

# HOW TO ENGAGE FOLLOWERS: CLASSIFYING FASHION BRANDS ACCORDING TO THEIR INSTAGRAM PROFILES, POSTS AND COMMENTS

Stefanie Scholz<sup>1</sup> and Christian Winkler<sup>2</sup>

<sup>1</sup>Department of Social Economy,  
Wilhelm Loehe University of Applied Sciences, Fuerth, Germany  
<sup>2</sup>datanizing GmbH, Schwarzenbruck, Germany

## **ABSTRACT**

*In this article we show how fashion brands communicate with their follower on Instagram. We use a continuously update dataset of 68 brands, more than 300,000 posts and more than 40,000,000 comments.*

*Starting with descriptive statistics, we uncover different behavior and success of the various brands. It turns out that there are patterns specific to luxury, mass-market and sportswear brands. Posting volume is extremely brand dependent as is the number of comments and the engagement of the community.*

*Having understood the statistics, we turn to machine learning techniques to measure the response of the community via comments. Topic models help us understand the structure of their respective community and uncover insights regarding the response to campaigns.*

*Having up-to-date content is essential for this kind of analysis, as the market is highly volatile. Furthermore, automatic data analysis is crucial to measure the success of campaigns and adjust them accordingly for maximum effect.*

## **KEYWORDS**

*Instagram, Fashion Brands, Data Extraction, Marketing, Analysis, Artificial Intelligence, Netnography, Descriptive Statistics, Visualization, Community Engagement, Artificial Intelligence, Unsupervised Learning, Topic Modelling*

## **1. INTRODUCTION**

Social media provide us with valuable information about users' behaviour, not only regarding (prospective) customers, but also in terms of brands' or organizations' online activities. Understanding the relationships between actions taken by companies and the corresponding reactions of the community is crucial for marketers to efficiently influence users' behaviour in an intended way.

Data is readily available and often consists of very large corpora –it is now decisive for market research to take advantage of this big data by handling it via AI-based approaches. However, care has to be taken on data quality. Often, AI projects fail because the underlying data is biased,

incomplete or not recent. It is therefore crucial to use statistics first to analyse the (meta) data of the corpus before starting with AI-based analysis. Additionally, interesting insights can be derived from the statistical information itself.

So far, marketers over the last decade increasingly used a netnographic approach [1]. By using quantitative methods, netnography supports qualitative analysis and navigates the researcher through the large corpus of data. Netnography is acknowledged as a useful research tool for collecting and analyzing online customer information [2], [3], [4]. Originally developed as a response to customers' increasing internet use [2], netnography is based on an ethnographic research approach to studying and understanding consumption-related aspects of customers' lives online [5]. In today's environment of digitalization, netnography is more relevant than ever before [6], [7], [8].

Now it is the time to merge the advantages of netnography and AI-based analyses to gain valuable and valid insights in large UGC datasets.

As the main content in netnography is text, text analysis is in focus of computational support [9]. In this paper we focus on topic modelling, sentiment analyses, text complexity based on deep learning models.

Instagram is one of the most popular social networks in the world [10]. Particularly among young people, there is no other platform which attracts as many daily active users (DAU) as Instagram. Moreover, most of the content on Instagram is fashion-related with fashion being the most popular hashtag [11] with over 850 million posts matching the corresponding hashtag. Thus, it has also become one of the premier marketing platforms for fashion brands. On that platform brands and users post and share outfits that are then commented and voted on by other users, thus also serving as an inspiration for the brands themselves. Furthermore, Instagram has strong visual components that connect perfectly with fashion brands, such as images, videos, boomerangs, layout, hyperlapse, stories, live [12] and since 2020 also reels; therefore, Roncha and Radclyffe-Thomas [13] argue that it is ideal for companies to market their products. In our study, we focus on top fashion brands as their marketing budgets tend to be highest and they work as a first mover in the market. However, all insights are directly applicable to other fashion brands and with some transfer also to other brands in common.

However, communication on Instagram is not exclusively outbound. Users can also send feedback via "likes", comments, or shares and can follow their favourite profiles and brands, which is ideal for brands to generate engagement with users [14]. Furthermore, this offers excellent measurement capabilities for the brands to quantify the success of their marketing campaigns.

There are different definitions of digital engagement; most of them have in common that it presumes some form of active online behaviour, which is characterized by high personal involvement with the content, organization, brand, or cause presented in an online public space like Instagram [15], [16], [17].

## **2. DATA COLLECTION**

### **2.1. Identifying Top Fashion Brands**

We identified 68 fashion brands and analysed those brands using Netnography methods. Our list was combined by using several sources from the Internet [18]. We analysed the corresponding

Instagram profiles with respect to followers, popularity etc. Afterwards, we categorized the brands according to luxury brands, mass-market brands and sportswear brands. As we will see later, this turns out to be an interesting classification as the different classes have a distinctively different behaviour.

For each fashion brand, Table 1 shows the number of followers, the number of profiles followed by those brands, their number of posts, and their corresponding fashion category. The data about posts and followers included profile activities like posts and comment for each brand since their respectively first post on Instagram until October 2020. At first sight, it is striking how much the number of profiles followed by the brands differs. Carhartt follows 2,273 other Instagram users compared to Balenciaga which does not follow any. Obviously, community engagement works entirely differently (or at least is managed in very different ways).

Table 1: Profile names and statistics for the top fashion brands

Brand profile name	Number of followers	Number of profiles followed	Number of posts	Category
adidas	25,925,972	202	799	Sportswear
American Eagle	3,626,506	1,502	1,507	Mass-market
Arc'teryx	868,659	681	2,706	Sportswear
Emporio Armani	16,824,301	8	7,356	Luxury
ASICS	895,053	459	2,563	Sportswear
Balenciaga	11,563,751	0	1,895	Luxury
A BATHING APE®	4,760,152	37	4,442	Mass-market
Bershka	9,327,305	24	3,548	Mass-market
Billabong	2,100,467	723	4,299	Sportswear
Bottega Veneta	2,445,828	0	186	Luxury
BVLGARI	9,416,411	27	3,633	Luxury
Burberry	17,588,966	2	5,250	Luxury
Calvin Klein	21,010,077	387	5,806	Luxury
Carhartt	863,448	2,273	1,376	Sportswear
Cartier Official	9,996,947	2	2,072	Luxury
Coach	4,845,644	380	5,219	Luxury
Comme des Garçons	2,122,301	0	446	Luxury
Converse	10,183,541	18	472	Sportswear
Dior Official	32,885,690	259	7,060	Luxury
Salvatore Ferragamo	5,456,683	105	5,206	Luxury
Gap	3,028,782	66	221	Mass-market
Gucci Official	41,396,760	249	7,205	Luxury
Gymshark	4,866,198	129	3,589	Sportswear
Hermès	10,477,323	0	2,013	Luxury
H&M	36,114,187	497	5,869	Mass-market
HUGO	1,827,704	191	2,322	Luxury
hurley	1,998,128	440	3,249	Sportswear
Jordan	21,246,134	134	307	Sportswear
Levi's	7,258,373	1,700	3,245	Mass-market
Louis Vuitton	39,925,873	5	4,630	Luxury
lululemon	3,447,659	175	3,004	Sportswear
Valentino	13,983,770	2	8,221	Luxury
Massimo Dutti	2,628,668	280	4,388	Mass-market
Michael Kors	15,992,659	301	5,456	Luxury
Moncler	3,277,777	52	3,486	Mass-market
Nautica	582,411	622	3,023	Sportswear
Nike	122,358,832	137	768	Sportswear

