Graphical user interface, application

Description automatically generated

The effect of brand crises on

consumer engagement in

online brand communities

An empirical study into the Automobile industry

Group 7

Khoi Gia Pham 523755

Nam Hoang Giap 534598

Izzuddin Idham 527363

Hong Minh Nguyen 532460

Track 1: Analytic Decision Makers

Supervisor: Martina Pocchiari

BSc International Business Administration

Rotterdam School of Management, Erasmus University

Background pattern

Description automatically generated with low confidence

Table of contents

[Chapter 1: Introduction 1](#_Toc108074669)

[1.1 Overview 1](#_Toc108074670)

[1.2 Research objective and Research question 3](#_Toc108074671)

[1.3 Empirical context and Research design 4](#_Toc108074672)

[1.4 Relevance of study 4](#_Toc108074673)

[Chapter 2: Literature Review 5](#_Toc108074674)

[2.1 Brand crises and product–harm crises 5](#_Toc108074675)

[2.2 Product recall in the Automobile industry 6](#_Toc108074676)

[2.3 Online brand communities 7](#_Toc108074677)

[2.4 Consumer’s engagement and its relationship with volume and sentiments of user-generated content (UGC) on an online brand community 8](#_Toc108074678)

[2.5 The moderating role of brand commitment 9](#_Toc108074679)

[2.6 Media coverage as a characteristic of brand crises 9](#_Toc108074680)

[2.7 Conclusion 10](#_Toc108074681)

[Chapter 3: Research Design 11](#_Toc108074682)

[3.1 Research objectives and Research questions 11](#_Toc108074683)

[Managerial problem 11](#_Toc108074684)

[3.2 Data collection plan 11](#_Toc108074685)

[Advantages and disadvantages of the data 13](#_Toc108074686)

[3.3 Variables description 13](#_Toc108074687)

[3.3.1 Independent variable 13](#_Toc108074688)

[3.3.2 Outcome variable 14](#_Toc108074689)

[3.3.3 Moderators 14](#_Toc108074690)

[3.4 Data analysis plan 15](#_Toc108074691)

[3.4.1 Overview 15](#_Toc108074692)

[3.4.2 Sentiment Analysis 15](#_Toc108074693)

[3.4.3 Linear Probability Model 16](#_Toc108074694)

[3.5 Limitations 18](#_Toc108074695)

[3.6 Conclusions 19](#_Toc108074696)

[Chapter 4: Research Results 20](#_Toc108074697)

[4.1 Recap of research design and methods 20](#_Toc108074698)

[4.2 Descriptive statistics 20](#_Toc108074699)

[4.2.1 Sample size 20](#_Toc108074700)

[4.2.2 Distribution of variables 20](#_Toc108074701)

[4.2.3 Graphical representation 21](#_Toc108074702)

[4.3 Results 23](#_Toc108074703)

[4.4 Robustness check 24](#_Toc108074704)

[Chapter 5: Conclusion 25](#_Toc108074705)

[5.1 Summary of research design and methods 25](#_Toc108074706)

[5.2 Summary of results 25](#_Toc108074707)

[5.3 Contributions to theory 26](#_Toc108074708)

[5.3 Contributions to practice 27](#_Toc108074709)

[5.4 Limitations 28](#_Toc108074710)

[5.5 Implications for future research 29](#_Toc108074711)

[Bibliography 30](#_Toc108074712)

# Chapter 1: Introduction

## 1.1 Overview

The present-day global business climate constantly forces companies to be innovative in every approach to stay prevalent. The Internet revolution is shifting physical business places onto virtual ones and creating new competitiveness sources for global competitors. One of them is the longitudinal concept of online brand communities (OBCs) - the mixture of physical brand communities and Web 2.0 technologies (Martínez-López et al., 2016). By definition, “A brand community is a specialized, non-geographically bound community, based on a structured set of social relationships among admirers of a brand” (Muniz & O’Guinn, 2001). Presently, through online brand communities, consumers can share their knowledge, connect with like-minded people, solve problems, and personalize their consumption experiences (Brogi, 2014). Meanwhile, companies may use OBCs to increase brand awareness, facilitate brand co-creation, and understand customer needs (current and potential) by monitoring the information exchanged (Meek et al., 2019). However, OBCs are also communicative platforms for customer uprising. On which customers can widely exchange information about brand negative events (so-called brand crises) (Chang et al., 2013).

Li & Wei (2016) defined brand crises as “unexpected events that threaten a brand’s perceived ability to deliver expected benefits”. Based on brand equity theory, brand crises are conceptualized into two broad types (Li & Wei, 2016). Performance-related crises, also known as product-harm crises, commonly involve defective products and negatively influence brands' ability to carry out functional outcomes (Dutta & Pullig, 2011). Values-related crises are caused by social and ethical issues surrounding the brand’s values (Dutta & Pullig, 2011). Between the two common types, online brand communities most frequently face product-harm crises – discrete events in which products are found defective and affect brand equity and market performance (Cleeren et al., 2017).

The effect of product-harm crises has been studied in a wide range of industries, and Automobile is among the most popular ones ( Rubel et al., 2011; Kalaignanam et al., 2013; Borah & Tellis, 2016; Li & Wei, 2016; Eilert et al., 2017; Liu et al., 2017). Past studies into the Automobile industry typically use product recalls to represent or operationalize product-harm crises. Thus, in this study, “product recalls” and “product-harm crises” are used interchangeably. However, though a product recall often results from a product-harm crisis, this is not a necessary condition due to tolerable imperfections or the legal environment involved (Cleeren et al., 2017). Moreover, in the US, the number of Automobile product recalls is greater than that in other industries combined (Chen et al. 2009). Thus, the automobile industry is an ideal context to study the product-harm crises. We prefer using the well-formulated knowledge and data from this popular industry to explore the crises effects on online brand communities.

As Online Brand Communities are becoming a source of competitiveness for many companies, their success has drawn great attention from marketing professionals and academics (McAlexander et al., 2002; Lin, 2008; Marzocchi et al., 2013). Prior literature denoted two attributes for this accomplishment. The nature of these communities (the commonalities comprise OBCs existence, OBCs classification, and OBC member’s interaction) on the one hand. And *consumer engagement* and the consequences thereof for brands, on the other hand (Woisetschläger et al., 2008). A comprehensive definition has been introduced by Brodie et al. (2013): consumer engagement is a set of personalized interactions between consumers and/or with the brand. With the confirmation from several studies, customer engagement in an OBC is likely to encourage participation in the OBC and increase satisfaction, trust, and commitment to the OBC (Casaló et al., 2007 Woisetschläger et al., 2008, Jang et al., 2008). Notably, while consumer engagement, by nature, is constructive (produce informative content, rally and support favorite brand), others may engage in a more distrustful manner (betray, give misleading advice) (Brodie et al., 2013). Therefore, members engagement and attitude bilateral correlated with the user-generated content on OBC (Mishra & Sharma, 2018).

User-generated content (UGC) is defined as original, brand-specific content created by one or many customers and published on various online platforms (Bowen & Ozuem, 2019). This study considers the volume and sentiment of user-generated content as measurable variables representing the constructs of *consumer engagement* in online brand communities. Past knowledge implied that, in the event of brand crises, the volume of user-generated content tends to surge(Yang et al., 2015)**.** And for a strong and trusted brand, health-related crises generate a larger number of positive than negative sentiments(Mishra & Sharma, 2018). However, the studied brand was in the food & beverage industry, and the online platform was Facebook. In this research, we want to examine whether a different result might happen among mid-level brands in the Automobile industry, focusing on the Reddit community.

We also study how two consumer types (loyal and experienced versus inexperienced consumers) moderate this relationship. Specifically, we want to test whether the influence of product recall on positive user-generated content volume (H1) will be more pronounced among the loyal, experienced consumers. According to Schmalz and Orth (2012), brand commitment is claimed to attenuate consumers’ reactions to negative information. On the other hand, Germann et al., 2014 suggested that negative information could be augmented by brand commitment. For instance, experienced consumers express deep disappointment or demonstrate loyalty and defense for the brand. We categorized online members based on their brand commitment level. And the measurement applied for determining brand commitment level is the number of weeks during which the author’s account was active before the crisis.

Next to brand commitment, media coverage also interacts with the negative impact of product-harm crises. (Liu & Shankar, 2015). Thus, we want to also study the moderating effect of the “media coverage” as a product-recall characteristic in the Automobile industry. Jolly and Mowen (1984) suggested that media reports are more trustworthy to consumers than firms' information about negative events than conventional means. Indeed, by making negative events more salient, media may either worsen the recalling firm’s performance (Ahluwalia et al., 2000) or create more recalling alerts (Berger et al., 2010). In this research, we evaluate the media’s coverage of the crisis via the users’ *reachability* towards those news sources. Further, media coverage through social media platforms created the so-called secondary crisis communication. “Secondary crisis communication” is the ability to share and forward crisis information to respective online communities (Shi et al., 2014). Additionally, social media allows consumers to have “followers”, those who subscribe and receive posts, as their audience. As a result, complexity in consciousness occurs when consumers attempt to impress and keep to the expectations of others (Zheng et al., 2020). People may then modify the content themselves (Zheng et al., 2020). Thus, the percentage of User-generated content with positive sentiments tends to decrease accordingly, which serves as the reason for our study.

