Erasmus School of Economics

Thesis

To obtain the academic degree of

Master of Science in Economics & Business

(Major in Marketing)

Exploring the impact of social influences on consumers' shopping behavior in crowded areas

Author: Marina Georgieva
Email address: 414358mg@eur.nl
Supervisor: Nuno Camacho
Study Program: Business Economics
Specialization: Marketing
Date: 23/08/2016

Table of Contents

Abstract

Store crowding is found to be a double-edged sword. On the one hand, seeing a lot of people is attracting you to the store, on the other hand, the presence of many other people poses different obstacles in achieving a purchasing goal, resulting in decreasing the shoppers' purchase likelihood. In a store, we are constantly under social influencers, including our companion, other people, salespersons, etc. Moreover, we can go to a store with a particular goal in mind or just to browse a bit. This thesis examines whether the social influences and shopping motives would help reduce the negative impact of store crowding on consumers' purchase incidence. Using high-quality behavioral data, this relationship is examined in the 'real-world'. The context involves four store located in an airport in Portugal. The results, however indicate that store density does not have any impact on the purchasing decision, neither do the social influences impact the purchase incidence in crowded areas. Discussed are future research directions for improvements of the data collection and implications.

Introduction

Social influences are found throughout every daily interaction. In a retail environment, consumers face many opportunities for social interaction such as shopping companion, salespersons interaction, and interaction with other shoppers or just the presence of other shoppers. Hence, at the same point of time consumers can be involved in multiple interactions while they are in the shop (Zhang, Li, Burke, & Leykin, 2014). These social influences are important determinants of consumer behavior in retail as they are found to have an effect on time spent in the store, satisfaction with the store, spending behavior and purchase deferral (Kurt, Inman, & Argo, 2011; Eroglu, Machleit, & Barr, 2005; Harrell, Hutt, & Anderson, 1980).

The current research investigates the joint impact of two social influencers, namely (i) shopping with a companion and (ii) "store crowding", on consumers' in-store decision making. Additionally, it examines the interaction effect of these social influencers with certain attitudes such as goal-directness.

In this paper, I rely on Latané's Social Impact Theory (referred to as SIT hereafter) which suggests that people's behavior, in different daily situations, is often guided by the "social impact" of others on our own behavior (Latané, 1981). SIT rests on three principles. The first principle postulates that the social impact is a multiplicative function of three forces: *size*, *proximity* and *social strength*. If the number of influencers (*size*) is higher, influencers are spatially and/or temporally *proximal*, or the relationship or bond between two parties is closer (e.g. family, friends), then the influence of one party over the other will be greater. The second principle in SIT (a psychosocial law) refers to the statement that "the first other person in a social force field should have greater impact than the hundredth" (Latane, 1981, p.344). That is, the impact will be highest in the difference between 0 and 1, and less as the number of sources increases. The third principle in SIT states that when the individual is not alone, with the increment of strength, immediacy or number of others, the social impact will be divided amongst all the targets. In other words, when other people are present, each individual will have less influence. For example, people are more likely to be influenced and listen to a speaker when in a small group than when in a larger group (Latane, 1981).

The present study relies on SIT's major principles and tests behavioral hypothesis derived from the principles underlying the theory.

The first social influencer investigated here is with regard to shopping with a companion. Researchers have shown that shopping with a friend or family influences product and brand decisions, the amount spent while shopping and is triggering more impulsive purchases (Childers

& Rao, 1992; Luo, 2005; Kurt et al., 2011; Hart & Dale, 2014). Moreover, shopping with a companion results in greater shopping value and an overall positive experience (Guido, 2006; Borges, Chebat, & Babin, 2010). Shopping with a companion can enhance the shopping experience by providing the opportunity to identify a need, ask for advice and/or expertise and enjoying a good company (Inman, Winer, & Ferraro, 2009).

The second social influencer, that is a subject of study of the present work, refers to "crowding". Stokols (1972) suggests that perceived crowding occurs when an individual feels certain constraints of limited space. In other words, there is a discrepancy between one's perceived space and the actual space available. Retail crowding has been studied for nearly thirty years. In 1976, Harrell and Hutt proposed an initial framework for the phenomenon of "perceived crowding" in a retail environment. The authors were motivated by recent advances in the American retail market in which superstores and shopping centers were quickly gaining popularity. Their model described that personal factors such as past experience, impatience, aggressiveness and time awareness determine the degree of perceived crowding per individual. Based on this perceived crowding, a consumer exploits different adaptation strategies which have a direct impact on several retail outcomes such as satisfaction with the store and the purchase, perceived store image with regard to price, quality, and assortment, etc.

Since the initial framework (Harrell & Hutt, 1976) was presented, many studies have been conducted exploring the concept of retail crowding. A stream of research explored the relationship of perceived crowding and different retail outcomes such as customer satisfaction, store atmosphere, and spending (Eroglu et al., 2005; Eroglu & Macheleit, 1990; Machleit, Eroglu, & Mantel, 2000; Harrell et al., 1980). Other studies have focused on examining the mediating and moderating effects of culture, adaptation and emotions on the retail outcome (Michon, Chebat, & Turley, 2005; Turley & Milliman, 2000; Pons & Laroche, 2007). A third flow of researchers focused on the antecedents of retail crowding, such as individual characteristics related to perceived risk, perceived control, shopping motives and time pressure (Van Rompay, Krooshoop, Verhoeven, & Pruyn, 2012; M. Hui & Bateson, 1991).

Generally, it has been found that although people are attracted by crowded areas, once in that area they are less likely to buy a product (Sam K. Hui, Bradlow, & Fader, 2009; Becker, 1991; Zhang et al., 2014; Harrell et al., 1980). Reasons for the negative impact of crowding on the purchase decision have been drawn to different adaptation strategies that consumers apply to cope with the psychological stress and physical barriers they encounter while shopping (e.g. delay the purchase; Harrell et al., 1980). Zhang et al., 2014 conclude that retailers should be careful when creating the store layout and design, and the overall store ambience so that they attract customers to crowded areas while at the same time provide them with the necessary resources (e.g. aisle width, fitting rooms, cashiers) so that the journey throughout the store is not prematurely terminated.

According to the third principle of the SIT, with the increase of number of people, the strength or immediacy in the social group, the influence of external and internal social forces over the group and the individual will decrease. Thus, the negative impact of crowding will be reduced with the increase of the social size of the target (the consumer) as it is divided among all of the targets (the consumer and his/hers companions). Research in other domains also suggests that a group can help reduce the negative impact of crowding as highly dense areas are perceived less crowded when groups are present (Baum, Harpin, & Valins, 1975). Groups also possess the power to become unaffected by the external world (Willis, 2008). Stokols (1972) argues that "*while the amount of space in a given area may appear limited to an outside observer, it will not inevitably seem inadequate to the occupants of the area, especially if their activities do not require a high degree of behavioral coordination, if their relationships with each other are cooperative and friendly, or if they have had much experience with living and working under conditions of limited space*". This suggests that shopping with a companion should reduce the negative impact of crowding on consumers' shopping behavior. Such information is valuable for the retailer as it will allow the creation of optimized marketing and promotional campaigns as well as designing better targeting approaches.

Despite the potentially important effect of shopping with a friend on consumers' purchase decisions under conditions of crowding, such effect has not been thoroughly examined in the literature. To this day, and to my best knowledge, only one study has empirically investigated the interplay between different social influencers in a retail environment. Namely, Zhang et al. (2014) explored the interaction between several social elements that could occur while shopping using video tracking data. Contrary to what psychological research and SIT imply, examining the "*influence of crowding, shopping group size, customer interaction, and salesperson interaction on product touch and purchase, controlling for in-store marketing activities, shopping path, and other environmental factors*", the authors find that crowding has a negative impact on the purchase likelihood and the larger the group shopping together group the less likely for them to buy when the store was crowded.

Therefore, an objective of this thesis is to investigate whether shopping with companion(s) - as opposed to being alone - under conditions of crowding, minimizes the negative effect of crowding and thus enhances the shopping trip resulting in positive retail outcomes such as purchase incidence as suggested by SIT and research in psychology. Moreover, exploring attitudinal variables such as goal-directness, the study will assess whether shopping with a friend has a stronger effect on purchase likelihood in dense retail areas than other variables.

Specifically, a “goal-directed” mindset implies that the consumer has a clear goal and uses certain heuristics to achieve it (Janiszewski, 1998). This behavior is also deemed as “utilitarian” as it is “ergic, task-related and rational” (Babin, Darden, & Griffin, 1994). On the other hand, the shopping experience can be exploratory, experiential or driven by certain hedonic benefits for the consumer (Babin et al., 1994; S. K. Hui, Fader, & Bradlow, 2009; Novak, Hoffman, & Duhackek, 2003). Research has shown that in dense areas the feeling of control is reduced, thus achieving a certain goal is obstructed, leading to negative experiences for the consumer and negative outcomes for the retailer (Ward & Barnes, 2001; M. Hui & Bateson, 1991). The paper will further explore how the shopping motive of the consumer (e.g. goal-directed or browsing) affects purchase decision under conditions of crowding. By adding this variable to the model, the study will provide a more comprehensive overview of what drivers consumer behavior in dense retail areas.

Based on the above considerations, the objectives of the current paper are summarised as follows:

(1) Validate whether crowding has negative influence on the purchase incidence;

(2) Explore whether shopping with a friend reduces the negative impact of crowding on the likelihood to purchase;

(3) Explore the effect of attitudinal variables, e.g. goal-directness, would be on the purchase incidence.

By means of high-quality behavioral passive data, the present research provides a more comprehensive view of the shopping behavior under conditions of crowding when with a companion. In other words, this thesis offers two key contributions to the marketing literature. First, this study analyzes the interaction between two distinct social influencers on the shopping trip, namely shopping with a friend and the presence of other shoppers, resulting in shopping in a dense area. To my best knowledge, only one past study (Zhang et al., 2014) explores the relationship between the above-mentioned influencers on consumers’ purchase behavior. Thus, the current study contributes to reducing the scarcity of the literature streams in this direction. The latter is also one of the few studies that utilizes video tracking data rather than relying on self-reported data. Another exception of using high quality behavioral data is the study conducted by Hui (2009), in which ‘behavioral hypothesis’ regarding consumers shopping behavioral are explored. Using high-quality behavioral data, this research can ensure high external validity of the tested hypothesis.

Second, this thesis contributes to the behavioral modeling literature, i.e. to the efforts of marketing scientists to test behavioral hypothesis using real-world secondary data with revealed consumer behavior, in this case in a retail context. Previous studies on crowding have used mostly self-reported data, which can be prone to biases as respondents can misinterpret their answers due do different motives (e.g. social desirability bias) (Manski, 2004). Here, the external validity of the previous findings using unbiased, complete and accurate data showing consumers' paths along their journey in a store has been tested.

In addition, this paper offers new managerial insights for retailers concerned with store crowding and its influences on consumers' store choice and in-store purchase decisions. In particular, retailers should be concerned with the fact that consumers may apply different adaptation strategies leading to non-desired actions (e.g. purchase deferral or delay) and crowding can be a significant determinant when evaluating a certain retailer (Herrington & Capella, 2015). The present thesis contributes toward retailers' understanding of crowding and offers three different directions that may contribute to improve retail management under crowding conditions. First, unraveling consumer shopping behavior in groups would allow retailers to execute proper promotional and advertising campaigns, even in situations where the store is crowded. Second, understanding the dynamics of crowding, owners can properly design the store layout resulting in limiting perceived feeling of human crowding while still attracting consumers when it looks a bit crowded. Third, recognizing the main drivers for conducting a purchase in a dense area (e.g. shopping motive and shopping group) can help retailers optimize their offerings and marketing campaigns, and adapt customer messages accordingly.

The thesis is structured in five chapters: 1) introduction; 2) theoretical background and hypothesis development; 3) research methodology; 4) data analysis and results; and 5) conclusions and directions for future research. The literature review consists of an overview of the three major theoretical and research streams used to derive the main hypothesis, namely shopping with a companion, crowding, and goal-directness. The econometric model which forms the basis of the present study, the input data needed for this model and the derivation of variables used for the purposes of the current analysis are described in Chapter 3. Chapter 4 consists of a detailed analysis of the main results and findings obtained from the performed investigations in order to answer the posed research questions. Chapter 5 presents the conclusions from the current studies and discusses the implications for retailers. In addition, suggestions for future research and limitations are presented in this chapter.

Theoretical background

Conceptual framework

In this thesis I adopt the *Theory of Social Impact* (Latane, 1981) and its main principles as the overarching theory I will use to develop all my hypotheses. I now discuss the tenets of Social Impact Theory (SIT) applied to the context of my thesis: retail and shopping behavior. First, I discuss shopping with a friend and its impact on consumer behavior and retail outcomes. Second, I review the concept of retail crowding. Third, with the aid of SIT and existing literature, I develop testable hypohteses regarding group shopping in conditions of crowding. Fourth, I discuss the shopping motives and what would their relationship look like under conditions of retail crowding.

