Improving Order Prioritization for the Allocation of Constrained Supply

By
Ashton David Imlay

B.S. Environmental Engineering, Tufts University, 2012

SUBMITTED TO THE MIT SLOAN SCHOOL OF MANAGEMENT AND THE CIVIL AND ENVIRONMENTAL ENGINEERING DEPARTMENT IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREES OF

MASTER OF BUSINESS ADMINISTRATION
AND
MASTER OF SCIENCE IN CIVIL AND ENVIRONMENTAL ENGINEERING

IN CONJUCTION WITH THE LEADERS FOR GLOBAL OPERATIONS PROGRAM AT THE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUNE 2019

Signature of Author: Signature redacted
Civil and Environmental Engineering, MIT Sloan School of Management
May 10, 2019

Certified by: Signature redacted
Dr. Stephen Graves, Thesis Supervisor
Professor, MIT Sloan School of Management

Certified by: Signature redacted
Dr. David Simchi-Levi, Thesis Supervisor
Professor, Civil and Environmental Engineering

Accepted by: Signature redacted
Dr. Heidi Nepf, Chair, Graduate Program Committee
Professor, Civil and Environmental Engineering

Accepted by: Signature redacted
Maura Herson, Assistant Dean, MBA Program
MIT Sloan School of Management

This page has been intentionally left blank

Improving Order Prioritization for the Allocation of Constrained Supply

By
Ashton Imlay

Submitted to MIT Sloan School of Management
on May 10, 2019 in Partial Fulfillment of the
requirements for the Degrees of Master of Business Administration
and Master of Science in Civil and Environmental Engineering

Abstract

For top wholesale retail companies, the demand for products from US-based customers (wholesale, digital, and direct to consumer) is extremely high. However, the available supply of a product is contingent upon the success of long-term forecasting, manufacturers across the globe, and intercontinental transportation. Therefore, there is not always enough supply to meet demand. In these situations, wholesale retailers must decide which orders to prioritize in the allocation of available supply.

This thesis presents a method for improving order prioritization by utilizing readily available data to wholesale retail companies and a method for predicting the effectiveness of the new prioritization methodology utilizing historical data. By prioritizing orders that meet certain characteristics deemed to be in-line with company strategy and simulating multiple conditions, it is possible to deliver improved service on a specific set of orders.

The impact of this work has been verified through a simulation model. The model was used to simulate three months of supply and demand and indicated a possible increase of 10-90% in the number of units made available to ship to specific marketplace segments.

Thesis Supervisor: Dr. Stephen Graves
Title: Abraham J. Siegel Professor of Management

Thesis Supervisor: Dr. David Simchi-Levi
Title: Professor, Civil and Environmental Engineering

This page has been intentionally left blank

Table of Contents

Acknowledgements

First I would like to thank my family: my parents and my sister. My parents encouraged me to take any path I wanted, to move anywhere in the world, and to tackle any new challenge that interested me. I have felt supported and loved unconditionally for my entire life, and that was as important through these few years of graduate school as it was for every year prior. My sister has constantly been a sounding board for any struggle I faced, personal and professional.

Next I would like to thank my girlfriend, Liza Bagley, who encouraged me to take a project that put 3000 miles between us, knowing that it was important for me to work with a company for which I had always dreamed of working. Her support and encouragement through the highs and lows of this project were invaluable.

I could not have succeeded in this project without the advice and help from my academic advisors, Dr. Stephen Graves and Dr. David Simchi-Levi. They grounded my approach in reasoning and careful analysis, and prevented me from skipping valuable steps en route to the completion of a final product, no matter how hard I tried to skip right to the end.

This project would not have been possible without the strategic guidance provided by my Nike manager, Angharad Porteous. Not only did she help me learn many analytical tools, tips, and tricks, she also taught me how to turn data into actionable insights, and how to properly present technically advanced methods in a business setting. Her friendship and willingness to spend countless hours coaching and counseling an intern will never be forgotten.

I would also like to thank my greater support network at Nike. The Global SCI team, specifically Gareth Olds, Mark LeBoeuf, Marcus Schneider, Liz Walker, and Arun Kumar made coming to work each day an inspiring, educational, and laughter-filled activity. Additionally, the LGO network at Nike welcomed me with open arms. Specifically, Jon Frommelt, my Nike mentor, taught me to look at the big picture that surrounds everything we do, be it around my project or around my career.

Additionally, I'd like to thank the North America Marketplace Operations team and others who supported this project – Kristen Hanson, Brian LeBlanc, Harpreet Mithra, Eric Reynolds, and David Taylor – who conceptualized such a wonderful internship project and provided guidance and enthusiasm for supply allocation from beginning to end.

Finally, I'd like to thank my LGO classmates, who constantly inspire me to grow, learn, and evolve.

Note on Nike Proprietary Information

In order to protect information that is proprietary to Nike, Inc., the data presented throughout this thesis has been modified and does not represent actual values. Data labels have been altered, converted or removed in order to protect competitive information, while still conveying the findings of this project.

List of Figures

List of Tables

1. Introduction and Background

For top wholesale retail companies, the demand for products from US-based customers (wholesale, digital, and direct to consumer) is extremely high. However, the available supply of a product is contingent upon the success of long-term forecasting, manufacturers across the globe, and intercontinental transportation. Therefore, although the wholesalers plan to meet the entirety of the demand, the volatility within a global supply chain creates situations where there is not always enough supply to meet demand. Retail wholesalers must prioritize and focus on the high impact marketplace segments: moving from "serve all marketplace segments well" to "serve some marketplace segments extremely well." This thesis investigates methods for improving delivery precision to wholesale customers through the adoption of data-driven internal processes. The research was done in cooperation between MIT Leaders for Global Operations and Nike, Inc.

This thesis is structured as follows: Section 1, this section, provides an overview of the project and how it fits into the overall Nike Supply Chain. Section 2 outlines the current state of retail manufacturers and how decisions related to the problem statement are made. Section 3 presents the methodologies utilized in the completion of this project and the relevant literature that guided their usage. Section 4 describes the models that were developed and how the methodologies were combined and utilized. Section 5 explains the results of the models and the successes of the project. Section 6 concludes the thesis and provides a look forward into future opportunities to continue this research. Section 7 provides the references used in the research.

1.1 Company Overview

It started with a handshake. Nike Inc., based in Beaverton, Oregon, USA, was founded in 1964 by co-founders Phil Knight and Bill Bowerman, initially as Blue Ribbon Sports. Blue Ribbon Sports was formed to distribute shoes made by Onituska, a Japanese shoe company. But in 1971, the relationship with Onitsuka had soured, and Blue Ribbon Sports sold its first shoes under a new shoe line: Nike.[1]

Today Nike is the largest sports apparel and footwear seller in the world. Nike is comprised of several brands, including Jordan, Hurley, and Converse. In 2018, Nike posted revenues of $36 billion with a 43.8% gross margin. Nike sells its products through three main channels: wholesale to customers like Dick's Sporting Goods and Foot Locker; retail through their Niketown and Nike Factory Store retail stores; and direct-to-consumer through Nike.com, Converse.com, and the mobile app based purchasing options. Although it is a US-based company, the US made up only 42% of 2018 revenues, with the remaining 58% coming from international markets. Nike's main competitors are Adidas, ASICS, lululemon athletica, Puma, and Under Armour. Nike footwear is manufactured in 124 footwear factories in 13 countries and Nike apparel is manufactured in 328 apparel factories in 37 countries, all through a contract manufacturing model (i.e., Nike does not own any of the factories making its products). [2]

1.1.1 Nike Supply Chain Overview

Nike's supply chain planning begins many months prior to an item ever landing on the shelf. Nike's designers begin work on products years in advance; however, the supply chain realistically begins once a customer places an order. A "customer" at Nike is typically be considered to be wholesale customers (e.g., Dick's Sporting Goods), but also includes Nike-owned stores (e.g., Niketown, Nike Factory Store) or individual Nike-sponsored athletes (e.g., Lebron James). This order placed by a customer is referred to within Nike as a "sales order." Under the "Futures" model, these sales orders are placed approximately six months prior to the requested delivery date, a type

[1] Knight, Phil, *Shoe Dog: A Memoir by the Creator of Nike*.
[2] Nike Inc., "2018 Form 10-K."

of order introduced by Nike to avoid inventory risk.[3] Once a sales order is received, it is entered into Nike's enterprise resource planning (ERP) software, where it is associated with the distribution center (DC) from which the items from the sales order will eventually be fulfilled. At this point, Nike places a "purchase order" to their network of factories to request that certain styles be manufactured so that they can be delivered on-time. The quantities of the orders placed in the purchase orders are determined by both existing and forecasted orders for those products.

Once the items are made at the factories (a majority of which are located in Asia), the items are transported to their local ports where they are consolidated into volumes large enough to ship in a shipping container. From there, the items travel by ocean freight to a US port (for US-based orders). In times of rushed schedules and the need to make up for upstream delays, some items will be shipped by air freight, however this volume generally makes up a very small percentage of Nike's annual volume. At the port, approximately 50% of items are transferred to intermodal transportation en route to Nike DCs. The rest of the items will be moved to a deconsolidation center, where they will be used to fulfill orders from different large retailers who have chosen to receive their orders directly from Nike, where Nike transports the items directly from the port to the customers' own DCs. Once the inventory is within Nike's DCs, the inventory is allocated to specific orders through an automated supply allocation process that runs nightly in the ERP, matching available inventory with unfilled orders.

1.1.2 Changing Strategy

In 2017, Nike announced a new corporate strategy, called the "Consumer Direct Offense." This new strategy is built upon the "Triple Double," an effort to double innovation, double speed, and

[3] Knight, Phil, *Shoe Dog: A Memoir by the Creator of Nike*.

double direct connections with customers. Doubling innovation means launching more innovative platforms like the Nike React foam or the Air VaporMax, a new version of a Nike Air sole. Doubling speed is Nike's commitment to cutting the duration of a product creation cycle in half. And doubling direct means directing more resources into Nike.com, Nike digital apps, and any other direct-to-consumer channels.[4]

This new strategy plans to drive growth in twelve key cities in ten countries across the world: New York, London, Shanghai, Beijing, Los Angeles, Tokyo, Paris, Berlin, Mexico City, Barcelona, Seoul, and Milan. These are cities that drive style throughout the world and where fashion trends are grown. By better serving these cities, through operations, merchandising, and marketing efforts, Nike expects these cities and their countries to represent over 80% of Nike's projected growth through 2020.[5]

1.2 Project Overview

1.2.1 Project Summary

The primary goal of this research was to develop an improved priority list (or lists) of marketplace segments that will be used as part of the existing algorithm to match available supply of a product with the orders for that product to increase the level of service to specific marketplace segments. To achieve this, an analytical approach involving optimization and simulation was developed to both create new lists of prioritized marketplace segments and compare the performance of the current state with the proposed state, utilizing the number of shippable units made available for orders from each marketplace segment as the basis of the comparison.

