Exploring the value of large databases of transaction data for business intelligence

Customer Behavior Analysis in an E-commerce Setting

Vincent Meulenbroek
S1007424
University of Twente
The Netherlands
Master Business Administration

ABSTRACT

Business intelligence is the process of collecting and transforming data into knowledge, to positively influence decision-making, and is the main subject in this study. Market intelligence is part of business intelligence, and customer insight is part of market intelligence. Small to medium-sized enterprises often face large databases but do not know how to use or exploit the databases because of different reasons. This study tries to find the value of large databases of transaction data for creating business intelligence. Concepts as relationship management, repurchase behavior and customer life cycle are discussed thoroughly. The focus is on small to medium-sized enterprises, a case is added to give the research its power; Webprint.Data of more than 37.000 customers are analyzed, from their first to third purchase, in order to find the value of transaction data for business intelligence.Customer behavior patterns are exposed to find the value of the large database.

Supervisors:

Prof. Dr. A.J. Groen
Drs P. Bliek

Date: 13-02-2017

Preface

This thesis concludes and thereby ends my time as a student, a period of which I enjoyed myself most of the time. The study is the last phase of the Master program Business Administration, after I already finished the Bachelor program International Business Administration late 2014. My college time has taught and brought me a lot, not only in academic perspective but I have also experienced my college time as pleasant and enjoyable.

During the process of the research I have had a lot of help conducting and processing the thesis. Therefore I would especially express my gratitude towards my first supervisor Prof. Dr. A.J. Groen. You have helped me by providing useful feedback and advise. In my opinion we have worked together without any major problems and made sure that I was able to finish the thesis in time. But, together with my first supervisor, my second supervisor; Drs P. Bliek, also provided me with meaningful feedback and advise. I would like to thank you for your help.

Besides the help from the University of Twente I would also like to thank Webprint Group for the opportunity to graduate at the company in question. Not only the opportunity but also all the feedback and advise I have received from different people within the organization are worth mentioning. Thank you very much for all the help I was privileged to receive.

At last, I would like to thank everyone else who I might have forgotten to thank earlier in the preface. Thank you for participating in my thesis in any kind of way.

With kind regards,

Vincent Meulenbroek

Oldenzaal, 13-02-2017

Contents page

List of tables and figures

Tables

Figures

1. Introduction

Electronic commerce or e-commerce is a concept that includes any type of business, or commercial transaction, that involves the transfer of information across the internet. It covers a range of different types of businesses, from consumer based retail sites, through auction or music sites, to business exchanges trading goods and services between corporations (Mahadevan, 2000; NetworkSolutions, 2015). It is currently one of the most present partsof the internet to emerge. E-commerce has experienced a huge growth in the last decade even though it already existed since the seventies, when money was exchanged by use of the computer network.

Companies in the e-commerce business are swamped with lots of information, particularly because of the ease of collecting this data. Purchase data of customers are collected continuously and often only pile up without using the data purposeful. Especially small to medium-sized enterprises deal with lots of unused customer behavior data because they do not have the time, knowledge, or resources to transform these data into value for the company(McAfee A. , Brynjolfsson, Davenport, Patil, & Barton, 2012).

Customer data which is collected, but often not used, remains in large, untouched databases. In most cases the data involves information which is collected after transactions are made: transaction data (Foerster & Karolyi, 1998). Analyzing large databases of transaction data demands much specialized attention, analytical skills, and, if the databases are huge, even resources such as specialized software. Large amounts of transaction data are often labeled as big data, and big data, which is not surprising, can outstrip the capacity of manual analysis (Provost & Fawcett, 2013).

Because of big data, managers can measure radically more about their businesses, and directly translate that knowledge into improved decision making and performance(McAfee A. , Brynjolfsson, Davenport, Patil, & Barton, 2012). Big data analysis provides more insight in customer behavior and can expose customer behavior patterns. Not surprisingly, companies that are able to analyze big data successfully are able to gain a competitive advantage. Big data analysis is the key analytical component in business intelligence (Chen, Chiang, & Storey, 2012).

Nowadays, improving business intelligence is a core activity in large companies. It is often entangled in the company, and influences decision making. Business intelligence refers to transforming hard data into knowledge. Online businesses acknowledge that the understanding of their data is crucial to their competitive position (Negash, 2004).

This study tempts to find out what the value is of large databases of transaction data for business intelligence. A case is used to illustrate the surplus of analyzing big data and, as a consequence, improve business intelligence. The case which is used is Webprint, a Dutch online photo service which is online for more than 10 years. It is a medium-sized company. Especially small to medium-sized companies tend to have unused databases of transaction data. The answer to the research question will be sought in the discipline of customer behavior, because the case data is best fitted for this field.

The company wants to improve its use of transaction data, funded by theoretical and practical rationale. It is in line with the focus of this research and completes the problem statement of this study.

The following research questions and sub questions are formulated:

What is the value of large databases of transaction data for creating business intelligence?

1. What is business intelligence, what does it consist of and what can be achieved by improving business intelligence?

2. What are the characteristics of customer repurchase behavior in e-commerce and how to stimulate it?

What customer behavior patterns are present at Webprint.nl and what points of action do the findings unveil?

3. How can retaining customers be pushed in the desired direction?

4. How can customer relationship management (CRM) influence customer behavior?

5. How can the customer life cycle be used for decision-making?

6. Does socio-economic status (SES) influence order value?

In this study, the case is used to find out the value of large databases of transaction data. The first two sub questions are set up to provide a theoretical foundation on which the second research question, and the last four sub questions, are based. The second research question considers the case Webprint, and focuses on three main subjects coming from the theoretical foundation; customer retention, customer relationship management and the customer life cycle. The sub questions formulated to support the second research question tend to help the case improve their business intelligence by use of transaction data.

The first chapter contained the introduction. The second chapter offers a theoretical foundation in relation to all valuable and relevant concepts, which will provide a theoretical framework for the remaining part of the thesis. The third chapter contains the methodology and case information relevant to the study performed. It includes the research design, data collection methods and data analysis. The fourth chapter reflects the results and findings from the statistical part of the study. Finally the thesis will end with a discussion of the results, the limitations, further research and the conclusion.

2 Literature review

2.1 Business intelligence

2.1.1 Big data as input and data-driven decision making as output

Nowadays vast amounts of data are available, which makes companies in almost every industry focus on exploiting data for competitive advantage. “The volume and variety of data have far outstripped the capacity of manual analysis, and in some cases have exceeded the capacity of conventional databases” (Provost & Fawcett, 2013, p. 51). The term big data appears when businesses are collecting more data than they know what to do with (McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012). Because of big data, managers can measure radically more about their businesses, and directly translate that knowledge into improved decision making and performance. Big data analytics represents the key analytical component in business intelligence(Davenport, 2006). “Big data and big data analytics have been used to describe the data sets and analytical techniques in applications that are so large that they require advanced and unique data storage, management, analysis, and visualization technologies” (Chen, Chiang, & Storey, 2012, p. 1166). Big data analytics has its roots in the database management field, which is part of this study. Analyzing big data correctly may offer companies great opportunities for competitive advantages, especially for online businesses. Online businesses acknowledge that the understanding of their data is crucial to their competitive position (McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012). On top of that, big data does not erase the need for vision or human insight, because big data cannot be analyzed whatsoever without specialized and skilled human resources (McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012; Provost & Fawcett, 2013). According to McAfee et al. (2012) and Provost and Fawcett (2013) data-driven decisions are decisions on the basis of evidence rather than intuition, and are therefore better than other types of decisions (figure 1). Finding and translating patterns in data provides useful information to embrace evidence-based decision making. It means that professional marketers (e.g.) do not take decisions purely on their experience of the business (intuition), but have to integrate analytical knowledge on databases. Hence, data-driven decision making is not an all-or-nothing practice, as it can combine intuition and data knowledge.McAfee et al. (2012) conducted a study in which the degree of influence of data-driven decision making is tested towards firm performance. The results were noteworthy, as they show statistically that the more data-driven a firm is, the more productive it is, even controlling for a wide range of possible confounding factors. Big data information is a structured input for business intelligence.

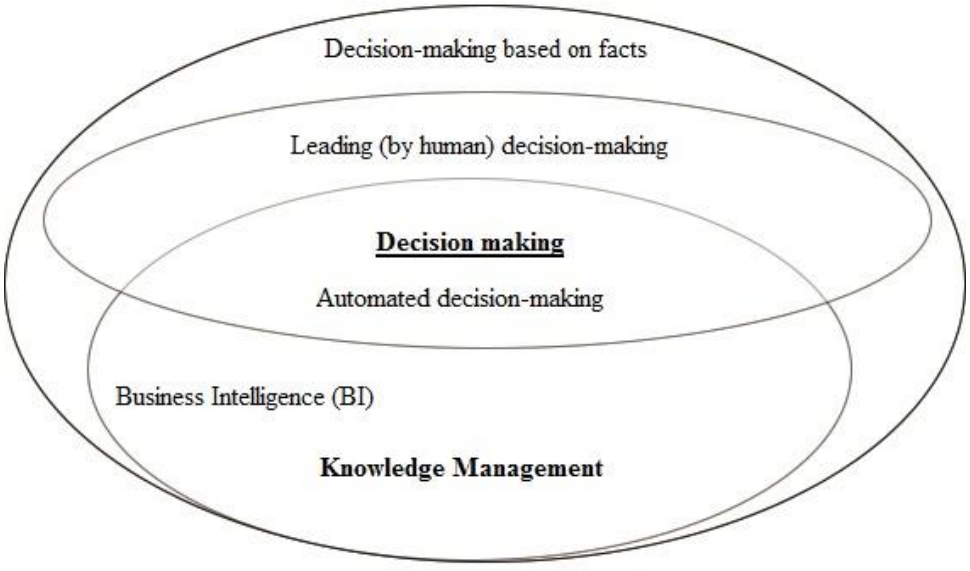


Figure 1: Data-driven (automated) decision-making and business intelligence(Hakanen, 2014)

2.1.2 Business intelligence as overarching theme of other intelligence concepts

Business intelligence can be referred to as the process of collecting and transforming data into knowledge, with as goal to positively influence decision-making, which can improve and uphold performance. It can be seen as part of the overarching theme of information and knowledge management. Business intelligence is viewed as being proactive and supports decision makingon different levels. Negash (2004) illustrates that business intelligence demands unstructured and structured inputs of information to provide intelligence needed in decision making. Different information systems, such as geographic information systems and data mining systems, are part of the structured part of the input. The unstructured part of the input consists of, among others, conversations, graphics and spreadsheets. "The emergence of the data warehouse as a repository, the advances in data cleansing that lead to a single truth, the greater capabilities of hardware and software, and the boom of internet technologies that provided the prevalent user interface all combine to create a richer business intelligence environment than was available previously" (Negash, 2004, p. 179).Only getting the data in a business intelligence system is not enough as is emphasized here: "Getting data in delivers limited value to an enterprise; only when users and applications access the data and use it to make decisions does the organization realize the full value from its data warehouse. Thus, getting data out receives most attention from organizations" (Watson & Wixon, 2007, p. 96). The activity of getting the data out after getting data in is often referred to as business intelligence(Chaudhure, Dayal, & Narasayya, 2011).As mentioned before, business intelligence can be seen as part of the overarching theme of information and knowledge management. But, business intelligence can also be divided. Related intelligence concepts include competitive intelligence, competitor intelligence, customer intelligence, market intelligence and strategic intelligence. These other concepts focus mainly on external environment and are seen as subgroups of a more extensive term, business intelligence (Hakanen, 2014). Figure 2 reveals the different phases of business intelligence and creates an overall picture of the issues that have to be considered to the companies' business intelligence. Strategic intelligence has the broadest scope of information and depends on internal and external information. Competitive intelligence depends on external information which makes the scope of information at the same time a little less than the scope of strategic intelligence. Market intelligence compels customer and competitor intelligence and also depend on external information, the scope of information declines and narrows. To set focus in this research, market intelligence is discussed further.

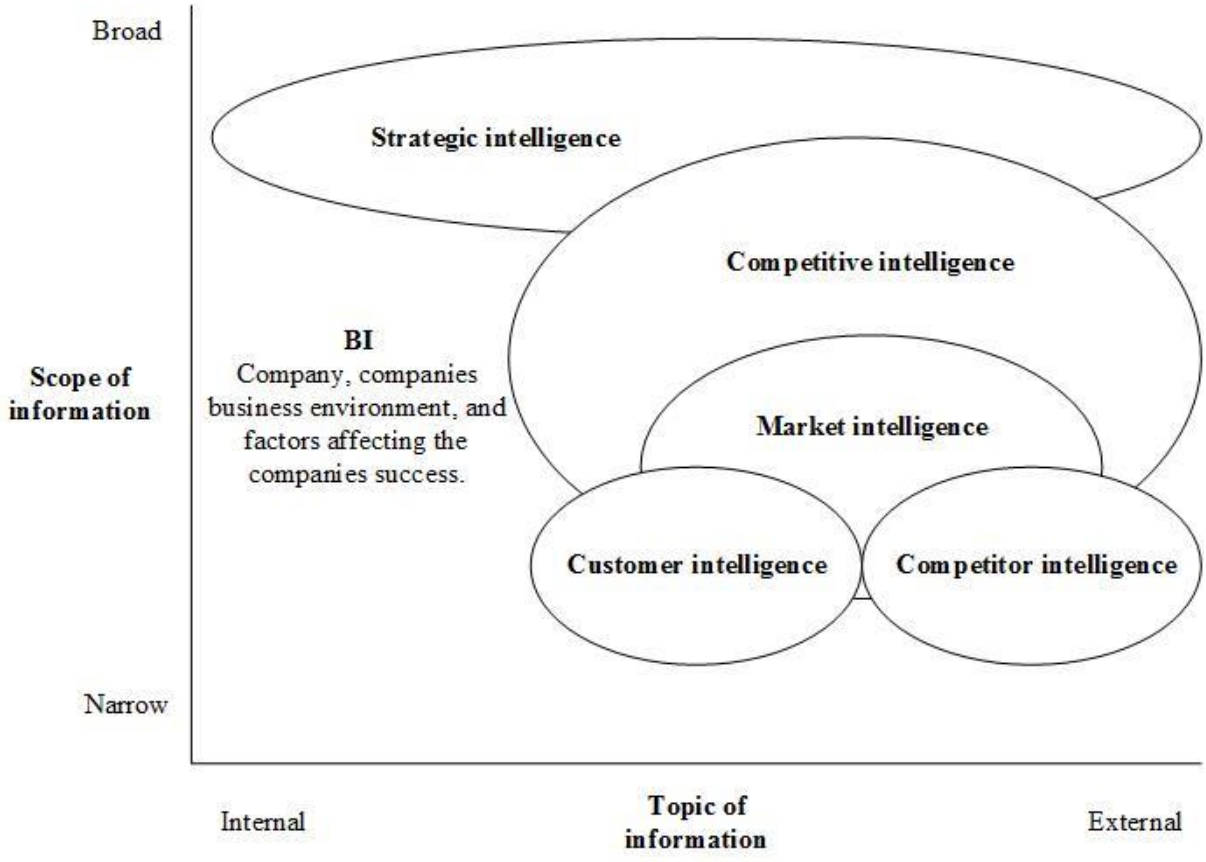


Figure 2: Business intelligence and intelligence concepts (Hakanen, 2014)

2.1.3 Market intelligence

Market intelligence is concerned with information relevant to the markets the company is selling its products in, while business intelligence is concerned with information from the overall environment the company is surrounded by. The model by Hakanen (2014) can be further examined, by which market intelligence is most closely related to information streams relevant for day-to-day processes. According to Crowley (2007) market intelligence consists of four cornerstones; competitor intelligence, product intelligence, market understanding, and customer understanding. Figure 3 illustrates this mindset in a pyramid shape. The model by Crowley (2007) can be seen as an extension and deepening of the model by Hakanen (2014) and displays concrete activities or themes by which information streams originate from. These activities or themes provide leads for studies. The difference between both models lies in the view, the model by Hakanen (2014) is using a top-down view, while the model by Crowley (2007) is a bottom-up view. This means that Crowley's model is much more customer-oriented instead of the strategic focus the model of Hakanen (2014) starts with. The customer-oriented focus of this study makes the model of Crowley more applicable to this study.

The case Webprint has limitations concerning available data which means a focus needs to be set on the area of intelligence. This study uses transaction data coming from, as self-evident, customers' transactions. Therefore, competitive insight is not investigated in the research, since that data is not readily available (technical limitations). Secondly, because the company's intention is to gain more insight into customer transaction data, and therefore into their database, customer insight is the cornerstone this study is built around.

Market intelligence, and in particular customer insight, is the main subject in this study. The subject compels customer behavior literature, case information and customer transaction data to test the research questions. Webprint, to be precise the Dutch part of the Webprint Group, is the case by which customer insight, as part of market intelligence and therefore as part of business intelligence, is examined. The conclusion of the research includes the completion of the research questions. The overall research provides insight into the database of transaction data, substantiated by relevant customer behavior literature.

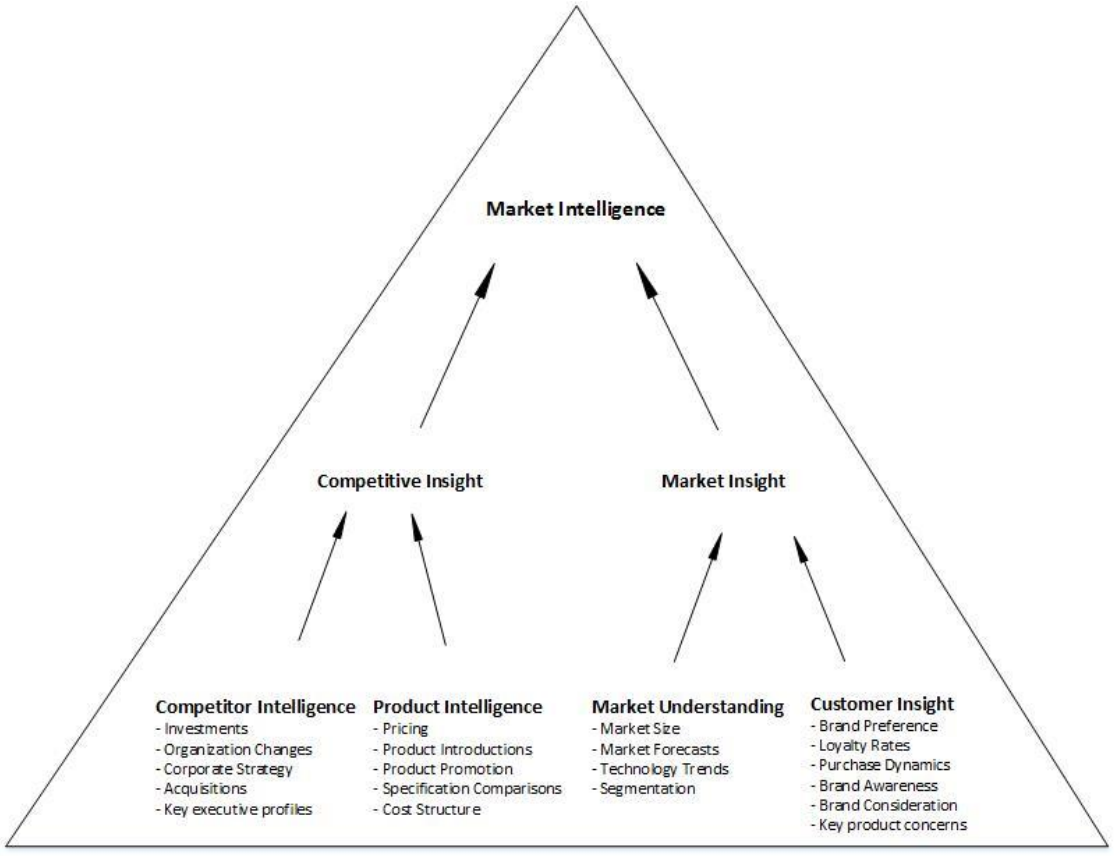


Figure3: Market Intelligence Model (Crowley, 2007)

2.2 Customer behavior

Since business intelligence is narrowed by focus in the previous chapter, relevant literature on customer behavior is discussed next. The literature in this part contains the subjects relationship management, the customer life cycle, and repurchase behavior. Socio-economic influences on customer behavior are discussed last.

2.2.1 Necessity of relationship management for repurchase behavior

Because of the revolution on information technology, especially the World Wide Web, companies have the opportunity to choose how they interact with their customers. “By combining the abilities to respond directly to customer requests and to provide the customer with a highly interactive, customized experience, companies have a greater ability today to establish, nurture, and sustain long-term customer relationships than ever before” (Winer, 2001, p. 89). Buying behavior on the web refers to two stages: encouraging people to purchase online which is the first stage and the second stage encouraging customers to repurchase; the essence of e-commerce. Attracting new customers instead of retaining existing customers requires more effort and time (Zhang, Fang, Wei, Ramsey, McCole, & Chen, 2011). Therefore, online customer retention attracted considerable attention in recent years, but also because it is a means of gaining competitive advantage(Tsai & Huang, 2007). Understandable, a focus on customer retention reduces the pressures associated with attracting new customers (De Bock & Poel, 2011). In order to focus on customer retention, relationships with customers have to be formed and maintained. Relationship management is therefore been present in management for a long time past but has not always postulated attention.

The importance of Customer Relationship Management (CRM) nowadays increased with regard to the past and it is often much more expensive to acquire new customers than to keep them (Phan & Vogel, 2010). CRM in e-commerce is concerned with retaining the existing customers on a long-term basis, motivating them to come back to shop for more and to talk positively to their peers about the products and services provided (Ozok, Oldenburger, & Salvendy, 2007). Ozok et al. (2007) emphasize it is five times as expensive to attract a new customer as it is to keep an existing one, customers who have relationships with companies are more loyal and less likely to switch when a better price is discovered, satisfied customers are more likely to refer new customers, and customer feedback can be effective (p.285). A good relationship in B2C e-commerce between buyer and seller is developed only when buyers feel satisfied and have trust in their relationship with the vendor; most of the literature agrees trust and customer satisfaction are the key sub-constructs of relationship quality (Zhang, Fang, Wei, Ramsey, McCole, & Chen, 2011).

In the information society in which we now live companies are continuously searching for knowledge to understand customer behavior, and at the same time the focus of attention shifted to customers that can deliver long-term profits. Initially marketers focused on acquiring customers, both new ones that have not bought the product before or those who are currently competitors’ customers. Winer (2001) is a pioneer onCRMand developed its own model in which he states that CRM can be measured by satisfaction and is influenced by different customer retention programs: customer service, frequency / loyalty programs, customization, rewards programs, and community building (p.98). Winer’s CRM model is years later refurbished and adapted to the current economic situation or environment. N’Goala and Cases (2012) identify eight drivers allowing an e-commerce site to enhance customer relations: 1) Aesthetic design: emphasizes on the beauty, pleasure and / or entertainment provided by the website, 2) Choice: the ability of the e-tailer to offer a wide range of products / services, 3) Contact: the contact between client and e-tailer in order to create familiarity and keep clients informed, 4) Privacy Policy: the intentions of the e-tailer to respect the privacy of its customers with regards to personal information, 5) Customization: the intentions of the e-tailer to adapt communication and commercial offers according to the customer profile, 6) Interactivity: the intentions of the e-tailer to facilitate navigation on its website and offer some interactivity, 7) Commitment: absence of errors, customer service and advice, reliability and other obligations e-tailers need to meet during the transaction, 8) Community: crucial in customer relationship management and implies the free exchange between clients and the sharing of experiences on the website. The presented framework is, however, largely comparable to the study of Srinivasan, Anderson, and Ponnavolu (2002) who investigated the antecedents and consequences of customer loyalty in e-commerce. The drivers allowing an e-commerce site to act on customer relations, in both frameworks, refer to the core of the commercial service (choice and commitment), the website properties (design and interactivity / navigability), the practices of managing contacts (respect for privacy, customization of messages and frequency of contact) and social devices (community) (N'Goala & Cases, 2012).

Customer relations can be enhanced by different drivers, like the eight c's, moreover CRM also comprehends certain functions (Xiong & Liu, 2011). CRM can be used to improve the service. Customers can receive personalized information based on their historical information as a service. Personalization is defined as personal interaction that is being practiced by the store to enhance customer relationships (Julian, Ahmed, Wel, & Bojei, 2015). Efficiency can be improved by CRM: enterprises can integrate the available resources, improve the overall response capacity and processing capabilities of sales department, improve the working quality and marketing efficiency of the company. CRM can reduce costs compared with traditional marketing methods by modern network technology that reduces operating costs of marketing. Sales of companies can increase by increasing the success rate and the customer satisfaction that will probably be stimulated by customer relationship management(Xiong & Liu, 2011). Customers' needs can be predicted, targeted products and services can be provided, and the accuracy of the sales business can be improved. Customer relationship management is an important management tool to manage customer relationships and the ability to extend and repeat the customer life cycle.