## 1.2 Research objective and Research question

Overall, this research project aims to study the effect of product recalls on consumer engagement in online brand communities, focusing on the Automobile industry. In addition, two indirect objectives are: to investigate the moderating role of media coverage over consumers' behavior and examine the differential effects crises have on the online behavior of loyal, experienced consumers versus less experienced ones.

Hence, the main research question is formulated as follows.

*“To what extent do brand crises affect consumer engagement in online brand communities and whether this relation is moderated by brand engagement and media coverage within the Automobile industry?”*

To experimentally answer the question, the corresponding hypotheses were initiated:

* + **H1:** The occurrence of product-harm crises decreases the probability of positive user-generated content.
  + **H2:** Greater number of authors’ active weeks pre-crisis amplifies the negative influence of product-harm crises on the probability of positive user-generated content (H1).
  + **H3:** Greater media coverage amplifies the negative influence of product-harm crises on the probability of positive user-generated content (H1).

Diagram

Description automatically generated



Figure 1. Illustration for relations between the IV, DV, and Moderators

## 1.3 Empirical context and Research design

We aim to study the effect of product recalls on consumer engagement in online brand communities, focusing on the Automobile industry. The window of analysis covers the before and during each crisis period while neglecting the after phase. Five automobile companies facing crises from 2017 to 2019 are analyzed, including General Motors Co (GM), Lexus, McLaren Group, Tesla Inc (formerly Tesla Motors Inc), and Volkswagen UK. Our dataset includes information about seven product-harm crises, along with consumer discussions in 23 Reddit communities over the timeframe denoted.

Our study follows a secondary data analysis approach involving both quantitative and qualitative data. We use deductive logic to reason our findings. The Linear Probability Model and Sentiment analysis are the statistical methods we used to test the hypotheses formulated above. And we utilize R programming language and RStudio to manipulate the datasets and perform the model.

## 1.4 Relevance of study

This study contributes to the literature on the impact of negative publicity on consumer behavior in virtual platforms, including social media and online communities. Next, we added further empirical knowledge to literature in marketing, sociology, and network science that study the factors leading to the success or failure of OBC. Considering the interaction effects, we also examine the moderating roles of media coverage. Practically, media management skills are highly crucial for a brand to monitor the consequences before, during, and after the crisis (Eilert et al., 2017).  Knowledge about online consumer engagement and brand commitment in the event of crisis help firms to improve Crises Management and develop the Online brand community into a source of competitiveness.

# Chapter 2: Literature Review

## 2.1 Brand crises and product–harm crises

According to Li & Wei (2016), the proposed definition of brand crisis goes as follows: “brand crises are unexpected events that threaten a brand’s perceived ability to deliver expected benefits”. In order to categorize brand crises into different kinds, researchers base their foundation on the perspective of two major theories, which are the attribution theory and brand equity theory (Li & Wei, 2016). From the perspective of attribution theory, Coombs (2007) has come up with three types of crises with the foundation of crisis responsibility’s attributions. First is the victim crisis, such as workplace violence and natural disasters, whose crisis responsibility’s attributions are weak, and the brand is seen as the victim in this adverse event (Coombs, 2007). Secondly, an accidental crisis consisting of crisis responsibility’s attributions staying at the minimal level has such events as accidents and product-harm crises related to a technical error (Coombs, 2007). Additionally, the brand’s crisis is uncontrollable (Coombs, 2007). Finally, an intentional crisis, which is believed to happen purposely by the brand, consists of “very strong attributions of crisis responsibility” and includes such accidents related to human error and organizational misdeed (Coombs, 2007).

Moreover, based on the foundation of the brand equity theory perspective, brand crises are conceptualized into two broad types (Li & Wei, 2016). Performance-related crises, also known as product-harm crises, commonly happen with the involvement of defective products, which might negatively influence the ability to carry out functional outcomes of brands (Dutta & Pullig, 2011). Values-related crises have a different direct involvement of product attributes as the above, but they are caused by social and ethical issues surrounding the brand’s values (Dutta & Pullig, 2011). The above type of crisis is more prone to question the brand’s capability to bring benefits from the symbolic and psychological perspective (Dutta & Pullig, 2011). Overall, performance-related crises generate impacts on the brand’s confidence towards functional benefits and symbolic benefits for values-related crises (Dutta & Pullig, 2011).

With the foundation of brand crises definition and types from different perspectives, several pieces of research have been conducted with various aspects to understand the effects such as product recalls, brand scandal, brand misconduct, and brand failure (Liu et al., 2017; Huber et al., 2010; Michelle et al., 2006; Cheng et al., 2012). In most previous academic papers, the main research and findings were concentrated on the impacts of brand crises on the marketing finance interface of brands, advertising expenditure, stock price, brand’s reputation, and learning curves of brands on how to handle the crisis with as least damages as possible (Haunschild & Rhee, 2004; Gao et al., 2015; Ruble et al., 2011). Thus, with the effects of different stakeholders, which have been studied and researched along with brand crises, the effects of brand crises on online brand communities receive little attention and are still open for research.

Considering the main research objective is to develop further understanding of the effects of brand crises on online communities, product-harm crises appear to be the suitable type of crises. Indeed, the participation of consumers in a brand online is not only about seeking useful information but also about a brand community’s requirements based on product characteristics (Shiyong et al., 2022). Brands that have product feature-specific functions can emphasize their functional values and provide information, functions, and intangible values for customers later on (Shiyong et al., 2022). Product-harm crises are considered performance-related crises, which could generate both functional and symbolic impacts on customers, directly affecting the brand's online communities.

## 2.2 Product recall in the Automobile industry

Numerous research papers allocate their concentration to the effects of product-harm crises in a broad range of industries (Cleerence et al., 2017). Automobile appears to be a famous and rich topic for many researchers (Cleerence et al., 2017). In essence, in the automobile industry, a product recall typically happens to a defect in one or more parts, which could indirectly relate to one specific feature. (Liu & Shankar, 2015). It should be noteworthy that a product-harm crisis is often the reason, but not necessarily the condition for the product recall in the distribution system due to tolerable imperfections of the legal environment (Cleerence et al., 2017).

The impacts generated by product-harm crises on brand equity are researched carefully as to the fragility of brand equity since it is founded on consumers' beliefs (Dawar & Pillulta, 2000). Thus, as product-harm crises, customers can access new but possibly negative information about brands, which might cause a significant shift outside of management control (Dawar & Pillulta, 2000). Product-harm crises appear to be one of the most popular topics for study and research of marketing crises. According to Clark’s findings (1988), product-harm crises consist of the following characteristics: (1) product-harm crises act as a threat to marketing goals, (2) they lead to a decrease in the marketer’s ability to execute tasks relating to controlling or directing the marketing environment, and (3) when coping with product-harm crises, both the period for the decision-making process and the response time span is short.

         Additionally, product-harm crises are categorized into two kinds based on the origins and degree of each adverse event falling into this crisis, which are initiated by firms and government separately (Gao et al., 2015). For recalls initiated by the firm itself, the manufacturer is the one who decides whether and/or when to issue the recall after executing their inspection procedures to check for the option of safety defects (Gao et al., 2015). Meanwhile, the government (or NHTSA – National Highway Traffic Safety Administration), the investigation process takes a longer period to fully execute and analyze the situation (Gao et al., 2015).

Nevertheless, papers about product recalls in Automobile mainly look into the effects on the stock market, brand reputation and learning curves, suggestions on how to alleviate damages to the least with product recalls with timing, advertising expenditure (Haunschild & Rhee, 2004; Gao et al., 2015; Ruble et al., 2011; Eilert et al., 2017). Therefore, the area for researching the impacts of brand crises on online brand communities is widely open for discussion in the field of the Automobile industry. Combined with a great amount of prior research, the foundation for studying Automobile brand crises and online communities appears to be prospective.

## 2.3 Online brand communities

During the Web 1.0 era, Online brand communities were first constructed by companies or independent consumers through portals (Jang et al., 2008). Only until the dawn of the Social Web and social media platforms did users and companies start to participate and organize in OBC (Laroche, Habibi, Richard, & Sankaranarayanan, 2012).

