|  |
| --- |
| **Optimizing Media Strategies** |
| **Synergistic effects in the customer journey; can companies benefit from it?** |



**18th of June 2018**

By Faber de Groot

**Optimizing Media Strategies: Synergistic effects in the customer journey; can companies benefit from it?**

F.J. (Faber) de Groot

Hoornsediep 28a

9725HK, Groningen

Tel: 0648141678

e-mail: faber-vlad@hotmail.com

Student number: S3269841

University of Groningen

Faculty Economics and Business

Department of Marketing

Master thesis: Marketing Management & Intelligence

Date: 18-08-2017

Supervisor: Peter van Eck

Second supervisor: Peter Verhoef

**Abstract**

**Preface**

Contents

[**1.** **Introduction** 6](#_Toc516250228)

[**2.** **Theory** 9](#_Toc516250229)

[**2.1** **Firm-Initiated Contacts (FICs) & Customer-Initiated Contacts (CICs)** 9](#_Toc516250230)

[2.1.1 Brand-owned touchpoints 10](#_Toc516250231)

[2.1.2 Partner-owned touchpoints 10](#_Toc516250232)

[2.1.3 Customer-owned touchpoints 10](#_Toc516250233)

[2.1.4 Social/External touchpoints 10](#_Toc516250234)

[**2.2** **Customer Journey** 11](#_Toc516250235)

[**2.3** **Synergistic Effects** 13](#_Toc516250236)

[2.3.1 Synergy 13](#_Toc516250237)

[2.3.2 Firm’s Influence on CICs 14](#_Toc516250238)

[**2.4** **Conceptual Model** 15](#_Toc516250239)

[**3. Methodology** 16](#_Toc516250240)

[**3.1**  **Data Description** 16](#_Toc516250241)

[**3.2** **Variable computation** 19](#_Toc516250242)

[3.2.1 Control variables 19](#_Toc516250243)

[3.2.2 Time-variable 20](#_Toc516250244)

[**3.3** **Sample Description** 21](#_Toc516250245)

[3.3.1 Data Transformation I 21](#_Toc516250246)

[3.3.2 Data Transformation II 21](#_Toc516250247)

[3.3.3 Outliers & Missing Values 21](#_Toc516250248)

[3.3.4 Non-Balanced Dataset 22](#_Toc516250249)

[**3.4** **Data Exploration** 22](#_Toc516250250)

[**3.5** **Model Building** 25](#_Toc516250251)

[3.5.1 Functional Form 25](#_Toc516250252)

[3.5.2 Lagged Variables 26](#_Toc516250253)

[3.5.3 Interaction Effects 26](#_Toc516250254)

[3.5.4 Model Specification 27](#_Toc516250255)

[**4.** **Results** 28](#_Toc516250256)

[**4.1** **Assumptions** 29](#_Toc516250257)

[4.1.1 Binary Outcome 29](#_Toc516250258)

[4.1.2 Linearity Assumption 29](#_Toc516250259)

[4.1.3 Influential Values 30](#_Toc516250260)

[4.1.4 Multicollinearity 31](#_Toc516250261)

[4.1.5 Parallel Lines/Proportional Odds 31](#_Toc516250262)

[**4.2** **Estimation models** 32](#_Toc516250263)

[4.2.1 Purchase at any company 33](#_Toc516250264)

[4.2.2 Purchase at Focal Company 36](#_Toc516250265)

[4.2.3 Purchase Stages 37](#_Toc516250266)

[**4.3** **Modelling Issues** 39](#_Toc516250267)

[4.3.1 Touchpoint Frequency 40](#_Toc516250268)

[4.3.2 Customer Heterogeneity 40](#_Toc516250269)

[**5.** **Conclusion** 40](#_Toc516250270)

[**5.1** **Discussion** 41](#_Toc516250271)

[5.1.1 The effectiveness of initiated contacts 41](#_Toc516250272)

[5.1.2 The effectiveness on the customer journey 42](#_Toc516250273)

[5.1.3 The existence of synergies 42](#_Toc516250274)

[**5.2** **Recommendations and Contribution** 42](#_Toc516250275)

[**5.3** **Limitations and Future Research** 44](#_Toc516250276)

[**6.** **Appendix** 44](#_Toc516250277)

[Literature 50](#_Toc516250278)

1. **Introduction**

Creating a strong customer experience is becoming more and more a leading management objective in the constantly growing digital age we’re living in. Research done by Accenture (2015; in cooperation with Forrester) shows that when executives were being asked about their top priorities for the upcoming year improving the customer experience was mentioned most often (72%) (Lemon & Verhoef, 2016). According to LaSalle and Britton we have entered the experience economy (2003, p. 18); “An economy giving a challenge for facility managers to create environments that enhance the user experience” in which an experience can be described as “a consumer’s beliefs about how a touchpoint fits into his/her life”. Different consumers can have different experiences with the same content (Calder, Malthouse, & Schaedel, 2009). As the elements of a customer experience are critical factors in retaining or attracting customers, it is important to both identify and measure them (Africa, 2010).

Due to the rise of the Internet the customer experience has been changing a lot and gained a new dimension. Compared to the traditional customer experience, the e-commerce environment has enabled consumers to search for information and purchase products or services through direct interaction with the online store. Many online retailers have developed Web-based information systems as a response to this growing e-commerce. With the help of these systems insights can be achieved on the browsing histories and purchase records of individual customers (Verhoef, Kooge, & Walk, 2016). It is very important for firms to deal with this huge amount of data that is available to them. Using the data in an efficient way can result in an effective marketing strategy, which is essential for firms (Bergemann & Bonatti, 2010).

With the arrival of the mobile channel, tablets, social media and the integration of these new channels in online and offline retailing, we are moving from a multi-channel to a omni-channel retailing model. Meaning that there are more channels available but also that the distinction between offline and online channels will disappear, turning the world into ‘a showroom without walls’ (Verhoef, Kannan, & Inman, 2015). Due to these changes, it becomes much harder for companies to monitor the ´customer journey´; “A multistage process that results in a purchase” (Anderl, Schumann, & Kunz, 2015, p. 187). Customers nowadays face a sequence of several touchpoints of different companies in multiple channels and media, making the customer journey more complex (Lemon & Verhoef, 2016).

This study will try to extend the current literature about the online customer journey by using in depth analysis for different types of touchpoints. Moreover, research will be done towards the most effective combination of customer touchpoints, in terms of purchase conversion, and find out to what extent these touchpoints are manageable by the focal company. Furthermore, a customer journey can be divided into several stages: a cognitive stage, an affective stage and a conative stage. This study takes into account that in every stage customers face different kinds of touchpoints and thus show different kinds of behavior. These different types of touchpoints are divided into ‘firm-initiated contacts’ (FICs) and ‘customer-initiated contacts’ (CICs). Since customers prefer to interact with a company on their own terms, CICs are becoming much more important than FICs (Wiesel, Pauwels, & Arts, 2011). This study will also check for synergy effects between touchpoints, since some touchpoints might not be of great effectiveness on their own, but do have a strong effect on others. In this way, these touchpoints can still be very useful (Naik & Raman, 2003). The main research question will therefore be:

*‘Which (combination of) customer touchpoints results in the highest purchase conversion and to what extent is the company able to influence these touchpoints?’*

Besides, this research questions gives several sub-questions:

*‘Which single customer touchpoints result in the highest purchase conversion?’*

*‘Which combination of customer touchpoints result in synergy benefits, leading to a higher purchase conversion?’*

*‘How can different customer touchpoints be assigned to different purchase stages?’*

*‘To what extent can touchpoints be influenced by the focal company?’*

The academic contribution to the current literature is the combination of research done about the customer journey and research done about synergy effects between advertising mediums. Till now, the effectiveness of FICs and CICs has been measured in several ways. However, no distinction has been made between the different stages in the customer journey when measuring the purchase conversion of these touchpoints. Besides, any synergy effects has not been taken into account. Furthermore, new insights will be generated about the influence firms might have on CICs. This study will therefore add to some previous research and generate new insights in the customer journey.

In order to do this research we will use data about a travel agency in the Netherlands, provided by research company GFK.

The structure of the paper is as follows: we will start with a literature review about the different concepts of the customer journey. With the help of the literature different hypotheses and a conceptual model can be developed. Next, a description is given about the used methods for our research followed up by the results of this research. Finally, the findings of our research will be discussed and a conclusion will be formed together with some possible managerial implications and advices towards further research.

1. **Theory**

In this section an overview is given of the current knowledge of the paper’s topic. Based on the literature, we come up with some hypotheses. At the end of this part the hypotheses are represented in a conceptual model.

* 1. **Firm-Initiated Contacts (FICs) & Customer-Initiated Contacts (CICs)**

The customer experience can be defined as a combination of elements that encourage or inhibit consumers during their contact with a company and can be either non-controllable or controllable (Berman & Evans, 1998). Controllable elements are usually named as firm-initiated contacts (FICs). FICs are the traditional marketing communication activities with the focus on pushing messages on to consumers (e.g. television, radio & e-mail). They are defined as “any contact with the customer initiated by the firm” (Wiesel, Pauwels, & Arts, 2011, p. 605).

Next to FICs there are the uncontrollable elements of the customer experience, the customer-initiated contacts (CICs), which are defined as “any contact with a firm that is initiated by a customer or prospective customer” (Wiesel, Pauwels, & Arts, 2011, p. 605). The internet has given consumers the possibility to get in contact with companies on behalf of their own initiative (e.g. organic/paid search, price comparison sites, referrals & retargeting) (Bowman & Naryandas, 2001).

FICs and CICs are customer experience touchpoints, which can be categorized into four categories: brand-owned, partner-owned, customer-owned and social/external. By categorizing touchpoints a framework can be developed to understand potential leverage points in the customer experience. In this way, firms can recognize the controllable touchpoints and be aware of the less- or non-controllable touchpoints (Lemon & Verhoef, 2016).

* + 1. Brand-owned touchpoints

These touchpoints are ways of interactions between the customer and the firm that are developed by the focal firm and under its control. They are a combination of brand-owned media (e.g., advertising, websites, loyalty programs) and brand-controlled elements of the marketing mix (e.g., packaging, service, price).

* + 1. Partner-owned touchpoints

These touchpoints are ways of interactions between the customer and the firm that are jointly developed and controlled by the focal firm and one or more of its partners, such as marketing agencies, multichannel distribution partners, multivendor loyalty program partners and communication channel partners.

* + 1. Customer-owned touchpoints

These touchpoints are ways of interactions between the customer and the firm that are part of the customer experience but are not under control by the focal firm or one of its partners. For example, customers thinking about their needs and desires during the pre-purchase phase (will be discussed later on).

* + 1. Social/External touchpoints

These touchpoints show the influence of third parties in the customer experience. Important external factors can be other customers, peer influences, independent information sources and environments.

|  |  |  |
| --- | --- | --- |
| **Type of Touchpoint** | **FIC** | **CIC** |
| *Brand-owned* | Brand Advertising, Loyalty Programs | Brand Websites |
| *Partner-owned* | Retailer Advertising, Loyalty Programs | Retailer Websites |
| *Customer-owned* | Sponsored Search | Search Engines |
| *Social/External* | Affiliates | WOM, Peers |

Table 1: Categorization of types of touchpoints.

The level of responses to factors under a firm’s control varies across CICs and for this reason firms can get an advantage when they adapt their processing. Instead of providing an overall answer to all CICs, guidelines can be given on how to adjust different responses to various CICs and find a way to improve the efficiency and effectiveness of firms’ CIC management efforts (Bowman & Naryandas, 2001).

According to Blattberg et al. (2008) FICs are becoming less wanted. Controversial, CICs show a lot of potential and have become an important element of firm’s marketing strategies. For example, banner ads have been perceived by many consumers as being annoying (resulting in low conversion rates) (Manchanda, Dube, Goh, & Chintagunta, 2002), while the global paid search advertising market is predicted to have a 37% compound annual growth rate, to more than $33 billion in 2010 (Ghose & Yang, 2009). Moreover, results from research done by Sarner and Herschel (2008) predict that the response rates for CICs are about 15 times higher than for FICs, since this type of contact is less intrusive and is based on customers’ own terms (Shankar & Malthouse, 2007).

Conclusively, it is expected that CICs are the most effective touchpoints in terms of conversion, which leads to the following hypothesis:

*H1: CICs lead to a higher purchase conversion than FICs*

* 1. **Customer Journey**

The customer journey is the cycle of the relationship/buying interaction between the customer and the organization. During this journey the value of customers change.

