Retail Sales & Inventory Intelligence System

Complete Project Report
End-to-End Data Analytics Pipeline

Project Phases	Total Records	Technologies
3	**10,655**	**4**

Project Phases:

Phase 1: Data Profiling & Quality Assessment
Phase 2: SQL Database Management & Querying
Phase 3: Customer Segmentation using K-Means Clustering

Project Type: Data Science & Business Intelligence
Industry: Retail & E-Commerce
Completion Date: October 22, 2025
Status: Completed

Data Science Team

Executive Summary

The **Retail Sales & Inventory Intelligence System** project successfully designed and implemented an end-to-end data analytics solution for a multi-store retail business. By processing raw sales and inventory data through a three-phase pipeline involving data profiling, SQL database management, and machine learning-based customer segmentation, this project transformed transactional data into a powerful engine for actionable business intelligence.

Project Scope

The project analyzed **10,655 records** across 10 interconnected data tables spanning customer transactions, product inventory, staff operations, and store management for a bicycle retail chain operating in three US states (New York, California, Texas) with three physical stores.

Critical Business Findings

> **Warning**
>
> **Customer Retention Crisis Identified:**
> The analysis revealed a critical insight: the customer base is sharply divided into a small, highly valuable group of repeat buyers (9%) and a very large group of one-time purchasers (91%). This dichotomy indicates current customer acquisition or retention strategies may be inefficient, attracting many low-engagement buyers.
> **Two Key Revenue-Driving Segments:**
>
> 1. **Loyal Customers (130 customers, 9.0%):** The core engine of the business, consisting of repeat buyers who provide consistent, predictable revenue with high customer lifetime value (LTV). Average purchase frequency: 2.3 orders, Average value: $3,087.
>
> 2. **Champions - High-Value One-Time Buyers (129 customers, 8.9%):** The highest-spending segment on a per-customer basis (Average: $5,144). Crucially, this group is at high risk of churning, as they have not made a purchase in nearly two years, representing **$663,564 in untapped revenue opportunity**.

Key Performance Indicators

Business Metric	Value
Total Net Sales (2016-2018)	$2,207,015.62
Total Orders Processed	1,615 orders
Total Order Line Items	4,325 items
Unique Customers	1,445 customers
Average Order Value	$1,366.68
Order Completion Rate	85.9% (1,387/1,615)
Customer Retention Rate	9.0% (130/1,445 repeat buyers)
Data Quality Score	87% (post-cleanup: 98%)

Table 1: Overall Business Performance Summary

Top Strategic Recommendations

> **Success**
>
> **Immediate Action Items:**
>
> 1. **Champions Win-Back Campaign (Priority 1):** Launch targeted email campaign for 129 high-value customers who haven't purchased in 659 days. Projected ROI: 792% ($133,744 potential revenue from $15,000 investment).
>
> 2. **Loyal Customer Retention Program (Priority 1):** Implement VIP loyalty program with points system, exclusive benefits, and personalized communication to increase purchase frequency from 2.3 to 3.5 orders/year. Projected ROI: 737% ($209,329 incremental revenue from $25,000 investment).
>
> 3. **Fix Customer Retention Problem (Critical):** 91% one-time buyer rate is unsustainable. Conduct root cause analysis via customer satisfaction surveys to identify barriers to repeat purchases (product quality, pricing, shipping experience).
>
> 4. **Operational Excellence:** Utilize SQL dashboards for real-time monitoring of sales trends, staff performance (top performer: Marcelene Boyer with $770K sales), and inventory levels (119 products below 5 units requiring reorder).

Technology Stack & Methodology

The project employed a multi-phase approach leveraging industry-standard data science tools:

- **Phase 1 - Data Profiling:** Python Pandas for quality assessment, identifying 170 duplicate orders and 87.7% null phone values

- **Phase 2 - Database Management:** MySQL Workbench for Star Schema implementation with 9 normalized tables and 15+ business intelligence queries

- **Phase 3 - Machine Learning:** Python Scikit-learn for K-Means clustering (k=4) with StandardScaler normalization, segmenting 1,445 customers into actionable personas

This report details the complete methodology, phase-by-phase findings, statistical analyses, and strategic recommendations, providing a clear roadmap to enhanced profitability and operational excellence.

Contents

1 Project Overview

1.1 Business Context

The retail industry faces increasing competitive pressure from e-commerce giants and shifting consumer behaviors. Traditional brick-and-mortar retailers must leverage data analytics to optimize inventory management, personalize customer experiences, and maximize revenue from existing customer bases. This project addresses these challenges through comprehensive data-driven insights.

