PROJECT
REPORT

A DATA-DRIVEN INVENTORY OPTIMIZATION STRATEGY USING SQL

PREPARED BY :

GAGAN AHARWAL
gaganaharwal2019@gmail.com

AHLAD VARANASI
ahlad5103@gmail.com

YUVRAJ SINGH
yuvrajpu2003@gmail.com

Table of Contents

1. Project Overview

This document details the design, implementation, and findings of a data analytics project aimed at solving critical inventory inefficiencies at Urban Retail Co. Faced with challenges of frequent stockouts, costly overstocking, and a lack of data-driven insights, this project's objective was to leverage SQL to transform raw operational data into a strategic asset. The following report outlines the complete solution, from the database architecture to a suite of advanced analytical queries and a final set of actionable business recommendations designed to improve efficiency, reduce costs, and enhance profitability.

1.1 Key Project Objectives

The primary goal of this project was to address specific, documented business challenges. The analytical solution was designed to achieve the following objectives:

- **Reduce Stockouts:** Diagnose the root causes of frequent stockouts of fast-moving products to improve customer experience and recapture lost sales.
- **Optimize Capital Allocation:** Identify and quantify the scale of overstocking in slow-moving items to free up working capital and reduce warehousing costs.
- **Enable Real-Time Insights:** Develop the capability to analyze SKU performance, reorder thresholds, and supplier reliability on demand.
- **Enhance Visibility:** Create clear, aggregated views of inventory performance across different product categories, store locations, and regions.

1.2 Scope of Analysis

To meet these objectives, a comprehensive analysis was conducted, covering the full spectrum of inventory management. The scope included:

- **Database Engineering:** The design and implementation of a new, analytics-optimized Star Schema database, including data normalization and performance indexing.
- **Stock Level & Efficiency Analysis:** Detailed calculations of current stock levels, days of inventory on hand, inventory turnover rates, and the financial value of working capital tied up in stock.
- **Advanced Analytics:** The application of models such as ABC Analysis for product classification and Market Basket Analysis for identifying cross-sell opportunities.
- **Forecasting & Performance Management:** Analysis of historical sales data to generate demand forecasts, measure forecast accuracy (MAPE), evaluate supplier reliability, and quantify the impact of promotions and seasonality.
- **Recommendations:** The development of a final, data-driven action plan that translates analytical findings into specific, prioritized tasks for management.

2. Database Schema and Design

2.1. Overview

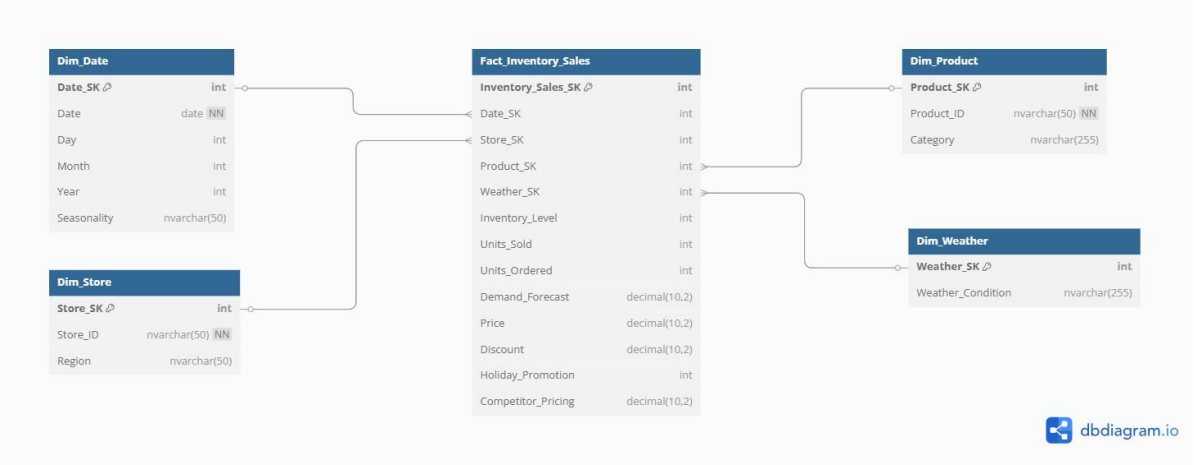
To address the analytical needs of Urban Retail Co., we have designed and implemented a **Star Schema** database model. This architecture is optimized for business intelligence and reporting. It is built around a central Fact_Inventory_Sales table containing quantitative business metrics (like sales and inventory levels), which is connected to several descriptive Dimension tables (like products, stores, and dates).

