IJCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Vol.-7, Special Issue-14, May 2019 E-ISSN: 2347-2693

Retail Assortment Planning Using Data Science

S.K. Begur[1*], K. Gururaj[2], K.K. Bhowmik[3], K. Kumar[4], K. Malik[5], S.A. Rabbani[6], N. Tengli[7]

[1,2,3,4,5,6,7]Kigyan Techno Solutions, 887(23), Tulsi Arcade, 28th Main, Jayanagar 9th Block Bangalore, India
School of Computer Science & IT, REVA University, Bangalore, India

Corresponding Author: kiran@kigyan.com

DOI: https://doi.org/10.26438/ijcse/v7si14.1217 | Available online at: www.ijcseonline.org

***Abstract*—** This project is based on providing solution to retail giants to address their current Assortment strategy and increase their profit. Assortment AI is an AI project under the trade name reMark which is a social analytical platform of Kigyan Techno Solutions that provides retail marketing solutions. The concept of the project: Today majority of the giant retail companies are facing a lot of issues in their current assortment planning of their products. These include the total dump of products being 45%. This wrong assortment planning leads to products being out of stock which causes loss to businesses and major customer dissatisfaction, also this assortment planning requires a lot of manual strategies which are very costly and hence these assortment strategies then turn out to be costly, time taking, biased and working on mostly non relevant data. Due to the above factors the retail companies have understood that the current assortment planning strategies are not working and are only causing business loss and customer attrition. Hence there is a need of new assortment plan. The project that we are currently working on is building Assortment AI Module in a retail market and so the project is an Artificial Intelligence block which will process the current situation and will build a proper assortment plan. This would then address various problems that the retail companies were facing and present them with proper assortment plan which would include unbiased assortment strategy, the dump would be reduced to less than 20%, increase the customer satisfaction and reduce the customer attrition.

***Keywords*—**Kigyan Techno Solutions,data science,attrition,retail,assortment dumpstrategy

I. INTRODUCTION

A retailer's assortment is defined by the set of products carried in each store at each point in time. The goal of assortment planning is to specify an assortment that maximizes sales or gross margin subject to various constraints, such as a limited budget for purchase of products, limited shelf space for displaying products, and a variety of miscellaneous constraints such as a desire to have at least two vendors for each type of product.

Clearly the assortment retailer carries an enormous impact on sales and gross margin, and hence assortment planning has received high priority from retailers, consultants and software providers. However, no dominant solution has yet emerged for assortment planning, so assortment planning represents a wonderful opportunity for academia to contribute to enhancing retail practice. Moreover, an academic literature on assortment planning is beginning to emerge.

This new release centers around three pivotal territories of retail store network the board: (1) experimental investigation so retail market(2) variety and stock arranging and (3) coordinating value advancement into retail inventory network choices. This paper has been completely refreshed, developing the distinctive highlights of the first, while offering three new sections on late themes which reflect regions of extraordinary intrigue and pertinence to the scholastic and expert networks alike – stock administration within the sight of information errors, retail workforce the executives, and quick style retail procedures. The developments, exercises for training, and new innovative answers for overseeing retail supply chains are imperative in retailing, however, offer essential bits of knowledge and methodologies for a definitive powerful administration of supply chains in different enterprises also.

The retail business has risen as an interesting decision for scientists in the field of store network the board. It introduces a huge swath of animating difficulties that have since quite a while ago gave the setting of a significant part of the exploration in the region of activities research and stock administration. Be that as it may, as of late, progresses in registering abilities and data innovations, hyper-rivalry in the retail business, rise of numerous retail arrangements and appropriation channels, a consistently expanding pattern towards a comprehensively scattered retail organizes, and a superior comprehension of the significance of coordinated effort in the all-encompassing store network have prompted a flood in scholarly research on points in retail inventory network the board. Many production network advancements (e.g., merchant oversaw stock) were first considered and effectively approved in this industry and have since been received in others. Alternately, numerous retailers have rushed to receive bleeding edge rehearses that originally began in different enterprises. Retail Supply Chain Management: Quantitative Models and Empirical Studies, second Ed. is an endeavor to outline the cutting edge in this examination, just as offer a viewpoint on what new applications may lie ahead.