[4] Nike Inc., "NIKE, Inc. Announces New Consumer Direct Offense: A Faster Pipeline to Serve Consumers Personally, At Scale."
[5] Nike Inc.

1.2.2 Motivation and Statement of Problem

The motivation for this research project came from the desire within Nike's Global Operations and Logistics department to identify any method to increase delivery precision and to enact the Consumer Direct Offense. Several Nike employees within this group identified the marketplace segment prioritization methodology for supply allocation as a potential "low hanging fruit" opportunity to improve delivery precision to certain marketplace segments. This project is seen as low hanging fruit because it could potentially improve delivery precision without requiring significant investments or process improvements.

When limited supply is allocated to orders from customers, a prioritized list of marketplace segments is used as an important factor into how the supply is allocated. There is an opportunity for Nike to leverage their readily available data and in-house analytics skills to understand how changing the prioritized list of marketplace segments could improve their delivery precision. The specific opportunities being addressed by this research is that: (i) Nike could better utilize the massive amounts of data they have collected over time regarding marketplace segment ordering patterns to improve the allocation of supply; and (ii) Nike could better understand how altering the prioritized list of marketplace segments could be used to improve delivery precision to various marketplace segments.

1.2.3 Anticipated Outcome

The anticipated outcome of this research project is three-fold: (i) a methodology to create a prioritized list of marketplace segments based on customer data to be used in the supply allocation process; (ii) a simulation model to simulate historical scenarios to test whether a specific prioritized list of marketplace segments would have performed better (as measured by quantity of units made shippable through each list) than another; and (iii) an understanding of whether or not the marketplace segment prioritization list can be manipulated to improve delivery precision to marketplace segments considered important to Nike. Additionally, a time schedule will be developed which will outline when and how often to use the new methodology and simulation model to update the priority list over time.

1.2.4 Project Approach

The project consisted of four distinct phases in order to achieve the above stated goals: current state analysis (including analysis of data gaps), development of the marketplace segment prioritization algorithm, development of the allocation simulation model, and the testing and running of the algorithm and model.

The current state analysis involved interviews with employees familiar with the allocation process and the delivery precision metrics. Additionally, a review of available data regarding the supply allocation process was analyzed. Once the current data was understood, data gaps were identified and processes were put in place to collect the data required.

After the current state was understood, the prioritization algorithm and the allocation simulation model were developed concurrently, as they are separate except for the fact that the allocation simulation model requires output from the prioritization algorithm. The approach for the prioritization algorithm was developed initially, and the marketplace segment characteristics were selected over time after numerous meetings with multiple stakeholders within several Nike departments where business needs and priorities were identified. The simulation model was developed iteratively.

Once the algorithm was finalized and the model was functioning, they were used to identify three strategies for potential implementation. At this point, a preliminary implementation plan was developed. However, the handoff to a full-time Nike employee was seen as the ultimate culmination of the work. This project was seen as a proof of concept; thus, a successful handoff indicated that the project was a success, and efforts to develop the next iteration of this project, an operational version, was begun.

2 Current State

Retail manufacturers must prioritize orders in some fashion, perhaps through a simple first-in, first-out system, or through custom algorithms that can be generically described as using pre-determined demand and supply matching logic (DSML). The DSML typically represents a small, but impactful, part of the process to assign units of a specific item (referred to as a Stock Keeping Unit, SKU) to a specific order placed by a customer.

2.1 Current Allocation Prioritization

The process for allocation prioritization at a retail manufacturer begins when an order is placed by a customer and is entered into the retailer's ERP software system. The customer will typically be guided into placing an order that only includes products from a single product engine (e.g., apparel, footwear, or equipment) if the retailer segments its products into different DCs. This is done such that the order can be fulfilled from a single DC, as a product engine is not split across multiple distribution centers. Alternatively, if DCs typically hold multiple different types of products within the same DC, this requirement is not necessary. The order will be a specific type of order (e.g., a Futures order, indicating the order was placed months prior to the planned delivery date, or an At-

Once order, indicating the order is to be fulfilled as soon as possible upon order placement), indicating when it is placed with respect to the requested delivery date, if the order will be received all at once (a regular order) or kept in the distribution center (a contract order) to be available when the customer asks for it (a customer request from a contract order) or other different ways to deliver an order.

At that point, the order is present within the retail manufacturer's ERP and the next step in the supply allocation process doesn't typically occur until the month prior to the month that contains its scheduled delivery date. At this point, the DSML process will review all orders within the ERP along with the complete picture of the retail manufacturer's inventory, and begin to sort through the data to match supply with demand.

A sample DSML process, run on a daily basis, works as follows:

1. Select a specific location where inventory is held.
2. Select all orders that are scheduled to be fulfilled from that location.
3. Select orders with delivery dates that range from the next month, the current month, or any past month (i.e., orders that are already late).
4. The orders are sorted according to the sales order type, with order types typically sorted in a custom hierarchical manner.
5. Within the buckets created by the previous steps, each order is then sorted by month of the delivery date, from earliest date to latest date (e.g., if today is 5/1/2015, an order to be delivered 5/5/2015 will be higher on the list than an order to be delivered 6/6/2015).
6. Within the buckets created by the previous steps, each order is typically then sorted by:
    a. An internal ranking of marketplace segments, usually decided by company strategy.

b. Order Sub-Type (i.e., further delineation of orders that fall into the "Futures" or "At-Once" order type; retail manufacturers may have 5-20 sub-types of orders within an order type).

c. Date at which the inventory must be present at the inventory storage location to be able to be delivered on-time.

d. The identifier for the order assigned by the ERP, often used as a final tie-break for orders which have very similar order characteristics. This typically indicates that of two extremely similar orders, the order placed first will be sorted above the order placed second.

7. Select all inventory stored within, en route to, or assigned to be delivered to the same inventory location as selected above.

8. Sort the inventory such that inventory present in the inventory location is first, inventory en route to the inventory location is second, and inventory still at the factory but assigned to the delivered to the inventory location is last.

9. Now there is a sorted list of inventory (the supply), and a sorted list of orders (demand). It is important to remember that each order typically has multiple SKUs (multiple products and multiple sizes and colors of each product). Starting with the order at the top of the list, select the order and review the inventory. Once an item of inventory that matches an item on the order (e.g., a specific women's shoe in size 7 in the red/white colorway) is found, match the inventory to the order.

   a. If there is enough inventory at the location to fill the entirety of the order (the entire volume of all SKUs included on the order):

i. The order is then indicated as "shippable" and will be set aside to be dropped to the warehouse management system (WMS) of the selected distribution center once the entire DSML process is complete. This will tell the DC that the order can be shipped. At this point, the inventory assigned to this order is fully matched, and cannot be assigned to another order.

ii. The order will then be planned by the DC as to when it will be picked, packed, and shipped.

b. If there is not enough inventory at the location to fill the entirety of the order:

i. The inventory available is matched to the order and will not be matched to any further orders reviewed that day. For example, if Company A orders 500 units of SKU X and 100 units of SKU Y, and the location has 300 units of SKU X and 300 units of SKU Y, Company A's order will be matched to 300 units of SKU X and 100 units of SKU Y. Company A's order is not marked as shippable because the entire volume of SKU X is not present. However, the DSML will then consider the location to only contain 0 units of SKU X and 200 units of SKU Y. Therefore, the 300 units of SKU X and 100 units of SKU Y are still sitting in the inventory location but will not be matched to a different order. This logic means that there could be many units sitting within the inventory location each day that could have been matched to a different order that could have then been shipped, but due to the sorting logic, that will not happen.

10. Once the matching of supply to the demand occurs, all orders indicated as shippable and sent to the WMS for fulfillment are removed from the set of input data of outstanding

demand. All orders that were not fully matched to available stock are disassociated from the stock they were matched with: the orders once again go through the DSML when the process is run again the next day, and the stock previously matched to that order is considered available to be matched again the next day. Only when an order is fully matched is the order and stock considered to be a "hard allocation," all others are "soft allocations" and will be undone at the end of that day's allocation process.

a. A caveat to this step may exist for contract orders. Contract orders are contracts between the retail manufacturer and their customers which state that as of a certain date, the retail manufacturer will have the full volume of the customer's order present in the inventory location ready to deliver. But the delivery will not occur until the customer requests a delivery and indicates when the delivery should take place and the quantity of the delivery, pulling stock from the contracted volume they indicated in the contract. The customer requests can occur at any time (within a set period of time as outlined in the contract, typically a full season).

   The caveat to Step 10 for contract orders is that when any on-hand stock is matched to a contract order, the stock is often considered to be permanently allocated to the order, even if the entire volume of the contract order was available and therefore the order was not considered to be "shippable."

2.2 Current Marketplace Segment Service Metrics

There are numerous metrics by which retail manufacturers' supply chains measure performance. The most applicable metrics to this project are: Shipped In Full On Time (SIFOT) and Delivered In Full On Time (DIFOT). Both metrics are often utilized as umbrella metrics to determine the effectiveness of the supply chain overall. The downside of using these metrics is that SIFOT is dependent on the entirety of the supply chain that occurs upstream of distribution center outbound and DIFOT is dependent on the entire supply chain. Therefore, when undertaking projects that are designed to improve specific portions of the supply chain, SIFOT and DIFOT measures are generally inadequate to accurately measure the improvements from a specific project due to their lack of granularity.

3 Methods and Literature Review

The methods outlined within this section were utilized within the development of the marketplace segment prioritization algorithm. The logical and/or mathematical backing for each technique is presented below.

