



Market Basket Analysis: Retail Intelligence System

Data Science & Machine Learning

By: Az-Zukhrufu Fi Silmi Suwondo



Business Problem & Context

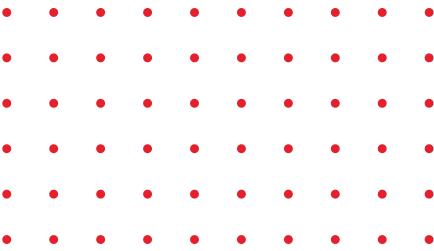
UK-based online retailer generating £7.14M annually across 17,533 transactions, but lacking systematic approach to maximize revenue per customer through intelligent product recommendations and strategic merchandising.

Primary Goal

Deploy Association Rule Mining to uncover hidden product relationships and translate findings into 3-5 actionable strategies that drive measurable revenue growth through intelligent cross-selling, strategic bundling, and optimized product placement.

Success Criteria

- Generate statistically validated product association rules
- Identify patterns covering $\geq 25\%$ of transactions
- Ensure model stability ($\geq 95\%$) for production deployment
- Deliver actionable merchandising strategies



Data Overview



Data Quality
17,816 Records



Customer Base
4,265 Customers



Revenue Performance
£7.14M Total



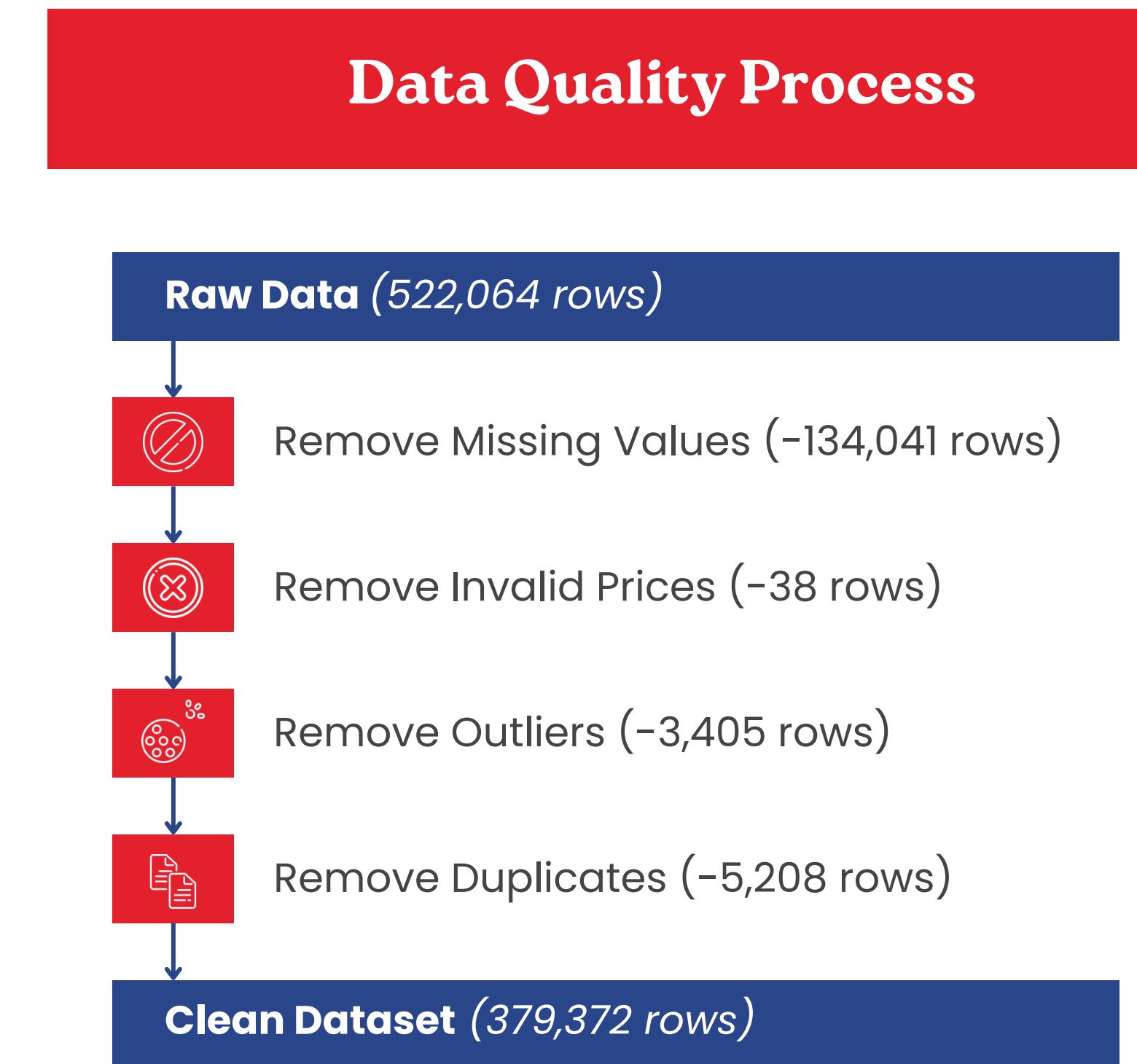
Geographic Coverage
39 Countries



Product Catalog
3,663 Products

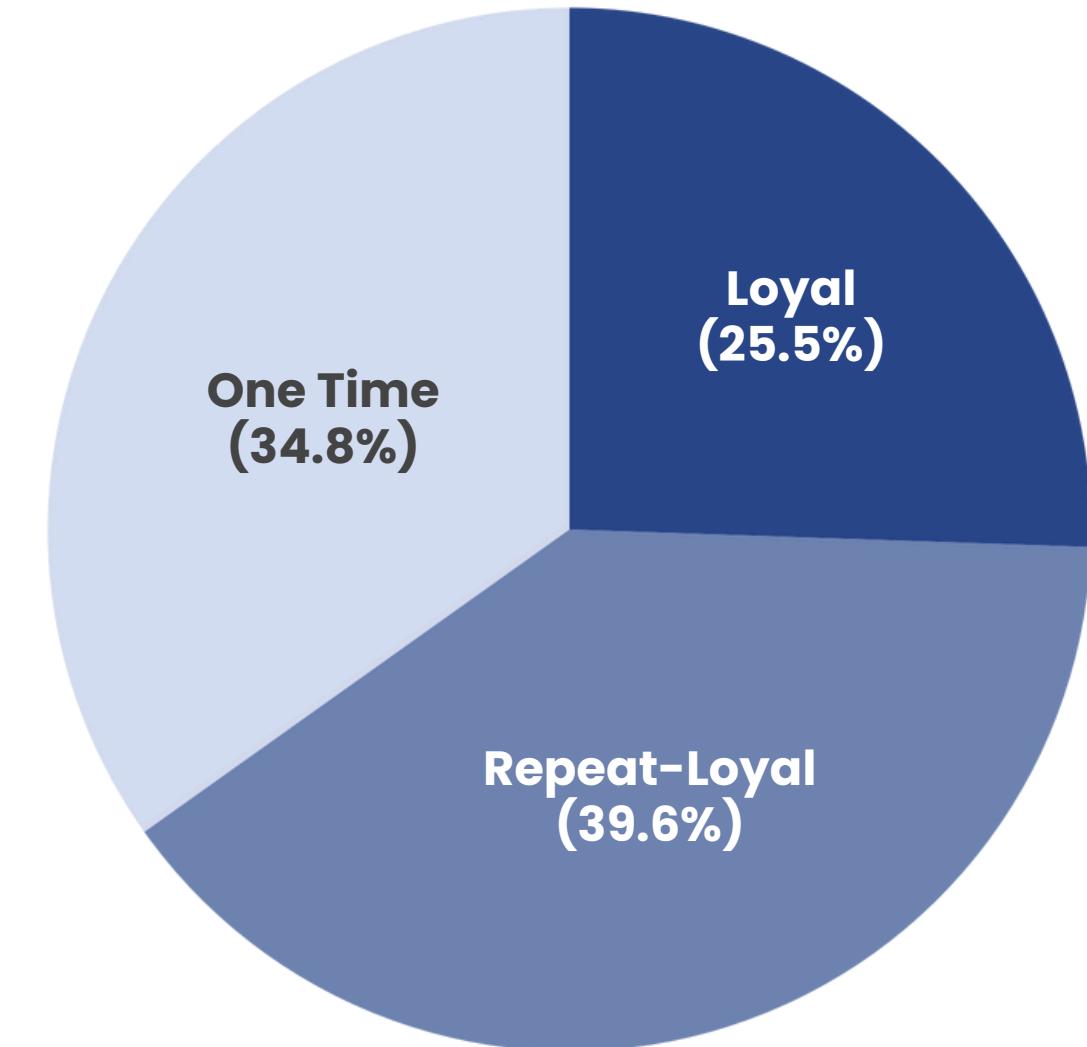


Time Period
Dec '10 - Nov '11



Customer Behavior Segmentation

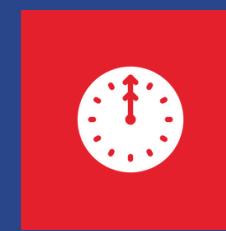
Given that loyal customers generate 5.3x higher revenue per customer (£4,444 vs £839), the primary focus should be on retaining the 1,089 loyal customers (25.5% of the base), alongside targeted re-engagement initiatives to convert 1,486 one-time buyers into repeat purchasers.