The conceptual framework presented in Figure 1, represents the relationships and the drivers of purchase behavior in a crowded area, namely shopping with friend(s) and goal-directness, while controlling for culture, shopping time and path (note: control variables not shown in Figure 1). The goal of this research is to understand the factors that affect consumer behavior in a dense area, leading to a purchase decision. More precisely, the paper will examine which factors could minimize the negative effect of retail crowding so that the purchase decision is executed.

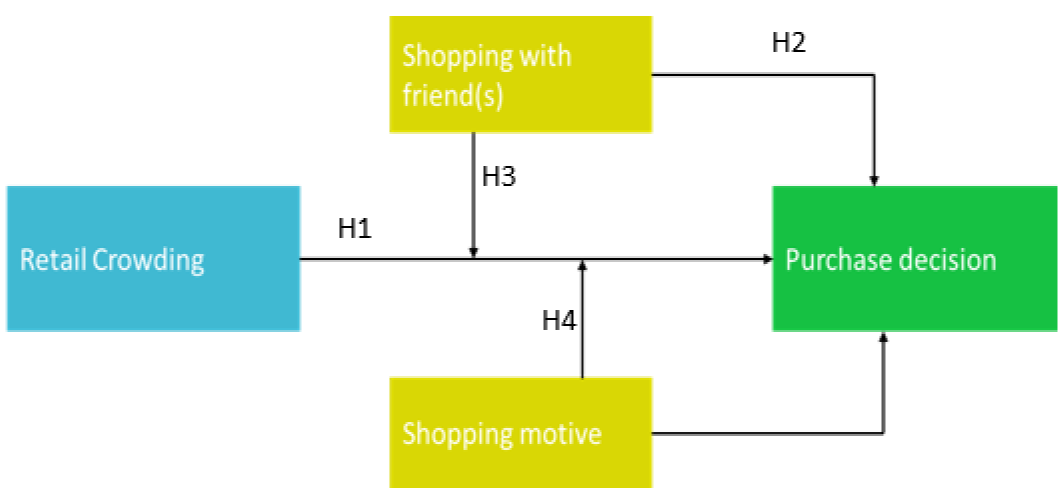

Figure 1, Conceptual Framework

Retail Crowding and shopping behavior

Kotler (1973) argues that atmospherics is an important marketing tool that can have a significant impact on consumers' purchase decision. In fact, "in some cases, the place, more specifically the atmosphere of the place, is more influential than the product itself in the purchase decision." Crowding as induced by density can be considered an important atmospheric tool with a notable effect on consumer behavior (Harrell et al., 1980).

Defining Crowding

In line with the academic literature in marketing, I define retail crowding from both a physical and a psychological perspective. Feelings of crowding are detected when the personal space of an individual is restricted (Stokols, 1972). Such feelings typically occur when the overall space available is limited. This definition implies that crowding is more an experiential state due to limited space rather than simply being in a dense area. Retail density, the number of consumers in a given space, is thus, a required condition for an individual to experience crowding, but not sufficient. Additional factors such as personal characteristics, environmental and social elements influence consumers' retail crowding perceptions (Eroglu et al., 2005). Research has shown that generally higher levels of retail density increase the perceptions of retail crowding (Eroglu & Macheleit, 1990). Crowding has been measured as both function of the number of people in a given space (density) and as an assessment of the "psychological state of the individual" (perceived crowding) (Harrell et al., 1980).

Retail crowding can further be divided into two dimensions: *human crowding* and *spatial crowding* (Karen a. Machleit, Kellaris, & Eroglu, 1994). *Human crowding* refers to feelings of crowding based on the number of people in a closed area. *Spatial crowding*, on the other hand is based on feelings of crowding due to the amount of merchandise, aisles, and their configuration within the store. Such nonhuman elements and their arrangement within the retail area could induce feelings of spatial crowding. For the purpose of this paper, spatial crowding won't be observed, however, it should be noted that it is an essential component when constructing the store layout which could ultimately lead to human crowding.

Framework

In 1976, Harrell and Hutt described an "initial framework" for buyer behavior under conditions of retail crowding. The model created was used as a benchmark for future research as it incorporates both antecedents and consequences for shopping in a crowded area. The authors describe a framework that utilizes several adaptation strategies based on consumers' personal characteristics. The different adaptation strategies determine the possible retail outcomes. *Error! Reference source not found.* depicts the framework. Serving as a benchmark for future research, Harrell and Hutt’s framework defined future research streams on retail crowding in three directions: 1) determining the antecedents; 2) observing the outcomes on shopping behavior and 3) exploring different mediating and moderating variables. Research on crowding extended the framework by accounting for additional personal factors, adaptation strategies and examining different retail outcomes.

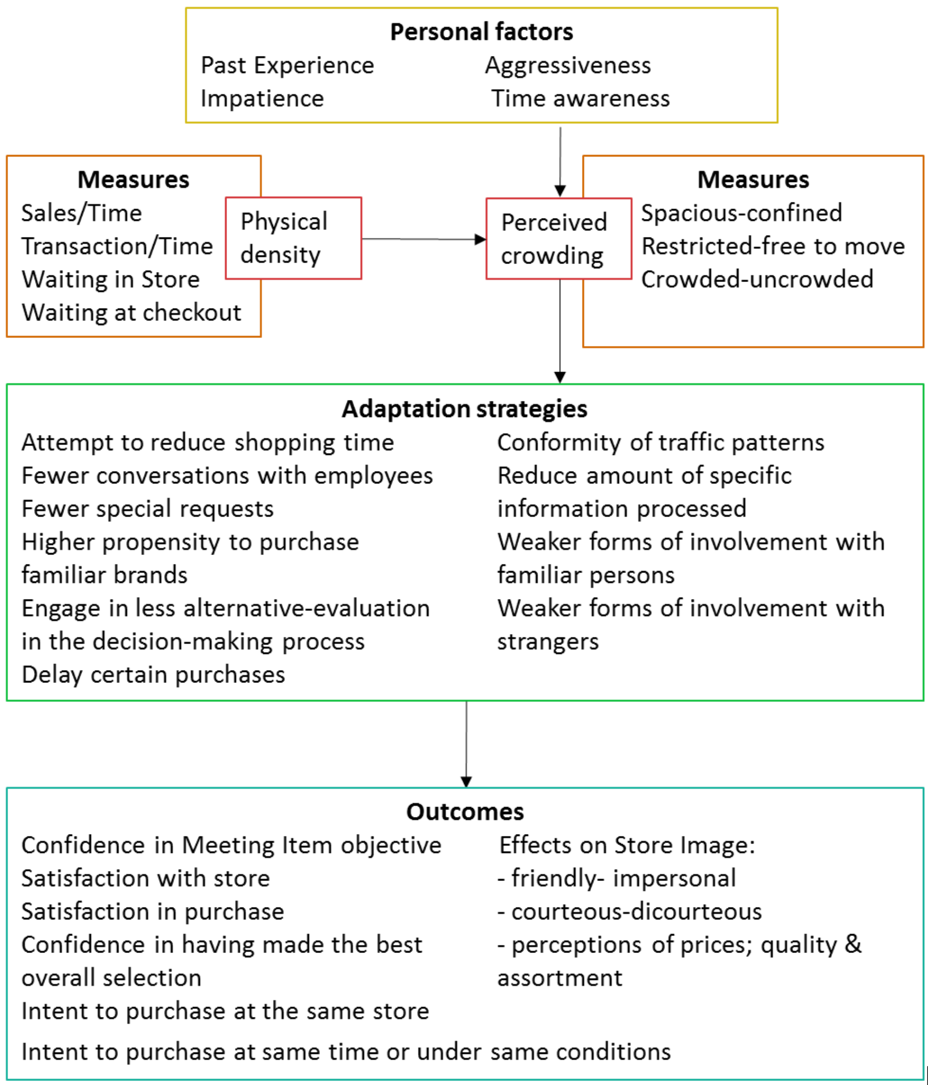

Figure 2, Initial Framework Crowding (Harrell and Hutt, 1976)

Influence of retail crowding on consumers’ purchase likelihood

Marketing scholars have explored different streams of research in order to understand the impact of retail crowding on consumers’ shopping behavior. Researchers have studied different *antecedents* of retail crowding including shopping motives, perceived control, perceived risk

(Mehta, 2013; Eroglu & Macheleit, 1990; Van Rompay, Krooshoop, Verhoeven, & Pruyn, 2012; Dion, 2004). For example, goal-directed consumers would feel more restrained in their attempt to conduct a purchase, which would negatively influence the retailer (Eroglu & Macheleit, 1990). When constrained by perceived risk and time pressure, goal-directed shoppers will experience elevated levels of crowding. Dion (2004) has shown to what extent personal control influences the crowding perceptions. His findings suggest that people with low levels of control will perceive high retail crowding.

Another stream of research involves studying the *process* through which retail crowding affects shopping behavior, i.e. different mediating and moderating variables that could affect the perception of human crowding. Machleit et al. (2000) proposed emotions as a mediator between the relationship of crowding and customer satisfaction. Their results indicated that emotions partially moderate the relationship along with personal expectations and tolerance for crowding. Eroglu et al. (2005) implemented emotions when examining the influence of crowding on consumer satisfaction and interestingly enough found that accounting for emotions under crowding has a positive effect on shopping satisfaction. Furthermore, Eroglu et al. (2005) examined the impact of human and spatial crowding on consumers' shopping value (e.g. hedonic and utilitarian) revealing that hedonic shopping mode is negatively influenced by spatial crowding. Zhang et al. (2014) found that salesperson's effectiveness is decreased in conditions of crowding while Mattila & Wirtz (2008) indicated that salespersons' assistance reduces the negative impact of crowding. Moreover, culture was found to significantly affect the levels of crowding (Pons & Laroche, 2007). For instance, Canadian consumers perceived a dense area as more crowded compared to Mexican people. Male consumers have less tolerance to crowded areas as compared to women (Eroglu & Macheleit, 1990).

Researchers have also been interested in determining the *outcomes* of crowding for retailers. More specifically, studies has been conducted examining customer satisfaction, evaluations, purchasing and spending behavior (Kim & Kim, 2012; Argo, Dahl, & Manchanda, 2005; Eroglu & Macheleit, 1990; Eroglu et al., 2005; Zhang et al., 2014; Maeng, Tanner, & Soman, 2013). The results suggest that customer satisfaction can be positively and negatively influenced by crowding depending on different environmental and personal factors. For instance, Machleit, Eroglu, & Mantel (2000) found that the higher the personal tolerance for crowding the higher was the shopper satisfaction. Moreover, Li, Kim, & Lee (2009) found that positive feelings can be exerted in crowding areas depending on the store, e.g. in discount shops pleasure and dominance were experienced in crowded areas. On the other hand, when consumers are time constrained in a crowded area their satisfaction with the retailer decreases (Eroglu & Macheleit, 1990). Maeng et al. (2013) demonstrated when consumers are affected by high levels of density, they are more likely to avoid such areas and opt for a more safety-oriented place.

Generally, is has been demonstrated that purchasing behavior is negatively influenced by retail crowding. Two distinct forces influencing consumer behavior in such an environment have been examined. On the one hand, a “herd behavior” is observed. People are attracted to crowded areas and are motivated to visit it as it might indicate that the products there are good offers (Hui, Fader, & Bradlow, 2009; Harrell et al., 1980; Becker, 1991; Zhang et al., 2014). Argo, Dahl, & Manchanda (2005) argue that the social presence of others in crowded areas creates feelings of “belongingness” which exerts positive emotions. Additionally, the presence of others in a crowded store might be a sign for high-quality products which increases the motivation to visit it (Cialdini & Goldstein, 2004). On the other hand, studies have shown that once in that area consumers are less likely to buy a product. Reasons for the negative impact of crowding on the purchase decision have been drawn to different adaptation strategies consumers apply to cope with the psychological stress and the physical barriers (e.g. delay the purchase) (Harrell et al., 1980; Zhang et al., 2014). Feeling that their space is restricted, consumers are more likely to reduce their shopping time and defer their purchases (S. K. Hui et al., 2009).

In sum, prior research supports the notion that purchase incidence is reduced when an individual perceives high levels of crowding. This may occur because crowding may “decrease of feelings of comfort and increase of negative emotions”, suggesting that the increase in the size of social presence would induce negative emotions in shoppers leading to negative outcomes for the retailers (Argo et al., 2005). Therefore, I hypothesize that:

H_1: Crowding decreases the likelihood of purchasing.

