Retailer's Dilemma: Personalized Product Marketing to Maximize Revenue

Ryan Ferrera
School of Data Science
University of Virginia
Charlottesville, U.S.
re4np@virginia.edu

John Mark Pittman
School of Data Science
University of Virginia
Charlottesville, U.S.
jmp3fqb@virginia.edu

Martin Zapryanov
School of Data Science
University of Virginia
Charlottesville, U.S.
mz2fe@virginia.edu

Oliver Schaer
Darden School of Business
University of Virginia
Charlottesville, U.S.
os9vu@virginia.edu

Stephen Adams
Department of Engineering Systems and Environment
University of Virginia
Charlottesville, U.S.
sca2c@virginia.edu

***Abstract*—Companies face many challenges when it comes to increasing revenue, but one of them is how to turn low or no-revenue customers into high revenue customers. When surveying their opportunities to do so, companies often turn to marketing. However, when deciding which among the many options to market to an existing customer, with only finite resources to do so, companies must make a choice rooted in the expected value that reflects the customer's interest in the offer and the business value of that product or feature. This paper explores techniques to identify customers and study product allocations that allow to increase revenue by nudging customers from lower-revenue groups to higher-revenue groups by recommending the next product to market. Our approach utilizes k-means clustering to identify customer segments based on the recency, frequency, and monetary value (RFM) of their purchases. Further, we demonstrate that an association analysis technique called Market Basket Analysis (MBA) can be extended to not only identify products commonly purchased with the products a specific customer already has, but also to identify which products are associated with higher-revenue customer behavior. We close with a discussion on how these two techniques (clustering and association analysis) can be combined to optimally nudge customers from low-revenue groups to high-revenue groups by incrementally marketing products that more-closely align with the purchasing behavior of higher-revenue customers.**

Keywords—personalized marketing, customer segmentation, market basket analysis, customer relationship management

I. Personalized Marketing

Personalized marketing involves crafting marketing experiences that target a specific customer or type of customer. The arguments in favor of a personalized marketing strategy are twofold. First, it improves the customer experience by only delivering relevant promotions. Second, since customers are served more relevant promotions, they are more likely to generate revenue for the firm. According to estimates by McKinsey & Company, personalized marketing can reduce customer acquisition costs by as much as 50 percent and increase revenue by 5 to 15 percent [1]. The research shows that it enhances customers' lives and increases engagement and loyalty. Personalized marketing is expected to drive competitive advantage and companies across a variety of industries have started to transform their customers' journey through personalization.

Online shopping is a canonical example of personalized marketing strategies at work. Developing and implementing this personalization requires proper collection and analysis of customer data; however, as literature and industry practice indicate, a successful personalized marketing capability can add enormous value to the firm.

A. Related Work

Personalized marketing in customer relationship management (CRM) has been researched in the past as an area which can lead to greater customer satisfaction and higher profits. Arora, *et al.* describe personalization as a form of one-to-one marketing as part of which the firm decides, typically on previously collected customer data, what marketing mix is suitable for the individual [2]. In their textbook *Relationship marketing: Theory and Practice*, Buttle emphasizes the long-term financial benefits of relationship marketing [3]. They explain that relationship marketing is based on two economic arguments. First, it is more expensive to win a new customer than it is to retain an existing customer. Second, the longer the relationship between the business and the customer, the more profitable it is for the firm. This leads us to the conclusion that personalized marketing has two goals. One, to determine which customers should be marketed and how to reach them. Two, what products or features to market to them.

Another study by Rust and Verhoef further explains that personalization in CRM can take two forms [4]. One approach companies can take to differentiate their marketing efforts is based on customer differences in their personal characteristics and historical behavior with the company. Probably one of the best-known examples as described by Nunes and Kambil is Amazon [5]. Amazon, as well as the e-commerce arena in general, finds patterns in customer's previous choices and/or demographics to recommend other products for purchase. Another approach companies can use to differentiate their marketing efforts is based on customer differences in response. This method models heterogeneity in responsiveness to marketing efforts including psychographic data such as lifestyle, opinions and interests of customers. In their research of optimizing the marketing intervention mix, Rust and Verhoef also developed a model which models based on not only responsiveness to the marketing efforts but also past behavior and personal characteristics [4].