Transaction Pattern



Avg Basket Size
21.3
items



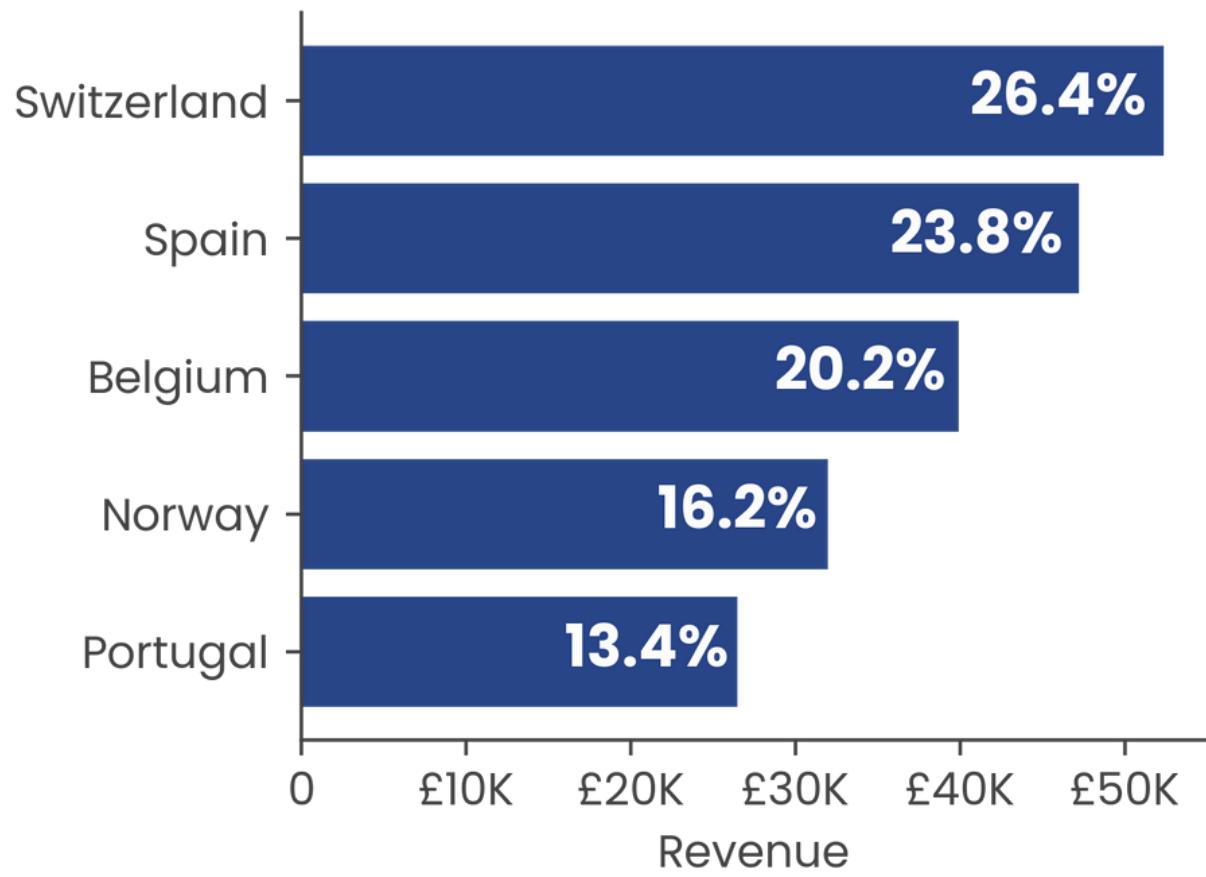
Peak Hour
12:00
noon



Peak Season
Q4
36.7% revenue

Geographic Revenue Analysis

Revenue breakdown by market segment and geographic location



Market Concentration Risk

85.5% revenue dependency on UK market creates vulnerability to country-specific economic shocks. Limited geographic diversification despite operational presence in 9 international markets. Risk mitigation requires accelerated international expansion.



Premium Segment Discovery

Netherlands (£25K CLV) and Australia (£12K CLV) demonstrate exceptional customer value from minimal customer counts (9 each). Pattern suggests B2B premium segment opportunity that justifies higher customer acquisition costs and international shipping complexity.



Next Action: Replicate Netherlands B2B playbook to Germany & France markets

Machine Learning Approach

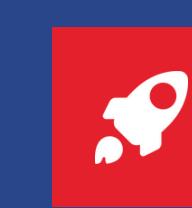


Testing 150 parameter combinations of FP-Growth was conducted to find the optimal balance between too few patterns and too much noise. The results are a minimum support of 2%, confidence level of 30%, and minimum lift of 2.0x, producing meaningful patterns without excessive noise.

