Steelcase Inc.

Final Report

Team 17

Team Members

Whitman Blass

Federico Ceccagnoli

Christian Mattei

Eli Katz

Evan Zappa

Aveed Matinfar

Grace McDonough

Luca Kurtz

Team Liason

Christian Mattei

cmattei22@gatech.edu

Client Contact

Edward Vander Bilt

evanderb@steelcase.com

Team Advisor

Dr. Gamze Tokol-Goldsman

gamze.tokol-goldsman@isye.gatech.edu

This report has been created as a part of a student design project at the Georgia Institute of Technology

Executive Summary

The client, Steelcase, is a multinational architecture, furniture, and technology manufacturer well known for their office products, including desks and chairs. Each of their product lines has its own unique supply chain and set of suppliers they use to provide different components. This project is focused on improving their supplier selection process with a goal of minimizing supplier lateness and volatility.

After analyzing over 65,000 supplier deliveries from 2018 – 2022, it became apparent that many of Steelcase's key performance categories were in decline. On average, Steelcase's orders went from two days early to over a day late, and during the same period, the percentage of orders that arrived on time per week dropped from 95% to 88%. While supply chain challenges have become more pronounced in the aftermath of COVID, the team learned that the current process for selecting suppliers relies heavily on employee intuition. Currently, Steelcase classifies suppliers by the material group they belong to, which is based on the materials that they supply. When looking at the allocation Steelcase gives to certain suppliers in each material group, the team noticed that in many cases high performing suppliers were being underutilized.

To find optimal supplier allocations, the team designed an algorithm that gives Steelcase a list of portfolios for each material group[1] with varying levels of lateness and risk. A portfolio in this context refers to the suppliers and the percentage of the materials in that material group that each supplier will supply. While intuition would say to pick the supplier with the best performance, diversifying suppliers allows Steelcase to achieve similarly high performance with lower risk. The algorithm takes in aggregated monthly supplier lateness data to create a probability mass function for each of the suppliers in the portfolio. A linear program is then used to find the optimal allocation weights[2] for suppliers that minimizes lateness severity and volatility.

In addition, the team designed a web app that will enable Steelcase to better visualize and monitor their supply chain. The web app is composed of three main pages: a supply chain health dashboard, an analytics page, and a supplier portfolio recommendations page. The supply chain health dashboard provides a high-level overview of key supply chain metrics. The analytics page allows for more in-depth analysis of supply chain performance at varying levels of detail using pre-built filters and interactive charts and visuals. Once problems are identified, decision makers can use the supplier portfolio recommendations page to analyze and compare possible solutions.

With the aid of the team's deliverables, Steelcase can discover problems in their supply chain sooner and make more efficient data-driven decisions surrounding supplier allocations. As a result, supply chain performance will improve, and costs associated with late deliveries will be reduced.

[1] Material Group: A group of suppliers capable of producing the same material.
[2] Weights: The percentage of a material demand that is sourced from a supplier.

Table of Contents

I. Overview of the Project

1.1 The Existing System

Steelcase is a multinational architecture, furniture, and technology manufacturer well known for their office products, including desks and chairs. By volume, Steelcase is the largest such company – leading the world in pieces of furniture supplied to offices, schools, and hospitals. Steelcase's functions span around eighty countries, 11,800 employees, and $2.8 billion in revenue the past fiscal year. This network of supplier organizations will be further referred to as the "supply chain." Suppliers who provide material directly to Steelcase are known as Tier 1 suppliers. By extension, suppliers who provide inputs to a Tier 1 supplier are known as Tier 2 suppliers. The most immediate way to impact the performance of the supply chain is to change Steelcase's behaviors with the Tier 1 suppliers, with whom they maintain business partnerships. These interactions will be further referred to as the procurement process.

The team will focus specifically on the Kentwood, Michigan plant. The procurement process begins while a new Steelcase product is being developed. Steelcase must first determine which materials are required to build the new product. They then choose a supplier for each material and sign a contract with that supplier for a specified period. Oftentimes, more than one supplier can produce a single material, so Steelcase further groups suppliers into "material groups." Within each material group, suppliers are allocated differently depending on current contracts, performance, and their capacity to produce those materials. In each material group, it is assumed that all suppliers can provide all the materials in that specific group meaning suppliers can be swapped with each other. *Figure 1* shows an example material group portfolio with 8 suppliers and the "weight" of each supplier, which is the percentage of materials that each supplier supplies in that group.

Supplier	Number of Orders	Ownership % ("Weight")
Supplier 1	2	00.1%
Supplier 2	180	09.3%
Supplier 3	32	16.6%
Supplier 4	211	10.9%
Supplier 5	1,250	64.8%
Supplier 6	246	12.6%
Supplier 7	5	00.3%
Supplier 8	2	00.1%
Total	**1,928**	**100%**

Figure 1: Material Group for Urethane Products

Once the item is in production, Steelcase tracks inventory of input materials in the enterprise resource planning software SAP. SAP then advises the Materials Resource Management team when to place purchase orders (POs) to restock inventory of input materials when they fall below a specific threshold. These POs include decisions regarding how many units of the input material to order, what lead time to give the supplier, and which supplier to use. The supplier chosen for a specific PO is often dependent on contractual obligations. When supplier contracts are up for renewal, there is an opportunity to switch to a different supplier. *Figure 2: Current Steelcase Procurement Process* outlines this process below.

