BSc. Computer Science & Economics

A Machine Learning and XAI approach for creating personalized retention strategies in subscription-based businesses.

Maartje Vennema

First supervisor: Dr. Guus Ramackers
Second supervisor: Dr. Jan N. van Rijn
Company supervisor: Wesley van Ham

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl
10/07/2022

Abstract

The company sells men's boxer shorts on a subscription basis, a new digital business model that has disrupted traditional ways of doing business. This study provides a data-driven method for defining member segments, and aims to predict profitability and loyalty in order to develop personalised marketing strategies. In particular, this research focuses on how Dutch subscribers to the company can be classified based on member value. The classification method developed will be used to discover member characteristics and behaviour in order to assign personalised retention strategies to current members.

The existing RFM marketing model that segments customers into subsets based on their indicators has been adapted to the subscription business model resulting in a FB-model. This FB-model classifies subscribers based on profitability and loyalty. Our research shows that the LightGBM classifier performs best for this classification machine learning problem with an F1-score between 0.53 and 0.61 per segment for a balanced dataset. By subsequently applying explainable artificial intelligence techniques, this research concludes that engagement strategies are essential member characteristics of subscribers. Additional features that impact classification are buying differentiated 'add-on' products and payment characteristics. Primarily payments that are charged back, payments paid after the deadline, as well as product returns significantly decrease the chance of a customer becoming a long-term 'premiere' customer. In addition, our research shows the importance of referral programs.

Based on the insights provided by our research, the company hosted a company presentation and discussion on how to implement the results and recommendations into their marketing strategies and design a basis for personalised customer retention strategies. This method will subsequently be applied to customers based in the growth market of Great Britain, as a springboard to derive the characteristics and behaviour of members in other European countries.

Contents

1 Introduction **1**
1.1 Context . . . 1
1.2 Problem statement . . . 1
1.3 Goal of this research . . . 2
1.3.1 Research question . . . 3
1.3.2 Hypothesis . . . 3
1.3.3 Deliverables . . . 4
1.4 Overview of this research . . . 4

2 Background **5**
2.1 Background knowledge . . . 5
2.1.1 Customer retention . . . 5
2.1.2 RFM-model . . . 5
2.1.3 Interpretability & explainability . . . 5
2.1.4 Class imbalance . . . 5
2.1.5 Cross-validation . . . 5
2.1.6 Boosting classifiers . . . 6
2.1.7 Automated machine learning & hyperparameter tuning . . . 6
2.2 Related work in customer retention . . . 6
2.2.1 Contribution . . . 7

3 Method **8**
3.1 Data . . . 8
3.2 RFM-model . . . 8
3.2.1 Adapting the RFM-model to subscription-based businesses . . . 9
3.3 Labeling . . . 10
3.4 Features . . . 11
3.5 Algorithms . . . 12
3.5.1 Class imbalance . . . 12
3.5.2 Multiclass classification . . . 13
3.5.3 Cross-Validation . . . 13
3.6 Reviewing models . . . 13
3.7 Interpretability . . . 14

4 Results **16**
4.1 Target . . . 16
4.2 Multiclass classification . . . 16
4.2.1 Overview of algorithms . . . 16
4.2.2 Selection of algorithms — zooming in on boosting classifiers . . . 17
4.3 Automated machine learning . . . 18
4.4 Binary classification . . . 19
4.5 Different methods to optimize performance . . . 20
4.5.1 Redefining segments . . . 20

4.5.2 Time-sensitive features . . . 21
4.6 Interpretability . . . 21
4.6.1 Most important features for classification . . . 22
4.6.2 Most significant features regarding the business strategies . . . 22
4.6.3 Secondary results . . . 30
4.6.4 Summary of most important interpretable results . . . 31

5 Recommendations **32**
5.1 Strategy . . . 32

6 Limitations **34**
6.1 Critical reflection . . . 34
6.1.1 Data . . . 34
6.1.2 Algorithm . . . 34
6.2 Future work . . . 35

7 Conclusions **36**

References **42**

A Model usage **43**

B Features **44**
B.1 Features with description and type . . . 44
B.2 Feature transformation — a description of the values . . . 46

C Summary plots **53**
C.1 Summary plot of SHAP values for Sleeping segment . . . 53
C.2 Summary plot of SHAP values for Potential segment . . . 54
C.3 Summary plot of SHAP values for Premiere segment . . . 55

D Dependence plots **56**

Acknowledgements

I wish to acknowledge the supervision provided by Dr. Guus Ramackers and Dr. Jan N. van Rijn of LIACS and Wesley van Ham on behalf of the company. My deep appreciation for Dr. Guus Ramackers for guiding me through the process that started with an idea and resulted in a thesis. You set a high bar and it made me move boundaries.
I would like to thank Dr. Jan N. van Rijn for the time he has given to supervise this research, although your time was the reason for withholding me in the first place. Providing me with your feedback and expertise raised the quality of this thesis.
Lastly, I am grateful for the chance the company gave me to realize my ambition to implement data science in subscription-based businesses for marketing resources. It was a pleasure working with all my colleagues, and in particular with Wesley van Ham.

Maartje Vennema

Delft, July 2022

1 Introduction

1.1 Context

One of the new digital business models that have impacted consumer focused business as well as business-to-business for the last decade is the subscription-based business model [1]. The Cambridge Business English Dictionary states that *subscription-based* is used to describe a website or television station that customers pay to use [2]. But as the industry has rapidly changed, so did the definition of subscription-based. More recent definitions of subscription-based state that this business model can be characterised by selling a service or selling a product as a service 'to receive monthly or yearly recurring subscription revenue' [3]. In the fourth quarter of 2020, subscription-based businesses demonstrated revenue growth of 21% while the revenues of Standard and Poor's 500 (S&P 500) companies reached revenue growth of 3% [4]. The origin of this trend can be explained by three factors.

First is the investment for buyers that play a role in the early adaption of subscriptions. When computers and software were introduced, they were exclusive to large enterprises due to their high cost. By offering software as a subscription-based service (SaaS), companies were selling the use of their product hosted by a third party instead of running software locally on in-house servers. This resulted in more affordable services and broader adaptation since not only large enterprises were able to pay for the product [5].

The second and more recent cause is the continuous digitalisation of organizations, which has been accelerated by the COVID-19 pandamic [6]. Subscription-based businesses are the product of digitalisation. For companies to stay competitive, their business models must fit the market.

The last trend can be defined by the evolution of marketing strategies including the shift from product-based industries to service-based industries. Subscription-based business are primarily operating as service-based businesses. Selling the product as a service touches customers' operations and feelings in which the product is being used. The strategy of operating service-based is the product of the rise of digital marketing and the evaluation of the importance of customer behaviour, as understanding and predicting customer behaviour is key to build customer loyalty and advocacy [7]. It can be created by stimulating engagement through personalisation. With such engagement strategies, a long-term relationship with the customer will be created where the company can profit in terms of recurring revenue [8]. A personalised experience is equally as important as convenience for commodity items according to McKinsey. It is the most important reason to continue the subscription for beauty and apparel items. [9].

1.2 Problem statement

All newly acquired customers are the result of marketing and sales efforts. The acquisition costs can be computed by dividing the total costs of marketing and sales divided by the number of customers acquired. The strategy of acquiring new customers to relatively high acquisition costs is common for starts- and scale-ups, as gaining market share is one of the main goals in this stage. Subsequently, in case of business success, these businesses will eventually reach their maximum marginal returns on acquisition costs [10]. In addition to the assertion that states the importance of retention strategies, according to *Reinartz W et al* (2005)., underspending on retention by 25%(with the assumption that other factors are optimal) can decrease the return on investment (ROI) with $-55.29. In contrast, underspending on acquisition costs by 25% can lower the ROI with $-3.031, with the assumption that other factors are optimal [11]. Thus, the focus needs to be distributed between acquiring new customers without losing sight of retaining existing customers. Retention strategies have been established to retain existing customers. Those strategies can be drafted ad-hoc, but would benefit from a data-driven approach.

The company is a Dutch scale-up, founded in 2016, that sells boxer shorts on subscription. The company reported a revenue growth of 361% between 2018 and 2020. By 2021, the annual revenue has reported to be up to 35 million euros [12] [13]. As of now, 350.000 active members receive a new pair of boxer shorts every month and the company aims to reach 1 million active members in three to five years. The company' sales area is expanding rapidly by serving nine countries at the beginning of 2022 and they plan on expanding to four more countries within a two years.
The company calls their customers members as the word customer implies a negative relationship [14]. Members should not be confused with users. The user owns the account, but can have multiple memberships. A membership can have only one product, whereas a user can have multiple memberships and so multiple product subscriptions.

The problem we adress is that the focus of the company is on acquisition and that the current retention strategy of the company focuses on preventing members to churn with non-personal strategies. This non-personal strategy comes down to offering the member something extra when the member is planning on terminating his subscription, a so-called *please stay* offer. Because the core value of service-based businesses is customer-centricc[7], The company must have a clear sight of the behaviour and needs of its members. This implies that all customer contact needs to be equal to the customers' needs, including the retention strategy. Moreover, it must need to find an optimal balance between aqcuisition and retention [10].

In order to lower the churn by offering personalised retention strategies, the company should gain insight into their member's behaviour and needs. By optimizing the retention strategy based on this behaviour and needs focusses on keeping existing members and retargeting hot customers, instead of only targeting new (warm or cold) customers. It will offer a service closer to the needs of the member. As every member has unique characteristics and behaviour, the best strategy varies for every person. The company can prevent members from churning by identifying customer segments based on member behaviour and needs with data-driven methods.

1.3 Goal of this research

The proposed solution direction for data-driven personalised retention strategies is machine learning based. Developing a machine learning model contributes to segmenting members by profitability and loyalty. Besides the segmentation by machine learning, explainable artificial intelligence (XAI) helps making data-driven decisions based on feature value impact whereas the company is now using gut feeling with regard to the the impact of feature values for decision making. Every member behaves differently according to their needs and personal information. To retain members, the impact of feature values derived by XAI should be used to formulate retention strategies. Since those strategies will be based on member characteristics, the strategies are customer-centric and increase profitability and loyalty.

The majority of methods used in XAI research focus on explainability of the algorithm instead of interpretability for the intended users [15]. This research will especially use XAI to produce interpretable results for the marketing team of the company.

1.3.1 Research question

This research aims to classify Dutch members of the company into segments based on profitability and loyalty in order to define data-driven and personalised retention strategies. The number of members and possible differences in behaviour and operational systems differs per country. Therefore, the scoping of this research is limited to Dutch members only.

The research question derived from this goal is:

"How can Dutch members of the company be classified based on the member value, and discover corresponding member characteristics and behaviour to assign personalised retention strategies to current members?"

This research question can be divided into the following sub-questions:

- In which segments can members be allocated based on their profitability and loyalty and how can profitability and loyalty be measured best?
- How can segments of current members be predicted?
- What are the characteristics of the least valuable members?

1.3.2 Hypothesis

The hypothesis is that segments can be created by using a RFM-model that is recreated according to the subscription-business. Using machine learning and XAI, a model is created which predicts the segment of a member based on personal member information and member behaviour. Using XAI, features that represent the purchase of 'add-on' products are derived as key to classify members and will be most used in order to create personalised retention strategies.

- **In which segments can members be allocated members based on their profitability and loyalty and how can profitability and loyalty be measured best?**
  Classification of members proceeds by assigning members to categories based on RFM segments which are the target label. With frequency, the profitability of members can be captured, as the frequency is in proportion to the revenue. By summing up the number of breaks in months between orders, member loyalty is derived. Loyal members return and stick to the brand, and will therefore not have many breaks between orders.

- **How can segments of current members be predicted?**
  Members can be allocated to segments by creating a machine learning model that predicts the segment of a member. By training and testing the model based on churned members, the model is able to predict segments for current members. Since comparing low valuable members with high valuable members is important, multiple classes have to be used as target and evaluation must be done per category. High valuable members are expected to show more engagement than low valuable members. The expectation is that the error of the prediction of high valuable members is lower than the error of the prediction of low and average valuable members.

  As not every causality for member value is captured in the model, significant scores on performance metrics are scarcer than the power of extracting important features and feature values out of the model. Implication of this expectation is that the company benefits less from classifying current members into segments due to reliability.

- **What are the characteristics of the least valuable members which characteristics can be used to draft retention strategies?**
  With an explainable model, insights into customer characteristics and behaviour about their value can be drawn. As literature showed that engagement has a positive impact on customer value [8], the hypothesis is that least valuable customers are not heavily engaged with the company. A retention strategy would therefore benefit from the use of engagement strategies.

1.3.3 Deliverables

The result consist of two deliverables. The first deliverable is the machine learning model that the company can employ to predict the segment of a member and user instructions (Appendix A). The second deliverable is the recommendation (Section 5) regarding the feature values that contribute to the allocation of the least valuable member segment.

1.4 Overview of this research

The thesis has been split up into seven sections. The introduction has been discussed. Sector 2 provides a background on the topic, Sector 3 follows with a description of the methods, Sector 4 provides the results of the thesis and in Sector 5 the recommendations are shared. The limitations are considered in Section 6. The last sector, Sector 7, closes with conclusions.

This bachelor thesis has been created under the supervision of assistant professors at the Leiden Institute of Advanced Computer Science, dr. Guus Ramackers, and dr. Jan N. van Rijn.

2 Background

2.1 Background knowledge

2.1.1 Customer retention

One of the key performance indicators of a subscription company is the retention rate. Retention describes the preservation of customers. It is part of relationship marketing which is described as "attracting, maintaining and enhancing customer relationships" [16].

2.1.2 RFM-model

A major challenge in marketing is targetting customers. A method to segment customers is using the RFM-model. The RFM-model defines the profitability of customers based on recency, frequency, and monetary amount. Table 1 explains the metrics.

Recency	Period in months since the last purchase
Frequency	Number of prior purchases
Monetary amount	Total amount of money spent on purchases

Table 1: Definition of recency, frequency and monetary amount indicators [17]

2.1.3 Interpretability & explainability

Interpretability and explainability are two factors that have taken on greater importance over the last few years. Although both are commonly used interchangeably, both have different perspectives on machine learning. Interpretability is the degree to which a human can consistently predict the model's result [18]. On the other hand, explainability describes the processes of the algorithms that have come to a prediction. Both have the purpose to be explained in human terms [19]. Both concepts are comprehended in XAI [15].

2.1.4 Class imbalance

Class imbalance is defined by the event that one class is represented by a large number of examples while the other class is represented by a few. The threat of using an imbalanced dataset for classification is that the minority class is often the one with the most interest. Some algorithms can not perform well on an imbalanced dataset. The minority class can be easily ignored, seen as outlier or rare occurrence. This results to misclassification of the minority class compared to the majority class [20] [21].

2.1.5 Cross-validation

The problem of testing a model on the test set is that the test set may not be a good representation of the entire dataset. Cross-validation is a method for generalizing the performance of the model. A different validation set and test set is created by reviewing in k iterations. In typical cross-validation, the training and validation sets must cross over in successive rounds such that each data point has a chance of being validated against[22]. The default number of folds is ten. With a total of ten folds, in every iteration, one-tenth of the training set is being used for validation while the other nine folds are being used for training [23].

2.1.6 Boosting classifiers

Boosting refers to a set of weak machine learning classifiers that are assembled to form a strong classifier. A method is called weak if the learner can produce a hypothesis that performs slightly better than random guessing [24]. The top 5 most used boosting classifiers are: Gradient Boosting Classifier, XGBoost, LightGBM, CatBoost, and ADABoost.

2.1.7 Automated machine learning & hyperparameter tuning

Hyperparameter tuning is adjusting the hyperparameters of an algorithm to improve the learning process. These hyperparameters vary for every algorithm. The difference between parameters and hyperparameters is that parameters are learned and hyperparameters are set.
Setting these hyperparameters takes human effort. With automated machine learning (AutoML), algorithms are used to define the optimal set of hyperparameters to reduce human effort. *FLAML* [25] is a fast and lightweight library that optimizes for low computational resources in finding accurate models and is used in this research to tune hyperparameters. It performs by balancing search cost and error in the search space and is therefore useful in tasks with low computational space [26].

