Success Factors of Social Media Content

-

The example of Accenture Facebook Posts

Research Paper
University of St. Gallen
Master Course 7,008
"RPV: Data Science - Methods and
Technologies for Data-driven Business Models"

Course Director: Prof. Dr. Jochen Wulf
Institute of Information Management (IWI-HSG)

Valentin Kahn, 12-613-378
valentin.kahn@student.unisg.ch

St. Gallen, 21.12.2016

Table of Content

Abstract

Social media has gained increasing importance in recent years as an instrument of branding, customer relations and sales initiation. The areas of application vary. Accenture uses social media as a channel increasing the general brand awareness, staging business deals by presenting the company as a competent and innovative advisor and raising awareness of the company amongst top talents.

With the help of data science, large amounts of historical social media data can be analysed in order to determine a company's success factors when it comes to reaching their goals on social media. In this study, Facebook feed data of the last six years from Accenture was analysed in order to determine these success factors. The study identified likes, comments, shares and user emotions as important variables describing the social media success of a company like Accenture. Several factors have been tested for their relation to and correlation with these variables. Amongst the factors tested are post length, the time focus of a post, the occurrence of key words within a post, the inclusion of links, questions and calls to action in these posts. The study worked with the full dataset and two subsets containing first only posts that received comments and second the most successful posts by means of the variables described above. The data was analysed using SQL (Structured Query Language) in two settings - a virtual server setting and a normal database setting.

The results show that some conclusions derived from scholars regarding success factors of social media activities can be reproduced using Accenture's Facebook feed data while others cannot. For several factors like "post length" or "time focus" there was no evidence for their correlation with the success of a Facebook post. Other factors turned out to have negative effects, as observed for questions, links and most call-to-action elements. Photos and videos as media elements however were shown to have a major positive impact on both the number of likes and user engagement measured by the number of comments.

As Accenture represents a company operating in B2B sectors exclusively, the conclusions derived from this study might not be successfully replicated when applied to consumer brands, because both the way that Accenture manage their brand and the way that users engage with that brand look different.

Figures

1. Introduction

Counting around 375'000 employees and ranking no. 312 in Forbes' "Global 500" list, Accenture plc seated in Dublin, Ireland is the largest consulting firm in the world (Time Inc., 2016). The technology consultancy advises clients in digitalising their business, changing and outsourcing their operational technology. Accenture does what could be described as "end-to-end consulting". They do not only build new strategies for their clients, but also support them in realising these new strategies (Accenture, 2016a; Vault, 2016).

In recent years, companies have started to discover the benefits of social media marketing. Accenture is no different. Accenture's Facebook Page counts almost 450'000 likes (Facebook, 2016a). Their focus is on shaping the brand around Accenture and raising awareness around it. But how successful are their posts? The aim of this study is to find out which factors support the success of Accenture Facebook posts through the analysis of Facebook data from Accenture.

1.1 Data Science & Social Media Management

The main research question of this study is: "What do Accenture posts that create the most attention, the biggest discussion and the best emotions have in common?".

Data science is a relatively young field of science, originating from the development of data software and the availability of ever cheaper yet higher performing computer hardware. According to "Moore's Law", the number of transistors of an affordable CPU – serving as a representation for processor speeds of computers – doubles within two years (MemeBridge, 2016). Yet the reason why data science is becoming ever more prominent now is the emerging availability of huge datasets. Never in the history of mankind has so much data been available to be explored, analysed and requested for wisdom. 90% of all our data has been generated in the last two years (Daily Science, 2013).

Big corporations increasingly place high importance on the management of social media channels. The large majority of Fortune 500 companies is active on Facebook, Twitter and YouTube (Slegg, 2013), heavily invest in social media efforts and engage with their customers via social media (Chung et al., 2014). Social media offers the opportunity for corporations to form online communities that might positively affect branding, product development, sales or customer service functions of the company (Chung et al., 2014). Chung et al. (2014, p.16) have found out that "social media efforts can improve consumer engagement and attention, which in turn, result in higher market performance". Goh et al. show that a firm's engagement in a brand community positively affects sales (Goh et al., 2013).

In the specific case of Accenture's social media efforts, three main goals can be identified: increasing the general brand awareness, staging business deals by presenting the company as a competent and innovative advisor and raising awareness of the company amongst top talent. For the latter, Accenture has created their own career Facebook page. According to Accenture, the main purpose of the additional page is to update potential candidates about job opportunities and career events and answer their open questions (Accenture, 2016b). Additionally, Accenture runs several country-specific pages (such as "Accenture India" which received almost the same number of likes than the main page) and brand-specific pages (such as "Accenture Digital", "Accenture Strategy" or "Accenture Interactive"). However, most of these pages by far underperform the main page in terms of number of likes and user engagement (Facebook, 2016b).

Stephen et al. (2015) mention the following six key factors in terms of content for a firm's social media posts:

- Arousal orientation (measured by positive emotions and humour)
- Persuasion orientation, separated into relevance, message clarity and advertising tone
- Product-, brand- or value-related information
- Calls to action (e.g. the call to read an article)
- References to not brand-related content (e.g. to holidays)

- Media elements (including posts containing references to other media elements, such as links to images or videos).

