Identifying Inventory Scrap Reduction Opportunities

Felix Morales[a], Lia Cappellari[a], Miteb Aloqab[a]

[a] Industrial & Systems Engineering, Shiley Marcos School of Engineering, University of San Diego

***Abstract*— This report presents a comprehensive solution to address Thermo Fisher Scientific Inc.'s significant scrap costs associated with their diverse stock keeping units (SKUs). The project aims to lower Thermo's annual scrap costs valued at tens of millions of dollars through effective scrap management strategies. Initial approaches involve identifying SKUs with high scrap value and quantity, analyzing their usage patterns, and exploring variable influences. To streamline the complexity of scrap management, a final strategy utilizing K-Means clustering is proposed. This technique groups SKUs based on shared characteristics to identify clusters with similar scrap quantities, while outliers are determined using statistical methods. A prioritization approach is employed, creating a table of top outlier SKUs with high opportunities for scrap reduction. Financial opportunity, calculated by multiplying unit cost value by the delta (the difference between outlier and non-outlier scrap quantities), quantifies potential savings. A conservative approach is adopted to address skewed distributions and reduce the impact of outliers on Thermo Fisher. The report suggests a future tool for automated screening, allowing dynamic cluster creation and customization options. This tool aims to save time and resources by replacing manual clustering methods. Insights into potential savings through various percentages of scrap reduction within the top 10 outlying SKUs in each cluster are provided. For instance, a 15% reduction in scrap for these SKUs would save Thermo Fisher $1.1 million.**

Keywords:
Scrap
Inventory
Data Analytics

I. INTRODUCTION

THE prominent supplier in the field of scientific instrumentation, Thermo Fisher Scientific Inc., offers a wide range of products including scientific instruments, reagents, consumables, and software services. With a diverse and expansive product portfolio consisting of tens of millions of unique stock keeping units (SKUs), Thermo Fisher faces a substantial challenge regarding the generation of significant scrap, resulting in immense costs that amount to tens of millions of dollars annually. This technical report aims to address this challenge by proposing a comprehensive solution to reduce scrap costs and optimize Thermo Fisher's operations.

Received May 20, 2023.

The problem at hand highlights the overlooked cost and impact of scrap within Thermo Fisher's inventory management. As SKUs are ordered in large quantities by customers, the production of large batches is necessary to meet demand. However, the economic worth of certain SKUs diminishes over time, rendering them no longer cost-effective to hold in inventory. Consequently, these products are discarded as scrap, contributing to the financial burden and inefficiencies faced by Thermo Fisher.

Previous research in the field has shed light on the significance of scrap management within supply chain operations. Studies have emphasized the detrimental effects of high scrap quantities and costs on company profitability, inventory control, and overall operational efficiency. However, the unique context of Thermo Fisher's extensive SKU portfolio necessitates a tailored approach to address the specific challenges and opportunities present within the organization.

To effectively tackle the problem of excessive scrap costs, this report builds upon previous research while presenting innovative methodologies and approaches. By analyzing historical data encompassing variables such as SKU types, usage patterns, and reasoning behind scrap occurrences, the aim is to identify commonalities, trends, and potential areas for improvement. The utilization of quantitative clustering algorithms, specifically K-Means clustering, offers a robust framework to group SKUs based on shared characteristics and identify outliers with high scrap quantities.

The hypothesis underlying this report is that by implementing a clustering-based approach and prioritizing efforts based on financial opportunity, Thermo Fisher can significantly reduce its overall scrap costs. By targeting outlier SKUs with the highest potential for cost reduction, substantial savings can be realized, leading to improved financial performance and operational efficiency for the organization.
The subsequent sections of this technical report will delve into the methodology employed, the analysis of findings, and the proposed strategies to streamline scrap complexity. By leveraging innovative techniques and automation tools, Thermo Fisher can revolutionize its scrap management processes and mitigate the financial and operational impact of excessive scrap within its supply chain.

II. METHODOLOGY

This study utilized historical data provided by Thermo Fisher to develop various strategies for analyzing scrap data. The dataset consisted of gigabytes of information, including scrap data, usage data, and inventory data. To handle the large dataset, SQL was employed for data manipulation and analysis.

