Presented by Olorunfemi Adeyemo

CAPSTONE PRESENTATION

Optimizing E-Commerce Business and Marketing Strategy with Customer Segmentation and Predictive Analytics

PROBLEM STATEMENT

E-commerce businesses face persistent challenges in retaining customers, driving sales growth, and delivering personalized marketing experiences. Despite significant investment in customer acquisition, a large proportion of buyers make only a single purchase and never return. At the same time, high product return rates erode margins, and limited visibility into customer behavior prevents targeted engagement.

Without data-driven insights, strategic decisions remain reactive, leading to wasted marketing spend, missed revenue opportunities, and declining customer loyalty. A systematic approach is required to understand customer segments, predict future behavior, and implement personalized strategies that maximize lifetime value and reduce churn.

PROJECT GOAL

The problem Statement means two things:

- Segmentation (Unsupervised Learning): Who are the different customer groups?

- Prediction (Supervised Learning): What will these customers do next?

THE 3 W'S

Goal 1

Where we are now

Goal 2

Who our customers are

Goal 3

What should we do?

THE DATA

Step 1

The data was sourced from Google BigQuery's thelook_ecommerce public dataset, which simulates a real-world e-commerce platform.

Step 2

Key tables extracted included:

- users – demographics
- orders – transaction history
- order_items – product-level purchases
- products – product details
- inventory_items – pricing and inventory metadata
- distribution_centers – logistics information

Step 3

Data Preparation

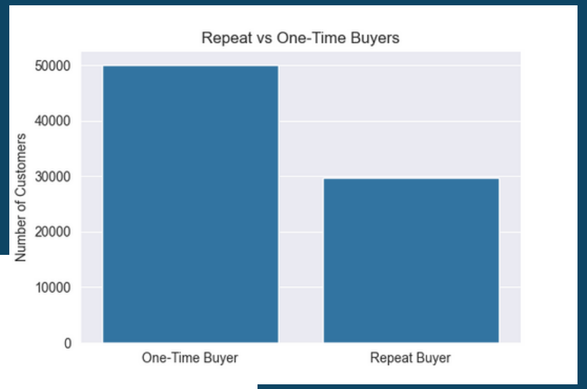
All tables were cleaned and merged to remove duplicates, handle missing values, and align keys such as user_id, order_id, and product_id.