Stephen et al. (2015) conclude that when it comes to optimising Facebook communication, how information is provided in Facebook posts (e.g. through persuasion or arousal) is more relevant than the actual (e.g. brand-, product- or value-related) information. According to Stephen et al. (2015), the inclusion of media elements such as photos or videos does not have a huge effect on user engagement.

In this study the focus is on those factors from above which are suitably measurable based on the Facebook feed data provided: arousal orientation (measured by positive and negative emotions), brand-related information (measured by brand-related words, such as "Accenture"), calls to action (measured by exclamation marks, "Learn" and "Read" statements) and media elements (with a focus on links). Additionally, the impact of length and time focus of posts was also subject to investigation in this study.

Stephen et al. (2015) mention "likes", "comments" and "shares" as actions by consumers that are suitable indicators to measure attitudinal responses and predict marketing outcomes. While likes and shares (at least on Facebook) can only signal positive reactions, comments are valenced which means they can represent both a positive or a negative reaction. Stephen et al. do recognize the limitations of a measurement approach based on these interaction patterns which only represent a "low-level form of engagement" (Stephen et al., 2015, p.20).

De Vries et al. (2012) define certain variables that trigger the number of likes and comments to a post. They emphasize that content related to the brand is more successful than content which is unrelated to the brand. According to them, user engagement (=comments) can be triggered by asking questions. However, in their study they find out that questions have a negative impact on the number of likes of a post. Taking up positive former comments in a next post triggers the number of likes of this post. Taking up former comments (both positive and negative) in a next post positively affects the number of comments of this post. According to their study, posting links turns out to have negative effects on the number of comments.

1.2 Research Question & Hypotheses Development

The main research question of this study is: "What do Accenture posts that create the most attention, the biggest discussion and the best emotions have in common?"

Reviewing the literature and setting the focus on mentioned factors, the following hypotheses emerge:

H1: Posts with brand-related content (measured by key words) have a higher success (measured in likes and shares) than posts without brand-related content
H2: Questions and calls to action can trigger user engagement, measured by the number of comments
H3: Questions have a negative impact on the number of likes of a post
H4: Posts that include links have less comments and shares than posts that do not contain links
H5: Length and time focus of posts also have impact on their success, measured by the number of likes, shares and comments.
H6: The inclusion of media elements in a post such as photos or videos does not have a huge effect on user engagement (measured by the number of comments).

2. Methodology

The explanatory research study was performed based on available Facebook feed data from Accenture. The data was provided by the course director, Prof. Dr. Wulf who performed the whole data collection process. The feed data was then analysed in a database setup.

2.1 Research Setup

In order to analyse the data, SQL (Structured Query Language) was used. For the analysis, two database setups have been used:

- The first setup was based on a virtual machine (Oracle VM VirtualBox) and Apache Hadoop for data mining. The query editor Apache Impala then was used to analyse the data. This setup allows to simulate working with a large dataset (Big Data). Apache Hadoop is a framework that allows distributed processing of large datasets among several clusters of computers (Apache, 2016a). The "MapReduce" function of Apache Hadoop allows for parallelised processing of large amounts of data (Langit, 2013). Apache Impala increases the functionality and performance of SQL queries in Apache Hadoop (Apache, 2016b).
- Hue serves as the web interface for Apache Hadoop, allowing the user to perform loading, viewing, processing and visualizing the data in Apache Hadoop (Gethue.com, 2016).
- In order to perform the most complex queries quicker, a second setup was used. These queries were performed on MySQL Workbench, using MariaDB.

2.2 Research Process

First, the dataset was scanned for data manipulation, foremost content by non-human accounts (or "bots") and intentionally deleted content. For this, several key word searches were performed. Neither bots nor any sign of deleted content was found. Several incomplete data rows and data rows containing extreme values were checked and excluded or handled with caution.

The dataset was analysed looking at the full dataset and two subsets:

- Analysis of the full dataset (11'352 rows)
- Analysis of a subset only containing posts that have comments (2'140 rows)
- Analysis of a second subset of "most successful" posts based on number of likes, number of shares, number of comments and emotions of comments (168 rows)

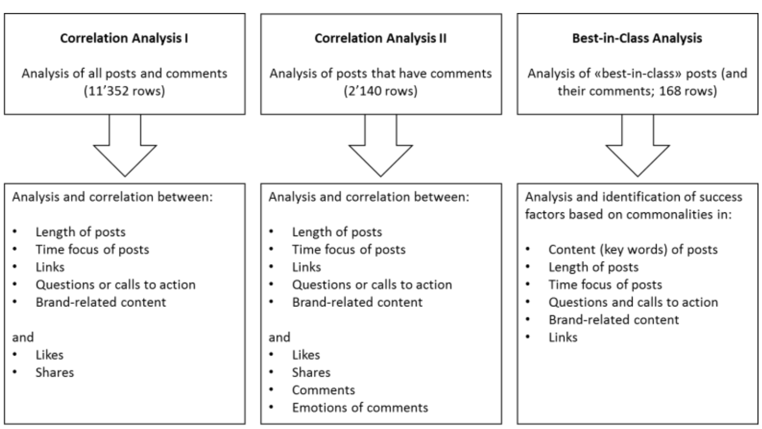

Figure 1: Approach to the data analysis for this study

In order to select the most successful posts, criteria for "success" had to be defined. The following criteria were used:

- Number of likes > 5
- Number of comments > 0
- Number of times shared > 0
- Highly positive emotions based on the given factor "posemo": Posemo > 5
- No occurrence of negative emotions based on the given factor "negemo". Negemo = 0
- Comments of post on average having posemo > 5
- Comments of post on average having negemo = 0

The subset of the most successful posts (168 rows) then was further analysed with the aim to find commonalities which serve as success factors for Accenture posts. The focus of the analysis was put on:

- The content of the posts defined as the occurrence of key words
- The content of the links within the posts defined as the occurrence of key words within the link names
- The time focus of the posts based on the three given factors "focuspast", "focuspresent" and "focusfuture", describing the respective time focus
- The length of the posts
- The inclusion of questions or calls to action in the posts.

