A Hierarchical Approach and Analysis of Assortment Optimization

Toolika Agrawal, Utkarsh Bajaj, Aaron Chen, Prerana Das, Manandeep Gill, Dhruv Shrivastava, and Matthew Lanham

MS Business Analytics & Information Management

Krannert School of Management, Purdue University

Abstract

Assortment planning is one of the most important and challenging applications of analytics in retail. Often retailers use a two-stage approach where in the first stage they run thousands of prediction experiments to identify what best captures expected demand. In the second stage, they decide which combination of products will lead to the best sales for a particular store – a classic knapsack-type problem. This work focuses specifically on combinatorial assortment optimization (or second stage) and how the hierarchical nature of the decisions and analysis that needs to occur can lead to drastically different outcomes in-store performance. Using data such as inventory, historical sales, wait times, geographical activity, budgetary constraints, product variety, and shelf space various linear and integer programming models were formulated to demonstrate how the assortment can change using sensitivity analysis on the constraint parameters. The client was presented with a strategy to understand how the constraint parameters are set in the assortment optimization process to achieve strategic revenue outcomes. This optimization exercise was performed using the **CVXPY** python package on Purdue Universities' high-performance Bell cluster, which is one of the top 500 HPCs in the world.

Keywords: *Assortment Optimization, Linear Programming, Mixed Integer Programming, Retail Analytics, Substitution Effect*

A Hierarchical Approach and Analysis of Assortment Optimization

A ubiquitous problem faced across many industries pertains to optimizing the assortment of your offerings. Many industries such as retail, online advertising, and security services are posed with this question periodically. A firm usually offers its products to many customers who decide to purchase them. Whether to make the purchase is a function of the assortment of products offered by the retailer and consumer needs, which are usually constrained by shelving space, store budget, and consumer budget considerations. Hence, the retailer can gain significant leverage by optimizing its assortment of offerings to maximize its business goals, which are, in most cases, either revenue or profit. This relationship is complicated because revenues are realized only after the assortment decision has been made using a forecasting model. Customers may walk away because they cannot find the product they intend to purchase or end up buying the closest substitute to it. Therefore, while making the assortment decision, incorporating the substitution effect of each SKU is also an important decision criterion.

Recent developments in today's cut-throat retail marketplace make the assortment decision more important than ever. A recent report by (McKinsey & Company, 2019) states that there are four reasons for this growing pressure on assortments: an ever-increasing array of SKUs, limited physical shelf space, growing supply-chain complexity, and the need for localization as per the local demographics around the store. The potential downsides for incorrect assortment decisions are manifold. Apart from lost sales, not having a particular product in the store can cause a highly unpredictable estimated demand curve, leading to chronic issues such as supply chain inefficiency and burgeoning transportation and inventory carrying costs for the company. This can cause severe financial mismanagement within an organization. In addition to this, it can also cause customer attrition.

This paper deals with one such problem in assortment optimization. The automotive aftermarkets industry is the secondary market for automobiles and mainly consists of manufacturing, retailing, and selling automotive accessories after purchasing the automobile from an original equipment manufacturer (OEM). The current assortment and optimization model of the client uses SKU ranking based on forecasted data (revenue and profit). This 1-D analysis requires enrichment of store-level inventory data, upon which the results are compared. The goals of the study are to secure maximum value (firm revenue or profit) by finding and maintaining an optimal assortment for a group of SKUs per store and market cluster based on SKU demand, SKU coverage (by cluster), acquisition cost, and shelving space while balancing the high sales volume & low-profit SKUs and low volume & high-profit SKUs while conducting the overall optimization exercise.

The design chosen to meet our study goals is prescriptive analytics. Linear and Integer Programming (MIP) is attempted based on the aforementioned constraints. Furthermore, a sensitivity analysis is conducted for space versus profit and revenue forecasts. This is done to ensure that the overall goal of the product placement optimization exercise is solved. The MIP is to be performed using the python package cvxpy.org for convex optimization problems and mixed-integer programs. To further work on this, an exploratory data analysis was performed to understand the spread and quality of currently available data, helping us to clearly define the decision variables, constraint equation, and objective function. The cvxpy.org package on python is used to solve it. The modeling is performed for a couple of iterations to find out the best possible solution for our proposed business problem for the client (a Fortune 500 automotive aftermarkets parts provider).