B. Our Contribution

We explore two methods of extracting insights from customer transactional information that can be used to develop personalized marketing strategies. Sourcing exogenous customer information, whether demographic or psychographic, can be impractical or cost prohibitive, and, in certain industries, expressly forbidden. We choose to focus on using historical information that the firm can easily collect as a part of its normal business in order to preserve broad applicability into other business settings.

Our two approaches each focus on one of the personalized marketing sub-objectives. First, we explore techniques that help the firm determine which customers to market to and derive insights to guide engagement strategies. We propose a customer lifetime value (CLV) focused approach to customer segmentation. This allows the firm to identify desirable behavior groups within a customer base, impose a firm-designed hierarchy to those groups, and finally to use this information as a basis for overall marketing strategy.

Second, we introduce association analysis techniques to identify relevant products or features to promote to individual customers. We then extend this approach by providing the firm a mechanism with which to estimate the incremental revenue that the firm can expect to realize by promoting a given product or set of products. In doing so, we enable the firm to evaluate promotional costs against expected revenues, which allows the firm to ultimately use profit and loss to guide its marketing decisions.

While the two approaches can be implemented independently of one another, when taken together, they represent a decision-making framework that fits customers into groups based on revenue generating potential. We then encourage movement from less to more desirable groups by recommending the next product to market.

II. The Online Retail Dataset

For our data we used a public data set of 25,900 online retail transactions for a UK-based online retailer occurring between December 2010 and December 2011 [6]. It included a total of 4,372 customers across 38 different countries and a mix of 4,223 unique products. The median number of items per basket is 100, consisting of 10 unique products and total revenue of $207.54.

The primary challenge with this data set is that it does not include any record of marketing interventions. As such, we cannot explore the down-stream impacts of marketing interventions from the data and instead focus our paper on demonstrating a technique to address the challenge of personalized marketing, while leaving its evaluation for further research.

III. Customer Segmentation

Recall that our purpose in performing customer segmentation is to group customers in a way that allows the firm to develop an internal hierarchy of those groups. This hierarchy will be used to guide marketing decisions with the objective of encouraging customers, in the form of marketing interventions, to move from segments with low revenue generating potential to one with greater potential.

A. *Theoretical Foundations*

Customer lifetime value has been defined as the present value of all future profits generated by a customer over the course of his or her relationship with the firm (Gupta, *et al.*) [7]. The recency, frequency, and monetary value (RFM) model is a common and robustly supported method of representing CLV (Bult and Wansbeek) [8]. It decomposes CLV as a function of customer purchase recency, customer purchase frequency, and customer purchase monetary value. Numerous methods have been developed to derive the individual customer attributes and to combine them to estimate CLV. Thus, by using the RFM model as a basis for customer segmentation, we group customers based on their estimated value to the firm. Furthermore, the RFM inputs are easily derived using the transactional data contained in the Online Retail dataset.

Our specific process for deriving the recency, frequency, and monetary value features will be discussed in greater detail in Section B; however, in general, the RFM inputs can be defined as follows [8]:

- **Recency:** either time periods since last order or the number of consecutive marketing engagements without a response.
- **Frequency:** number of purchases made over a given time period.
- **Monetary Value:** amount of money spent during a given time period.

From a practical standpoint, the time period can be treated as a hyperparameter that may vary based on use case. For example, if a firm typically develops an annual marketing strategy that is fine-tuned on a quarterly basis, then that firm might consider a 12-month time period over which to calculate the RFM measures. Those annual measures could then be recalculated at the start of each quarter to guide strategic adjustments.

The RFM model is not without its criticisms, chief amongst them being that past behaviors are not necessarily the best predictors of future behavior (Kumar, *et al.*) [9]. However, the RFM model is able to identify behavior with sparse customer-level data, which is often the case in practical settings.

B. *Data Preparation*

We created a robust data cleaning and feature engineering pipeline to coerce the Online Retail dataset into the RFM measures. We first removed all observations with missing Customer IDs, since we ultimately have to aggregate the dataset at the customer-level. Second, we identified observations that contained either extremely high Order Quantity absolute values or unusual item descriptions such as 'POSTAGE' or 'MANUAL.' After removing these problematic observations, the resulting dataset contained 4,361 customer IDs out of the original 4,737.