As a further step of the analysis, only posts that have comments were analysed (2'140 rows). The following factors have been included in the analysis, testing for correlation with the success variables:

- The length of the posts
- The time focus of the posts (described by "focuspast", "focuspresent" and "focusfuture")
- Number of likes of the posts
- Number of shares of the posts
- Number of comments of the posts
- Questions and calls to action in the posts (described by question marks, exclamation marks and key words indicating calls to action)
- Links in the posts
- Brand-related content in the posts (described by brand-related words, such as "Accenture")
- The occurrence of key words in the posts
- Positive sentiments of the comments (described by "posemo")
- Negative sentiments of the comments (described by "negemo")

All results then were further analysed and illustrated using the functionalities of Apache Impala and Microsoft Excel.

Based on the results (next chapter), theoretical triangulation based on the social media research presented in the "Introduction" chapter was performed in order to validate the findings (see "Discussion" part), as suggested by Müller et al. (2016).

2.3 Challenges

Most challenges that arised during this study were of technical nature. The first challenge arised during the setup of the virtual machine (Oracle VM VirtualBox). The feature of my laptop to distribute cores was disabled. The problem was solved by reconfiguring the settings in the BIOS. After the virtual machine was spun up, the data mining with Hadoop started.

A major challenge of Hadoop was the speed. Executing the queries in this setup took a lot of time. First, I tried to avoid the problem by creating smaller tables and working on those. Eventually I had to make the decision to migrate the data also on MySQL Workbench in order to save a lot of time when executing the most complex queries (e.g. the word count). For the migration, I had to perform some modifications on the data (.csv) file in order to be able to import it to MySQL Workbench. I then could further analyse the data in the Workbench using MariaDB.

Although I had first experiences using SQL when analysing user statistics for a start-up, I knew that I would have to upgrade my knowledge in order to build the right queries for the data analysis. Some queries demanded for further research and a deeper understanding of SQL's functionality. This was especially true for the query counting the occurrence of unspecified words in the messages analysed (for further information, see "Key Queries" in the "Appendix").

3. Results

Just as stated in theory, a clear correlation between number of likes and number of shares was observed:

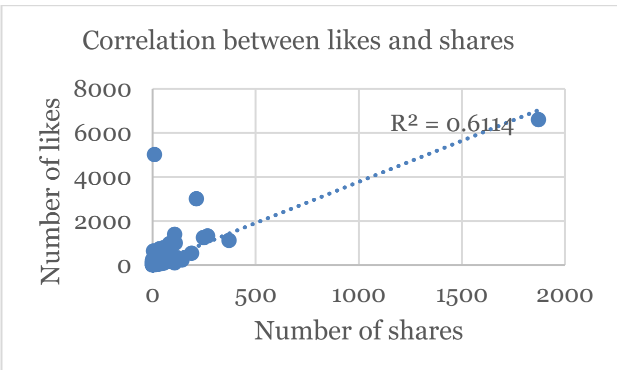

Figure 2: Correlation between likes and shares in Accenture Facebook posts

In regards to the hypotheses built, the following implications emerged, divided into the subsections of "Content", "Questions & Calls to Action", "Links, Length & Time Focus" and "Media Elements".

3.1 Content

H1: *Posts with brand-related content (measured by key words) have a higher success (measured in likes and shares) than posts without brand-related content*

From the word count, the following brand-related key words have been identified (all among the top 50 words that occurred in general): "accenture" (occurred 3'260 times), "IT" (occurred 1'347 times), "business" (occurred 993 times), "companies" (occurred 865 times), "digital" (occurred 783 times), "technology" (occurred 720 times), "services" (occurred 666 times), "accenture's" (occurred 506 times).

Posts which include at least one of these key words count 32.13 likes (σ = 114.36) and 2.768 shares (σ = 25.39) on average. Posts that do not include these key words count 36.32 likes (σ = 117.07) and 2.717 shares (σ = 11.01) on average.

The analysis of "Best-in-class" posts (the "most successful" 168 posts as defined above) however resulted in several of these words showing up in the top 50. "Accenture" occurs 62 times in these posts. Further "Learn" (24 times), "companies" (23 times), "business" (21 times), the hashtag "#GreaterThan" (the only hashtag in the top 50; 21 times), "energy" (19 times), "services" and "technology" (17 times) are amongst the significant results.

3.2 Questions & Calls to Action

***H2:** Questions and calls to action can trigger user engagement, measured by the number of comments*

Posts neither having a question nor any call to action received 0.81 comments on average. Posts including both at least one question mark and at least one exclamation mark received an average number of comments of 0.67. Posts including at least one question mark but no exclamation mark received 0.73 comments. Posts including at least one exclamation mark but no question mark received 1.07 comments on average, thus more comments than posts without neither question nor exclamation mark. Posts that include the "Read" statement ("Read" occurred 861 times in all posts) on average received 0.70 comments. Posts that include the "Learn" statement received 0.74 comments on average.