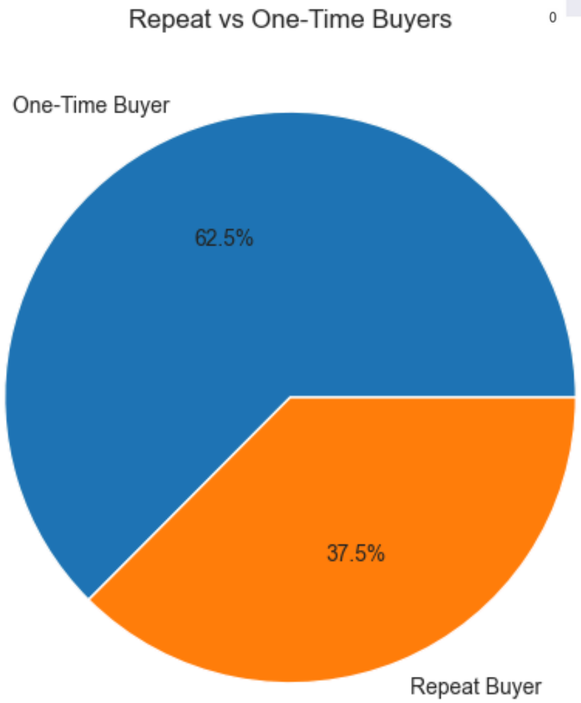
Step 4

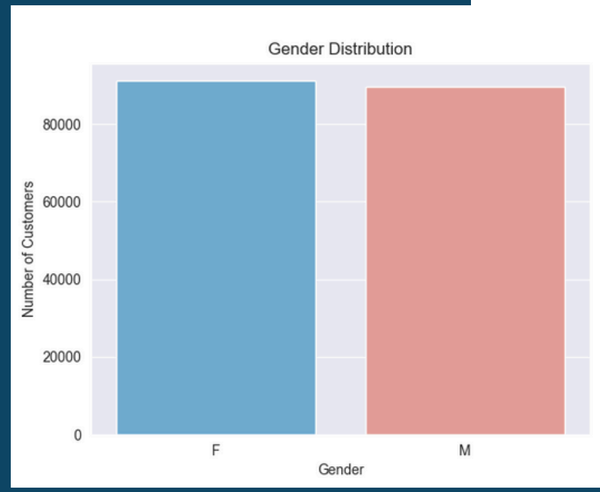
Analytical Dataset

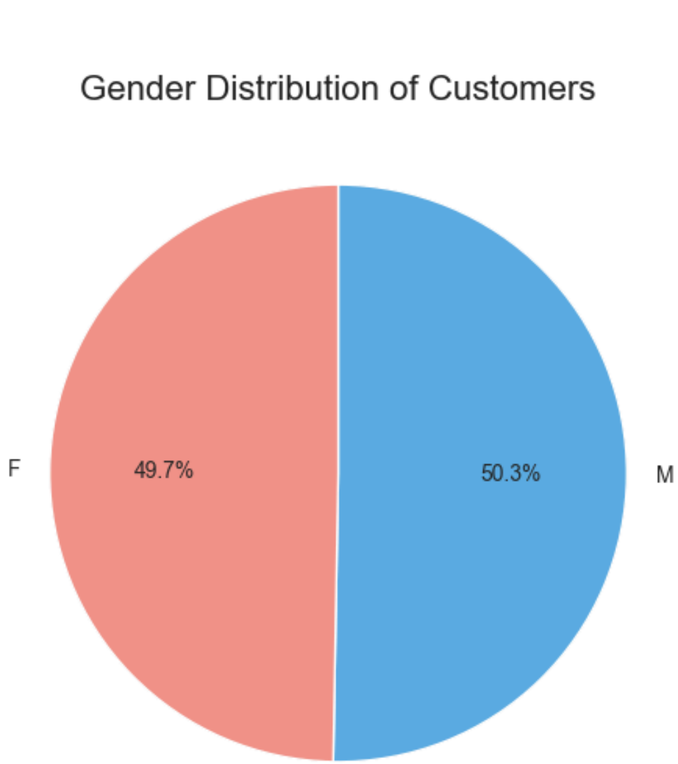
The result was a customer-level dataset (one row per customer) containing behavioral and demographic features, ready for segmentation and predictive modelling and an item-level dataset used in EDA.

WHERE WE ARE NOW





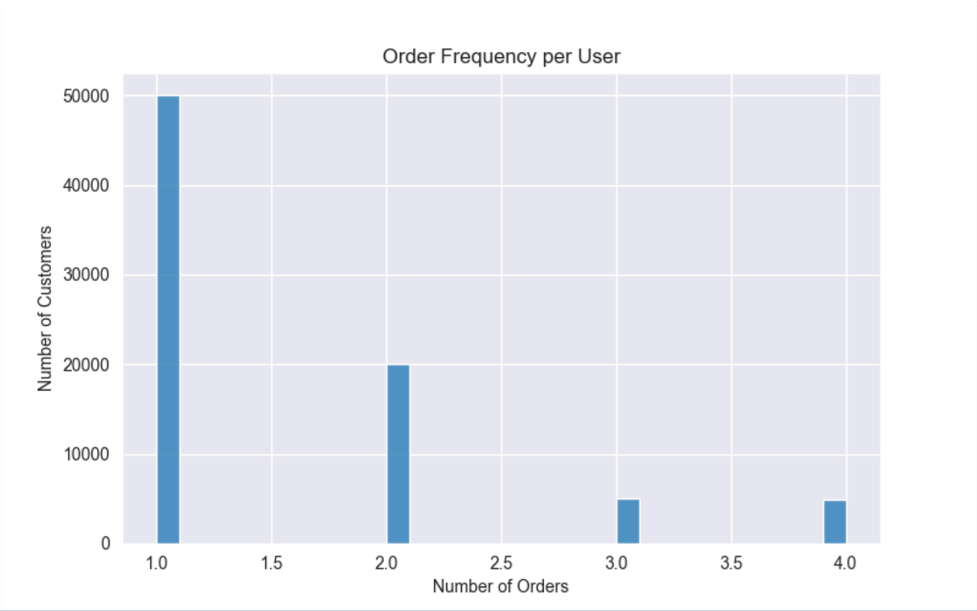




Order Frequency:

- More than half of customers only purchase once.
- Repeat buyers make up just ~37% of the customer base.
- This highlights a critical retention challenge.
- The distribution of both males and females is ~ 50 %.

CUSTOMERS BUY ONCE, BUT THEY GET THEIR ORDERS FAST.

