Optimising marketing strategies by customer segments and lifetime values, with A/B testing

Received in revised form 12th June, 2021

Paromita Guha
Co-founder and Data Scientist, Axiomatic Data, USA

Paromita (Paro) Guha is co-founder and data scientist at Axiomatic Data, a FinTech data startup. Paro has a PhD in economics with experience in econometrics, machine learning, predictive modelling and experimentation. She has collaborated with cross-functional teams, building end-to-end FinTech data products from ideation to market launch.

Axiomatic Data, 505 West End Ave, New York, NY 10024, USA
E-mail: paro@axiomaticdata.com

Christina Echagarruga
Data Scientist, Facebook, USA

Christina Echagarruga is a data scientist at Facebook. She has experience in signal processing, experimentation, machine learning, data cleaning methods and analytics to create and implement solutions with business value. Chris has a PhD in biomedical engineering from Penn State University.

Facebook, 1 Hacker Way, Menlo Park, CA 94025, USA
E-mail: chrisecha@fb.com

Eva Qi Tian
Data Scientist, Vanguard Group, USA

Eva Qi Tian is a data scientist at Vanguard Group, responsible for customer experience analytics. She has eight years' research and working experience in experimentation, behavioural decision research, econometric modelling and machine learning. She holds an economics PhD from Michigan State University.

Vanguard Group, 100 Vanguard Blvd, Malvern, PA 19355, USA
E-mail: eva_tian@vanguard.com

Abstract Every customer has different needs and purchasing behaviour. This paper shows how data science tools such as machine learning, artificial intelligence and A/B testing enable marketers to segment their target market, identify the most loyal high-value customers and their purchasing patterns, and calculate the lifetime value of these customer segments to optimise marketing strategies and campaigns. The paper also argues that A/B testing helps marketers make unbiased data-driven decisions, making it the gold standard for identifying the best marketing strategy.

KEYWORDS: predictive analytics, customer, segmentation, lifetime value (LTV), experimentation, A/B testing

INTRODUCTION

Over the last couple of decades, marketing strategies have evolved rapidly in response to changing trends and challenges in digital e-commerce, to become reliant on Big Data and computing power.[1] From designing the right marketing content for specific customer segments to measuring the attribution of various channels, marketing today is much more science and analytics than ever before. Indeed, today's companies need to leverage both the science and the art of data science (machine learning, artificial intelligence and experimentation) tools to derive actionable insights from available customer data and drive business outcomes.

Machine learning (ML) — a part of artificial intelligence (AI) — is the study of predictive models based on sample data (training data) to help machines make predictions or decisions without being programmed to do so. AI on the other hand, is machines imitating human intelligence. ML and AI tools must be employed in digital marketing to help marketers use campaign data to drive decision making that boosts customer engagement and loyalty through customer segmentation and personalisation.[2] For example, while AI tools may be used to predict a customer's next purchase and recommend products using natural language processing (NLP) to process reviews, predictive ML tools help to identify patterns in a customer's past purchases to make data-driven forecasts of future purchases.

A data-driven marketing planning process must, however, consider the product, collaborators, customers, competitors and the context in which the business operates. The art of asking the right questions to design actionable hypotheses and key measurable metrics is equally imperative. Meanwhile, the decision of whether to use ML or AI to personalise marketing content or an A/B experiment to choose marketing campaigns and strategies depends largely on the size, life cycle and scale of the company. Most importantly, it depends on the business questions and available data. Consequently, marketing strategies, such as customer segmentation, personalisation and brand positioning using the product, pricing, promotions, place (distribution channels) and people (customer support), should be optimised using those tools that align best with the company's resources and key performance indicators (KPIs).

To drive the KPIs in the modern era of digital marketing, analytics professionals must answer such questions as 'What internal and external data do I need to measure and improve my key metrics?'; 'Integrating data from every campaign, how do I use ML algorithms for customer segmentation, customer lifetime value calculation and channel attribution?'; 'How do I A/B test to find the winning marketing campaign?'; 'How do I design, execute and interpret A/B test results to optimise my marketing content by customer segments?'

