DATA SCIENCE CHALLENGE

Inventory Optimization & Supply Chain Analysis

Slooze Take-Home Challenge

Technical Documentation

Comprehensive Analysis Report

February 2026

Table of Contents

1. Executive Summary

Project Overview

This analysis presents a comprehensive inventory optimization system for Slooze, a retail wine and spirits company operating across multiple locations. Using **2.37 million purchase records**, **1.05 million sales transactions**, and inventory data from 2016, we implemented four core analytical frameworks to transform raw transactional data into actionable business intelligence.

Key Deliverables

- **ABC Analysis:** Classified 7,658 products by revenue contribution using the 80/20 rule (Pareto principle)
- **Demand Forecasting:** Built Facebook Prophet time-series models for top 5 A-Class products
- **Reorder Point Analysis:** Calculated optimal inventory triggers with 95% service level safety stock
- **Lead Time Analysis:** Evaluated 120 vendors across $321.9 million procurement spend

Critical Finding

80% of revenue comes from only 19.6% of products (A-Class), yet none of these critical products use Premium suppliers. This creates a significant supply chain vulnerability that requires immediate attention.

$33.1M	7,658	128	79
Total Revenue Analyzed	Unique Products	Vendors Evaluated	Store Locations

2. Problem Statement & Objectives

2.1 Business Context

Slooze manages millions of transactions across sales, purchases, and inventory records spanning 79+ store locations. Traditional spreadsheet-based analysis is inadequate for this data volume, creating risks of:

- **Stockouts** of high-revenue items leading to lost sales
- **Excess inventory** carrying costs tying up working capital
- **Missed optimization opportunities** from lack of data-driven insights
- **Supplier inefficiencies** going undetected

2.2 Core Objectives

Table 1: Project Objectives

Objective	Description	Success Metric
Inventory Optimization	Determine ideal stock levels for different product categories	Reduced stockouts + minimized carrying costs
Sales & Purchase Insights	Identify trends, top-performing products, and supplier efficiency	Clear product/vendor segmentation
Process Improvement	Optimize procurement and stock control to minimize financial loss	Data-driven reorder triggers

Analytical Tasks Completed

1. **ABC Analysis** - Product classification by revenue contribution
2. **Demand Forecasting** - Time-series prediction using Prophet
3. **Reorder Point Analysis** - Inventory trigger calculations with safety stock
4. **Lead Time Analysis** - Supplier performance evaluation

3. Dataset Overview

Data Sources (6 Files)

Table 2: Dataset Files Summary

File Name	Size	Records	Purpose
SalesFINAL12312016.csv	127.86 MB	1,048,575	Sales transactions ($33.1M revenue)
PurchasesFINAL12312016.csv	401.75 MB	2,372,474	Purchase orders and procurement
BegInvFINAL12312016.csv	19.31 MB	206,529	Beginning inventory (Jan 1, 2016)
EndInvFINAL12312016.csv	21.00 MB	224,489	Ending inventory (Dec 31, 2016)
InvoicePurchases12312016.csv	591 KB	5,543	Invoice records with payment dates
2017PurchasePricesDec.csv	1.16 MB	12,261	Product price reference

Data Quality Summary

Dataset Characteristics

- **Date Range:** January 1 - December 31, 2016 (with February anomaly)
- **Geography:** 79-81 store locations across multiple cities
- **Products:** ~7,658 unique SKUs
- **Vendors:** 128 significant suppliers (more than 10 purchase orders)
- **Total Records:** 3.87 million rows analyzed

4. Methodology

4.1 Phase 1: Data Exploration & Cleaning

Data Loading Process

The dataset was automatically downloaded using KaggleHub (61.9 MB compressed). All 6 CSV files were loaded into pandas DataFrames with appropriate data type optimization for memory efficiency.

Data Cleaning Steps

1. Standardized date formats (PODate, ReceivingDate, SalesDate)
2. Calculated derived metrics (LeadTime_Days, Revenue)
3. Removed invalid lead times (negative or zero values)
4. Converted 2,372,474 purchase records to datetime format
5. Stripped whitespace from string columns
6. Validated referential integrity across datasets

Key Relationships Mapped

- **Sales to Inventory:** Connected via Brand + Store combination
- **Purchases to Vendors:** Linked via VendorNumber
- **Prices to Products:** Matched via Brand identifier

4.2 Phase 2: ABC Analysis

Objective

Classify inventory by revenue contribution to prioritize management attention and allocate resources effectively.

