NOVA IMS Information Management School

MEGI

Mestrado em Estatística e Gestão de Informação
Master Program in Statistics and Information Management

Relation between client opinion (Net Promoter Score) and transactional data: A Practical Example in Retail at WORTEN

Intership Report Proposal

Guilherme Filipe Corgas da Cruz Esteves Paixão

Internship report proposal presented as partial requirement for obtaining the Master's degree in Statistics and Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

RELATION BETWEEN CLIENT OPINION (NET PROMOTER SCORE) AND TRANSACTIONAL DATA: A PRATICAL EXAMPLE IN RETAIL AT WORTEN

by

Guilherme Filipe Corgas da Cruz Esteves Paixão

Internship report proposal presented as partial requirement for obtaining the master's degree in Statistics and Information Management, with a specialization in Marketing Research and CRM.

Advisor: Prof. Teodora Szabo-Douat, Ph.D.

November of 2022

ABSTRACT

This professional internship took place at Worten, in Lisbon, with a duration of 9 months in the year 2021/2022 in order to apply and consolidate, in a practical context, the theoretical knowledge acquired in the 1st and 2nd semester of the Master with guidance and supervision, with the to complete the master's degree and gain experience in the area.

The main objective of this study was to try to understand customer behaviour considering their opinion given in the NPS (Net Promoter Score) process, trying to measure, classify and predict the customer's transactional behaviour in the company. Although this metric has been criticized by the academic community due to its poor predictive sales performance, NPS remains the most notorious metric in the market adopted by managers as a metric of consumer mindset. This internship report validates that NPS is a bad predictor of Sales in the long term, but a good predictor of frequency of purchase in the short term.

This report also emphasizes the significance of conducting a segmented and in-depth analysis of each business area in order to identify the areas that are harming the company the most and those that may have potential churners. Finally, this report offers a comprehensive view of the company and its relationship with the NPS metric.

Keywords: Customer loyalty; Brand health; Net Promoter Score; Marketing; Worten; Analytics

1. INTRODUCTION

Single-question customer metrics have been adopted by companies facing a growing competitive market, they intend to measure loyalty and focus on customer satisfaction. According to a study made by National Retail Federation in US, 418 executives from 137 retail companies say that customer satisfaction is their top priority (Geller, 2008). Of all the customer survey metrics an organization can use, there is one that attracts a lot of attention in terms of business results, the metric of net promoter score (NPS).

Fred Reichheld introduced the Net Promoter Score in 1993. Eventually in 2003, Fred collaborated with Bain & Company, which implemented the NPS to forecast client purchasing intention. NPS also gained a lot of notoriety due to the article published in the Harvard Business Review by Reichheld in 2003 stating that "You need only one question to determine the status" of a customer (Reichheld, 2003).

Rajasekaran (2018) also says that "With the help of the NPS, the company will be able to track recommendation rates for the service provided and estimated tools to identify the focus area by which will further help the company to improve the score" (Rajasekaran, 2018, p. 980).

Since 2003, the NPS metric has been adopted by most companies that aim to measure the loyalty of their customers and has become the most well-known loyalty metric in the market. With the attention received, this metric was put to the test by different academics who wanted to validate Reichheld's claims, such as, that NPS is "The best predictor of top-line growth" (Reichheld, 2003). Despite the fact that this strategy has been employed for many years and by numerous companies, Fisher (2019) states that "The NPS metric and the NPS system are still internally focused", he critiques NPS system because it "puts loyalty to the firm over a firm's loyalty to the customer". They use the costumer value management (CVM) as a better alternative, differentiating that "NPS is measuring what customers do for you" while "CVM is measuring what you do for customers" (Fisher, 2019).

It is important to note that this metric was created almost 20 years ago and the ability of companies to collect customer data has increased dramatically with the use of Big Data. However, 20 years ago most companies did not have access to loyalty metrics as they do today and, as a result, many of them did not base their decisions on data, in other words, they did not have a data-driven approach. This lack of knowledge on the part of companies made the NPS stand out, because many of the attempts to measure loyalty at the time were quite complex and difficult to analyse. For a manager, the choice between adopting a single-question survey used by almost the entire market or using a set of questions that proved ineffective was quite easy.

Something that also influenced the rapid growth of this metric in the market is the fact that it is quite simple for the customer to respond and understand. In fact, Reichheld was right when he said that some attempts to measure loyalty using surveys may have been too complex for those taking the questionnaires. However, managers today wonder why they employ this metric and whether the costs associated with NPS are repaid.

The goal of this paper is to answer this question providing a useful manual for businesses on how to use this metric by demonstrating its potential, highlighting its advantages, and harmonizing opinion with transactional data.

2. NET PROMOTER SCORE (LITERATURE REVIEW)

Reichheld sold his method on the basis of a critique of the orthodox way of doing research: “They t end to be long and complicated, producing low response rates and ambiguous implications that are difficult for operational managers to act on” (Reichheld, 2003), concluding that the results usually do not correlate with reality. His solution was much easier and according to his studies the results were much more promising.

According to a study conducted by Reichheld, which involved six industries, where the purpose of the study was to determine which research questions had the strongest statistical correlation with consumer behaviour, there was one question that was better for the majority of industries: On a scale of 0 to 10, "How likely is it that you would recommend [company X] to a friend or colleague?” that came in first or second place in 11 out of the 14 case studies (Reichheld, 2003). If a customer responded with a score of 10 or 9, they were categorized as promoters, if they responded with a score of 8 or 7, they were categorized as passive, and if they responded with a score of 6 to 1, they were categorized as detractors.

According to Reichheld, NPS can accurately predict sales growth; nonetheless, this metric is more widely accepted as a measure of customer intent rather than sales forecasting. NPS is basically a metric that tries to measure market word-of-mouth, this is due to the question format.

Keiningham et al. (2007) agreed that there is a positive link between NPS and Word of Mouth, however they were quite sceptical of the link between word-of-mouth and sales growth, stating that “there is no peerreviewed research that longitudinally examines the relationship between word-of-mouth activity and firm-level financial outcomes” (Keiningham, Cooil, Andreassen, & Aksoy, 2007).

Supporting the study by Keiningham et al. (2007), Grisaffe (2007) analysed in detail the NPS metric based on a social science perspective, where they concluded that this metric could deceive marketing managers with false information. This is because the NPS is limited to a single question, Grisaffe also adds that recommendations alone cannot lead a company to success.

Keiningham et al. (2007) presented a study in which it’s also noted that using a single measure may not be enough to assess customer loyalty effectively, suggesting that a multi-metric is a better alternative. The same conclusion was made by Pollack and Alexandrov (2013) in a study that intended to assess the ability of the NPS to measure consumer loyalty, the results did not support Reichheld's thesis that NPS is the best metric to measure customer loyalty. They also added that the use of multi-metrics is a more viable strategy and yields better outcomes when forecasting loyalty behaviours.

Zaki et al. (2016) presented another longitudinal study in which the NPS is examined and compared to other tools, particularly big data, demonstrating that the NPS falls short on a large scale when measuring customer loyalty to the company.

Baehre et al. (2022) also confirmed the study by Keiningham et al. (2007) that NPS is not a good loyalty metric, in a longitudinal study in the U.S. sportswear industry.

Kristensen and Eskildsen (2011) also managed to demonstrate that the NPS is a poor predictor of consumer loyalty, however, this study used satisfaction surveys from one in the insurance industry in Denmark. The fact that it is a unique business model in a single country may influence the results of the study.

Despite some criticism of this metric, Haan et al. (2015), argues that contrary to the study by Keiningham et al. (2007), NPS is one of the best indicators of customer retention. After a study of customers from 93 companies in 18 industries, they concluded that NPS is a strong predictor of customer retention.

Mecredy et al. (2018) also found a relationship between NPS and an increase in customer spending, in a study of almost 5 years where it is perceived that those with a better NPS level spend more money in the company in the year following the survey.

While NPS has some notoriety in the current market, the academic community is divided on NPS's ability to measure short-term customer intent. Something that Reichheld also mentioned is that the NPS is capable of predicting future sales, in this case the academic community agrees that the NPS is not the best predictor of future sales, at least in the long term.

Keiningham et al. (2007), in a study with 5 different types of industry, mostly banking, state that the Reichheld method does not produce good results as a sales predictor. Instead, NPS shows no evidence of being a good predictor of customer growth. Concluding that many managers may need to reconsider some of the metrics they use.

Although some studies by Pingitore et al. (2007) and Van Doorn (2013) demonstrate that this measure can predict growth in sales, these authors are skeptical that the NPS is the best measure for this purpose. For instance, they named their study "The Single Question Trap," while Doorn et al. (2013, p. 317) stated that "the predictive ability of customer metrics such as NPS for future sales growth... is limited."

Finally, Baehre et al. (2022) concluded that the NPS is a good predictor of short-term sales, but only under certain conditions. These conditions are associated with the industry in question, such as short buying cycles.

After carrying out some research on NPS, it may be concluded that a substantial proportion of the academic community doubts Reichheld's claims. However, the vast majority agree that the NPS is a useful tool that enables to understand how each company is being discussed in the market, making it a reliable source of information about word-of-mouth behaviour. Additionally, the community as a whole agrees that because NPS can only measure recommendations, and for that reason it is a poor predictor of loyalty. In a response to that problem, other studies use a variety of metrics to measure loyalty rather than relying solely on one.

However, Reichheld's claims regarding the short-term customer intention and the NPS's capacity to be an accurate sales forecast in short amounts of time are still the subject of much debate in the scientific community.

3. CONCEPTUAL FRAMEWORK

As was mentioned above, businesses are placing an increasing emphasis on customer satisfaction with their products and brands. The explanation is simple, consumers who spend more money and make more frequent purchases are also the ones who are most satisfied with brands.

Although it might seem obvious, businesses need to know if their consumers are happy regardless of the fact that many of them are reluctant to express it. The NPS metric comes into play here. A corporation will never be able to comprehend what is going on in the minds of its customers if it does not use a loyalty metric. Although a loyalty measure cannot tell you how your customers are feeling, it can give you a general notion of their thoughts.

