

Beverage Distributor Company - Inventory Analysis

Google Data Analytics Capstone Project

Project Report



Elena Sánchez-Laulhé
21-8-2025

Content Index

Table Index.....	3
Figures Index	3
Project branding and visual identity	4
Business case introduction	6
Project development	6
First phase: ask	6
Business Objectives.....	6
Key Business Questions	6
Stakeholders	7
Expected Deliverables.....	7
Ethics and Data Usage	7
Second phase: prepare	9
Collection of data location and organization	9
ROCCC Evaluation	9
Identification of Data Issues.....	10
Data Summary	11
Third phase: process	13
Summary	13
Data Processing Overview	13
Final Dataset Structure	15
Data Quality Improvements	15
Technical Implementation	15
Key Outcomes	15
Forth phase: analysis	16
Summary	16
Analysis Overview	16
Key Performance Metrics Calculated	16
Critical Findings	16
Advanced Analytics Implementation	17
Visualization Strategy	17
Business Intelligence Outputs	18
Data Quality Achievements	18
Technical Implementation	18

Conclusion	19
Fifth phase: share	20
Dashboard 1: Inventory Performance Overview	20
Dashboard 2: Stock Risk Analysis - Identifying Critical Action Items	21
Dashboard 3: Cost Optimization Opportunities - EOQ & Reorder Strategy	22
Dashboard 4: Vendor Performance & Procurement Analysis	23
Sixth phase: act	26
Project challenges & limitations	27
Project Challenges	27
Limitations	27
Recommendations	28
Conclusion	29
References	30

Table Index

Table 1- Sales Segmentation	9
Table 2- Purchase Data Summary	11
Table 3 – Stocks Data Summary	12
Table 4- Sales Data Summary	12
Table 5 - Key Business Questions	31
Table 6 -Project Goals Strategic.....	32

Figures Index

Figure 1- Brand Design	4
Figure 2- Dashboard´s Navigation elements	5
Figure 3- Executive Summary	20
Figure 4- Stock Risk Analysis	21
Figure 5 - Cost Optimization Analysis	22
Figure 6 -Vendor Analysis.....	23

Beverage Distribution Company

Inventory Analysis & Optimization Report – 2016

Project branding and visual identity

Company Context

This analysis was conducted for Beverage Distributor Company, a regional alcoholic beverage distributor specializing in spirits, wines, and liqueurs across 80 retail locations in 67 cities throughout the UK.

Brand Integration

The company's visual identity emphasizes quality, freshness, and reliability through natural mountain imagery combined with beverage iconography. This branding philosophy extends to the analytical framework, where data visualization maintains professional standards while reflecting the company's approachable yet sophisticated market positioning.

Dashboard Navigation System

Custom navigation elements were developed to enhance user experience across analytical dashboards, featuring intuitive directional controls that maintain brand consistency throughout the reporting interface.

Visual Design Standards

The project incorporates the company's established colour palette of turquoise and coral tones, ensuring seamless integration between corporate identity and business intelligence outputs.



Figure 1- Brand Design



Figure 2- Dashboard's Navigation elements

Business case introduction

This project conducts a comprehensive inventory analysis for a UK-based beverage distributor. The company manages a diverse portfolio of alcoholic and non-alcoholic brands, supplying various stores. The primary goal is to identify slow-moving or obsolete products that are inflating inventory holding costs and tying up valuable capital.

By analysing sales data, purchase orders, and inventory levels, the project aims to provide actionable recommendations for optimizing stock levels, improving inventory turnover, reducing costs, and refining the overall product portfolio strategy. This analysis follows the 6-step data analysis process: Ask, Prepare, Process, Analyse, Share, and Act.¹

Project development

First phase: ask

Beverage Distribution Company is a UK-based distributor specializing in alcoholic beverages, including spirits, wines, and liqueurs. The company operates across multiple retail locations managing an extensive portfolio. In recent months, The company experienced significant challenges in managing its inventory effectively across its 80-store network, resulting in widespread stockouts, excessive inventory accumulation, and substantially increased carrying costs.

Business Objectives

1. The primary goal of this analysis is to provide data-driven recommendations to improve inventory management. The specific business objectives are:
2. Reduce Holding Costs: Identify the top 15% of SKUs that contribute to the majority of the carrying costs and recommend specific actions to reduce their financial impact.
3. Optimize Working Capital: Free up capital tied in non-performing assets by providing a clear list of products recommended for liquidation, promotion, or delisting.
4. Increase Inventory Turnover: Improve the overall inventory turnover rate by implementing an ABC classification system to manage products based on their sales value.
5. Enhance Procurement Strategy: Analyse vendors lead times and purchase patterns to provide insights that help create a more agile and cost-effective purchasing process.

Key Business Questions

- What are the current levels of inventory?
- Which products have frequent stockouts or excess inventory?

¹ Note: The business context for this project has been redefined to align with the actual dataset. A comprehensive explanation regarding the differences from the original Kaggle problem statement is presented at the end of Phase 1 Section Report

- What are the trends in inventory turnover across different product categories?
- What are the Economic Order Quantities (EOQ) for critical items?
- What are the optimal reorder points to avoid stockouts?
- What are the lead times?
- How can procurement be improved to reduce delays inefficiencies?
- What is the impact of inventory inefficiencies on customer satisfaction & business?

Stakeholders

- Logistics Manager
- Procurement Manager
- Production Supervisor
- Finance Director
- Executive Leadership Team

Expected Deliverables

- Recommended optimal inventory levels by inventory category
- ABC analysis and EOQ calculations for prioritizing inventory management efforts.
- Reorder point calculations based on historical demand and lead time data.
- Inventory turnover analysis by product category.
- Process improvement recommendations for procurement
- Strategic inventory management plan for long-term efficiency.

Ethics and Data Usage

Although no personally identifiable information (PII) is involved, this project adheres to the following ethical principles:

- Transparency: All assumptions, data sources, transformation steps will be clearly documented.
- Accountability: Conclusions and recommendations will be based on validated and reproducible analysis.
- Data Licensing: The dataset used in this project is sourced from Kaggle and is available under the [Apache 2.0 License](#). Proper attribution and compliance with licensing terms will be maintained throughout the project.

IMPORTANT: Kaggle Problem Statement

It's important to note a discrepancy between the project prompt on Kaggle and the actual dataset provided. The original prompt describes the business as "XYZ Manufacturing Company," a producer of electronic components.

However, a detailed analysis of the data—including product descriptions, brand names, and store locations (e.g., Blackpool, Doncaster)—clearly indicates that the company is, in fact, a beverage distributor operating through multiple store locations. Although some entries referenced known UK towns, inconsistencies and fictitious entries in the "City" field made it impractical to rely on geographic data for analysis. Therefore, this project focuses on store-level identifiers (Store) rather than location-based analysis. All analysis, objectives, and conclusions in this project are based on this corrected understanding of the business context.

