Short Summary – E-commerce Customer Churn Analysis

- **Dataset:** 50,000 customers, 25 features (behavioral, transactional, demographic). Target variable: **Churned** (≈29% churn rate).
- **Data Cleaning:**
  - Handled missing values using **median imputation** for numerical columns.
  - **Outliers capped** using IQR method.
  - Categorical variables encoded using **Label Encoding** and later **One-Hot Encoding**.
- **EDA Insights:**
  - Churn is influenced by **login frequency, membership years, days since last purchase, lifetime value, age**, and engagement metrics.
  - Lower engagement and longer inactivity strongly correlate with churn.
- **Modeling:**
  - Trained and compared **10 models** (Logistic Regression, Random Forest, Gradient Boosting, XGBoost, etc.) using scaled features.
- **Best Models (by ROC-AUC):**
  - **Gradient Boosting:** ROC-AUC **0.927** (best overall)
  - **Random Forest:** ROC-AUC **0.925**
  - **XGBoost:** ROC-AUC **0.925**
- **Final Performance (Top Models):**
  - Accuracy ≈ **91%**
  - F1-score ≈ **0.84**
  - Recall for churned customers ≈ **0.79**
- **Feature Importance:**
  - Key drivers: **Lifetime Value, Login Frequency, Days Since Last Purchase, Total Purchases, Session Duration**.
- **Conclusion:**
  - The model effectively predicts churn with strong discrimination.
  - **Engagement and recency-based features** are the most critical for churn prediction.