Social Influences in Crowded Stores: The Theory of Social Impact

The three main principles

The first principle of SIT infers that the social impact is a multiplicative function based on the number of people, strength and immediacy. Functionally, this is presented as follows:

SI= f(SZ, PX, ST),

Where:

SI is the social influence;

F is the social influence function which is a multiplicative function of size (SZ, proximity (PX) and strength (ST);

Thus, it can be said that when the number of people is higher (size), the social strength within two parties is high, e.g. when they have a closer relationship and the influencers are spatially and/or temporally proximal (immediacy), the influences of one over the other would be greater.

SIT's second principle, *the psychosocial law*, implies that the presence of only one person will have the most impact on a certain individual. Increasing the group in size will have relatively less impact. Mathematically this is expressed in the following way:

$SI = sSZ^t$, t<1

Meaning that the social impact, *SI*, is a function of some power, *t* , of the number of people, *N* multiplied by a scaling constant, *s*. The value of *t* should be less than one, meaning that "impact will increase in proportion to some root of the number of people present" (Latane, p.344, 1981). In other words, increasing the number of people will lead to a small increase of the impact they have on an individual.

In a retail context, the second principle suggests that when a person is shopping with friends, the most influence will be exerted from that one friend. Increasing the group of friends will still have a larger impact, but to a lesser extent.

SIT's third principle, named "multiplication versus division of impact" defines a situation in which as the individual is accompanied by other people (e.g. is not alone), the influencers coming from outside this group will be divided among the members of the group. I now rely on SIT's principles to develop hypotheses for the social influences in retail.

SIT's principles in retail

In a retail context, research has been scarce when it comes to exploring the interaction between crowds and groups. The only notable study is this of Zhang et al. (2014), where the authors explore the interplay of different social influencers in a retail environment using video tracking data on the likelihood to buy a product and the touch frequency. The study investigates how crowding, groups, salesperson contacts and discussions within groups have an impact on one's propensity to commit a purchase and engage in touching the product. More specifically, the study explores whether shopping groups would lead to increased purchase likelihood as compared to shopping alone under conditions of crowding. The study also investigated the role of discussions within the shopping groups and their impact on purchase behavior. It confirms the hypothesis that crowding lowers the propensity to commit a purchase. Interestingly enough, the results, contrary to SIT and previous research in other domains, indicate that larger groups are less likely to engage in touching the products and buy them when the store was crowded. These results suggest that crowding evokes negative feelings regardless of whether one shops alone or not, showing that group size does not moderate the relationship between crowding and purchase intent.

Shopping with a friend: An SIT perspective

It happens that frequently the decision-making regarding a certain choice is done within groups (Aribarg, Arora, & Kang, 2010). Research has shown that this choice deviates from the choice an individual makes when alone (Corfman, 1991; Ariely & Levav, 2000). The deviation can be caused as individuals, in group settings, tend to engage in “impression management” in an attempt to present themselves in a certain positive way to others (Ariely & Levav, 2000). For example, back in the 1960s, Stafford (1966) conducted an exploratory study aiming to identify whether informal social groups have an influence on the brand preferences of their members. The results suggested that consumers are in fact influenced by their social group and their brand loyalty is most influenced by the informal leader of the group. Since then many studies and theories were executed attempting to explain the direction and strength of social influences in the consumer decision-making process.

Research has shown that when shopping with a companion, individuals are more likely to engage in impulsive purchases, spend more money and time, feel more confident regarding their purchase and thus, reduce the perceived risk of making the purchase decision alone (Bell, Corsten, & Knox, 2011; Sommer, Wynes, & Brinkley, 1992; Hart & Dale, 2014). For example, Luo (2005) found that when consumers picture themselves shopping with friends they are more likely to engage in impulse purchases, being with a family, however, decreases their spending. Mangleburg, Doney, & Bristo (2004) have also found that teens’ shopping together with friends are more likely to spend more than when they are alone. Studies suggest that whether the influence of shopping with a friend is positive or negative depends on gender (Kurt et al., 2011; Hart & Dale, 2014). That is, males tend to spend significantly more while shopping with a peer than when being alone while for females the opposite was true.

Moreover, shopping with a friend has a positive effect on the overall shopping experience (e.g. Guido, 2006). Shopping together can offer consumers more “social and informational value” to the shopping trip (Hart & Dale, 2014). Consumers benefit from their companion as they are given the opportunity to ask for advice, expertise and generally to enjoy a good company. Within-group discussion have also been found to have a positive influence on the purchase likelihood (Zhang et al., 2014) as it could help consumers identify a need (Inman et al., 2009) and reduce the perceived risk of the purchase (Willis, 2008).

According to the first principle of SIT, when the social strength within two parties is high, e.g. when they have a closer relationship, the influences of one over the other would be greater. Given this and results from previous research, which indicate that shopping with a friend enhances the shopping trip, first I will investigate whether there is generally a positive relationship between shopping with someone and the purchase likelihood. Thus, my first hypothesis is as follows:

H_2: Shopping with a friend increases the probability of committing a purchase rather than being alone.

The Moderating Effect of Shopping with a friend on the Effect of Retail Crowding on Purchase Likelihood

Please recall that SIT's third principle (the "multiplication versus division of impact" principle) suggests that when a consumer is not alone, social influence is diminished. That is, the impact of each social influencer coming from outside of the "core" group (the shopper and her or his friend(s)) will be reduced as the social impact is diluted among the members of this "core" group. *Figure 3* depicts such a situation[1].

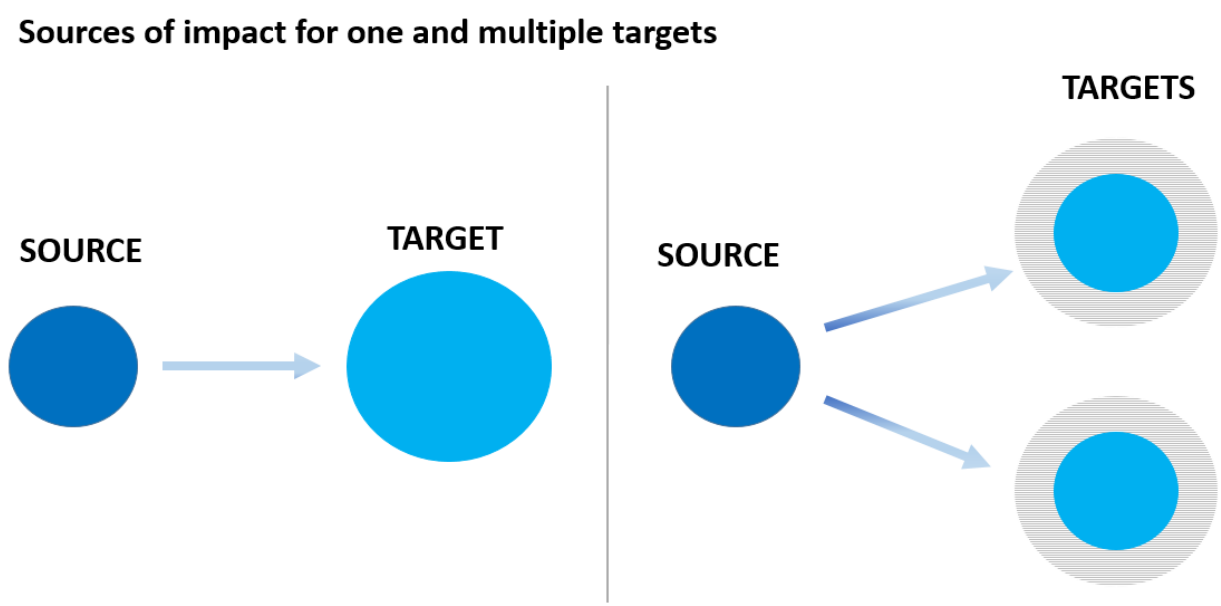

Figure 3, Sources of Impact

[1] Source: Latane (1981)

The figure represents the notion that "in such situations, increasing the strength, immediacy, or a number of other targets should lead to a division or reduction of impact, with each person feeling less than he or she would if alone" (Latane, 1981).

Extending SIT to a retail environment, this would imply that when consumers shops together with friend(s) the negative impact of crowding would be divided amongst all members and thus, be reduced as compared to the consumer being alone.

This notion is supported by Stokols (1972), who argues that feelings of crowding are evoked not only by the spatial restriction caused by high-density levels but also by his/hers relationship with the group.

Early research of urbanization and crowding confirms that view. For example, Baum, Harpin, & Valins (1975) found that groups help reduce the harmful effect of high density, and individuals in groups settings are less likely to experience crowding. Exploring the effects of groups in crowded residential areas, the authors found that cohesive groups reduce the psychological stress caused by crowded areas. In crowded pedestrian areas, the presence of a group is also found to be beneficial, as it can mitigate the negative impact of crowds and help "smooth" the pedestrian flow (Vizzari, Manenti, & Crociani, 2013). Maeng et al. (2013) found that when the group is composed of "in-group (vs. out-group)" members, negative feelings of crowding are reduced. Thus, groups are more likely to overcome external negative influences.

Hence, I hypothesize:

H_3: Shopping with friend(s) will reduce the negative impact of crowding on shoppers purchase.

The Moderating Effect of Shopping Motive on the Effect of Retail Crowding on Purchase Likelihood

When shopping, consumers often have different motives and goals. For example, shopping can be a purely utilitarian experience such that the consumer has purchased a product in a planned and efficient way (Babin et al., 1994). In that sense, such task-oriented shoppers have predetermined goal, spend less time shopping and generally are not influenced by contextual factors as long as they are not obstructed from achieving their goal (Korgaonkar, 1981). Consumers would exert high utility when the task is completed successfully. Hedonic shopping, on the other hand, does not need to be goal-directed. It involves exploration or simply "browsing". However, this does not mean that consumers do not intent to make a purchase

(Jarboe and McDaniel, 1987), rather it implies that consumers have abstract shopping goals and are engaging in a shopping trip for different personal and social motives. Such shoppers are found to be more responsive to contextual factors (Eroglu & Macheleit, 1990).

As mentioned above, it has been shown that the shopping motive (goal-directed vs browsing) affects the way consumers perceive dense areas (Eroglu & Macheleit, 1990). Results indicate that task-oriented shoppers have a low tolerance for crowded areas and perceive crowding greater than non-task oriented consumers resulting in low customer satisfaction. Other studies have also confirmed that task shopping increases the notion of perceived crowding (e.g. Baker & Wakefield (2012)). This behavior is due to reduced feelings of control which negatively corresponds with the shopping experience (Ward & Barnes, 2001; Baker & Wakefield, 2012). For example, Baker & Wakefield (2012) found that task-shoppers perceive density as crowding while social shoppers regard density in a positive way. Relying on existing literature, the last hypothesis is as follows:

H_4. Task-oriented shoppers are less likely to purchase a product in a crowded area as compared to browsing shoppers.

Data and Methodology

This section provides an overview of the data and the methodology used to test the hypotheses. In the first part I describe the data collection and cleaning processes, the variables derivation and summary statistics of them. The second part reports on the logistic regression used to analyze the data. Finally, described is the model that is applied to find evidence for the hypotheses described in Chapter II.

Data

Data description

This paper uses behavioral data collected from the period between 15.01.2012 to 29.01.2012 in the airport of Porto, Portugal. The data is collected by the Portuguese start-up Movvo, which specialized in path data collection. The company simply uses a set of antennae to track consumers' mobile devices in an enclosed area. Path-data includes information regarding shoppers' movement "in a spatial configuration" (Hui et al., 2009). Currently, there is rapid growth of behavioral data collection providers, supplying retailers, shopping malls and marketing professionals with location tracking data[2]. These new suppliers use diverse methods for configuring location tracking. The rapid growth in that domain is due to the development of technology which enable professionals to better analyze consumer behavior, investigate latent factors influencing the decision-making process and generally, observe how shoppers behave in the store (Hui et al., 2009). Using such data, researchers are able to observe consumer in-store dynamics allowing them to deeply understand human behavior without relying on self-reported data. In that sense, the paper will contribute to the usage of behavioral data by providing more complete and accurate insights on consumers' in-store behavior without relying on consumers' own memory.