2.2.2 Different phases in customer life cycle demands different approaches

"Customer Life Cycle is one of the most classic models in customer relationship management and refers to the whole process that a customer goes through when considering, purchasing, using, and maintaining loyalty to a product or service"(Yue & Xiang, 2012, p. 22). The customer life cycle is separated into four stages: initial period, formative period, stable period, and degenerated period (figure 4).In the initial period the relationship is very fragile and new. Vendor and buyer do not know each other and have not been able to bond yet. The formative period is a rapid stage of customer relationships, trust is gained and the relationship starts to develop quickly. The stable period is the optimal level regarding to customer relationships, whereby the customers are very loyal to the company and purchasing levels are high. In the degeneration period, customers can be lost because of different reasons, of which lowcustomer satisfaction is the main reason. Customers loyalty can decrease because of low switching costs, changed personal preferences, increased competition, or saturation of the product or service.

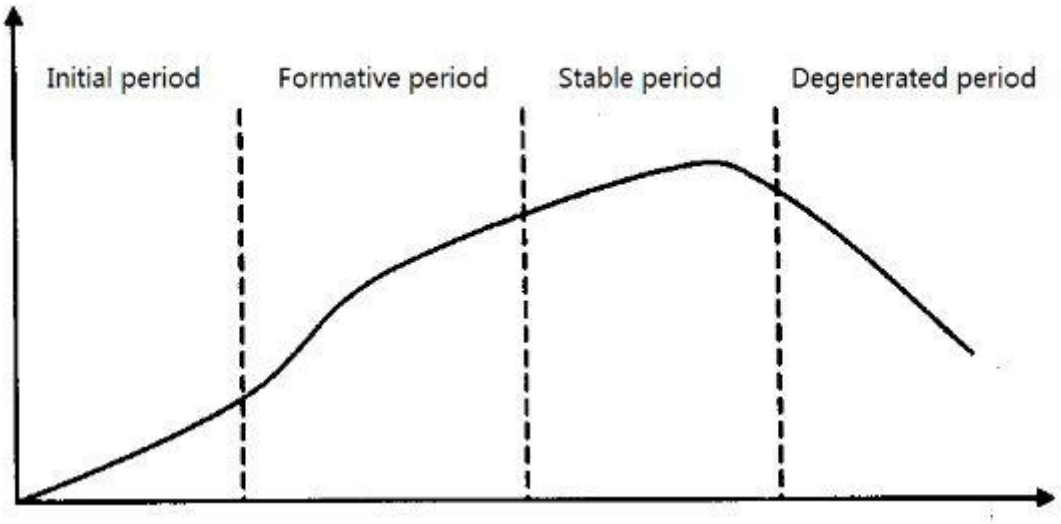


Figure 4: Customer life cycle (Yue & Xiang, 2012)

Acquiring customers is expensive in each industry, but is necessary in order to attain customer relationships. The early periods of the customer life cycle are pricey but costs of serving loyal customers fall in later years because of the rise of the volume of their purchases; big returns can be generated (Reichheld & Schefter, 2000). Reichheld and Schefter (2000) emphasize costs for acquiring customers are considerably higher in e-commerce than in traditional retail channels. On top of that, the authors state that new customers tend to cost twenty to forty percent more for internet companies than for traditional retailers with both physical and on-line stores. The customer life cycle can also be broke down in different steps instead of periods, by which the goal of CRM is to get the customer to move through the cycle again and again. The first step is reaching the customer, then to get their attention and show what the company can offer. Thirdly the potential target should be transformed into a paying customer, which is followed by keeping them as a loyal customer whose satisfaction with the product or services urges other customers to join the cycle.

The customer life cycle (CLC) can be divided into steps or periods, however a different point of view comes from splitting up different phases by customer characteristics, and consists of six phases(Gao, 2014). The first phase of the CLC is filled up by prospective customers. In this phase prospects' (potential customers) interest is peaked and the next task is to provide trust, security, and incentives to ensure prospects to overcome any barriers or obstacles and become a registered or paying user or customer (customer acquisition). The second phase consists of new customers or users. Vendor and buyer try to develop a long-term relationship because ideally you do not want a customer who purchases one time and then never returns. The key is to interact with the customer in order to develop the relationship wanted. The third phase consists of active customers or users. A prospective customer has been converted into a paying or registered customer. This phase is not only about showing the customer the quality of the products or services offered but focuses on building a relationship as well. Showing customers they are important; communicate with respect and offer them interesting promotions.The fourth phase of the CLC is made up by repeat or loyal customers or users. This phase is all about offering incentives to customers to remain loyal since they are already repeat customers; you want them to keep buying even more (often). Different channels can help to keep an ongoing dialogue with your loyal customers in order to retain them. The fifth phase consists of lapsed customers or users. A customer receives the tag "lapsed" whenever contact or interest is loosed by a loyal or repeat customer; the time to transform into a lapsed customer is defined by the company itself and can therefore differ between companies. Key in this phase is to identify these customers by the time they did not purchase, then you can decide what actions are appropriate and necessary to get them buying again. The last phase includes inactive or abandoned customers or users. Lapsed customers can turn into inactive or abandoned customers that in fact do not have any purchases or contact with your company. By the use of different campaigns some customers might be coerced to get back to phase one and the cycle starts all over again. If you know the likely life cycle or pattern of a customer, changes can be made to customer communications or marketing strategy to try to optimize the length of time and the value that a customer brings to the business (Gao, 2014).

The customer life cycle can be spectated from different angles as discussed in the paragraphs above. The different views examine the life cycle differently but do emphasize the same: the focus on customer retention. Each of the points of view confine the same starting point of the CLC; that is customer acquisition. Nevertheless CLC focuses on customer retention because of one main reason that is backed up by different authors: customer acquisition is much more expensive than focusing on retaining customers. Yue and Xiang (2012) argue that customer retention is most important in the third period, in other words the stable period. In this period the customer relationships are optimal and each customer's value is optimized. Secondly, Reichheld and Schefter's (2000) CLC consists of different steps and the fourth step comprises customer retention. In this step customers become loyal and satisfied and even stimulate other potential customers to join the cycle. Finally, Gao (2014) focuses on different phases of the CLC of which the fourth phase comprehends customer retention. Keeping an ongoing dialogue through all marketing communication channels possible is the best advise Gao gives to businesses that want to retain customers. Repeat customers regain profit for any venture because they are less sensitive to price, have greater spending capacity, can be served at a lower cost, and pass on positive recommendations to others (Reichheld & Sasser, 1990; Gupta & Kim, 2007). Companies therefore focus on long-term relationships in order to make customer more likely to purchase repeatedly. Nevertheless, companies often do not understand the expectations of online customers properly, which means that online stores fail to maintain repeat sales (Manning, Bodine, Temkin, & Amato, 2005).

2.2.3 Influencing repurchase behavior

Much research in the area of customer retention is done in an offline context. Gustafsson, Johnson & Roos (2006) studied customer retention at telecommunications services and consider three prominent drivers of customer retention: overall customer satisfaction, affective commitment, and calculative commitment. Customer satisfaction is an overall evaluation of performance to date, affective commitment captures the trust and reciprocity in a relationship, and calculative commitment captures the existence of switching costs or lack of viable alternatives (p.215). Another study by Verhoef (2003) on customer retention with data from financial services in an offline setting displays positive influence of affective commitment, satisfaction, and loyalty programs on customer retention. Julian, Ahmed, Che Wel, & Bojei (2015) test the influence of different variables on customer retention which include: customer service, loyalty/rewards programs, customization, personalization, and brand/store community. The results indicate that customer retention is influenced by brand/store community, personalization, customization, customer service, and loyalty/rewards programs (p.199). Increasing switching costs increases the threshold of customers to leave, and makes it more likely customers stick to the same company. Different retention strategies make this concept implementable, such as loyalty programs or discounts for long time customers. "A loyalty program is a structured marketing effort which rewards, and therefore encourages, loyal behavior of customers, which is presumably beneficial to the vendor" (Enzmann & Schneider, 2005). A loyalty program is a good tool for increasing customer's purchases and for high customer retention levels (Julian, Ahmed, Wel, & Bojei, 2015).

"Companies that succeed in their e-business initiatives are adept at creating and maintaining a long-term sustainable relationship with loyal customers" (Cheung, Chan, & Limayem, 2005, p. 5).This exemplifies why research in the consumer continuance behavior (repurchase behavior) has risen sharply, and why research on this subject becomes increasingly salient. Moreover, the increased competition in the online shopping business compels companies to focus on maintaining existing customer bases and therefore cherish long-term relationships (Kim, Galliers, Shin, Ryoo, & Kim, 2012). Buchanan and Gillies elucidate that long-term customers are more profitable for six reasons: 1) regular customers place frequent, consistent orders and, therefore, usually cost less to serve; 2) longer-established customers tend to buy more; 3) satisfied customers may sometimes pay premium prices; 4) retaining customers makes it difficult for competitors to enter a market or increase their share; 5) satisfied customers often refer new customers to the supplier at virtually no cost; 6) the cost of acquiring and serving new customers can be substantial (p. 524). Repeated customer purchasing behavior and a long-term relationship focus reflect customer loyalty. An increase in customer loyalty in most cases means increased customer shopping value and drives customer retention. Tsai and Huang (2007) investigated the drivers of customer retention in an online setting executing an empirical study. The authors found that customers are motivated to remain with a provider due to one or more of four attachments: constraint-based (they "have to"), desire-based (they "want to"), customization-based (they can "specify modifications"), community-based (they "flock to") (Bansal, Irving, & Taylor, 2004; Bendapudi & Berry, 1997; Burnham, Frels, & Mahajan, 2003; Jones, Mothersbaugh, & Beatty, 2000; Wathne, Biong, & Heide, 2001). Tsai and Huang found that community-based drivers, desire-based drivers, and constraint-based drivers significantly influence repurchase intentions, while customization-based drivers only indirectly influence repurchase intentions. Community building, overall satisfaction, switching barriers and customization were respectively tested. Online business seems to have advantages over offline business but the emergence of e-commerce can make the retention of existing customers even more difficult (Tamaddoni Jahromi, Sepehri, Teimourpour, & Choobdar, 2010). Competition is just "one click away" and unprecedented customer empowerment, the churn rate of customers is likely to amplify (Peng, Quan, & Zhang, 2013).

Buchanan & Gillies (1990) emphasize that for a customer retention program to achieve maximum effect, communication towards their employees is crucial instead of only focusing on retention rates. Product/service quality can enhance customer loyalty and purchase intention, but it cannot always persuade a customer to purchase. High quality products are not always enough because customers' feelings are crucial before they close a transaction. "A customer's purchase intention depends on his or her positive evaluation of the overall net value of a product or service" (Kim, Galliers, Shin, Ryoo, & Kim, 2012, p. 375). Critical for a customer to shop online is the generated shopping value. Shopping value can be classified into two psychological categories; utilitarian and hedonic shopping value. Utilitarian shopping value refers to the degree customers feel their shopping goals have been accomplished, while hedonic shopping value is all about the fun and excitement of the shopping experience (Kim, Galliers, Shin, Ryoo, & Kim, 2012). Cheung, Chan, & Limayem (2005) reviewed literature written on online consumer behavior. They found certain theories that dominate the publications in this matter.

The Theory of Reasoned Action (TRA)(Ajzen & Fishbein, Belief, attitude, intention and behavior: An introduction to theory and research, 1975), Technology Acceptance Model (TAM)(Davis, 1989), Theory of Planned Behavior (TPB)(Ajzen, 1985) and to a lesser extent Innovation Diffusion Theory (IDT)(Rogers, 1995), and Expectation – Confirmation Theory (ECT)(Oliver, 1977) control the area of online consumer behavior research. ECT focuses on post-purchase behavior and is, particularly, in online repeat purchases literature a common theme. Oliver (1977) developed the concept of this model and suggests that if the perceived performance meets one's expectations, confirmation is formed. Confirmation, as part of utilitarian shopping value, will eventually lead to consumer satisfaction. On top of that both hedonic and utilitarian shopping value are assigned to influence customer satisfaction. Zeithaml (1988) states that individuals' satisfaction with online shopping is the overall assessment of what is received and what is given, and Kotler (1994) already emphasized that customer satisfaction is the key to customer retention (p.20).

2.2.4 Socio-economic influences on customer behavior

Socio-economic status reflects the position people have on the social ladder in combination with prestige and status. The societal status of a region is depending on several aspects of the civilians living in that region: education, income, and position on the labor market (Volksgezondheidenzorg, 2016). In scientific literature, different socio-economic factors influence customer behavior, which mainly has to do with income distribution. Nevertheless, income distribution on its own compels different causes, like residential location, education, and class position. The Marxist theory already stated that class position has a pervasive and systematic impact on income determination, which is at least as important as race, education, occupational status, or seks(Wright, 1979).

According to Leonhardt (2013) location plays a major role in income distribution. Even though location much less mattered to well-off families than for middle-class and poor families. Perotti (1996) divided society in equal and unequal societies, in which unequal societies tend to be politically and socially unstable. As a result lower rates of investment are signalized and therefore lower rates of growth. Lower rates of growth relate to lower rates of income.On the other hand, more equal societies have lower fertility rates and higher rates of investment in education, which is a stimulant for higher income (Perotti, 1996). Education gives the possibility to accumulate higher income, and therefore spending possibilities (Spilimbergo, Londoño, & Székely, 1999).

It is likely that customers from low valued socio-economic locations purchase orders with lower values than customers who come from high valued socio-economic locations. Simply because the expenditure possibilities are higher in regions with better socio-economic circumstances. Looking at the location, it is possible to check above-mentioned assumption by postal codes. Volksgezondheidenzorg, a Dutch institution, is responsible for mapping the SES score in the Netherlands. The SES score is the Social-Economic Status score which is an indication for the geographical distribution of socio-economic status.

Even though data at Webprint is not organized optimal, for this study, data is accessible for the variables "Order value" and "Order Postal Code". In order to give this study an explaining load or power as well, the relation between postal code and order value is tested. Further clarification on the structure of this test is accommodated in chapter three; methodology.

2.3 Theoretical framework

Small to medium-sized companies often tend to have large databases of transaction data without the knowledge, capacity or will to effectively analyze the database. This study tends to improve market intelligence, as part of business intelligence by analyzing the big data within these databases. In the light of focus, the field of market intelligence is tackled to improve business intelligence. The field of market intelligence is strongly associated with the discipline of customer behavior. Customer behavior insight should be improved by analyzing transaction data, results are supported and can be improved by different marketing concepts. These concepts are discussed thoroughly; relationship management, customer life cycle and customer retention.

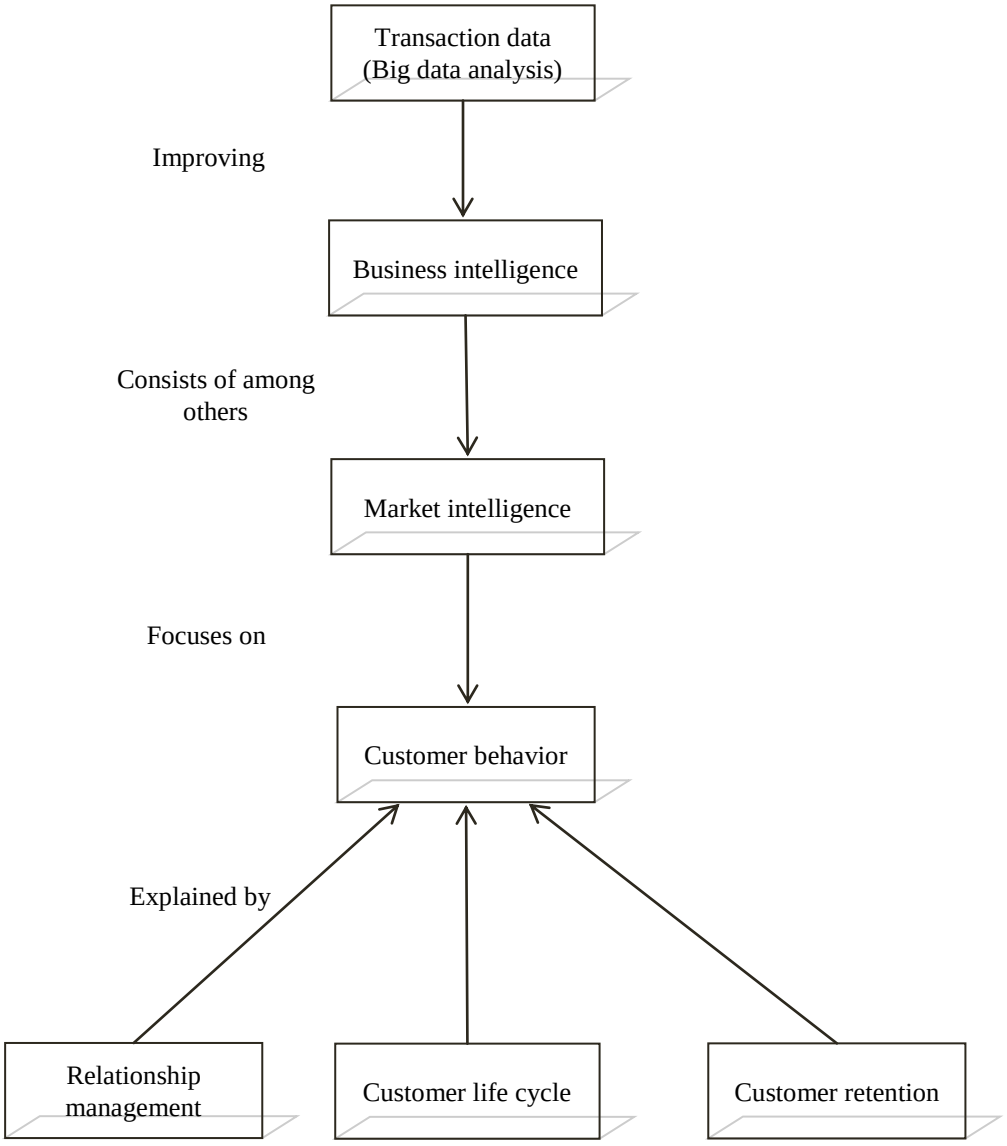


Figure 5: Theoretical framework

2.4 Research questions developed from the theoretical framework

The literature discussed the value of large databases of transaction data for creating business intelligence from a theoretical, non-practical point of view. It is known that customer retention, the customer life cycle and relationship management influence customer behavior in different ways. To find out how large databases of transaction data can improve business intelligence, customer behavior is a practical concept to be investigated. As mentioned before, market intelligence can improve by examining customer behavior more closely. At the same time, increasing market intelligence knowledge increases business intelligence. It completes the theoretical model which states that big data analysis improves business intelligence, which consists of among others; market intelligence. Market intelligence focuses primarily on customer behavior, which is influenced by three concepts; relationship management, the customer life cycle and customer retention, which also overlap mutually. Increasing knowledge, by big data analysis, on these three concepts can eventually improve business intelligence. As a consequence of the theoretical model, this study added a case which raises challenges or matters on the field of customer behavior, with the focus on improving business intelligence.

Customer retention is a comprehensive concept and focuses on repeat purchases. There are many different situations, and starting point from which to investigate the broad field of customer retention. In this case, the limited insight on important data such as order information, is the cause of research. Order information from customer retention after the first purchase. Understanding of important data can be crucial to the position of a company. As soon as deeper understanding of important order information is created, it would be preferable to know how to steer customers in a certain wanted direction (type of purchase). Therefore the following research question is to be investigated by means of the case:

How can returning customers be pushed in the desired direction?

Deeper understanding of customer behavior quickly points towards exploring the customer life cycle. Logically, companies like to know of each of their customers the phase of the customer life cycle their customers are located in. Since, insight in the customer life cycle can improve customer retention and, at least as important, improve and endure customer loyalty and satisfaction(Yue & Xiang, 2012). Customer profiling allows customers to be grouped by similar characteristics, which makes analyzing customers and locating them in the customer life cycle. The customer life cycle can improve the companies decision-making. Therefore the following research question is set up to be investigated though the case:

How can the customer life cycle be used for decision-making?

In order to establish solid relationships between companies and customers, customer behavior patterns from past purchases can be analyzed. After analyzing previous customers who have returned to converse again (customer retention), more insight can be created. As a consequence, newly acquired customers can be treated better, since the probabilities of the possible reactions are (more) clear. CRM influences customer behavior by, among others, creating more customer insight. Analyzing customer behavior from previous customers, newly acquired customers can be welcomed better and current customers can be treated better. As a result, the following research question is set up to be investigated by means of the case:

How can CRM influence customer behavior by analyzing past customer behavior patterns?

In addition to customer profiling, the socio-economic status is investigated to give the research its causal research part. Customer behavior depends, as self-evident, partly on the purchasing power of the relevant customer. It is plausible that customers who do not have the same level of socio-economic status, which covers among others education and income, will have different customer behavior patterns. That is exactly the reason why the next research question is formulated:

Does socio-economic status influence customer behavior?

The four research questions above will be answered by the help of some practical questions, which focus on the case used in this study. The case allows access to big data, which explains the introduction of practical questions (methodology). In the logical continuation of entering the practical side of the research, the case is examined in the next section.

3 Methodology

The methodology is divided into three parts. The first part considers the research design, divided in structure and the data collection methods. The first part focuses on the overall setup and intention of the research and is therefore more general of nature. The second part includes the introduction of the case Webprint and the third part focuses on the case-oriented methodology. Practical problems at the case are linked to the theoretical research questions. The section includes detailed descriptions and further deepening of the methodology used to solve or improve relevant situations at Webprint.

3.1 Research design

3.1.1 Structure

A research design is the framework or plan for a study, used as a guide in collecting and analyzing data. In fact there are three types of research designs; exploratory, descriptive and causal. In this study the emphasis is on the discovery of ideas and insight, which is directly linked to exploratory research (Babbie, 1998). It attempts to find the value of large databases of transaction data for creating business intelligence. On top of that, the study also includes descriptive research, since it attempts to find out how insight in customer behavior can be improved to eventually create business intelligence.

There are different types of exploratory research, such as literature research, experience surveys, focus groups and analyses of selected cases (Babbie, 1998). During this study exploratory research is executed in the form of literature search. It is a quick way of spitting in existing research published by highly reputable and respected authors. Different findings from several publications have been merged into one continuing framework.

Complementary, the descriptive part of the research contains cross-sectional analysis. Principally this means that variables of interest are viewed at a single point in time, in contrast to longitudinal studies. In this research different variables will be examined at a single point in time. To illustrate, customers will be analyzed at their first, second and third purchase, the frequencies will be described with regards to, for example, order value. More details on this subject can be found in section 4.2. To conclude, in longitudinal research the distances between the different purchases do have to be equal, which means this part of the research is not longitudinal.

A causal study typically takes the form of experiments in order to determine cause and effect. A causal relationship between two variables is investigated in the last research question; whether socio-economic status influences customer behavior. Each of the three research designs participate in this study, although the causal research design is by far the smallest part.

While qualitative research mainly deals with non-numerical data, quantitative research focuses on information that can be converted into numbers. This research includes quantitative research and qualitative research. The theoretical framework is set up by performing qualitative research. To provide more insight in customer behavior to increase business intelligence, quantitative research is executed. The research questions concerning the case involve quantitative research. Statistical findings are backed up or explained by literature, which again emphasizes the presence of qualitative research.

3.1.2 Quantitative data collection methods with secondary data

The research design is discussed and explained. The purpose of the research is to create and improve business intelligence by analyzing large databases of transaction data.

Secondary data is used in this study, which means the case did not gather the data for the immediate study at hand but for other purposes. Primary data, in contrast, are originated by the researcher for the purpose of the investigation at hand (Babbie, 1998). Since data from Webprint is stored in databases held for different purposes, the data used in this study can be labeled as secondary data. One of the reasons to use this kind of data is the cost and time advantage, but the main argument is the added value for the company by using their own data. After all, the company aims to improve data knowledge; business intelligence. Secondary data often does not fit immediately, since the data is not collected to serve research goals. The secondary data used can be labeled as internal data because of the source. "Internal data are those found within the organization for whom the research is being done, whereas external data are those obtained from outside sources" (Churchill G. A., 1999, p. 221).