Product & Assortment Decisions

Product selection and inventory have always been essential business decisions for the retailer. Retailers face constant pressure to broaden their assortment in the wake of growing consumer heterogeneity and competition between brands. Numerous studies have shown the negative impact that suboptimal assortments can have on a retailer's P&L. According to a study by Quelch and Kenny (1994), the number of products in the market grew by 16% per year between 1985 and 1992, while the growth of shelving space during this period was only 1.5%. To tackle this problem, retailers depend on two strategies: maximizing total profit by removing low-selling products and then maximizing the profit from each category of products. Numerous studies prove that various levels have become so excessive that profit/ revenue does not decrease with growing SKUs. The current literature is replete with information on the economics of assortment planning. According to Fisher and Vaidyanathan (2015), there are broadly three considerations that retailers have to look into while planning assortment: product variety, product line, and product quantity. Product variety can be defined as the different number of brands for the same product that the retailer has to include in its assortment to serve the consumers' varied tastes, the product line is defined as the different products of the same brand that the retailer needs to include in its assortment to keep the manufacturer happy, and lastly, product quantity is the number of units for each product that the retailer has to carry to meet the consumer demand. Hence, since a given assortment decision can impact numerous levers in supply-chain, sales and distribution, and inventory, a careful decision of product lines is needed for optimizing profit and revenue.

Multi-item Inventory Models

Multi-item inventory problems are critical to addressing an assortment optimization problem. It is essential to study how to manage the inventory of multiple products in one shelf space that is constrained to budgets and space availability to bring out the maximum revenue. Lagrangian multipliers in operations literature and solutions and heuristic approximation to the multi-period period sales losses are some methods where the demand of products is independent of other products' inventory, assuming no substitution.

Another exogenous demand method for multi-product inventory optimization is based on stock-out substitution, where stocking depends on one given selection but not the entire product selection. The total demand is formulated from the sum of initial demand and the substitution demand of substitutes. Substitution demand of any product is a factor of its unsatisfied demand. Over time researchers have worked on its multi-product version, decentralization, heuristic algorithm, products under centralized and decentralized management, and many more. It is impossible to obtain a specific, definitive solution due to its restrictive and complex nature. The decentralized routine allows having more inventory than the centralized one because of competition. This works when customer substitution is dynamic and a multinomial logit choice model is used. Researchers also devised an infinite horizontal version based on centralized and competitive scenarios, a single period under decentralized management taking aggregate demand as a random variable and individual firm demand concerning different rules of initial allocation and reallocation of excess demand. Other substitution models such as one done by observing demand before inventory allocation and the other by upgrading to higher quality products. Here, reallocation is solved similarly to that of a transportation problem.

Similarly, demand for individual components depends on the demand for finished goods, hence the inventory. This is majorly used in the case of an online retailer's order fulfillment strategy. This is called an assemble-to-order system.

Stochastic Knapsack Model

The problem of accepting and blocking calls to a circuit-switched telecommunication system that supports a variety of traffic types (e.g., voice, video, facsimile, etc.), each with different bandwidth requirements and holding-time distributions, motivated the development of the stochastic knapsack model. The challenge of efficiently accepting calls to maximize average income is identical to the stochastic knapsack problem when the communication channel is modeled as a knapsack, the traffic types as object classes, and the bandwidth requirements as object volumes. Other uses for the stochastic knapsack model include parallel processing, where jobs demand a different number of processors depending on their class and shared memory, where jobs require different amounts of memory. By setting the arrival rates for each of the classes to infinity, the classical knapsack issue can be regarded as an extended instance of our stochastic model. In this instance, a complete partitioning (CP) policy, in which the knapsack is partitioned and each object class has exclusive usage of its dedicated portion of the knapsack volume, would be best. On the other hand, if all of the arrival rates are "small," complete sharing (CS) would be the best option, where an object is always provided access anytime adequate volume is available.

Shelf Space Allocation Models

In some product segments such as grocery and pharmaceuticals, how much shelf space is allocated to a given product category is an important component of the assortment planning process. This view seems especially relevant for fast-moving products whose demand is sufficiently high that a significant amount of inventory is carried on the shelf.

This contrasts with other categories, e.g., shoes, music, books, where only one or two units are carried for most SKUs. For example, Transworld Entertainment carries 50,000 SKUs in an average store but stocks more than one of only the 300 bestsellers. Hence the amount of inventory and shelf space is not critical decisions at the product level.

In an influential paper, Corstjens and Doyle (1981) suggest a method for allocating shelf space to categories. They perform store experiments to estimate sales of the product I as is βi, si(, where si is the space allocated to a product I, βi is own space elasticity, and si(s are the cross-space elasticities. Cost functions are also estimated from the experiments. The problem of profit maximization with a shelf space constraint is solved within a geometric programming framework. Their results are significantly better than commercial algorithms that allocate space proportional to sales or gross profit by ignoring interdependencies between product groups. The estimation and optimization procedures cannot be applied to large problems; hence, they work with product groups rather than SKUs. Bultez and Naert (1988) apply the Corstjens and Doyle (1981) model at the brand level assuming symmetric cross elasticities within product groups. Their model is tested at four different Belgian supermarket chains, encouraging results.

