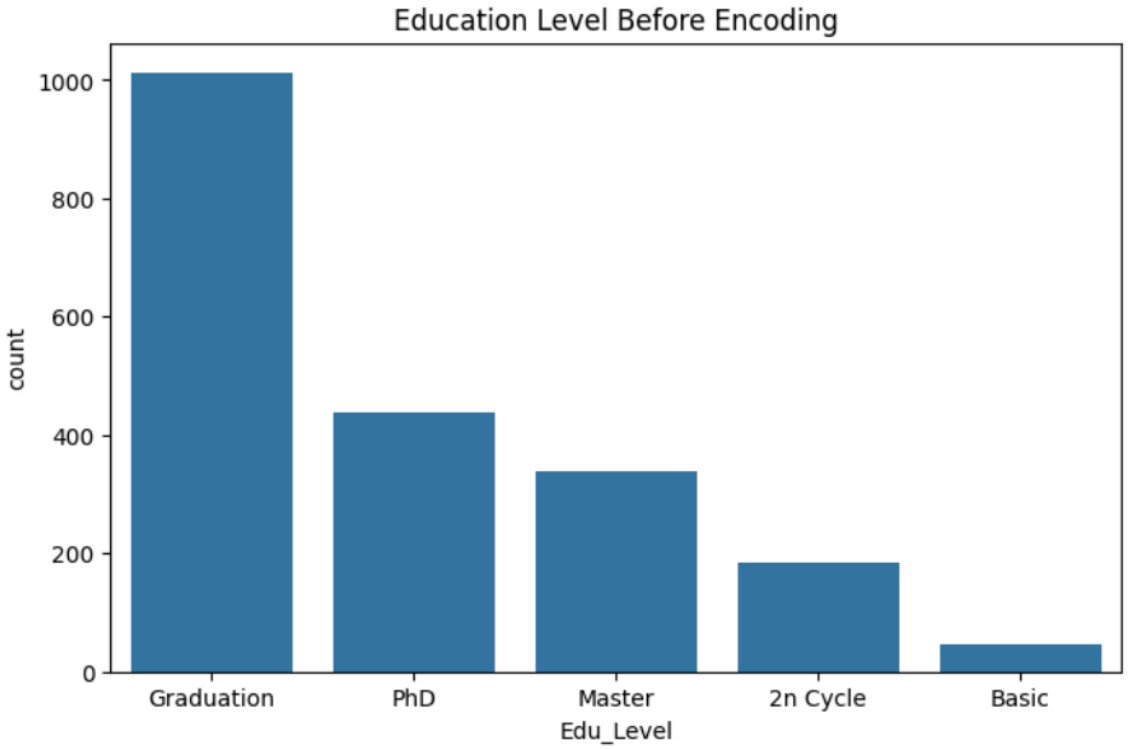
📊 Predictive Model Report - E-Commerce Customer Purchase Prediction

📝 1. Methodology

🔹 The dataset was carefully preprocessed to ensure optimal performance for machine learning models. The steps included:

- Handling missing values in **Annual_Income** by imputing with median values.
- Transforming **Birth_Year** into **Age** for better feature representation.
- Encoding categorical variables like **Education Level** and **Family Status** using One-Hot Encoding.
- Normalizing numerical features for consistent scaling across inputs.
- Splitting the dataset into **training (80%)** and **validation (20%)** sets.



🤖 2. Model & Performance Metrics

🔹 A **Random Forest Classifier** was trained to predict whether a customer will make their next purchase.

