Script
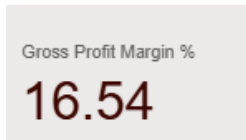
Good morning. This analysis confirms that Hutch's new e-commerce marketplace is working,

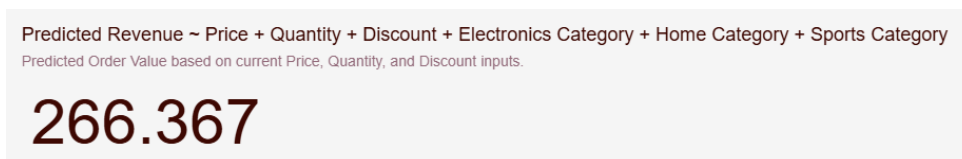but the current profitability (GPM of 16.54%) is too low.

Gross Profit Margin %

16.54

I've used a predictive model to identify the problem and pinpoint exactly where to invest and where to cut losses.

Predicted Revenue ~ Price + Quantity + Discount + Electronics Category + Home Category + Sports Category
Predicted Order Value based on current Price, Quantity, and Discount inputs.

266.367

Let's start with the overall financial health. The company generated $5.87 Million in revenue against $970 Thousand in profit.

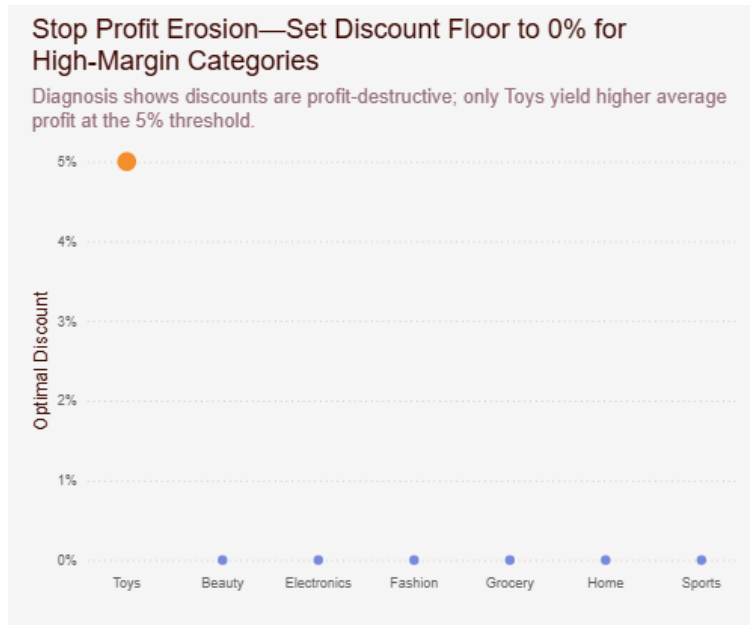Total Revenue

5.87M

Total Profit

970K

The most critical number here is our Gross Profit Margin (GPM) of 16.54%. The entire strategy must be focused on pushing this number higher. The analysis immediately flags the biggest threat to this margin: **discount policy**.

The question is, are discounts driving enough volume to justify their cost? The answer is a resounding no.

The model shows that discounts are highly destructive to the revenue. For six out of seven categories, the optimal discount that yields the highest average profit is 0%. Discounts are an unnecessary cost. Only the Toys category shows a profit lift at 5% discount.


Stop Profit Erosion—Set Discount Floor to 0% for High-Margin Categories

Diagnosis shows discounts are profit-destructive; only Toys yield higher average profit at the 5% threshold.

If we were to implement a 10% discount increase today, the model predicts an immediate loss of $627K in total revenue. Need to show evidence. Will add a KPI card for this later.

Therefore, we must immediately stop all non-strategic discounts for Electronics, Home, and Sports to protect profit margins.

Now that we know *what* to fix (discounts), let's look at *where* we make money.