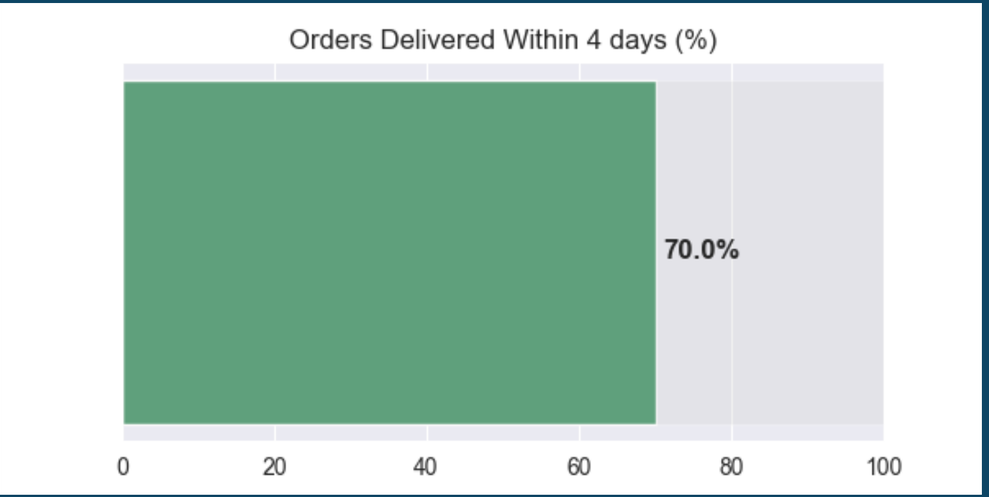


Delivery Times:

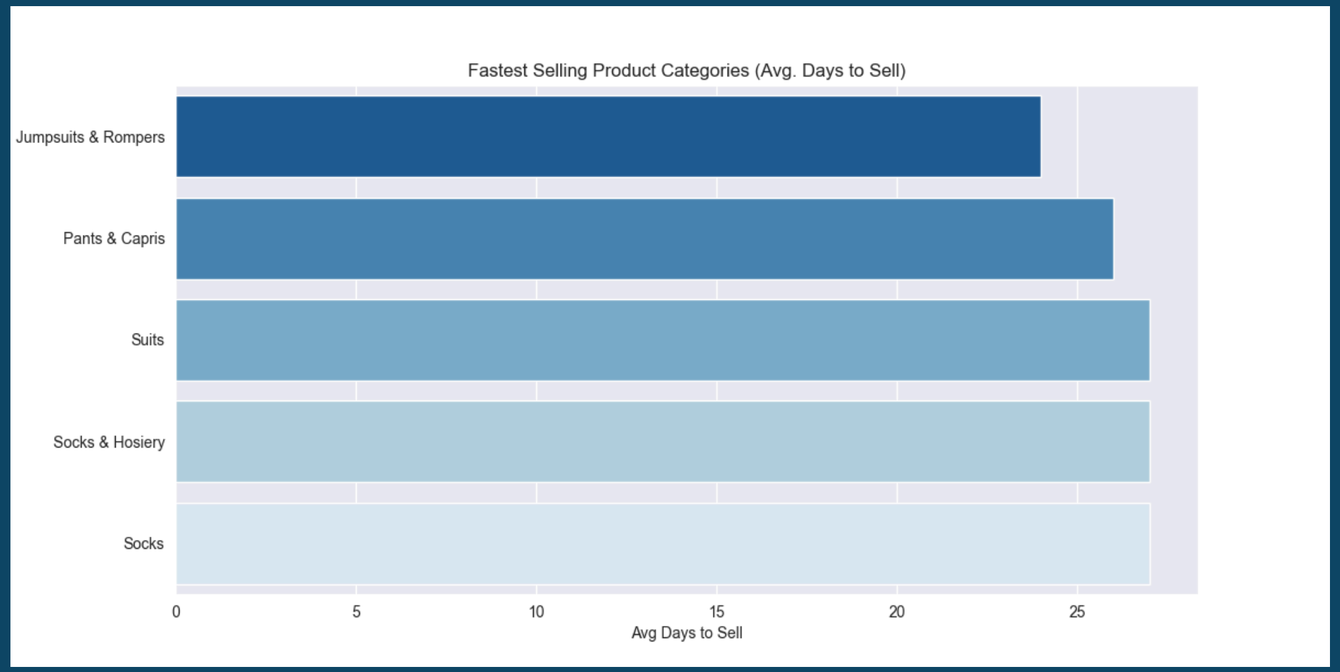
- The majority of orders are delivered within 2–4 days.
- A strength that can be marketed as “fast delivery within 48 hours.”
- This is a competitive advantage to leverage.