The source of the secondary data used in this study is Copernica Marketing Software. "Copernica is an online multichannel platform to create professional marketing campaigns and engage the audience. Copernica offers manners to store data and really understand the customer" (Copernica, 2016). Disadvantageous, the system only allows limited possibilities of exporting data, and only in certain formats. In return, this means that the data exported demands refinement in order to be able to use data in this research. Otherwise it is not able to convert data to the statistical analysis program SPSS (quantitative research), in which issues are tested or checked and insight can be improved.

As mentioned before, the research will combine qualitative and quantitative research. Quantitative data collection methods can contain different types of research instruments. Records are used as research instrument during this study. A record refers to all numbers and statistics that institutions, organizations and people keep as a record of their activities. Sources can be educational records, hospital/clinic records and census data (Nalzaro, 2012). Census data is collected and offers multiple advantages; (1) records often are unbiased, (2) records often cover a long period of time, and (3) records are often inexpensive. The IT-department at the case secured that the data used is unbiased. The records are covered a long period of time (more than three years). But, the database will still not adapt itself to the research. In other words, refinement is needed to use the database in this study. It makes it possible to measure data.

Measurement is the process by which the researcher assigns specific numbers to the collected data (Doebelin, 1990). In scientific methodology literature four different levels of measuring are separated; nominal, ordinal, interval and ratio. Ratio level is defined as the highest level of measurement (Babbie, 1998). Section 4.2 allocates the different variables to measurement levels. There are two key elements in judging the instrument of measurement; reliability and validity. Validity refers to the extent to which an instrument measures what it is designed to measure (Carmines & Zeller, 1979).

In this study the first three orders of customers of Webprint are collected. It includes the variables order date, product category and order value, and is extracted from Copernica. There is an insignificant chance Copernica did not track customers sharp and precise, since the platform is globally accounted as credible. Partners of Copernica include Magento, Shopware, Lightspeed, Microsoft Dynamics and other reputable e-commerce platforms and CRM-systems. Validity is therefore guaranteed to be sufficient. Reliability refers to the degree of consistency and accuracy with which an instrument measures a variable. Customer data does not come from a sample, since all customer ID's stored in Copernica, from a certain time span, are included. Data is continuously collected and recorded, it therefore seems that reliability is subserved and adequate. The study meets the requirements of reliability and validity.

3.2 Exploring Webprint

Webprint is founded in 2004 by a private person originating from Saasveld, the Netherlands. The founder realized the process of printing personal photo's was too expensive, and change was established by founding Webprint. During the heat of the financial crisis (2009-2013) Webprint went through very profitable years. In order to cope with strong growth the need for financial injection was inevitable. As a consequence Webprint was incorporated by investment organization Infestos (and still is). "Infestos is an independent investment organization focused on sustainable investment of capital in participations, (monumental) real estate, project development and minority interests" (Infestos, 2016).

Webprint's main area of distribution is the Netherlands but the company has different consuming markets. Webprint.de, Webprint.fr, and prentu.co.uk are different domains in different countries. Each of the products sold are manufactured in Oldenzaal, where the company is currently located. Production is done in-house, the same counts for sales, marketing, design, and ICT. All activities are done in-house for different labels, including labels that do not carry the same name (print-things.nl). Webprint is also an external partner for different companies; these relationships characterize on two different levels. The first possibility for an external company is to have their own website, sell products, and send orders directly to Webprint who produces the products (API-production). Secondly, Webprint also maintains websites and manufactures products for other companies, so called white label websites. Under this construction the website-holder is also responsible for the traffic the web site generates; Trekpleister is an example of a white label construction with Webprint.

Webprint attempts to be recognized as the most complete online photo service in the Netherlands. The company is able to offer customers a high quality / high price - ratio because of the low manufacturing costs. Most of the products are produced automatically and mechanically from the start to the end of the production process. Exclusively activities that cannot be mechanized are manually produced by employees. Webprint has grown to a strong and multi-faceted online photo service that had substantial revenue growth rates; with a compound annual growth rate (CAGR) over the last four years of 24%. At the start of 2016 the company employed around 80 FTE. Traditionally the moment of the extreme sale values are to be forecasted; after the summer vacation and school breaks, during the November and December month, and around holidays as Valentine's Day and Eastern. During the peak times the company runs (close to) full capacity, in contrast to periods where circumstances are not extreme or exceptional. Producing on full capacity can lower product costs (Lipsey & Chrystal, 2007).

The company tracks (aggregately) how customers entered the website after purchasing a product. Channels as television, radio and retail-partnerships are labeled as offline channels. Online channels include online sales and are managed by Google Analytics. The e-commerce platform distinguishes six different channels: (1) e-mail, (2) organic, (3) SEA (Search Engine Advertising), (4) direct traffic, (5) referrals, and (6) affiliates. Each of the channels are managed differently and enquire other decision-making processes. Communication towards customers by e-mail does not enhance the same expectations, processes and strategy as communication by referrals, which illustrates the dynamic of the channels. The ability to track customers by Google Analytics is a tool to understand customer behavior better, in so customers can be targeted better.

The company puts a lot of attention to research which is mostly executed by students from the college or university nearby (Saxion University of Applied Sciences, Enschede and University of Twente). Recent studies were mostly related to efficiency on the work floor and brand awareness. In that case the focus on repeat purchases is unique to the company and value-added. While most studies focus on finding solutions for certain problems, this study will focus on analyzing data and declare the patterns in the data. Webprint offers a platform for students to develop oneself and at the same time complex issues are (partially) solved. Students can perform marketing- and market intelligence research at Webprint.

3.3 Webprint's products

Since customer behavior is studied at a case, the importance of product assortment knowledge is necessary. To start, the online photo service, Webprint, classified their products by six different categories (or worlds, as they are called on the work floor): photobooks, wall decoration, photo prints, photo gifts, calendars and agendas, and greeting cards. Each of the categories have different specifications and, as obvious, accommodate different products. Different statistics regarding product type are tracked. Firstly, the different products each have their own unique identification code, also a stock keeping unit (SKU) code. By the use of these codes Google Analytics is able to track different statistics concerning revenues, selling quantities, conversion rates, bouncing rates, and so on.

Figure 9 (Appendix I) displays different diagrams from each of the six categories. First of all, the stakes of the six categories in relation to each other are shown. The photo book category, the wall decoration category, and the photo print category each demand approximately 25% of the total revenues. As self-evident, this means these three product categories provide for 75% of total sales revenues. Greeting cards, photo gifts, and calendars and agendas provide the remaining 25% of the revenue streams displayed. In order to take a closer look at the specific categories and to support figure 9 (Appendix I), table1 is brought to life.

Table 1 reflects which stock keeping units are of great importance to the revenue share of the category in question. Landscape photo books do generate most of the revenue within the photo books category. Product-wise, the proportion of revenue coming from sales of the photobook landscape XL with basic paper is the largest. Canvas prints are generating the most revenue in the wall decoration category. Specifying further, two dimensions of canvas prints are most popular: 20x30x2 cm and 40x60x2 cm. Photo prints have the biggest quantity share in total sales while the revenue share in total sales is equal to wall decoration and photo books. Especially the photo print 13x10 cm is contributing hugely to the revenue stream coming from the photo prints. Photo gifts, calendars and agendas, and greeting cards are the remaining 25% of total revenues. Beer coasters are most popular based on revenues in the photo gifts category. Within calendars and agendas the calendars is the best-performing subcategory, and the calendar A4 Matte is the best performing calendar. Looking at the greeting cards product category there are two types of subcategories; postcards and greeting cards. Greeting cards are much more popular, in particular during December, because of the Dutch habit of sending greeting cards with Christmas. The size 15x15 cm in a package of ten pieces is most contributing to the revenues in the cards product category.

In order to gain more insight in the selling processes of Webprint, I would recommend to read the additional information, which can be found in Appendix II, section A, B, and C. In these sections the online channels are discussed thoroughly, and it will become clear how products are sold to customers in an online setting.

Insofar the study of the case. The case-oriented methodology is discussed next.

Product category	**Subcategory**	**Product**
Photo books	Landscape	Photo book landscape XL basic paper
Wall decoration	Canvas prints	Canvas 20x30x2 cm / Canvas 40x60x2 cm
Photo prints	Photo prints	Photo print 13x10 cm
Photo gifts	Beer coasters	Beer coasters (24pc)
Calendars and agendas	Calendars	Calendar A4 Matte
Cards	Greeting cards	Greeting cards 15x15 cm (10pc)

Table 1: Best sold products per product category

3.4 Case-oriented methodology

3.4.1 Case data

The data is generated from Copernica, as mentioned before. It lists around 407.000 different orders from customers who were registered and bought products from Webprint.nl. Customers who did not have three orders, in other words less than three, were excluded. The file was reduced to around 260.000 different orders. This file still included customers' fourth and older purchases. In order to set focus, these particular orders were excluded as well. The dataset is reduced to 37.668 unique customers. Each of the customers' first three orders were displayed, together with the order date, product category and order value. In conclusion, the dataset included 113.001 purchases made by 37.668 unique customers. At last, the residential information of the customers is generated from Magento database.

Copernica was only able to export data to a certain degree, which means that the format of the data was not prepared for the tests in this study. Transforming data into usable data was time-consuming and not convenient. Macro's in Microsoft Excel and calculations in formulas enabled data to be structured appropriately. Columns and rows needed to be exchanged, and data was continuously filtered in order to work with a complete and correct file.

3.4.2 Units of analysis

Each of the research questions, besides the non-case related research questions, include the same units of analysis. It involves customers from Webprint.nl, which means only from the Dutch domain, not from Germany, Belgium, the UK or France. The customer needs to be registered at Webprint.nl and will therefore have a customer account and profile. Besides that, to be part of this research, customers need to have fulfilled at least three separate purchases. The time span in which the units of analysis are observed or tracked is around three years. There are no other constraint variables for customers to be part of this research. To compete in this study the following criteria need to be matched:

- Customers from Webprint.nl
- Customer need to have an account at Webprint.nl, registration is necessary
- Customers need to have at least three purchases at Webprint.nl since June 2013

3.4.3 Tackling issues at the case

The different research and sub questions will be answered by help of the case Webprint. This study attempts to answer the research questions by examining actual and practical problems at the case. Product type and order value will be the main research variables. In the following sections it will become clear how the case Webprint is used to answer the research questions. Figure 6 shows the research structure graphically.

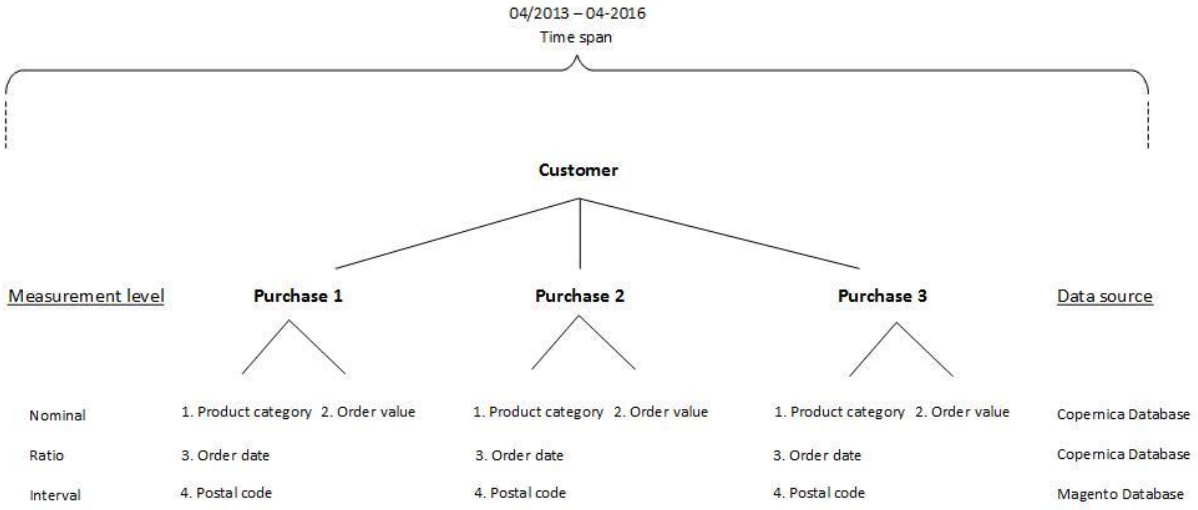


Figure 6: Research structure

How can returning customers be pushed in the desired direction?

The research question as outlined above will be answered by investigating customer behavior patterns at Webprint. One of the goals of Webprint is to increase their photobook sales. The company aims to sell more photobooks because this product type associates with the values of the company, and should become the core business. As mentioned before, the data covers the first three purchases of the customers. Therefore, the purchases before the purchase of a photobook will be analyzed. This means when the second purchase is a photobook, the preceding purchase is examined. At the same time, when the third purchase contains a photobook, the preceding two purchases are analyzed. As a consequence, cross tables will be set up which increase insight into the customer behavior before purchasing a photobook. As soon as the customer behavior before purchasing a photobook is analyzed, it might become more clear for Webprint to steer customers towards the purchase of a photobook (desired direction). Knowledge on the field of customer retention is gained.

How can CRM influence customer behavior by analyzing past customer behavior patterns?

Previous customers' behavior patterns will be analyzed to uncover the influence of CRM on customer behavior. To address this research question, attempted is to find the customer roads with the lowest and highest frequency. In the context of this study, a customer road refers to the possibilities customers have in the first three purchases on the variable product type. As self-evident different combinations are possible and by determining the least and most used customer roads, retention can be enhanced. The customer roads provide insight in the motivation of customers to retain, it offers CRM a chance for a long-term relationship. CRM motivates customers to shop for more and talk positively to their peers about the products provided (Ozok et al, 2007).

How can customers be profiled to position them in the customer life cycle?

After examining customer roads more closely, the case will select customer behavior patterns which are the most interesting for their business. Webprint can choose this pattern with different motives, which can be in the light of revenues or the future direction of the company. As a consequence customers, selected for their interesting behavior, are profiled on base of average subtotal, coupons codes used, average amount paid and residence. It offers a starting point for locating customers in the customer life cycle, which creates deeper understanding of customer behavior and eventually improves business intelligence.

Does socio-economic status influence order value?

In addition to customer profiling, a test is enhanced to examine the relationship between socio-economic status and order value. A random sample of n=50 is taken, which is declarable. The number of units investigated in this test is not high, because this test is performed to give an indication of the relationship. It attempts to show what the database is capable of generating. The socio-economic status (SES) statistics for the Netherlands, per postal code, is available at the Sociaal en Cultureel Planbureau. The output contains scores that are determined by a few characteristics; education, income and position on the labor market. To be more precise, the SES score is calculated by four percentages: average income per postal code, average percentage of residents with low income, average percentage low educated residents and the average percentage of residents who do not work. By the use of a factor-analysis of these characteristics the final SES score is determined. A status score of 0 (SES = 0) implies an average score; the average in the Netherlands overall tracked is zero. The higher the SES the better the societal status, reversely the lower the SES the worse the societal status.

3.4.4 Drawback of the dataset

The exported data from Copernica includes orders with one or more products. Technically it is not possible to filter the orders, mainly due to the size of the sample. But, this would be preferable since an order could then be assigned to one product category, instead of more than one. The data export shows orders with more products from different product categories as one. In each of the cases this will be the most expensive product. That is a drawback on the data used in this research. However, studies at Webprint emphasize that most of the orders include one product. In the last study performed around 80 percent of the orders only included one product. In conclusion; this drawback is to be taken into account but only a small percentage of the data can be distorted.

Secondly, the most expensive product is put forward, this product will be in most cases the motivation behind the order. It downsizes the influence of the drawback discussed above.

4 Results

4.1 Structure of the results section

The dataset consists of 37.668 unique customers who have at least three purchases in the store Webprint.nl. The main variable in this study is product category, whereby order value, order date and postal codes are the other three variables. Findings are shown on the base of product category frequencies from purchase one, two, and three. Statistics around the main research variable provide insight and understanding of the purchase rank and purchase content of customers at their first three acquisitions. A structured way of increasing customer insight is developed and can be translated in different chapters, categorized by customer retention, CRM and the customer life cycle. The structure of the results are discussed next, starting with customer retention.

Customer retention
In section 4.2.1 purchase one is fixed; calculated is the possibility the second order contains products from the same product category as in purchase one. Section 4.2.2 applies a different approach, and examines the other possibilities besides the same product category. This means the section tends to find out what the other possibilities are, if the second purchase does not contain products from the same product category as purchase one. Cross tabulations are used to gain customer insight and database understanding.

The same counts for the visual representation of the structure of the calculations for sections 4.2.3 and 4.2.4. In section 4.2.3 and 4.2.4 the same happens as in section 4.2.1 and 4.2.2. However, there is one difference, the first sections show findings on purchase one and two, while the latter focuses on purchase two and three. Cross tabulations are used to gain customer insight and database understanding. It is tested for each of the seven product categories.

Section 4.2.5 combines different sections in so the possibility a customer buys three products from identical product categories can be calculated. This is calculated for each of the seven product categories.

Section 4.2.6 focuses on the photo books category. Photo books are set up as fixed in the second purchase, from there on purchase one (the preceding purchase) is variable. It means this section attempts to find out the distribution of purchases before the purchase of a photo book. In section 4.2.7 the same is tested, but focuses on purchase two and three, in which purchase three is fixed (photo book category) and purchase two is variable.

Customer relationship management
Section 4.3 examines the most frequent customer roads from the first three purchases. This is also done for the least frequent customer roads. Ten of the most and least frequent customer roads are displayed in section 4.3. By the use of SPSS the function AGGREGATE is used. The function creates variables holding statistics over cases (Van den Berg, 2013). It enables analyses on combinations of different variables with different measurement levels. Purchase one, purchase two, and purchase three were assigned to be BREAK variables, in so the customer road were displayed.

The customer life cycle
Section 4.4 shows findings based on customer profiles from the most interesting customer roads found. It is an addition to section 4.3 in which the customer roads are revealed.

Section4.5, in which the order value of purchase one, two, and three is summed up, investigates if socio-economic status influences order value.

4.2 Calculations in the field of customer retention

4.2.1 Customers who buy the same product in purchase one and purchase two

To calculate the amount and shares of possibility of equality between purchase one and two on the base of product category, cross tabulations are set up (appendix III). In appendix III, two cross tabulations are displayed; both tables have the same content but table 13 shows row percentages while table14shows column percentages.

The percentage of the cases displayed in table 2, shows the possibility of a second purchase to be a product from the business category if the first purchase contained a business item. Self-evidently, this counts for the percentage of the cases of each of the seven product categories. The amount of cases includes the statistical presence of the second purchase to be the same (on the base of product category) as the first purchase.

Product category		Amount of cases showing purchase two after one	Percentage of the cases
Purchase one	Purchase two		
Business	Business	126/198	50,4
Photobook	Photobook	5.214/8396 etc	62,1
Wall decoration	Wall decoration	3.577	53,7
Photo	Photo	12.312	72,3
Photo gift	Photo gift	584	41,4
Calendar and agenda	Calendar and agenda	390	44,2
Greeting cards	Greeting cards	1.699	56,0

Table 2: Calculations on purchase one and two

Five out of seven assumptions are above 50%, two assumptions showed results less than 50%. The results of customers who bought products from the same products category in purchase one and two were: business (50,4%), calendars and agendas (44,2%), cards (56%), photo books (62,1%), photo gifts (41,4%), photos (72,3%), and wall decoration (53,7%). The strongest relationship was displayed by the photos category.

4.2.2 Customers who do not buy the same product in purchase one and two

This section shows results per product category from customers that bought a fixed first purchase, but the second purchase is variable and not the same as the first purchase. Results of the seven product categories are shown in appendix III, section B, and in table three below. Findings worth mentioning are summed up next. But, before mentioning results on this topic, the following underlying statistics has to be kept in mind:

In overall order frequencies and in order frequencies per purchase, photos are on top, secondly photo books, and thirdly wall decoration items.

In the following results, purchase one is fixed and second purchases that contain an item from the same product category as in purchase one are excluded. This also counts for the displayed percentages.

Business items are by far the least ordered product, both in terms of order frequencies and in terms of revenues, this is why business items are only considered in the analysis as fixed first purchase, and not as variable second purchase.

Looking at the business product category as first purchase, wall decoration items are the most ordered products in the second purchase (29,8%). Secondly products from the photos category (23,4%), and thirdly products from the photo books category (20,2%). Calendars and agendas are the least popular buy in purchase two after buying a business product in purchase one (2,4%).

Examining purchase two after buying a product from the calendars and agendas item in purchase one, products from the photos category have been ordered in most cases (31,8%). Secondly products from the photo books category (25,6%), and thirdly products from the wall decoration category (22,1%). Photo gifts are the least popular buy in purchase two after buying a calendars and agendas item in purchase one (6,5%).

Considering purchase two after buying a product from the greeting cards category in purchase one, products from the photos category are ordered the most in purchase two (34,7%). Secondly, products from the photo books category (27,9%), and thirdly products from the wall decoration category (22,5%). Photo gifts are the least popular buy in purchase two after buying a product from the greeting cards category in purchase one (6,4%).

Investigating purchase two after buying a product from the photo books category in purchase one, products from the photos category are by far the most ordered products in purchase two (45,7%). Secondly, products from the wall decoration category (26,5%), and thirdly products from the greeting cards category (16,3%). Photo gifts are the least bought products in purchase two after buying a photo book in purchase on (5,4%), although calendars and agendas as well have a low possibility to be bought in purchase two (5,6%).

Studying purchase two after buying a product from the photo gifts category in purchase one, products from the photos category are bought the most in purchase two (28,7%), but the difference with second place photo books is very small (27,9%). Products from the wall decoration category are the third best ordered products after buying a photo gift in purchase one (25,3%). Calendars and agendas are the least bought products in purchase two after buying photo gift in purchase one (5,1%).

Examining purchase two after buying a product from the photos category in purchase one, products from the photo books category have been ordered in most cases (40,6%). Secondly, wall decoration items (31,6%), and thirdly products from the greeting cards category (15,6%). Calendars and agendas showed the least possibility to be bought in purchase two after buying a photo in purchase one (5,1%).

Looking at the wall decoration product category as first purchase, photos are the most ordered products in the second purchase (41%). Secondly, products from the photo books category (31,1%) and thirdly are the greeting cards (14,2%). Calendars and agendas were the least ordered products in purchase two after buying a wall decoration product in purchase one (5,3%).

In short, products from the photos category were the most ordered second purchase after buying a calendar and agenda (31,8%), greeting card (34,7%), photo book (45,7%), or a wall decoration item (41%). On top of that, wall decoration items were the most ordered second purchase after buying a business product in the first purchase. After buying a photos item in the first purchase, photo books were mostly ordered in the second purchase (40,6%). Calendars and agendas and photo gifts were the least popular products in the second purchase, after every product category in purchase one. Photo gifts performed better overall, except for greeting cards as first purchase, photo gifts showed 6,4% as possibility for a second purchase and calendars and agendas 7,5%.

		If purchase one exists of products from this category:						
	Product category	Business	Greeting cards	Calendars and agendas	Photo books	Photo gifts	Photos	Wall decoration
Purchase two exists of:	Business	x	1%	0.6%	0.5%	0.6%	0.9%	0.7%
	Greeting cards	16.9%	x	13.4%	16.3%	12.4%	15.6%	14.2%
	Calendars and agendas	2.4%	7.5%	x	5.6%	5.1%	5.1%	5.3%
	Photo books	20.2%	27.9%	25.6%	x	27.9%	40.6%	31.1%
	Photo gifts	7.3%	6.4%	7.3%	5.4%	x	6.3%	7.7%
	Photos	23.4%	34.7%	23.4%	45.7%	28.7%	x	41%
	Wall decoration	29.8%	22.5%	29.8%	26.5%	25.3%	31.6%	X
Total		100%	100%	100%	100%	100%	100%	100%

Table 3: Calculations on non-identical products in purchase one and two

4.2.3 Customers who buy the same product in purchase two and purchase three

Does the third purchase contain an item from the same product category than the second purchase?

To calculate the amount and shares of possibility of equality between purchase two and three on the base of product category, cross tabulations are set up (appendix III). In appendix III, two cross tabulations are displayed; both tables have the same content but table 15 shows row percentages while table 16 shows column percentages.

The percentage of the cases displayed in table 4, shows the possibility of a third purchase to be a product from the business category if the second purchase contained a business item. Self-evidently, this counts for the percentage of the cases of each of the seven product categories. The amount of cases includes the statistical presence of the third purchase to be the same (on the base of product category) as the second purchase.