Result



Rules Generated
147
high-quality



Average Lift
10.2x
strong correlation



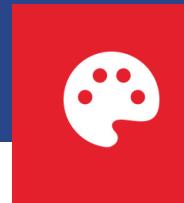
Coverage
27%
of transactions



Validation
98.5%
stability



Discovered Patterns Analysis



Color/Design Variants

Customers buying one color variant are highly likely to purchase other colors of the same product, indicates collection buying behavior

Action: Display color variants side-by-side in-store and online. Add "Available in other colors" carousel. Offer bundle discount: "Buy 2+ variants, save 15%".

21.8×

Avg Lift

65.4%

Avg Conf



Theme Collections

Products from same theme (e.g., Scandinavian Christmas) show strongest association, seasonal or themed bundling opportunity

Action: Create dedicated themed displays and landing pages. Bundle pricing: "Buy 2+ from collection, save 20%". Add gift packaging option.

30.0×

Avg Lift

74.7%

Avg Conf



Complementary Products

Different products with similar function purchased together, indicates gift shopping or multi-user household behavior

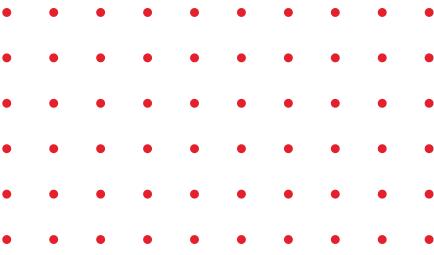
Action: Add "Frequently bought together" widget at checkout. In-store shelf talkers: "Customers also bought". Post-purchase email suggestions.