After gaining an understanding of the customer's feelings, it is still necessary to determine whether the opinion and transactional data are related. Depending on that answer, you can then proceed to predict transactional behaviour taking into account your opinion.

The purpose of this study is to show how businesses can use NPS data and combine it with transactional customer data to predict future behaviour. But above all, it aims to operationalize such opinion data in order to add value based on the customer's NPS, using a real-world example from my time at Worten. This report has a statistically relevant component in addition to having results that are statistically significant for the research topic because this type of cross-sectional analysis has never been done before.

This study aims to assist businesses in making the most of this measure by providing information on which business areas are having the greatest impact on the company as well as which areas are losing the most money as a result of consumer feedback. It will be possible to forecast which business areas will eventually lose more money or experience a lower purchase frequency once a relationship between NPS and customer transactional behaviour is discovered.

This is only possible if in each business area there is a significant sample, more than 50 questionnaires answered, and there must be an interval before and after the day the NPS survey was answered. This interval must coincide with the time period specified for a Churn customer, in the case of Worten, a customer who has Churned is a customer who has not bought once during the space of approximately one year. “Customer churn (also known as customer attrition) refers to when a customer … ceases his or her relationship with a company … businesses typically treat a customer as churned once a particular amount of time has elapsed since the customer’s last interaction with the site or service” (optimove, 2022).

The relationship between opinion and transactional behaviour can only be measured in this way.

Finally, it is necessary to ask how much money a negative or positive opinion of the company will cost. To properly respond to this question, it is essential to assess the NPS's ability to measure customer intention as well as its capability to forecast future customer transactional behaviour. This brings us to the research issue of this report: “Is the Net Promoter Score a good measure of loyalty and a good predictor of short-term transactional behaviour?”

In in order to respond to the research question, several hypotheses were developed in the hope of obtaining clear answers to the issue at hand.

The first topic that will be addressed is the NPS customer segmentation capabilities, according to Kristensen and Eskildsen (2012) the use of three clusters does not make sense, they said that this method can generate some loss of information in the process and provide misleading information about some customers. Mid-range rating is one cause, as a customer who wants to rate an experience as average can end up being a detractor. The study by Kristensen and Eskildsen (2012) noted that this factor also heavily depends on cultural differences, which is contradictory given that the NPS was developed and tested in the United States, whereas the study by Kristensen and Eskildsen in 2012 was conducted in Denmark, two nations with substantial cultural distinctions (Eskildsen et al., 2010).

A crucial perspective is provided by Keiningham et al. (2007), who present a variety of literature on cultural difference and how they can impact consumer behaviour. Greenleaf (1992) manages to perceive that there are different response styles and the presence of bias on the part of some consumers, this same finding is shared by Varki and Rust (1997). To determine whether there was a substantial difference between cultural differences and survey responses, more research was conducted, and it is a similar finding across all studies that varied response patterns are connected to cultural differences (Vandenberg 2002; Wong, Rindfleisch, and Burroughs 2003)

Fisher also says that “there is no such thing as a passive client” (Fisher, 2019) stating that the passive cluster is a bad cluster. Kristensen and Eskildsen (2012) present some solutions to this problem with different ratings that proved to be very effective, also concluding that NPS “is a very poor predictor of customer loyalty”. The provided solutions, however, are not accessible to Portugal.

Because this metric was introduced to the market some time ago, organizations like Worten were able to refine it and expand its potential to meet their demands. To solve the problem of NPS clustering, Worten introduced an additional category of NPS called a Super Detractor (defined as a customer who rates their overall experience a 1 or 2). Later on, we will see that this cluster will help a lot in identifying "upset" customers with the brand.

After some criticism of the NPS metric for its ability to cluster customers taking into account their loyalty to the brand, it is now the turn to help the NPS in this regard and adapt the metric to the company's needs. As we saw above, the super detractors were a case where Worten adapted to the metric.

So how can NPS be complemented in this case?

The K-means algorithm is used as a complementary tool for cluster segmentation to address this issue, the K-means clustering model divides data points into groups and uses an algorithm that maximizes inter-cluster distances while minimizing intra-cluster distances (Lloyd, 1982).

A K-means analysis before and after the process associated with the purchase allows us to validate the NPS segments, for example, if the customer is in a very high frequency and value cluster, but after the NPS survey moves to a lower frequency and value cluster then, according to Reicheld's claims that customer was a detractor at the time of the NPS inquiry.

The first hypotheses were built on this notion:

H1: "The K-means analysis together with NPS allows us to verify if there is a relationship between purchase frequency and value and NPS surveys."

K-means can be used to segment customers based on their frequency (it is not necessary to segment by value because frequency and value are highly correlated), and NPS may be used to segment customers based on their opinions.

Given that the K-means cluster helps to segment customers based on frequency and value and the NPS helps to segment customers taking into account their opinions, the existence of a relationship between the two denotes a relationship between Worten opinion data and transactional data. We hope that it will be able to determine "what the relationship between transactional data and opinion is" in this way.

A different method of clustering customers is to assess the recency, frequency, and monetary spend (RFM) of your customers. RFM was introduced by Cullinan in 1997, but only later was it considered a significant factor in predicting customer lifetime value (CLV) as well as customer behaviour and churn (Ballings, Poel, & Verhagen, 2012).

This analysis will provide customers ratings on a scale of 1 to 5 (with 1 being the lowest rating and 5 being the best) based on their frequency, recency, and money spent. This tool is a great way to determine whether your customers' purchasing habits have changed as a result of their interactions with the business. This leads to the second hypothesis:

H2: "RFM is a better method than NPS for gauging short-term customer loyalty."

This analysis will also make it possible to create a control group more effectively, giving us a better picture of the characteristics of respondents to the questionnaires. Which leads to the third hypothesis:

H3: “Customers who answer the NPS questionnaires are more loyal to the brand.”

With the creation of a control group, this report stands out as it is one of the few to create a control group for this purpose. With the results obtained from the control group, it will be possible to understand which are the business areas that have more relationship between opinion and transactional behaviour.

Another element we need to assess is the likelihood that certain businesses won't be able to access an advanced statistical component. As a result, this study will provide a reasonably accessible and effective method for determining the influence of a customer's opinion in an NPS survey. How many customers are lost as a result of an NPS evaluation will be addressed by a churn analysis.

H4: “Churn analysis enables companies to determine whether there is a relationship between customer transactional behaviour and NPS.”

In this report, a customer who churns is defined as someone who doesn't make purchases 365 days after their last purchase. The time window can be adapted to the needs of the company, for example, where purchases occur less frequently, it is wise to designate a shorter time frame.

After a more descriptive analysis of the data, it’s now needed to validate Reichheld's arguments that NPS is a good sales predictor, and this is how it will be accessed its predictive capacity, using logistic regression.

There are many models and algorithms that can be used in conjunction with NPS to predict customer behaviour, but the main goal of this report will be to show that NPS is a good indicator of customer intention over the short term. To do this, it will be examined for its capacity to predict lost customers (churners) after one year of responding to an NPS survey. This is how a logistic regression will assist in comprehending the significance, magnitude, and impact of the NPS variable in relation to short-term customer intent (churn). This brings us to the final hypothesis:

H5: “NPS accurately predicts the client's future short-term transactional behaviour.”

4. METHODOLOGY

The first section of this report focused on identifying which techniques might be used to supplement the NPS in order to carry out a better analysis. After reviewing the NPS, the second part will focus on combining customer input with client data that Worten has acquired from January 1, 2021, through August 16, 2022. In the third section, the use of RFM models will help to create scores for each customer based on their frequency, amount, and financial expenditures with the company. A k-means cluster analysis will also be performed in order to understand whether there is a link between NPS and transactional behaviour.

After the data processing, it is necessary to validate the results using significance tests in order to evaluate any type of dependencies between the selected variables. On this basis, it is possible to confirm using statistical validity if there are dependencies or not. Finally, the use of big data techniques is essential for predicting future customer behaviour.

It is essential to understand the steps that each consumer took in order to better understand their behaviour. Customers who submitted an inquiry related to a Marketplace purchase, for instance, would always have a different experience from those who went through the repair process. To prevent the generalisation of experiences as being equivalent, it was necessary to divide experiences into processes. This technique is very useful for companies not only because they can have access to more reliable data related to only one process, but also to have a more holistic view of the data.

The data were combined in order to identify the factors that led to the customer's opinion, and a time period of one year before and one year after the opinion survey was established. With this straightforward action, it will be possible to quickly distinguish between churners and new or recovered clients. It is obvious that in this instance, we define churning customers as those who haven't made a purchase in a year, but it still makes sense to align this definition with the business you use.

This analysis broken down by time periods is essential to compare the relationship between opinion and customer transactional data. In the case of some processes in which the opinion is triggered by a purchase, such as the Marketplace, this analysis allows us to identify where the purchase was made, when it was made and, above all, to analyse which business areas create the worst experiences (detractors) and which create the best experiences (promoters). However, Worten understood that it was critical to include another metric to understand customer satisfaction with the purchase, not just with the company, as the NPS metric measures the customer's perception of an organization. This measurement is very much in line with the academic community's recommendation to employ a multi-metric approach (Keiningham et al. 2007; Pollack and Alexandrov 2013). To measure customer satisfaction with purchases, the NSS (Net Satisfaction Score) metric was developed, which results from a simple but effective question of asking customers how satisfied they are with their purchase. For this specific analysis, the NPS metric will be used since the objective of this study focuses on the evaluation of the NPS metric.

The first segmentation analysis performed was the RFM analysis, with 5 being the highest score and 1 the lowest for each RFM class. It was discussed whether recency needed to be eliminated at the beginning, as it could lead to misleading conclusions, since the analysis was done in a limited period of time. This issue was resolved by using SAS code to generate recency. We can determine recency if we know what the customer's most recent purchase was before the date we want to know.

During the recency generation process, the idea arose of developing an RFM segmentation (independent RFM binning) for each previously defined time period (12 months before the NPS process and 12 months later). An independent RFM binning analysis is simple to understand, ratings are assigned to each score as explained earlier, however the nested binning is calculated a little differently. The algorithm assigns a rating to recency and after 5 cells representing the scores are created for the recency class, only then does it start to analyse the frequency. The difference in this method comes when frequency bins are calculated for each recency score and after 25 frequency bins are calculated, the algorithm will create 125 more by ending the analysis. In this case, as the data were not enough for 125 clusters, it was decided that the independent method was the most favourable for the sample we had. In this way, each RFM class will also be analysed in more detail, for example, there may be a greater correlation when we talk about frequency and NPS.