Initially, the goal was to identify specific SKUs with high scrap values. R scripting and data management tools, such as SQL, were used for this purpose. The data was imported into RStudio, and libraries such as RPostgres, DBI, dplyr, and tidyr were utilized to analyze SKUs with the highest display value sums. This analysis provided insights into the frequency and quantity of scrap occurrences, the year of scrap, reasons for scrap (e.g., forecasting), and additional characteristics like branch code and fiscal year. Notably, SKUs such as 100103376 and A48105 exhibited scrap values exceeding 1.4 million (Figure 1).

Another initial approach involved creating tables and ordering SKUs based on scrap quantity rather than value. This enabled the identification of SKUs with high scrap volumes, providing an understanding of the quantity of a specific SKU being scrapped. For example, SKUs such as MT36814, MT34080, and 100094872 displayed scrap quantities exceeding 50 million (Figure 2).

To overcome the limitations of the first two initial strategies, we explored the relationship between scrapped SKUs and their usage to determine proportionality. It was observed that although highly scrapped items were identified, their scrap values were relatively low compared to other SKUs. To enhance the analysis, a sequence of filtering steps was implemented.

First however, to analyze the reasoning behind scrap, the 185 unique decisions recorded for scrap were transformed into a manageable number of scrap reasons. Using Excel, the 185 reasons for scrap were extracted and manually grouped into 13 categories based on similarity. This grouping facilitated more meaningful analysis. After creating the excel and turning it into a csv file. We were able to input the csv into RStudio and then use the “Mutate” function to give each of our SKUs in the data frame a reason for being scrapped. This would be later useful after a sequence of filter as follows:

The data was filtered by specific product group codes to focus on consistently scrapped product groups over the past five years. We considered and chose subsets of a SKUs variable such as branch code US03 (The Carlsbad location), “C” SKUs in terms of sales rank, and the “Z-Stranger” group of the WXYZ variable. This filtering process aimed to identify specific opportunities for scrap reduction or potential ways to save money. By looking at the US03 branch alone, we reduced the number of observations from 385,000 to 93,000 rows. This number decreased by 67% after looking at “Z-Stranger” and “C” ranked SKUs alone. It should be noted that at this stage SKU “A33502” was removed due to its immense influence on our analysis given its scrap value of 282 million. The 30,000 remaining rows underwent further processing by looking at the top 3 reasons for scrap only (Forecast, BOM, Overbought), bringing the total number of observations to 15,000.

Following the series of filtering steps aimed at refining our focus on opportunities for scrap reduction and potential cost savings, we proceeded to generate data frames encompassing the five-year period. Subsequently, we organized the data in descending order based on scrap values. Then we joined the 5 data frames into one and displayed the data by the highest count of years a product group was scrapped from 1 to 5. After doing this, the only 3 product groups that were scrapped all 5 years were “1C7”, “ANT”, and “FCR.” This brought the data frame down to 10,000 observations (Figure 3). The analysis then focused on the FCR product group code, which consistently exhibited high scrap values compared to other product groups. The data was filtered to include only FCR-coded SKUs, resulting in a final data frame of approximately 9,000 observations.

Our goal of comparing FCR-coded SKUs' scrap and usage over the last five years was to identify opportunities to lower safety stock. The analysis also aimed to assess the on-hand inventory of FCR-coded SKUs in relation to their usage and scrap rates. However, complications arose due to the limited availability of inventory data, which only dated back to 2022.

In addition to the previously described strategies, an additional initial approach included a sequence of examining variables over time to spot trends. This strategy involved visually interpreting the scrap data and its various characteristics using the ggplot2 package. Specifically, graphs were created with scrap quantity on the y-axis and fiscal year on the x-axis, as well as scrap value on the y-axis and fiscal year on the x-axis. These graphs were further analyzed by incorporating variables such as WXYZ and reasoning for scrap.