18.5×

Avg Lift

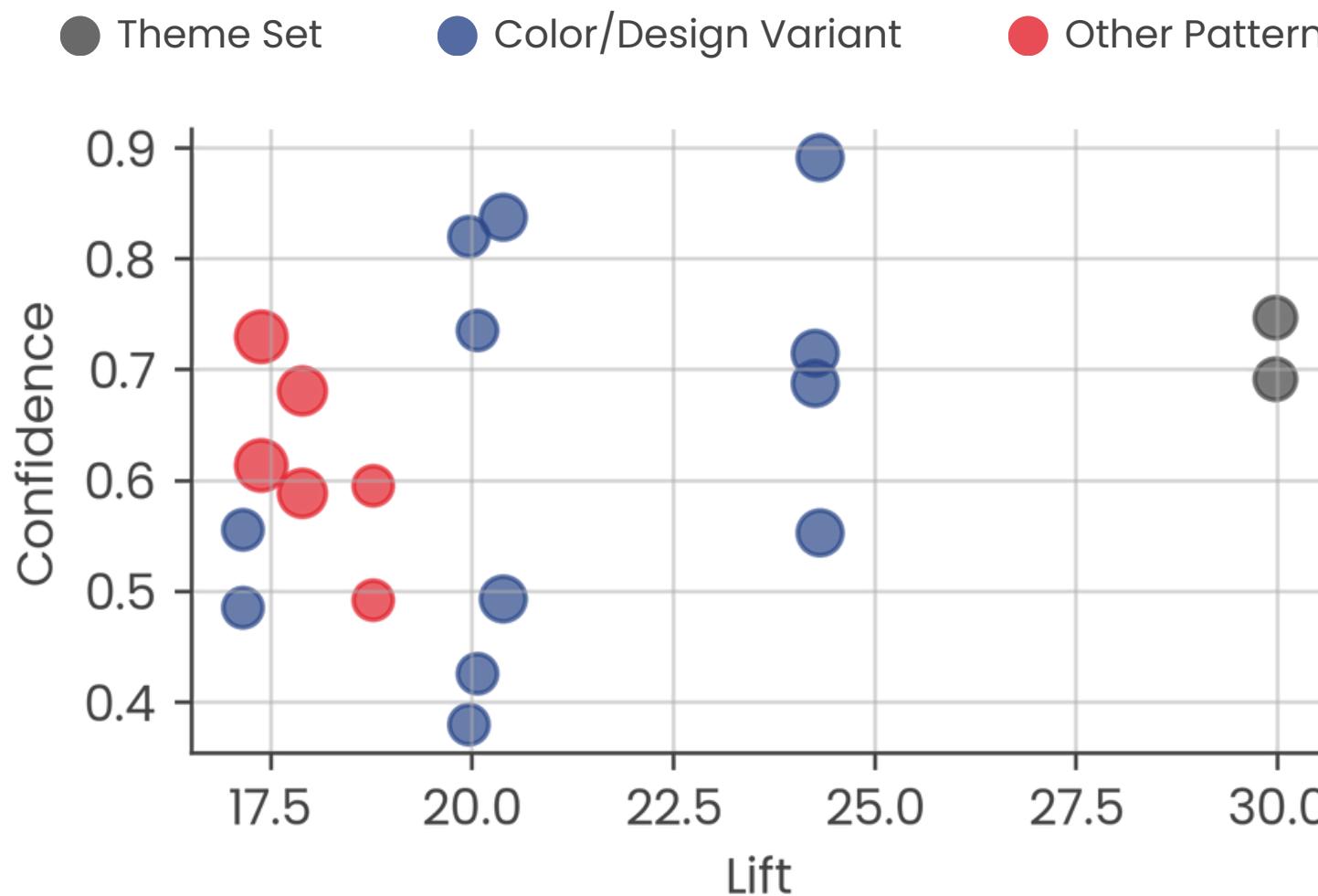
61.2%

Avg Conf



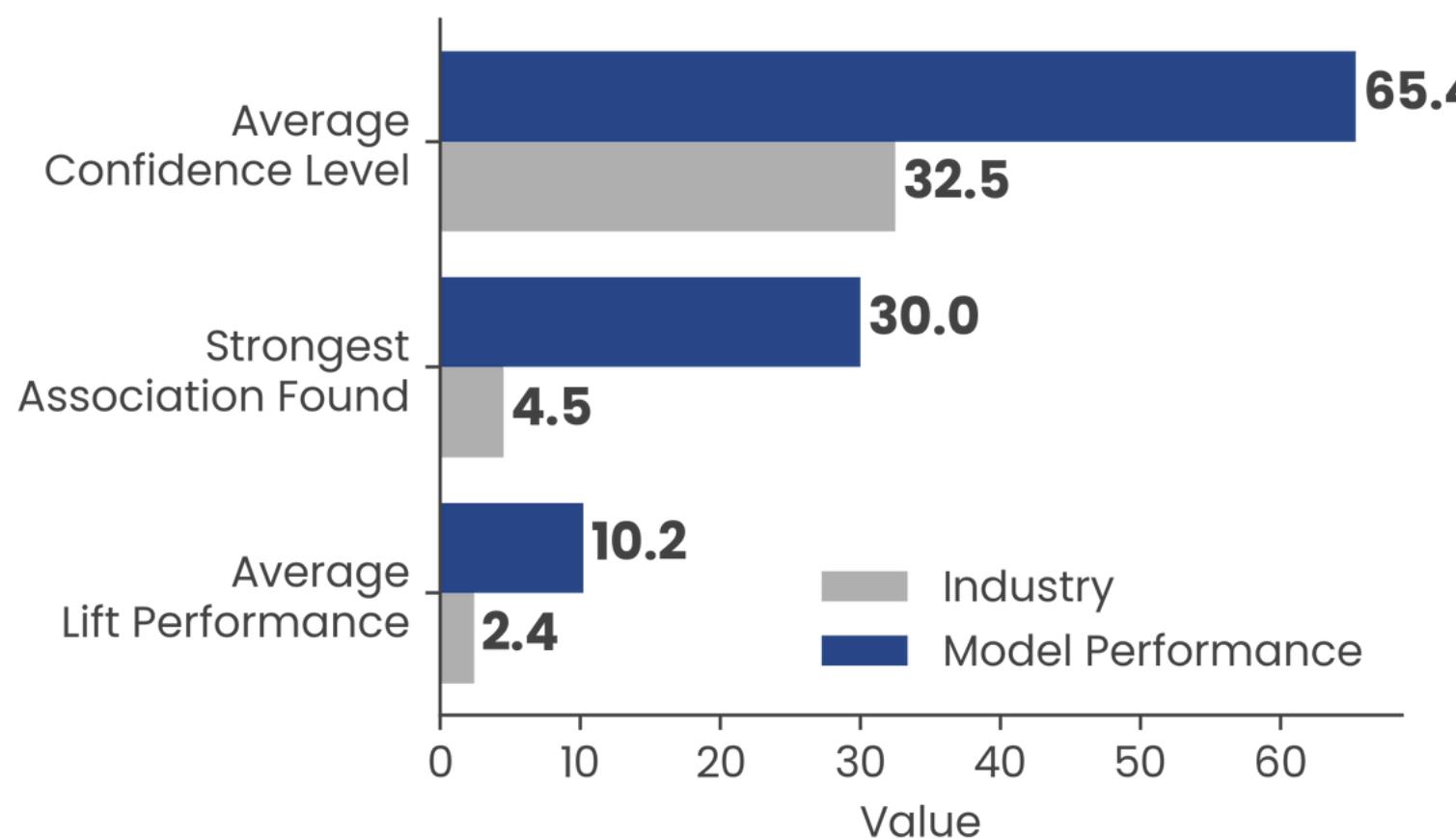
Rule Quality Assessment

From 194 initial rules generated by FP-Growth, filtered to 165 actionable rules (High + Medium quality), then removed 18 redundant rules to get 147 final rules. Train-test validation (80-20 split) confirms 191/194 rules stable with average confidence difference of 3.2%.