Then a cluster analysis was performed using k-means and customer clusters were created and classified into good, bad and average. A cluster analysis needs input variables to perform customer segmentation. And in this case, as the purpose of cluster analysis is to segment customers according to their transactional data, the variables most indicated in this specific case are recency, frequency and amount spent. In this way, you can save time and effort by using the frequency and value variables used for RFM analysis as well.

However, a problem arose, and it was one that had been anticipated. K-means is extremely sensitive to outliers, in a sample research an observation that deviates significantly from the others is considered an outlier (Ghosh & Vogt, 2012), resulting in clusters with very few customers and dispersed in terms of frequency and value, as can be seen in photo 1 (Scatter plot of all customers on the X axis represents the frequency in the year prior to the survey and the y axis the amount spent in the year before the survey) and in photo 2 (Scatter plot of all customers on the X axis represents the frequency in the year after the survey and the y axis the amount spent in the year after the survey).

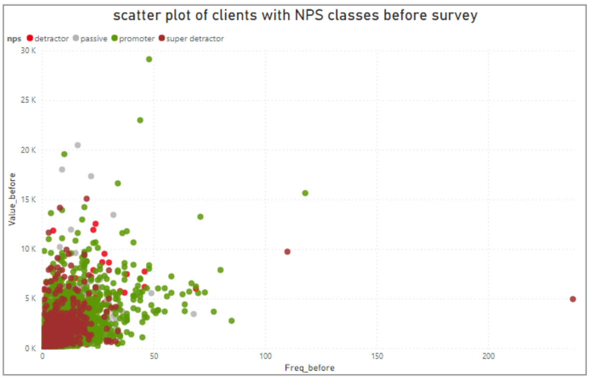


Figure 1

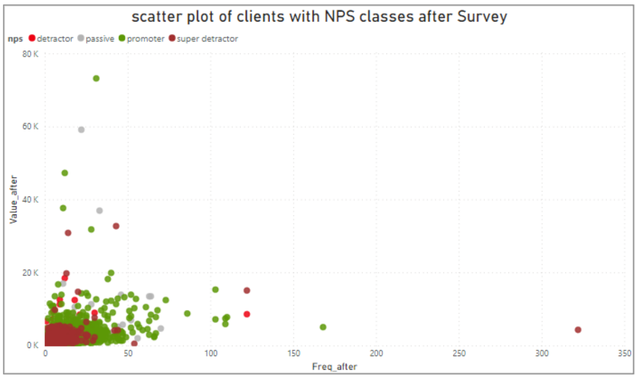


Figure 2

The truth is that, in this case, the k-means failure proved to be quite advantageous, as this analysis allowed us to determine which were the best clients. However, for this study, it was important to include all customers in the same analysis, and the RFM variables were transformed using a logarithmic function as a solution for outliers. The logarithmic function makes working with very large numbers easier, converting them into a smaller and more understandable version, not only for users but also for the K-means algorithm that no longer finds exceptional cases in its data source. With this transformation, it was possible to create clusters without loss of information, which is also a factor to take into account.

In order to understand the impact that the use of the logarithmic function has, two graphs were created (figures 3 and 4) that allow us to visualize customers in a point cloud, where the X axis represents the frequency that the customer spent before/after of the NPS survey and the Y axis represents the amount the customer spent before/after the NPS survey.

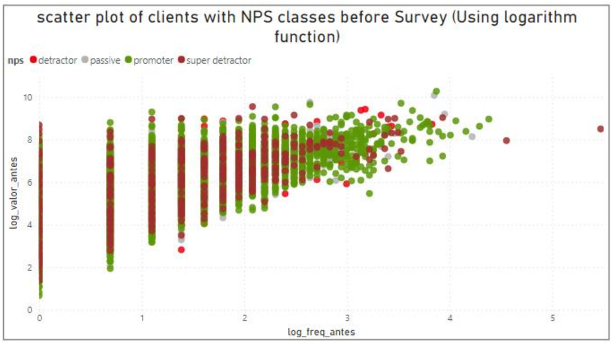


Figure 3

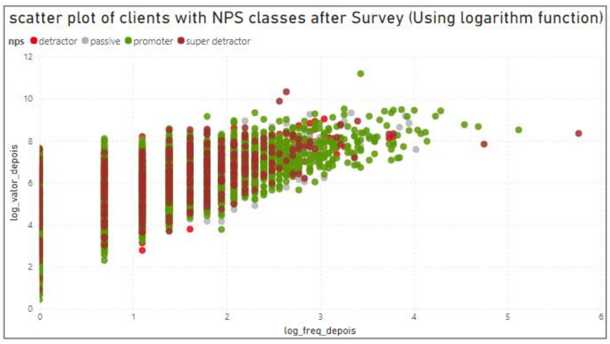


Figure 4

It is worth noting that after employing logarithmic variables to create clusters, the altered variables will have extremely low values as a result of the transformation. To solve this problem it is necessary to use the exponential function, which is the inverse of the logarithmic function, to recover the results of the original variables.

A key topic in this report will address the ability to form a control group to recognize the impact of customer feedback on transactional data, and to get to know better the customer who responds to the surveys.

After the control group's analyses are complete, it will be necessary to explore the data more thoroughly and try to identify patterns or relationships between the client's opinion and their transactional data. In this way, some statistical analyses will be created, namely an analysis of Pearson's correlation coefficient and an analysis of variance (ANOVA). Since we will be able to have both transactional data before the NPS survey and after the survey, it is possible to have a better understanding of the relationship between opinion and transactional data.

After having an exploratory analysis of the data collected, it is necessary to evaluate a topic that is much discussed in the academic community, and where there is much debate, the ability of the NPS to access customer loyalty. To this end, a logistic regression model will be created that will aim to predict the ability of the NPS and the RFM analysis to predict future behaviour (churn).

5. FINAL RESULTS

5.1. DESCRIPTIVE:

Regarding the overall analysis, because customers had to cross-reference their email addresses to acquire transactional data, only 23336 customers who responded to the email questionnaires were considered for this analysis. Only e-mail responses were used in this study, because, according to Scheuren (2004), people who are submitted to a questionnaire provide more truthful answers to e-mail surveys.

Product sales processes, such as Marketplace, received the largest number of customers, which is understandable considering that it is Worten's primary industry. However, it is necessary to keep in mind that some experiences that are not related to sales can have a great influence on the customer's intention.

After the creation of the K-means cluster, it was found that the clusters of the period before and after the survey changed accordingly to their survey process, with the Average cluster having the largest number of customers and the Bad cluster the smallest number of customers.

Regarding the type of customers who respond to the surveys, it was found that the customers who responded to the questionnaires are, in fact, customers with monetary values and purchase frequency higher than those in the database of all customers, with more customers with low value in the general database than there are in the customer opinion base.

This means that the customer who responded to the corporate surveys tend to be better customers, this finding is very important for Worten, and creates the need to create a control group, to validate these findings, something that we will do in the next phase.

5.2. CONTROL GROUP

The creation of the control group (CG) is something important in this topic, this is because this is the only study in which a CG is created based on customer frequency and value to understand the relationship between opinion and transactional behaviour. For a customer to be in the CG it needs never to have had an approach with the company or that the company has had an approach with the customer, in this way all customers of the target group (TG) are excluded from the base that will serve to create the CG, as well as all customers who have already made a complaint, etc. Due to the time constraints, it was not possible to conduct a more thorough investigation of Worten's digital and social media channels, including in-store customer service and social media comments, for this research. However, the data we have is sufficient to conduct a meaningful analysis.

After excluding any customers who had interactions with Worten other than during the purchase process, we needed to determine the time frame for those transactions. In the case of the CG, customer purchases must be made starting in 2020 (one year before the date of the first NPS responses) and continuing through July 27, 2022. This time frame must be the same for the control group.

The next step is to assign fictitious dates for a survey, because in the TG it is analysed one year before and one year after the opinion survey and in order to have a comparison term it is necessary to create a fictitious date. For this, after some discussion with colleagues, it was concluded that the best dates to assign to CG clients are the same dates as the TG surveys.

After gathering all the TG dates, a simple random probabilistic sampling is carried out (in the SAS program the Ranuni function was used) which allows assigning a fictitious date to each of the CG clients. With the fictitious inquiry's data, we need to analyse the period one year before and after, measuring the frequency of their purchases and the amount they spent during those times.

Because we need to compare clients with identical frequency and value in order for the comparison to be fair, and because an RFM analysis was performed in the GA, we now need to use it for our benefit. To understand the impact of the opinion, we need to use the RFM analysis of the period before and after the survey answered and then see the differences between the GA and the CG.

The process of matching customers with equal frequencies and amount spent starts by assigning each RFM class a number of customers, for example in GA the RFM class 111 contains 1215 customers and class 555 contains 580 customers, now it is necessary to assign customers from the GC to the already created RFM classes.

In order for RFM classes to be comparable across groups, the CG has to comply with the limits of each TG RFM class, for example, a TG customer that is within RFM class 2 had to comply with the limits of that same class. class. Frequency class 2, in this specific case, has a minimum limit of one purchase, during the year, and a maximum limit of two purchases, so all CG customers to enter frequency class 2 need to comply with this limit. The same applies to the other frequency, value and recency classes.

In the case of Worten, what happened was that there are RFM classes, such as 251, which only had 6 customers and the CG did not have enough customers to match that number. For the analysis to be fair it had to be subjected to a loss of information, in the case of the RFM 251 class, three TG clients had to be left behind for the analysis to be fair. The same is likely to happen with many companies.

5.3. CONTROL GROUP (BASELINE)

The control group will make it possible to compare the transactional behaviour of customers who go through a purchase process at Worten and customers who go through a purchase process at Worten and give their opinion in the NPS surveys. The results of this analysis will make it possible to understand the profile of customers who respond to surveys, because two groups with the same characteristics will be analysed in the same period of time and it will be understood what happens after the trigger (NPS survey) is activated. However, it will not be possible to understand whether these customers are Worten's best customers or not. For this purpose, it is necessary to create a group of customers that is representative of Worten's base, using simple random probabilistic sampling. In this study, the group was called the Baseline Group.