Product category		Amount of cases	Percentage of the cases
Purchase two	Purchase three		
Business	Business	118	52
Photobook	Photobook	487	61,5
Wall decoration	Wall decoration	1.951	53,5
Photo	Photo	5.441	73,2
Photo gift	Photo gift	564	39,9
Calendar and agenda	Calendar and agenda	11.649	43,5
Greeting cards	Greeting cards	3.517	54,5

Table 4: Calculations on purchase two and three

Five out of seven assumptions are above 50%, two assumptions are below 50%. The results of customers who bought products from the same products category in purchase two and three were: business (52,0%), calendars

and agendas (43,5%), cards (54,5%), photo books (61,5%), photo gifts (39,9%), photos (73,2%), and wall decoration (53,5%). The strongest relationship was displayed by the photos category.

4.2.4 Customers who do not buy the same product in purchase one and two

This section shows results per product category from customers that bought a fixed second purchase, but the third purchase is variable and not the same as the second purchase. Results of the seven product categories are shown in appendix IV, section B, and in table five below. Findings worth mentioning are summed up next. But, before mentioning results on this topic, the following underlying statistics has to be kept in mind:

In overall order frequencies and in order frequencies per purchase, photos are on top, secondly photo books, and thirdly wall decoration items.

In the following results, purchase two is fixed and third purchases that contain an item from the same product category as in purchase two are excluded. This also counts for the displayed percentages.

Business items are by far the least ordered product, both in terms of order frequencies and in terms of revenues, this is why business items are only considered in the analysis as fixed second purchase, and not as variable third purchase.

Looking at the business product category as second purchase, photos are the most ordered products in the third purchase (29,4%). Secondly, products from the wall decoration category (28,4%), and thirdly products from the greeting cards category (20,2%). Calendars and agendas are the least bought products in purchase three after buying a business item in purchase two (2,8%).

Examining products from calendars and agendas category in purchase two, purchase three mostly consists of products from the photos category (30,8%). Although photo books are not far behind (29,2%). Wall decoration is third, and shows an percentage of 17,4%, which is hardly more than the percentage greeting cards (15,5%). Photo gifts are by distance the least ordered products in purchase three, after buying a calendar and agenda's item in purchase two (6,8%).

Considering purchase two, including a greeting card, photos are ordered most frequently in purchase three (36,4%). Secondly are the photo books (28%) and thirdly wall decoration items (20,7%). Photo gifts are ordered the least as third purchase, given that the second purchase contains a greeting card (5,8%).

Investigating items from the photo book category as second purchase, the third purchase is in most cases photos (45,6%). Secondly, wall decoration items with 26,1% of the third purchases. Thirdly, the greeting cards category with 16,1% of purchase three. Photo gifts are again the least bought product in the third purchase (5,2%).

Studying purchase three after buying a product from the photo gifts category in purchase two, products from the photos category are bought the most in purchase three (27,9%), but the difference with second place photo books is very small (26,6%). Products from the wall decoration category are the third best ordered products after buying a photo gift in purchase two (26,6%). Calendars and agendas are the least bought products in purchase three after buying photo gift in purchase one (6,6%).

Examining purchase three after buying a product from the photos category in purchase two, products from the photo books category have been ordered in most cases (38,4%). Secondly, wall decoration items (31,6%), and thirdly products from the greeting cards category (17,1%). Calendars and agendas showed the least possibility to be bought in purchase three after buying a photo in purchase two (6,0%).

Looking at the wall decoration product category as second purchase, photos are the most ordered products in the third purchase (43,1%). Secondly, products from the photo books category (30,0%) and thirdly are the greeting cards (15,2%). Calendars and agendas were the least ordered products in purchase three after buying a wall decoration product in purchase two (5,3%).

In short, products from the photos category were the most ordered third purchase after buying a business item (29,4%), calendar and agenda (30,8%), greeting card (36,4%), photo book (45,6%), photo gifts (27,9%), or wall decoration item (43,1%). On top of that, photo books were the most ordered third purchase after buying a photos product in the second purchase. Calendars and agendas and photo gifts were the least popular products in the third purchase, after every product category in purchase two. Calendars and agenda's performed better as third purchase after the second purchase as a greeting card (8,1% - 5,8%) and after photo books as second purchase (6,3% - 5,2%).

		If purchase two exists of products from this category:						
	Product category	Business	Greeting cards	Calendars and agendas	Photo books	Photo gifts	Photos	Wall decoration
Purchase three exists of:	Business	x	1%	0.3%	0.7%	0.9%	0.6%	0.6%
	Greeting cards	20.2%	x	15.5%	16.1%	12.5%	17.1%	15.2%
	Calenders and agendas	2.8%	8.1%	x	6.3%	6.6%	6%	5.3%
	Photo books	13.8%	28%	29.2%	x	25.5%	38.4%	30%
	Photo gifts	5.5%	5.8%	6.8%	5.2%	x	6.3%	5.9%
	Photos	29.4%	36.4%	30.8%	45.6%	27.9%	x	43.1%
	Wall decoration	28.4%	20.7%	17.4%	26.1%	26.6%	31.6%	x
Total		100%	100%	100%	100%	100%	100%	100%

Table 5: Calculations on non-identical products in purchase two and three

4.2.5 Customers who buy the same product in purchase one, two and three

In the previous two sections, calculations are made on purchase one and two, and on purchase two and three. In this section the results of the two previous chapters will be combined. Results will show which product category displays the strongest relation from purchase one, to purchase three. The calculation can be made, since the possibilities of identical product categories in purchase one to purchase two are known, and the same counts for purchase two and three. The following calculations are made on the base of table 2 and table 4 in section 5.2.1 and 5.2.3:

Product category			Calculations	
Purchase one	Purchase two	Purchase three	Computation	Outcome
Business	Business	Business	0.504 x 0.520	0,262
Photobook	Photobook	Photobook	0.621 x 0.615	0,382
Wall decoration	Wall decoration	Wall decoration	0.537 x 0.535	0,287
Photo	Photo	Photo	0.723 x 0.732	0,529
Photo gift	Photo gift	Photo gift	0.414 x 0.393	0,163
Calendar and agenda	Calendar and agenda	Calendar and agenda	0.442 x 0.435	0,192
Greeting cards	Greeting cards	Greeting cards	0.560 x 0.545	0,305

Table 6: Calculations on purchase one, two and three

If someone buys a product from the business product category, the possibility the customer will buy the same in purchase two and purchase three is 0,262. This means that if a customer buys a business item in purchase one, in 26,2% of the cases the same customer will buy a business product in the second and third purchase.

The photo product category displays the strongest outcome: 52,9%. Products from the photo gifts category displays the lowest outcome of 16,3%. In short, customers who buy photo products as first purchase are most inclined to buy the products from the same category in purchase two to purchase three. Form each of the seven categories, photo product buyers are most tended to buy products from the same product category in purchase two and three. While, photo gift buyers are most tended to buy products from another product category in the follow-up purchases.

4.2.6 The purchase preliminary to the purchase of a photo book (purchase one and two)

In this section results will be shown on the base of the product category photo books. In order to find out the preceding purchase of the purchase of a photo book, cross tabulations are set up, appendix V (section A and table 17) displays the complete results. The calculations are made to find out what purchase is made before the purchase of a photo book, this means that in first instance purchase one and two are discussed. In part 5.5.2 the finding from purchase two and three are provided.

In contrast to the previous calculations, where the first purchase was set up as a fixed product category, in this case the last purchase is set up as fixed. This means that in this section the second purchase is fixed (photo books category), and the first seven possibilities are calculated. The results have been checked and remarkable findings are shown next.

In 59% of the time the first purchase was a product from the photo books category if the second purchase contained a photo book. In section 5.2.1 this relation is already shown, which means this result is not that surprising. But, much more interesting might be, to know what customers buy before buying a photo book in the second purchase, if we exclude the first purchase of a photo book. Of the remaining 41%, the stake of the photos category was 21,6%. Only 10,8%, which is quite a difference, consisted of wall decoration items in the first purchase. The remaining part is filled up by greeting cards (4,2%), photo gifts (2,6%), calendars and agendas (1,4%), and business items (0,3%).

4.2.7 The second purchase preliminary to the third purchase of a photo book

In contrast to the previous calculations, where the first purchase was set up as a fixed product category, in this case the last purchase is set up as fixed. This means that in this section the third purchase is fixed (photo books category), and seven possibilities are calculated for the second purchase (preliminary to the third purchase). The results have been checked and remarkable findings are shown next (Appendix V, section B, table 16).

In 61,4% of the time the second purchase was a product from the photo books category if the second purchase contained a photo book. In section 5.3.1 this relation is already shown, which means that this result is not new or surprising. But, much more interesting might be, to know what customers buy before buying a photo book in the second purchase, if we exclude the first purchase of a photo book. The photos product category showed 18,5% stake in the second purchase preceding the third purchase containing a photo book. Wall decoration scored 10,3% and the remaining product categories as following: greeting cards (5,1%), photo gifts (2,4%), calendars and agendas (2,1%), and business items (0,2%).

In comparison to section 5.5.1 the stake the photos category has in the preceding purchase is in the third purchase lower than in the second purchase. Wall decoration items are reduced by half looking at the different tests. On top of that, photo gifts and business items remained stable, while photo gifts showed a stronger interest.

But, we have to keep in mind that on the test with fixed product category purchase three does not say anything about the first purchase, since that purchase is left out of consideration in section 5.5.2.

4.3 Calculations in the field of customer relationship management (CRM)

4.3.1 Customer roads with the lowest and highest frequency

In the context of this study the term customer road refers to the possibilities customers have in the first three purchases. This means customers have the option to choose between products from seven product categories in each of the three purchases. This means there are in total 7^3 possible customer roads a customer can take, which are 343 possible customer roads. The ten highest and lowest customer roads found are examined next, results can be found in appendix VI, table 19 and 20.

The highest frequency of purchases is sequenced by Photos - Photos - Photos. This string is followed up by Photo book – Photo books – Photo books, which is followed up by Wall decoration – Wall decoration – Wall decoration. In the ten highest customer road frequencies the product category photos is expressly evident. On top of that, many combinations with photo books and photos are present. Possible explanations for these results are discussed in the discussion part (chapter six).

The lowest frequency of purchases is sequenced by Calendars and agendas – Photos – Wall decoration. What is notable is that products from the calendars and agendas category are well represented in this list. In most of the ten lowest frequencies of purchases three different product categories are noticed. In short, it does not seem reasonable to think that calendars and agendas, photos and wall decoration items will be ordered in this sequence (in the first three purchases). On top of that, besides the explicit presence of calendars and agendas in this list, photo gifts are also well represented in the list. Possible explanations for these results are discussed in the discussion part (chapter six).

4.4 Calculations in the field of the customer life cycle

In the previous section the customer roads with the highest and lowest frequency are displayed. The goal of Webprint is to increase the share of photo books in total sales. Therefore it is interesting to find out how the most interesting customers are identified, and what the characteristics of these customers are. Most interesting, in this case, relates to the terminus of the purchase of a photo book. In this case, the strings photos – photo books – photo books, and photos – photos – photo books appear to be often present. The first-mentioned string is found 807 times and the second one is found 899 times. These purchase ranks are interesting to check out whether other remarkable findings on behalf of the identity of customers can be found. The complete results are found in table 22 and 23, located in appendix VII.

Photos – Photos – Photo books	Purchase one	Purchase two	Purchase three
Average subtotal	€ 11,13	€ 8,47	€ 26,21
Amount of coupon codes used	20	16	9
Average amount paid	€ 13,49	€ 10,78	€ 21,38

Table 7:Averages on sample string 1

Photos – Photo books – Photo books	Purchase one	Purchase two	Purchase three
Average subtotal	€8,80	€ 29,20	€ 37,57
Amount of coupon codes used	22	36	11
Average amount paid	€ 10,56	€ 25,07	€ 29,09

Table 8: Averages on sample string 2

The next figures include the distribution of the sample over the different location within the Netherlands. Two figures are set up to increase insight in to the customers in the sample. The sample consists of fifty customers, each of them are living in the Netherlands.

Table 9 shows the location of the customers who were sampled per province.

Photos - Photos - Photo books	
Province	**Amount of customers**
Zuid-Holland	9
Noord-Holland	10
Noord-Brabant	8
Overijssel	0
Gelderland	6
Groningen	3
Friesland	1
Drenthe	0
Flevoland	3
Zeeland	3
Utrecht	5
Limburg	2

Table 9: Postal codes string 1

Table 10shows the location of the customers who were sampled per province.

Photos – Photo books - Photo books	
Province	**Amount of customers**
Zuid-Holland	14
Noord-Holland	1
Noord-Brabant	8
Overijssel	5
Gelderland	5
Groningen	0
Friesland	3
Drenthe	2
Flevoland	3
Zeeland	1
Utrecht	3
Limburg	5

Table 10: Postal codes string 2

Table 9 and 10 are quite different to each other and provide interesting insights which are discussed in the discussion part of the research. Remarkable at first insight is the dominance of Zuid-Holland in the light of customers in the first and second purchase rank. On top of that, the low presence of customers in the sample from Drenthe and Overijssel in purchase rank one stood out, while the second purchase rank shows low presence of customers located at mainly Groningen and Zeeland.

4.5 Socio-economic Status and order value

The customer data in our study, coming from Webprint, is generated from April 2013 to April 2016. The most recent SES-scores are from 2014, and do not change heavily in a short period, according to the institution. Therefore, in this section the SES-scores of 2014 are compared to the customer data from Webprint. The complete data used in this sample can be found in Appendix VIII, table 24.

A single regression analysis is performed, with p=0,05. Order value functions as dependent variable and SES-score as independent variable, since the test is about uncovering the influence of the SES-score on order value. The following hypotheses are set up:

H0: There is no relation between the SES-scores and order value.

H1: There is a relation between the SES-scores and order value.

Model	Sum of squares	Df	Mean Square	F	Sig.
Regression	23,965	1	23,965	,169	,683
Residual	6795,736	48	141,578		
Total	6819,701				

Table 11: ANOVA-table, dependent variable

The ANOVA-table makes clear that the model is not significant. It means that the model is no addition to the study, if there was no model the situation would be the same, it makes no difference. In short, the sig = ,683 this means the model is not significant.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(constant)	18,865	1,752		10,767	,000
	SES	,627	1,525	,059	,411	,683

Table 12: Coefficients-table, dependent variable

The coefficients-table declares the influence of the independent variable to the dependent variable. The influence of the SES-score on order value is slightly positive because of the positive Beta.

It can be stated with 95% certainty that there is no effect from the SES-score on the order value. H0 is accepted, and H1 is not accepted.

5 Discussion

This part will discuss the findings at the case Webprint with regards to the research questions. These questions have been set up based on three fields as influencers of customer behavior: customer retention, customer relationship management (CRM) and the customer life cycle. On each of the three fields the practical findings and implications will be discussed, and will be attached to the theory.

5.1 The field of customer retention

5.1.1 Relevant problems on identical purchases

Chapter five, section one, discusses the first fourteen assumptions gathered from the practical questions above. An overview of the fourteen relevant problems that are checked are shown in figure 7.The findings are discussed next.

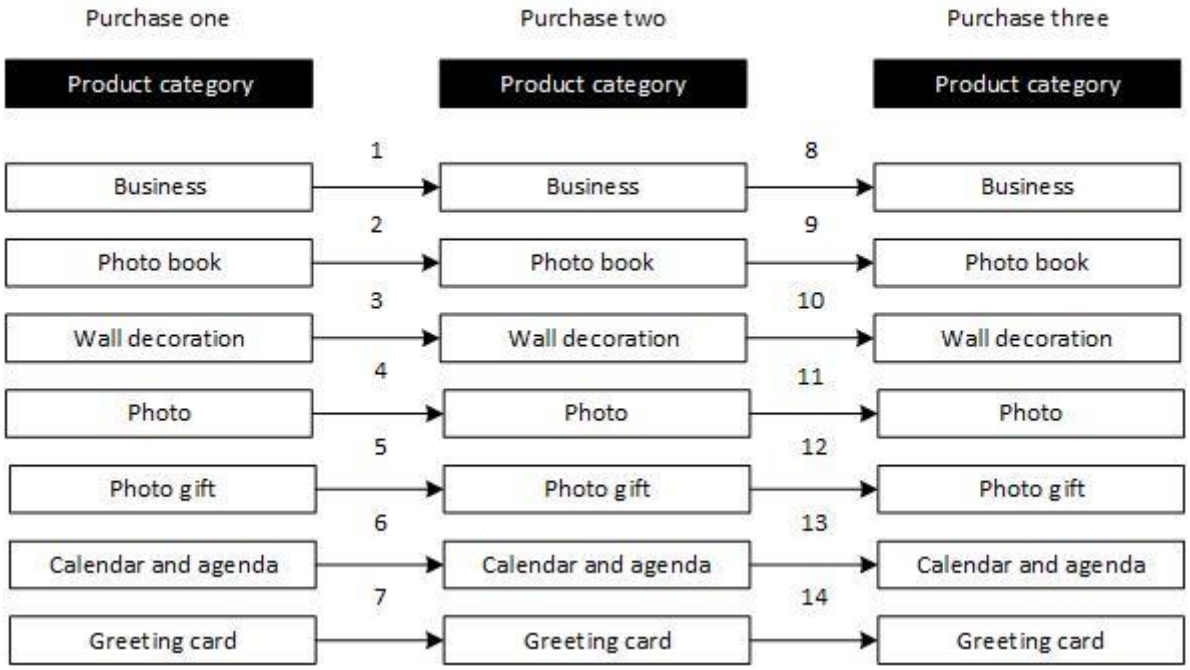


Figure 7: An overview of assumptions

As mentioned before in the results part, fourteen relevant problems were tested from purchase one to two, and from purchase two to three. Numerically, the share of photos category is 25%, the same counts for the photo book category and for the wall decoration category. This means that the share of photo gifts, greeting cards, business products, and calendars and agendas is covering up the remaining 25%. Knowing this, it is remarkable to see that products from the business category, and products from the cards category do in most cases show an identical follow-up purchase, whether this is the second or third purchase. The photo gifts category showed a weak relation on behalf of the identical products, the same counts for the photo calendars and agendas category. Customers of these two categories do not seem to buy identical products in the follow-up purchase as often as the other product categories.

Section 4.4 provides insight in the possibility a customer buys three products from identical product categories in the first three purchases. Together with the discussion above, it is not surprising that that chance is the highest at the photo product category. Remarkably is the relative high outcome of the business product category, the customers only have very limited product options in this category, but on the other hand, business buyers are different from ordinary customers, since these products are not for personal use.

5.1.2 Discussing findings on non-identical purchase one and two and purchase two and three

If customers buy products from the calendars and agendas category as first purchase, numerically photos is on top at the second purchase, which is not remarkable since the dominant number of photos sold. But, looking at the percentage of photos column wise, even though the total photos sold in purchase two after buying a calendars and agendas item (1,0%) is not very high. In terms of percentages column wise photo gifts are most popular after buying a calendar or agenda. If customers buy products from the cards category as first purchase, numerically photos is on top at the second purchase, which is not remarkable since the dominant number of photos sold. In column percentages it is observed that photos definitely does not have the top spot. Photo gifts (6,1%) and calendars and agendas (9,0%) do have the highest column wise percentages in purchase two. Remarkably, photos has in percentages the lowest column wise stake in purchase two after buying a cards item. If customers buy products from the photo books category as first purchase, numerically photos is on top at the second purchase, which is not remarkable since the dominant number of photos sold. In percentages column wise, calendars and agendas (15,9%) and cards (14,4%) are at the top. If customers buy products from the photo gifts category as first purchase, numerically photos is on top at the second purchase, which is again not remarkable since the dominant number of photos sold. In percentages photos is the lowest again, just as by calendars and agendas and cards as first purchase, while calendars and agendas has the highest stake column wise (3,8%). If customers buy products from the photos category as first purchase, numerically photo books is on top at the second purchase, while in percentages wall decoration items (22,7%), photo books (21,6%) and calendars and agendas (21,6%) are at top. If customers buy products from the wall decoration category as first purchase, numerically photos is on top at the second purchase, which is not remarkable since the dominant number of photos sold. In percentages photo gifts (16,7%) and calendars and agendas (14,6%) are at top after buying a wall decoration item as first purchase.

If customers buy products from the calendars and agendas category as second purchase, numerically photos is on top at the third purchase together with photo books (table 13),which is not remarkable since the dominant number of photos,and secondly photo books, sold. In percentages column wise the photo gifts category is on top (3,2%). If customers buy products from the cards category as second purchase, numerically photos is on top at the third purchase together with photo books, which is again not remarkably. In percentages column wise the calendars and agendas category, just as in purchase one and two, is on top(10,1%). If customers buy products from the photo books category as second purchase, numerically photos is on top at the third purchase together with wall decoration, which is again not remarkably. In percentages column wise calendars and agendas is on top, just as in purchase one and two (16,5%). If customers buy products from the photo gifts category as second purchase, numerically photos, wall decoration, and photo books are on top at the third purchase. In percentages calendars and agendas are on top with a stake of 4,3% column wise. If customers buy products from the photos category as second purchase, numerically photo books and wall decoration are on top at the third purchase. In percentages calendars and agendas are on top with a stake of 4,3% column wise. If customers buy products from the wall decoration category as second purchase, numerically photos are on top at the third purchase. In percentages photo gifts are on top with a stake of 13,5% column wise.

5.1.3 Discussion on the photo books category

An important observation has to be made before deeply analyzing the purchases before the second purchase of a photo book. Namely to state that the photo books, photos, and wall decoration category counts for around 90% of all first, second, and third purchases made. It is therefore plausible that Webprint will profit the most of an analysis based on these three product categories.

If purchase two is a photo book, in most cases (59% of the time) a photo book has been ordered in purchase one. In 21,6% of the time, photos items have been ordered in purchase one. Finally, in 10,8% of the time, a wall decoration item is ordered in purchase one. In order to compare; purchase three as a photo book in 61,4% precedes a photo book in purchase two. In 18,5% of the time a photos product is ordered in purchase two preceding a photo book in purchase three.

10,3% of the third purchases including a photo book contained a wall decoration item in purchase two. The percentage of photos stake in purchase two preceding purchase three as a photo book is lowered from 21,6% to 18,5%. On top of that, if customers bought a photo book in their first order, and buy another one in their second

purchase (59% of the time), the urge to buy another one rises. Since, the stake of purchase two as a photo book in purchase three as a photo book has increased to 61,4%, which is a rise of three percent points. If the first purchase of customers is a photos item, in 21,6% of the cases a photo book will be ordered in the second purchase. While if the second purchase is a photos item the chance of buying a photo book is lowered to 18,5% of the cases.

5.1.4 Discussing theory in addition to the case

By analyzing customer behavior patterns information on customer satisfaction is generated. It is plausible that unsatisfied customer will probably not order again, and customers which are analyzed in this study have at least three purchases. It implies the units of analyses have been satisfied for at least these three purchases. If a customer purchases photos in the first order, there is a high chance the second order will be photos as well. When the second purchase includes photos, there is a high chance the third order will be photos again. In conclusion this means that by looking closely at the most common patterns, satisfied customers can be identified. As a consequence, the company is able to analyze which products it aims to sell. By gaining knowledge on customer behavior patterns the case is aware of the highest chance a customer is going to buy a certain product, at least on the variable of product category. It can therefore steer customers towards the sale of products from preferred product categories, after analyzing customer behavior patterns before the purchase of a product from the preferred product category.

High customer satisfaction will positively influence customer loyalty. Customer loyalty can be improved by use of customer retention programs. Customer retention programs improve loyalty and therefore improve customer retention (Enzmann & Schneider, 2005). Loyal and long-term customers are more profitable for multiple reasons according to Buchanan and Gillies (1990); they tend to buy more and these customers may pay premium prices. Analyzing customer behavior patterns can positively influence customer satisfaction, which gives insight into customer behavior and imply how to steer customers into a certain direction.

