



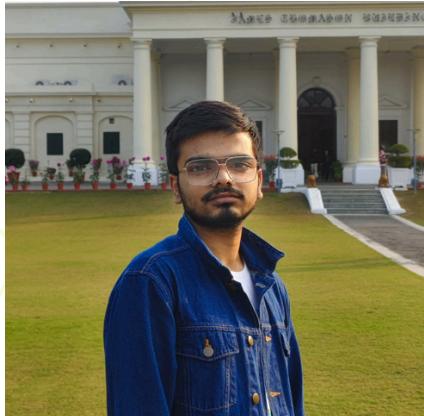
Team StockSmart

Summer Projects 2025

Project Report



Meet Our Team



Arpit Kumar

SQL & Data Analytics Specialist

Led SQL development, data cleaning.
Designed business dashboard using
plotly and streamlit.



Lakshay Bansal

Data Visualization & Business Analyst

Derived actionable insights from SQL
outputs and led inventory analysis.



Shikhar Bhandari

Machine Learning Engineer

Validated ML-ready datasets and initiated
forecasting model planning for future
implementation.



Consulting & Analytics Club
IIT Guwahati



SCAN THIS QR CODE



TO ACCESS DASHBOARD



Problem Statement

Solving Inventory Inefficiencies Using Advanced SQL Analytics

- **Problem Summary :**

Urban Retail Co., a growing retail chain, is facing inventory management inefficiencies due to a lack of integrated analytics. Common issues include frequent stockouts, overstocking, poor SKU visibility, and reactive decision-making. Despite having rich sales and logistics data, these datasets are underutilized.

- **Project Objective**

Design and implement a SQL-driven solution to monitor and optimize inventory operations across stores and warehouses using structured queries and analytics. The goal is to empower real-time insights and improve inventory accuracy, turnover, and service levels.

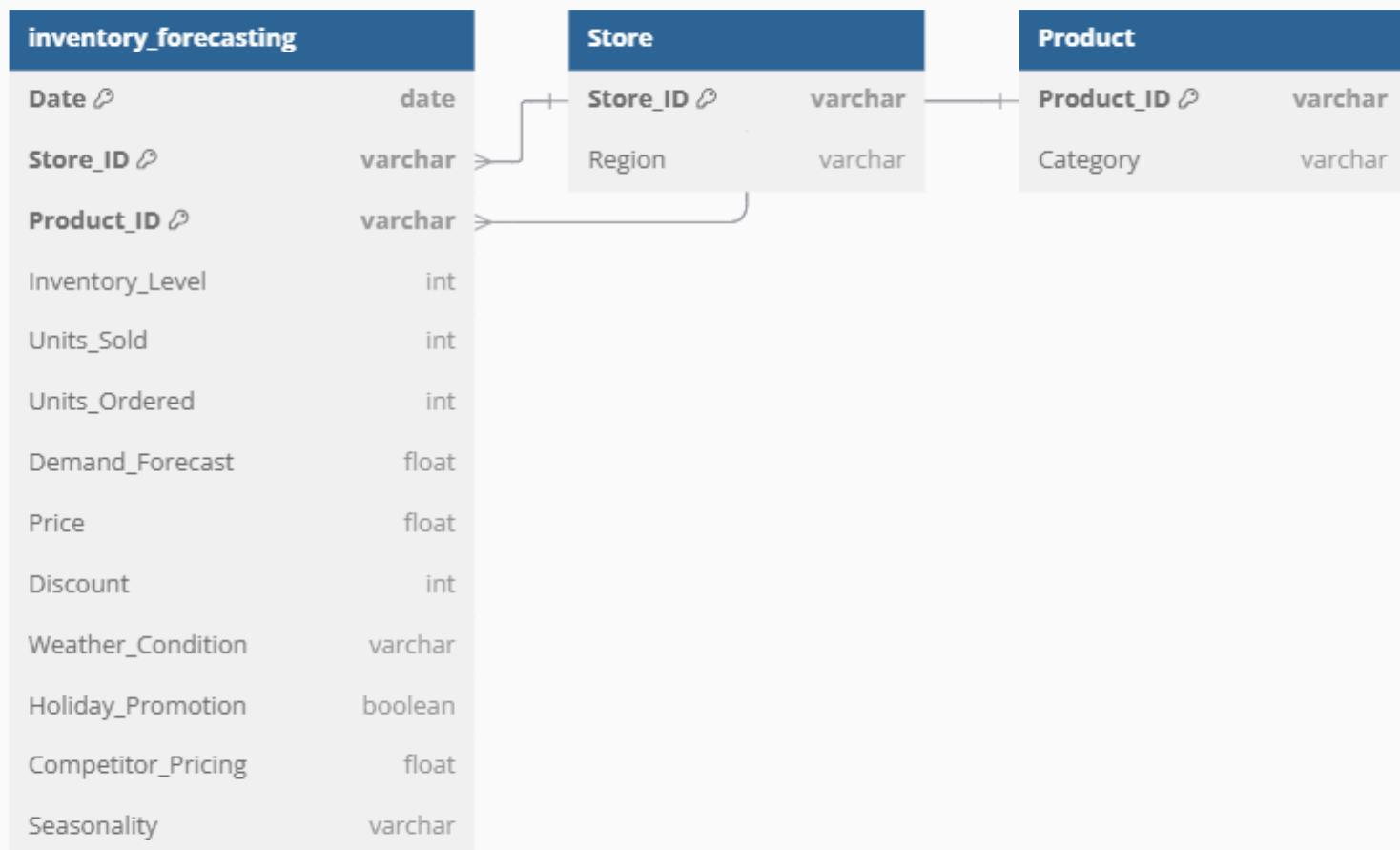
- **Expected Outcomes**

- Smarter inventory decisions using historical and real-time data
- Reduction in stockouts and excess inventory
- Improved supply chain efficiency
- Enhanced customer satisfaction and profitability



Database Design

The project used a pre-integrated dataset (`inventory_forecasting.csv`) that contained transactional, contextual, and forecast-related attributes in a single denormalized format. Instead of decomposing this into multiple normalized tables, the dataset was ingested directly into SQL to enable analytical query development and performance monitoring.





SQL Queries

This section highlights key SQL queries developed to extract actionable insights from the inventory dataset, covering stock levels, forecast deviations, and operational inefficiencies.

Query 1: Daily Inventory Snapshot

Get a date-wise view of inventory levels for each product at each store. This dataset will serve as a solid time-series reference for stock trends.

```
1   <use inventory;
2   <SELECT
3       date,
4       store_id,
5       product_id,
6       inventory_level,
7       units_sold,
8       units_ordered
9   FROM inventory_forecasting
10  ORDER BY date, store_id, product_id;
```



Query 2: Low Inventory Alerts

Identify products where current inventory < 7-day avg units_sold

```
1  use inventory;
2
3  WITH avg_demand AS (
4      SELECT
5          store_id,
6          product_id,
7          AVG(units_sold) AS avg_units_sold
8      FROM inventory_forecasting
9      WHERE date BETWEEN DATEADD(DAY, -7, '2022-12-24') AND '2022-12-31'
10     GROUP BY store_id, product_id
11 )
12
13  SELECT
14      inv.date,
15      inv.store_id,
16      inv.product_id,
17      inv.inventory_level,
18      inv.units_sold,
19      inv.units_ordered,
20      avg.avg_units_sold,
21      inv.price,
22      inv.discount,
23      inv.demand_forecast,
24      inv.weather_condition,
25      inv.seasonality,
26      inv.holiday_promotion,
27      inv.competitor_pricing
28  FROM inventory_forecasting inv
29  JOIN avg_demand avg
30      ON inv.store_id = avg.store_id
31      AND inv.product_id = avg.product_id
32  WHERE inv.date = '2022-12-24'
33      AND inv.inventory_level < avg.avg_units_sold
34  ORDER BY avg.avg_units_sold DESC;
```

Query 3: Inventory Turnover Rate

Measure how efficiently inventory is being sold. It shows how fast stock is moving for each product in each store.

