Inventory Management Analysis: Addressing Deadstock and Stock Discrepancies

Final Submission for the BDM Capstone Project

Submitted by

Name:

ARVIND M

Roll number:

23f2005701

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

1. Executive Summary 3

2. Explanation of Analysis Process 4

3. Results and Findings 7

4. Interpretation and Recommendation 17

1 Executive Summary

The project deals with Jothi Polymers Private Limited which manages Business-to-Business activities related to engineering polymers.

The major problem that downgraded the company's operational efficiency, as stated earlier was deadstock. The cleaned data was used for several analysis techniques. One of them was ABC analysis which tells us about distribution of inventory into three major categories. This helps in prioritizing the resources. As deadstock is time series based, ARIMA analysis was used to forecast the future accumulation based on historical pattern. Next, RFM analysis was done to examine the inventory based on recency, frequency and monetary value. This was helpful in identifying the risk factor. Finally, K-means clustering, a popular unsupervised ML algorithm was employed for revealing unexpected patterns in how items become deadstock analyzing multiple factors simultaneously.

The second problem faced by the company was stock mismatches. For this, two techniques were adopted. First was Decision Tree Analysis, which is a supervised ML model. It was helpful in identifying which reason significantly contributed to mismatches wherein identifying threshold values. Secondly, Association Rule Mining was used to discover hidden relationships that occur together in stock mismatches. It was efficient in identifying multi-item relationships that led to stock problems.

This project helped in identifying which stock type becomes obsolete also considerably identified the financial impact created by them. Also, the analysis methods used were helpful in drawing insights and provide recommendations to the company. It also helped in bringing out suggestions to avoid stock mismatches by discovering the underlying cause which is discussed further.

Python libraries like pandas, NumPy, statsmodels, matplotlib, scikit-learn, seaborn and mlxtend were primarily used to build these analysis models. Data present in Excel was easier and convenient in drawing insights and for processing.

This project as a whole, was useful in identifying the cause of the problems and helpful in examining how to avoid them in the future. This also can have a positive impact on the company's profits and further boost their market position.

Link: dataset

2 Explanation of Analysis Process

In this analysis, a hybrid mechanism was adopted. Several methods were used to analyze and draw insights from the data. Based on the problem, particular strategy was adopted. For example, for deadstock ARIMA analysis was used. These methods uncovered causes, financial impact, forecasting patterns etc. Prior to this, data cleaning was already done to avoid anomalies and abnormalities in data. For building these models, necessary python libraries were installed and imported. Data used for deadstock analysis consisted of 9 columns and 775 rows. Stock mismatch data consisted of 11 columns and 300 rows. The analysis methods and processes involved are discussed below:

2.1. ABC Analysis (Deadstock):

This analysis' aim was to distribute the inventory value and item count across three categories A, B and C. By doing so, value concentration, item distribution and resource allocation becomes easier.

Steps:

1. Total value for each inventory item was calculated as Quantity x Price.
2. All items were sorted in descending order of their value
3. Cumulative value and cumulative percentage of total inventory was calculated
4. Then classification was done according to 80-15-5 rule in which items were classified into

(i). **Category A:** Items that contribute to 80% of total value.

(ii). **Category B:** Items that contribute to next 15% value.

(iii). **Category C:** Items that contribute to 5% value.

5. Then a pie chart showing distribution of items across different categories was constructed.

2.2. ARIMA Time Series Analysis (Deadstock):

It accounted for finding stationarity of data. It forecasted deadstock pattern for the subsequent months. It was useful in identifying short and long-term fluctuations.

Steps:

1. Plotted monthly deadstock value to visualize pattern and trends.
2. Test for stationarity was done using Augmented Dickey-Fuller test.
3. ACF and PACF plots were generated to obtain p and q values.
4. Optional parameter was selected based on above results.
5. Fit this ARIMA model to historical deadstock data.
6. Validated the model by checking residuals.

7. With this, future deadstock values for 3 months were forecasted.
8. Visualization of historical and forecasted value in time series plot was done.

2.3. **RFM Analysis (Deadstock):**

It provided multi-dimensional risk assessment by combining movement Recency, Frequency and Monetary value. With this high-value and slow-moving inventory with respect to finance was identified. It also identified inventory requiring immediate action.

Steps:

1. Stock quantities, movement dates, and values were extracted.
2. Calculated recency metrics based on days since last movement.
3. Frequency metric was determined by counting transaction per item.
4. Monetary value computed using item values.
5. Created a scoring system for each R, F and M dimension.
6. Higher weights for recency and monetary value were assigned
7. Combined RFM score was calculated and segment inventory into risk categories.
8. Across the segments, distribution of inventory value was analyzed.
9. Finally, percentage of items and value in each risk segment was visualized.