Off-White™	10,832,684	35	7,163	Mass-market
oldnavy	2,420,428	691	3,793	Sportswear
OMEGA	2,992,910	45	1,948	Luxury
On	518,764	1,057	1,471	Sportswear
PALACE	1,715,894	432	1,731	Mass-market
Patagonia	4,594,390	638	2,689	Sportswear
Prada	24,536,567	6	5,968	Luxury
Primark	8,670,829	802	12,408	Mass-market
PUMA	11,541,062	9	3,007	Sportswear
Ralph Lauren	12,061,735	7	5,453	Sportswear
Ray-Ban	5,281,356	78	102	Mass-market
Reebok	2,481,959	732	2,657	Sportswear
Reef	626,987	248	4,002	Sportswear
ROLEX	11,164,320	168	1,021	Luxury
SKECHERS USA, Inc.	544,567	320	975	Sportswear
Stüssy	4,032,604	1	2,596	Mass-market
Supreme	13,896,829	113	1,573	Sportswear
Swatch	1,279,593	36	1,011	Mass-market
TAG Heuer	2,488,827	136	5,442	Luxury
The North Face	4,714,665	413	348	Sportswear
Tiffany & Co.	11,636,286	175	1,591	Luxury
Timberland	3,034,485	64	3,562	Sportswear
Tommy Hilfiger	13,692,548	299	2,834	Luxury
Under Armour	8,086,500	433	3,227	Sportswear
UNIQLO Global	2,361,115	409	1,975	Mass-market
vans	17,851,394	299	3,991	Sportswear
VEJA	521,242	776	4,398	Mass-market
Victoria's Secret	68,756,526	249	1,223	Mass-market
Volcom	1,486,942	409	1,267	Sportswear
SAINT LAURENT	8,226,590	1	1,446	Luxury
ZARA Official	41,237,203	57	3,137	Mass-market

## 2.2. Collect Post Data Via Instaloader

To analyse metadata and content of posts and comments, we need to get the data. Unfortunately, Instagram does not offer a public API for that, maybe due to the data scandal of its parent company Facebook in 2018 [19]. Therefore, we had to find different ways how to acquire the content.

We use a customized version of Instaloader [20] for downloading the Instagram posts. Unfortunately, Instaloader stopped working on cloud servers as Instagram blocks the corresponding IP addresses.

To overcome this limitation, we set up an infrastructure of distributed proxy servers using Squid [21] on Raspberry PIs [22] which is distributed among different home networks. For easier handling, they connect to a central OpenVPN [23] hub which hosts another Squid proxy working as client with the Raspberry PIs as parents.

This distributed infrastructure has proved to work very well. New exit nodes can be setup very easily and connect automatically to the VPN extending the capacity of the network.

### **2.3. Collect Profile Data Via HTML Profile Page Download**

Unfortunately, Instaloader was not able to download all profile information which we needed. Therefore, we implemented a proprietary solution for downloading the Instagram profiles. It works by downloading the profile page and parsing the JSON contained on the page itself. This makes the process stable against frequently occurring layout changes by Instagram.

As the profile pages suffer from the same limitation as the post pages, the data is collected via our private proxy network.

### **2.4. Update Regularly**

The data set is updated regularly. Profiles are downloaded daily to have an always up-to-date number of followers and follow information. This information is kept to have a history and allow reports regarding successful marketing campaigns.

Posts and comments are downloaded weekly going back four weeks since the download date. This means that comments which go to older posts will not be downloaded to save Instagram requests.

### **2.5. Data Formats**

All data is saved in its original format. The HTML profile pages are kept for safety reasons (format might have changed or additional information might be available). The posts, comments etc. are saved in an intermediate JSON [24].

After each successful download, data is transferred to an SQLite [25] database for easier access.

### **2.6. Data Volume**

Our data was collected over a large time interval. The first post was created at Feb 9, 2011 (shortly after the start of Instagram) by Michael Kors [26]. The latest post was from Oct 20, 2020; in total we downloaded 312,131 posts over this time interval.

Apart from posts, we also collected the corresponding comments which in total amount to 41,399,076.

## **3. ANALYSIS PROCESS**

### **3.1. Clean Data**

Bots are a very common phenomenon in Instagram. Bots are mainly used for creating visibility in their own account or promoting third party accounts.

For most of our analysis, bots are not really relevant as their impact vanishes statistically. However, for some KPIs like seconds to first comment, it is essential to remove them.

For this, we searched for all profiles which have ever posted comments in less than 60 seconds after the post itself. Additionally, we calculated the frequency of these events and the average number of seconds it took them to write a comment. Combining these values, we could eliminate almost all of the bots.

### **3.2. Data Analysis**

We performed our data analysis in Jupyter notebooks [27] using the Python [28] programming language and several libraries like Pandas [29], SciPy [30] and Scikit Learn [31]. These are standard libraries which have been tested extensively.

The Jupyter notebook connects directly to the SQLite database.

All analysis run automatically and generate (SVG) diagrams as artefacts. This is essential in our process as data changes continuously. We will keep an updated version of the diagrams on Github.

Our process completely avoids Excel for intermediate data and therefore is completely reproducible and safe against regressions.

## **4. RESULTS**

### **4.1. Descriptive Data Analyses**

To start the exploration of the big data set the following analyses are descriptive focusing on the number of posts in total, comments and average number of posts per day. In order to clearly visualize differences between fashion categories (see Table 1), each category was assigned to a colour.

#### **4.1.1. Number of Posts and Average Comments**

As shown in Figure 1 Primark has by far the most posts. As shown in Figure 1, other mass market-brands are evenly distributed regarding the number of posts (mean value = 3469, standard deviation = 2920). Whereas most of the luxury brands are located in the upper field of posts (mean value = 3980, standard deviation = 2333). Concerning sportswear brands the number of posts has an average of 2289 (standard deviation = 1275), which is significantly lower ( $p = 0.03$ ) than the two other categories.

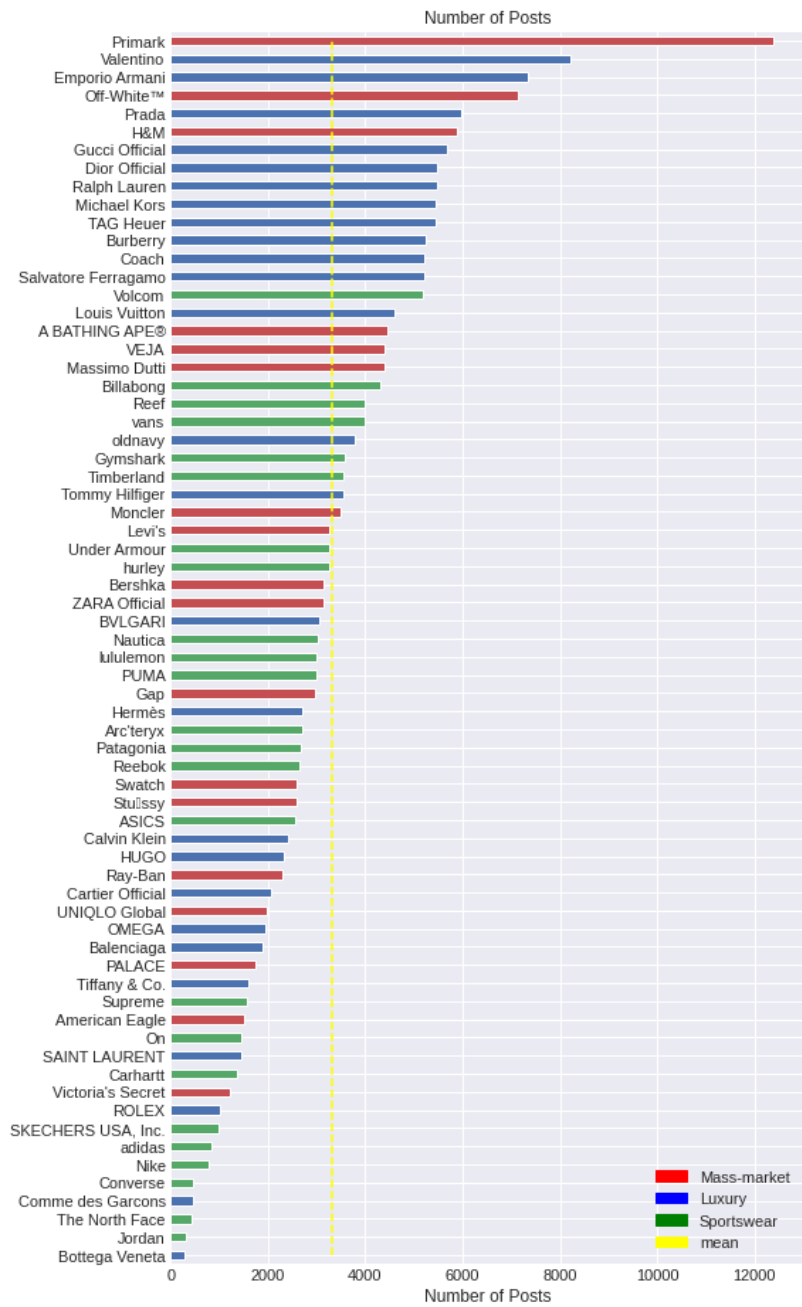


Figure 1: Average number of posts for each brand

Of course, the reason for this could be that Primark is using Instagram much longer than the other brands. This turns out not to be true given the number of posts per day (Figure 2). The distribution is similar to the number of posts in total. The three categories differ significantly ( $p = 0.004$ ), with sports brands posting mainly below average per day and in total (mean = 0.9, standard deviation = 0.3) and with great distance to luxury (mean = 1.5, standard deviation = 0.9) and mass-market brands (mean = 1.6, standard deviation = 1.1).

