FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

ASSESSING THE IMPACT OF INFLUENCERS ON BRAND ENGAGEMENT ACROSS DIFFERENT CATEGORIES ON INSTAGRAM

Aantal woorden / Word count: 28.009

Hanne Vandermarliere
Stamnummer / student number : 01506165

Mélanie Vercaempt
Stamnummer / student number : 01510435

Promotor / supervisor: Prof. Dr. Dirk Van den Poel

Masterproef voorgedragen tot het bekomen van de graad van:
Master's Dissertation submitted to obtain the degree of:

Master in Business Engineering: Data Analytics

Academiejaar / Academic year: 2019-2020

UNIVERSITEIT
GENT

Abstract

In current literature, the effects of influencer marketing with respect to brand engagement are vastly understudied. This thesis tries to remedy the lack of research in this area by looking at the impact influencers and their post characteristics have on brand engagement on the social media platform Instagram. Moreover, it is investigated whether there are significant differences in performance across the three most used Influencer marketing product categories: fashion, beauty and food. While previous studies have investigated the influence of different post characteristics of the brand on engagement on SNSs, this is the first research conducting an analysis with a comprehensive set of features, including both brand and influencer characteristics, on Instagram. In this paper, a total of 83 brands and 4162 influencers are taken into account. The analysis shows that influencers, both in terms of influencer characteristics and post influencer characteristics, have an additional impact on brand engagement. Specifically, we find that the fashion industry should focus on finding influencers with less than 200K followers, whereas food influencers only require a high engagement ratio and activity level. As for the beauty sector, all these three influencer characteristics are important. The findings further show that fashion influencers should be positive while beauty and food influencers should be honest to enhance brand engagement. In addition, interaction - defined as the use of second pronouns, questions, actionable verbs, hashtags and the mention of the brand in the influencer's caption - has a considerable impact on brand engagement. For the fashion industry, the level of interaction is only important in the influencer's posts, while for beauty and food it is important for both the influencer's and brand's posts. The relationships of the various interaction aspects are different regarding the impact on likes and comments across the categories. The only interaction aspect that is positively related to both likes and comments for all categories is mentioning the brand in the influencer's caption. The findings of this study are of great importance to managers because the brand engagement can be optimized by formulating appropriate guidelines for influencers.

Confidentiality Clause

We declare that the content of this Master's Dissertation may not be consulted nor reproduced.

Name students: Hanne Vandermarliere and Mélanie Vercaempt

Signature:

Preface

This section is dedicated to express our gratitude to everyone who helped us complete this research project or sparkle our interest for it. Without their influence this research would not have existed.

This master's dissertation has been a huge learning experience for both of us. It was certainly a challenge, but we are very grateful that we had the opportunity to discover the area of influencer marketing, a subject that really interests us.

We want to thank Prof. dr. Dirk Van den Poel for mentoring us throughout the thesis, his guidance has been of great value in completing this study. Furthermore, we would like to express our gratitude towards him for teaching us the interesting subjects of the data world. Data Analytics was relatively new to us two years ago, but we can now say that we have built a strong foundation for our first work experience. Especially an extra thank you for scraping and monitoring the gigantic volume of Instagram data and for delivering it in a fast and structured way. We are aware of the fact that this must have been a very time consuming task.

We also want to thank Sarah Carron for putting us on the road by providing us the json inread files. These files immediately pushed us in the right direction and saved us a lot of time. Besides, we want to thank her for the insights she provided and her guidance at the first stages of our journey.

Finally, we would like to thank our friends and family for their encouragement and interest in our master dissertation. We would like to thank our parents for all the chances and opportunities they have given us. Without them, we would not be able to finish our Business Engineering study. Especially, we want to thank Alexander and Thomas for proofreading our thesis and their continuous support.