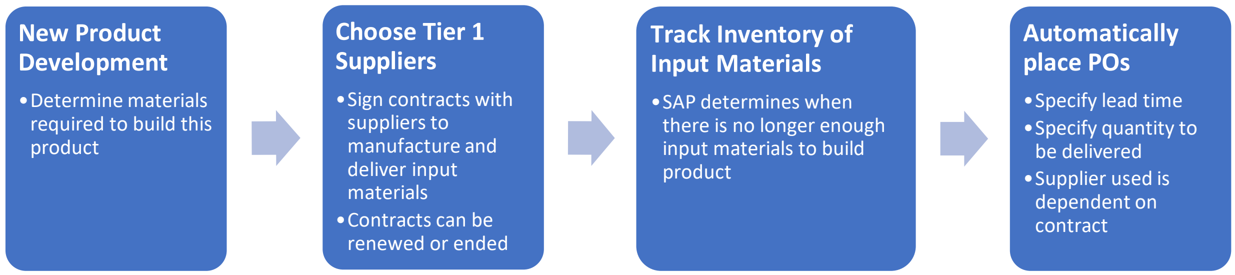


Figure 2: Current Steelcase Procurement Process

1.2 The Worsening Problem

With over 130 distinct Tier 1 suppliers delivering to the Kentwood facility at various intervals, the supply chain is complex and difficult to track or predict. The immediate aftermath of the Covid-19 pandemic brought sizable disruptions to the supply chain. In the two years since its initial outbreak, global economic shutdowns, labor shortages, and malfunctioning seaports have continually shocked the Kentwood facility's supply chain. From May of 2019 to March of 2022, **Steelcase's orders have steadily descended from two days early on average to over a day late** (see *Appendix E: Daily PO Delays over Time*). Further, during the same period **the percentage of POs that arrived on time per week dropped from 95% to 88%**. From these analyses and client discussions, it is evident that the supply chain has deteriorated over the past two years[3] and is trending in the wrong direction.

This observed trend is unsustainable and could lead to larger issues for Steelcase if left unchecked. Delayed input materials are very expensive, and lead to increased costs in three distinct ways:

1. A product cannot be made unless all materials are present. When one material is late all other materials sit in storage incurring inventory carrying costs.
2. Delayed inputs cause Steelcase to delay manufacturing products, ultimately resulting in customers getting products late.
3. Less trust in the supply chain necessitates a buildup of safety stock to avoid the two aforementioned problems, further increasing inventory costs.

1.3 The Opportunity

Steelcase's Kentwood facility has set a goal of receiving 99% of their POs on time. The facility's current on-time delivery percentage is 93%, well shy of their goal. Of the four steps in the procurement process mentioned before (see Figure 2), steps two and four "Choosing Suppliers" and "Placing POs" most directly impact whether a PO will arrive on time. Breaking these two steps down further, one can see that there are really three decisions that can be made that will affect the PO's timeliness: which supplier to choose, how much lead time to give the supplier, and what quantity of a material to order.

Logically, it seems that these three decisions and the on-time delivery of a PO should be highly correlated. This hypothesis was tested by using a logistic regression model with an ordinary least squares loss function to find relationships between the independent variables (supplier, lead time, and quantity)

[3] All data is for the Kentwood Facility only

and the dependent variable: on-time delivery. Each independent variable was regressed against on-time delivery percentage separately; predictors were scored by calculating the percentage of binary predictions that the fitted model was able to accurately guess. It is worth noting that before modeling, the data was under-sampled to produce more meaningful results.

INDEPENDENT VARIABLE	**DEPENDENT VARIABLE**	**SCORE**
SUPPLIER	On-Time (Y/N)	**0.72**
EXPECTED LEAD TIME	On-Time (Y/N)	**0.57**
QUANTITY	On-Time (Y/N)	**0.48**

Figure 3: Relationship between Procurement Decisions and On-Time Delivery

Figure 3 reveals that **the supplier chosen has a strong relationship with the on-time delivery** of a PO, predicting accurately whether a PO would be late 72% of the time. Expected lead time and Quantity were able to predict whether a delivery would be late 57% and 48% of the time respectively. Accurately predicting a binary outcome only 57% of the time is marginally better than a random guess, and 48% is worse. Hence, it seems that supplier selection is the biggest opportunity to improve Steelcase's performance.

Steelcase does not currently employ a data-driven supplier selection process and currently relies on system knowledge to make these decisions. For example, when looking at the Urethane products material group in *Figure 1*, it is evident that certain high performing suppliers are underutilized. Supplier 2 supplies approximately 9% of the material group, but its late percentage is significantly lower than the average. The team found many similar examples where Steelcase was sourcing large percentages of materials from poor performing suppliers.

From the current state of the system and the above relationships, the following opportunities for improvement are presented:

1. Develop a methodology for supplier selection that will improve the performance of the material group as a whole
2. Inform decision makers about the problems they are currently facing or are expected to face, and make it easy to act on these problems

To adequately address each opportunity, the team creates two deliverables:

1. Supplier recommendation algorithm to optimize material group performance by minimizing order delays and risk
2. Real-time dashboard that monitors supply chain health and forecasts performance

It is important to note that there are non-technical end users interacting with each deliverable in different ways. These deliverables and their outputs must be easy to understand and interact with. Furthermore, the methodology used must be well defined so that it can be repeated with new supply chains once the deliverables are handed over to Steelcase.

It is for this reason that the supplier selection process and forecasting suite will be automated and come packaged as part of a custom software application. The end user will simply download a package, and then run the application to find what they are interested in. To this end, the team has elected to build the deliverables in Streamlit[4], an open-source app framework for python. This will streamline the process of developing a graphical user interface (GUI) while also enabling easy data ingestion, modeling, and visualization capabilities.

II. Design Strategies and Deliverables

2.1 A Recommending Suppliers

As shown previously, selecting the suppliers to use for a material group is the biggest lever that Steelcase possesses to improve performance and reduce costs. Therefore, the team's first deliverable is a data-driven approach for selecting the most optimal supplier portfolio for each material group.