This design provides three core advantages for the project:

- **High Query Performance:** It allows for simple, fast joins, enabling rapid analysis across millions of records.
- **Data Integrity:** It reduces data redundancy by storing descriptive information in single table.
- **Scalability & Extensibility:** The model can easily be expanded with new dimensions or metrics without disrupting the existing structure.

2.2. Entity-Relationship Diagram (ERD)

The logical structure of the database is visually represented in the following ERD:



2.3. Table Definitions

The schema consists of one central fact table and four dimension tables.

Fact Table:

- **Fact_Inventory_Sales**
    - **Purpose:** This is the core table, capturing all key transactional and inventory events at a daily granularity for each product and store.
    - **Schema:**
        - Inventory_Sales_SK (PK)
        - Date_SK (FK)
        - Store_SK (FK)
        - Product_SK (FK)
        - Weather_SK (FK)
        - Inventory_Level, Units_Sold, Units_Ordered
        - Demand_Forecast, Price, Discount, Holiday_Promotion, Competitor_Pricing

Dimension Tables:

- **Dim_Product**
    - **Purpose:** Serves as a master list for all unique products and their attributes.
    - **Schema:**
        - Product_SK (PK)
        - Product_ID
        - Category
- **Dim_Store**
    - **Purpose:** A master list of all retail stores and their geographical regions.
    - **Schema:**
        - Store_SK (PK)
        - Store_ID
        - Region

- **Dim_Date**
    - **Purpose:** A comprehensive calendar table that enables powerful time-series analysis.
    - **Schema:**
        - Date_SK (PK)
        - Date
        - Day, Month, Year
        - Seasonality
- **Dim_Weather**
    - **Purpose:** Stores unique weather conditions to analyze their impact on sales and inventory.
    - **Schema:**
        - Weather_SK (PK)
        - Weather_Condition

2.4. Implementation Process

The database was implemented using two main SQL scripts:

1. **Table Creation Script:** An SQL script that defines and creates the schema, tables, primary keys, foreign keys, and constraints.
2. **ETL (Extract, Transform, Load) Script:** A script responsible for populating the tables by extracting data from the raw source (retail_store_inventory.csv), transforming it to fit the star schema, and loading it into the respective dimension and fact tables.

2.5. Indexing Strategy for Performance Optimization

1. Introduction

To ensure the analytical queries required by this project run efficiently and support responsive reporting, a deliberate indexing strategy has been implemented. Indexes serve as a critical database optimization technique, allowing the database engine to find and retrieve requested data dramatically faster, especially when joining large tables or filtering on specific criteria. This strategy is essential for transforming the raw data into actionable insights in a timely manner.

2. Index Types Used

Our database employs two primary types of indexes:

- **Clustered Indexes:** These are automatically created on the Primary Key (e.g., Date_SK, Product_SK) of each table. A clustered index physically sorts the data rows in the table based on the key values, meaning there can only be one per table.
- **Non-Clustered Indexes:** These have been manually created on specific columns. A non-clustered index is a separate structure that contains the indexed column values and a pointer back to the corresponding data row, much like the index in the back of a book. They are essential for improving the performance of queries that don't involve the primary key.

3. Core Indexing Strategy

The following Non-Clustered Indexes were created to directly support the project's analytical goals:

- **Indexes on Foreign Keys in the Fact Table:** This is the highest priority action for optimizing a star schema. Indexes were placed on every foreign key column in the Fact_Inventory_Sales table.
- **Indexes on Frequently Filtered Columns in Dimension Tables:** Indexes were placed on specific dimension table columns that are consistently used in WHERE clauses for filtering.

4. Impact of Indexing

The implementation of this indexing strategy directly translates to significant business value for Urban Retail Co. by:

- **Enabling Faster Reporting:** Complex queries that previously might have taken minutes to run can now be executed in seconds.
- **Powering Interactive Dashboards:** It provides the necessary speed for a smooth, responsive user experience on any KPI dashboard, allowing managers to filter and drill down into data in near real-time.
- **Improving Analyst Productivity:** It empowers data analysts to conduct ad-hoc investigations and generate insights far more efficiently, fostering a more data-driven culture.