Unlike the Control Group, in the Baseline group it is necessary to match groups of customers with the same frequency and value class, because we do not want to compare customers with the same value and frequency characteristics. We just need to get a representative group of all Worten customers. However, it is necessary to go through the date creation process (the same as the CG), so that the comparison has the same time intervals.

This baseline control group will allow us to understand a more detailed notion of each process through which the customer goes through, for example in the images (5, 6, 7 and 8) we can see that a customer who goes through a process of installing a product in home and a customer who goes through a process of buying Worten products is quite different.

Not only can you see the type of customer they were before the survey, but also the impact of the opinion on the customer's purchase frequency after going through a process and giving their opinion respectively in that process.

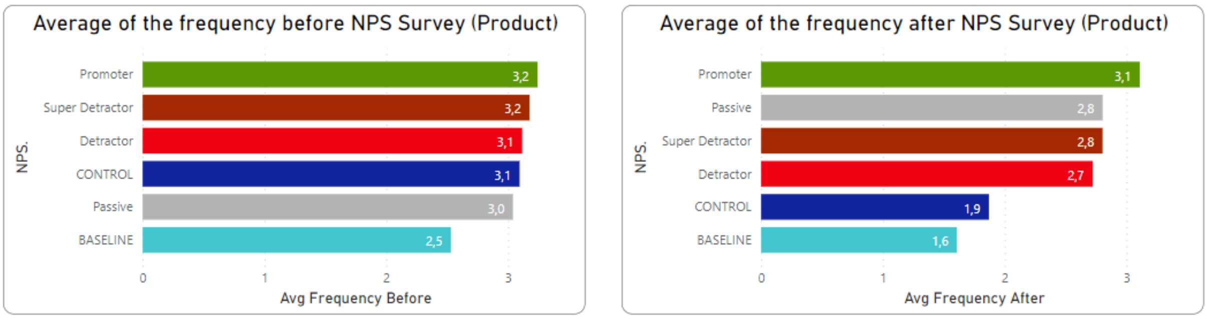


Figure 5

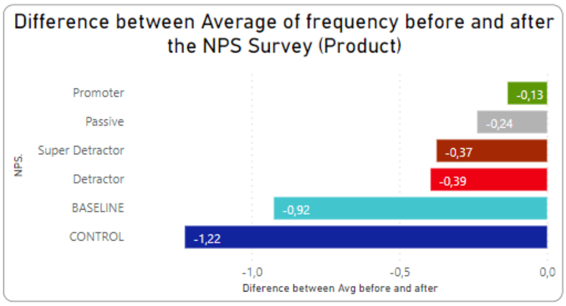


Figure 6

In the purchase process for Worten products, we can see that the type of customer that went through it made an average of three purchases, whereas the remainder of the Worten customer base made an average of only 2.5 purchases. We can see from this that a customer who responds to inquiries about the purchasing procedure for Worten products is a loyal customer and makes purchases more frequently than Worten's average customer.

When analyzing the year after the purchase process, we noticed a correlation between opinion and purchase frequency. The frequency of the client who has a positive experience with Worten stabilises one year after the completion of their treatment, but the frequency of the client who has a negative experience decreases. The graphs of the control groups were greatly unexpected because both significantly reduce their frequency, which leads us to believe that the customers are loyal customers who continue to make purchases regardless of their opinions.

Let's now understand the impact that an analysis by process has, with the next graphs.

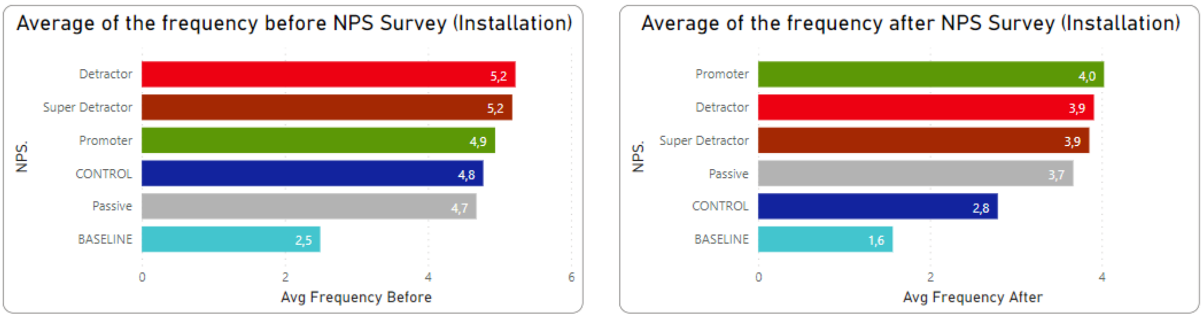


Figure 7

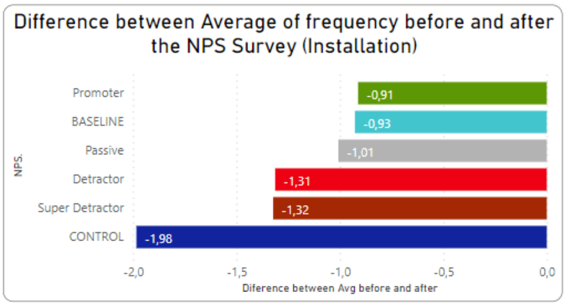


Figure 8

The installation procedure has a much more direct correlation between opinion and frequency of purchase, as shown in the figure 8 graph. For customers who had a poor experience, there was a difference in frequency before and after the procedure of 1.31 purchases, whereas for customers who had a good experience, there was only a difference of 0.91 purchases. The significant difference becomes apparent when the baseline group is analysed, and it is discovered that the group's median difference is quite similar to that of the promoters' clients. This analysis informs us that Worten's detractor and super detractor customers are below average, which is extremely worrying.

However, what surprised us was the Control group, as it allowed us to reject the null hypothesis (H3) that the control group can show that customers who respond to surveys are more loyal. However, by investigating more on this topic, we were able to create a baseline control group and reach much more enlightening conclusions about the importance of separating NPS analysis by processes.

One thing that stands out is the behaviour of the control group, which is without a doubt a group who behaves abnormally in relation to the target group and the baseline group. These outcomes are not isolated cases because this behaviour does apply to all processes. After much deliberation, it became clear that there is something peculiar about the client who responds to NPS questions, namely the expectation.

Customers who have very high expectations are customers who have been buying from the brand for some time and therefore it is necessary to realize that their expectations are different from customers who buy for the first time.

Expectation determines whether or not a person responds to a questionnaire, in the case of customers in the control group, the impact of the experience may have been so bad that they did not want to answer, or to get more involved with the company.

A Customer life cycle (CLC) analysis was carried out for customers from both the control group and the target group and what was discovered was that customers who have been shopping for more than three years at Worten occupy almost the same percentage of customers. In the control group, 53% of this group's base are customers who have been buying for more than three years. In the target group, customers who have been shopping at Worten for more than three years occupy 55% of this group's base. However, when we evaluate the Churn percentage, there is a difference of almost 10%, with customers in the control group who have been shopping at Worten for more than 3 years have a churn rate of 27% while customers in the target group who shop at Worten for more than 3 years have a churn rate of 19%. Among CLC classes, the group of customers who have made purchases at Worten for more than 3 years and the group of customers who have made purchases at Worten for 2 years are the groups where the most difference in the Churn rate is noticed, while new and recovered customers have a very similar Churn rate.

This may not be a clear answer to the question of why customers in the control group have a higher churn rate, but it is an indication that cannot be ignored.

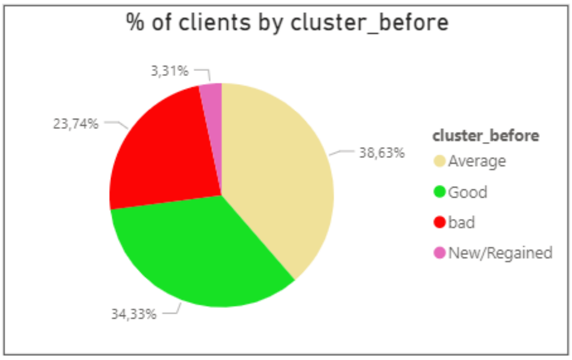
There is also the possibility that customers who do not respond to the questionnaires have no connection with the brand, “in general, participants tend to have a slightly favorable attitude toward the company or product for both unpaid and paid surveys” (Sauro, 2019)

This connection with the brand can also suggest that the customers who answer the questionnaires are the ones who feel valued, these “respondents believe their opinions are valued and that their answers will be put to good use” (Mayfield, 2013)

These cases end up validating that the customers who answer the questionnaires are the best customers, because they are the customers who continue to be loyal to the brand.

5.4. SEGMENTATION ANALYSIS:

After having a general understanding of the data and realizing how crucial it is to construct a control group and divide the analyses into distinct business processes, it is now required to add statistical bases to support the findings and confirm the original hypothesis. This enables a more thorough examination of the connection between NPS and transactional behaviour. The first hypothesis says that K-means clusters help to understand whether there are relationships between purchase frequency and value with NPS surveys. To make it easier to visualize the data, two pie charts were created to understand how the database customers are distributed by clusters.