2.2 Related work in customer retention

With a subscription-based model, companies offer home-delivered goods or (online) streaming services. The strategy of selling a subscription is not new as Charles Dickens already attended charity dinners that were used to raise money by subscription lists [27]. Besides that, telecommunication providers have been using this business model for decades. This is reflected in research about churning customers in subscription-based businesses. In the preparation of this report, three different studies were evaluated.
Ezenkwu et. all. [28] have applied a k-means algorithm that clusters customers based on big data. This method aims to segregate customers into unnamed groups to show similar market behaviour or characteristics. *Blank & Hermansson* [29] have used machine learning to predict the chance of a customer to churn. This regression technique provides a percentage but does not tell why the customer is tagged with that percentage. Other research focuses on customer churn prediction, use classification algorithms to predict whether or not a customer is churning [30]. This method is mostly used in customer retention research. However, similar research is still in its infancy.

Research in the context of customer segmentation has already matured over the past couple of years. Researchers have found different approaches for segmentation and developed many hands-on tools to acquire the benefit. These tools include customer lifetime analysis, recency, frequency, and monetary (RFM) analysis, behavioural and demographic analysis [31].

More recent studies discuss customer preferences for explanations of algorithmic decisions. It discovered that despite the popularity of explainable models, counterfactual explanations are not popular among users, unless they follow a negative outcome [32]. Moreover, the majority of methods used in XAI research focus on explainability of the algorithm instead of interpretability for the intended users [15]. While the very first AI systems were easily interpretable, the last years have witnessed the rise of opaque decision systems such as Deep Neural Networks which are considered as black-box models [33].

SHAP is one of the many methods to include XAI into research. It is used to explain the prediction of an instance by computing the contribution of each feature to the prediction [34].

2.2.1 Contribution

This research contributes by adapting the RFM-model to subscription-based businesses, resulting in a FB-model. Moreover, the FB-model is used to classify subscribers based on profitability and loyalty to create retention strategies by machine learning and XAI methods.

3 Method

3.1 Data

The data that the company collects is dynamic, meaning the data is updated every minute. Results may therefore differ for every period. The dataset has monthly order as rows. A member can have multiple orders and therefore multiple rows. With 200 new Dutch members a day, more than 200 orders are added to the database as rows every day. Combined with 150.000 active Dutch subscribers, more than 155.000 new rows are added to the database monthly. To enhance reliability, the dataset consists of all orders before the sixteenth of June, 2022, where the member is churned at the moment of running the script. The member status indicates if a member is active or has churned. It displays the current status and does not keep track of historic member status. If the algorithm used for training and testing is used after the sixteenth of June, 2022 there is a chance that a member has churned between the sixteenth of June, 2022 and time of running the script. If so, the dataset at the time of running is different than the dataset at the sixteenth of June, 2022. Database architecture constraints cause this implication. To reduce the overflow of members, the algorithm and results are all generated between the sixteenth of June, 2022 and the seventeenth of June, 2022. Between this time interval, the dataset listed a growth of 0,22%.

Scoring for members based on the RFM-model will be improperly low for new, maybe valuable members since those members are subscribed not long enough to be valuable. The dataset will therefore enclose only churned members. The churned members have an established score, whereas the score of current members can develop over time. With this approach, low-scoring members in the present but high-scoring members in the future will not be confused with high-scoring members in the present.

With the gradual expansion of the sales area throughout Europe, it has become apparent that customer behaviour differs per country. Considering the potential volatility of behaviour and member characteristics considering customer value, the dataset excludes all countries but the Netherlands.

The dataset consists of paid memberships only. A new registered user always tries the first boxer short for free. They have a ten-day trial period in which the boxer short can be tested. If unsatisfied, members need to cancel their subscription within this trial period. If not, members are automatically signed up for the paid membership and payment is debited monthly.

Finally, the data is labeled in Section 3.3, as the pre-defined segments serve as target. This marks the experiment as supervised learning [35], more specific as classification. The total number of members represented in the dataset is approximately 300.000.

3.2 RFM-model

Using classification states that the members the company can be segmented in different ways based on value. A method for classifying customers based on profitability is the RFM-model, see Table 1.

The monetary amount does not add value to the model and is left out as is it proportional to the frequency in the subscription business. This statement however can be put into perspective since the company sells extra products *(cross-selling)*, like swimming trunks and special edition boxer shorts as limited editions. Because this research specifically focuses on revenue due to subscriptions, the limited editions, so-called *specials*, are used as features instead of as part of the target.

Besides monetary amount, recency is also eliminated from the model due to the decision to only take churned members into account. To give a member a permanent segment, the assigned segment is not supposed to change over time. When recency is considered, a recent churned member will have a higher score for recency than a member that has churned years ago. However, the recency of the just churned member will lower with time since the recency goes up with time. The second reason to eliminate recency is that some features may be correlated with time. To illustrate the implication, referral programs have only been used for a period of time, just like the usage of polls where members can vote for the new design. An algorithm can incorrectly recognize the patterns between the recency and the promotion codes based on the time instead of the correlation between the intrinsic value of the poll or referral features and the value of the customer. Table 2 shows the correlation between recency and referral features compared to frequency and breaks. The definition of the referral features can be found in Appendix B.

	recency	**frequency**	**breaks**
friend	-0.11	0.09	0.06
invitation_sent	-0.22	0.18	-0.01

Table 2: Correlation of time-sensitive features with recency

3.2.1 Adapting the RFM-model to subscription-based businesses

To apply the RFM-model to the subscription business, the indicators have been recreated according to the industry. Table 3 shows the opposed model suited for subscription-based businesses, the FB-model. The frequency is the total number of boxer shorts orders. The break indicator is defined as the number of paused months plus the number of months wherein the payment has not been fulfilled plus the time in months wherein members are churned and subscribed again later on. The FB-model is used for labeling the dataset in Sector 3.3.

Frequency	Number of orders (free trial + paid)
Breaks	Number of paused, open payment months and time between not ordered

Table 3: FB-model

3.3 Labeling

The FB-indicators have been added to the dataset by computing the values for frequency and break score per member. The value of the indicators can be statistically divided into quartiles to assign a score. Table 4 shows the ranges of these quartiles. Every member can be given a score for frequency and breaks separately that corresponds with the quartile of the indicator [36]. The indicator generates a maximum score of two as the quartiles of the breaks are not unique, see Table 4. Ultimately, the maximum score that can be given to a member comes down to six and the minimum score can not be lower than two.

	frequency	**breaks**
mean	20	1
std	13.6	0.8
min	1	0
25%	10	0
50%	20	1
75%	28	1
max	150	32

Table 4: Characteristics frequency and breaks indicators

The score given to members describes the member profitability in a range of two to six. Since the difference between these four groups is not significant enough for the company, the members are labeled into fewer categories than the number of different scores a member can have. The labels given to members are used as target for supervised learning. This research focuses primarily on detecting non valuable customers and secondarily on high profitable customers, therefore these target groups need to be isolated. Members with a score of 5 and 6 are labeled as *Premiere*, scores of 3 and 4 are labeled as *Potential*, and scores of 2 are labeled as *Sleeping*. An alternative method for allocating scores to segments is discussed in Sector 4.5.1.
The names of the categories are in correspondence with three out of seven of the original RFM-model [17]. Characteristics of these categories are displayed in Table 6 and Table 5. The three segments are:

1. *Premiere*: Most valuable customer segment
2. *Potential*: Customers that have the ability to become valuable members
3. *Sleeping*: Specifies the least valuable members

Frequency	**Minimum**	**1st quartile**	**Mean**	**3rd quartile**	**Maximum**	**Group size**
Premiere	18	21	28	33	150	39.622
Potential	1	1	5	8	28	249.812
Sleeping	2	3	5	7	10	18.510

Table 5: Characteristics of frequency per segment

Breaks	**Minimum**	**1st quartile**	**Mean**	**3rd quartile**	**Maximum**	**Group size**
Premiere	0	0	1	1	2	39.622
Potential	0	1	1	1	6	249.812
Sleeping	2	2	4	4	32	18.510

Table 6: Characteristics of number of breaks per segment

3.4 Features

The features used for this research are:

- product_name
- postal_code
- please_stay
- special
- payment_charged_back
- payment_succeed_late
- return
- newsletter
- number_of_members
- user_socks
- payment_refund
- active_payment_contract_past
- is_original_member
- changed_product
- size
- changed_size
- player_email
- poll_vote
- friend
- invitation_sent
- first_order_hour
- first_order_day
- first_order_month
- first_order_year
- first_order_weekday
- promotion_code

Appendix B.1 represents the list of features obtained from the database with a short description. The features are quantitative as well as qualitative, nominal as well as ordinal. With feature engineering, the data has been transformed into informative and compatible features. Features like *changed size*, and *changed product* have been created out of the data by measuring the change in input, in other words, customer behaviour. On the contrary, *product name*, *return*, and *size* have been created by extracting particular rows from the dataset. Feature engineering tools that have been applied include handling missing data and one hot encoding.

Some features have missing data. For example, in 2021, the company created a game in addition to the monthly orders. This online game was available for members as well as potential members.

Since not every member competed in the online game, or was not a (potential) member by that time, the data has missing values. This has been handled by defining whether the member has competed in the online game, which indicates the feature as binary. This method has, amongst others, also been applied to *friend*, *invitation_sent*, *poll_vote* and *player_email* too. An alternative method for handling with those time-sensitive features is discussed in Sector 4.5.2.

One hot encoding is applied to extract payment statuses. Different codes are given for different payment behaviour and those codes correspond with the monthly order it relates to. The features that are created define if the member has once *charged the payment back*, *paid late*, or has a *payment refunded* to extract the unusual categorical values out of the successful payments.

Accordingly, to fit the data in the machine learning model, all features are transformed to integer or binary types. Appendix B.2 displays the features with the values and interpretation of the values.

Finally, because the rows of the dataset are all monthly orders, the dataset needs to be grouped by member. During grouping, some features are be aggregated as mode and some as number of unique values, minimum or maximum value. For features like *size*, *product name*, and *postal code*, the mode is used as the value. For some features that represent behaviour, the number of times that behaviour occurs is captured. Values that are aggregated as maximum or minimum represent a binary indication. After grouping the rows with member information corresponding to the same member, the row is called an instance and represents an unique member.

3.5 Algorithms

The algorithms used for this research are supervised, as the dataset has a clear target and the interest of this research lies in classifying members into pre-defined segments.

The functionality of using regression with value in euros as target is not convenient in the context of the company. As the target will be in a range of 0-1500 euros (maximum frequency, according to Table 5, times the sales price), a difference of 50 euros is not significant for decision making. Besides that, a difference of 50 euros in customer value is not significant as the company is not planning on creating individually personalised retention strategies. Use of regression has therefore been rejected.

The scope of algorithms used for segment prediction is limited to the algorithms included in the open-source software scikit-learn [37].

3.5.1 Class imbalance

Table 6 and Table 5 show the sizes of the segments used for training and testing. The different proportions of those classes are displayed in Table 7. *Sleeping* and *Premiere* members occur in small numbers in the dataset whereas *Potential* members occur in large numbers in the dataset. This phenomenon is defined as *class imbalance*. Possible implications of this occurence are described in Section 2.1.4.

	Group size	Proportion
Premiere	39.622	0.13
Potential	249.812	0.81
Sleeping	18.510	0.06

Table 7: Group sizes and proportions in data set per segment

	Group size	Proportion
Premiere	18.510	0.33
Potential	18.510	0.33
Sleeping	18.510	0.33

Table 8: Group sizes and proportions in data set per segment after majority under-sampling

A solution for this problem is to use a sampling strategy that randomly selects the same number of members for every segment [38]. This under-sampling technique is used to create a training- and testset. The dataset contains equal proportions of classes after under-sampling, as shown in Table 8. The dataset after majority under-sampling is used for model training and testing.

3.5.2 Multiclass classification

The classification of members into subsets is required to provide different retention strategies for each customer segment. As discussed in subsection 3.3, the dataset consists of three different labels. This makes the dataset multiclass [39]. According to *Sánchez-Maroño et. al.* (2010) [40]: 'There are two basic approaches to deal with classifying multiple classes: one is to use classification algorithms that can deal directly with multiple classes, and the alternative is to divide the original problem into several binary classification problems.' The choice of the best method for classification has to be determined with respect to the research question. All considered, multiclass classifiers are used since usability and interpretability are of main importance in this research and binary classification takes an extra dimension of interpretability for non-data scientists.

A member in the dataset can only have one label, meaning a member cannot be located into two categories. Multi-labeled classification does therefore not suit the business case.

3.5.3 Cross-Validation

The performance of the algorithms is based on 10-fold cross-validation. A bias is reduced due to the use of all data as training and test data with cross-validation. The reliability of the performance increases with bias reduction [41].

3.6 Reviewing models

The performance of the algorithm can be captured with the pre-defined scoring method. The folds are averaged by cross-validation to aggregate and average the results [23].
Different scoring methods are suitable for classification problems. Selecting the best-performing metric is key for the viability of the algorithm in usage. The selection of the best-suited performance

metric starts with understanding the computation of these performance methods and aligning the metric with the goal of the research [42]. A good model must have a balance between bias and variance such that it minimizes the margin of error. The performance of the classification algorithm can be split into the performance of the classification algorithm per category. The focus of the performance needs to be on the *Sleeping* segment and its segment in relation to the *Premiere* segment as the aim of the research is to create distinct retention strategies. For this per-category evaluation, three performing metrics that have proven to be effective or this research.

- *Precision*: the ratio of the number of correctly classified items (True Positives) to the total items in that category (True Positives & False Positives), see Equation 1 [43].
- *Recall*: the number of correctly classified items (True Positives) to the total times that item occurs (True Positives & False Negatives), see Equation 2 [43].
- *F1-score*: the harmonic mean of precision and recall, see Equation 3 [44].

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F1 - score = \frac{2 \times precision \times recall}{precision + recall} \tag{3}$$

3.7 Interpretability

Segments of members are created based on member value and a model is created to predict the segment of current members. With this model and its segments, the company can develop retention strategies. However the company has no clue what drives the allocation of members to segments. For optimal usability, understanding and interpretability, the company must understand and conform with the model and its outcome.

Besides the usability for the company, model interpretation is used as error analysis to detect spurious correlations and to review the model.

SHAP is used to interpret and analyze the results of the black-box model. It is one of the many methods to include XAI into research. It explains the prediction of an instance by computing the contribution of each feature to the prediction [45]. When combining feature importance with feature effects, an overview of the relationship between the value of a feature and the impact prediction per category is given.

Two SHAP interpretation techniques are used.

1. **Summary plot**. The impact of binary and ordinal values on classification are derived with this summary plot. Features in these plot are displayed in order of importance.

2. **Dependence plot**. Summary plots are maximally informative with numerical or ordinal feature values. But as some of the features are nominal, the summary plots are not maximally informative in this research. Dependence plots are used to expose the correlation with the target for every feature value. Correlations and differences between distinctive feature values can be clarified by illustrating their SHAP value in the dependence plot. The SHAP value is the importance of the feature value for classification [45] [46]. The SHAP value on the y-axis of the dependence plot and the x-axis of the summary plot together show the contribution of that feature value to the prediction. Each point in the dependence plot represents an instance with its contribution which can be displayed as a coordinate. The SHAP value corresponding the instance defines the weight of that feature value to the prediction of that member.

Figure 1 shows an example of how SHAP values are added up to compute the chance of being a *Sleeping* member. It shows that the member is invited by a friend and that this feature value has the biggest influence on the prediction for the segment of this member. The base value is the average model output over the training dataset we passed [**?**]. The features displayed in red increased the chance of being classified as *Sleeping*, while the features in blue decreased the chance of being classified as *Sleeping*. Together, the chance that this member is a *Sleeping* member is predicted as -0.89.

All in all, the SHAP value on the y-axis of the dependence plot and the x-axis of the summary plot together show the contribution of that feature value to the prediction of all members summarized.