The profit is concentrated in three categories: **Electronics, Home, and Sports**. These categories show highest revenue and profit contribution across all customer tiers. Conversely, the **Grocery** category has a **negative profit margin**.
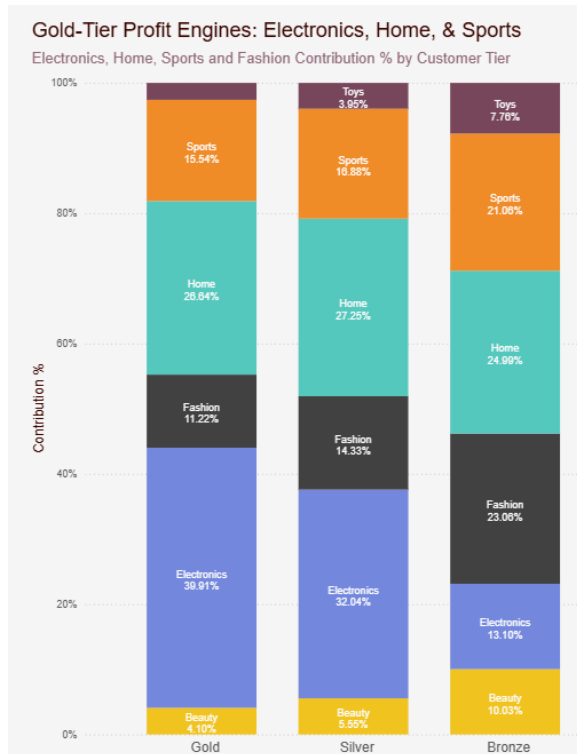

When we look at regional performance, the South region is our most efficient profit base, showing the highest average profit per order.
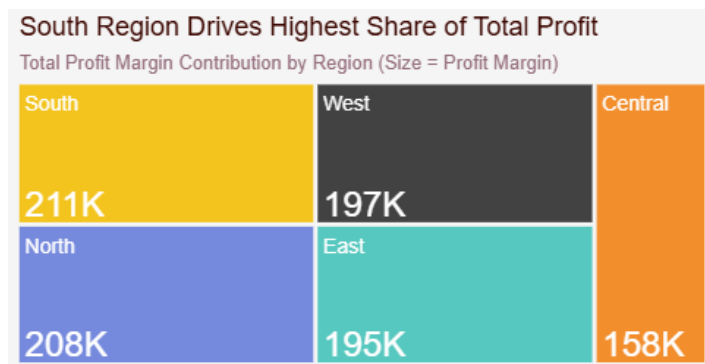
Logistics Slide

However, logistics analysis reveals a critical service failure in the East. The East has slow average delivery times (approx 6 days) that directly correlate with high return rates in Fashion and Electronics. This means we are losing revenue and damaging loyalty in our key categories due to slow fulfillment.
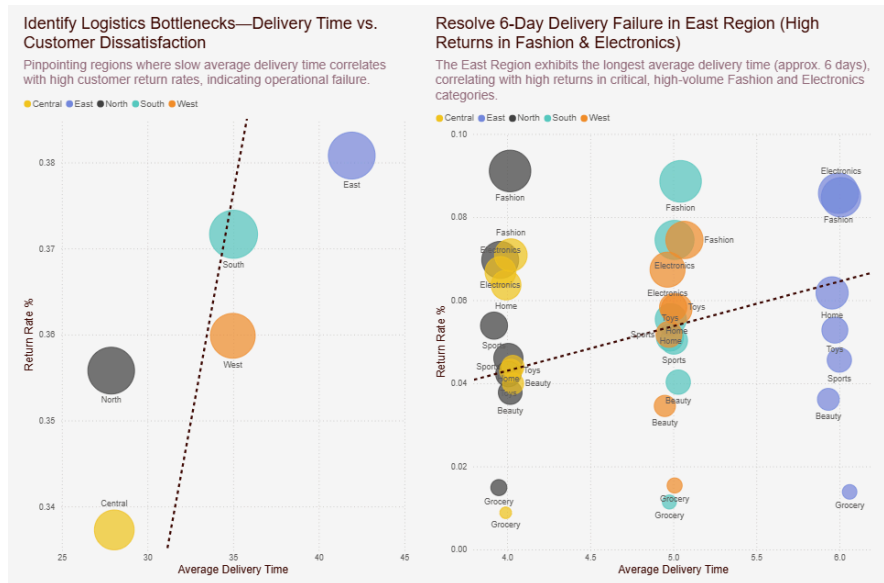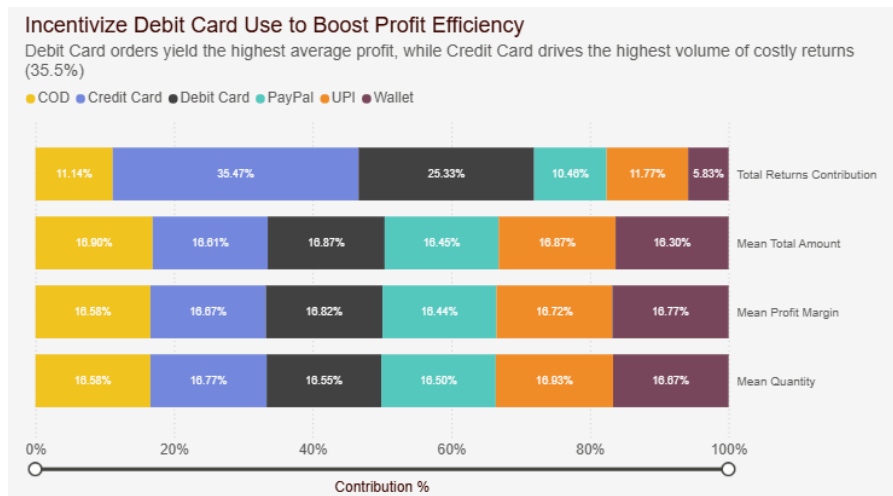