Industry Context & Performance

Benchmarking Model Performance Against Retail Industry Standards



Market Basket Analysis is common in retail, but results vary significantly. The "lift" metric measures how much more likely products are bought together versus separately, a 2x lift means twice as likely. Industry research shows typical retail models achieve 1.5-3.3x lift, with 3x+ considered strong. This analysis achieved 10.2x average lift (30x maximum), meaning these product associations are 3-10x stronger than industry standard, indicating highly reliable patterns for merchandising decisions.

-  **Highly reliable patterns for decision-making**
3-10x above industry benchmark
-  **Low risk of false positives**
validated on 76K test transactions
-  **Ready for production implementation**
98.5% model stability confirmed

Key Takeaways

27%

Transaction Coverage
Patterns explain over 1 in 4
customer purchases

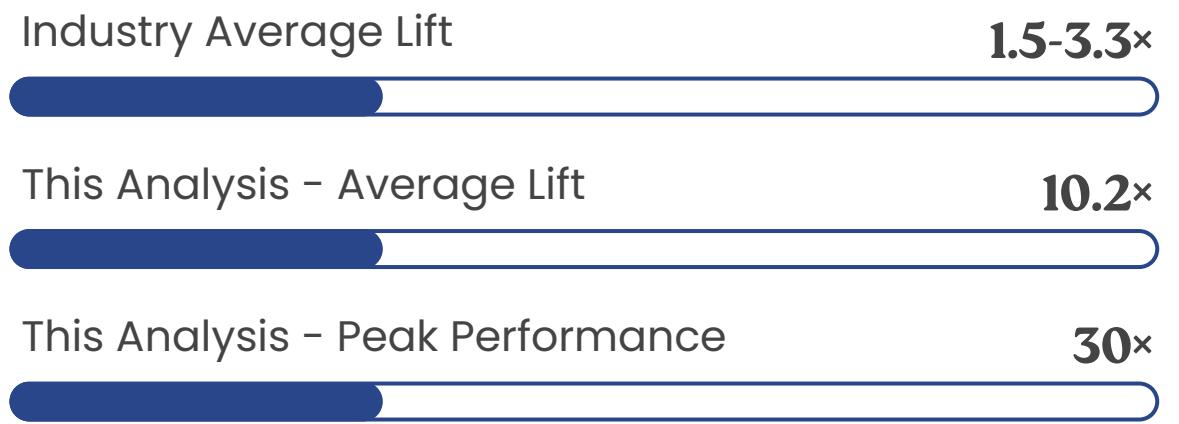
10.2×

Average Lift
3-10× stronger than industry
standard (1.5-3.3×)

98.5%

Model Stability
191 out of 194 rules remain
stable on test data

Performance vs Industry Benchmark



Project Execution Highlights

Data Quality Control

Cleaned 522K raw records to 379K validated transactions, removing 27% noise

Model Validation

Tested 150+ parameter combinations with train-test split, achieving 98.5% stability

Strategic Outcomes

Generated 147 high-quality rules across 3 patterns, covering 27% of transactions

Technical Rigor

Parameter combinations tested	150+
High-quality rules generated	147
Transaction coverage	27%

Business Translation

Color variant displays + bundles	21.8×
Seasonal theme collections	30×
Cross-sell recommendations	18.5×



Thank You

I'm passionate about turning data into actionable insights. Feel free to reach out if you'd like to connect or collaborate.



afsilmis@gmail.com



linkedin.com/in/az-zukhrufu-fi-silmi-suwondo/



github.com/afsilmis/