## Computer Science &amp; Information Technology (CS &amp; IT)

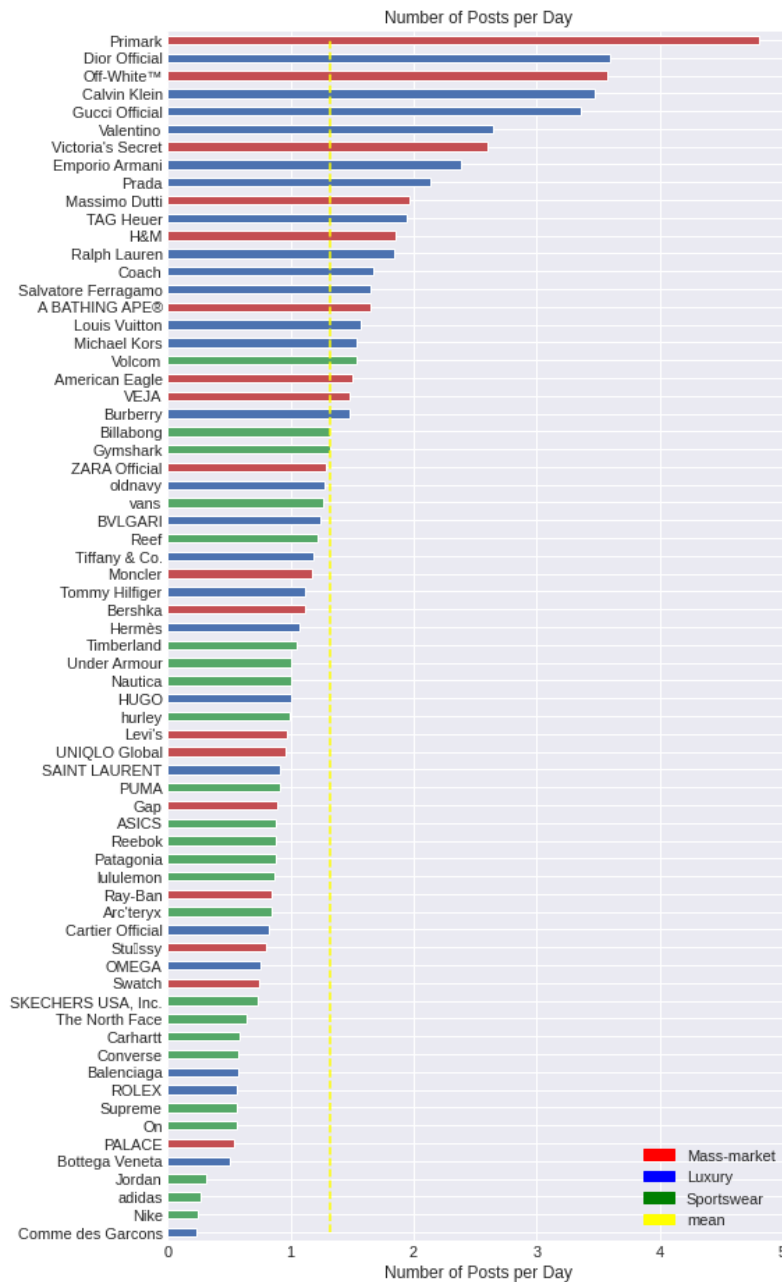


Figure 2: Average number of posts per day for each brand

Primark is a very active poster with almost five posts per day. Compared to the top sports brands with roughly one post each five days, which is more than 20 times the post volume. It will be interesting to see the effects of these extremes on community engagement.

Does the number of posts also have direct implications on the number of comments? In other words, is the engagement of the community stronger if they can read more posts? This could well be possible as users tend to look more often at profiles with very frequent changes

To abstract this from the number of posts, we now take a look at the average number of comments per post:



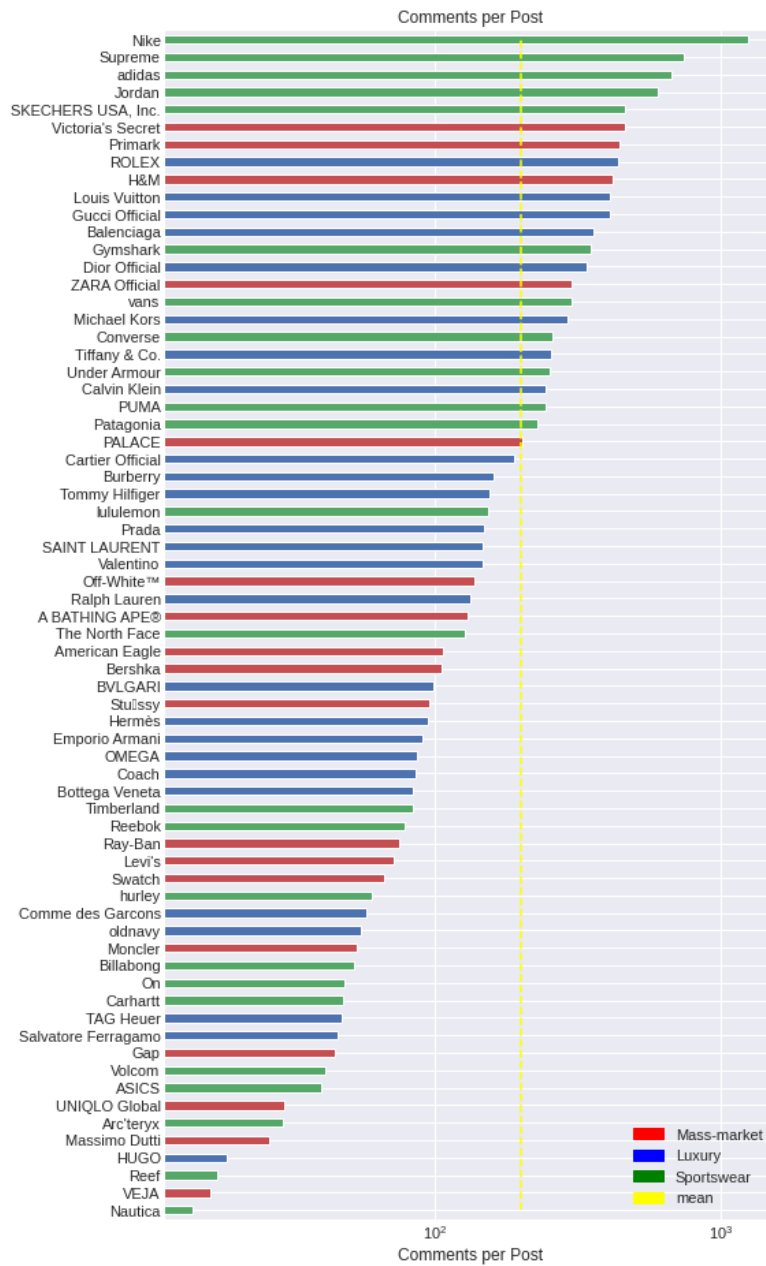


Figure 3: Average comments per Post for each brand

The results show a completely different ranking with Nike on the top generating 1253 comments per post on average

Contrary to the results concerning the posts of brands themselves, concerning community engagement sportswear brands have the most active users with 257 comments per Post on average (standard deviation = 302). Posts of luxury brands trigger 177 comments on average (standard deviation = 127), whereas mass-market brands follow with 155 comments per post (standard deviation = 149).