To help marketers answer such questions, this paper illustrates the end-to-end marketing analytics process, using ML models and experimentations to:

- optimise marketing strategies by customer segments using ML algorithms based on recency, frequency and monetary value of customer purchase;
- calculate customer lifetime value (LTV), predict future lifetime value segments of customers to adapt marketing strategies to retain customers with high LTV;
- design A/B tests, ie randomised control experiments to determine the winning marketing strategies.

CUSTOMER SEGMENTATION

Segmentation helps marketers better understand customers' needs in order to optimise and tailor campaigns to those customer segments most likely to purchase the company's products.

To demonstrate how ML tools help marketers to segment and target customers, the present study uses a dataset containing all transactions conducted during the period from 12th January, 2010 to 12th September, 2011 by a UK-based online retail firm that specialises in selling unique gift items to wholesalers.[3] That the customers are wholesalers is noteworthy because customers are more likely to purchase multiple units of the same product.

The field 'InvoiceDate' (ie when the transaction occurred) is used to identify the recency and the frequency of the purchases made by the customer, identified by 'CustomerID' (ie unique customer identifier). The fields 'Quantity' (ie quantities of each product per transaction) and 'UnitPrice' (ie product price per unit) for each CustomerID are used to compute the monetary value of purchases. The calculated recency (R), frequency (F) and monetary value (M) of customer purchases is used to assign an RFM score to each purchase.

Using unsupervised ML algorithms such as K-means and hierarchical clustering for the dataset (Table 1), the customers are partitioned according to patterns in their purchasing behaviour. For K-means, which use centroid or partition-based clustering, the silhouette score or elbow method is used to identify the optimum number of clusters to which customers should be assigned based on their overall RFM score.[4] The centroid is then used to form the customer segments. Hierarchical clustering, by contrast, takes a top-down or bottom-up approach and adopts an agglomerative form of customer segmentation using dendrograms. The results of the two clustering methodologies are similar except for a very large dataset, when the shape of clusters may differ a little.

Table 1: Retail data for InvoiceNo. 536365 (CustomerID 17850, 1st December, 2010)

StockCode	Quantity	UnitPrice
85123A	6	$2.55
71053	6	$3.39
84406B	8	$2.75
84029G	6	$3.39
84029E	6	$3.39

Customers purchasing most recently, most frequently and spending more are clustered into high-value active customer segments with a high RFM score, separated from other customer segments[5] based on the Euclidean distance between the centroids of the clusters with different RFM scores. Once the high, middle and low-valued customer segments are identified, marketing strategies and campaigns are tailored for each customer segment based on the overall value score.

Based on Figure 1, customers that are assigned to cluster 1 have a higher frequency, monetary value and more recent purchases, hence cluster 1 is the high-value customer segment. The mid-value customer segment assigned to cluster 3 has a similar frequency and monetary value of purchases as the low-value cluster 2 but more recent purchases than cluster 2. The marketing campaigns are typically tailored by clusters, with cluster 1 consisting of the most sought-after customers and cluster 2 the least valued ones.

In the hierarchical clustering algorithm, points that are closest to each other are organised on the basis of Euclidean distance and the proximity of points dictates the clusters of a dendrogram. This assigns customers to different value clusters based on the similarities and differences in their purchasing patterns.

OPTIMISING MARKETING STRATEGY BY CUSTOMER SEGMENTS

It is evident from the scatter plots (Figure 2) of the RFM value-based customer segmentation that marketing campaigns and targeted strategies must be tailored for the