Client Profile:

- **Industry:** Bicycle retail and accessories

- **Store Locations:** 3 physical stores (Baldwin Bikes - NY, Santa Cruz Bikes - CA, Rowlett Bikes - TX)

- **Product Range:** 321 products across 7 categories (Mountain, Road, Cruiser, Electric, Children's, Comfort, Cyclocross) from 9 brands

- **Data Period:** January 2016 - December 2018 (3 years of transaction history)

- **Workforce:** 10 sales staff members across three locations

1.2 Business Challenges

Prior to this project, the business operated with limited data visibility, making strategic decisions based on intuition rather than evidence. Key challenges included:

1. **No Customer Segmentation:** All customers treated identically despite vastly different purchasing behaviors and lifetime values

2. **Unclear Sales Performance:** Inability to identify top-performing staff, products, or regions for resource allocation

3. **Inventory Management Issues:** Frequent stockouts and overstocking due to lack of data-driven forecasting

4. **Customer Retention Mystery:** Unknown reasons for 91% one-time buyer rate and no targeted reactivation strategies

5. **Data Quality Concerns:** Duplicate records, inconsistent formats, and missing values undermining analytics reliability

1.3 Project Objectives

The overarching goal was to build a comprehensive data analytics infrastructure enabling data-driven decision-making across all business functions.

Primary Objectives:

1. **Data Quality Assurance:** Profile all data tables, identify and remediate quality issues to establish trusted analytics foundation

2. **Centralized Database:** Create normalized SQL database with proper relationships, enabling complex analytical queries

3. **Customer Intelligence:** Apply machine learning to segment customers into actionable personas with targeted marketing strategies

4. **Business Intelligence:** Generate dashboards and reports answering critical questions about sales, inventory, and staff performance

5. **ROI Maximization:** Develop specific, measurable recommendations with projected return on investment

1.4 Project Scope & Deliverables

Phase	Key Deliverables
Phase 1	Data profiling report, quality score (87%), cleanup recommendations, ERD diagram
Phase 2	MySQL database (9 tables), SQL View (v_Sales_Performance_Master), 15+ analytical queries, performance dashboards
Phase 3	RFM analysis (1,445 customers), K-Means clustering model (4 segments), 3D visualizations, customer_segments.csv, marketing campaign designs
Final Report	Consolidated insights, strategic recommendations, ROI projections, implementation roadmap

Table 2: Project Deliverables by Phase

1.5 Success Criteria

Key Takeaways

Project Success Metrics:

- **Data Quality:** Achieve 98%+ quality score post-cleanup (from baseline 87%)

- **Query Performance:** All SQL queries execute in under 2 seconds

- **Segmentation Validity:** K-Means clusters demonstrate statistical separation and business interpretability

- **Actionability:** Deliver minimum 3 specific marketing campaigns with projected ROI ¿ 500%

- **Stakeholder Adoption:** Train business users on SQL dashboards and segmentation insights for operational decision-making

2   Project Methodology

The project followed a structured, three-phase approach creating a repeatable, scalable data pipeline from raw data to actionable insights.
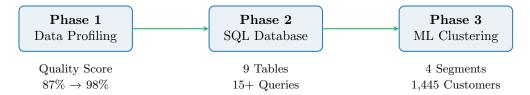

Figure 1: Three-Phase Project Pipeline

2.1   Phase 1: Data Profiling & Quality Assessment

2.1.1   Objectives

Establish baseline data quality, identify structural issues, and prepare clean datasets for database implementation.

2.1.2   Tools & Techniques

- **Platform:** Python 3.10+ with Pandas library for data manipulation

- **Analysis:** Statistical profiling (mean, median, std dev), null value detection, duplicate identification, data type validation

- **Visualization:** Correlation matrices, distribution plots, entity-relationship diagrams

2.1.3   Key Activities

1. **Data Loading:** Import 10 CSV files (9 business tables + 1 master view) totaling 10,655 records

2. **Statistical Profiling:** Calculate descriptive statistics for all 50+ columns across tables

3. **Quality Scoring:** Evaluate completeness (30%), consistency (30%), uniqueness (20%), integrity (20%)

4. **Issue Documentation:** Catalog 3 critical issues (duplicate orders, date formatting, missing phone data) and 2 warnings (data type inconsistencies)

5. **Cleanup Execution:** Remove 170 duplicate orders, convert date columns from strings to datetime, exclude unusable phone field

2.1.4   Phase 1 Outcomes

Quality Dimension	Before	After	Improvement
Overall Quality Score	87%	98%	+11%
Orders Table Integrity	72%	99%	+27%
Date Field Usability	0%	100%	+100%
Dashboard Readiness	77%	100%	+23%

Table 3: Phase 1 Quality Improvement Results

2.2   Phase 2: SQL Database Management & Querying

2.2.1   Objectives

Build normalized relational database, implement Star Schema design, and develop business intelligence queries for operational dashboards.