2.4. **Decision Tree Analysis (Stock mismatch):**

This analysis provided clarity on which factors led to mismatches. It also captured non-linear relationships among variables. It tells about relative importance of each factor in quantifiable form.

Steps:

1. Opening stock, Rate and Closing Stock were taken as necessary features.
2. Mismatch status was converted to binary form (1 for mismatch,0 for no mismatch).
3. Dataset was then split into test and train set.
4. A decision tree classifier model was trained and was evaluated.
5. Then the model calculated the importance of each feature.
6. Each feature importance was visualized as bar chart.
7. A decision tree was visualized showing all decision nodes and leaf nodes with decision paths based on the above three features.

2.5. **K-means Clustering Analysis (Deadstock):**

This was used to understand supplier-specific behavior for each inventory. It segmented inventory into distinct clusters based on multiple characteristics.

Steps:

1. Key features: Quantity, Rate and Days since last dispatch were selected for clustering.
2. Features were standardized using StandardScaler to ensure uniformity.
3. Elbow method was used to identify optimal number of clusters by plotting Within-Cluster Sum of Squares (WCSS) against different number of clusters (1-10) and based on this 4 was selected as optimum.
4. K-means with clusters=4 was applied to inventory items.
5. Each inventory was assigned to a particular cluster based on this.
6. Statistic values for each cluster were calculated.
7. Finally, a visualization was created highlighting distribution of clusters across top 10 suppliers.

2.6. **Association Rule Mining with Apriori Algorithm (Stock mismatch):**

It was used to identify non-obvious relation between product and mismatch reason. It was also helpful in quantifying them in metrics like support, confidence and lift. It was also useful in deriving actionable insights.

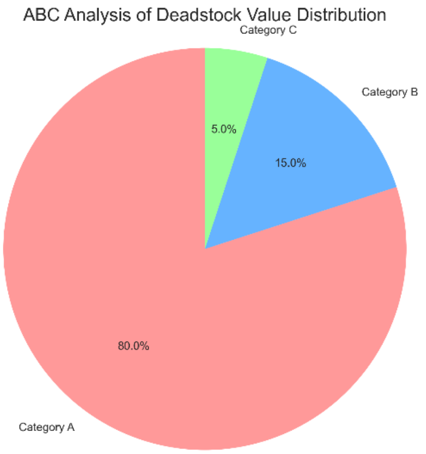
Steps:

1. A binary transaction matrix was created where each row represents transaction and column represent items.
2. Support threshold was set as 0.5%.
3. Apriori algorithm was applied to identify frequent item sets that met minimum support threshold.
4. Association rules were created from frequent item sets and support, confidence and lift metrics were calculated for each rule.
5. Only meaningful rules were retained.
6. Sorted rules by lift value to show the strongest associations.
7. A scatter plot of support vs confidence was plotted along with a network graph showing connections.
8. Finally, a pattern was identified revealing specific product consistently associated with specific reasons.

3 Results and Findings

After implementing the above methods for analysis, the following results and insights were obtained. All these methods were executed in python in notebook format. As discussed earlier necessary libraries such as pandas, NumPy, sklearn, statsmodel etc. were used to build the models. The graphs presented are visualized using matplotlib and seaborn. Data present in excel was imported to the python notebook.

3.1. ABC Analysis



This graph shows the distribution of item and inventory count across three categories.

Findings:

Category A:

1. Contains 304 items comprising 39.28% of the total inventory count which accounts for around 80% total deadstock value.
2. Follows Pareto principle, where small percent of items account for larger value.
3. These are high deadstock value items that require immediate actions.

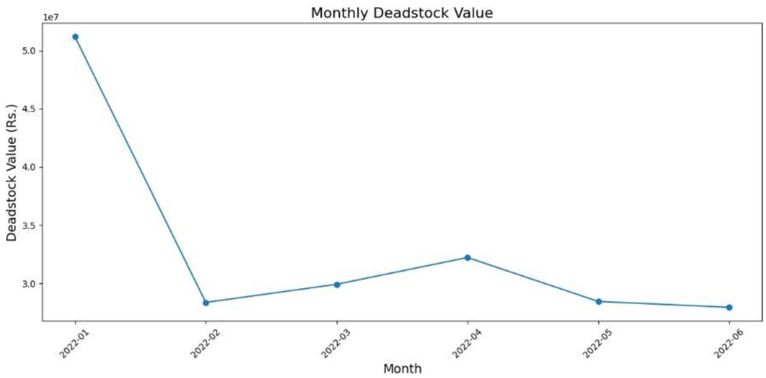
Category B:

1. Contains 175 items comprising 22.61% of the total inventory count accounting for 15% of total value.
2. These are mid-value items that needs regular monitoring but not as critical as category A.