To get insights in and information about the customer journey with its customer touchpoints companies can develop a customer journey map (CJM). A CJM can be defined as “a visual representation of the customer journey and experience in using a service or space” (Marquez & Downey, 2015, p. 7). By visualizing the complete journey in a map the various stages, steps and touchpoints a consumer goes through can be highlighted and understood (Marquez, Downey, & Clement, 2015). In this way, dependent on the level of complexity, the map visualizes opportunities, pain points and calls to action (Risdon, 2011).

Thus, when customers move toward a purchase they go through a series of stages, including a cognitive (need recognition and information search), an affective (evaluation of alternatives) and, ultimately, a conative stage (choice/purchase) (Wiesel, Pauwels, & Arts, 2011). According to Lemon & Verhoef (2016) the cognitive and the affective stage can be assigned to the pre-purchase phase. The conative stage is similar to the purchase phase. Moreover, they mention another phase, the phase of post-purchase. This phase is about all the interactions of the customer with the brand and its environment after buying. Behaviors during this phase can be characterized as usage and consumption, post-purchase engagement and service requests.

In the online funnel signal web visits (information requests) the beginning of the thought process (cognitive stage). Followed by request for quotes, implying that the customer is evaluating the offer (affective stage). The (final) conative stage is reached when an online order is placed on the website (Wiesel, Pauwels, & Arts, 2011).

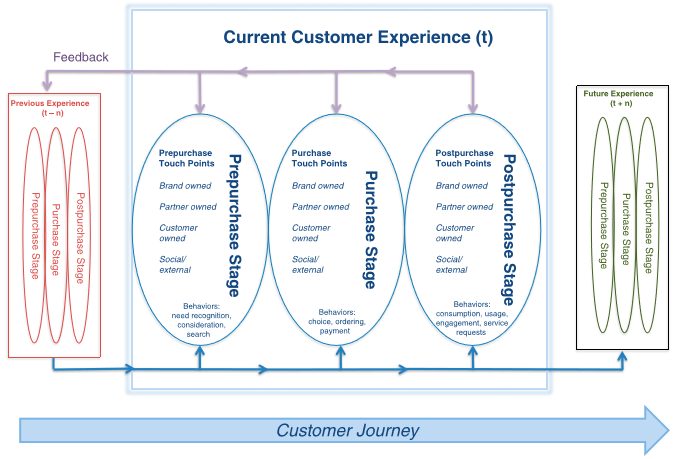


Figure 1: Customer Experience (Lemon & Verhoef, 2016).

During a customer journey, multiple touchpoints or stages are typically required before doing a purchase and these touchpoints have different effects on the purchase likelihood. Furthermore, the strength/importance of touchpoints may differ per stage, depending on the nature of the product/service or the customer’s own journey. FICs can reach consumers that have not yet recognized a need for a certain product category, while CICs assist consumers who have. As a result, CICs should be more directly sales effective than FICs (Haan, Wiesel, & Pauwels, 2016). Moreover, when an initial session was firm-initiated and the following session is customer-initiated it shows a progression towards the purchase decision, while staying in the same channel signals stagnation (Anderl, Schumann, & Kunz, 2015).

Conclusions from prior research give two controversial perspectives about the effects of marketing efforts on the different stages of the ´purchase funnel´. On one hand is said that impersonal marketing communication activities do feed the funnel and bring in prospective customers. However, on the other hand is mentioned that these activities will not feed the funnel but will stuck in the customer´s mind and can be beneficial in later stages of the purchase funnel (Fulgoni & Morn, 2009).

As a customer goes through different stages of the purchase funnel, he faces multiple types of touchpoints. Thus we expect the following:

*H2a: A transition from the cognitive stage to the affective stage is most often caused by FICs*

*H2b: A transition from the affective stage to the conative stage is most often caused by CICs*

* 1. **Synergistic Effects**

This chapter elaborates on the theory of synergistic effects and explains why these effects can be important for companies tracking the customer journey.

* + 1. Synergy

Synergy emerges when the combined effect of two media exceeds their individual effects on the outcome measure (Naik, Integrated Marketing Communications, 2007). The word is derived from the word *synergia* which is ancient Greek and means ‘working together’ (Hindle, 2008). Naik and Peters (2009) describe in their paper two types of synergy effects; within-media synergies (e.g. intra-offline) and cross-media synergies (e.g. offline-online). The difference is that there can be synergy effects between offline media components such as television and print (within-media) but also synergy effects between offline media and online media (cross-media).

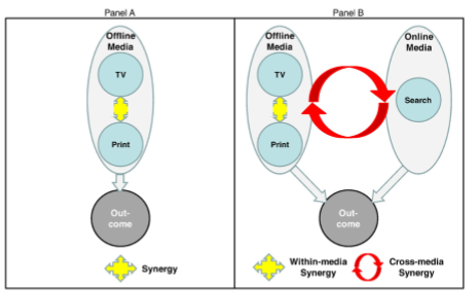


Figure 2: Synergy effect (Naik & Peters, 2009).

When within-media synergy exists, the total media budget should be increased and more than fair share should be given to the less effective medium, which will reinforce the more effective medium (Naik & Raman, 2003). Moreover, when uncertainty is included it results in the catalytic effect; “A non-zero amount should be allocated to media even if their own effectiveness is zero, provided they exhibit positive synergy with other media in the communications mix” (Raman & Naik, 2004, p. 11). However, if the various mediums are equally effective, the media budget should be allocated equally as well among the mediums, regardless of the size of synergy (Naik, 2007).

Until now, synergy effects are only described for offline and online mediums. However, these effects can also be found in a situation wherein media activities are categorized as CICs and FICs. As mentioned earlier, to benefit from synergies a company needs to be able to influence both of the mediums. Therefore, we need to find out to what extend companies are able to have an influence on CICs.

* + 1. Firm’s Influence on CICs

While there are different types of CICs, search engines are the types used mostly by customers. A distinction can be made in the way customers search; brand search or generic search. Brand search already shows brand awareness, since the customer is searching for a specific brand. Generic search however, does not show brand awareness, since the customer is searching for general information (Ghose & Yang, 2009). Rutz and Bucklin (2008) showed that there are spillover effects between the two types of search, as many customers start with a generic search to gather information and use a brand search to finish their transaction.

Although search engines are categorized as CICs, firms can influence the customers’ search behavior. First of all, firms can pay search engines to be on top of the search list, also known as sponsored search advertising. Firms that are on top of the list have an advantage over firms appearing lower at the list (Arbatskaya, 2007). A large proportion of consumers believes that a company higher in the listed search results sells products of higher quality than the companies listed below (Animesh, Ramachandran, & Viswanathan, 2010). Furthermore, visiting the website of a company is a CIC, but the appearance of the website can be fully designed by the company itself. In this way, companies can have a lot of influence on the CICs. Another argument is that effective FICs (e.g. banners) can be an incentive for customers to start their search behavior (CICs). In this way, the quality of FICs should have an influence on CICs.

Since firms appear to have an influence on CICs, synergies between touchpoints do matter. Therefore the next hypotheses can be established:

*H3a: Within-contact synergies exist for FICs*

*H3b: Within-contact synergies exist for CICs*

*H3c: Cross-contact synergies exist for FICs & CICs*

In case we find synergy effects, we expect the following:

*H3d: Cross-contact synergies are stronger than within-contact synergies*

*H3e: Within-contact synergies of CICs are stronger than within-contact synergies of FICs*

* 1. **Conceptual Model**

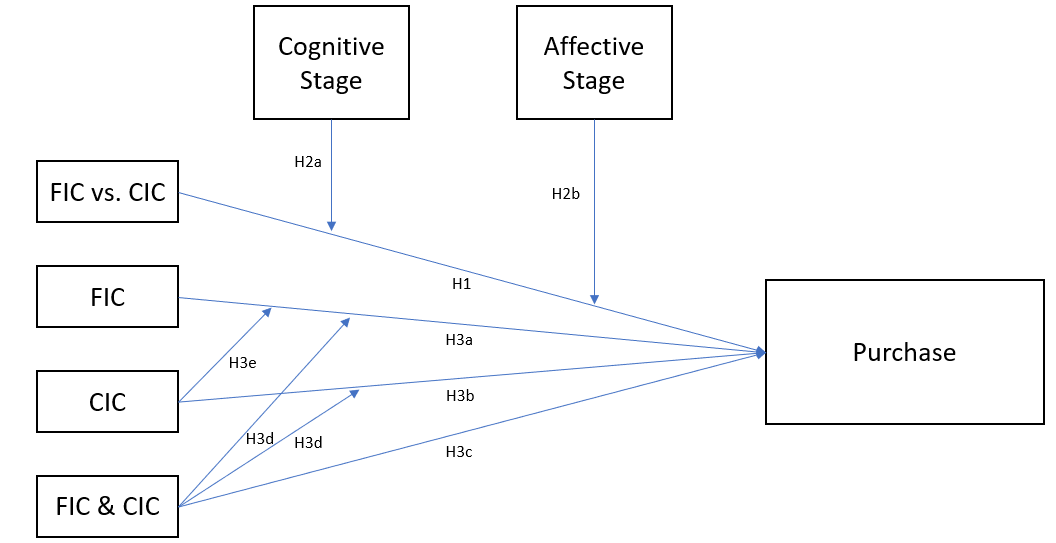


Figure 3: Conceptual framework initiated contacts influencing purchase behavior

**3. Methodology**

This chapter discusses the data that is used to answer the research question. First, a description of the data in this study is given. Secondly, the different variables included in the model will be defined and explained. Furthermore, after doing data transformations a sample description is given combined with some graphs of the data exploration. Lastly, to test the established hypotheses a model specification is built.

## **3.1 Data Description**

For this study the data of a Dutch travel agency is used, collected and provided by research company GFK in the period from June 2015 until the end of September 2016. Research company GFK collected the data of the online customer journey of a Dutch travel agency by using an intelligent system (a plug-in) that analyses the internet usage of panelists. This plug-in is called a browser-extension and has three functions; (1) all the URL’s calls of all the household users will be registered in all the PCs, (2) the advertisement shown to the user is identified, (3) retrieval queries in advertising-relevant search engines are registered. Together, the collected data is delivered to the GFK server. In this way, for every single person the customer journey can be tracked, from exposure to purchase behavior.

To obtain the data for this dataset quantitative research is done containing information about different customer events, purchase information and related demographics. The dataset is event-based, meaning that the data is encountered on daily basis and that they are collections of ordered sequences of events (Vrotsou, 2010). Each of these events has an own starting time and an own duration. The first measured event starts on May 31th in 2015 and the last measured event ends on November 30th 2016. Thus, measurements of events are done within a period of 1.5 years. The event data shows the type of touchpoints a customer (*UserID*) faces during a certain customer journey (*PurchaseID*) at a certain moment in time. There are 22 different types of touchpoints that can be faced, categorized into CICs and FICs. The purchase information shows whether a certain purchase journey led to a purchase (Yes/No) and if this purchase was made at the focal company or at one of the competitors. The time variable reflects the date and time that the touchpoint has been seen. Additionally, the demographics reflect the information of age, gender and income of the prospect.

When customers face different types of touchpoints and move towards doing a purchase, they go through a set of purchase stages. To increase the chance of purchasing, it is important for a focal company to recognize the stage a customer is in. The table below shows the 22 different types of touchpoints, wherein also a column for different purchase stages is added. These stages are categorized based on the type of touchpoint that is given.

|  |  |  |  |
| --- | --- | --- | --- |
| **# in Dataset** | **Type of Touchpoints** | **FIC/CIC** | **Purchase Stage** |
| 1 | Accomodations Website | CIC | Affective/Conative Stage |
| 2 | Accomodations App | CIC | Affective/Conative Stage |
| 3 | Accomodations Search | CIC | Affective Stage |
| 4 | Information / comparison Website | CIC | Affective Stage |
| 5 | Information / comparison App | CIC | Affective Stage |
| 6 | Information / comparison Search | CIC | Affective Stage |
| 7 | Touroperator / Travel agent Website Competitor | CIC | Affective/Conative Stage |
| 8 | Touroperator / Travel agent App Competitor | CIC | Affective/Conative Stage |
| 9 | Touroperator / Travel agent Search Competitor | CIC | Affective Stage |
| 10 | Touroperator / Travel agent Website Focus brand | CIC | Affective/Conative Stage |
| 12 | Touroperator / Travel agent Search Focus brand | CIC | Affective Stage |
| 13 | Flight tickets Website | CIC | Affective/Conative Stage |
| 14 | Flight tickets App | CIC | Affective/Conative Stage |
| 15 | Flight tickets Search | CIC | Affective Stage |
| 16 | generic search | CIC | Cognitive Stage |
| 18 | AFFILIATES | FIC | Cognitive Stage |
| 19 | BANNER | FIC | Cognitive Stage |
| 20 | EMAIL | FIC | Cognitive Stage |
| 21 | PREROLLS | FIC | Cognitive Stage |
| 22 | RETARGETING | FIC | Affective Stage |

Table 2: Type of Touchpoints

Since FICs are mostly types of contacts that give an incentive to start a customer journey, except for retargeting, they are assigned to the cognitive stage of the purchase journey. The cognitive stage does include search behavior as well and for that reason generic search is also assigned to this stage. In the next stage, the affective stage, customers exhibit the behavior of evaluating of alternatives. The information/comparison website is an example of an initiated contact during this stage. However, not every initiated contact belongs to only one stage, since some of them can belong to more stages. The accommodations website for instance can be used for evaluating as for purchasing behaviors. In this way, it is not always easy to derive what exact stage the customer is in. Therefore, the third stage is categorized as a combination between the affective and the conative stage.