3.1 Customer Prioritization

Retail manufacturing companies, or wholesale companies more broadly, can have thousands upon thousands of customers. When companies choose to interact directly with their customers, rather than relying on middle-man distributors, they have a choice to make: treat all customers equally or prioritize specific customers over others. It has become commonplace in business for companies to select strategic customers, whom they treat differently from the average customer.[6] In marketing for example, customer prioritization is shown to increase customer satisfaction and profitability from prioritized customers.[7] Further, if a company is able to prioritize customers such that the

[6] Lacoste, "From Selling to Managing Strategic Customers - a Competency Analysis."
[7] Homburg, Totzek, and Droll, "All Customers Are Equal, but Some Are More Equal Should Firms Prioritize Their Customers ?"

prioritized customers receive higher rates of on-time delivery, transaction volume and unit prices for units sold to those customers could increase.[8]

3.2 List Comparison

Prioritizing marketplace segments produces ordinal data: a list of items in which the order is important. In situations with two ordinal lists, a rank correlation coefficient can be used to determine the similarity between the two lists. For example, if two people were to rank their favorite fast food restaurants, the similarity between the two lists could be calculated with a rank correlation coefficient. Two of the most popular coefficients are Kendall's τ and Spearman's ρ.[9]

Kendall's τ is defined as:

$$\tau = \frac{2S}{n(n-1)}$$

where n = length of each of the two lists and S = the number of concordant pairs minus the number of discordant pairs.[10] The notion of concordance is explained as follows:

> If (x_j, y_j) and (x_k, y_k) are two elements of a sample $\{(x_i, y_i)\}_{i=1}^{n}$ from a bivariate population, one says that (x_j, y_j) and (x_k, y_k) are concordant if $x_j < x_k$ and $y_j < y_k$ or if $x_j > x_k$ and $y_j > y_k$ (i.e., if $(x_j - x_k)(y_j, -y_k) > 0$); and discordant if $x_j < x_k$ and $y_j > y_k$ or if $x_j > x_k$ and $y_j < y_k$ (i.e., if $(x_j - x_k)(y_j, -y_k) < 0$). There are $\binom{n}{2}$ distinct pairs of observations in the sample, and each pair (barring ties) is either concordant or discordant.[11]

[8] Peng and Lu, "Exploring the Impact of Delivery Performance on Customer Transaction Volume and Unit Price."
[9] Agresti, *Analysis of Ordinal Categorical Data*.
[10] Nelsen, "Kendall Tau Metric."
[11] Nelsen.

For example, the List A is "A, B, C, D, E, F, G, H, I, J, K, L" and list B is "A, B, D, C, F, E, H, G, J, I, L, K." Table 1 outlines how the Kendall's τ coefficient is calculated between Lists A and B:

Table 1. Kendall's τ Example

Item	Location of Item in List A	Location of Item in List B	Concordant Pairs	Discordant Pairs
A	1	1	11	0
B	2	2	10	0
C	3	4	8	1
D	4	3	8	0
E	5	6	6	1
F	6	5	6	0
G	7	8	4	1
H	8	7	4	0
I	9	10	2	1
J	10	9	2	0
K	11	12	0	1
L	12	11		
		Total:	61	5

$$\tau = \frac{2(61-5)}{12(12-1)} = 0.85$$

A perfect relationship between lists would have a score of τ = 1, and lists with no relationship would have a score of τ = 0. The example presented above has a Kendall's τ value of 0.85, therefore the lists are considered to be quite similar.

Spearman's ρ is defined as:

$$\rho = 1 - \frac{6\sum_{i=1}^{n}(R_i - S_i)^2}{n(n^2-1)}$$

Where $R_i = rank(x_i)$ and $S_i = rank(y_i)$ from the sample $\{(x_i, y_i)\}_{i=1}^{n}$ and assuming all ranks are distinct integers.[12] Utilizing List A and B described above, the Spearman's ρ for those lists would be calculated as follows:

Table 2. Spearman's ρ Example

Item	Location of Item in List A (R_i)	Location of Item in List B (S_i)	($R_i – S_i$)	$(R_i – S_i)^2$
A	1	1	0	0
B	2	2	0	0
C	3	4	1	1
D	4	3	1	1
E	5	6	1	1
F	6	5	1	1
G	7	8	1	1
H	8	7	1	1
I	9	10	1	1
J	10	9	1	1
K	11	12	1	1
L	12	11	1	1
		Total:	10	10

[12] Nelsen, "Spearman Rho Metric."

$$\rho = 1 - \frac{6\sum_{i=1}^{n}(R_i - S_i)^2}{n(n^2-1)} = 1 - \frac{6 * 1}{12 * (12^2 - 1)} = 0.9965$$

Similar to Kendall's τ, a perfect relationship between lists would have a score of Spearman's $\rho = 1$, and lists with no relationship would have a score of $\rho = 0$. The example presented above has a Spearman's ρ value of 0.9965, therefore the lists are considered to be quite similar. This matches the conclusion drawn from the Kendall's τ example.

3.3 Clustering Methodology

Clustering is the process of splitting data into groups of similar data points such that the groups are dissimilar from one another.[13] Clustering can be either hierarchical or partitional. Hierarchical clustering employs a tree-like structure through which the data set is broken down into more and more clusters. Partitional clustering employs an iterative approach, where clusters are attempted and compared to previous iterations of clusters.

"K-means" is an example of a simple and popular method of partitional clustering.[14] The "k" stands for the number of clusters that are used to represent the data. Unlike hierarchical clustering, partitional clustering requires the user to decide what value of k is appropriate for the data. However, one limitation of the k-means algorithm is that it cannot receive as input a pre-calculated distance matrix (a matrix of similarities between different data points).

A similar method to k-means that can receive a pre-computed distance matrix as input is "k-medoids." A medoid is defined as "the object of a cluster, whose average dissimilarity to all

[13] Caruso et al., "Cluster Analysis as a Decision-Making Tool."
[14] Macqueen, "SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS."

the objects in the cluster is minimal i.e. it is a most centrally located point in the given data set."[15] This is different than the mean or the centroid of a cluster, as they are not necessarily associated with a specific object in the cluster. The k-medoids algorithm is less sensitive to outliers than k-means, as an outlier can drastically change the mean of a set of data but the medoid will be less affected by a similar outlier.[16] The k-medoids algorithm, specifically the "Partitioning Around Medoids" heuristic (PAM), functions as follows:

Input:

k: The number of clusters

D: A data set containing n objects

Output: A set of k clusters that minimizes the sum of the dissimilarities of all the objects to their nearest medoid.

Method: Arbitrarily choose k objects in D as the initial medoids. Assign each remaining object to the cluster with the nearest medoids. Calculate the total sum of distances (all distances between the items of each cluster and the medoids of that cluster, for every cluster).

Repeat: Randomly select one non-medoid object, swap it with any one of the initial medoids. Once the swap was made, re-assign all objects into clusters by assigning them to the closest medoid, then calculate the total sum of distances (all distances between items in cluster and medoid of cluster). If the total sum of distances for all clusters increases,

[15] Patel and Singh, "New Approach for K-Mean and K-Medoids Algorithm."
[16] Park and Jun, "A Simple and Fast Algorithm for K-Medoids Clustering."

undo the swap and try again.[17] If the total sum of distances is lowered by swapping one non-medoid object with one of the initial medoids, that object will be the medoid of a cluster from then on, until a better medoid is found.

3.3.1 Selecting Optimal Number of Clusters

As stated above, partitional clustering methods typically require the user to select the number of clusters, k, that best represents the data. To do this, the elbow method is typically employed. Thorndike presented the idea of the elbow method in 1953, indicating that the optimal number of clusters for a data set can be found by plotting the sum of squared errors (error indicated as the distance between each item in a cluster and the representative center/centroid/medoid/etc. of that cluster) against the number of clusters, k.[18] The essence of this idea is that the number of clusters should be selected such that adding an additional cluster does not drastically improve the representation of the data.[19]

4 Model Development

The analysis completed within this thesis follows the following process: (i) identify important marketplace segment characteristics; (ii) prioritize marketplace segments through an algorithm that takes the marketplace segment characteristics in as inputs; (iii) select a few prioritized lists of marketplace segments to test to see if they perform better than the current state (a specific prioritization of marketplace segments utilized by a retail manufacturer today); and (iv) simulate a DSML process to determine which list (either the current state list or one of the few newly created

[17] Velmurugan, "Computational Complexity between K-Means and K-Medoids Clustering Algorithms for Normal and Uniform Distributions of Data Points."
[18] Thorndike, "Who Belongs in the Family?"
[19] Bholowalia, "EBK-Means: A Clustering Technique Based on Elbow Method and K-Means in WSN."

prioritized lists) performs better (a list will be determined to be "best" if it increases the total number of units that are able to be shipped within the time period for which the simulation was run). Assumptions and simplifications were made within the simulation to reduce computation time.

4.1 Demand and Supply Matching Logic Simulation

The development of the allocation simulation involved replicating the example DSML process (outlined in Section 2.1) into Python code. Several simplifications and assumptions were made in the modeling process. The first simplification that was made was reducing the scope of the demand and supply that were being analyzed by reducing the number of inventory locations that were analyzed, thus reducing the complexity and computation time required. Another similar assumption and simplification was made: rather than simulate the allocation of all inventory (regardless of its status as at-factory, in-transit, or on-hand in the DC) and all demand, the simulation was limited to just on-hand inventory. This simplification prevents the simulation from providing insights and KPIs that are tracked in the months leading up to the delivery date of an order, such as to "coverage" (defined loosely as the percentage of an order's volume that is estimated to be delivered on-time based on inputs ranging from under-ordering to transit delays). However, along with the simplification also comes the ability to simulate the allocation process with a reasonable amount of computation time. Therefore, since the project motivation was how to improve delivery precision to the marketplace segment, the sacrifice of such forward looking KPIs for the sake of reduced computation time was in-line with the project goal.

Another assumption was related to the input data that was collected. Demand data was collected from demand that was present in the system each day that data was collected. Inherently, since the demand data was collected once per day (data used in this project was collected daily at approximately 8am Pacific Daylight Time), any at-once orders that were placed and fulfilled in a single day were not captured in the dataset. Additionally, the first day in which an order was identified in the data, the order was assumed to exist in that state until either the order is filled or the simulation period is completed without the order being filled. This means that cancelled orders, returned orders, and orders in which the quantity was changed would not be captured in the simulation.

