Customer Shopping Behavior Analysis: Optimizing Operational Expenditure (Opex)

1. Summary: Driving Margin Restoration

This project provides a comprehensive operational analysis focused on efficiency and profitability by identifying areas of wasted spending.

The key finding is the **inefficiency of blanket discount strategies**, which are eroding margin without increasing Average Order Value (AOV). This provides an immediate, actionable opportunity to restore profit margins.

Metric	Result	Strategic Implication (Opex Control)
Discount Erosion	**[1.42]% Margin Lost**	**IMMEDIATE Opex Reduction:** Stop blanket promotions; discounts do not increase basket size.
Value Concentration	**Top 25% of customers** drive **[38]% of revenue**	**TARGETED Opex:** Reallocate retention budget exclusively to the high-value segment.
Total Revenue Lost	**$[1427.77]**	**QUANTIFIED Opex Cost:** This is the immediate financial amount to be saved by optimizing the discount strategy.

2. Business Problem & Project Goals

Business Problem

How can the company strategically restructure its Operational Expenditure (Opex) on marketing and promotions to maximize profit margins and increase Customer Lifetime Value (CLV) by identifying the most inefficient spending areas and the most valuable customer groups?

Project Goals

1. **Quantify Discount Inefficiency:** Prove or disprove the effectiveness of current discount strategies.
2. **Segment Customer Base:** Develop a robust CLV segmentation model to identify high-value customers.
3. **Provide Actionable Strategy:** Deliver recommendations for **Promotional Opex** (pricing) and **Marketing Opex** (targeting).

3. Data & Methodology

A. Tools & Technologies

- **Data Wrangling & Modeling:** Python (Pandas, NumPy)
- **Visualization:** Matplotlib, Seaborn, (Optional: Altair for interactivity)
- **Database Integration:** MySQL (via MySQL Workbench)
- **Presentation:** Jupyter Notebook, GitHub

B. Data Cleaning and Feature Engineering

The initial dataset was cleaned to ensure integrity and consistency across columns.

- **Handling Missing Data:** Missing `Review_Rating` values were imputed using the **median rating per product category**.
- **Frequency Mapping:** The categorical `Frequency_of_Purchases` (e.g., 'Weekly', 'Monthly') was converted into a **numerical score** for use in the CLV proxy. (This step required robust error handling for missing values using the mode).
- **Column Standardization:** All column headers were standardized for database compatibility.

C. Customer Lifetime Value (CLV) Proxy Segmentation

To segment the customer base, we created an RFM-like model based on key purchase metrics:

1. **Total_Spend** (`Purchase_Amount_(USD)` SUM)
2. **Purchase_Frequency** (Mapped numerical score MEAN)
3. **Loyalty** (`Previous_Purchases` MEAN)

Segment	Contribution	Strategic Focus
High Value	Top 33% of total spenders	**Retention, Upselling** (Focus Marketing Opex)
Mid Value	Middle 33% of spenders	**Growth, Cross-selling** (Focus Acquisition Opex)
Low Value	Bottom 33% of spenders	**Reactivation, Cost Optimization** (Minimize Retention Opex)

4. Key Analytical Findings

A. Discount Efficiency (Promotional Opex) - The Most Critical Insight

The analysis confirms that the current discount strategy is fundamentally flawed and drives down profitability.

Metric	Result	Undiscounted AOV	Discounted AOV
Average Order Value (AOV)	**AOV is higher without a discount.**	**$[60.13]**	**$[59.28]**

Conclusion: Discounts are *not* successfully driving customers to increase their basket size; they primarily serve to **pull future purchases forward or erode margin** on items customers were already willing to buy.

B. Customer Value & Targeting (Marketing Opex)

The segmentation confirms the Pareto Principle is active in the business:

- The **High Value** customer segment (top 25% by spend) contributes **[38]%** of the total revenue.
- **Age:** Customers aged **40+** contribute the highest total revenue, establishing them as the core base to be maintained.

C. Product & Inventory (Procurement/Service Opex)

- **Core Category:** The **Clothing** category generates the highest overall revenue.
- **Quality Control:** Specific low-rated product and color combinations were identified, which are likely contributors to high **Service Opex** (returns, customer support).
- **Shipping:** Transactions using **Standard Shipping** are the highest volume, but expedited methods need scrutiny to ensure cost coverage.

---

5. Actionable Recommendations

Based on the analysis, here are the targeted actions to control Opex and boost margin:

1. Optimize Promotional Opex (Pricing Strategy)

- **Action: Cease blanket store-wide discounts immediately.** This will restore the [1.42]% margin erosion identified by the analysis.
- **Strategy:** Implement highly **targeted discounts** only in specific categories (e.g., Footwear) where Opex analysis shows the discount *actually* increases Average Order Value (AOV) above the undiscounted baseline.

2. Reallocate Marketing Opex (Targeting Strategy)

- **Action: Concentrate 60% of high-cost retention budget** exclusively on the **High Value** customer segment.

- **Strategy:** For the **40+ Age Group** (the steady core), focus marketing efforts on stable, high-value product offerings rather than high-risk growth campaigns.