Whereas there is no touchpoint of purchasegiven in the data, there is assumed that the purchase takes place right after the last touchpoint that is registered in the data. Other touchpoints that are observed very likely but are not included in the dataset are brand search, own searchandcompetitor search. For this reason, only one ‘search touchpoint’ initiated by the customer is included in the dataset.

There is assumed that the last touchpoint in time led to the purchase done, so we need to track very carefully which touchpoint was the last touchpoint of the users customer journey. To be able to do this, a certain customer journey can be recognized in the dataset by its purchase ID. Moreover, since one user can do multiple purchases but also no purchase it can happen that a single user has been tracked in multiple customer journeys over time, so that he or she has multiple purchase IDs. Therefore, the dataset contains much more Purchase IDs than User IDs.

In summary, the focus of this paper is the effectiveness of single- and combinations of touchpoints on moving to a next purchase stage or purchasing. To measure the effectiveness of these touchpoints lag effects and interaction effects (synergy) will be included.

## **Variable computation**

* + 1. Control variables

Next to the main variables that are included in the model to test for the different hypotheses, a few control variables are added to test for their effect on the response variable. These control variables can explain variance of the model that otherwise would have been caught by other predictor variables or be left without any explanation (Leeflang, 2015). The control variables that are included in our model are the length of the customer journey, gender, age and income (see *Appendix*). The length of the customer journey can be calculated in two ways. One way is by measuring the number of touchpoints that led to a purchase. A higher touchpoint frequency increases brand awareness and might influence the brand attitude (Yaveroglu & Donthu, 2008). The other way is by creating a new time variable out of the original time variable that is already in the dataset. The original time variable views both the date and the time. By calculating the average length in time of all customer journeys together, we can create a new categorical variable of the length of the customer journey with the following ordinal categories; ‘short’, ‘normal’ and ‘long’. The same can be done for the first method, but then by calculating the average frequencies.

Since the type of touchpoints show what stage a customer is in, we have chosen to determine the length of the customer journey by measuring the number of touchpoints. Moreover, companies support a quick transition to the next stage and this method gives better insights why a customer is stuck in a certain stage.

The other three variables are demographic factors that are expected to have an influence on the response variable. As a result, a higher income for instance could decrease the search behavior, since we expect people who have more money to be less price sensitive (Soba & Aydin, 2012). Therefore this effect can shorten the customer journey that might otherwise have led to biased results and conclusions. In this way, the length of customer journey can be seen as a mediator variable with the demographic variables as control variables influencing this mediator and the response variable. For clarification, in *figure ‘.’* below are the description and the index of each of the control variables given.

|  |  |  |
| --- | --- | --- |
| **Variable** | *Description* | Index |
| **GenderID** | *Gender* | 1: Male, 2: Female |
| **Age** | *Age* | Age in Years |
| **BAS\_bruto\_Jaarinkomen** | *Gross-Income* | 1: < €12.900 (minimum) |
|  |  | 2: €12.900 <= 27.000 (below average) |
|  |  | 3: €27.000 <= 33.500 (almost average) |
|  |  | 4: €33.500 <= 40.000 (average) |
|  |  | 5: €40.000 <= 67.000 (between 1 and 2 times average) |
|  |  | 6: €67.000 <= 79.900 (2 times average) |
|  |  | 7: > 79.900 (above 2 times average) |
|  |  | 8: don't know / don't want to say |
| **JourneyLength** | *Length of the Customer Journey* | 1: 1 <= 20 (short) |
|  |  | 2: 20 <= 100 (middle) |
|  |  | 3: > 100 (long) |

Table 3: Control Variables

Adding these variables in our model will increase our model fit but may result in overfitting. This reduces the generalizability of the model beyond the data on which the model is fit (Leeflang, 2015). For this reason, the control variables will be added one at a time.

Since there is no data of pricing or about the content of the different touchpoints available, the model does not account for the effect of these variables. Information about these variables would increase the explained variance of the model, but it is not of significant importance for our research.

* + 1. Time-variable

In the dataset the time variable contains the date and the time of a certain touchpoint. To be able to make use of this variable, it is split up in a date and a time variable. From here, it becomes easier to derive reliable results from the dataset, since we can order the dataset first by Purchase ID, next by date and last by time. From here, two necessary data transformations are possible.

## **Sample Description**

* + 1. Data Transformation I

In this paper we want to investigate the effect of the last or the last two touchpoints on purchasing. To be able to conduct this research and get the wanted results the dataset should be transformed. First, the Purchase IDs that only contain one touchpoint usually don’t result in a sale and more importantly don’t give us the possibility to check for synergy effects. As a result, these Purchase IDs are removed from the dataset. Second, for every Purchase ID all information is removed except for the information about the last two touchpoints per Purchase ID. In this way, the effect of the last touchpoint and the effect of the combination between the last and the second last touchpoint on purchasing can be measured.

* + 1. Data Transformation II

To investigate for the effect of ICs that cause a transition to another purchase stage another transformation is done. For this transformation all the data is removed except for the observations that led to a transition to the next purchase stage within the customer journey. Now, we can get insights in the effectiveness of ICs on transitioning to the next stage of the customer journey.

### Outliers & Missing Values

Outliers and missing values show wrong or no values for an observation and for this reason the outcome of a conducted analysis can be biased. In the end, to get reliable results, it’s necessary to check for these outliers and missing values.

The control variables *‘Age’, ‘Gender’* and *‘BAS\_bruto\_jaarinkomen’* that will be discussed later contain missing values (NA’s). To be able to get reliable results from these variables the observations of the missing values are removed from the dataset.

The dataset contains one Purchase ID that consists of 64,503 number of touchpoints. Since this biases our results of the control variable ‘*Journeylength’,* this particular Purchase ID is removed from the dataset.

The variable *‘BAS\_bruto\_jaarinkomen’* is categorized in 8 different classes. The last class (#8) contains all the users that *‘don’t know’* or *‘don’t want to tell’* their salary. Therefore, this class can be seen as missing values and will be removed from our dataset.

* + 1. Non-Balanced Dataset

In the total dataset 3,674 purchases have been done from which 192 with the focal company, while there are 12,251 Purchase IDs. Before transformation I, each Purchase ID contained at least one but most often multiple touchpoints/observations. This is expected since a customer journey usually contains several touchpoints before it is converted into a purchase, if it already converts. If a certain customer journey led to a purchase, the variable of *‘purchase\_own’* or/and *‘purchase\_any’* is given the value of 1 during the whole journey. Nevertheless, both of the purchase variables mostly contain zero values. Moreover, as can be seen in the graph below, even after transformation I the dependent variables contain a lot more zero’s than one’s.

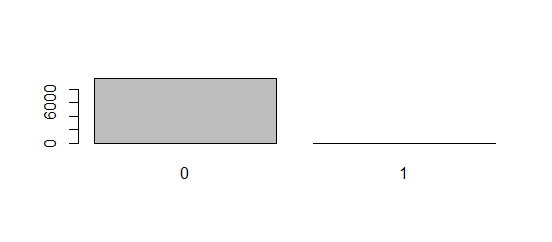
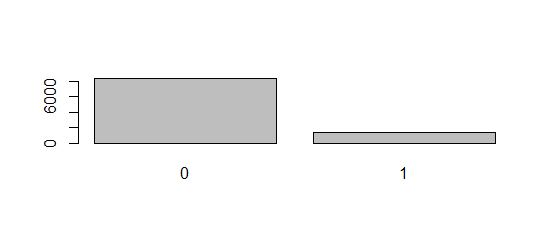


Figure 5: Amount of Purchases (1st: purchase\_any, 2nd: purchase\_own)

The mass of zero’s will cause reliability problems when we want to make predictions of the data. This problem is known as the corner solution, in which the quantity of one of the arguments in the maximized function is zero (Leeflang, 2015). To solve this problem, we can make use of down or up sampling (Leeflang, 2015). With this method we create a balanced dataset, wherein the number of customers who did purchase is set equal to the number of customers who didn’t purchase. However, since we prefer a representative dataset over a predictive dataset we don’t make any changes. In this way, our results are realistic but are not of good use for further predictions.

* 1. **Data Exploration**

To gain some further insights in the data, a few graphs are plotted. In this chapter, these graphs will be shown and analyzed.

As can be seen in *figure 6*, the database contains 22 different types of touchpoints. However, touchpoint ‘11’ and touchpoint ‘17’ were never used and for that reason have no name (see *table 2*). Furthermore, the first 16 touchpoints are CICs and the last 6 touchpoints are FICs in which the accommodations website (‘1’) and the tour operator website of the competitor (‘7’) are seen the most, both being CICs. Additionally, the FIC that has been used the most is retargeting (‘22’).

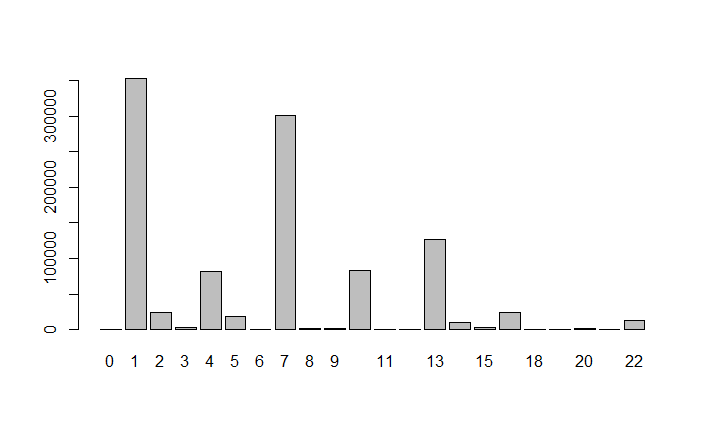
****

Figure 6: Type of Touchpoints 2

To get more valuable results, a new variable is created that makes a distinction between the CICs and FICs within the dataset. This categorical variable is called *‘IC’* which stands for initiated contact. It is created because it divides the 22 different types of touchpoints into 2 groups, *customer initiated contacts* and *firm initiated contacts*. By creating this variable we can analyze the effects of initiated contacts on purchasing and make the model less complex.

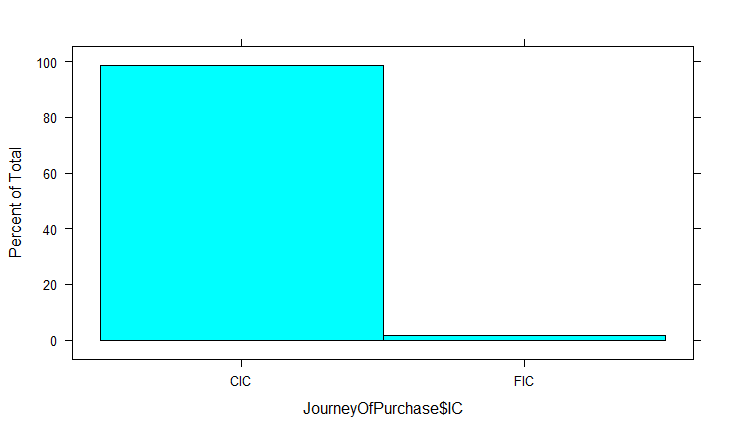


Figure 7: Initiated Contacts (after transformation I)

*Figure 7* shows that after transformation of the data more that about 99 percent of the initiated contacts are CICs. In fact, of the 24,504 observations left, 24,238 are CICs and 266 are FICs. Thus, when formulating conclusions of the results we have to pay attention to the percental difference in usage.