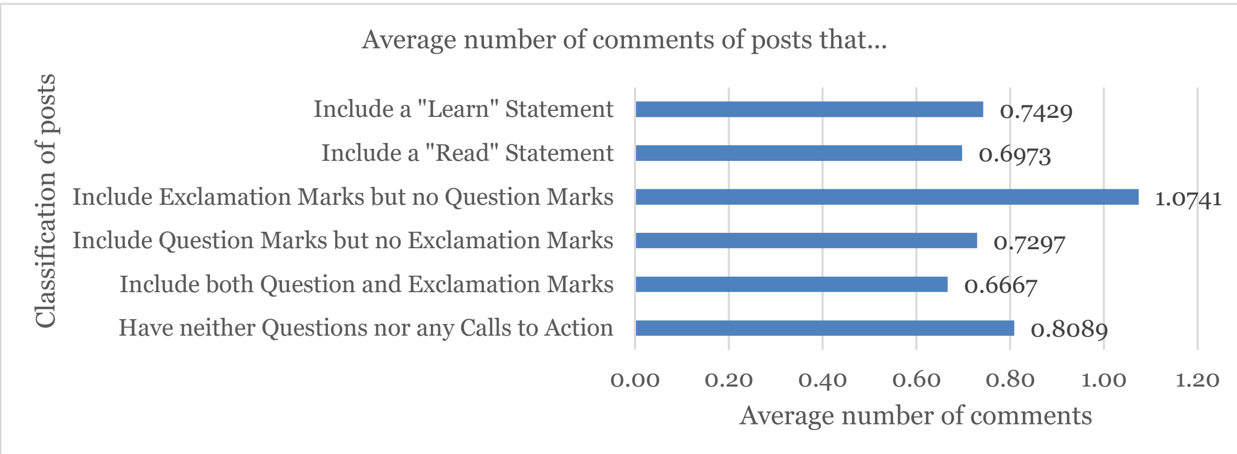

Figure 3: Relation between questions and calls to action and number of comments

***H3:** Questions have a negative impact on the number of likes of a post*

Posts that include a question mark on average received around 24 likes ($\sigma = 24.45$). Posts without question marks received on average around 35 likes ($\sigma = 130.45$).

3.3 Links, Length & Time Focus

***H4:** Posts that include links have less comments and shares than posts that do not contain links*

Posts that do not contain a link on average received 1.71 comments ($\sigma = 3.80$) and were shared 3.66 times ($\sigma = 26.27$). Posts containing a link only received 0.76 comments ($\sigma = 1.77$) on average and only got shared 2.71 times ($\sigma = 24.01$).

In terms of the number of likes, posts having a link are slightly less successful (receiving 32.66 likes on average, $\sigma = 115.50$) than posts not containing links (receiving 35.44 likes on average, $\sigma = 134.21$).

***H5:** Length and time focus of posts also have impact on their success, measured by the number of likes, shares and comments.*

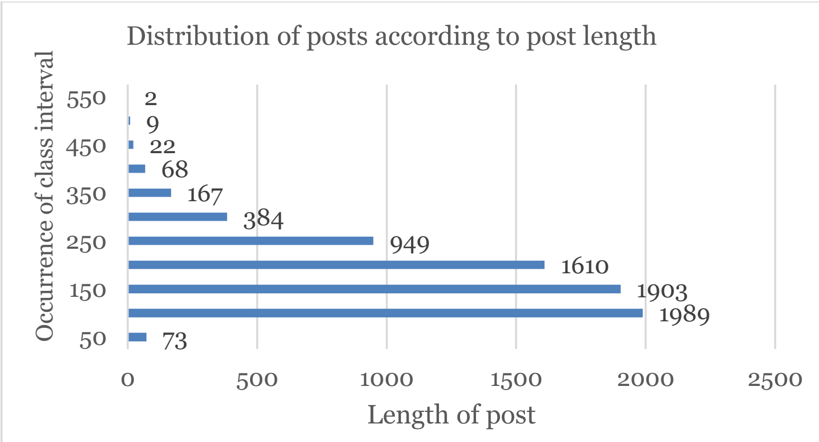

Figure 4: Distribution of posts according to post length

There was no significant correlation between length (independent variable), number of likes, number of comments or number of shares (dependent variables).

There was also no significant correlation between any of the time foci (independent variables) and any of the dependent variables.

Despite a missing correlation, there seems to be an optimal post length range resulting in the biggest number of likes. This range lies in between about 80 and 250 characters. Resulting implications should be weakened by the fact that many posts fall within that length range, naturally resulting in a higher number of posts potentially receiving a high number of likes.

