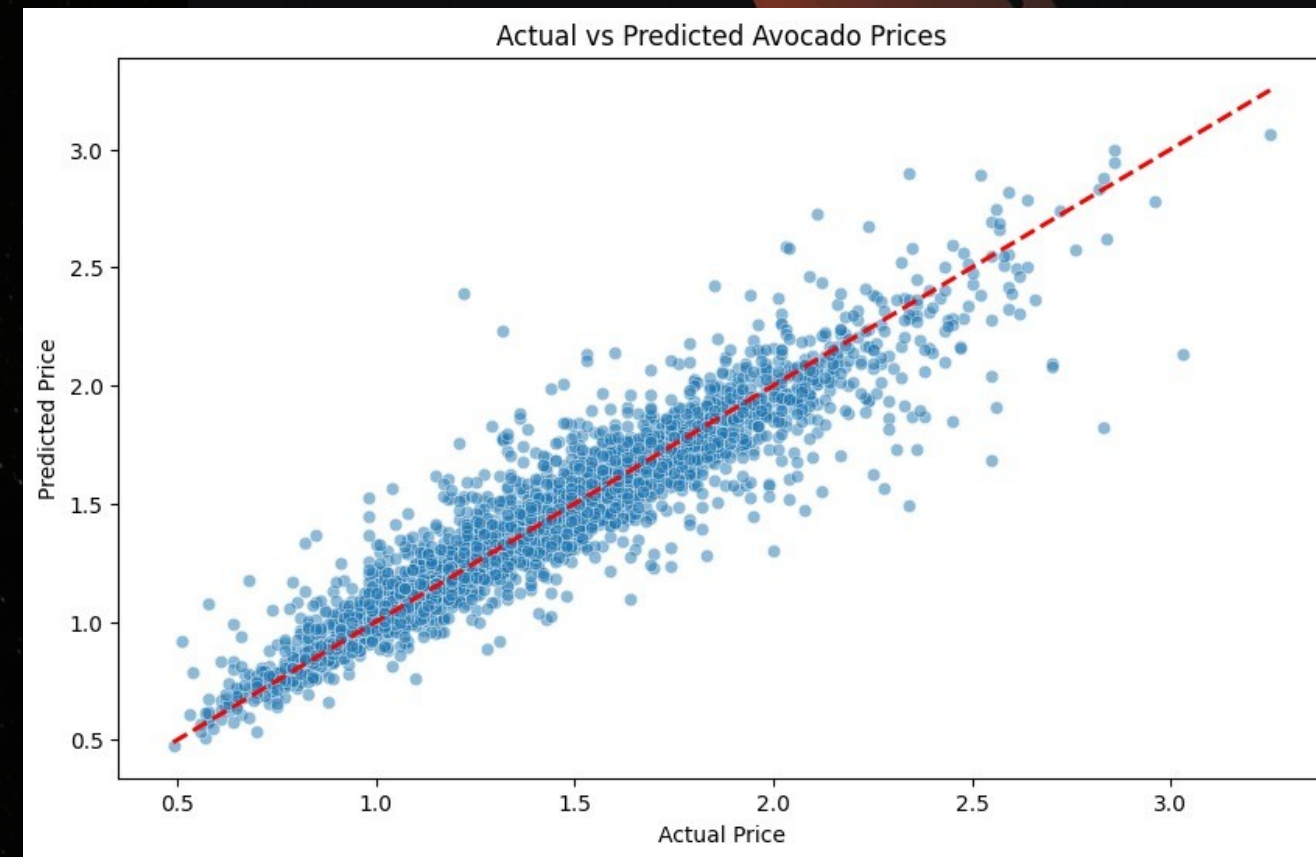
Conclusion & Next Steps

- **SVR: Our Top Performer**

  The Support Vector Regressor (SVR) emerged as the most effective model for avocado price prediction on this dataset.

- **Actionable Insights**

  These predictive insights empower growers and retailers to proactively anticipate price trends, optimizing their strategies.

- **Future Enhancements**

  Preprocessing & feature engineering significantly improved predictions