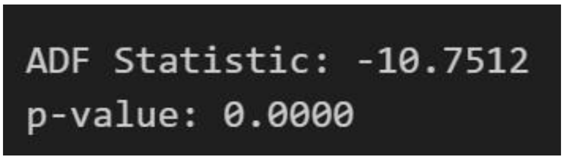
Category C:

1. Contains 203 items comprising 26.23% of the total inventory count accounting for 5.03% of total value.
2. These are low value items and contribute less to deadstock accumulation.

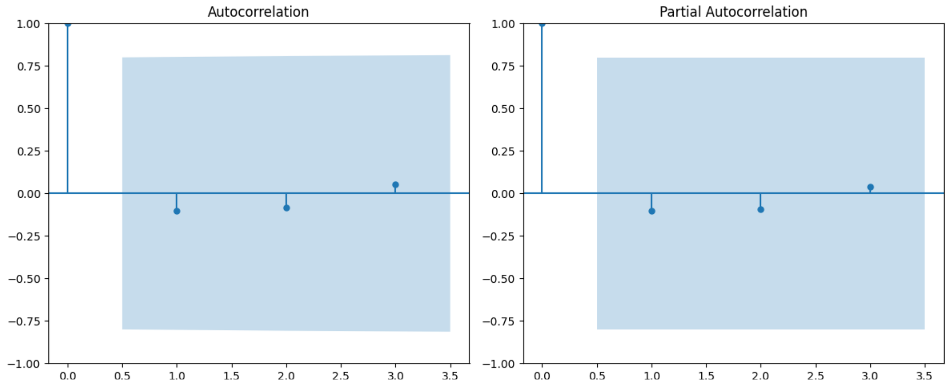
3.2 **ARIMA Time Series Analysis:**



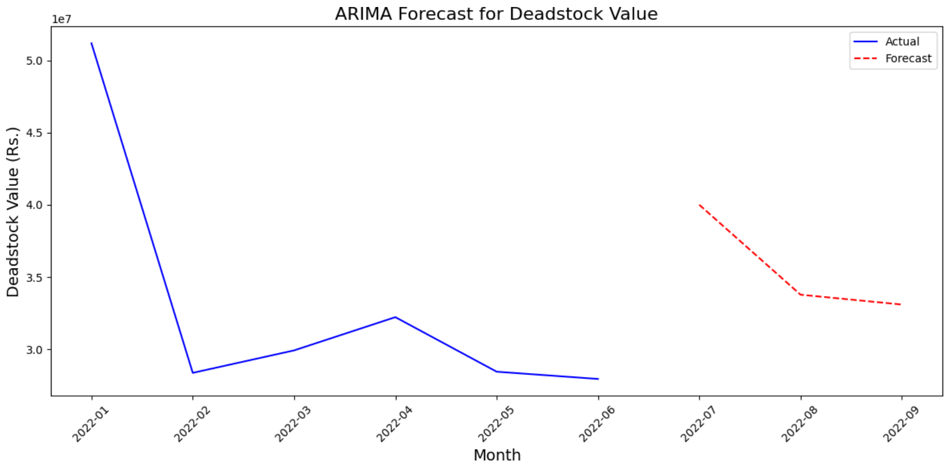
The above plot shows a sharp decrease from January 2022 (approximately ₹52 million) to February 2022 (approximately ₹28 million), which is followed by a relatively stable value around ₹28-32 million till June 2022.



The ADF statistic of -10.7512 with a p-value of 0.0000. This strongly rejects the null hypothesis of non-stationarity and thus concludes that the time series is stationary, which is a prerequisite for ARIMA modeling.



The above plots show minimal correlation at all lags (except lag 0, which is always 1), with values staying within the confidence intervals (blue shaded areas). This suggests the series has no significant autocorrelation patterns.



The red dotted line shows forecasted value for months from July to September and follows a decent pattern from the historical pattern data.