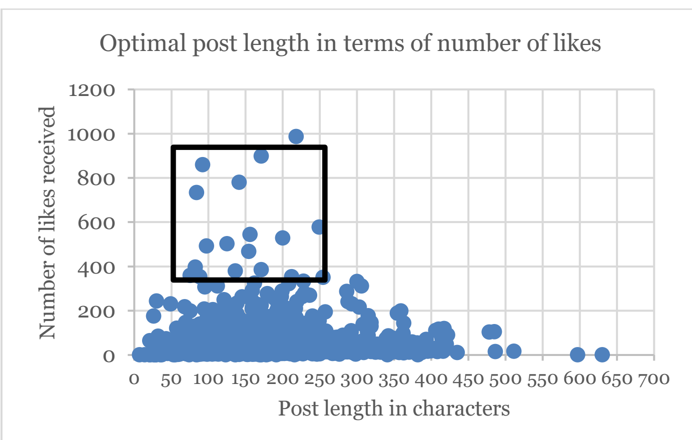

Figure 5: The range of the optimal post length in terms of number of likes

3.4 Media Elements

H6: *The inclusion of media elements in a post such as photos or videos does not have a huge effect on user engagement (measured by the number of comments).*

There is a clear positive effect on the number of likes emerging from the presence of photos and videos. Photos on average received 127.21 likes, videos 55.15 likes and posts not containing these elements only 28.21 likes.

The same effect applies for the number of comments. Posts with photos receive 1.599 comments on average, posts containing videos receive 1.277 comments and post without these media elements only receive 0.722 comments.

After all, 16 of 25 posts that received more than 400 likes contained a photo.

4. Discussion

The research seeks to understand which factors determine the success of Facebook posts of Accenture. It aims to check whether the conclusions from contemporary literature can be reproduced working with Accenture Facebook data. The study shows that several conclusions from contemporary literature cannot be reproduced when analysing the Facebook data of Accenture. However, several other significant success factors were identified.

4.1 Theoretical Implications

H1: *Posts including brand-related content (measured by key words) have a higher success (measured in likes and shares) than posts without brand-related content*

Posts including brand-related key words do only have marginally more shares and even significantly less likes than posts that do not include these key words. Thus, based on key words, the statement can be falsified in the case of Accenture.

<u>Possible explanation:</u> In a world of information overflow and user attention remaining to be limited to the wake phase during the day, it has become much harder to catch the user's attention. Superior or extreme content receiving more attention and likes many times has no relation to a brand.

H2: *Questions and calls to action can trigger user engagement, measured by the number of comments*
In the case of Accenture, the statement can be verified for calls to action measured by exclamation marks only and must be falsified for questions measured by question marks and calls to action measured by the inclusion of "Learn" or "Read" statements.

<u>Possible explanation:</u> As questions, "Read" and "Learn" statements many times refer to an external page or external content, the user and his scarce attention are occupied consuming this content, rather than writing comments to a clear statement in the post.

H3: *Questions have a negative impact on the number of likes of a post*

The statement can be verified in the case of Accenture. Note that extreme values in the number of likes in two cases (that included a question mark) were excluded.

<u>Possible explanation:</u> Questions and calls to action are missing a clear, focused statement. User prefer clear and focused information and rate it with a like. Questions and calls to action however receives attention at another level (i.e. the user reads external content).

H4: *Posts that include links have less comments and shares than posts that do not contain links*

In the Accenture dataset, the statement can be numerically verified.

<u>Possible explanation:</u> The same applies as for H2. The user is occupied consuming external information, leaving him less time to comment or share. Additionally, technical content will be shared less than "fluffier", easier consumable information.

H5: *Length and time focus of posts also have impact on their success, measured by the number of likes, shares and comments.*

The statement can be falsified in the case of Accenture for both factors (length and time focus of posts).

Possible explanation: The user has a limited span of attention and thus tends to only read a certain length of text anyways, leaving the factor "post length" irrelevant in the end.

H6: *The inclusion of media elements in a post such as photos or videos does not have a huge effect on user engagement (measured by the number of comments).*

In conclusion, the theory of Stephen et al. from Saïd Business School / Oxford University can be falsified in the case of Accenture.

Possible explanation: As humans are both social and visual "animals", photos and videos – especially when reaching a high graphical quality – tend to touch the user more and make him engage in a discussion.

4.2 Managerial Implications

Success factors for user engagement

User engagement was measured by the number of comments. Based on this study, the inclusion of media elements in Facebook posts is the most promising success factor when it comes to driving user engagement. Posts that contain photos receive more than double as much comments than posts without media elements. Posts including video links or videos also perform significantly better.

In the dataset of Accenture there is no clear correlation between questions or calls to action and user engagement. In fact, these factors tend to lower user engagement rather than increasing it. As an exception, the inclusion of exclamation marks in posts had a positive effect on user engagement. However, the research did not focus on grouping posts including exclamation marks, thus the factor cannot be exclusively defined to be a call of action.

The inclusion of links also tends to lower user engagement. In fact, posts including one or more links did not even receive half the number of comments compared to posts not having any link.

Interestingly, neither the length nor the time focus of the posts has any significant, measurable impact on user engagement.

Success factors for the number of likes

Media elements majorly drive the number of likes. Posts containing photos received four-and-a-half times (!) as many likes as post without media elements. Posts that worked with videos almost received twice the number of likes than posts without media elements.

Publishing brand-related content (in this study measured by the inclusion of key words) is not a success factor when it comes to driving the number of likes. Posts that work with brand-related key words on average received around 11% less likes.

The assumption that questions have a negative effect on the number of likes of a post was verified in this study. Surprisingly, the impact of this factor even turned out to have major significance. On average, posts working with questions had more than 30% less likes than posts without questions.

Conversely, the factors "length" and "time focus" do not seem to have any impact at all on the number of likes of a post, as measured in this study.