The technology behind's Movvo solution integrates beacons (or antennas) that can be installed in a given store, which monitor customers' mobile devices by detecting radio frequency signals. A computer makes a real-time calculation of the distance of the device to the different base stations by using a trilateration algorithm. The real-time calculation naturally assumes real-time data collection and storage. The data is then represented through dashboards. In other words, radio frequencies emitted from mobile devices are transformed into data, which is then turned into real-time KPIs that can be viewed on dashboard. The process is visualized in *Figure 4.*

[2] https://www.quora.com/Location-Based-Services-LBS-Who-are-the-major-players-in-indoor-tracking-and-retail-analytics

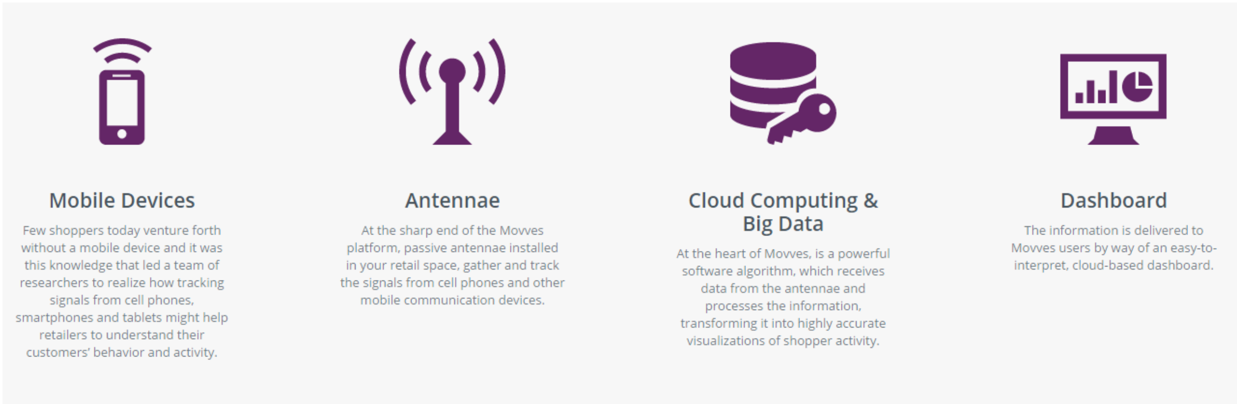

Figure 4, Path Data Collection

This solution provides quite precise information regarding one's location (1m^2) and time spent in it.

The analyzed dataset in this paper relies on airport data, collected in a two-week period as stated above. The airport in Porto, the third busiest airport in Portugal, accommodated roughly 6 million passengers in 2011/2012. Passengers are tracked after they have left the security zone and before leaving the duty free zone (see *Figure 5*

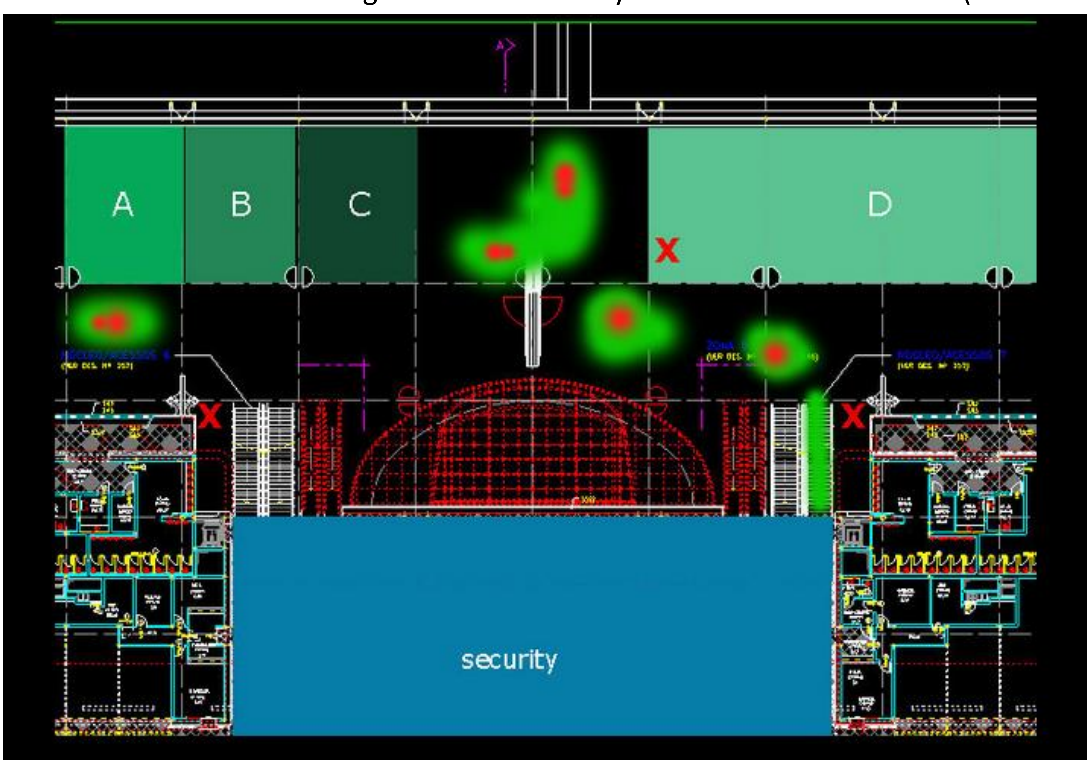

Figure 5, Airport layout).

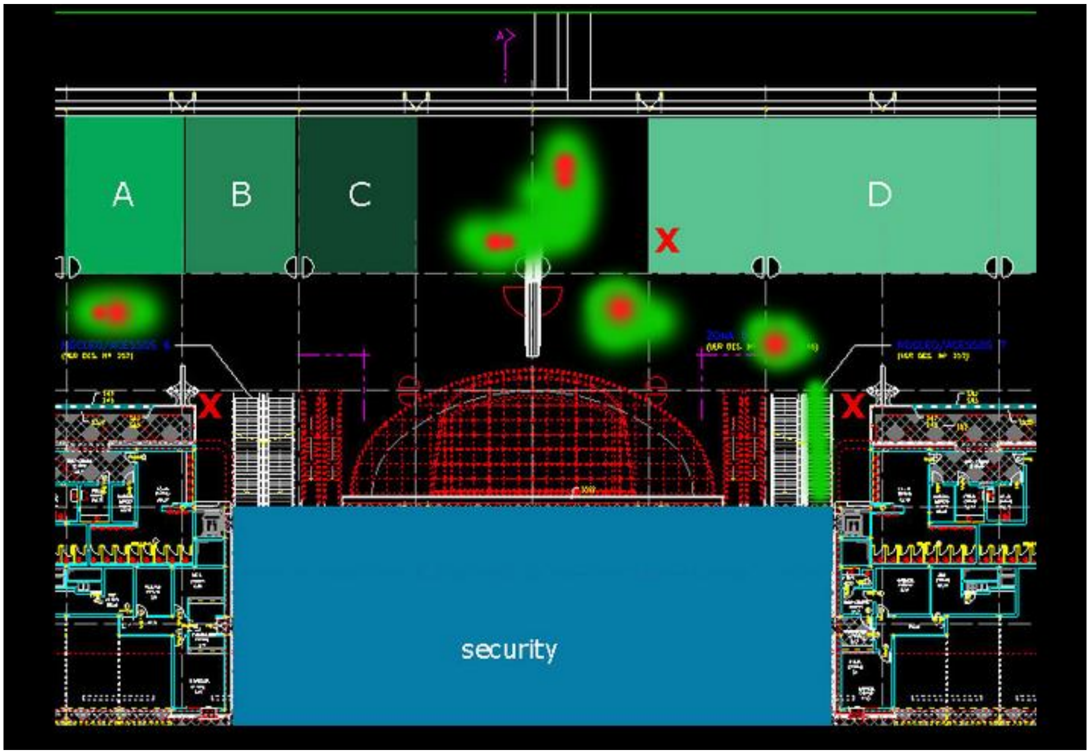

Figure 5, Airport layout

A, B, C and D indicate the 4 distinct shops in the airport[3], green areas represent frequently visited areas and the red areas are the ones visited the most. X are the antennae spread across the airport.

Data cleaning

The dataset consists of 45,980 distinct observations, each coming from a different person, thus, there are no returning visitors. The passengers are from 5 different countries, the Portuguese comprising of about two thirds of the all observations.

Since the data consists of a lot of observations, I investigate whether all of them are needed for the purpose of the analysis. First of all, it can be seen that 62% of the passengers did not enter any of the stores, making them redundant for this analysis. Thus, new subset was created removing these observations from the original dataset leaving **17,292** observations. Additionally, some other observations were removed from the analysis with respect to their buying behavior. It was seen that there were **4,470** observations who purchased an item spend less than 30 seconds in the store. Such observations are regarded as inconsistent and removed from the dataset as well. It could be noted that the visitors that spent less than 30 seconds in the stores

[3] Note: The number of stores and the retailers present in Porto's airport have, in the meantime, changed.

and made a purchase were mostly coming from stores A and B. These stores are more 'open' and it could have been the case that subjects stayed in the front of the store and having indeed made a purchase within 30 seconds. Thus, it might have happen that their signals were recorded a bit later when they were in front of the stores. As the reasons for this discrepancy are unclear, such respondents are ultimately removed from the analysis. Finally, for some subjects, information regarding their store path was missing. After the dataset was clean it consisted of **12,773** observations (each observation being a shopping path of a consumer, or passenger, on a specific date). Stores description after data cleaning can be seen in *Table 1*.

	Store A Sibarium	Store B El Corte Ingles	Store C Travellers-Luxury	Store D Taste of Portugal
Shop description	Regional products	Fashion clothing	Fashion &branded accessories	Utilitarian gifts and products such as chocolates, tobacco & liquor. Hedonic gifts such as cosmetics and perfumes
Visitors	2,594	2,595	4,573	10,795
Purchases	-	-	1,022	3,951
Time spend (minutes)	1	1	2.4	6.7
% shoppers	**0%**	**0%**	**22%**	**37%**

Table 1, Stores Summary Statistics

Table 1 provides an overview of the visitors and the purchases committed within each of the stores. It is visible that in store A and store B no purchases were made and the time spent per store is low, thus these two stores would not be considered and examined further. Store D is the most visited one with the highest average time and percentage of shoppers.

Variables and Descriptive Statistics

Variables' computation

For the purposes of the analysis, several additional variables are needed. Additional data processing is necessary in order to come up with the desired econometric model. The new variables include shopping with a companion, browsers or goal-directed shoppers and in-store density, each described below:

1) **Shopping with a companion.** The variable was constructed based on the time at which a consumer entered a store, the date, the dwell time of each store (if more than one) and the path. Using this information, the new variable looked whether someone's path had more than one occurrence across the sample. That is, if two people were to shop together, they'd probably

entered at the same time (accounting for a couple of seconds difference) in a same day, spend the same amount of time together in a shop and entered the same shops (e.g. visited Store D, C, B and A in that order). The biggest group of shoppers consisted of 4 people, yet 75% of the subjects shopped alone.

2) Another variable based on the **shopping motive** was constructed. It relied on whether the consumer visited only one or numerous shops. If a person has visited only one shop he/she was deemed as "goal-directed" (1), while visiting more shops indicated "browsing" behavior (0).

3) The **in-store density** at a given time per store also needed to be computed. As density was found to be the major component of creating feelings of crowding (Harrell et al., 1980) with a correlation of 0.58, perceived crowding will be measured as "density" which refers to the number of shoppers in an area at a given time. Additionally, in their study, Zhang et al. (2014) also measure crowding as a direct function of density in their models. Thus, crowding will be presented as in-store density in a given time.

The density was computed per store. The data contains information regarding one's dwell and entry time per store. This information was used to further compute subject's exit time allowing to create a time interval that shows how long a person was in a particular store on a particular day. Each subjects' time interval was compared to all the rest, allowing to observe how many people have overlapping intervals. The number of occurrences of the time intervals overlaps is used as a measure for density. Thus, in-store density is a continuous variable which accounts for the number of people inside the store in a given time range.

Descriptive statistics

Before proceeding with the analysis, frequencies across the different variables were computed as to allow for better overview of the data and what could be expected out of the analysis.

What we can see is that on average people shopping alone tend to commit more purchases than when with companion. With the increment of number of people shopping together there is a sharp decrease in purchase incidence (see *Table 2)*. This can already indicate that the hypothesis looking for specific effect as the group size increased might not be confirmed. The split between shop C and D is similar. Observed is also that the average dwell time in store C is far less than the average dwell time in store D independent on whether the shopper is alone or not (this is also in line with the total average described in *Table 1*). There is no decrease in dwell time when the number of shoppers increase, however, there is no increase either. That is, shopping with friends results in spending the same amount of time as when shopping alone.

When we look at *Table 3,* it is evident that no browsers committed a purchase in store C. As of this the analysis for store C would not include the shopping motive as an explanatory variable. Interestingly enough, the shoppers spend on average 30 seconds, while the browsers spent 10 minutes in the store without purchasing anything. The goal-directed shoppers in store D also prevail.

Not surprisingly, on average the density was highest for Store D (see *Table 4*).

