Regional Sales Analysis — Comprehensive Project Report

Executive Summary

This comprehensive exploratory data analysis (EDA) and business intelligence project analyzed Acme Co.'s 5-year USA regional sales dataset (2014–2018) comprising 64,104 transactions across 175 customers, 30 products, and 4 US regions. The analysis uncovered critical patterns in seasonality, product performance, channel profitability, geographic concentration, and customer segmentation that inform strategic recommendations for revenue optimization and risk mitigation.

Five Key Findings:

1. **Pronounced seasonality** with January peaks ($124M average) and April troughs ($95M), indicating strong seasonal recovery opportunities
2. **SKU concentration risk** where Products 26 & 25 drive ~25% of total revenue, requiring portfolio diversification
3. **Channel margin divergence** with Wholesale capturing 54% of volume but Export channels achieving ~38% average margins
4. **Geographic dominance** with California alone generating 7.6K orders ($230M+), representing concentration risk
5. **Customer revenue clustering** where Aibox Company and State Ltd disproportionately drive profitability

---

Problem Statement

Sales teams often lack a clear, data-driven understanding of regional performance, making it difficult to identify growth opportunities and optimize resources. This project addresses three core business challenges:

Business Challenges Addressed:

- **Inconsistent revenue and profit performance** across US regions despite 5 years of historical data
- **Lack of visibility** into seasonal swings, top-performing SKUs, channel profitability, and margin drivers
- **Resource optimization gaps** preventing informed decisions on inventory, promotions, pricing, and regional investment

Objective: Leverage 5 years of historical transaction data to pinpoint growth levers, evaluate regional and channel strategies, and optimize sales policies and operational planning (warehouse preparation, staffing, inventory management aligned with seasonal trends).

Project Approach

Two-Phase Solution Framework

Phase 1: Exploratory Data Analysis (EDA)

- Dive into historical sales, margins, products, channels, and regions
- Surface trends, outliers, and relationships through statistical analysis and visualization
- Generate actionable insights for business decision-making

Phase 2: Interactive Dashboard & Strategic Recommendations

- Build a live Power BI dashboard for business users to self-serve insights
- Enable ad-hoc slicing by time, product, region, and channel
- Present strategic findings supporting tactical and strategic planning

Data Overview & Architecture

Source Data Structure

The project consolidates data from six interconnected Excel sheets:

Sheet	Records	Purpose
Sales Orders	64,104	Transaction-level data: order number, date, customer, channel, product, quantity, pricing, cost
Customers	175	Customer master: customer index and names
Products	30	Product master: product index and names
Regions	994	Geographic data: city, state, coordinates, population, income, demographics
State Regions	49	State-to-US region mapping: state code, state name, US region (West, South, Midwest, Northeast)
2017 Budgets	30	Product-level budget targets for 2017

Final Dataset Composition

After integration, the analysis dataset contains 64,104 records with 21 dimensional and analytical attributes:

Identifiers: order_number, order_date, customer_name, channel, product_name

Financials: quantity, unit_price, revenue, cost, total_cost, profit, profit_margin_pct

Calendar: order_month_name, order_month_num, order_month

Geography: state (code), state_name, us_region, latitude, longitude

Planning: budget (2017), budget vs. actual variance

Data Processing & Feature Engineering

Data Quality Assessment

- ✅ **No missing values** in transaction records
- ✅ **No duplicate rows** detected
- ✅ **Consistent schema** across all data sources
- ✅ **Date integrity** verified for temporal analysis

Pre-Processing Steps

1. **Header standardization** – Set proper headers for State-Region mapping table
2. **Multi-source consolidation** – Merged Sales Orders with Customers, Products, Regions, State-Region, and Budget tables using key relationships
3. **Redundancy elimination** – Dropped duplicate and unnecessary columns post-join
4. **Column naming** – Standardized all column names to lowercase with underscores for consistency
5. **Column selection** – Retained only analytical and dimensional columns relevant to analysis
6. **Alias creation** – Renamed cryptic columns to sensible names (e.g., "Line Total" → "revenue", "Total Unit Cost" → "total_cost")
7. **Feature engineering** – Calculated derived columns:
   - profit = revenue – total_cost
   - profit_margin_pct = (profit / revenue) × 100

Dataset Readiness Verification

All preprocessing steps validated with zero data loss. Final dataset ready for statistical analysis, visualization, and insight generation.