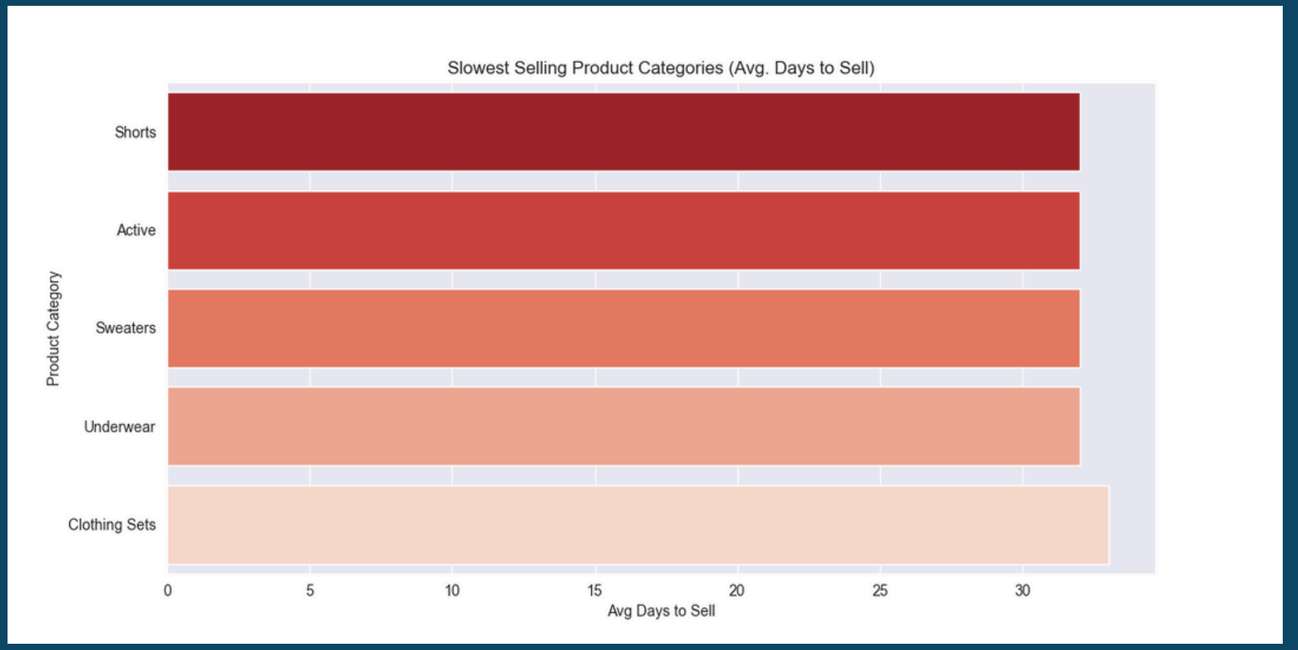
However,

- A large share of customers purchase only once.
- This suggests a need for personalized marketing outreach

