Customer Behavior Analysis: Predicting Order Cancellations

---

1. Business Problem Statement

Customer order cancellations impact both revenue and operational efficiency. The objective is to predict whether an order will be cancelled or completed using historical order data, and identify the key factors influencing cancellations. This will help the business optimize fulfillment strategies and reduce cancellation rates.

---

2. Objectives

- Predict if an order will be **Cancelled** or **Completed**.

- Understand key drivers contributing to cancellations.

- Provide actionable recommendations to reduce cancellations.

---

3. Dataset Overview

- **Size:** 20,000+ orders

- **Target Variable:** Order Status (Cancelled / Completed)

- **Features Used:**

    - Demographics: Age Group, Gender, Loyalty Member

    - Order details: Product Type, Quantity, Add-on Total, Total Price

○ Purchase info: Payment Method, Shipping Type, Purchased Month, Purchased Weekday, Purchased Year

○ Rating feedback: Rating Category

---

4. Data Preprocessing

● Converted categorical columns using Label Encoding and One-Hot Encoding.

● Mapped weekday names to numeric day numbers.

● Created **Order Status Binary** column for classification (1 = Completed, 0 = Cancelled).

● Split dataset using stratified train-test split (80%-20%).

---

5. Exploratory Data Analysis (Key Findings)

● **Higher Cancellation Trends Observed In:**

○ **Shipping Type:** Standard Shipping

○ **Age Group:** Older customers

○ **Payment Method:** Credit Card

○ **Product Type:** Mobile

● **Cancellations were frequent on weekends** and during specific months.

---

6. Modeling

Models Used:

- Logistic Regression (class_weight='balanced')

- Random Forest Classifier (class_weight='balanced')

Best Performance:

- **Random Forest**

    - **Accuracy:** 64.5%

    - **Precision (Cancelled):** 68%

    - **Recall (Cancelled):** 53%

    - **F1 Score (Cancelled):** 60%

---

7. Feature Importance (From Random Forest)

Top influencing features:

1. **Shipping Type**

2. **Product Type**

3. **Age Group**

4. **Payment Method**

5. **Total Price**

---

8. Business Recommendations

Based on model insights and EDA:

1. **Revise Shipping Policies:** Prioritize fast or premium shipping options for mobile devices or older customers.

2. **Payment Gateway Alerts:** Monitor and possibly require verification for credit card payments in high-risk segments.

3. **Product-Specific Interventions:** Run targeted engagement or follow-up for mobile-related orders before processing.

4. **Customer Segmentation:** Introduce loyalty incentives or confirmations for older customer segments.

5. **Proactive Order Confirmation:** Add a pre-shipment confirmation step for high-risk orders.

---

9. Limitations

- Imbalance in class distribution despite using class_weight='balanced'

- No information on real-time customer feedback or delivery issues

- Potential biases in self-reported demographic or rating data

---

10. Conclusion

The predictive model provides reliable insights into order cancellation behavior, helping the business take proactive steps in improving customer satisfaction and reducing revenue loss due to cancellations.