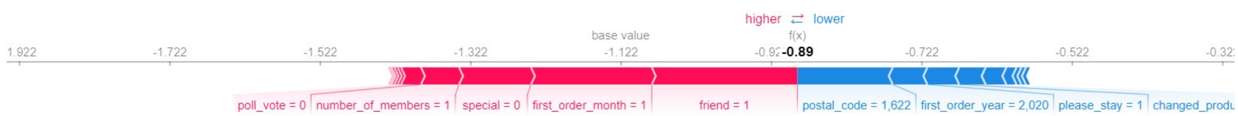

Figure 1: Example of the influence of each feature value to segmentation of a member in the *Sleeping* segment according to SHAP values

4 Results

4.1 Target

The first goal of this research is to define member segments based on profitability and loyalty. The second goal of this research is to predict the member segment according to the labels of the original RFM-model. At last is to derive the feature characteristics of the segments. The ambition of the company is to develop three different retention strategies based on the FB-segments. A member is labeled with *Premiere*, *Potential*, or *Sleeping*. Corresponding characteristics of the frequency and breaks per label can be found in Table 5 and 6. The per-category evaluation shows the performance per class as a result of the difference in importance of the three segments.

4.2 Multiclass classification

4.2.1 Overview of algorithms

Table 9 displays a group of multiclass classification algorithms and their F1-scoring for the different segments. The selection of classifiers is based on the classification algorithms sci-kit learn provides [37]. Since we have dealt with class imbalance, the class proportions are equal per segment.

The best performing algorithms based on F1-score are the boosting classifiers. The Gradient Boosting Classifier and ADABoost are two of the boosting classifiers implemented by sci-kit learn [37] but many more boosting classifiers exist. Because of the outstanding performance of the boosting classifiers; XGBoost, LGCMClassifier, and CatBoost have been added to the selection of algorithms in Section 4.2.2.

	Premiere	**Potential**	**Sleeping**
Class proportion	*0.33*	*0.33*	*0.33*
Gradient Boosting	0.59	0.57	0.53
ADABoost	0.59	0.57	0.50
Random Forest	0.59	0.54	0.49
Extra Trees	0.58	0.53	0.49
Bernoulli NB	0.53	0.56	0.46
Linear Discriminant	0.54	0.58	0.51
Decision Tree	0.52	0.48	0.43
Gaussian NB	0.52	0.57	0.33
Ridge	0.55	0.58	0.49
k-Neighbours	0.37	0.42	0.28
Nearest Centroid	0.40	0.01	0.41

Table 9: F1-scoring for sci-kit learn classification algorithms

4.2.2 Selection of algorithms — zooming in on boosting classifiers

The search is widened by including more boosting classifiers since the boosting technique is promising. Table 10 shows the classification matrix for the XGB Classifier [47], LightGBM Classifier [48] and the Cat Boost Classifier [49] besides the Gradient Boosting Classifier. ADABoost is eliminated from this list as the Gradient Boosting Classifier outperforms ADABoost.
The matrices show the precision and recall per segment. A closer look at the results shows that the algorithms commit a similar result on all segments with regard to the metrics. However, the LightGBM scores better on the *Sleeping* segment.
The Gradient Boosting Classifier however comes with deficiencies as it only supports the log odds estimator for binary classification [50]. Due to the importance of interpretability and the fact that the Gradient Boosting Classifier in multiclass classification is not supported by SHAP, this algorithm is not suitable.

Gradient Boosting Classifier

	Precision	**Recall**	**F1-score**	***Class proportion***
Premiere	0.62	0.57	0.59	*0.33*
Potential	0.61	0.54	0.57	*0.33*
Sleeping	0.49	0.58	0.53	*0.33*

XGBoost

	Precision	**Recall**	**F1-score**	***Class proportion***
Premiere	0.59	0.62	0.61	*0.33*
Potential	0.61	0.53	0.57	*0.33*
Sleeping	0.51	0.55	0.53	*0.33*

LightGBM

	Precision	**Recall**	**F1-score**	***Class proportion***
Premiere	0.59	0.62	0.61	*0.33*
Potential	0.62	0.53	0.57	*0.33*
Sleeping	0.51	0.56	0.53	*0.33*

CatBoost

	Precision	**Recall**	**F1-score**	***Class proportion***
Premiere	0.58	0.63	0.61	*0.33*
Potential	0.62	0.52	0.56	*0.33*
Sleeping	0.51	0.55	0.53	*0.33*

Table 10: Performance of boosting classifiers

4.3 Automated machine learning

All suitable algorithms are tuned on hyperparameters in search for the best algorithm since Table 10 showed that the margins on performance are limited. The LightGBM Classifier established the highest macro-F1-score when hyperparameter optimization by FLAML [25] is applied jointly with 10-fold cross-validation. The hyperparameters used for FLAML hyperparameter tuning are:

- metric: macro_f1
- time-budget: 10.000
- task: classification
- estimator_list: auto
- n_jobs: -1
- ensemble: False
- verbose: 3
- retrain_full: True
- split_type: auto
- hpo_method: auto
- starting_points: static
- seed: None
- n_concurrent_trials: 1
- early_stop: False
- append_log: False
- auto_augment: False
- min_sample_size: MIN_SAMPLE_TRAIN
- use_ray: False
- metric_constraints: [ ]
- custom_hp: None

Table 10 presents the results for LightGBM after and before hyperparameter optimization. After applying the best hyperparameters, the classification repots shows that the F1-score of the *Potential* segment increases with 1 percentage-point by the use of AutoML. Other segments remain unchanged. The best parameters for LightGBM in this model are:

- n_estimators: 72
- num_leaves: 26
- min_child_samples: 8
- learning_rate: 0.06456013493444863
- log_max_bin: 8
- colsample_bytree: 0.740903127598881
- reg_alpha: 0.036114468962103394
- reg_lambda: 0.02544819127338174

LightGBM

	Precision	Recall	F1-score	*Class proportion*
Premiere	0.59	0.62	0.61	*0.33*
Potential	0.62	0.53	0.57	*0.33*
Sleeping	0.51	0.56	0.53	*0.33*

LightGBM after hyperparameter optimization

	Precision	Recall	F1-score	*Class proportion*
Premiere	0.60	0.61	0.61	*0.33*
Potential	0.62	0.54	0.58	*0.33*
Sleeping	0.50	0.57	0.53	*0.33*

Table 11: Classification report for the LightGBM Classifer with and without hyperparameter optimization

4.4 Binary classification

With binary classification, one of the target segments is put in contrast to the other segments combined. This method is commonly preferred over multiclass classification since binary algorithms are less complex due to dimensionality [51]. Section 3.5.2 discussed why the usage of multiclass classification is preferred over binary class classification. Table 12 shows a difference between binary classification scorings and multiclass classification scorings. Despite using cross-validation, performance of binary classification may differ every run by small percentages towards multiclass classification. If the this research would solely focus on predicting member value, binary class classification is promising. However, multiclass classification is more accessible in terms of interpretability. This research is highly focused on providing interpretable results. Multiclass classification is therefore favoured relative to binary classification although binary class classification is commonly preferred over multiclass classification.

	Precision	Recall	F1 - score	*Class proportion*
Premiere	0.70	0.80	0.75	*0.33*
other label	0.77	0.66	0.71	*0.66*
Potential	0.77	0.58	0.66	*0.33*
other label	0.67	0.83	0.74	*0.66*
Sleeping	0.64	0.69	0.66	*0.33*
other label	0.67	0.62	0.64	*0.66*

Table 12: Scoring for binary classification with LightGBM

4.5 Different methods to optimize performance

Two methods are used to optimize the model. Both methods will be compared with the algorithm after hyperparameter tuning stated in Table 11. This model will be used as baseline.

4.5.1 Redefining segments

As seen in Figure 5, the minimum number of months a *Sleeping* member is subscribed, is higher than the minimum number of months a *Potential* member is subscribed. A member that has a frequency of 1, without having any breaks, can therefore also be classified as *Potential*. This circumstance is based on the defined FB-model in Section 3.2 and based on the allocation of scores to segments. Section 3.3 states the maximum score that can be given to a member is six and the minimum score that can be given to a member is two. The members with a score of 5 and 6 are labeled as *Premiere*, scores of 3 and 4 are labeled as *Potential*, and scores of 2 are labeled as *Sleeping*.

Tables 13 and 14 show the characteristics of members with a score of 5 and 6 are still labeled as *Premiere*, but scores of 4 are labeled as *Potential*, and scores of 3 and 2 are labeled as *Sleeping*. Table 13 shows that the frequency of *Sleeping* and *Potential* members is more distinct than in Table 5.

Frequency	**Minimum**	**1st quartile**	**Mean**	**3rd quartile**	**Maximum**	**Group size**
Premiere	18	21	28	33	150	39.622
Potential	10	11	14	16	28	42.130
Sleeping	1	1	3	5	18	226.108

Table 13: Characteristics of frequency per segment after re-allocating scores

Breaks	**Minimum**	**1st quartile**	**Mean**	**3rd quartile**	**Maximum**	**Group size**
Premiere	0	1	1	1	2	39.622
Potential	0	0	1	1	3	42.130
Sleeping	0	0	1	1	32	226.108

Table 14: Characteristics of number of breaks per segment after re-allocating scores

Using a dataset with better distinction in labels does not result in a better performing model compared to the baseline. Table 15 shows that the performance of classifying *Sleeping* members has improved. However the performance of *Potential* has decreased significantly. Allocating scores to segments according to Section 3.3 has proven most successful.

	Precision	Recall	F1-score	*Class proportion*
Premiere	0.52	0.72	0.60	*0.33*
Potential	0.44	0.36	0.40	*0.33*
Sleeping	0.65	0.51	0.57	*0.33*

Table 15: Classification report for the LightGBM Classifer with and without hyperparameter optimization

4.5.2 Time-sensitive features

Engagement features include *poll_vote*, *friend*, *invitation_sent* and *player_email*. Those engagement strategies are used for a pre-defined period of time only. This means that not all members can have participated in an engagement strategy. A new dataset is created out of the original dataset to train the algorithm only on members that have had a chance to participate in engagement strategies. The timeline shows the date ranges that are extracted from the original dataset. For the evaluation of such time-sensitive features, the *friend*, *poll* and *player_email* features are summarized as a feature named *engagement*. This binary feature defines if a member participated in at least one of the engagement strategies.

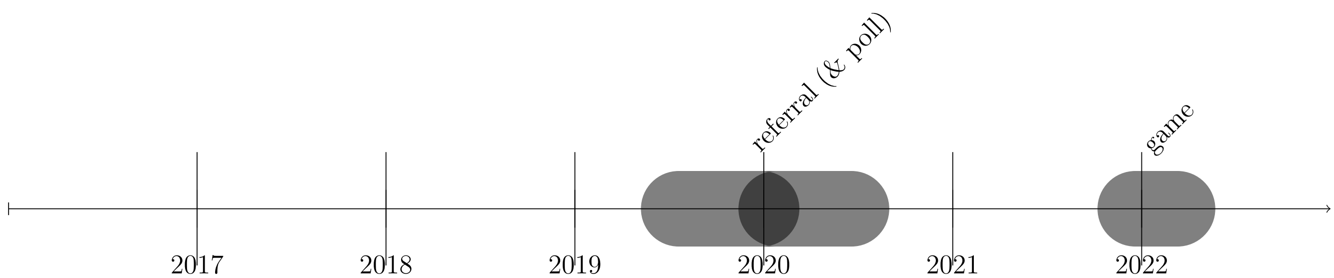

Table 16 shows the performance of grouping engagement strategies and eliminating time sensitivity. Eliminating time sensitivity by only using time periods where engagement strategies were used does not increase the performance of the model compared to the baseline in Section 11. In particular the performance of the *Sleeping* segment decreased relative to the baseline model.

	Precision	Recall	F1-score	*Class proportion*
Premiere	0.49	0.65	0.56	*0.33*
Potential	0.55	0.46	0.50	*0.33*
Sleeping	0.50	0.42	0.46	*0.33*

Table 16: Classification report for the LightGBM Classifer by grouping engemenent strategies and eliminating time sensitivity

4.6 Interpretability

Answering the third sub-question contributes to the usability, understanding, and trust of the algorithm. The feature importance is showed in Figure 20. The summary plots are enclosed in Appendix Section C.

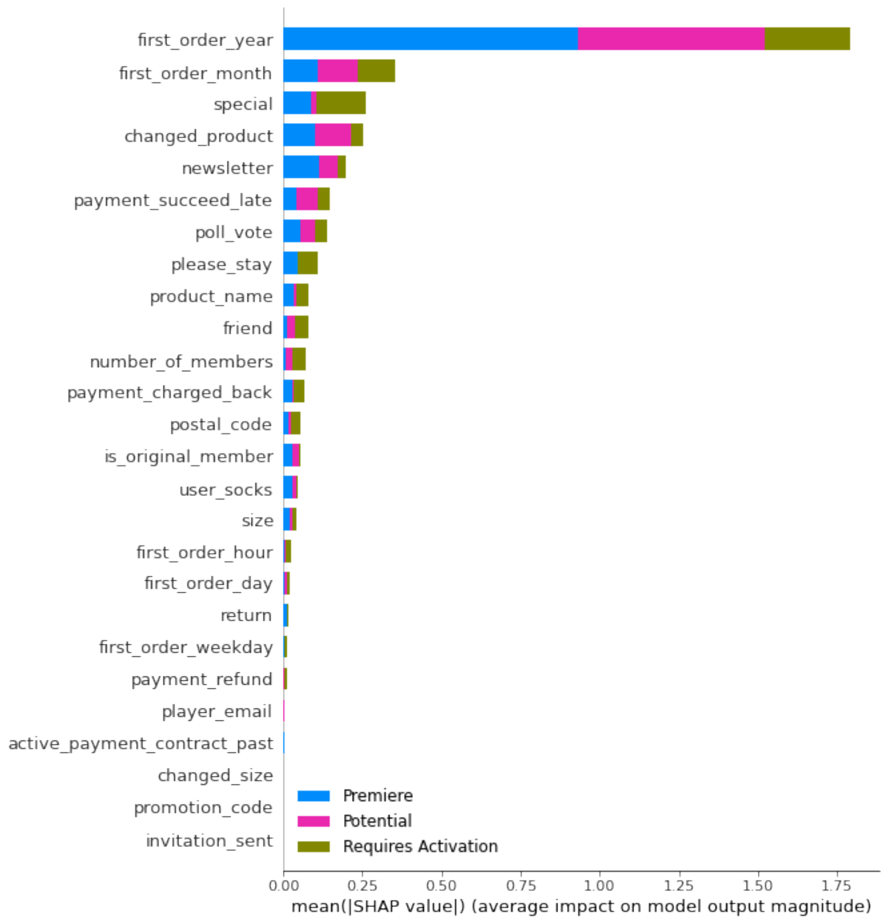

Figure 2: Cumulative feature importances

4.6.1 Most important features for classification

Figure 20 shows the cumulative feature importance broken down in segments. As shown in Figure 20, the *first_order_year*, *special*, *changed_product* are the three most important features for segmentation. Although those features are key to predict the segment of member, the features are not necessarily practically significant. As stated in Sector 1.3.3, this research delivers a machine learning model for member segment prediction and a list of recommendations regarding the feature values that contribute to the allocation of the *Sleeping* segment. Sector 4.6.2 defines the features that have a major impact on the segmentation of *Sleeping* members and the potential retention strategies drafted by the company.

4.6.2 Most significant features regarding the business strategies

Product name — A significant difference can be derived between the product types *Men Originals* (print), *Men One* (without print), *Boys Originals* (kids sizes with print). According to Figure 3, the contribution of Men One to the *Sleeping* segment is higher than the other product types to the *Sleeping* segment. This is in line with the negative contribution of Men One to the *Premiere* segment.