Exploratory Data Analysis: Detailed Findings

1. Monthly Sales Trend Over Time (2014–2018)

Key Findings:

- Sales demonstrate a **consistent annual cycle** ranging from $24M to $26M per month
- **Seasonal peaks** occur in late spring and early summer (May–June), suggesting promotional or demand-driven patterns
- **Annual low point** in January, indicating post-holiday demand saturation
- **Notable outlier** in early 2017 shows sharp revenue drop, warranting investigation into external factors (market disruption, operational issue, or data anomaly)
- **Cyclical predictability** enables proactive inventory and staffing planning

Business Implication: Strong seasonality enables targeted recovery campaigns in low-performing periods (April) and amplified promotions during natural peaks.

---

2. Product Performance: Revenue Leaders

Top Product Revenue Drivers:

- **Products 26 & 25** dominate, collectively driving ~25% of total sales
- **Mid-tier performers** (Products 5, 13, 14, 15) show similar revenue (~$15–18M range each)
- **Bottom cluster** (Products 1, 2, 3, 4) generate significantly lower revenue, indicating either low demand, poor positioning, or margin constraints

Strategic Recommendation:

- Double down on revenue leaders (26 & 25) for inventory optimization
- Grow mid-tier performers through targeted marketing and bundle strategies
- Re-evaluate low performers: consider pricing optimization, repositioning, or phase-out if margin deficient

---

3. Product Profitability: Margin Leaders

Top 10 Products by Average Profit Margin:

- **Premium tier** (Products 18 & 28): $8.1K–$8.4K average profit per unit
- **Strong performers** (Products 5, 11, 12, 26): $7.5K–$7.8K range
- **Entry-level** (Products 1, 4, 16, 21): ~$7.3K range
- **Consistent threshold:** All top 10 products exceed $7.3K margin, indicating robust pricing control and cost management

Business Insight: Margin distribution suggests strong price-to-cost alignment across the portfolio. Focus attention on bottom performers outside top 10 for margin recovery.

---

4. Sales Distribution by Channel

Channel Performance Breakdown:

- **Wholesale:** 54.1% of total sales (largest volume contributor)
- **Distributor:** 31.3% of total sales (significant secondary channel)
- **Export:** 14.6% of total sales (smaller, specialized channel)

Margin Insight: While Wholesale dominates volume, Export channels typically achieve superior average margins (~38%), suggesting:

- Opportunity to incentivize Export partnerships for margin improvement
- Volume-based pricing strategies appropriate for Wholesale to maintain volume
- Balanced channel strategy required for revenue vs. profitability optimization

5. Geographic Sales Performance: State Analysis

Top Sales States (Revenue):

- **California:** Highest total sales ($230M+), 7,600+ orders
- **Texas, Florida, Illinois:** Significant sales contributors, each $150M+
- **Other states:** Moderate to low sales, showing wide variation

Geographic Pattern: Higher sales concentrated in larger states and coastal regions, suggesting:

- Population and economic activity correlation
- Existing market penetration in major metros
- Expansion opportunity in underserved secondary markets

Risk Assessment: Geographic concentration in California (>7% of total orders) represents moderate concentration risk requiring diversification.

6. Regional Performance: US Regions

Region Revenue Hierarchy:

- **West:** Highest sales region, strong market influence
- **South:** Major sales contributor, key market area
- **Midwest:** Steady sales performance, moderate market size
- **Northeast:** Lowest sales, suggests untapped potential or underinvestment

Strategic Implication:

- Replicate West region's success factors in other regions
- Northeast underperformance warrants deeper market analysis (competitive landscape, pricing, positioning)
- South and Midwest show growth potential through increased marketing and sales effort

7. Average Order Value (AOV) Distribution

Distribution Characteristics:

- **Low AOV frequency:** Majority of orders fall in lower AOV bands
- **Right-skewed distribution:** Long tail of high-value orders, but infrequent
- **Multiple clusters:** Distinct order value groupings suggest different customer segments (SMB vs. enterprise)
- **High-value scarcity:** Premium order values (>$50K) represent small transaction count but meaningful revenue

Implication: Segmented sales strategy appropriate—volume plays for small orders, relationship management for high-value transactions.

---

8. Top State Performance: Revenue vs. Order Count

State Performance Matrix:

- **California leads:** Highest in both revenue ($230M+) and order count (7,600+)
- **High performers** (Illinois, Florida, Texas): Consistently ranked in top revenue and top order volume
- **Strong correlation:** Revenue and order count show positive relationship, indicating market penetration drives both metrics
- **Other top states:** Lower aggregate contribution, suggesting secondary market positions

Insight: Market penetration (order count) correlates with revenue generation—expansion in new states creates future revenue uplift.

---

9. Customer Segmentation: Revenue Analysis

Top 10 vs. Bottom 10 Customers:

- **Aibox Company:** Top revenue generator, significantly outperforms peer group
- **State Ltd & peers:** Substantial but lower contribution than Aibox
- **Bottom 10 cluster:** Generate $4–5M each in revenue (5-10x lower than top tier)
- **Revenue concentration risk:** Top customers drive disproportionate revenue share, requiring account risk management and diversification efforts

Strategic Risk: Customer concentration creates vulnerability to account loss. Recommended: account protection strategies, mid-market growth programs, and new customer acquisition targets.