Methodology

Classification Criteria:
A-Class: Top 80% cumulative revenue (High priority)
B-Class: 80-95% cumulative revenue (Medium priority)
C-Class: Bottom 5% cumulative revenue (Low priority)

The analysis followed these steps:

1. Calculated total revenue per product (Brand + Description combination)
2. Sorted products in descending order by revenue
3. Computed cumulative percentage of total revenue
4. Applied Pareto classification thresholds

4.3 Phase 2.2: Demand Forecasting

Objective

Predict future demand for top A-Class products using time-series analysis to enable proactive inventory planning.

> **Data Limitation**
>
> Only 60 days of reliable data available (January-February 2016). February showed a 90% sales drop, likely a data anomaly or seasonal effect, which constrained forecasting accuracy.

Model Configuration

Implemented Facebook Prophet models with the following parameters:

- Daily and weekly seasonality enabled
- Multiplicative seasonality mode
- 14-day forecast horizon
- MAE and RMSE calculated for model validation

4.4 Phase 2.3: Reorder Point Analysis

Objective

Determine optimal inventory levels to trigger replenishment orders and avoid stockouts.

Reorder Point Formula:

```
ROP = (Average Daily Demand x Lead Time) + Safety Stock

Where:
Safety Stock = Z x Standard Deviation of Demand x sqrt(Lead Time)
```

Parameters

Table 3: Reorder Point Parameters

Parameter	Value	Description
Service Level	95%	Z-score = 1.65 (industry standard)
Lead Time	7.3-7.6 days	Vendor-specific average
Safety Stock	Variable	Based on demand variability

4.5 Phase 2.4: Lead Time Analysis

Objective

Assess supplier efficiency and identify procurement risks through lead time evaluation.

Methodology

1. Calculated LeadTime_Days = ReceivingDate - PODate
2. Filtered to 128 significant vendors (more than 10 purchase orders)
3. Classified vendors using median thresholds

Vendor Classification

Table 4: Vendor Classification Criteria

Classification	Median Lead Time	Standard Deviation	Risk Level
Premium	Less than or equal to 7.7 days	Less than or equal to 2.2 days	Low
Fast but Variable	Less than or equal to 7.7 days	Greater than 2.2 days	Medium
Slow but Steady	Greater than 7.7 days	Less than or equal to 2.2 days	Low-Medium
High Risk	Greater than 7.7 days	Greater than 2.2 days	High

5. Key Results

ABC Analysis Results

Table 5: ABC Classification Summary

Category	Products	% of SKUs	Revenue	% of Revenue	Avg per Product
A-Class	1,502	19.6%	$26.51M	80.0%	$17,649
B-Class	1,813	23.7%	$4.97M	15.0%	$2,742
C-Class	4,343	56.7%	$1.66M	5.0%	$382

Top 5 Revenue Products

Table 6: Top 5 Products by Revenue

Rank	Product	Revenue	% of Total
1	Captain Morgan Spiced Rum	$444,811	1.34%
2	Ketel One Vodka	$357,759	1.08%
3	Jack Daniels No 7 Black	$344,712	1.04%
4	Absolut 80 Proof	$288,135	0.87%
5	Tito's Handmade Vodka	$275,163	0.83%

Reorder Point Results

Table 7: Reorder Points for Top 5 Products

Product	Daily Demand	Lead Time	Safety Stock	Reorder Point	Current Stock	Status
Captain Morgan	579	7.3 days	1,452	5,676	16,769	Healthy
Ketel One	360	7.4 days	953	3,616	16,770	Healthy
Jack Daniels	273	7.4 days	597	2,620	15,047	Healthy
Absolut	310	7.5 days	672	2,978	12,268	Healthy
Tito's	292	7.0 days	769	2,811	14,018	Healthy

Inventory Status

All products are currently healthy with current stock levels 2-4x above reorder points. No immediate stockout risk detected.

Lead Time Analysis Results

Table 8: Vendor Distribution by Classification

Tier	Vendors	Percentage	Spend	Risk Level
Premium	26	21.7%	$50.6M	Low
Fast but Variable	34	28.3%	$165.0M	Medium
Slow but Steady	35	29.2%	$82.9M	Low-Medium
High Risk	25	20.8%	$23.5M	High

Critical Supplier Risk

All top 5 revenue products use "Fast but Variable" or "Slow but Steady" vendors. None use Premium vendors, creating supply chain vulnerability.

- Vendor 3960 (Diageo): Supplies Captain Morgan + Ketel One ($802K combined revenue)
- Vendor 1128 (Brown-Forman): Supplies Jack Daniels
- **Concentration Risk:** 40% of A-Class revenue depends on single vendor

6. Assumptions & Limitations

Data Limitations

1. **Temporal Scope:** Only 60 days of reliable data (January 2016). February showed 90% sales drop (likely anomaly).
2. **Missing Costs:** No carrying cost or ordering cost data prevented EOQ (Economic Order Quantity) calculation.
3. **Vendor Names:** Some vendors only had ID numbers; names extracted from InvoicePurchases where available.