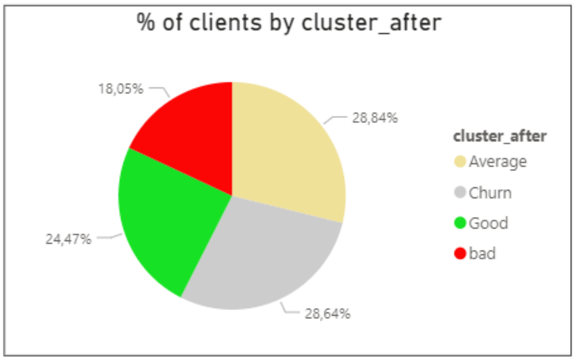
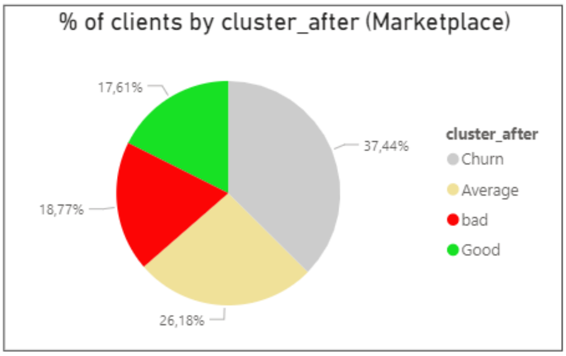


Figure 9	Figure 10

The graph in Figure 9 shows that the clusters of frequency and value are distributed similarly, with the cluster "Average" occupying the largest percentage of the customer base, with a median frequency of purchase of 2.54 and a median value spent of 308,60€. The cluster "Good" has a median purchase frequency of 7.52 and a median purchase value of 1460,17€, while the cluster "Bad" has a median purchase frequency of 1.32 and a median purchase value of 41,83€.

The New/Regained customer was the customer who did not buy anything a year before the NPS survey, which is why it does not present data relative to frequency and value, and because there are only a few cases it only occupies 3.31% of the database.

In the graph of Figure 10 we can see that the Churn percentage of the Database is 28.63%. However, to reinforce again the importance of analysis by processes, it is necessary to understand if the customer base in each process behaves the same after the NPS survey. What is observed is that the processes have different churn rates, for example, the Marketplace process has a churn rate of 37.44% (Figure 11) while the helpline process only has a churn rate of 14.11% (Figure 12).



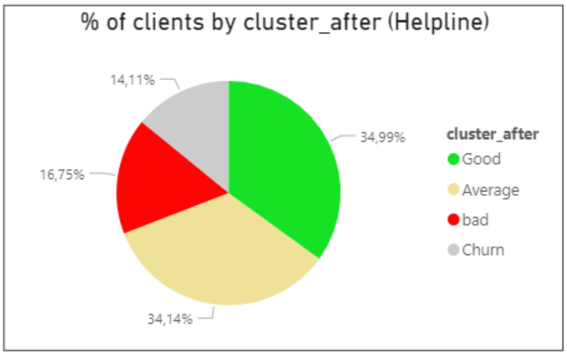


Figure 11

Figure 12

Although we can have these insights regarding churn customers, it is still not possible to understand if there is any kind of relationship between the NPS classes and the customer's transactional data. For this purpose, the use of statistical analyzes is crucial to prove that a K-means analysis together with the NPS allows us to understand if there is a relationship between frequency and monetary value spent with the opinion given by the customer in NPS surveys. In this manner, a Spearman correlation analysis was performed between these two variables; however, in order to perform this analysis, the variables in question must be ordinal, interval, or ratio.

"Testing the equality of two population correlation coefficients when the data are bivariate normal and Pearson correlation coefficients are used as estimates of the population parameters is a straightforward procedure" (Myers & Sirois, 2004).

To modify the variable referring to the clusters, the number 1 was assigned to the cluster "Bad" the number 2 to the cluster "Average," and the number 3 to the cluster "Good." To use the NPS variable, the responses to each questionnaire (from 1 to 10) were used; this allowed us to have two ordinal variables to perform the correlation.

In Figure 13, we can see a table extracted from an SAS output; these results show that there is a relationship between the cluster after the survey response (cluster_D_N) and the NPS variable (nps_score). We can see this by looking at the result of 0.0725, which, while not statistically significant, shows that it is quite close to the value of 0.05. However, there is still some discussion in the academic community over whether a p-value less than 0.10 should be considered significant. The value of 0.05 is simply an alpha limit that we may provide; in this case, we can see that instead of observing a relationship in 5% of the cases seen, we only see it in 7% of the cases. Although the value has been studied, it is important to respect its value, in order to simplify the interpretation of this result, the designation of marginally significant will be assigned.

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations	
	nps_value - N
cluster_A_N	0.00737 0.1410 39894
cluster_D_N	0.01047 0.0725 29442

Figure 13

For a better visualization of the data and a better understanding of what they are saying, a linear graph (figure 14) was created where we can see the NPS Score (percentage of promoters minus the percentage of detractors) for each cluster of K-means after the survey was completed.

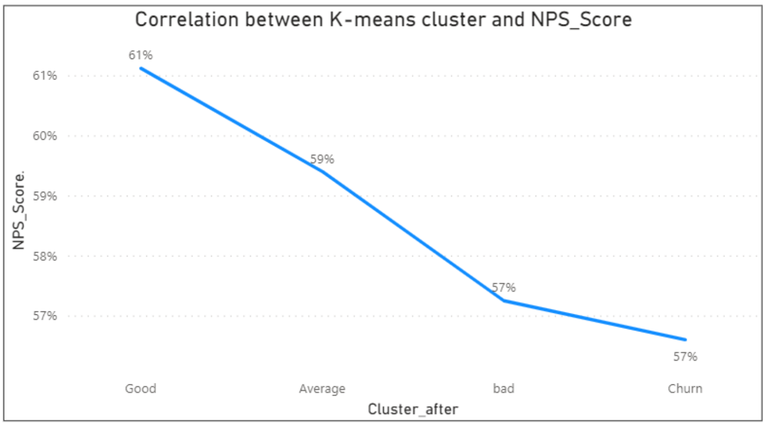


Figure 14

In spite of everything this is an enlightening result. What this analysis tells us is that there is no correlation between clusters in the period prior to the NPS survey, but after the survey is answered, opinion has a relationship with the frequency and value of the customer. This tells us that the NPS (opinion) only presents a relationship with transactional data after this opinion has been presented to the company, if this relationship was also significant for the year before the NPS process, this analysis would lose its value.

Given these results, the conclusion reached is that the null hypothesis is rejected. H1: "The K-means analysis together with NPS allows us to verify if there is a relationship between purchase frequency and value and NPS surveys".

Something to keep in mind is that this analysis was only possible because two time slots were created, and two cluster analyses were performed for the period before and after the NPS survey.

5.5. LOYALTY

After realizing that there is a correlation between Frequency of purchase and monetary value spent with the NPS metric in the short term, it remained to prove whether NPS can measure customer loyalty. According to Reichheld (2003), “Loyalty is the willingness of someone—a customer, an employee, a friend—to make an investment or personal sacrifice in order to strengthen a relationship”, Despite the fact that Reichheld defined loyalty in a very straightforward manner, it is important to realise that, while acceptable, this definition is very general and lacking in specifics, ruling out any business applications.

Although there is much discussion in the academic community on this topic, loyalty is quite complex and difficult to define. According to Bowen (2001), there are three different ways of measuring loyalty, with customer behavioural measures, with measures related to attitude and finally, with composite measures. The last measure presented by Bowen can be evaluated using a multi-metric approach, however in this study only the first two measures will be evaluated in order to answer the research question of the report.

However, the adoption of the third measure can be very advantageous for companies looking for new ways to measure loyalty. Following the pillars of Bowen and Chen (2001), it is essential to evaluate the customer's behaviour and attitude, which is what will be done with the collected data. Purchase frequency analysis will be performed to analyse customer behaviour, and the NPS metric will be employed to determine consumer attitude. But in the meantime, it's important to establish whether each of these variables can accurately forecast loyalty. In this study, it will be determined whether a client is loyal whether they remained and continued to make more or fewer purchases in the year after the survey.

After defining two variables that represent loyalty, some hypotheses will be tested to see if they support the previously defined second hypothesis, H2: “RFM is a better method than NPS for gauging short-term customer loyalty" as well as the fourth hypothesis, H4: “Churn analysis enables companies to determine whether there is a relationship between customer transactional behaviour and NPS."

The first test that was done was Pearson's correlation test to see if there was any kind of relationship between the Churn variable and the NPS variable. In the image of Figure 11, a table extracted from an output of the SAS program can be seen. These results show us that there is a very significant relationship between opinion results and customer churners. We can validate this statement by looking at the area highlighted in red in figure 11, which tells us that there is a strong positive relationship between the two variables, thus observing a significant p-value.

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations	
	churn
nps_score	-0.03397 <.0001 41258

Figure 11

In order to visually observe these results, a graph was created (figure 12) that illustrates the results of the NPS classes and their respective Churn rate. In this way it can observed the relationship more clearly.

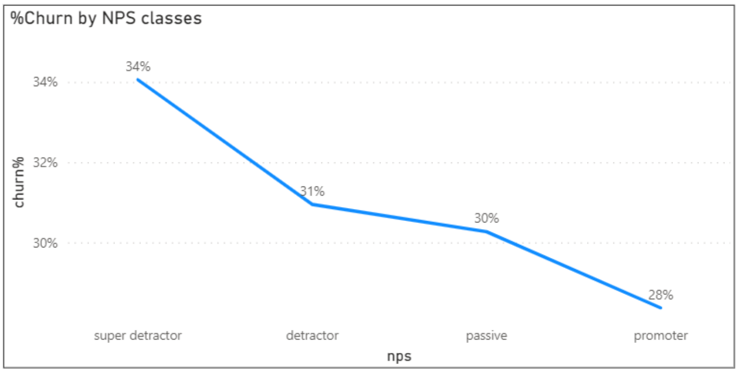


Figure 12

The advantage of splitting this analysis into two distinct time periods, before and after the NPS query, gives companies greater scope in evaluating their data against the NPS metric. In this case, it was decided to analyse the influence of customer frequency over time leading up to the survey and see how this related to the churn rate.

Another Pearson correlation test was created in order to determine whether frequency during the period before to the NPS inquiry had any effect on the rate of churn. The results of this test can be seen in Figure 13.

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations	
	churn
a_frequency_score	-0.42502 <.0001 46568

Figure 13

In the table in Fig. 13 there is a strong correlation between the frequency values from the previous period and the churn rate because the result obtained was statistically significant for the p-value. A line graph was created in figure 14 so that the relationship between these two variables could be more easily visualized.