We next began the feature engineering step of the process. We defined a new variable called Item Value that is defined as the product of Item Quantity and Unit Price. Next, since our dataset records transactions at the item-level, we aggregate at first the invoice and then customer level. Along the way, we calculate the following features for each customer:

- **Last order date:** date of most recent invoice
- **Order frequency:** total number of all invoices
- **Monetary Value:** mean dollar value spent per order

- **Unit price:** mean unit price of items ordered
- **Quantity:** mean number of items ordered at a time

Notice that the Order Frequency and Monetary Value features are ready for input into the RFM model; however, we must still derive a Recency measure. In practice, this measure will vary depending on when the data is collected, and will be updated on an ongoing basis. In our case, we defined an index date to be one day after the last observed transaction, such that

- **Recency:** days between Last Order Date and December 10, 2011

We then take logarithms of all variables and standardize them by subtracting variable mean from each observation divided by the standard deviation, such that all values are on a [-1, 1] scale for subsequent clustering.

For segmentation, we used the RFM input variables and K-Means Clustering for its simplicity and interpretability. To select the optimal number of groups, k-value, we used an iterative process that performed K-Means clustering with k-values ranging from 1 to 21. At each step, we recorded the 'within-cluster distance' and generated a simple line plot depicted in **Figure 1**, and selected k=6 as the number of customer groups moving forward.

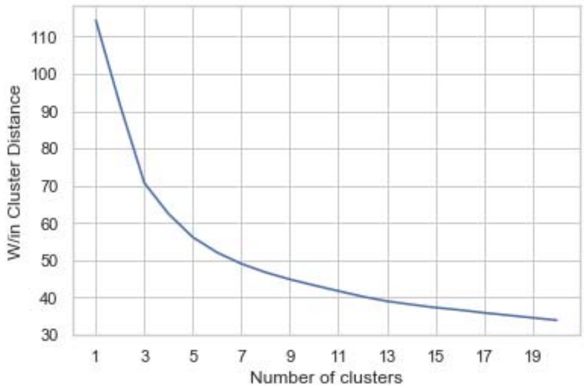

Fig. 1. Within-cluster distance as a functinon of k.

C. *Results and Discussion*

Having grouped customers by behavior, we can now analyze the resulting segments and begin to develop an overarching engagement strategy. **Table 1** shows summary statistics for each customer segment.

TABLE I. CUSTOMER SEGMENT SUMMARY

Group	Mean Recency	Mean Freq.	Mean Order Vol.	# Cust. in Group	Avg. Qty
To Re-engage	63	5.5	$319	1183	192
Returners	242	1.5	$0	41	NA
Churners	228	1.5	$277	1164	166
Mild Majors	7	3.9	$322	544	197
Rare Spenders	47	1.4	$365	919	240
MVPs	8	19.7	$393	510	227

We immediately notice a distinct hierarchy in the customer segments and begin to make inferences about the customers that belong to each segment. For example, the customer segment that we call the *MVPs* would most likely be the segment that we want to nudge customers into. They gave the highest order *frequency* and *monetary value* and low average *recency* score. By contrast, both the *Rare Spenders* and *To Re-engage* segments have relatively strong *monetary value* scores, but are worse on the *recency* scale. Together they represent almost half of the customer base. The firm stands to gain significant revenue by re-engaging encouraging these customers to behave more like the *MVP* segment, in terms of *recency* and *frequency*.

The firm can also use this segmentation to tailor its marketing approach according to customer behavior. For the *Rare Spenders* customers (low *frequency*), the firm might explore promotional offers that target this feature of interest. For the *Churners* or *To Re-engage* segments, the firm may simply want to bring those customers back by offering say a 25% off discount on their next order. Thus, as we have demonstrated, customer segmentation using the RFM model can be used to develop internal hierarchies of customer behavior, frame marketing objectives, and inform specific, customer-centric marketing tactics.