Hanne Vandermarliere and Mélanie Vercaempt

Ghent, 2 June, 2020

Abbreviations

AUC Area Under the ROC Curve

BB Brand Basetable *(Specific to this study)*

CFB Compressed Final Basetable *(Specific to this study)*

e-WOM Electronic Word-Of-Mouth

FPR False Positive Rate

IB Influencer Basetable *(Specific to this study)*

MB Merged Basetable *(Specific to this study)*

MDGI Mean Decrease in Gini Index

PDP Partial Dependence Plot

ROC Receiver Operating Characteristic

SNS Social Network Site

TPR True Positive Rate

UGC User Generated Content

WOM Word-Of-Mouth

List of Figures

List of Tables

Table of Contents

1 Introduction

1.1 Social media

Social media nowadays has a profound impact on people's lives (Alalwan et al., 2017; Dwivedi et al., 2015). On the one hand, users of social networking sites (SNSs) generate content like written posts, photos or videos and conversations between users. On the other hand, the technological platform also includes the connection, creation, sharing, tagging and editing of content online (Boyd & Ellison, 2007).

The expansion of SNSs compared to ten years ago can be supported by the statistics of DataReportal (2020) stating that more than three billion social media users are currently active around the world, representing about 50% of the world population. Several studies have shown that SNSs have an important impact on communication. They have become the most important element in developing an online network (Lee et al., 2015). Not to mention, SNSs bring substantial and omnipresent changes in the way organizations, communities and individuals communicate (Kietzmann et al., 2011). In addition, since smartphones are ubiquitous, people are connected with SNSs at any time and at any place.

This has been confirmed more than ever during the COVID-19 pandemic in which many countries are confronted with a lockdown. During this lockdown, the streets are empty and people are bored. To counter the boredom, people make challenges on social media and tag their friends to join these challenges (Müller, 2020). This gives social media the opportunity to expand further. Already 21% more people are using social media as of April 2020 (Statista, 2020a). Moreover, companies are forced to let their employees work from home. This way, SNSs become more important than ever in order to keep in touch and organize online meetings. Unfortunately, COVID-19 is not yet out of the world. Dr. Bruce Aylward[1] claims in the online newspaper Time (2020) that the scenario where COVID-19 might disappear completely is very unlikely. That means, social media is and will retain an important role in communication.

Given the aforementioned reach and impact of SNSs, companies are scrambling to figure out how to use SNSs in order to reach the millions of (potential) consumers who use it on a daily basis. For brands, SNSs offer the potential to 1) advertise and therefore enhance brand awareness, 2) observe and analyze user generated content (UGC), and 3) interact with end customers to let the consumer

[1]Senior adviser to the Director-General of the World Health Organization (WHO)

have influence on the product development (Richter et al., 2011). However, in a world cluttered with information coming from a variety of sources like television, newspapers and SNSs, it becomes a great challenge for brands to stand out. Fortunately, influencer marketing has proven to be a valid channel for brand promotion.

1.2 Influencer Marketing

Since the emergence of Web 2.0, consumers are able to share their experiences and opinions of products and brands in the form of an online review on SNSs (Filieri, 2015). This refers to the concept of e-WOM which is essential for brands. This is because potential customers are visiting websites and reading peer-to-peer reviews to earn more information about a product before making an actual purchase decision (Doh & Hwang, 2009). With influencer marketing, the brand can positively manipulate this online buzz.

Influencers are social media users with a wide reach, who share their opinions about products and/or services of a particular brand on their SNS account. This is done by referencing the brand in the influencer's posts or by participating in larger advertisement campaigns and events. In either way, influencers try to convince their followers to buy the product or service of the brand they have a collaboration with (Abidin, 2016; Domingues Aguiar et al., 2018). There are various methods by which influencers can promote products or services. Some of the most used are: product placement, discount codes and recommendations (Zeren & Gökdağlı, 2020).

The advantage of influencer marketing compared to earlier traditional marketing techniques is that influencers are seen as "regular people" being personal, authentic, trustworthy and easy-to-relate-to (De Veirman et al., 2017; Djafarova & Rushworth, 2017; Lou & Yuan, 2019; Schouten et al., 2019; Oliveira et al., 2020).