---

10. Customer Segmentation: Revenue vs. Profit Margin Analysis

Margin Patterns by Customer Revenue Tier:

- **$10M+ customers with 35–40% margins:** Uniform margins indicate strong pricing power and cost control with large accounts
- **$10M+ customers with <36% margins:** Discount hotspots—recommend margin recovery through pricing review and terms renegotiation
- **$6–9M customers with >40% margins:** High-value mid-market segment, ideal candidates for targeted upsell and expansion
- **Smaller accounts:** Typically maintain 30–35% margins, suggesting volume discounting or operational efficiency variance

Margin Recovery Opportunity: 5–10% of large accounts showing suppressed margins (<36%) represent immediate margin improvement opportunity through strategic account reviews.

---

11. Correlation Analysis: Key Metric Relationships

Critical Correlations:

Relationship	Correlation	Interpretation
Unit Price ↔ Cost	0.94	Very strong; pricing strategy directly cascades to cost (procurement strategy or pass-through model)
Unit Price ↔ Revenue	0.91	Very strong; pricing power is primary revenue driver
Unit Price ↔ Profit	0.79	Strong; pricing controls profitability
Revenue ↔ Profit	0.87	Strong; revenue expansion directly increases profitability
Cost ↔ Profit	0.58	Moderate; margin improvement requires cost focus beyond volume
Quantity ↔ Financial Metrics	≤0.34	Weak; volume plays secondary role to pricing strategy

Critical Insight: Unit pricing is the primary lever for revenue and profit improvement. Volume quantity has minimal correlation with profitability—strategic emphasis should focus on:

1. Price optimization for high-demand products
2. Cost structure improvement through procurement efficiency
3. High-margin channel incentives (Export) vs. volume channels (Wholesale)

Key Business Insights Summary

1. **Pronounced Seasonality:** January revenues average $124M, dipping 24% to $95M in April. Seasonal patterns enable proactive operational planning and promotional strategy calibration.
2. **SKU Concentration:** Products 26 & 25 together represent ~25% of total sales. Portfolio diversification and mid-tier product growth strategies critical for risk mitigation.
3. **Channel Trade-Off:** Wholesale captures 54% of volume; Export leads with ~38% average margin. Balanced channel strategy required for revenue-profit optimization.
4. **Geographic Dominance:** California alone logs 7,600 orders ($230M). Geographic expansion in secondary markets (Northeast, non-coastal states) represents growth

opportunity and concentration risk mitigation.

5. **Top Customer Concentration:** Aibox Company and State Ltd are the most valuable customers in terms of revenue. Account protection and mid-market growth programs essential for customer portfolio diversification.
6. **Margin Drivers:** Unit pricing (0.94 correlation with cost, 0.91 with revenue) is the primary lever. Quantity plays secondary role (0.34 correlation). Pricing strategy and cost management take precedence over volume.
7. **Customer Segmentation Opportunity:** $6–9M mid-market customers with >40% margins represent high-value upsell candidates. Large accounts with <36% margins warrant pricing review.

Strategic Recommendations

1. Seasonal Promotions & Demand Planning

Action: Launch recovery campaigns in April and amplify January offers to smooth revenue swings.

- **Target:** Achieve 5–8% revenue uplift in low-period months through targeted promotions
- **Impact:** Smoothed revenue enables optimized warehouse and staffing capacity planning
- **Owner:** Sales & Marketing

2. SKU Optimization & Portfolio Management

Action: Double down on top products (26 & 25) and re-evaluate pricing or phase out low-margin SKUs.

- **Analysis:** Review bottom quartile products for margin recovery or discontinuation
- **Strategy:** Allocate marketing budget proportional to margin contribution, not revenue volume
- **Target:** Shift product mix toward higher-margin offerings, improving overall portfolio margin by 2–3%
- **Owner:** Product Management & Marketing

3. Channel Expansion & Profitability Focus

Action: Incentivize Export partnerships for high margins and introduce volume deals in Wholesale.

- **Export strategy:** Expand export channel partnerships with margin-based pricing
- **Wholesale strategy:** Volume discounts acceptable given 54% channel share; negotiate cost reductions
- **Target:** Grow Export channel from 14.6% to 18–20% of total sales within 12 months
- **Owner:** Sales & Channel Management

4. Regional Investment & Market Expansion

Action: Replicate California's success in other regions and boost marketing in Northeast & Midwest.

