

E-Commerce Growth Intelligence & Forecasting System

Nova Retail Co. | End-to-End Data Analytics Case Study

Python | SQL | Tableau | Business Intelligence | Time-Series Forecasting

NET RETAINED REVENUE	COMPLETED ORDERS	AVERAGE ORDER VALUE	RETURN RATE
\$367,943	5,622	\$65.45	6.7%

Fictional business case study created to demonstrate an end-to-end analytics consulting workflow.

Executive Summary

Nova Retail Co., a U.S.-based online home and lifestyle retailer, needed a clearer view of what was driving revenue, profitability, customer value, marketing performance, returns, and future sales. The source data was fragmented across orders, customers, products, and marketing tables and contained realistic quality issues that had to be investigated before analysis.

I built an end-to-end analytics workflow that audited and cleaned the data, standardized business logic, used SQL and Python for analysis, compared forecasting approaches using holdout validation, and translated the results into an executive Tableau dashboard and business recommendations.

Business Questions

- What is driving revenue and profit?
- Which products and categories are strongest or weakest?
- Which customers and acquisition sources appear most valuable?
- Which marketing channels show the strongest reported efficiency and scale?
- Are returns, discounts, shipping costs, or other operational factors hurting profitability?
- What trends should management monitor, and what is a defensible near-term sales baseline?

Data & Quality Work

The project used four core datasets: orders, customers, products, and marketing. The order file began with 6,203 rows. The audit identified duplicate transactions, missing identifiers and revenue, inconsistent categories/statuses, an invalid quantity, an invalid discount, a high shipping-cost outlier, missing product-master values, an orphan product ID, and marketing anomalies. I documented 21 issues in a data-quality log and separated source preservation from analysis-safe corrections.

Dataset	Quality Treatment
Orders	Removed 3 exact duplicates; standardized status/date values; repaired one missing revenue value from valid transaction fields; retained unresolved anomalies with analysis-safe nulls where appropriate.
Customers	Standardized state values; treated an implausible age as missing; retained incomplete email/acquisition fields rather than inventing values.
Products	Standardized category/supplier values; repaired one missing unit cost from consistent transaction history; left an unprovable list price missing; retained orphan transaction revenue outside product-master analysis.
Marketing	Corrected one clearly misclassified channel row; retained and flagged missing spend, negative attributed revenue, and a click/conversion anomaly rather than silently rewriting them.

Analytical Workflow

1. Data audit & cleaning: Profiled completeness, duplicates, ranges, category consistency, and cross-table integrity; created cleaned and analysis-ready datasets.

2. SQL analysis: Loaded cleaned tables into SQLite and queried revenue, status, category, product, customer, and marketing performance.

3. Python analysis: Calculated KPIs, profitability, return rates, discount bands, customer metrics, acquisition-source performance, and marketing efficiency.

4. Forecast validation: Compared Seasonal Naive and Holt trend models on a six-month holdout period using MAE and MAPE.

5. Executive visualization: Built a Tableau executive dashboard with KPIs, actual-vs-forecast revenue, top product profitability, and marketing performance.

Key Findings

Revenue growth came from transaction value, not order growth.

2025 completed revenue reached \$192,934 versus \$175,009 in 2024, a 10.24% increase. Completed orders increased only 0.36%, while AOV increased 9.84% to \$68.51. The growth story is therefore primarily higher transaction value rather than materially higher order volume.

Seasonality is meaningful.

Both years show stronger late-year revenue, especially November and December. This pattern mattered enough that a seasonal baseline outperformed the tested Holt trend model on the holdout period.

Product profitability is healthy overall, but margin problems are concentrated.

Most products generated healthy contribution margins. Two products were negative on contribution profit and two more were below a 10% contribution margin. Shipping cost was a major pressure point for several low-priced SKUs.

Deeper discounts correlate with weaker contribution economics.

No-discount orders produced a 49.80% contribution margin, compared with 33.75% for the 21-30% discount band. AOV also declined as discount depth increased. This does not prove causality, but it supports testing whether deeper promotions generate enough incremental demand to justify the margin sacrifice.