Success factors for the number of shares

The inclusion of brand-related content in the case of the number of shares signs responsible for a slightly positive impact. Posts including brand-related key words counted slightly more shares than their reference group without brand-related key words.

Just as for comments and likes, length and time focus do not have a major impact on the number of shares that a post receives.

Links do have a significant negative impact on the number of shares of a post. Posts including links on average received more than 25% less shares than posts without links.

4.3 Limitations & Further Research

This study encountered several limitations which inherently offer potential for further research.

As Accenture is a company that exclusively operates in B2B sectors, a lot of recommendations and strategies from scholars targeting end consumers are not valid for Accenture. On the other hand, the conclusions derived from this study might not be successfully replicated when applied to consumer brands, because both the way that Accenture manage their brand and the way that users engage with that brand are different. There is relatively few literature in this rather novel research field about exploring these differences in brand management and consecutively brand engagement in social media between B2C and B2B. The same applies for differences amongst specific industries. Because the research field is relatively new, it is also not yet clear how consciously social media strategies are applied in the companies. Certainly, qualitative studies including interviews with company representatives would offer valid insights into behavioural factors like intentionality or path dependence related to social media efforts.

The study checked several variables in order to find success factors, but excluded the factor of the day period. Further research could consider what day (or night) time posts are most successful. Just like many other social media studies so far, the analysis of engagement variables like “likes”, “comments” and “shares” remains being a vague approach to estimate consumer engagement. Further studies should look more detailed into the relation between these low-engagement variables and actual consumer engagement, i.e. in a buying process or crowdsourcing activities. Further, no multifactor regression was performed in order to check for several variables at once. This could also offer further insights into success factors based on combinations of variables. A lot of conclusions in this study were derived based on key words. Key words offered a suitable opportunity to explore the occurrence of brand-related content (through brand-related key words), questions (through question marks) and calls to action (through exclamation marks and key words indicating calls to action). This approach however excludes potentially relevant content that does not contain these key words.

References

Accenture. (2016a). *About Accenture*. Retrieved December 7, 2016, from https://www.accenture.com/ch-en/company

Accenture. (2016b). *Accenture Karriere*. Retrieved from https://www.accenture.com/ch-de/careers/social-media

Apache. (2016a). *Apache Hadoop*. Retrieved December 8, 2016, from https://hadoop.apache.org/

Apache. (2016b). *Apache Impala (incubating) - Overview*. Retrieved December 8, 2016, from https://impala.apache.org/overview.html

Chung et al., S. (2014). *Firms' Social Media Efforts, Consumer Behavior, and Firm Performance: Evidence from Facebook.*

Daily Science. (2013, May 22). *Big Data, for better or worse: 90% of world's data generated over last two years*. Retrieved December 8, 2016, from https://www.sciencedaily.com/releases/2013/05/130522085217.htm

De Vries, L., Gensler, S., & Leeflang, P. S. (2012). *Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing. Journal of Interactive Marketing, 26(2), 83-91.*

Facebook. (2016a). *Accenture*. Retrieved December 7, 2016, from https://www.facebook.com/accenture/?fref=ts

Facebook. (2016b). *"Accenture" Pages*. Retrieved from https://www.facebook.com/search/pages/?q=accenture

Gethue.com. (2016). *Let's big data. Hue is a Web interface for analyzing data with Apache Hadoop*. Retrieved December 8, 2016, from http://gethue.com/

Goh, K.-Y., Cheng-Suang, H., & Zhijie, L. (2013). *Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content. Information Systems Research 24.1 (2013): 88-107.*

Langit, L. (2013, May 18). *Hadoop MapReduce Fundamentals 1 of 5*. Retrieved December 8, 2016, from https://www.youtube.com/watch?v=7FcMhTTG1Cs

MemeBridge. (2016). *Moore's Law*. Retrieved December 8, 2016, from http://www.mooreslaw.org/

Slegg, J. (2013, July 25). *Fortune 500 Social Media: 77% Active on Twitter; 70% on Facebook*. Retrieved December 8, 2016, from https://searchenginewatch.com/sew/study/2284930/fortune-500-social-media-77-active-on-twitter-70-on-facebook

Stephen, A., Sciandra, M., & Inman, J. (2015). *Is It What You Say or How You Say It? How Content Characteristics Affect Consumer Engagement with Brands on Facebook.* Saïd Business School.

Time Inc. (2016). *Fortune Global 500* . Retrieved December 7, 2016, from http://beta.fortune.com/global500/list/filtered?searchByName=accenture

Vault Inc. (2016). *Accenture*. Retrieved December 7, 2016, from http://www.vault.com/company-profiles/management-strategy/accenture/company-overview.aspx

Appendix

Key Queries

```
SELECT status_message as Content FROM accenture
    WHERE type = "post"
    and num_likes > 5
    and num_comments > 0
    and num_shares > 0
    and posemo > 5
    and negemo = 0
    and post_id
IN
    (SELECT Distinct original_post_id from accenture
        where posemo > 5
        and negemo = 0
        and type = "comment");
```

Figure 6: SQL query – selection of "best-in-class" posts

```
SELECT status_message, length(status_message), focuspast, focuspresent, focusfuture, num_likes, num_shares, t.pos, t.neg
   FROM accenture
INNER JOIN
   (SELECT avg(posemo) as pos, avg(negemo) as neg, original_post_id
       FROM accenture
          WHERE type = "comment"
GROUP BY original_post_id)
as t
ON accenture.post_id_summary = t.original_post_id;
```