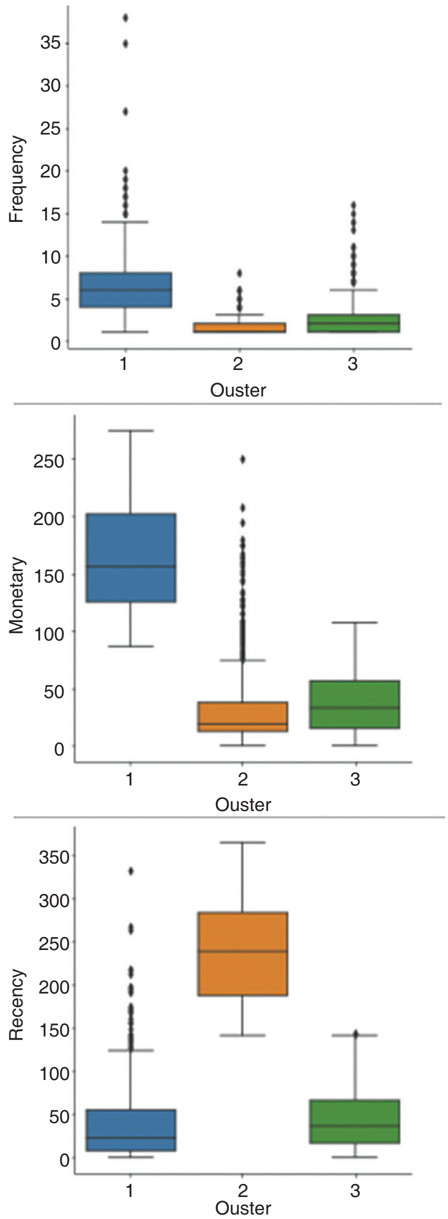

Figure 1: Customer segments using RFM

individual customer segments[6] to increase conversion rate and customer retention. High-value customers in Cluster 1 have more frequent and recent purchases of higher monetary value. These insights help the marketing team create targeted CRM marketing campaigns, regular low-cost promotions and great customer service to retain the high-value loyal customers as brand ambassadors.

Mid-value customers in Cluster 3 have had relatively infrequent but recent purchases of lower monetary value, which suggests that this segment should be targeted with bigger promotional offers and discounts to increase the conversion rate. Low-value customers seem to be the churned customers with purchases made a while ago, which were less frequent and of lower value. Marketing campaigns should be designed to win them back. That said, spending marketing dollars and resources on this segment might not provide a high return on investment (ROI).[7]

CUSTOMER LIFETIME VALUE

The next phase of optimising the marketing strategy with customer segmentation is using a lifetime value (LTV) model to estimate the net present value of a customer's contributions to future profits over a certain time period. The basis of marketing strategies is to invest in the customers (with promotions, ads, discounts, etc) to create long-term revenue gains. Hence, the historical LTV calculation of *Total Revenue – Total Cost* for each customer becomes an important metric in driving and optimising marketing strategy. LTV calculation has a multitude of use cases, while providing three primary benefits:

- business intelligence about the current value of its existing users;
- a valuable metric in selecting winning variants from A/B tests; and
- intelligent trade-offs with respect to cost-spending treatments such as offering a discount or targeting specific segments when paired with payout prediction.

While customer segmentation helps identify the high-value cluster of customers, customer LTV estimation aims to project or forecast customer values by cluster. This lends insights for marketers to make the trade-off between marketing expenditure