The forecasted values for these months are as follows:

```
Forecasted deadstock for 2022-07: Rs.40,001,654.26
Forecasted deadstock for 2022-08: Rs.33,778,181.94
Forecasted deadstock for 2022-09: Rs.33,098,998.98
```

3.3 **RFM Analysis:**

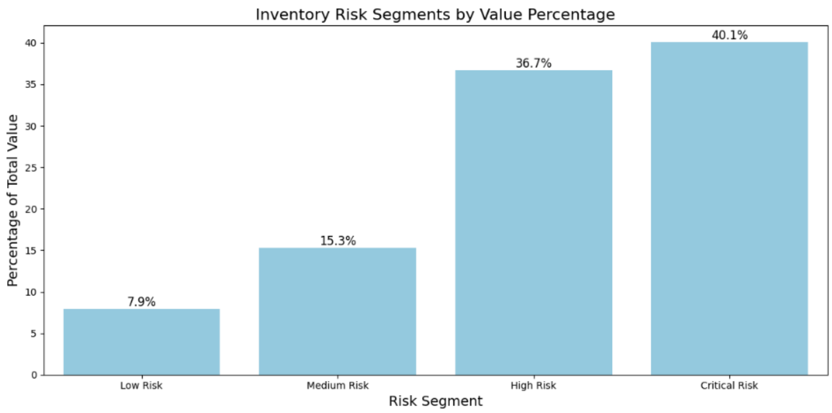
RFM Segmentation Results:

Segment	Count	Total Value (Rs.)	Avg Days Since Movement
Low Risk	227	15687750.46	711.022026
Medium Risk	164	30315203.32	548.073171
High Risk	204	72713703.06	448.024510
Critical Risk	179	79378134.35	282.072626

Segment	Percentage of Items	Percentage of Value
Low Risk	29.328165	7.919315
Medium Risk	21.188630	15.303382
High Risk	26.356589	36.706519
Critical Risk	23.126615	40.070783

The inventory has been classified into four segments namely Low, Medium, High and Critical Risk based on the RFM scoring.

The above breakdown includes count, total value, average days since movement, percentage of items, and percentage of value for each segment, providing multiple statistic values for analysis.



This bar chart shows percentage of deadstock value with respect to each risk segment. It can be inferred that specific concentration is needed for high-risk segment as it accounts for 36.7% and is close to critical risk segment.

Top 10 Highest Risk Items (Critical Risk):

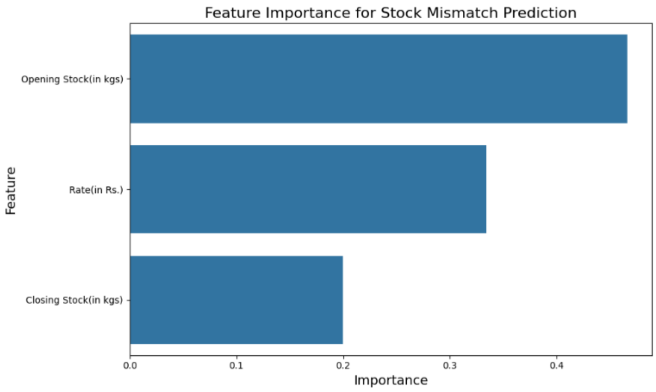
	Grade	Recency	Monetary	RFM_Score
688	M.2407C 651566 BK150W	187	361200.00	4.0
641	ULTRAMID® A218 V35 BLACK 21N	184	442520.00	4.0
762	ULTRAMID A218 V35 NATURAL	141	392968.65	4.0
767	ULTRAMID® A218 V30 BLACK 21NS	158	466375.00	4.0
639	ULTRAMID® A218 V30 BLACK 34NG-K	124	349575.00	4.0
635	ULTRAMID A 402 H1 NATURAL	253	1031355.00	4.0
621	OMNIX 4050 BK 000 S25 P1000	226	865117.00	4.0
660	B. T85XF 900307 BBS910 BK 150W	176	519750.00	4.0
610	NGC 3001 NATURAL	124	327525.00	4.0
389	B. FR 3010 500018 BK 150W	73	396930.00	3.8

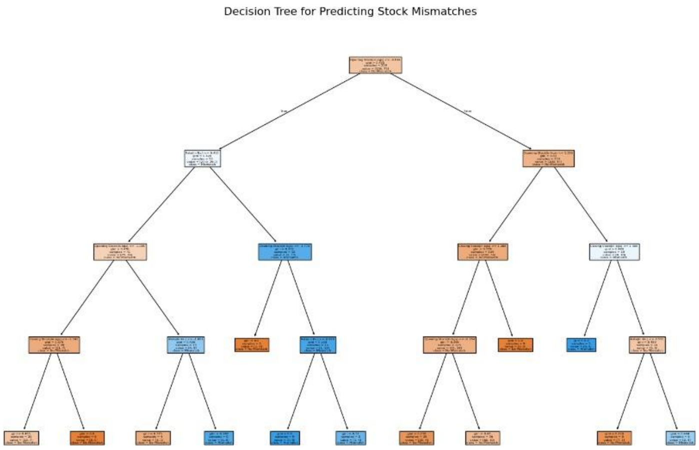
3.4 **Decision Tree Analysis:**

Feature Importance for Mismatch Prediction:

	Feature	Importance
1	Opening Stock(in kgs)	0.465899
2	Rate(in Rs.)	0.334290
0	Closing Stock(in kgs)	0.199811

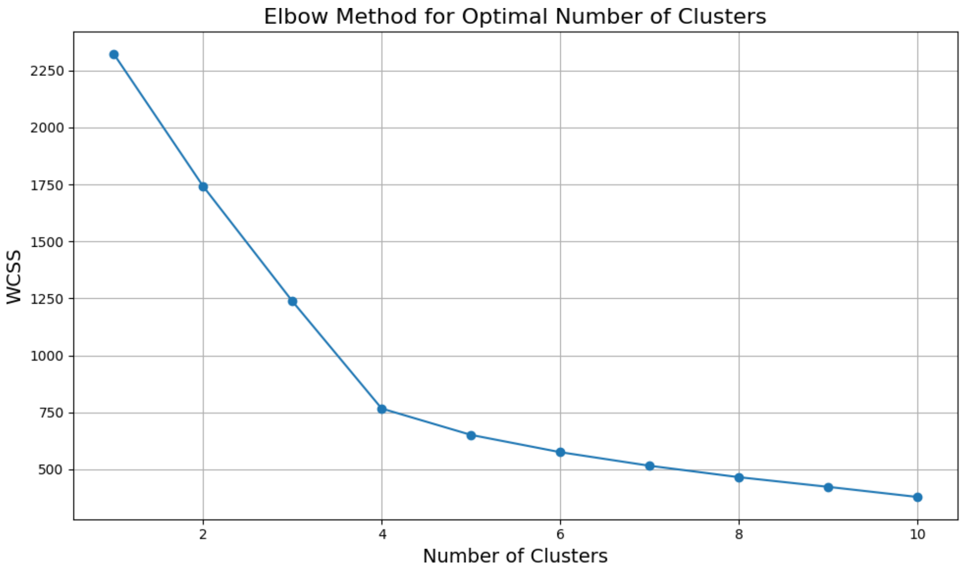
The above distribution shows that opening stock is important in predicting mismatches, followed by rate with closing stock being the least critical.





The tree's root node is based on opening stock indicating it is the most important factor. The tree has different branching paths with leaf color orange for mismatch and blue for no mismatch. Tree has depth of 4, indicating a balanced approach. It can also be noted that tree has more orange nodes implying mismatch prediction on pathways with higher opening stock values.

3.5 **K-means Clustering:**



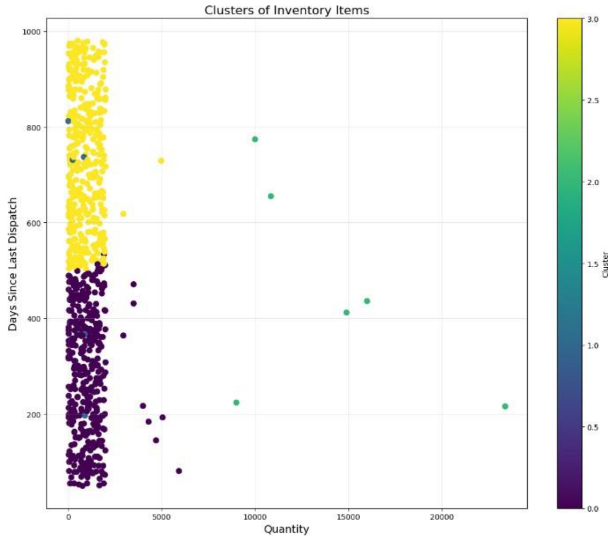
The elbow method suggests 4 is the optimal cluster choice as their significant reduction in WCSS till here post which reduction becomes gradual.

Cluster Analysis:

Cluster	Count	Value(in Rs.)	Quantity	Rate(in Rs.)
0	382	84684941.73	1003.654450	215.248272
1	6	6909540.00	551.000000	2090.000000
2	6	24551965.11	14025.000000	299.645000
3	380	81948344.35	881.434211	239.708605

Cluster	Days Since Last Dispatch	Percentage of Items	Percentage of Value
0	284.057592	49.354005	42.749706
1	535.500000	0.775194	3.487997
2	452.833333	0.775194	12.394049
3	733.510526	49.095607	41.368248

This analysis segmented the inventory into 4 different clusters each with unique characteristics in terms of quantity, value and days since last dispatch.



The scatter plot of inventory items clearly shows the separation between clusters, particularly in terms of quantity and days since last dispatch.