3. Refine Subscription Value Proposition

- **Action:** Pivot the subscription benefits away from monetary discounts.
- **Strategy:** Since subscribers do not inherently spend more, focus the value proposition on **reducing their Opex** (e.g., guaranteed free, faster shipping or an exclusive loyalty points multiplier) to drive long-term retention.

---

6. Technical Appendix: SQL Queries (Proof of Database Skill)

The final cleaned and segmented data was exported to a MySQL database table named `final_table` (or your actual table name) to prove data integrity and analysis capability using SQL.

Here are examples of the core Opex analysis queries:

A. Margin Analysis (Query 1 & 2)

Query Goal	Output
Discount Margin Impact (The core finding)	**Undiscounted_AOV** = [60.13] **Discounted_AOV** = [59.28] **Margin_Erosion_Pct** = [1.42]%
Total Revenue Loss (Quantification)	**Total_Revenue_Lost** = **$[1427.77]**

SQL Query (The Margin Proof):

```
-- This query proves discounts erode margin by calculating the difference between the two AVGs.

WITH AOV_Calculation AS (

    SELECT

        AVG(CASE WHEN Discount_Applied = 'No' THEN `Purchase_Amount_(USD)` END) AS
undiscounted_aov,
```

```
    AVG(CASE WHEN Discount_Applied = 'Yes' THEN `Purchase_Amount_(USD)` END) AS
discounted_aov
  FROM final_table
)
SELECT
  undiscounted_aov AS Undiscounted_AOV,
  discounted_aov AS Discounted_AOV,
  ROUND(((undiscounted_aov - discounted_aov) / undiscounted_aov) * 100, 2) AS Margin_Erosion_Pct
FROM AOV_Calculation;
```

B. Strategic Segmentation and Product Queries

Query Goal	Sample SQL Query
Revenue per Payment Method (Payment Opex)	**SELECT Payment_Method, SUM(`Purchase_Amount_(USD)`) AS total_revenue FROM final_table GROUP BY Payment_Method ORDER BY total_revenue DESC;**
AOV by Shipping Type (Shipping Opex)	**SELECT Shipping_Type, ROUND(AVG(`Purchase_Amount_(USD)`), 2) AS average_order_value FROM final_table GROUP BY Shipping_Type;**
Lowest Rated Products (Service Opex Risk)	**SELECT Item_Purchased, AVG(Review_Rating) AS avg_rating FROM final_table GROUP BY Item_Purchased ORDER BY avg_rating ASC LIMIT 5;**
Subscription Conversion by Location (Targeting Opex)	**SELECT Location, (COUNT(CASE WHEN Subscription_Status = 'Yes' THEN 1 END) * 100.0 / COUNT(Customer_ID)) AS subscription_rate_pct FROM final_table GROUP BY Location ORDER BY subscription_rate_pct DESC LIMIT 5;**

The Business Problem: Operational Expenditure (Opex) Control

The core business challenge is:

> **"How can the company strategically restructure its Operational Expenditure (Opex) on promotions and marketing to maximize profit margins and increase Customer Lifetime Value (CLV)?"**

This problem centers on **inefficient spending** that is eroding profitability, specifically:

- **Promotional Opex:** Costs associated with giving unnecessary discounts.
- **Marketing Opex:** High retention costs that are not strategically focused on the most valuable customers.

Optimization Strategy: Margin Restoration & Targeted Marketing

The solution involves a two-pronged strategy derived from your data analysis:

A. Pricing Strategy (Restoring Margin)

The analysis proved that the current discount strategy is **ineffective** because transactions *without* a discount yielded a **higher Average Order Value (AOV)** ($\$[60.13]$) than discounted transactions ($\$[59.28]$). This marginal erosion of **$[1.42]\%$** confirms that discounts are wasting money without driving basket size.

B. Targeting Strategy (Focusing Opex)

Your CLV Segmentation confirmed the **Pareto Principle**: the top $\sim 25\%$ of customers contribute $\sim 38\%$ of the total revenue. This justifies **reallocating high-cost retention efforts** away from low-value customers and focusing them exclusively on this high-value cohort.

Actionable Recommendations

These are the immediate steps to optimize Opex and leverage your findings:

1. Reduce Promotional Opex (Pricing)

- **Action: Cease blanket, store-wide discounts immediately.** This stops the **$[1.42]\%$** margin erosion on unnecessary purchases.
- **Recommendation:** Switch to highly **targeted promotions** only for specific goals (e.g., clearance, or to reactivate inactive *High Value* customers).

2. Focus Marketing Opex (Targeting)

- **Action: Reallocate at least 60% of the retention budget** toward the **High Value** customer segment (top 25% of spenders).
- **Recommendation:** Focus marketing efforts on **Age 40+** customers with stable, high-value product offerings, as they are the highest revenue-generating demographic.

3. Strengthen Subscription Value

- **Action:** Pivot the subscription benefits away from monetary discounts.
- **Recommendation:** Re-package the subscription to offer high-value, low-Opex benefits like **guaranteed faster, free shipping** or exclusive loyalty points multipliers, which drive long-term retention without cutting into product margin.