Figure 7: SQL query - selecting the average of "posemo" and "negemo" of the comments to a post

```
select avg(num_comments) from accenture where type = "post" and status_message not like "%!%" and status_message not like "%?%"
    and status_message not like "%Read%" and status_message not like "%Learn%"
union
select avg(num_comments) from accenture where type = "post" and status_message like "%!%" and status_message like "%?%"
union
select avg(num_comments) from accenture where type = "post" and status_message like "%?%" and status_message not like "%!%"
union
select avg(num_comments) from accenture where type = "post" and status_message like "%!%" and status_message not like "%?%"
union
select avg(num_comments) from accenture where type = "post" and status_message like "%Read%"
union
select avg(num_comments) from accenture where type = "post" and status_message like "%Learn%";
```

Figure 8: SQL query - exploring the relation between questions, calls to action and user engagement

```sql
SELECT SUM(total_count) as total, value
    FROM
        (SELECT count(*) AS total_count, REPLACE(REPLACE(REPLACE(x.value,'?',''),'.',''),'!','') as value
            FROM
                (SELECT SUBSTRING_INDEX(SUBSTRING_INDEX(t.status_message, ' ', n.n), ' ', -1) value
                    FROM
                        (SELECT status_message FROM accenture
                            WHERE type = "post"
                            and num_likes > 5
                            and num_comments > 0
                            and num_shares > 0
                            and posemo > 5
                            and negemo = 0
                            and post_id IN
                                (SELECT Distinct original_post_id from accenture
                                    where posemo > 5
                                    and negemo = 0
                                    and type = "comment")) t
    CROSS JOIN
        (SELECT a.N + b.N * 10 + 1 n
            FROM
                (SELECT 0 AS N UNION ALL SELECT 1 UNION ALL SELECT 2 UNION ALL SELECT 3 UNION ALL SELECT 4
                UNION ALL SELECT 5 UNION ALL SELECT 6 UNION ALL SELECT 7 UNION ALL SELECT 8 UNION ALL SELECT 9) a
               ,(SELECT 0 AS N UNION ALL SELECT 1 UNION ALL SELECT 2 UNION ALL SELECT 3 UNION ALL SELECT 4
               UNION ALL SELECT 5 UNION ALL SELECT 6 UNION ALL SELECT 7 UNION ALL SELECT 8 UNION ALL SELECT 9) b
                ORDER BY n) n
                    WHERE n.n <= 1 + (LENGTH(t.status_message) - LENGTH(REPLACE(t.status_message, ' ', '')))
                    ORDER BY value) AS x
                    GROUP BY x.value) AS y
                        GROUP BY value
                        ORDER BY total
                        DESC LIMIT 50;
```

Figure 9: SQL query – word count (top 50) in "best-in-class" posts

```sql
select avg(num_shares), avg(num_likes) from accenture where type = "post"
    and num_likes < 1396069215
    and
    (status_message like "%X%" or
    status_message like "%accenture%" or
    status_message like "%IT%" or
    status_message like "%business%" or
    status_message like "%companies%" or
    status_message like "%digital%" or
    status_message like "%technology%" or
    status_message like "%services%" or
    status_message like "%accenture's%")
union
select avg(num_shares), avg(num_likes) from accenture
    where type = "post"
    and num_likes < 1396069215
    and (status_message not like "%accenture%"
    and status_message not like "%IT%"
    and status_message not like "%business%"
    and status_message not like "%companies%"
    and status_message not like "%digital%"
    and status_message not like "%technology%"
    and status_message not like "%services%"
    and status_message not like "%accenture's%")
    ;
```

Figure 10: SQL query – exploring the relation between brand-related key words and number of likes and shares while excluding extreme data points