Next, another categorical variable is created called *´Purchase Stage´*. For this variable the touchpoints are split up in a *cognitive stage*, an *affective stage* and a *conative stage,* as mentioned earlier. With the help of this variable we can get better insights about the purchase stage the customer is in and transitioning to. In the end, the effectiveness of different ICs in different purchase stages can be measured.

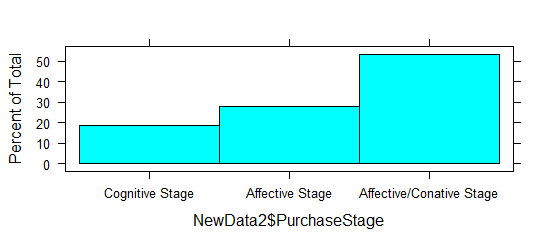


Figure 8: Purchase Stages (after transformation II)

As mentioned earlier, it’s not possible to categorize all touchpoints to one particular stage of the customer journey. For that reason, we combined the affective stage and the conative together as the final stage. Conclusively, after transformation II most of the touchpoints seen, belong to the affective/conative stage (see *figure 8*).

A third categorical variable that is added to the dataset is called *‘JourneyLength’*. This variable is created by counting the number of touchpoints within a certain customer journey (PurchaseID). A customer journey with less than 20 touchpoints is categorized as *short*, between 20 touchpoints and 100 touchpoints is categorized as *normal* and more than 100 touchpoints is categorized as *long.* These numbers are based on the mean, the median, the maximum and the minimum of the number of touchpoints.

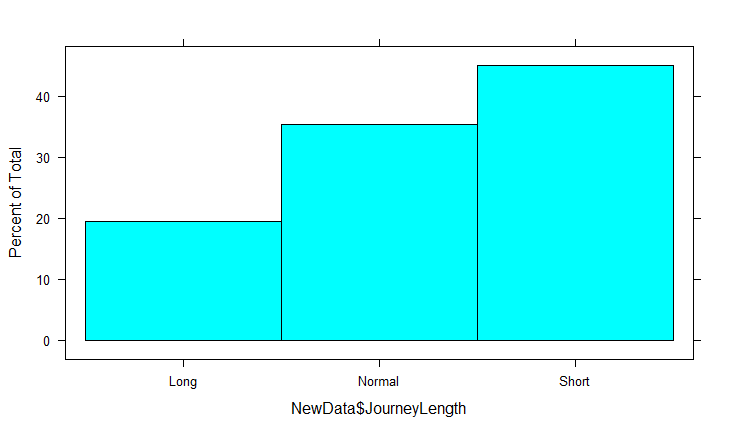


Figure 9: Length of the Journey (after transformation I)

*Figure 9* shows that most of the customer journeys, even after transformation I (= deleting all Purchase IDs that only contain one touchpoint), are short ones. Therefore customers will most often need less than 20 touchpoints to decide if they will convert their journey into a purchase or not.

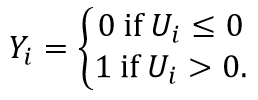
* 1. **Model Building**

When building a model different choices should be made. First we have to make a choice between a descriptive, predictive or normative model. For this research a descriptive model is used, since we want to investigate the effects of touchpoints on purchasing for this particular travel agency.

Furthermore, there are four steps to take when building a model; (1) specification, (2) estimation, (3) validation and (4) use (Leeflang, 2015). In this section we discuss the specification of the model.

* + 1. Functional Form

For this research the response variable is based on whether a purchase is done yes or no. Whether people purchase and what they purchase depends on their utility. People purchase if their utility of purchasing is larger than their utility of not purchasing (Upurchase > Unot purchasing). However, people’s utility is not observable. For that reason, we are making use of probabilities (Leeflang, 2015). Accordingly, the dependent variable is a binary variable and should therefore contain a value within the boundaries of 0 and 1. The latent utility and the observed decision are linked as follows:



Hence, we have to make use of a logistic or probit regression. The relationship between the response variable (Y) and the predictor variable (Xi) for this model is sigmoidal (S-shaped), rather than a straight line. Moreover, instead of choosing parameters that minimize the sum of squared errors as in OLS (Ordinary Least Squares), we are choosing parameters that maximize the likelihood of observing the sample values (Maximum Likelihood Estimation) (Hosmer, Lemeshow, & Sturdivant, 2013). Since there are no relevant differences between a logistic and probit function we choose to use a logistic function which gives easier interpretation possibilities.

The logistic regression has a different interpretation than a normal distributed regression. The coefficient of the logistic regression is called the ‘Odds Ratio’ and is equal to EXP(. The odds ratio “compares the odds that an outcome will occur given a particular exposure to the odds that an outcome will occur in the absence of that particular exposure” (Hosmer, Lemeshow, & Sturdivant, 2013, p. 157).

* + 1. Lagged Variables

To account for synergistic effects between touchpoints, lagged variables are included in the model. A lagged variable is a past period version of a predictor variable in the current period, which can be projected in the following way:

A regression model including lagged variables is called a distributed lag model.

By including a lagged variable of *‘IC’*, the current chance of purchasing can be predicted by using both the current type and the past period type of initiated contacts. In this way, we can check for not only the effect of the last touchpoint on purchasing but also the effect of touchpoints that were seen before.

* + 1. Interaction Effects

Since we want to check for the presence of synergy effects between types of touchpoints, interaction effects should be included in our model. If the effect of one predictor variable at a particular level of the other predictor variable changes, we can speak of an interaction effect. If the coefficient of the interaction term is statistically significant it would suggest that the effect of the two mutations is synergistic. Nonetheless, if the coefficient is not significant we can conclude that the effect is additive (Assari, 2014). An interaction effect between a variable and its lagged version can be added in the following way:

Because of using interaction effects, we have to create two models. One model with the allowance for interaction effects and one without. For both of the models the main effects are kept in the model, even if not statistically significant. For the model without an interaction effect, the value of the parameter of the interaction effect is set to 0.

Furthermore, a fifth model is created to get answers on the hypotheses about the purchase stages. This model shows the effect of touchpoints during the different stages and is less focused on whether a purchase is done or not.

* + 1. Model Specification

Eventually, five models are built and estimated; First, (1) a model for **purchasing at any company** without the interaction effect between ICs and (2) a model for **purchasing at any company** with the interaction effect between ICs. Secondly, (3) A model for **purchasing at the focal company** without the interaction effect between ICs and (4) a model for **purchasing at the focal company** with the interaction effect between ICs. Lastly, (5) a model for measuring the effects **on transitioning between stages**.

This gives us the following model specification:

**(1 & 2):**

**(3 & 4):**

**(5)**

**Index:**

*=* Unique user ID

*=* The moment in time (day, hour, minute)

= Lagged variable of the moment in time (day, hour, minute)

= Constant

= Parameter estimate

= Binary variable of a purchase at any company for user i at a moment in time t

= Binary variable of a purchase at focal company for user i at moment in time t

= Ordinary variable of purchase stage that user i is in at moment in time t

= Dummy variable of initiated contact for user i at a moment in time t

Dummy variable of the length of the customer journey for user i at a moment in time t

Age of user i at a moment in time t

Dummy variable of gender for user i at a moment in time t

Dummy variable of gross income for user i at a moment in time t

= Disturbance term, for user i at moment in time

The dataset consists of a big amount of user IDs, each with an own combination of customer touchpoints. Therefore the data is structured in a pooled format; all the observations of different Purchase IDs are grouped in one data frame. With this model, all parameters are assumed to be the same across different users (Leeflang, 2015).

Furthermore, the statistical analyses used for this research are a multiple logistic regression and a ordered logistic regression. For this reason, the parameters are estimated with maximum likelihood estimation (MLE).

To check if the MLE estimates are correct we need to check for several assumptions (Leeflang, 2015). Violation of these assumptions causes wrong estimates of the parameters and of the variance of the parameters. If the assumptions of MLE are violated the model has to be re-specified and the pooled model can’t be used.

1. **Results**

Within this chapter we discuss step 2 (estimation) an step 3 (validation) of building a model. Since we want to be able to answer our hypotheses in the estimation section, we start with checking for our assumptions (validation).

* 1. **Assumptions**

The logistic regression has 4 assumptions; binary outcome, linearity assumption, influential values and multicollinearity. For the ordered logistic regression one other assumption is tested; parallel lines. The results of these tests are described below.

* + 1. Binary Outcome

For the first four models, the dependent variable can be either ‘Purchase’ or ‘No Purchase’ (e.g. 0 & 1, ‘pos’ & ‘neg’). To test for this assumption, the predicted probabilities of the first logit model (for other models, see *Appendix*) can be calculated. The first six observations give the following outcomes:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 |
| "Neg" | "Neg" | "Neg" | "Pos" | "Neg" | "Neg" |

Table 4: Predicted Probabilities (model 1)

This result shows that the dependent variable can be positive or negative. In other words, the dependent variable has a binary outcome. Furthermore, since most of the observations didn’t result in a purchase, the probability of purchasing will most often be negative.

* + 1. Linearity Assumption

A logistic regression assumes linearity of independent variables and log odds. This does not require the dependent and the independent variables to be related linearly, but it does require the independent variables to be linearly related to the log odds. Since *‘Age’* is the only numeric independent variable in our models, we only test this assumption for one variable.

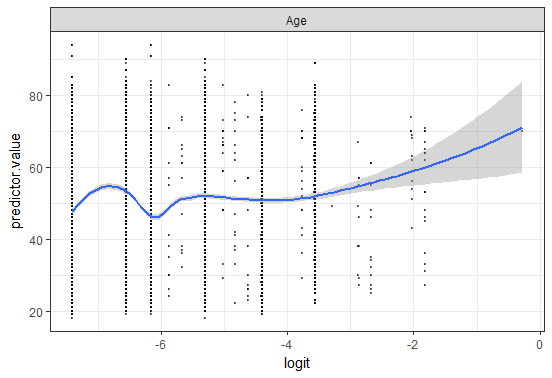


Figure 10: Linearity Assumption for Age (model 4)

*Figure 10* shows a non-linear relationship for the variable *‘Age’* (*model 4*)*,* as the scatterplot shows a random distribution instead of a linear one (the dots are distributed in straight vertical lines instead of a linear line). To solve this problem, we can include 2 or 3-power terms, fractional polynomials or spline functions. However, since *‘Age’* is only a control variable in our model we don’t necessarily need to turn our model around. The variable *‘Age’* on itself will not change our conclusions and therefore we don´t make any changes.

* + 1. Influential Values

Influential values are extreme individual data points that can alter the quality of the logistic regression model, like outliers. To detect which outliers should be removed we can use Cook’s distance. By visualizing Cook’s distance values the most extreme values are considered to merit closer examination. *Figure 11* visualizes the top three most extreme values labeled (of *model 4*) with its observation number. However, not all outliers are influential values and therefore we also plot the absolute standardized residuals.

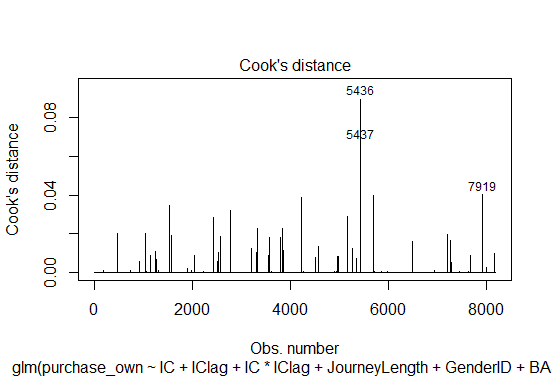
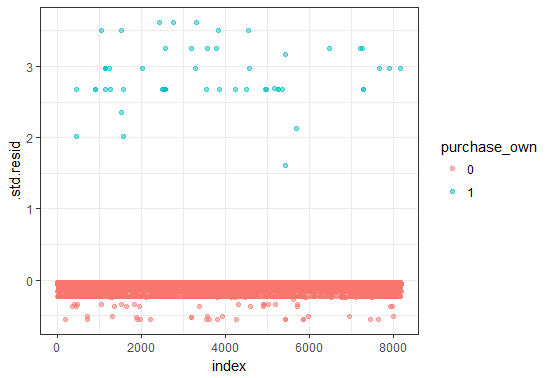
 

Figure 11: Cook’s Distance (model 4) Figure 12: Standardized Residuals (model 4)

The standardized residual is “a measure of strength of the difference between observed and expected values” (Stefanie, 2017, p. 1). Data points with an absolute standardized residual value above 3 may represent an outlier and need further investigation. As can be seen in *figure 12*, some of the observations (17 to be precise) of the standard residuals actually have a value higher than 3. Consequently, these values are possible outliers and have a large effect on the slope of a regression line fitting the data. For that reason, these observations are removed from our dataset.