An interesting paper by Borin and Farris (1995) reports the sensitivity of the shelf space allocation models to forecast accuracy. They compare the solution with correct parameters to that with incorrect parameter estimates. Even when the error in parameter estimates is 24%, the net loss in category return on inventory is just over 5% compared to the optimal allocation based on true estimates. This proves the robustness of these models to estimation errors. Like these shelf space allocation papers, but using an inventory theoretic perspective, Urban (1998) models the own and cross-product effects of displayed inventory on-demand rate in a mathematical program and solves for shelf space allocation and optimal order-up-to quantities. He reports that, on average, a greedy heuristic yields solutions within

1% of a solution obtained by genetic programming.

Irion et al. (2004) extend the Corstjens and Doyle model to study the shelf space allocation problem at the product level. Demand for each product is a function of its own and other products' shelf space through own and cross-shelf space elasticities. The cost for each product consists of linear purchasing costs, inventory costs from an economic order quantity model, and a fixed cost of being included in the assortment. The objective is to allocate (integer) the number of facings to each product to maximize profits under a total shelf space availability constraint and lower and upper bounds on the number of facings for each product. The problem is transformed into a mixed-integer program (MIP) with linear constraints and objective function through a series of linearization steps. The linearization framework is general enough to accommodate several extensions. However, there is no empirical evidence that product level demand can be modeled as a function of the shelf space allocated to the product itself and competing products via own and cross-space elasticities.

Shelf space allocation papers do not explicitly address assortment selection and inventory decisions and ignore the stochastic nature of demand. In collaboration with a national automotive aftermarket parts provider, access to the data related to three base product groups was provided: batteries, brakes, and filters. Prediction inputs and demand predictions for these product groups in this data were also provided.

In addition to this, data was granular at multiple levels. There was information on the level of Stock Keeping Units (SKU) with the space measurements (e.g., length, height, width), retail price, and store acquisition cost. These observations were related to fixture space with category and store-wise current and maximum space constraints. Data were also available at the store level with information related to quantity sold and gross sales (in USD) for each SKU. The inventory file was at the store and SKU level with on-hand stock quantity. Each variable has been described in the following table.