The item varieties convey a colossal effect on deals and gross edge, which is the reason exact arrangement arranging is such a high need. Shockingly for most retailers, the genuine variety arranging process brings out sentiments of fear and tension.

It's a complex and tedious procedure with numerous factors to consider. Previously, retailers depended exclusively on human judgment to settle on collection choices, which hasn't generally yielded positive outcomes.

The key is to deliberately settle on which level item varieties should be separated while restricting the multifaceted nature and cost of separating groupings to address the issues of your neighborhood customers. The answer for growing such a technique is in picking the fitting combination arranging programming to furnish you with dependable information and a method for deciphering the information.

Retail examination ought to be utilized to disentangle and streamline the variety arranging process; this will prompt augmenting deals and edges while diminishing the time it takes to design.

Inthispresentation,we'llclarifyhowretailinvestigation can enable your store to accomplish and keep up a focused edge. We've additionally incorporated a fast manual for helping you pick the correct collection arranging programming for your needs.

Combination arranging includes the measure of stock decision accessible and retail examination will enable you to figure out what stock assortment and grouping will prompt the most deals.

Combinationarrangingprogrammingwillbolsteryouin settlingonkeyassortmentandcollectionrelatedchoices, for example,

- Store and channel grouping
- Depth and broadness ofextents
- Creating concordance among global andrestricted itemgoes
- Space-obligedgroupings

Utilize hard information to enable you to decide assortment, variety, and item accessibility, while anticipating how clients will respond to these combination changes. Retail Analytics, given by the product, will support you:

- Comprehend your clients
- Expand on their wants
- Audit their shopping designs
- Concentrate your classes' past execution
- Take a gander at patterns in the market

To make the most out of your item variety, the product should take all the obtained information into thought and detail a firm intend to convey the correct stock, at the correct cost and at the perfect time, to the correct clients.

II. RELATED WORK

This section briefly discusses normative studies on assortment selection. For extensive reviews of this literature, we refer to Köketal. (2009) and Mantralaet al. (2009). Table 1 offers a number of key characteristics of the studies.

A first point of differentiation between studies isthe type of data that has been used. Borinetal(1994) and Borin and Farris (1995) consider a supermarket assortment selection based on synthetic model parameters. Their objective is to maximize the return on inventory subject to space constraints. They solve a “small” problem (6 SKUs) and a “large” problem (18 SKUs) using a simulated annealing heuristic. Using the same data as Borinetal(1994),Urban(1998) extends this methodology

by proposing agreed and genetic heuristic to solve the problem of jointly optimizing item selection, space allocation, and inventory policy. McIntyre and Miller (1999) consider an assortment selection problem based on data from an individual choice experiment regarding backpacks. They find a solution to their problem by applying an exhaustive search to a set of eight backpacks. The studies based on synthetic (Borin and Farris 1995, Borin et al. 1994, Smith and Agrawal 2000) or experimental (McIntyre and Miller 1999, Miller etal. 2010)data raise the issue of external validity. To overcome this issue, several of the studies reviewed in Table 1(including ours) use empirically observeddata. We structure the rest of the discussion alongside the six key challenges that retailers face when optimizing assortments. These challenges were briefly mentioned in 1 but are discussed in more detail below. The columns in Table 1 correspond closely with these challenges, which highlight the points of differentiation between studies.

1.Choosing among large numbers of SKUs: A typical product category contains many dozens or hundreds of SKUs (Bucklin and Gupta 1999). Modeling a large set of SKUs imposes challenges for the demand model. We must construct a parsimonious sales model to predict the sales for all SKUs in an assortment, including low-selling ones. UsingSKU-specific parameter (which is what most studies in Table 1 do) means that every additional SKU needs extra parameters (e.g., an intercept). To mitigate this problem, we adopt the attribute-based approach that replaces intercepts by attribute dummies (Fader and Hardie 1996).A large set of items complicates not only model estimation but also the optimization problem. It becomes very hard to solve, and exhaustive search is infeasible. Therefore, this study develops new heuristics that solve the problem within a reasonable amount of time.