Further Statistics

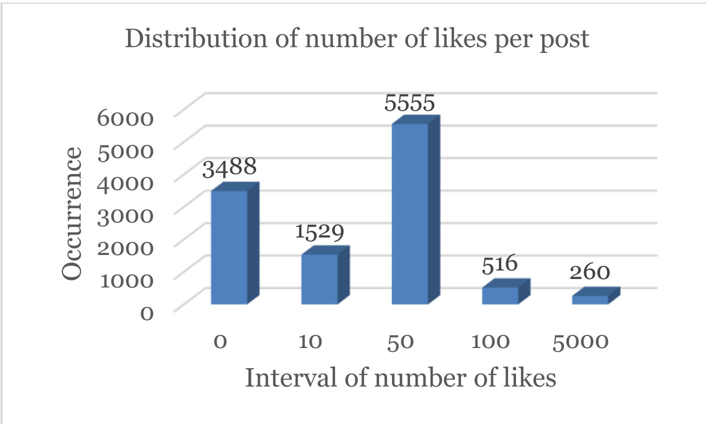

Figure 11: Distribution of number of likes per post

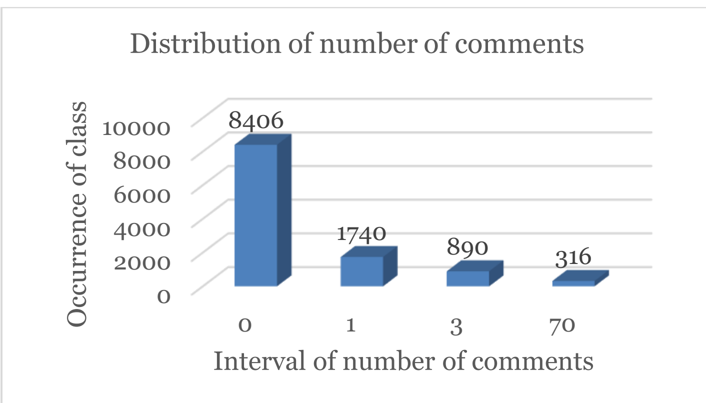

Figure 12: Distribution of number of comments in interval classes

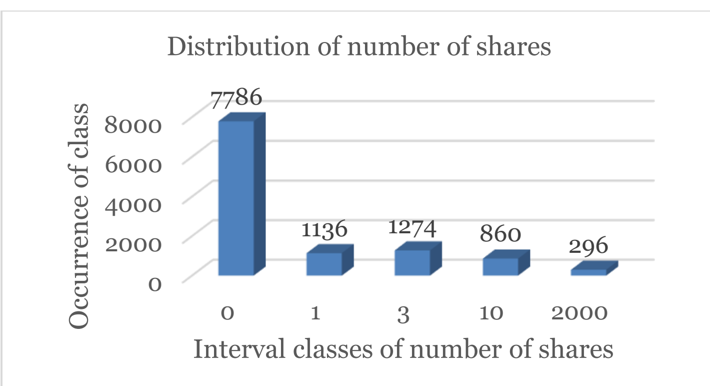

Figure 13: Distribution of number of shares in interval classes

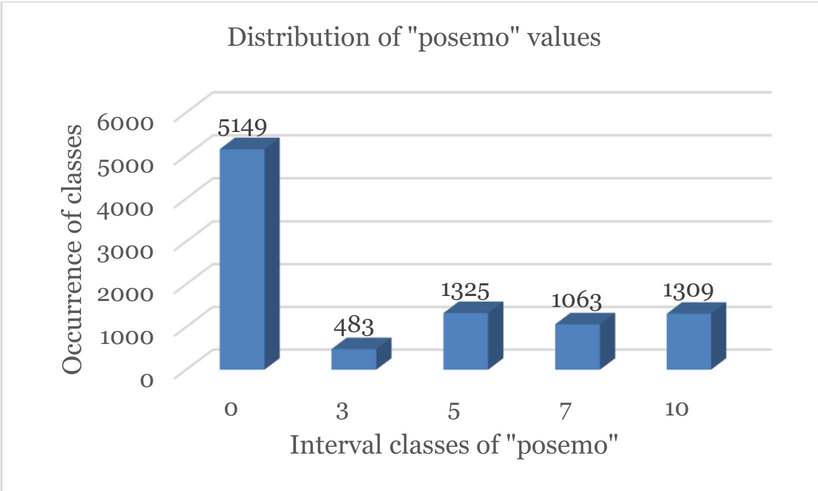

Figure 14: Distribution of "posemo" values in interval classes

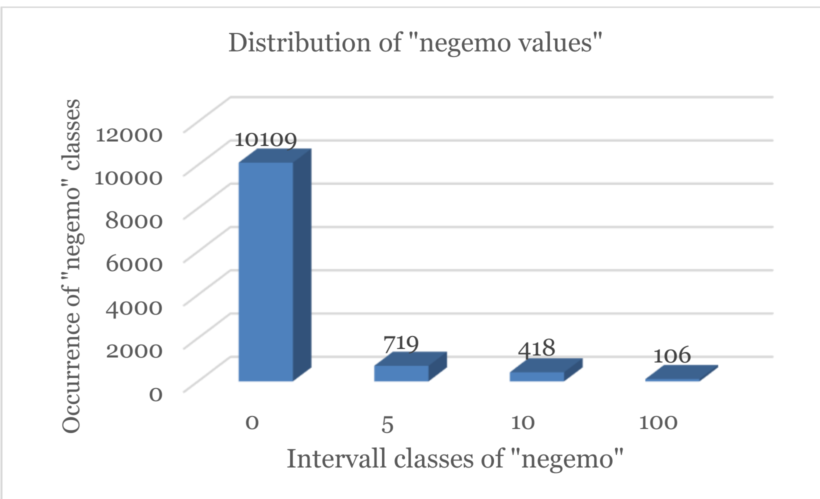

Figure 15: Distribution of "negemo" values in interval classes

Declaration of Authorship

"I hereby declare

- that I have written this thesis without any help from others and without the use of documents and aids other than those stated above,
- that I have mentioned all the sources used and that I have cited them correctly according to established academic citation rules,
- that the topic or parts of it are not already the object of any work or examination of another course unless this has been explicitly agreed on with the faculty member in advance and is referred to in the paper,
- that I will not pass on copies of this work to third parties or publish them without the University's written consent if a direct connection can be established with the University of St. Gallen or its faculty members,
- that I am aware that my work can be electronically checked for plagiarism and that I hereby grant the University of St. Gallen copyright in accordance with the Examination Regulations in so far as this is required for administrative action.
- that I am aware that the University will prosecute any infringement of this declaration of authorship and, in particular, the employment of a ghostwriter, and that any such infringement may result in disciplinary and criminal consequences which may result in my expulsion from the University or my being stripped of my degree."

St. Gallen, 21.12.2016

Valentin Kahn, 12-613-378