IV. MARKET BASKET ANALYSIS

A. *Technique Overview*

Market Basket Analysis (MBA) is a popular data mining technique to identify interesting relationships among data sets. The use of the word "sets" here is very deliberate as association analysis looks for common sets or "baskets" of items of some kind. An example of a basket reflecting a single grocery transaction would be {Bread, Milk}. This set reflects that Bread and Milk were purchased together in this transaction. This item set would be considered frequent if these items appeared together in a certain proportion of all transactions (called itemset Support) greater than a user- set threshold. MBA goes one step further to determine the probability of certain frequent disjoint itemsets occurring together. These interesting relationships between item sets are represented in the form of association rules. A canonical example of such a rule is {Antecedent} → {Consequent}: {Diapers} → {Beer}. This rule suggests that when people buy Diapers (the antecedent set), they also frequently buy Beer (the consequent set). These rules are generated and determined to be "interesting" if the rule's probability metric exceeds a user-set threshold. Common probability metrics used at this stage include (Tan, *et al.*) [10]:

- **Support**: the probability of both the antecedent and consequent sets occurring in the same basket
- **Confidence:** the conditional probability of the consequent set being in a basket if the antecedent set is already in the basket
- **Lift:** how much more often the antecedent and consequent occur together than if they were statistically independent (lift = 1 would reflect independence, lift > 1 would reflect a positive association)

The output of an association analysis is a list of {Antecedent} → {Consequent} rules that satisfy a user-set threshold that reflects a rule being "interesting". The consequents of these rules of interest are the candidates for recommendation to the customer. An example of how this could be done is through a mechanism like "*People like you also have product X and use feature Y. Click here to learn more.*"

B. An Extension of Market Basket Analysis for Company Decision-Making

While the MBA addresses the first part of the two-sided problem by identifying candidate sets, it must be extended to examine which of these candidate sets would also be financially beneficial to the company. To do this, we developed an algorithm to go through each candidate rule ({A} → {C}) and create two customer populations: (p1) customers who have made at least one order including antecedent items, but none of the consequent items and (p2) customers who have made at least one order with antecedent and consequent items. We then go on to examine the behavior and financial metrics for these two customer populations. The metrics of interest will differ based on the specific application, so for the online retail data set we compared total revenue, average order revenue, and number of orders per month across all customer activity (not just the transactions specifically matching the antecedent or antecedent + consequent transaction(s)) for the two populations.

Since what we are primarily interested in is an estimate of potential return for the company by a customer adopting additional products or product features, we are able to engineer an estimate of the increase in revenue over a one-year period:

$$1Y\ Rev\ \Delta = 12\ \times\ (AvgOrderRev_{p2} \times OrdersPerMo_{p2} - AvgOrderRev_{p1} \times OrdersPerMo_{p1}) \qquad (1)$$

This metric captures the difference in average monthly revenue for the two customer populations, those exhibiting the same behavior as your target customer and those exhibiting the behavior you are considering nudging your target customer towards (in this case, making an order that includes the consequent items).

This can be extended further to calculate the expected value of offering a consequent set to a customer with only an antecedent set if you were to consider the confidence of the association rule (the conditional probability that the consequent appears with the antecedent) as the probability that a customer will accept your offer. The expected value of the change in one-years' worth of revenue would be:

$$E(1Y\ Rev\ \Delta) = 1Y\ Rev\ \Delta \times Confidence \qquad (2)$$

The implication of this extension for company decision-making would be to consider marketing consequents to the target customer that provide an expected value greater than the expected cost to reach and convert the customer to that product or product feature.

C. Data Preparation

We will now walk through the application of this approach on the online retail data set. We started by formatting the data to specifically suit it for MBA. This includes dropping rows with missing product description or customer ID values and stripping extra white spaces from the product descriptions to ensure uniformity. For simplicity, we also removed returns from the dataset to focus solely on products purchased together and removed the few erroneous orders like the order for 81,000 [PAPER CRAFT, LITTLE BIRDIE] items. These orders were determined to be erroneous because they were accompanied by a corresponding return order.

Once the data was cleaned and prepared for MBA, we began generating frequent itemsets. Since there are 4,223 unique products and 25,900 transactions, we set a low item set support threshold of 0.005 (itemset must occur in minimum ~130 transactions) to maintain a broad set of potential itemsets. This keeps open the possibility of finding strong product associations, even if they occur infrequently. From these frequent itemsets, we then created a list of interesting association rules. The threshold we used to determine an "interesting" association rule is a lift >1, meaning only association rules that exhibit a positive association (i.e. the antecedents and consequents appear together more frequently than expected given their individual frequencies). This yielded 12,988 interesting association rules, which are all possible combinations of antecedents and consequents that have a lift >1.