Methodology

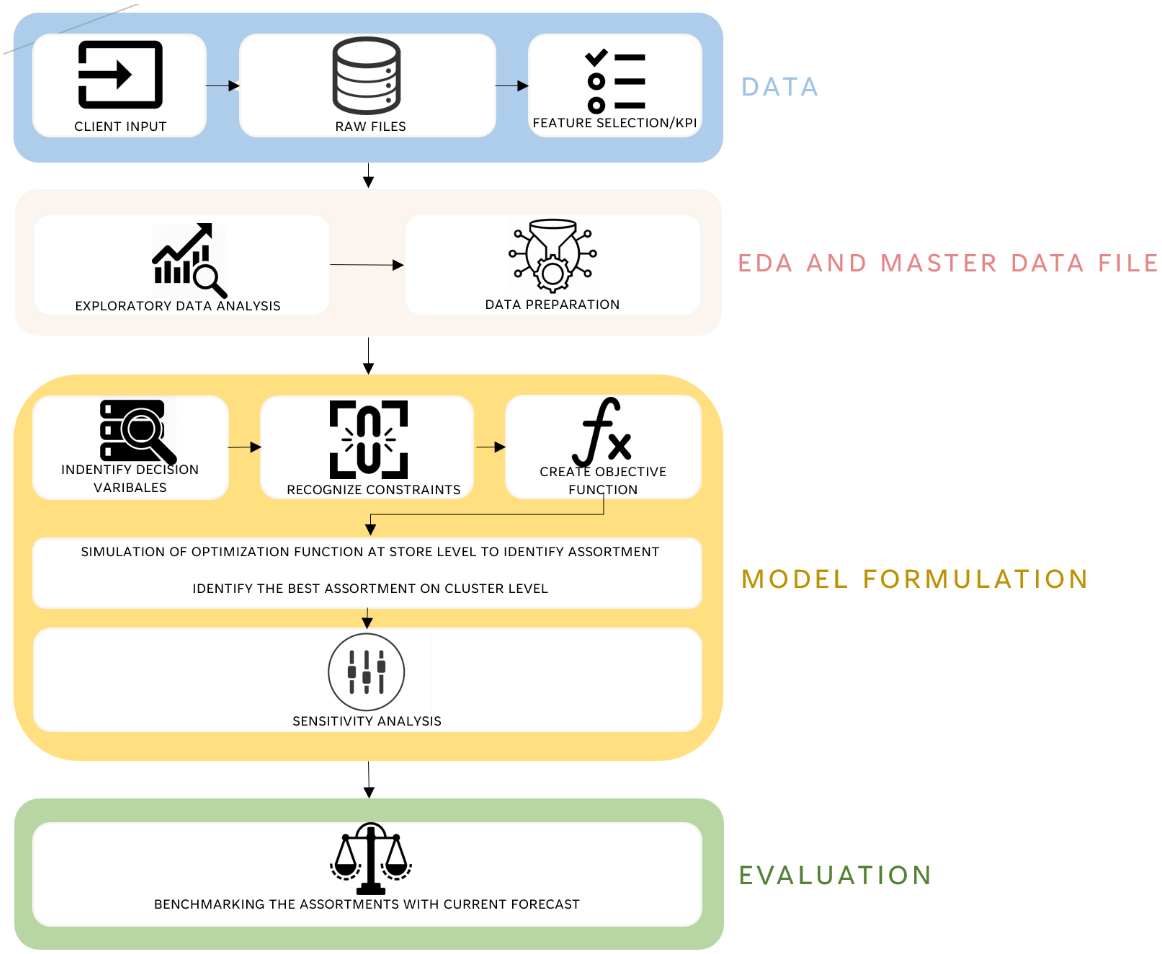


Table 1: Data used in the study

Variable	Type	Description
store_number	Numeric	Unique number for a particular store
sku_number	Numeric	A unique number to track SKU
open_date	Date	Date from which the store was operational
market_share	Numeric	Market share for a particular store

market_class	Categorical	Market class for a particular store
latitude	Numeric	Latitude value for a particular store
longitude	Numeric	Longitude value for a particular store
platform_cluster_name	Categorical	Cluster category for a particular store
mpog_description	Categorical	Description of a part of SKU hierarchy / category
sku_description	Categorical	Description of a SKU
length	Numeric	Length of a particular SKU
width	Numeric	Width of a particular SKU
height	Numeric	height of a particular SKU
cubic_inches	Numeric	The volume of a particular SKU
retail_price	Numeric	Price of a particular SKU
unit_cost	Numeric	Unit Cost of a particular SKU
min_app	Categorical	Variable related to SKU stock
discontinued_flg	Binary	Flag describing if the stock is discontinued or not
stocking_location	Categorical	Location at which stock is located
store_acquisition_cost	Numeric	Cost of store acquisition
sales_ID	Numeric	Unique identifier for a particular sale
qty_sold	Numeric	Quantity sold at a particular store and SKU level
gross_sales	Numeric	Gross sales at a particular store and SKU level
sales_cost	Numeric	Cost of Sales at a particular store and SKU level
fiscal_year	Categorical	Years for which sales data is present
fiscal_period	Categorical	Period for which sales data is present

Inv_ID	Numeric	Unique identifier for a particular inventory
inv_year	Categorical	Years for which inventory data is present
inv_period	Categorical	Period for which inventory data is present
smaxi	Categorical	Condition for checking if SKU is stocked at store
sesq	Categorical	Condition for on-hand stock
scoq	Categorical	Quantity of on-hand stock in stores
job_id	Numeric	Unique id for a fixture space
mpog_id	Numeric	Part of SKU hierarchy / category
category_current_space	Numeric	Category wise space available per SKU
category_max_space	Numeric	Category wise space allocated per SKU
store_current_space	Numeric	Maximum space available per store
store_max_space	Numeric	Available space available per store

Data and Model

Data Characteristics

With regards to running the optimization algorithm, two types of data are of critical importance: the sales data for the current year (in this case, 2020-21) and demand forecasts for the next year (in this case, 2021-22). The current sales data would provide the assortment and profitability/ revenue benchmark that the optimized assortment for the respective store has to attempt to beat, and the demand forecasts for the coming year shall be used to calculate the profitability of the assortment predicted by the optimization function. Hence, in essence, the analysis compared the sales performance of the current assortment with the predicted sales performance of the assortment given by the objective function. The sales data has been provided for 10295 SKUs whereas the demand forecast data has been provided for 12121 SKUs, both across 5450 stores. These SKUs are grouped into twenty-one categories (called MPOGs hereafter), which are further grouped into three base product groups (called BPGs hereafter). The three base product groups are filters, brakes, and batteries.