* + 1. Multicollinearity

The fourth assumption for logistic regressions is the assumption of multicollinearity. This assumption is violated when the included predictor variables are not independent of each other. If multicollinearity is found within the model, it can lead to biased parameters and too large variances. By calculating Variance Inflation Factors (VIF) of the predictor variables we can test for multicollinearity in our model.

**Variables GVIF Df GVIF^(1/(2\*Df))**

IC 1.162 1 1.078

IClag 1.104 1 1.051

JourneyLength 1.008 2 1.002

GenderID 1.127 1 1.062

BAS\_bruto\_jaarinkomen 1.126 6 1.010

Age 1.134 1 1.065

IC:IClag 1.267 1 1.126

Table 5: GVIF Scores (model 2)

If any terms in an unweighted linear model have more than 1 degrees of freedom the Generalized Variance Inflation Factor (GVIF) is calculated instead of the VIF (Fox & Monette, Generalized Collinearity Diagnostics, 1992). The GVIF is interpretable as “the inflation in size of the confidence ellipse for the coefficients of the predictor variable in comparison with what would be obtained for orthogonal data” (Fox & Weisberg, 2011, p. 188). All the GVIF-scores have a value below 5 which indicates that the multicollinearity assumption has not been violated. Therefore, multicollinearity is not an issue and no changes to our model have to be made.

* + 1. Parallel Lines/Proportional Odds

The ordered logit model assumes all regression lines to be parallel. In this way, the intercepts should be different from one another but all the independent variables should have the same coefficients in all equations. Therefore, we test for this assumption by graphing predicted logits from individual logistic regressions with a single predictor where the outcome groups are defined by either ‘*PurchaseStage’* >= 2 or *‘PurchaseStage’* >= 3. If the difference between predicted logits for varying levels of a single predictor are almost the same whether ‘Y>=2’ or ‘Y>=3’, the parallel lines assumption holds (Liao, 1994).

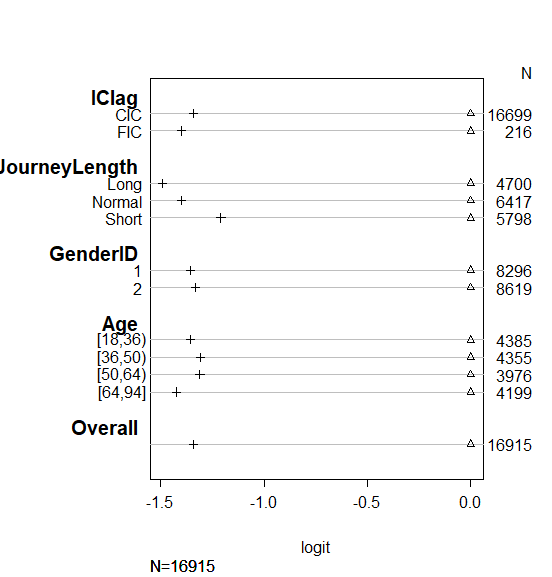


Figure 13: Parallel Lines

By using one predictor variable at a time, the values (plus signs) in *figure 13* are linear predictions from a logit model, used to model the probability that the dependent variable is greater than or equal to a given value. *‘IC’* is missing in the graph above, because there are no FICs in the third stage of the customer journey and therefore it has an infinite outcome for (Y=>3) (see *Appendix Parallel Lines)*.

In *figure 13* the ‘N’ displays the number of observations per variable level. Furthermore, when the plus signs (in *figure 13*) for the different levels of a particular variable are close to each other the parallel lines assumption is not violated. The outcomes show that the plus signs for every level of a predictor are close and therefore the parallel lines assumption holds. This indicates that for example the effect of being a male or a female is the same for a transition from the cognitive stage to the affective stage as from the affective stage to the affective/conative stage. Accordingly, we can conduct an ordered logistic regression.

# **Estimation models**

In this section the conducted models, after validation, are estimated and the results are given. Based on the results the hypothesis are accepted or rejected.

As we analyze the odds of a certain variable, the values of the other variables are held constant so that we don’t make any wrong conclusions. This doesn’t count for the interaction variables, since it doesn’t make sense to fix the interaction variable at a certain value and still allow one of the two combined variables to change.

### Purchase at any company

By estimating the first logit model (without interaction effect) the effect of initiated contacts and it’s lagged variable on purchasing at any company can be investigated. Furthermore, to account for heterogeneity between consumers control variables are added to the model. Since these variables are added one at a time, we can check for the relevance of these variables. We observe a higher AIC of the model and significance for the variables ‘*JourneyLength’, ‘Gender’* and *‘BAS\_bruto\_jaarinkomen’*, while the variable *‘Age’* is not significant and doesn’t improve our model. For this reason, the control variable *‘Age’* will be left out of the first model.

The estimates and the significance of the estimates remain about the same for the first and the second logit model. Therefore, *table 3* shows the outcomes of the second logit model including the interaction effect. The model without the interaction effect can be found in the *Appendix* (*Logit model 2*).

*Table 6* shows us that the categorical variable *‘IC’* is highly significant (p=0.003), with a lower or negative effect for FICs as for CICs. Furthermore, the lagged variable *‘IClag’* is not significant (p=0.822) and does not significantly affect purchasing.

From the results of the control variables we can conclude that the length of the customer journey significantly affects (p=0.000) purchasing. Moreover, the longer the journey the higher the chance of purchasing. Secondly, the variable of *‘GenderID’* is significant (p=0.044), with a higher chance of purchasing for males as for females. Lastly, the income variable shows significant effects for higher levels of income (#5, #6, #7), with the highest chance of purchasing for income class 6.

In the model below the interaction effect is added. Since *‘IClag’* is not significant the interaction effect *‘IC:IClag’* is not significant either. Therefore we can’t conclude that there are synergies between the last and the second last touchpoint for any company.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variables** | **Estimate** | **Standard Error** | **Z-value** | **Pr(>|z|)** |
| *(Intercept)* | -0.863 | 0.172 | -5.023 | 5.08e-07 \*\*\* |
| ICFIC | -1.277 | 0.436 | -2.927 | 0.003 \*\* |
| IClagFIC | 0.016 | 0.351 | 0.045 | 0.964 |
| JourneyLengthNormal | -1.226 | 0.074 | -16.469 | < 2e-16 \*\*\* |
| JourneyLengthShort | -2.565 | 0.010 | -26.053 | < 2e-16 \*\*\* |
| GenderID2 | -0.141 | 0.070 | -2.016 | 0.0438 \* |
| BAS\_bruto\_jaarinkomen2 | 0.054 | 0.185 | 0.291 | 0.771 |
| BAS\_bruto\_jaarinkomen3 | 0.227 | 0.185 | 1.232 | 0.218 |
| BAS\_bruto\_jaarinkomen4 | 0.239 | 0.178 | 1.338 | 0.181 |
| BAS\_bruto\_jaarinkomen5 | 0.547 | 0.170 | 3.212 | 0.001 \*\* |
| BAS\_bruto\_jaarinkomen6 | 0.695 | 0.189 | 3.670 | 0.000 \*\*\* |
| BAS\_bruto\_jaarinkomen7 | 0.641 | 0.202 | 3.177 | 0.001 \*\* |
| ICFIC:IClagFIC | 0.926 | 1.220 | 0.759 | 0.450 |
| *Signif. Codes: 0 '\*\*\*'* | *0.001 '\*\*'* | *0.01 '\*'* | *0.05 '.'* | *0.1 ' ' 1* |

Table 6: Logit model 2 (purchase\_any)

There are 3 ways to access the impact of the independent variables of a multiple logistic regression: (1) interpretation of the coefficients, (2) interpretation of the odds ratio and (3) interpretation of the marginal effects (Leeflang, 2015). Above, we already used the first interpretation. The odds ratio is the most useful interpretation when investigating relationships between variables while marginal effects are more useful when investigating the effect of one extra unit. Since both interpretations give extra information, they will both be discussed.

The odds ratio shows us the likelihood of ‘happening vs. not happening’. When the odds ratio has a value higher than one it indicates a positive relation, while an odds ratio with a value lower than one indicates a negative relation. Consequently, for the interpretation of the values below one, the factor should be inversed.

In *table 7* the odds ratio for the categorical variable *‘IC’* when it is an FIC is 0.279, expecting that the second last touchpoint *(IClag)* is an CIC (due to the inclusion of the interaction effect). This means that when someone’s last touchpoint is an FIC, the odds for this customer to make a purchase increase by factor 0.279 compared to someone who’s last touchpoint is an CIC. In other words, the likelihood of purchasing is about 3.5 times (calculation: 0.279 ≈ 28/100, inverse = 100/28 ≈ 3.5) smaller if the last touchpoint is an FIC compared to when it would have been an CIC. We can’t draw any conclusions if the last and the second last touchpoint are an FIC because the interaction effect is not statistically significant. Secondly, the odds for the length of the journey show that the probability of purchasing is about 3.5 times smaller if the journey is normal and 13 times smaller if the journey is short, compared to when the journey would have been long. Furthermore, if the customer is a male he is 1.15 times more likely to do a purchase compared to if the customer would have been a female. Lastly, if the customer has an income within the fifth, sixth or seventh income level he is about 1.7 to 2 times more likely to do a purchase than when he would have had an income of level one.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables** | **Odds** | **dF/dx** | **P>|z|** |
| *(Intercept)* | 0.422 |  |  |
| ICFIC | 0.279 | -0.072 | 4.205e-07 \*\*\* |
| IClagFIC | 1.016 | 0.001 | 0.964 |
| JourneyLengthNormal | 0.293 | -0.101 | < 2.2e-16 \*\*\* |
| JourneyLengthShort | 0.077 | -0.235 | < 2.2e-16 \*\*\* |
| GenderID2 | 0.868 | -0.013 | 0.043 \* |
| BAS\_bruto\_jaarinkomen2 | 1.055 | 0.004 | 0.774 |
| BAS\_bruto\_jaarinkomen3 | 1.255 | 0.022 | 0.247 |
| BAS\_bruto\_jaarinkomen4 | 1.269 | 0.023 | 0.206 |
| BAS\_bruto\_jaarinkomen5 | 1.728 | 0.055 | 0.003 \*\* |
| BAS\_bruto\_jaarinkomen6 | 2.003 | 0.079 | 0.003 \*\* |
| BAS\_bruto\_jaarinkomen7 | 1.899 | 0.073 | 0.009 \*\* |
| ICFIC:IClagFIC | 2.525 | 0.121 | 0.567 |
| Signif. codes: 0 '\*\*\*' 0.001 '\*\*' | 0.01 '\*' | 0.05 '.' | 0.1 ' ' 1 |

Table 7: Odds ratio and Marginal effects (model 2)

The marginal effects show the same significance as the earlier calculated estimates. The marginal effects can be interpreted as follows; the significant variable *‘ICFIC’* with a value of -0.072 means that if the last touchpoint was an FIC the probability that a purchase has been done is about 7 percentage points lower than in a situation where the last touchpoint was a CIC. The variable *‘JourneyLengthShort’* gives a value of -0.235 and is significant as well. This means that if the customer journey was short (<20 touchpoints) the probability that a purchase has been done is about 24 percentage points lower than in a situation where the customer journey was long (>100).

Conclusively, to determine which model is the most optimal for predicting a purchase at any company goodness-of-fit measurements are used.

**Model Type AIC LL** **Pseudo R2**

Including Age, no Interaction Effect 5751.5 -2862.7 0.144

Without Interaction Effect, no Age 5750.2 -2863.1 0.144

Including Interaction Effect, no Age 5751.7 -2862.9 0.144

Null model 6688.3 -3343.1 0.000

Table 8: Goodness-of-fit measurements model 1 & 2

The models don’t differ much and therefore the goodness-of-fit measurements differ not much either. Since we are investigating synergy effects, it is preferred to include the interaction effect. Therefore we choose to use the third model of *table 4.*

Because of these results we can conclude that H1 is accepted for any company since CICs do have a significant positive effect on purchasing compared to FICs. Furthermore, we found that the interaction effect is not significant and therefore we can’t conclude that there are any synergistic effects between the last two touchpoints on purchasing at any company. Therefore, if the dependent variable is any company H3a,b,c are rejected.