We illustrate the results of our analysis based on an example that sets the initial basket with the two most popular items {LUNCH BAG SUKI DESIGN, LUNCH BAG RED RETROSPOT}. Therefore, all subsequent presented associations are with regards to our sample basket. By passing this basket/antecedent to our decision model, our model runs through all of the relevant association rules and calculates the *1Y Rev* Δ and E(*1Y Rev* Δ) metrics discussed in the prior section.

D. Results and Discussion

From the 12,988 interesting association rules, 71 were relevant to the example basket, i.e. 71 had antecedents equal to this basket. These 71 rules are then displayed by E(*1Y Rev* Δ) in descending order. That is, the consequent expected to yield the greatest one-year increase in revenue is at the top. The top three association rules returned by our decision model for our example basket can be seen in **Table 2**.

TABLE II. DECISION MODEL OUTPUT[a]

Consequents	Support	Conf.	Lift	1Y Rev△	E(1Y Rev△)
[JUMBO BAG RED RETROSPOT]	0.84%	34%	3.99	$1,308.90	$450.84
[JUMBO STORAGE BAG SUKI]	0.72%	30%	7.13	$1,350.97	$402.29
[JUMBO SHOPPER VINTAGE RED PAISLEY]	0.52%	21%	5.00	$1,752.35	$373.83

a. Example for [LUNCH BAG SUKI DESIGN, LUNCH BAG RED RETROSPOT] basket/antecedent.

The consequents seen in **Table 2**, sorted by the expected value of the one-year revenue delta, do not display all of the options available to the company. In **Figure 2** we illustrate the expected revenue decay for all 71 association rules that help decision makers to determine the number of consequents to recommend to the customer. An example interpretation for the top consequent would be as follows (from the first association rule in **Table 2**), *c1*: {JUMBO BAG RED RETROSPOT}. On average over a 12-month period, customers who purchased *b1*: {LUNCH BAG SUKI DESIGN, LUNCH BAG RED RETROSPOT, JUMBO BAG RED RETROSPORT} all together yielded $1,308.90 more in revenue from all of their purchases (not just those including these items) than customers who purchased *a1*: {LUNCH BAG SUKI DESIGN, LUNCH BAG RED RETROSPOT} alone. That is to say, customers who exhibited a certain behavior (purchasing the full basket) exhibited more favorable outcomes (greater revenue) across all of their purchases than customers who exhibited another behavior (purchasing only the antecedent basket). The goal would be to convert the target customer to this other behavior with the expectation that they would begin to exhibit the more favorable outcomes as well.

If we were to consider the expected value of recommending *c1* to an average customer with *a1*, we would expect an average revenue return of $450.84 per recommendation. This is based on the confidence of the association rule which shows that *c1* occurs with *a1* 34% of the transactions where *a1* occurs. As an input to decision-making, this could be interpreted as a net positive recommendation to make to a customer if the cost of converting the customer to the full basket is less than $450.84.

A plot of expected revenue decay like that in **Figure 2** can be used by a decision-maker to consider how many consequents to consider recommending to the customer.

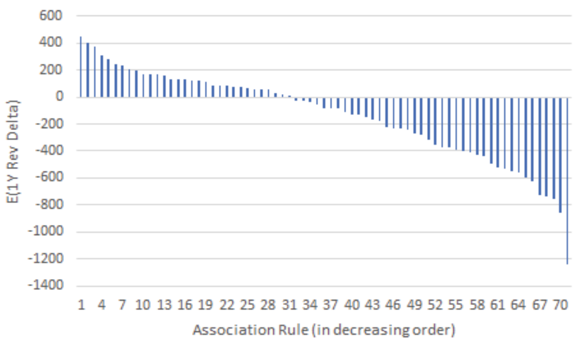

Fig. 2. Expected revenue decay.

As can be seen from **Figure 2**, the first 30 consequents yield a positive expected one-year increase in revenue, but these must be considered alongside the cost of marketing to the customer. If the marketing cost to convert a customer to include these consequent items is $200, for example, then it may only make sense for the company to look at the first 10 association rules, as those are the only ones that expected to yield a revenue increase of $200 or more. A company could consider offering customers special coupons for those 10 consequent items in the hope that the customer elects to add at least one.