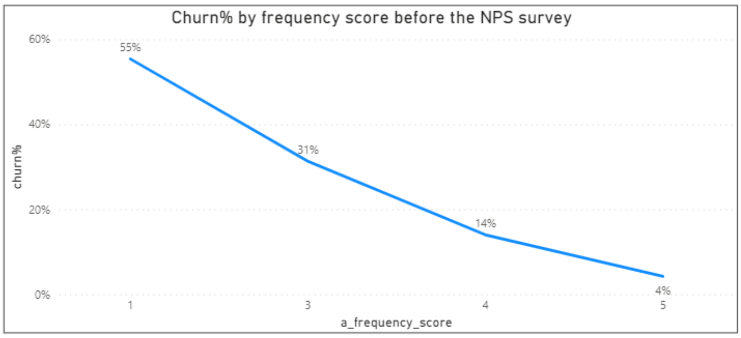


Figure 14

This discovery shows that customers with the highest frequency of purchases (class 5) in the year before to the NPS survey have the lowest churn rates across all classes, whereas customers with the lowest frequency of purchases (class 1) have the highest churn rates.

In view of these results, it is possible to access the fourth hypothesis of this article and reach the conclusion that we reject the null hypothesis that “a churn analysis enables companies to determine whether there is a relationship between customer intention and NPS”. With this analysis, it was also possible to recognise that there is a relationship between churn rate and NPS, just as there is a relationship between frequency prior to the NPS inquiry and churn rate.

Despite achieving these results, it remains to be determined whether the NPS has predictive power in relation to client loyalty, to address this issue, an analysis was performed on customer transaction data one year after the NPS survey to see if there is any relationship to the NPS metric. To obtain statistical validation of this finding, we will perform an ANOVA analysis (Analysis of Variance).

This analysis is a measure of dispersion, it indicates how far the values are from the mean, that is, how far our data are from the expected values (average). To use this measure, it is necessary to have a numerical variable (frequency class in the period after the survey) and a categorical variable (NPS). In this case, ANOVA will evaluate and compare the variance within each group and between groups (promoter, passive, detractor and super detractor).

To perform this analysis, the SAS program was used, and the first output of this interpretation can be seen in the table in figure 15.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	114.14937	38.04979	16.98	<.0001
Error	29438	65947.70458	2.24022		
Corrected Total	29441	66061.85395			

Figure 15

Although there are other insights that we can take from this table, it is important to focus on the p-value of this analysis, which gives a statistically significant value, less than 0.05. And since the ANOVA analysis focus on the difference in variances (Kim, 2017), this means that the difference in variances between the different NPS groups, with purchase frequency values in the period after the survey, is different.

Despite the fact that the variation amongst NPS groups is different, it is now necessary to determine which groups differ from one another. In this way, companies are able to understand which of the NPS classes are the most distant from each other, and which opinions have the most impact on customer transactional data.

"An ANOVA test can tell you if your results are significant overall, but it won't tell you exactly where those differences lie. After you have run an ANOVA and found significant results, then you can run Tukey's HSD to find out which specific groups's means (compared with each other) are different." (Glen, 2022)

The Tukey's test must now be evaluated to discover which NPS classes differ from one another; for this purpose a SAS output it's presented with the table in Figure 16.

Comparisons significant at the 0.05 level are indicated by ***.				
nps Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
promoter - detractor	0.07718	0.00064	0.15372	***
promoter - super detractor	0.12191	0.04583	0.19800	***
promoter - passive	0.14437	0.08652	0.20223	***
detractor - promoter	-0.07718	-0.15372	-0.00064	***
detractor - super detractor	0.04474	-0.05510	0.14458	
detractor - passive	0.06719	-0.01956	0.15395	
super detractor - promoter	-0.12191	-0.19800	-0.04583	***
super detractor - detractor	-0.04474	-0.14458	0.05510	
super detractor - passive	0.02246	-0.06389	0.10881	
passive - promoter	-0.14437	-0.20223	-0.08652	***
passive - detractor	-0.06719	-0.15395	0.01956	
passive - super detractor	-0.02246	-0.10881	0.06389	

Figure 16

The table in Figure 16 shows the comparison between the various NPS classes that can be found in the column "nps comparison," but it is vital to note the top of the table, which states that significant comparisons are marked in the last column on the right with the signal indicated.

These results show that of the four NPS classes created, the class of promoters has a significant difference from the other classes. This means that the customers who are passive, detractors, and super detractors do not have significant differences in terms of purchase frequency in the period following the response to the survey.

Although the NPS result is better than expected, given the criticisms of the academic community, it is necessary to respond to the previously established hypothesis that an RFM analysis is a better technique than the NPS to measure customer loyalty. For this, we need to use the frequency classes before the process instead of the NPS classes and see if there is any relationship with the frequency after the process. It was only used frequency before the process because the objective was to assess the client's loyalty, which is the best of the three RFM classes and even better than using a nested RFM analysis, because it assesses the number of times the client purchased at Worten, which is the clearest indicator of loyalty that can be presented. A Churn analysis was also used because, if customers do not return, it indicates that they are not transactional loyal.

5.6. LOSGISTIC REGRESSION:

Whether or not NPS is a reliable indicator of loyalty is the last important question that needs to be addressed. For this, it is necessary to compare the results of its predictive capacity with another variable, in this case, the frequency before the NPS survey.

As previously stated, the ability to predict loyalty will be assessed using just the binary variable, Churn. In other words, if a customer goes an entire year without purchasing after completing an NPS survey, he is not classified as a loyal customer.

On this basis, a logistic regression will be used to understand the predictive capacity of the two independent variables (NPS and Frequency), thus making the dependent variable, Churn. The variable being classified as 1 for customers who churn and 0 for customers who do not churn.

The logistic regression "(also known as logit model) is often used for classification and predictive analytics. Logistic regression estimates the probability of an event occurring…, based on a given dataset of independent variables. Since the outcome is a probability, the dependent variable is bounded between 0 and 1." (IBM, 2021)

There are three types of learning models in machine learning: supervised learning, unsupervised learning, and reinforcement learning. In the case of a logical regression, the learning is monitored, with the goal of predicting the next result while taking previous data into account.

All predictive models tend to have different assumptions that we must obey, "the advantages of logistic regression" are that there are a "general lack of assumptions required in a logistic regression analysis…, logistic regression does not require linear relationships between the independent variables and t dependent variables" (Hair, 2009)

In order to understand if the two variables can accurately predict churn customers, it is necessary to create two groups of data, the training dataset and the test dataset. The group of data for training will exist to, as the name implies, train the model with the assistance of an algorithm that reads and learns from patterns; this group of data is typically larger than the group of data for testing. In this case, our dataset had 43392 observations, then the database is split it into two groups, the training group, that took a simple random probability sample of 70% of the base, resulting in a data group of 34027 observations and the test group that got 14365 observations, 30% of the base.

After having created two sets of data, it's now possible to apply a logistic regression to understand what impact the independent variables (frequency and NPS) have on the prediction of the dependent variable (Churn). For this purpose, a logistic regression was performed in the SAS program, where the output of this analysis can be seen in figure 17.

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	5347.9218	3	<.0001
Score	5083.9377	3	<.0001
Wald	4412.8276	3	<.0001

Figure 17

As can be seen in in the output of Figure 17, all of the p-values are significant, and they all compare models based on best-fit criteria in the case of logistic regression. In this case, we will just focus on the Likelihood Ratio test, which allows us to determine whether the model we created is superior to the base model. Because the p-value from the chi-squared distribution is highly significant, this indicates that our model is superior.

Now it is mandatory to know within our model which of the two variables contributed better to a loyalty prediction. For this we need to evaluate the p-value of each variable within the table of analysis of maximum likelihood estimates, in figure 18.

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.4280	0.0369	134.3716	<.0001
freq_antes	1	-0.4262	0.00704	3665.3712	<.0001
NPS	1	-0.0195	0.00405	23.2430	<.0001

Figure 18

Looking at the p-value, it tells us whether the variable is statistically significant or not, and consequently whether or not it has an effect on the dependent variable. In this case it is possible to see that being the two variables statistically significant, both help to predict churn.

The estimate column shows what the direction of the effect is, in other words an increase in the frequency class (achieved through RFM analysis) decreases the probability of churn. However, we cannot understand what its impact is, for that it is necessary to use the odds ratio. The odds ratio shows how much the chance of success will increase or decrease based on the value of the independent variable.

To observe the output of the odds ratio estimates table, the figure 19 with the SAS output in relation to the odds ratio analysis is presented.

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
freq_antes	0.653	0.644	0.662
NPS	0.981	0.973	0.988

Figure 19

As mentioned before, if the objective is to know the impact that the independent variables have on the dependent variable, the odds ratio must be used, which is the exponential of beta. To evaluate the Wald confidence intervals, it can be observed in the results of "95% Wald Confidence Limits" that between the two limits of the two variables there is no 1, which means that the independent variable has an impact on the dependent variable.

To interpret these values and understand the magnitude of the impact of the dependent variables, it is necessary to use both the odds ratio and the analysis of maximum likelihood estimates outputs. In this way, if the variable freq_antes (frequency of purchase before the process) is analysed it is possible to see from the estimate in figure 18 that it has a negative impact on the churn forecast, in other words, with the addition of a value in freq_antes, the probability of making churn increases.

Finally, it is necessary to understand the impact and for that it is necessary to analyse the odds ratio which shows that for each increase in frequency, the probability of churn decreases by 65%, keeping the other variables constant.

To assess the impact of the NPS variable, it is necessary to perform the same calculation as for the frequency. In summary, by analysing the odds ratio it is possible to observe direction, magnitude, and significance.

Although NPS has a big impact on predicting churn, it is important to consider the processes that customers go through, for example, if we use this model for the installation process, we can see that the significance values change. This example can be seen in Figure 20.

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.2020	0.2945	16.6630	<.0001
NPS	1	-0.0421	0.0293	2.0662	0.1506
a_frequency_score	1	-0.6926	0.0599	133.7883	<.0001

Figure 20

As can be seen from the p-value of the NPS, this variable is not statistically significant for the model, which means that it does not have any kind of predictive value to detect churn customers. This is not an isolated case, as it happens in 2 different processes. Highlighting once again the importance of analysis by processes within the company

In this case, it is possible to assess that the purchase frequency variable, obtained through an RFM analysis, is a better loyalty predictor than NPS, thus accepting the second hypothesis H2: “RFM is a better method than NPS for gauging short-term customer loyalty”.