The rewards the influencers get range from free products to a substantial monetary compensation. Influencers in the USA get on average around 2000 dollars per advertisement post (Statista, 2019). Although influencers can ask high prices per post, influencer marketing is considered being the most cost-efficient and -effective marketing tool (Harrison, 2017; Patel, 2016; Talaverna, 2015). The report of Launchmetrics (2019) shows that nearly 76% of survey respondents declare better sales results because of their collaboration with influencers. As reported by many companies of that study, influencers can help enhance brand awareness, reach new target audiences and increase sales growth. The use of influencers has shown to have positive results in terms of consumer persuasion

(Booth & Matic, 2011) and return on investments (Solutions, 2016). However, it is essential for brands to identify influencers that match their brand image in order to achieve those benefits. (Booth & Matic, 2011; Huang et al., 2014; Naumanen et al., 2017).

Given these numbers and advantages, it is obvious that influencer marketing nowadays can be seen as a vital part of advertising. While it turned out to have a significant impact on sales and return on investments, little research has been conducted regarding its influence on online brand engagement. This is rather surprising as brand engagement appears to be the most used measure of influencer effectiveness (Launchmetrics, 2019). To fill this void, this quantitative research will cover how effective influencers are regarding brand engagement, in terms of likes and comments on Instagram.

1.3 Instagram

Instagram is a photo and video based social platform founded in October 2010 by Kevin Systrom and Mike Krieger. After one week it already hit a hundred thousand followers. Although Facebook (2.45 billion users) and Youtube (2 billion users) have more users, Instagram (1 billion users) stands out compared to other SNSs, because it is the fastest growing social media platform (Statista, 2020b). Moreover, Instagram was considered the most appropriate and popular platform for influencer marketing in 2019, as reported by Launchmetrics (2019) and Mediakix (2019). Besides that, the statistics of Brandwatch (2019) tell us that of the top 100 brands in the world, 90% has an Instagram account. Furthermore, they show that 70% of the most used hashtags are branded.

The success of Instagram regarding influencer marketing can be explained by the originality and the simplicity in which this platform enables users to share and recognize each other's lives through photos or videos. Instagram helps brands to give customers a "behind the scenes" view. This enhances their relationship with the company and thus, defines marketing-relationship goals to improve the brand (Pittman & Reich, 2016). Furthermore, its built-in filter tool allows users to add their own stamp on their photos which distinguishes Instagram from all other SNSs. Photos and videos on Instagram can easily be uploaded, shared and manipulated in order to get noticed. Moreover, with the use of hashtags, tags, mentions and the discovery page, all Instagram accounts can be connected (Lee et al., 2015).

For these reasons, Instagram is used to assess the impact of influencers on brand engagement. Moreover, it is investigated whether or not this impact differs across the categories fashion, beauty and food.

The remainder of this research is organized as follows. In Section 2, an overview of the current literature and research regarding influencer marketing and brand engagement is provided. In addition, it is explained how this thesis contributes to the existing material. Section 3 and 4 provide more information on respectively the data and the methodologies used in this study. The results and managerial implications can be found in Section 5 and 6. Section 7 contains a brief conclusion of this study and finally, this research is closed with some limitations and opportunities for further work in Section 8.

2 Literature Study

This section discusses relevant theories about influencers and their impact on brand engagement on Instagram. In this research, brand engagement refers to the user engagement with regard to brand posts on Instagram, measured in number of likes and comments. Because Instagram is a relatively new platform, less research has been conducted in this field compared to Twitter and Facebook. Several studies have already been carried out, including L. De Vries et al. (2012), Cvijikj and Michahelles (2013), Sabate et al. (2014) and Luarn et al. (2015), to determine which factors are considered important to generate an engaged brand post on Facebook. Compared to the works mentioned above, this research has one major difference: we presume that the posts of the brand are not solely responsible for user engagement. With the rise of influencer marketing on Instagram, the brand page is viewed more easily. In this study, the dependent variable, brand engagement, remains the same but the independent variables differ because influencer variables are taken into consideration.