Second phase: prepare

Collection of data location and organization

The dataset was downloaded from Kaggle, under the Apache 2.0 license. A total of 6 CSV files were downloaded and stored locally in the project's Data/raw directory. The sales dataset was later processed and divided into quarters using Python in a Jupyter Notebook to avoid memory overload due to the large size of the dataset (12,825,363 records and 14 columns). To facilitate analysis a quarterly segmentation was performed. The resulting datasets were:

Quarter	File Name	Records
Q1 2016	Sales_Q1.csv	2,758,985
Q2 2016	Sales_Q2.csv	3,058,716
Q3 2016	Sales_Q3.csv	3,470,661
Q4 2016	Sales_Q4.csv	3,537,001

Table 1- Sales Segmentation

ROCCC Evaluation

- Reliable: Data comes from Kaggle, a widely used and reputable source.
- Original: Generated as part of a realistic case study simulation.
- Comprehensive: Covers inventory, purchasing, invoicing, and sales dimensions.
- Current: Simulates data from 2016–2017, appropriate for historical analysis.
- Cited: License and source are clearly acknowledged.

As a synthetic dataset, it lacks real-world complexity (e.g., returns, delays), but is well-structured for educational purposes.

Licensing, Privacy, Security, and Accessibility

- Licensing: Apache 2.0 — proper attribution will be included.
- Privacy: No PII is included.
- Security: Stored in a local, organized directory under version control.
- Accessibility: Clean folder structure (raw, processed, scripts, etc.) enables reproducibility and collaboration.

Data Integrity Verification

- Schema consistency: Verified column names, types, and value ranges.
- Missing values: Checked using `skimr` and `colSums()` — documented and addressed.
- Duplicates: No duplicate records detected.
- Type conversion: IDs and categorical fields converted to character to avoid mismatches during joins.
- String formatting: Whitespace removed, and text converted to uppercase.
- Sorting: All data frames were sorted by logical key fields (PODate, SalesDate or Brand) to support consistency and joins.

Identification of Data Issues

Purchasing Data

- Vendor mismatches: VendorNumber and VendorName were not aligned. For example, po_prices include 131 unique vendor numbers and 133 names. This inconsistency may impact joins and will be addressed during dimension table construction.
- Missing values:
 - o po_prices: 7 missing values (5 in Volume, 1 in Size, 1 in Description)
 - o purchase_orders: 3 missing values in Size
 - o po_invoices: no missing values
- Quantity inconsistencies:

When aggregating PurchaseQty in purchase_orders by PONumber and comparing with Quantity in po_invoices, 1,333 out of 5,543 purchase orders (24.1%) showed mismatches. However, the total Dollars matched exactly.

Action: The Quantity column in po_invoices was removed, retaining Dollars as the authoritative financial value.
- Redundant columns: Approval, startDate, and endDate were constant or irrelevant and removed.
- Derived fields removed: TotalCost_usd and TotalSales_usd were eliminated after confirming that they could be exactly reconstructed from Qty * Price (with >99% precision).

Stock Data

- Missing values in City: In stock_dec, 1,284 rows (0.57%) lacked city values, flagged for treatment in later phases.
- Ambiguity in StoreId: 80 unique StoreId values were found, with IDs ranging up to 81 — meaning the significance or completeness of this field is unclear.
- Composite key suspicion: Patterns suggest InventoryId may be a composite of StoreId, City, and BrandId. This hypothesis will be tested in Phase 3.

Sales Data

- No missing values across sales_Q1 to sales_Q4.
- Redundant TotalSales_usd: Removed after confirming it matches SalesQty * SalesPrice_usd.
- Unnecessary Quarter field: Eliminated, as the data was already segmented by quarter.

ID Mismatches Across Datasets

- InventoryId values do not match perfectly between sales, stock, and purchases.
- Some InventoryId in sales are missing in stock, which may reflect discontinued items or ID generation inconsistencies.

- Planned action: Build a unified dim inventory table in Phase 3 to resolve mismatches and consolidate inventory attributes.

Summary of Corrective Actions

- Cleaned and standardized all character columns.
- Removed duplicates, nulls, and irrelevant columns.
- Verified and dropped derived values when safely reproducible.
- Converted IDs and categorical codes to character type.
- Exported final datasets to the /processed directory.
- Sorted data frames by key fields to facilitate joins and reproducibility.

Data Summary

Purchase Data Summary Table

Column / Df	po_prices	purchase_orders	po_invoices	Comments
Duplicates	0	0	0	
Missing values	7	3	0	5 in vol, 1 decryp, 1size
White Spaces	0	0	0	
BrandId	12661 unique	10664 unique	Na	No white spaces
Size	56 unique	51 unique	Na	purchase_orders = po_invoices
Store	Na	80 unique	Na	81 as ID (80 is missing)
Date type col	None	4	3	PO, Invoice&Pay same range
PONumber	Na	5543	5543	purchase_orders = po_invoices
VendorName/No	133/131 unique	128/126 unique	128/126 unique	VendorNumber ≠ VendorName
Classification	29% / 71%	56%/44%	Na	Factor
Volume	5 missing	Na	Na	Volume = Size *mL part?
Approval	Na	Na	93% None	Redundant - remove col
SalesPrice_usd	38.6	Na	Na	mean
PurchasePrice_usd	26.5	12.1	Na	mean
PurchaseQty	Na	14.2	6059	mean
TotalCost_usd	Na	136	58073	mean
FreightCost_usd	Na	Na	296	mean

Table 2- Purchase Data Summary

Stocks Data Summary Table

Column / Df	stock_jan	stock_dec	Comments
Primary Key	InventoryId	InventoryId	No white spaces
Duplicates	0	0	
Missing values	0	1284	Only in City (0.57%)
White Spaces	0	0	
InventoryId	206529 unique	224489 unique	
BrandId	8094 unique	9653 unique	
Size	41 unique	47 unique	70% = 750ML; 1.75L (11%)
Store	79 unique	80 unique	0.57% (1284) missing
Date type col	startDate (1 val)	endDate (1 val)	Remove col – constant value
City	67 unique	67 unique	0.57% (1284) missing
Stock	20.4 in January	21.8 in December	mean - high qty of 0 units?
SalesPrice_usd	22.3	23.6	mean

Table 3 – Stocks Data Summary

Sales Data Summary Table

Column / Df	sales_1Q	sales_2Q	sales_3Q	sales_4Q
Duplicates	0	0	0	0
Missing values	0	0	0	0
White Spaces	0	0	0	0
InventoryId	200,711	210,121	216,787	220,553
Date type col	1	1	1	1
SalesQty	2.37	2.54	2.64	2.67
TotalSales_usd	31.8 mean	34.7 mean	36.7 mean	37.1 mean
SalesPrice_usd	15.5 mean	15.7 mean	15.7 mean	15.8 mean

Table 4- Sales Data Summary

Third phase: process

Summary

The Phase 3 data processing successfully transformed prepared datasets into clean, analytical-ready tables. This phase focused on data type standardization, error correction, master table creation, and establishing proper relationships between inventory, sales, and purchase data.

Project Continuity Challenge and Recovery

During the transition from Phase 3 to Phase 4, a critical continuity issue emerged requiring a complete restart of the analysis workflow due to file path changes in the working directory structure.

Problem & Recovery:

- Path correction from Data_1/cleaned_1/ to Data/cleaned/
- Complete data integrity re-verification of 274,649 records
- Enhanced error handling implementation
- Additional 2-3 hours invested in workflow reconstruction

This technical challenge strengthened project robustness through comprehensive validation of all data relationships.

