Model Evaluation Report

Objective: To compare and evaluate the performance of five regression models to predict US regional sales, and select the best-performing model for deployment.

🔍 1. Overview of the Evaluation Strategy

To ensure a robust evaluation, each model was tested using the following metrics:

- **Mean Squared Error (MSE)**: Measures the average squared difference between actual and predicted values. Lower is better.
- **Root Mean Squared Error (RMSE)**: Square root of MSE; provides error in the same units as the target variable.
- **Mean Absolute Error (MAE)**: Measures the average magnitude of the errors in predictions. Less sensitive to outliers than RMSE.
- **R-squared (R^2)**: Indicates the proportion of the variance in the dependent variable explained by the model. Closer to 1 means better fit.

All models were trained and evaluated on the same training and testing splits to ensure fairness in comparison.

⚙️ 2. Models Evaluated

Model	MSE	RMSE	MAE	R^2 Score
XGBoost	94,829.28	307.94	184.98	0.9913
Gradient Boosting	116,016.12	340.61	192.12	0.9894
Random Forest	209,040.68	457.21	241.95	0.9809
Decision Tree	580,253.89	761.74	405.24	0.9470
AdaBoost	2,896,720.22	1701.98	1377.80	0.7356

📈 3. Analysis of Results

✅ XGBoost

- **Best performance** across all metrics.
- Extremely low MSE, RMSE, and MAE, indicating very close predictions.
- **R^2 of 0.9913**, meaning it explains over 99% of the variability in sales data.
- Benefited from **boosting techniques** that reduce bias and variance.

- Well-tuned using hyperparameters such as `n_estimators`, `max_depth`, and `learning_rate`.

⚠️ Gradient Boosting

- Strong performance, second to XGBoost.
- Close RMSE and MAE, but slightly worse generalization power ($R^2 = 0.9894$).
- Slower training time compared to XGBoost.

⚖️ Random Forest

- Performs well but with noticeable error margins.
- Higher MSE and RMSE suggest reduced accuracy.
- **R^2 of 0.9809**, still considered a strong fit.

❌ Decision Tree

- Much higher errors due to overfitting.
- Lacks ensemble robustness.
- R^2 dropped to 0.947, making it unsuitable for precise predictions.

❌ AdaBoost

- Poor performance in this regression context.
- Extremely high MSE and MAE.
- $R^2 = 0.7356$ suggests poor model fit and generalization.

🧠 4. Insights Gained

- **Ensemble methods outperform single estimators** significantly.
- Boosted models (XGBoost and Gradient Boosting) consistently lead to better generalization.
- Tuning hyperparameters played a crucial role in XGBoost's performance edge.
- Model choice should consider both performance and scalability—XGBoost excels in both.

📌 5. Model Training and Hyperparameter Tuning

- **Cross-Validation**: 5-fold cross-validation was used to mitigate overfitting and ensure model generalization.

- **GridSearchCV / RandomizedSearchCV**: Employed to optimize hyperparameters for each model, especially for XGBoost (`max_depth`, `n_estimators`, `learning_rate`, `subsample`, etc.).
- **Early Stopping** (XGBoost): Prevented overfitting by monitoring validation error and halting training when improvement plateaued.

🏁 6. Final Recommendation

Given its **superior performance**, **XGBoost** is selected as the final model for predicting US regional sales. It strikes the best balance between **accuracy, generalization, and computational efficiency**, making it highly suitable for both batch and real-time sales prediction applications.

📊 7. Visual Evidence