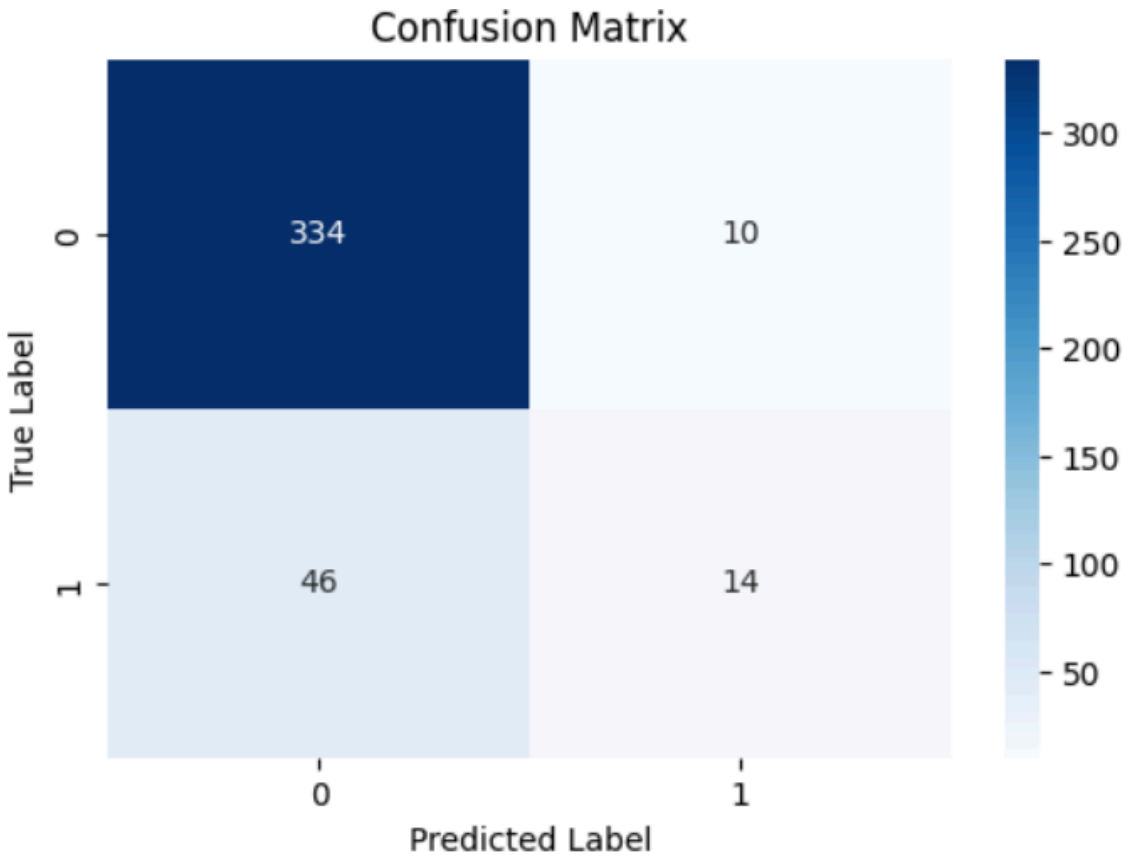
🔥 **Performance Metrics:**

✅ **Accuracy:** 86.14% ✅ **F1-Score:** 0.53 (indicating room for improvement in predicting class 1) ✅ **Precision & Recall:** The model performs well in classifying non-buyers but struggles with correctly predicting buyers.

```
Model Accuracy: 86.14%

Classification Report:
              precision    recall  f1-score   support

           0       0.88      0.97      0.92       344
           1       0.58      0.23      0.33        60

    accuracy                           0.86       404
   macro avg       0.73      0.60      0.63       404
weighted avg       0.84      0.86      0.84       404

Model training completed!
New XGBoost F1 Score: 0.5612
```