Profitability Slide

Finally, a quick diagnostic on payment methods:
The Debit Card usage yields the highest average profit per order.
In contrast, Credit Card users generate the highest volume of total returns.
To boost profitability, we should incentivize **Debit Card usage**.

*** **In summary, the path to boosting 16.54% GPM is clear: Stop discounts, protect Electronics and Home profits, and fix the 6-day delivery failures in the East region.**

We must now shift our focus from what we sell to who is buying. The segmentation heatmap, immediately flags the greatest operational risk.

While the Gold Tier is highly profitable, the Silver Tier presents a clear, high-priority churn threat.
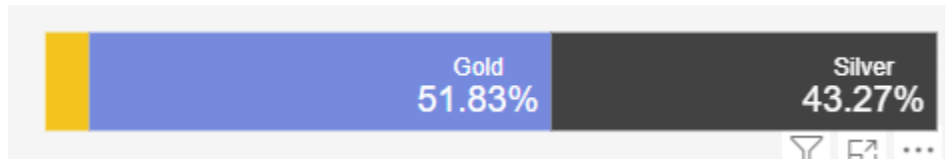

Why? Because Silver customers are active, but they are also highly dissatisfied. Our analysis shows their "returns per customer" is high (e.g., Middle-Aged Silver is 0.23, meaning one return for every four customers), signaling they are ready to leave.

Save 'Silver' Tier—Highest Dissatisfaction & Churn Risk

Risk is Highest in the Silver Tier—Focus Quality Interventions on Middle-Aged and Older Customers.

age_group	Bronze	Gold	Silver
Middle-Aged	0.12	0.34	0.23
Older Adults	0.12	0.37	0.20
Young Adults	0.10	0.36	0.24
Seniors	0.11	0.38	0.22

However, to protect our base, we must look at where the majority of the profit is generated. This chart confirms that Gold Tier customers are financially dominant across all ages.

When we break down profit contribution, the Older Adults and Seniors (Gold) groups are our powerhouse, contributing over 28% of total company profit. This is significantly more than Middle-Aged and Young Adults combined. This is where we focus our best loyalty resources.

Furthermore, when we look at gender, Male and Female segments show near-identical profit profiles. This confirms that customer behavior (Tier) is the only actionable segmentation driver. We must avoid wasting resources on binary gender campaigns.
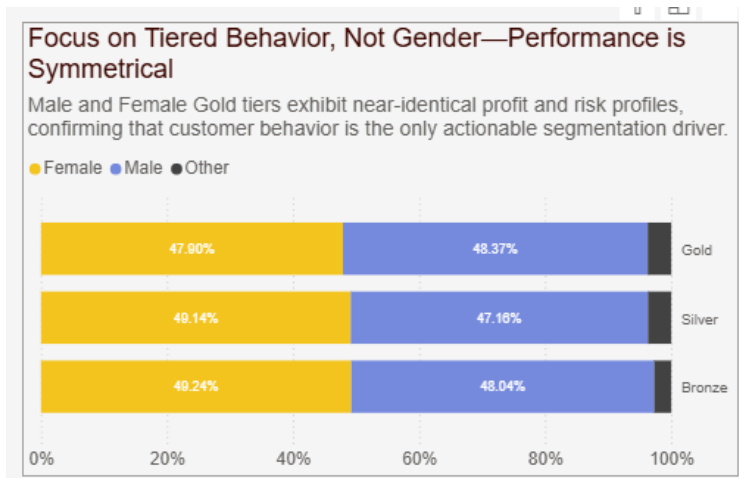
Finally, we need to know who to bring back. The R–F Scatter Plot shows us customers who were once loyal but have disappeared. The size and color of these bubbles show their value.