Table 2, Shopping with companion

# of friends	# purchases Store C	# purchases Store D	Dwell Time Store C	Dwell Time Store D
None	73%	73%	2.4	6.8
1	20%	21%	2.3	6.8
2	6%	5%	2.3	6.9
3	1%	1%	2.6	6.7
4	0%	0%	1.8	7.5

Table 3, Shopping Motive

Shopping motive	# purchases Store C	# purchases Store D	Dwell Time Store C	Dwell Time Store D
Goal directed	100%	82%	0.3	6.7
Browser	0%	18%	10.6	7.1

Table 4, Density per store

Density	Density Store C	Density Store D
Minimum # of people	5	5
Average	41	91
Maximum # of people	94	184

Methodology

Econometric model

Using the newly created variables, the model can be constructed. The following variables are determined:

1) Dependent variable – **purchase incidence**, showing if a consumer purchased (1) or did not (0).

2) Independent variables:

i. Density, a continuous variable indicating the overlap of shoppers in a given time range.

ii. Shopping with friends indicating whether the consumer shopped alone (0), with 1 friend, 2 friends, etc.

iii. Shopping motive, taking value 1 if the customers were browsers and 0 if they were goal-directed.

3) Control variables:

i. Time spend in a store

ii. Culture

iii. Contact with sales-person

Given that the purchase incidence takes values 1 (bought) and 0 (did not buy), logistic regression is to be used as typically applied when the dependent variable is binary (Gelman & Hill, 2007) to model the probability that the dependent variable falls within the given categories. This paper will investigate the probability of consumers making a purchase (1) or not (0) given the predictors, thus:

$$y_i \begin{cases} 1 \; if \;\; a \; consumer \; made \; a \; purchase \\ \qquad\qquad 0 \; otherwise \end{cases}$$

The probability that y=1 is modeled as:

$$\Pr(y_i = 1) = logit^{-1}(X_i\beta)$$

Using the $logit^{-1}(x) = \frac{e^x}{1+e^x}$ function it is possible to transform a continuous signal into interval between 0 and 1, such that it falls within a proper probability range (Gelman & Hill, 2007). Using this model, one can "derive a binary choice model from underlying behavioral assumptions" (Verbeek, 2008). That is, the binary choice model observes an unobservable utility (latent utility) to determine the probability of *y* given *x:*

$$y^* = logit^{-1}(X_i\beta),$$

where y^* is unobserved, latent variable. The assumption here is that an individual makes a purchase, when y^*>1. Thus, we have y=1, if y^*>1 and y=0 otherwise. That means that an individual makes a purchase when he derived higher utility of buying than from not buying. The model is transformed into:

$$\Pr(y_i = 1) = \Pr\{y^* > 0\} = \mathrm{P}\{X_i\beta + \; \varepsilon_i > 0\} = Pl\{-\varepsilon_i \leq X_i\beta\} = F(\;\; X_i\beta),$$

where F() is defined as the logit function.

This general model is based on the conceptual framework and includes the listed variables. The model is based on 12773 observations and has the following form:

$$\Pr(y_i = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 * x1 + \beta_2 * x1^2 + \beta_3 * x2 + \beta_4 * x3 + \beta_5 * x4 + \beta_6 * x4\text{^}2 + \beta_6 * x5 + \beta_7 * x1 * x2 + \beta_8 * x1 * x3}},$$

Where:

X1 - > density in store

X2 -> shopping with friends

X3 -> shopping motive

X4 - > dwell time in store

X5 - >culture

Density in store and dwell time are modelled as quadratic terms. That is, for those two variables I do not assume that there is a linear relationship with the probability of committing a purchase, rather there are curvilinear effects that should be taken into account. Assumptions of the logistic regression were satisfied (see Appendix A).

Analysis

In this chapter, I will look into the results of the logistic regression as well as model fit. Discussed are the confirmation of each of my hypothesis as well as general results drawn from the logit model. The findings are discussed only for the restricted model, in which excluded variables are mainly control variables that do not add any value to the model.

Model fit

Goodness-of-fit is a summary statistic that indicates how well the model fits the data (Verbeek, 2008). Typically, such measures can be obtained by comparing the model with a model containing nothing else but the intercept. Performing such a comparison in R is quite straight forward. Running the model provides me with a table indicating the estimates, their standard errors, z-scores and significance. Moreover, I have the null and residual deviance that can be further used to evaluate how well the model fits the observed data.

First, based on the null and residual deviance we can compare how good the model is against a model consisting only of an intercept by computing chi-square statistic. The null deviance provides information on how well the data is predicted based only on a model that contains a constant. The residual deviance shows how well the data is predicted based on the variables included in the model. Performing the chi-square test shows that overall the model fits the data well ($p<0.05$).

Looking at the model estimates however, observed is that not all variables are significant, thus next I test for joint significance of the control variables 'culture' and 'salesperson contact' (see *Table 5, Full model*). For this purposes F-statistic is calculated. Based on the F-test, the new variables are not jointly significant at 5% significance level. Thus, these variables does not need to be added to the model. However, to see whether they should be excluded I can compare which model fits the data better – the model using all the variables or the new model in which culture and salesperson contact are excluded. This can be done by looking at the AIC scores. Akaike Information Criterion scores penalized the log likelihood with the number of parameters. Typically, the smaller the AIC the better the fit. Looking at the AIC scores for both models the AIC score is quite high, but a bit lower for the restricted model (see *Table 6*). Chi-square statistic was calculated again for the restricted model in order to see whether the model can fit the data. The results indicate that the new model fits the data well ($p<0.05$), thus the analysis will be done with the restricted model as it fits the data and potentially (although slightly better) have better predictive power.

Table 5, Full model

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.42	0.25	-1.63	0.10	
Dwell time	**-12.80**	**4.26**	**-3.00**	**0.00**	**
Dwell time ^2	5.49	3.30	1.67	0.10	.
Shopping motive 1	**-0.68**	**0.15**	**-4.58**	**0.00**	***
In-store density	0.00	0.00	-0.48	0.63	
Shop with 1 friend	0.03	0.15	0.20	0.84	
Shop with 2 friends	**-0.76**	**0.31**	**-2.47**	**0.01**	*
Shop with 3 friends	-0.21	0.71	-0.30	0.77	
Shop with 4 friends	0.86	1.82	0.47	0.64	
In-store density ^2	0.00	0.00	0.87	0.38	
salesperson contact	0.00	0.00	0.24	0.81	
salesperson contact ^2	0.00	0.00	-0.31	0.76	
Culture - DE	0.10	0.17	0.63	0.53	
Culture - ES	0.15	0.16	0.94	0.35	
Culture - FR	0.15	0.17	0.91	0.36	
Culture - PT	0.14	0.15	0.89	0.37	
Culture - UK	0.17	0.17	1.01	0.32	
In-store density * shopping motive	0.00	0.00	0.58	0.56	
In-Store density* shop with 1 friend	0.00	0.00	-0.24	0.81	

In-Store density* shop with 2 friends	**0.01**	**0.00**	**2.50**	**0.01**	*
In-Store density* shop with 3 friends	0.00	0.01	0.57	0.57	
In-Store density* shop with 4 friends	0.00	0.02	-0.23	0.82	
In-store density* Salesperson Contant	0.00	0.00	-0.96	0.34	

Note: '***': p < 0.001; '**': p <0.01; '*': p < 0.05; '.': p < 0.1

Null deviance: 14180 on 10794 degrees of freedom

Residual deviance: 13989 on 10772 degrees of freedom

(1978 observations deleted due to missingness)

AIC: **14,035**

Table 6, AIC scores

	AIC
Full model	14,035
Restricted model[4]	14,027

[4] The restricted model excludes culture and salesperson contact

Results

As I already have determined which model I will use and base my conclusions, I further look at the estimates provided by running the logistic regression (see **Error! Reference source not found.**). Based on this information I can see which variables are significant and the 'direction' of their significance, e.g. whether the probability to commit a purchase will decrease or increase. However, at this point I cannot say anything about the magnitude of their effect. For this purpose marginal effects need to be computed.

Looking at the results, I see that only a few of the variables included in the model explain the variation of purchase incidence, yet most of them do not comprehend the main effect this paper is trying to explore. Using this information I can already conclude that most of the hypotheses posed in this paper are not confirmed.

First, in-store density is not significant at 5% significance level. My first hypothesis states that crowding decreases the likelihood of purchasing. Given the data I cannot conclude this, thus this hypothesis is rejected.

Second, I am interested to see whether shopping with a friend increases the probability of committing a purchase rather than being alone as posed in hypothesis 2. To check for the effect of shopping with companion on the purchase incidence, I use two models. The first one uses shopping with companion(s) as a categorical variable, while the second uses a dummy variable that indicates whether the person shopped alone or not. Looking at the results, listed in *Table 7*, shopping with 2 friends has a negative effect on the purchase incidence, but overall shopping with friends is not found to have a significant impact on the purchase incidence. The first result is quite an interesting insight as I see that shopping with only one more person or with more than 2 people does not have any effect on the purchase likelihood, but only when the person was shopping with 2 more people. Looking at the impact of the shopping with a companion overall, it can be concluded that whether a person shops alone or not does not really have an impact on the purchasing behavior. Given the data, this hypothesis can be regarded as rejected as no effect is found on the purchase behavior and shopping with 2 friends results in the opposite effect than initially assumed. This merits further theoretical and empirical examination, which are further discussed in the next section. As the first model produces significant results in regards to the effect of a shopping companion over the purchase incidence, I am interested in exploring what the magnitude of their effect would be, thus in the following section I would use Model 1 to estimate the marginal effects of the estimates.

Table 7, Models comparison

	Model 1 (shop with friends as a categorical variable)	Model 2 (Shop with friends as a dummy variable)
Dwell time	**-17.271*** **	**-17.322*** **
Dwell time ^2	**5.977***	**6.082***
Shopping motive 1	**-0.693*** **	**-0.696*** **
In-store density	-0.003	-0.003
Shop with 1 friend	0.025	
Shop with 2 friend	**-0.770** **	
Shop with 3 friend	-0.217	
Shop with 4 friend	0.781	
Shop alone or not		*-0.122*
In-store density ^2	0.00002	0.00002
In-Store density* shop with 1 friend	-0.0003	
In-Store density* shop with 2 friend	**0.007** **	
In-Store density* shop with 3 friend	0.004	
In-Store density* shop with 4 friend	-0.003	
In-Store density shop alone or not*		*0.001*
Observations	10,795	10,795
Log-likelihood	-6,998	-7,003
AIC	14,027	14,023

Note: '***': $p < 0.001$; '**': $p < 0.01$; '*': $p < 0.05$; '.': $p < 0.1$

Third, I investigate the interaction effect of shopping with someone in a dense area. I investigate whether shopping with friends will reduce the negative impact of crowding on shoppers purchase incidence. Interestingly enough here I find that the only significant effect in on shopping again with 2 more people in a dense area. When three people are shopping together they are not affected by the density of the area. This, however, is not true when someone is shopping with one or more than two people. Model 2 indicates that there is shopping alone or not in a dense area does not influence in any way the purchase decision. Thus, so far hypothesis H3 can be regarded as partially confirmed.

I lastly investigate whether goal-directed shoppers are less likely to purchase a product in a crowded area as compared to browsing shoppers. Unfortunately, here observed again is that there is no effect on the purchase incidence at 5% significance level. However, there is negative effect on being a 'browser' on the purchasing likelihood. The magnitude of the effect is analyzed next.

Other conclusions that can be drawn based on the model so far indicate that there is an effect on the time spent per store on the purchase incidence.

Overall, I can conclude so far that **four** of my hypothesis are rejected and not confirmed by the data.

Table 8, Hypothesis

H1: Crowding decreases the likelihood of purchasing.	**Rejected**
H2: Shopping with a friend increases the probability of committing a purchase rather than being alone.	**Rejected**
H3: Shopping with friend(s) will reduce the negative impact of crowding on shoppers purchase;	**Partially confirmed**
H4. Task-oriented shoppers are less likely to purchase a product in a crowded area as compared to browsing shoppers.	**Rejected**

Marginal effects

Further I look into the magnitude of the significant effects. To see what the actual impact of the independent variables is, I estimate the marginal effects (see *Table*). Marginal effects are computed to measure how the predicted probabilities change dependent on the exploratory variables.