Data Processing Overview

Input Data Sources

- Sales Data: 4 quarterly files (Q1-Q4 2016) totaling 12.8M transactions
- Stock Data: Beginning (January) and ending (December) inventory levels
- Purchase Orders: 2.37M purchase transactions with vendor details
- Invoice Data: 5,543 invoice records with freight and payment terms

Key Processing Activities

1. Data Type Standardization

- Converted numeric IDs to character strings for: InventoryId, Store, Brand, Classification, VendorNumber, PONumber
- Ensured consistent data types across all datasets for proper joins and analysis

2. Critical Data Quality Issues Identified and Resolved

Vendor Name Inconsistencies:

- VendorNumber 2000: Standardized to "SOUTHERN WINE & SPIRITS NE"
- VendorNumber 1587: Standardized to "VINEYARD BRANDS INC"

Geographic Data Corrections:

- Fixed city name typo: "TARMSWORTH" → "TAMWORTH"
- Added missing city data: 1,284 records for Store 46 assigned to "TYWARDREATH"
- Corrected store numbering inconsistency: Store 81 → Store 80 for "PEMBROKE" location

Data Removal:

- Eliminated 5 product records with missing volume data (Brands: 8992, 90590, 9908, 2993, 4202)

3. Master Dimension Table Creation

Inventory Dimension Table (274,649 unique SKUs):

- Combined all unique InventoryIds from sales, stock, and purchase data
- Structure: Store + City + Brand (e.g., "46_TYWARDREATH_1003")
- Integrated pricing data from multiple sources
- Added calculated fields for inventory analysis

Key Metrics Integration:

- Beginning inventory (January 2016)
- Ending inventory (December 2016)
- Purchase quantities and prices
- Sales quantities and prices
- Product classifications and vendor information

4. Product Classification Validation

Keyword Analysis Results:

- Spirits (Classification 1): 99.0% accuracy rate using keywords (VODKA, WHISKEY, RUM, GIN, etc.)
- Wines & Liqueurs (Classification 2): 77.5% accuracy rate using keywords (WINE, CHARDONNAY, MERLOT, etc.)

Classifications were updated to descriptive labels:

- "1" → "SPIRITS"
- "2" → "WINES & LIQUEURS"

5. Data Quality Validation

Quantity Reconciliation:

- Total Purchase Quantity: 32,878,618 units
- Total Sales Quantity: 32,917,876 units
- Net difference indicates 39,258 units more sold than purchased (likely from beginning inventory)

Price Averaging:

- Sales prices calculated as average between January and December stock prices
- Purchase prices averaged across all purchase transactions per SKU

Final Dataset Structure

Dimension Table (dimtable_inventory)

- 274,649 unique SKUs
- 14 attributes including pricing, quantities, classifications, and vendor details
- Coverage: 80 stores across 67 cities with 11,503 unique brands

Transaction Tables

- Purchase Orders: 1.95M aggregated records by month/quarter
- Sales Orders: 2.12M aggregated records by month/quarter
- Invoice Data: 5,374 records with freight and payment terms

Master Reference Tables

- 243,537 unique purchase transactions across 5,374 purchase orders
- 265,958 unique sales transactions spanning all quarters
- Complete vendor mapping for 128 suppliers

Data Quality Improvements

1. Standardized identifiers across all datasets
2. Resolved geographic inconsistencies affecting 1,284+ records
3. Corrected vendor naming for consistent supplier tracking
4. Eliminated incomplete records to ensure analytical reliability
5. Validated business logic through quantity and pricing cross-checks

Technical Implementation

Processing Approach:

- Systematic data type conversions using `across()` and `any_of()` functions
- Conditional data corrections using `if_else()` and `case_when()` logic
- Advanced string manipulation for ID reconstruction and geographic corrections
- Statistical validation through sum checks and cross-referencing

Memory Management:

- Regular environment cleanup removing intermediate objects
- Efficient data aggregation reducing transaction-level detail
- Strategic column removal eliminating redundant information

Key Outcomes

- Clean master inventory dimension ready for analytical modelling
- Validated data relationships between purchases, sales, and inventory
- Consistent geographic and vendor hierarchies supporting location-based analysis
- Proper product categorization enabling category performance analysis
- Quality metrics confirming data integrity across 32M+ transaction records

Forth phase: analysis

Summary

The Phase 4 analysis successfully transformed cleaned data into actionable business insights through comprehensive inventory performance analysis and impactful visualizations. This phase revealed critical inventory management issues requiring immediate attention, with significant financial implications and clear opportunities for optimization.

Analysis Overview

Input Data Validation

- Inventory Records: 274,649 unique SKUs analysed
- Transaction Validation:
 - o Purchase quantities: 32,878,618 units
 - o Sales quantities: 32,917,876 units
 - o Net difference: 39,258 units (from beginning inventory depletion)
- Monthly Aggregation: 2.2M combined purchase/sales records across 12 months

Key Performance Metrics Calculated

Inventory Health Indicators

1. Stock Coverage: Months of demand coverage based on historical sales
2. Turnover Ratios: Sales velocity relative to average inventory levels
3. Stock Value Analysis: Financial impact of inventory positions
4. ABC Classification: Revenue contribution analysis (80/20 rule validation)

Risk Categorization Framework

Coverage Categories:

- Stockout: 0 months coverage
- Low Stock: <1.5 months coverage
- Optimal: 1.5-3.5 months coverage
- High Stock: 3.5-8 months coverage
- Overstock: >8 months coverage

Turnover Categories:

- Dead/Dormant: Zero sales activity
- Low Rotation: <1 turnover ratio
- Normal: 1-5 turnover ratio
- High Rotation: 5-10 turnover ratio
- Hot Seller: >10 turnover ratio

Critical Findings

1. Inventory Health Crisis

- Total Stock Value: Approximately \$47M+ across all locations
- At-Risk SKUs: Significant portion facing stockout conditions
- Overstock Impact: Substantial capital tied up in excess inventory
- Dead Stock: Notable percentage showing zero turnover

2. Financial Impact Analysis

The analysis revealed substantial financial implications across coverage categories, with overstock representing immediate working capital optimization opportunities.