Instead of recommending a single supplier, the team recommends a group of suppliers that Steelcase can order from, along with the percentage of materials that each recommended supplier should supply. Ordering from a few suppliers, rather than relying on one, reduces risk in a couple of key categories. First, having multiple suppliers for a material group reduces performance risk. Using a group of suppliers, it is possible to achieve near optimal on-time delivery rates with much lower variance. Ordering all materials from the best performing supplier in the material group would achieve the best on-time delivery rate. However, if that one supplier has a poor spell of performance, the entire material group's performance would drop. Additionally, blending suppliers reduces strategic risk. If a supplier were to increase prices, having other suppliers available mitigates the cost damage done to the overall material group. Similarly, if a supplier goes bankrupt or is unable to process purchase orders, having other suppliers would prevent the entire material group from failing.

To create the optimal material group portfolios, we first tabulate data of supplier performance aggregated monthly for the given material group. For this process, the Average Late Days (ADL) of a supplier per month is used instead of the supplier's on-time delivery rate. ADL is used instead of on-time percentage because, despite the client's stated goal being to improve on-time percentage, the total number of late days in a month is more connected to cost incurred. Therefore, the team feels that ADL is a better metric to focus on improving, knowing that an improvement in ADL will roughly correlate to an improvement in on-time percentage. In the example below, we analyze a material group with three suppliers who have supplied this material group at some point in the past. In the table below, each row represents the ADL for that supplier in the given month for the given material group.

[4] More information about the Streamlit technology can be found here: https://streamlit.io.

Month	Supplier 1	Supplier 2	Supplier 3
Nov. 21	6.22	0.00	3.50
Dec. 21	7.78	1.11	0.06
Jan. 22	9.33	3.82	1.10
Feb. 22	10.89	1.42	1.39
…	…	...	…

Figure 4: Urethane Products Material Group Performance

Next, we take this table and transform it into binned data with three bins—0, 1, and 2—where each bin represents the severity of lateness, with bin 0 being the best, or least severe, and bin 2 being the most severe. The datapoints are assigned to bins as follows: an average days late value of 0 is assigned to bin 0, average days late between 0 and 2 is assigned to bin 1, and average days late greater then 2 is assigned to bin 2. We then find the frequency of each supplier performance combination by dividing the number of times that combination occurred by the total number of months. For example, if supplier 1 had a lateness severity (bin value) of 0, supplier 2 had a lateness severity of 2, and supplier 3 had a lateness severity of 0 in the same month twice over 50 months of data, the probability would be: $\frac{2}{50} = .02$ (row four of *Figure 5* on the right).

Month	Supp. 1	Supp. 2	Supp. 3
Nov. 21	2	0	1
Dec. 21	2	1	1
Jan. 22	2	1	1
Feb. 22	2	1	1
…	…	...	…

Comb.	Supp. 1	Supp. 2	Supp. 3	Frequency	Prob.
1	0	0	0	0	0.00
2	0	0	1	3	0.03
3	0	0	2	1	0.01
4	0	1	0	2	0.02
…	…	...	…	…	…

Figure 5: Binning Data to Calculate Probabilities

With the data formatted this way, we can find the combined performance of all suppliers, Z, and the probability that the combined performance would occur. To do this, we simply add up their lateness severity values, i.e., $Z = p_0 + p_1 + p_2$ where p_i is the binned value of supplier i. Again, the smaller Z is, the better the material group performance was for that month, and vice versa. For example, the first row is saying that suppliers 1, 2, and 3 all averaged zero days late, which translates to a Z value of 0 (0 = 0 + 0 + 0). On the other hand, if all suppliers had an average of over 2 days late (bin 2), their Z value would be 6 (6 = 2 + 2 + 2).

After transforming the data, we solve for the optimal weights. Before solving for the weights, let's look at what the probability distribution of the combined lateness severity (Z) would look like for an arbitrary set of weights. For this example, we will use the weights 0.2, 0.3 and 0.5 for suppliers 1, 2, and 3, respectively, which can be interpreted as follows: supplier 1 supplies 20% of the materials, supplier 2 supplies 30%, and supplier 3 supplies 50%. Using these weights, we can calculate the weighted Z value as shown in *Figure 6*.

Comb.	W 1	Supp. 1	W 2	Supp. 2	W 3	Supp. 3	Freq.
1	20%	0	30%	0	50%	0	0
2	20%	0	30%	0	50%	1	3
3	20%	0	30%	0	50%	2	1
4	20%	0	30%	1	50%	0	2
...		...		...		...	...

Σ (Weights x Suppliers) = Z
(0.2 x 0) + (0.3 x 0) + (0.5 x 0) = 0.0
(0.2 x 0) + (0.3 x 0) + (0.5 x 1) = 0.5
(0.2 x 0) + (0.3 x 0) + (0.5 x 2) = 1.0
(0.2 x 0) + (0.3 x 1) + (0.5 x 0) = 0.3
...

Figure 6: Calculating the Weighted Z Score

These weights produce the probability distribution below with mean 0.96 binned late days and standard deviation 0.63.

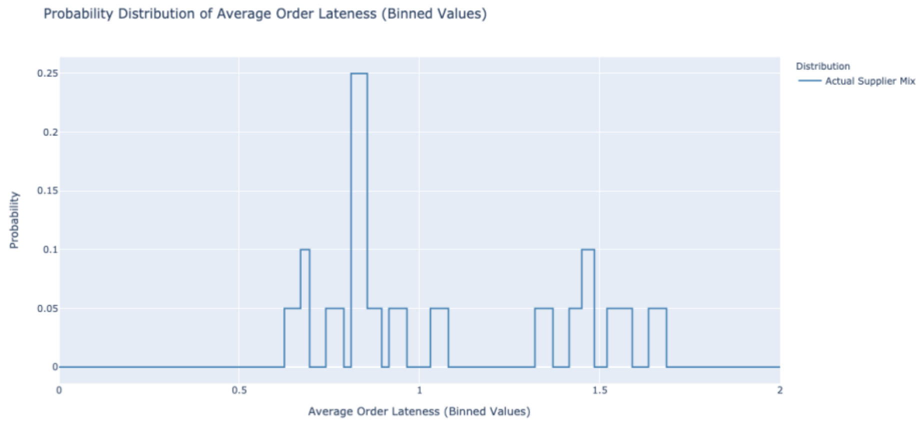


Figure 7: Probability Distribution for the Portfolio Weights (0.2, 0.3, 0.5)

However, with the different set of weights for the suppliers (0.0, 0.9, 0.1), the probability distribution has a mean of 0.15, and standard deviation of 0.06, which is a significant improvement to both mean and variance.