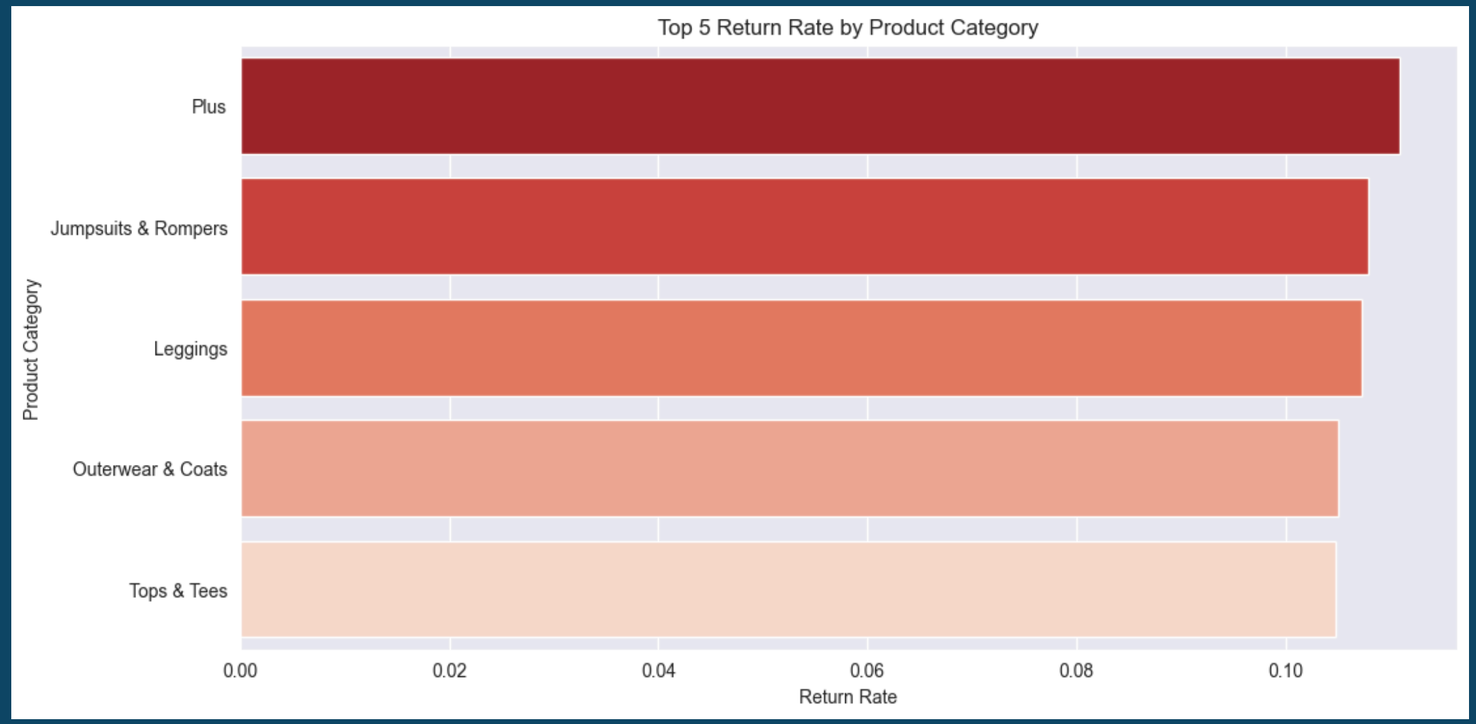


PRODUCT VELOCITY





PRODUCT VELOCITY



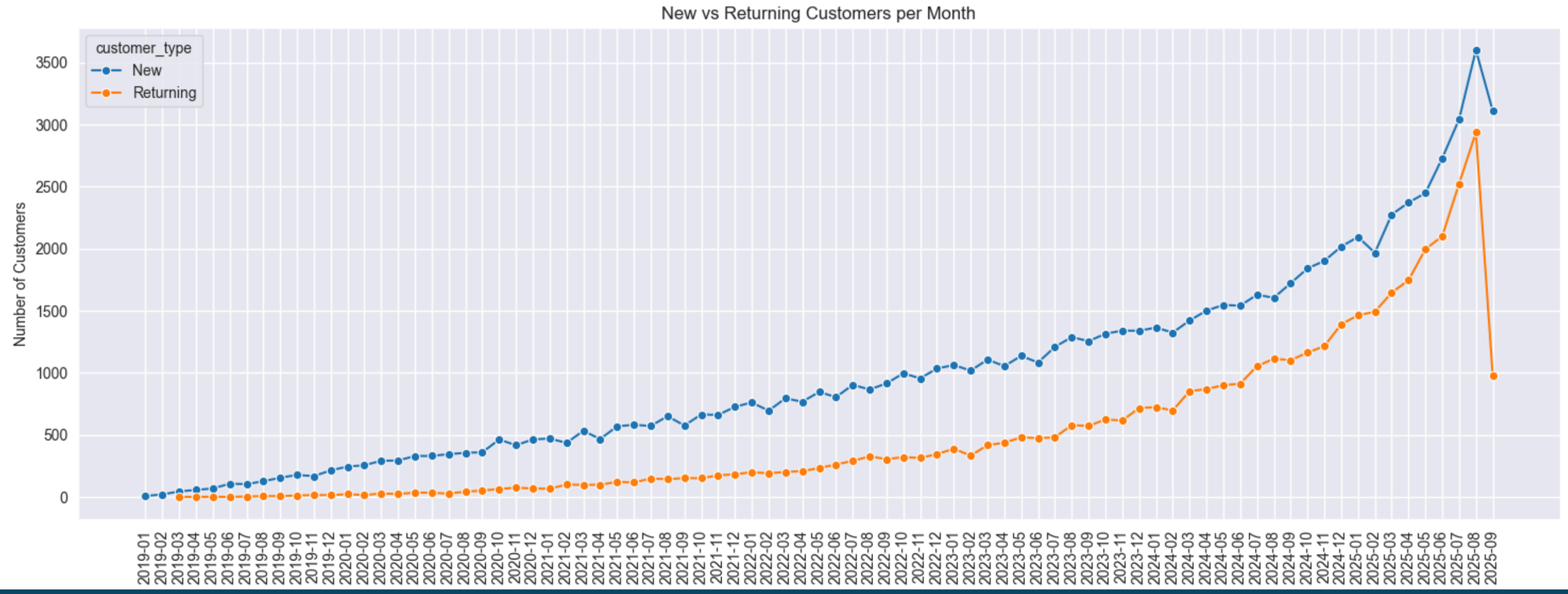
Fast-Selling but Risky Products

- Jumpsuits & Rompers are among the fastest-selling categories (high product velocity).
- However, the same category also shows up with high return rates, signalling potential customer dissatisfaction or sizing/fit issues.
- Other fast movers include Pants & Capris, Suits, and Hosiery — but without the same return intensity.
-

Recommendations

1. Investigate Returns: For Jumpsuits & Rompers, analyze reasons for returns (fit, quality, style mismatch).
2. Product Improvement: Work with suppliers/brands (e.g., Allegra K) to improve fit guides, product descriptions, and sizing charts.
3. Targeted Marketing: Promote categories that combine fast sales + low return rates (e.g., Suits, Pants & Capris) to maximize profitability.

STRONG ACQUISITION, WEAK RETENTION — LOYALTY IS THE NEXT GROWTH LEVER



- New customers have grown consistently, peaking at approximately 3,600 per month by mid-2025.
- Returning customers also increased but remain below new acquisitions (approx. 3,000 per month).
- Marketing efforts are effective at acquisition.

Recommendations

- Leverage personalized campaigns, rewards, and improved post-purchase experience to close the gap between new and returning customers.

WHO OUR CUSTOMERS ARE

Beyond descriptive insights, we used machine learning to uncover hidden patterns and make predictions.
Using *Random Forest*

Why Random Forest?