2.Allowing for similarity effects: Key to any assortment optimization exercise is that the demand model accounts for similarity effects. Similarity effects imply that items whose attributes are more similar are more likely to compete for demand (Rooderkerk et al. 2011,Tversky 1972). Although some of the reviewed studies in Table 1 account for similarity effects, none of them allows for substitution patterns that are governed by attributes. We use choice theory to allow for attribute-based similarity effects in the demand model. Thus, we extend the Fader and Hardie (1996) approach by not only modeling preferences (intercepts) as a function of attributes but also modeling substitution patterns and cross-marketing mix effects.

3.Controlling for the marketing mix: Many studies ignore the role of (part of) the marketing mix (e.g., price, shelf space, promotional support) during demand estimation and/or assortment optimization. However, the marketing mix instruments have a profound effect on the demand for individual SKUs and need to be accounted for.

4.Accounting for assortment and price endogeneity: Retailers are likely to include SKUs in the assortment that (are expected to) sell well. Similarly, prices are set based on demand shocks that can be observed by the retailer but not by the researcher. Hence, assortments and prices are likely to be endogenous. Accommodating endogeneity is needed for consistent parameter estimates in the sales response function that are used in the optimization. To our knowledge this study presents the first method accounting for assortment and price endogeneity simultaneously.

5.Store level optimization: Differences in store characteristics and demographics of the trade area challenge retailers to customize their marketing mix to the store level (Bambridge 2007, 2008; Campoet al. 2000; Mantrala et al. 2009; Montgomery 1997), including assortments. Leading retailers such as Macy’s have realized that a “one size, one style fits all” strategy does not work and have begun tailoring a substantial part of their assortment to the local level (O’Connell 2008). Hence, Mantrala et al. (2009) label store-level customization asone of the key challenges that demands more research attention. To address this issue, we develop a model that allows for heterogeneity in parameters across stores, and we then conduct store- specific assortment (and price) optimization.

6.Joint assortment and price optimization: Besides the assortment, another important element of the retailer’s marketing mix is SKU prices. Personal communication with retailers has made it clear to us that most retailers use a two-stage optimization approach. First, they optimize the assortment. Next, they optimize the prices of the available SKUs. There seems to be potential to jointly optimize the assortment and SKU prices, and our approach offers this.

To conclude, Table 1 clearly shows that whereas some papers address some of the challenges, our paper addresses all of them. The key contribution is that we develop an implementable and scalable assortment optimization method that allows for theory-based substitution patterns yet is feasible to estimate and to optimize for real-life, large-scale assortments. Our new method includes (i) an attribute-based demand model to capture preferences, substitution patterns, and cross- marketing mix effects;and (ii) heuristics that optimizeretailer category profit subject to constraints such as the amount of available shelf space. The next section details the methodology of our approach

DESCRIPTION ABOUT LINEAR REGRESSION

Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things: (1) does a set of predictor variables do a good job in predicting an outcome (dependent) variable (2) Which variables in particular are significant predictors of the outcome variable, and in what way do they–indicated by the magnitude and sign of the beta estimates–impact the outcome variable? These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. The simplest form of the regression equation with one dependent and one independent variable is defined by the formula $y = c + b*x$, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.

Linear regression was the first type of regression analysis to be studied rigorously, and to be used extensively in practical applications.[4] This is because models which depend linearly on their unknown parameters are easier to fit than models which are non-linearly related to their parameters and because the statistical properties of the resulting estimators are easier to determine.

Linear regression has many practical uses. Most applications fall into one of the following two broad categories: If the goal is prediction, or forecasting, or error reduction,[clarification needed] linear regression can be used to fit a predictive model to an observed data set of values of the response and explanatory variables. After developing such a model, if additional values of the explanatory variables are collected without an accompanying response value, the fitted model can be used to make a prediction of the response.