Another assumption related to demand involves contract orders. As explained in Section 2.1 Current Allocation Prioritization, contracts and their associated customer requests require the customer to request the delivery. As such, contract orders have a higher return rate than a typical order. For the simulation, since this proof of concept project was planned to be used to simulate allocation over one to three-month simulation periods, it was assumed that the cancelled volume from contract orders would not be a significant volume of stock, and was therefore ignored.

Similarly, assumptions had to be made with the supply data. Data was collected as daily snapshots of inventory (location and quantity of each SKU). However, this data could not be directly input into the allocation simulation because it is inherently affected by the existing allocation process that occurs daily. Therefore, we also collected the inbound supply information, indicating what SKUs were arriving at the DCs each day. With this information, the allocation simulation can utilize the snapshot of inventory associated with the first day of the simulation, subtract the volume that is matched to orders and considered to be "shippable" by the allocation simulation, and then add in the quantity of each SKU that arrived at the DC that day. This assumption, however, does not take into account returned orders that were returned to the available pool of on-hand inventory.

4.2 Marketplace Segment Characteristic Weighting and Ranking

The marketplace segment characteristic weighting and ranking process is the main method developed for this thesis. This section explains how characteristics of marketplace segments can be used to create a better prioritization of the segments to be used within the DMSL process. It is broken into three steps: (i) select marketplace segment characteristics; (ii) collect the data; and (iii) run an algorithm to rank the marketplace segments.

4.2.1 Selecting Marketplace Segment Characteristics

The first step to prioritizing marketplace segments requires a subjective decision to be made: what characteristics will you include to prioritize the segments. This decision is a subjective one because it requires a careful understanding of the company's overall strategy. The question, "what defines an important marketplace segment?" must be answered first. However, a company's strategy can, and should, change over time. From one season to another the priority of the company may change. As a result, the algorithm was built modularly, such that marketplace segment characteristics can be changed each time the marketplace segments are re-prioritized.

For this proof of concept project, eight marketplace segment characteristics were selected, as explained below.

1. Different parts of the United States are viewed differently by retail manufacturers, in terms of strategic, financial, and operational factors. Therefore, geographic importance was included as marketplace segment characteristic.

2-3. Retail manufacturers typically have tiers of products, which incorporate price, design, and other similar factors. These tiers were incorporated into the model to represent the different types and price points of products.

4-8. Marketplace segments for retail manufacturers are represented by various characteristics: if they sell online; if they are one-stop shop department stores or specialized boutiques; if they are internally branded stores (e.g., a Columbia Sportswear store selling only Columbia Sportswear products); if they order in a specific manner; if they are considered higher end stores or bargain stores; if they are in malls, strip malls, or stand-alone stores; if they market towards specific demographics; etc. Five characteristics similar to those mentioned above were included within the analysis.

4.2.2 Collecting Marketplace Segment Data

Satisfactory data availability is a key assumption for successful implementation of a project like this. When selecting which characteristics to use for the marketplace segment prioritization, not only must the company strategy be considered, but also the ease of acquiring the information. For example, product sell-through information by customer could be a valuable characteristic to utilize when prioritizing marketplace segments. However, that information was not available for inclusion into this proof of concept project.

For this project, marketplace segment characteristics selected were available through Nike's existing pipeline of supply chain data. A script was developed to collect, calculate, or decipher the marketplace segment characteristics from available data and import them into a format readable by the prioritization algorithm.

4.2.3 Marketplace Segment Prioritization Algorithm

Once the characteristics for marketplace segment prioritization are determined and the data is collected, the data is run through an algorithm to turn the collected data into a prioritized list of marketplace segments. The recommendation presented by this thesis is a simple optimization-based approach for prioritizing marketplace segments using simple marketplace segment characteristics. This approach was utilized to promote understanding amongst future project stakeholders and, therefore, the likelihood of adoption of the process. This approach was also selected because within multiple business units of a retail manufacturer it is often difficult to identify which characteristics of the marketplace segments are more important. This approach collects the top characteristics and then uses an algorithmic approach to weight each characteristic differently, thus eliminating potential biases of different business units. The optimization was formulated as follows:

<u>Objective Function</u>

$$\max \sum_{i} \sum_{j} \left(\frac{X_{ij}}{j}\right) [w_1 * a_{i_1} + w_2 * a_{i_2} + \cdots + w_k * a_{i_k} w]$$

<u>Constraints</u>

$$X_{ij} = binary$$
$$\sum_{i} X_{ij} \leq 1 \; \forall j$$
$$\sum_{j} X_{ij} \leq 1 \; \forall i$$

<u>Decision Variables</u>

$$X_{ij} = \begin{cases} 1 & if\ marketplace\ segment\ i\ has\ rank\ j \\ 0 & otherwise \end{cases}$$

<u>Parameters</u>

$$w_k = weight\ for\ each\ characteristic$$
$$a_{i_k} = characteristic\ for\ marketplace\ segment\ i$$

This optimization formulation will output a sparse matrix, where each row and column contains a single non-zero entry: a "1" representing that a marketplace segment (column), i, is associated with a specific rank (row), j. This optimization function can also be considered to be a sorting function, which will place the marketplace segment with the highest score (each weight multiplied by each marketplace segment characteristic) at the top of the list and the marketplace segment with the lowest score at the bottom of the list.

With a formulation such as this, the weights assigned to each marketplace segment characteristic are just as important as the marketplace segment characteristics themselves in determining the overall score for each marketplace segment. There are two main options for assigning weights: (i) set the inputs according to relative importance of each characteristic as defined by the business leaders championing the project; or (ii) attempt to find the weights that provide the best prioritized list of marketplace segments (explanation of how the "best" list of marketplace segments is determined is discussed in Section 4.3 Selection of Weights Through Prioritization List Comparison and Reduction). The answer to this choice is dependent on the status of the project. As the proof of concept, each idea utilized in this project was, in some way, a novel idea for retail manufacturers. In this case, the idea of using specific marketplace segment characteristics (some which are traditionally captured in performance management reporting) to prioritize marketplace segments, is the next step in toward adoption of data and willingness to trust and rely on data-driven insights within retail manufacturers. Therefore, at the time this project was conducted, there was not a clear understanding of which of the eight marketplace segment characteristics used was more important than another. As an illustrative example, if two marketplace segment characteristics used were (i) how many pairs of socks a marketplace segment ordered and (ii) how many of their orders were requested to be delivered on a Wednesday, it was not immediately clear if a higher weight should be applied to the number of socks or the number of deliveries on a Wednesday. While both had been deemed important and in line with current corporate and operational strategies, there was not a clear hierarchical relationship between the two. Therefore, the weights assigned to each characteristic must be determined by the overall algorithm, rather than being directly assigned.

This method does not allow for two marketplace segments to receive the same rank. It is likely that orders from two marketplace segments with identical ranks would be sorted due to one of the many order characteristics that the DSML process sorts upon, thus indicating that allowing ties in the ranking process is not a hindrance to proper demand/supply matching. However, not allowing ties provided a simpler environment for the programming of the algorithm.

4.3 Selection of Weights Through Prioritization List Comparison and Reduction

As explained in Section 4.2.3 Marketplace Segment Prioritization Algorithm, the marketplace segment prioritization algorithm requires the calculation of the "best" weights to be assigned to each marketplace segment characteristic. For the sake of completeness, the ideal method of determining the "best" weights would be to create every possible list of marketplace segments available with the marketplace segment prioritization formulation and the requirement that the weights sum to one, a set of lists created with every possible combination of weights, and to run the allocation simulation with each list. Then, with the output in hand from every possible combination of weights, one list would have performed better, at least marginally, than all others. The weights that were used to create that list could be assumed to be the best weights and that list would be operationalized moving forward.

However, the computation time required with currently available resources and the current iteration of the allocation simulation model would be infeasible. At the culmination of the project, the allocation simulation model took approximately one hour to simulate the DSML process for one day. When considering the desire within retail manufacturers to simulate entire months or entire seasons, the time required for simulation would be infeasible. Therefore, a more practical and less time intensive solution was required.

While there were many marketplace segments that were being prioritized, it was assumed that if every combination of possible weights were used, there would be several clusters of lists that were similar to one another. This assumption allowed the weights to be selected as follows:

1. Randomly select N combination of weights (in this case, eight numbers that summed to one).
    a. The N combinations of weights were each calculated with a random number generator to select a number between one and ten, and then each weight was divided by the sum of the eight weights such that the values all summed to one.
2. Run the marketplace segment prioritization algorithm with each set of weights (as discussed in Section 4.2.3) to create N lists of marketplace segments.
3. Create a N x N matrix where each entry represents the similarity (using Kendall's τ or Spearman's ρ) between the two lists being compared (the diagonal should be 1, indicating that each list compared with itself is found to be perfectly similar.
    a. Spearman's ρ was utilized as it was determined to be faster and produce similar results to Kendall's τ.
4. Run a k-medoids clustering algorithm on the matrix of similarities M times, where each iteration was given a random set of values as the initial guesses for cluster medoids (as required by the pyclustering python package used to implement the k-medoids clustering method).
    a. The initial set of values for cluster medoids were calculated with a NumPy random choice function from all possible lists without replacement.

b. Clustering was completed using the k-medoids function from the pyclustering package in Python.

5. Review the M clustering outputs and identify the most common output of clustering medoids

    a. M was typically in the range of 10^3-10^4, therefore the output was large enough to identify the set of medoids that was output most commonly.

6. Identify the k clusters, the medoid of each cluster, and the standard errors of the clustering (the difference between each medoid and all lists considered to be clustered around it).

7. Repeat steps 4-6 for all values of k to be analyzed.

8. Select the optimal value of k.

    a. As explained in Section 3.3.1 Selecting Optimal Number of Clusters, the optimal value of k for a clustering algorithm can be found by "finding the elbow" in the plot of errors. However, an additional element was added to the modeling process: a review of the number of times a particular set of medoids was output by the clustering algorithm. If the optimal value of k per the location of the elbow were k=5, but with k=5 there was no clear set of medoids that was output more often than the others, then k=5 would not be a valid solution for this data. Therefore, both the traditional elbow and the medoid output results were reviewed to determine the optimal value of k for this dataset.

9. Run the allocation simulation k times, once per medoid (i.e. prioritized list of marketplace segments) that was discovered.

10. Identify the "best" list of the k lists that were run through the allocation simulation and the weights that were used to create that list.