Four attributes led to OBC’s success, according to McWilliam (2020): “A forum for sharing common interests; A sense of place with behavioral codes; Agreeable and stimulating dialogues that lead to relationships based on trust; A setting where everyone is encouraged to participate actively, not only an exclusive few”. In a more straightforward description, Fuller, Jawecki, and Muhlbacher (2007) defined an online brand community (OBC) as an Internet-mediated group of people whose members would interact with each other based on a virtual setting. Members in this type of community have a “common interest, admiration and sympathy for a brand” (Brogi, 2014). Muniz and O’Guinn (2001) mentioned three key markers for each online brand community’s social identity. The first marker is the consciousness of kind, indicating members’ connection to many members in the community as well as the brand; therefore, the first marker also involves people’s sense of membership in their community. The second marker is the shared rituals and traditions, which imply members’ effort in maintaining and embracing the community’s history, norms, values, culture, and language. The third marker is moral responsibility which refers to the attitude of members to retain the existing ones and to attract and integrate new members into the community to enhance the overall community experience.

According to Bussgang and Bacon (2020), the success of an online community is determined based on seven essential factors: “(1) shared purpose and values, (2) accessible value consumption, (3) navigable value creation, (4) clearly defined incentives and rewards, (5) accountability, (6) healthy, diverse participation driven by good leadership, and (7) open and objective governance and evolution”. Whether an online brand community can achieve these factors depends on the characteristics of the community itself and its members’ attitudes and interactions.

Besides, Ozuem et al. (2021) suggest four main types of users in online brand communities, including the Judgementalist, the Sugar-coater, the Bias situation, and the Rationalizer. Among them, judgementalists and sugar-coaters contribute to the community’s activities in an irrational way based solely on their thoughts and existing perspectives of the brand since they already know about the brand. On the other hand, bias situators and rationalizers base their contributions on a “reason-based strategy”. The paper also proposes that despite rationality, sugar-coaters and judgementalists are not likely to be influenced by other members’ opinions. Hence, sugar-coaters are superior to judgementalists*.*

## 2.4 Consumer’s engagement and its relationship with volume and sentiments of user-generated content (UGC) on an online brand community

As presented, Online Brand Communities (OBCs) are becoming a source of competitiveness for many companies. Consumers, through online communities, can perform knowledge-based and interactive communication, connect with like-minded people, make more informed purchase decisions, and personalize their consumption experiences (Brogi, 2014). This has drawn the attention of marketing professionals and academics to study the system that contributes to OBC’s success (McAlexander et al., 2002; Lin, 2008; Marzocchi et al., 2013). Prior literature denoted two main focuses of this topic. On the one hand, the nature of these communities (the commonalities comprising OBCs existence, OBCs classification, and OBC member’s interaction) and consumer engagement and the consequences thereof for brands, on the other hand. (Woisetschläger et al., 2008).

Despite operating at a theoretical level, the importance of user engagement is the decisive factor to allocate the emphasis (Brodie et al., 2013, Wirtz et al., 2013). In terms of the brand community’s success, there has been little attention paid at a practical level (Casaló et al., 2007, Kuo and Feng, 2013, Relling et al., 2015). A comprehensive definition has been introduced by Brodie et al. (2013): consumer engagement is distinctive interactions between consumers and/or with the brand. The definition helps differentiate engagement from participation and involvement concepts (Brodie et al., 2013). And it confirms findings on Woisetschläger et al.’s findings on engagement outcomes.

With the confirmation from several studies, customer engagement in an OBC has the tendency to encourage participation in the OBC continuously and increase satisfaction, trust, and commitment to the OBC (Casaló et al., 2007 Woisetschläger et al., 2008, Jang et al., 2008). As a result, the user-generated content includes positive attitudes towards the brand, recommendations among members, and loyalty expression in their purchasing decisions.

By definition, user-generated content is considered as the published materials of content on various online platforms with numerous users (Bowen & Ozuem, 2019). User-generated content is noted to be any form of content that is published freely on an associated online platform by users of a serivce and system (Bowen & Ozuem, 2019). We consider the volume and sentiment of user-generated content to be measurable variables representing consumer engagement. Moreover, different types of consumers would generate different viewpoints for user-generated content (Yang et al., 2015). Moreover, though the number of positive sentiments is larger in product-harm crises, negative sentiments should also be reviewed due to the crisis's contrasting impacts (Mishra & Sharma, 2018).

## 2.5 The moderating role of brand commitment

The term brand commitment is most commonly used to describe the behavioral point of view, measured in repurchase rates, and the attitude, measured by purchase intention (Dubois & Laurent, 1999). This does not paint the whole picture as other studies show how it lacks the cognitive and affective attitudes of customers, which encompass the emotion and sentiments of consumers (Dick & Basu, 1994).  In this sense, brand commitment can be better redefined as the degree to which the consumers maintain a positive attitude towards the brand and how likely they intend to buy it in the future (Dick & Basu, 1994)

We want to study how brand commitment level, reflected by two types of consumers, affects customers’ responses to a product recall. On the one hand, brand commitment is claimed to attenuate consumers’ reactions to negative information (Schmalz & Orth, 2012). On the other hand, negative information could be augmented. For instance, loyal, experienced consumers express deep disappointment (Germann et al., 2014). However, it is also noteworthy to mention that the high-commitment consumers have been seen to counter-argue and discount negative information (Ahluwalia et al., 2000).) Furthermore, the main reason to investigate brand commitment as a moderator is that the negativity effect is absent at higher levels of responsibility. Still, positive information is more preferred than negative information for different customers at different commitment levels.  This ambiguity of the negativity between the different studies shows how much commitment moderates communication effects. Hence, we will use the differences between loyal, more experienced customers and compare them to less loyal consumers.

## 2.6 Media coverage as a characteristic of brand crises

Next to brand commitment, event severity, prior customer impression, and media coverage also interact with the negative impact of product-harm crises. (Liu & Shankar, 2015). We focus on media coverage since it is relevant to our dependent variable - user-generated content, and limited study attention has captured this crisis characteristics (Cleeren et al., 2017). By making negative events more salient, media may either worsen the recalling firm’s performance (Ahluwalia et al., 2000) or increase the awareness of brands being recalled (Berger et al., 2010). Practically, media management skills are highly crucial for a brand to monitor the consequences before, during, and after the crisis (Eilert et al., 2017).

Further, media coverage through social media platforms created the so-called secondary crisis communication. The term is unfamiliar in the conventional mass media as “secondary crisis communication” present the ability to share and forward the crisis information to respective online communities (Shi et al., 2014). Additionally, social media allows consumers to have “followers”, those who subscribe and receive posts, as their audience. As a result, complexity in consciousness occurs when consumers attempt to impress and keep to the expectations of others (Zheng et al., 2020). Thus, people may modify the content themselves (Zheng et al., 2020). It is necessary for further empirical research on the moderating effects of media coverage.

## 2.7 Conclusion

The literature review serves as preliminary research for our Bachelor Project on the theme of “How do brand crises make or break consumer communities online?”.  Online brand communities are communicative platforms where consumers, who share common interests and sympathy for a specific brand, can exchange knowledge and express attitudes toward such brands. Members can be categorized into four groups based on their behaviors: Judgementalist, Sugar-coater, Bias situation, and Rationalizer. These users, to different extents, can be affected by different types of brand crises. Based on the two major perspectives, the attribution theory and brand equity theory perspective, brand crises can be categorized into several types. The above phenomenon has opened the opportunities to develop research for discovering the effects of brand crises on online brand communities.

Among various types of brand crises, the product-harm crisis is our main focus, also called product recalls, as the independent variable. Next, consumer engagement, which consists of the volume and sentiment of user-generated content, acts as the dependent variable. We believe consumer engagement and its outcomes are decisive factors in the well-being and sustainability of an OBC. As for the moderators, brand commitment and media coverage are our choices. The inter-related relationship between brand commitment and online community commitment are proven and positively correlate to each other (Kim et al., 2010). But still, controversial impacts on different types of customers in the case of a negative event are open for our research. Meanwhile, media coverage also shows diverse effects of a product-harm crisis (Liu & Shankar, 2015). Nevertheless, Linear Regression model and are our research methods. This secondary analysis knowledge from our past Research Project course, combined with the use of R programming tool, help us discover the relationship between brand crises and online brand communities.

# Chapter 3: Research Design

## 3.1 Methodology

As mentioned in section 1.2, this research project aims to study the effect of product recalls on consumer engagement in online brand communities, with a focus on the Automobile industry. Additionally, we investigate the moderating role of media coverage on consumers' behavior; and examine the differential effects of crises on the online behavior of experienced consumers versus less experienced ones.

Hence, the main research question is formulated as follows.

*“To what extent do brand crises affect consumer engagement in online brand communities and whether this relation is moderated by brand engagement and media coverage within the Automobile industry?”*

To experimentally answer the question, the corresponding hypotheses were initiated:

* + **H1:** The occurrence of product-harm crises decreases the probability of positive user-generated content.
  + **H2:** Greater number of authors’ active weeks pre-crisis amplifies the negative influence of product-harm crises on the probability of positive user-generated content (H1).
  + **H3:** Greater media coverage amplifies the negative influence of product-harm crises on the probability of positive user-generated content (H1).