If the goal is to explain variation in the response variable that can be attributed to variation in the explanatory variables, linear regression analysis can be applied to quantify the strength of the relationship between the response and the explanatory variables, and in particular to determine whether some explanatory variables may have no linear relationship with the response at all, or to identify which subsets of explanatory variables may contain redundant information about the response.

Linear regression models are often fitted using the least squares approach, but they may also be fitted in other ways, such as by minimizing the "lack of fit" in some other norm (as with least absolute deviations regression), or by minimizing a penalized version of the least squares cost function as in ridge regression (L2-norm penalty) and lasso (L1-norm penalty). Conversely, the least squares approach can be used to fit models that are not linear models. Thus, although the terms "least squares" and "linear model" are closely linked, they are not synonymous.

IV METHODOLOGY

Before formulating our model, we look at a significant challenge encountered when developing a sales model: modeling SKU sales while retaining parsimony. We explain how modeling SKU sales at the store level using attribute-based modeling helps overcome this challenge. We develop a model for SKU sales at the store level.

We choose store-level scanner data (as opposed to household-level scanner data) because they are often readily available, cost relatively little, and provide a census of sales of all SKUs in a store. This is especially importantforlow-selling items. Reliablesales measurementfortheseitems(whichcouldbe problematicwithhouseholddata)iscrucialto assortment optimization, because these are the items that are possibly eliminated.

A critical issue when modeling sales at the SKU level is parsimony. A typical product category consists of many SKUs, which means that a large set of SKU-specific intercepts would have to be estimated (Hardie et al. 1998). A parsimonious approach to overcome this problem is using an attribute-based way of modeling, proposed by Fader and Hardie (1996). This approach is motivated by the assertion that consumers do not form preferences for each individual SKU in a particular product category but that these preferences are derived from preferences for the underlying attributes (e.g., size, flavor, color). Theoretical justification for this approach is offered in economics (Lancaster 1971) and psychology (Fishbein 1967).

Our model thus replaces SKU-specific intercepts by SKU attributes as in Fader and Hardie (1996). We take the approach one step further to accommodate the substitution between SKUs, both because of the mere presence of other SKUs and because of their marketing mix activities. Using SKU-specific parameters would increasethe number of cross-effect parameters quadratic ally in the number of SKUs. Consequently, the number of parameters would quickly grow too large to estimate. To overcome this problem, we model cross-SKU substitution and cross-SKU marketing mix effects based on attribute-based similarity between SKUs.

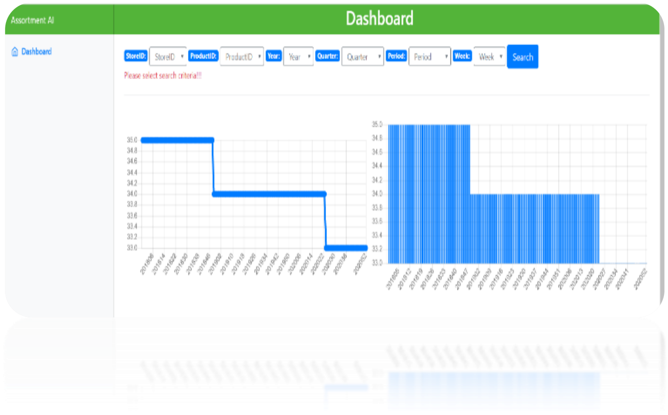
Pseudo code

```
Step 1: Start program
Step 2: Import all the necessary libraries
Step 3: Connect to MySQL database using credentials
Step 4: Write a sql query to select distinct store id from the transaction_summary tableand execute the query
Step 5: Write a sql query to select distinct product id from the  transaction_summary table and execute the query
Step 6: Write a sql query to select s_week id, sum(qty) from the transaction_summary table and execute the query
Step 7: Select X axis as week_id and Y axis as predicted quantities
Step 8: To find alpha beta values we apply a formula
numer += (X[i] - mean_x) * (Y[i] - mean_y)
            denom += (X[i] - mean_x) ** 2

          b1 = numer / denom
          b0 = mean_y - (b1 * mean_x)
          print(b1,b0)
          n = 158
         for j in range(n)
Step 9: To find the prediction values we use the formula as stated below:
y_pred = b0 + b1 * j
Step 10: If the algorithm is working fine the values are inserted in predication master table
Step 11: Else not Failed to insert into MySQL table
Step 12: Finally, the connection is successfully closed to my sql
Step 13: Finish
```