5 Model Results

5.1 Selecting the Marketplace Segment Lists to Simulate

The first task of the model was to specify which prioritized lists of marketplace segments should be simulated with the allocation simulation portion of the model. After creating 10,000 lists with 10,000 sets of the eight randomly selected weighting factors, the lists were run through the process outlined in Section 4.3 Selection of Weights Through Prioritization List Comparison and Reduction. Figure 1 presents the sum of squares of the errors (squared sums of the Spearman's ρ differences between each medoid and every list in its respective cluster). Figure 2 presents the sum of absolute errors and was included to provide another view of the possible elbow location.

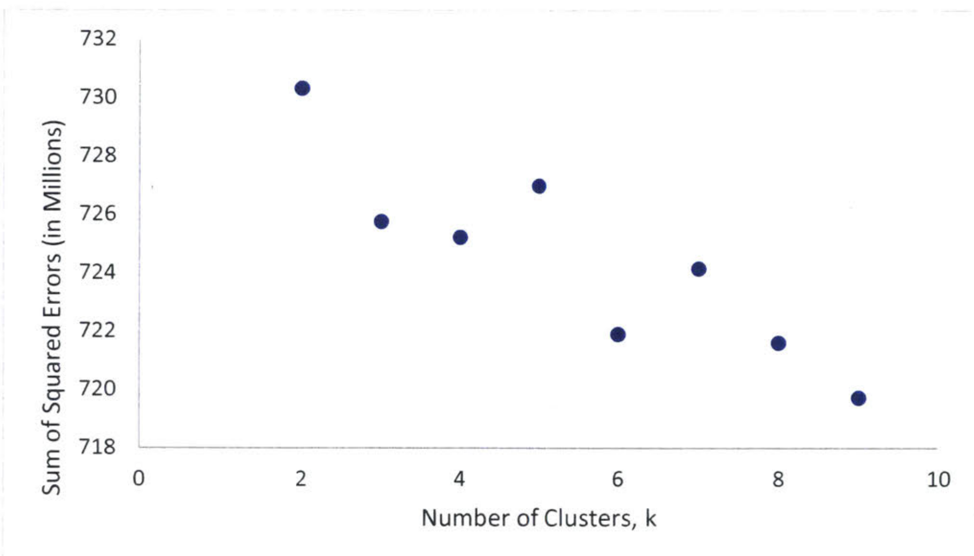

Figure 1. Sum of Squared Errors by k-value

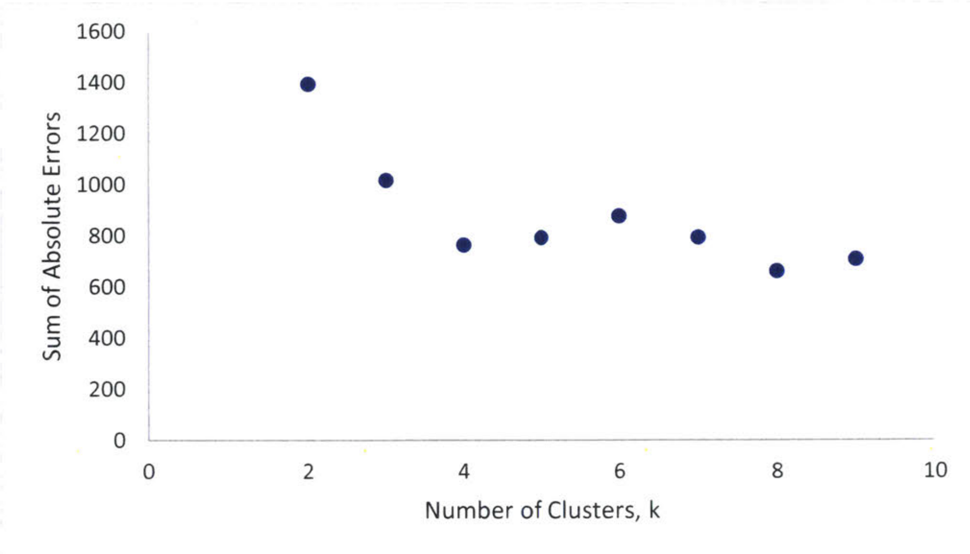

Figure 2. Sum of Absolute Errors by k-value

As shown in Figure 1, error values increased when the numbers of clusters were increased from four to five and from six to seven. This runs counter to the logic of clustering which typically assumes error will decrease when the number of clusters is increased, with diminishing returns on the error reduction as the number of clusters increases. These two instances of error increasing when a cluster is added is most likely explained by a small sample size. If a larger number of iterations was run, the errors would most likely adhere to the typical pattern of decreasing error with increasing values of k. Alternatively, the changes in error after k=4 could be considered to be noise based on the low percentages changes (~0.5%) as k increases.

Additionally, as stated in Section 4.3, due to the bias in the k-medoids clustering method towards selecting the initial guesses as the output medoids, the number of times each set of medoids was output per k value was also investigated. Figure 3 and Figure 4 present the histograms of each set of medoids that was output by the analysis and how many times each list was output for k=3 and k=4, respectively. Each unique set of medoids that was output by the algorithm was assigned a unique identifying number, shown on the x-axis of Figure 3 and Figure 4. It was assumed that the best set of medoids to represent the data would be the set that was output by the clustering method the most times.

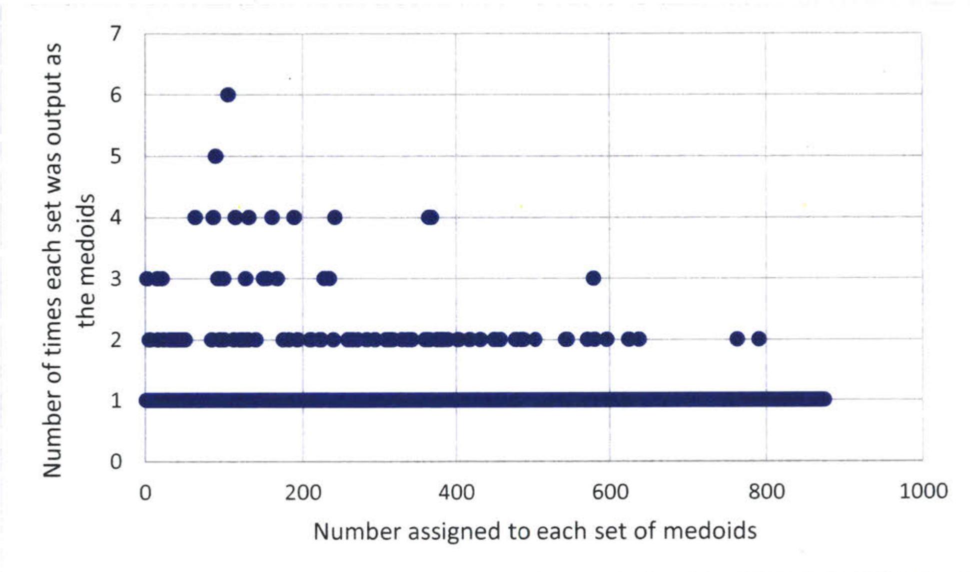

Figure 3. Histogram of Medoid Output, k=3

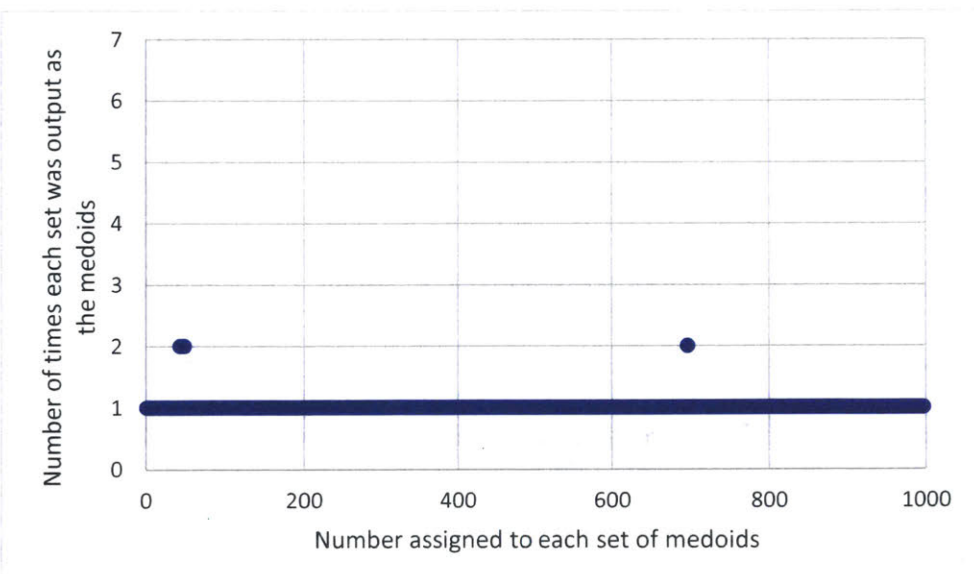

Figure 4. Histogram of Medoid Output, k=4

While the elbow in Figure 1 and Figure 2 could be considered to be either at k=3 or k=4, the two histograms presented in Figure 3 and Figure 4 indicate that k=3 is a better fit to this data. With k=3, one set of medoids was selected six times by the algorithm, another set was selected five times, and nine sets were selected four times. With k=3 there is a clear indication that the set of medoids that was output six times is likely to be the best representation of the data. However, with k=4, three different sets of medoids were each output twice by the clustering algorithm, providing no clear indication that any set of medoids is better than another. For these reasons, as well as the fact that the elbow in the sum of squares error plot is at k=3, k=3 was used in the remainder of the analysis.

5.2 Analysis of the Lists of Marketplace Segments to be Simulated

The three (since the data was determined to be best represented with k=3) selected lists can be reviewed in two ways: (i) reviewing the weights that were used to create the list; and (ii) understand the list by identifying where particular types of marketplace segments are located in the list. The first method is visualized in Figure 5, which identifies what weighting value was applied to each marketplace segment characteristic to create each of the lists. The three lists are identified as Strategy A, Strategy B, and Strategy C.

As indicated in Section 4.2.1 Selecting Marketplace Segment Characteristics, eight marketplace segment characteristics were selected. Characteristics 1, 2, 5, 6, and 8 are each binary characteristics. Characteristics 3, 4, and 7 represent average values, ratios, or other such statistics regarding marketplace segment behavior.