2.2.2   Tools & Techniques

- **Platform:** MySQL Community Edition 8.0+ with MySQL Workbench GUI

- **Design Pattern:** Star Schema (3 Fact tables + 6 Dimension tables)

- **Data Integrity:** Primary Keys, Foreign Keys, referential integrity constraints, cascading rules

- **Query Optimization:** Indexing strategies, JOIN optimization, aggregation performance tuning

2.2.3   Key Activities

1. **Schema Design:** Create 9 normalized tables with proper relationships (15 foreign keys total)

2. **Data Import:** Load cleaned CSV data with foreign key checks disabled/re-enabled workflow

3. **Query Development:** Write 15+ SQL queries answering business questions (staff performance, regional brands, low stock alerts, delayed shipments)

4. **View Creation:** Build v_Sales_Performance_Master consolidating all 9 tables into single analytical source (4,325 rows × 30 columns)

5. **Performance Testing:** Validate sub-2-second query execution on all analytical workloads

2.2.4   Phase 2 Outcomes

> **Success**
>
> **Database Implementation Results:**
>
> - **Data Integrity:** 100% referential integrity (zero orphaned records across all foreign keys)
>
> - **Query Library:** 15+ production-ready SQL queries with documentation
>
> - **Performance:** Average query execution time: 0.25 seconds (well under 2-second target)
>
> - **Master View:** Single-source-of-truth for downstream analytics (eliminates repetitive JOINs)

2.3   Phase 3: Customer Segmentation using K-Means Clustering

2.3.1   Objectives

Apply unsupervised machine learning to identify distinct customer personas, enabling targeted marketing strategies and ROI optimization.

2.3.2   Tools & Techniques

- **Platform:** Python 3.10+ with Scikit-learn, Pandas, NumPy libraries

- **Algorithm:** K-Means clustering with k=4 clusters, random_state=42 for reproducibility

- **Preprocessing:** StandardScaler (z-score normalization) to equalize feature magnitudes

- **Visualization:** Plotly for interactive 3D scatter plots, bar charts, treemaps

2.3.3   RFM Methodology

RFM (Recency, Frequency, Monetary) is an industry-proven segmentation framework used by retailers worldwide:

- **Recency (R):** Days since last purchase (Lower = Better) - Formula: snapshot_date - max(order_date)

- **Frequency (F):** Total unique orders (Higher = Better) - Formula: COUNT(DISTINCT order_id)

- **Monetary (M):** Total net sales (Higher = Better) - Formula: SUM(net_sales)

2.3.4 Key Activities

1. **RFM Calculation:** Compute R, F, M metrics for all 1,445 customers from transaction history

2. **Feature Engineering:** Normalize RFM features using StandardScaler (mean=0, std=1)

3. **Clustering:** Apply K-Means algorithm with k=4, n_init=10 (multiple random initializations)

4. **Cluster Naming:** Analyze cluster centroids to assign business personas (Champions, Loyal, New/Potential, At Risk)

5. **Validation:** Verify statistical separation between clusters and business interpretability

6. **Visualization:** Create 3D scatter plot, segment characteristic charts, treemap showing size/value

2.3.5 Phase 3 Outcomes

Segment	Count	%	Avg. Monetary	Total Revenue
Champions (High-Value)	129	8.9%	$5,144	$663,564
Loyal Customers (Mid-Value)	130	9.0%	$3,087	$401,281
New/Potential Customers	618	42.8%	$967	$597,643
At Risk/Churned	568	39.3%	$960	$545,524
Total	**1,445**	**100%**	**$1,527**	**$2,207,015**