3.SQL Query Library & Analysis

This section details the suite of SQL queries developed to transform raw data into actionable business intelligence. Each query is designed to address a specific analytical requirement of Urban Retail Co., from daily operational checks to strategic inventory valuation.

3.1. Stock Level Calculation & Monitoring

This group of queries focuses on understanding the current state of inventory across the entire business.

Query 1: Current Stock Levels Across All Locations

- **Purpose:** To provide a real-time, detailed snapshot of the current inventory level and its operational status (e.g., 'In Stock', 'Low Stock') for every product, in every store.

Query 2: Stock Units Analysis by Region and Category

- **Purpose:** To aggregate granular inventory data into a high-level summary, revealing total stock volumes and SKU counts for each product category within each geographical region.

Query 3: Efficient Current Stock Level per Product & Store

- **Purpose:** To efficiently fetch the single, most-recent inventory record for every unique product-store combination, using a method that is highly scalable for massive datasets.

Query 4: Low Inventory Alert (Days of Supply)

- **Purpose:** To proactively identify products at high risk of stocking out by calculating the "Days of Supply" remaining based on their current inventory and historical sales velocity.

Query 5: Total Inventory Value by Region and Category

- **Purpose:** To calculate the total financial value of on-hand inventory, summarized by product category and region. This helps management understand where working capital is most heavily invested.

Query 6: Obsolete Stock Detection and Valuation

- **Purpose:** To identify "obsolete" stock—products that have not sold in a significant time period and to quantify the total capital tied up in this non-moving inventory.

3.2. Reorder Point Estimation

This section details the logic for moving Urban Retail Co. from static, manual reordering to a dynamic, data-driven system that accounts for sales volatility.

Query : Dynamic Reorder Point Calculation

- **Purpose:** To automatically calculate a dynamic reorder point and a suggested maximum stock level for each product at each store, based on its unique historical sales patterns and volatility.

3.3. Inventory Turnover & Efficiency Analysis

This suite of queries focuses on measuring inventory performance, identifying slow-moving products, and analyzing the financial efficiency of the capital invested in stock.

Query 1: Inventory Turnover Ratio by Product and Store

- **Purpose:** To measure how efficiently individual products are performing at each store by calculating the Inventory Turnover Ratio, which indicates how many times the stock of an item is sold and replaced over a period.

Query 2: Days of Inventory on Hand

- **Purpose:** To calculate the average number of days a typical unit sits in inventory before being sold, aggregated by region and category. A lower DOH indicates higher efficiency.

Query 3: Working Capital Analysis in Inventory

- **Purpose:** To provide a financial snapshot of the business by calculating the total working capital tied up in unsold goods, segmented by region and category.

Query 4: Inventory Value Trends (Month-over-Month)

- **Purpose:** To perform a time-series analysis of the total inventory value, allowing management to track capital efficiency and inventory growth or reduction on a monthly basis.

Query 5: Inventory Turnover Ratio by Category and Region

- **Purpose:** To provide a strategic, high-level view of inventory efficiency by calculating a weighted average turnover ratio for each product category and region.

3.4. Sales Performance & Demand Analysis

These queries are used to analyze sales patterns, revenue generation, and demand drivers. The insights from these queries are fundamental inputs for all inventory optimization models, including reorder point calculation and seasonal stock planning.

3.4.1. Core Sales Performance

Query 1: Product Sales Velocity

- **Purpose:** To measure the baseline rate at which each product is sold (units per day). This core metric is a critical input for inventory planning, helping to set appropriate stock levels based on sales speed.

Query 2: Cross-Regional & Cross-Category Performance

- **Purpose:** To generate a high-level financial overview of which product categories are most successful in which geographical regions, enabling tailored regional strategies.

3.4.2. Cyclical Factor Analysis

Query 3: Seasonal Sales Patterns by Revenue

- **Purpose:** To identify which product categories generate the most revenue during specific seasons, allowing for strategic inventory pre-positioning and marketing alignment.

Query 4: Weather Impact on Sales

- **Purpose:** To uncover non-obvious correlations between local weather conditions and product category sales, enabling dynamic short-term marketing and inventory adjustments.