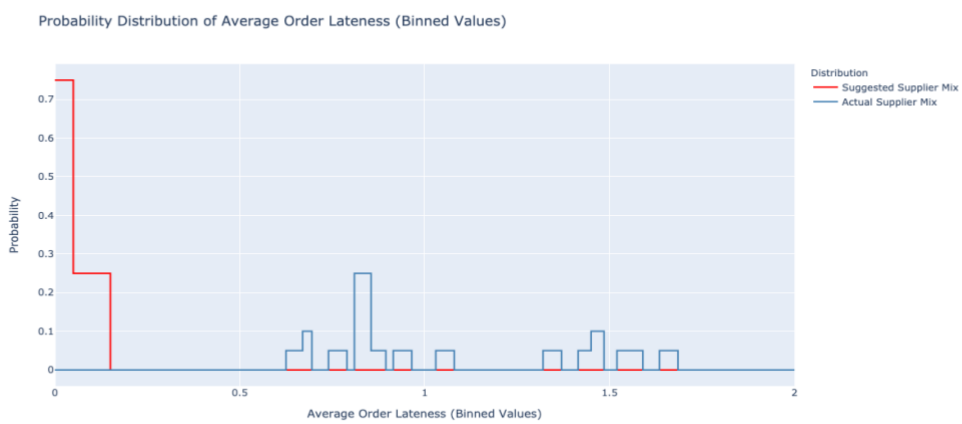


Figure 8: Probability Distribution for the Portfolio Weights (0.0, 0.9, 0.5)

As you can see, altering the weights changes the probability distribution. Knowing this, we can create a simple linear program (LP) to solve for the optimal set of weights that minimizes the expected average days late Z_i of a portfolio. In the LP, the weights a_j are the decision variables, and x_{ij} refers to the weighted bin value of combination i and supplier j. The full LP formulation can be found in Appendix D.

2.1B Supplier Recommendations Results and Limitations

To test the recommendation algorithm, we try recommending suppliers at different points in the past and compare the performance of recommended suppliers to that of Steelcase's actual suppliers. It is important to note that supplier performance is not correlated to the volume of deliveries that supplier makes, so the team feels that back-testing in this way is a fair way to evaluate recommendations. We first show what this looks like for material group number 17, Urethane Products, by trying our recommendation algorithm in August 2021 using all data leading up to that date. At the time, three suppliers supported the material group, with weights 20%, 30%, and 50%, respectively. Using the algorithm, we recommend the weights 0%, 90%, and 10%. If we assume that it takes one month to fully make the switch to the algorithm's recommended suppliers, we can begin analyzing the results in September 2021.

We can see that allocating 90% of the portfolio to supplier B and 10% to supplier C outperforms Steelcase's original portfolio. When this recommendation is implemented, the average days late days per month decreases from 53 to 30. Not only do materials arrive sooner on average, but they also arrive at more consistent times, making the material group performance more predictable. See Appendix F for further results from this experiment.

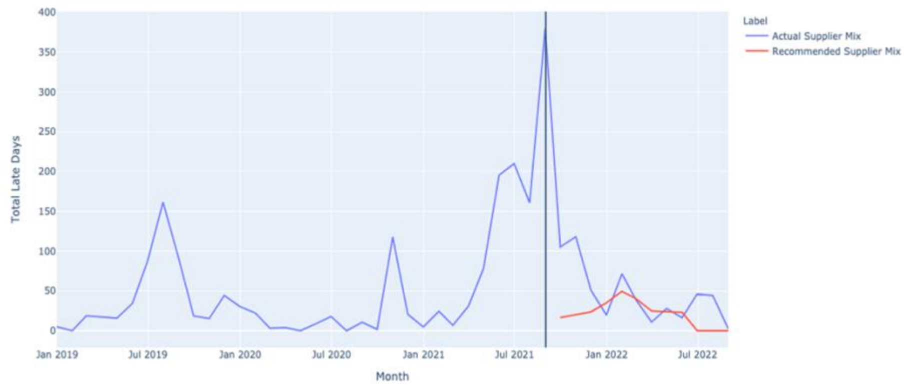


Figure 9: Back Testing the Recommendation Algorithm for Material Group 017

This example only tests the algorithm for one material group at a single point in time. To better understand the algorithm's performance, the team ran the same test for the 8 largest material groups at 6 different points in time. These points in time, or cutoff dates, are selected to ensure that both training and testing sets have enough data to run tests.[5] The results are shown below.

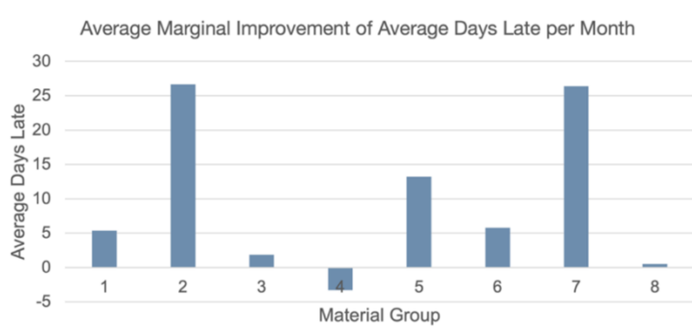


Figure 10: Back Testing the Recommendation Algorithm for 8 Material Groups

[5] The training set refers to all available data before the cutoff date while the testing set refers to all available data after the cutoff date. The tests use the same methodology as the test performed on material group 017.