Formula :

Inventory Turnover = Total Units Sold over a Period ÷ Average Inventory Level during that Period

We'll compute it per product-store over a 7-day window.

```
1  |> use inventory;
2  |> SELECT
3      store_id,
4      product_id,
5      SUM(units_sold) AS total_units_sold,
6      ROUND(AVG(inventory_level), 2) AS avg_inventory_level,
7      ROUND(
8          CAST(SUM(units_sold) AS FLOAT) / NULLIF(AVG(inventory_level), 0),
9          2
10     ) AS inventory_turnover
11     FROM inventory_forecasting
12     WHERE date BETWEEN DATEADD(DAY, -7, '2022-12-24') AND '2022-12-31'
13     GROUP BY store_id, product_id
14     ORDER BY inventory_turnover DESC;
```

Query 4: Forecast Deviation

Compare the actual units sold with the expected demand (forecast) to identify where the forecast was off – either underestimating or overestimating demand. This helps Analysts evaluate forecast accuracy and Inventory teams improve planning for promotions, pricing, etc.

```
1  use inventory
2
3  SELECT
4      date,
5      store_id,
6      product_id,
7      units_sold,
8      demand_forecast,
9      (units_sold - demand_forecast) AS forecast_deviation,
10
11     CASE
12         WHEN (units_sold - demand_forecast) > 20 THEN 'Underforecasted'
13         WHEN (units_sold - demand_forecast) < -20 THEN 'Overforecasted'
14         ELSE 'Accurate'
15     END AS forecast_accuracy_flag,
16
17     price,
18     discount,
19     Holiday_Promotion,
20     weather_condition,
21     seasonality
22 FROM inventory_forecasting
23 ORDER BY forecast_deviation DESC;
```



Query 5: Stockouts

Identify all instances where a product had zero inventory at a store on a given date – meaning it was out of stock. This helps Flag fulfillment issues

```
1  SELECT
2      date,
3      store_id,
4      product_id,
5      inventory_level,
6      units_sold,
7      units_ordered,
8
9      CASE
10         WHEN units_sold > 0 AND units_ordered = 0 THEN 'No Replenishment'
11         WHEN units_sold = 0 THEN 'No Demand'
12         ELSE 'Likely Stock Miss'
13     END AS stockout_reason,
14
15     price,
16     discount,
17     demand_forecast,
18     weather_condition,
19     holiday_promotion,
20     seasonality,
21     competitor_pricing
22
23     FROM inventory_forecasting
24     WHERE inventory_level = 0
25     ORDER BY date, store_id, product_id;
26
```



Query 6: Overstock

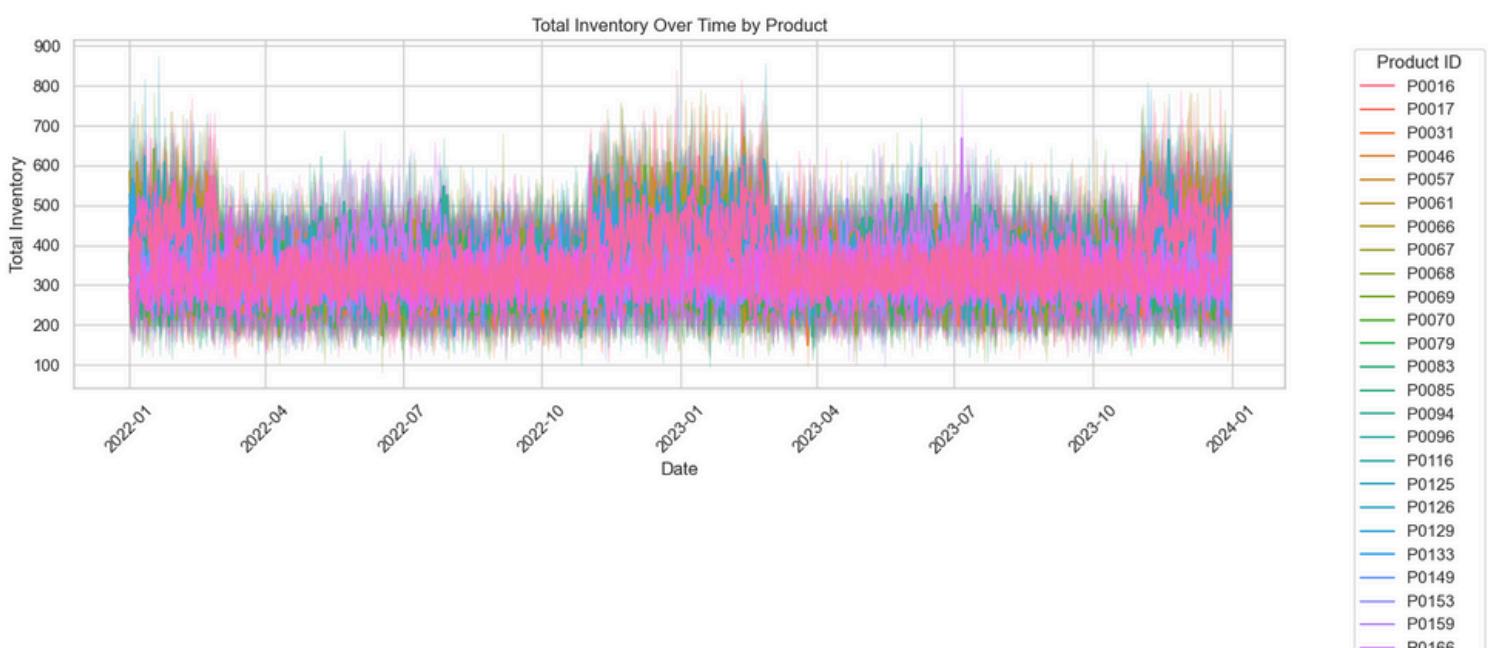
Identify product-store pairs that consistently have high inventory but low movement (i.e., low units_sold), indicating overstock or aging inventory. This helps Prevent wastage or product expiry , Free up storage space, Inform clearance sales or promotions

```
1  SELECT
2      date,
3      store_id,
4      product_id,
5      inventory_level,
6      units_sold,
7      units_ordered,
8
9      CASE
10         WHEN units_sold > 0 AND units_ordered = 0 THEN 'No Replenishment'
11         WHEN units_sold = 0 THEN 'No Demand'
12         ELSE 'Likely Stock Miss'
13     END AS stockout_reason,
14
15     price,
16     discount,
17     demand_forecast,
18     weather_condition,
19     holiday_promotion,
20     seasonality,
21     competitor_pricing
22
23     FROM inventory_forecasting
24     WHERE inventory_level = 0
25     ORDER BY date, store_id, product_id;
```



Inventory Trend Analysis

1. Total Inventory Over Time by Product



This line plot visualizes the temporal trends of total inventory levels for various products (identified by Product_ID) across the time range of *2022 to 2024*. Each product is represented by a distinct color in the legend.

Time Frame

- The x-axis spans from early 2022 to the beginning of 2024, showing over two years of inventory evolution.
- Inventory values are plotted for each product over time, allowing trend identification at daily or weekly resolution.
- The y-axis represents the total inventory, calculated as the sum of multiple warehouse zones (Inventory_1 through Inventory_4).
- The values fluctuate between 100 to 800 units, depending on the product and period.



Key Insights

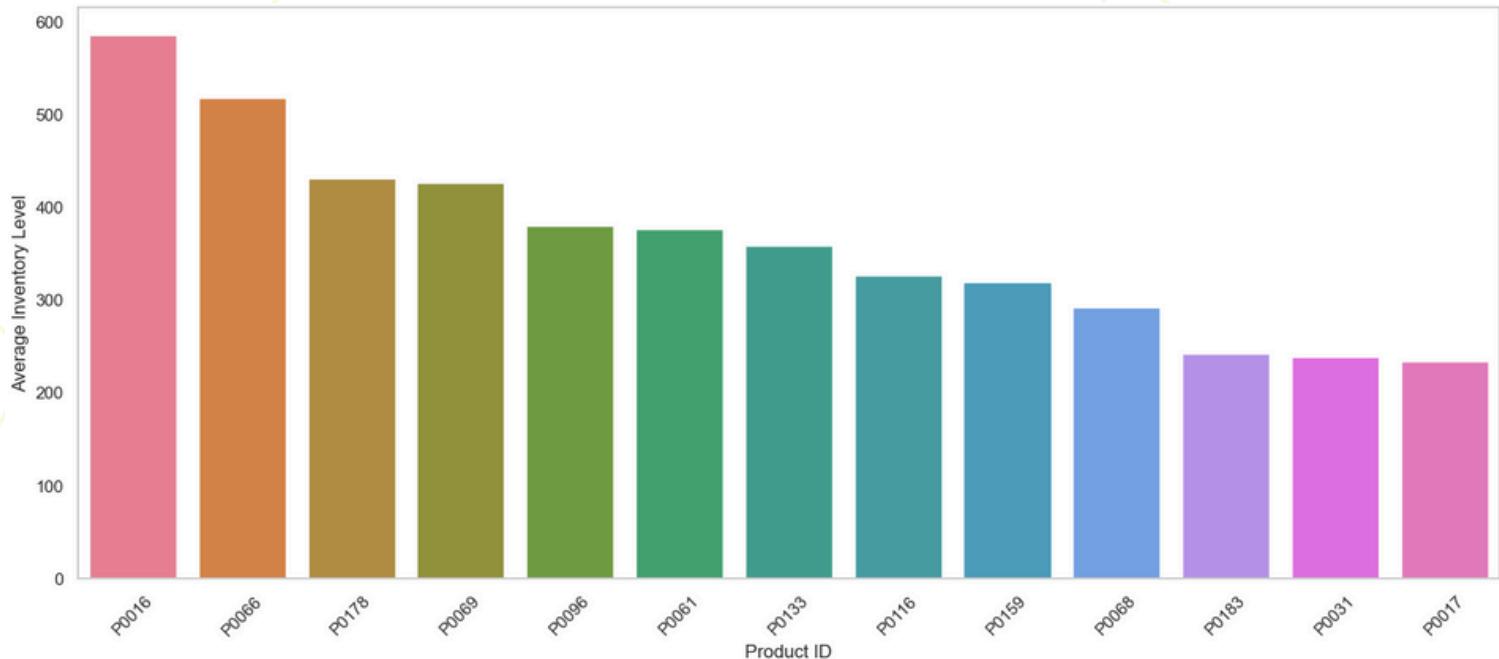
- **Seasonality and Demand Cycles :** Spikes in inventory can be observed at multiple intervals, particularly End of Q1 2022, Q4 2022, and Q3 2023. These suggest seasonal restocking patterns or anticipated high-demand periods(e.g., festive sales, end-of-year promotions).
- **Product Diversity :** Over 25 different products (e.g., P0016, P0096, P0133) are tracked. Some products (like *P0096, P0066, P0153) consistently show higher inventory levels, possibly indicating Popular SKUs* with high demand, Strategic overstocking due to supplier cycles or promotion dependency.
- **Inventory Volatility :** Products show rapid oscillations in inventory levels—indicating Frequent stock replenishment cycles, Possible demand forecasting inaccuracies, Varying sales rates across the products.
- **Stockout Risk Identification :** Periods of sudden drops in inventory (below \~200) may point to: Stockouts or delays in restocking, High consumption rates without matching replenishment.
- **Anomaly Detection :** Certain *sharp peaks and dips* (e.g., in early 2023 and late 2023) suggest potential Promotional stocking events*, or Supply chain disturbances