Objective Function

The objective function for an assortment optimization/ knapsack problem is a mathematical equation that describes the assortment output target that corresponds to the maximization of firm value concerning the assortment. In this case, the profit generated per store was used as an indicator for the firm value, but it can also be extended to cases where revenue and shareholder's value might be the more appropriate variable in question. Therefore, the objective function for a particular store is as follows:

$$Maximize\ profit\!: \Pi_i = \sum_{j=1}^{n} \beta_{ij} * Y_{ij} * (P_j - C_{ij})$$

Assumptions, notations, and definitions:

1. $There\ are\ total\ m\ stores\ \forall\ i \in \{1,\ 2,\ 3 \ldots\ldots i,\ i+1, \ldots . m-1,\ m\}$

$For\ each\ i^{th}\ store:$

2. $There\ are\ total\ n\ SKUs\ \forall\ j \in \{1,\ 2,\ 3 \ldots\ldots j,\ j+1, \ldots . n-1,\ n\}$

3. $There\ are\ total\ p\ categories\ \forall\ k \in \{1,\ 2,\ 3 \ldots\ldots k,\ k+1, \ldots . p-1,\ p\}$

4. $There\ are\ total\ t\ base\ product\ groups\ \forall\ r \in \{1,\ 2,\ 3 \ldots\ldots r,\ r+1, \ldots . t-1,\ t\}$

5. $Each\ k^{th}\ category\ has\ total\ n_k\ SKUs\ \forall\ n_1 + n_2 + n_3 \ldots\ldots + n_k + n_{k+1} \ldots\ldots + n_{p-1} + n_p = n$

6. $Each\ r^{th}\ base\ product\ group\ has\ total\ p_r\ categories\ \forall\ p_1 + p_2 + p_3 \ldots\ldots + p_r + p_{r+1} + \ \ldots\ldots + p_{t-1} + p_t = p$

7. $\beta_{ij}\ \forall\ \beta_{ij} \in \{0{,}1\}\ for\ all\ j\ (1\ if\ included\ in\ the\ assortment,\ 0\ if\ not\ included\ in\ the\ assortment)$

8. $P_j\!: Retail\ Price\ for\ j^{th}\ SKU$

9. Y_{ij}: *Forecasted Demand for* j^{th} *SKU in* i^{th} *store*

10. C_{ij}: *Cost of Acquisition for* j^{th} *SKU in* i^{th} *store*

11. V_j: *Volume occupied by* j^{th} *SKU*

The decision variable of interest is Bij, which is a binary variable denoting whether the jth SKU is included in the assortment of the ith store (Bij = 1) or not (Bij = 0). The objective function has to be optimized under certain constraints which are defined for every store. These constraints are broadly defined on three aspects: category volume assigned for the SKU within the store, the budget allocated for the Base Product Group that the SKU belongs to, and the minimum coverage of the SKUs.

Constraints

1. *Non negativity constraint for* j^{th} SKU: $Y_{ij} \geq 0$ *for all j*

2. *Binary constraint for* j^{th} *SKU*: $\beta_{ij} \in \{0,1\}$ *for all j* $\left(0\ if\ included\ in\ the\ assortment,\ 1\ if\ not\ included\ in\ the\ assortment\right)$

3. *Category max space constraint for* k^{th} *category*: $\sum_{j=1}^{n_k} V_j * Y_{ij} * \beta_{ij} \leq \theta * CMS_{ki}\ \forall\ \theta \in \left\{0.90, 0.95, \underline{1.00}, 1.05, 1.10\right\}$ *for all j*

4. *BPG budget constraint for* r^{th} *BPG*: $\sum_{k=1}^{p_r} \sum_{j=1}^{n_k} C_{ij} * Y_{ij} * \beta_{ij} \leq \alpha * (\sum_{k=1}^{p_r} (\sum_{j=1}^{n_k} P_j)/n_k)/p_r * 20 * \sum_{k=1}^{p_r} \sum_{j=1}^{n_k} Y_{ij}\ \forall\ \alpha \in \left\{0.90, 0.95, \underline{1.00}, 1.05, 1.10\right\}$ *for all j*

5. *Category coverage constraint for* k^{th} *category*: $\sum_{j=1}^{n_k} \beta_{ij} \geq \lambda * n_k\ \forall\ \lambda \in \left\{0.30, 0.40, \underline{0.50}, 0.60, 0.70\right\}$ *for all j*