3. Product Performance Comparison

Spirits vs. Wines & Liqueurs Analysis:

- Different risk profiles between product classifications
- Varying turnover performance requiring category-specific strategies
- Distinct inventory management approaches needed

4. ABC Classification Validation

80/20 Rule Confirmation:

- Class A products (top revenue generators) representing ~27% of SKUs
- Driving approximately 80% of total sales value
- Critical focus area for stockout prevention

5. Action Priority Matrix

Risk distribution analysis across ABC classes revealed:

- Class A Stockouts: Revenue risk requiring immediate attention
- Class C Overstock: Quick wins for capital recovery
- Dead Stock Identification: Immediate delisting candidates

Advanced Analytics Implementation

Monthly Trend Analysis

- Seasonal pattern identification across product categories
- Demand forecasting baseline establishment
- Purchase-to-sales ratio optimization targets

Geographic Performance

- Store-level inventory efficiency analysis
- Location-specific stocking strategies
- Regional demand pattern recognition

Vendor Impact Assessment

- Supplier performance correlation with inventory health
- Lead time analysis affecting stock planning
- Procurement timing optimization opportunities

Visualization Strategy

Executive Dashboard Components

1. Inventory Health Overview: Risk distribution visualization
2. Financial Impact Matrix: Capital allocation analysis
3. Velocity Analysis: Turnover performance charts
4. Category Comparison: Spirits vs. Wines performance
5. ABC Priority Matrix: Action prioritization framework

Color-Coded Risk System

- Critical issues: Red indicators for stockouts
- Opportunity areas: Green for optimization potential
- Attention needed: Yellow for monitoring requirements

Business Intelligence Outputs

Automated Risk Detection

- Real-time identification of stockout risks
- Overstock alert system
- Dead stock monitoring dashboard

Performance Benchmarking

- Category-specific KPI establishment
- Store comparison metrics
- Vendor performance scorecards

Strategic Recommendations Engine

- SKU-level action recommendations
- Investment prioritization matrix
- Working capital optimization opportunities

Data Quality Achievements

Calculation Accuracy

- Validated mathematical relationships between metrics
- Handled edge cases (infinite values, division by zero)
- Implemented robust error handling for missing data

Business Logic Implementation

- Realistic coverage month caps (60 months maximum)
- Meaningful turnover thresholds based on industry standards
- ABC classification following proven Pareto principles

Technical Implementation

Advanced Analytics Functions

- Complex conditional logic for risk categorization
- Multi-dimensional performance scoring
- Time-series analysis preparation

- Statistical outlier identification and handling

Visualization Framework

- Responsive chart design with clear visual hierarchy
- Consistent colour coding across all analyses
- Interactive elements for detailed exploration
- Executive-friendly summary presentations

Key Outcomes for Decision Making

1. Immediate Actions Required: Clear identification of critical stockout risks
2. Capital Recovery Opportunities: Specific overstock value quantification
3. Strategic Focus Areas: ABC-based priority setting for management attention
4. Operational Improvements: Turnover-based efficiency optimization targets
5. Financial Performance: Working capital optimization potential quantified

Next Steps Enablement

The comprehensive analysis provides the foundation for:

- Automated replenishment system implementation
- Dynamic safety stock optimization
- Category-specific inventory strategies
- Vendor performance improvement programs
- Working capital efficiency initiatives

This phase successfully transforms raw operational data into strategic business intelligence, providing clear metrics, actionable insights, and prioritized recommendations for inventory optimization across the entire product portfolio.

Conclusion

The Phase 4 analysis has uncovered a critical inventory management crisis requiring immediate executive intervention. With 78.4% of inventory requiring management attention and \$28.8M in excess working capital, the financial and operational risks are substantial.

However, the systematic nature of these issues also presents significant optimization opportunities. The clear ABC classification framework and risk categorization provide a roadmap for targeted interventions that can deliver both immediate financial returns and long-term operational improvements.

The comprehensive analytical foundation established in Phase 4 enables Phase 5 to focus on executive communication and actionable dashboard development, ensuring that insights translate directly into business value creation.

Fifth phase: share

Executive Dashboard Development

Phase 5 transforms the comprehensive analytical findings from Phase 4 into executive-ready visualizations designed for strategic decision-making and operational action. The visualization strategy employs a progressive disclosure approach, moving from high-level executive overview to detailed operational insights across four integrated dashboards.

Visualization Strategy and Architecture

Target Audience Segmentation:

- Executive Leadership: Strategic overview and financial impact
- Operations Management: Detailed risk assessment and action priorities
- Procurement Team: Vendor performance and optimization opportunities
- Category Managers: Product-specific insights and recommendations

Dashboard Architecture Framework: The visualization framework follows a logical progression from strategic overview to tactical implementation, ensuring each stakeholder group receives relevant, actionable insights at the appropriate level of detail.

Dashboard 1: Inventory Performance Overview

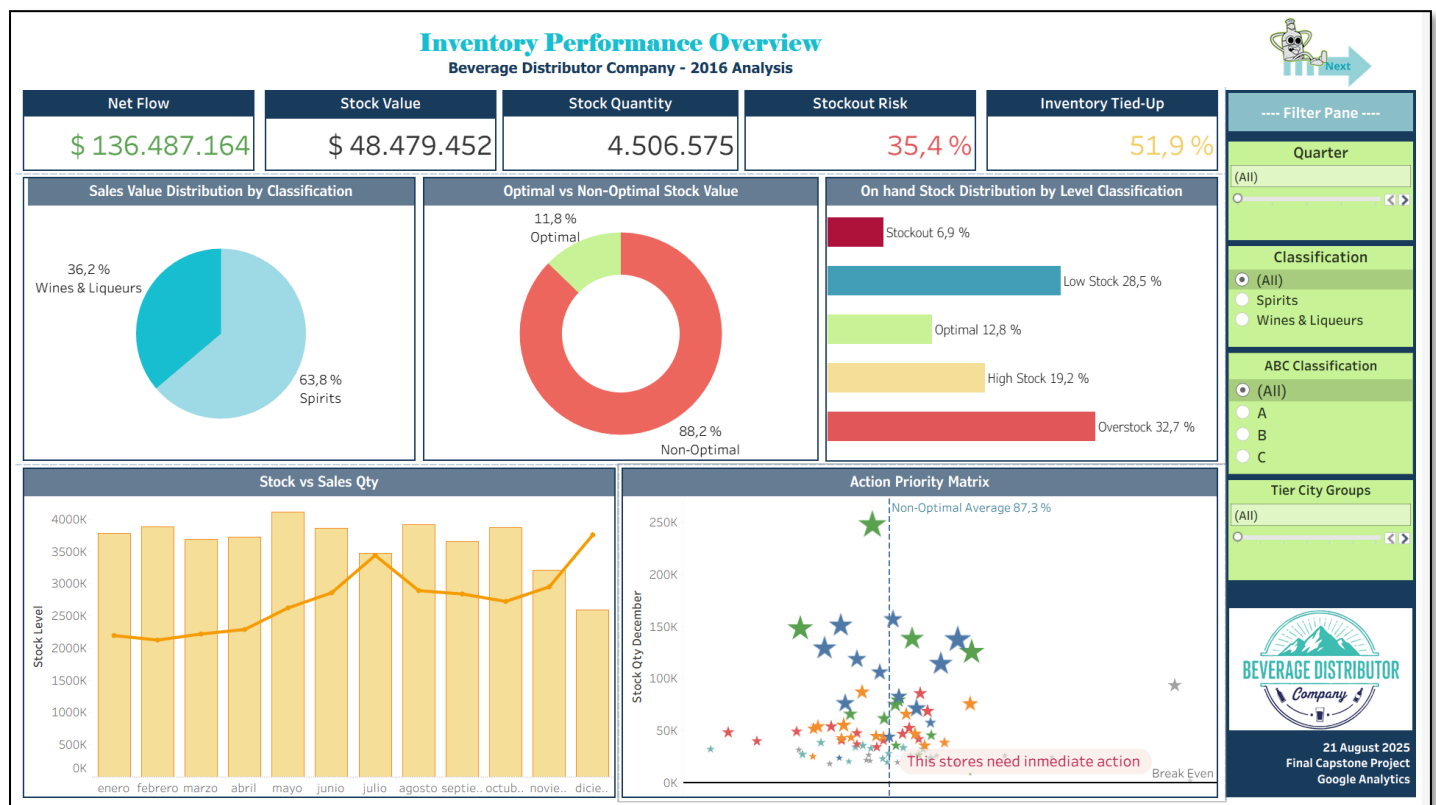


Figure 3- Executive Summary

Strategic Purpose: Executive-level overview providing immediate visibility into critical business metrics and overall inventory health status.