5.2 The field of customer relationship management (CRM)

5.2.1 Discussing the customer road

Additionally to the goal of this research of providing more insight into the database, the customer road from purchase one to purchase three has been mapped. Table 17 reveals that the most frequent consecutive relationship between these purchases is: Photos- Photos – Photos. Secondly are the consecutive purchases of Photo Books – Photo Books – Photo Books. Thirdly, are the consecutive purchases of Wall Decoration – Wall Decoration – Wall Decoration. These three findings are quite expectable in relation to the analyzed data before, it is shown that photo books and wall decoration, and mainly photos had a strong relationship. 59% of the customer who bought a photo book in the first purchase bought a photo book in the second purchase. On top of that, customer who bought a photo book in purchase two bought in 61,4% of the cases a photo book in purchase three. Wall decoration on identical purchases in purchase one and two offers 54,5%, and in purchase two and three 54,5%. Photos on the other hand shows a stronger relationship; 77,3% in purchase one and two, and 74,8% in purchase two and three. It illustrates the top is covered by the photos category in this matter. The fourth most frequent consecutive purchase string is Cards – Cards - Cards. It makes clear that customers who buy a card, wall decoration, photo book, or photo item in most cases do want to buy a product from the same product category in the next two purchases. It also means that customers who have bought one of these products will in most cases be interested in the same products. Emails sent by Webprint often include two promotions, based on the results above customers who have ordered one of these products, should always receive a main promotion with these products. Customers who have bought a photo book as first purchase at Webprint, should receive a main promotion or news item regarding photo books, since that is, according to these statistics, what they are interested in. On top of that, photos buyers and photo books buyers are interchangeable according to the customer road. Customers are able to switch easily from photos to photo books, and from photo books to photos, no matter the order of purchases. This indicates that photo books buyers are at any time interested in photos, and photos buyers are at any time interested in photo books. Email marketing can use this information to steer customers the direction wanted.

On the contrary, there are customer roads with the lowest frequency. In this case Calendars and Agendas – Photos – Wall Decoration and Calendars and Agendas – Photos – Photo Books are the least frequent purchase orders. Based on table 18 it can be concluded that customers who have made the first two purchases as in this table will most probably not buy the third one. To be precise, if customers bought a calendars and agendas item and secondly a photos item, they will probably not be interested in photo books or wall decoration items as third purchase in history. If two consecutive purchases as in table 18 are detected, communication towards these customers should not be related to the third purchase in this figure; simply because customers do not take this path in most cases. Customer do therefore not prefer the routes in table 18. Communication in different stages can influence the pattern.

5.2.2 Discussing theory in addition to the case

In addition to customer retention, CRM is concerned with creating a long-term relationship with customers, and motivating them to come back to shop for more (Ozok et al; 2007). It focuses on the basic need of establishing a relationship with customers, otherwise customers will not retain. The most common customer roads are identified and analyzed. The theory of CRM also emphasizes that the focus of marketing must shift from the realization of transactions to building a customer relationship, focusing on maintaining and enhancing the relationship. It will create customer satisfaction and loyalty(Xiong & Liu, 2011). Critical in creating customer relationships is the treatment, or communication. Customers will then feel appreciated and start connecting to the brand or company.

The case studied provides information on the customer roads offered. It enhances deeper understanding of the possibility of starting a relationship with customers, whether it is useful and worthwhile. In the case part, many customer roads have been displayed with the according frequencies. It provided insight in the communication after buying certain products. If a certain customer road is rarely ended with a wall decoration item, then the customer, with the same first two purchases, should not be communicated towards buying that item next. The customer roads provide information on communication, in the light of possible scenarios. It is more likely that a customer road which often appears has more potential to a relationship than a customer road which hardly appears. This is because there is no data available on the last customer road, which makes it hard to imagine what the customer desires communication-wise. Eventually, customer insight is created by focusing on customer relationship management.

5.3 The field of the customer life cycle

5.3.1 Customer profiling

A random sample (n=50) is used to gain more insight into profiles of customers at two of the most interesting purchase ranks or customer roads found. In this case that is Photos – Photos – Photo books and Photos – Photo books – Photo books. This choice is made because of the central role the photo books category plays in this research, since the company wishes to increase the share of photo books in total sales. Section 4.5 shows that both above-mentioned customer roads were found to be frequently present and both illustrated the goal of photo books in purchase three.

First of all customer profiles relating the Photos – Photos – Photo books customer road are discussed. The subtotal of order one is € 11,13 and decreased in order two to € 8,47. As we know, in order one and in order two products from the photos category are ordered, this means that this development is notably. The amount paid did decrease as well from € 13.49 to € 10.78, while the coupon codes used are less; from 20 in purchase one to 16 in order two. This indicates that the amount paid should increase, but the amount paid was simply lower which diminishes the effect of the coupon codes in order one, and made the amount paid in purchase to decrease. In purchase three even less coupon codes are used, this is the purchase of a photo book. This could be due to the fact that coupon codes of photos are more issued than coupon codes for photo books. Nevertheless, a decrease of coupon codes from purchase one to purchase three is present in this customer road. The amount paid substantially increased to € 21,38 which is not surprising since photos are beneath 20 cents per piece while photo

books are available from a price of € 4,95. With regards to the location of customers in this customer road, most of the customers are located at Noord-Holland (20%). Remarkably, not even one customer is located at the province Drenthe, and only one in Friesland. Besides Noord-Holland (20%), Noord-Brabant (16%), and Zuid-Holland (18%) are well presented in the random sample of the customer road Photos –Photos – Photo books.

Secondly customer profiles relating to the Photos – Photo books – Photo books customer road are examined. The subtotal of photos is € 8,80 which is lower than the in the previous customer road's purchase one (€ 11,13), but higher than (or comparable to) the previous customer road's purchase two (€8,47). Looking at the last two purchases of photo books, in order two and order three of the Photos - Photo books – Photo books customer road, the subtotal increased together with the amount paid. In any regular case this would be a normal development, but in the case of Webprint this is worth mentioning because of the use of coupons. Remarkably is the decrease of coupon codes used in purchase three, from 36 to 11. Examining the postal codes of customers in the random sample, the presence of Zuid-Holland increased in comparison to the previous customer road (from 9 to 14). But, more striking is the lack of customers from Noord-Holland. If the random sample is representable we can say that there are hardly any customers from Noord-Holland having the purchase pattern Photos - Photo books – Photo books, which is quite striking because of the high density of civilians in this province. The north of the country is not presented well either, as is also shown in the previous customer road. Drenthe and Zeeland are also not evidently present. Most of the customers, based on the both customer roads tested, are located at Zuid-Holland and Noord-Brabant.

5.3.2 Discussing theory in addition to the case

The customer life cycle is one of the most classic models in customer relationship management and refers to the whole process that a customer goes through when considering, purchasing, using, and maintaining loyalty to a product or service (Yue & Xiang, 2012). The customer life cycle contains of four stages with each stage having different characteristics of customer relationships. If customers are profiled based on their customer roads, it can be analyzed which customer road is most interesting due to order value, location or coupon codes used. If a certain uninteresting customer road often occurs at a certain location, we might be sure that there has not been much bounding yet. Since the customer would have been directed towards a more interesting customer road, after its first purchase.

Preferably, customers stay in the formative and stable period as long as possible, since customers are then very loyal to the company and purchasing levels are high, the same counts for customer satisfaction. If high coupons are used it is almost sure that customers are located in the initial period, since high promotions are often used to attract new customers. It is often an expensive way to attain new customer relationships. Customer behavior insight can be created by profiling customers in so customer roads get divided by customer characteristics, and can be identified in the customer life cycle.

5.4 Other findings, limitations, future research and contribution

5.4.1 The influence of socio-economic status on order value

The socio-economic status (SES) score is calculated on the base of education, income, and position on the labor market. The position on the labor market mainly depends of education and income. Spilimbergo et al. (1999) concluded education accumulates higher income, and higher income would increase spending. In this case it is clear that there is no relation between SES-scores and the spending of money on products from Webprint.

This can be due to many causes. First of all, it is not unrealistic to think that people do not have to spend their money towards products of Webprint, or products from the photo-service industry, since they can spend it on other products. Therefore, it would not really matter whether someone has much money to spend. Secondly, Webprint does not even have ten per cent market share, which means that if people wanted to buy photo-products it would be plausible people spent it at another photo-service company.

Thirdly, the sample size can be the cause to the absence of a relation. Since the results showed a small link from the SES-score towards the order value, it is also likely that if the sample size would be larger, the results would be different. The results are not purposeless, since it encourages further research on this subject, because of the absence of an explanation and the lack of a large sample.

5.4.2 Limitations and future research

There are different limitations that need to be reported. First of all in this study only the first three purchases are considered. It allows to set a focus in a research but also limits the study since the goal is to have customers repeat purchases as often as possible. Mainly due to technical restrictions this choice is made. Moreover, only the variable product category is deeply discussed. In first instance, different variables as promotions, order value, and the influence of channels was to be discussed. In the context of focus and the possibilities of (or lack of) Copernica to match data with Google Analytics and Magento the boundaries were set to product categories, order value, order data, and postal codes.

In addition to the last limitation, for future studies it is interesting to eventually match data between different systems or platforms in so deeper analyses can be made. It enables different variables to participate in the study, such as the influence of promotions. Furthermore, different or complementary insights can be provided if more than three consecutive purchases are provided. It enables to map the complete customer road, and provides even more customer behavior patterns than this study attempted to do.

5.4.3 Contribution to previous research

Different articles are already written on the subject of business intelligence. This thesis distinguishes itself not from the rest not only by adding an e-commerce case in a niche market, but the theoretical concepts which are used, relationship management, customer life cycle and repurchase behavior added a remarkable dimension. And, because of the tight focus market intelligence implies, the research adds value to previous research. This research demonstrates the importance of big data analysis for business intelligence in a medium-sized e-commerce company, where previous research mainly focused on large enterprises. Thereby, the study also shows that with relatively low resources customer behavior patterns can be exposed, which can improve customer retention.

6 Conclusion

What is the value of large databases of transaction data for creating business intelligence?

This study illustrates that the value of large databases of transaction data lies in uncovering customer behavior patterns in large databases. Transaction data, as the term indicates, contains lots of customer specific data. The customer specific data can be analyzed in order to improve customer behavior insight. Customer behavior insight is recognized as part of market intelligence (Crowley, 2007), and market intelligence is part of the overarching business intelligence (Hakanen, 2014). As the literature already stated, big data often hides patterns or relations between different variables and, online businesses acknowledge that understanding their data is crucial to their competitive position (McAfee et al., 2012). Big data analysis creates customer behavior insight, to improve customer retention.

In order to improve customer retention, relationships with customers have to be formed and maintained. Relationship management is therefore an important concept. The customer life cycle is one of the key aspects of relationship management. It attempts to classify customers on base of their purchases. It tends to maintain and improve loyalty to a product or service (Yue & Xiang, 2012). Analyzing transaction data will gather information on the customer life cycle and can contribute to customer retention, and even customer acquisition. The customer life cycle classifies customers in order to be able to find the best approach to groups of customer with the same customer behavior. In the case Webprint, customers are also classified on base of their purchase behavior, the analysis is done to improve customer retention. The field of customer retention, customer relationship management and the customer life cycle are key to answering the second research question.

What customer behavior patterns are present at Webprint.nl and what points of action do the findings unveil?

The theoretical framework insist that transaction data can improve business intelligence. Business intelligence consists of market intelligence, which focuses on customer behavior. Customer behavior can be explained by three concepts. The three concepts are applied to the case and offers results which can be achieved by analyzing transaction data. On top of that, each of the following results can be backed up by doing tests. For example, AB tests could enforce the results.

Customer retention
In each of the seven product categories it can be stated that the consecutive purchase has the highest chance, whether this is the second or third purchase, to include products from the same product category as the purchase before. This relation was the strongest at the photos category and the least strong at the photo gifts category. In short; customers tend to buy a product from a product category they ordered the purchase before. It is therefore advisable for Webprint to at least communicate the same products from the same product category as the preliminary purchase. Simply, because most customers buy products from the same product category as the preceding purchase. The next pattern worth mentioning is related to the purchase of a photo book. The strength of the urge to buy a photo book only increases more after a photo book has been bought in the first purchase. The urge of a customer, who bought a wall decoration item or photo in the first or second purchase, only declines in the following purchases. Nevertheless, the same increases for calendars and agendas, and for greeting cards, and remains equal for items from the photo gift category.

To advise Webprint, it would be best to communicate non-customers photo book products, because the urge to buy a photo book only increases after buying a photo book in the first purchase. By communicating this, fixed photo book buyers are created. Besides that, if a customer's first purchase is a calendar and agenda item, or a greeting card, the urge to buy a photo book only increases as well. This means that these groups of buyers are sensitive to the purchase of a photo book in the second and third purchase. Communication toward these groups, on the area of photo books, will influence their ability of buying photo books in the following purchases positively.

By analyzing customer behavior patterns information on customer satisfaction is generated. By gaining knowledge on customer behavior patterns, the case is aware of the highest chance a customer has to buying a certain product. It can therefore steer customers towards the sale of products from preferred product categories, after analyzing customer behavior patterns before the purchase of a product from the preferred product category.

Customer relationship management
CRM focuses on the basic need of establishing a relationship with customers, otherwise customers will not retain. The most common customer roads are identified and analyzed. Especially communication could improve by analyzing the different frequencies of customer roads. Webprint should focus on customer roads which often appear, such as Photos-Photos-Photos, since this customer road has more potential to sturdy customer relationships than roads which hardly occur. By improving the relationship between companies and customers, customer insight is created. The eventual goal is to create a solid customer relationship.

The customer life cycle
If customers are profiled based on their customer roads, it can be analyzed which customer road is most interesting due to order value, location or coupon codes used. The theory tries to position customers in the life cycle in order to maintain, create or improve customer relationships. Preferably, customers stay in the formative and stable period as long as possible, since customers are then very loyal and purchasing levels are high. If uninteresting customer roads often occur at a certain location, there has not yet been any bounding. These customer roads do not deserve the focus. Especially the customer road Photos-Photobooks-Photobooks is very interesting due to the increasing order value. It indicates these customers have reached the formative period in the customer life cycle. It means that customers are loyal and their purchasing levels are high. Customers with these customer roads deserve the companies attention.

On top of that, results reveal that socio-economic status does not affect order value significantly. A small positive relationship has been found, but did not appear to be significant. Spilimbergo et al. (1999) implicate that better education possibilities positively related to better income, which would increase spending. This concept is accommodated in the SES-score, which effect is tested on order value. The findings suggest the alternative hypotheses to be applied, which means there is no significant relationship.

Core competence of Webprint
Moreover, the advantage or core competence Webprint has in comparison to smaller and new companies is the access to lots of customer data. This competitive advantage is according to this study not exploited enough. This is why further research should focus on restructuring the different databases into a more structured database, in which enables Webprint to analyze customers precisely and individually. It will increase the insight the company has in customer behavior, the life cycle can be increased and customer life time value can be optimized. Gao (2014) emphasizes: “If you know the likely life cycle or pattern of a customer, changes can be made to customer communications or marketing strategy to try to optimize the length of time and the value that a customer brings to the business”. Webprint can improve its customer behavior insight with relatively low resources, as this research showed. It should focus on analyzing customer behavior patterns in so customer retention can be improved. By analyzing these patterns customers can be classified, as the customer life cycle theory implies. But, before that, the databases should be structured and centralized in one database.

Implications
First of all, there was not any research to be found in the online photo service industry. The core concepts used in this study are customer retention, customer relationship management and the customer life cycle as influencers of customer behavior insight. It offers a unique combination with the online photo service industry, because these concepts are mainly used to study other business. The unique perspective is created by the combination of the three concepts and the e-commerce.

Besides that, there has not been any research which focused this deeply, with big data, on customer behavior patterns in an e-commerce setting. This study has shown that customer behavior patterns can be studied by help of customer retention, CRM and the customer life cycle as support bars. The theory in this research offers an unique self-made framework, which guides the whole study. Much more comprehensive frameworks are often created as a guidance for the study, which can complicate the whole story.

The study provided different points of attention for further research. On each of the three fields discussed, lots of deeper studies can be performed. Especially because the different databases are not centralized yet, and data seem to be unused. More than three purchases per customer should be investigated, variables as order value should be engaged to each of the fields, profiling customers should engage more variables except the three used, and finally socio-economic status can be investigated by taking a larger sample. It will give the study a better fundament.

To conclude, by analyzing large databases of transaction data more closely, important aspects of customer retention, CRM and the customer life cycle can be identified, which are valuable for creating business intelligence. Webprint should handle their data differently, in so the data gets more structure and can be much more valuable to the company.

7. References

Ailawadi, K. L., Neslin, S. A., & Gedenk, K. (2001). Pursuing the value-conscious consumer: Store brands versus national brand promotions. *Journal of Marketing, 65(1)*, 71-89.

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. *Action Control*, 11-39.

Ajzen, I., & Fishbein, M. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research.* Reading, MA.: Addison-Wesley.

Arnold, T. J., Fang, E., & Palmatier, R. W. (2011). The effects of customer acquisition and retention orientations on a firm's radical and incremental innovation performance. *Academy of Marketing Science*, 234-251.

Babbie, E. R. (1998). *The practice of social research.* Belmont: CA: Wadsworth publishing company.

Ball, D., Coelho, P., & Vilares, M. (2006). Service personalization and loyalty. *Journal of Services Marketing, 20(6)*, 391-403.

Bansal, H., Irving, P., & Taylor, S. (2004). A three-component model of customer commitment to service providers. *Journal of the Academy of Marketing Science 32(3)*, 234-250.

Bashara, R. (2015, 04 20). *What Is Referral Traffic in Google Analytics.* Retrieved 03 09, 2016, from Small Business: smallbusiness.chron.com

Bendapudi, N., & Berry, L. (1997). Customers' motivations for maintaining relationships with service providers. *Journal of Retailing 73(1)*, 15-37.

Berman, B. (2006). Developing an effective customer loyalty program. *California Management Review, 49(1)*, 123-148.

Buchanan, R., & Gillies, C. (1990). Value Managed Relationships: The Key to Customer Retention and Profitability. *European Management Journal*, 523-526.

Burnham, T., Frels, J., & Mahajan, V. (2003). Consumer switching costs: a typology, antecedents, and consequences. *Journal of the Academy of Marketing Science 31(2)*, 109-126.

Carmines, E. G., & Zeller, R. A. (1979). Reliability and validity assessment. *Sage publications (Vol. 17)*.

CBS. (2015). *CBS: Ruim 10 miljoen online shoppers*. Retrieved from CBS: http://www.cbs.nl/nl-NL/menu/themas/vrije-tijd-cultuur/publicaties/artikelen/archief/2015/ruim-10-miljoen-online-shoppers.htm

Chaudhure, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM, 54 (8)*, 88-98.

Chen, H., Chiang, R. H., & Storey, V. (2012). Business intelligence and Analytics: From Big Data to Big Impact. *MIS quarterly, 36 (4)*, 1165-1188.

Cheung, C., Chan, G., & Limayem, M. (2005). A Critical Review of Online Consumer Behavior: Empirical Research. *Journal of Electronic Commerce in Organizations*, 1-19.

Churchill, G. (1999). Data Analysis: Discriminant, Factor, and Cluster Analysis. In G. Churchill, *Marketing Research: Methodological Foundations* (pp. 1-991). Fort Worth: The Dryden Press.

Churchill, G. A. (1999). *Marketing Reserach.* Madison: The Dryden Press.

Copernica. (2016, 01 01). *Mission and Values*. Retrieved 04 07, 2016, from Copernica: www.copernica.com/en/about-us/mission-and-values

Cornish, R. (2007). *Statistics: 3.1 Cluster Analysis.* Mathematics Learning Support Centre.

Davenport, T. H. (2006). Competing on Analytics. *Harvard Business Review (84:1)*, 98-107.

Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly 13(3)*, 319-340.

De Bock, K., & Poel, D. (2011). An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. *Expert Systems with Applications, 38*, 12293-12301.

Dobney. (2016, 2 2). *Market Intelligence.* Retrieved 2 2, 2016, from Dobney: www.dobney.com/market_intelligence.htm

Doebelin, E. O. (1990). *Measurement systems: application and design.* New York: Mc Graw-Hill.

Enzmann, M., & Schneider, M. (2005). Improving Customer Retention in E-Commerce through a Secure and Privacy-Enhanced Loyalty System. *Information Systems Frontiers*, 359-370.

Gao, K. (2014, 1 1). *The Ultimate Guyid to Email Marketing.* Retrieved 2 29, 2016, from Comm100: http://emailmarketing.comm100.com/email-marketing-ebook/

GoogleAnalytics. (2016, 1 4). *Google Analytics Help*. Retrieved 1 4, 2016, from Google: http://www.google.com/intl/nl_ALL/analytics/features/index.html

Gresty, M. (2014). Market intelligence gathering in executive search firms. *Business Information Review*, Vol. 31(4), 206-211.
Gupta, S., & Hee-Woong, K. (2007). The moderating effect of transaction experience on the decision calculus in on-line repurchase. *International Journal of Electronic Commerce (12)1*, 127-158.
Gupta, S., & Kim, H. (2007). The moderating effect of transaction experience on the decision calculus in on-line repurchase. *International Journal of Electronic Commerce*, 127-158.
Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends. *Computer & Operations Research, 34*, 2902-2917.
Hakanen, P. (2014, September 13). *Tag Archives: Knowledge management.* Retrieved Juni 14, 2016, from Strategic Design: www.hakanen.eu
Hennig-Thurau, T., & Klee, A. (1997). The impact of customer satisfaction and relationship quality on customer retention: A critical reassessment and model development. *Psychology & Marketing*, 737-764.
Hwang, H., Jung, T., & Suh, E. (2004). An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Expert systems with applications*, 181-188.
Infestos. (2016, 2 2). *Infestos Sustainable invenstment*. Retrieved 2 2, 2016, from Infestos: www.infestos.com
Jones, M., Mothersbaugh, D., & Beatty, S. (2000). Switching barriers and repurchase intentions in services. *Journal of Retailing 76(2)*, 259-274.
Jonger, J., Piersma, J., & Van den Poel, D. (2004). Joint optimization of customer segmentation and marketing policy to maximize long-term profitability. *Expert Systems with Applications*, 159-168.
Julian, C., Ahmed, Z., Wel, C., & Bojei, J. (2015). Discriminant Analysis of Antecedents of Customer Retention in Malaysian Retailing. *Journal of Transnational Management*, 190-204.
Kim, C., Galliers, R., Shin, N., Ryoo, J., & Kim, J. (2012). Factors influencing Internet shopping value and customer repurchase intention. *Electronic Commerce Research and Applications, 11(4)*, 374-387.
Komiak, S., & Benbasat, I. (2006). The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS Quaterly, 30(4)*, 941-960.
Kotler, P. (1994). *Marketing management. Analysis, planning, implementation, and control.* Englewood Cliffs: NJ: Prentice-Hall.
Kushwaha, T., & Shankar, V. (2013). Are Multichannel Customers Really More Valuable? The Moderating Role of Product Category Characteristics. *Journal of Marketing*, 67-85.
Lee, E.-J., & Park, J. (2009). Online service personalization for apparel shopping. *Journal of Retailing and Consumer Services, 16*, 83-91.
Leonhardt, D. (2013). In climbing income ladder, location matters. *New York Times*, 23.
Lipsey, R., & Chrystal, K. (2007). *Economics.* Oxford: Oxford University Press.
Liu, Y. (2007). The long-term impact of loyalty programs on consumer purchase behavior and loyalty. *Journal of Marketing, 71*, 19-35.
Magento. (2016, 1 4). *About Magento*. Retrieved 1 4, 2016, from Magento Inc.: http://magento.com
Mahadevan, B. (2000). Business models for Internet-based e-commerce. *California management review*, 55-69.
Manning, H., Bodine, K., Temkin, B., & Amato, M. (2005, 03 01). *Best and worst of site design.* Retrieved 03 04, 2016, from Forrester Research: www.forrester.com/research/document/excerpt/0,7211,36074.html
McAfee, A., Brynjolfsson, E., Davenport, T., Patil, D., & Barton, D. (2012). Big data. *The management revolution. Harvard Bus Rev, 90(10)*, 61-67.
Mitra, A., & Mitra, S. (2012, 07 12). Applied Multivariate Analysis. *Mod-01 Lec-26 Cluster Analysis*. Kanpur: IIT Kanpur.
Muhamad, A. (2013, 11 22). *Quantitative Data Analysis*. SlideShare.
Mulvenna, M., Anand, S., & Büchner, A. (2000). Personalization on the Net using Web mining: introduction. *Communications of the ACM*, 122-125.
Nalzaro, L. (2012, 06 09). *Methods of Data Collection*. SlideShare.
Negash, S. (2004). Business intelligence. *The communications of the Association for Information Systems, 13(1)*, 54.
NetworkSolutions. (2015). *What is Ecommerce?* Retrieved 12 29, 2015, from networksolutions: http://www.networksolutions.com/education/what-is-ecommerce/
N'Goala, G., & Cases, A. S. (2012). Monitoring Customer Relationship in E-Commerce: Which Drivers and Which Effects on Buying Behavior. *Recherche et Applications en Marketing (English Edition)*, 95-117.
Oliver, R. L. (1977). Effect of Expectation and Disconfirmation on Postexposure Product Evaluations - an Alternative Interpretation. *Journal of Applied Psychology, 62(4)*, 480.
Ozok, A., Oldenburger, K., & Salvendy, G. (2007). Impact of Consistency in Customer Relationship Management on E-Commerce Shopper Performances. *Journal of Organizational Computing and Electronic Commerce, 17(4)*, 283-309.