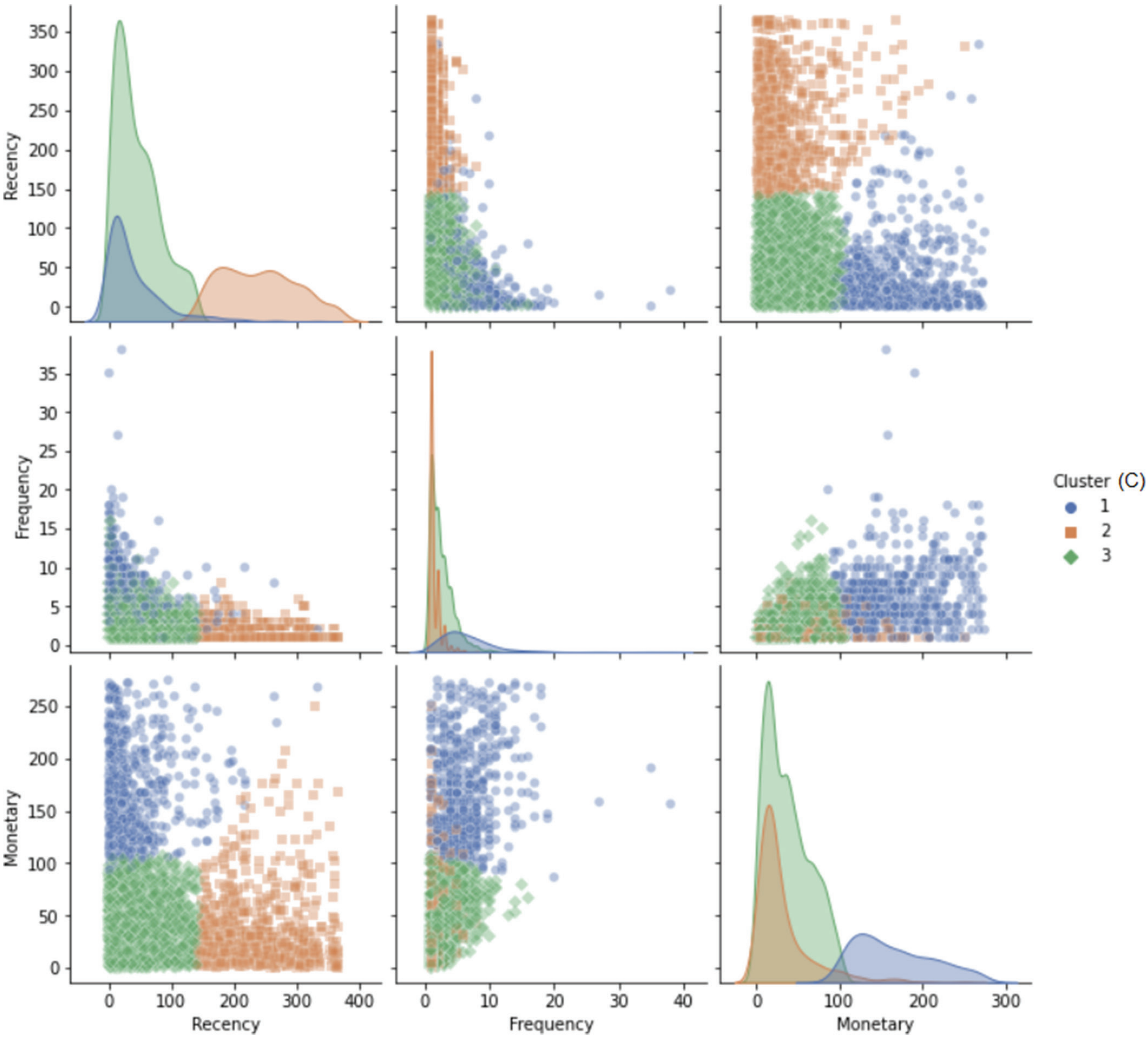

Figure 2: Optimise marketing strategy by customer segments
High-value: C1; Mid-value: Cr2; Low-value: C3

and expected returns by each cluster. The goal is to target a user base with a high intent to become long-term customers, or to drive product changes to convert those with low LTV into high LTV while balancing the costs of retaining and acquiring customers. There are always customers that are less valuable in terms of net returns.

The key to targeting the high-intent users is to recognise behaviours and patterns of the users and segment them accordingly (based on frequency of product use, region, the device used, specifics of the type of orders, etc). Based on these sub-segments, an LTV model can be built that uses multiple user dimensions to predict each customer's expected revenue. The steps in customer LTV calculation[8] are as follows:

1. *Determine the most relevant time period.* 'Lifetime value' is a misnomer as lifetime value is usually calculated for a discrete time period that might be anything from three months to five years. The longer the time period, the less accurate the prediction, because historical data going back five years will not be available for most products. The time period of choice is dependent on many factors, including the business model, industry, frequency of product usage by the customer, etc. Oftentimes, many models are built for different time periods and each of these values becomes important for decision making.
2. *Define features that are indicative of high-intent customers in the future.* This is where segmentation becomes important. The more granular the sub-segments, the more accurate the predictions. However, this also means more intense ML modelling training conditional on the availability of data and research resources. More granular sub-segments also imply more customised campaign strategies with fewer customers in each segment — often requiring more campaign investment. For this reason, a balance must be sought between the accuracy of revenue prediction, the ML modelling training efforts and campaign investment. In the above section, three clusters are defined based on their transaction behaviours.
3. *Train an ML model to calculate the LTV for each segment of customers.* This model can be as simple as linear regression, where the features and prior data make up the coefficient values, or calculating survival curves, where the value over time for each sub-segment is plotted and used for future calculations of LTV.
4. *Use the LTV model to make predictions for future campaigns.*
5. *Leverage the model to calculate the LTV for each segment of customers who interface with the product.*
6. *Tailor the marketing strategy to target customers based on the LTV and marketing costs for each segment.* Campaign strategies can also be made to convert customers from low LTV to high.

A/B TESTING

With the insights generated from customer segmentation and LTV estimation, the next phase is to plan a marketing strategy to optimise the business metrics. A/B (aka randomised control) testing plays a critical role in experimenting with new ideas and enabling product owners to make unbiased data-driven decisions and move forward rapidly. In an A/B experiment, customers are randomly assigned to two or more variants of the testing subject, and statistical analysis is used to determine which variation performs better for the predefined goal.

While the key to a successful future in digital marketing relies on a 'growth mindset' of frequent A/B testing to iterate quickly, fail quickly and pivot quickly, wrongly executed experiments that violate assumptions and misinterpret A/B test results yield misguided decisions that hurt company goals.

This section discusses assumptions, methodology, requirements and caveats for the design and execution of successful online experiments to better target the most valued customer segments.[9]

In the previous sections of this paper, customer segmentation modelling is used to associate customer purchase behaviours with customer clusters of high to low monetary value, and the customer LTV estimation helps calculate the three months to five years LTV of each cluster. The marketer can therefore customise the campaign for each cluster to maximise the business opportunities.

One example might be to design a customer relationship management (CRM) marketing campaign based on sending e-mails to high-value loyal customers in cluster 1 to increase brand loyalty and retention, while sending promotional offers or discount coupons to low-value customers in clusters 2 and 3 to increase conversion and decrease churn. A/B testing should be the gold standard to measure the effectiveness of these marketing strategies.

Once the decision regarding a marketing strategy has been made, A/B testing[10] can be conducted according to the steps outlined below.

Define the experiment

To run a successful experiment, the first step for the product (or marketing) team is to clarify and commit to the following definitions.

Define the metrics

First, marketing professionals need to define the goal metrics that the experiment is testing. The goal metric should be specific and quantifiable from the available customer database. Most importantly, it should capture the ultimate success the company is striving for. This is usually a single metric or a few metrics. Oftentimes, business revenue is the top-line goal, but different companies might have key metrics that are indicators of success/progress towards revenues. For instance, in the early stages of a company, growth is the highest priority, so the number of retained users will be the general business goal. However, in the later stages of the company's life cycle, long-term business revenues can be the topmost priority. In such cases, both the expected revenues from each user and the number of retained/active users will be the priority.

Clarifying the business goal and the associated north-star metric with the product owner and decision-maker is the most critical part of a successful A/B testing exercise. In the example of online retail used throughout this paper, customer retention is the business priority. Consequently, the percentage of existing customers purchasing within 30 days of the marketing campaign is chosen as the goal or north-star metric. The KPI is calculated as follows:

> no. unique customers purchasing within 30 days of the campaign / no. unique existing customers receiving the campaign.

Together with the goal metric, a few monitoring/guardrail metrics should also be set up to ensure that the company moves toward business success without violating other business constraints. For instance, marketing professionals should monitor the amount of money each purchased customer spends to monitor whether the coupon campaign nudges purchase on the coupon face-value or beyond.