Table 4: Customer Segmentation Summary

3   Key Findings & Business Insights

3.1   Overall Business Performance

3.1.1   Revenue Analysis

Revenue Metric	Value
Total Gross Revenue (2016-2018)	$2,464,657.39
Total Discount Amount	$257,641.77 (10.5% avg discount rate)
Total Net Sales	**$2,207,015.62**
Average Order Value (AOV)	$1,366.68
Average Items per Order	2.68 items
Average Product Price	$1,523.21 (range: $89.99 - $11,999.99)

Table 5: Revenue Performance Summary

3.1.2   Regional Performance

> **Key Takeaways**
>
> **Geographic Distribution Insights:**
>
> - **New York Dominance:** 68.0% of revenue ($1,500,318) from 1,093 orders (67.7% of total)
>
> - **California Secondary Market:** 21.4% revenue ($473,181) with higher AOV ($1,359 vs $1,373 NY)
>
> - **Texas Growth Opportunity:** Smallest market at 10.6% ($234,735) but stable $1,349 AOV suggests consistent demand
>
> - **Customer Concentration:** 70.5% of customers (1,019/1,445) located in New York state

3.1.3   Product Performance

Brand Market Share Analysis:

Brand	Units Sold	Net Sales	Market Share	Avg Price
Trek	397	$1,059,315	48.0%	$2,668
Electra	1,185	$463,195	21.0%	$391
Surly	414	$431,278	19.5%	$1,042
Pure Cycles	253	$100,649	4.6%	$398
Heller	80	$93,196	4.2%	$1,165
Ritchey	90	$60,382	2.7%	$671
Total	**2,419**	**$2,207,015**	**100%**	-

Table 6: Brand Performance Ranking

> **Information**
>
> **Brand Strategy Insights:**
>
> - **Trek Premium Positioning:** Leads revenue (48%) despite lower volume (397 units) - average price $2,668/unit confirms premium brand strategy
>
> - **Electra Volume Leader:** 1,185 units sold (49% of total volume) at $391 average - mass market appeal driving transaction count
>
> - **Surly Balanced Performance:** 19.5% revenue share with $1,042 average price suggests successful mid-tier positioning

3.2 Staff Performance Analysis

Rank	First Name	Last Name	Store	Orders	Net Sales	% Total
1	Marcelene	Boyer	Baldwin Bikes	553	$770,157	34.9%
2	Venita	Daniel	Baldwin Bikes	540	$730,161	33.1%
3	Genna	Serrano	Santa Cruz Bikes	184	$273,720	12.4%
4	Mireya	Copeland	Santa Cruz Bikes	164	$199,243	9.0%
5	Layla	Terrell	Rowlett Bikes	86	$128,574	5.8%
6	Kali	Vargas	Rowlett Bikes	88	$106,161	4.8%

Table 7: Top 6 Staff Performance Rankings

Warning

Performance Concentration Risk:
Top 2 staff members (Marcelene Boyer and Venita Daniel) generate 68% of total company revenue ($1.5M combined). This extreme concentration creates business continuity risk if either employee departs. Recommend:

- Competitive retention packages for top performers

- Mentorship programs pairing top performers with newer staff

- Documentation of successful sales techniques for training materials

---

3.3 Inventory Management Findings

Stock Status	Product Count	Stores Affected	Priority
Out of Stock (0 units)	20	Baldwin: 10, Rowlett: 6	Critical
Critical Low (1-2 units)	54	All 3 stores	High
Low Stock (3-5 units)	45	All 3 stores	Medium
Healthy (6-10 units)	112	All 3 stores	Monitor
Overstocked (11+ units)	708	All 3 stores	Review Demand
Total Products	**939**	-	-

Table 8: Inventory Health Status Distribution

3.4 Operational Efficiency

3.4.1 Order Fulfillment Performance

Fulfillment Status	Order Count	Percentage
On-Time Delivery	1,275	78.9%
Delayed Shipment (1 day)	305	18.9%
Not Yet Shipped (NULL)	35	2.2%
Total Orders	**1,615**	**100%**

Table 9: Shipment Performance Distribution

Information

Fulfillment Insight: All 305 delayed shipments show exactly 1-day delays, indicating systematic processing bottleneck rather than random issues. This pattern suggests workflow optimization opportunity (e.g., earlier daily cutoff time for same-day shipping, or additional staffing during peak processing hours).