- Handles Non-Linearity: Unlike Logistic Regression, Random Forest captures complex, non-linear relationships in customer behavior.
- Robust to Outliers & Noise: Works well with messy e-commerce data (returns, order anomalies, etc.).
- Feature Importance: Provides rankings of the most important features (e.g., total_spent, avg_order_value) to explain business drivers.
- Class Imbalance: With class_weight="balanced", it adjusts for skewed target distributions (important for churn and high-value customers).

Predict loyalty

If a customer's order is greater than 1

Predict top spenders
Using Pareto Principle, roughly 80% of consequences come from 20% of causes

Predict the likelihood of inactivity
Target = 1 if the customer hasn't purchased in 120 days (Median * 3)

Target Feature	Accuracy	Key Drivers	Business Insight
Repeat Buyer	100%	Total Spend, Avg Order Value	Spending behavior is the strongest signal of loyalty.
High-Value Customer	100%	Avg Order Value, Number of Orders	High-value customers combine big basket sizes with repeat purchases.
Churn Risk	70%	Total Spend, Avg Order Value, Age	Model flags churned customers well (86% recall) but struggles with active ones due to imbalance.

WHY SHOULD WE CARE?

Supervised Learning Contribution

The supervised models answer the “predictive” part of the problem statement.

- Repeat Buyer Prediction (100% accuracy)
  - Identifies which first-time buyers are likely to return.
  - Business value: Marketing can focus retention campaigns on those at risk of not returning.

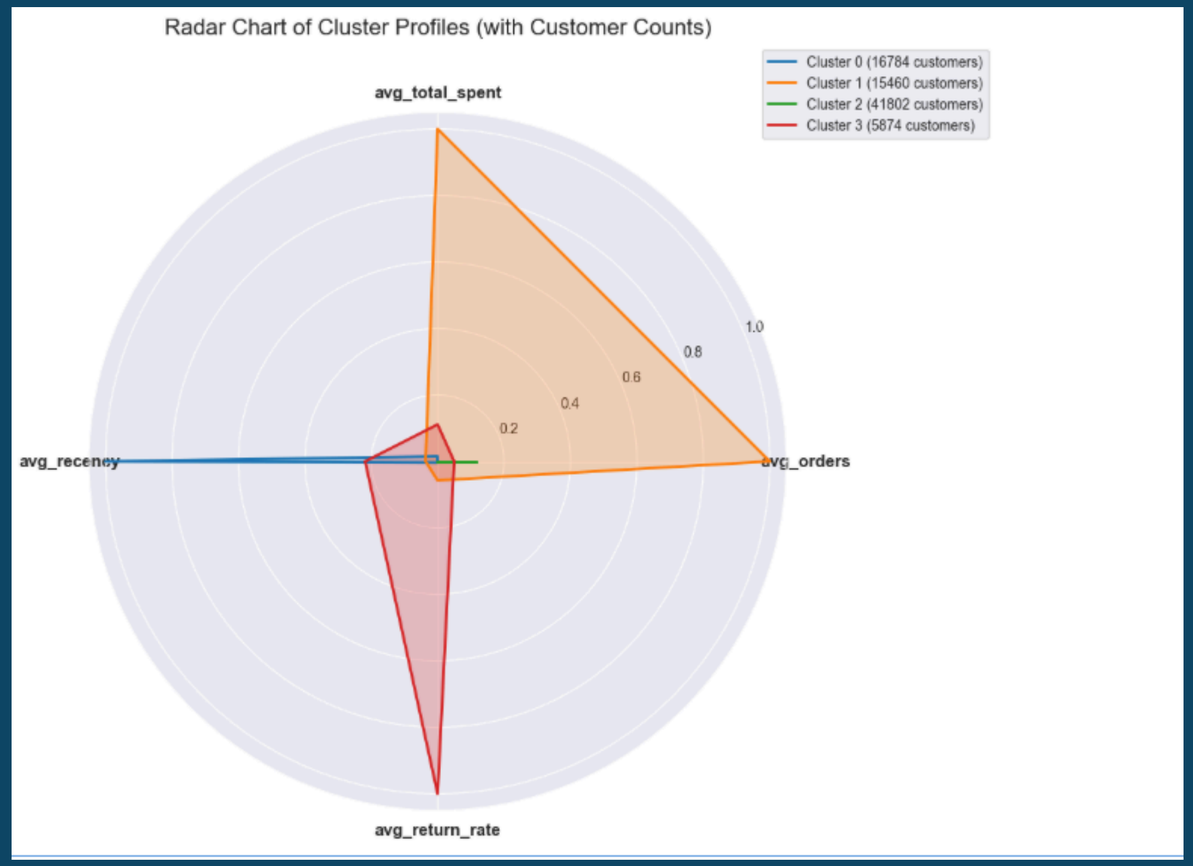
- High-Value Customer Prediction (100% accuracy)
  - Flags potential VIP customers early.
  - Business value: Loyalty programs can target these individuals before competitors win them over.