Define the treatment

Marketing analytics professionals should define the treatment of the experiment or the new marketing campaign to be tested in addition to the old/control marketing strategy. For this example, the control strategy is no segment customised marketing campaign, while the new strategy is to distribute campaign e-mails customised to the customer clusters/segments. Note that the example here entails two-variant testing, with one control group and one testing group. A/B testing can also be extended to test multiple groups.

Define the hypothesis to be tested

The hypothesis of the statistical testing is: by introducing the new campaign strategy, the KPI will increase. Additionally, the team needs to further clarify with the decision-makers the minimum lift in KPI the new campaign is expected to target so that a decision criterion is committed before observing any results. The minimum lift should be decided based on the baseline

KPI and, more importantly, the product owners' judgments about the scale of lift that is worth launching the new marketing campaign.

Once the hypothesis and minimum lift are defined, the decision whether the marketing strategy should be scaled depends entirely on whether or not the difference in KPI between the treatment and control is statistically greater than the minimum lift defined.

Calculating sample size

The required sample size for the experiment is calculated through power analysis and the associated run time for the experiment is estimated based on this. The mean and standard error of the KPI, the minimum required lift, the statistical testing method of choice (eg one-sided versus two-sided testing, t-testing versus F-testing) and the associated statistical criteria (ie type I and type II errors) will all affect the required sample size. Furthermore, if the experiment entails testing multiple variants, the required minimum sample size will also be significantly increased. Note that the run time and the decision criteria should be clearly defined before starting to run the experiment and should not be adjusted in the middle of an experiment in the hope of obtaining any desired statistical testing results.

Experiments should generally run for at least a week or two in order to capture the day-of-week effect or seasonality pattern of customer behaviour. In circumstances where the required sample size is likely to be obtained too quickly (eg with large companies with considerable online traffic), the scope of experiment should be limited to a portion of the company's online traffic only. The running time of the experiment will also need to be sufficiently long to control the novelty effect, yet short enough to avoid confusing the treatment effect with the long-term trend.

Last, but not least, although business success is the ultimate goal of any marketing campaign, it should never be achieved at the cost of compromising ethical/moral standards or providing a bad customer experience. With this in mind, marketing professionals should always ask whether there are any emotional or financial risks for the customers participating in the experiment; likewise, the marketers conducting the experiment must also be mindful of any risks to the customer experience and whether these risks would be compensated by the expected improvements to the user experience in the long run. If there are any risks or concerns, marketing professionals should ensure that customers will be notified and have the option to opt out of the experiment. This safeguard must be in place before launching any experiment.

Run the experiment

To start an experiment, marketers need to set up the experiment distribution. This requires collaboration with the engineering team (in this case, the e-mail distribution team) to clarify the experiment distribution population, randomisation unit, the randomisation approach and experiment version tracking strategy. The experiment should be distributed among the population that the market is targeting, or the cohort that testing should start with. The distribution unit should be consistent with the goal metric's denominator. In the case of online retail (Table 1), the unique CustomerID will be the distribution unit. The experiment distribution should be randomised.

Depending on the scales of customer volumes and minimum required sample size, a decision will have to be made whether the experiment should include a complete or partial population.

The selected experiment traffic will be randomly split into the control and treatment versions in a 1:1 ratio. Depending on the company's choices, either external campaign management tools or in-house engineering teams will play critical roles in campaign distribution to guarantee that the campaign ID is correctly firing for each customer and customers receive consistent campaigns over the experimental period. Last, but not least, the A/B version tracking should be defined and the associated database tested to make sure the data will be sent back correctly in order to enable the statistical testing.

Once the experiment has been officially launched, it is important to conduct daily base monitoring on the splitting of the samples to ensure the randomisation and 1:1 split distribution are successful and working as expected. Likewise, the goal and guardrail metrics will need to be monitored frequently to avoid any large and unexpected negative impact on the business or customer experience.