Still, Primark has a solid number of comments per post. So, the inverse is also not true; the community does not get bored by many posts.

#### 4.1.2. Time to First Comment

In the next step, we considered the time it takes each profile to attract the first comment as some brands take this as an indicator for their community's engagement. This turns out to be a difficult approach as there are many bots on Instagram which automatically comment on popular posts to gain popularity and increase their reach.

As shown in Figure 4 comments appear fastest for luxury brands with about 1442 seconds (24 minutes) in average (standard deviation = 2364 sec. – 39 minutes), followed by mass-market (mean = 1742 sec. – 29 minutes, standard deviation = 2175 sec. – 45 minutes). It seems that the sportswear community is slowest in terms of commenting on posts (mean 1908 sec. – 32 minutes, standard deviation = 1890 sec. – 32 minutes).

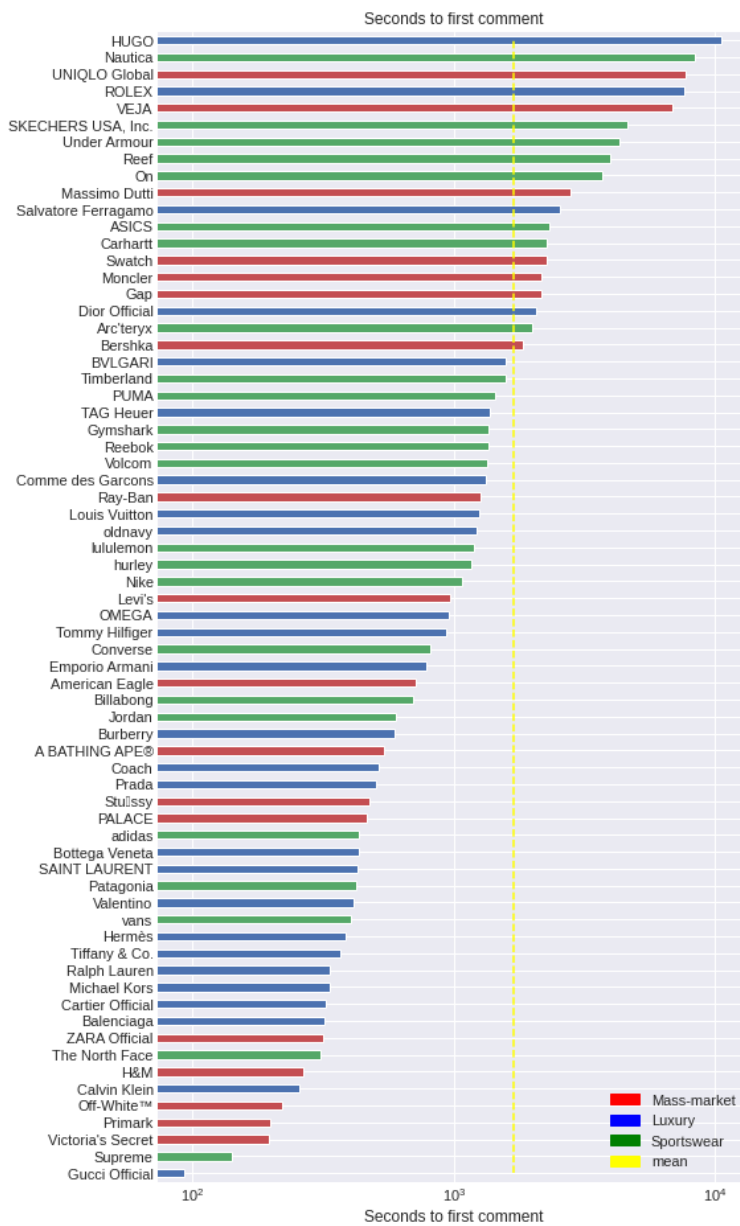


Figure 4: Seconds to first comment after post for each brand

4.1.3. Number of Likes Per Post and Follower

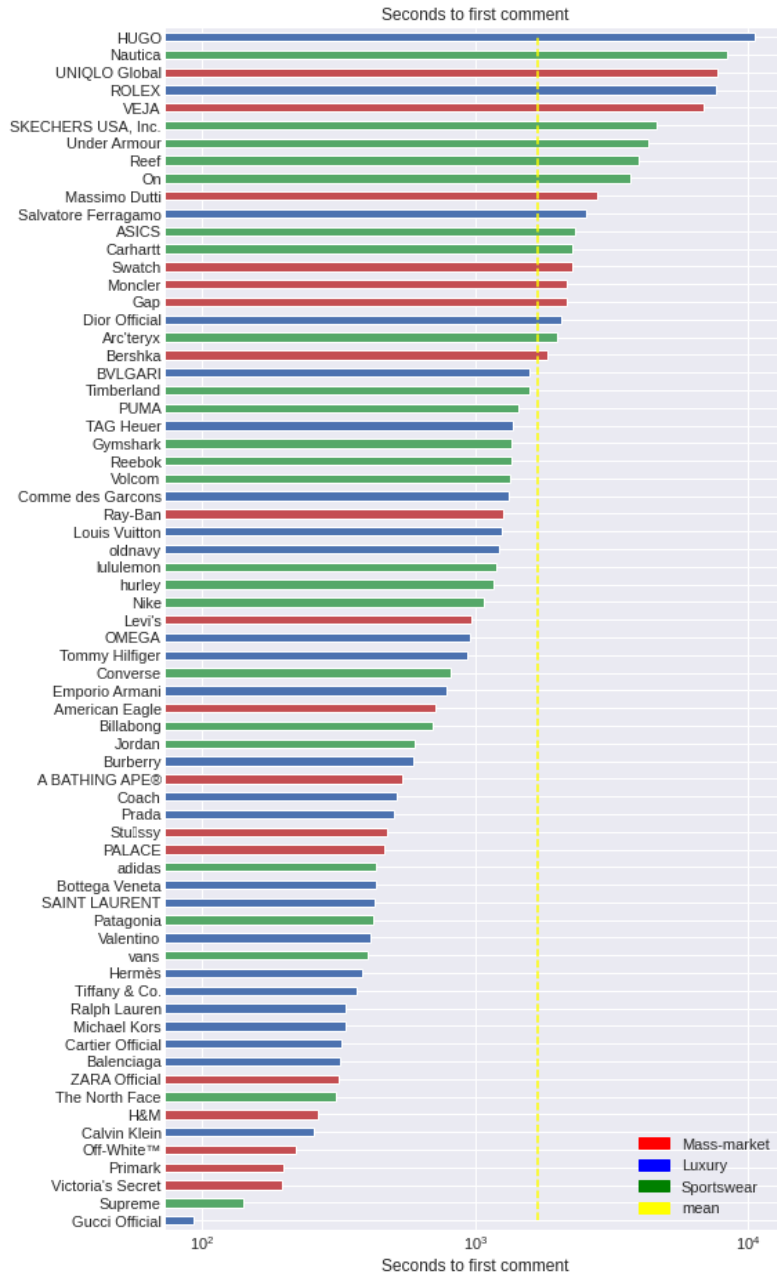


Figure 5: Average likes per post and follower for each brand

The last descriptive analysis refers to a normalized indicator by putting likes in relation to posts and followers for each brand. Figure 5 shows clearly that sportswear brands have the most active community (mean = 0.005, standard deviation = 0.002), followed by mass-market brands (mean = 0.004, standard deviation = 0.002) and luxury brands (mean = 0.003, standard deviation = 0.001).

Finally over all descriptive analyses the high variances limit the informative value concerning average indicators. Therefore we proceed with a deep dive regarding the Instagram corpus of our selected brands by analysing correlations and AI-based text analyses.

## 4.2. Correlation Analyses

### 4.2.1. Relation between Number of Followers and Time to First Comment

So, time to first comment must somehow depend on the community and engagement itself, not just on its size? It must also be taken into account that the historical time to the first of course depends on the number of followers at that time which is not available to us.