V. Conclusions

In this paper, we have endeavored to explore techniques to assist retail companies with identifying the right products to market to individual customers that are agreeable to the customer, but also maximally beneficial to the company in terms of revenue. We outlined two separate approaches that segments customers based on behavior (recency, frequency, and monetary value) and utilized Market Basket Analysis to identify the next product to market. We demonstrated that by extending the Market Basket Analysis, one can identify the best products to recommend based on expected future revenue increases. These two approaches, taken together, form a decision-making framework that can be used to guide the development of personalized marketing strategies.

VI. Further Research and Testing

In this paper, we focused on developing the segmentation and association analysis techniques we foresaw as relevant to the personalized marketing challenge. However, we have presented them here largely as functionally independent of one another. We would recommend further research as to the integration of these techniques for a more robust decision model. As an example as to how they could be integrated, consider that the MBA assesses association rules in a naive fashion, meaning they rules are mined from the entire data set. Since our intent is to nudge customers from low revenue to high revenue groups, it is possible to mine association rules for the high revenue groups identified through segmentation and then select the relevant rules (those with antecedents matching the low revenue target customer basket) from that rule list. By recommending relevant products to the antecedent basket that more closely align with higher revenue behavior, it is possible that this will lead to a more effective nudging strategy.

Our model could be further improved by incorporating better mechanisms for dealing with new customers. In these situations, the RFM model struggles to deliver meaningful insights, and this should be considered an area for future research.

Since there are no marketing interventions to track and measure in this dataset, we would recommend testing and validating our approach through A/B testing similar to what Bradlow, *et al.* [11] used to successfully demonstrate the validity of their techniques. We would recommend taking the population of customers with a target basket (perhaps a popular basket from a low-revenue group) and split them into two groups, one as a hold out and one as a testing group. For the hold out group, we recommend doing nothing and observing their revenue over time. For the test group, we recommend marketing one or more of the top consequents/products from our decision model and monitoring the percentage of adoption and the ensuing revenue for that group over time. By comparing these two groups over time, the company would then be able to validate whether the inferences about revenue increases over time were accurate.

Acknowledgment

We would like to thank our University of Virginia advisors for the insight and guidance on the use case for this paper.

REFERENCES

[1] B. Gregg, H. Kalaoui, J. Maynes, and G. Schuler, "Marketing's Holy Grail: Digital personalization at scale", November 2016, Mckinsey & Company.

[2] N. Arora, X. Dreze, A. Ghose, J. D. Hess, R. Iyengar, *et al.* "Putting One-to-one Marketing to Work: Personalization, Customization, and Choice", Marketing Letters, vol. 19 (3), pp. 305-321, December 2008.

[3] F. Buttle, "Relationship Marketing: Theory and Practice", 1st ed., SAGE Publications, 1996, Chapter 1.

[4] R. Rust and P. Verhoef, "Optimizing the Marketing Interventions Mix in Intermediate-Term CRM", Marketing Science, vol. 24 (3), pp. 477-489, Summer 2005.

[5] P. F. Nunes and A. Kambil, "Personalization? No thanks", Harvard Business Review, vol. 79 (4), pp. 32-34, March 2001.

[6] D. Chen, , Online Retail II Data Set, UCI Machine Learning Repository, 2019, archive.ics.uci.edu/ml/datasets/Online+Retail+II.

[7] S. Gupta, D. Hanssens, B. Hardie, W. Kahn, V. Kumar, *et al.*, "Modeling Customer Lifetime Value," Journal of Service Research, vol. 9 (2), pp. 139–155, Novermber 2006.

[8] J. Bult, T. Wansbeek, "Optimal Selection for Direct Mail," Marketing Science, vol. 14 (4), pp. 378-392, 1995.

[9] V. Kumar, R. Venkatesan, T. Bohling, D. Beckmann, "The Power of CLV: Managing Customer Lifetime Value at IBM" Marketing Science Vol. 27 (4), pp. 585-599, July-August 2008.

[10] P. N. Tan, M. Steinbach, A. Karpatne, V. Kumar, "Introduction to Data Mining", 2nd ed., Pearson, 2018, Chapter 5

[11] E. Bradlow, M. Gangwar, P.Kopalle, S. Voleti, "The Role of Big Data and Predictive Analytics in Retailing," Journal of Retailing, vol. 93 (1), pp. 79-95, March 2017.