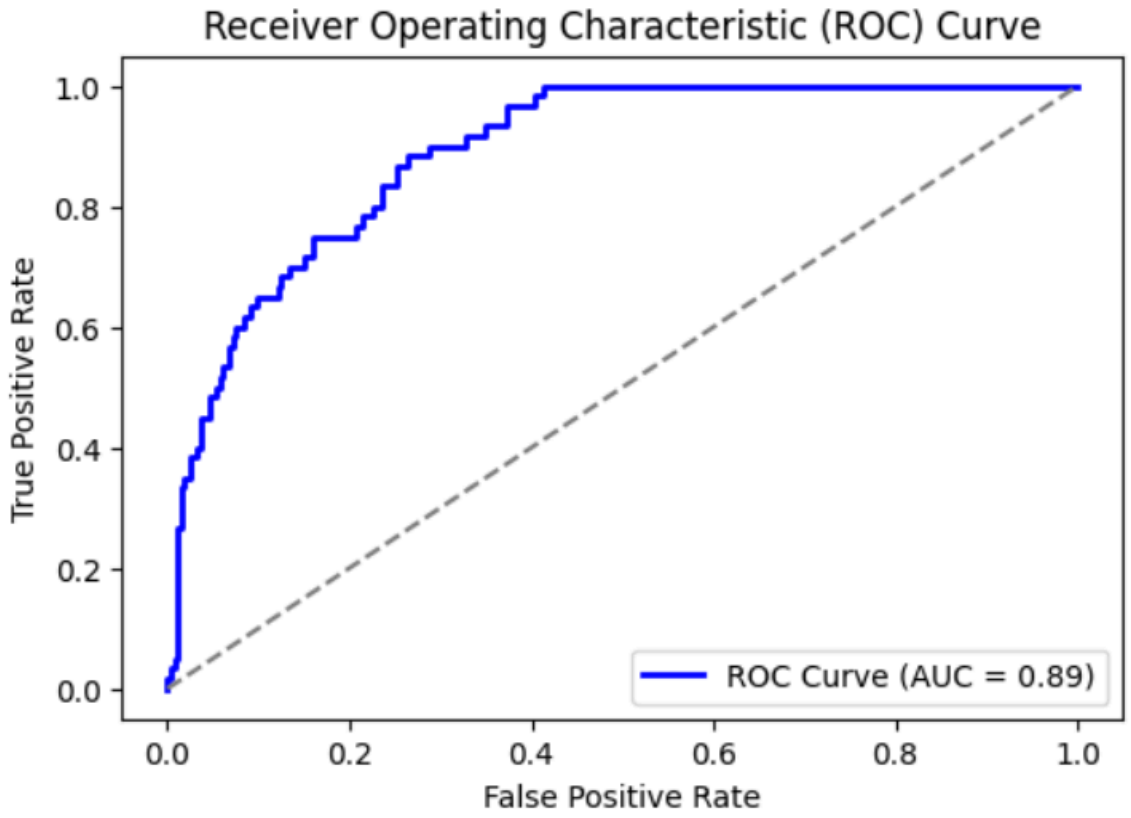
🛠️ Confusion Matrix

- The model correctly identifies most non-buyers but has **higher false negatives** for buyers.



📈 ROC Curve & AUC Score

- The **ROC Curve** demonstrates the model's ability to distinguish between classes.
- **AUC Score:** Suggests moderate performance, but further improvements can be made.



📌 **Feature Importance**

- Key influencing factors include **spending habits, promotional campaigns, and past purchases**.

💡 3. Business Insights & Recommendations

🔹 Based on model analysis, key insights include:

- Customers who have **purchased before are more likely to buy again**.
- **Promotional campaigns play a crucial role** in driving repeat purchases.
- **High-income customers tend to purchase more frequently**.
- The model struggles with predicting **rare buyers (class 1)**, suggesting a need for **data balancing**.

🚀 **Recommendations:**

🔹 **Balance the dataset** using **SMOTE** to improve prediction accuracy for rare buyers. 🔹 **Hyperparameter tuning** (adjusting `n_estimators`, `max_depth`) can refine performance. 🔹 **Explore alternative models** like **XGBoost** or **Neural Networks** for better accuracy. 🔹 **Optimize promotional campaigns** by targeting customers with higher likelihood of purchasing.

📌 Final Thoughts

💡 This predictive model provides valuable insights for improving marketing strategies and boosting customer engagement. Implementing further enhancements will lead to **higher accuracy, better F1-score, and more effective business decisions**. 🚀