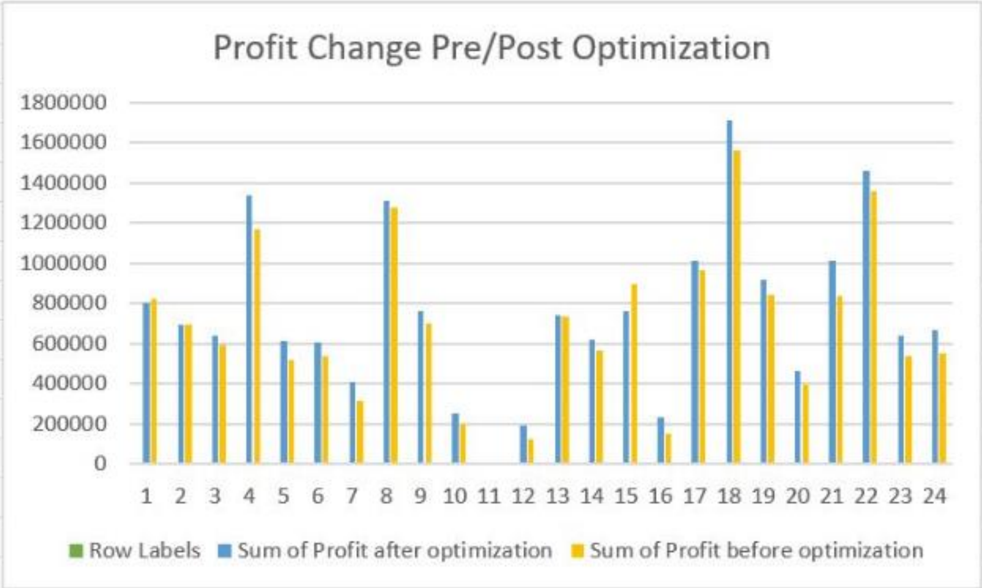
6. *Store coverage constraint*: $\sum_{j=1}^{n} \beta_{ij} \geq \rho * n\ \forall\ \rho \in \left\{0.50, 0.625, \underline{0.75}, 0.875, 1.00\right\}$ *for all j*

Analysis and Results

Results were derived for a particular store and these were benchmarked with the historical assortments. This benchmarking helped us realize the precision of the numbers derived from our optimization function. The results were obtained at multiple granularities which were category level, store level, and cluster level. The results for a particular store were extrapolated to the cluster level that assisted in the decision making for category, store, and area managers.

Change in Profitability

The following figure shows the total increase in profits per store for the first 24 stores. As can be seen from the bar charts, there is an average increase of 11.24% post-optimization across the 5450 stores.



Planogram Visualization

The following two figures represent the change in BPG-wise space allocation before and after the optimization function was applied to the data. As can be seen, the total space allocated to Batteries as a BPG has increased by 4% whereas the total space allocated to Brakes and Filters has decreased by 2% and 3% respectively. Hence, it can be concluded that Batteries on average are more profitable than Brakes and Filters.