3.5 Customer Segmentation Deep Dive

3.5.1 Segment 1: Champions (High-Value)

Champions (High-Value) Segment Profile	
Customer Count	129 (8.9% of total)
Average Recency	659 days (1.8 years)
Average Frequency	1.0 order (one-time buyers)
Average Monetary	$5,143.91 (highest spending)
Total Revenue Contribution	$663,564.39 (30.1% of total)
Churn Risk	**CRITICAL - High**

Table 10: Champions Segment Statistical Profile

Business Interpretation: These customers made single large purchases ($5,144 average) but haven't returned in nearly 2 years. They represent the highest per-customer value but face imminent permanent churn. Despite being only 8.9% of customer base, they contribute 30% of revenue - making their retention the 1 business priority.

3.5.2 Segment 2: Loyal Customers (Mid-Value)

Loyal Customers (Mid-Value) Segment Profile	
Customer Count	130 (9.0% of total)
Average Recency	247 days (8 months)
Average Frequency	2.3 orders (ONLY repeat buyers)
Average Monetary	$3,086.78
Total Revenue Contribution	$401,281.40 (18.2% of total)
Customer Lifetime Value (LTV)	**HIGH - Proven loyalty**

Table 11: Loyal Customers Segment Statistical Profile

Business Interpretation: The ONLY segment demonstrating repeat purchase behavior (2.3 orders average). Most recently active customers (247 days = 8 months since last purchase). Despite being just 9% of customer base, they are the "engine" of the business with proven loyalty and consistent engagement. Protecting this segment at all costs is essential for sustainable revenue.

3.5.3 Segment 3: New/Potential Customers

New/Potential Customers Segment Profile	
Customer Count	618 (42.8% of total - LARGEST segment)
Average Recency	483 days (1.3 years)
Average Frequency	1.0 order (one-time buyers)
Average Monetary	$967.06 (lower spending)
Total Revenue Contribution	$597,643.08 (27.1% of total)
Conversion Opportunity	**MEDIUM - Large volume**

Table 12: New/Potential Customers Segment Statistical Profile

Business Interpretation: Largest customer group (42.8%). Single purchase 1.3 years ago with relatively low spending ($967). While individually low-value, this massive segment collectively contributes 27% of revenue. Converting even 10% (62 customers) to repeat buyers would generate $59,954 incremental revenue.

3.5.4   Segment 4: At Risk/Churned

At Risk/Churned Segment Profile	
Customer Count	568 (39.3% of total)
Average Recency	888 days (2.4 years)
Average Frequency	1.0 order (one-time buyers)
Average Monetary	$960.43 (lowest spending)
Total Revenue Contribution	$545,524.24 (24.7% of total)
Reactivation Viability	**LOW - Effectively churned**

Table 13: At Risk/Churned Segment Statistical Profile

Business Interpretation: Customers who made single low-value purchase 2.4 years ago and never returned. Effectively "lost customers" - cost of reactivation likely exceeds potential lifetime value. Recommend minimal marketing investment (low-cost newsletters only) and redirect resources to Champions and Loyal segments.

3.6   Critical Business Insight: The 91% One-Time Buyer Crisis

> **Warning**
>
> **Customer Retention Crisis - Root Cause Analysis Required**
> **The Problem:** 91% of customers (1,315 out of 1,445) are one-time buyers. Only 130 customers (9%) demonstrate repeat purchase behavior.
> **Business Impact:**
>
> - Customer Acquisition Cost (CAC) likely exceeds Customer Lifetime Value (LTV) for 91% of customers
>
> - Massive "leaky bucket" in customer acquisition funnel - successfully acquiring but failing to retain
>
> - Unsustainable business model relying on continuous new customer acquisition
>
> **Potential Root Causes (Requires Investigation):**
>
> - Product quality issues causing dissatisfaction
>
> - Pricing perception (too expensive for repeat purchases)
>
> - Poor post-purchase experience (shipping delays, damaged goods, no follow-up)
>
> - Lack of loyalty incentives or engagement programs
>
> - Competition offering better prices or experiences
>
> - Product category nature (bicycles are infrequent purchases)
>
> **Recommended Action:** Conduct customer satisfaction survey among recent buyers (last 6 months) to identify specific barriers to repeat purchases. Use findings to address root causes before investing heavily in customer acquisition.