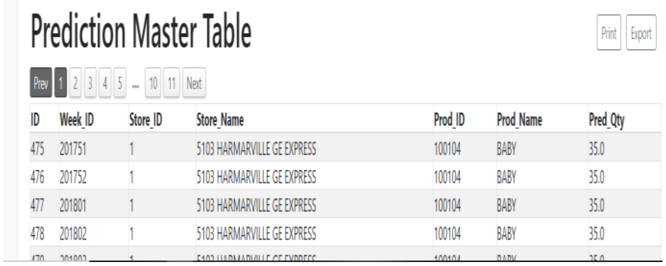
V. RESULTS AND DISCUSSION

Predictive Validity: We have estimated two versions of the sales model: the focal model with store-specific parameters and a model with homogeneous parameters across stores. In-sample fit (for the full three years of data) of the models was determined by computing the deviance information criterion (DIC; see Spiegelhalter et al.2002), which balances model fit and complexity (a lower DIC is better). In addition, both in- sampleand out- of-sample fit (predicting the last12year of data based on the first 212years of data) were established by computing the log marginal density (LMD), the correlation between actual and predicted sales,1 minus Theil's inequality index (1 – Theil's U), the mean absolute error (MAE), the mean absolute percentage error (MAPE), and the root mean squared error (RMSE). For the DIC, MAE, MAPE, and RMSE measures, lower values are more preferredOutcome.



Please select search criteria!!!

Outcome



Prediction Master Table

ID	Week_ID	Store_ID	Store_Name	Prod_ID	Prod_Name	Pred_Qty
475	201751	1	5103 HARMARVILLE GE EXPRESS	100104	BABY	35.0
476	201752	1	5103 HARMARVILLE GE EXPRESS	100104	BABY	35.0
477	201801	1	5103 HARMARVILLE GE EXPRESS	100104	BABY	35.0
478	201802	1	5103 HARMARVILLE GE EXPRESS	100104	BABY	35.0

VI CONCLUSION AND FUTURE SCOPE

The traditional no-search model utilizes the multinomial log it models of consumer choice without explicitly considering consumer search. The other two models incorporate consumer search into the traditional model.

The first search model, which we call the independent search model, presumes that there is an essentially unlimited pool of variants, which implies a low probability that the same product variant is carried in the assortment of two retailers. The second model, called the overlapping search model, presumes a limited pool of variants so that overlap between retailers is likely. The value of search to a consumer in the independent model is limited by the possibility of not finding an even better product if search is chosen. However, search is quite valuable to a consumer in the overlapping model because in that situation the consumer is never worse with search. Hence, consumer choice is most sensitive to search in the overlapping model and least sensitive to search in the independent model.
With each of the three assortment models we derive the optimal assortment and present a procedure for estimating the necessary parameters to implement the models in practice. Nevertheless, a retailer may choose to use the no-search assortment planning model even though search influences consumer choice. We presume

a retailer would apply the model iteratively: anassortment is chosen, then updated parameters are estimated using the sales data with that assortment, an updated assortment is chosen, etc. A potential issue with this method is that the no-search model does not distinguish between consumers who abstain from purchase because they do not like the variants in the current assortment in absolute terms (i.e., no variant exceeds their no-purchase utility) and consumers who do not purchase because they engage in search (i.e., no variant exceeds their no-search threshold).

Because the no-search model does not distinguish between the two reasons for a consumer to choose to not purchase any variant, we demonstrate that the estimated parameters with the no-search model are not independent of the assortment chosen and the iterative application of that model can lead to a heuristic equilibrium: an assortment that is optimal given the parameters estimated from the data observed with that assortment. We show that the heuristic equilibrium assortment never contains more variants than the optimal assortment and may very well contain fewer variants.