Figure 6 shows the correlation between the number of followers and the time to the first comment for each brand which is  $r = -0.24$  ( $p = 0.05$ ). That indicates that there is a significantly weak negative relationship between the both variables.

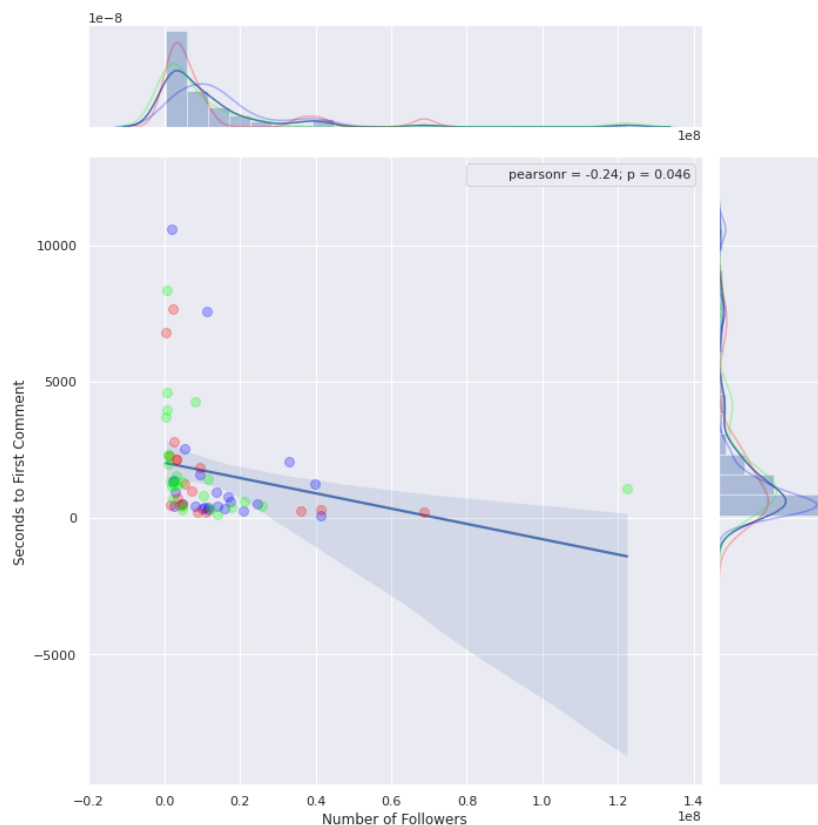


Figure 6: Jointplot correlation between number of followers and seconds to first comment

### 4.2.2. Relation between Likes and Comments

The correlation between the number of comments per post and the number of likes per post is strongly positive and highly significant ( $r = 0.86$ ,  $p = 3.3e-21$ ), which is quite intuitive: the more likes a post generates, the more comments will be made and vice versa.

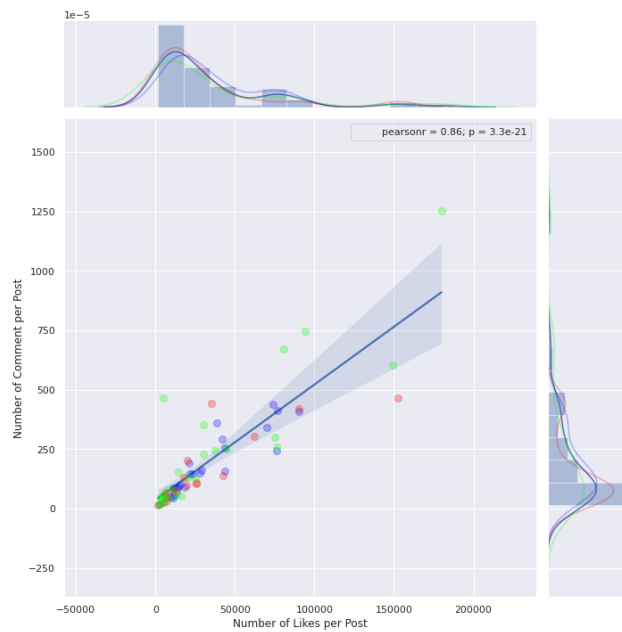


Figure 7: Jointplot of correlation between number of comments per post and numbers of likes per post

#### 4.2.3. Relation between Number of Followers and Comments and Likes

As comments and likes are more or less linearly related, it suffices to consider comments. Likes are the same.

The strong positive correlation ( $r = 0.77$ ,  $p = 9.3e-15$ ) between number of followers and number of comments per post is also quite intuitive: The more followers a brand has the more comments can be expected per post.

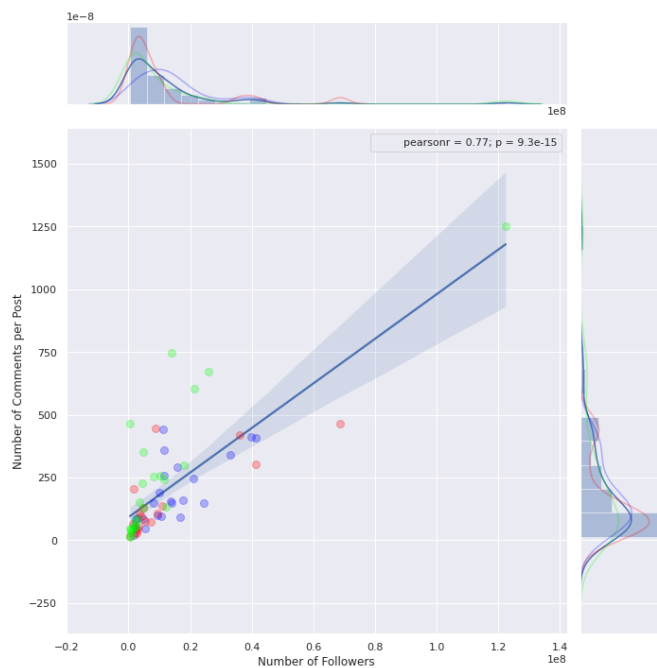


Figure 8: Jointplot of correlation between number of followers and number of comments per post

As there is one outlier brand in the category sportswear (i.e. Nike), we separated this category for a closer look on the relationship number of followers with number of likes per post:

The correlation is stronger ( $r = 0.9$  vs.  $r = 0.8$ , both highly significant) excluding Nike as an outlier regarding the number of followers. That leads to the interpretation that Nike despite the high number of followers has a substantive amount of followers who is not active at all (regarding likes or posting comments).

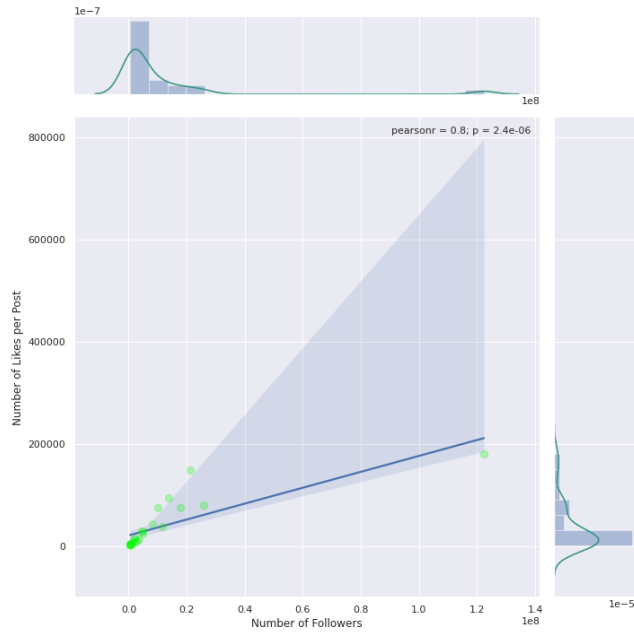


Figure 9: Jointplot of correlation between number of followers and number of likes (including outlier Nike)