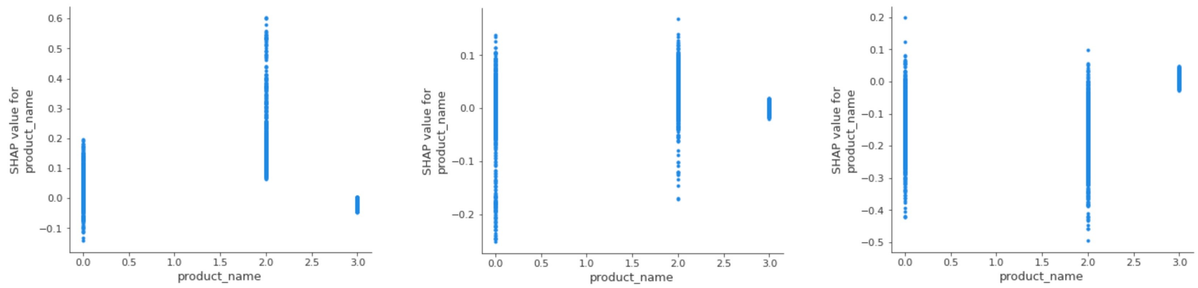

Figure 3: Dependence plots of *product_name* for relatively the Sleeping, Potential, and Premiere segment.

Please stay — A positive case for please stay indicates that the member has accepted an offer to please stay under a special condition, concerning the dependence plots in Figure 4. Members that have accepted a please stay offer are more likely to be classified *Sleeping*, or at least as *Potential*. This is in line with expectations since members that received a please stay offer already had an incentive to churn.

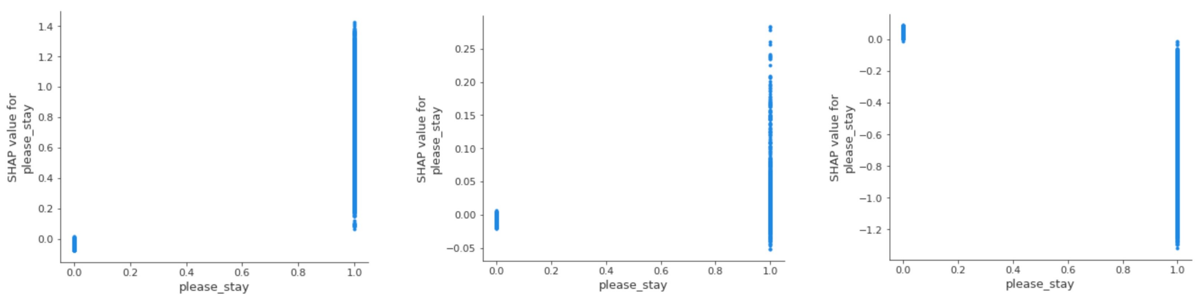

Figure 4: Dependence plots of *please_stay* for relatively the Sleeping, Potential, and Premiere segment.

Special — The purchase of an extra limited edition boxer short, specials, decreases the probability to be classified as *Sleeping*. It is furthermore a strong indication that the member will be a *Premiere* customer, as can be seen in Figure 5.

Payment charged back — When payments are charged back by the member, it increases the possibility that the member belongs in the *Sleeping* segment, as can be seen in Figure 6. This implies that if a member succeeds his payment late it is no indication the member becomes a *Sleeping* member, despite the presumption.

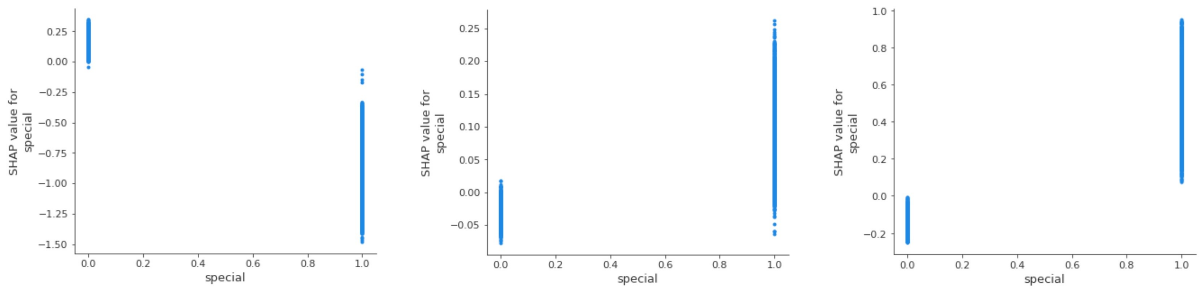

Figure 5: Dependence plots of *special* for relatively the Sleeping, Potential, and Premiere segment.

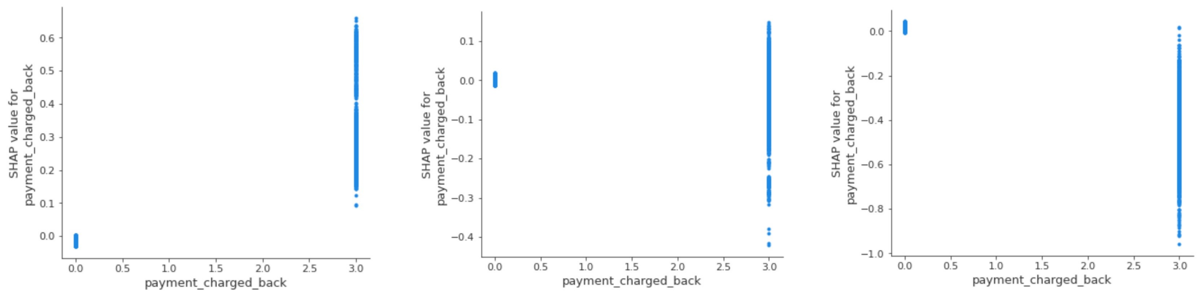

Figure 6: Dependence plots of *payment_charged_back* for relatively the Sleeping, Potential, and Premiere segment.

Payment succeed late — When a payment has been completed but after the payment deadline, it could suggest the member being a *Sleeping* member. However, the dependency plot in Figure 7 shows that if the payment succeeds late, it is of marginal contribution to being a *Sleeping* member.

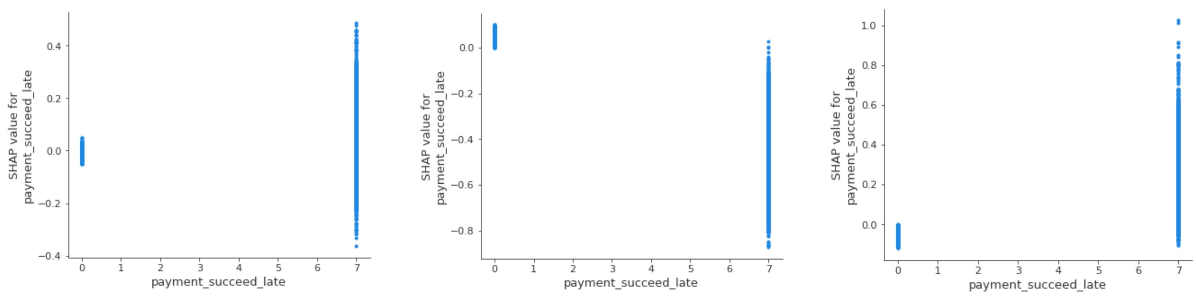

Figure 7: Dependence plots of *payment_succeed_late* for relatively the Sleeping, Potential, and Premiere segment.

Payment refund — Figure 8 shows that a payment refund significantly increases the probability that the member becomes a *Sleeping* member. This finding can potentially be declared as payment refunds are made only if the member rejects a replacement which implies dissatisfaction.

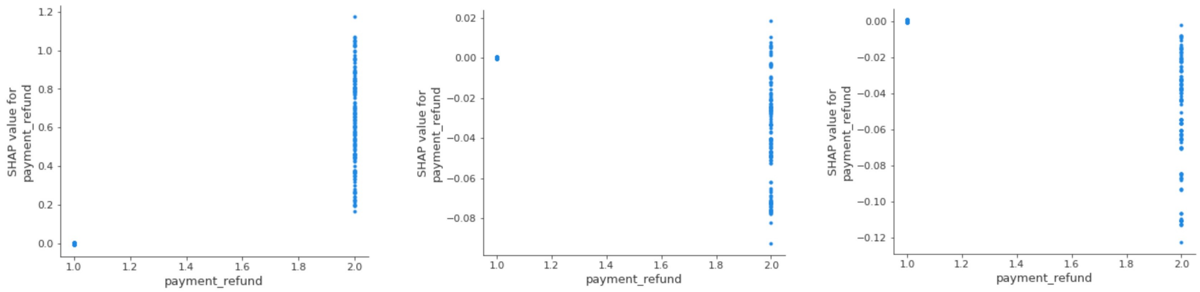

Figure 8: Dependence plots of *payment_refund* for relatively the Sleeping, Potential, and Premiere segment.

Return — Returns differ from payment refunds as returns could result in either payment refunds or product replacements. In general, returns do not positively contribute to being classified as *Sleeping*, according to Figure 9. If a return is being made, the member is not likely to be a *Sleeping* member. The chance of returning the product is presumably higher when the member has a higher frequency. The dependency can therefore best be summarized by the fact that returns do not contribute to the member being as *Sleeping*. The contribution of returns to the model are opposite from the contribution of payment refunds.

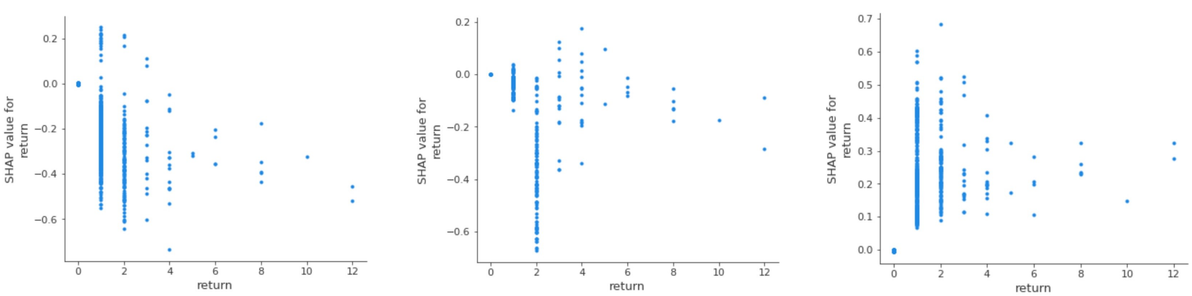

Figure 9: Dependence plots of *return* for relatively the Sleeping, Potential, and Premiere segment.

Newsletter — According to Figure 10, a member that is subscribed to the newsletter shows a higher probability of being classified as *Premiere*. This shows that engagement with the company contributes to being a valuable member.

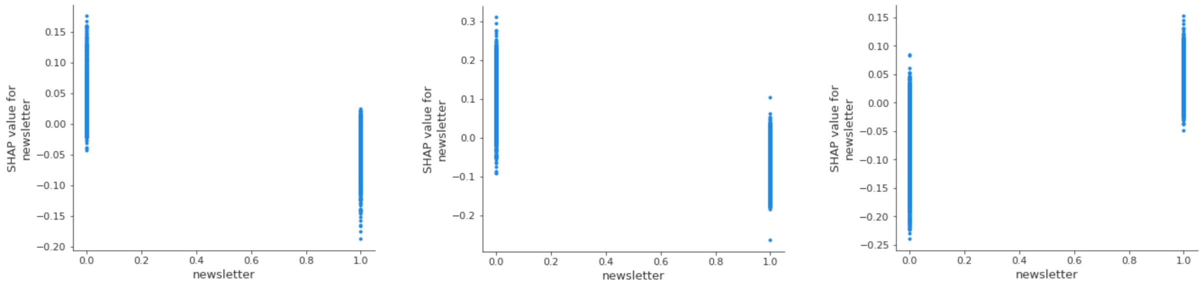

Figure 10: Dependence plots of *newsletter* for relatively the Sleeping, Potential, and Premiere segment.

Number of members — The more members to one user account, the more chance of being a *Sleeping* user, according to Figure 11. Only the first member added to a user account profits from the free trial, so adding an extra membership to profit from the free-trial period is futile. The maximum number of boxer shorts memberships on one account is five, as set by the company. The maximum number of socks memberships on one account is also five.

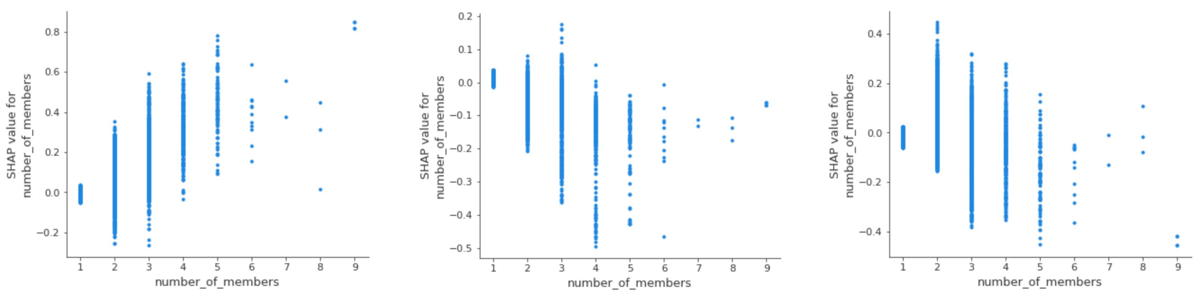

Figure 11: Dependence plots of *number_of_members* for relatively the Sleeping, Potential, and Premiere segment.

Socks — As can be seen in Figure 12, a socks membership decreases the chance to be classified as a *Sleeping* user and increases the chance of being classified as *Premiere* user.

Is original member — Figure 13 does not demonstrates a significant difference in segmentation. Being the first member of the user account barely contributes to segmentation.

Poll vote — The same influence on the classification is found in the feature that captures the votes of members in a poll. The dependence shown in Figure 14 indicates that if a member participated in the poll, the member has a higher probability of being classified as *Premiere* instead of *Sleeping*.

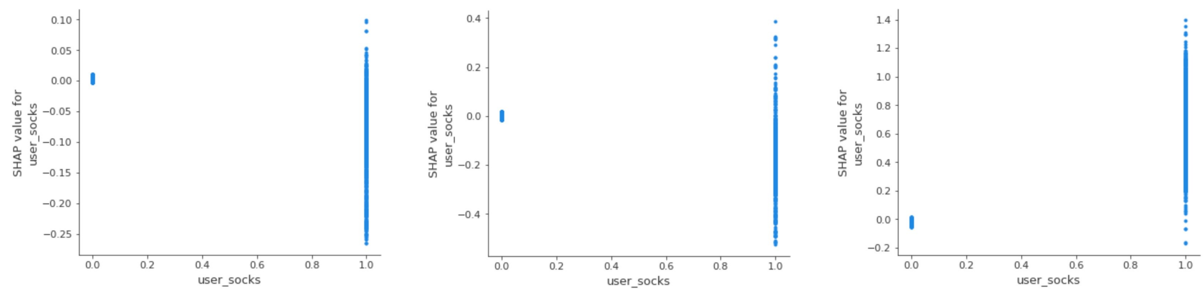

Figure 12: Dependence plots of *user_socks* for relatively the Sleeping, Potential, and Premiere segment.

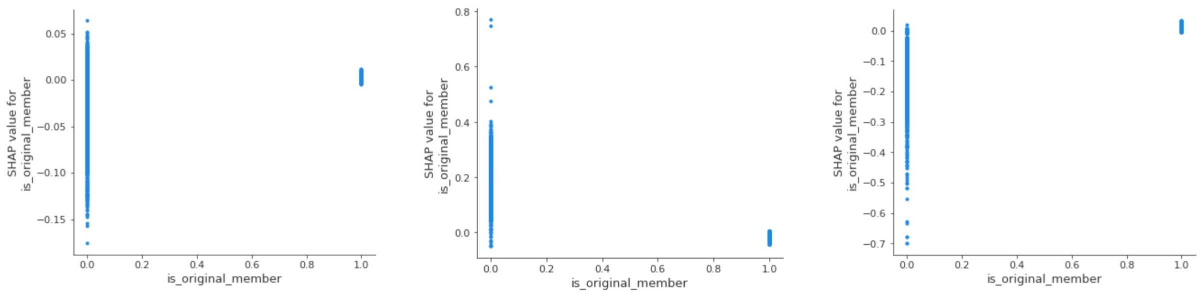

Figure 13: Dependence plots of *is_original_member* for relatively the Sleeping, Potential, and Premiere segment.