Interpretation for Business Strategy

- High fluctuation products require tighter demand planning and buffer stock policy.
- Stable inventory SKUs might allow leaner warehousing strategies.
- Temporal clusters of high inventory may need better coordination with *seasonal promotions and clearance planning.



2. Low Inventory Analysis



The bar chart illustrates the *average inventory level* maintained across different products in the dataset. It aims to highlight which products consistently maintain higher or lower stock levels, serving as a critical input for inventory optimization and demand planning.

Key Insights

- Top Inventory Products :** Product P0016 has the *highest average inventory level, nearing **590 units, indicating it is either a **high-demand product* or *overstocked* due to conservative replenishment policies. P0066 and P0178 follow, both maintaining over 500 and 430 average units, respectively. These could be essential or frequently moving items that require buffer stock.
- Lower Inventory Products :** P0017, P0031, and P0183 are among the lowest in average inventory, each hovering around 230–250 units. This might reflect low turnover, tight replenishment cycles, or minimal shelf space requirements. It could also indicate a potential risk of stockouts if demand spikes.



- **Product Inventory Balance :** The inventory spread shows diverse stocking behavior across products, with some being stocked almost 2.5× more than others. This hints at product-level demand variability, emphasizing the importance of customized inventory policies per SKU rather than a one-size-fits-all approach.

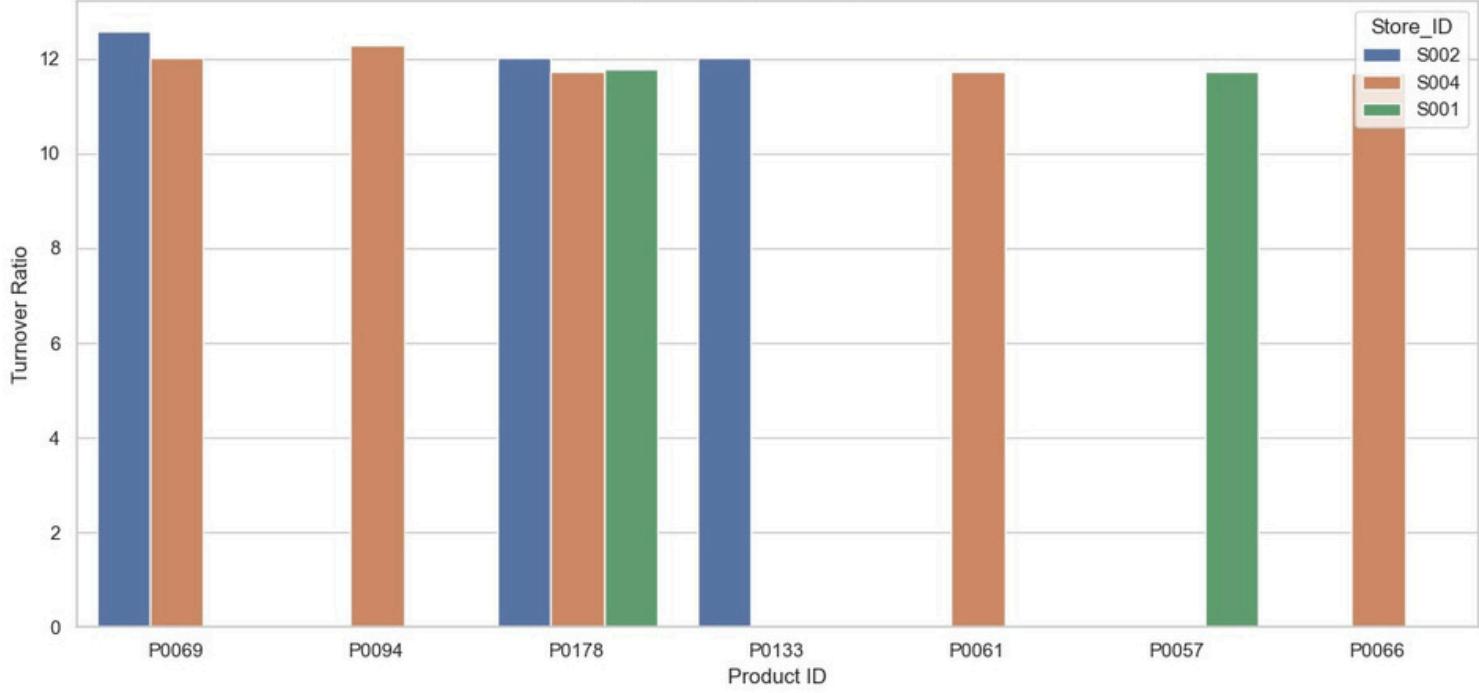
Strategic Recommendations

- Re-evaluate stocking policies for products like P0016, P0066, and P0178 validate whether high levels are demand-driven or simply due to overstocking.
- Monitor low-stock products like P0017, P0031, and P0183 for potential stockout risks. Analyze their historical demand trends and consider slight buffer increases if needed.



3. Top 10 SKUs by Inventory Turnover

Top 10 SKUs by Inventory Turnover



This bar chart presents the top 10 Stock Keeping Units (SKUs) with the highest inventory turnover ratios across different stores. The inventory turnover ratio is a critical performance indicator that reflects how efficiently a product is sold and replaced over a given period.

Key Insights

- **Top Performing SKUs :** Product **P0069 (S002)** and **P0094 (S004)** both exhibit turnover ratios **above 12**, indicating they are fast-moving items with high sales velocity relative to their inventory holding. **P0178**, appearing in both S004 and S001, also **maintains high turnover values (~12)**, indicating **consistent demand across stores** and possibly a **lean, efficient replenishment cycle**.
- Other **high performers** include:
 1. P0133 (S002)
 2. P0061 (S004)
 3. P0057 (S001)
 4. P0066 (S004)



- **Store-Specific Dynamics :** **Store S004** has 4 products in the top 10 list (P0069, P0094, P0061, P0066), indicating **strong inventory management or high customer traffic**. **Store S002** showcases P0069 and P0133, both high-turnover SKUs, **indicating efficient product movement**. **Store S001** has only P0057 and P0178, but they still compete well in turnover—**suggesting focused efficiency on fewer SKUs**.

Interpretation of Turnover Ratios

A higher turnover ratio (>12) generally indicates:

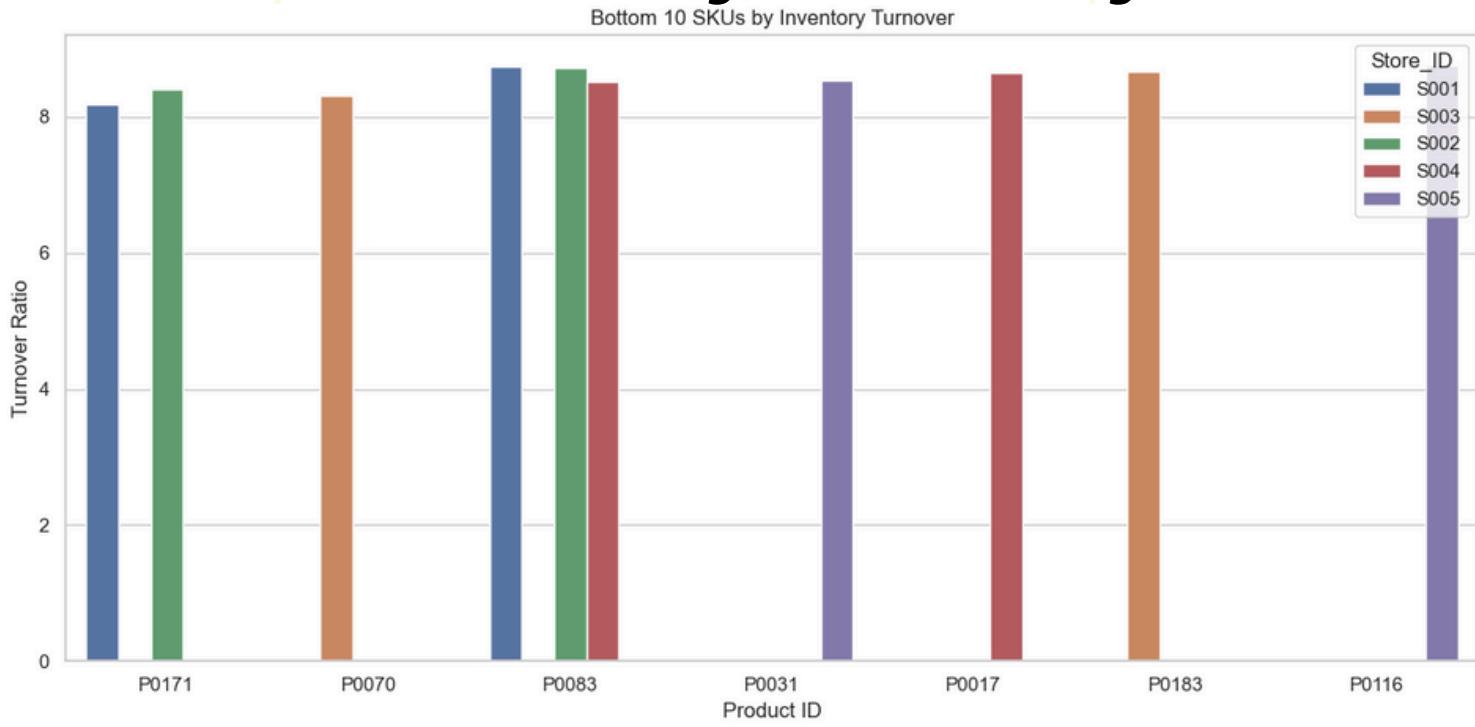
- Better product demand and sell-through rate.
- Lower risk of inventory holding costs or obsolescence.
- However, excessively high turnover may also hint at low safety stock, risking stockouts if not balanced with demand forecasting.

Strategic Recommendations

- **Replicate SKU strategy of S004 across other stores**, its dominance in high-turnover products suggests **best practices that could be transferred to S001 and S002**.
- **Review replenishment policies** for these top SKUs to ensure inventory remains **optimized and aligned with actual consumption**.
- **Align marketing and promotions** around **top-turnover SKUs** to drive further **efficiency and customer satisfaction**.



4. Bottom 10 SKUs by Inventory Turnover



This chart visualizes the 10 SKUs (products) with the **lowest inventory turnover ratios** across multiple stores, representing slow-moving or potentially overstocked items. The inventory turnover ratio measures how many times a product's inventory is sold and replaced over a given time period.

Key Insights

- Underperforming SKUs :** Products like **P0171, P0070, P0083, P0031, P0017, P0183, and P0116** consistently show **lower turnover ratios** (~8.2 to 8.8), indicating **Slower sales performance**, Possible **overstocking or misaligned demand forecasting and Suboptimal shelf rotation**.
- Store-Specific Breakdown :** These underperforming products appear in multiple stores:
 1. P0171– S001, S002
 2. P0070– S003
 3. P0083– S001, S002, S004
 4. P0031– S005



5. P0017- S004
6. P0183- S003
7. P0116- S005

- This wide store spread suggests the issue **may not be just location-specific** but could relate to the SKU itself, **such as:**

 1. **Ineffective pricing.**
 2. **Poor promotion.**
 3. **Reduced customer interest.**

Interpretation of Turnover Ratios

A low turnover ratio (<9) implies:

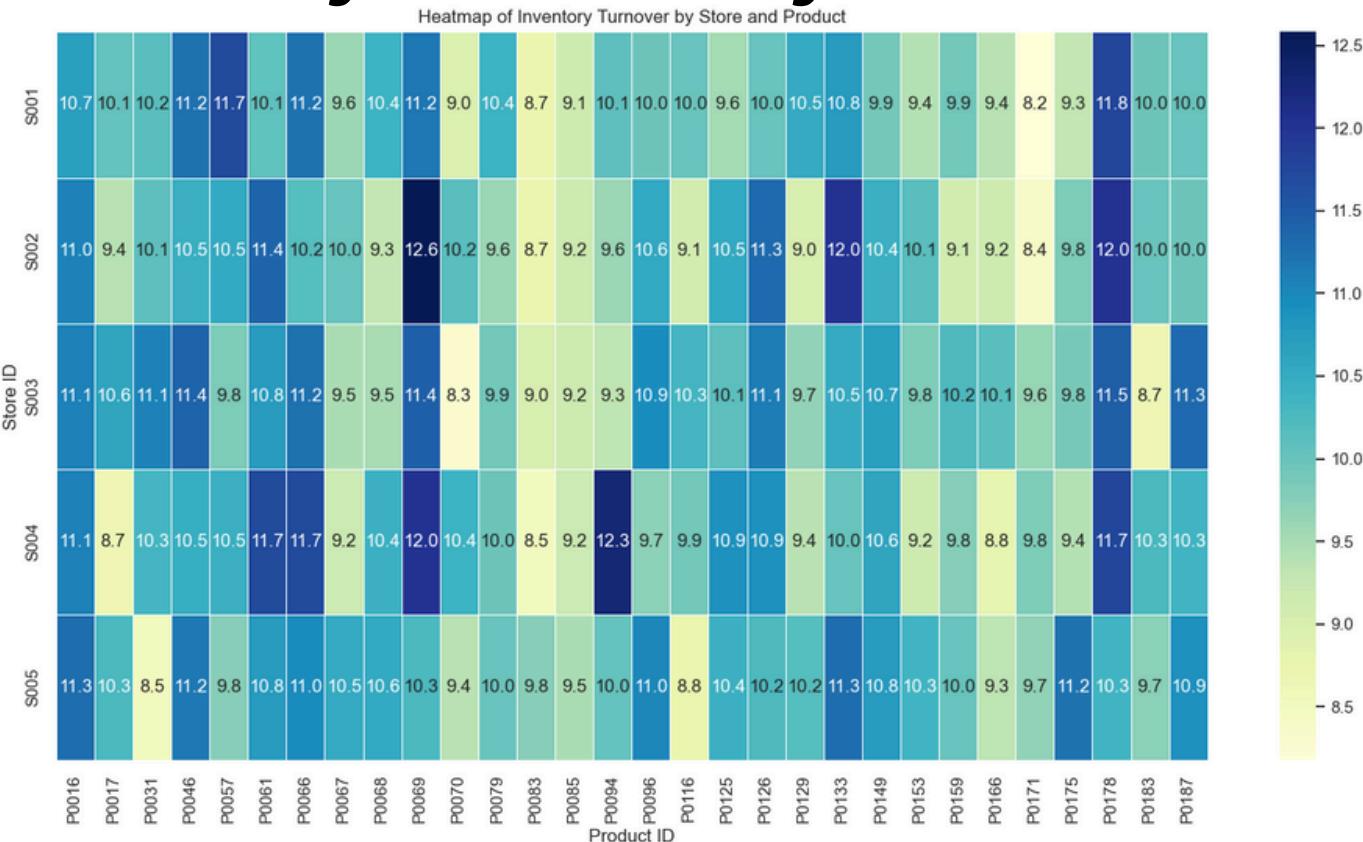
- **Excess inventory** leading to storage and holding costs.
- **Potential product obsolescence**, especially **in seasonal or fast-changing categories.**
- If these SKUs are essential but slow-moving, **it may be a sign to adjust stocking thresholds and demand forecasts.**

Strategic Recommendations

- **Demand Analysis :** Investigate why these SKUs aren't selling quickly. Are they priced right? Is marketing sufficient?
- **Inventory Rebalancing :** Consider transferring excess stock to higher-performing locations or bundling slow movers with fast movers.
- **Review Reordering Rules :** These products may benefit from just-in-time stocking rather than static thresholds.
- **SKU Rationalization :** If consistently underperforming, evaluate whether certain SKUs should be phased out or replaced.



5. Inventory Turnover by Store and Product



This heatmap visualizes the *inventory turnover ratio* across different combinations of *Store IDs* and *Product IDs*. The turnover ratio measures how often inventory is sold and replenished, making it a key indicator of inventory performance, efficiency, and potential demand.