Table 9, Marginal effects[5]

Marginal Effects:	Marginal effect	Std. Error	z value	Pr(>|z|)	
Dwell time	**-4.00**	**0.84**	**-4.73**	**0.00**	***
Dwell time ^2	**1.38**	**0.73**	**1.89**	**0.06**	.
Shopping motive 1	**-0.15**	**0.03**	**-5.03**	**0.00**	***
In-store density	0.00	0.00	-0.88	0.38	
Shop with 1 friend	0.01	0.04	0.16	0.87	
Shop with 2 friend	**-0.16**	**0.05**	**-2.95**	**0.00**	**
Shop with 3 friend	-0.05	0.15	-0.32	0.75	
Shop with 4 friend	0.19	0.45	0.43	0.67	

[5] The marginal effects for Model 2 are quite similar, thus the estimates used are only from Model 1

In-store density ^2	0.00	0.00	0.95	0.34	
ShoppingMotive*In-store density	0.00	0.00	0.65	0.52	
In-Store density* shop with 1 friend	0.00	0.00	-0.21	0.84	
In-Store density* shop with 2 friend	**0.00**	**0.00**	**2.52**	**0.01**	*****
In-Store density* shop with 3 friend	0.00	0.00	0.59	0.56	
In-Store density* shop with 4 friend	0.00	0.00	-0.19	0.85	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Looking at the magnitude of each variable, I can further conclude that hypothesis H3 can also be rejected. That is, shopping with two friends where the density of the area increases has a positive influence on the purchasing behavior at 5% significance level. However, the magnitude of the effect is too small to have an actual significance on the purchasing behavior, currently it implies that with the purchase likelihood will increase by 0.001 percentage points when a person is shopping with three people in a dense. Such a small effect, although significant, is not of interest. Thus, it can be disregarded.

Shopping with friends is shown to have a significant impact on the purchasing behavior only when a person is shopping together with two friends. The effect of one shopping companion is negligible, but shopping together with two other people decreases the probability of committing a purchase by 16 percentage points. Contrary to previous research, this notion suggests that 3 people shopping together are rather a large group and less likely to commit a purchase. A possible explanation for this effect might be that in the context of an airport, there could be families that travel together and previous research indicates that shopping together with family members decreases the likelihood to purchase (Borges, Chebat, & Babin, 2010). Moreover, it has been found that females shopping together are also less likely to commit a purchase (Kurt, Inman, & Argo, 2011). One of the limitations of the data is that such information is missing, thus it cannot be concluded definitevely why exactly is this effect significant. The insignificant effect of people shopping with more than 2 people can be due to the low sample size (around 1% of the sample).

The effects of dwell time and shopping motive are found to be both significant. The dwell time was initially inputted in the model as a quadratic term. This suggest that there could be a 'turning point' in consumers behavior. For example, as people spend too little time in the store it is most likely that they will not commit a purchase, however with the increase in time they will become more likely to actually purchase something from the store. Such effect is observed in this dataset. The dwell time is negative at first, meaning that shoppers spending too little time in the store are less likely to purchase something, however, the quadratic term of dwell time is positive and this value actually indicates the turning point at which consumers are now more likely to commit a purchase.

Being a 'browser' has a negative effect on the purchase incidence. Browsing while shopping will decrease the probability of committing a purchase by 15 percentage points. This is consistent with theory indicating that such consumers have rather abstract goals and might be engaged in the shopping trip due to different personal and social motives (Eroglu & Macheleit, 1990). Shoppers at the Porto airport also exhibit such behavior. They have entered all four shops in the airport, yet they committed less purchases than the goal-oriented shoppers.

The majority of the hypotheses are rejected due to insignificant effects. Hypothesis H2 is rejected as it has partially confirmed the opposite effect.

H1: Crowding decreases the likelihood of purchasing.	**Rejected**
H2: Shopping with a friend increases the probability of committing a purchase rather than being alone.	**Rejected**
H3: Shopping with friend(s) will reduce the negative impact of crowding on shoppers purchase;	**Rejected**
H4. Task-oriented shoppers are less likely to purchase a product in a crowded area as compared to browsing shoppers.	**Rejected**

Conclusions

General Discussion

The main goal of this paper was to explore the social influences on consumers' in-store decision making. More precisely, the main objectives of this thesis was to investigate the joint impact of shopping with a companion and store crowding on shoppers' purchase likelihood. My proposition was that when shoppers are accompanied by someone in a dense store area, their purchase incidence won't be affected by the store atmospherics. Moreover, the paper investigated whether certain behaviors such as goal-directness will be affected when the store is crowded. Behavioral data was used for testing. The data was already obtained for an airport in Porto and it consisted of approximately 50,000 observations. Another objective of the thesis was to use the data in order to get robust insights rather than utilizing self-reported data, which can be prone to different biases.

The research was performed with the aid of high-quality behavioral passive data as it aimed to provide a more extensive view of the social influences during shopping in a crowded area. The data covered shopping behavior in an airport in Porto including 4 different shops. The collected data had information in regards to one's shopping time, dwell time per store, path between the stores, whether subjects had made a purchase or not, etc. Using this information it was possible to construe the additional independent exploratory variables that were used in the analysis. The data analysis was done in two steps. First, I looked into the descriptive statistics as to get a more comprehensive view of different patterns within the data. Based on the frequencies, only one shop was used for the main analysis as for the others there was insufficient information. Second, logistic regression was performed to evaluate what are the main dependencies on the purchase incidence based on my model.

In regards to the main effects, I find no support for my hypotheses. Despite previous research indicating that store density has a negative effect on the shopping behavior, I find no significant results on that. Since the data was collected at an airport, it might be the case that this effect just do not occur in such a setting. People at airports might have different shopping goals than people shopping in a shopping center for example. Thus, further research should focus on examining such interactions in another setting, e.g. a department store. Moreover, the research did not find significant results in regards to the effect of shopping with a companion versus shopping alone on the purchase behavior. This again can be due to the setting in which the analysis was performed. Additionally, omitted from the model are important control variables such as demographics which can further help explain some of the variance. Future research can also adopt a joint methodology in which both the behavioral data but also self-reported data of a random part of the sample is collected.

There are some other interesting insights from the research.

First, contrary to one of my hypothesis, shopping with 2 more people has a negative effect on the purchase incidence. This effect is present without accounting for the store's density. Thus, on average across the sample shopping with two more shoppers can decrease the purchase likelihood by 16%. Previous research indicates that the purchase likelihood is affected by the gender of the shopping group (Kurt, Inman, & Argo, 2011). That is, when males are shopping together they tend to spend more, while when women are shopping together they tend to spend less. Furthermore, previous studies indicate that the purchase incidence decreases if a shopper is together with family rather than friends (Borges, Chebat, & Babin, 2010). Unfortunately, the behavioral data did not provide information about demographic characteristics of the subjects or information about the shopping companion as to investigate further the reasons for such a result.

Second, browsing in an shopping area rather than having a shopping goal in mind has again a negative impact on the purchase incidence. This coincides with previous research and theory that such shoppers have rather abstract goal, but that does not necessarily mean they will not commit a purchase (Eroglu & Macheleit, 1990).

Finally, the dwell time has at first a negative impact on the purchase likelihood, and afterwards a positive. That is consumers spending too little time in the store are less likely to purchase something, but there is also a turning point at which consumers become more likely to commit a purchase.

Contributions & Implications

The study contributions and implications are in three directions.

First, the thesis contributes to the marketing literature. More precisely, the study analyzes the social interaction between different forces on the shopping behavior using high-quality behavioral data. The marketing literature suggests that people are less inclined to commit a purchase in a dense area. However, the marketing literature is scarce when exploring the relationship between different forces, e.g. shopping with a friend or goal-directness in a dense area on the purchase likelihood. By exploring the interaction of shopping with a friend and crowding, my thesis adds to that literature stream.

Second, this paper adds up methodologically to the emerging stream of literature on behavioral modelling. The paper tested behavioral hypothesis using real-world secondary data in a retail context. Previous researchers have used mostly self-reported data, however utilizing and analyzing passive behavioral data has not been widely applied. Typically such data is deemed as more accurate, complete and unbiased. Although the method utilized in the thesis is not novel, the paper employs different data processing techniques in order to come up with the needed variables used in the main statistic method. The challenges in that direction include having sufficient technical skills to process the data and devoting more time to that procedure. Thus, when researchers are limited due to time and technical abilities, these considerations should be taken into account in order to make a fair trade-off between conducting a research using such rich passive data and exploiting the “known” self-reported data.

Third, the thesis offers several managerial insights for retailers. Shopping with two friends turns out to have a negative impact on the overall purchase intention in an airport. Although, future research should examine what forces stop a group of 3 people from committing a purchase, managers can lure such shoppers with different promotional techniques and marketing campaigns. For example, managers could make use of promotions that allows to buy multiple products for less, e.g. 2 for 1. This can attract people shopping together as the cost over a given purchase might be diversified among them. Additionally, retail managers might think of how to attract browsers to commit more purchases and be more in ‘goal-directed mode’. Promotional offers would also be well applied here.

Limitations and future research

One limitations was in regards to inconsistent data. There were a lot of subjects that had stayed in the store less than 30 seconds, yet they made a purchase. Such respondents were exclusively shopping in stores A and B, however they were removed from the analysis due to inconsistency. Typically if they have spent 30 second in the cash register, it might be the case that they did spent a bit more in the store. Moreover, the path data suggested that subjects only entered store D or entered all of the stores. This distinction was taken into account when creating the variable ‘shopping motive’. Although respondents were not removed based on this, it is a strange behavior that might have been coded wrongly beforehand.

Another limitation is with regards to the measurement of the variables and the assumptions behind them. As it is behavioral passive data, not collected for the purpose of the analysis, a lot of assumptions are being made in order to assess it and further use it. For instance, dividing people over browsers and goal-oriented might be done differently if it is done using self-reported data. It could have been that people entering store D, which was the largest store in the airport consisting of different area that offer a variety of products, might have also been browsers and not goal-oriented. Moreover, people shopping together were regarded as those that entered the same shops in the same time. However, it might have been the case that those are not two distinct persons rather one person caring two phones.

Third limitation for this particular analysis was the context. People shopping in airports might be in a different mindset than people shopping in a regular department store or shopping malls. The insignificance of the main effect, shopping in a crowded area, might have not affected shoppers in the airport store as they experience different emotions before travelling. For example, they might be excited about their trip, nervous when flying, etc. Thus, the negative effect of crowding might have already been eliminated as of such emotions not due to different social interactions.

Future research should replicate the design in a different context that might be more appropriate to capture shopper behavior more accurately. Using such data there might be an abundance of opportunities to discover consumers behavior by examining a large amount of data patterns.

Moreover, the data would have more predictive power if along with it self-reported data is collected for a proportion of the shoppers. That is, future research could aim at obtaining as much information as possible by merging two datasets – namely, behavioral data along with survey data. In such a way, personal information as well as purchase intentions and products usage might be captured to compliment the behavioral data. Such design might be more insightful as it would allow for many different variables to be collected and explored. Exploiting two datasets can results in a model that fits the data better and explain a lot of the variance of the dependent variable, thus leading to greater insights.

References

Argo, J. J., Dahl, D. W., & Manchanda, R. V. (2005). The Influence of a Mere Social Presence in a Retail Context. *Journal of Consumer Research*, *32*(2), 207–212.

Aribarg, A., Arora, N., & Kang, M. Y. (2010). Predicting Joint Choice Using Individual Data. *Marketing Science*, *29*(1), 139–157. doi:10.1287/mksc.1090.0490

Ariely, D., & Levav, J. (2000). Sequential Choice in Group Settings: Taking the Road Less Traveled and Less Enjoyed. *Journal of Consumer Research*, *27*(3), 279–290. doi:10.1086/317585

Babin, B. J., Darden, W. R., & Griffin, M. (1994). Work and/or Fun: Measuring Hedonic and Utilitarian Shopping Value. *Journal of Consumer Research*, *20*.

Baker, J., & Wakefield, K. L. (2012). How consumer shopping orientation influences perceived crowding, excitement, and stress at the mall. *Journal of the Academy of Marketing Science*, *40*(6), 791–806. doi:10.1007/s11747-011-0284-z

Baum, A., Harpin, R. E., & Valins, S. (1975). The Role of Group Phenomena in the Experience of Crowding. *Environment and Behavior*, *7*(2), 185–198. doi:10.1177/001391657500700204

Becker, G. S. (1991). A Note on Restaurant Pricing and Other Examples of Social Influences on Price. *Journal of Political Economy*. doi:10.1086/261791

Bell, D. R., Corsten, D., & Knox, G. (2011). From Point of Purchase to Path to Purchase: How Preshopping Factors Drive Unplanned Buying. *Journal of Marketing*, *75*(1), 31–45. doi:10.1509/jmkg.75.1.31

Borges, A., Chebat, J.-C., & Babin, B. J. (2010). Does a companion always enhance the shopping experience? *Journal of Retailing and Consumer Services*, *17*(4), 294–299. doi:10.1016/j.jretconser.2010.02.007

Childers, T. L., & Rao, A. R. (1992). The Influence of Familial and Peer-Based Reference Groups on Consumer Decisions. *Journal of Consumer Research*, *19*(2), 198. doi:10.1086/209296

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: compliance and conformity. *Annual Review of Psychology*, *55*(1974), 591–621. doi:10.1146/annurev.psych.55.090902.142015

Corfman, K. P. (1991). Perceptions of Relative Influence: Formation and Measurement. *Journal of Marketing Research*, *28*(2), 125. doi:10.2307/3172802

Dion, D. (2004). Personal control and coping with retail crowding. *International Journal* *of Service Industry Management*, *15*(3), 250–263. doi:10.1108/09564230410540926

Eroglu, S. a., & Macheleit, K. A. (1990). An Empirical Study of Retail Crowding: Antecedents and Consequences. *Journal of Retailing*, *66*(2), 201–221.