Figure 1: Pre-Optimization TreeMap for all stores

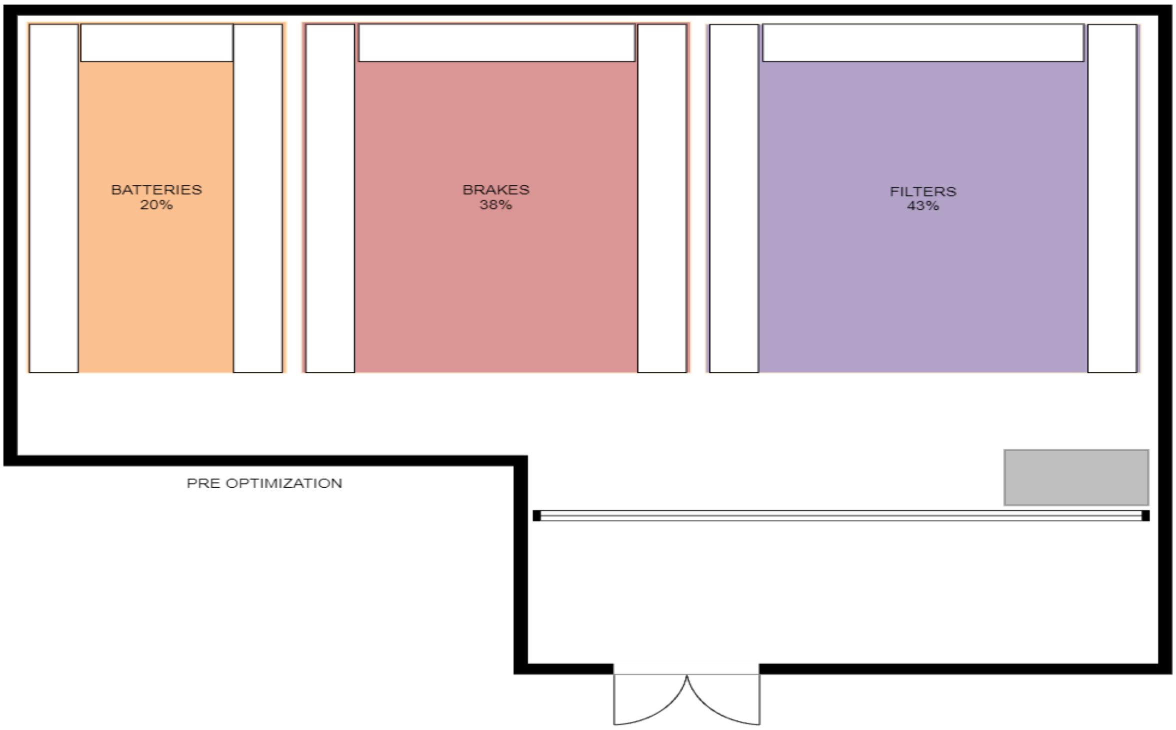
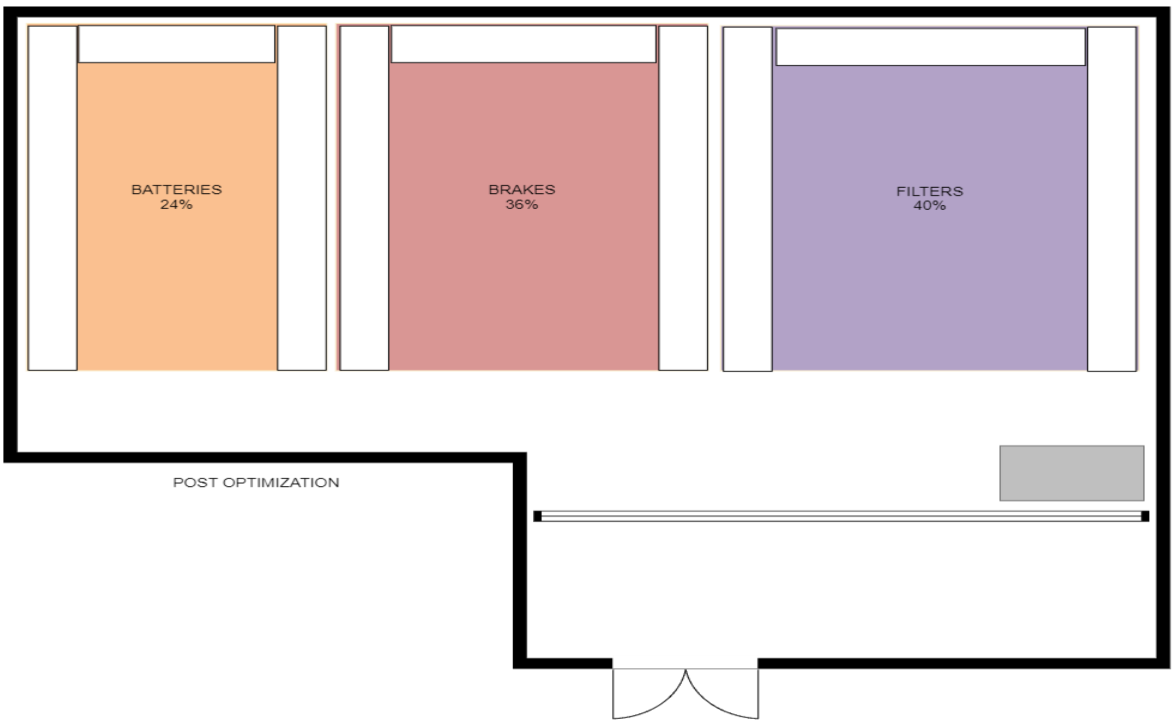


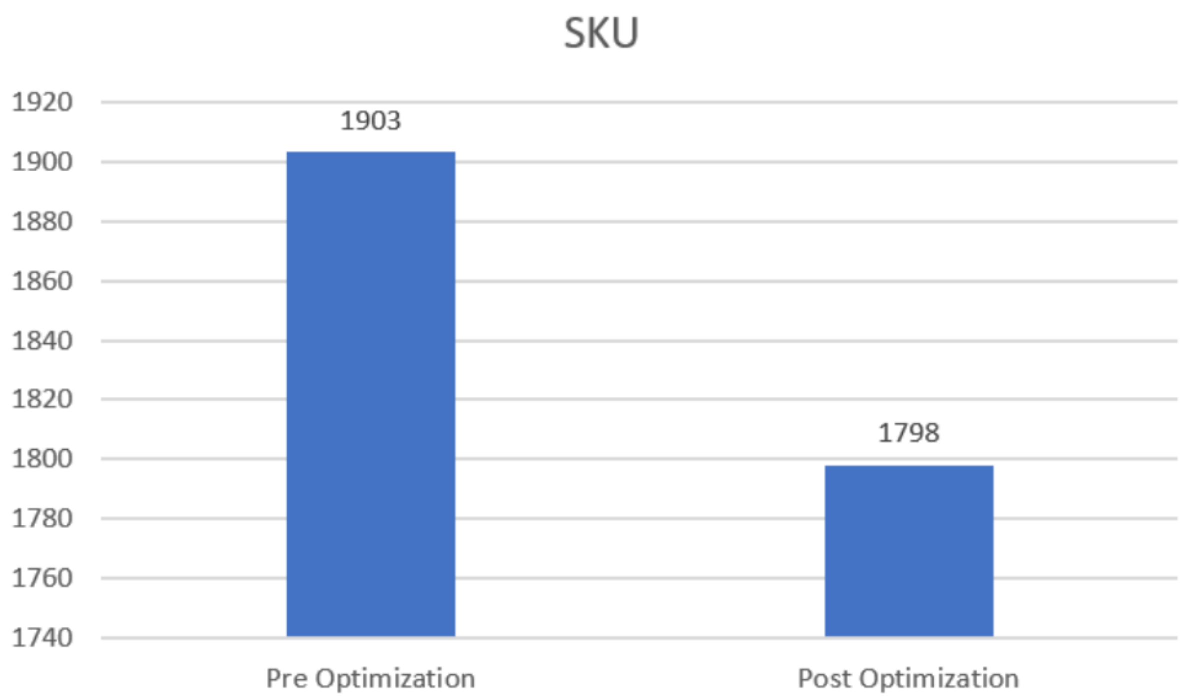
Figure 2: Post-Optimization TreeMap for all stores