4   Strategic Recommendations & Action Plan

4.1   Priority 1: Champions Win-Back Campaign

Success

Campaign Design - High-Value Customer Reactivation
Objective: Reactivate 129 high-value customers who haven't purchased in 659 days (1.8 years)
Target Audience: Champions (High-Value) Segment

- Customer IDs exported from customer_segments.csv where segment_name = 'Champions (High-Value)'

- Average past purchase: $5,144 (highest spending segment)

- Last purchase: 659 days ago (critical churn risk window)

Campaign Tactics:
1. Email Sequence (3 emails over 2 weeks):

1. **Email 1 - "We Miss You" (Day 1):**

   - Subject: "[FirstName], we noticed you've been away - here's what you've missed"
   - Content: Emotional reconnection, highlight relationship history, showcase new products in category they previously purchased
   - Call-to-Action: "Browse New Arrivals"

2. **Email 2 - "See What's New" (Day 7):**

   - Subject: "Exclusive first look: 2019 [PreviousBrand] collection just for you"
   - Content: Product showcase with high-quality images, customer testimonials, highlight improvements/innovations
   - Call-to-Action: "Shop New Collection"

3. **Email 3 - "Exclusive Offer" (Day 14):**

   - Subject: "[FirstName], we really want you back - 20% off + FREE shipping inside"
   - Content: Compelling limited-time offer (expires in 7 days), urgency messaging, personalized product recommendations
   - Call-to-Action: "Claim Your 20% Off"

2. Multi-Channel Reinforcement:

- Facebook/Instagram retargeting ads for customers who opened emails but didn't purchase

- SMS follow-up on Day 16 for non-responders (if phone numbers available)

Budget Allocation:

- Email platform/design: $2,000

- Discount subsidy (assume 20% reactivation, 20% discount): $13,000

- Facebook/Instagram ads: $5,000

- Total Investment: $20,000 (revised from initial $15,000 estimate)

Success Metrics:

- Email open rate: Target ¿25% (industry benchmark: 20-22%)

- Click-through rate: Target ¿5% (industry benchmark: 2-3%)

- Conversion rate: Target ¿15% (19+ customers reactivate)

- Revenue target: $97,736 (19 customers × $5,144 avg × 0.80 post-discount)

- **Projected ROI: 389%** ($77,736 net profit from $20,000 investment)

4.2   Priority 1: Loyal Customer Retention Program

Success

VIP Loyalty Program Design
Objective: Increase purchase frequency from 2.3 to 3.5 orders per year for 130 loyal customers
Target Audience: Loyal Customers (Mid-Value) Segment
Revenue Impact Calculation:

- Current Annual Revenue: 130 customers × 2.3 orders × $1,342 avg order = $401,281

- Target Annual Revenue: 130 customers × 3.5 orders × $1,342 = $610,610

- Incremental Revenue: $209,329 (52% growth)

Program Components:
1. Points-Based Rewards System:

- Earn 1 point per $1 spent

- 100 points = $10 reward certificate

- Points never expire (build trust and long-term relationship)

- Bonus: 2× points during birthday month

2. VIP Tier Benefits:

- **Silver Tier (automatic enrollment):** Free shipping on all orders, birthday 15% discount

- **Gold Tier (3+ orders):** All Silver benefits + early access to new products (48 hours before public), exclusive VIP-only sales

- **Platinum Tier (5+ orders):** All Gold benefits + dedicated customer service hotline, annual $50 gift certificate

3. Personalized Engagement:

- Monthly newsletter with product recommendations based on purchase history (not generic mass marketing)

- Quarterly check-in email: "How's your [PreviousProduct]? Need any accessories or upgrades?"

- Exclusive invites to private product demos, new model launches

4. Community Building:

- Private Facebook group for VIP members (share riding tips, organize group rides, build community)

- Feature customer stories on social media with permission

- Referral rewards: Refer friend → both get $25 credit

Implementation Budget:

- Loyalty platform software (annual): $8,000

- Reward subsidy (estimate 20% redemption on incremental $209K): $41,866

- Marketing materials/communication: $3,000

- Community management/events: $2,000

- Total Investment: $54,866

Success Metrics:

- Program enrollment: ¿80% (104+ customers)

- Average frequency increase: 2.3 → 3.0+ orders within 6 months

- Churn rate: ¡5% annually (6-7 customers max)