1. **Cluster 0 (High Volume, Medium Age):**

   i. Has 49.35% of items and 42.75% of total inventory value.

   ii. Clustered by high quantity (avg. 1003.65 units) and medium age (284 days since last dispatch).

   iii. Represents the bulk of the inventory in both count and value.

2. **Cluster 1 (Low Volume, High Value, Medium Age):**

   i. Only 0.78% of items but 3.49% of total value.

   ii. Very high average rate (Rs. 2090) but low quantity (551 units).

   iii. Represents high-value, specialized items with moderate turnover (535 days since last dispatch).

3. **Cluster 2 (Very High Volume, High Value, Medium-High Age):**

   i. Another 0.78% of items but a significant 12.39% of total value.

   ii. Extremely high average quantity (14,025 units) with medium rate (Rs. 299.65).

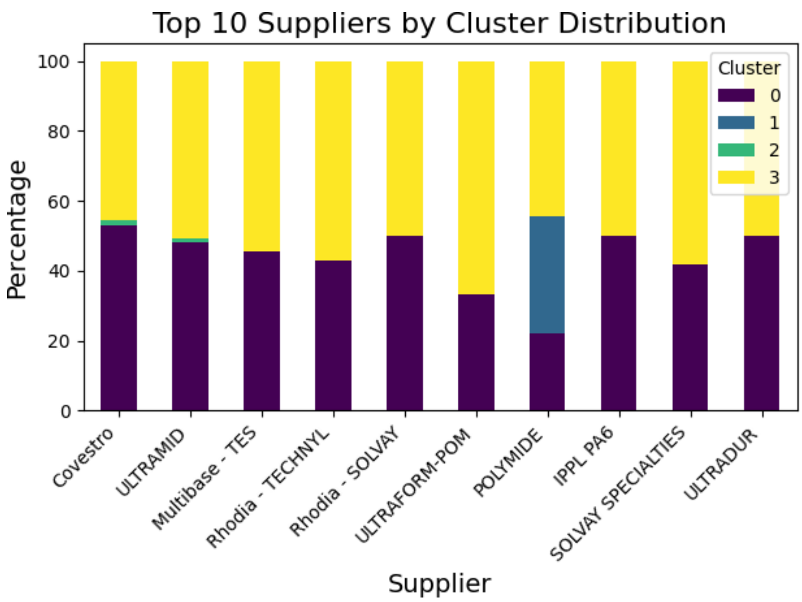
   iii. Represents bulk items with slower turnover (452 days since last dispatch).

4. **Cluster 3 (Medium Volume, High Age):**

i. 49.10% of items and 41.37% of total value.

ii. Medium quantity (881 units) but very high age (733 days since last dispatch).

iii. Represents potential deadstock or very slow-moving items.



3.6 **Association Rule Mining with Apriori Algorithm:**

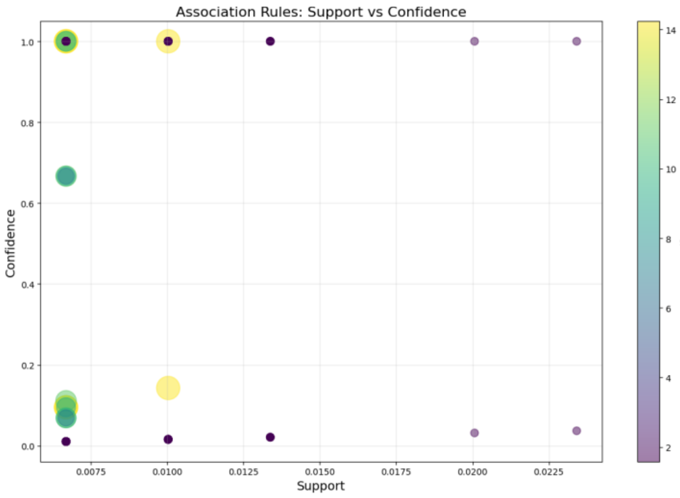
```
APRIORI ASSOCIATION RULE MINING FOR INVENTORY PATTERNS
Minimum support threshold: 0.005
Found 142 frequent itemsets.
Found 122 association rules.
```

The Apriori algorithm successfully identified 142 frequent itemset and 122 association rules with a minimum support threshold of 0.005, indicating that meaningful patterns exist in the inventory data.