* + 1. Purchase at Focal Company

For the third and fourth logit model the dependent variable changes. Instead of measuring the effects of purchasing at any company we are now measuring the effects of purchasing at our own company. By adding control variables one at a time, we find no significant effects for *‘Age’* and for *‘BAS\_bruto\_jaarinkomen’.* Moreover, the AIC and the pseudo R2 do not improve when adding these variables. Thus, adding these variables to the model will not improve our model and are therefore left out of the model.

Again, the estimates and the significance of the model with and without the interaction effect are about the same. Therefore the model without an interaction effect can be found in the *Appendix.* The estimate of the variable *‘IC’* is significant (p=0.001) and positive for FICs. This means that the chance of purchasing at the focal company is higher when the last initiated contact was an FIC, given that *‘IClag’* was an CIC. Additionally, for purchasing at the focal company *‘IClag’* is significant (p=0.038) as well and has a positive estimate for FICs. The length of the journey and the gender of the customer have about the same significance and estimates as in the models for purchasing at any company. Additionally, the estimate of *‘Gender’* is more negative for women as in the previous models so that men in comparison are even more likely to buy. The interaction effect is not significant (p=0.971) and can therefore not be interpreted.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variables** | **Estimate** | **Standard Error** | **Z-value** | **Pr(>|z|)** |
| *(Intercept)* | -3.560 | 0.197 | -18.071 | < 2e-16 \*\*\* |
| ICFIC | 1.745 | 0.548 | 3.183 | 0.001 \*\* |
| IClagFIC | 1.551 | 0.749 | 2.070 | 0.038 \* |
| JourneyLengthNormal | -1.747 | 0.338 | -5.173 | 2.30e-07 \*\*\* |
| JourneyLengthShort | -3.001 | 0.528 | -5.682 | 1.33e-08 \*\*\* |
| GenderID2 | -0.845 | 0.308 | -2.743 | 0.006 \*\* |
| ICFIC:IClagFIC | -0.054 | 1.467 | -0.037 | 0.971 |

Table 9: Logit model 4 (purchase\_own)

To access the impact of the independent variables we can check again for the odds ratio and the marginal effects of these variables. For *‘IC’* the odds ratio is 5.726, meaning that if the last touchpoint is an FIC a customer is 5.7 times more likely to do a purchase at the focal company (if the second last touchpoint was a CIC) than when the last touchpoint would be an CIC. Additionally, if the second last touchpoint (*IClag*) is an FIC a customer is 4.7 times (odds = 4.714) more likely to do a purchase at the focal company (if the last touchpoint would be an CIC) than when the second last touchpoint would have been an CIC. Thirdly, if the length of the customer journey is long a customer is about 6 times (calculation: 0.174 ≈ 17/100, inverse = 100/17 ≈ 6) more likely to purchase than if the length would have been normal and 20 times more likely than if the journey would have been short. Lastly, if the customer is a male he is about 2.5 times more likely to do a purchase than if the customer would have been a female.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables** | **Odds** | **dF/dx** | **P >|z|** |
| *(Intercept)* | 0.028 |  |  |
| ICFIC | 5.726 | -0.073 | 1.774e-07 \*\*\* |
| IClagFIC | 4.714 | -0.002 | 0.955 |
| JourneyLengthNormal | 0.174 | -0.104 | < 2.2e-16 \*\*\* |
| JourneyLengthShort | 0.050 | -0.241 | < 2.2e-16 \*\*\* |
| GenderID2 | 0.430 | -0.022 | 0.001 \*\*\* |
| ICFIC:IClagFIC | 0.948 | 0.109 | 0.593 |
| Signif. codes : 0 '\*\*\*' 0.001 '\*\*' | 0.01 '\*' | 0.05 '.' | 0.1 ' ' 1 |

Table 10: Odds ratio and Marginal effects (model 4)

As can be seen in *table ‘.’* *‘IClag’* does not have a significant marginal effect and therefore we can not interpret its value. In contrary to the odds, the probability of purchasing decreases with 7 percental points if the last touchpoint was an FIC.

To determine which model is the most optimal, goodness-of-fit measurements are used again.

**Model Type AIC LL** **Pseudo R2**

Without Interaction Effect, no Age & Income 561.73 -274.86 0.141

Without Interaction Effect, with Age & Income 563.4 -268.70 0.160

Including Interaction Effect, no Age & Income 563.73 -274.86 0.141

Null model 641.98 -319.99 0.000

Table 11: Goodness-of-fit measurements model 3 & 4

Based on the goodness-of-fit measurements given above, the second model, without interaction effect but with the variables of *‘Age’* and *‘BAS\_bruto\_jaarinkomen’,* is preferred the most since it has the lowest Log-Likelihood and the highest pseudo R2. The AIC punishes for a higher amount of variables and has therefore not the most optimal value out of the four models.

In the results we found that for the focal company an FIC as a last touchpoint has a more positive effect on purchasing than an CIC. Therefore, for the focal company we reject H1. Furthermore, as for second logit model, we didn’t find any significant effect between the last and the second last touchpoint. For that reason we have to reject H3a,b,c as well for purchasing at the focal company.

* + 1. Purchase Stages

To measure the effectiveness of predictor variables on transitioning to another stage in the purchase journey an ordered logistic regression is used. In the assumptions chapter we already tested the parallel lines assumption, in which we proved that all estimates have similar values for different levels of the predictor variables but different intercepts. This result allows us to continue with the ordered logistic regression; a logit model using an ordered dependent variable with more than two levels. Our dependent variable is categorized into three categories ordered as follows; *‘a cognitive stage’, ‘an affective stage’* and *‘an affective/conative stage’*. Nevertheless, the interpretation of this model is somewhat difficult, since it is only directly possible in terms of latent scale. Therefore, we can only say something about the direction of the estimate and not about the probabilities. Thus for instance, if we find positive estimates for a certain customer, the more likely this customer is to be in a higher stage of the customer journey.

For the fifth logit model (see *table 12*), *‘BAS\_bruto\_jaarinkomen’* isn’t significant and doesn’t improve the model. Hence, this variable has been left out of the model. The other predictor variables all remain significant and improve our model. First of all, the predictor *‘IC’* is significant (p=0.000) and demonstrates a negative estimate. If the last touchpoint is an FIC the customer is less likely to be in a higher stage of the customer journey than if the last touchpoint is an CIC. Conversely, for the second last touchpoint the opposite effect is visible, but less strong. Furthermore,a remarkable difference with the coefficients of the previous models is the significance of *‘Age’* with a p-value of 0.000. If a customer becomes older the more likely he or she is to be in a higher stage in the purchase journey. Next, the length of the customer journey seems to have a different effect on transitioning to another stage as that it did on purchasing. If the customer has a short journey the higher is his chance to be in a higher stage of the journey. Lastly, females seem to have a significantly higher likelihood to be in a higher stage than males.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Variables** | **Estimate** | **Standard Error** | **Z-value** | **Pr(>|z|)** | | ICFIC | -2.611 | 0.010 | -26.392 | < 2e-16 \*\*\* | | IClagFIC | 0.324 | 0.139 | 2.339 | 0.019 \* | | JourneyLengthNormal | 0.390 | 0.037 | 10.667 | < 2e-16 \*\*\* | | JourneyLengthShort | 0.788 | 0.039 | 20.383 | < 2e-16 \*\*\* | | GenderID2 | 0.116 | 0.031 | 3.781 | 0.000 \*\*\* | | Age | 0.011 | 0.001 | 11.870 | < 2e-16 \*\*\* | | *Signif. Codes: 0 '\*\*\*' 0.001 '\*\*'* | *0.01 '\*'* | *0.05 '.'* | *0.1 ' '* | *1* | |  |  |  |  |

Table 12: Logit model 5 (PurchaseStage)

To better interpret the estimates the odds ratio and the marginal effects are calculated (see *table 13*). For FICs as a last touchpoint, the odds of being in the highest stage (affective/conative) versus the combined middle (affective) and low (cognitive) stage are about 14 times (calculation: 0.071 ≈ 7/100, inverse = 100/7 ≈ 14) less likely than for CICs as a last touchpoint. Likewise, the odds of the combined high and middle stage versus low stage is 14 times lower for FICs compared to CICs. On the other hand, the results show that if a customer’s second last touchpoint (*IClag*) was an FIC the odds of being in the highest stage versus the combined middle and low stage are about 1.38 times higher. Again, the odds of the combined high and middle stage versus low stage is 1.38 times higher for FICs compared to CICs. A normal length of the journey has a factor of 1.48 and a short length has a factor 2.2 compared to when it would have been a long journey. Females have a factor of 1.12 compared to males. Age is a continuous variable and has therefore another approach. For a one year increase in age, the odds of being in a high stage versus the combined middle and low stages are 1.01 times greater. Similarly, for a one year increase in age, the odds of the combined high and middle stage versus low stage are 1.01 times greater.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables** | **Odds** | **dF/dx** | **P >|z|** |
| Cog|Aff Stage | 0.543 |  |  |
| Aff|Aff/Con Stage | 2.307 |  |  |
| ICFIC | 0.071 | -0.483 | < 2.2e-16 \*\*\* |
| IClagFIC | 1.383 | 0.047 | 0.000 |
| JourneyLengthNormal | 1.476 | -0.024 | 0.010 \* |
| JourneyLengthShort | 2.200 | -0.047 | 1.479e-07 \*\*\* |
| GenderID2 | 1.123 | 0.016 | 0.004 \*\* |
| Age | 1.011 | 0.002 | < 2.2e-16 \*\*\* |
| *Signif. codes: 0 '\*\*\*' 0.001 '\*\*'* | 0.01 '\*' | 0.05 '.' | 0.1 ' ' 1 |

Table 13: Log odds and Marginal effects (model 5)

Since we also want to know which initiated contact is more effective in the beginning of the purchase journey, we turned the model around by changing the order of the dependent variable. The table with the results can be found in the *Appendix (model 5)*. The results show a significant positive estimate (factor 13.5) for FICs compared to CICs, meaning that FICs are more effective in the earlier stages of the customer journey.

We already concluded that the variable *‘BAS\_bruto\_jaarinkomen’* should be left out of the model. Moreover, we don’t need to compare any models. Therefore we would choose to use the discussed model above.

From here, we can conclude that our 2nd hypothesis can be accepted. FICs seem to be more sales effective in earlier stages of the customer journey, while CICs seem to be more sales effective in later stages of the customer journey.

* 1. **Modelling Issues**

This chapter discusses some of the issues that are taken account of in our models by including control variables.

* + 1. Touchpoint Frequency

One of the issues that can affect our logistic purchasing model is the number of touchpoints faced. People differ in the length of their search behavior before doing a purchase, especially when spending large amounts of money, like at a travel agency. Since a higher amount of touchpoints can lead to better informed customers it can also change their purchase behavior. Therefore, to get more accurate estimates and a higher explained variance a *‘JourneyLength’* variable is added to the model.

* + 1. Customer Heterogeneity

Exploration of the customer-related data reveals heterogeneity between the observed possible customers. Heterogeneity between customers leads to different preferences and can affect someone’s purchase decision making process. As an example, being a male might shorten the decision making process and therefore affect the dependent variable. To account for the heterogeneity between customers the variables *‘Age’, ‘Gender’* and *‘BAS\_bruto\_jaarinkomen’* are included in our models.

1. **Conclusion**

The aim of this study was to analyze the effects of (combinations of) initiated contacts on purchasing. To do this analysis, multiple logit models were conducted on an event-based dataset of a Dutch travel agency, provided by GFK. By including lagged variables and interaction effects we tested for the presence of synergy effects between initiated contacts. Furthermore, an ordered logistic regression was conducted to analyze which type of initiated contacts cause a transition to a next stage in the purchase journey. To control for heterogeneity between customers some control variables were also added to the models.

*Table 14* shows all the tested hypotheses and whether they were accepted or rejected. The remaining part of this chapter will start with a discussion of the results, followed by theoretical and managerial implications of this study and in the end the limitations and future directions of this research will be discussed.

|  |  |  |
| --- | --- | --- |
| **#** | **Hypotheses** | **Reject/Accept** |
| H1 | *CICs lead to a higher purchase conversion than FICs* | R |
| H2a | *A transition from the cognitive stage to the affective stage is most often caused by FICs* | A |
| H2b | *A transition from the affective stage to the conative stage is most often caused by CICs* | A |
| H3a | *Within-contact synergies exist for FICs* | R |
| H3b | *Within-contact synergies exist for CICs* | R |
| H3c | *Cross-contact synergies exist for FICs & CICs* | R |
| H3d | *Cross-contact synergies are stronger than within-contact synergies* | NA |
| H3e | *Within-contact synergies of CICs are stronger than within-contact synergies of FICs* | NA |

Table 14: Hypotheses

* 1. **Discussion**

In this part the results are discussed and a more clear answer will be given on the acceptance or rejection of our hypotheses and research question.