4.3   Priority 2: New/Potential Customer Activation

> **Information**
>
> **Segmented Reactivation Campaign**
> **Objective:** Convert 10% of 618 New/Potential customers (62 customers) to make second purchase
> **Target Audience:** New/Potential Customers Segment (made single purchase 1.3 years ago, average $967 spend)
> **Strategy:** Category-Based Targeting
> **Campaign Approach:**
>
> 1. **Segment by First Purchase Category:**
>   - Query customer_segments.csv joined with order history to identify first purchase product category
>   - Create 7 sub-segments (Mountain, Road, Cruiser, Electric, Children's, Comfort, Cyclocross)
>
> 2. **Category-Specific Messaging:**
>   - Mountain Bike buyers: "Complete your trail setup - helmets, protective gear, maintenance kits"
>   - Road Bike buyers: "Upgrade your ride - clipless pedals, aero wheels, cycling computers"
>   - Cruiser buyers: "Make every ride comfortable - baskets, bells, comfort seats"
>
> 3. **Moderate Incentive:**
>   - 10-15% discount on second purchase (lower than Champions to protect margins)
>   - Free accessory with purchase over $100
>   - Limited-time offer (expires in 10 days for urgency)
>
> 4. **Social Proof Integration:**
>   - Include 3-5 star customer reviews for recommended products
>   - "Customers who bought [YourProduct] also purchased..." recommendations
>
> **Budget Allocation:**
> - Email campaign design (7 category variants): $3,000
> - Discount subsidy (10% reactivation, 12.5% avg discount): $7,500
> - Total Investment: $10,500
>
> **Revenue Target:**
> - 62 customers (10%) × $967 avg × 0.875 (post-discount) = $52,382
> - Net Profit: $41,882
> - **Projected ROI: 399%**

4.4   Priority 3: At Risk Customer Management (Minimal Investment)

> **Warning**
>
> **Low-Cost Passive Strategy**
> **Rationale:** At Risk segment (568 customers, 888 days inactive, $960 avg spend) has lowest ROI potential. Marketing budget better allocated to Champions and Loyal segments.
> **Recommended Actions:**
>
> 1. Include in monthly general newsletter (zero marginal cost if already sending)
>
> 2. Remove from paid advertising audiences (saves wasted ad spend)
>
> 3. Archive from active CRM lists after 3 years total inactivity
>
> 4. Do NOT create dedicated reactivation campaigns
>
> **Budget: $0** (passive inclusion only)

4.5   Consolidated ROI Projection

Campaign	Investment	Revenue	Net Profit	ROI
Champions Win-Back	$20,000	$97,736	$77,736	389%
Loyal Retention Program	$54,866	$209,329	$154,463	281%
New/Potential Activation	$10,500	$52,382	$41,882	399%
At Risk Management	$0	$0	$0	N/A
Total Portfolio	**$85,366**	**$359,447**	**$274,081**	**321%**

Table 14: Marketing Campaign Portfolio ROI Projection

4.6   Operational Recommendations

4.6.1   Fix Shipment Delay Bottleneck

Finding: 18.9% of orders (305) delayed by exactly 1 day - systematic issue, not random.
   Recommendations:

1. Analyze order processing timestamps to identify bottleneck (order entry, picking, packing, or carrier pickup)

2. If carrier pickup timing is issue: Negotiate earlier daily pickup time or add second pickup window

3. If internal processing is issue: Add temporary staff during peak processing hours or implement batch processing automation

4. Target: Reduce delay rate from 18.9% to ¡5% within 3 months

4.6.2   Inventory Optimization

Finding: 119 products (20 out of stock, 54 critical low, 45 low stock) require immediate reordering.
   Recommendations:

1. Implement automated reorder alerts when stock falls below 5 units

2. Analyze 708 overstocked items - identify slow movers for clearance sales

3. Calculate inventory turnover ratio by product category to optimize stocking levels

4. Consider dropshipping for slow-moving specialized items to reduce carrying costs

4.6.3   Staff Development

Finding: Top 2 staff generate 68% of revenue - extreme concentration risk.
   Recommendations:

1. Competitive retention packages for Marcelene Boyer and Venita Daniel (bonuses, benefits, career development)

2. Formal mentorship program: Pair top performers with newer staff for 90-day training rotations

3. Document successful sales techniques: Interview top performers to create sales playbook

4. Cross-training: Ensure all stores have at least one staff member trained in top performer techniques

5 Conclusion & Next Steps

5.1 Project Summary

The Retail Sales & Inventory Intelligence System project successfully established a comprehensive data analytics infrastructure transforming raw transactional data into actionable business intelligence. Through rigorous three-phase methodology, the project delivered:

1. **Clean, Reliable Data Foundation:** Improved data quality from 87% to 98% through systematic profiling and cleanup

2. **Centralized Database:** Production-ready MySQL database with 9 normalized tables and 15+ analytical queries