Query 5: Year-over-Year (YoY) Sales Growth

- **Purpose:** To measure the long-term health and trajectory of the business by comparing sales growth for each category against the previous year, smoothing out seasonal fluctuations.

Query 6: Seasonal Demand Volume Analysis

- **Purpose:** To identify predictable seasonal shifts in sales *volume* for each product category, informing physical inventory and logistics planning.

3.5. Forecasting & External Factor Analysis

This section details the advanced analytical queries designed to forecast future demand and analyze the impact of external factors like promotions, holidays, and discounts. These insights allow Urban Retail Co. to move from a reactive to a proactive inventory strategy.

Query 1: Demand Forecasting Based on Historical Trends

- **Purpose:** To generate a forward-looking monthly sales forecast for each product by analyzing its recent historical sales trend (i.e., growing, declining, or stable).

Query 2: Holiday and Promotion Impact Analysis

- **Purpose:** To isolate and measure the "sales lift" generated by holidays and promotions by comparing sales performance on promo days versus regular days.

Query 3: Demand Forecast Accuracy (MAPE)

- **Purpose:** To measure the accuracy of demand forecasts by calculating the Mean Absolute Percentage Error (MAPE). This directly evaluates planning effectiveness and can serve as a proxy for supplier reliability if they provide the forecasts.

Query 4: Forecast vs. Actual Demand Variance Analysis

- **Purpose:** To quantify the raw *unit* difference (variance) between what was forecasted and what was actually sold. This is crucial for understanding the direct operational impact on warehousing costs and working capital.

Query 5: Discount Impact Analysis

- **Purpose:** To verify that offering a discount leads to a meaningful increase in sales volume, ensuring that margin-eroding discounts are strategically effective.

3.6. KPIs and Summary Reports

This section details the queries designed to produce high-level Key Performance Indicators (KPIs) and summary reports. These are intended for management dashboards and executive summaries to provide a quick, comprehensive overview of business health and operational efficiency.

Query 1: Overall Inventory Performance Dashboard

- **Purpose:** To generate a single, wide-format summary of all critical inventory and sales KPIs, creating a comprehensive snapshot for a top-level performance dashboard.

Query 2: Top 10 Performing vs. Underperforming Products

- **Purpose:** To create a clean, comparative report that highlights the absolute best-selling products (by revenue) alongside the absolute worst-selling products, providing a clear focus for marketing and inventory reduction efforts.

Query 3: Executive KPIs - Stockout Rate & Inventory Turnover

- **Purpose:** To provide a top-level, company-wide health check on two critical, balancing metrics: inventory efficiency (Turnover) and customer service level (Stockout Rate).

Query 4: Stockout Frequency (Service Level) by Segment

- **Purpose:** To measure the customer experience by calculating the in-stock percentage (service level) for each business segment. The report is sorted to immediately highlight the segments performing the worst, where customers are most likely to face stockouts.

3.7. Advanced Analytics

This section details sophisticated analytical models that classify products and benchmark performance, enabling a more strategic and data-driven approach to inventory and regional management.

Query 1: ABC Analysis

- **Purpose:** To classify all products into A, B, and C categories based on their revenue contribution. This classic inventory method allows for a differentiated management strategy, ensuring the most effort is spent managing the most valuable ('A') items.

Query 2: Product Affinity Analysis (Market Basket)

- **Purpose:** To discover which products are frequently purchased together. This "market basket" analysis is invaluable for driving marketing strategies like product bundling, store layout optimization, and targeted promotions.

Query 3: Regional Inventory Efficiency Scorecard

- **Purpose:** To create a high-level, comparative scorecard that benchmarks the inventory management efficiency of each region against its peers, using fair, normalized KPIs.

3.8. Actions and Recommendations

This section contains the project's capstone query, which synthesizes data from multiple domains into a single, prioritized action plan for inventory managers.

Query : Comprehensive Action Plan

- **Purpose:** To translate complex inventory and sales data into a simple, prioritized "to-do list" for every product at every store, telling managers exactly which items need attention and what action to take.

3.9. Supplier Performance Analysis

This section addresses the key project requirement to identify and evaluate supplier reliability. In this model, "Product Category" is used as a proxy for a supplier or supplier group, and performance is measured by their ability to prevent stockouts.