During the interpretation of scrap quantity over time filled by WXYZ variables, a notable observation was made. The "Z-Stranger" group of SKUs exhibited a significant increase of 30 million in scrap quantity over the last five years (Figure 4). However, the scrap value for the "Z-Stranger" category decreased over the same period (Figure 5). To gain further insights, the reasons for scrap specific to the Z-Stranger category were examined. The same reasoning column used in the previous strategy was utilized to create these visuals. It was discovered that Z-Strangers were predominantly overbought, which explained the substantial increase in scrap quantity since 2018 (Figure 6). However, this finding did not account for the decrease in scrap value for Z-Strangers. Analysis of the scrap value over time filled by reasoning for scrap indicated that the overbought group of Z-Stranger SKUs experienced an increase from 2018 to 2022 (Figure 7). Thus, the conclusion drawn from this strategy was that the quantity of overbought scrap was not proportional to its value.

Finding commonalities in scrapped SKUs was challenging due to high variance in scrap quantity, SKU types, reasoning, and other variables over time. Some of our initial strategies also did not consider the relative relationship between scrap quantity and usage. Although highly scrapped items were identified, their scrap value was found to be very low. Moreover, SKUs could have multiple variants of the same variable, such as WXYZ. This made our visual interpretations over time difficult to follow.

From these initial approaches, several key takeaways did emerge, however. Firstly, there was no single reason for high scrap values and quantities, as these factors varied over time. It was also observed that certain SKUs were highly scrapped despite having minimal scrap value. Approximately 23% of scrapped SKUs had missing values (NAs) for the reasoning column. The top three reasons for scrap value were Forecasting, NULL, and Overbought, collectively accounting for over 61% of all scrap value (Figure 8). Similarly, the top three reasons for scrap quantity were Overbought, Raw Material, and Defects, representing quantities exceeding 80 million and accounting for 62% of all scrap quantity (Figure 9). Importantly, these findings shifted our focus towards identifying ways to prevent scrap by detecting abnormal behavior.

Once we realized that our initial approaches were not the best way to go about this problem, we decided to change our method and create a prioritized list of SKUs utilizing k-means clustering. K-means clustering is a machine learning algorithm that groups together data points based on similarities in their characteristics. Our hypothesis here was that if a group of SKUs belong to the same cluster, that means that they share similarities, and therefore should have similar amounts of scrap.

To create the clusters, we used five variables: branch code, sales rank code, WXYZ, usage from the previous year, and a BOM dependency number. Normalization of the data was crucial in preparing it for clustering analysis. By standardizing the values of the branch code, sales rank code, WXYZ, usage from the previous year, and the BOM dependency number, we ensured that each variable contributed equally to the clustering process. This normalization step eliminated any potential biases arising from differences in measurement scales and allowed for a fair comparison between the variables, enabling us to uncover meaningful patterns and structures within the dataset. We used the usage of the previous year because we found that there was a stronger correlation between the prior year's usage of a SKU and that SKU's scrap quantity of the current year than using the current year's usage. The BOM dependency number that we used is how many parent SKUs a single component SKU has. In other words, it is the number of other SKUs that one SKU goes into. For example, if one SKU is a part for 20 other SKUs, then that BOM dependency number would be 20.

Once we created the clusters, the next step was determining which SKUs we would consider to be outliers. We set the upper limit to be the third quartile value plus 1.5 times the interquartile range, and any SKU with a scrap quantity above this limit would be considered an outlier.

Figure 10 shows cluster 1 of 2022 as an example. The black dots are the non-outlier SKUs, or "normal" SKUs. The red dots on the graph are the outlier SKUs, or SKUs who have a high scrap quantity compared to the other SKUs within that same cluster.

The next step was to identify which of the outlier SKUs Thermo Fisher should look at. In order to do this, we first calculated a delta value which we defined to be the difference between an outlier SKU's scrap quantity and the 95th percentile value of the non-outlier SKUs. Originally, we attempted to calculate delta as the difference between an outlier SKU's scrap quantity and the mean of the non-outlier SKUs. However, upon further investigation, we realized that the data for the non-outlier SKUs was skewed to the right. This means that the overall average was being brought down to a lower number due to the high frequency of low scrap quantities. This was causing our delta value to be extremely large and making it unachievable for Thermo Fisher to close that gap. We decided to shift and take a more conservative approach by using the 95% value rather than the mean. The 95% value means that 95% of the data points fall to the left of that value. Our idea with delta was to quantify how much scrap needed to be reduced for certain SKUs so that it can still fall on the higher end of scrap, without being an outlier and causing issues.