Each bar represents the average marginal improvement to the average days late of the recommended portfolio compared to the actual portfolio across all 8 cutoff dates. While material group 4 experienced a slight increase in average days late, there is a clear overall decrease in average days late. On average, Steelcase would have saved 10.9 late days per month, which translates to 1,046 late days a year. More data on marginal improvement can be found in Appendix C.

This approach does not come without its limitations. First, suppliers with limited data cannot be considered as options for portfolios. This means that suppliers who might end up being the best option, would never be recommended because of their lack of data. Additionally, the relationship between average days late and cost savings is unclear. While we know that more cost is incurred as the order delay increases, the exact relationship is unknown. Therefore, the exact cost savings of the team's recommendations could not be predicted with high accuracy. Furthermore, Steelcase is locked into long-term supplier contracts meaning that suppliers cannot be swapped at any moment. Steelcase would have to wait until the end of the contract to make the switch, which would delay the cost savings due to the improved portfolios. Finally, certain recommended portfolios could be infeasible for a few reasons. For one, a supplier might not be able to handle supplying an increased number of materials due to an unknown capacity constraint. Similarly, a supplier might not be willing to ship a lower number of materials due to a minimum capacity constraint. Lastly, Steelcase might not want to reduce the number of materials shipped by a supplier to maintain a strong strategic relationship that could outweigh the benefits of swapping out that supplier.

While there are limitations, the team is confident that the recommended portfolios will add value to Steelcase. For example, new suppliers are likely to have the least amount of data. Since they are locked in until the end of their contract, enough data will be collected by the time the algorithm can be run. Most importantly, this tool is not intended to be followed blindly. It simply presents Steelcase with a list of portfolios to choose from that are likely to outperform their current selection. Steelcase supply chain leads with more knowledge of suppliers and strategic relationships can rule out recommendations based on their understanding of the constraints that the team does not have. Even with certain recommendations ruled out, there are many portfolios that only slightly alter supplier weights that would still save a significant amount of money due to fewer delays.

2.2A Descriptive Analytics

Steelcase's supply chain leads want to better understand the current state of their supply chains. Their current processes for understanding their supply chains are time-consuming and require various manual applications. Many non-technical team members are unable to produce their desired visualizations and metric calculations and have been internally requesting an easy-to-use analytical tool for over 3 years. The team's goal was to create a centralized location where supply chain leads can browse extensive metrics and visualizations to better understand their supply chains' performance. Furthermore, by integrating forecasting models, the team can show the expected future performance of various metrics across the entire supply chain.

The descriptive analytics portion is focused on improving visibility in the procurement process to make it easier to quickly identify issues. Users can select how to view and aggregate key data however they see fit with options to filter data by supplier, material group, material, time frames, etc. Furthermore, performance (by supplier, material group, etc) is put into context by allowing users to compare various performance metrics across all levels of aggregation (ex: supplier A vs supplier B). These performance metrics are also shown in tandem with the Global Supply Chain Pressure Index, a time series performance metric created by the Federal Reserve Bank of New York tracking global supply chain performances. This allows Steelcase to better understand how their suppliers are performing in relation to other global supply chains.

Driven by the Steelcase team, the team integrated a vast array of intuitive visualizations and performance metrics into Streamlit. Streamlit is an open-source app framework made for Data Science applications which are being used to centralize all the deliverables. The descriptive analytics are located on the Homepage and Analytics pages of the Streamlit app.

The Homepage provides a high-level overview of the supply chain performance by material group. The Analytics page goes into more detail about performance and contains forecasted metrics.

Figure 11 below is taken from the Streamlit Analytics page showing material group performances in a scatter view where material groups that meet Steelcase's performance criteria are highlighted in green. The size of the bubble corresponds to the number of orders placed per material group. Red bubbles indicate a material group that is not meeting Steelcase's standards and prompts further investigation.

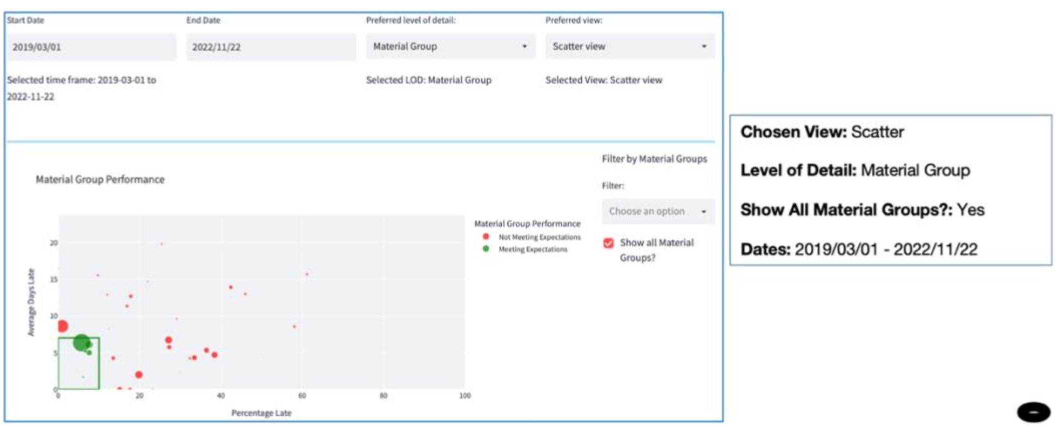


Figure 11: Streamlit Analytics page – Scatter View

Rather than only including analytics on past performance, forecasting allows supply chain leads to predict future performance metrics. The forecasts apply an auto-regressive integrated moving average (ARIMA) model to predict the near future (five months forward) based on historical data. The methodology for the ARIMA modeling is included in section 2.2B and Appendix A.