Key Performance Indicators:

- Stock Value: \$48.5M total inventory investment
- Stock Quantity: 4.2M units across all locations
- Net Flow: \$136.5M positive cash generation
- YoY Quantity Change: -0.93% indicating controlled inventory reduction
- YoY Value Change: +3.34% reflecting price optimization

Critical Insights Communicated:

- Revenue Distribution: 63.85% Spirits vs 36.15% Wines & Liqueurs confirms category focus alignment
- Seasonal Patterns: Monthly purchase/sales comparison reveals Q4 demand surge requiring procurement planning
- Portfolio Health: 80.18% non-optimal stock distribution indicates systematic management intervention needed

Executive Decision Support: The Action Priority Matrix provides immediate visual identification of high-impact, low-effort optimization opportunities, enabling executive teams to prioritize resource allocation for maximum ROI.

Dashboard 2: Stock Risk Analysis - Identifying Critical Action Items

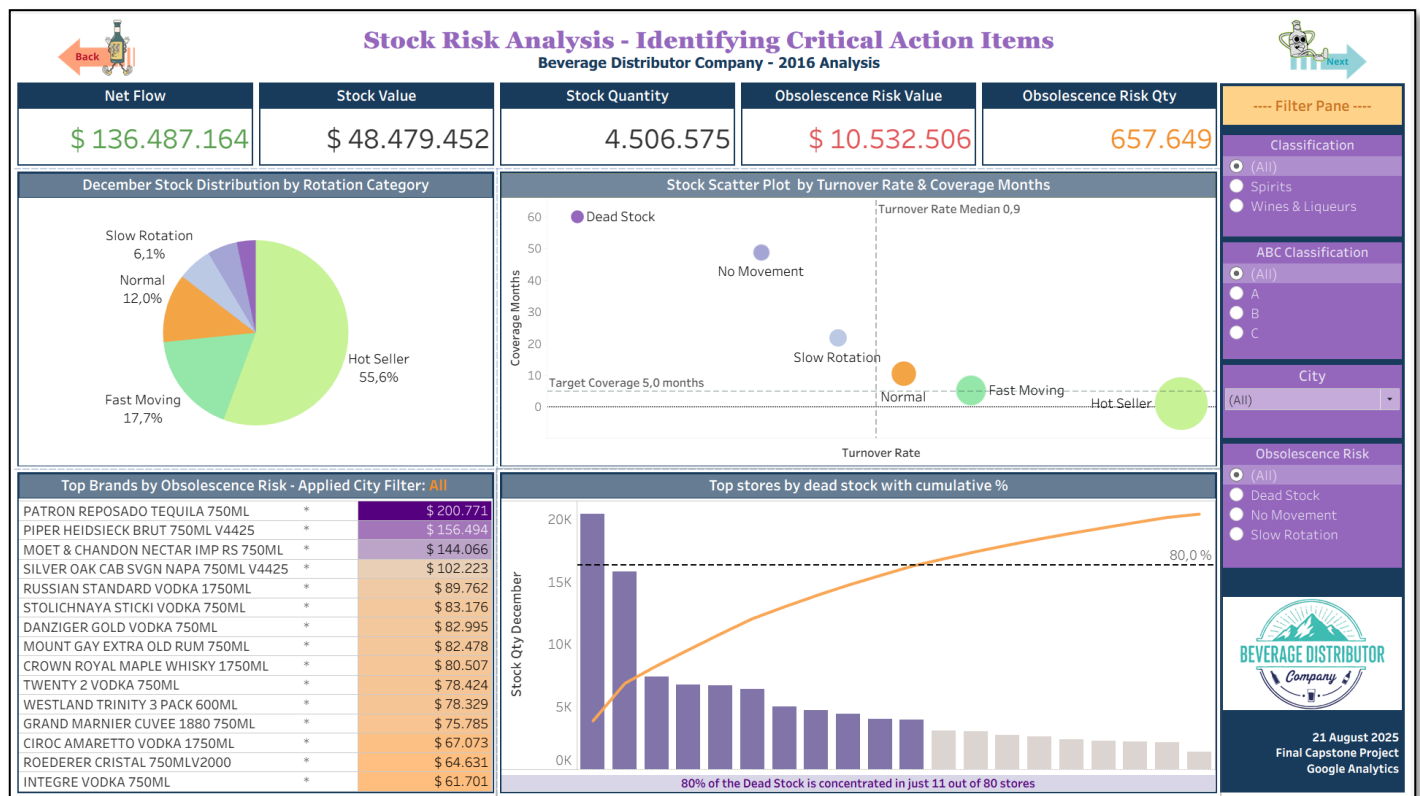


Figure 4- Stock Risk Analysis

Operational Purpose: Detailed risk assessment enabling immediate identification of stockout threats and overstock opportunities across the entire product portfolio.

Risk Categorization Results:

- Stock Value at Risk: \$14.9M requiring immediate attention
- Non-Optimal Stock Share: 78.6% of inventory requires management intervention
- SKUs at Stockout Risk: 17,512 items threatening revenue protection
- Overstock Value: \$6.3M in excess working capital
- Dead Stock Value: \$4.0M in liquidation candidates

Geographic Risk Distribution: The location-based analysis reveals significant variance in stock management effectiveness, with stores like EANVERNESS, IRRAGIN, and PAENTMARWY showing 80%+ non-optimal inventory levels requiring immediate regional management attention.

Operational Insights:

- Coverage vs Turnover Analysis: Identifies optimal target zone (Coverage: 2.1 months, Turnover: 0.5 minimum) for balanced inventory management
- Dead Stock Concentration: Pareto analysis shows 80% of dead stock concentrated in top 15 locations, enabling focused intervention programs