### Managerial problem

The three hypotheses are tested to better study the impact of negative publicity on consumer behavior in online communities. As discussed in chapter 2, there is theoretical knowledge about how brand crises affect consumer behavior and their negative effects on brand values, revenues, and marketing strength. Meanwhile, OBC is a growing concept, and there is a need to answer whether members will positively or negatively react to the occurrence of brand crises. Additionally, the findings take into account the moderating effects of media coverage and brand commitment level. Knowledge about online consumer engagement and brand commitment in the event of crisis help firms to improve their approach to Crises Management and develop their Online brand community into a source of competitiveness.

## 3.2 Data collection plan

We aim to study the effect of product recalls on consumer engagement in online brand communities, focusing on the Automobile industry. The analysis window covers the before, during, and after the crisis period. Five automobile companies facing crises **from 2017 to 2019** are analyzed, **including General Motors Co (GM), Lexus, McLaren Group, Tesla Inc (formerly Tesla Motors Inc),**and Volkswagen UK. Our secondary dataset includes information about seven product-harm crises, along with consumer discussions in 23 Reddit communities over the timeframe denoted (Baumgartner et al., 2020)

The optimal data that could be used in the studies are those that can identify all our variables immediately. For loyalty, conducting surveys would give the most insights since it is more in-depth, qualitative data (Punniyamoorthy & Prasanna Mohan Raj, 2007). Substitutes to these are those from online databases, such as YouGov, which is a secondary data collection instead. YouGov is a market research company collecting data on brand information, such as brand strength (YouGov, 2020). The data collected can help analyze the changes in consumer behavior and perception (Backhaus & Fischer, 2016). The final data used is all the user interactions on online platforms, including company websites and (more significant) social media such as Twitter, Facebook, Meta, and Reddit. An analysis using this would give us the volume and sentiment of all OBC interactions. To encapsulate all user behavior, we would need to conduct text mining of user posts, comments, and several interactions for sentiment analysis. These types of data have been used in methodologies of similar studies to examine the popularity of a brand (Khobzi and Teimourpour, 2014) and measure the change in consumer attitude towards a brand after a crisis (Mishra & Sharma, 2019).

We use secondary data to achieve the objectives of this research. The first dataset consists of information about user-generated content relating to the issues of a product recall on Reddit from 2017 to 2019 (Baumgartner et al., 2020). Each record (post/comment) is associated with a risk ID, risk story ID, the news date, related issues, related topic tags, related UNGC principles, country, severity, novelty, reach, company, subreddit, year, and self-text. The risk ID and risk story ID are unique identifiers for the crisis and the news piece covering the crisis, respectively. The news date implies the date when the news covers the crisis (*RepRisk AG ESG News Dataset*, 2019). Each record can also be associated with related issues, topic tags, and UNGC (United Nations Global Compact) principles, including ten principles from human rights, labor, and environment, to anti-corruption (UN Global Compact, 2022). The severity specifies the level of seriousness of the crisis consequences, with three values (1, 2, 3) implying “low,” “medium,” and “high,” respectively. Each record is also associated with the reach level, either international (reach = 3) or national and local (reach = 1 and 2). In terms of the company: Our research focuses on five companies in the automobile industry, namely General Motors, Lexus, McLaren Group, Tesla, and Volkswagen UK. The dataset contains the content of posts and comments for each record in a word format. Each one belongs to an online community on Reddit, specified under “subreddit.” The second dataset includes the information regarding the subreddit, the reference scandal date, the exact date and time that the post/comment was created, the author's name, the ID, and the name of the company for each of the records. We will conduct the research using seven elements from these datasets, including the risk ID, news date, country, severity, reach, reference scandal date, and self-text. Overall, the final dataset we used contains both quantitative and qualitative data, with 1,714,725 records (*RepRisk AG ESG News Dataset*, 2019)

### Advantages and disadvantages of the data

The given data sets contain longitudinal data collected for three years, enabling a potential increase in the statistical power and a greater range in conditional probabilities compared to cross-sectional data. The use of secondary data also allows for potential justification of future research on the same matter (Pederson et al., 2020). Additionally, the YouGov database contains data regarding the weekly brand ratings with more than 700 responses, which can help reduce sampling errors. Moreover, as mentioned above, the data can be dissected weekly, allowing for detecting changes in brand perception triggered by single events such as certain brand crises. The collective use of user-generated content on Reddit and other online platforms such as Facebook and Twitter can increase the objectivity of the input data, improving the validity and reliability of the study.

On the other hand, there are several limitations to the data used in our research. Regarding the Reddit data, a number of records contain missing information in some of the fields, for instance, the information about countries, related topic tags, and UNGC principles, which may affect the validity of the outcome. Specifically, missing data can lead to lower statistical power, biased estimations, and lower sample representativeness (Kang, 2013). In addition, we focus on only three years’ data from 2017 to 2019 of five companies in the automobile industries, which may impose a threat to the reliability of the research in case of potential continuation for future research due to the possibility of obsolescence of the data and the changing nature of technologies and online communities.

## 3.3 Variables description

### 3.3.1 Independent variable

*The occurrence of product-harm crises* is the extent of whether brand crises happen to companies or not. In this study, we consider the occurrence of product-harm crises as a binary variable. Throughout the provided timeframe, we will base our analysis of the brand crises’ occurrence on the news date about the crises and the timestamp of content creation within the brand’s online communities. For each timestamp of the crisis event, based on the dates of news released about the brand crises, we will mark it down and evaluate whether the above timestamp coincides or overlaps any period of the timeframe of user-generated content. The period starting from the overlapped data of the news date and content creation date is deemed to be during the troubles. Thus, we will be able to identify different patterns of product-harm crisis occurrence, then study the impacts on related variables.

### 3.3.2 Outcome variable

As for the dependent variable, we consider the volume and sentiment of user-generated content to be outcome variables illustrating the constructs *consumer engagement* in online brand communities.

*Volume of user-generated content* as a variable is measured as the magnitudes of user-generated content that community members create in different phases, before, during, and after the occurrence of brand crises. To measure the volume of UGC, we rely on descriptive information about posts and recorded activities of all community members to count the number of content-generated rows in the communities’ dataset. The information relating to scandals date, content created date, author, subreddit topic, and companies in each data row will help define the number of data rows via collecting similar UGC due to the occurrence of brand crises of a particular company. Also, some findings indicate that in the event of brand crises, the UGC volume tends to go upwards (Yang et al., 2015). Accordingly, our study of the Automobile industry will help us define the pattern of UGC distribution with the occurrence of product-harm crises and its effects on the volume of UGC.

*Sentiment of user-generated content* is a binary variable which identifies whether communities’ members contribute UGC to the platform with the attitude being positive or negative. Our approach is to use the sentiment analysis with R to evaluate whether UGC is positive or not. The sentiment index will be the benchmark for the evaluation, which is equal to the numeric difference of positive words count minus the negative ones. Finally, if the sentiment index of a UGC is equal or larger than 0, it is considered as positive. And if the sentiment score is less than 0, it is marked as negative.

### 3.3.3 Moderators

Number of active weeks pre-crisis. Schmalz and Orth (2012) claimed that brand commitment attenuates consumers’ reactions to negative information. The measurement to determine brand commitment level is the number of weeks the author’s account was active before the crisis. Hence, within the product-harm crisis occurrence, the volume and sentiment of UGC are possibly affected by the brand commitment level to a certain extent. Therefore, to identify the moderating effects of consumer types, our approach evaluates the volume of positive UGC based on the author’s account active weeks pre-crises. To calculate the active weeks of each author, we firstly grouped our dataset by subreddit. Then we subtract the oldest UGC timestamp of that account from the related crisis news date. And finally, we divide the number of days by seven and take the integer part.

Media coverage level. We control the media coverage by evaluating the reachability level of the crises via social and traditional media channels on the scale regarding international, national, and local levels. According to Liu & Shankar (2015), media coverage also contributes to the negativity of product-harm crises. With the information about adverse events and conventional means, media reports are claimed to gain more trust from consumers than firm-initiated channels (Jolly & Mowen, 1984). Media coverage is measured via the news’ reachability on three levels: 1 (regional), 2 (national), and 3 (international). Mitchell and colleagues (2018) claimed that news close to home tends to gain more interest from people. Around 86% of participants follow national news or 76% about their cities, while the attention is 57% on international communication (Mitchell et al., 2018). Thus, the media coverage level relating to product-harm crises would need to be further studied since the effects on each level of coverage are possibly different. With each additional level of media coverage, the volume of positive UGC is considered and evaluated to define the influences of media coverage on positive UGC distribution.