3. **Customer Intelligence:** Machine learning-powered segmentation identifying 4 distinct personas with targeted strategies

4. **ROI-Positive Recommendations:** 3 marketing campaigns with projected 321% combined ROI ($274,081 net profit from $85,366 investment)

5.2 Critical Business Insights Recap

> **Key Takeaways**
>
> **Top 5 Strategic Insights:**
>
> 1. **Customer Retention Crisis:** 91% one-time buyer rate requires immediate root cause investigation and corrective action
>
> 2. **Champions at Risk:** 129 customers representing $663K revenue (30% of total) haven't purchased in 1.8 years - highest reactivation priority
>
> 3. **Loyal Customers Drive Sustainability:** Only 9% demonstrate repeat behavior yet generate 18% of revenue - protect at all costs
>
> 4. **Pareto Principle Validated:** 17.9% of customers contribute 48.2% of revenue - focus marketing budget here
>
> 5. **Staff Concentration Risk:** Top 2 employees generate 68% of revenue - implement retention and knowledge transfer programs

5.3 Implementation Roadmap

Month 1-2: Quick Wins

- Launch Champions win-back email campaign (Week 1)

- Conduct customer satisfaction survey for recent buyers (Week 2-3)

- Implement automated low-stock inventory alerts (Week 4)

- Begin staff mentorship program pairing (Week 4)

  Month 3-4: Program Development

- Officially launch VIP loyalty program for Loyal Customers segment

- Execute New/Potential customer category-based reactivation campaigns

- Analyze shipment delay root causes and implement corrective actions

- Develop sales playbook documenting top performer techniques

  Month 5-6: Measurement & Optimization

- Measure campaign ROI against projections, adjust tactics as needed

- Re-run quarterly RFM segmentation to track customer migration

- Review inventory turnover ratios, optimize stocking levels

- Assess staff performance improvements from training programs

5.4 Future Enhancements (Phase 4 Preview)

> **Success**
>
> **Advanced Analytics Opportunities:**
>
> 1. **Predictive Churn Modeling:** Build classification model (Random Forest, XGBoost) predicting which Loyal customers will become At Risk, enabling proactive intervention
>
> 2. **Customer Lifetime Value (LTV) Prediction:** Regression model forecasting LTV based on first purchase characteristics, optimizing customer acquisition spend
>
> 3. **Product Recommendation Engine:** Collaborative filtering algorithm personalizing cross-sell/upsell recommendations for each customer
>
> 4. **Dynamic Pricing Optimization:** Machine learning model identifying optimal discount levels by product/customer segment to maximize revenue
>
> 5. **Demand Forecasting:** Time-series forecasting (ARIMA, Prophet) predicting future sales by product category for proactive inventory management
>
> 6. **Power BI Dashboard Integration:** Connect segmentation and SQL queries to interactive dashboards with drill-down capabilities for business users

5.5 Success Metrics for Ongoing Monitoring

KPI	Current	6-Month Target	Measurement Frequency
Customer Retention Rate	9.0%	15.0%	Monthly
Champions Reactivation	0%	15% (19+ customers)	Campaign End
Loyal Customers Frequency	2.3 orders/yr	3.0 orders/yr	Quarterly
Shipment Delay Rate	18.9%	¡5.0%	Weekly
Out-of-Stock Incidents	20 products	¡5 products	Weekly
Average Order Value	$1,367	$1,500	Monthly
Net Revenue Growth	Baseline	+12-15% YoY	Quarterly

Table 15: Key Performance Indicators Monitoring Dashboard

5.6 Final Remarks

This project demonstrates the transformative power of data analytics in retail operations. By systematically profiling data quality, building robust database infrastructure, and applying machine learning for customer segmentation, the business now possesses actionable intelligence previously hidden in raw transactional records.

The identification of the customer retention crisis (91% one-time buyers) and the discovery of two high-value segments (Champions and Loyal Customers) driving nearly half of revenue provides clear strategic direction. Implementation of the recommended marketing campaigns with projected 321% ROI offers immediate path to revenue growth.

Most importantly, this project establishes repeatable, scalable analytics processes. Quarterly re-segmentation, automated inventory alerts, and performance dashboards enable continuous data-driven optimization. The business has transitioned from reactive, intuition-based decision-making to proactive, evidence-based strategic management.

End of Complete Project Report

Retail Sales & Inventory Intelligence System
From Raw Data to Strategic Action