The impact of influencers on brand engagement is based upon two aspects: the pure influencer characteristics and their Instagram post characteristics. Therefore, the first research question is formulated:

> **RQ1**: *Does the use of influencers in terms of influencer characteristics (a) and influencer post characteristics (b) have an impact on brand engagement on Instagram?*

Table 1 gives an overview of the research that has already been conducted in the field of influencer marketing and brand engagement. For each author, it is indicated which features in the categories *influencer characteristics*, *post influencer characteristics* and *post brand characteristics* they have examined. In addition, the legend below the table shows on which dependent variable and SNS the research was carried out. It is clear from Table 1 that brand engagement is only investigated in the context of *post brand characteristics*. None of the *influencer characteristics* and *post influencer characteristics* have been examined in terms of brand engagement. Hence, a link between influencers and their impact on brand engagement seems necessary. In addition, none of the studies discussed has taken into account more than four features. Some aspects, like multiple post, have not even been investigated yet, as far as we know. It is therefore necessary to conduct a comprehensive quantitative analysis that takes all these features into account while making the bridge between influencer marketing and brand engagement on Instagram. Moreover, the last column *category* of

Table 1 shows that only the study of Schultz (2017) takes into account two different categories. The other studies either have taken product category as a control variable, focused on only one category or did not even mention the used category. Therefore, this comprehensive quantitative study is conducted for the most popular categories regarding influencing: fashion, beauty and food (Zeren & Gökdağlı, 2020). In the remainder of this literature review, each facet of Table 1 is explained in more detail with respect to their previous findings. This literature study is subdivided into four sections. First of all, the existing literature on the characteristics of influencers is examined. Secondly, specific aspects about posts of an influencer are scrutinized, followed by specific aspects about posts of the brand. General aspects of a post, which apply to both brands and influencers, can also have an effect on brand engagement. This is discussed in the third section. Finally, existing literature is examined regarding the three most employed product categories in influencer marketing: fashion, beauty and food.

The aim of this study is to make the connection between mainly qualitative research on influencers and quantitative studies solely focusing on brand engagement. This is done by conducting a quantitative analysis taking into account Instagram data of 83 brands and 4162 influencers. For each category, a separate analysis is conducted to discover differences.

2.1 Influencer Characteristics

Number of Followers

The number of followers of an influencer is probably the first characteristic that comes to mind when people think of influencers. The more followers an influencer has, the more engagement in terms of absolute number of likes and comments he or she will get (Bakhshi et al., 2014). Instagram influencers with a large number of followers are often considered more likeable, because they are perceived as being more popular (De Veirman et al., 2017) and reliable (Jin & Phua, 2014). These influencers are believed to be the best to collaborate with when it comes to purchase intentions (Jin & Phua, 2014) and even actual purchases (Y. Zhang et al., 2018a). In contrast, the study of De Veirman et al. (2017) concludes that a large number of followers rarely leads to perceived opinion leadership and may even have a negative impact on brand attitude.

The report of Launchmetrics (2019) claims that brands consider micro-influencers to be the most effective, followed by mid-tier influencers. They tend to have a better connection with the audience and possess authenticity. Only luxury brands consider mega-influencers and all-star influencers to be more useful[2].

Research has been done to assess the impact of the number of followers on influencer engagement, brand attitude, purchase intentions and sales. However, hardly any research has been done in the area of brand engagement on Instagram. In order to fill this gap in the literature, the following research question is formulated.

> **RQ2**: *Does the number of followers of an influencer have an impact on brand engagement on Instagram?*

Engagement Ratio

Although the number of followers of an influencer gives a swift overview of the number of people that directly sees the influencer's post, it does not tell how many of these followers effectively like the post and engage with this particular influencer. Therefore, it is important to look at the efficiency of influencers, which can be approximated by the engagement ratio. This ratio is an important KPI in influencer marketing, because 35,2% of the businesses choose influencers based on this ratio (Launchmetrics, 2019). In a very recent study of Oliveira et al. (2020), the engagement ratio of Instagram users is defined as "the number of likes and comments divided by the number of followers multiplied by 100" (p.119). The following example proves that it is better to also focus on the engagement ratio instead of merely the number of followers:

> *"An influencer with 500K followers may charge the brand €3.000 per post. This influencer has an engagement ratio of 1,2% resulting in approximately 6.000 interactions per post at a cost of €0,50 per engagement. However, if an influencer has only 20K followers and charges €750 per post but has a higher engagement ratio of 15%, the brand can expect then 3.000 interactions at an average cost of €0,25 per engagement."*