Eroglu, S. a., Machleit, K., & Barr, T. F. (2005). Perceived retail crowding and shopping satisfaction: The role of shopping values. *Journal of Business Research*, *58*(8), 1146–1153. doi:10.1016/j.jbusres.2004.01.005

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. *Cambridge*. doi:10.2277/0521867061

Guido, G. (2006). Shopping motives, big five factors, and the hedonic/utilitarian shopping value: An integration and factorial study. *Innovative Marketing*, *2*(2), 57–67. Retrieved from http://www.businessperspectives.org/journals_free/im/2006/im_en_2006_02_Guido.pdf

Harrell, G. D., & Hutt, M. D. (1976). Buyer Behavior under Conditions of Crowding: an Initial Framework. *Advanced in Consumer Research*.

Harrell, G. D., Hutt, M. D., & Anderson, J. C. (1980). Path Analysis of Buyer Behavior under Conditions of Crowding. *Journal of Marketing Research*, *17*(1), 45–51.

Harrell, Frank E., Kerry L. Lee, and Daniel B. Mark. "Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors." Statistics in medicine 15 (1996): 361-387.

Hart, P. M., & Dale, R. (2014). With or without you: The positive and negative influence of retail companions. *Journal of Retailing and Consumer Services*, *21*(5), 780–787. doi:10.1016/j.jretconser.2014.06.004

Hui, M., & Bateson, J. E. G. (1991). Perceived Crowding on Control and and Consumer the Effects Choice of the Experience. *Journal of Consumer Research*, *18*(2), 174–184.

Hui, S. K., Bradlow, E. T., & Fader, P. S. (2009). Testing Behavioral Hypotheses Using an Integrated Model of Grocery Store Shopping Path and Purchase Behavior. *Journal of Consumer Research*, *36*(3), 478–493. doi:10.1086/599046

Hui, S. K., Fader, P. S., & Bradlow, E. T. (2009). Path Data in Marketing: An Integrative Framework and Prospectus for Model Building. *Marketing Science*, *28*(March 2015), 320–335. doi:10.1287/mksc.1080.0400

Inman, J. J., Winer, R. S., & Ferraro, R. (2009). The Interplay Among Category Characteristics, Customer Characteristics, and Customer Activities on In-Store Decision Making. *Journal of Marketing*, *73*(5), 19–29. doi:10.1509/jmkg.73.5.19

Jarboe, Glen R., and Carl D. McDaniel. "A profile of browsers in regional shopping malls." Journal of the Academy of Marketing Science 15.1 (1987): 46-53.

Janiszewski, C. (1998). The Influence of Display Characteristics on Visual Exploratory Search Behavior. *Journal of Consumer Research*, *25*(3), 290–301. doi:10.1086/209540

Kim, J.-E., & Kim, J. (2012). Human factors in retail environments: a review. *International Journal of Retail & Distribution Management*, *40*, 818–841. doi:10.1108/09590551211267593

Korgaonkar, P. (1981). Shopping Orientations of Catalog Showroom Patrons. *Journal of Retailing*, *57*(1), 78–90.

Kotler, P. (1973). Atmospherics as a Marketing Tool. *Journal of Retailing*, *49*(4), 48–65. doi:10.1016/j.obhdp.2011.03.002

Kurt, D., Inman, J. J., & Argo, J. J. (2011). The Influence of Friends on Consumer Spending: The Role of Agency– Communion Orientation and Self-Monitoring. *Journal of Marketing Research*, *48*(4), 741–754. doi:10.1509/jmkr.48.4.741

Latane, B. (1981). The psychology of social impact. *American Psychologist*, *36*(4), 343–356. doi:10.1037/0003-066X.36.4.343

Li, J.-G. T., Kim, J.-O., & Lee, S. Y. (2009). An empirical examination of perceived retail crowding, emotions, and retail outcomes. *The Service Industries Journal*, *29*(5), 635–652. doi:10.1080/02642060902720121

Luo, X. (2005). How Does Shopping With Others Influence Impulsive Purchasing? *Journal of Consumer Psychology*, *15*(4), 288–294. doi:10.1207/s15327663jcp1504_3

Machleit, K. a., Kellaris, J. J., & Eroglu, S. a. (1994). Human versus spatial dimensions of crowding perceptions in retail environments: A note on their measurement and effect on shopper satisfaction. *Marketing Letters*, *5*(2), 183–194. doi:10.1007/BF00994108

Machleit, K., Eroglu, S., & Mantel, S. (2000). Perceived Retail Crowding and Shopping Satisfaction: What Modifies This Relationship? *Journal of Consumer Psychology*, *9*(1), 29–42. doi:10.1207/s15327663jcp0901_3

Maeng, A., Tanner, R. J., & Soman, D. (2013). Conservative When Crowded : Social Crowding and Consumer Choice. *Journal of Market Research*, *L*(December), 739–752. doi:10.1509/jmr.12.0118

Mangleburg, T. F., Doney, P. M., & Bristol, T. (2004). Shopping with friends and teens' susceptibility to peer influence. *Journal of Retailing*, *80*(2), 101–116. doi:10.1016/j.jretai.2004.04.005

Manski, C. (2004). Measuring expectations. *Econometrica*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0262.2004.00537.x/abstract

Mattila, A. S., & Wirtz, J. (2008). The role of store environmental stimulation and social factors on impulse purchasing. *Journal of Services Marketing*, *22*(7), 562–567. doi:10.1108/08876040810909686

Mehta, R. (2013). Understanding perceived retail crowding: A critical review and research agenda. *Journal of Retailing and Consumer Services*, *20*(6), 642–649. doi:10.1016/j.jretconser.2013.06.002

Michon, R., Chebat, J. C., & Turley, L. W. (2005). Mall atmospherics: The interaction effects of the mall environment on shopping behavior. *Journal of Business Research*, *58*(5), 576–583. doi:10.1016/j.jbusres.2003.07.004

Novak, T. P., Hoffman, D., & Duhackek, A. (2003). The influece of Goal-Directed and Experiental Activities on Online Flow Activities. *Journal of Consumer Psychology*. doi:10.1207/S15327663JCP13-1

Pons, F., & Laroche, M. (2007). Cross-cultural differences in crowd assessment. *Journal of Business Research*, *60*(3), 269–276. doi:10.1016/j.jbusres.2006.10.017

Sommer, R., Wynes, M., & Brinkley, G. (1992). Social Falicitation Effects in Shopping behavior. *Environment and Behavior*, *24*(3), 285–297. doi:0803973233

Stafford, James E. "Effects of group influences on consumer brand preferences." Journal of marketing Research (1966): 68-75.

Stokols, D. (1972). On the distinction between density and crowding: some implications for future research. *Psychological Review*, *79*(3), 275–277. doi:10.1037/h0032706

Turley, L. W., & Milliman, R. E. (2000). Atmospheric Effects on Shopping Behavior: A Review of the Experimental Evidence. *Journal of Business Resea*, *211*(1990), 193–211.

Van Rompay, T. J. L., Krooshoop, J., Verhoeven, J. W. M., & Pruyn, A. T. H. (2012). With or without you: Interactive effects of retail density and need for affiliation on shopping pleasure and spending. *Journal of Business Research*, *65*(8), 1126–1131. doi:10.1016/j.jbusres.2011.08.005

Verbeek, M. (2008). *A Guide to Modern Econometrics*. *Text*. doi:10.1017/CBO9781107415324.004

Vizzari, G., Manenti, L., & Crociani, L. (2013). Adaptive pedestrian behaviour for the preservation of group cohesion. *Complex Adaptive Systems Modeling*, *1*(1), 1–29. doi:10.1186/2194-3206-1-7

Ward, J. C., & Barnes, J. W. (2001). Control and affect: The influence of feeling in control of the retail environment on affect, involvement, attitude, and behavior. *Journal of Business Research*, *54*(2), 139–144. doi:10.1016/S0148-2963(99)00083-1

Willis, M. (2008). Shopping East and Shopping West—Chinese Consumer Behavior in Two Worlds. *Journal of East-West Business*, *14*(3-4), 271–298. doi:10.1080/10669860802530400

Zhang, X., Li, S., Burke, R. R., & Leykin, A. (2014). An Examination of Social Influence on Shopper Behavior Using Video Tracking Data. *Journal of Marketing*, *78*(5), 24–41. doi:10.1509/jm.12.0106

Appendix

Appendix A

Assumptions of logistic regression

1. Linearity

Linearity between the independent and dependent variables is not assumed in logistic regression, rather linearity is assumed between the odds rations of the independent variables and the dependent variable (Harrell, Lee & Mark, 1996). One way for accounting for linearity would be to transform the X variables into "multiple terms". Given the model, this would mean adding squared terms or even higher terms. The model I assumed already has transformed the density in store and the dwell time in store as quadratic terms so that curvilinear relationship is assumed between the propensity to buy and in-store density and the time spend in the store.

2. Normal distribution of the residuals

When the data sample is large enough the residuals do not necessarily need to be normally distributed. Non-normality does not result in inconsistent estimates (Verbeek, 2008).

3. Heteroscedasticity

Homoscedasticity of the variances is not needed for logistic regression. Homoscedasticity implies that the variance of the independent variables should be the same for all observations across the values of the dependent variable. However, in logistic regression this assumption is not needed as we cannot assume that the variance of the two events (have bought and not bought) is the same for all observations.

Model fit however and independence of observations are still important indicators. Due to the data collection procedure itself, it can be concluded that the observations are independent of one another.