2.2B Descriptive Analytics Analysis

Unique ARIMA forecasting models were created to forecast the average monthly on-time percentage for each of Steelcase's suppliers, materials, material groups, and supplier/material group combinations. In total, 157 forecasting models were created using the following methodology and validation techniques.

1. Parameter optimization

a. Every model was created from a unique filtered data frame. Using a training/testing data split (training data = first 68% of the time period), each data frame's training dataset was fit to 30 ARIMA forecasting models using a grid-search methodology for the ARIMA parameters p, d, q. The grid-search outputs the parameters that minimized mean-square-error when testing the forecasted results against the test dataset. See Appendix A for code.

2. Residual Analysis
    a. Residual Analysis was used to confirm that the models performed well. Residuals are the error calculated from the difference between the predicted values and the actual values in the test dataset. Residuals must be normally distributed with a mean of 0 or the model may be over or under forecasted. Further residual analysis was performed including autocorrelation analysis, error analysis, analysis at each timestep, etc. Forecasts which did not meet expectations (residuals weren't normally distributed, high autocorrelation, high mean-squared-error, etc) were discarded and are not included in the forecasting suite. See appendix A for excerpts of residual analysis.
3. Confidence Interval Creation
    a. 95% confidence intervals were calculated based on the residual analysis. 95% confidence intervals are simply calculated from: $predicted\ value\ \pm 1.96 \times stdev$ where stdev stands for the standard deviation of residuals. The team believed that this would be the best way to let Steelcase supply chain leads intuitively understand the quality and confidence of these predictions. Wide confidence intervals let the end user understand that a model may not be highly accurate. Appendix B contains the confidence intervals for a forecast.

III. Evaluating the Project

Because the financial impact of late deliveries is not readily available, the team used reduction in days late as the performance metric for evaluation. While late percentage is another key metric that Steelcase uses, it fails to capture the severity of orders being late. Because the relationship between late deliveries and cost is undefined but present, days late was used as the key performance metric. To quantify the impact of the solution, the team ran the supplier selection algorithm across the 8 largest material groups in supply chain. If Steelcase had used the recommended supplier portfolios, they would have saved a total of 1,046 late days a year for an average of 10.9 days per month.

In addition, Steelcase has not had an easily accessible data platform to analyze and visualize key supply chain metrics. The team's web app allows Steelcase supply chain leads the opportunity to perform analysis on one unified dataset that can help easily monitor and diagnose critical supply chain issues. Through client meetings, the team learned that this tool is something that has been requested for several years but has never been implemented. Going forward, it will directly help Steelcase through its insights but also serve as a steppingstone for future functionalities to be built on top of it. Because the tool is built from a python package, it is highly customizable and easy to work with making it an ideal tool for long-term use.

IV. Appendix

Appendix A

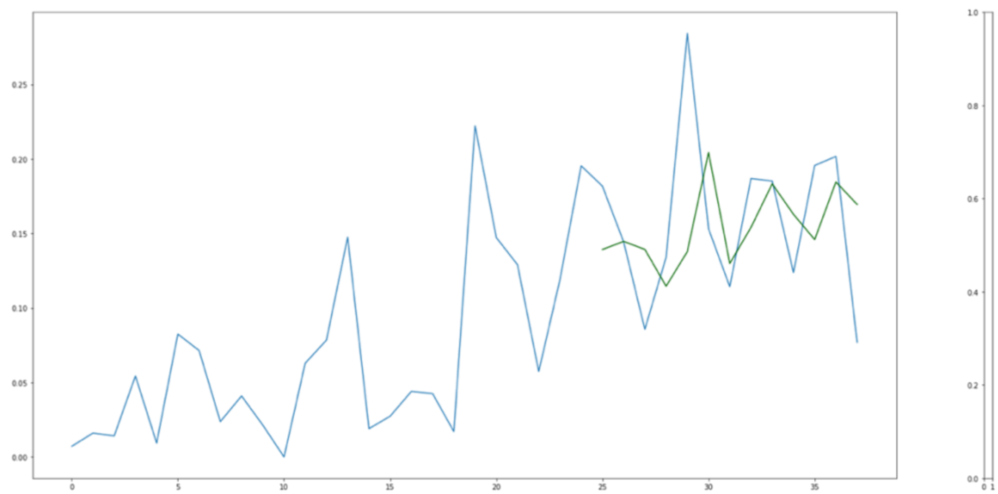
ARIMA Forecasting Parameter Optimization Code