Key Insights

- **High-Performing Zones**
1. **P0178:** Consistently shows **high turnover ratios across multiple stores**, e.g., S002 (12.0), S004 (11.7), S005 (11.2).
 2. **P0069 and P0094:** High turnover across all stores, **peaking at 12.6 (S002) and 12.3 (S004) respectively.**
 3. These SKUs are fast-moving products **ideal candidates for frequent restocking and prioritization in inventory planning.**



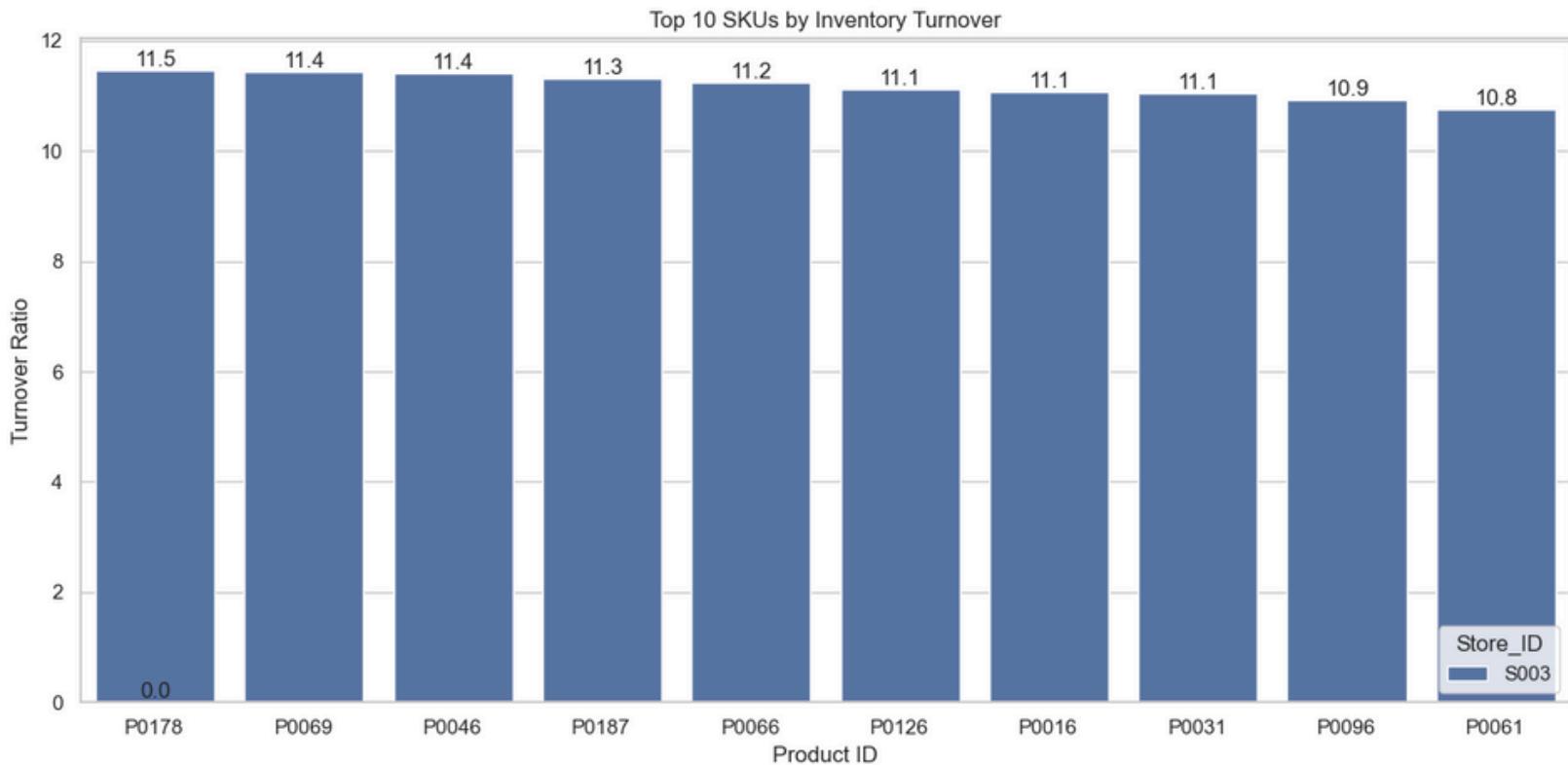
- **Low-Performing Zone**
 1. P0031 (S005 – 8.5) and P0171 (S001 – 8.2) show the lowest turnover, indicating slow movement.
 2. P0083 also records below-average turnover across most stores (8.5–9.2).
 3. These SKUs **may be suffering from low demand, ineffective promotions, or poor placement in stores.**
- **Store-Level Observations***
 1. **S002 & S004** have more frequent darker cells, **indicating efficient inventory management and faster turnover.**
 2. **S001 & S003** show a mix of high and low turnover, **suggesting opportunities for optimization.**
 3. **S005**, while mostly average, **reveals a few standout performers and underperformers, e.g., high for P0066 (11.2), low for P0031 (8.5).**

Strategic Recommendations

- **Optimize Stock Levels :**
 1. High-turnover items (e.g., P0069, P0094) should be **restocked more frequently** to avoid stockouts.
 2. Low-turnover items may require **reduced ordering frequency** to avoid overstock and holding costs.
- **Targeted Store Interventions :**
 1. Investigate stores with lower turnover (e.g., S001 for P0171) ,**reassess pricing, placement, or local demand factors.**
 2. Stores like **S002 and S004** can serve as benchmarks for **replication of stocking and sales strategies.**
- **SKU Rationalization :**
 1. Persistent underperformers across stores (e.g., P0031, P0083) **could be candidates for promotion strategies or even phase-out decisions.**



6. Top 10 SKUs by Inventory Turnover – Store S003



The chart displays the inventory turnover ratio for the top 10 SKUs in Store S003. Inventory turnover measures how efficiently stock is sold and replaced over a period, a higher ratio indicates better sales performance and leaner inventory management.

Key Insights

- High Inventory Turnover Across All Top SKUs**

1. The narrow range (10.8–11.5) suggests uniformly high-performing products, with consistent replenishment cycles.
2. High turnover implies that these SKUs are fast movers, likely contributing a large portion of revenue and shelf activity.

- SKU Strategy***

1. These items may be:
 - **Frequently promoted**
 - **Core essentials or high-repeat-purchase items**
 - **Seasonless or staple items with *stable year-round demand**



- **Operational Implications**

1. These products are being replenished *approximately every 32 days (assuming monthly operations).
2. Indicates efficient stock handling, minimal holding costs, and strong product-market fit.
3. There's likely a well-synchronized supply chain supporting these SKUs — little deadstock, high predictability.

- **Strategic Recommendations**

1. Prioritize High-Turnover SKUs in Replenishment Planning
2. Ensure automatic reordering or dynamic safety stock levels for these items.
3. Maintain buffer inventory for top SKUs during peak seasons or promotions.

- **Replicate Success Across Stores**

1. Conduct cross-store performance analysis:
 - **Are these SKUs high-turnover across all stores, or unique to S003?**
2. If unique, examine local factors (demographics, marketing, layout) that may be replicable.

- **Use Turnover Ratios for Inventory Benchmarking**

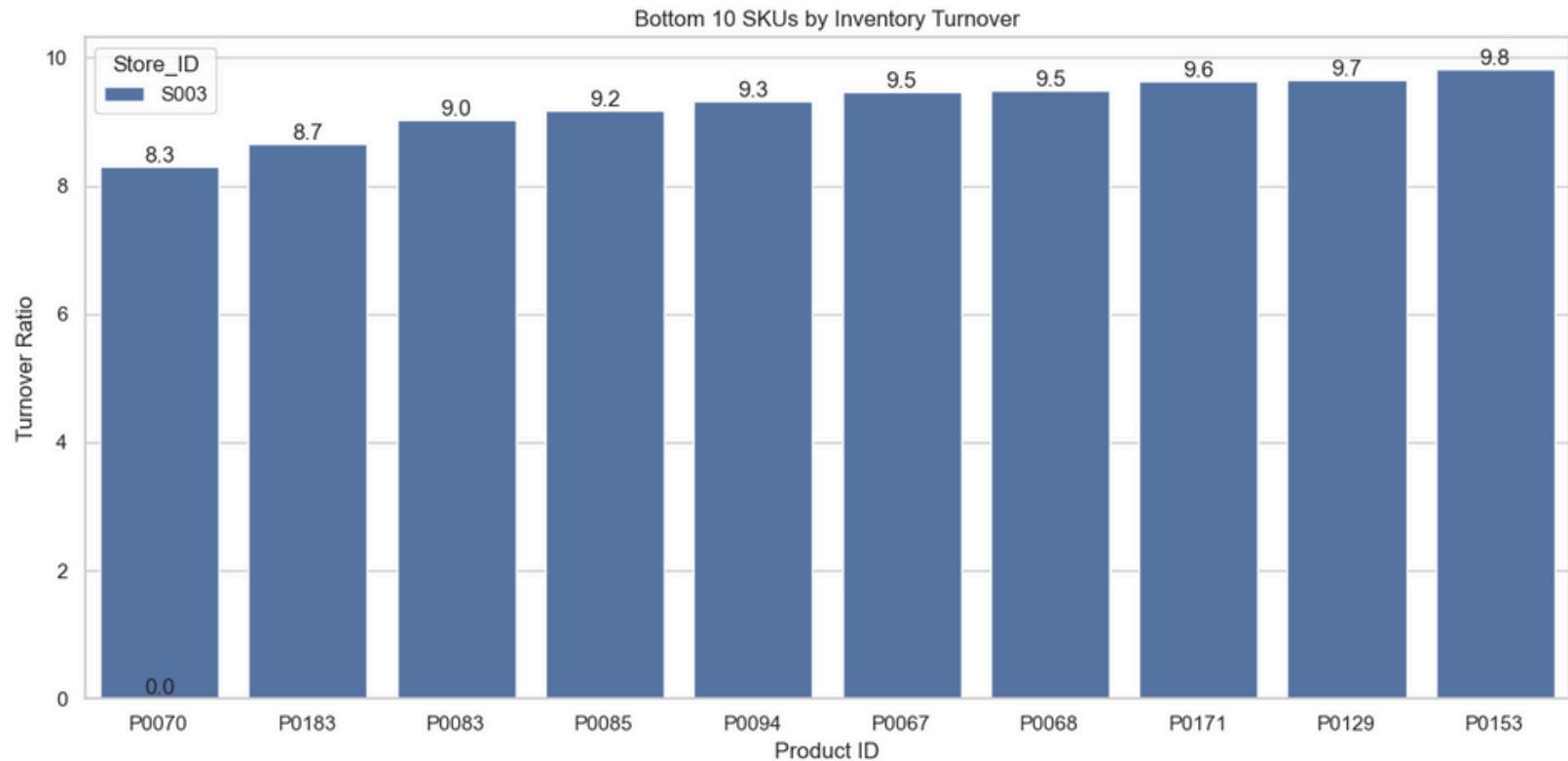
1. Classify products into ABC categories based on turnover (A = top 20%).
2. Use this classification to drive:
 - **Assortment decisions**
 - **Shelf-space optimization**
 - **Vendor negotiations**