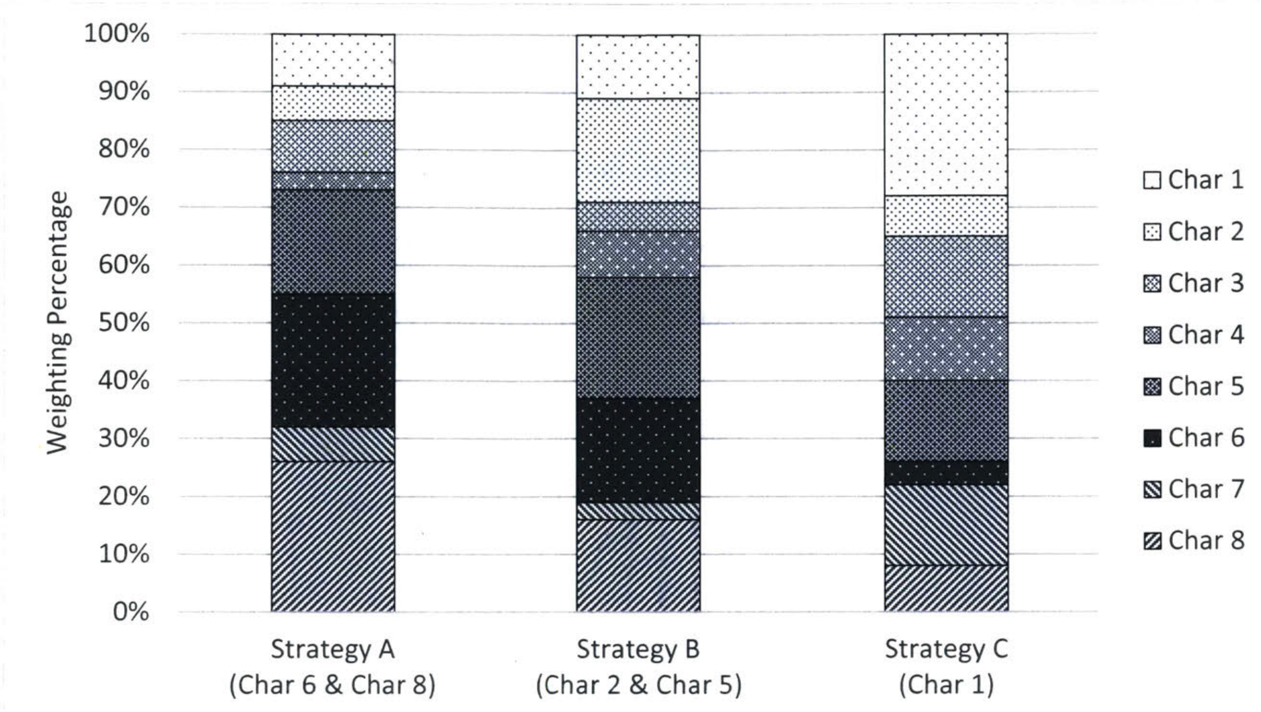

Figure 5. Weighting Factors Used to Create Each List of Marketplace Segments

Strategy A is most heavily weighted on characteristics 6 and 8, Strategy B is most heavily weighted on characteristics 2 and 5, and Strategy C is most heavily weighted on characteristic 1. Therefore, we would expect to see marketplace segments who either fall into the binary categories (i.e. have a value of 1 rather than 0) or have larger numbers for the non-binary characteristics to be positioned near the top of the list for the strategies where those characteristics are heavily weighted. Figure 6 displays the three strategies as number lines with the scale presented as a percentage, where 0% is the highest ranked marketplace segment and 100% is the lowest ranked marketplace segment. The number line is annotated to identify the location of the average rank for each of the marketplace segments that fall into the binary characteristics. For example, in Strategy B, the average rank for all marketplace segments that have characteristic 1 is approximately in the middle, around 50%.

There is a connection between Figure 5 and Figure 6: if a characteristic were to be heavily weighted, it is found to be near the top of the number line. For example, Strategy A heavily weighted characteristics 6 and 8, and the average rank for marketplace segments which fall into either of those categories is atop of the list in Figure 6. We see similar relationships with characteristics 2 and 5 in Strategy B and characteristic 1 in Strategy C.

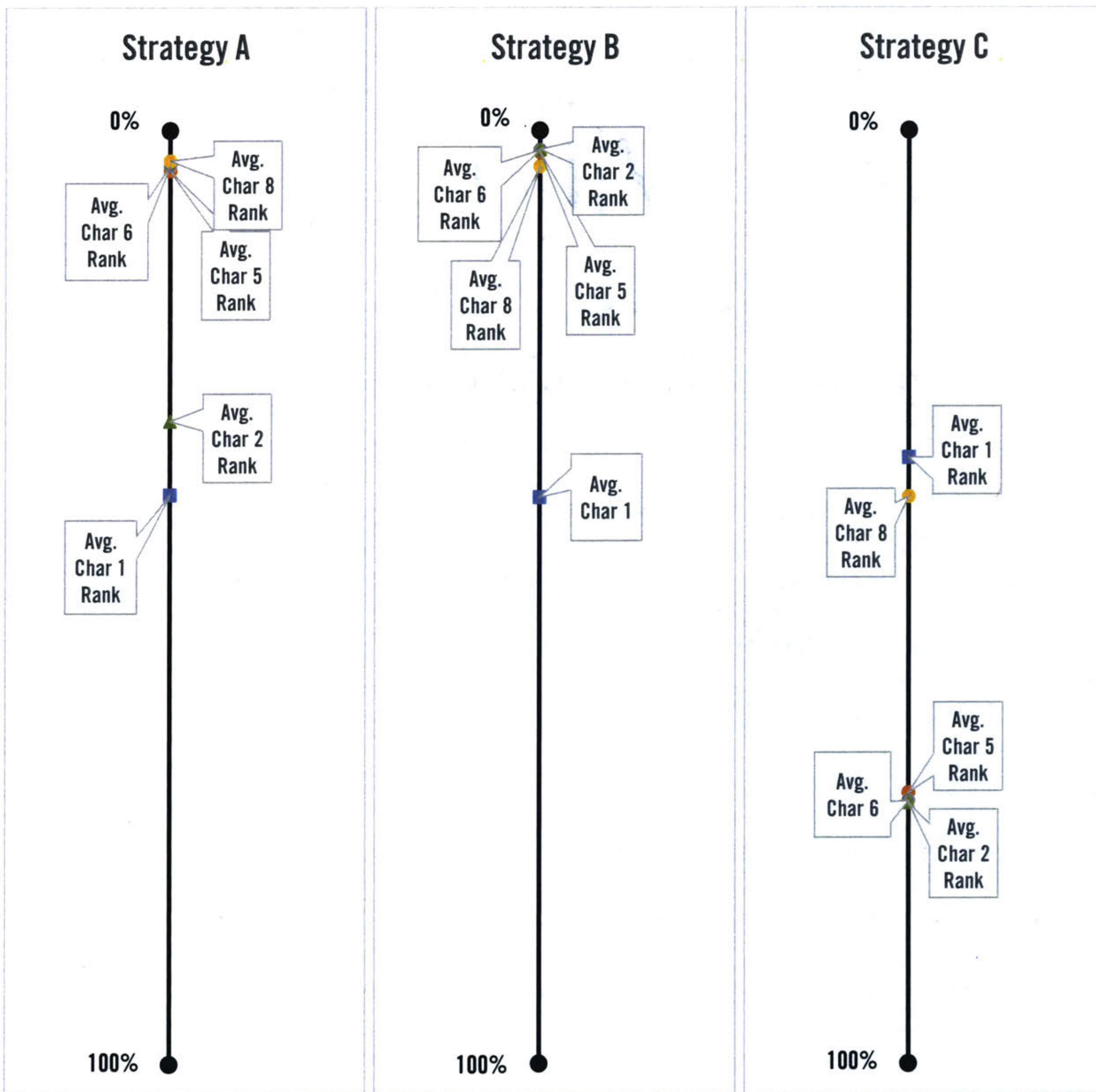

Figure 6. Marketplace Segment Location within Lists Created by Strategies A, B, and C

One item to note is that Strategy B, when shown to decision makers within retail manufacturers familiar with allocation prioritization and the DSML process, was indicated as similar to what a retail manufacturer may consider doing if they were to use a manual process to re-rank all of their marketplace segments today. That is because in Figure 6, the average rank of marketplace segments that fall into characteristics 2, 5, 6, or 8 sit at the very top of the list for Strategy B, thus making it seem similar to a manual decision to prioritize all of the marketplace segments with those characteristics. Therefore, the results of this proof of concept project became slightly more poignant, as they could be used to quantify the potential effects of a possible future state if a retail manufacturer were to maintain the art-based, manual process of re-prioritizing marketplace segments.

5.3 Allocation Simulation Results

The allocation simulation model was run twelve times: the four lists (Strategies A-C and the marketplace segment list currently being used) were each run over three different time periods. Three time periods were selected, each consisting of approximately 30 days. These periods were selected for two reasons: (i) data availability for those months; and (ii) the three months selected provide a representative sample of different business months within the retail industry. The historical data needed to run the simulation model was collected daily, beginning at the beginning of month one, which therefore became the earliest possible day we could simulate. Similarly, the simulation models were run a few months after data collection began, so data from month four and beyond was not yet available.

The results of the simulations are presented in Figure 7, Figure 8, **Error! Reference source not found.**, **Error! Reference source not found.**, and Figure 10. The x-axis of the figures shows the different strategies (i.e. different prioritized lists of marketplace segments) that were simulated.

The y-axis represents the number of shippable units output by the simulation of each strategy divided by the number of shippable units output by the simulation of the current state. So a positive number means that the shippable units increased compared to the current state, and a negative number means that the shippable units decreased compared to the current state. The figures are further broken down into specific categories of marketplace segments, as specified by the specific characteristics they have. Figure 7 presents the overall results; the results for each type of marketplace segment that were represented by binary characteristics 1, 2, 5, 6, and 8 (i.e., each marketplace segment that received a value of 1, rather than 0, for that binary characteristic); and the results for a set of marketplace segments, S, of particular interest to retail manufacturers.