## 3.4 Data analysis plan

### 3.4.1 Overview

Overall, our study follows a secondary data analysis approach and involves both quantitative and qualitative data from the Automobile industry. Our analysis window covers the before and during phase of each crisis. Five automobile companies facing crises from 2017 to 2019 are analyzed, including General Motors Co (GM), Lexus, McLaren Group, Tesla Inc (formerly Tesla Motors Inc), and Volkswagen UK. Our dataset includes information about seven product-harm crises and consumer discussions in twenty-three Reddit communities over the denoted timeframe. We use deductive logic through testing the described hypotheses to reason our findings. The Linear probability model and Sentiment analysis are the statistical methods we used to test the hypotheses formulated above. We set the significance level at 5% (Quinn, 2021). And to manipulate our datasets and perform the model, we utilize programming tools R and RStudio. The following paragraphs describe our methodology choices in detail.

### 3.4.2 Sentiment Analysis

When people read a pile of text, they often emotionally characterize that text using the understanding of the emotional intent of words. In this study, so as to programmatical evaluate whether User-generated content is positive or negative, we performed Sentiment analysis with R. To begin with, we used the **tidytext** package to transform User-generated content into tidy text format - a table with one token per row (Silge & Robinson, 2017). Each token we stored represents a meaningful English word extracted from the corresponding user’s post or comment. Secondly, we use *get\_sentiment()* function along with bing lexicon to classify words into binary (positive and negative) categories (Zhang et al., 2018). We then join (using *inner\_join()*) the text data from User-generated content with the lexicon and count the number of positive or negative words involved. Thirdly, we use *pivot\_wider()* function from **tidyr** package to split negative and positive UGC into different columns and compute the sentiment index (sentiment = positive – negative) (Silge & Robinson, 2017). Lastly, based on the sentiment index, we classified User-generated content to be positive if its Sentiment score is equal to or larger than 0, and negative otherwise. Often the post or comments are in sentence-sized or paragraph-sized text, which suits this sentiment analysis more than big-chunk text (those in books, reports, or literature) (Silge & Robinson, 2017). Figure 2 below summarizes the typical text sentiment analysis flowchart that we followed. In the original paper, the final step involves Visualizations using ggplot2 package. However, we omitted the part since it does not serve the objective of this research.

Diagram

Description automatically generated

*(Source: Chapter 2 Sentiment analysis with tidy data | Text Mining with R by Silge & Robinson, 2017)*

Figure 2. A flowchart of a typical text analysis that uses tidytext for sentiment analysis

However, there are certain drawbacks present in our Sentiment analysis. First, the lexicons from tidytext package contain only English words' measurements (Silge & Robinson, 2017). Hence, we cannot touch upon the User-generated content in smaller regional Online-brand communities where languages are spoken. Second, the lexicon does not capture every English word since many of them are neutral (Silge & Robinson, 2017). The method also neglects qualifiers before a word, for example, “not bad” or “no good”. Third, there is content in pictures, sounds, or videos format, which cannot be analyzed using this approach. Fourth, we gave an absolute polarity (either 1 or -1) to the Sentiment score of each word. This is intuitively imprecise since some words can have different degree of emotion (e.g “fine” vs “excellent”). And finally, sentimental analysis faces other challenges arising from tone, irony, sarcasm, idioms, and negations (Inc, 2021).

### 3.4.3 Linear Probability Model

In session 13 of the Research Project course, block 7, Quinn (2021) introduced the Linear probability model as a regression model where the outcome variable is a binary variable, and one or more explanatory variables (binary or continuous) are used to predict the outcome. Explanatory variables can be binary or continuous. The model is suitable for our study for the following reasons. First, after the classification between positive and negative sentiment from the Sentiment Analysis above, we now treat our DV: *user-generated content* (UGC.n)as a dummy binary variable (positive = 1 and negative = 0). Meanwhile, our IV is the binary occurrenceof *a brand crises event*. Our experiment also accounts for the moderating variables: *media coverage* and *consumer types*. The two moderators are measured by the range of news (international reach or national reach) and brand commitment level (the number of days in which the author’s account was active). The steps for testing the Hypotheses are discussed as follows.

As for H1, we test whether, in the occurrence of product-harm crises, the percentage of positive user-generated content is higher than that for negative ones. The initial sentiment (DV) segmentation step is illustrated in detail in section 3.4.3, Sentiment analysis. Next, as our analysis window covers the before and during crisis phase, we exercised lubridate R package and marked a timestamp for each crisis event. The IV *occurrence of crisis event* (CR) is then binary manipulated (*crisis -* intervention and *no*\_*crisis* - control). Afterward, we run the Linear Probability model on R: **lm(formula = UGC.n ~ CR, data)**. As a result, the intercept will represent the chances of positive user-generated content without brand crises (Quinn, 2021). Meanwhile, the coefficient 𝛽 will represent the additional chances of positive user-generated content in the occurrence of brand crises (Quinn, 2021). H1 is confirmed based on two criteria: the regression coefficient 𝛽1 is positive, and the effect is significant (p-value is smaller than 0.05) (Quinn, 2021).

Considering H2, we want to test whether the occurrence of brand crises (IV) on the percentage of positive sentiment in the user-generated content is positively moderated by *the number of weeks the author’s account was active before the crisis* (AW). The call for this Linear Probability model on R: **lm(formula = UGC.n ~ CR \* AW , data)** (Quinn, 2021). To interpret the result, the intercept captures the impact of *no\_crisis* with zero active day. The *crisis* coefficient (𝛽1) presents the extra probability of positive sentiment in the occurrence of a *crisis*, compared to *no\_crisis* condition. The *AD* coefficient (𝛽2) presents the extra probability of positive sentiment given by each additional active day in the *no\_crisis* condition. The *crisis:AW (*𝛽3) presents the extra probability of positive sentiment given by each additional active day in the occurrence of *crisis,* compared to *no\_crisis* condition. The effect of *the number of active days of the author’s account* (AW) on the percentage of positive sentiment, in the occurrence of crisis event is determined by summing coefficients 𝛽2 and 𝛽3. Finally, H2 is confirmed based on two criteria: the total value of 𝛽2 and 𝛽3 and positive, and the effect is significant (p-value is smaller than the significance level of 0.05).

The same approach is applied for H3, where the moderator *media coverage* (reach) is measured from 1 to 3 (regional, national, and international) while its moderation effect is negative. The Linear Probability model we called in R is: **lm(formula = UGC.n ~ CR \* reach , data).** Correspondingly, the intercept captures the impact of *no\_crisis* with no media coverage. The *crisis* coefficient (𝛽1) presents the extra probability of positive sentiment in the occurrence of a *crisis*, compared to *no\_crisis* condition. The *AW* coefficient (𝛽2) presents the extra probability of positive sentiment given by each additional media coverage level, in the *no\_crisis* condition. The *crisis:AW (*𝛽3) presents the extra probability of positive sentiment given by each additional media coverage level in the occurrence of *crisis,* compared to *no\_crisis* condition. The effect of *media coverage* (MC) on the percentage of positive sentiment in the occurrence of a crisis is determined by summing coefficients 𝛽2 and 𝛽3. Finally, H3 is confirmed based on two criteria: the total value of 𝛽2 and 𝛽3 and negative, and the effect is significant (p-value is smaller than the significance level of 0.05) (Quinn, 2021).

However, there are three caveats involved in the Linear Probability model. The predicted probability, in our study, is the chance of positive User-generated content, which can be below 0 or above 1. This is a mathematically impossible prediction (Quinn, 2021). There are also assumptions about linearity between variables. For example, an increase in media coverage level from regional to national (1 to 2), has the same effect as national to international (2 to 3) (Quinn, 2021). And the final drawback relates to the presence of Heteroskedasticity in all Linear Probability models (Quinn, 2021).

## 3.5 Limitations

In our research, we tried to capture as many observations as possible by selecting all companies presented under the automobile category from our secondary datasets (Baumgartner et al., 2020; *RepRisk AG ESG News Dataset*, 2019). However, the choice of companies is limited and cannot represent the whole market. Such a limited sample size may help to make specific conclusions regarding the automobile industry; however, it may result in a low generalization ability when applied to other industries. Future studies can improve upon this by including a wider array of companies and industries. The first thread to validity is the *number of account’s active weeks* variable we use to analyze the level of commitment/loyalty for each community user. This variable is might not the optimal one to explain the construct customer engagement. Moreover, we compute the weeks by subtracting that account's oldest UGC timestamp from the related crisis news date (which is limited to only 90 days). We then divide the number of days by seven and take the integer part. Thus, data on accounts with active weeks larger than 12 (active days larger than 84) is insufficient and unreliable. And finally, in terms of *Media coverage level*, we assume each level of news reachability has the same distance to another. However, this is not the case in reality, since the marginal effect of local reach news might differ from that of continental or global reach news.