- **Explore Upsell and Cross-Sell Opportunities**

1. Bundle top SKUs with lower-turnover items to boost overall inventory flow.
2. Leverage them in promotional campaigns — e.g., “Buy Product P0178, Get 10% Off on Accessories.”



7. Bottom 10 SKUs by Inventory Turnover – Store S003



This chart identifies the least efficient SKUs in terms of inventory turnover for Store S003. A lower turnover ratio implies that the product is selling slower, leading to higher holding costs and potential overstocking.

Key Insights

- **Lowest Performing SKU**

1. P0070 stands out with an inventory turnover of 0.0, which is either:
2. A data issue (e.g., zero recorded sales) Or a genuine stock stagnation, which is a red flag
3. This requires urgent investigation and may suggest discontinued demand or stockout problems.

- **Slow-Moving SKUs**

1. P0183 (8.7) and P0083 (9.0) are among the lowest active performers.
2. They pose inventory management challenges and likely tie up capital without yielding fast returns.



- **Moderate Yet Low Turnover SKUs**
 1. Products like **P0085, P0094, P0067, P0068, P0171, P0129, and P0153** have turnover ratios between **9.2 to 9.8**.
 2. While not extremely poor, these SKUs still rank among the **bottom 10 and should be monitored closely.**
 3. **Potential causes include:**
 - Poor shelf placement
 - Low demand trends
 - Over-purchasing relative to demand

Strategic Recommendations

- **Product Rationalization**

Reassess the SKU strategy* for P0070 and P0183. If they remain stagnant, consider markdowns or removal from inventory.

- **Sales Stimulation**

Apply promotions or bundling strategies to push underperforming SKUs like P0083 and P0085.

- **Inventory Optimization**

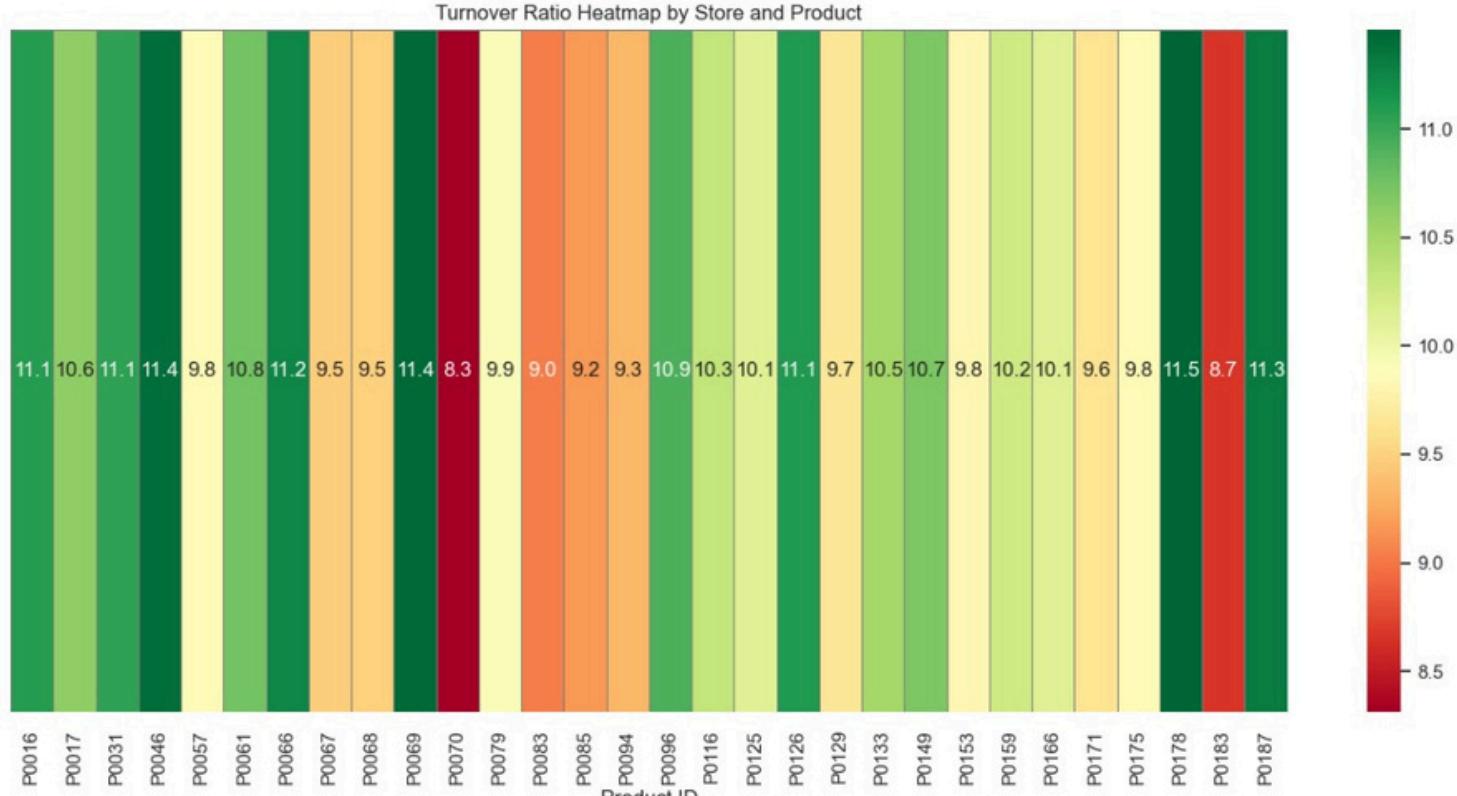
These SKUs may require reduced stock levels or just-in-time restocking to avoid holding excess inventory.

- **Forecasting Improvement**

Historical turnover should inform more accurate demand forecasting, especially for borderline SKUs (P0067–P0153).



8. Turnover Ratio Heatmap – Store S003



This heatmap provides a **color-coded view** of how efficiently each product (SKU) is being sold and replaced in *Store S003. The inventory turnover ratio reflects **how quickly inventory is sold and replenished?** higher numbers indicate better turnover.

Key Insights

- **High Turnover Products (Efficient)**

1. These products show **turnover ratios above 11.0**, highlighted in **dark green**:

 - P0016 : 11.1
 - P0031 : 11.1
 - P0066 : 11.2
 - P0126 : 11.1
 - P0178 : 11.5
 - P0187 : 11.3



2. These SKUs are **high performers in terms of sales and stock movement**. These should be **prioritized for restocking and promoted more aggressively**.

- **Moderate Turnover Products**

1. These range between 10.0 and 11.0, represented in light green :
P0046, P0061, P0094, P0096, P0125, P0133, P0149, etc.

2. These items are performing reasonably well. They maintain a healthy balance of supply and demand, and do not require urgent intervention.

- **Low Turnover Products (Underperforming)**

1. These products are marked in **yellow to red** and have a turnover ratio **below 10.0**: **P0070** : 8.3 (lowest) , **P0183** : 8.7 , **P0079** : 9.9 **P0083, P0085** : \~9.0–9.2.

2. These are potential **slow-moving SKUs**. They might indicate Overstocking, Low demand, Pricing or marketing issues

3. They should be evaluated for *promotions, discontinuation, or price adjustments*.

Strategic Recommendations

- **Boost top performers**

Ensure consistent availability of products like P0178, P0066, P0031.
Consider bundling them or offering as part of promotions.