Once we had all of the delta values calculated for the outlier SKUs, we were able to make a table as seen in figure 11. This table shows the top 10 outlier SKUs from cluster 1 of 2022 ranked by their potential financial opportunity. We calculated the financial opportunity by multiplying each delta value by the unit cost of each SKU. This value is how much money Thermo Fisher could save if they were able to reduce their scrap quantities for these outlier SKUs to fall within that 95% range. Now that we had a prioritized list of SKUs to look into, the next step was to dig deeper into these specific SKUs. We picked two of these top 10 SKUs to explain further: 100022934 and PPC1007. SKU 10022934 was scrapped over 8,000 times in 2022 alone all for being expired. Figure 12 shows the on-hand inventory for this specific SKU over the last couple of months. Looking at the circled region, you can see that the inventory is growing rapidly in recent months, and this could be an opportunity for Thermo Fisher to reduce their inventory growth in order to limit the number of SKUs that are scrapped for being expired.

Next, looking at SKU PPC1007, figure 13 shows the reason for this SKU being scrapped over the past 5 years and the quantity. You can see that there seems to be a large welding issue with this SKU as it has been scrapped due to welding over 100 times each year since 2018. This could be an indicator for Thermo Fisher to look into the welding process and see what may be causing this. This process can be continued for the rest of the outlier SKUs for each cluster to determine where changes can be made.

Since we know that it is impossible to completely reduce scrap by 100%, we created the table that shows the potential savings across all 10 clusters, only reducing the top 10 outliers, by 1%, 5%, 15%, and 25% to show that even a small reduction in scrap can lead to large savings (figure 14).

III. RESULTS AND DISCUSSION

Our initial approaches fell short due to several reasons. Finding commonalities in a scrapped SKU was difficult due to high variance in scrap quantity, types of SKUs, reasoning, and other variables over time. Additionally, our first strategies didn't account for a scraps quantity being relative to its usage and despite being able to identify highly scrapped items, their scrap value was very low. Another reason was that a SKU can be multiple variants of the same variable such as WXYZ. Lastly, overlap in reasons for scrap, such as forecasting and overbought, further complicated our analysis.

We learned that although scrap cannot be completely eliminated, there are many opportunities to reduce it. Clustering automatically groups SKUs into similar categories, which allows for more accurate prediction of outliers. Using the clusters we created, we were able to narrow down the search for SKUs and create a prioritized list of problematic SKUs for Thermo Fisher to investigate.

Taking our clustering method one step further, we created a dynamic tool to pinpoint abnormally highly scrapped SKUs. We utilized the "shiny" package in R Studio to create an interactive user interface for Thermo Fisher. This tool automatically creates the clusters for Thermo Fisher and allows the user to filter by year, sales rank code, and WXYZ. This tool then creates the outlier SKUs table with the potential monetary savings and will help Thermo Fisher save time and resources rather than manually clustering. Even with a slight reduction in scrap, Thermo Fisher could save millions annually.

IV. CONCLUSION

In conclusion, this report provides a comprehensive solution to address Thermo Fisher Scientific Inc.'s significant scrap costs associated with diverse stock keeping units (SKUs). Through the implementation of effective scrap management strategies, the project aims to lower Thermo's annual scrap costs valued at tens of millions of dollars. The proposed solution involves identifying high-value scrap SKUs, analyzing usage patterns, and applying K-Means clustering for better decision-making. By prioritizing efforts based on financial opportunity, Thermo Fisher can focus on reducing scrap costs in SKUs with the highest potential for savings. The utilization of automated screening tools in the future can further enhance efficiency in scrap management processes, resulting in substantial financial savings and operational improvements for Thermo Fisher.

In summary, this report offers a tailored approach to tackle Thermo Fisher's challenge of excessive scrap costs. By leveraging data analytics and advanced clustering techniques, Thermo Fisher can optimize its scrap management processes and improve overall financial performance. The identification of top outlier SKUs with high opportunities for scrap reduction enables targeted efforts and quantifies potential savings. With the proposed solution, Thermo Fisher is equipped to streamline its scrap management, drive cost reduction, and achieve operational excellence.

V. ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to Thermo Fisher for their invaluable support and collaboration throughout this study. Their provision of access to the historical data, including scrap data, usage data, and inventory data, played a vital role in enabling us to conduct our analysis and derive meaningful insights. We would also like to extend our heartfelt appreciation to our esteemed professor, Dr. Kim, from the Shiley Marcos School of Engineering at the University of San Diego, for guiding and supervising us during our senior design project. Dr. Kim's expertise, knowledge, and unwavering support were instrumental in shaping our methodologies, refining our strategies, and ensuring the overall success of this research endeavor. Furthermore, we would like to acknowledge the entire team at Thermo Fisher for their guidance, expertise, and responsiveness throughout the project, as well as the individuals within Thermo Fisher who generously shared their knowledge and insights with us. Our sincere appreciation also goes out to the Shiley Marcos School of Engineering at the University of San Diego for providing us with the opportunity to undertake this senior design project, and to our dedicated research team for their hard work, collaboration, and collective efforts. This project has been a tremendous learning experience, and we are honored to have had the opportunity to work with such a distinguished organization, exceptional individuals, and a supportive academic institution.

VI. REFLECTION PAPER

The engineering senior design project at the University of San Diego has been an exceptional opportunity to apply the comprehensive education received throughout our journey as industrial and systems engineering majors. This reflection paper aims to demonstrate how our senior design team effectively integrated knowledge and insights from various disciplines to design and evaluate an innovative solution to a real-world problem. Specifically, we will discuss how our project considered the impact of our proposed engineering solution in global, economic, environmental, and societal contexts. Furthermore, we will highlight how the synthesis of information from diverse fields outside our engineering major influenced our final design.

Our team recognized the significance of incorporating principles from various core courses to effectively define the user requirements when designing our engineering solution. Courses such as ENGR 101: Introduction to Engineering and COMP 110: Computational Problem Solving equipped us with problem-solving skills and a systematic approach to identify and address the needs of Thermo Fisher. The understanding gained from these core courses enabled us to assess and measure whether our project met these established standards, thus ensuring the quality and reliability of our final design.

To comprehensively evaluate the impact of our proposed engineering solution, our team leveraged knowledge from business and economics courses. Drawing from courses like ISYE 220: Engineering Economics, we weighed the various alternatives and their potential benefits. By considering factors such as cost-effectiveness, feasibility, and return on investment, we made informed judgments that not only addressed technical aspects but also aligned with economic objectives and considerations. This integration of business and economics principles allowed us to develop a solution that not only met technical requirements but also exhibited viability and sustainability from an economic standpoint.

Throughout the project, our team identified common standards and protocols relevant to our engineering solution. We carefully measured and discussed whether our project met these standards to ensure the highest level of communication, teamwork, and compliance to company standards. By adhering to established oaths such as the Pledge of Professionalism, we aimed to deliver an engineering solution that surpassed expectations. This attention to standards not only reflects our commitment to professional responsibility but also ensures that our solution meets the required specifications and guidelines.

In line with our ethical and professional responsibilities, we conducted a comprehensive risk assessment to identify and mitigate potential risks associated with our project. For example, we targeted scrap opportunities in lower ranked SKUs in order to minimize negative outcomes from implementing our method of scrap reduction.

The engineering senior design project served as a culmination of our USD education, where we integrated knowledge and insights from diverse fields to develop an ethical and innovative engineering solution. By drawing upon core courses, business and economics principles, and considering standards and risks, our team successfully designed and evaluated a solution that met technical requirements while also addressing global, economic, environmental, and societal contexts. By embracing the synthesis of information from different areas outside our engineering major, we broadened our perspectives and ensured a holistic approach to problem-solving. Our senior design project exemplifies the power of interdisciplinary collaboration and highlights the importance of considering multiple perspectives in creating meaningful engineering solutions that positively impact the world.