Additionally, the chosen analysis methods also impose several limitations. On the one hand, Sentiment analysis can only be used to analyze the content that contains only words, more specifically, English words. The method ignores words in other languages and content in picture, sound, or video formats. Additionally, the sentiment analysis also ignores other hidden verbal elements included in the content, such as irony, sarcasm, tone, idioms, and negations (Inc., 2021). On the other hand, Linear Probability poses drawbacks related to mathematically impossible predictions, assumptions in linearity between variables, and Heteroskedasticity (Quinn, 2021).

(*RepRisk AG ESG News Dataset*, 2019). For instance, Volkswagen UK is associated with three online communities, among which “golfgti” is the most active one with more than 21,000 contents generated within the chosen timeframe; while McLaren Group is associated with only one community with 620 generated contents within the specified period. The activity levels also vary among communities of the same brand (e.g, Volkswagen UK is connected to 3 communities with 119, 32, and 21,000 generated contents, respectively) (*RepRisk AG ESG News Dataset*, 2019). Additionally, the neglected variables, such as demographic variables, may have affected the relationship between our research variables both directly and indirectly as they can interact and impact other factors, such as users’ perceptions and contents (Kwon, 2002). Replicated studies using different sets of online brand communities and demographic elements might not give the same results due to the variance in the scale of our chosen set of communities and the effects of demographics (Golafshani, 2003).

## 3.6 Conclusions

In brief, this research project follows a secondary data analysis approach, with both quantitative and qualitative data involved. We use deductive logic to reason our findings by testing the described hypotheses. In our model, DV: is the volume and sentiment of User-generated content, IV is the occurrence of product-harm crisis, and moderators are the number of active days pre-crisis and media coverage level. The Linear Probability Model and Sentiment analysis are the statistical methods we used to test the hypotheses formulated above. And to manipulate our datasets and perform the model, we utilize programming tools R and RStudio.

On the one hand, this study is beneficial from longitudinal data, which allows for increasing statistical power and potential justification of future research on the same topic. Moreover, the collective use of user-generated content on Reddit, reports, and public databases can strengthen the objectivity of the input data, thereby improving both the validity and reliability of the research. On the other hand, our study focuses on a limited number of companies in the automotive industry within a certain period, which may not correctly represent the whole market and are sufficiently used for future research. The chosen analysis methods are limited in their ability to analyze particular contents and other problems such as mathematically impossible predictions and heteroskedasticity. Lastly, our model measures user commitment level based solely on the number of active days pre-crisis and ignores all other factors that may be related, including the demographic factor.

# Chapter 4: Research Results

## 4.1 Recap of research design and methods

Overall, this research project aims to study the effect of product recalls on consumer engagement in online brand communities, with a focus on the Automobile industry. The research question is formulated as follows:

*“To what extent do brand crises affect consumer engagement in online brand communities and whether this relation is moderated by brand engagement and media coverage, within the Automobile industry?”*

And the corresponding hypotheses were initiated:

* + **H1:** The occurrence of product-harm crises decreases the probability of positive user-generated content.
  + **H2:** Greater number of authors’ active weeks pre-crisis amplifies the negative influence of product-harm crises on the probability of positive user-generated content (H1).
  + **H3:** Greater media coverage amplifies the negative influence of product-harm crises on the probability of positive user-generated content (H1).

We use deductive logic to reason our findings from testing the described hypotheses. The Linear probability model and Sentiment analysis are the statistical methods we used. We set the significance level at 5%. And to manipulate our datasets and perform the model, we used programming tools R and RStudio.

Our datasets contain secondary data, both quantitative and qualitative. The first dataset consists of information about user-generated content relating to the issues of a product recall on Reddit from 2017 to 2019 (Baumgartner et al., 2020). The second data set provides an overview of the risk incident (news) on a monthly basis over the same period (*RepRisk AG ESG News Dataset*, 2019).

## 4.2 Descriptive statistics

### 4.2.1 Sample size

We tested the hypotheses from a sample of five automobile companies facing product harm crises from 2017 to 2019: General Motors Co (GM), Lexus, McLaren Group, Tesla Inc (formerly Tesla Motors Inc), and Volkswagen UK. In total, this sample includes 1,714,725 observations.

### 4.2.2 Distribution of variables

As for the dependent variable, we consider the *volume and sentiment of user-generated* *content* to explain the construct *consumer engagement* in online brand communities. In our sample, when there is no crisis, the positive and negative content percentages are 69.28% and 30.72%, respectively. In contrast, when product-harm crises hit, the proportion slightly shifted to 68.76% over 31.24%. Regarding the independent variable - *the occurrence of product-harm crises*, we treated this variable as a binary data type. Of all the 1,714,725 observations, 887,734 rows are marked with “crisis”, and the other 826,991 rows are marked with “no crisis”. Correspondingly, the ratio is 51.78% “crisis” and 48.22% “no-crisis”.

In terms of moderation effects, the two moderators Number of active weeks pre-crisis and Media coverage level present the following statistics. First, the number of user accounts’ active weeks prior to the crisis within each subreddit ranges from 0 to 12 weeks. The median value is 8 weeks, while the mean value averages at 6.756 weeks. 1st quartile is 0 weeks, and 3rd quartile is 12 weeks. Second, for the Media coverage level when there is no crisis, the percentage of positive UGC associated with local (1) and global (2 and 3) reach news 69,13%, 69,47%, and 69,08%, in the order given. On the contrary, in the event of crises, these figures reduced to 68,38%, 68,89%, and 68,73%, respectively.

### 4.2.3 Graphical representation

The relations between the IV, DV, and Moderators have been illustrated in Figure 1 in section 1.3 Empirical context and Research Design, chapter 1. In this section, we further present three descriptive plots that demonstrate these variables' relationships related to the three hypotheses. For H1, Figure 3 compares positive and negative content percentages between no crisis and crisis conditions. The two columns seem identical to each other. Next, for H2, Figure 4 represents the percentages of positive UGC created by the authors’ account with different active weeks pre-crisis, in both crisis and no-crisis conditions. In general, the proportion of positive content in the event of the crisis remained below that in no crisis event. And the gap between the two columns became smaller when the account’s active weeks pre-crises increased. Finally, for H3, the proportion of positive content among three media coverage levels, between crisis and no crisis conditions, is compared in Figure 5. Similarly, the positive content ratio for crisis conditions is still lower than that for no-crisis. The lowest positive content ratio is observed concerning crisis news with local reach (level = 1). Also, in this group, the gap between the two conditions is larger than crises news in the global reach group (levels = *2* and *3*).

Chart, bar chart

Description automatically generated  
Figure 3. Percentages of positive and negative content in no crisis and crisis conditions

Chart, bar chart

Description automatically generated

Figure 4. Percentages of positive content created by users’ account with different active weeks prior crises

Chart, bar chart

Description automatically generated  
Figure 5. Percentages of positive content among three media coverage levels.

## 4.3 Results

Table

Description automatically generated  
Table 1. Results of Linear Probability Model for testing the hypotheses

Table 1 above reports results from the Linear Probability model in testing the effect of brand crises on consumer behavior in OBCs. The first column (model for H1), which presents results for the main effect, is significant (R2 = 0.00031, p-value < 0.01). The constant equals 0.6928, meaning a 69.28% chance of UGC is positive when there is no brand crisis (p-value < 0.01). The coefficient 𝛽1 (CRcrisis) is -0.005160, suggesting that the occurrence of a brand crisis decreases the probability of users generating positive content by 0.5160% (p-value < 0.01).

We turn to column 2 (model for H2) to test the moderation effect of the *number of author’s account active weeks pre crises (AW)*. This model is significant as well (R2 = 0.000626, p-value <0.01). The intercept captures the impact of “no crisis” condition with 0 active weeks on the probability of positive sentiment, which is equal to a 70% extra chance (p-value < 0.01). Coefficient 𝛽1 (CRcrisis) = -9.160e-03, suggests an extra 0.916% probability of positive sentiment in a crisis, compared to “no crisis” condition. Next, 𝛽2 (AW) = -8.969e-04 suggests that in the 'no crisis' condition, each additional active week decreases almost 0.1% probability of a positive UGC (p-value < 0.01). Meanwhile, 𝛽3 (CRcrisis:AW) = 8.603e-05 suggests that in the 'crisis' group, each additional active week insignificantly increases 0% probability of a positive UGC (p-value > 0.05), compared to each additional active week in the ‘no crisis’ group. Summing 𝛽2 and 𝛽3 gives us no or very limited marginal effect of the *number of active weeks pre-crisis* for the ‘crisis’ group. Moreover, since the interaction term 𝛽3 is already insignificant, the null hypothesis H0: there is no moderation effect of the *number of active weeks pre-crisis* cannot be rejected.