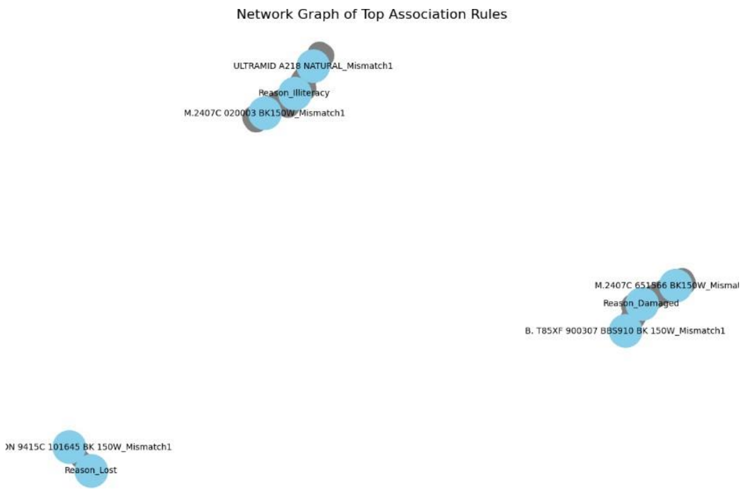
Top 10 Association Rules:

	antecedents	consequents	support	confidence	lift
0	(B. T85XF 900307 BBS910 BK 150W_Mismatch1)	(Reason_Damaged)	0.006689	1.000000	14.238095
1	(Reason_Damaged)	(B. T85XF 900307 BBS910 BK 150W_Mismatch1)	0.006689	0.095238	14.238095
45	(M.2407C 651566 BK150W_Mismatch1)	(Reason_Damaged)	0.010033	1.000000	14.238095
44	(Reason_Damaged)	(M.2407C 651566 BK150W_Mismatch1)	0.010033	0.142857	14.238095
28	(M.2407C 020003 BK150W_Mismatch1)	(Reason_Illiteracy)	0.006689	1.000000	14.238095
29	(Reason_Illiteracy)	(M.2407C 020003 BK150W_Mismatch1)	0.006689	0.095238	14.238095
107	(Reason_Illiteracy)	(ULTRAMID A218 NATURAL_Mismatch1)	0.006689	0.095238	14.238095
106	(ULTRAMID A218 NATURAL_Mismatch1)	(Reason_Illiteracy)	0.006689	1.000000	14.238095
71	(MAKROLON 9415C 101645 BK 150W_Mismatch1)	(Reason_Lost)	0.006689	0.666667	11.074074
70	(Reason_Lost)	(MAKROLON 9415C 101645 BK 150W_Mismatch1)	0.006689	0.111111	11.074074

The top rules show extremely high lift values (14.24 for many rules), indicating very strong associations between specific product mismatches and their reasons.



This plot effectively illustrates the distribution of rules, with point size and color representing lift values.



This graph illustrates connections between specific products and mismatch reasons.

4. Interpretation of Results and Recommendation:

4.1 ABC Analysis:

Interpretation:

i. It is found that relatively small number of items account for larger deadstock value.

ii. There is even distribution of item counts across categories and category A has most items.

iii. This shows that resources need to allocated with efforts towards category A items.

Recommendation:

1. **Category A Focus:**

i. Conduct weekly reviews of these items.

ii. Consider implementing just-in-time inventory practices for these products.

2. **Category B Management:**

i. Develop moderate control measures for Category B items:

ii. Implement cycle counting procedures.

3. **Category C Efficiency:**

i. Implement automated reordering systems

ii. Consider outsourcing management of these items or exploring vendor-managed inventory options.

4.2 **ARIMA Time Series Analysis:**

Interpretation:

i. There was sharp decrease in deadstock value between January and February and from February to June there was stabilization period.

ii. The model predicts initial increase in July similar to pattern observed in January and decrease in subsequent months.

iii. Through the ACF and PACF plots no seasonal patterns were observed.

Recommendation:

i. **Prepare for July Increase:** Plan resources for potential increased deadstock in July 2022 (approximately ₹40 million) as forecasted by the model.**Implement Preventive Measures:** To overcome the forecasted increase, schedule targeted inventory reduction initiatives in June and July, such as special sales promotions or return- to-vendor programs for aging inventory.

ii. **Investigate January Spike**: Conduct a detailed analysis of what caused the January 2022 spike (₹52 million) to determine if it took place only once or if similar conditions might occur in future.

4.3 **RFM Analysis:**

Interpretation:

i. Although high and critical risk segment represent 49.48% of items, they account for 76.78% of the inventory value.

ii. Critical risk items represent only 23.13% items but account for around 40% inventory with being static for an average of 282 days.

iii. Low risk represents 29.33% items but account for only 7.92% inventory value showing efficient movement.

iv. There exists an inverse relation between risk level and days since last movement.

v. The top 10 critical risk items identified include may ULTRAMID products, with individual values ranging from ₹327,525 to over ₹1 million. ULTRAMID A 402 H1 NATURAL has a value of ₹1,031,355, indicating a serious deadstock concern.