Analytical Assumptions

1. **Service Level:** 95% used for safety stock calculations (Z = 1.65, industry standard)
2. **Lead Time Distribution:** Assumed normal distribution for safety stock formula
3. **Demand Stability:** Used January 2016 averages (ignoring February anomaly)
4. **Product Classification:** ABC classification based on 2-month snapshot; annual data would be more robust

Technical Constraints

1. **Prophet Forecasts:** Limited by short time series (60 observations)
2. **External Regressors:** No weather, holidays, or promotions included in forecasting
3. **Vendor Classification:** Thresholds based on median (data-driven) rather than business SLAs

7. Business Recommendations

Immediate Actions (0-30 Days)

1. Dual-Source Vendor 3960

40% of A-Class revenue depends on a single supplier (Diageo). Implement dual-sourcing strategy to mitigate concentration risk.

2. Renegotiate SLAs

$85 million spent with vendors slower than 7.7 days. Implement penalty clauses for late deliveries and incentive programs for early delivery.

3. Increase Safety Stock

Apply 1.5x multiplier for "Fast but Variable" vendors (4 of 5 top products) to buffer against delivery variability.

4. Consolidate Orders

Top 3 vendors account for 618,000 orders. Negotiate volume discounts and preferred customer status.

Strategic Initiatives (30-90 Days)

1. Supplier Development

Move A-Class products to Premium vendors (currently 0 of 5). Develop relationships with reliable suppliers.

2. Inventory Rationalization

Review 4,343 C-Class products (56% of SKUs, only 5% of revenue) for potential discontinuation.

3. Data Infrastructure

Collect full year of data for seasonal forecasting and improved demand prediction.

Risk Mitigation Priorities

Table 9: Risk Summary

Risk Factor	Exposure	Mitigation
Vendor Concentration	Only 21.7% of vendors are Premium tier	Diversify supplier base
Variable Suppliers	$165M (51.2%) of spend with unreliable vendors	Increase safety stock buffers
Lead Time Buffer	7.6-day average with 2.2-day variability	Maintain safety stock coverage

8. How to Run the Code

Prerequisites

- Python 3.8 or higher
- Google Colab (recommended) or Jupyter Notebook
- 16GB RAM (for large CSV processing)

Installation

```
pip install pandas numpy matplotlib seaborn plotly prophet scikit-learn kagglehub
```

Execution Steps

1. **Download Dataset:** Automatically via KaggleHub
2. **Phase 1:** Data Loading & Cleaning (Cells 1-10)
3. **Phase 2:** ABC Analysis (Cells 11-18)
4. **Phase 2.2:** Demand Forecasting (Cells 19-23)
5. **Phase 2.3:** Reorder Point Analysis (Cells 24-31)
6. **Phase 2.4:** Lead Time Analysis (Cells 32-39)

Output Files Generated

- ABC_Analysis_Results.csv
- Reorder_Point_Analysis.csv
- Vendor_Performance_Scorecard.csv
- AClass_Vendor_Analysis.csv
- Demand_Forecast_Detailed.csv

Sample Code Structure

```
# Phase 1: Data Loading
import kagglehub
import pandas as pd

# Download dataset
path = kagglehub.dataset_download("sloozecareers/slooze-challenge")

# Load data
sales = pd.read_csv(f"{path}/slooze_challenge/SalesFINAL12312016.csv")
purchases = pd.read_csv(f"{path}/slooze_challenge/PurchasesFINAL12312016.csv")
```

9. Conclusion

This analysis transformed Slooze's raw transactional data into actionable inventory intelligence. Through ABC classification, we identified that **20% of products drive 80% of revenue** - yet these critical items rely on variable suppliers, creating significant supply chain vulnerability.

The reorder point system provides data-driven triggers for procurement, while vendor analysis reveals **$85 million in spend with suboptimal suppliers**. By implementing the dual-sourcing and safety stock recommendations, Slooze can protect high-revenue streams while optimizing working capital across the portfolio.

Key Deliverables Provided

- Automated data pipeline with KaggleHub integration
- Statistical product classification (ABC Analysis)
- Predictive demand models with uncertainty quantification
- Operational reorder triggers with safety stock buffers
- Strategic vendor scorecards with risk classification

> **Final Note**

> All code is modular, documented, and ready for production deployment. The analysis framework can be extended to include additional data sources and more sophisticated forecasting models as the data infrastructure matures.

--- *End of Documentation* ---