- Churn Risk Prediction (70% accuracy, 86% recall for churners)
  - Detects customers likely to stop buying.
  - Business value: Triggers win-back offers and personalized re-engagement campaigns.

WHO OUR CUSTOMERS ARE

Unsupervised Learning(KMeans Clustering k = 4)

The business question wasn't just "Can we predict behavior?" but also "Who are our different customer types



High-Value Buyers – profitable & loyal

High Returners – low spend, costly returns

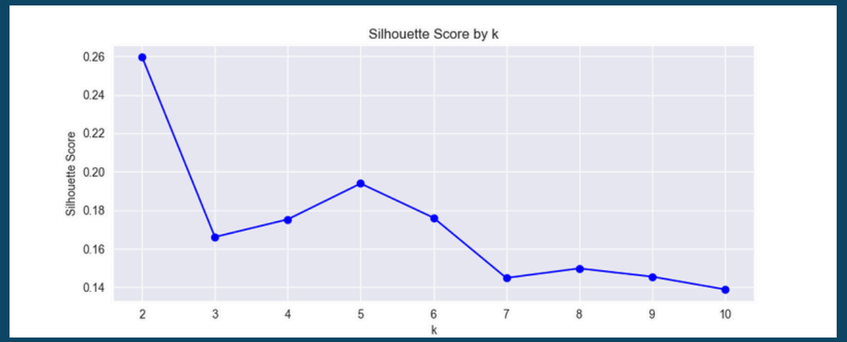
Lost Customers – dormant, low ROI

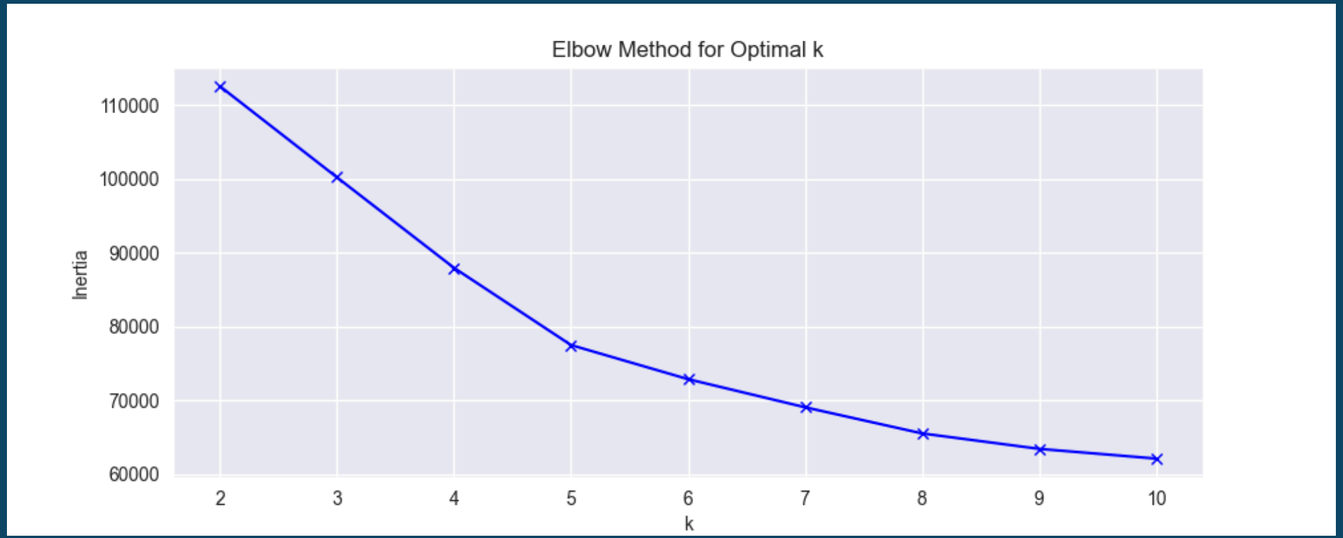
Low-Spend One-Timers – large group, growth potential