The third column (model for H3) follows the same estimation method but with *media coverage* (or reach*) level* as the moderator. This model is also significant (R2 = 0.000033, p-value <0.01). The intercept reveals the impact of “no crisis” condition with 0 reach level on the probability of positive content, which is equal to 69.55% (p-value < 0.01). Coefficient 𝛽1 (CRcrisis) = - -0.009892, suggests a minus 0.9892% chance of positive sentiment in the occurrence of a crisis, compared to “no crisis” condition (p-value < 0.01). Next, 𝛽2 (reach) = -0.001224 suggests that in the 'no crisis' condition, each additional reach level insignificantly decreases 0.1224% probability of a positive UGC (p-value > 0.05). And 𝛽3 (CRcrisis:reach) = 0.002107 suggests that in the 'crisis' group, each additional active week increases 0.2107% probability of a positive UGC (p-value < 0.05), compared to the additional reach level in the ‘no crisis’ group. Summing 𝛽2 and 𝛽3 gives us 0.883% extra chance of positive UGC for each additional reach level in the ‘crisis’ condition.

## 4.4 Robustness check

The number, size and activity level vary across communities within the same brand and across different brands, assuming that the size and activity level of each community are measured by the number of generated contents within such community. In our Reddit data, each company has a different number of associated OBCs. The number of communities for each company may also be different on other platforms such as Facebook groups. Additionally, there are also variances in sizes and activity levels among communities within the same brand (*RepRisk AG ESG News Dataset*, 2019). Therefore, the same results cannot be guaranteed when using a different set of communities about the same brand. Nevertheless, users can be part of OBCs on different platforms at the same time. Future studies should consider users’ demographics to minimize the sensitivity of the results.

The variance also occurs among communities across different companies (*RepRisk AG ESG News Dataset*, 2019). Hence, changing the set of companies would not drive our results. However, we focused on the automobile industry solely. Hence, the sensitivity of the result can increase when the study includes brands from other industries. Furthermore, given the limitations of our research models, Sentiment Analysis and Linear Probability, the ignored elements’ effects on the text content and the interactions between them were not captured. Adding these elements requires more complex research models, such as Support-Vector Machine and Sarcasm Detection (Srivastava & Bhambhu, 2010; Berasategi, 2020), which may cause a change in our research results. Lastly, our data contains 1,714,725 records in a period of 3 consecutive years from 2017 to 2019, which is considered extensive.

# Chapter 5: Conclusion

## 5.1 Summary of research design and methods

In this study, our goal has been to study the effect of product recalls on consumer engagement in online brand communities, with a focus on the Automobile industry. We additionally investigate the moderating role of media coverage on consumers' behavior and examine the differential effects of crises on the online behavior of experienced consumers versus less experienced ones. Hence, the main research question was formulated:

*“To what extent do brand crises affect consumer engagement in online brand communities and whether this relation is moderated by brand engagement and media coverage, within the Automobile industry?”*

And the corresponding hypotheses were initiated:

* + **H1:** The occurrence of product-harm crises decreases the probability of positive user-generated content.
  + **H2:** Greater number of authors’ active weeks pre-crisis amplifies the negative influence of product-harm crises on the probability of positive user-generated content (H1).
  + **H3:** Greater media coverage amplifies the negative influence of product-harm crises on the probability of positive user-generated content (H1).

We used deductive logic to reason our findings from testing the described hypotheses. The Linear probability model and Sentiment analysis are the statistical methods we used. We set the significance level at 5%. And to manipulate our datasets and perform the model, we used programming tools R and RStudio.

Our datasets contain secondary data, both quantitative and qualitative. The first dataset consists of information about user-generated content relating to the issues of a product recall on Reddit from 2017 to 2019 (Baumgartner et al., 2020). The second data set provides an overview of the risk incident (news) on a monthly basis over the same period (*RepRisk AG ESG News Dataset*, 2019). We tested the hypotheses from a sample of five automobile companies facing product harm crises from 2017 to 2019: General Motors Co (GM), Lexus, McLaren Group, Tesla Inc (formerly Tesla Motors Inc), and Volkswagen UK. In total, this sample includes 1,714,725 observations.

## 5.2 Summary of results

Overall, based on our findings in the previous chapter, we propose the following answer to the research question: The event of a brand crisis decreases consumer engagement in online brand communities. And while brand engagement does not moderate this relationship, greater media coverage reduces the negative influence of product-harm crises on the probability of positive user-generated content.

Results from Sentiment analysis show that, generally, when there is no crisis, the positive and negative content percentages are 69.28% and 30.72%, respectively. And when product-harm crises hit, the proportion slightly shifted to 68.76% over 31.24%. In more detail, hypotheses testing results from Linear Probability Model are summarized in table 2 below.

|  |  |
| --- | --- |
| Hypothesis | Results |
| H1 | **Supported**: the occurrence of a brand crisis decreases the probability of users generating positive content. |
| H2 | **Rejected**: the null hypothesis H0: there is no moderation effect of the *number of active weeks pre-crisis* cannot be rejected. |
| H3 | **Rejected**: each additional reach level increases the chance of positive UGC in the event of a product-harm crisis. |

Table 2. Summary of testing the hypotheses

## 5.3 Contributions to theory

Prior papers concentrated on the impacts of brand crises on the marketing finance interface of brands, advertising expenditure, stock price, brand’s reputation, and learning curves of brands on how to handle the crisis with as least damage as possible (Haunschild & Rhee, 2004; Gao et al., 2015; Ruble et al., 2011). Very limited attention has been given on the direct impact of the brand crisis on online communities. Thereof, our study contributes to the literature on the impact of negative publicity on consumer behavior in these virtual brand communities. Furthermore, from the hypotheses testing results, our study contributes to the existing research in three additional ways.

First, past knowledge has implied that in the event of brand crises, the volume of user-generated content tends to surge(Yang et al., 2015)**.** Our result supports this implication. However, we also showed that the proportion of positive over negative content decreases in the event of product harm crises (H1), which contradicts Mishra & Sharma (2018) findings. Second, we added further empirical knowledge to literature in marketing, sociology, and network science that study the factors leading to the success or failure of OBC. With confirmation from several studies, customer engagement in an OBC is likely to encourage participation in the OBC and increase satisfaction, trust, and commitment to the OBC (Casaló et al., 2007 Woisetschläger et al., 2008, Jang et al., 2008). According to Schmalz and Orth (2012), brand commitment is claimed to attenuate consumers’ reactions to negative information. Meanwhile, Germann et al., 2014 suggested that brand commitment could augment negative information. Yet, our result from testing H2 proposes that *brand engagement* does not impact the relationship between brand crisis and customer engagement. Thus, it supports neither of the two viewpoints. Third, our study reveals that the relationships between brand crisis and customer engagement do not stand separate from those between the brand crisis and the degree of media coverage. Our findings, in the specific context of Reddit communities, indicate that greater media coverage urges online community users to protect their brand through positive content. And this result confirms Eilert et al., (2017) and Zheng et al., (2020) findings.

## 5.3 Contributions to practice

Online brand communities are considered an effective source of competitive advantage for firms because they foster customer interactivity. Firms may use OBCs to increase brand awareness, facilitate brand co-creation, and understand customer needs (current and potential) by monitoring the information exchanged (Meek et al., 2019). In correspond to section 5.2 above, our research findings offer three managerial implications for those who build and maintain online brand communities, particularly on Reddit, and those involved in crises and media management.

First, our results confirm a negative relationship between brand crisis and consumer engagement in the online brand community. Therefore, we suggest managers in the Automobile industry conduct further study into the company’s online brand community, specifically, how the communities interpret the company product recalls. This approach provides companies means of assessing the brand community's value and effective plans to initiate any product recall. For example, the company may use different advertising campaigns during and after a recall, strategically initiate recalls, and conscientiously organize post-recall actions to mitigate the negative effects of a brand crisis on consumer engagement in online brand communities (Liu et al., 2017).

Second, companies can indirectly design the brand communities to consist of features and content that effectively influence the members’ emotions toward the brand. As we have revealed that the number of author’s account active weeks pre crises have no moderation effect, companies can reshape communities’ brand preferences without being restricted to the duration of consumers have participated in the online community. This approach creates a sense of brand belonging and fosters consumers' positive attitudes in a crisis (Brodie et al., 2013).

Third, Automobile companies should promote media coverage about their product recalls so as to receive a higher probability of positive UGC. We assumed that instead of restricting the news reachability, spreading news would portray the morality, responsibility, and transparency of the Automobile firms regarding their product recalls.