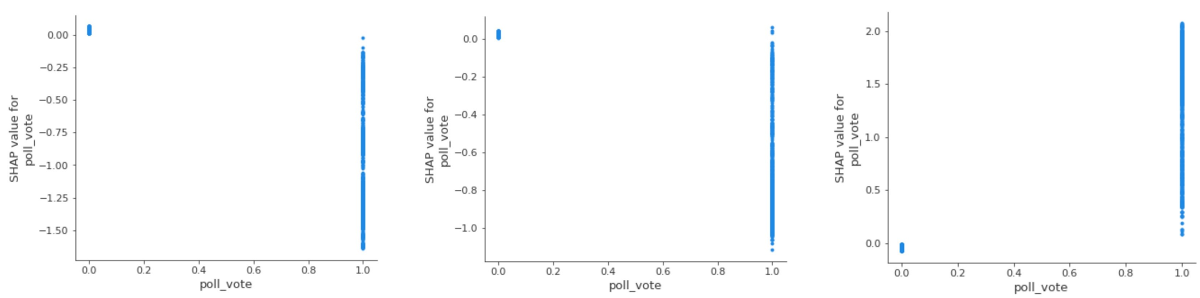

Figure 14: Dependence plots of *poll_vote* for relatively the Sleeping, Potential, and Premiere segment.

Invitation sent — Members were able to invite friends in the past. Since the company is planning on launching a new referral program, the performance of the referral features is relevant. The dependence plots in Figure 15 shows us that the member sending the invitation to a friend is more likely not to be classified as *Sleeping*. However, the plot for the *Premiere* segment does not show a significant probability of the member being classified as *Premiere*.

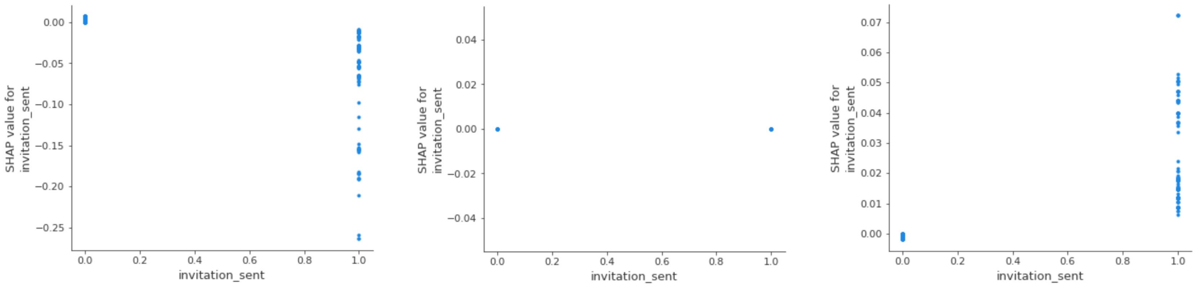

Figure 15: Dependence plots of *invitation_sent* for relatively the Sleeping, Potential, and Premiere segment.

Friend — Apart from sending an invitation, some members have received an invitation before joining the company. Although sending an invite slightly suggests being classified at least as a *Premiere* member, receiving an invite does not particularly create valuable members. The dependency plot in Figure 16 even shows that the probability of being classified as *Sleeping* is higher than the probability of being classified differently.

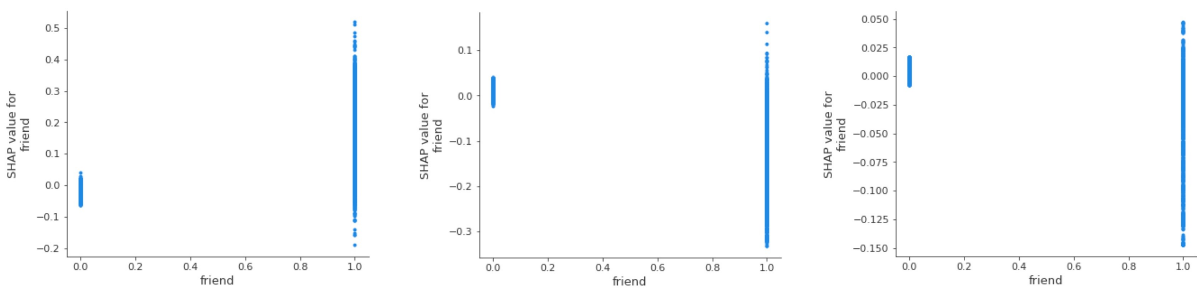

Figure 16: Dependence plots of *friend* for relatively the Sleeping, Potential, and Premiere segment.

First order weekday — A distinction in days of the week can be made between weekdays and weekend, according to Figure 17. Weekends have a higher potential to members being classified as *Potential* or *Premiere*. Especially creating an account on Sunday decreases the chance of being classified as *Sleeping*.

First order month — Concerning the dependency graph in Figure 18, a member is more likely to be classified as *Potential* when the membership is create in November or December. The probability that a member is *Premiere* if the account is created in November of December is significantly small.

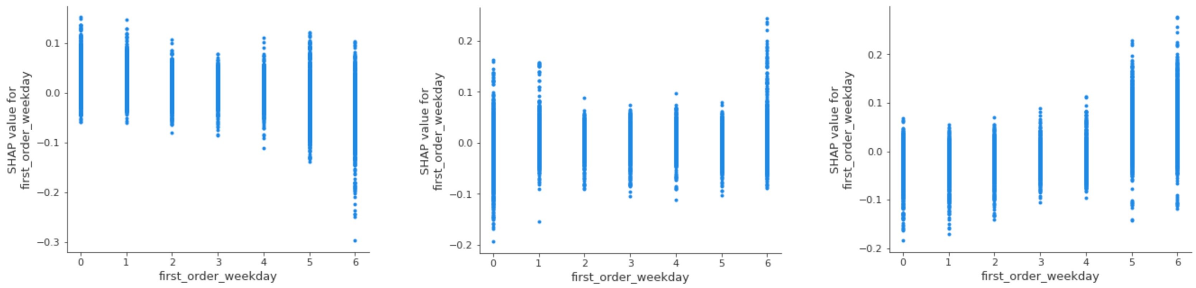

Figure 17: Dependence plots of *first_order_weekday* for relatively the Sleeping, Potential, and Premiere segment.

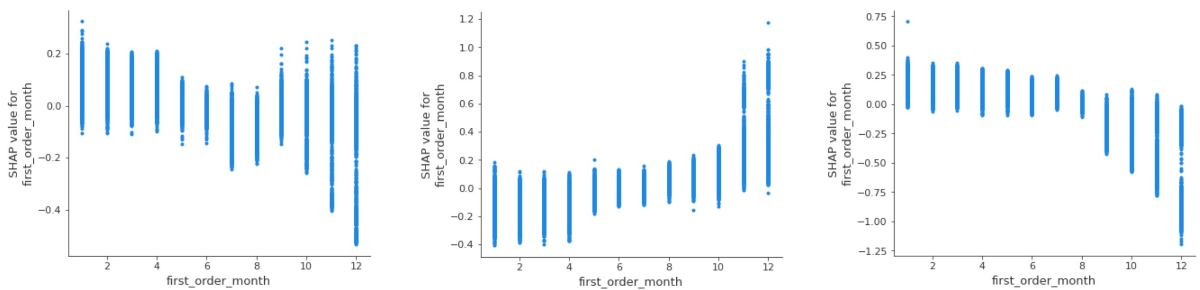

Figure 18: Dependence plots of *first_order_month* for relatively the Sleeping, Potential, and Premiere segment.

Active payment contract past — If the member ever had an inactive payment contract, it is a very good indication that the member is not a *Premiere* member, according to Figure 19.

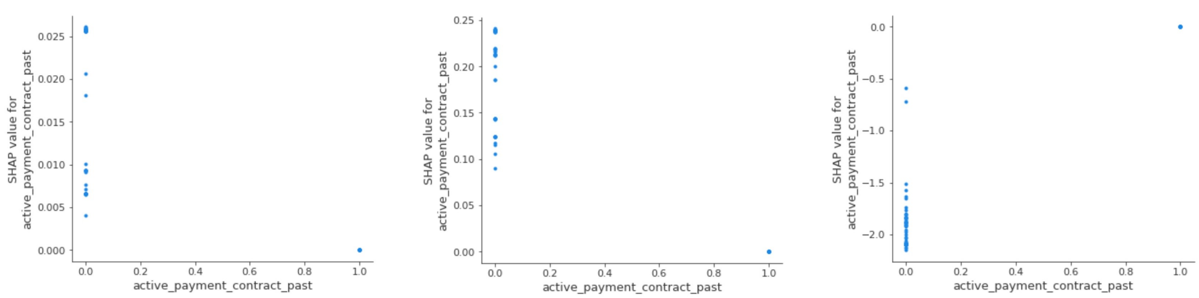

Figure 19: Dependence plots of *active_payment_contract_past* for relatively the Sleeping, Potential, and Premiere segment.

4.6.3 Secondary results

This Section describes the features that do not have a practical significant impact on defining potential retention strategies.

Postal code — The dependence of postal code in relation with the member segments, shows some variation. However, interpretation is still hard, see Figure 23. Due to the complexity, a more detailed research is outside the scope of this research.

Changed product — Making a change in product type increases the chance of not being classified as *Potential* but as *Sleeping* or *Premiere*. With the results from Figure 24 mind, we can state that changing the product does not significantly contribute to segmenting *Sleeping* from *Premiere* members.

Size — The size of the boxer short does not make a clear distinction in contribution to the classification problem, as seen in Figure 25.

Changed size — If the member changed the size of the boxer short, it appears in Figure 26 that the member is not a *Sleeping* member. This can be explained by the *Sleeping* member having an average frequency of only 1 month and therefore not being likely to change size.

Player email — Different than other features measuring member engagement, members that competed in an online game do not have higher probability to be allocated to a segment. This is shown in Figure 27.

First order hour — The dependency between the target and the hour in which the member was created shows a slight positive trend on the first 12 hours of the day for the *Sleeping* segment. The trend indicates that gradually, less valuable members are created as the first 12 hours of the day have passed by. The most valuable members are created between 4 AM and 6 AM, according to Figure 28.

First order day — For the days an user account has been made, a slight periodicity can be seen in the dependence graph for *Sleeping* members. The last and first days of the month are the top and the middle days of the month are valley points in this dependence graph. This indicates that the first and last days of the month state the highest probability of being a *Sleeping* member. Figure 29 shows a sharp fall around the twentieth of the month for *Premiere* members. This indicates a decrease in probability of being classified as *Premiere*. All in all, some distinction can be made according to the dependence plots in Figure 29, but differences are minimal.

First order year — The feature that captures the first year the member subscribed, shows a biased result. Since *Premiere* members have an average frequency of 28 months, *Premiere* members can, by the seventeenth of June 2022, not have been created in 2021 or 2022. When looking at the years till 2021, members created in 2018 have the highest probability of being classified *Sleeping*.

Promotion code — The summary plot in Appendix C shows that the influence of promotion codes is relatively small in relation with other features. The dependence plots in Figure 31 confirm this statement.

4.6.4 Summary of most important interpretable results

To summarize the dependence of the feature values in relation to the segmentation, we can state that:

- The contribution of Men One to the *Sleeping* segment is higher than the other product types.
- Members that have accepted a please stay offer are more likely to be classified Sleeping.
- The purchase of a special increases the chance that the member will be a *Premiere* member.
- When payments are charged back by the member, it increases the possibility that the member belongs in the *Sleeping* segment
- If a member succeeds his payment late, it is no indication the member is a *Sleeping* member
- If a member has had a payment refund, it increases the chance of being a *Sleeping* member.
- If a return is made, the member is not more likely to be a *Sleeping* member.
- A member that is subscribed to the newsletter shows a higher prediction of being classified as *Premiere*.
- The more members to one user account, the more chance of being a *Sleeping* user.
- A socks membership decreases the chance to be classified as a *Sleeping* user and increases the chance of being classified as *Premiere* user.
- Being the first member of the user barely contributes to segmentation.
- If a member participated in the poll, the member has a higher probability of being classified as *Premiere* instead of *Sleeping*.
- The member sending the invitation to a friend is more likely not to be classified as *Sleeping*, However, the chance the member is *Premiere* is not significant.
- The probability of a member invited by a friend being classified as *Sleeping* is higher than the probability of being classified differently.
- Weekends have a higher potential to members being classified as *Potential* or *Premiere*. Especially creating an account on Sunday decreases the chance of being classified as *Sleeping*.
- A member is more likely to be classified as *Potential* when the membership is create in November or December. The probability that a member is *Premiere* if the account is created in November of December is significantly small.
- If the member ever had an inactive payment contract, it is a very good indication that the member is not a *Premiere* member.

5 Recommendations

The goal of this research is to classify Dutch current members of the company into segments based on profitability and loyalty and to predict the value of a new member, in order to behave towards personalised retention strategies. The second part of the research purpose is to uncover which feature values contribute to the classification to improve understanding and usability. Instructions on how to use the model are enclosed in Appendix A.

5.1 Strategy

The retention strategy can be designed in cooperation with the influence of feature values on segmentation the predictions per member. The recommendations is drawn up with regard to the *Sleeping* segment.

The first points of interest are the features that represent strategies used in the past or used infrequently. Those features include *friend*, *invite*, *poll_vote* and *player_email* and are one of the few features that represent member engagement. As stated in Section 1, member engagement is key to profit from customers. Engagement is currently also captured *newsletter*, *special* and *user_socks*, and shows significant results. As seen in Figures 5, 10, 12, 14 and 16, engagement strategies are key to segment members. The recommendation for the company is therefore to implement marketing strategies based the features representing member engagement again or to deploy them more frequently.

One of the engagement strategies that is planned to relaunch according to the company is the referral strategy. This referral strategy is caputred in the friend/invite features. The results in Figure 15 and 16 showed us that although sending an invite suggests being classified as *Potential* or *Premiere*, receiving an invite tends not to create valuable members. In relation to the soon-to-be-launched referral program, this conclusion should be taken into account when defining the terms and conditions for inviting a friend. Literature states that referral programs are the final step in the ' Grow Customers' strategy that comes after the ' Get Customers' strategy and ' Keep Customers' strategy [52]. It implies that referral strategies should be used for loyal customers, so-called brand ambassadors. One of the possible conditions that can be drafted according to the results and the literature can be found in the minimal length of the subscription or the degree of engagement. It implies that the company should restrict the option to send an invite to *Premiere* members only, since those members have a higher brand loyalty than *Sleeping* members. In any case should the company monitor the behaviour of friends that have become members.

Extra attention is also beneficial for members that ask for a refund. Since Figure 8 shows that the odds of being classified as *Sleeping* increase when a payment refund is requested, customer service must define a strategy for situations where a member asks for a refund in order to increase satisfaction and customer lifetime.

Another recommendation regarding the feature value impact on segmenting members is about the please stay strategy. A member that has once received and accepted such an offer, is likely to be classified as *Sleeping* anyway according to Figure 4. The recommendation is to look at this strategy with a critical eye and to calculate the surplus profit. If this surplus is positive, the revenue of a member accepting a please stay offer is higher than the cost of the extra goods sold. In case of a positive surplus, the strategy is profitable.

The size of the member and the behaviour of changing the size does not to have a evident impact on the segmentation, according to Figure 25 and 26. That result along with ethical concerns, no marketing strategy should be drafted with regard to the size of the member.

The features *product_name*, *postal_code number_of_members*, *payment_charged_back*, *socks*, *returns*, *first_order_month* and *first_order_weekday* also show significant results in Section 4.6.2 and 4.6.3. To define the most optimal strategy regarding those features, brainstorming and discussions with the marketing team would be the best next step in order to define personalised retention strategies.

At last, the algorithm is currently suited for Dutch members only. Since the sales area has expanded to over 9 countries, according to Section 1.2, the model used in this research is only significant to a specific member base. Therefore, the methods used in this research must also be applied to other countries to derive the possibility to create data-driven retention strategies in all nine countries.

6 Limitations

6.1 Critical reflection

6.1.1 Data

One of the limitations of this research is the potential skewness from limited features. A feature not included in the database and therefore not in the dataset is age. To speed up the enrollment, this data field is removed from the sign-up page in the past. Since age is a commonly used metric in marketing, it is likely that this feature would be of influence in the model, but verification is not possible at the moment. Despite age being a protected feature, as it enables individuals to be treated unfairly, including age as input field in the enrollment process should be considered. In addition. the data in this research does not include social media and e-mail traffic data. This data is stored in a customer data platform. Extracting data from this platform was too time-consuming for this research.