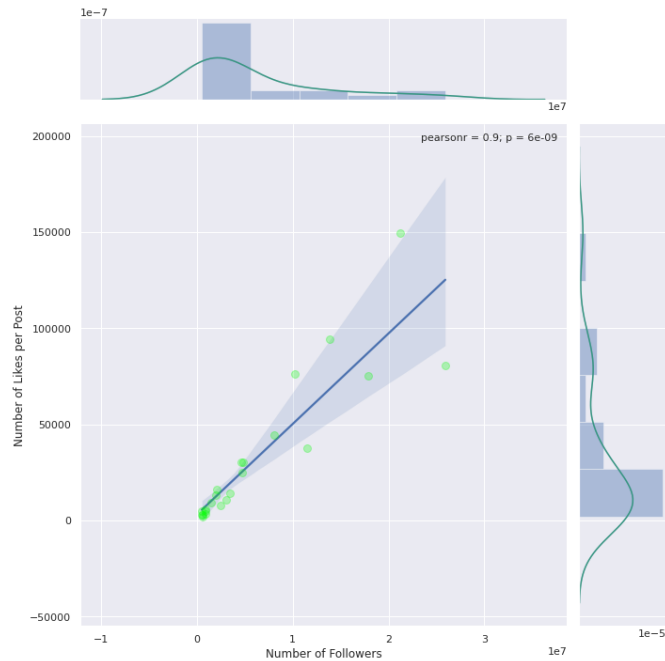
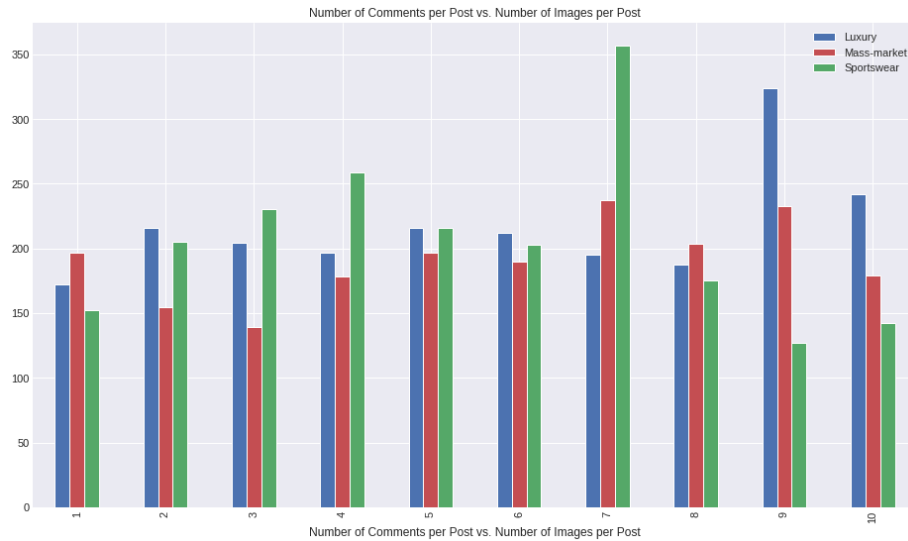


Figure 10: Jointplot of correlation between number of followers and number of likes per post (excluding outlier Nike)

#### 4.2.4. Relation between Number of Comments and Video Vs. Images

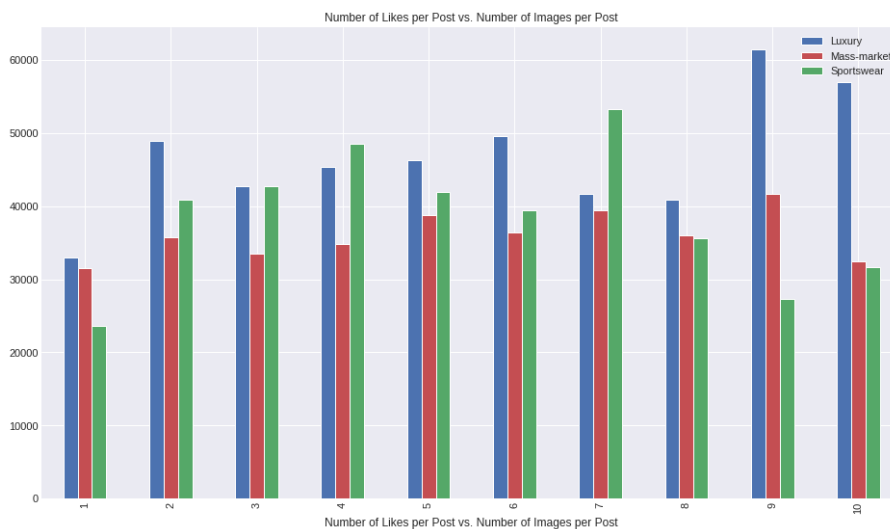
As first noted by Bonilla et al. [32] the number of videos and images in a post has a significant effect on popularity and community interaction.

To see this effect in our dataset, we first consider the number of images in a post plotted against the average number of comments:



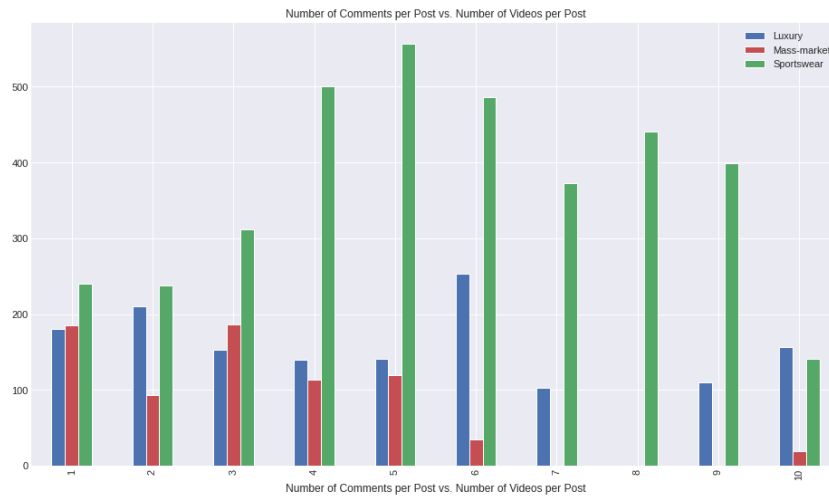
Interestingly, the results depend on the category of the fashion brand. The sportswear community honours the number of images and comments much more frequently on posts with multiple images, reaching a peak at seven images per post. The commenting behaviour in the other categories is a bit more erratic with no clear trend visible.

The story for likes is a bit different:



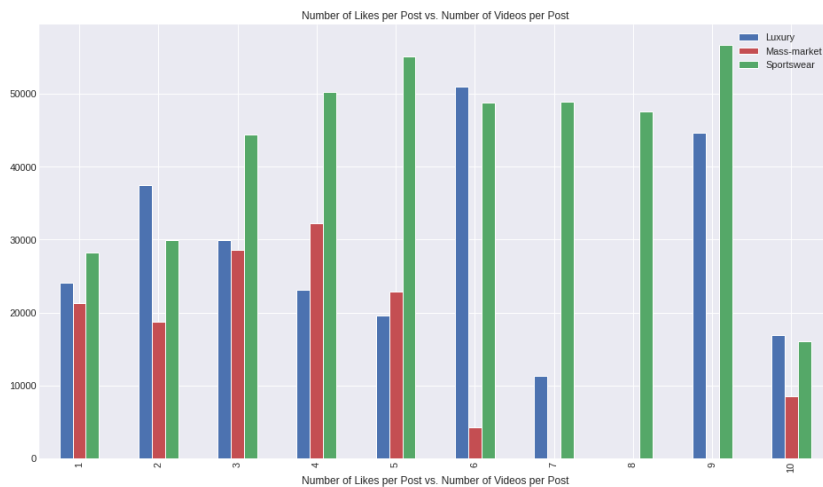
The mass market brands do not get many more likes with increasing number. But both sportswear and luxury brands have a positive effect for likes when posting with a larger number of images.

Videos are much more complicated and expensive to produce and have a similar effect depending on the category of the brand:



Comments increase specifically in the sportswear category with no clear trend in the other categories.

Likes are increased both in sportswear and in to a lesser extent in the luxury category:



### 4.3. Topic Modelling of Comments

Topic modelling is a technique of artificial intelligence to uncover the structure of a large text corpus [33]. There is a large variety of methods available [33]; for our analysis using non-negative matrix factorization [33] proves to be efficient and provides good results.

