Early Predictors of High-Value Customers

An end-to-end e-commerce analysis using SQL, Python, and Tableau

Problem Statement

Customer value is usually measured too late.
By the time we know who our best customers are, budget, discounts, and acquisition costs have already been spent.

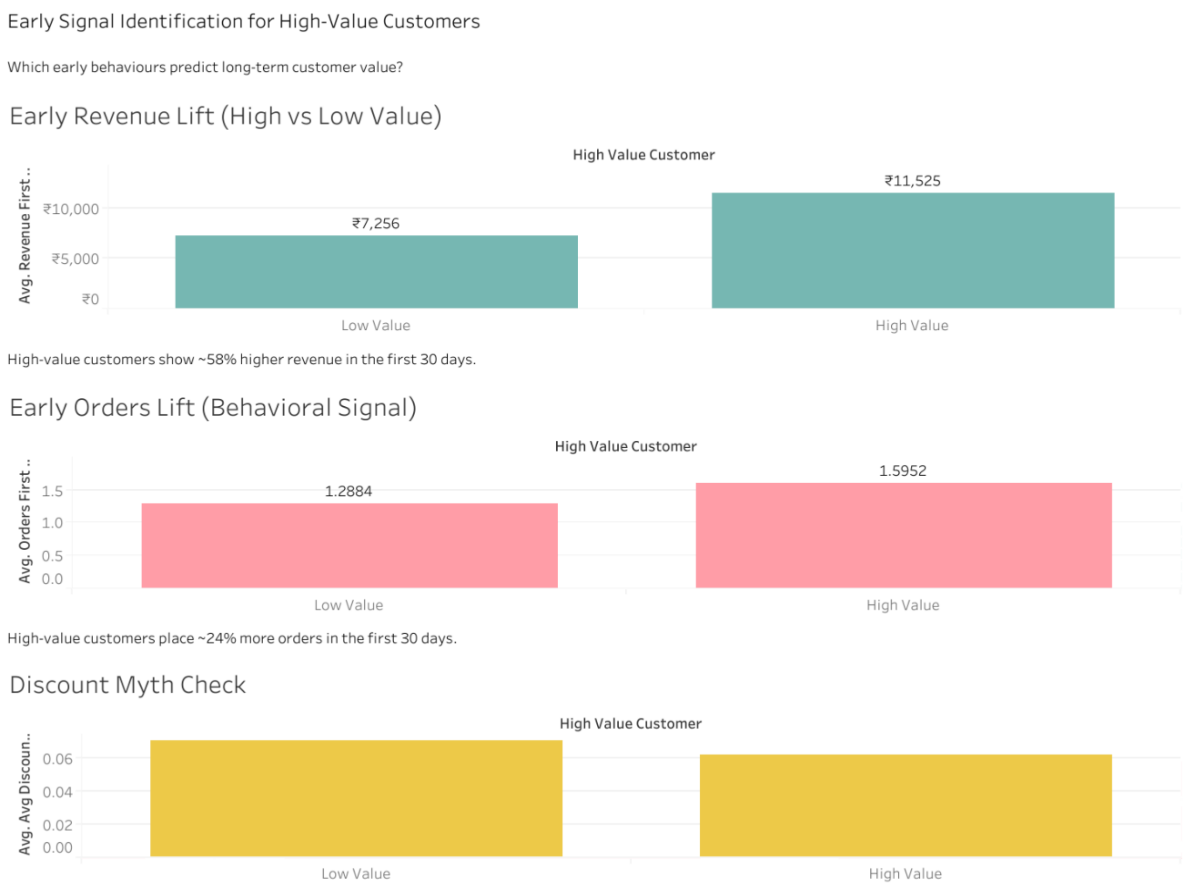
Can we identify high-value customers early using observable behaviour in their first 30 days?

Data & Approach

Using transactional and customer-level data from Shopsphere:

- Aggregate order-level data to the customer level
- Engineered early behaviour features (first 30 days):
  - Orders first 30 days
  - Revenue first 30 days
  - Average discount first 30 days
- Defined High-Value Customers as top 25% by total revenue
- Conducted statistical lift analysis + significance testing

Early Predictors of High-Value Customers



Link: https://tinyurl.com/early-high-value-dashboard

Project artifacts: SQL schema & queries, Python analysis, Tableau dashboard.

Findings

- Early Revenue is a Strong Signal
  High-value customers generated ~59% higher revenue in their first 30 days.
  Highly statistically significant ($p < 0.0001$).
  → High-value customers start strong.

- Early Order Volume Matters
  High-value customers placed ~24% more orders in their first 30 days.
  Statistically significant.
  → Frequency signals lifetime value.

- The Discount Myth
  Average discount depth in the first 30 days showed no statistically significant difference between high and low-value customers.
  → Heavy discounts do not create high-value customers.
  → Discounts are not a reliable early signal of long-term value.

Business Interpretation

The strongest predictors of long-term value are:

- Early revenue intensity
- Early purchase frequency

Not discount exposure.
This suggests customer value is driven more by intrinsic engagement or product-market fit than by incentive intensity.

Strategic Takeaways

This analysis suggests four operational shifts:

- Build early value scoring models
  - Flag high-potential customers within 30 days
  - Prioritize CRM personalization for this cohort
- Shift growth spend from discounting to engagement
  - Invest in onboarding experience
  - Improve early repeat triggers
- Use first-month revenue as a performance KPI
  - Track 30-day revenue per acquisition cohort
  - Use it as an early proxy for CLV
- Align retention strategy with behavioural strength, not discounts
  - High-value customers are behaviourally different, not incentive-dependent

Business Takeaway

Early engagement — not discounts — predicts customer value.

Revenue and frequency in the first 30 days provide statistically significant signals of long-term value, enabling earlier, smarter customer differentiation and lifecycle investment.

Project artifacts: SQL schema & queries, Python analysis, Tableau dashboard.