Appendix B

Variables & Data Processing in R

```
setwd('C:/Users/m.georgieva/Documents/uni')

AirportData <- read.csv("data_v1.csv", header = TRUE, sep = ',')

#install.packages("lubridate")

library(lubridate)

#install.packages("plyr")
```

```
library(plyr)

#install.packages('sqldf')

library(sqldf)

AirportData$Date=as.Date(AirportData$Date, '%m/%d/%Y')

summary(AirportData$Date)

colnames(AirportData) <- gsub("\\.","_",colnames(AirportData))

colnames(AirportData)

#recode variables

AirportData$A_entry__Sec_=as.numeric(AirportData$A_entry__Sec_)

AirportData$B_entry__Sec_=as.numeric(AirportData$B_entry__Sec_)

AirportData$C_entry__Sec_=as.numeric(AirportData$C_entry__Sec_)

AirportData$D_entry__Sec_=as.numeric(AirportData$D_entry__Sec_)

AirportData$A_dwell=period_to_seconds(hms(AirportData$A_dwell))

AirportData$B_dwell=period_to_seconds(hms(AirportData$B_dwell))

AirportData$C_dwell=period_to_seconds(hms(AirportData$C_dwell))

AirportData$store_D__cosmetics__Dwell=period_to_seconds(hms(AirportData$store_D__cosm
etics__Dwell))

AirportData$store_D__wine___others__Dwell=period_to_seconds(hms(AirportData$store_D__
wine___others__Dwell))

AirportData$Security_Time=period_to_seconds(hms(AirportData$Security_Time))

#summary(AirportData$Security_Time)


#create new clean dataset

AirportDataCleaned = sqldf("select * from AirportData where X_Stores>0")

AirportDataCleaned$Store_C_contact=period_to_seconds(hms(AirportDataCleaned$Store_C_co
ntact))

AirportDataCleaned$Store_D_Contact=period_to_seconds(hms(AirportDataCleaned$Store_D_C
ontact))

#create browsers=1 if x_stores>1, goal-oriented=0 if x_stores=1

AirportDataCleaned$shopping_motive="N/A"
```

```
AirportDataCleaned$shopping_motive[which(AirportDataCleaned$X_Stores > 1)]= "1" #browsers

AirportDataCleaned$shopping_motive[which(AirportDataCleaned$X_Stores == 1)]= "0" #goal-directed

#AirportDataCleaned$Shopping_Motive<- ifelse(AirportDataCleaned$X_Stores==1,"0", "1")

time_of_purchase=sqldf("select Date, count(Date) as AirportDate from AirportDataCleaned group by 1 order by 1")

#adjsut the margins of the chart

par(mar=c(2,2,1,1))

barplot(time_of_purchase$AirportDate, names.arg = time_of_purchase$Date)

#1) do comparisons - compare the hypothesis against alternative

#2) show causality

#3) show multivariate data - as much data as possible on one chart

#4) integrate the evidence

#5) sources from where the data came from - to make it credible

#time per store

AirportDataCleaned$timeSpent <- AirportDataCleaned$A_dwell + AirportDataCleaned$B_dwell+ AirportDataCleaned$C_dwell + AirportDataCleaned$store_D__cosmetics__Dwell+ AirportDataCleaned$store_D__wine___others__Dwell

#hist(AirportDataCleaned$timeSpent, breaks=500)


#REMOVE SPENT LESS THAN 30 SEC And No PAth

AirportDataCleaned2 = sqldf("select * from AirportDataCleaned where timeSpent>30 AND Path<>''")

#hist(AirportDataCleaned2$timeSpent, breaks=500)


#create 'shopping with a companion' variable

#Shop_Together=sqldf("select DISTINCT Date, Path from AirportDataCleaned ")

test = sqldf("select * from AirportDataCleaned where Time_HHMMSS < '82320' and Time_HHMMSS > '81320'")

test2 <- AirportDataCleaned2

test2$Time_HHMM <- gsub('.{2}$', '', test2$Time_HHMMSS)
```

```
test2 <- transform(test2,id=as.numeric(factor(paste(test2$Time_HHMM, test$Date, test$Path))))


var1df <- ddply(test2,.(test2$id),nrow)
colnames(var1df)[1] <- "id"
agg <- aggregate(data=df, type ~ color, function(x) length(unique(x)))
dataSetFinal <- merge(test2, var1df, by="id", all=TRUE)
count(dataSetFinal$V1)
colnames(dataSetFinal)[42] <- "Shop_with_frieds"
drop(dataSetFinal$Shop_Together)
drop (dataSetFinal$In_store_Density_A)
drop (dataSetFinal$In_store_Density_B)
drop (dataSetFinal$In_store_Density_C)
drop (dataSetFinal$In_store_Density_D)


ddply(AirportDataCleaned2,.(AirportDataCleaned2$Date, AirportDataCleaned2$Time),nrow)
testDublicated = sqldf("sELECT Path, Ent_St1, Date, timeSpent, count(timeSpent) FROM AirportDataCleaned2 GROUP BY timeSpent order by Date")
test = sqldf("select * from AirportDataCleaned2 where Path='D'")
names(dataSetFinal)[2] <- "resp_id"



#create 'density' variable
dataSetFinal$D_dwell=dataSetFinal$store_D__cosmetics__Dwell+ dataSetFinal$store_D__wine___others__Dwell
finalDS <- dataSetFinal
finalDS$Path1 <-NULL
finalDS$Path2 <-NULL
finalDS$Path3 <-NULL
finalDS$Path4 <-NULL
```

```
write.csv(finalDS, "final.csv")

summary(dataSetFinal$Store_D_Contact)

dataSetFinal$Path1 <- lapply(strsplit(as.character(dataSetFinal$Path), "\\-"), "[", 1)

dataSetFinal$Path2 <- lapply(strsplit(as.character(dataSetFinal$Path), "\\-"), "[", 2)

dataSetFinal$Path3 <- lapply(strsplit(as.character(dataSetFinal$Path), "\\-"), "[", 3)

dataSetFinal$Path4 <- lapply(strsplit(as.character(dataSetFinal$Path), "\\-"), "[", 4)

dataSetFinal$store_A_ENT <- ifelse(dataSetFinal$Path1 == "A", paste(dataSetFinal$Date,dataSetFinal$Ent_St1),
        ifelse(dataSetFinal$Path2 == "A", paste(dataSetFinal$Date,dataSetFinal$Ent_St2),
        ifelse(dataSetFinal$Path3 == "A", paste(dataSetFinal$Date,dataSetFinal$Ent_St3),
        ifelse(dataSetFinal$Path4=="A", paste(dataSetFinal$Date,dataSetFinal$Ent_St4),0))))


#strptime(paste(dataSetFinal$Date[1],dataSetFinal$Ent_St1[1]), format="%Y-%m-%d %H:%M:%S")

dataSetFinal$store_B_ENT <- ifelse(dataSetFinal$Path1 == "B", paste(dataSetFinal$Date,dataSetFinal$Ent_St1),
        ifelse(dataSetFinal$Path2 == "B", paste(dataSetFinal$Date,dataSetFinal$Ent_St2),
        ifelse(dataSetFinal$Path3 == "B", paste(dataSetFinal$Date,dataSetFinal$Ent_St3),
        ifelse(dataSetFinal$Path4 == "B", paste(dataSetFinal$Date,dataSetFinal$Ent_St4),0))))


dataSetFinal$store_C_ENT <- ifelse(dataSetFinal$Path1 == "C", paste(dataSetFinal$Date,dataSetFinal$Ent_St1),
        ifelse(dataSetFinal$Path2 == "C", paste(dataSetFinal$Date,dataSetFinal$Ent_St2),
        ifelse(dataSetFinal$Path3 == "C", paste(dataSetFinal$Date,dataSetFinal$Ent_St3),
        ifelse(dataSetFinal$Path4 == "C", paste(dataSetFinal$Date,dataSetFinal$Ent_St4),0))))


dataSetFinal$store_D_ENT <- ifelse(dataSetFinal$Path1 == "D", paste(dataSetFinal$Date,dataSetFinal$Ent_St1),
```

```
                    ifelse(dataSetFinal$Path2 == "D", paste(dataSetFinal$Date,dataSetFinal$Ent_St2),
                    ifelse(dataSetFinal$Path3 == "D", paste(dataSetFinal$Date,dataSetFinal$Ent_St3),
                    ifelse(dataSetFinal$Path4 == "D",
paste(dataSetFinal$Date,dataSetFinal$Ent_St4),0))))


dataSetFinal$store_A_EXT <- as.POSIXct(dataSetFinal$store_A_ENT) + dataSetFinal$A_dwell
dataSetFinal$store_B_EXT <- as.POSIXct(dataSetFinal$store_B_ENT) + dataSetFinal$B_dwell
dataSetFinal$store_C_EXT <- as.POSIXct(dataSetFinal$store_C_ENT) + dataSetFinal$C_dwell
dataSetFinal$store_D_EXT <- as.POSIXct(dataSetFinal$store_D_ENT) + dataSetFinal$D_dwell
dataSetFinal$store_A_INT <- interval(dataSetFinal$store_A_ENT, dataSetFinal$store_A_EXT)
dataSetFinal$store_B_INT <- interval(dataSetFinal$store_B_ENT, dataSetFinal$store_B_EXT)
dataSetFinal$store_C_INT <- interval(dataSetFinal$store_C_ENT, dataSetFinal$store_C_EXT)
dataSetFinal$store_D_INT <- interval(dataSetFinal$store_D_ENT, dataSetFinal$store_D_EXT)
#test123 <- as.data.frame(dataSetFinal$resp_id)
#for(var1 in 1:nrow(test123)){
#test123$test123[var1] <-
int_overlaps(dataSetFinal$store_A_INT[1],dataSetFinal$store_A_INT[var1])
#  count(test123$test123)[2,2]
dataSetFinal$store_A_OLAP <- lapply(dataSetFinal$store_A_INT, FUN = function (x) if (!is.na(x)){
as.numeric( count(int_overlaps(x,dataSetFinal$store_A_INT))[2,2])}else{NA})
dataSetFinal$store_B_OLAP <- lapply(dataSetFinal$store_B_INT, FUN = function (x) if (!is.na(x)){
as.numeric( count(int_overlaps(x,dataSetFinal$store_B_INT))[2,2])}else{NA})
dataSetFinal$store_C_OLAP <- lapply(dataSetFinal$store_C_INT, FUN = function (x) if (!is.na(x)){
as.numeric( count(int_overlaps(x,dataSetFinal$store_C_INT))[2,2])}else{NA})
dataSetFinal$store_D_OLAP <- lapply(dataSetFinal$store_D_INT, FUN = function (x) if (!is.na(x)){
as.numeric( count(int_overlaps(x,dataSetFinal$store_D_INT))[2,2])}else{NA})
dataSetFinal$store_A_OLAP <- unlist(lapply(dataSetFinal$store_A_OLAP, FUN= function(x)
as.numeric(as.character(x))))
dataSetFinal$store_B_OLAP <- unlist(lapply(dataSetFinal$store_B_OLAP, FUN= function(x)
as.numeric(as.character(x))))
dataSetFinal$store_C_OLAP <- unlist(lapply(dataSetFinal$store_C_OLAP, FUN= function(x)
as.numeric(as.character(x))))
```

```
dataSetFinal$store_D_OLAP <- unlist(lapply(dataSetFinal$store_D_OLAP, FUN= function(x) as.numeric(as.character(x))))

summary(dataSetFinal[60:63])

#purchases

count(dataSetFinal$A_bought)

count(dataSetFinal$B_bought)

count(dataSetFinal$C_bought)

count(dataSetFinal$D_Bought)

#time spend per shop

summary(dataSetFinal$A_dwell)

summary(dataSetFinal$B_dwell)

summary(dataSetFinal$C_dwell)

summary(dataSetFinal$store_D__cosmetics__Dwell)

summary(dataSetFinal$store_D__wine___others__Dwell)

summary(dataSetFinal$D_dwell)

#visitors per store

library(Hmisc)

describe(dataSetFinal[60:63])

dataSetFinal <- as.data.frame(dataSetFinal)

#install.packages("interplot")

library(ggplot2)

library(Rcpp)

#fit <- glm(Fprmula~x1+x2+x3,data=mydata,family=binomial())

dataSetFinal$shopping_motive <-  as.factor(dataSetFinal$shopping_motive)

dataSetFinal$Shop_with_frieds <-  as.factor(dataSetFinal$Shop_with_frieds)

logitTest <- glm(dataSetFinal$D_Bought ~

   #       dataSetFinal$D_dwell +

           poly(dataSetFinal$D_dwell,2) +

           dataSetFinal$shopping_motive +
```

```
                dataSetFinal$shopping_motive* dataSetFinal$store_D_OLAP +
                dataSetFinal$Shop_with_frieds +
                I(dataSetFinal$store_D_OLAP^2) +
                dataSetFinal$Shop_with_frieds* dataSetFinal$store_D_OLAP,
               #I(dataSetFinal$Store_D_Contact^2) +
               # dataSetFinal$Store_D_Contact*dataSetFinal$store_D_OLAP +

      #dataSetFinal$Country,
              data = dataSetFinal,
              family = binomial(link="logit") )
#assumptions
#grapths
summary(logitTest)
par(mfrow = c(2, 2))
plot(logitTest)
#modelfit
chidiff <- logitTest$null.deviance - logitTest$deviance
ddiff <- logitTest$df.null - logitTest$df.residual
pchisq(chidiff, ddiff, lower.tail = FALSE)
library(BaylorEdPsych)

PseudoR2(logitTest)

correct <- logitTest$fitted.values
binarycorrect <- ifelse(correct<0.385,0,1)
table(dataSetFinal$D_Bought, binarycorrect)
summary(logitTest)
dataSetFinal$shopping_motive <- as.factor(dataSetFinal$shopping_motive)
```

```
dataSetFinal$Shop_with_frieds <- as.factor(dataSetFinal$Shop_with_frieds)
logitmfx(formula = dataSetFinal$D_Bought ~
          poly(dataSetFinal$D_dwell,2) +
          dataSetFinal$shopping_motive +
         dataSetFinal$shopping_motive* dataSetFinal$store_D_OLAP +
          dataSetFinal$Shop_with_frieds +
          I(dataSetFinal$store_D_OLAP^2) +
        dataSetFinal$Shop_with_frieds* dataSetFinal$store_D_OLAP,
          #poly(dataSetFinal$Store_D_Contact,2),# +
     #    dataSetFinal$Store_D_Contact*dataSetFinal$store_D_OLAP +
     #   dataSetFinal$Country,
        data = dataSetFinal)


library(mfx)
logitmfxTest <- logitmfx(formula = dataSetFinal$D_Bought ~
               #          dataSetFinal$D_dwell +
               poly(dataSetFinal$D_dwell,2) +
               dataSetFinal$shopping_motive +
               dataSetFinal$shopping_motive* dataSetFinal$store_D_OLAP +
               dataSetFinal$Shop_with_frieds +
               dataSetFinal$store_D_OLAP +
               I(dataSetFinal$store_D_OLAP^2) +
               dataSetFinal$Shop_with_frieds* dataSetFinal$store_D_OLAP,
              # dataSetFinal$Store_D_Contact +
              # I(dataSetFinal$Store_D_Contact^2) +
              # dataSetFinal$Store_D_Contact*dataSetFinal$store_D_OLAP +
              # dataSetFinal$Country,
             data = dataSetFinal)
```

```
var.test(dataSetFinal$Country, dataSetFinal$Store_D_Contact)
```