The second limitation in data occurs with taking churned and paid members only as input for training. To repeat, paid members are members that succeed in at least one payment, in contrast to trial members, who cancel their subscription after the 10-day trial period. The decision to not take current members as input for training is that current members are still in transit as to in which segment they belong. If a new member, with a low frequency, would be classified as *Sleeping* but is in fact in transit of becoming a *Premiere* member, the model would be inadequate. Only paid members are included in this research since the research is focused on value over months, assuming the member-generated revenue, instead of generating revenue or not.

Finally, some features are built upon temporal events in the past. *Player_email* is a feature built upon an online game that the company once introduced in combination with a boxer short. *Friend* and *invitation* are extracted from a marketing campaign in 2019, just like *poll_vote*. When those campaigns are not recurring, the features lose value over time. Moreover, since these features were used over some time in the past, most of the members did not have a chance to decide whether or not to respond to these marketing strategies. This implication has partially been resolved by extracting only the time periods where engagement features have been used. However, this method decreases the size of the trainingsset and the performance.

6.1.2 Algorithm

The constraint regarding the algorithm is the finding that the Gradient Boosting Classifier is not supported for multiclass classification in SHAP [50] as SHAP only supports the log odds estimator for binary classification. In Section 4.2.2, the Gradient Boosting Classifier is outperformed by other classifiers, but differences are minimal.

At last, we discuss the performance of the algorithm with regard to the business problem. To prevent a member classified as *Sleeping* from churning, the algorithm must be capable to predict the segment before this member churns. As the member provides more data over the length of the subscription, predicting a segment is easier as months pass by. This can result in not being able to classify members correctly, change in the segment of a member over time, or not being able to classify members before they are churned.

6.2 Future work

In the Critical Reflection in Segment 6.1, the research is stated not to include trial members, only trial months of paid members. In future research, trial members could be incorporated in search of whether a member will be a paid member or not. However, the limited subscription period of trial members potentially lacks data.

Regarding the results in Section 4, future work could include more extensive search into the relation between postal codes and the segmentation of members. Useful information that could be derived from this feature is the extent to which subscription-based businesses are adapted in different regions and possibly in different income groups.

Extracting data from the customer data platform to add social media and e-mail traffic data would be of high value for customer analysis, as social media and e-mail campaigns are the most used marketing tools at the company and e-commerce in general [53]. Extracting this data from this platform requires an API to connect with the separate SQL database that stores information. Including data from the customer data platform potentially makes the model more robust.

One of the key values throughout this research has been interpretability and usability. To stimulate these key performances, future research could find methods to include data science in business operations. An interview with stakeholders, requirement specification, and literature research could lead to building an interface that is used throughout the organisation or other methods on how to include data science in business operations.

7 Conclusions

Digitalisation, software as a service, and marketing strategies have evolved over the past years. With this evolution, so raised the adaptation and acceptance of subscription-based models in society. With data including demographic data, payment data, and behavioural data, so comes the need of processing this data for decision making. At the company, key performance indicators like the cost of acquisition per member and customer lifetime value continuously must improve in order to stay competitive. To improve the customer lifetime, retention strategies need to be personalised to the needs of the member to enhance prior aspects of subscription-based business, operating customer-centric. The problem we adressed is that the focus of the company is primarily on acquisition and that the current retention strategy of the company focuses on preventing members to churn with non-personal strategies. This research has specifically been set up to discover the characteristics and behaviour of members with regard to their profitability to create persinalised retention strategies.

The goal of this research was to classify Dutch members of the company based on value and discover corresponding member characteristics and behaviour to assign personalised retention strategies. The first objective aimed to discover a method for customer segmentation based on profitability and loyalty. By using the RFM-model as a foundation for measuring profitability and loyalty, this model was reshaped into a FB-model with frequency and breaks as metrics to better suit subscription-based businesses. These metrics were defined by grouping the metrics in quartiles based on the distribution.

The derived labels given to the segments based on the FB-model in order of least valuable to most valuable are *Sleeping*, *Potential* and *Premiere*. Our results showed that the use of multiclass classification with boosting algorithms suits the problem best by demonstrating a F1-score between 0.53 and 0.61 per segment.

With the results in terms of the performance of the algorithms, the substantiation has been derived to shed light on what has driven the algorithm to allocate members into particular segments.
The reputation of the company is mostly built upon the bold and colourful designs which are present in the Men Original subscription. The results showed us that members with a subscription on Men Originals (print), have a higher probability to be classified as *Premiere*. Time based features showed us that weekend days have a higher potential to members being classified as *Premiere*. Other time based feature values show minimal distinctions in probability.
With regard to the soon-to-be-launched referral program, the positive relationship between inviting a friend or receiving an invite with regard to profitability and loyalty is vital. A member sending the invitation to a friend is more likely not to be classified as *Sleeping*. Receiving an invite however, tends not to create valuable members since the probability that these members are classified as *Sleeping* is higher than the probability of being classified *Potential* or *Premiere*.
The purchase of a special is a strong indication that the member will be a *Premiere* customer, just like a socks membership increases the chance of being a *Premiere* member too.
The referral program and purchase of extra products can together with participation in polls, online games, newsletter subscription be summarised as engagement strategies. Those strategies show significant results for segmenting members, especially to extract *Premiere* members.
When payments are charged back by the member or refunded, it increases the possibility that the member belongs in the *Sleeping* segment. In contrast to payments that are charged back or

refunded, returns do not increase the potential of being a *Sleeping* member.
More members on one account increase the chance of being a *Sleeping* member and decreases the chance of being *Premiere* member.
At last, members that accept a please stay offer, are more likely to be classified as *Sleeping*.

Limitations regarding the data aspect of this research include missing data or lack of data and training on churned members only. Algorithmic constraints were found in the poor integration of multiclass classification with the Gradient Boosting classifier in combination with SHAP. The biggest limitation of this research is the lack of performance to classify members correctly resulting in the chance of the segment of a member over time or not being able to classify members before they are churned. That would be possible as the member provides more data the over the length of the subscription and predicting a segment is easier as months pass by.

The applied methods and results have proven to be beneficial for the company and invite further research.

References

[1] H. Long, A. Van Dam, Everything's becoming a subscription, and the pandemic is partly to blame (06 2021).
URL https://www.washingtonpost.com/business/2021/06/01/subscription-boom-pandemic/

[2] Cambridge University Press, Subscribtion-based.
URL https://dictionary.cambridge.org/us/dictionary/english/subscription-based

[3] E. Tarver, S. Anderson, Subscription Business Model (05 2021).
URL https://www.investopedia.com/ask/answers/042715/how-do-subscription-business-models-work.asp

[4] T. S. Institute, The subscription economy index, Tech. rep.
URL http://info.zuora.com/rs/602-QGZ-447/images/Zuora_SEI_2021_web.pdf

[5] McKinsey, Delivering Software as a service, Tech. rep. (05 2007).
URL https://abs.in/sites/default/files/Delivering%20Software%20as%20a%20Service.pdf

[6] J. Amankwah-Amoah, Z. Khan, G. Wood, G. Knight, Covid-19 and digitalization: The great acceleration, Journal of Business Research 136 (2021) 602–611. doi:https://doi.org/10.1016/j.jbusres.2021.08.011.
URL https://www.sciencedirect.com/science/article/pii/S0148296321005725

[7] K. Meesilapavikkai, Marketing Communication and Consumer Behavior in Digital Era, Journal of Communication and Innovation NIDA 2016 (3) (2016).

[8] D. Kindström, C. Kowalkowski, T. Alejandro, Adding services to product-based portfolios: An exploration of the implications for the sales function, Journal of Service Management 26 (3) (2015) 372–393. doi:10.1108/JOSM-02-2014-0042.

[9] McKinsey, Thinking inside the subscription box: New research on e-commerce consumers, Tech. rep.
URL https://www.mckinsey.com/~/media/McKinsey/Industries/Technology%20Media%20and%20Telecommunications/High%20Tech/Our%20Insights/Thinking%20inside%20the%20subscription%20box%20New%20research%20on%20ecommerce%20consumers/Thinking-inside-the-subscription-box-New-research-on-ecommerce-consumers.pdf

[10] P. E. Pfeifer, The optimal ratio of acquisition and retention costs, Journal of Targeting, Measurement and Analysis for Marketing 13 (2) (2005) 179–188.

[11] W. Reinartz, J. S. Thomas, V. Kumar, Balancing acquisition and retention resources to maximize customer profitability, Journal of Marketing 69 (1) (2005) 63–79. arXiv:https://doi.org/10.1509/jmkg.69.1.63.55511, doi:10.1509/jmkg.69.1.63.55511.
URL https://doi.org/10.1509/jmkg.69.1.63.55511

[12]

[13]

[14] S. Denning, Don't Call Me A 'Customer'! (04 2013).
URL https://www.forbes.com/sites/stevedenning/2013/02/20/dont-call-me-a-customer/

[15] T. Miller, P. Howe, L. Sonenberg, Explainable AI: beware of inmates running the asylum or: How I learnt to stop worrying and love the social and behavioural sciences, CoRR abs/1712.00547 (2017). arXiv:1712.00547.
URL http://arxiv.org/abs/1712.00547

[16] L. L. Berry, Relationship marketing of services perspectives from 1983 and 2000, Journal of Relationship Marketing 1 (1) (2002) 59–77. arXiv:https://doi.org/10.1300/J366v01n01_05, doi:10.1300/J366v01n01_05.
URL https://doi.org/10.1300/J366v01n01_05

[17] S. Moro, P. Cortez, P. Rita, Using customer lifetime value and neural networks to improve the prediction of bank deposit subscription in telemarketing campaigns, Neural Computing and Applications 26 (1) (2014). doi:10.1007/s00521-014-1703-0.

[18] B. Kim, R. Khanna, O. O. Koyejo, Examples are not enough, learn to criticize! criticism for interpretability, in: D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Vol. 29, Curran Associates, Inc., 2016.
URL https://proceedings.neurips.cc/paper/2016/file/5680522b8e2bb01943234bce7bf84534-Paper.pdf

[19] f. given i=R., given=152, Machine learning explainability vs interpretability: Two concepts that could help restore trust in ai.
URL https://www.kdnuggets.com/2018/12/machine-learning-explainability-interpretability html

[20] A. Ali, S. M. Shamsuddin, A. L. Ralescu, Classification with class imbalance problem, Int. J. Advance Soft Compu. Appl 5 (3) (2013).

[21] N. Japkowicz, The class imbalance problem: Significance and strategies, in: Proc. of the Int'l Conf. on artificial intelligence, Vol. 56, Citeseer, 2000, pp. 111–117.

[22] P. Refaeilzadeh, L. Tang, H. Liu, Cross-Validation, Springer US, Boston, MA, 2009, pp. 532–538. doi:10.1007/978-0-387-39940-9_565.
URL https://doi.org/10.1007/978-0-387-39940-9_565

[23] P. Refaeilzadeh, L. Tang, H. Liu, Cross-Validation, Encyclopedia of Database Systems (2009) 532–538565 doi:10.1007/978-0-387-39940-9\{_}565.

[24] R. Schapire, The strength of weak learnability, Machine Learning 5 (2) (1990) 197–227. doi:10.1023/A:1022648800760.

[25] C. Wang, Q. Wu, M. Weimer, E. Zhu, Flaml: A fast and lightweight automl library, in: MLSys, 2021.

[26] C. Wang, Q. Wu, M. Weimer, E. Zhu, Flaml: A fast and lightweight automl library (2019). doi:10.48550/ARXIV.1911.04706.
URL https://arxiv.org/abs/1911.04706

[27] M. H. Miller, Charles dickens at the english charity dinner, Quarterly Journal of Speech 47 (2) (1961) 143–149. arXiv:https://doi.org/10.1080/00335636109382468, doi:10.1080/00335636109382468.
URL https://doi.org/10.1080/00335636109382468

[28] S. A. Qureshi, A. S. Rehman, A. M. Qamar, A. Kamal, A. Rehman, Telecommunication subscribers' churn prediction model using machine learning (2013) 131–13doi:10.1109/ICDIM.2013.6693977.

[29] C. Blank, T. Hermansson, A machine learning approach to churn prediction in a subscription-based service (2018).

[30] C. Junior, G. Dinis, Churn analysis in a music streaming service : Predicting and understanding retention, 2017.

[31] B. Cooil, L. Aksoy, T. L. Keiningham, Approaches to customer segmentation, Journal of Relationship Marketing 6 (3-4) (2008) 9–39. arXiv:https://doi.org/10.1300/J366v06n03_02, doi:10.1300/J366v06n03_02.
URL https://doi.org/10.1300/J366v06n03_02

[32] Y. Ramon, T. Vermeire, O. Toubia, D. Martens, T. Evgeniou, Understanding consumer preferences for explanations generated by xai algorithms, arXiv preprint arXiv:2107.02624 (2021).

[33] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, F. Herrera, Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai, Information Fusion 58 (2020) 82–115. doi:https://doi.org/10.1016/j.inffus.2019.12.012.
URL https://www.sciencedirect.com/science/article/pii/S1566253519308103

[34] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Vol. 30, Curran Associates, Inc., 2017.
URL https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

[35] P. Cunningham, M. Cord, S. J. Delany, Supervised Learning, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 21–49. doi:10.1007/978-3-540-75171-7_2.
URL https://doi.org/10.1007/978-3-540-75171-7_2

[36] S. Prabasankar, Machine Learning Algorithm — RFM Mod (06 2020).
URL https://sprabasankar.medium.com/machine-learning-algorithm-rfm-model-efd543228a1e

[37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.

[38] G. Lemaître, F. Nogueira, C. K. Aridas, Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning, Journal of Machine Learning Research 18 (17) (2017) 1–5.
URL http://jmlr.org/papers/v18/16-365.html

[39] Skikit Learn, Multiclass and multioutput algorithms.
URL https://scikit-learn.org/stable/modules/multiclass.html

[40] N. Sánchez-Maroño, A. Alonso-Betanzos, P. Garcia-Gonzalez, V. Bolón-Canedo, Multiclass classifiers vs multiple binary classifiers using filters for feature selection, 2010, pp. 1–8. doi:10.1109/IJCNN.2010.5596567.

[41] Cross-validation: evaluating estimator performance (2011).
URL https://scikit-learn.org/stable/modules/cross_validation.html#a-note-on-shuffling

[42] D. M. W. Powers, Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation (2020). doi:10.48550/ARXIV.2010.16061.
URL https://arxiv.org/abs/2010.16061

[43] M. Junker, R. Hoch, A. Dengel, On the evaluation of document analysis components by recall, precision, and accuracy, Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR '99 (Cat. No.PR00318) (1999). doi:10.1109/icdar.1999.791887.

[44] Y. Sasaki, The truth of the F-measure, Tech. rep. (10 2007).
URL https://www.cs.odu.edu/~mukka/cs795sum09dm/Lecturenotes/Day3/F-measure-YS-26Oct07.pdf

[45] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems 30, Curran Associates, Inc., 2017, pp. 4765–4774.
URL http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predicti pdf

[46] C. Molnar, Interpretable Machine Learning, 2nd Edition, 2022.
URL https://christophm.github.io/interpretable-ml-book

[47] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, Association for Computing Machinery, New York, NY, USA, 2016, p. 785–794. doi:10.1145/2939672.2939785.
URL https://doi.org/10.1145/2939672.2939785

[48] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, Lightgbm: A highly efficient gradient boosting decision tree, in: NIPS, 2017.

[49] A. V. Dorogush, V. Ershov, A. Gulin, Catboost: gradient boosting with categorical features support, CoRR abs/1810.11363 (2018). arXiv:1810.11363.
URL http://arxiv.org/abs/1810.11363

[50] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable ai for trees.
URL https://github.com/slundberg/shap/blob/b9e71b7d71e04d835d5ce6a000a5b7ea211c7d21/shap/explainers/tree.py#L472-L477

[51] B. Kijsirikul, N. Ussivakul, Multiclass support vector machines using adaptive directed acyclic graph, Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290) 1 (2002) 980–985 vol.1.