We are not so much interested in the posts of the brands because they focus on their own names and values. Rather, we would be interested in the much larger number of comments which are given by the followers (and fans) of the respective brands.

The number of comments is much too large for calculating a complete topic model over all comments. Additionally, that would not be fair as it has a bias toward the brands with a lot of comments. Therefore, we chose a stratification approach and took only 10,000 comments from



each brand amounting to a total of 680,000 comments which is enough for calculating valid topic models.

We chose to calculate 10 topics using the NMF method. The topics are shown as wordclouds in the figure below:



Figure 11: Topic Modelling

Instagram is not a platform which is renowned for extensive content. As expected, the topics contain a lot of generic words, but still offer quite distinctive feedback regarding the products of the respective brands.

In the next step, we wanted to know the top 5 brands for each of the topics. For this, we calculated the topic distribution for each brand and took only the top 5 brands for each topic:

Table 2: Ranking of Topics and corresponding brands

	top 1	top 2	top 3	top 4	top 5
love	SKECHERS USA, Inc.	VEJA	Michael Kors	American Eagle	Gucci Official
nice	Emporio Armani	OMEGA	Nautica	TAG Heuer	Stüssy
beautiful	Dior Official	Victoria's Secret	Tiffany & Co.	Calvin Klein	BVLGARI
cool	hurley	Nike	Volcom	Ray-Ban	Stüssy
want	A BATHING APE®	Supreme	Primark	PALACE	Swatch
like	A BATHING APE®	Jordan	PALACE	Supreme	Under Armour
amazing	Arc'teryx	hurley	The North Face	Billabong	Patagonia
awesome	Billabong	Patagonia	Arc'teryx	hurley	Volcom
cute	oldnavy	Primark	American Eagle	Gap	ZARA Official
great	On	Arc'teryx	The North Face	TAG Heuer	Patagonia

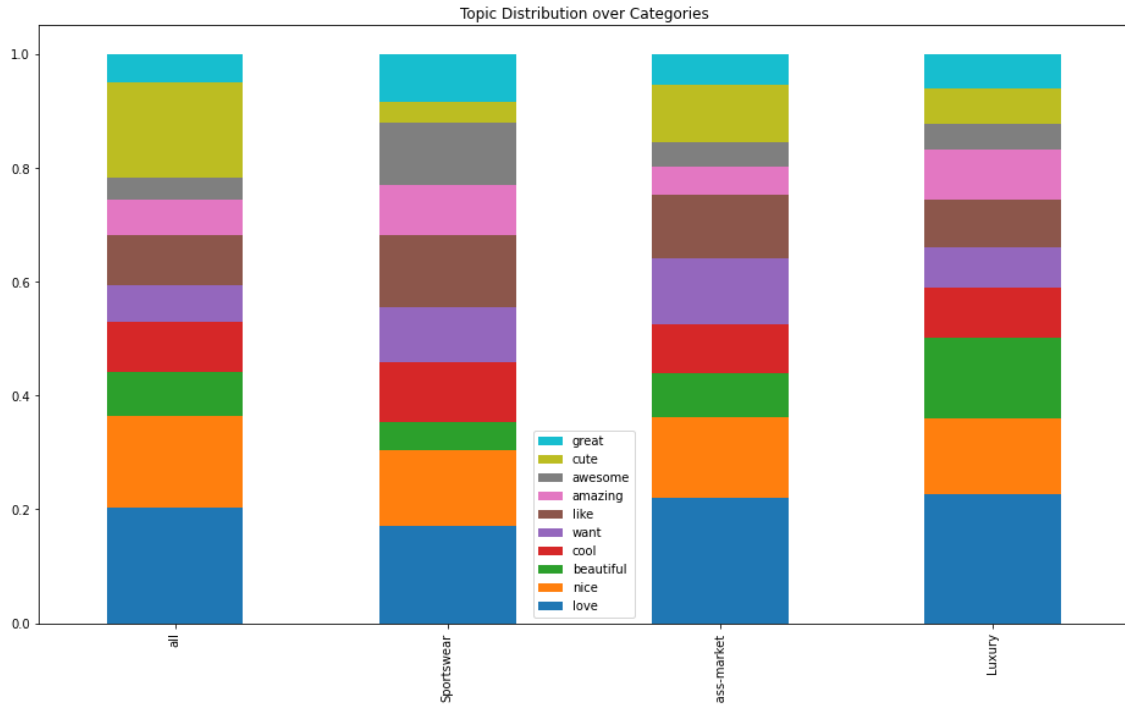
The data can also be visualized more qualitatively but in more detail using a heatmap. The order of the brands has been adjusted to have the luxury brands first, then the mass market and finally the sports brands:



Figure 12: Heatmap of topics and their appearance in brands' UGC

In this representation, it is quite easy to see that cute is most prominent in oldnavy, love is most popular for SKECHERS etc.

In a final step, we calculated the topic distribution of the different categories:



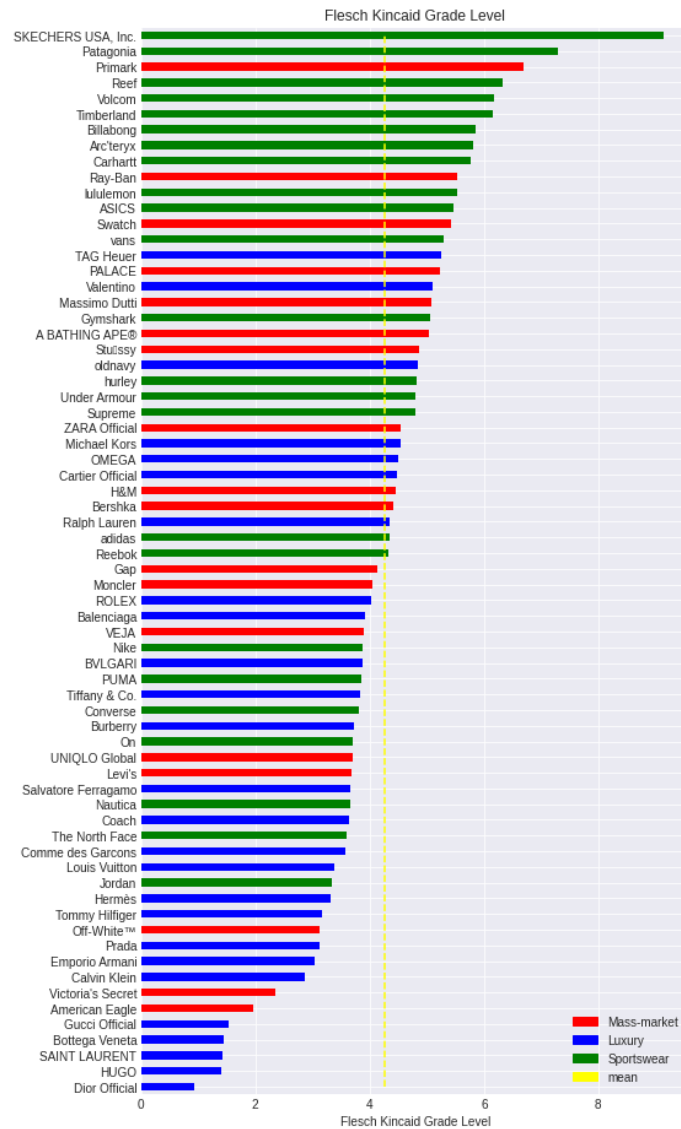
As expected, the “cute” topic is not much associated with “sportswear” but rather with mass-market. The “beautiful” topic is mainly present in the luxury brands, which is also according to our expectations. All in all it looks like the different categories reach their target audience quite well with the feedback going into the same direction as the primary

#### 4.4. Text Complexity in Comments

For analysing the complexity we used the Flesch-Kincaid readability tests [33]. The calculation was performed using the Python package textacy [34]. As the necessary linguistic analysis is quite expensive, we used the same stratified dataset as in topic modelling.