Dashboard 3: Cost Optimization Opportunities - EOQ & Reorder Strategy

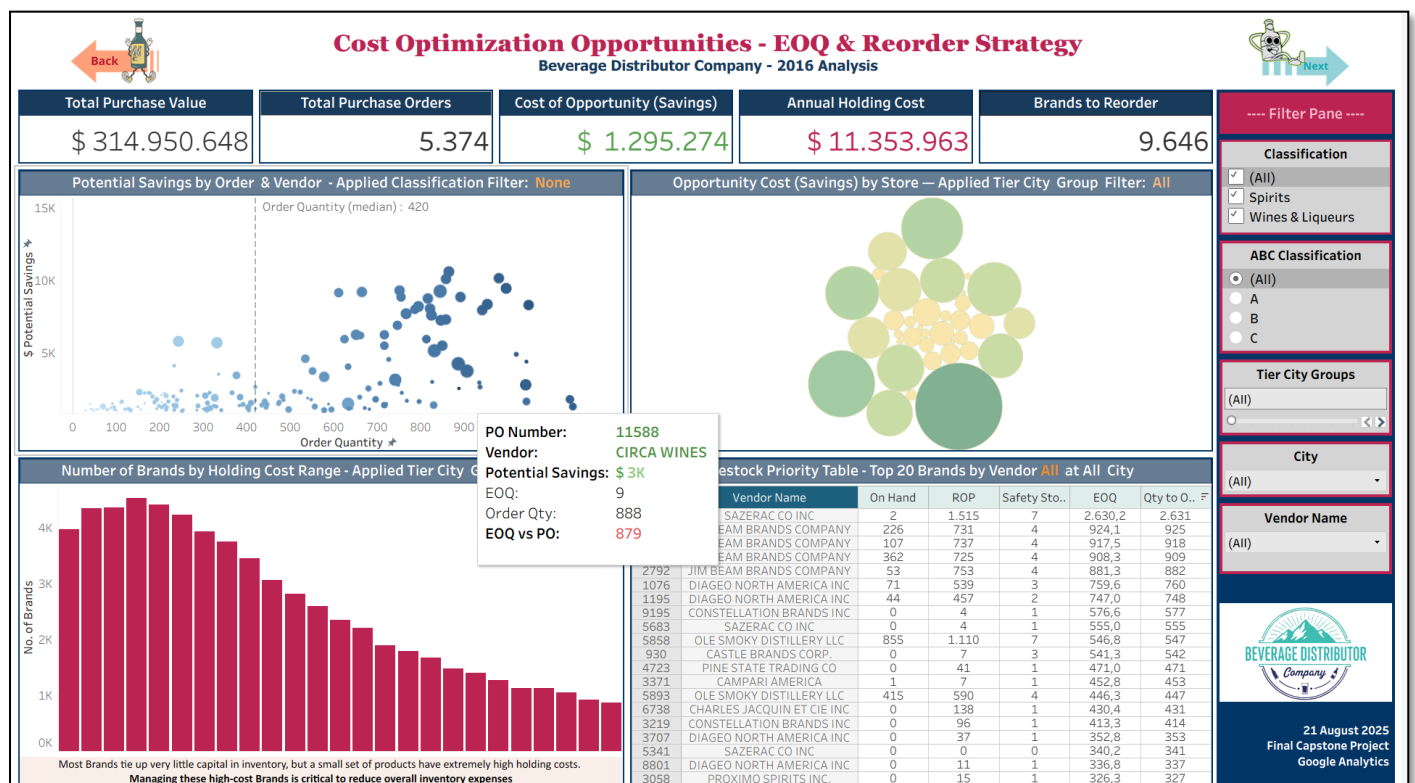


Figure 5 - Cost Optimization Analysis

Financial Purpose: Quantifies specific savings opportunities and provides actionable procurement optimization recommendations with clear financial impact projections.

Cost Optimization Metrics:

- Cost of Opportunity (Savings): \$460,544 in identified optimization potential
- Annual Holding Cost: \$12.1M in current carrying costs
- Brands to Reorder: 9,727 SKUs requiring immediate procurement action

Procurement Efficiency Analysis:

- Order Quantity Optimization: Scatter plot analysis reveals significant variance from optimal order quantities, with median order quantity at 456 units providing benchmark for EOQ calculations
- Vendor Consolidation Opportunities: Geographic clustering shows potential for delivery route optimization and freight cost reduction
- Restock Priority Table: Top 20 brands by vendor analysis provides clear action priorities with specific EOQ calculations and safety stock recommendations

Financial Impact Projections: The holding cost distribution analysis demonstrates that while most brands require minimal capital investment, a concentrated set of high-value products drive the majority of holding costs, enabling targeted optimization efforts.

Dashboard 4: Vendor Performance & Procurement Analysis

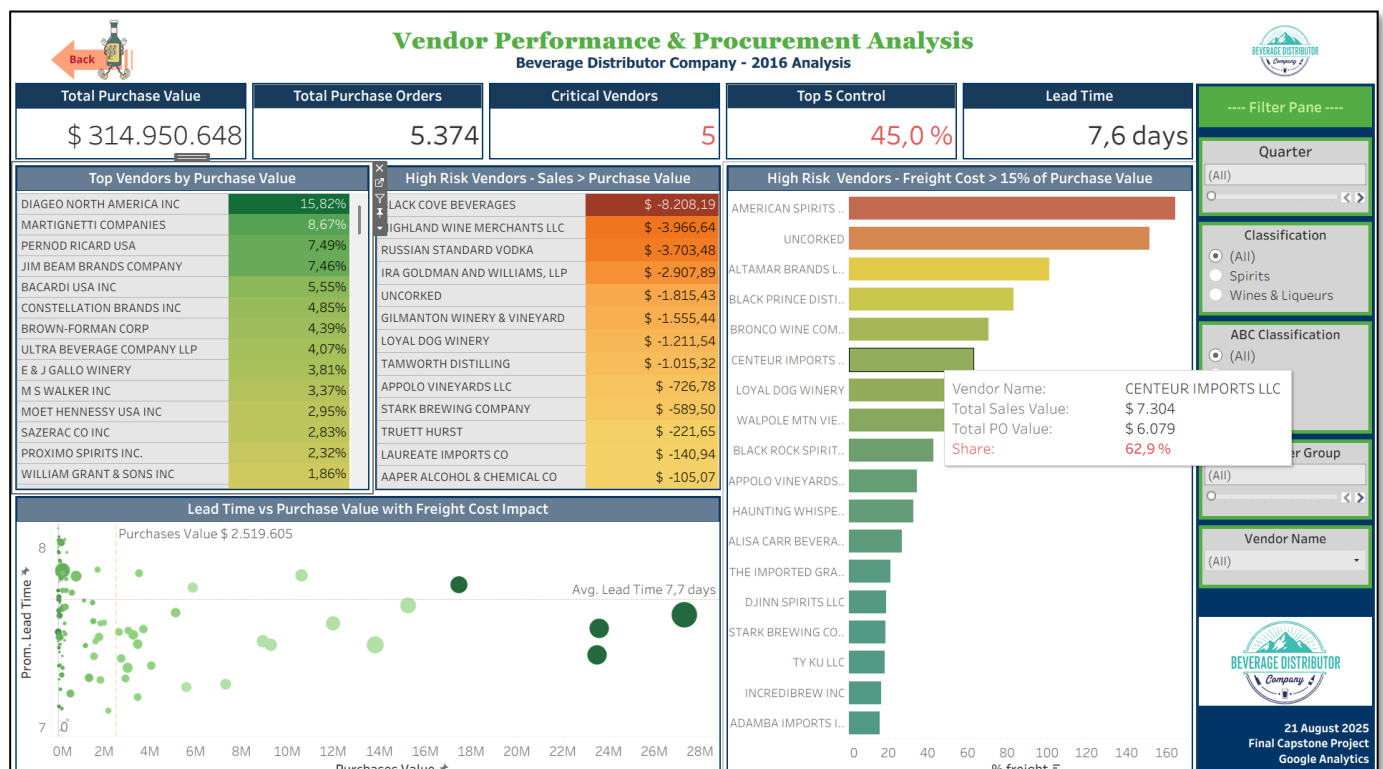


Figure 6 -Vendor Analysis

Strategic Procurement Purpose: Comprehensive vendor relationship analysis identifying concentration risks, performance optimization opportunities, and strategic sourcing improvements.