VII. FIGURES

SKU Number	Display Value	Scrap Quantity	Fiscal Year	Cause of Scrap	Branch Code
A33502	$282,622,673	2,188,112	2019	Null	US03
100103376	$ 1,669,466	70,000	2021	Null	LT01
A48105	$ 1,460,730	2,001	2022	Quality Issue	GB01

Figure 1: Creating Tables and Ordering by Scrap Value

SKU Number	Display Value	Scrap Quantity	Fiscal Year	Cause of Scrap	Branch Code
MT36814	$ 0	199,605,000	2021	Overbought	US15
MT34080	$ 9990	99,898,341	2018	Overbought	US15
100094872	$ 333,555	66,300,000	2022	Misalignment	LT01

Figure 2: Creating Tables and Ordering by Scrap Quantity

Product Group	2018	2019	2020	2021	2022
1C7	$ 242,693	$ 180,820	$ 138,725	$ 174,794	$ 295,997
ANT	$ 1,106,337	$ 526,835	$ 633,667	$ 256,692	$ 1,172,656
FCR	$ 2,129,282	$ 1,593,957	$ 3,316,388	$ 1,593,147	$ 1,487,461

Figure 3: Relating Scrapped SKUs to a SKUs Usage to Determine Proportionality

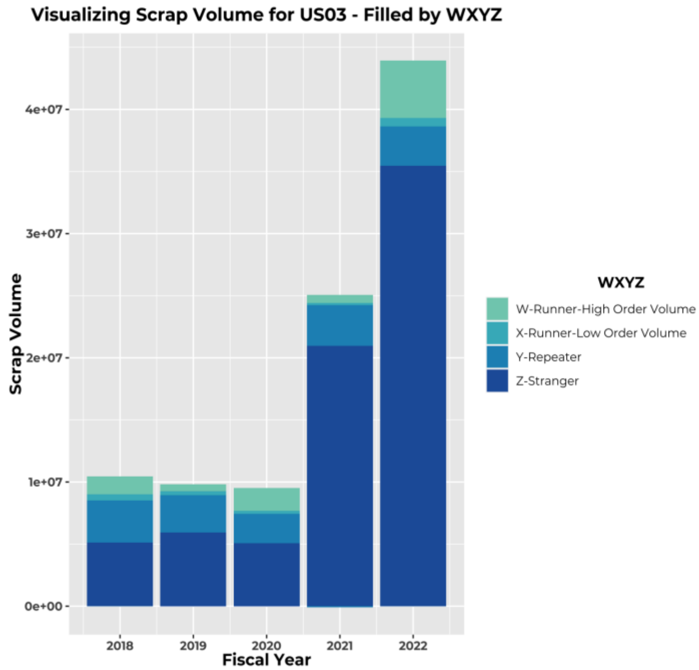

Figure 4: Examining Scrap Quantity Over Time Filled by WXYZ to Spot Trends

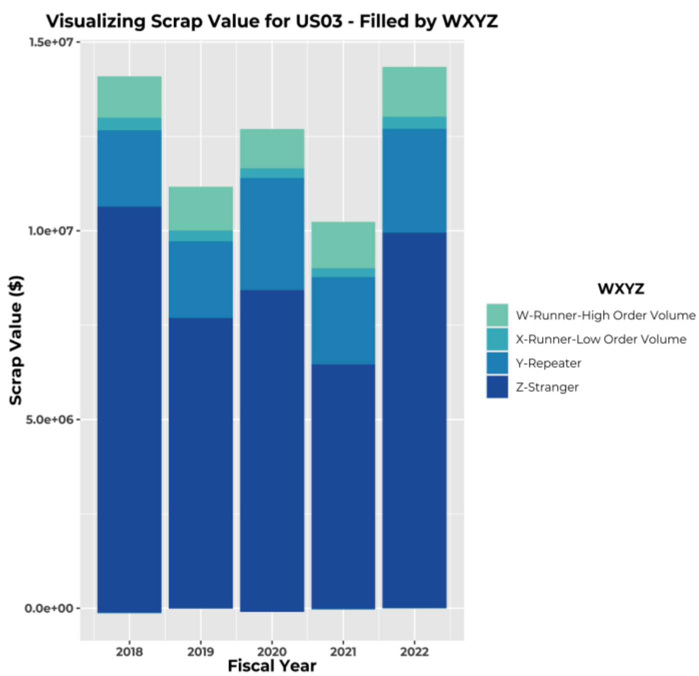

Figure 5: Examining Scrap Value Over Time Filled by WXYZ to Spot Trends

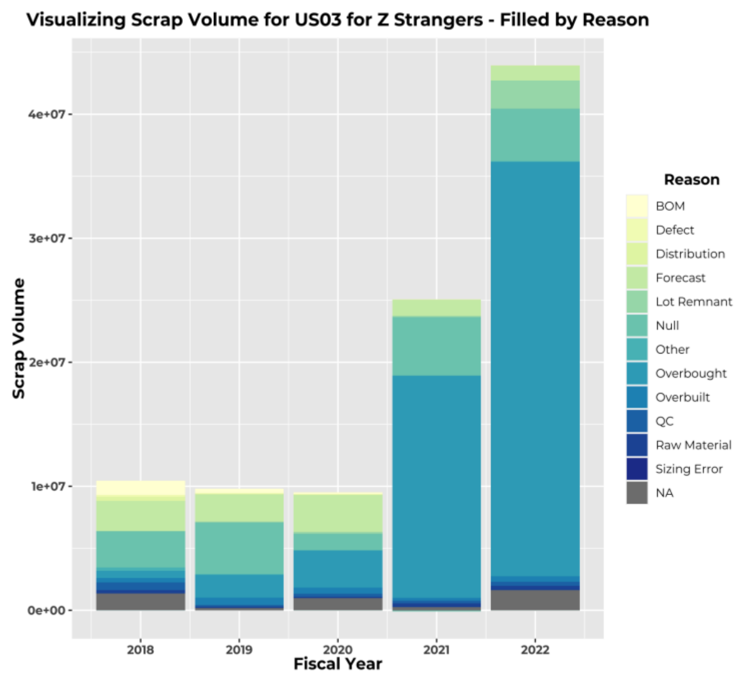

Figure 6: Examining Scrap Quantity Over Time Filled by Reason to Spot Trends

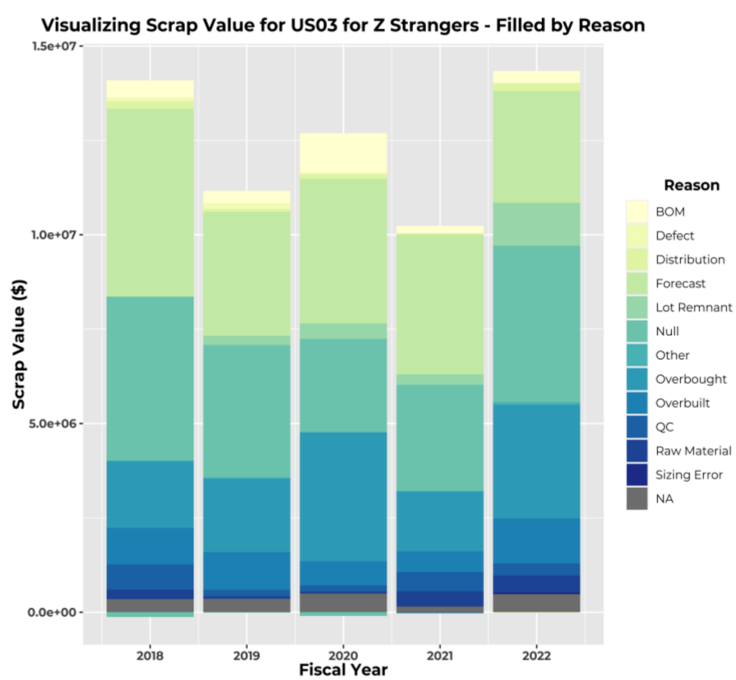

Figure 7: Examining Scrap Value Over Time Filled by Reason to Spot Trends

Reason	Value	Percentage Value	Percentage Quantity
Forecast	$ 107,198,744	35%	7%
Null	$ 41,319,387	14%	7%
Overbought	$ 36,043,700	12%	42%

Figure 8: Reasons for Top 3 Scrap Values From 2018-2023

Reason	Quantity	Percentage Value	Percentage Quantity
Overbought	431,565,031	12%	42%
Raw Material	113,386,438	2%	11%
Defect	88,860,208	5%	9%