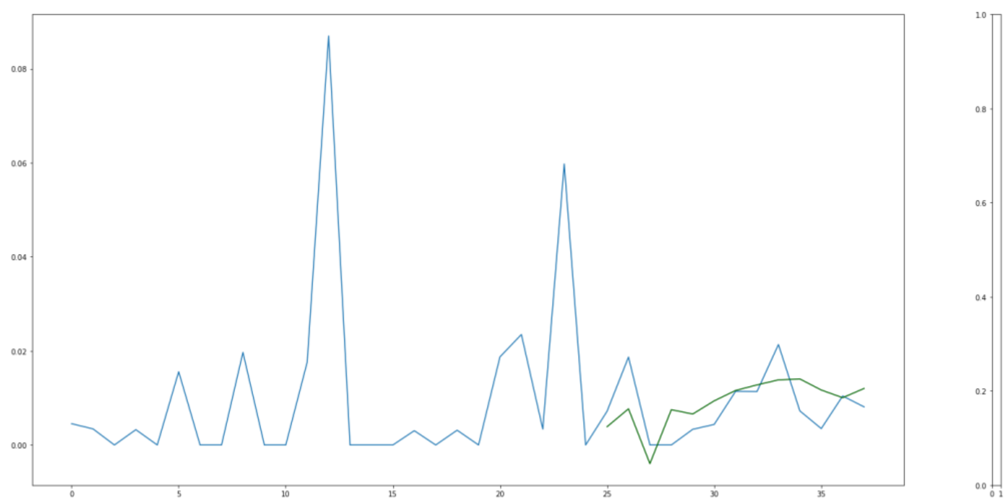
```
def evaluate_arima_model(X, arima_order):
    # prepare training dataset

    train_size = int(len(X) * 0.680)
    train, test = X[0:train_size], X[train_size:]
    tr = train.tolist()
    te = test.tolist()
    history = [x for x in tr]
    #return(test[0])
    # make predictions
    predictions = list()
    for t in range(len(te)):
        model = ARIMA(history, order=arima_order)
        model_fit = model.fit()
        yhat = model_fit.forecast()[0]
        predictions.append(yhat)
        history.append(te[t])
    # calculate out of sample error
    #return(history)
    error = mean_squared_error(test, predictions)
    return error

#print(evaluate_arima_model(df3['latepercentage'].squeeze(),(1,1,1)))



# evaluate combinations of p, d and q values for an ARIMA model
def evaluate_models(dataset, p_values, d_values, q_values):
    dataset = dataset.astype('float32')
    best_score, best_cfg = float("inf"), None
    for p in p_values:
        for d in d_values:
            for q in q_values:
                order = (p,d,q)
                try:
                    mse = evaluate_arima_model(dataset, order)
                    #print((order,mse))
                    if mse < best_score:
                        best_score, best_cfg = mse, order
                    #print('ARIMA%s MSE=%.3f' % (order,mse))
                except:
                    continue
    #print('Best ARIMA%s MSE=%.3f' % (best_cfg, best_score))
    return(best_cfg, best_score)

# load dataset
#series = df3['latepercentage'].squeeze()

# evaluate parameters
p_values = [0, 1, 2, 4, 6, 8, 10]
d_values = range(0, 3)
q_values = range(0, 3)
warnings.filterwarnings("ignore")
```

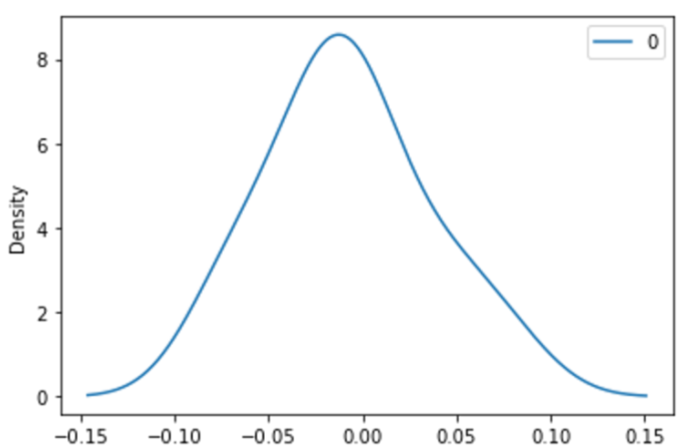
Appendix B

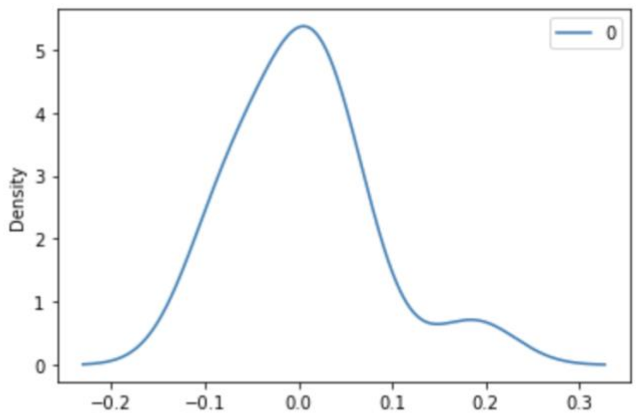
Residual Analysis and Confidence Intervals for Forecasts

The first two images show the forecasted vs. actual values for 2 ARIMA models. Those images are followed by density plots of the residuals. As you can see, they seem to be normally distributed at 0. The final graphs are showing the confidence intervals of the 2 models.