## 5.4 Limitations

Certain limitations present in our study. Initially, the choice of companies the whole market. Such a limited sample size may help to make specific conclusions regarding the automobile industry; however, it may result in low generalizability when applied to other industries. Next, the first thread to validity is the *number of account’s active weeks* variable we use to analyze the level of commitment/loyalty for each community user. This variable is might not the optimal one to explain the construct ‘customer engagement’. Moreover, we compute the weeks by subtracting that account's oldest UGC timestamp from the related crisis news date (which is limited to only 90 days). We then divide the number of days by seven and take the integer part. Thus, data on accounts with active weeks larger than 12 (active days larger than 84) is insufficient and unreliable. And finally, in terms of *Media coverage level*, we assume each level of news reachability has the same distance to another. However, this is not the case in reality since the marginal effect of local reach news might differ from that of continental or global reach news.

Additionally, the chosen analysis methods also impose several limitations. On the one hand, Sentiment analysis can only be used to analyze the content that contains only words, more specifically, English words. The method ignores words in other languages and content in picture, sound, or video formats. Additionally, the sentiment analysis also ignores other hidden verbal elements included in the content, such as irony, sarcasm, tone, idioms, and negations (Inc., 2021). On the other hand, Linear Probability poses drawbacks related to mathematically impossible predictions, assumptions in linearity between variables, and Heteroskedasticity (Quinn, 2021).

Regarding the risk of reliability, each company in our sample has different numbers, sizes and activity levels of online brand communities (*RepRisk AG ESG News Dataset*, 2019). In our Reddit data, each company has a different number of associated OBCs. The number of communities for each company may also be different on other platforms such as Facebook groups. Additionally, there are also variances in sizes and activity levels among communities within the same brand (*RepRisk AG ESG News Dataset*, 2019). Therefore, the same results cannot be guaranteed when using a different set of communities about the same brand. Nevertheless, users can be part of OBCs on different platforms at the same time. Future studies should consider users’ demographics to minimize the sensitivity of the results.

The variance also occurs among communities across different companies (*RepRisk AG ESG News Dataset*, 2019). Hence, the sensitivity of the result can increase when the study includes brands from other industries. Furthermore, given the limitations of our research models, Sentiment Analysis ignored hidden elements in text content, as well as contents in other formats such as picture or video.

## 5.5 Implications for future research

The section limitations has proposed several improvement suggestion for future studies. Regarding our research design, future research into identical objectives may overcome the above limitations. Data over a wider array of companies and industries should be collected for better generalizability. Accordingly, investigations into brand communities on other platforms such as Facebook and Twitter are also advised to improve robustness. Future studies may also include better variables such as purchase behaviors, reviews, or retention rates to analyze each community user's brand commitment level. Regarding our research methods, future research may develop the system and implement more advanced models such as Support-Vector Machine and Sarcasm Detection (Srivastava & Bhambhu, 2010; Berasategi, 2020) to overcome mentioned limitations in our Sentiment Analysis approach.

Our study findings have also suggested the directions for future research into the relationships between brand crises, customer engagement, brand commitment, and media coverage. However, a valuable empirical extension would be explicitly considering the brand strength within a specific industry. Since communities belonging to a strong brand would present different characteristics from a weaker brand (Germann et al., 2014). Other than that, product segmentation should also be considered because how users interpret high-end products is another story compared to regular product crises (Liu et al., 2017).

# Bibliography

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020). The Pushshift Reddit Dataset. *Proceedings of the International AAAI Conference on Web and Social Media*, *14*, 830–839.

Berger, J., Sorensen, A. T., & Rasmussen, S. J. (2010). Positive Effects of Negative Publicity: When Negative Reviews Increase Sales. *Marketing Science*, *29*(5), 815–827. https://doi.org/10.1287/mksc.1090.0557

Borah, A., & Tellis, G. J. (2016). Halo (Spillover) Effects in Social Media: Do Product Recalls of One Brand Hurt or Help Rival Brands? *Journal of Marketing Research*, *53*(2), 143–160. https://doi.org/10.1509/jmr.13.0009

Brodie, R. J., Ilic, A., Juric, B., & Hollebeek, L. (2013). Consumer engagement in a virtual brand community: An exploratory analysis. *Journal of Business Research*, *66*(1), 105–114. https://doi.org/10.1016/j.jbusres.2011.07.029

Brogi, S. (2014). Online Brand Communities: A Literature Review. *Procedia - Social and Behavioral Sciences*, *109*, 385–389. https://doi.org/10.1016/j.sbspro.2013.12.477

Casaló, L., Flavián, C., & Guinalíu, M. (2007). The impact of participation in virtual brand communities on consumer trust and loyalty. *Online Information Review*, *31*(6), 775.

Chang, A., Hsieh, S. H., & Tseng, T. H. (2013). Online brand community response to negative brand events: The role of group eWOM. *Internet Research*, *23*(4), 486–506. https://doi.org/10.1108/IntR-06-2012-0107

Cleeren, K., Dekimpe, M. G., & van Heerde, H. J. (2017). Marketing research on product-harm crises: A review, managerial implications, and an agenda for future research. *Journal of the Academy of Marketing Science*, *45*(5), 593–615. https://doi.org/10.1007/s11747-017-0558-1

Eilert, M., Jayachandran, S., Kalaignanam, K., & Swartz, T. A. (2017). Does it Pay to Recall your Product Early? An Empirical Investigation in the Automobile Industry. *Journal of Marketing*, *81*(3), 111–129. https://doi.org/10.1509/jm.15.0074

Germann, F., Grewal, R., Ross, W. T., & Srivastava, R. K. (2014). Product recalls and the moderating role of brand commitment. *Marketing Letters*, *25*(2), 179–191. https://doi.org/10.1007/s11002-013-9250-5

Inc, R. (2021, June 11). *Sentiment Analysis Challenges: Everything You Need to Know*. https://www.repustate.com/blog/sentiment-analysis-challenges-with-solutions/

Jang, H., Olfman, L., Ko, I., Koh, J., & Kim, K. (2008). The Influence of On-Line Brand Community Characteristics on Community Commitment and Brand Loyalty. *International Journal of Electronic Commerce*, *12*(3), 57–80. https://doi.org/10.2753/JEC1086-4415120304

Kalaignanam, K., Kushwaha, T., & Eilert, M. (2013). The Impact of Product Recalls on Future Product Reliability and Future Accidents: Evidence from the Automobile Industry. *Journal of Marketing*, *77*(2), 41–57. https://doi.org/10.1509/jm.11.0356

Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology*, *64*(5), 402–406. https://doi.org/10.4097/kjae.2013.64.5.402

Li, M., & Wei, H. (2016). How to Save Brand after Crises? A Literature Review on Brand Crisis Management. *American Journal of Industrial and Business Management*, *6*(2), 89–96. https://doi.org/10.4236/ajibm.2016.62008

Lin, H.-F. (2008). Determinants of successful virtual communities: Contributions from system characteristics and social factors. *Information & Management*, *45*(8), 522–527. https://doi.org/10.1016/j.im.2008.08.002

Liu, Y., Shankar, V., & Yun, W. (2017). Crisis Management Strategies and the Long-Term Effects of Product Recalls on Firm Value. *Journal of Marketing*, *81*(5), 30–48. https://doi.org/10.1509/jm.15.0535

Marzocchi, G., Morandin, G., & Bergami, M. (2013). Brand communities: Loyal to the community or the brand? *European Journal of Marketing*, *47*(1/2), 93–114. https://doi.org/10.1108/03090561311285475

Meek, S., Ogilvie, M., Lambert, C., & Ryan, M. M. (2019). Contextualising social capital in online brand communities. *Journal of Brand Management*, *26*(4), 426–444. https://doi.org/10.1057/s41262-018-00145-3

Quinn, M. S. (2021). *Session 13: Logistic regression with a binary DV*. Research Project Block 7, Erasmus University Rotterdam. https://canvas.eur.nl/courses/38153/files/59057923?module\_item\_id=752507

*RepRisk AG ESG news dataset*. (2019). RepRisk AG ESG. https://www.reprisk.com/news-research/year:2019#news-and-media-coverage

Rubel, O., Naik, P. A., & Srinivasan, S. (2011). Optimal Advertising When Envisioning a Product-Harm Crisis. *Marketing Science*, *30*(6), 1048–1065. https://doi.org/10.1287/mksc.1110.0679

Schmalz, S., & Orth, U. R. (2012). Brand Attachment and Consumer Emotional Response to Unethical Firm Behavior. *Psychology & Marketing*, *29*(11), 869–884. https://doi.org/10.1002/mar.20570

Silge, J., & Robinson, D. (2017). *Text Mining with R: A Tidy Approach*. O’Reilly Media, Inc.

Woisetschläger, D. M., Hartleb, V., & Blut, M. (2008). How to Make Brand Communities Work: Antecedents and Consequences of Consumer Participation. *Journal of Relationship Marketing*, *7*(3), 237–256. https://doi.org/10.1080/15332660802409605

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *WIREs Data Mining and Knowledge Discovery*, *8*(4). https://doi.org/10.1002/widm.1253