### The effectiveness of initiated contacts

Our first hypothesis states that CICs lead to a higher purchase conversion than FICs. Therefore we tested the effect of the last seen touchpoint during a customer journey on whether a person did a purchase or not. We found evidence that for any company CICs have a more positive effect than FICs on purchasing (factor 3.5). However, when purchasing from the focal company was set as a dependent variable we found a more positive effect for FICs (factor 5.6) than for CICs. This difference can be declared by the fact that the FICs that were observed for this dataset were owned by the focal firm. In this way, these FICs will have a more negative effect for competitors but a more positive effect for the focal company.

Based on this reasoning, FICs might be more sales effective for the focal company. This is in contradiction with previous findings, which proofs that CICs are 15 times more sales effective than FICs (Serner & Herschel, 2008). Though, this conclusion can be refuted. When the dataset would’ve contained FICs of the focal company as from the other companies it would give different estimates and possibly different conclusions. Moreover, other declarations are that the differences in outcomes can be caused by the fact that for this research we only measured the effect of the last touchpoint. So we didn’t include the effect of the other touchpoints on purchasing. Additionally, there is a big contrast in use of initiated contacts. As we have seen before, our dataset consists of way more CICs as FICs, especially in the last stage of the customer journey. This huge difference can bias our results and therefore lead to wrong or less accurate estimates.

### The effectiveness on the customer journey

The second hypothesis is split up in two sub hypotheses (a & b). We tested, by conducting an ordered logistic regression, the effects of initiated contacts on transitioning to another stage in the purchase journey. We found that FICs (factor 13.5) are more likely to be found and to be effective in earlier stages of the purchase journey, while CICs (factor 14) are more likely to be found and to be effective in later stages of the purchase journey. This result is in agreement with our expectations and therefore both of our hypotheses can be accepted. Moreover, the results showed that if *‘IClag’* is an FIC a customer is more likely to be in a higher stage (factor 1.38) as when the second last touchpoint would have been an CIC. This is in agreement with previous findings, as Anderl, Schumann & Kunz (2015) said that switching between channels signals progression while staying in the same channel signals stagnation. Thus, when a FIC is followed up by a CIC the prospective customer is more likely to be in a higher stage of the customer journey.

* + 1. The existence of synergies

The third and last hypothesis is split up in 5 sub hypotheses (a, b, c, d & e). By conducting a multiple logistic regression we tested for the presence of synergies with purchasing as a dependent variable. For *‘purchase\_any’* we did find a significant effect for *‘IClag’* but for neither *‘purchase\_any’* as *‘purchase\_own’* we found significance for the interaction effect between *‘IC’* and *‘IClag’.* Therefore we conclude that there are no synergies between the last and the second last touchpoint for purchasing at either any or at the focal company. Consequently, we rejected our first three hypotheses and couldn’t test for the last two. As the literature shows proof for synergies between online and offline mediums, we can’t find any proof for synergies between initiated contacts. Again, just as for the first hypothesis, this might has to do with the fact of only taking the last two touchpoints in account. Another argument is that the testing for synergies between offline and online mediums are both fully controllable by the customer while this is not the case for initiated contacts.

* 1. **Recommendations and Contribution**

This research was conducted to advice a particular Dutch travel agency in managing their touchpoints. Moreover, also academics can learn from this study. This research shows the effectiveness of touchpoints on the purchase journey.

As we already explained in the theoretical framework, although CICs are characterized as contacts initiated by the customer, firms do have influence on CICs. For example, a quality FIC can start someone’s search behavior and therefore start using CICs. Moreover, a company can optimally design its own website, while the website is categorized as a CIC. Although no synergy effects were found between the last two touchpoints (categorized as initiated contacts) of the customer journey, there might be synergies between particular types of touchpoints. Additionally, for the focal company a more positive effect was found for FICs than for CICs, leading to a higher purchase conversion. For that reason, we recommend to optimize spending on FICs, but also investigate in research about how to better influence CICs.

Furthermore, we found that FICs are more effective during the cognitive stage, while CICs become more effective during later stages in the purchase journey. This again proves that FICs can incentivize search behavior. Therefore, it is recommended that the travel agency tracks carefully what purchase stage a customer is in. Based on that knowledge, the company can decide how to make purchasing more attractive for this particular customer.

By tracking clickstream-data and customer info, the company can also account for customer heterogeneity. Our models showed that Gender, Age and Income all had a certain influence on transitioning or purchasing. Hence, based on this information, initiated contacts or types of touchpoints can be adjusted to a certain type of customer. Later on, this will increase the purchase conversion. Moreover, we saw that the longer the journey lasts, the higher the chance of purchasing. Whereas a company has the possibility to lengthen the customer journey, it can increase the likelihood of purchasing.

The main focus during this study is on one particular Dutch travel agency, but we expect that the outcomes of the study are generalizable for similar companies from the same branch. As we think that the gained insights about the effects of customer touchpoints and the combinations between them will have the same effect in similar situations. Some of these new insights are in contrast with the existing literature, wherein Blattberg et al. (2008) mentions that FICs are becoming less wanted. Our research shows that FICs are of great value to incentivize a customer’s search behavior and increases the company’s purchase conversion. Furthermore, where Naik & Raman (2003) proved that synergy effects exist between online and offline advertising mediums, we didn’t find any synergy effects in our research between initiated type of contacts. It doesn’t mean that synergy effects don’t exist between initiated contacts, but in our research a significant effect on purchasing wasn’t found.

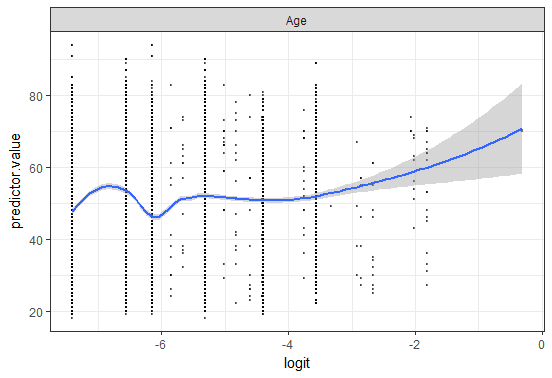
* 1. **Limitations and Future Research**

Our study possesses several limitations that can be interesting opportunities for further research. First, instead of gaining insights in the type of touchpoints we categorized the touchpoints into two groups, namely FICs and CICs. Due to this categorization, we missed out on a lot of deeper insights. Future research can gain more insights in particular types of touchpoints and testing for synergies between them. Secondly, due to the huge amount of zeroes in the dependent variables *‘purchase\_any’* and *‘purchase\_own’*  our results can’t be used to do further predictions. In future research it is advised to use over or down sampling and develop a balanced dataset. It makes the outcome less realistic, but the model can be used to predict. Thirdly, in our research we only included a binary variable for purchasing. Since customers are price sensitive, adding a continuous variable for pricing or sales can also explain why and what a customer bought at the competitor and not at the focal company. In most purchasing scenarios, especially for travel agencies, prices play a huge role. Lastly, the different types of touchpoints are, based on theory, categorized to different purchase stages. However, since a purchase stage is quite an abstract phenomena, it isn’t a totally fair distribution. Therefore, the results might be somewhat biased. For further research we advise to determine purchase stages in a different way (e.g. tracking behaviors more accurately).

1. **Appendix**

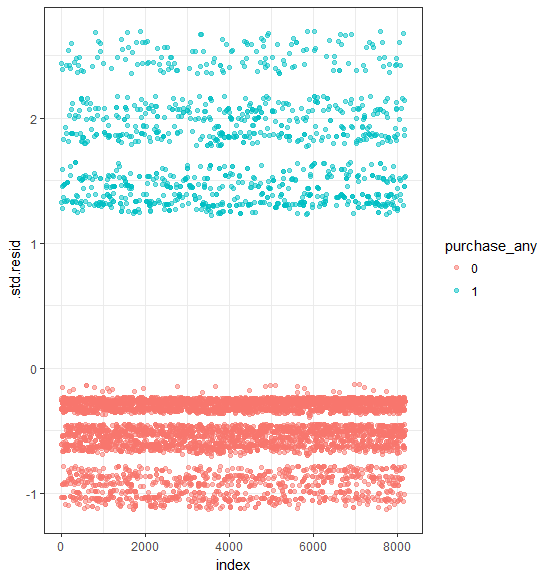
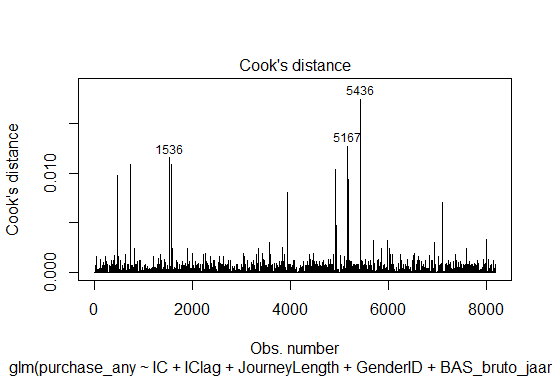
**Binary outcome:**

|  |  |  |
| --- | --- | --- |
| Model 2:   |  | | --- | | 1 2 3 4 5 6 | | “neg” “neg” “neg” “pos” “neg” “neg” |     Model 3: |
| |  | | --- | | 1 2 3 4 5 6 | | “neg” “neg” “neg” “neg” “neg” “neg” |   Model 4:   |  | | --- | | 1 2 3 4 5 6 | | “neg” “neg” “neg” “neg” “neg” “neg” | |  |   **Linearity assumption:**  Model 1:    Model 2:    Model 3: |

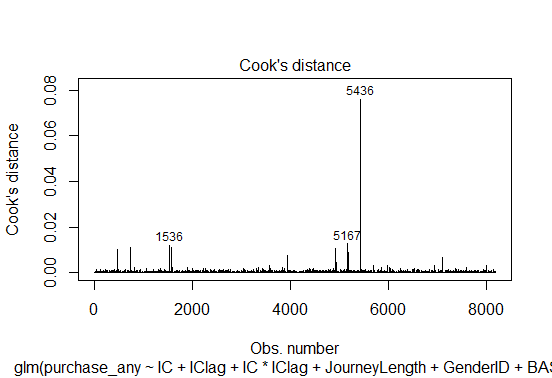
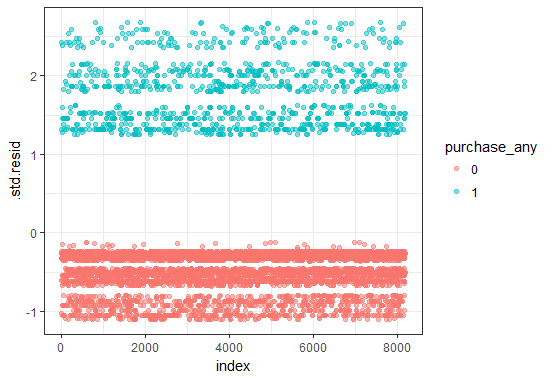
****

**Influential values:**

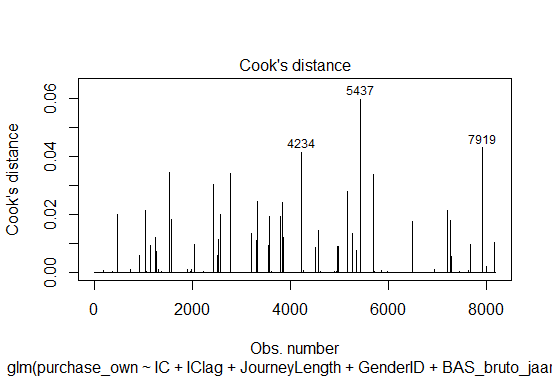
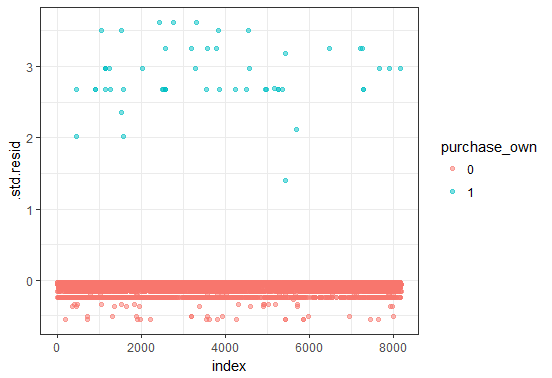
Model 1:



Model 2:

Model 3:

**Multicollinearity:**

Model 1:

|  |
| --- |
| GVIF Df GVIF^(1/(2\*Df)) |
| IC 1.005407 1 1.002700 |
| IClag 1.004028 1 1.002012 |
| JourneyLength 1.007205 2 1.001796 |
| GenderID 1.053973 1 1.026632 |
| BAS\_bruto\_jaarinkomen 1.058458 6 1.004746 |

Model 3:

|  |  |  |  |
| --- | --- | --- | --- |
| GVIF Df GVIF^(1/(2\*Df)) |  |  |  |
| IC 1.059852 1 1.029491 | | | |
| IClag 1.060699 1 1.029902 | | | |
| JourneyLength 1.017975 2 1.004464 | | | |
| GenderID 1.097986 1 1.047848 | | | |
| BAS\_bruto\_jaarinkomen 1.075915 6 1.006116 | | | |
| Age 1.114739 1 1.055812 | | | |

Model 4:

|  |
| --- |
| GVIF Df GVIF^(1/(2\*Df)) |
| IC 1.225686 1 1.107107 |
| IClag 1.435772 1 1.198237 |
| JourneyLength 1.032883 2 1.008121 |
| GenderID 1.098759 1 1.048217 |
| BAS\_bruto\_jaarinkomen 1.086345 6 1.006925 |
| Age 1.116873 1 1.056822 |
| IC:IClag 1.713710 1 1.309088 |

**Parallel lines assumption:**

|  |
| --- |
| | | |N |Y>=1|Y>=2|Y>=3 | |
| +-------------+-------+-----+----+----+---------+ |
| |IC |CIC |16394|Inf |0 |-1.389382| |
| | |FIC | 521|Inf |0 | -Inf| |
| +-------------+-------+-----+----+----+---------+ |
| |IClag |CIC |16699|Inf |0 |-1.342239| |
| | |FIC | 216|Inf |0 |-1.399816| |
| +-------------+-------+-----+----+----+---------+ |
| |JourneyLength|Long | 4700|Inf |0 |-1.490340| |
| | |Normal | 6417|Inf |0 |-1.400758| |
| | |Short | 5798|Inf |0 |-1.208909| |
| +-------------+-------+-----+----+----+---------+ |
| |GenderID |1 | 8296|Inf |0 |-1.355591| |
| | |2 | 8619|Inf |0 |-1.330642| |
| +-------------+-------+-----+----+----+---------+ |
| |Age |[18,36)| 4385|Inf |0 |-1.356204| |
| | |[36,50)| 4355|Inf |0 |-1.307808| |
| | |[50,64)| 3976|Inf |0 |-1.310177| |
| | |[64,94]| 4199|Inf |0 |-1.423638| |
| +-------------+-------+-----+----+----+---------+ |
| |Overall | |16915|Inf |0 |-1.342870| |

**Conducted models:**

Model 1:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Coefficients: | |  |  |  |  |  |  |
| Estimate Std. Error z value Pr(>|z|) | | | | | | | |
| (Intercept) -0.86505 0.17183 -5.034 4.79e-07 \*\*\* | | | | | | | |
| ICFIC -1.18518 0.40589 -2.920 0.00350 \*\* | | | | | | | |
| IClagFIC 0.07604 0.33713 0.226 0.82154 | | | | | | | |
| JourneyLengthNormal -1.22499 0.07443 -16.458 < 2e-16 \*\*\* | | | | | | | |
| JourneyLengthShort -2.56403 0.09843 -26.050 < 2e-16 \*\*\* | | | | | | | |
| GenderID2 -0.14097 0.07002 -2.013 0.04410 \* | | | | | | | |
| BAS\_bruto\_jaarinkomen2 0.05380 0.18478 0.291 0.77092 | | | | | | | |
| BAS\_bruto\_jaarinkomen3 0.22916 0.18456 1.242 0.21435 | | | | | | | |
| BAS\_bruto\_jaarinkomen4 0.24004 0.17824 1.347 0.17807 | | | | | | | |
| BAS\_bruto\_jaarinkomen5 0.54708 0.17027 3.213 0.00131 \*\* | | | | | | | |
| BAS\_bruto\_jaarinkomen6 0.69533 0.18932 3.673 0.00024 \*\*\* | | | | | | | |
| BAS\_bruto\_jaarinkomen7 0.64178 0.20190 3.179 0.00148 \*\* | | | | | | | |
| --- |  |  |  |  |  |  |  |
| Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1 | | | | | | | |

Model 3:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Coefficients: | |  |  |  |  |  |  |
| Estimate Std. Error z value Pr(>|z|) | | | | | | | |
| (Intercept) -3.5594 0.1956 -18.195 < 2e-16 \*\*\* | | | | | | | |
| ICFIC 1.7375 0.5088 3.415 0.000639 \*\*\* | | | | | | | |
| IClagFIC 1.5365 0.6465 2.377 0.017472 \* | | | | | | | |
| JourneyLengthNormal -1.7485 0.3363 -5.199 2.00e-07 \*\*\* | | | | | | | |
| JourneyLengthShort -3.0014 0.5281 -5.683 1.32e-08 \*\*\* | | | | | | | |
| GenderID2 -0.8450 0.3081 -2.743 0.006091 \*\* | | | | | | | |
| --- |  |  |  |  |  |  |  |
| Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1 | | | | | | | |

# Literature

Africa, N. (2010). *Measuring the controllable variables in the customer experience in convenience stores at filling stations.*

Anderl, E., Schumann, J. H., & Kunz, W. (2015). Helping Firms Reduce Complexity in Multichannel Online Data: A New Taxonomy-Based Approach for Customer Journeys. *Journal of Retailing*.

Animesh, A., Ramachandran, V., & Viswanathan, S. (2010). Quality Uncertainty and the Performance of Online Sponsored Search . *Information Systems Research*, 190-201.

Arbatskaya, M. N. (2007). Ordered search. *Journal of Economies*, 119-127.

Assari, S. (2014). Synergistic Effects of Lifetime Psychiatric Disorders on Suicidal Ideation Among Blacks in the USA. *Journal of Racial and Ethnic Health Disparities*, 275-282.

Bergemann, D., & Bonatti, A. (2010). *Targeting in advertising markets: implications for offline versus online media.* London: Centre for Economic Policy Research.

Berman, B., & Evans, J. (1998). *Retail management: a strategic approach.* NJ: Englewood Cliffs.

Blattberg, R. C., Kim, B. D., & Neslin, S. A. (2008). *Database Marketing: Analyzing and Managing Customers.* Heidelberg, Germany: Springer.

Bowman, D., & Naryandas, D. (2001). Managing Customer-Initiated Contacts with Manufacturers: The Impact on Share of Category Requirements and Word-of-Mouth Behavior. *Journal of Marketing Research*, 281-297.

Calder, B. J., Malthouse, E. C., & Schaedel, U. (2009). An Experimental Study of the Relationship between Online Engagement and Advertising Effectiveness. *Journal of Interactive Marketing*, 321-331.

Fox, J., & Monette, G. (1992). Generalized Collinearity Diagnostics. *Journal of the American Statistical Association*, 178-183.

Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression.* Sage Publications.

Frambach, R. T., Roest, H. C., & Krishnan, T. V. (2007). The Impact of Consumer Internet Experience on Channel Preference and Usage Intentions across the Different Stages of the Buying Process. *Journal of Interactive Marketing*, 26-41.

Fulgoni, G. M., & Morn, M. P. (2009). Whither the click? How online advertising works. *Journal of Advertising Research*, 134-142.

Ghose, A., & Yang, S. (2009). An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets. *Journal of Marketing Science*, 1605-1622.

Haan, E. d., Wiesel, T., & Pauwels, K. (2016). The Effectiveness of Different Forms of Online Advertising for Purchase Conversion in a Multiple-Channel Attribution Framework. *International Journal of Research in Marketing*, 491-507.

Hindle, T. (2008). *The Economist Guide to Management Ideas and Gurus.* John Wiley & Sons.

Hosmer, D. j., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression. Third Edition.* New Jersey: John Wiley and Sons.

Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukopadhyay, T., & Scherlis, W. (1998). Internet paradox: A social technology that reduces social involvement and psychological well-being? *American Psychologist*, 1017-1031.

LaSalle, D., & Britton, T. A. (2003). Priceless. Turning Ordinary Products into Extraordinary Experiences. *Harvard Business School Press*.

Leeflang, P. S. (2015). *Modeling Markets: Analyzing Marketing Phenomena and Improving Marketing Decision Making.* New York: Springer.

Lemon, K. N., & Verhoef, P. C. (2016, November 1). Understanding customer experience throughout the customer journey. *Journal of Marketing*, pp. 69-96.

Liao, T. (1994). *Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models.* Thousand Oaks, CA: Sage Publications, Inc.

Manchanda, P., Dube, J., Goh, K., & Chintagunta, P. (2002). The Effect of Banner Advertising on Internet Purchasing. *Journal of Marketing Research*, 98-108.

Marquez, J. J., Downey, A., & Clement, R. (2015). Walking a Mile in the User's Shoes: Customer Journey Mapping as a Method to Understanding the User Experience. *Internet Reference Services Quarterly*, 135-150.

Marquez, J., & Downey, A. (2015). Service Design: An Introduction to a Holistic Assessment Methodology of Library Services. *Journal of Library User Experience*.

Montgomery, A. L., li, S., Srinivasan, K., & Liechty, J. C. (2004). Modeling Online Browsing and Path Analysis Using Clickstream Data. *Journal of Marketing Science*, 579-595.

Naik, P. A. (2007). Integrated Marketing Communications. In G. J. Tellis, & T. Ambler, *The Sage Handbook of Advertising* (pp. 35-53). London: Sage Publications.

Naik, P. A., & Peters, K. (2009). A Hierarchical Marketing Communications Model of Offline and Online Medium Synergies. *Journal of Interactive Marketing*, 288-299.

Naik, P. A., & Raman, K. R. (2003). Understanding the Impact of Synergy in Multimedia Communications. *Journal of Marketing Research*, 375-388.

Peterson, R. A., & Merino, M. C. (2003). Consumer Information Search Behavior and the Internet. *Journal of Psychology and Marketing*, 99-121.

Raman, K. R., & Naik, P. A. (2004). Long-term Profit Impact of Integrated Marketing Communication Programs. *Review of Marketing Science*.

Risdon, C. (2011, November 11). *The Anatomy of an Experience Map | Adaptive Path*. Retrieved from http://www.adaptivepath.org/ideas/the-anatomy-of-an-experiencemap/

Rutz, O., & Bucklin, R. (2008). *From generic to branded: A model of spillover dynamics in paid search advertising.* New Haven, CT: Yale School of Management.

Serner, A., & Herschel, G. (2008). *A Checklist for Evaluating an Inbound and Outbound Multichannel Campaign Management Application.* Stamford, CT.: Gartner.

Shankar, V., & Malthouse, E. C. (2007). The growth of interactions and dialogs in interactive marketing. *Journal of Interactive Marketing*, 2-4.

Soba, M., & Aydin, E. (2012). The Role of Income Level on Sensitivity Levels for Similar Product: A Purchasing Behavior Study . *International Journal of Humanities and Social Science* , 177-181.

Stefanie. (2017, October 15). *Standardized Residuals in Statistics: What are They?* Retrieved from Statistics How To: http://www.statisticshowto.com/what-is-a-standardized-residuals/

Tutaj, K., & Reijmersdal, E. A. (2012). Effects of Online Advertising Format and Persuasion Knowledge on Audience Reactions. *Journal of Marketing Communications*, 5-18.

Verhoef, P. C., Kannan, P. K., & Inman, J. J. (2015, Februari). From Multi-Channel Retailing to Omni-Channel Retailing. *Journal of Retailing*, pp. 174-181.

Verhoef, P. C., Kooge, E., & Walk, N. (2016). *Creating Value with Big Data Analytics: Making Smarter Marketing Decisions.* London: Taylor & Francis Group.

Vrotsou, K. (2010). *Everyday Mining: Exploring sequences in event-based data.* Norrkoping: Linkoping University.

Wiesel, T., Pauwels, K., & Arts, J. (2011). Marketing's Profit Impact: Quantifying Online and Off-line Funnel Progression. *Marketing Science*, 604-611.

Yao, S., & Mela, C. F. (2011). A Dynamic Model of Sponsored Search Advertising. *Journal of Marketing Science*, 447-468.

Yaveroglu, I., & Donthu, N. (2008). Advertising Repetition and Placement Issues in Online Environments. *Journal of Advertising*, 31-44.