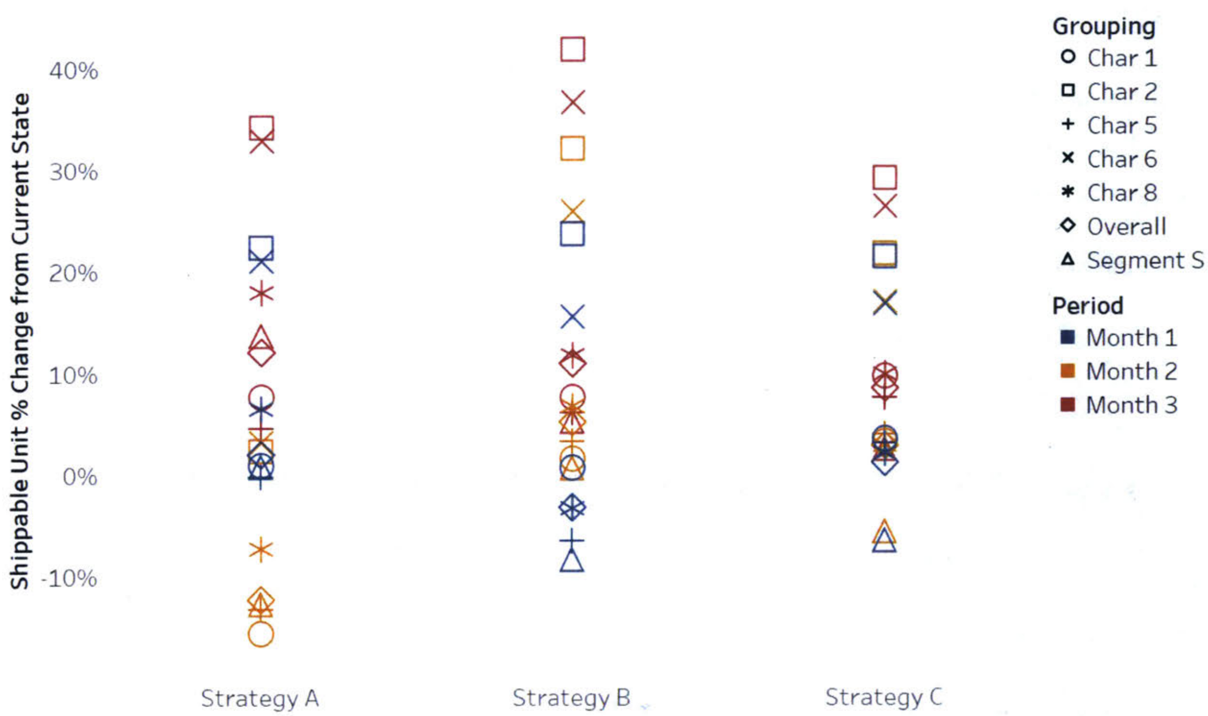

Figure 7. Results from Allocation Simulation - All Groupings

The results shown in Figure 7 indicate an achievement of one of the project goals: to prove that a re-prioritization of marketplace segments could increase delivery precision. All icons located above the 0% line represent groups of marketplace segments for which more shippable units are being made available than under the current state. Although the simulation is backward looking in nature, we at least can see that if we had employed different marketplace segment prioritization strategies in the past, the marketplace segment groups would have received more units faster.

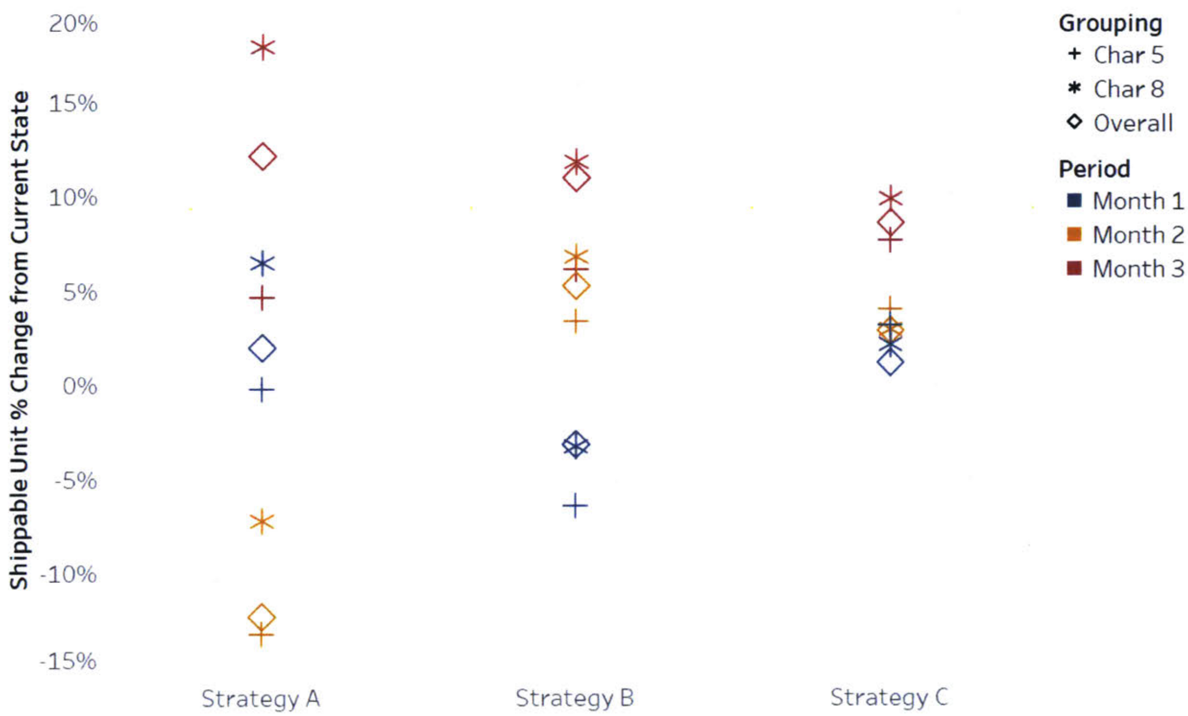

Figure 8. Results from Allocation Simulation - Characteristics 5 & 8, Overall

Figure 8 is included to allow for further review of the overall results and for how marketplace segments who fell into characteristics 5 and 8 fared under the different simulations. For example, let's pay particular attention to Strategy B. As discussed in the previous section, Strategy B is a good approximation of what a retail manufacturer could have put in place under current methodologies. We see that Strategy B would have performed well in months two and three but would have performed worse than the current state in month one. Similarly, Strategy A would have performed poorly for month two.

When reviewing Figure 7 and Figure 8, Strategy C appears to be a very attractive strategy, as the only marketplace segment groupings and time periods that created fewer shippable units than the current state were the S subset of marketplace segments in months one and two.

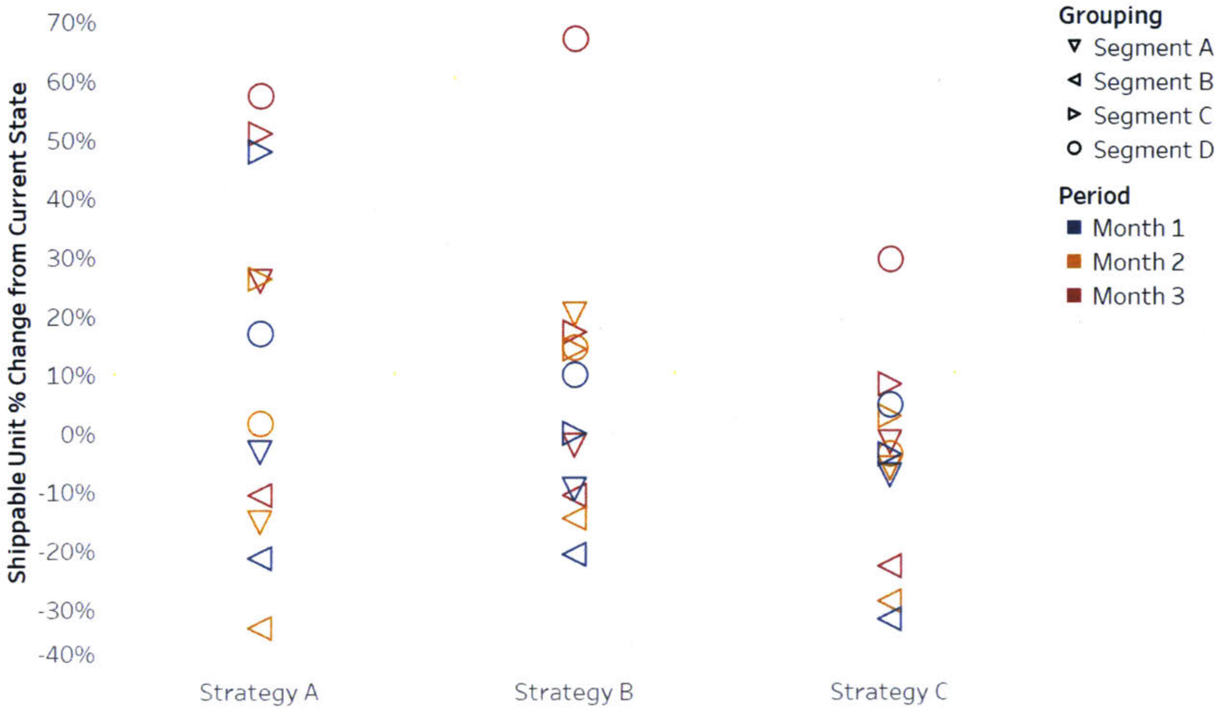

Figure 9. Results from Allocation Simulation - Four Marketplace Segments in Segment S

However, looking further into several marketplace segments within the larger Segment S, we see from Figure 9 that the effects within Segment S are not standard. If a retail manufacturer were to select Strategy C as the marketplace segment prioritization method moving forward, they would need to particularly review all members of Segment S to understand if they are willing to create fewer shippable units for some marketplace segments such that others within Segment S, and all members of the other groups analyzed above, can have a higher rate of shippable units made available to them.