Despite realizing that the NPS is a good predictor of loyalty for the general model, it is necessary to realize that in companies with several processes a general analysis of the data may not always be a good measure, as it can mislead many of the results of the predictive model. Thus concluding that in electronic retail, with relative long periods of time between purchases, the NPS falls short compared to the frequency variable. However, NPS emerged as a good predictive variable in most of Worten's processes, making it possible to accept the fifth hypotheses. H5: “NPS accurately predicts the client's future short-term transactional behaviour.”

Something curious that was found was that in NPS surveys where the purchase is associated with a higher monetary value, as in the case of home appliances, the NPS has no predictive power, but when the process is associated with relatively short times between purchases the NPS emerges as a good predictor. Meaning that when the time period between purchases is shorter the NPS stands out as a good predictor, this finding is quite similar to the finding of Baehre et al. (2022), who found that the NPS can be used as a predictive variable depending on the circumstances, specifically in industries where there are short interpurchase cycles. Concluding that in electronic retail, with a purchase frequency of 2/3 purchases per year (Galego, 2014) the NPS is a good predictor of short-term customer behaviour. In view of this result, all companies are advised to employ a suitable predictive model for each process within the company, in this way it is possible to understand where the NPS metric is being well employed.

5.6.1. Goodness of fit:

To understand whether the employed model has a good fit, it is necessary to perform a Goodness of Fit (GOF) test, which is a test that measures how effectively the model fits a set of observations (Maydeu-Olivares et al. 2010). For that we need to use a latent variable, that “is a variable that cannot be observed. The presence of latent variables, however, can be detected by their effects on variables that are observable” (Richard Wagner, 2012)

Finally, the latent variable will be classified as the predictor variable and will be compared with the values observed in the test model, creating a matrix of predicted values vs. observed values. With this matrix it will be possible to understand what the correct percentage of the classification is, in figure 21 the matrix is displayed with the values that must be filled in to be able to create the graph of the ROC curve which is a useful tool in the assessment of the performance of a classification model (Mandrekar and Jayawant 2010).

		Observed Values		Percentage of Correct Classification
		Churn = 1	Churn = 0	
Predicted Values	Predicted Churn = 1	True Positives (TP)	False Positives (FP)	Specificity = TN / (TN + FP)
	Predicted Churn = 0	False Negatives (FN)	True Negatives (TN)	Sensitivity = TP / (TP + FN)

Figure 21

After obtaining the specificity and sensitivity values, the X and Y axes can be created, where the X axis is the value of the false positive rate (specificity 1) and the Y axis, which are the values of the true rate positive (sensitivity). In order to better visualize the ROC curve, a graph was created in SAS, represented in figure 22.

In the case of our model, the area under the curve is quite good, thus achieving a very good predictive value. If the area under the curve was smaller and therefore closer to the line called the “random classifier” (which is the horizontal line that is shown in black on the graph) then it is because your prediction is random, and that meant that our model had no predictive value.

In conclusion, the ROC curve represents sensitivity and specificity for all thresholds of the model under consideration.

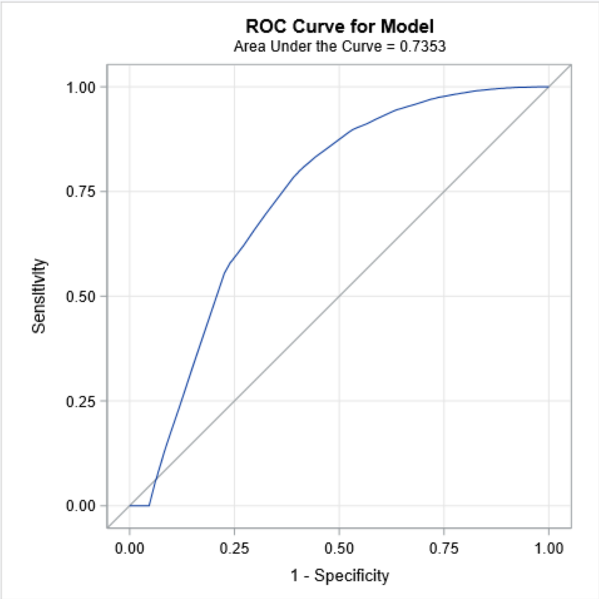


Figure 22

Something that was also possible to notice, is that if companies want to create a model for each process, it is possible if a logistic regression model is created for each process. In this case it is necessary to validate the model's quality from the ROC curve, which in the case of Worten it is possible to validate that all models created for each different process always had an area under the curve of more than 0.70, which is more than acceptable in the context of Marketing. This detail is very important because there may be processes where the NPS metric does not add any predictive value and other processes where it can undoubtedly be a variable to take into account.

6. DISCUSSION AND CONCLUSIONS:

The results of this study aim to answer a research question presented at the beginning of the report, with the aim of providing companies with greater detail about a tool they use, the NPS. To answer the research question of this study, it was necessary to create 5 hypotheses, which aim to explore the ability of this metric to measure loyalty and predict future transactional behaviour. With the help of big data tools, it was possible to reach the first conclusion that there is a relationship between transactional behaviour and NPS surveys at Worten.

Second, and in line with the academic community (Fisher, 2019; Keiningham et al., 2007); Pollack and Alexandrov, 2013; Zaki et al., 2016; Grisaffe, 2007), this study shows that the NPS is not the best loyalty predictor, presenting a viable solution, in the frequency of one year before the NPS survey.

Third, and with a disruptive theme in the domain of questionnaires, it was discovered that the customers that responded to the questionnaires in this study were actually the best customers on a transactional level.

This realization calls into question all the NPS studies carried out by Worten, this because if the base of customers that are being analysed are in reality customers with transactional characteristics much better than normal customers, this means that the base of NPS customers it is not representative of Worten's total customer base. This theoretical and practical implication is likely to be one of the most important findings of this report.

One of the most distinctive aspects of this report is the ability to help small companies to understand the relationship that NPS has with the customer's transactional behavior, through an analysis of customer churn. However, this analysis is difficult to operationalize, as it is necessary to obtain at least one year (or another time interval associated with a customer churner) to analyze the customers' transactional results. Despite all of this, this analysis enables significant inferences to be made, particularly if it is in line with the recommendation provided in this study to break down NPS surveys into company processes.

The examination of processes also enables acquiring a crucial understanding of which areas accumulate more positive (Promoters) and negative (Super detractor/ Detractor) client opinions. With this straightforward research it is possible to observe that poor financial performance in some areas can be linked to the accumulation of unhappy or happy customers.

This study, like few others in the academic community (Kristensen and Eskildsen, 2012; Fisher, 2019; Zaki et al., 2016), presents an adaptation to the NPS metric, namely in customer clustering. The fact of being able to further segment customers is always a good idea, especially when this segmentation presents more enlightening conclusions, which is why another advice presented in this study is to add the class of super detractors in customer segmentation.

One of the final contributions of this study was to realize that the NPS has a good predictive capacity of transactional behaviour. It is crucial to remember that if an analysis is conducted via company processes, this metric will have a significantly higher predictive ability, in this way, it is possible to have a holistic view of the company and use the full potential of the NPS. This research adds to the notion that, under some circumstances, NPS is truly an excellent predictor of future consumer behaviour, which is in line with the finding made by other academics (e.g. Shaw, 2008).

Finally, it can be concluded that there is a lot of potential in the NPS metric, not only if we apply Big Data tools, but also analyses used by almost all companies, such as a simple churn analysis. It is therefore safe to say that NPS has its limitations, but it is undoubtedly a metric that deserves to be explored more carefully by companies.

7. LIMITATIONS AND SUGGESTIONS

While there are several noteworthy insights throughout this article, the research has left certain significant subjects unexamined. The impossibility of following the insight gained from the analysis of the control group exists only because this discovery was something new and disruptive. The fact that the customers who respond to the questionnaires are the customers who have a superior transactional performance than the rest of the company is data that is important to analyse and verify in several companies and not just Worten. The fact that this study only used one firm's database limits the analysis in general; however, it is recommended that all companies use this analysis and approach it with caution, as the NPS analysis can generate misleading information within the company.

In line with many other analyses carried out in the academic community (Kristensen and Eskildsen, 2012; Vandenberg 2002; Wong, Rindfleisch; Burroughs 2003; Eskildsen et al., 2010) is the fact that this study only has data from a single country. The fact that the cultural differences first mentioned by Keiningham et al. (2007) can have an impact on the analysis as a whole, so is important for all companies that only use this metric in a country or in a population with few cultural differences to pay close attention to the results. Therefore, it is important to keep in mind that the NPS is highly susceptible to cultural changes and a change in NPS classes may be necessary, Worten was able to adapt the NPS metric to better serve its needs, which is why it is advisable for many companies to adapt a metric that was launched on the market almost 20 years ago.

Another limitation that prevented other analyses from going forward was the collection of demographic data, which would allow companies to better categorize their customers and, eventually, understand which areas are most significant in each sector. These variables can offer value to any firm, as evidenced by the unique analyses performed by Zaki et al (2016). Demographic data enables businesses with Big Data resources to better explore this tool and reach crucial business insights.

REFERENCES

Baehre, S. O. (2022). The use of Net Promoter Score (NPS) to predict sales growth: insights from an empirical investigation. *Journal of the Academy of Marketing Science*, 67-84.