While the no-search assortment planning model yields biased parameter estimates, if the true model is independent search, we then find that the iterative application of the no-search assortment planning model generally leads to good assortments: out of 2868 scenarios, the average profit loss was only 0.29% and the maximum profit loss relative to the optimal profit was only 10.26%. However, the same does not hold if the no- search assortment planning model is applied when the overlapping search model is appropriate: out of 910 scenarios, the average profit loss was 10.26% and the maximum profit loss was 100%. Furthermore, the average profit loss across the tested scenarios increases as consumer search cost decreases.

We conclude that retailers should explicitly consider consumer search in their assortment planning process in markets with a limited pool of variants (overlapping search) and low consumer search costs. But in markets that are better characterized by the independent search model, the traditional no-search assortment planning model yields good results. We believe our model addresses an important issue in the retailing industry, yet several interesting questions remain unexplored. In this paper, we consider a static setting that ignores the competitive interactions associated with a retailer's assortment. Thus, an extension of our model couldexplore assortment planning with competing retailers in the presence of consumer search. Using retail data to empirically compare estimates from our model with the traditional no-search estimates also seems to be an interesting avenue for future research.

VII ACKNOWLEDGMENT

In performing our technical paper, we had to take the help and guideline of some respected persons, who deserve our greatest gratitude. The completion of this assignment gives us much Pleasure. We would like to show our gratitude to Mr. Santoshkumar Begur, Technical Manager at Kigyan Techno Solutions and Mr. Nikhil Tengli, Assistant Professor, School of C & IT, REVA University for giving us a good guideline for assignment throughout numerous consultations. We would also like to expand our deepest gratitude to Mr. Kiran Gururaj, Executive Director of Kigyan Techno Solutions, and Mr. Kanchan Kumar Bhowmik, Data Science Trainer all others who have directly and indirectly guided us in writing this technical paper.

Many people, especially our classmates and team members itself, have made valuable comment suggestions on this proposal which gave us an inspiration to improve our assignment. We thank all the people for their help directly and indirectly to complete our assignment.

REFERENCES

[1] K. M. Murali, L. Michael, E.K. Barbara, "Why is AssortmentPlanning so Difficult for Retailers? A Framework and Research Agenda", Journal of Retailing, vol. 85, no. 1, pp. 71-83, 2009.Show Context CrossRef Google Scholar.

[2]M. K. Mantrala, K. Manfred, S. Leopold, Krafft Manfred, MuraliK. Mantrala, "Entrepreneurship in Retailing: Leopold Stiefel's 'Big Idea' and the Growth of Media Markt-Saturn" in Retailing in the 21st Century: Current and Emerging Trends, Berlin/Heidelberg/New York:Springer., 2008.

[3]K. Wong, Major Retailers Rev Up Green Campaigns, February 2008.

[4]P.D. Larson, R.A. DeMarais, "Psychic stock: An independent variable category of inventory", International Journal of Physical Distribution and Logistics Management, vol. 20, no. 7, pp. 28-34, 1990.

[5]Albuquerque P, Bronnenberg BJ (2009) Estimating demand heterogeneity using aggregated data: An application to the frozen pizza category. Marketing Sci. 28(2):356–372. [6] David A. Freedman *(2009). Statistical Models: Theory and Practice*. Cambridge University Press. *p. 26. A simple regression equation has on the right hand side an intercept and an explanatory variable with a slope coefficient. A multiple regression equation has two or more explanatory variables on the right hand side, each with its own slope coefficient*

.[7] *Hilary L. Seal (1967). "The historical development of the Gauss linear model". Biometrika. 54 (1/2): 1–24.* doi*:*10.1093/biomet/54.1-2.1. JSTOR 2333849.

[8] *Yan, Xin (2009),* Linear Regression Analysis: Theory and Computing, *World Scientific, pp. 1–2*, ISBN 9789812834119, *Regression analysis ... is probably one of the oldest topics in mathematical statistics dating back to about two hundred years ago. The earliest form of the linear regression was the least squares method, which was published by Legendre in 1805, and by Gauss in 1809 ... Legendre and Gauss both applied the method to the problem of determining, from astronomical observations, the orbits of bodies about the sun.*