Peng, J., Quan, J., & Zhang, S. (2013). Mobile phone customer retention strategies and Chinese e-commerce. *Electronic Commerce Research and Applications*, 321-327.
Perotti, R. (1996). Growth, income distribution, and democracy: what the data say. *Journal of Economic growth, 1(2)*, 149-187.
Phan, D., & Vogel, D. (2010). A model of customer relationship management and business intelligence systems for catalogue and online retailers. *Information & Management, (47)2*, 69-77.
Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data, 1(1)*, 51-59.
Rangaswamy, A., & Van Bruggen, G. H. (2005). Opportunities and challenges in multichannel marketing: An introduction to the special issue. *Journal of Interactive Marketing, 19(2)*, 5-11.
Reichheld, F., & Sasser, W. (1990). Jr. Zero defections: Quality comes to services. *Harvard Business Review*, 105-111.
Reichheld, F., & Schefter, P. (2000). E-Loyalty: Your secret weapon on the Web. *Harvard Business Review, 78(4)*, 105-113.
Rogers, E. (1995). *Diffusion of innovations.* New York: 12.
Rust, R., & Zahorik, A. (1993). Customer satisfaction, customer retention, and market share. *Journal of retailing, 69(2)*, 139-215.
Spilimbergo, A., Londoño, J. L., & Székely, M. (1999). Income distribution, factor endowments, and trade openness. *Journal of development Economics, 59(1)*, 77-101.
Srinivasan, S., Anderson, R., & Ponnavolu, K. (2002). Customer loyalty in e-commerce: an exploration of its antecedents and consequences. *Journal of retailing*, 41-50.
StatArena. (2012, 07 15). Cluster Analysis In SPSS (Hierarchical, Non-hierarchical & Two-step).
Tamaddoni Jahromi, A., Sepehri, A., Teimourpour, B., & Choobdar, S. (2010). Modelling customer churn in a non-contractual setting: the case of telecommunications service providers. *Journal of Strategic Marketing*, 587-598.
Tsai, H.-T., & Huang, H.-C. (2007). Determinants of e-repurchase intentions: An integrative model of quadruple retention drivers. *Information & Management 44*, 231-239.
Tushman, M. L., & O'Reilly, C. (2002). *Winning through innovation.* Cambridge: MA: Harvard Business School Press.
Van den Berg, G. R. (2013, 10 07). *SPSS AGGREGATE Command.* Retrieved 05 03, 2016, from SPSS Tutorials: http://www.spss-tutorials.com/spss-aggregate-command/
Verhoef, P. (2003). Understanding the Effect of Customer Relationship Management Efforts on Customer Retention and Customer Share Development. *Journal of Marketing*, 30-45.
Volksgezondheidenzorg. (2016). *Sociaaleconomische status regionaal.* Retrieved 06 27, 2016, from volksgezondheidenzorg: https://www.volksgezondheidenzorg.info/onderwerp/sociaaleconomische-status/cijfers-context/opleiding#definities
Wathne, K., Biong, H., & Heide, J. (2001). Choice of supplier in embedded markets: relationship and marketing program effects. *Journal of Marketing 65(2)*, 54-67.
Watson, H. J., & Wixon, B. H. (2007). The current state of business intelligence. *Computer, 40(9)*, 96-99.
Winer, R. (2001). A framework for customer relationship management. *California management review 43(4)*, 89-105.
Wright, E. O. (1979). *Class structure and income detern~ii~ation.* New York: Academic.
Xiong, Y., & Liu, H. (2011). Research on Customer Relation Management in e-Commerce Environment. *International Conference on Management and Service Science*.
Yue, W., & Xiang, C. (2012). A multi-interests model of Recommendation System based on Customer Life Cycle. *Intelligent Computation Technology and Automation*, 22-25.
Zhang, Y., Fang, Y., Wei, K.-K., Ramsey, E., McCole, P., & Chen, H. (2011). Repurchase intention in B2C e-commerce - A relationship quality perspective. *Information & Management (48)*, 192-200.

8. Appendices and additional information

I. Product assortment Webprint

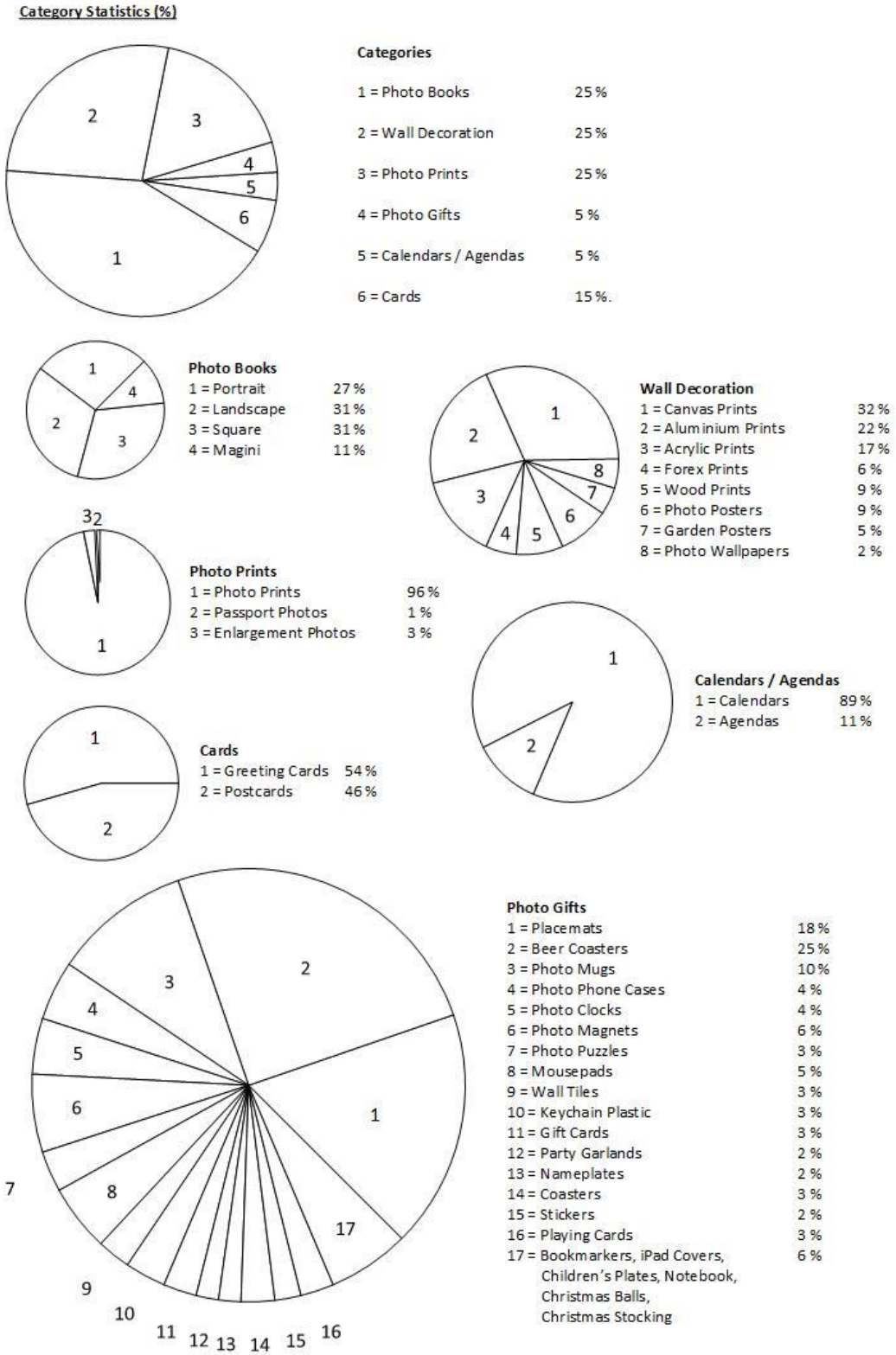


Figure 8: Revenue share per product group (01-01-2015 / 31-12-2015) - *Data retrieved from Google Analytics*

II. Channels

A. Six online channels are tracked by Google Analytics and characterize differently

Multichannel marketing in literature relates to the use of online and offline channels, but within Webprint, and therefore in this context, a different description is used. To be more precise, for the greater part revenues are generated at Webprint by online channels, while offline channels contribute for a small part in revenues. Multichannel marketing in this case, relates to the six online channels originating from Google Analytics. Single channel customers only use one channel at all stages of their decision process, mainly because that channel is the only one available or within reach. But, during past years technologies have changed and developed. The change enables interaction between company and customer and reversely through different channels; multichannel communication. "Multichannel marketing enables firms to build lasting customer relationships by simultaneously offering their customers and prospects information, products, services, and support (or any combination of these) through two or more synchronized channels" (Rangaswamy & Van Bruggen, 2005, p. 6). Customers can find pre-purchase information, like product properties, on the website, compare prices at another website, purchase at an offline store, and finally communicate with the accompanying call center to get post-purchase information by telephone, mail, WhatsApp or another channel (figure 9).

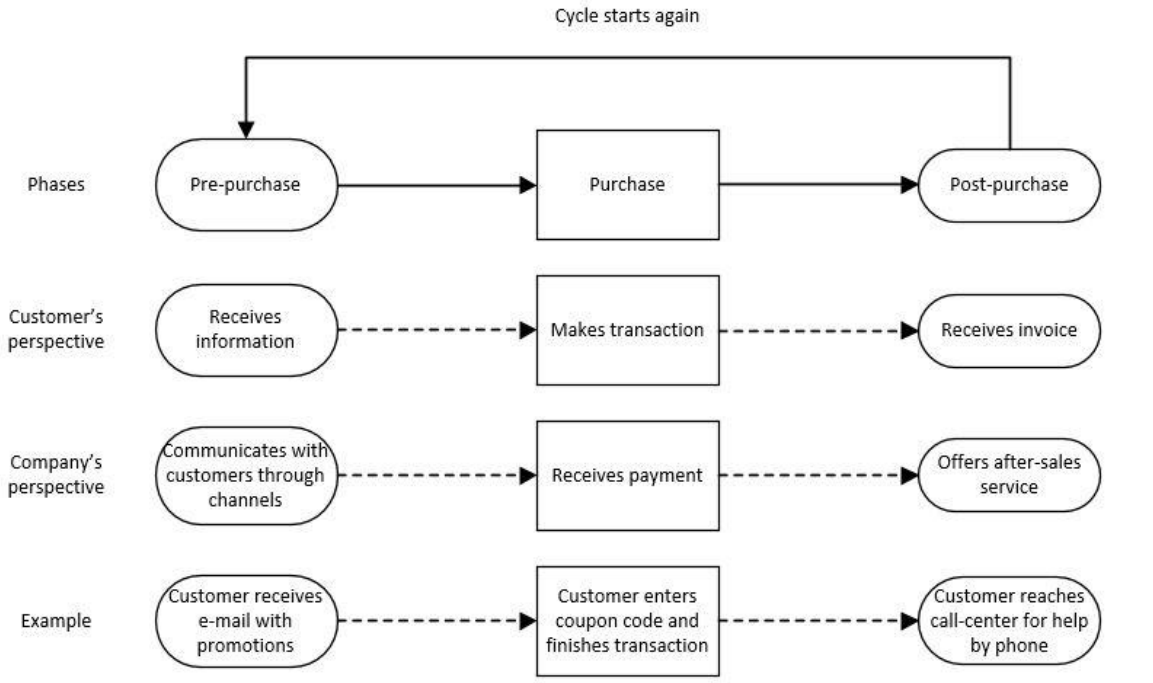


Figure 9: Multichannel process

Multichannel marketing, as the word clearly indicates, refers to the use of different channels communicating with customers, organizations or any enterprise. Complexity can increase using multichannel marketing but that does not have to be the case. Customers can be better reached and tracked, and it depends on the company how they interact with their target population, and in different industries different channels might be more appropriate. Companies can track customer behavior across channels, and improve their understanding of customer's decision making. This can improve relationships with customers and create advantages for customer retention. But, companies do not control each source of information, which has to be taken into consideration. Comparison websites show the pros and cons of products at different suppliers (sometimes) without the permission of the product supplier.

Traditional channel literature mainly identifies three channel categories: offline channels such as a retail store, an electronic or digital channel such as a website, and multiple channels (Kushwaha & Shankar, 2013). But, in reality an online retailer can have more detailed channels instead of just enhancing an electronic channel.

Webprint generates sales partially by offline partnerships and offline deals, however most of the sales are generated by online multichannel marketing. At the photo-related product retailer online sales are sourced by different channels, tracked by e-commerce platform Google Analytics. This platform translates online sales generated through six different channels (also named funnels): 1) Search Engine Optimization, 2) Search Engine Advertising,3) Direct Traffic, 4) Referrals, 5) Affiliates, and 6) E-mail. Prospects reach the website through one of these funnels, and are tracked by Google Analytics. This platform follows prospects through every step on the website, this can be done by UTM-tags or campaigns, even though there are many other technical capabilities. Companies can track customers through Google Analytics based on their traffic on the website. Sessions per page or user, bouncing percentages, new users, transactions, revenues, and conversion percentages are a small selection of the possibilities to track customer behavior by this platform. Different dimensions can be added in order to gain other points of view. Each of the six e-channels or funnels track customer behavior on the website and have different purposes such as attracting new customers, or provide inactive customers with appealing promotions, to mention just a few. Google Analytics also embraces a major disadvantage which is the lack of availability to track customers individually, instead the data is collectively and not dividable per customer. Although, different dimensions can be added to the customer dataset to get a more detailed point of view.

Search engine optimization is the process of getting traffic from the free, organic, editorial or natural search results on search engines. Major search engines such as Google and Yahoo have primary search results, where web pages and other content such as videos or local listings are shown and ranked based on what the search engine considers most relevant to users. Search engines, like Google, show two types of search results: organic results, which are on the left side, and sponsored results, which are on the right side or above the organic results with a yellow background. If an internet user decides to click on a sponsored link, the search engine advertising operation is working. The user can be led to a landing page, which may be especially tailored to the sponsored ad, or to the general homepage of the advertiser. SEA compels different cost measurement methods such as cost-per-mile (CPM), cost-per-click (CPC), cost-per-sale (CPS), and cost-per-lead (CPL). A short and easy definition of direct traffic is often used, which states that direct traffic refers to any visitor that came directly to the website either by typing in the domain name or selecting a bookmark. But, a better definition is any traffic that does not have a referral passed to the website. Referral traffic refers to a recommendation from website to another. This funnel in Google reports visits that came to the website from sources outside of its search engine, for example coming from social media. When someone clicks on a link to a website, the browser sends a request to the current server. The request includes a field with data about the last place the person visited. Google Analytics then captures this data and reports it as a referral domain (Bashara, 2015). Affiliate marketing is basically about promoting other people's products while receiving a certain commission if a sale is generated. This means that affiliates are rewarded for each customer they bring; a performance based marketing tool. Amazon initially introduced the concept in the mid 90's when the company paid commissions to affiliates when their direct links to the Amazon homepage resulted in sales. Affiliate marketers sign up with affiliate networks, that display affiliate programs of merchants. The merchants pay commissions if the targeted traffic (in the form of clicks, buys, or a measurement of choice) is reached. Traffic from e-mail marketing is generated by sending e-mails to prospects, in order to gain sales by offering promotions, news, or information. Customer press the link, or image and can therefore be tracked by Google Analytics. To offer an insight into proportions based on different variables per channel, figure 9 has been set up. Sessions, transactions, and conversion rates are only the proverbial tip of the iceberg of the tracking possibilities Google Analytics allows us to observe. To indicate why observing these statistics is necessary to generate a steady and growing stream of revenue for your website, some remarkable proportions are displayed. Each of the six channels demand a different kind of attention since (causal) relationships between variables are not the same per channel. As depicted, SEA displays the highest number of absolute sessions, the highest number of transactions, but does not peak in conversion rate, since it shows the second lowest conversion rate. On the other hand, the affiliates channel demonstrates the second lowest amount of sessions, the second lowest amount of transactions, but displays as well the highest conversion rate. The variable the company focuses attention on per channel depends on the industry, product group, in more detailed sense; the direction it prefers to go, and the goals it has set and therefore differs per company and even per product or service. To give an example of a strategy companies might refer to, the e-mail channel is used. Companies can choose who to send the e-mail to, what the content of the e-mail is, and what landing page the e-mail links to. Different rational motives determine the content of the mail, the receivers of the mail, and the goal of the mail. This can be to entice customers that have not ordered for at least a year by sending this group an email with large promotions on, in general, popular products. Customers can be tempted to order again, and once

they did, they might get the taste of it and feel connected to the company again. The next time they want to buy a similar product online, hopefully that will happen at the concerning website. However, as the conversion rates in figure 10 already imply, the largest part of the visitors do not make a transaction when visiting websites (in this case Webprint). The figure below only shows data from one year, even though data originally will be analyzed hourly, or daily, in order to get a detailed reflection of reality in the statistics. Google Analytics allows companies to track and report website traffic, and is an web analysis tool. Webprint makes full use of the tool to acquire, retain, and inform (potential) customers. Segmentation can play an important role in communication between companies and customers.

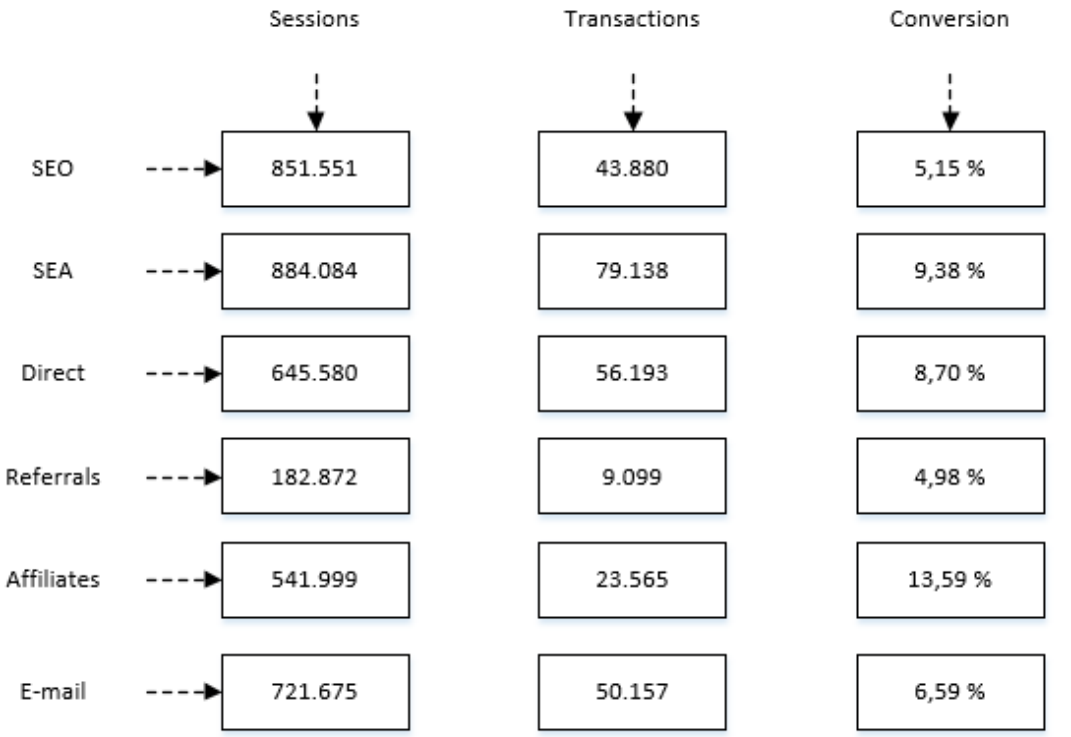


Figure 10: Google Analytics Data Webprint.nl (period 01-01-2015 / 31-12-2015)

The next chapters contain statistical figures including data exported from Google Analytics. Only data from 2015 is displayed since recent data is needed to ensure sharpness and correctness of the numbers. From Google Analytics only channel information is used in the analysis phase of this research. Channel information is not sensitive to the passage of time; unlike revenues and quantity shares. Therefore, channel information can be appropriately used as variable in this research, while revenues or quantity shares would be less reliable. This also means that the data displayed in the figures in the next chapters, is mainly to support and explain processes of Webprint. It enriches in-depth statistical knowledge on product assortment, channels, and revenues. Understanding and insight in the research variables is provided.

B. Comparing statistics between different channels

The diagrams in figure 11 display different variables concerning channel traffic; sessions, new users, revenues, and conversion rates. Time-wise statistics are displayed regarding the whole year 2015, in order to have recent and up-to-date data. These statistics are commonly used as measurement for different purposes inside Google Analytics and therefore are treated in the figure. Overall, six different channels are displayed in the diagrams, SEA, Organic, Email, Direct, Affiliate, and Referral, whose content and definitions have already been discussed. Looking at the sessions concerning channel traffic, it is remarkable that half of the circle is filled with SEA and Organic, which are both search engine tools for advertising or search results. Secondly email generates lots of sessions, a little less than Organic and SEA, which could be due to the amount of people receiving an email is less than the amount of people finding Webprint through search engines. Moreover, not everyone uses search engines, as is shown by the amount of direct traffic regarding sessions, which is a considerable amount. Affiliate and referral traffic score the worst percentages, which is probably because attention on these channels is not equally distributed; they are emphasized less within the company because they demand close, intensive, committed, and time-consuming cooperation with other companies. Screening channel traffic concerning new users it is obvious that Organic brings in the most new users. SEA and direct traffic are respectively second and third, followed by email, referral and affiliate. The amount of email receivers are made up of mostly existing users, which explains the low amount of new users percentage in this channel, while Organic and SEA have another scope. Most of the revenues in these statistics are generated out of search engine advertising (30%). However, we have to keep in to mind that this channel is the most pricy, Organic or direct traffic for example does not enhance any additional costs, besides the costs running the channel, such as wages. Referral and affiliate traffic are, again, the minority channels concerning revenues. Email positions the second place, with 21%, which implies it has a large stake in revenues. This could be explained by the fact that emails are sent very targeted, which means that people receiving an email experience a personalized email which is aligned specifically towards them. Therefore the revenues could be substantially higher than other channels, in this case lower than SEA but higher than Organic. Exploring the conversion rates from the different channels, the affiliate channel is displaying a detectable higher rate than the other channels. Since affiliate marketing is dependable of other companies or organizations for conversion through this channel, the choice of the right collaborations is very important. The high conversion rate can be declared by the fact that companies who are collaborating receive payments based on the amount of customers reaching or conversing (on) the Webprint.nl domain. Therefore partners might be more stimulated to make sure their customers also reach the website of the other partner to enrich themselves. Hence it is remarkable to note that email channel does not have a high conversion rate in comparison. But, on the other hand it is understandable because not everyone opens the email, which means that conversion percentages of emails which have been opened would be a more reliable number. Overall, Organic and SEA seem to deliver the best results based on the statistics in figure 9. But, as self-evidently, the way customers handle depends on the context which cannot always be checked. Google Analytics attempts to transparent human or customer behavior as far as possible, by offering analyses as those in figure 10. On top of that, companies try to control customer behavior by the use of these analyses.

Statistics retrieved from Google Analytics

Channel Traffic - Sessions

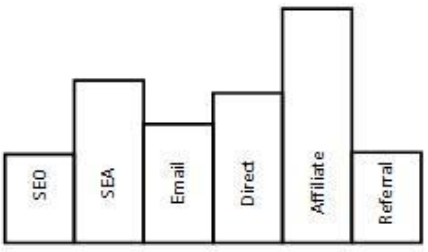
SEO =	25 %
SEA =	25 %
Email =	21 %
Direct =	19 %
Affiliate =	5 %
Referral =	5 %

Channel Traffic – New Users

SEO =	32 %
SEA =	24 %
Email =	11 %
Direct =	23 %
Affiliate =	4 %
Referral =	5 %

Channel Traffic – Revenues

SEO =	18 %
SEA =	30 %
Email =	21 %
Direct =	19 %
Affiliate =	9 %
Referral =	3 %



Channel Traffic – Conversion Rates

SEO =	5,15 %
SEA =	9,46 %
Email =	6,91 %
Direct =	8,72 %
Affiliate =	13,63 %
Referral =	5,27 %

Figure 11: Channel Traffic - (01-01-2015 / 31-12-2015) - *Data retrieved from Google Analytics*

C. Overlap between channels

There is another characteristic of channels which has to be taken into account. Channels often overlap, which means that customers can have different routes preliminary to a purchase. A customer can be informed by email on a promotion but might decide not to order immediately. Days later the same customer receives an advert supported by the SEA channel on which the person clicks. It decides to give it a shot and orders the product, even though the customer was already informed by the promotion by email. Nevertheless, the purchase is assigned to the SEA channel, since that channel was the link the customer reached the website by. This means that the last channel preceding the purchase, is the channel the purchase is assigned to. That does not alter the fact that, the customer is influenced by the email channel. Obviously, assigning channels to purchases in this manner enlarges proportional differences between channel revenues. Ideally, the purchase should be assigned to both channels. To illustrate this phenomenon, figure 12 is set up.