Ballings, M., Poel, D. V., & Verhagen, E. (2012). *Improving Customer Churn Prediction by Data Augmentation Using Pictorial Stimulus-Choice Data.* Obtido de Semantic Scholar: https://www.semanticscholar.org/paper/Improving-Customer-Churn-Prediction-by-Data-Using-Ballings-Poel/5ccb4254e55c2a471278c1173892da29f8d31e11

Borle, S., Dholakia, U. M., Singh, S. S., & Westbrook, R. A. (September-October de 2007). *The Impact of Survey Participation on Subsequent Customer Behavior: An Empirical Investigation.* Obtido de ProQuest: https://www.proquest.com/docview/212297902?pq-origsite=gscholar&fromopenview=true

Bowen, J. T.-L. (1 de September de 2001). The relationship between customer loyalty and customer satisfaction. *International journal of contemporary hospitality management*, 213-217. Obtido de academia: https://www.emerald.com/insight/content/doi/10.1108/09596110110395893/full/html

Burnham, T. A., & Wong, J. A. (December de 2018). *Factors Influencing Successful Net Promoter Score Adoption by a Nonprofit Organization: A Case Study of the Boy Scouts of America.* Obtido de ResearchGate: https://www.researchgate.net/publication/328148966_Factors_Influencing_Successful_Net_Promoter_Score_Adoption_by_a_Nonprofit_Organization_A_Case_Study_of_the_Boy_Scouts_of_America

Caren, A. (01 de November de 2018). *Why we often remember the bad better than the good.* Obtido de The Washington Post: https://www.washingtonpost.com/science/2018/11/01/why-we-often-remember-bad-better-than-good/

Daniel, S., Matt, B., Randall, T., & Jon, K. (June de 2008). *Measuring Customer Satisfaction and Loyalty: Improving the 'Net-Promoter' Score.* Obtido de semanticscholar: https://www.van-haaften.nl/images/documents/pdf/Measuring%20customer%20satisfaction%20and%20loyalty.pdf

De Haan, E. V. (June de 2015). The predictive ability of different customer feedback metrics for retention. *International Journal of Research in Marketing*, 195-206. Obtido de Science Direct: https://www.sciencedirect.com/science/article/pii/S0167811615000324?casa_token=XEvH8jF56lMAAAAA:qt4rN_AsfbgEZXdX6DdZtC4MzI2KOLAOpK5VZqT6afH8r6VhxeLKVLQPg7ytkZpXWpLb3xhHbQs

Eskildsen, J. K. (2010). The relationship between job satisfaction and national culture. *The TQM Journal*, 369-378.

Evans, G. (2002). *Emerald Insight.* Obtido em 25 de June de 2021, de Measuring and managing customer value:

https://www.emerald.com/insight/content/doi/10.1108/00438020210424262/full/pdf?title=measuring-and-managing-customer-value

Fisher, N. (January de 2019). *Good and bad market research: A critical review of Net Promoter Score.* Obtido de Wiley Online Library: https://onlinelibrary.wiley.com/doi/abs/10.1002/asmb.2417

Galego, M. S. (November de 2014). *Desenvolvimento do e-service no retalho eletrónico.* Obtido de ISCTE University Institute of Lisbon: https://repositorio.iscte-iul.pt/handle/10071/10236

Geller, M. (16 de January de 2008). *Customer satisfaction top U.S. issue in 2008: survey.* Obtido de Reuters: https://www.reuters.com/article/us-nrf-ibm-idUSN1555013620080116

Ghosh, D., & Vogt, A. (2012). Outliers: An evaluation of methodologies. *Joint Statistical Meetings*, (pp. 3455-3460). San Diego. Obtido de Semantic Scholar: https://www.semanticscholar.org/paper/Outliers-%3A-An-Evaluation-of-Methodologies-Ghosh-Vogt/dea2a8a665de3605d075c22a0a897ec7d20a14a0

Gina Pingitore, N. A. (2007). The Single-Question Trap. 9 -13. Obtido de web.p.ebscohost.com.

Glen, S. (2022). *Tukey Test / Tukey Procedure / Honest Significant Difference.* Obtido de Statistics How To: https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/post-hoc/tukey-test-honest-significant-difference/

Greenleaf, E. (1992). Improving Rating Scale Measures by Detecting and Correcting Bias Components in Some Response Styles. *Journal of Marketing Research*, 176–88.

Grisaffe, D. B. (2007). Questions about the ultimate question: conceptual considerations in evaluating Reichheld's net promoter score (NPS). *Journal of Consumer Satisfaction, Dissatisfaction, and Complaining Behavior*, 36–53.

Gruca, T. S. (2005). Customer satisfaction, cash flow, and shareholder value. *Journal of marketing*, 115-130.

Hair, J. F. (2 de February de 2009). *Multivariate Data Analysis.* Obtido de DigitalCommons: https://digitalcommons.kennesaw.edu/facpubs/2925/

IBM. (2021). *What is logistic regression?* Obtido de IBM: https://www.ibm.com/topics/logistic-regression

Jacob Eskildsen, K. K. (15 de June de 2010). *The relationship between job satisfaction and national culture.* Obtido de emerald insight: https://www.emerald.com/insight/content/doi/10.1108/17542731011053299/full/pdf?title=the-relationship-between-job-satisfaction-and-national-culture

Jenny van Doorn, P. S. (September de 2013). *International Journal of Research in Marketing.* Obtido de Science Direct: https://pdf.sciencedirectassets.com/271657/1-s2.0-S0167811613X00047/1-s2.0-S0167811613000463/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEOr%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCXVzLWVhc3QtMSJGMEQCIEZBPocBG81LMPd6zZO3Ldv6a9x0bWUpfXoLCiBGuhnEAiAFDmYN6j4q5ih2

John Jonides, R. L., & Moore, K. S. (01 de April de 2014). *The Mind and Brain of Short-Term Memory.* Obtido de PMC PubMed Central: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3971378/

Keiningham, T. L. (2008). A holistic examination of Net Promoter. *Journal of Database Marketing & Customer Strategy Management*, 79–90.

Keiningham, T., Cooil, B., Andreassen, T., & Aksoy, L. (1 de July de 2007). *A Longitudinal Examination of Net Promoter and Firm Revenue Growth.* Obtido de Journal of Marketing: https://journals.sagepub.com/doi/full/10.1509/jmkg.71.3.039

Kim, T. K. (February de 2017). *Understanding one-way ANOVA using conceptual figures.* Obtido de PMC US National Library of Medicine National Institutes of health: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5296382/

Kotler, P. (2017). Customer Value Management. *Journal of Creating Value*, 170-172.

Kristensen, K., & Eskildsen, J. (January de 2012). *Is the NPS a trustworthy performance measure?* Obtido de ResearchGate: https://www.researchgate.net/publication/258770491_Is_the_NPS_a_trustworthy_performance_measure

Lloyd, S. P. (1 de March de 1982). *Least squares quantization in PCM.* Obtido de Semantic Scholar: https://www.semanticscholar.org/paper/Least-squares-quantization-in-PCM-Lloyd/9241ea3d8cb85633d314ecb74b31567b8e73f6af

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, 1315-1316.

Maydeu-Olivares, A. a.-F. (2010). Goodness-of-fit testing. *International encyclopedia of education 7.1*, 190-196.

Mayfield, A. (19 de March de 2013). *Survey psychology: The psychology of respondents.* Obtido de OW: https://blog.optimalworkshop.com/the-psychology-of-survey-respondents/

McHugh, M. L. (June de 2013). *The Chi-square test of independence.* Obtido de Research Gate: https://www.researchgate.net/publication/253336860_The_Chi-square_test_of_independence

Mecredy, P. M. (2018). Are promoters valuable customers? An application of the net promoter scale to predict future customer spend. *Australasian marketing journal, 26(1)*, 3-9.

Morgan, N. A., & Rego, L. L. (September de 2006). *The Value of Different Customer Satisfaction and Loyalty Metrics in Predicting Business Performance.* Obtido de ResearchGate: https://www.researchgate.net/publication/257527259_The_Value_of_Different_Customer_Satisfaction_and_Loyalty_Metrics_in_Predicting_Business_Performance

Myers, L., & Sirois, M. J. (July de 2004). *Differences between Spearman Correlation Coefficients.* Obtido de ResearchGate: https://www.researchgate.net/publication/286294834_Differences_between_Spearman_Correlation_Coefficients optimove. (11 de 2022). *Customer Churn Prediction and Prevention*. Obtido de optimove: https://www.optimove.com/resources/learning-center/customer-churn-prediction-and-prevention

Pollack, B. L. (2013). Nomological validity of the Net Promoter. *Journal of Services Marketing*, 118–29.

Rajasekaran, M. N. (2018). HOW NET PROMOTER SCORE RELATES TO ORGANIZATIONAL. *International Journal of Creative Research Thoughts (IJCRT)*, 980.

Reichheld. (December de 2003). *The One Number You Need to Grow.* Obtido de Harvard Business Review: https://hbr.org/2003/12/the-one-number-you-need-to-grow

Richard Wagner, P. T. (December de 2012). *Latent Variable.* Obtido de SAGE researchmethods: https://methods.sagepub.com/reference/encyc-of-research-design/n213.xml

Sauro, J. (23 de January de 2019). *What Motivates People to Take Free Surveys?* Obtido de Measuring U: https://measuringu.com/free-surveys/

Scheuren, F. (2004). *What is a Survey.* Alexandria: American Statistical Association. Obtido de Library: https://1library.net/document/y4x27mkz-what-is-a-survey-by-fritz-scheuren.html

Shaw, R. (2008). Net promoter. *Journal of Database Marketing & Customer Strategy Management*, 138-140.

Taylor, J. (2013). *Delivering Customer Value Faster.* Obtido em 25 de june de 2021, de Decision Management Solutions: https://brblog.typepad.com/files/deliveringcustomervaluefasterwithbigdataanalytics-update-june-6.pdf

Vandenberg, R. J. (2002). Toward a Further Understanding of and Improvement in Measurement Invariance Methods and Procedures. *Organizational Research Methods*, 58-139.

Varki, S. a. (1997). Satisfaction Is Relative: Apply Analysis of Variance Techniques to Determine if Your CSM Scores Measure Up. *Marketing Research*, 15–19.

Verhoef, P., & Lemon, K. (2011). *Msi - Marketing Science Institute.* Obtido em 25 de june de 2021, de Customer Value Management: https://www.msi.org/wp-content/uploads/2020/06/MSI_FF_11-700.pdf

Wong, N. A. (2003). Do Reverse-Worded Items Confound Measures in Cross-Cultural Consumer Research? The Case of the Material Values Scale. *Journal of Consumer Research*, 72–91.

Zaki, M. K.-K. (2016). The fallacy of the net promoter score: Customer loyalty predictive model. *Cambridge Service Alliance*, 1-25.