Recommendation:

i. Immediate Action for Critical Risk Items: Devise a 60-day action plan targeting the top 10 Critical Risk items, particularly ULTRAMID A 402 H1 NATURAL (₹1,031,355) and OMNIX 4050 BK 000 S25 P1000 (₹865,117).

ii. Segment-Specific Inventory Policies: Implement different inventory management policies for each risk segment:

   a. Critical Risk: Weekly review, aggressive price discounts, consider write-offs for items over 1 year old

   b. High Risk: Bi-weekly review, targeted sales promotions, reduced reorder quantities

   c. Medium Risk: Monthly review, regular pricing strategies

   d. Low Risk: Quarterly review, maintain current management approach

iii. ULTRAMID Product Line Review: Conduct a comprehensive review of the ULTRAMID product line, which appears frequently in the Critical Risk segment, to identify underlying demand issues or opportunities for product consolidation.

4.4 **Decision Tree Analysis:**

Interpretation:

i. Opening stock is the most dominant predictor (46.59% importance).

ii. Rate also has significant importance leading to mismatches.

iii. Trees's multiple branches shows that there is complex relationship and is not linear.

Recommendation:

i. Tiered Verification Protocols: Implement graduated verification procedures based on opening stock quantities, with more rigorous checks for quantities above the thresholds identified in the decision tree splits.

ii. Value-Based Control Measures: Develop specialized handling procedures for items with rates above certain thresholds, as identified in the Rate-based splits in the decision tree.

iii. Automated Flagging System: Create an automated system that flags transactions matching the high-risk patterns identified in the orange-colored (mismatch) terminal nodes of the decision tree.

4.5 **K-means Cluster Analysis:**

Interpretation:

i. Majority of suppliers have inventory across different clusters

ii. Certain suppliers like ULTRAMID, ULTRAFORM-POM have a higher proportion in Cluster 3, indicating more slow-moving inventory.

iii. Suppliers like POLYMIDE and SOLVAY SPECIALTIES show more balanced distribution across clusters.

Recommendation:

1. **Targeted Inventory Reduction for Cluster 3:** Implement aggressive markdown strategies or consider liquidation for items in Cluster 3, which represents nearly half of all items but has extremely low turnover.

2. **High-Value Item Management (Cluster 1):** Develop specialized handling and forecasting procedures for the high-value, low-volume items in Cluster 1 to ensure optimal stock levels and prevent overstocking.

3. **Bulk Item Optimization (Cluster 2):** Review ordering and storage strategies for the high-volume items in Cluster 2. Consider negotiating better terms with suppliers or implementing just-in-time inventory practices to reduce holding costs.

4.6 **Association Rule Mining with Apriori Algorithm:**

Interpretation:

1. **Product-Specific Mismatch Patterns:**

   a. B. T85XF 900307 BBS910 BK 150W has a perfect (100%) association with damaged goods issues

   b. M.2407C 651566 BK150W has a perfect association with damaged goods issues

   c. M.2407C 020003 BK150W has a perfect association with illiteracy issues

   d. ULTRAMID A218 NATURAL has a perfect association with illiteracy issues

   e. MAKROLON 9415C 101645 BK 150W has a strong (66.7%) association with lost items.

2. **Bidirectional Relationships:**

   a. The rules show high confidence in both directions for many product-reason pairs.

   b. For example, when B. T85XF 900307 BBS910 BK 150W has a mismatch, it's always (100%) due to damage

   c. Conversely, 9.52% of all damage-related mismatches involve this specific product.

3. **Reason-Specific Concentrations:**

   a. Illiteracy issues are strongly associated with specific products (ULTRAMID A218 NATURAL and M.2407C 020003 BK150W)

   b. Damage issues are strongly associated with other specific products (B. T85XF 900307 BBS910 BK 150W and M.2407C 651566 BK150W)

   c. Lost items are strongly associated with MAKROLON 9415C 101645 BK 150W

4. **Support Levels:**

a. Individual rules have relatively low support (0.67% to 1.00%), indicating these are specific patterns rather than widespread issues

b. However, the extremely high lift values (11.07 to 14.24) confirm these arestatistically significant and non-random associations.

Recommendation:

1. **Product-Specific Handling Procedures**: Implement specialized handling protocols for the identified high-risk products:

a. B. T85XF 900307 BBS910 BK 150W: Enhanced packaging and handling to prevent damage

b. M.2407C 651566 BK150W: Improved protective measures during transport and storage

c. ULTRAMID A218 NATURAL: Simplified documentation and pictorial guides for warehouse staff.