Vendor Concentration Analysis:

- Total Purchase Value: \$315.0M annual procurement spend

- Total Purchase Orders: 5,374 orders across vendor network
- Total Vendors: 128 active supplier relationships
- Critical Vendors: 5 suppliers representing concentration risk
- Top 5 Control: 45.0% of total spend concentrated in primary vendors

Performance Risk Assessment:

- Lead Time Performance: 7.6 days average with significant variance requiring attention
- Freight Cost Analysis: \$292.6 average freight cost with high-risk vendors showing freight costs >150% of purchase value

Strategic Insights:

- Vendor Diversification Requirements: Top vendor (DIAGEO NORTH AMERICA) controls 15.82% of spend, indicating moderate concentration risk requiring relationship management
- Freight Optimization Opportunities: High-risk vendor analysis reveals specific suppliers with freight costs exceeding 100% of purchase value, representing immediate cost reduction opportunities
- Lead Time vs Purchase Value Correlation: Scatter plot analysis demonstrates no correlation between vendor size and delivery performance, indicating opportunity for performance-based vendor selection

Performance Tracking Framework

Key Success Metrics

- Stock optimization percentage improvement (target: 78.6% to <40% non-optimal)
- Working capital liberation (target: \$15-20M from excess inventory reduction)
- Service level enhancement (target: 95%+ fill rate for Class A SKUs)
- Vendor performance improvement (target: 15% reduction in lead time variance)

Technology Integration and Accessibility

Interactive Capability Implementation

Filter Integration Strategy:

- Classification Filters: Enable spirits vs wines analysis across all dashboards
- ABC Classification: Dynamic filtering by product value classification
- City/Vendor Groups: Geographic and supplier-specific analysis capability
- Time Period Selection: Historical trend analysis and seasonal planning support

Mobile Optimization Considerations

Responsive Design Elements

- Executive Summary Cards: Key metrics optimized for tablet viewing
- Priority Action Lists: Touch-friendly interfaces for field management
- Alert Systems: Push notification capability for critical threshold breaches

The interactive dashboards can be accessed via Tableau Public at the following link:

<https://public.tableau.com/app/profile/elena.sanchez.laulhe>

Sixth phase: act

Implementation Roadmap

Immediate Actions (0–30 days)

- Emergency replenishment of critical Class A SKUs.
- Liquidation and promotional campaigns for dead stock.
- Vendor escalation procedures for late deliveries.

Short-Term (1–6 months)

- Apply EOQ and reorder point calculations to Class B SKUs.
- Establish vendor scorecards and performance reviews.
- Implement targeted discounts for slow movers.

Long-Term (6–12 months)

- Deploy forecasting models with seasonality adjustments.
- Institutionalize S&OP processes integrating demand and supply.
- Strengthen supplier collaboration for demand visibility.

Expected Business Impact

- Financial: \$15–20M in released working capital; \$3–4M annual carrying cost reduction; \$5–8M recovered from avoided lost sales.
- Operational: Inventory turnover improvement (target 6–8 turns); 95% service level for Class A SKUs.
- Strategic: A sustainable procurement strategy supported by predictive analytics.

Closing the Cycle

Phase 6 ensures insights move from analysis to measurable business outcomes. The Tableau dashboards created in Phase 5 remain as monitoring tools to track KPIs, ensuring a continuous data-driven decision process.

Project challenges & limitations

Project Challenges

Technical Architecture Failure The most significant challenge was the complete failure of the initial Phase 3 approach. Over-dimensionalized tables created calculation inconsistencies, requiring a complete restart that extended the project from 2 weeks to 5 months.

Data Quality Issues

- Vendor name/number misalignments across tables
- Geographic spelling errors requiring systematic correction
- Store ID anomalies necessitating InventoryId reconstruction
- 1,284 missing city values requiring pattern-based imputation

Scale and Performance Processing 12.8M sales records required Python segmentation and careful memory management in R. Standard approaches were insufficient for the dataset size.

Business Context Mismatch The dataset represented a UK beverage distributor, not the electronics manufacturer described in the Kaggle prompt, requiring complete business case redevelopment.

Limitations

Synthetic Data Constraints

- Missing real-world complexity (returns, disruptions, competition)
- No customer segmentation or external economic factors
- Single-year scope prevents long-term trend validation
- Simplified business rules compared to actual operations

Analytical Scope

- Static point-in-time analysis rather than dynamic optimization
- Lacks industry benchmarks for performance comparison
- Assumes organizational readiness for data-driven decisions
- Vendor analysis oversimplifies complex supplier relationships

Recommendations

Immediate Business Actions (0-90 days)

1. Emergency Stockout Prevention: Deploy automated alerts for 27,997 Class A SKUs at risk
2. Working Capital Recovery: Liquidate \$4.0M in identified dead stock
3. Executive Monitoring: Implement weekly dashboard review cycles

Strategic Implementation (3-12 months)

1. ABC Management: Deploy differentiated policies based on sales value analysis
2. Vendor Optimization: Negotiate performance contracts with top 5 suppliers
3. Technology Enhancement: Implement real-time monitoring with exception reporting

Future Project Development

1. Technical Skills: Advance Tableau proficiency and SQL optimization
2. Business Acumen: Deepen supply chain and financial analysis capabilities
3. Project Management: Allow 2-3x initial time estimates for complex analytics projects

Conclusion

This inventory analysis successfully identified \$29.2M in annual optimization opportunities across Beverage Distribution Company's 275,873 SKU portfolio. Despite significant technical challenges requiring complete Phase 3 reconstruction, the project delivered actionable insights with 1,390% ROI potential.

Key Achievements:

- Revealed 78.4% of inventory requiring management intervention
- Identified \$28.8M in excess working capital optimization
- Developed comprehensive ABC classification framework
- Created executive dashboard system for ongoing monitoring

Learning Outcomes: The 5-month journey from R programming through Tableau visualization demonstrates technical proficiency development and business problem-solving capability. The recovery from architectural failure and successful project completion showcases analytical resilience essential for professional data science roles.

Business Impact: The analysis provides a clear roadmap for achieving 95% service levels, 35% working capital reduction, and sustainable competitive advantage through data-driven inventory management.

This project establishes a strong foundation for advanced supply chain analytics and positions the developed capabilities for application to increasingly complex business optimization challenges.