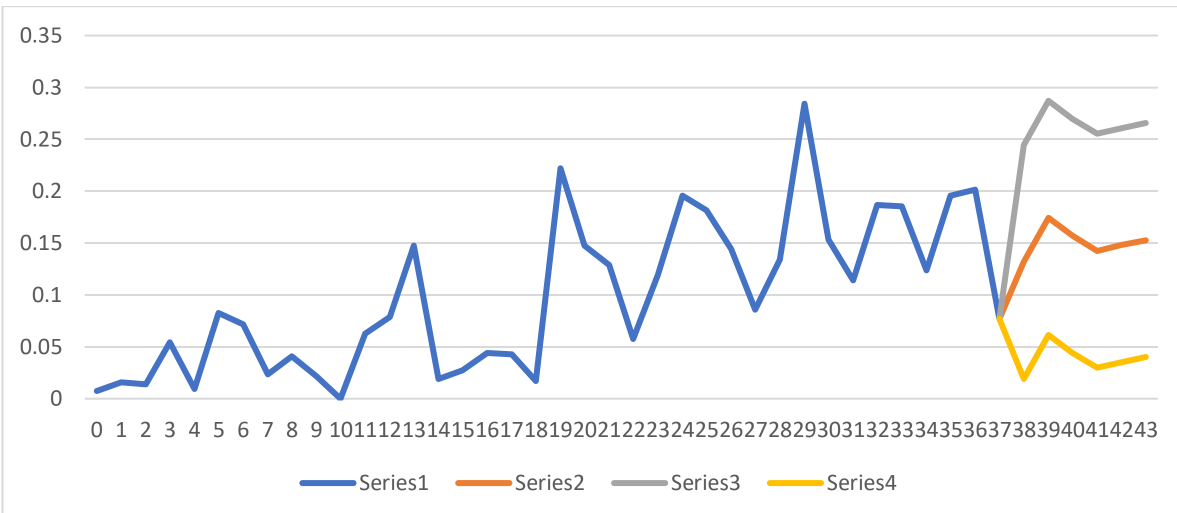




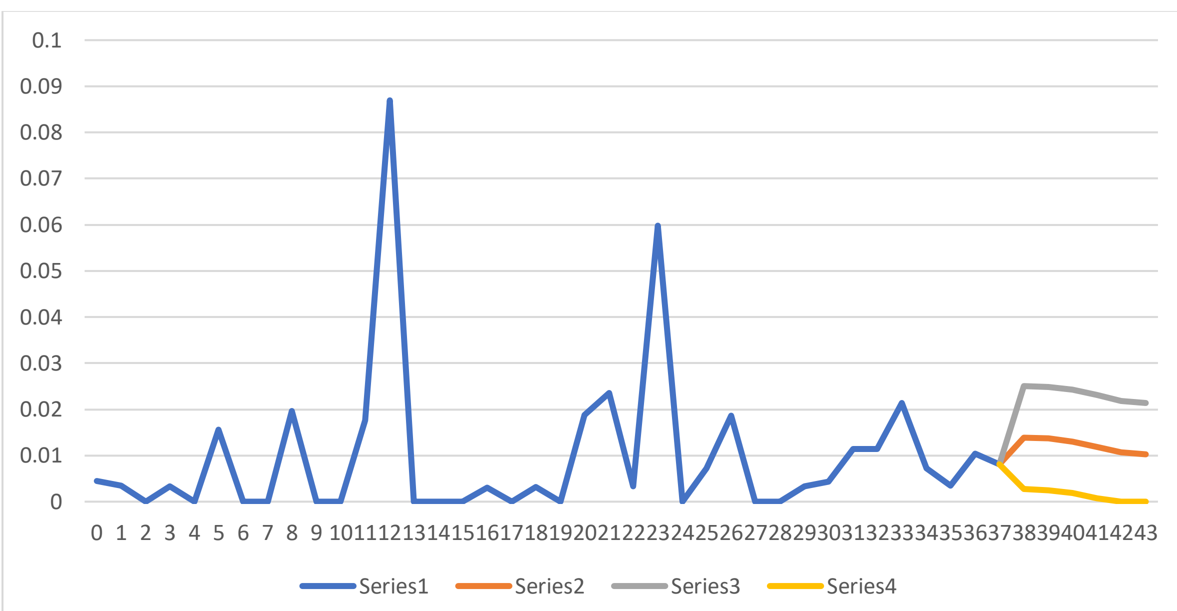




Density
0
−0.2
−0.1
0.0
0.1
0.2
0.3



Series1
Series2
Series3
Series4



Series1
Series2
Series3
Series4

Appendix C

Material Group Improvement

The below table shows us the average number of late days avoided per month when the recommended weights are followed. The implementation date is the date that was used to cutoff the training data from the test data. In general, performance is quite good. That said, one material group, '032', performs marginally worse on average when the recommendation is followed. This is due in large part to the surprising negative turn in performance of one of largest suppliers in September of 2021.

Implementation Date	073	013	017	032	012	014	060	018
5/1/21	3.13	28.55	27.13	0.39	17.01	5.25	27.64	0.01
6/1/21	7.28	29.48	15.91	0.42	17.86	5.58	27.87	0.00
7/1/21	7.04	30.46	21.07	-0.36	18.26	4.62	26.38	0.01
8/1/21	6.94	28.08	-10.14	-10.02	-10.70	4.98	25.31	0.00
9/1/21	4.17	23.04	-18.12	-11.05	14.05	6.72	24.53	0.01
10/1/21	3.66	20.40	-24.82	0.81	22.84	7.33	26.63	0.01
Average Monthly Late Days Saved	5.37	26.67	1.84	-3.30	13.22	5.75	26.39	0.01

Appendix D

The LP Formulation

The below breaks down the objective, decision variables, and constraints used in the LP to find optimal supplier weights. The number of decision variables grows linearly with the number of suppliers in a material group, and the number of constraints grows approximately exponentially with the number suppliers.

It is important to note that the weights found from this formulation are used to validate the algorithm during back-testing, but when we make recommendations to the client, we show them the optimal values along with many other combinations of the top-performing weight values. This is to enable them to keep an element of optionality.

Objective Function:

$\min \sum Z_i \times p_i$ where Z_i is the average days late per month binned value and P_i is the probability of that value occurring.

Decision Variables:

$a_j \ \forall\ j \in \{Suppliers\}$ where a_j is the weight given to supplier j

Constraints:

$Z_i = \sum a_j \times x_{i,j} \ \forall\ i \in \{Combinations\ in\ joint\ distribution\}$ – Taking the weighted sum of individual supplier performances for possible combination i

$\sum a_j = 1$ – Weights must sum to 1

$a_j \geq 0 \ \forall\ j \in \{Suppliers\}$ – Weights must be non-negative

$a_j \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\} \ \forall\ j \in \{Suppliers\}$ – Weights must be in increments of 0.1

Appendix E

Daily Order Delays over Time

The below chart shows average days late in orange and the rolling average of days late in blue. It is evident that post Covid there are significantly more disruptions as seen by the rolling average changes.

Appendix F

Marginal Improvement for Material Group 17

This bar plot shows the actual average late days per month, the average late days per month following the recommended weights, and the marginal average late days saved. Each column represents a different cutoff date for which the algorithm was tested on. All data is for material group 17.