- **Investigate low performers**

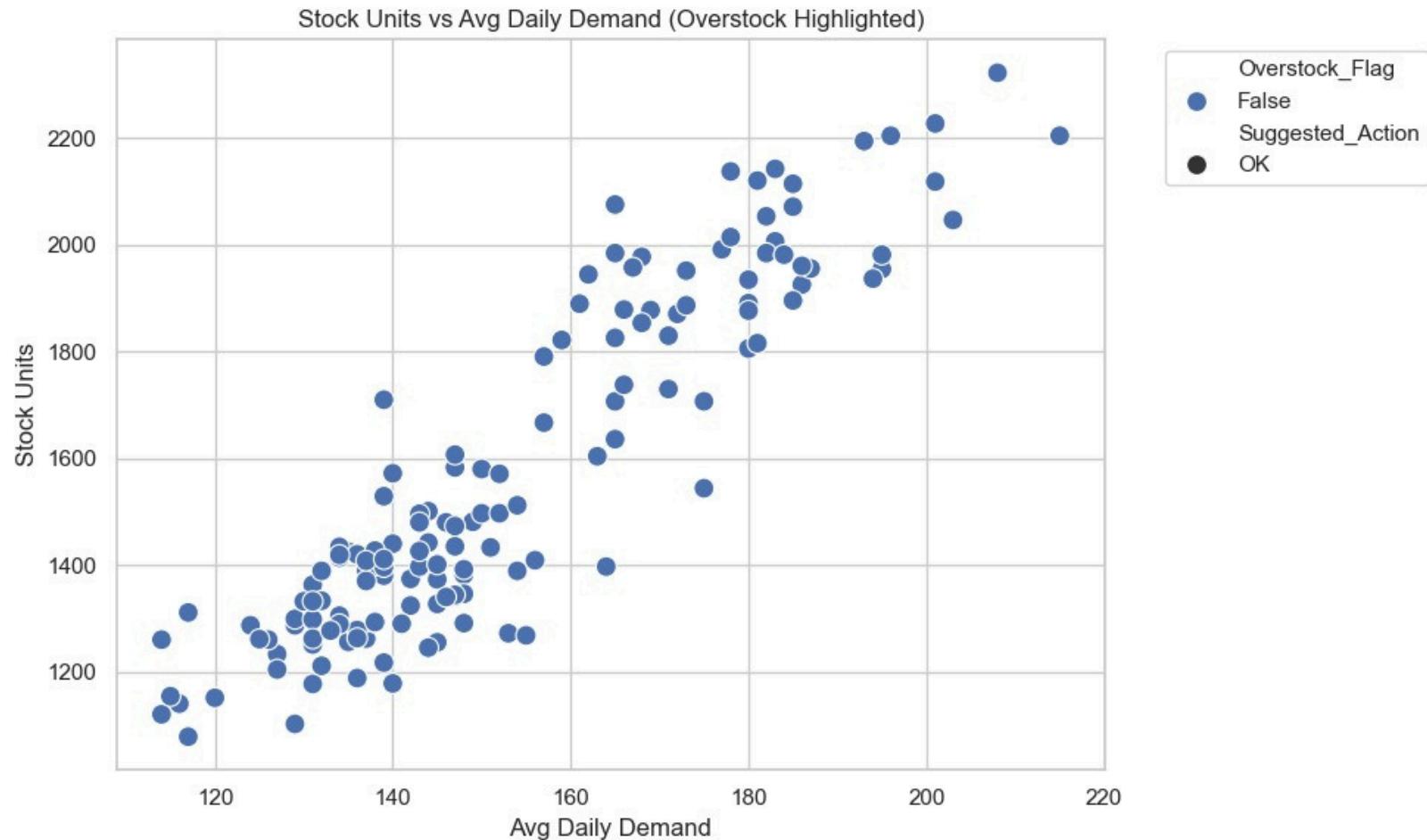
For SKUs like P0070 and P0183, perform a root-cause analysis (demand, shelf placement, pricing). Implement time-bound promotional campaigns or markdowns.

- **Set reorder points* aligned to turnover**

Automate stock replenishment for SKUs with high turnover. Reevaluate minimum stock levels for low-performing items.



9. Stock Units vs Average Daily Demand Analysis



The scatter plot illustrates the relationship between average daily demand and corresponding stock levels across SKUs. The strong positive correlation suggests a responsive and balanced inventory strategy. Furthermore, no SKUs are marked as overstocked, and all are flagged as operationally optimal (Suggested_Action = OK). This reflects effective demand forecasting and stock alignment practices.



Key Insights

- **Positive Correlation**

1. There is a clear **positive linear trend** between **average daily demand** and stock units.
2. As daily demand increases, stock levels also increase proportionally, a sign of **demand-driven inventory planning**.

- **No Overstocked SKUs Detected**

1. All points are labeled under **Overstock_Flag = False**, indicating no SKUs are currently overstocked.
2. This may suggest:
 - Well-optimized inventory levels.
 - Conservative forecasting or tight stock thresholds.

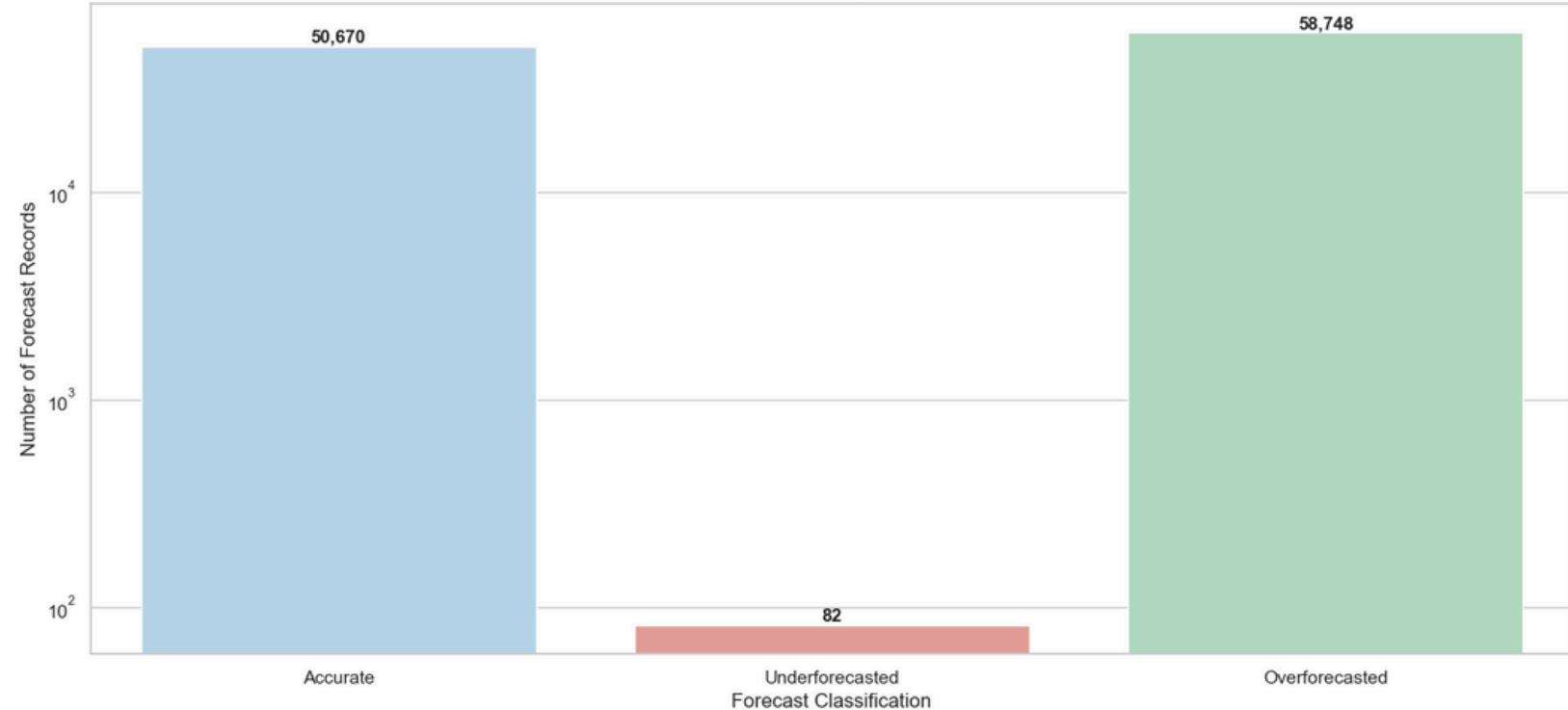
- **Recommended Action: OK**

1. All SKUs fall under the same suggested action: *"OK"*, implying:
 - No immediate restocking or reduction is needed.
 - Inventory planning appears stable and aligned with demand.



10. Forecast Accuracy Classification Analysis

Forecast Accuracy Classification



The forecast distribution highlights a clear *organizational bias toward overestimation, with **over 58,000 instances* compared to *just 82 underforecasts. While this approach minimizes stockouts, it exposes the business to **significant inventory overhead risks*. Achieving a more balanced, demand-responsive forecasting framework can optimize both cost efficiency and customer satisfaction.