[52] S. Blank, B. Dorf, The startup owner's manual: The step-by-step guide for building a great company, John Wiley & Sons, 2020.

[53] M. Hudák, E. Kianičková, R. Madleňák, The importance of e-mail marketing in e-commerce, Procedia Engineering 192 (2017) 342–347, 12th international scientific conference of young scientists on sustainable, modern and safe transport. doi:https://doi.org/10.1016/j.proeng.2017.06.059.
URL https://www.sciencedirect.com/science/article/pii/S187770581732605X

A Model usage

The model needs to run on a dataset with current members for the company to predict a member's segment. All packages must be installed prior to running the model. The first part of the script is focused on model training. To extract data from the SQL database, the user has to fill in their username and password. Then, the SQL query extracts all data needed for model training. Pre-processing of the data results in a dataframe with feature values per member.The FB-model to label members is built next. Members are automatically labeled while running the script. The LightGBM Classifier has proven to be the most suitable algorithm for the classification problem. The model is trained on this algorithm and the classification report shows the performance of the algorithm. It is important to keep track of the performance as the performance of the algorithm decreases over time and may not meet requirements anymore in the future. When the model is trained, summary and dependence plots are printed to enhance interpretability.

When training is done, the model can be applied to current operations. Extracting data of current Dutch members out of the database needs other data since the goal is to predict the segment of current members. Again, the user must fill in their username and password. When the model is executed, the outcomes are printed in the terminal and exported to `predictions.cvs`. This file needs to be cleared at the start of every run. The outcome represents a dataframe with *member_id*, *user_id*, all features, and the predicted segment as columns.

B Features

B.1 Features with description and type

Feature	Description	Type
product_name	Men Originals, Men One or Boys Originals	integer
postal_code	Postal code of member.	integer
please_stay	Extended subscription after retention strategy.	binary
special	Special edition, not included in the membership.	binary
payment_charged_back	If the payment of the member has once been charged back after the member disputes the transaction.	binary
payment_succeed_late	If the payment of the member once succeeds, but after the payment deadline.	binary
return	Has the product once been returned.	binary
newsletter	If the member is subscribed to the newsletter in more months than the member is not subscribed to the newsletter.	binary
number_of_members	The account is made by a user. A user can have multiple members, every member is a subscription. Number of members stands for how many memberships the user has.	integer
user_socks	If the user has socks. Socks memberships are different memberships, therefore they can only be related to users and not to members.	binary
payment_refund	Has the payment once been refunded.	binary
active_payment_contract_past	If the member ever had an inactive payment method.	binary
is_original_member	If the member is the first member of the user.	binary
changed_product	There are three different key products, Men Originals, Men One and Boys Originals. Did the member changed his product type once.	binary
size	Most ordered size of the subscription.	integer
changed_size	If the member ever changed the size.	binary
player_email	If the user participated in an online the company game.	binary
poll_vote	If the user participated in an poll created by the company.	binary
invitation_sent	If the user has sent an invitation to join the company to a friend.	binary
friend	If the user is invited to join the company by a friend.	binary
first_order_hour	The hour that the member has created the first order.	integer
first_order_day	The day in which the member has created the first order.	integer
first_order_month	The month in which the member has created the first order.	integer
first_order_year	The year in which the member has created the first order.	integer

first_order_weekday	The day of the week in which the member has created the first order.	integer
promotion_code	If used a promotion code when creating the user account.	binary

B.2 Feature transformation — a description of the values

Feature	Value	Interpretation
product_name	0	Boys Originals as most ordered product
	2	Men One as most ordered product
	3	Men Originals as most ordered product
postal_code	diverse	Number indicated the first 4 digits of the members' postal code
please_stay	0	Member has not been offered a retention strategy or has not accepted the offer
	1	Member has accepted to retain after retention strategy
special	0	Member has not bought a special
	1	Member bought at least one special
payment charged back	0	Payment has not been charged back
	3	At least one payment has been charged back
payment succeed late	0	No payment succeed late after the payment deadline
	7	At least one payment succeed after the payment deadline
extended_trial	0	Trail period (10 days) has not been extended by customer service
	1	Trial period (10 days) has been extended by customer service
return	0	Order has not been returned
	1	At least one order has been returned
newsletter	0	Member is not subscribed to the newsletter at least more months than the member is subscribed to the newsletter.
	1	Member is subscribed to the newsletter at least more months than the member is not

number of members	1	User has one membership
	>1	User has more than one membership
user_socks	0	User does not have a socks membership
	1	User has a socks membership
payment_refund	0	At least one payment has been refunded.
	1	At least one payment has been refunded.
active_payment_contract_past	0	Member has had an inactive payment contract
	1	Member has not had an inactive payment contract
is_original_member	0	Member is the the first member of the user
	1	Member is not the first member of the user
changed_product	0	Member never changed the product
	1	Member has changed from product
size	0	104
	1	110-116
	2	122-128
	3	134-140
	4	146-152
	5	3XL
	6	L
	7	M
	8	S
	9	XL

	10	XS
	11	XXL
	12	nan
changed_size	0	Member never changed the size
	1	Member at least once changed the size
player_email	0	User did not compete in the company game
	1	User did compete in the company game
poll_vote	0	User did not voted in poll
	1	User voted in poll
invitation_sent	0	User has not sent an invitation to join the company to a friend
	1	User has sent an invitation to join the company to a friend
friend	0	User is not invited to join the company by a friend.
	1	User is invited to join the company by a friend
promotion_code	0	User has not used a promotion code at user creation
	1	User has used a promotion code at user creation
first_order_hour	0	Activated membership between 00:00AM and 01:00AM
	1	Activated membership between 01:00AM and 02:00AM
	2	Activated membership between 02:00AM and 03:00AM
	3	Activated membership between 03:00AM and 04:00AM
	4	Activated membership between 04:00AM and 05:00AM
	5	Activated membership between 05:00AM and 06:00AM
	6	Activated membership between 06:00AM and 07:00AM

	7	Activated membership between 07:00AM and 08:00AM
	8	Activated membership between 08:00AM and 09:00AM
	9	Activated membership between 09:00AM and 10:00AM
	10	Activated membership between 10:00AM and 11:00AM
	11	Activated membership between 11:00AM and 12:00AM
	12	Activated membership between 12:00AM and 01:00PM
	13	Activated membership between 01:00PM and 02:00PM
	14	Activated membership between 02:00PM and 03:00PM
	15	Activated membership between 03:00PM and 04:00PM
	16	Activated membership between 04:00PM and 05:00PM
	17	Activated membership between 05:00PM and 06:00PM
	18	Activated membership between 06:00PM and 07:00PM
	19	Activated membership between 07:00PM and 08:00PM
	20	Activated membership between 08:00PM and 09:00PM
	21	Activated membership between 09:00PM and 10:00PM
	22	Activated membership between 10:00PM and 11:00PM
	23	Activated membership between 11:00PM and 12:00PM
first_order_day	1	Activated membership on the first day of the month
	2	Activated membership on the second day of the month
	3	Activated membership on the third day of the month
	4	Activated membership on the fourth day of the month
	5	Activated membership on the fifth day of the month
	6	Activated membership on the sixt day of the month

7	Activated membership on the seventh day of the month
8	Activated membership on the eight day of the month
9	Activated membership on the nineth day of the month
10	Activated membership on the tenth day of the month
11	Activated membership on the eleventh day of the month
12	Activated membership on the twelfth day of the month
13	Activated membership on the thirteenth day of the month
14	Activated membership on the fourteenth day of the month
15	Activated membership on the fifteenth day of the month
16	Activated membership on the sixteenth day of the month
17	Activated membership on the seventeenth day of the month
18	Activated membership on the eighteenth day of the month
19	Activated membership on the nineteenth day of the month
20	Activated membership on the twentieth day of the month
21	Activated membership on the twenty-first day of the month
22	Activated membership on the twenty-second day of the month
23	Activated membership on the twenty-third day of the month
24	Activated membership on the twenty-fourth day of the month
25	Activated membership on the twenty-fifth day of the month

	26	Activated membership on the twenty-sixth of the month
	27	Activated membership on the twenty-seventh day of the month
	28	Activated membership on the twenty-eighth day of the month
	29	Activated membership on the twenty-ninth day of the month
	30	Activated membership on the thirtieth day of the month
	31	Activated membership on the thirty-first day of the month
first_order_month	1	Activated membership in January
	2	Activated membership in February
	3	Activated membership in March
	4	Activated membership in April
	5	Activated membership in May
	6	Activated membership in June
	7	Activated membership in July
	8	Activated membership in August
	9	Activated membership in September
	10	Activated membership in October
	11	Activated membership in November
	12	Activated membership in December
first_order_year	2016	Activated membership in 2016
	2017	Activated membership in 2017
	2018	Activated membership in 2018

	2019	Activated membership in 2019
	2020	Activated membership in 2020
	2021	Activated membership in 2021
	2022	Activated membership in 2022
first_order_weekday	0	Monday
	1	Tuesday
	2	Wednesday
	3	Thursday
	4	Friday
	5	Saturday
	6	Sunday
promotion_code	0	Did not used a promotion code when creating the user account
	1	Used a promotion code when creating the user account