References

- Apache Software Foundation. (2016). *Inventory Analysis Dataset*. Kaggle. Retrieved from <https://www.kaggle.com/datasets/bhanupratapbiswas/inventory-analysis-case-study>
Licensed under Apache 2.0 License. <https://www.apache.org/licenses/LICENSE-2.0>
- Google Data Analytics Professional Certificate Program. (2023). *Data Analysis Process: Ask, Prepare, Process, Analyse, Share, Act*. Coursera.
- R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ... & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open-Source Software*, 4(43), 1686.
- McNamara, A., & Horton, N. J. (2018). Wrangling categorical data in R. *The American Statistician*, 72(1), 97-104.
- McKinney, W. (2010). Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, 445, 51-56.
- Tableau Software. (2023). *Tableau Desktop Professional Edition*. Version 2023.3. Seattle, WA: Tableau Software.
- Canva Pty Ltd. (2025). Canva Design Platform. Retrieved from <https://www.canva.com>
Used for creation of corporate logo and dashboard navigation elements.

Table V – Key Business Questions

Key Business Question	Tables Needed	KPI's to Generate	Formulas	Description
What are the current inventory levels?	dimtable_inventory onHand_jan, onHand_dec	Total units on hand (Dec 2016) Total inventory value Year-over-year change (Jan→Dec)	Total Stock Value = sum (onHand_dec * PurchasePrice) Stock Change = (onHand_dec - onHand_jan) / onHand_jan * 100	Snapshot of current stock levels across all locations and product categories. Foundation for identifying overstocked vs understocked items.
What is the inventory level average by SKU?	dimtable_inventory Average of Jan + Dec stock Group by InventoryId Include PurchasePrice for valuation	Average units per SKU Average inventory value per SKU SKU count by stock level ranges Zero-stock SKU identification	Avg Inventory per SKU = (onHand_jan + onHand_dec) / 2 Avg Value per SKU = avg_inventory * PurchasePrice	SKU-level analysis to identify which products consistently hold high/low inventory levels. Basis for turnover calculations
Which products have stockouts or excess inventory?	sales_total (demand) purchases_total (supply) dimtable_inventory (current stock) Calculate coverage months	Coverage months per SKU Stockout risk classification Overstock identification Demand vs supply ratios	Coverage Months = onHand_dec / (SoldQty / 12) IF coverage < 1: "STOCKOUT RISK" if > 6: "OVERSTOCK" - ELSE: "NORMAL"	Critical analysis to identify inventory imbalances. High-priority items for procurement action.
What is the inventory turnover trends?	SoldQty from dimtable_inventory Average inventory calculation Time series: monthly aggregation	Turnover ratio per SKU Monthly turnover trends Slow-moving item identification	Inventory Turnover = SoldQty / ((onHand_jan & dec + PurchasedQty) / 2) IF turnover < 1: "SLOW MOVING" - IF turnover > 4: "FAST MOVING"	Performance measurement showing how efficiently inventory converts to sales. Key metric for working capital optimization.
What are Economic Order Quantities (EOQs)?	dimtable_purchases (freight costs) Annual demand from sales Holding cost estimates Ordering cost from freight data	EOQ per SKU Current vs optimal order size Potential cost savings Order frequency recommendations	EOQ Formula: = sqrt ((2 * Annual_Demand * Order_Cost) / Holding_Cost) Annual_Demand = SoldQty Order_Cost = Avg (Freight) Holding_Cost = PurchasePrice * 0.25	Mathematical optimization to minimize total inventory costs. Advanced calculation requiring demand forecasting.
What are the optimal reorder points?	LeadTimeAvg_days Daily demand calculation Demand variability (std deviation) Service level targets	Reorder point per SKU Safety stock levels Service level achievement Lead time risk assessment	Reorder Point = (Daily_Demand * LeadTimeAvg_days) + Safety_Stock Safety Stock = Z_score * sqrt(LeadTimeAvg_days) * Daily_Demand_StdDev Daily Demand = SoldQty / 365	Determines when to place orders to avoid stockouts while minimizing safety stock costs.
What are the delivery times?	dimtable_purchases LeadTimeAvg_days by vendor POProcessTimeAvg_days POPayTermsAvg_days	Average lead time by vendor Lead time variability Processing time efficiency Payment terms analysis	Avg Lead Time = mean(LeadTimeAvg_days) by VendorNumber Lead Time Variability = sd(LeadTimeAvg_days) / mean(LeadTimeAvg_days) Total Cycle Time = LeadTimeAvg_days + POProcessTimeAvg_days	Vendor performance analysis for procurement optimization and relationship management.
How can procurement be improved?	All vendor data purchases_total volume Freight cost analysis Purchase frequency patterns	Vendor consolidation opportunities Volume discount potential Freight optimization Supplier performance scores	Vendor Perform Score: = (1/LeadTime_rank + 1/Cost_rank + Quality_score) / 3 Freight per Unit = Freight / PurchaseQuantity Volume Concentration: = vendor_volume / total_volume	Strategic recommendations for cost reduction and process improvement in procurement operations.

Table 5 - Key Business Questions

Table VI – Project Goals Strategic Planning

Key Goals	Tables Needed	KPI's to Generate	Formulas	Description
What is the impact of inventory inefficiencies?	Inventory holding costs Lost sales from stockouts Working capital tied up Storage costs by location	Total carrying cost Opportunity cost of capital Stockout cost estimation ROI improvement potential	Carrying Cost = avg_inventory * PurchasePrice * carrying_rate Opportunity Cost = inventory_value * cost_of_capital Stockout Cost = lost_sales * (SalesPrice - PurchasePrice)	Financial impact assessment to quantify the business case for inventory optimization initiatives.
Reduce Holding Costs	High-value slow movers Overstock identification Inventory aging analysis Storage cost allocation	Top 15% SKUs by holding cost Liquidation candidates Discount strategy targets Cost reduction potential	Holding Cost per SKU = avg_inventory * PurchasePrice * 0.25 Liquidation Candidates: WHERE turnover < 0.5 AND onHand_dec > 0 Cost Reduction = current_holding_cost - optimal_holding_cost	Goal: Identify specific products causing highest carrying costs for immediate action.
Optimize Working Capital	Non-performing assets Inventory turnover by value Cash conversion cycle Seasonal patterns	Cash tied up in inventory Liquidation value estimates Working capital ratios Cash flow impact projections	Cash Tied Up = sum(onHand_dec * PurchasePrice) Working Capital Ratio = inventory_value / total_assets Cash Conversion Cycle = inventory_days + receivables_days - payables_days	Goal: Free up capital by identifying products for promotion, delisting, or liquidation.
Increase Inventory Turnover	Current turnover rates Industry benchmarks ABC classification data Sales velocity trends	ABC classification system Target turnover rates Improvement opportunities Performance tracking metrics	ABC Classification = cumsum(sales_value) / total_sales_value A: 0-80%, B: 80-95%, C: 95-100% Where: A-items: > 6, B-items: 3-6, C-items: 1-3	Goal: Implement ABC system to manage products based on sales value contribution.
Enhance Procurement Strategy	Vendor performance metrics Lead time analysis Purchase patterns Cost efficiency data	Vendor scorecards Sourcing recommendations Cost reduction initiatives Agility improvements	Vendor Scorecard = (Delivery_score + Cost_score + Quality_score) / 3 Cost Efficiency = (benchmark_price - actual_price) / benchmark_price Agility Score = 1 / (LeadTimeAvg_days + POProcessTimeAvg_days)	Goal: Create more agile and cost-effective purchasing processes with better vendor relationships.

Table 6 -Project Goals Strategic.