- **Market analysis:** Conduct competitive and positioning analysis in underperforming regions
- **Northeast focus:** Investigate barriers to growth; consider targeted sales effort or market repositioning
- **Secondary markets:** Identify high-potential secondary cities and develop expansion playbooks
- **Target:** Increase Northeast region sales from current baseline by 15–20% within 18 months
- **Owner:** Regional Sales Management

5. Margin Monitoring & Cost Optimization

Action: Flag orders below 80% margin target and analyze cost drivers to uplift underperforming segments.

- **Pricing review:** Conduct quarterly account-level margin analysis; flag large accounts with suppressed margins
- **Cost structure:** Partner with procurement to analyze unit cost drivers and opportunities for reduction
- **Target:** Recover 1–2% margin improvement in accounts currently below 35% margin threshold
- **Owner:** Finance & Operations

6. Customer Portfolio Risk Management

Action: Implement account protection strategies for top customers and develop mid-market growth programs.

- **Account management:** Assign dedicated resources to Aibox Company and State Ltd for account retention
- **Mid-market growth:** Target $6–9M customers with >40% margins for expansion/upsell (high-value segment)
- **New customer acquisition:** Develop small/SMB acquisition programs to diversify revenue base
- **Target:** Reduce customer concentration from current peak by acquiring 15–20 new mid-market customers within 12 months
- **Owner:** Sales & Account Management

Dashboard Solution

Three-Page Interactive Power BI Dashboard

Page 1: Performance Summary

- KPI cards: Total Revenue, Total Profit, Average Margin %, Customer Count
- Monthly sales trend with seasonal pattern visualization
- Sales by channel (Wholesale, Distributor, Export) with margin breakdown

- Top 10 products by revenue and profitability
- Regional performance heatmap

Page 2: Customer Segmentation

- Revenue vs. profit margin scatter plot (customer bubble chart)
- Top 10 and bottom 10 customer lists with metrics
- Customer segmentation matrix (size, margin, growth potential)
- Customer acquisition trend over time
- Account health scorecard (concentration risk, margin trend)

Page 3: Revenue Scenarios

- State-level choropleth map with revenue and order count overlays
- Regional drill-down analysis
- Product performance matrix (revenue vs. margin)
- Channel profitability comparison
- Budget vs. actual variance analysis by product/region/channel

Self-Service Capabilities:

- Slicers for time period, channel, product, region, customer segment
- Drill-through capabilities for detailed transaction analysis
- Dynamic calculations for margin and variance metrics
- Real-time refresh capability for quarterly budget updates

Conclusion

This end-to-end EDA and interactive Power BI dashboard project successfully surfaced seasonality, SKU concentration, channel profitability, regional performance, and customer segmentation insights from Acme Co.'s 5-year sales history.

Deliverables completed:
✓ Comprehensive exploratory data analysis with 11 analytical dimensions
✓ Data quality validation and feature engineering
✓ Statistical correlation and segmentation analysis
✓ Interactive Power BI dashboard for stakeholder self-service
✓ Strategic recommendations aligned with quantified insights

Business Outcomes Enabled:
✓ Informed sales policies and operational planning (warehouse and staffing aligned with seasonal trends)
✓ Risk mitigation through geographic and customer diversification strategies
✓ Margin improvement through pricing strategy and cost optimization focus
✓ Growth acceleration through channel expansion and market prioritization

Path Forward:
Stakeholders can now confidently leverage the dashboard to:

- Monitor performance against KPIs and recommendations
- Conduct ad-hoc analysis by dimension (time, product, region, channel, customer)
- Onboard new datasets (quarterly updates, new regions, product lines) for ongoing strategic insight generation

- Measure progress against recommendation implementation targets

The foundation is established for data-driven decision-making, sustainable revenue growth, and reduced operational risk through informed resource allocation.

Technical Appendix

Tools & Technologies:

- **Data Processing:** Python (Pandas, NumPy), Google Colab
- **Visualization:** Matplotlib, Seaborn, Power BI
- **Data Sources:** Multi-sheet Excel workbook
- **Analysis Methods:** Exploratory Data Analysis, Univariate & Bivariate Analysis, Correlation Analysis, Customer Segmentation, Trend & Seasonality Decomposition

Data Lineage:
Raw data from six Excel sheets → Consolidated Python environment → Processed analysis dataset (64,104 records, 21 attributes) → Power BI for visualization and stakeholder engagement

Quality Assurance:

- Schema validation across all source systems
- Missing data verification (zero nulls in transaction records)
- Duplicate record detection (zero duplicates)
- Date integrity and temporal sequence verification
- Statistical sanity checks on derived metrics (margin %, unit costs)

Report Generated: November 28, 2025
Data Period: January 2014 – February 2018
Organization: Acme Co.
Project Scope: USA Regional Sales Analysis (5-year history)