Channel	% of total conversions
☑ Direct	64.29%
☑ Paid Search	37.69%
☑ Email	27.20%
☑ Organic Search	23.09%
☐ Display	16.71%
☐ Other Advertising	9.35%
☐ Referral	2.64%
☐ Social Network	0.14%

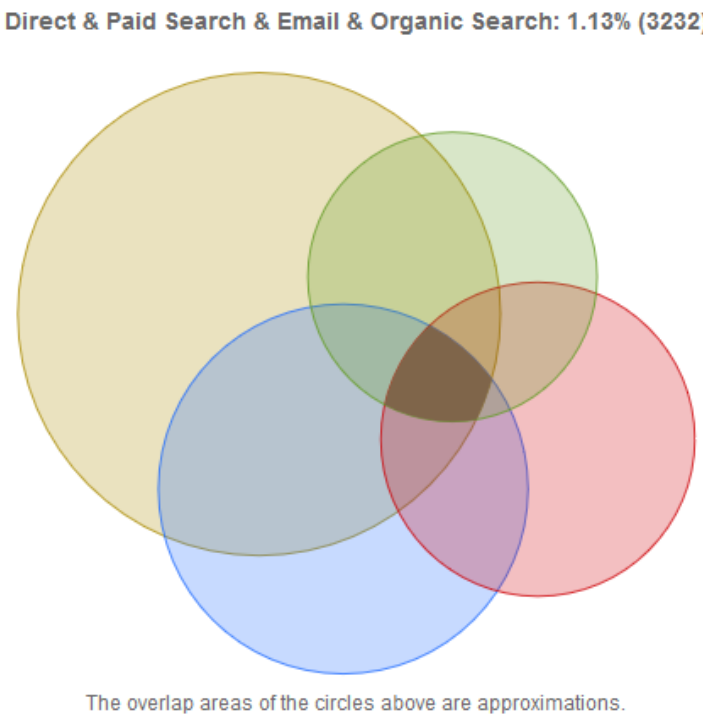


Figure 12: Overlap of conversion shares (01-01-2015 / 31-12-2015) - *Data retrieved from Google Analytics*

III. Calculations on purchase one and purchase two

A. Customers who buy the same product in purchase one and purchase two

Row percentages		**Purchase 2**							
	Product category	Business	Calendars and agendas	Greeting cards	Photo books	Photo gifts	Photos	Wall decoration	**Total**
Purchase 1	Business	126 (50,4%)	3 (1,2%)	21 (8,4%)	25 (10,0%)	9 (3,6%)	29 (11,6%)	37 (14,8%)	250 (100,0%)
	Calendars and agendas	3 (0,3%)	390 (44,2%)	66 (7,5%)	126 (14,3%)	32 (3,6%)	157 (17,8%)	109 (12,3%)	883 (100,0%)
	Greeting cards	14 (0,5%)	101 (3,3%)	1699 (56%)	375 (12,3%)	86 (2,8%)	467 (15,3%)	303 (10,0%)	3045 (100,0%)
	Photo books	15 (0,2%)	178 (2,1%)	517 (6,2%)	5214 (62,1%)	172 (2,1%)	1453 (17,3%)	841 (10,0%)	8390 (100,0%)
	Photo gifts	5 (0,4%)	42 (3,0%)	102 (7,2%)	230 (16,3%)	584 (41,4%)	237 (16,8%)	209 (14,8%)	1409 (100,0%)
	Photos	41 (0,2%)	242 (1,4%)	736 (4,3%)	1914 (11,2%)	295 (1,7%)	12312 (72,3%)	1492 (8,8%)	17032 (100,0%)
	Wall decoration	23 (0,3%)	164 (2,5%)	437 (6,6%)	959 (14,4%)	236 (3,5%)	1263 (19,0%)	3577 (53,7%)	6659 (100,0%)
	Total	227 (0,6%)	1120 (3,0%)	3578 (9,5%)	8843 (23,5%)	1414 (3,8%)	15918 (42,3%)	6568 (17,4%)	37668 (100%)

Table 13: Product category purchase one versus product category purchase two (row percentages)

Column percentages		**Purchase 2**							
	Product category	Business	Calendars and agendas	Greeting cards	Photo books	Photo gifts	Photos	Wall decoration	**Total**
Purchase 1	Business	126 (55,5%)	3 (0,3%)	21 (0,6%)	25 (0,3%)	9 (0,6%)	29 (0,2%)	37 (0,6%)	250 (0,7%)
	Calendars and agendas	3 (1,3%)	390 (34,8%)	66 (1,8%)	126 (1,4%)	32 (2,3%)	157 (1,0%)	109 (1,7%)	883 (2,3%)
	Greeting cards	14 (6,2%)	101 (9,0%)	1699 (47,5%)	375 (4,2%)	86 (6,1%)	467 (2,9%)	303 (4,6%)	3045 (8,1%)
	Photo books	15 (6,6%)	178 (15,9%)	517 (14,4%)	5214 (59,0%)	172 (12,2%)	1453 (9,1%)	841 (12,8%)	8390 (22,3%)
	Photo gifts	5 (2,2%)	42 (3,8%)	102 (2,9%)	230 (2,6%)	584 (41,3%)	237 (1,5%)	209 (3,2%)	1409 (3,7%)
	Photos	41 (18,1%)	242 (21,6%)	736 (20,6%)	1914 (21,6%)	295 (20,9%)	12312 (77,3%)	1492 (22,7%)	17032 (45,2%)
	Wall decoration	23 (10,1%)	164 (14,6%)	437 (12,2%)	959 (10,8%)	236 (16,7%)	1263 (7,9%)	3577 (54,5%)	6659 (17,7%)
	Total	227 (100,0%)	1120 (100,0%)	3578 (100,0%)	8843 (100,0%)	1414 (100,0%)	15918 (100,0%)	6568 (100,0%)	37668 (100%)

Table 14: Product category purchase one versus product category purchase two (column percentages)

B. Customers who do not buy the same product in purchase one and purchase two

If purchase one exists of products from the business category and the second purchase does not exist of products from the business category, then purchase two consists of:

Calendars and agendas	2.4 %
Cards	16.9 %
Photo Books	20.2 %
Photo Gifts	7.3 %
Photos	23.4 %
Wall Decoration	29.8 %
Total	**100 %**

If purchase one exists of products from the calendars and agendas category and the second purchase does not exist of products from the calendars and agendas category, then purchase two consists of:

Business	0.6 %
Cards	13.4 %
Photo Books	25.6 %
Photo Gifts	6.5 %
Photos	31.8 %
Wall Decoration	22.1 %
Total	**100 %**

If purchase one exists of products from the cards category and the second purchase does not exist of products from the cards category, then purchase two consists of:

Business	1 %
Calendars and agendas	7.5 %
Photo Books	27.9 %
Photo Gifts	6.4%
Photos	34.7 %
Wall Decoration	22.5 %
Total	**100 %**

If purchase one exists of products from the photo books category and the second purchase does not exist of products from the photo books category, then purchase two consists of:

Business	0.5 %
Calendars and agendas	5.6 %
Cards	16.3 %
Photo Gifts	5.4 %
Photos	45.7 %
Wall Decoration	26.5 %
Total	**100 %**

If purchase one exists of products from the photo gifts category and the second purchase does not exist of products from the photo gifts category, then purchase two consists of:

Business	0.6 %
Calendars and agendas	5.1 %
Cards	12.4 %
Photo Books	27.9 %
Photos	28.7 %
Wall Decoration	25.3 %
Total	**100 %**

If purchase one exists of products from the photos category and the second purchase does not exist of products from the photos category, then purchase two consists of:

Business	0.9 %
Calendars and agendas	5.1 %
Cards	15.6 %
Photo Books	40.6 %
Photo Gifts	6.3 %
Wall Decoration	31.6 %
Total	**100 %**

If purchase one exists of products from the wall decoration category and the second purchase does not exist of products from the wall decoration category, then purchase two consists of:

Business	0.7 %
Calendars and agendas	5.3 %
Cards	14.2 %
Photo Books	31.1 %
Photo Gifts	7.7 %
Photos	41 %
Total	**100 %**

IV. Calculations on purchase two and purchase three

A. Customers who buy the same product in purchase two and purchase three

Row percentages		Purchase 3							
	Product category	Business	Calendars and agendas	Greeting cards	Photo books	Photo gifts	Photos	Wall decoration	**Total**
Purchase 2	Business	118 (52,0%)	3 (1,3%)	22 (9,7%)	15 (6,6%)	6 (2,6%)	32 (14,1%)	31 (13,7%)	227 (100,0%)
	Calendars and agendas	2 (0,2%)	487 (43,5%)	98 (8,8%)	185 (16,5%)	43 (3,8%)	195 (17,4%)	110 (9,8%)	1120 (100,0%)
	Greeting cards	16 (0,4%)	132 (3,7%)	1951 (54,5%)	456 (12,7%)	94 (2,6%)	592 (16,5%)	337 (9,4%)	3578 (100,0%)
	Photo books	23 (0,3%)	216 (2,4%)	547 (6,2%)	5441 (61,5%)	178 (2,0%)	1551 (17,5%)	887 (10,0%)	8843 (100,0%)
	Photo gifts	8 (0,6%)	56 (4,0%)	106 (7,5%)	217 (15,3%)	564 (39,9%)	237 (16,8%)	226 (16,0%)	1414 (100,0%)
	Photos	26 (0,2%)	255 (1,6%)	729 (4,6%)	1640 (10,3%)	271 (1,7%)	11649 (73,2%)	1348 (8,5%)	15918 (100,0%)
	Wall decoration	17 (0,3%)	161 (2,5%)	463 (7,0%)	914 (13,9%)	181 (2,8%)	1315 (20,0%)	3517 (53,5%)	6568 (100,0%)
	Total	210 (0,6%)	1310 (3,5%)	3916 (10,4%)	8868 (23,5%)	1337 (3,5%)	15571 (41,3%)	6456 (17,1%)	37668 (100%)

Table 15: Product category purchase two versus product category purchase three (row percentages)

Column percentages		Purchase 3							
	Product category	Business	Calendars and agendas	Greeting cards	Photo books	Photo gifts	Photos	Wall decoration	**Total**
Purchase 2	Business	118 (56,2%)	3 (0,2%)	22 (0,6%)	15 (0,2%)	6 (0,4%)	32 (0,2%)	31 (0,5%)	227 (0,6%)
	Calendars and agendas	2 (1,0%)	487 (37,2%)	98 (2,5%)	185 (2,1%)	43 (3,2%)	195 (1,3%)	110 (1,7%)	1120 (3,0%)
	Greeting cards	16 (7,6%)	132 (10,1%)	1951 (49,8%)	456 (5,1%)	94 (7,0%)	592 (3,8%)	337 (5,2%)	3578 (9,5%)
	Photo books	23 (11,0%)	216 (16,5%)	547 (14,0%)	5441 (61,4%)	178 (13,3%)	1551 (10,0%)	887 (13,7%)	8843 (23,5%)
	Photo gifts	8 (3,8%)	56 (4,3%)	106 (2,7%)	217 (2,4%)	564 (42,2%)	237 (1,5%)	226 (3,5%)	1414 (3,8%)
	Photos	26 (12,4%)	255 (19,5%)	729 (18,6%)	1640 (18,5%)	271 (20,3%)	11649 (74,8%)	1348 (20,9%)	15918 (42,3%)
	Wall decoration	17 (8,1%)	161 (12,3%)	463 (11,8%)	914 (10,3%)	181 (13,5%)	1315 (8,4%)	3517 (54,5%)	6568 (17,4%)
	Total	210 (100,0%)	1310 (100,0%)	3916 (100,0%)	8868 (100,0%)	1337 (100,0%)	15571 (100,0%)	6456 (100,0%)	37668 (100%)

Table 16: Product category purchase two versus product category purchase three (column percentages)

B. Customers who do not buy the same product in purchase two and purchase three

If purchase two exists of products from the business category and the third purchase does not exist of products from the business category, then purchase two consists of:

Calendars and agendas	2.8 %
Cards	20.2 %
Photo Books	13.8 %
Photo Gifts	5.5 %
Photos	29.4 %
Wall Decoration	28.4 %
Total	**100 %**

If purchase two exists of products from the calendars and agendas category and the third purchase does not exist of products from the calendars and agendas category, then purchase two consists of:

Business	0.3 %
Cards	15.5 %
Photo Books	29.2 %
Photo Gifts	6.8 %
Photos	30.8 %
Wall Decoration	17.4 %
Total	**100 %**

If purchase two exists of products from the cards category and the third purchase does not exist of products from the cards category, then purchase two consists of:

Business	1 %
Calendars and agendas	8.1 %
Photo Books	28 %
Photo Gifts	5.8%
Photos	36.4 %
Wall Decoration	20.7 %
Total	**100 %**

If purchase two exists of products from the photo books category and the third purchase does not exist of products from the photo books category, then purchase two consists of:

Business	0.7 %
Calendars and agendas	6.3 %
Cards	16.1 %
Photo Gifts	5.2 %
Photos	45.6 %
Wall Decoration	26.1 %
Total	**100 %**

If purchase two exists of products from the photo gifts category and the third purchase does not exist of products from the photo gifts category, then purchase two consists of:

Business	0.9 %
Calendars and agendas	6.6 %
Cards	12.5 %
Photo Books	25.5 %
Photos	27.9 %
Wall Decoration	26.6 %
Total	**100 %**

If purchase two exists of products from the photos category and the third purchase does not exist of products from the photos category, then purchase two consists of:

Business	0.6 %
Calendars and agendas	6 %
Cards	17.1 %
Photo Books	38.4 %
Photo Gifts	6.3 %
Wall Decoration	31.6 %
Total	**100 %**

If purchase two exists of products from the wall decoration category and the third purchase does not exist of products from the wall decoration category, then purchase two consists of:

Business	0.6 %
Calendars and agendas	5.3 %
Cards	15.2 %
Photo Books	30 %
Photo Gifts	5.9 %
Photos	43.1 %
Total	**100 %**

V. Calculations in the photo book category

A. The purchase preliminary to the purchase of a photo book

If purchase two exists of products from the photo books category then the proportion of products from the business category in purchase one is 0.3%

If purchase two exists of products from the photo books category then the proportion of products from the calendars and agendas category in purchase one is 1.4%

If purchase two exists of products from the photo books category then the proportion of products from the cards category in purchase one is 4.2%

If purchase two exists of products from the photo books category then the proportion of products from the photo books category in purchase one is 59%

If purchase two exists of products from the photo books category then the proportion of products from the photo gifts category in purchase one is 2.6%

If purchase two exists of products from the photo books category then the proportion of products from the photos category in purchase one is 21.6%

If purchase two exists of products from the photo books category then the proportion of products from the wall decoration category in purchase one is 10.8%

Row percentages		**Purchase 1**							
	Product category	Business	Calendars and agendas	Greeting cards	Photo books	Photo gifts	Photos	Wall decoration	**Total**
Purchase 2	Photo books	25 (0,3%)	126 (1,4%)	375 (4,2%)	5214 (59,0%)	230 (2,6%)	1914 (21,6%)	959 (10,8%)	8843 (100,0%)
	Total	250	883	3045	8390	1409	17032	6659	37668 (100%)

Table 17: Photo books purchase two shares

B. The second purchase preliminary to the third purchase which includes a photo book

If purchase two exists of products from the photo books category then the proportion of products from the business category in purchase one is 0.2%

If purchase two exists of products from the photo books category then the proportion of products from the calendars and agendas category in purchase one is 2.1%

If purchase two exists of products from the photo books category then the proportion of products from the cards category in purchase one is 5.1%

If purchase two exists of products from the photo books category then the proportion of products from the photo books category in purchase one is 61.4%

If purchase two exists of products from the photo books category then the proportion of products from the photo gifts category in purchase one is 2.4%

If purchase two exists of products from the photo books category then the proportion of products from the photos category in purchase one is 18.5%

If purchase two exists of products from the photo books category then the proportion of products from the wall decoration category in purchase one is 10.3%

Row percentages		**Purchase 2**							
	Product category	Business	Calendars and agendas	Greeting cards	Photo books	Photo gifts	Photos	Wall decoration	**Total**
Purchase 3	Photo books	15 (0,2%)	185 (2,1%)	456 (5,1%)	5441 (61,4%)	217 (2,4%)	1640 (18,5%)	914 (10,3%)	8868 (100,0%)
	Total	227	1120	3578	8843	1414	15918	6568	37668 (100%)

Table18: Photo books purchase three shares

VI. Mapping the customer road

A. Customer roads with the highest frequency

	N_BREAK		
Frequency	*Purchase one*	*Purchase two*	*Purchase three*
672	Photo books	Photos	Photos
688	Photos	Photo books	Photos
690	Wall decoration	Photos	Photos
807	Photos	Photos	Wall decoration
830	Photos	Photo books	Photo books
899	Photos	Photos	Photo books
1175	Cards	Cards	Cards
2424	Wall decoration	Wall decoration	Wall decoration
3706	Photo books	Photo books	Photo books
9870	Photos	Photos	Photos

Table19: Most frequent customer roads

B. Customer roads with the lowest frequency

	N_BREAK		
Frequency	*Purchase one*	*Purchase two*	*Purchase three*
18	Calendars and agendas	Photos	Wall decoration
25	Calendars and agendas	Photos	Photo books
34	Photo gifts	Wall decoration	Photos
40	Photo gifts	Photo gifts	Photos
42	Photos	Photo gifts	Wall decoration
46	Photos	Photo gifts	Photo books
47	Cards	Cards	Calendars and agendas
48	Photos	Photo gifts	Photo gifts
49	Photo gifts	Photo gifts	Photo books
50	Wall decoration	Photo books	Cards

Table 20:Least frequent customer roads

VII. Customer profiling

Legend:

ST1	Subtotal Purchase 1
CC1	Coupon Code Purchase 1
AP1	Amount payed Purchase 1
DT1	Order Date Purchase 1
ST2	Subtotal Purchase 2
CC2	Coupon Code Purchase 2
AP2	Amount Payed Purchase 2
DT2	Order Date Purchase 2
ST3	Subtotal Purchase 3
CC3	Coupon Code Purchase 3
AP3	Amount Payed Purchase 3
DT3	Order Date Purchase 3

Table 21: Legend customer profiling

Customer road: Photos – Photos – Photo books

	Postal code	City	ST1	CC1	AP1	DT1	ST2	CC2	AP2	DT2	ST3	CC3	AP3	DT3
1	9501 PE	Stadskanaal	€ 67,52	€ 20,26	€ 55,20	31-1-2014	€ 3,00	x	€ 5,75	19-2-2014	€ 14,85	€ 11,85	€ 7,14	22-10-2014
2	6579 AP	Kekerdom	€ 2,35	x	€ 4,60	8-4-2013	€ 2,70	x	€ 5,45	4-9-2013	€ 53,90	€ 11,86	€ 47,98	12-12-2013
3	6591 ZE	Gennep	€ 5,88	x	€ 8,83	7-3-2014	€ 5,46	€ 1,37	€ 7,05	13-5-2014	€ 12,95	x	€ 15,89	20-5-2014
4	2291 RJ	Wateringen	€ 12,21	x	€ 15,84	4-4-2013	€ 1,05	x	€ 3,80	8-5-2013	€ 5,95	€ 1,79	€ 7,11	27-2-2014
5	1702 PD	Heerhugowaard	€ 14,28	x	€ 18,22	4-7-2014	€ 1,47	x	€ 4,41	1-10-2014	€ 16,61	€ 3,14	€ 17,91	17-10-2014
6	3437 NP	Nieuwegein	€ 14,56	€ 3,50	€ 14,69	23-8-2013	€ 19,25	€ 5,78	€ 17,42	14-7-2014	€ 37,90	€ 9,48	€ 34,37	21-1-2015
7	8457 EK	Gersloot	€ 8,75	x	€ 12,38	9-9-2013	€ 15,23	x	€ 18,67	14-2-2014	€ 70,53	x	€ 76,47	5-2-2015
8	2182 TG	Hillegom	€ 1,54	x	€ 4,29	27-8-2013	€ 9,21	€ 2,49	€ 10,35	6-1-2014	€ 20,90	x	€ 25,84	15-4-2014
9	8043 EP	Zwolle	€ 6,25	x	€ 9,88	26-7-2013	€ 4,25	x	€ 7,88	4-9-2013	€ 14,99	€ 2,23	€ 17,40	7-5-2014
10	6294 AW	Vijlen	€ 4,20	x	€ 7,15	11-12-2013	€ 1,52	x	€ 4,27	5-5-2014	€ 16,90	x	€ 20,44	30-10-2014
11	4255 VV	Nieuwendijk	€ 0,70	€ 2,02	€ 0,00	3-4-2013	€ 23,81	€ 5,24	€ 22,51	17-11-2014	€ 21,95	€ 21,95	€ 4,94	4-12-2014
12	1817 HW	Alkmaar	€ 11,30	x	€ 14,24	29-1-2015	€ 9,80	x	€ 12,74	30-1-2015	€ 11,45	€ 8,00	€ 7,39	17-2-2015

