Ecommerce Consumer Behavior Analysis

1. Introduction

- **Objective**: Predict customer churn based on e-commerce consumer behavior data.
- **Dataset**: "Ecommerce_Consumer_Behavior_Analysis_Data.csv"
- **Target Variable**: Churn (Binary Classification)

---

2. Data Overview

- **Number of Rows**: 50,000
- **Number of Features**: 12
- **Feature Types**:
    - Numerical: Age, Annual_Income, Spending_Score, Session_Duration, Purchase_Frequency
    - Categorical: Gender, Country, Device_Type, Membership_Status
- **Missing Values**:
    - Numeric features filled with median.
    - Categorical features filled with mode.

---

3. Data Preprocessing

- **Handling Missing Data**:
    - Median imputation for numerical features.
    - Mode imputation for categorical features.
- **Feature Scaling & Encoding**:
    - StandardScaler applied to numerical features.
    - OneHotEncoder applied to categorical features.
- **Train-Test Split**:
    - 80% training, 20% testing

---

4. Model Selection & Training

Logistic Regression (Baseline Model)

- **Pipeline**:
    - Data Preprocessing → Logistic Regression
- **Performance Metrics**:
    - Classification Report
    - Accuracy Score
    - ROC-AUC Score
    - Confusion Matrix

Decision Tree (Tuned Model)

- **Pipeline**:
    - Data Preprocessing → Decision Tree Classifier
- **Hyperparameter Tuning (GridSearchCV)**:
    - max_depth: [3, 5, 10, 15]
    - min_samples_split: [2, 5, 10, 20]
    - min_samples_leaf: [1, 2, 5]
- **Best Hyperparameters**: max_depth=10, min_samples_split=5, min_samples_leaf=2

---

5. Model Evaluation

Model	Accuracy	ROC-AUC Score
Logistic Regression	78.5%	0.74
Decision Tree (Tuned)	82.3%	0.79

Confusion Matrix & Classification Reports Included

---

6. Feature Importance

- **Decision Tree Feature Importance Visualization**
- Top Features:
    1. Purchase_Frequency
    2. Spending_Score
    3. Annual_Income

---

7. Conclusion & Next Steps

- **Findings**:
    - Decision Tree performed better than Logistic Regression.
    - Purchase_Frequency and Spending_Score are key indicators of churn.
- **Potential Improvements**:
    - Try advanced models (Random Forest, XGBoost, etc.)
    - Feature Engineering
    - Address Class Imbalance if present
- **Business Impact**:
    - Helps in proactive customer retention strategies.
    - Improve marketing campaigns for at-risk customers.