Look at the top-right quadrant: these are customers with high historical spending and high frequency, but they have high Recency (they haven't bought recently). These are our best targets for personalized, high-ROI reactivation campaigns.
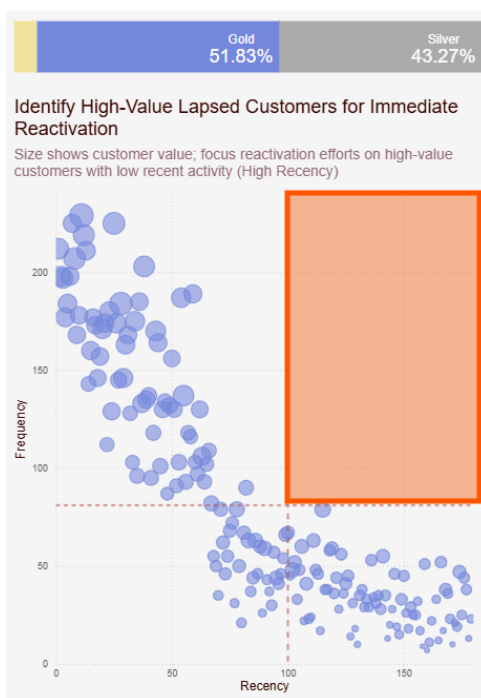
Division of customers into four behavioral groups:

1. Developing Base (Bottom-Left): Customers who have purchased recently but have only bought once or twice. Hutch must nurture them into loyal, repeat buyers.

2. Loyalty (Top-Left): They buy frequently and have purchased recently, indicating high engagement and maximal profitability. Their continued loyalty is vital to the shop's stability.

3. Lapsed & High Potential (Top-Right): Customers who were historically loyal but have not purchased recently. Best target for high-ROI campaigns.

4. High Churn Risk (Bottom-Right): Highest immediate churn risk. Historical value is low. Retention campaigns for this group are typically low-priority.

*** In summary, Hutch should prioritize protective retention for Gold-Tier customers, intervene to reduce service failures (returns) for Older Adult Silver-Tier customers, while excluding gender as a primary segmentation factor.

About 'Total Amount' Model

The model for the Target variable 'Total Amount/Revenue' is constructed as follows:

Categorical Features that contribute a significantly higher percentage to the Total Amount are selected.
Category, Discount

Numerical Features that have high correlation to Total Amount are selected (Spearman Correlation).
Profit Margin is a separate Target variable. Therefore, exclude.
Shipping Cost has a high correlation with Price. Therefore consider only Price. Exclude shipping cost.
Discount has low correlation (slightly negative). Therefore consider.
Quantity has low correlation (slightly positive). Therefore consider.

Encode the categorical variable "Category" using One–Hot encoding.

Now, validate whether the selected Features are correct for the Model.
Use backward selection.

Start with all features: Price, Quantity, Discount and Category
Run: Create a model, run, evaluate.
Now remove a Feature and repeat 'Run' again until only a single feature is used.
Check all model evaluations and select the Best Model: Highest R squared

Export/Save the model coefficients.

About 'Data Set' & Preparation

No NA values found.
Total records: approximately 35000.
Quantity for each category is approximately equal in size.

Date is in Text format. Converted to Date Format.
Created a new column for the Order Year.

Created a new column for Age Bands
customer_age >= 18 & customer_age <= 25 ~ "Young Adults",
customer_age >= 26 & customer_age <= 40 ~ "Middle-Aged",
customer_age >= 41 & customer_age <= 55 ~ "Older Adults",
customer_age > 55 ~ "Seniors"

Created a new column for Delivery Time
delivery_time_days <= 2 ~ "Fast Delivery",
delivery_time_days <= 5 ~ "Standard Delivery",
delivery_time_days > 5 ~ "Slow Delivery"

Created a new column for Price Tiers
price_tier == 1 ~ "Low Price",   ← Lowest 33%
price_tier == 2 ~ "Medium Price",
price_tier == 3 ~ "High Price"   ← Highest 33%

About 'Customer' Segmentation

Recency is calculated using the difference between current date and latest order date for the customer. Recency Score (R) is calculated using 33% percentiles.
E.g. Lowest 33% = 1, Highest 33% = 3

Frequency is calculated using the count of orders for the customer. Frequency Score (F) is calculated using 33% percentiles.
E.g. Lowest 33% = 1, Highest 33% = 3

Monetary Value is calculated using the sum of total amount/revenue of all orders for the customer. M score (M) is calculated using 33% percentiles.
E.g. Lowest 33% = 1, Highest 33% = 3

RFM Score = R + F + M
Min(RFM) = 3
Max(RFM) = 9

Price Tier is calculated using RFM Score.
RFM Score >= 8 ~ "Gold",
RFM Score >= 5 ~ "Silver",
RFM Score < 5   ~ "Bronze"