It’s very interesting to see that the complexity of the comments is much higher in sportswear compared to luxury brands. This can be attributed to many “empty” discussions just containing emojis or trying to gain popularity, which is much more common in the luxury regime.

Taking samples from sportswear, it became clear that the discussions there are much more centred around the presented products and commenters exchange tips regarding their gear. This of course contributes to much more complicated text.



## 5. CONCLUSIONS AND OUTLOOK

In this paper, we have shown how Instagram works for fashion brands as a marketing platform and how their individual success can be measured. Coming from descriptive statistics, we have used machine learning models like topic models to quantify the response of the community.

Comments and likes were shown to be good indicators of the success of various brands. We have shown that the engagement of the community depends heavily on the kind of brand with a totally different response from sportswear compared to luxury brands. Interestingly, the comments for sportswear brands have a much higher linguistic complexity compared to luxury brands. However, this has to be taken with a grain of salt as the Instagram comments are always quite short.

Our process was designed in a way to easily facilitate easy updates. All data is aggregated in a central database and used for automatically creating all diagrams. This leads to a scalable, reproducible analysis which is essential if you use the data to plan your (expensive) marketing campaigns.

In the future, we plan to extend the analysis to analyse the Emoji content of the posts and mainly the comments. New work regarding Emoji embedding [35] allows us to find uncommon Emojis and take a specific look at that. Additionally, we can use pre-classified Emojis [36] to have a hierarchy of feelings associated with individual Emojis.

Apart from starting with the brands, there are many other ways to approach Instagram content. We have already started with community detection to find special interests also within the fashion industry (e.g. sneakers). Experts will also convey weak signals and this can be used for very early trend detection.

Yet another possibility is given by local Instagram content. Many posts are annotated with a geolocation. Eventually, this will allow the detection of local and hyperlocal trends.

## REFERENCES

- [1] Heinonen, K., Medberg, G. (2018). "Netnography as a tool for understanding customers: implications for service research and practice", *Journal of Services Marketing*
- [2] Kozinets, R. V. (1999). „E-tribalized marketing?: the strategic implications of virtual communities of consumption“, *European Management Journal*, 1999, vol. 17, issue 3, 252-264
- [3] Bickart, B., & Schindler, R. M. (2001). „Internet Forums as Influential Sources of Consumer Information“. *Journal of Interactive Marketing*, 15, 31-40
- [4] Catterall, M., Maclaran, P., Stevens, L. (2002). "Consumption and Gender", in *GCB - Gender and Consumer Behavior Volume 6*, eds. Pauline Maclaran, Paris, France: Association for Consumer Research.
- [5] Kozinets, R. V. (2006). „Click to Connect: Netnography and Tribal Advertising“. *Journal of Advertising Research*, vol. 46, no. 3, 279–288
- [6] Simmons, G. (2008). „Marketing to postmodern consumers: Introducing the Internet chameleon“, *European Journal of Marketing* 42(3/4)
- [7] Tikkanen, H., Hietanen, J., Henttonen, T., Rokka, J. (2009). „Exploring virtual worlds: Success factors in virtual world marketing“. *Management Decision* 47(8):1357-1381
- [8] Rokka, J. (2010). „Netnographic inquiry and new translocal sites of the social“. *International Journal of Consumer Studies*, Volume34, Issue4, July 2010, 381-387
- [9] Musabirov, I., Bulygin, D. (2020). „Prototyping Text Mining and Network Analysis Tools to Support Netnographic Student Projects“. *International Journal of Emerging Technologies in Learning*, Vol. 15, No. 10 (2020)
- [10] Statista (2020), Most popular social networks worldwide as of July 2020, ranked by number of active users, [Online]. Available: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>
- [11] Later (2020), The Top 50 Instagram Hashtags of All Time, [Online]. Available: <https://later.com/blog/ultimate-guide-to-using-instagram-hashtags/#top>
- [12] Leaver, T., & Highfield, T. (2018). Visualising the ends of identity: Pre-birth and post-death on Instagram. *Information, Communication & Society*, 21(1), 30–45.
- [13] Roncha, A., & Radclyffe-Thomas, N. (2016). How TOMS' "One day without shoes" campaign brings stakeholders together and co-creates value for the brand using instagram as a platform. *Journal of Fashion Marketing and Management*, 20(3), 300–321.
- [14] Paine, K. D. (2011). *Measure what matters: Online tools for understanding customers, social media, engagement, and key relationships*. Hoboken, New York: John Wiley and Sons.
- [15] Dhanesh, G.S. (2017), "Social media and the rise of visual rhetoric: implications for public relations theory and practice", in Bridgen, E. and Vercic, D. (Eds), *Experiencing Public Relations*, Routledge, New York, NY, pp. 137-150.
- [16] Mersey, R., Malthouse, E.C. and Calder, B.J. (2010), "Engagement with online media", *Journal of Media Business Studies*, Vol. 7 No. 2, pp. 39-56.
- [17] Muntinga, D.G., Moorman, M. and Smit, E.G. (2011), "Introducing COBRAs: exploring motivations for brand-related social media use", *International Journal of Advertising*, Vol. 30 No. 1, pp. 13-46.

- [18] FashionUnited (2020), „Most valuable fashion brands“ [Online]. Available: <https://fashionunited.com/i/most-valuable-fashion-brands>
- [19] BBC (2019), Facebook staff 'flagged Cambridge Analytica fears earlier than thought', [Online]. Available: <https://www.bbc.com/news/technology-47666909>
- [20] GitHub (2020), Instaloader, [Online]. Available: <https://instaloader.github.io/>
- [21] Squid (2020), Optimising Web Delivery, [Online]. Available: <http://www.squid-cache.org/>
- [22] Raspberry Pi 4 (2020), [Online]. Available: <https://www.raspberrypi.org/>
- [23] OpenVPN (2020), Building a Strong Community, [Online]. Available: <https://openvpn.org>
- [24] JSON (2020), JavaScript Object Notation, [Online]. Available: <https://www.json.org>
- [25] SQLite (2020), What Is SQLite?, [Online]. Available: [www.sqlite.org](http://www.sqlite.org)
- [26] Instagram (2011, Profile Michael Kors, [Online]. Available: <https://www.instagram.com/p/Bfs65/>
- [27] Project Jupyter (2020), The Jupyter Notebook, [Online]. Available: [www.jupyter.org](http://www.jupyter.org)
- [28] Python (2020), Compound Data Types, [Online]. Available: [www.python.org](http://www.python.org)
- [29] NumFOCUS (2020), pandas, [Online]. Available: <https://pandas.pydata.org/>
- [30] NumFOCUS (2020), SciPy, [Online]. Available: <https://www.scipy.org/>
- [31] Scikit-Learn (2020), Machine Learning in Python, [Online]. Available: <https://scikit-learn.org/stable/>
- [32] María Del Rocío Bonilla, José Luis del Olmo Arriaga & David Andreu (2019) : The interaction of Instagram followers in the fast fashion sector: The case of Hennes and Mauritz (H&M), Journal of Global Fashion Marketing]
- [33] Albrecht, J., Ramachandran, S. and Winkler, C. Blueprints for Text Analysis Using Python, Sebastopol: O'Reilly, 2020
- [34] textacy 0.10.1 (2020). textacy: NLP, before and after spaCy, [Online]. Available: <https://pypi.org/project/textacy/>
- [35] Al-Halah, Z., Aitken, A., Shi, W., Caballero, J. (2019). „Smile, Be Happy :) Emoji Embedding for Visual Sentiment Analysis“ [Online]. Available: <https://arxiv.org/abs/1907.06160>
- [36] Unicode (2020). „Full Emoji List, v13.1“ [Online]. Available: <https://unicode.org/emoji/charts/full-emoji-list.html>

## AUTHORS

### Prof. Dr. Stefanie Scholz

Professor for social economy at Wilhelm Loehe University of Applied Sciences, Fuerth, Germany. Focussed on data driven marketing, netnography, consumer behaviour and customer empowerment



### Dr. Christian Winkler

Christian Winkler holds a PhD in Theoretical Physics. He has worked in software and AI for 20 years, specializing in intelligent algorithms for unstructured data and text. He is a frequent speaker at conference and author of many articles and tutorials.