C Summary plots

C.1 Summary plot of SHAP values for Sleeping segment

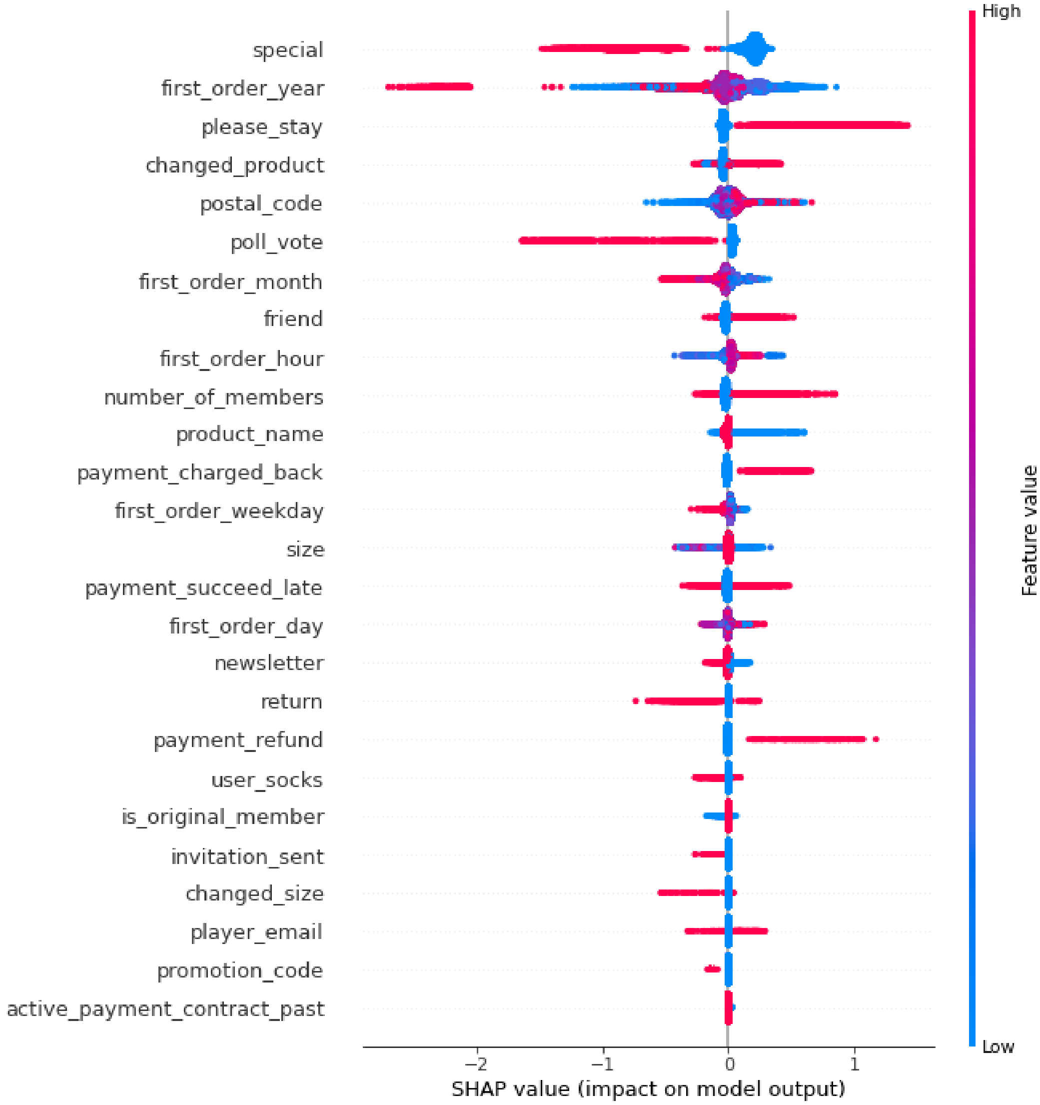

Figure 20: Summary plot of SHAP values for Sleeping segment

C.2 Summary plot of SHAP values for Potential segment

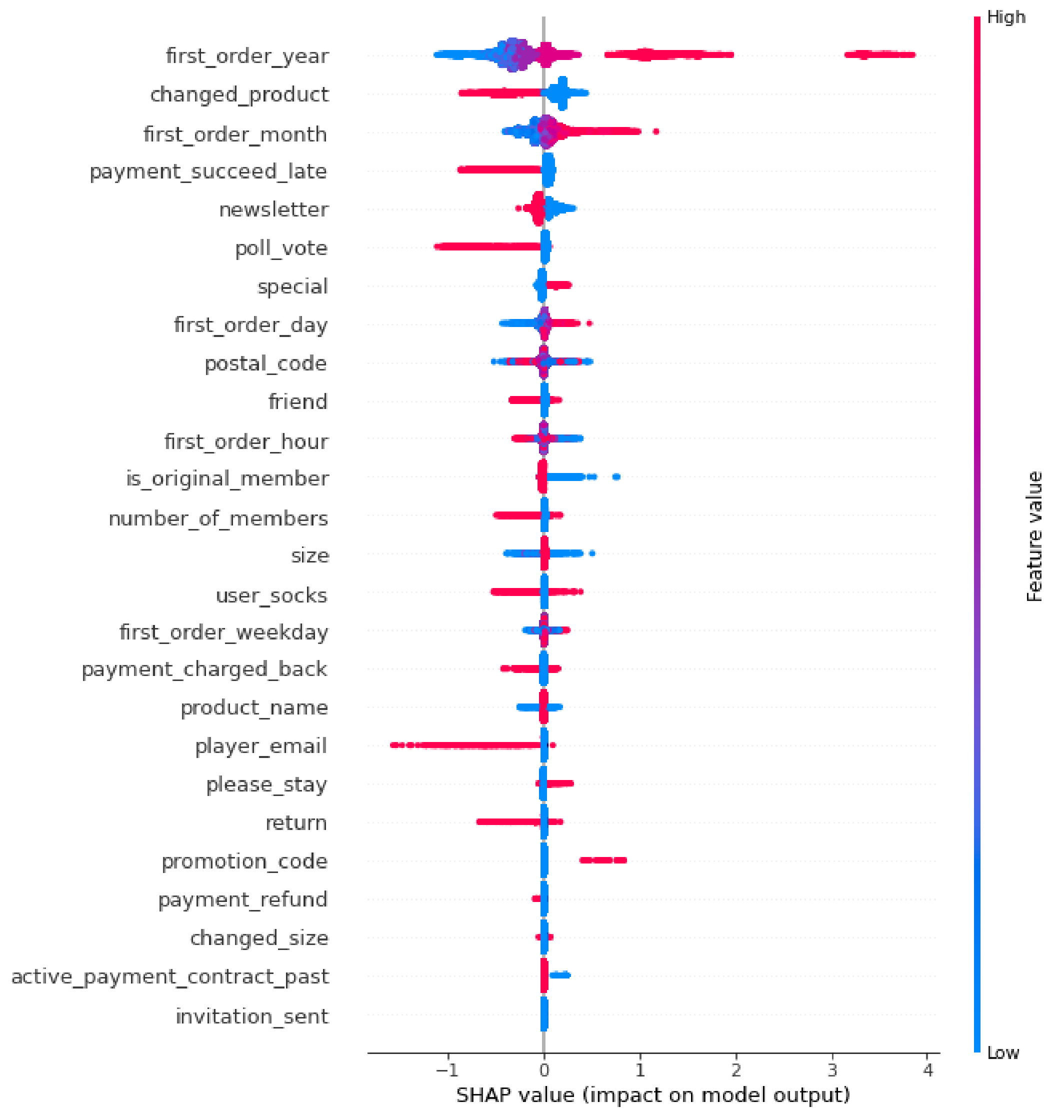

Figure 21: Summary plot of SHAP values for Potential segment

C.3 Summary plot of SHAP values for Premiere segment

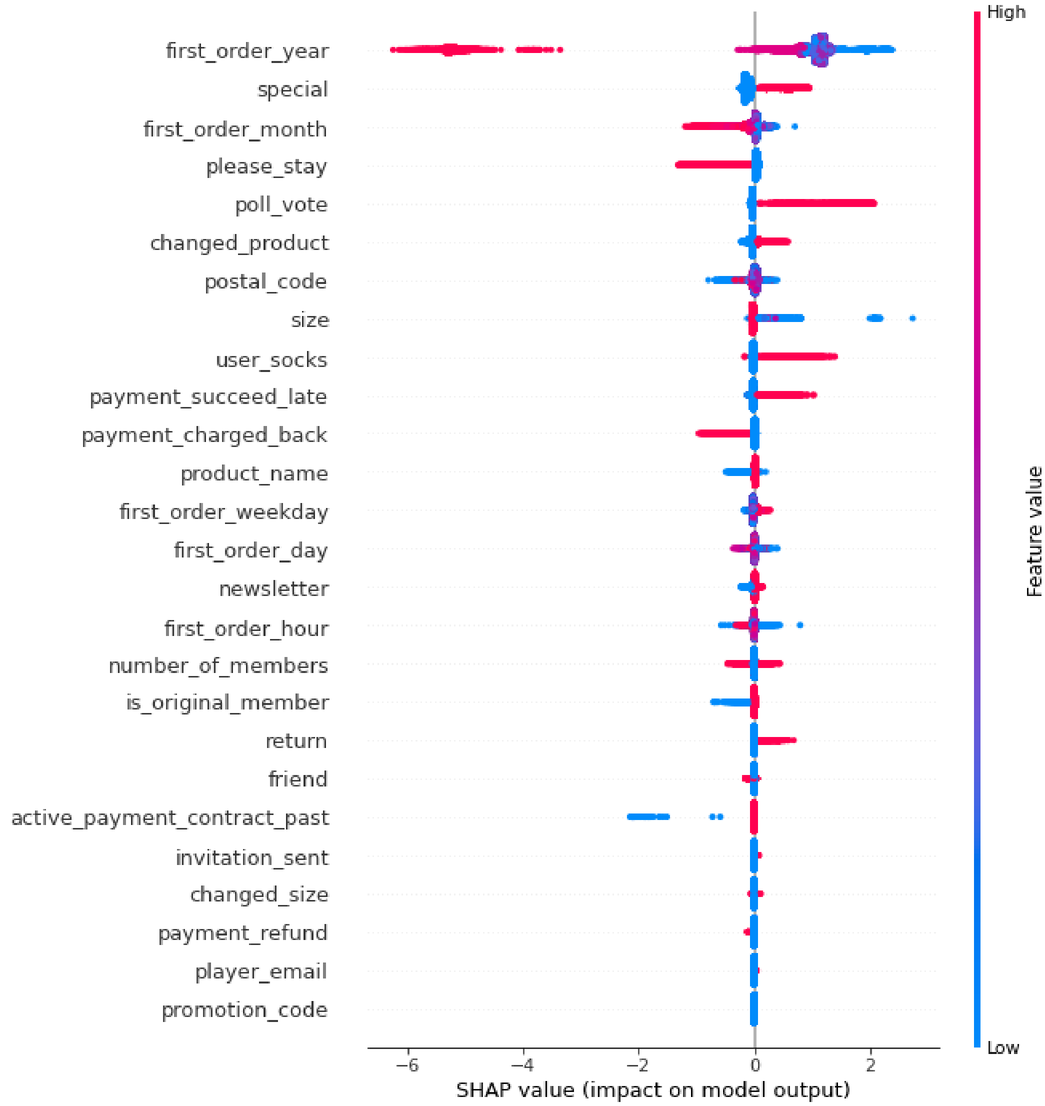

Figure 22: Summary plot of SHAP values for Premiere segment

D Dependence plots

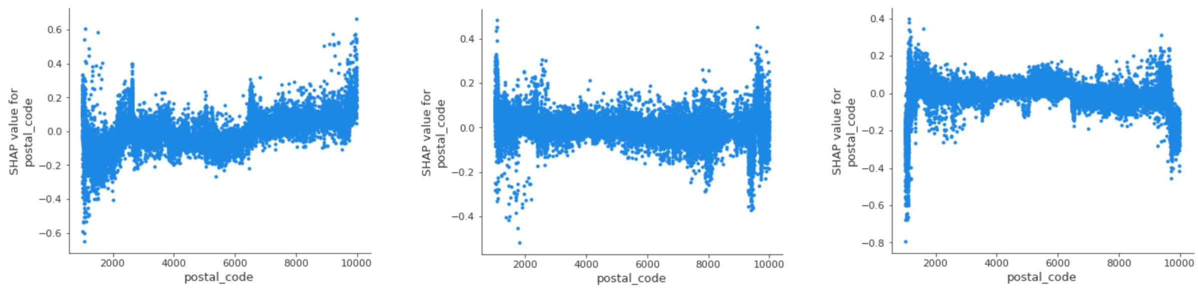

Figure 23: Dependence plots of *postal_code* for relatively the Sleeping, Potential, and Premiere segment.

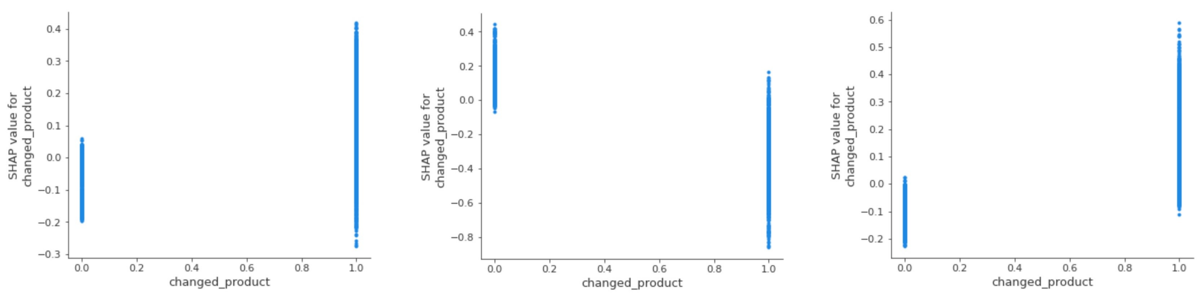

Figure 24: Dependence plots of *changed_product* for relatively the Sleeping, Potential, and Premiere segment.

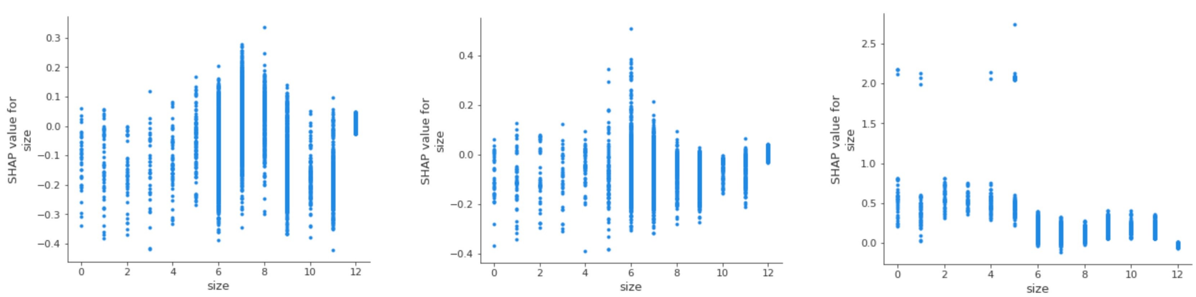

Figure 25: Dependence plots of *size* for relatively the Sleeping, Potential, and Premiere segment.

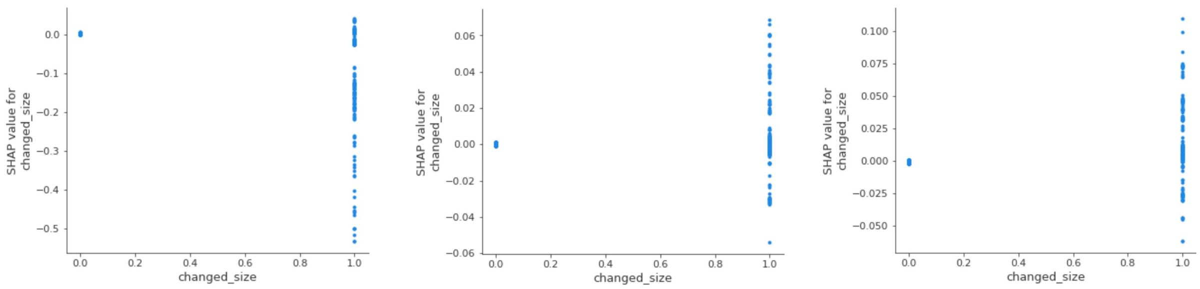

Figure 26: Dependence plots of *changed_size* for relatively the Sleeping, Potential, and Premiere segment.

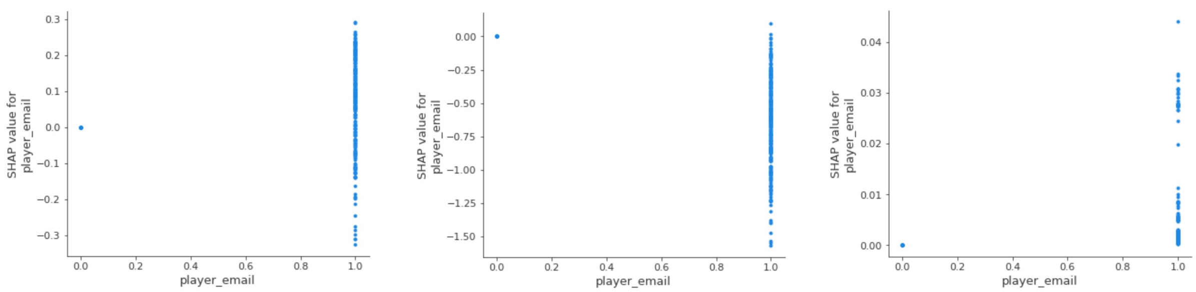

Figure 27: Dependence plots of *player_email* for relatively the Sleeping, Potential, and Premiere segment.

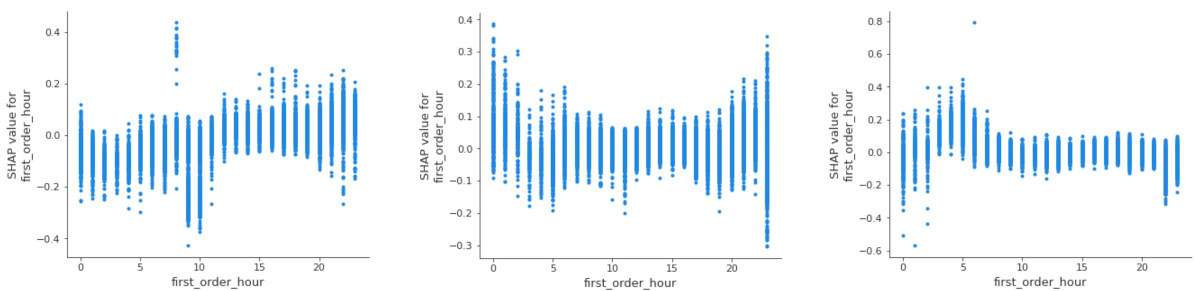

Figure 28: Dependence plots of *first_order_hour* for relatively the Sleeping, Potential, and Premiere segment.

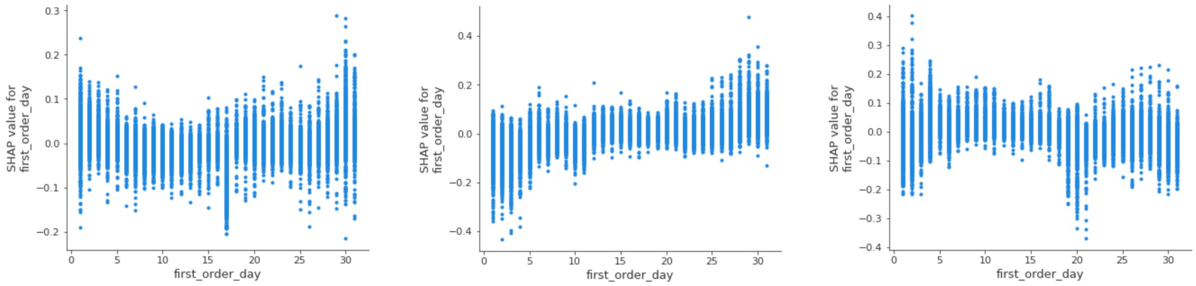

Figure 29: Dependence plots of *first_order_day* for relatively the Sleeping, Potential, and Premiere segment.

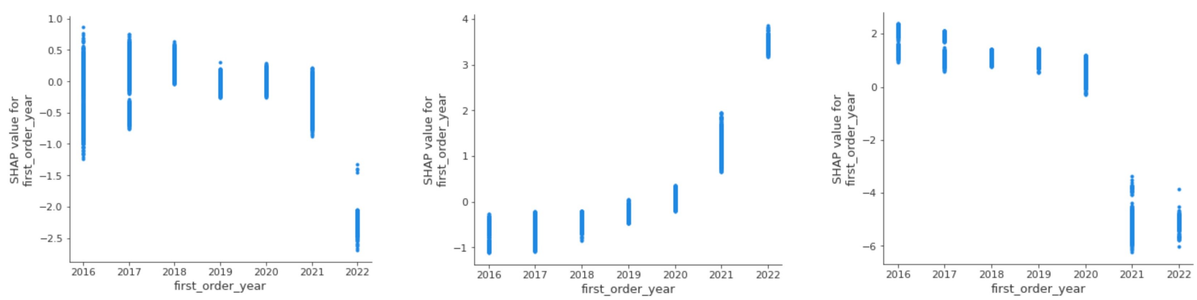

Figure 30: Dependence plots of *first_order_year* for relatively the Sleeping, Potential, and Premiere segment.

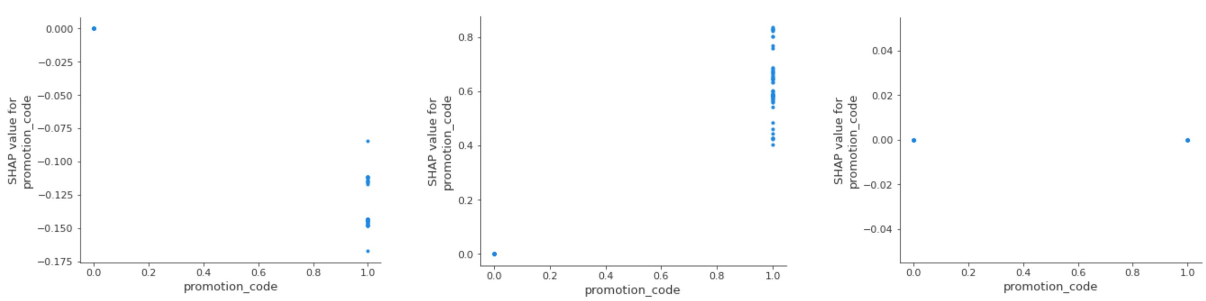

Figure 31: Dependence plots of *promotion_code* for relatively the Sleeping, Potential, and Premiere segment.