13	1551 SH	Westzaan	€ 17,29	x	€ 20,92	2-7-2013	€ 8,30	€ 1,25	€ 10,69	5-7-2013	€ 20,74	€ 0,45	€ 24,73	12-6-2014
14	1625 RL	Hoorn	€ 0,42	x	€ 2,73	7-11-2013	€ 0,91	€ 0,91	€ 2,31	9-1-2014	€ 69,45	€ 20,84	€ 52,55	2-1-2015
15	6923 AG	Groessen	€ 16,52	€ 3,50	€ 16,96	10-7-2014	€ 5,25	€ 1,31	€ 6,88	14-7-2014	€ 10,55	€ 2,64	€ 11,85	31-10-2014
16	2543 SR	Den Haag	€ 1,60	x	€ 3,91	9-10-2013	€ 1,70	x	€ 4,01	12-11-2013	€ 4,95	€ 4,95	€ 2,94	5-2-2014
17	5735 AM	Aarle-Rixtel	€ 16,17	x	€ 19,80	14-6-2013	€ 6,86	x	€ 9,81	24-12-2013	€ 22,75	€ 21,95	€ 4,74	3-3-2014
18	2024 TG	Haarlem	€ 3,43	€ 3,43	€ 2,75	13-5-2013	€ 2,57	x	€ 5,51	14-7-2014	€ 17,85	€ 4,46	€ 19,33	25-9-2014
19	3861 DD	Nijkerk	€ 5,88	x	€ 9,07	10-5-2013	€ 5,81	x	€ 8,76	9-9-2013	€ 29,35	€ 7,34	€ 26,95	18-10-2014
20	1688 CD	Nibbixwoud	€ 15,47	€ 2,32	€ 17,09	7-11-2014	€ 7,42	€ 1,86	€ 8,50	21-12-2014	€ 43,12	€ 10,78	€ 38,78	29-1-2015
21	3446 AC	Woerden	€ 14,61	€ 3,65	€ 17,40	29-3-2013	€ 54,39	€ 15,05	€ 47,28	21-8-2013	€ 72,45	€ 48,95	€ 28,44	15-12-2014
22	1216 KE	Hilversum	€ 4,12	€ 3,50	€ 3,37	17-7-2013	€ 10,72	€ 0,19	€ 13,97	6-6-2014	€ 6,95	€ 3,95	€ 5,94	16-12-2014
23	2712 BA	Zoetermeer	€ 0,56	x	€ 2,87	17-6-2013	€ 14,77	x	€ 18,40	27-1-2014	€ 30,75	x	€ 33,67	26-2-2015
24	3825 WK	Amersfoort	€ 7,83	x	€ 10,78	26-11-2013	€ 12,80	x	€ 15,75	7-4-2014	€ 8,35	€ 1,67	€ 10,62	27-11-2014
25	9746 CA	Groningen	€ 27,72	x	€ 35,66	15-11-2013	€ 19,92	x	€ 23,55	14-1-2014	€ 22,85	€ 13,90	€ 16,09	11-6-2014
26	2461 VJ	Ter Aar	€ 31,91	x	€ 39,65	8-4-2013	€ 0,91	x	€ 3,22	6-12-2013	€ 21,35	€ 21,36	€ 3,94	24-3-2014
27	5491 AB	Sint-Oedenrode	€ 5,23	x	€ 8,42	28-5-2013	€ 0,47	€ 0,07	€ 3,34	28-11-2014	€ 13,95	€ 3,49	€ 14,40	17-12-2014
28	2802 AP	Gouda	€ 26,29	x	€ 34,23	16-8-2013	€ 13,59	x	€ 17,22	24-12-2013	€ 76,50	€ 15,30	€ 69,54	26-2-2014
29	5492 PJ	Sint-Oedenrode	€ 0,49	x	€ 2,80	12-4-2013	€ 3,43	x	€ 6,18	29-5-2013	€ 41,20	€ 12,36	€ 33,58	5-9-2013
30	5541 DR	Reusel	€ 8,68	€ 1,91	€ 9,52	5-11-2013	€ 0,84	x	€ 3,78	20-11-2014	€ 14,85	€ 11,85	€ 7,14	16-12-2014
31	4885 AL	Achtmaal	€ 22,40	€ 0,00	€ 26,03	18-4-2013	€ 14,40	x	€ 18,03	14-4-2014	€ 9,90	€ 1,98	€ 11,46	25-6-2014
32	4708 HA	Roosendaal	€ 5,59	€ 3,50	€ 5,04	21-8-2013	€ 1,68	x	€ 4,43	28-8-2013	€ 22,85	€ 5,71	€ 21,08	14-7-2014
33	5472 LB	Loosbroek	€ 4,04	€ 1,05	€ 5,74	13-8-2013	€ 3,08	x	€ 6,02	18-9-2014	€ 31,55	€ 27,95	€ 8,54	2-3-2015
34	5863 AX	Blitterswijck	€ 0,98	x	€ 3,29	5-12-2013	€ 1,64	x	€ 4,39	24-4-2014	€ 7,95	€ 3,95	€ 7,94	5-6-2014
35	5502 JS	Veldhoven	€ 1,67	€ 1,67	€ 2,94	26-11-2014	€ 0,20	x	€ 3,14	3-12-2014	€ 64,05	x	€ 69,99	15-12-2014
36	8253 ZA	Dronten	€ 3,36	€ 0,50	€ 5,80	7-11-2014	€ 0,84	x	€ 3,78	21-1-2015	€ 22,75	€ 9,10	€ 18,59	11-2-2015
37	4791 KK	Klundert	€ 23,10	x	€ 31,04	12-8-2013	€ 13,93	€ 3,48	€ 14,39	2-7-2014	€ 4,95	€ 1,24	€ 6,65	31-10-2014
38	6903 RG	Zevenaar	€ 0,20	x	€ 2,51	7-6-2013	€ 1,82	x	€ 4,57	24-10-2013	€ 10,72	x	€ 15,16	15-6-2014
39	2728 NW	Zoetermeer	€ 14,14	€ 3,50	€ 14,27	30-4-2013	€ 1,33	x	€ 4,08	2-7-2014	€ 6,95	x	€ 9,89	4-7-2014
40	1504 HA	Zaandam	€ 3,57	x	€ 6,32	13-12-2013	€ 5,60	€ 1,68	€ 6,67	31-1-2014	€ 7,95	€ 3,95	€ 7,94	22-10-2014
41	2245 XB	Wassenaar	€ 2,94	x	€ 11,38	13-5-2013	€ 12,67	x	€ 16,30	6-6-2013	€ 26,85	x	€ 32,79	2-2-2015
42	4241 AP	Arkel	€ 3,50	€ 3,50	€ 2,75	12-2-2014	€ 5,35	x	€ 8,29	17-9-2014	€ 12,75	€ 3,95	€ 11,74	2-12-2014
43	9711 EE	Groningen	€ 10,63	€ 3,50	€ 11,07	3-12-2014	€ 1,88	x	€ 4,82	21-12-2014	€ 26,85	€ 13,43	€ 17,57	23-12-2014
44	2514 GG	s-Gravenhage	€ 12,34	€ 3,09	€ 12,20	21-5-2014	€ 0,70	€ 3,64	€ 0,00	6-6-2014	€ 33,35	€ 8,34	€ 29,95	4-2-2015
45	2983 BL	Ridderkerk	€ 11,18	€ 3,50	€ 11,31	6-4-2013	€ 2,38	x	€ 5,13	15-4-2013	€ 75,90	€ 33,95	€ 47,89	14-1-2015

46	8253 JD	Dronten	€ 24,01	€ 3,50	€ 27,45	22-1-2015	€ 14,21	€ 3,55	€ 14,60	2-2-2015	€ 35,15	€ 14,06	€ 26,03	18-2-2015
47	4132 HS	Vianen	€ 44,73	x	€ 52,67	5-8-2013	€ 60,41	x	€ 68,35	24-4-2014	€ 55,95	€ 50,00	€ 10,89	25-3-2015
48	1902 AH	Castricum	€ 3,51	x	€ 6,26	17-3-2014	€ 0,81	€ 0,18	€ 2,94	15-4-2014	€ 8,26	€ 3,95	€ 7,85	12-2-2015
49	15202 AE	Dongen	€ 7,25	x	€ 10,20	19-6-2013	€ 10,76	x	€ 14,39	16-12-2013	€ 13,08	€ 2,66	€ 13,96	17-10-2014
50	8084 GW	T Harde	€ 3,50	€ 3,50	€ 2,75	12-3-2014	€ 2,24	x	€ 4,99	18-3-2014	€ 5,75	€ 3,95	€ 4,74	3-12-2014

Table 22: Sample of customer road Photos – Photos – Photo books

Customer road: Photos – Photo books – Photo books

	Postal code	**City**	**ST1**	**CC1**	**AP1**	**DT1**	**ST2**	**CC2**	**AP2**	**DT2**	**ST3**	**CC3**	**AP3**	**DT3**
1	6851 VK	Huissen	€ 2,80	x	€ 5,55	28-3-2013	€ 33,55	x	€ 38,49	2-6-2014	€ 13,95	x	€ 17,89	15-10-2014
2	3256 AH	Achthuizen	€ 4,00	x	€ 6,75	31-12-2013	€ 21,35	x	€ 25,29	22-6-2014	€ 19,95	€ 15,95	€ 8,94	26-6-2014
3	4835 CA	Breda	€ 19,95	€ 15,95	€ 13,00	3-4-2013	€ 54,45	x	€ 58,33	7-6-2013	€ 108,90	€ 54,45	€ 59,39	30-6-2013
4	3264 XW	Nieuw-Beijerland	€ 5,75	x	€ 8,70	2-8-2013	€ 13,95	€ 2,79	€ 15,10	21-11-2013	€ 13,95	€ 2,09	€ 15,80	14-5-2014
5	2396 EL	Koudekerk aan den Rijn	€ 8,75	€ 3,50	€ 9,19	21-10-2014	€ 17,06	€ 2,56	€ 18,94	4-11-2014	€ 43,90	€ 13,17	€ 35,67	10-12-2014
6	5102 BG	Dongen	€ 11,13	x	€ 14,76	7-8-2013	€ 81,90	€ 40,95	€ 46,89	12-8-2013	€ 59,90	€ 17,97	€ 47,87	4-12-2014
7	3212 BE	Pellenberg	€ 3,50	€ 3,50	€ 2,94	8-7-2014	€ 51,80	€ 12,95	€ 43,59	9-7-2014	€ 25,50	€ 12,55	€ 16,49	18-10-2014
8	9075 NH	Westhoek	€ 8,40	x	€ 11,15	29-11-2013	€ 22,95	€ 6,89	€ 21,01	24-1-2014	€ 22,95	€ 5,05	€ 22,84	17-4-2014
9	4043 LW	Opheusden	€ 1,18	x	€ 3,93	10-3-2014	€ 34,65	€ 10,00	€ 29,59	23-1-2015	€ 31,85	€ 7,96	€ 28,83	4-2-2015
10	4283 HB	Giessen	€ 3,50	x	€ 6,25	11-11-2013	€ 5,95	€ 5,29	€ 4,61	7-2-2014	€ 29,75	€ 8,93	€ 28,77	27-2-2014
11	9404 EK	Assen	€ 0,77	€ 2,09	€ 0,00	3-4-2013	€ 8,15	x	€ 11,09	21-5-2013	€ 10,70	€ 4,90	€ 9,24	17-6-2013
12	4724 HC	Wouw	€ 10,50	x	€ 14,44	2-7-2014	€ 19,95	x	€ 24,89	25-9-2014	€ 19,95	x	€ 24,89	24-10-2014

13	2555 PN	s-Gravenhage	€ 2,40	€ 0,60	€ 4,55	26-5-2014	€ 7,95	€ 1,59	€ 10,30	24-6-2014	€ 35,95	x	€ 41,84	4-2-2015
14	2223XA	Katwijk	€ 46,90	€ 11,73	€ 38,81	28-5-2014	€ 27,95	€ 6,99	€ 25,90	6-6-2014	€ 43,55	€ 8,71	€ 39,78	29-6-2014
15	3068 RT	Rotterdam	€ 3,50	€ 3,50	€ 2,76	7-10-2013	€ 4,95	€ 3,95	€ 3,94	2-1-2014	€ 34,65	€ 10,40	€ 30,80	17-1-2014
16	3195 GV	Pernis Rotterdam	€ 9,08	€ 1,36	€ 11,66	27-10-2014	€ 4,95	€ 3,95	€ 3,94	28-10-2014	€ 10,70	€ 3,21	€ 11,03	17-12-2014
17	5491 XA	Sint-Oedenrode	€ 2,45	x	€ 10,40	6-5-2013	€ 6,95	€ 1,04	€ 8,85	15-7-2013	€ 6,95	€ 2,78	€ 7,11	1-8-2013
18	8103 AW	Raalte	€ 0,70	€ 2,02	€ 0,00	3-4-2013	€ 13,95	€ 13,95	€ 2,94	11-4-2013	€ 4,95	€ 4,95	€ 2,94	15-5-2013
19	7471 HK	Goor	€ 11,55	x	€ 15,18	28-5-2013	€ 19,95	€ 10,00	€ 14,89	6-1-2014	€ 29,95	€ 21,00	€ 14,79	28-3-2014
20	7531 ZG	Enschede	€ 43,15	€ 10,79	€ 40,80	21-5-2014	€ 23,95	€ 3,59	€ 25,30	26-5-2014	€ 71,85	€ 14,37	€ 64,42	23-6-2014
21	7771 AC	Hardenberg	€ 0,60	x	€ 2,91	8-7-2013	€ 14,95	€ 3,95	€ 14,94	15-6-2014	€ 11,95	€ 8,00	€ 7,89	17-2-2015
22	3771 HH	Barneveld	€ 0,54	x	€ 2,85	28-11-2013	€ 63,10	€ 18,93	€ 50,11	13-1-2014	€ 63,10	€ 31,55	€ 37,49	11-3-2014
23	8243 WN	Lelystad	€ 1,47	x	€ 4,28	29-4-2013	€ 14,85	€ 9,00	€ 9,79	10-6-2013	€ 85,85	€ 39,95	€ 52,84	1-10-2013
24	6035 BS	Ospel	€ 3,00	€ 0,99	€ 9,52	2-5-2013	€ 23,95	€ 11,98	€ 15,92	5-7-2013	€ 38,45	€ 19,95	€ 23,44	21-8-2013
25	3818 DB	Amersfoort	€ 26,77	€ 8,03	€ 27,18	6-5-2013	€ 7,15	€ 2,45	€ 7,64	10-5-2013	€ 37,93	€ 10,98	€ 31,70	4-6-2013
26	2681 KV	Monster	€ 17,57	x	€ 20,76	7-5-2013	€ 24,42	€ 0,59	€ 31,53	16-5-2013	€ 7,95	€ 2,39	€ 8,51	23-5-2013
27	9415 PH	Hijken	€ 9,59	x	€ 13,22	12-4-2013	€ 26,45	x	€ 30,39	17-5-2013	€ 76,65	x	€ 81,59	25-7-2013
28	8103 EX	Raalte	€ 13,23	x	€ 16,48	12-6-2013	€ 20,02	€ 9,98	€ 14,49	5-7-2013	€ 19,90	€ 9,95	€ 14,89	2-8-2013
29	5165 AD	Waspik	€ 3,50	€ 3,50	€ 2,75	7-10-2013	€ 4,95	€ 3,95	€ 3,94	14-5-2014	€ 4,95	€ 1,24	€ 7,65	17-10-2014
30	1328 EL	Almere	€ 1,40	€ 0,35	€ 7,60	3-4-2013	€ 89,80	€ 29,63	€ 67,11	1-5-2013	€ 89,80	€ 96,74	€ 0,00	13-5-2013
31	6116 BZ	Roosteren	€ 18,55	€ 3,57	€ 22,42	19-5-2013	€ 70,90	€ 35,45	€ 41,39	23-5-2014	€ 8,95	€ 1,79	€ 11,10	25-6-2014
32	5913 CL	Venlo	€ 5,03	€ 1,51	€ 6,32	25-11-2014	€ 24,21	€ 7,26	€ 22,39	6-4-2015	€ 67,55	€ 20,27	€ 52,22	31-12-2015
33	2165 AM	Lisserbroek	€ 55,61	€ 14,76	€ 49,79	7-5-2013	€ 27,95	€ 5,59	€ 27,30	18-7-2013	€ 4,95	x	€ 8,89	2-4-2014
34	6871 LS	Renkum	€ 4,06	x	€ 7,01	26-8-2013	€ 4,95	€ 4,95	€ 2,94	21-10-2013	€ 14,85	€ 3,95	€ 15,04	6-1-2014
35	3264 XR	Nieuw-Beijerland	€ 11,90	x	€ 14,85	3-7-2013	€ 23,95	x	€ 27,89	26-4-2014	€ 48,90	x	€ 53,44	29-4-2014
36	5011	Tilburg	€ 4,69	x	€ 7,63	14-7-2014	€ 113,70	€ 28,43	€ 91,22	23-10-2014	€ 149,45	€ 44,84	€ 115,56	19-9-2015

	BM													
37	9851 TH	Burum	€ 1,60	x	€ 3,91	20-6-2013	€ 13,95	x	€ 17,89	28-6-2013	€ 21,35	x	€ 25,29	11-10-2013
38	3994 DB	Houten	€ 0,28	€ 0,04	€ 2,55	24-6-2013	€ 87,05	€ 21,76	€ 71,23	14-10-2013	€ 95,05	€ 47,53	€ 52,47	25-6-2014
39	6595 AT	Ottersum	€ 12,62	€ 3,16	€ 13,10	28-5-2014	€ 14,70	x	€ 18,24	31-5-2014	€ 6,95	x	€ 9,89	1-7-2014
40	3344 GL	Hendrik-Ido-Ambacht	€ 0,70	€ 2,02	€ 0,00	3-4-2013	€ 39,95	€ 11,99	€ 32,91	1-5-2013	€ 27,95	€ 13,98	€ 19,46	9-7-2013
41	2914 BN	Nieuwerkerk AD Ijssel	€ 1,47	x	€ 4,22	5-7-2013	€ 35,45	€ 17,73	€ 22,67	23-5-2014	€ 19,10	€ 9,55	€ 13,09	17-9-2014
42	8939 ER	Leeuwarden	€ 16,24	x	€ 19,87	4-2-2014	€ 55,90	€ 27,95	€ 33,89	14-3-2014	€ 11,95	€ 3,95	€ 11,94	15-6-2014
43	5961 NL	Horst	€ 11,95	x	€ 15,58	28-3-2013	€ 13,95	x	€ 17,89	23-12-2013	€ 16,35	x	€ 20,29	8-1-2014
44	7006 JP	Doetinchem	€ 4,55	x	€ 7,50	9-8-2013	€ 17,90	x	€ 22,84	27-1-2014	€ 45,75	x	€ 50,69	12-6-2014
45	2596 XD	s- Gravenhage	€ 1,80	x	€ 4,74	3-7-2014	€ 43,35	€ 21,68	€ 26,62	21-11-2014	€ 9,90	€ 7,90	€ 5,54	15-12-2014
46	7007 ML	Doetinchem	€ 0,82	x	€ 3,13	8-4-2013	€ 4,95	x	€ 7,89	25-4-2013	€ 54,65	€ 27,33	€ 32,27	4-9-2013
47	5671 XG	Nuenen	€ 1,05	x	€ 3,80	18-9-2013	€ 36,35	€ 18,18	€ 23,11	24-10-2013	€ 22,85	€ 6,86	€ 21,54	4-2-2014
48	3531 HA	Utrechts	€ 3,50	€ 3,50	€ 2,75	11-11-2013	€ 9,90	€ 7,90	€ 5,54	1-5-2014	€ 111,30	€ 27,83	€ 89,42	14-1-2015
49	2312 AE	Leiden	€ 3,82	€ 1,15	€ 5,42	4-2-2014	€ 10,95	x	€ 10,95	18-2-2014	€ 21,95	€ 21,95	€ 3,94	25-3-2014
50	1335 CA	Almere	€ 2,31	x	€ 5,06	30-5-2013	€ 44,15	x	€ 49,09	11-10-2013	€ 42,36	x	€ 47,31	30-10-2013

Table 23: Sample of customer road Photos – Photo books – Photo books

VIII. Socio-economic status and order value

Number	Postal code	DT1	AP1	DT2	AP2	DT3	AP 3	AVG
1	1019 WG	16-9-2014	€ 12,54	1-10-2014	€ 24,59	30-11-2014	€ 24,39	€ 20,51
2	5161 HE	17-10-2013	€ 9,11	14-11-2013	€ 15,86	12-9-2014	€ 7,88	€ 10,95
3	2391 GZ	8-4-2013	€ 22,54	6-6-2013	€ 20,09	7-10-2013	€ 9,40	€ 17,34
4	8881 CM	16-10-2014	€ 6,65	5-1-2015	€ 1,39	17-2-2015	€ 1,49	€ 3,18
5	6218 BX	30-7-2013	€ 24,00	28-3-2014	€ 27,94	10-4-2014	€ 31,39	€ 27,78
6	1274 BC	22-8-2013	€ 16,89	19-12-2013	€ 12,15	27-2-2014	€ 21,29	€ 16,78
7	8084 VC	5-5-2013	€ 2,74	14-5-2013	€ 19,82	24-5-2013	€ 2,73	€ 8,43
8	6418 GH	17-7-2013	€ 35,79	20-10-2013	€ 7,89	22-10-2013	€ 3,94	€ 15,87
9	2461 TV	2-5-2013	€ 21,63	12-7-2013	€ 9,55	9-12-2013	€ 19,34	€ 16,84
10	3906 ZH	10-7-2013	€ 15,60	9-12-2013	€ 32,42	5-6-2014	€ 10,14	€ 19,39
11	4207 TL	8-4-2013	€ 15,74	14-5-2013	€ 34,28	21-5-2013	€ 10,15	€ 20,06
12	2802 VD	5-4-2013	€ 5,34	12-4-2013	€ 6,84	18-12-2013	€ 7,74	€ 6,64
13	2871 JV	12-6-2013	€ 4,78	1-11-2013	€ 16,72	24-1-2014	€ 12,06	€ 11,19
14	9937 TH	2-4-2013	€ 7,67	10-9-2013	€ 69,82	12-1-2015	€ 4,69	€ 27,39
15	4907 GV	20-11-2013	€ 24,89	10-1-2014	€ 6,41	2-6-2014	€ 19,94	€ 17,08
16	4874 KS	26-11-2013	€ 12,80	7-12-2014	€ 51,44	9-12-2014	€ 16,59	€ 26,94
17	1862 KN	16-5-2014	€ 5,55	2-6-2014	€ 12,55	16-6-2014	€ 4,34	€ 7,48
18	9678 RJ	11-9-2014	€ 4,94	21-12-2014	€ 4,94	24-12-2014	€ 4,94	€ 4,94
19	5825 BV	15-5-2013	€ 3,15	7-8-2013	€ 38,46	16-12-2013	€ 7,74	€ 16,45
20	7041 HL	11-11-2014	€ 24,65	27-1-2015	€ 11,49	27-1-2015	€ 19,89	€ 18,68
21	1781 CL	23-9-2014	€ 25,88	30-10-2014	€ 22,38	16-12-2014	€ 9,03	€ 19,10
22	3205 BV	30-10-2014	€ 22,97	8-1-2015	€ 53,06	25-2-2015	€ 70,92	€ 48,98
23	6865 EG	18-7-2013	€ 37,91	19-10-2013	€ 19,41	13-12-2013	€ 10,64	€ 22,65
24	2162 VB	20-9-2013	€ 32,56	27-11-2013	€ 31,06	17-12-2013	€ 21,49	€ 28,37
25	1442 VN	10-6-2014	€ 18,01	15-12-2014	€ 11,50	13-2-2015	€ 9,31	€ 12,94

26	4694 EP	9-12-2013	€ 28,04	13-12-2013	€ 28,04	23-4-2014	€ 5,74	€ 20,61
27	5065 SP	16-4-2013	€ 70,29	30-5-2013	€ 17,91	3-7-2014	€ 14,89	€ 34,36
28	3291 LN	11-3-2014	€ 51,90	17-3-2014	€ 6,63	10-4-2014	€ 57,74	€ 38,76
29	1108 HH	18-4-2013	€ 20,89	13-6-2014	€ 7,74	21-11-2014	€ 10,42	€ 13,02
30	6871 TP	24-5-2013	€ 3,15	23-8-2013	€ 15,25	16-10-2013	€ 19,41	€ 12,60
31	8244 DR	28-10-2013	€ 14,09	13-12-2013	€ 4,15	16-4-2014	€ 15,07	€ 11,10
32	6686 CS	10-4-2013	€ 37,92	27-2-2014	€ 8,54	2-5-2014	€ 57,89	€ 34,78
33	4336 BG	24-5-2013	€ 14,27	24-6-2013	€ 19,31	8-7-2013	€ 15,60	€ 16,39
34	8024 CH	13-5-2013	€ 2,94	13-8-2013	€ 3,41	13-9-2013	€ 18,05	€ 8,13
35	2026 XR	27-5-2013	€ 39,89	10-6-2013	€ 32,89	12-6-2013	€ 19,95	€ 30,91
36	8604 BG	11-2-2014	€ 8,95	3-6-2014	€ 17,89	4-11-2014	€ 25,91	€ 17,58
37	6582 AP	11-6-2013	€ 16,49	19-9-2013	€ 22,25	24-12-2013	€ 13,54	€ 17,43
38	4706 DM	28-3-2013	€ 4,50	22-10-2013	€ 27,43	6-2-2014	€ 3,71	€ 11,88
39	7311 GP	5-4-2013	€ 2,94	22-10-2013	€ 3,94	25-2-2014	€ 5,54	€ 4,14
40	4881 HC	26-8-2013	€ 16,89	22-11-2013	€ 4,64	6-1-2014	€ 9,92	€ 10,48
41	4208 BA	27-8-2013	€ 31,51	17-9-2013	€ 59,07	30-12-2013	€ 26,29	€ 38,96
42	2651 BS	28-6-2013	€ 4,01	28-8-2013	€ 37,55	30-12-2013	€ 38,40	€ 26,65
43	4756 CH	13-10-2014	€ 7,21	25-11-2014	€ 19,38	9-1-2015	€ 29,94	€ 18,84
44	5465 RL	15-5-2013	€ 66,61	3-7-2013	€ 18,11	31-7-2013	€ 23,35	€ 36,02
45	7245 AH	12-7-2013	€ 8,43	31-7-2013	€ 5,69	17-9-2013	€ 18,09	€ 10,74
46	7667 TH	16-4-2013	€ 22,40	12-3-2014	€ 82,89	19-1-2015	€ 4,81	€ 36,70
47	3812 CB	3-5-2013	€ 3,97	10-5-2013	€ 13,71	21-5-2013	€ 21,76	€ 13,15
48	6214 RH	29-10-2014	€ 8,33	19-11-2014	€ 4,26	21-12-2014	€ 4,25	€ 5,61
49	1187 WC	16-4-2013	€ 13,85	21-5-2013	€ 15,74	11-6-2013	€ 12,03	€ 13,87
50	9008 TP	17-9-2013	€ 20,04	28-2-2014	€ 14,50	4-3-2014	€ 19,81	€ 18,12

Table 24: Sample socio-economic status and order value