E-Commerce Customer Analysis Report

1. Executive Summary

This report presents a comprehensive analysis of an e-commerce customer dataset, covering data engineering, exploratory analysis, predictive modeling, and deployment strategies. The goal is to understand customer behavior, predict spending patterns, and classify customers into value tiers for targeted marketing.

---

2. Dataset Overview

- Source: `marketing_campaign.csv` (2,240 customers, 29 features)
- Key Features:
  - Demographics: `Year_Birth`, `Education`, `Marital_Status`, `Income`
  - Behavioral: `Recency`, `NumWebVisitsMonth`, `NumDealsPurchases`
  - Spending: `MntWines`, `MntFruits`, `MntMeatProducts`, etc.
  - Campaign Response: `AcceptedCmp1-5`, `Response`

Data Cleaning & Preprocessing

- Handled Missing Values: Imputed missing `Income` values with the median.
- Outlier Treatment: Capped extreme `Income` values at the 99th percentile.
- Feature Engineering:
  - `TotalSpend`: Sum of all spending categories.
  - `CustomerValueTier`: Classified customers into `Premium`, `Gold`, `Silver`, `Bronze`.
  - `CustomerAge`: Derived from `Year_Birth`.
  - `AverageSpending`: Mean spending across categories.

---

3. Exploratory Data Analysis (EDA)

Key Insights

1. Income Distribution:
   - Median income: $51,381
   - Right-skewed distribution (some high-income outliers).

2. Customer Segmentation:
   - `CustomerValueTier` Distribution:
     - Bronze: 903
     - Premium: 602
     - Gold: 393
     - Silver: 342

3. Spending Patterns:
   - Married customers spend more on wine and meat.

  - Older customers (60+) tend to spend more overall.

4. Correlations:
  - `Income` positively correlates with `TotalSpend`.
  - `NumWebVisitsMonth` negatively correlates with `NumStorePurchases`.

Visualizations
- Income & Age Distributions (Box plots & histograms).
- Spending vs. Age (Regression plot).
- Marital Status vs. Spending (Bar chart).
- Correlation Heatmap (Numerical variables).

---

4. Predictive Modeling
A. Regression Model (Predicting `TotalSpend`)
- Model: Linear Regression (with preprocessing pipeline).
- Performance:
  - RMSE: 341.11
  - R^2 Score: 0.67
- Interpretation:
  - The model explains 67% of variance in spending.
  - Predictions are within ±$341 of actual spend.

B. Classification Model (Predicting `CustomerValueTier`)
- Model: Random Forest (tuned via GridSearchCV).
- Best Parameters:
  - `max_depth=10`, `min_samples_split=2`, `n_estimators=200`
- Performance:
  - Accuracy: 69%
  - Precision/Recall:
    - Bronze: 0.78 / 0.92
    - Premium: 0.72 / 0.87
    - Gold & Silver: Lower performance (class imbalance).
- Key Features:
  - `Income` (Most important).
  - `CustomerAge`, `NumWebVisitsMonth`, `Recency`.

---

5. Model Deployment
Packaging & API Integration
- Saved Model: `customer_value_model.pkl` (Pickle format).
- Sample Prediction:

Predicted Customer Value Tier: Gold

- Deployment Plan:
  1. Microservice: Flask/FastAPI for API endpoints.
  2. Integration: Connect with CRM/marketing tools.
  3. Monitoring: Track model drift & retrain periodically.

---

6. Business Recommendations

1. Target High-Value Customers:
   - Focus on Premium & Gold tiers with personalized offers.
2. Improve Classification for Silver/Gold:
   - Collect more data or use oversampling techniques.
3. Campaign Optimization:
   - Customers with higher web visits but lower in-store purchases may need incentives.
4. Dynamic Pricing:
   - Adjust discounts based on predicted spending behavior.

---

7. Next Steps

- A/B Testing: Validate model impact on marketing campaigns.
- Feature Enhancement: Incorporate purchase frequency & customer lifetime value (CLV).
- Real-time Prediction: Deploy model in cloud (AWS/GCP) for scalability.

Conclusion

This analysis provides actionable insights into customer segmentation and spending behavior. The predictive models enable targeted marketing strategies, improving ROI on customer acquisition and retention efforts. Future work includes refining classification accuracy and deploying the model in a production environment.

Appendix: Full code

Prepared by: Daniel Muthama
Date: 9/8/2025