Figure 9: Reasons for Top 3 Scrap Quantities From 2018-2023

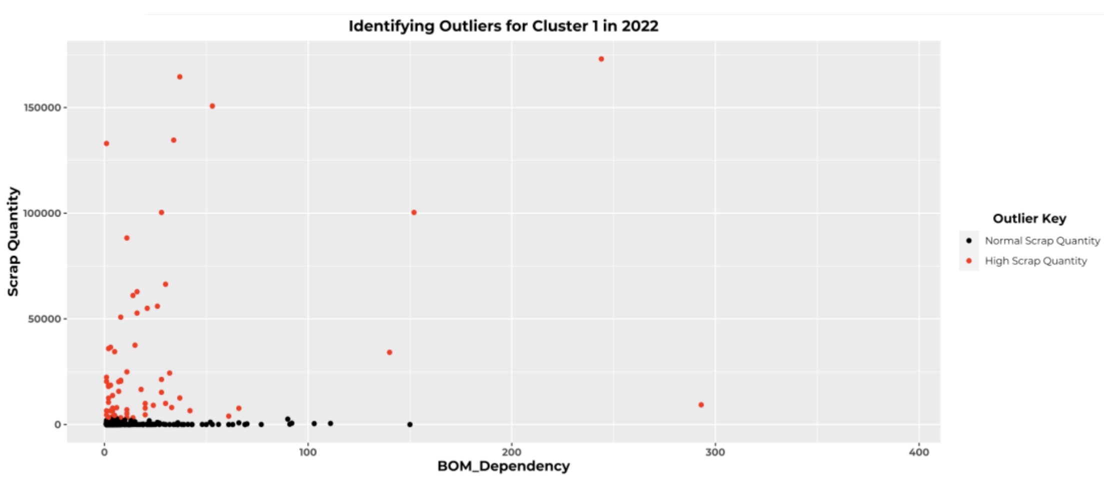

Figure 10: Clustering Outcome 1 for 2022

SKU Number	Unit Cost	Delta	Financial Opportunity
100015737	$ 18.71	8,177	$153,006
100040578	$ 1.80	48,993	$87,996
PPC1007	$ 0.82	32,333	$26,607
100022824	$ 1.99	10,740	$21,335
100021400	$ 1.14	18,676	$21,300
100020691	$ 16.56	1,127	$18,668
100026499	$ 0.87	19,513	$16,920
100036589	$ 1.19	11,912	$14,164
100038835	$ 2.70	4,664	$12,592
100022934	$ 0.65	18,591	$12,145

Figure 11: Top 10 Outliers of Cluster 1

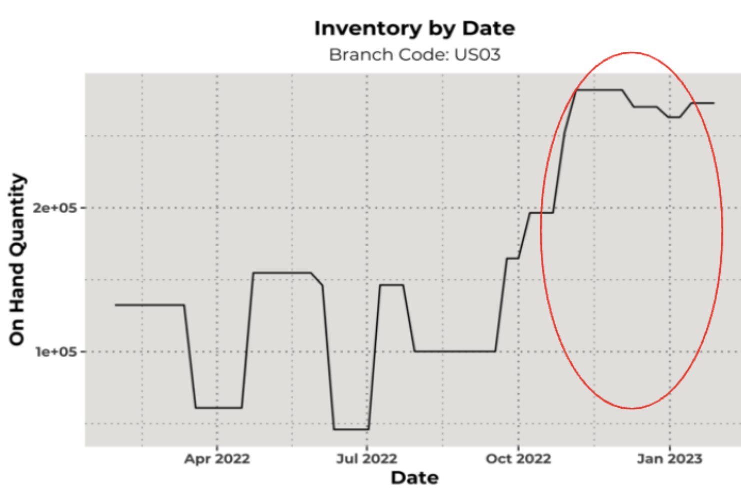

Figure 12: Investigating Top Outliers: SKU = 100022934

Year	Quality Issue	Dry Lines	Fibers	NULL	Welding
2018	2	1	1	15	148
2019	0	0	0	7	163
2020	0	0	0	3	132
2021	0	0	0	2	143
2022	0	0	0	2	145

Figure 13: Investigating Top Outliers: SKU = PPC1007

Cluster	Potential Savings for Top 10 SKUs	1% Savings	5% Savings	15% Savings	25% Savings
1	$384,732	$3,847	$19,237	$57,710	$96,183
2	$293,959	$2,940	$14,698	$44,094	$73,490
3	$282,018	$2,820	$14,101	$42,303	$70,505
...	...	...	...	...	...
Total	$7,777,899	$77,779	$388,895	$1,166,685	$1,944,475

Figure 14: Potential Savings Across Clusters