Returns are not evenly distributed.

The overall return rate was 6.7%. Bath had the highest category return rate at 7.95%, while Lifestyle was lowest at 4.71%. Soft Towel Set had the highest observed product return rate among the highlighted products at 14.06%. Return reasons were not available, so causes were not assumed.

Marketing efficiency and scale tell different stories.

Email had the strongest reported ROAS at 5.43, while Google Ads had the largest platform-attributed revenue at approximately \$537K. TikTok had the weakest immediate reported ROAS at 1.77, yet TikTok-acquired customers showed the highest observed revenue per customer among the paid acquisition sources in the customer data. This suggests channel decisions should not rely on platform ROAS alone.

Loyalty membership was not associated with higher observed value.

Observed revenue per customer and order frequency were very similar between loyalty members and non-members in the analysis period, suggesting the program should be evaluated for whether it is creating incremental behavior.

Forecasting & Model Selection

I held out July-December 2025 and compared two forecasting approaches. Seasonal Naive produced MAE of \$2,215.91 and MAPE of 11.52%, outperforming the Holt trend model (MAE \$3,161.41; MAPE 15.92%). Because the seasonal baseline had the lower validation error, it was selected for the portfolio forecast.

Model	MAE	MAPE
Seasonal Naive	\$2,215.91	11.52%
Holt Trend	\$3,161.41	15.92%

Selected baseline forecast, Jan-Jun 2026: \$83,365.20. This is a continuation-of-seasonality baseline, not a guarantee of future sales and not an assumption that the prior year's 10.24% growth rate will repeat.

Business Recommendations

- Protect high-contribution products and investigate pricing, bundling, fulfillment, and shipping strategy for low-priced SKUs where shipping erodes margin.
- Test discount depth rather than assuming deeper promotions are beneficial. Measure incremental orders/revenue against the observed contribution-margin sacrifice.
- Investigate Bath-category and high-return SKUs for product quality, descriptions, expectations, specifications, packaging, and fulfillment issues; collect return-reason data before assigning cause.
- Use Email as an efficiency benchmark and Google Ads as a scale channel, while monitoring Meta's deterioration and validating TikTok with cohort/customer economics rather than platform ROAS alone.
- Reassess the loyalty program's incentives and engagement strategy because membership was not associated with higher observed customer value in this dataset.
- Use the seasonal forecast as a planning baseline and refresh it as new monthly actuals arrive.

Executive Dashboard

The Tableau dashboard consolidates the decision layer of the project: executive KPIs, historical revenue with the validated 2026 forecast, top-10 product contribution profitability, and marketing performance by reported ROAS and platform-attributed revenue.

Executive KPI Row \$367,943 net retained revenue 5,622 completed orders \$65.45 AOV 6.7% return rate.	
Revenue Trend + Forecast 24 months of actual completed revenue (2024-2025) plus the selected Jan-Jun 2026 seasonal baseline.	Top 10 Product Profitability Ranks the strongest products by contribution profit after product cost and shipping.
Marketing Performance Compares reported ROAS with platform-attributed revenue to separate efficiency from scale.	Dashboard title Nova Retail Executive Performance

Technical Skills Demonstrated

- Python/pandas: data profiling, cleaning, feature creation, KPI calculations, analysis-ready exports
- SQL/SQLite: multi-table querying, aggregation, joins, reconciliation, and business metric analysis
- Statistics & forecasting: holdout validation, MAE/MAPE comparison, seasonal baseline selection
- Tableau: KPI cards, continuous time-series visualization, forecast handoff, profitability ranking, marketing comparison
- Business intelligence: metric definitions, source reconciliation, analytical caveats, executive recommendations
- Data quality: anomaly investigation, cross-table integrity checks, issue logging, and defensible treatment of uncertain values

Portfolio Summary

This project demonstrates my ability to move beyond dashboard creation and manage an analytics problem end to end: understand the business questions, audit imperfect data, establish defensible business logic, analyze performance with Python and SQL, validate a forecast, and communicate the findings through an executive BI deliverable.

E-Commerce Growth Intelligence & Forecasting System | Nova Retail Co.