Experiment analysis and decision-making

With successful experimental design and setup, the experiment analysis should be straightforward. A sanity check or A/A testing that statistically tests the experiment's irrelevant features is often a good start to verify that the experiment randomisation is successful. Although the sanity check is not always necessary, it is recommended the first few times a company runs A/B tests.

Following the predefined statistical testing method, the hypothesis that the new campaign strategy will increase KPI by the minimum lift is tested. The statistical testing result is either rejected or fails to be rejected at a predefined statistical level, such as 5 per cent significance or a Type I error in most cases.

If the hypothesis is rejected with statistical significance, it is reasonable to conclude that the new campaign strategy will increase the KPI to some extent. This testing will support the decision to scale the campaign to full traffic. Meanwhile, based on the above LTV estimation, marketing professionals can estimate the incremental financial returns that may be attributed to the campaign. Of course, even when the statistical testing supports the new campaign, it is still worth double-checking that any novelty effects, promotion effects or seasonal effects that might have affected the results of the experiment have been accounted for.

If, on the other hand, the hypothesis cannot be statistically rejected, the decision will be not to scale the marketing campaign. This does not mean that the experiment has not been useful; indeed, the results will almost certainly provide plenty of valuable insights. For instance, there might be significant differences in the KPI change among the subgroups of the sample, indicating that the campaign has greater impact among certain demographics (eg younger female customers). Consequently, marketers should consider running further experiments to explore their findings in greater depth.

Besides testing the goal metric, marketing professionals should also check the other guardrail metrics' response to the campaign to identify whether there are any negative impacts or opportunity costs of scaling the campaign (eg product cannibalisation).

SUMMARY

Used in concert, customer segmentation, LTV calculation and A/B testing can obtain the data necessary to create customised marketing strategies to target the most valuable customer segments in order to maximise market opportunities and revenue. Furthermore, by conducting statistical analysis with these predictive models, it is possible to quantify marketing impacts and ROI, to motivate decision-making and assist in channel attribution.[11]

References

1. Ma, L. and Sun, B. (2020) 'Machine learning and AI in marketing — Connecting computing power to human insights', *International Journal of Research in Marketing*, Vol. 37, No. 3, pp. 481–504.

2. Blanchard, T., Bhatnagar, P. and Behera, D. (2019) 'Data Science for Marketing Analytics: Achieve your marketing goals with the data analytics power of Python', S.l.: Packt Publishing Limited, Birmingham.
3. UCI Machine Learning Repository (2015) 'Online Retail Data Set', available at: http://archive.ics.uci.edu/ml/datasets/online+retail (accessed 31st July, 2021).
4. McCarty, J.A. and Hastak, M. (2007) 'Segmentation approaches to data mining: A comparison of RFM, CHAID, and logistic regression', *Journal of Business Research*, Vol. 60, No. 6, pp. 656–662.
5. Miller, T.W. (2015) 'Marketing Data Science: Modeling Techniques in Predictive Analytics with R and Python', Pearson Education, Inc, Upper Saddle River, NJ.
6. Hwang, Y. H. (2019) 'Hands-on Data Science For Marketing: Improve your marketing strategies with machine learning … using python and r', S.l.: Packt Publishing Limited, Birmingham.
7. Griva, A., Bardaki, C., Pramatari, K. and Papakiriakopoulos, D. (2018) 'Retail business analytics: customer visit segmentation using market basket data', *Expert Systems with Applications,* Vol. 100, pp. 1–16.
8. Hong, T. and Kim, E. (2011) 'Segmenting customers in online stores based on factors that affect the customer's intention to purchase', *Expert Systems with Applications*, Vol. 39, No. 2, pp. 2127–2131.
9. Webber, R. (2013) 'The evolution of direct, data and digital marketing', *Journal of Direct, Data and Digital Marketing Practice*, Vol. 14, pp. 291–309.
10. Kohavi, R., Tang, D. and Xu, Y. (2020) 'Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing', Cambridge University Press, Cambridge.