

Early Customer Signals & Long-Term Value

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

Problem Statement

Can we identify high-value customers early in their lifecycle using only initial behavioural signals?

Most companies wait months to segment customers by lifetime value.

This analysis tested whether early signals are statistically informative enough to predict long-term value.

Data & Approach

Built from the existing relational schema and customer summary table.

High-value customer defined as top 25% by total revenue

Cutoff: ₹40,700

Class distribution:

- 378 non-high-value
- 126 high-value

Early features engineered:

- Orders (first 30 days)
- Revenue (first 30 days)
- Avg. discount (first 30 days)
- Income band, gender, discount bucket

Logistic regression used as a statistical signal test.

Model Performance

Test Set Metrics:

- Accuracy: 82%
- ROC-AUC: 0.62

However:

- Low recall for high-value customers
- Better at eliminating low-value than identifying premium ones
- Predictive power is moderate

Interpretation:

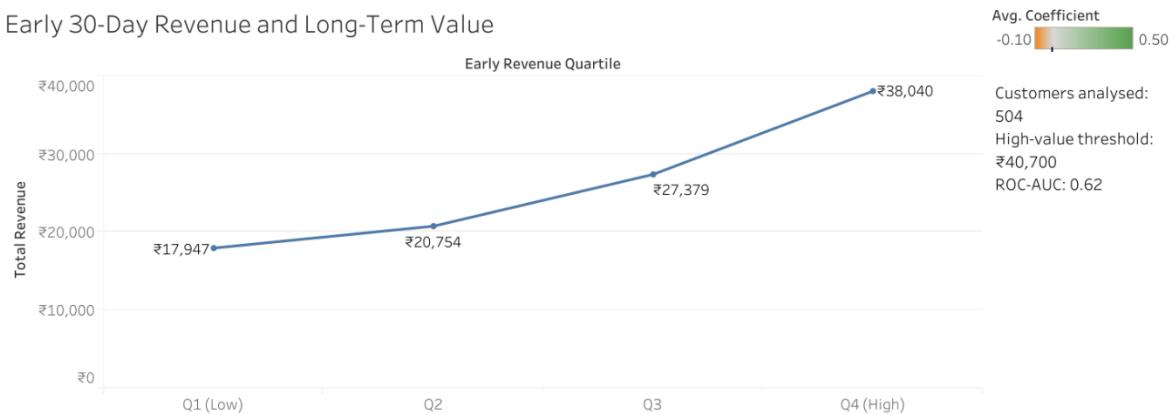
Early behaviour contains signal — but not enough for reliable classification alone.

Early Behaviour & Lifetime Value

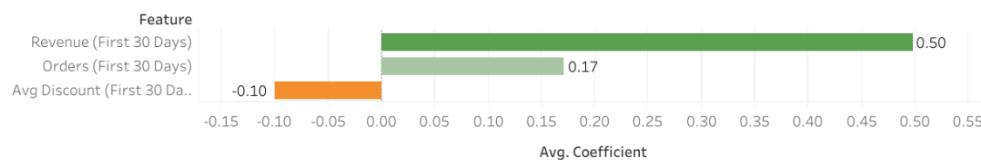
Predicting High-Value Customers from Early Behaviour

Using the first 30-day behaviour to estimate long-term revenue potential

Early 30-Day Revenue and Long-Term Value



Drivers of High Lifetime Value (Logistic Model)



Link: <https://tinyurl.com/early-behaviour-dashboard>

Coefficient Insights

Standardized Logistic Coefficients:

- Revenue (first 30 days): +0.50
- Orders (first 30 days): +0.17
- Average Discount (first 30 days): -0.10

Implications:

- Early spend intensity is the strongest predictor.
- Order frequency matters, but less.
- Discount exposure slightly reduces long-term value probability.

Strategic Conclusion

Across Models 1,2, and 3:

- Discounting does not structurally create high-value customers.
- Instead, long-term value appears driven by intrinsic purchase intent and early willingness to pay.

Strategic Decision:

- Identify strong early spenders
- Prioritize premium engagement
- Avoid over-subsidizing low-margin segments

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

Limitations

- Synthetic dataset
- Class imbalance affects recall
- No behavioural sequence modelling
- Limited to 30-day window

Business Takeaway

While early revenue is directionally predictive of long-term value, discounts do not meaningfully create high-value customers.

Targeting should prioritize identifying naturally high-intent customers rather than stimulating artificial demand through discount depth.