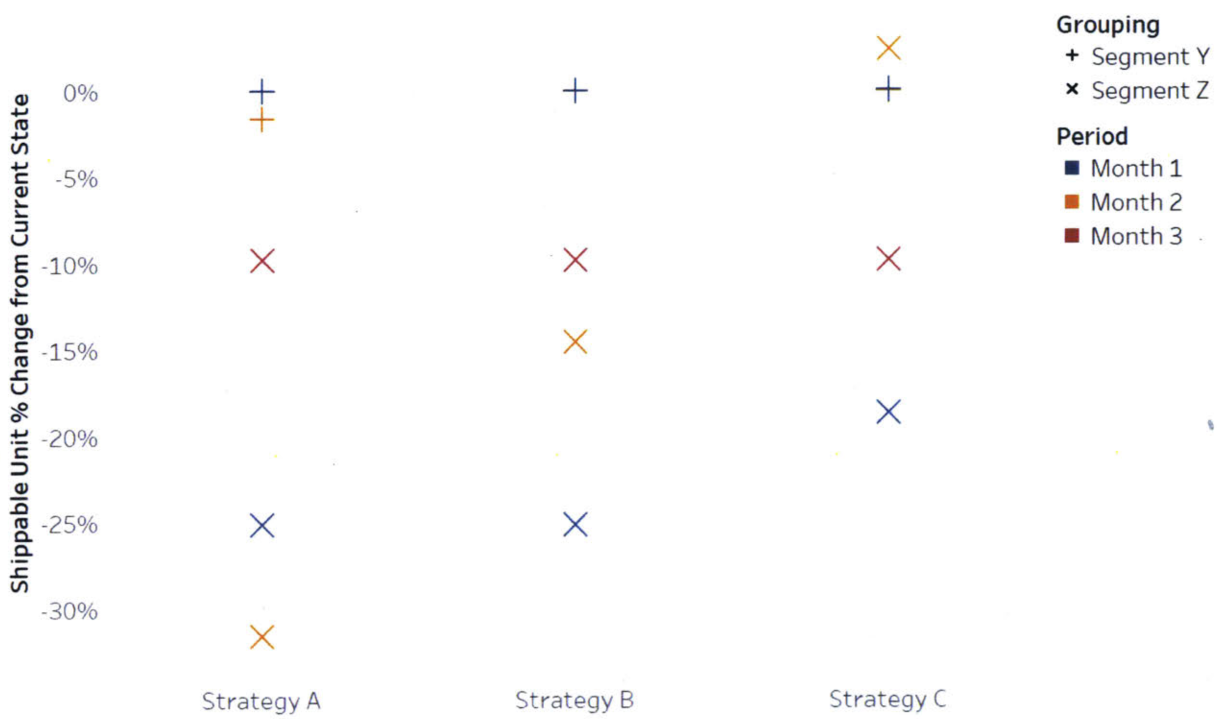

Figure 10. Results from Allocation Simulation - Value Segments

Figure 10 presents marketplace segments Y and Z, considered to be value segments (segments that purchase styles from previous seasons at steeply discounted prices). Since no extra demand or extra supply are being created through this process, it is inevitable that some segments will see the downside of this new approach to marketplace segment prioritization. Segments Y and Z are two examples of this phenomenon. These marketplace segments see either fewer shippable units or the same shippable units being attributed to their orders in all simulation periods but one, therefore indicating poorer delivery precision.

Figure 11. Results from Allocation Simulation - Segment K

Lastly, we reviewed the effects on a single marketplace segment with an important geographic location, Segment K, as a specific case study. This marketplace segment is typically seen as an important segment for retail manufacturers, so it was of particular importance to see how this segment fared under the three Strategies of marketplace segment prioritization. Under each of the three Strategies, the segment saw double digit increases in shippable units, even up to a 90% increase during month one.

6 Conclusion

6.1 Recommendations

The primary recommendation as a result of this project is that data-driven marketplace segment prioritization combined with an allocation simulation tool should be used to improve delivery precision from the current state, and therefore should be implemented into the current DSML processes for retail manufacturers. Strategy C was identified as a good option for a pilot program, from which it could be learned if the improvements suggested by the simulation tool would be realized in the real world. However, it is recommended that this proof of concept be improved upon and turned into an operational set of models and processes, which may identify a more effective strategy for marketplace segment prioritization than Strategy C.

6.2 Next Steps

6.2.1 Model Implementation

Although the results indicate that this proof of concept is able to improve delivery precision over the current state, the runtime and user interface for the newly created algorithms and models are not at a point that they can be included into typical business processes by employees without intricate knowledge of the modeling approaches and programming choices that were made. Simply put, the proof of concept developed for this thesis is not user-friendly (i.e. doesn't have a graphical user interface) and therefore requires knowledge of python and the data storage methods utilized by the retail manufacturer. The next steps for this project are therefore two-fold: (i) improve the model runtime; and (ii) create a valid user interface. Currently, the model takes approximately one hour per day simulated. While simulating a month at a time was valid for this proof of concept, retail manufacturers would benefit from simulating an entire season (three months), which at the current runtime would take approximately four days of computing time. Simulating an entire season would provide a better approximation of how units flow through the distribution network from the initial high volumes of the beginning of a season through to the lower volume expected at the end of the season. Therefore, to make this concept fully implementable, the runtime would need to be reduced dramatically. Alternatively, this concept could be turned into an analysis that is always running, such that employees of the retail manufacturer could access the results at any time, and use the readily available results to make real-time decisions.

Additionally, the algorithms and models created in this proof of concept were handed off as python script files (.py files). These files are only useful to an employee with experience executing python scripts. Therefore, to make these tools fully useful, they must be combined with a user interface that creates a barrier between the user and the code being executed.

6.2.2 Timeline for Model Usage

At the end of this proof of concept project, the use of the model and methodologies presented above require significant computation time. However, it is hypothesized that the more often the marketplace segments are re-prioritized to better match current business strategies, the better delivery precision to important marketplace segments. Therefore, as a balance between the computation time required and the potential benefits of consistent updates, it is recommended that the marketplace segments be re-prioritized once before each season and once in the middle of each season. This way, there is an opportunity to match the supply allocation to the current business strategy once before each season and then an opportunity to update or adjust the process in the middle of the season if the first half of the season produced delivery precision metrics below targets.

6.2.3 Dynamic Marketplace Segment Prioritization

This proof of concept project identified the opportunity to increase delivery precision through periodically updating how marketplace segments are prioritized, which leads to the next question: can we dynamically, on perhaps a daily basis, update which orders and which marketplace segments are prioritized? The next steps, beyond operationalizing the methods presented above, are to develop the ability to dynamically refresh marketplace characteristics and then use that information to update marketplace segment prioritizations in real-time.

Additionally, the largest constraint on this project was the example DSML process that was strictly adhered to. Rather than working around an imperfect process, the most effective method to improve delivery precision solely within the ERP environment would be to improve the overall DSML process, including building in the ability to dynamically prioritize marketplace segments. It is the hypothesis of the author that this would provide the best opportunity to increase delivery precision to marketplace segments through solely adjusting internal, ERP-based processes.

7 References

Agresti, Alan. *Analysis of Ordinal Categorical Data*. 2nd ed. Wiley Series in Probability and Statistics. Hoboken, N.J: Wiley, 2010.

Bholowalia, Purnima. "EBK-Means: A Clustering Technique Based on Elbow Method and K-Means in WSN." *International Journal of Computer Applications* 105, no. 9 (n.d.): 8.

Caruso, Giulia, Stefano Antonio Gattone, Francesca Fortuna, and Tonio Di Battista. "Cluster Analysis as a Decision-Making Tool: A Methodological Review." In *Decision Economics: In the Tradition of Herbert A. Simon's Heritage*, edited by Edgardo Bucciarelli, Shu-Heng Chen, and Juan M. Corchado, 618:48–55. Cham: Springer International Publishing, 2018. https://doi.org/10.1007/978-3-319-60882-2_6.

Homburg, Christian, Dirk Totzek, and Mathias Droll. "All Customers Are Equal, but Some Are More Equal Should Firms Prioritize Their Customers ?" *GfK Marketing Intelligence Review* 2, no. 1 (May 1, 2010): 16–25. https://doi.org/10.2478/gfkmir-2014-0067.

Knight, Phil. *Shoe Dog: A Memoir by the Creator of Nike*. First Scribner hardcover edition., 2016.

Lacoste, Sylvie. "From Selling to Managing Strategic Customers - a Competency Analysis." *Journal of Personal Selling & Sales Management* 38, no. 1 (January 2, 2018): 92–122. https://doi.org/10.1080/08853134.2018.1426991.

Macqueen, J. "SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS." *MULTIVARIATE OBSERVATIONS*, n.d., 17.

Nelsen, R.B. "Kendall Tau Metric." *Encyclopedia of Mathematics*, n.d. http://www.encyclopediaofmath.org/index.php?title=Kendall_tau_metric&oldid=12869.

———. "Spearman Rho Metric." *Encyclopedia of Mathematics*, n.d. http://www.encyclopediaofmath.org/index.php?title=Spearman_rho_metric&oldid=15466.

Nike Inc. "2018 Form 10-K," 2018. https://s1.q4cdn.com/806093406/files/doc_financials/2018/ar/docs/nike-2018-form-10K.pdf.

———. "NIKE, Inc. Announces New Consumer Direct Offense: A Faster Pipeline to Serve Consumers Personally, At Scale." *Nike News* (blog), June 15, 2017. https://news.nike.com/news/nike-consumer-direct-offense.

Park, Hae-Sang, and Chi-Hyuck Jun. "A Simple and Fast Algorithm for K-Medoids Clustering." *Expert Systems with Applications* 36, no. 2 (March 2009): 3336–41. https://doi.org/10.1016/j.eswa.2008.01.039.

Patel, Abhishek, and Purnima Singh. "New Approach for K-Mean and K-Medoids Algorithm." *International Journal of Computer Applications Technology and Research* 2, no. 1 (January 10, 2012): 1–5. https://doi.org/10.7753/IJCATR0201.1001.

Peng, David Xiaosong, and Guanyi Lu. "Exploring the Impact of Delivery Performance on Customer Transaction Volume and Unit Price: Evidence from an Assembly Manufacturing Supply Chain." *Production and Operations Management* 26, no. 5 (May 2017): 880–902. https://doi.org/10.1111/poms.12682.

Thorndike, Robert L. "Who Belongs in the Family?" *Psychometrika* 18, no. 4 (December 1953): 267–76. https://doi.org/10.1007/BF02289263.

Velmurugan. "Computational Complexity between K-Means and K-Medoids Clustering Algorithms for Normal and Uniform Distributions of Data Points." *Journal of Computer Science* 6, no. 3 (March 1, 2010): 363–68. https://doi.org/10.3844/jcssp.2010.363.368.