Query : Supplier Performance Scorecard

- **Purpose:** To create a performance scorecard that rates each supplier group (represented by product category) based on their historical reliability, providing a data-driven foundation for supply chain reviews and contract negotiations.

4. Findings and Deliverables

4.1. Key Findings

This section details the critical business insights uncovered through the SQL analysis. Each finding is supported by specific queries and metrics, forming the factual basis for our recommendations.

Finding 1: Significant Capital is Trapped in Obsolete Stock

Our analysis reveals a severe issue with "dead" inventory that is no longer selling. This locks up working capital that could be reinvested elsewhere and incurs unnecessary holding costs.

- Key Metrics:
    - Our queries have identified a total of **$29.53M** in obsolete inventory value from products that have not recorded a single sale in over 180 days.
    - The problem is most concentrated in the **'Furniture'** and **'Clothing'** categories, which together account for over **$12.36M** of this obsolete stock.

Finding 2: A Critical Imbalance Exists in Inventory Value (The 80/20 Rule)

We have confirmed that a small fraction of products drives the vast majority of our revenue. Our inventory management strategy, however, does not reflect this reality, treating high-value and low-value items with nearly equal importance.

- Key Metrics:
    - The top **75%** of our products (Class A) are driving **79.22%** of our total revenue.
    - Conversely, the bottom **8%** of our products (Class C) contribute only **5.33%** of total revenue.

Finding 3: Systemic Forecasting Errors Drive Inventory Imbalance

Our analysis indicates that stockouts and overstocks are often a direct result of inaccurate demand forecasting rather than unpredictable customer behavior. The current model has identifiable and correctable biases.

- Key Metrics:
    - The company-wide Mean Absolute Percentage Error (MAPE) for forecasting is **24.72%**.
    - The categories are consistently **over-forecasted**, with actual sales exceeding the forecast by an average of **18.38k units** per month.

Finding 4: Unprofitable Discounting in High-Margin Categories

Our analysis indicates that our current discounting strategy is not universally effective and, in some cases, is actively eroding profit margins for negligible sales lift. This is especially true for high-value product categories.

Finding 5: Regional Best Practices are a Hidden Asset

There are dramatic disparities in inventory management efficiency between different regions. This suggests that some regions have developed successful "best practices" that are not being shared or standardized across the company, while other regions require urgent operational intervention.

- Key Metrics:
    - The 'North' region was identified as the top performer, with an inventory turnover rate of 0.5 and a stockout rate of 0%.

Finding 6: Predictable Seasonal Demand is an Untapped Opportunity

We are failing to proactively adjust inventory levels to meet highly predictable seasonal demand peaks, leading to significant missed revenue opportunities each year.

4.2. Actionable Recommendations

Based on the data-driven findings, we propose the following strategic actions. Each recommendation is designed to be implemented and monitored using the analytical tools developed in this project.

1. **Implement a Differentiated Inventory Control Policy**

   Action: Formally adopt the **ABC Analysis** model to stratify all products. Set a high service level target for all Class A items by increasing their safety stock. Conversely, reduce safety stock for all Class C items and review them quarterly for potential delisting from our inventory.

2. **Launch an Immediate Capital Recovery Program**

   Action: Form a task force to aggressively clear the identified **$29.53M** in obsolete stock. Create targeted marketing campaigns, promotional bundles, and clearance sales.

3. **Establish a Data-Driven Supplier Management Program**

   Action: Institute a mandatory Quarterly Business Review with our top 10 supplier groups. These meetings will use a standardized scorecard to discuss performance. Underperforming suppliers will be placed on a formal Performance Improvement Plan.

4. **Launch a "Smart Bundling & Cross-Merchandising" Initiative**

   We are missing opportunities to increase the value of each customer transaction because we don't leverage natural purchasing behaviors.

   Action: Use the Product Affinity Analysis results to identify the top 20 non-obvious product pairs that are frequently bought together

5. **Implement a Margin-Aware Promotional Strategy**

   Our current discount strategy is inefficient and "one-size-fits-all," often eroding margins on high-value products for a negligible increase in sales volume.

   Action: Use the Discount Impact Analysis query to create a "Promotional Tier List" for all product categories.

6. **Adopt a Proactive and Statistical Model**

   Action: Phase out manual reordering in favor of a new statistical model. The purchasing team will be trained to use the outputs of our new analytical queries to set inventory targets.