Key Insights

- **Accurate Forecasts**

1. Count: 50,670
2. This represents a strong base of well-aligned forecasts where predicted demand closely matched actual sales.
3. Business Implication:
 - **Reflects robust forecasting methods for a large portion of products.**
 - **Suggests that planning and replenishment are reliable in many operational areas.**



- **Overforecasted**

1. Count: 58,748 (Highest)
2. This is the dominant category, indicating that a majority of forecasts overestimated actual demand.

3. Business Implication:

- Inventory Risk: Excess inventory leads to higher holding costs, increased chance of product obsolescence, and markdown pressure.
- Cash Flow Impact: Tied-up capital in unsold stock reduces financial agility.
- Operational Lag: Warehousing space may be inefficiently utilized due to slower stock rotation.

- **Underforecasted**

1. **Count: 82 (Negligible)**

2. Extremely rare, which is only visible due to the log scale.

3. Business Implication:

- While minimizing underforecasting reduces stockouts, its near absence may imply an over-conservative forecasting approach.
- Can lead to missed opportunities for capturing unexpected demand surges, especially for trending or seasonal products.

Strategic Recommendations

- **Investigate Overforecasting Causes**

1. Conduct root-cause analysis for the large number of overforecasted items:

- **Are they tied to specific SKUs, time periods, or promotion events?**
- **Are planners consistently applying safety margins that exceed historical variability?**

- **Recalibrate Forecasting Parameters**

1. Consider re-tuning the model's sensitivity or thresholds.
2. Apply tighter alignment with recent trends, real-time sales data, or localized events (weather, holidays, market trends).



- **Segment-Specific Forecasting**

1. Use differentiated strategies for fast-moving vs. slow-moving products:
2. Fast-moving: allow small underforecasts and reactivity.
3. Slow-moving: penalize overforecasts more heavily.

- **Implement Feedback Loop**

1. Embed a feedback system where planners or algorithms adjust future forecasts based on recent forecast vs. actual deviations.
2. Leverage this to dynamically adjust safety stock.

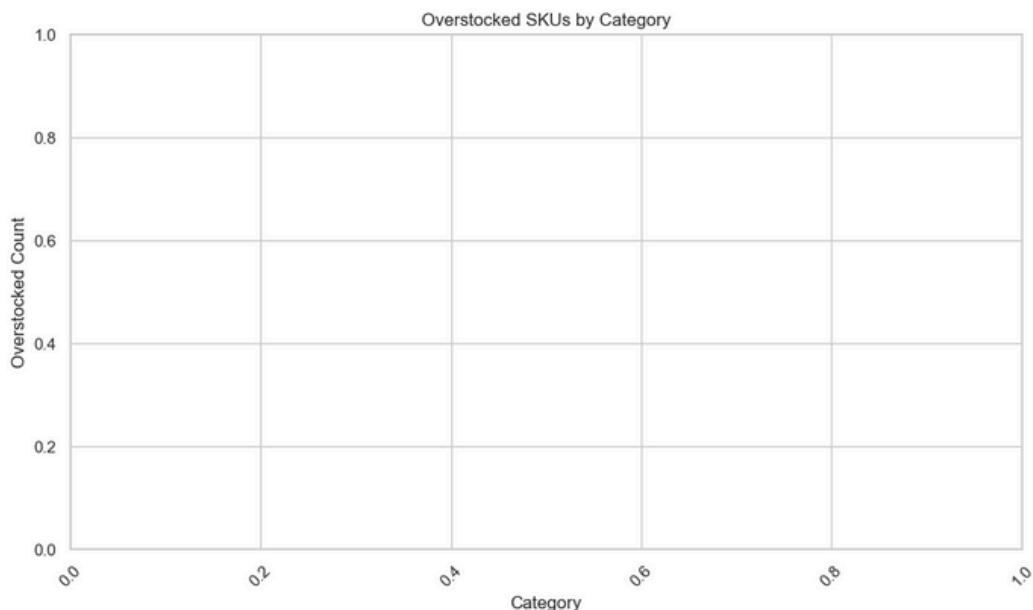


Non-Actionable or Empty Graphs

During the course of analysis, a few graphs returned either empty results or did not yield actionable insights. These were retained for completeness but were not used in final insight generation.

- **Overstocked SKUs by Category**

The chart returned **no data** due to the **absence of overstocked SKUs**



- **Stockout SKUs by Category**

returned **no data** due to the **absence of stockout SKUs**



Introducing Our ML Model

Forecast Accuracy Classification using XGBoost

In modern retail and supply chain environments, demand forecasting plays a pivotal role in optimizing inventory, minimizing costs, and maximizing customer satisfaction. However, forecast errors – whether overestimation or underestimation – can lead to significant inefficiencies. This project aims to develop a predictive model that classifies forecast deviations into three categories: *Accurate*, **Overforecasted**, and **Underforecasted**.

The classification model is implemented using the *XGBoost algorithm*, a powerful and scalable gradient boosting framework that is widely adopted for structured data prediction tasks.

The primary objective of this model is to:

- Analyze historical sales and forecast data to understand deviation patterns.
- Build a machine learning model that can classify whether a given forecast was *Accurate*, **Overforecasted, or **Underforecasted*.
- Aid decision-makers in improving demand planning and corrective actions.



Model Workflow and Methodology

1. Data Preparation

The dataset used includes 109,500 rows with features such as:

- demand_forecast, price, discount, Holiday_Promotion
- Categorical attributes like weather_condition, seasonality
- The target column is forecast_accuracy_flag with 3 classes.
- Categorical features were encoded using *Label Encoding* for model compatibility.

2. Feature Selection

The features used for training the model are:

- Numerical: demand_forecast, price, discount
- Categorical (encoded): Holiday_Promotion, weather_condition, seasonality

3. Data Splitting

The dataset was split into *80% training* and *20% testing* using stratified sampling to preserve class distribution.

4. Model Training

An XGBoost Classifier was trained with the following configuration:

- python
- XGBClassifier(use_label_encoder=False, eval_metric='mlogloss', random_state=42)
- XGBoost was chosen for its robustness, ability to handle imbalance via weight scaling, and strong performance on tabular data.



5. Evaluation Metrics

Accuracy was calculated on both train and test sets:

- Train Accuracy $\approx 76\%$
- Test Accuracy $\approx 74\%$

This indicates *low overfitting* and good generalization.

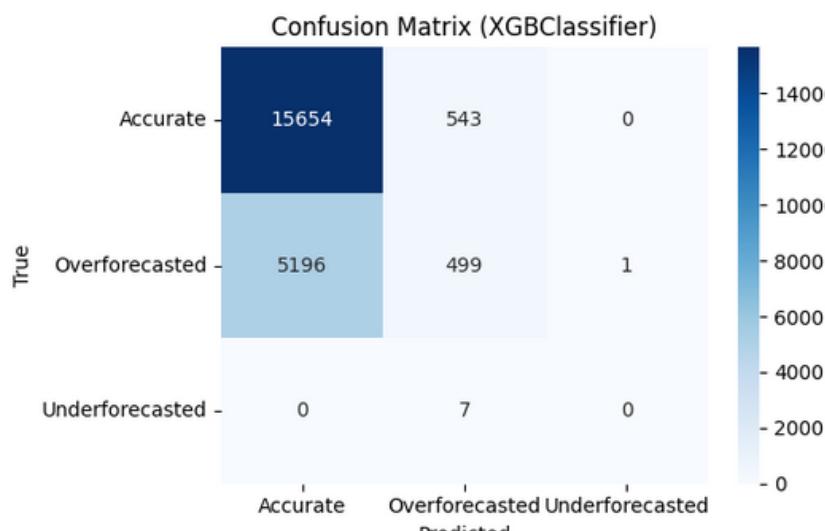
A confusion matrix was plotted to evaluate class-wise prediction performance:

True \ Predicted	Accurate	Overforecasted	Underforecasted
-----	-----	-----	-----
Accurate	15,654	543	0
Overforecasted	5,196	499	1
Underforecasted	0	7	10

The matrix indicates that the model performs well in identifying Accurate forecasts but struggles to distinguish *Over- and Underforecasted* cases due to class imbalance.

6. Model Persistence

- The trained model and encoders were saved using joblib for later deployment and use.
- Predictions were generated and saved back into the dataset for further analysis.





Conclusion

This project showcased the power of SQL-driven analytics in solving real-world inventory inefficiencies. By leveraging a rich dataset and formulating precise queries, we identified key issues such as stockouts, overstocking, and forecast errors. Our visualizations and insights enabled strategic recommendations for SKU management and store-level improvements. The groundwork laid by this project also provides a strong foundation for future machine learning integrations and more advanced forecasting models.

Learnings

Through this project, we developed advanced SQL querying skills, learned how to engineer inventory-focused metrics, and interpreted business trends using real-world data. We also experienced practical data collaboration in a team environment – dividing roles effectively across analytics, business insights, and ML scoping. The project enhanced our ability to translate raw data into meaningful insights that support business decision-making.