WHO OUR CUSTOMERS ARE

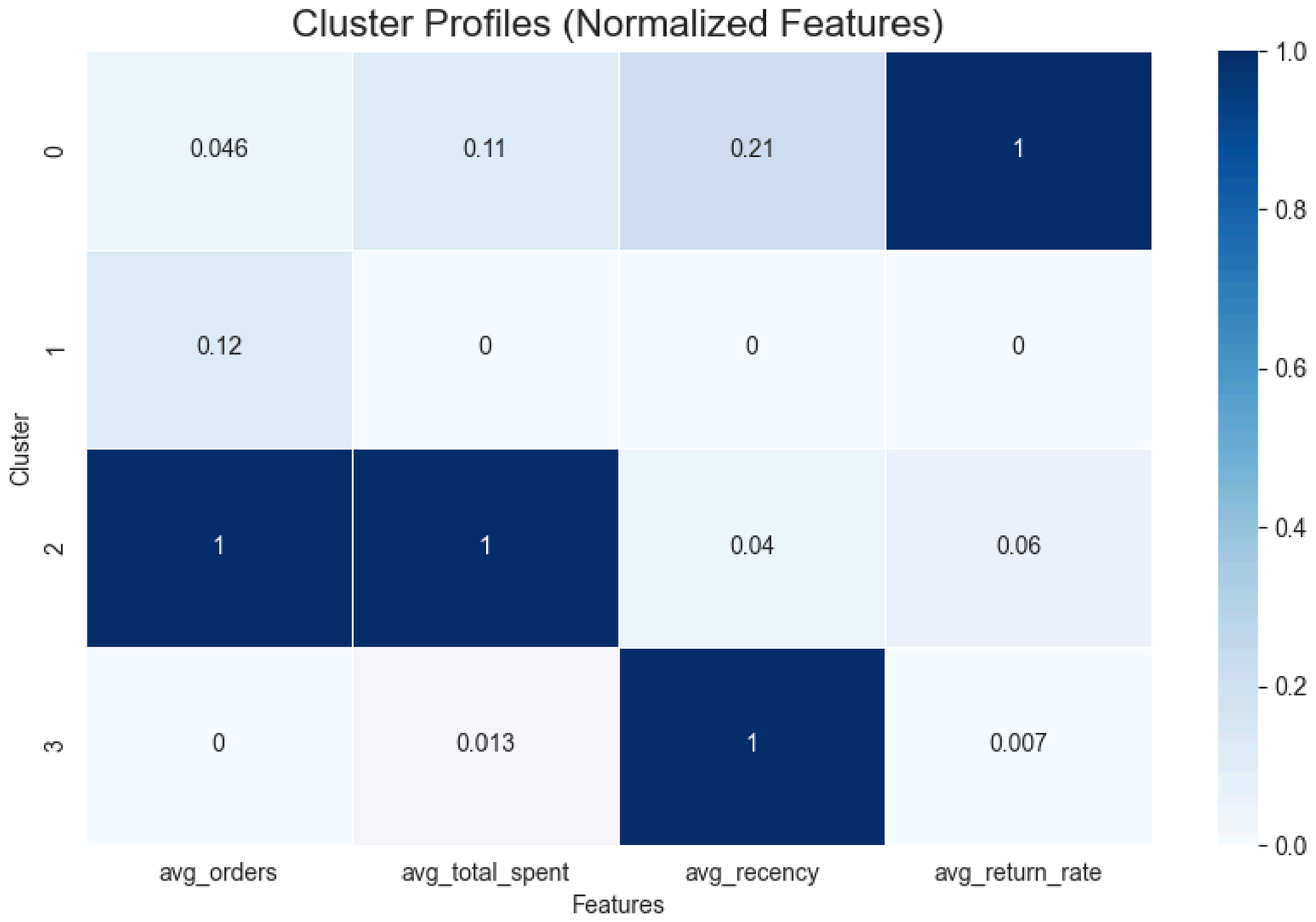
Unsupervised Learning(KMeans Clustering k = 4)

The business question wasn't just "Can we predict behavior?" but also "Who are our different customer types





WHO OUR CUSTOMERS ARE



Cluster	Customer Count	Traits	Business Insight
Cluster 2 – High-Value Buyers	~15,788	Multiple purchases, high spend, moderate returns	VIPs worth protecting and upselling
Cluster 0 – High Returners	~5,653	Low spend, very high return rate	Unprofitable segment, review policies
Cluster 3 – Lost Customers	~16,879	Very old recency, low spend	Dormant customers, low ROI for reactivation
Cluster 1 – Low-Spend One-Timers	~41,629	Largest group, one order, low spend	Opportunity for conversion into repeat buyers

RECOMMENDATIONS

1. Retention & Loyalty

- Focus retention efforts on the 37% repeat buyers, as they drive long-term value.
- Design loyalty programs (discounts, exclusive access) to encourage one-time buyers to return.

2. Product Strategy

- Jumpsuits & Rompers: High sales velocity but also high return rates.
  - Improve sizing guides, quality checks, and product descriptions to reduce returns.
- Fast movers with low returns (e.g., Suits, Pants & Capris) should be prioritized in marketing campaigns and stock planning.

3. Customer Segments

- High spenders (Cluster 1): Reward with personalized recommendations and premium offers.
- One-time/low spenders (Cluster 3): Convert through onboarding emails, discounts, and low-risk offers.
- Frequent buyers with returns (Cluster 0): Focus on product quality & expectation alignment to reduce dissatisfaction.

5. Churn Reduction

- With approx. 78% customers at churn risk (no purchase after ~120 days):
  - Launch re-engagement campaigns at the 60-day mark.
  - Use personalized emails or retargeting ads before customers fall inactive.


THANK YOU
I HOPE WE ALL WIN