SKU Reduction

The following figure shows the average reduction in the number of SKUs per store. As can be seen from the bar chart, there is an average decrease of 4.52% post-optimization across the 5450 stores. Hence, since the profit per store has increased by 11.24% and the number of SKUs has reduced by 4.52%, the average profit per SKU has improved. Thus, it can be said that the client will be making more profit per SKU post-optimization.

Figure 3: Reduction in SKUs on an average



Conclusion

The optimization model was created utilizing a variety of constraints and represents the retail stores' varied assortment issues. The model's objective function is to maximize profit per store. Shelf space, MPOG, total store budget, and minimum coverage of SKUs if they are stocked are all employed as limitations. The CVXPY Python library is used to run the model for 5450 stores across the United States. The objective function was able to get an optimal value with this model. Then, for each SKU, binary results were obtained to see which ones needed to be stocked based on factors like cost, space, and minimum number.

References

McKinsey & Company. (2019). *Analytical Assortment Optimization* . McKinsey & Company.

Quelch, J. A., & Kenny, D. (1995). Extend profits, not product lines. *The Journal of Product Innovation Management*, *3*(12), 249-250.

Kök, A. G., Fisher, M. L., & Vaidyanathan, R. (2015). Assortment planning: Review of literature and industry practice. , 175-236.

Borin, N., & Farris, P. (1995). A sensitivity analysis of retailer shelf management models. *Journal of Retailing*, *71*(2), 153-171.

Corstjens, M., & Doyle, P. (1981). A model for optimizing retail space allocations. *Management Science*, *27*(7), 822-833.

Fisher, M., & Vaidyanathan, R. (2014). A demand estimation procedure for retail assortment optimization with results from implementations. *Management Science*, *60*(10), 2401-2415.

Paat Rusmevichientong, Zuo-Jun Max Shen, David B. Shmoys, (2010) Dynamic Assortment Optimization with a Multinomial Logit Choice Model and Capacity Constraint. *Operations Research*, *58*(6) 1666-1680.

Guillermo Gallego, Huseyin Topaloglu (2014) Constrained Assortment Optimization for the Nested Logit Model. *Management Science*, *60*(10), 2583-2601.

James M. Davis, Guillermo Gallego, Huseyin Topaloglu (2014) Assortment Optimization Under Variants of the Nested Logit Model. *Operations Research 62*(2) 250-273.

Gallego and Topaloglu (2014)Constrained Assortment Optimization for the Nested Logit Model 2584, *Management Science 60*(10), 2583–2601, INFORMS

Agrawal, S., Avadhanula, V., Goyal, V., & Zeevi, A. (2019). Mnl-bandit: A dynamic learning approach to assortment selection. Operations Research, 67(5), 1453-1485.

Caro and Gallien: Dynamic Assortment with Demand Learning for Seasonal Consumer Goods Management Science 53(2), pp. 276–292, © 2007 INFORMS

Lee et al.: Stockout-Based Substitution and Inventory Planning in Textbook Retailing Manufacturing & Service Operations Management 18(1), pp. 104–121, © 2016 INFORMS

Hübner, A., Schäfer, F., & Schaal, K. N. (2020). Maximizing Profit via Assortment and Shelf-Space Optimization for Two-Dimensional Shelves. *Production and operations management*, *29*(3), 547-570.

Ross, K. W., & Tsang, D. H. (1989). The stochastic knapsack problem. IEEE Transactions on communications, 37(7), 740-747.