The importance of this ratio has even increased since the introduction of the new Instagram algorithm in 2018. Before this new algorithm, the posts on the Instagram feed were displayed in chronological order. However, with this new algorithm, this sequence is affected by other factors

[2] Micro-influencers: 10K-100K followers
Mid-tier influencers: 101K-500K followers
Mega-influencers: 501K-2M followers
All-star influencers: > 2M followers

	CHARACTERTISTICS INFLUENCER			POST CHARACTERITICS INFLUENCER								POST CHARACTERISTICS BRAND						CATEGORY
	Number of followers	**Engage-ment**	**Activity**	**Interaction**	**Giveaway**	**Image Content**	**Video**	**Multiple Post**	**Caption Sidedness**	**Discount**	**Ad Disclosure**	**Interaction**	**Giveaway**	**Image Content**	**Video**	**Multiple Post**	**Regram**	
De Veirman et al. (2017)	3a																	Unknown
Jin and Phua (2014)	5b																	Food
Y. Zhang et al. (2018a)	6d																	Unknown
E. L. De Vries (2019)		1a																Travel
Henderickx and De Wolf (2019)		1a																Unknown
Bodrunova et al. (2016)			2b															Politics
Brorsson and Plotnikova (2017)			1a	1a														Unknown
Ledbetter and Redd (2016)			1e															Unknown
Stephen et al. (2017)			2b															Unknown
Braatz (2017)									1a									Unknown
De Veirman and Hudders (2020)									3a		3a							Food
Korotina and Jargalsaikhan (2016)										5a								Fitness
Delrue and Sinigaglia (2017)											5a							Lifestyle
Evans et al. (2017)											3a							Food
Mathisen and Stangeby (2017)											5a							Unknown
Johnson et al. (2019)											5a							Control Variable
Boerman (2020)											2a							Fashion
Dhanesh and Duthler (2019)											5a							Unknown
Van den Abeele (2019)					2a													Unknown
Berger and Schwartz (2011)					2e													Control Variable
Hughes et al. (2019)			2c		2e													Unknown
Hellberg et al. (2015)														4a	4a			Unknown
L. De Vries et al. (2012)												4c	4c		4c			Control Variable
Schultz (2017)												4c	4c		4c			Fashion and Food
Lei et al. (2017)												4c	4c					Integrated Resorts
Suh et al. (2010)				2b								4b						Unknown
Cruz et al. (2017)												4c						Control Variable
Tafesse (2015)												4c						Automative
Cvijikj and Michahelles (2013)															4c			Food
Sabate et al. (2014)															4c			Travel
Kwok and Yu (2013)															4c			Restaurant
Swani and Milne (2017)															4c			Goods vs Service
Bakhshi et al. (2014)	2a		2a			2a								4a				Unknown
Z. Zhang et al. (2018b)			6d			2a								4a				Unknown
Ding et al. (2019)						2a								4a				Unknown
Vignisdóttir (2017)												4a		4a	4a		4a	Beauty
Tresna and Wijaya (2015)																	3a	Mall X
Our Study	**4a**	**4a**	**4a**	**4a**	**4a**	**4a**	**4a**	**4a**	**4a**	**4a**	**4a**	**4a**	**4a**	**4a**	**4a**	**4a**	**4a**	**Fashion, Beauty, Food**

Legend:

a *Instagram*
b *Twitter*
c *Facebook*
d *Weibo*
e *SNS general*

1 Influencer Credibility and Trustworthiness
2 Influencer Engagement (likes, comments on influencer post)
3 Brand Attitude
4 Brand Engagement (like, comments on brand post)
5 Brand Purchase Intention
6 Actual Sales

Table 1: Overview Literature Study

like engagement. Several influencers in interviews with Van den Abeele (2019) and Henderickx and De Wolf (2019) have indicated to be negatively influenced by this algorithm as only a part of their followers now see their posts. To make their posts more visible, influencers have admitted that they use several practices such as posting stories with the announcement of a new post or participating in instapods. An instapod is a group of influencers in which every member is obligated to comment on all other members' posts (Henderickx & De Wolf, 2019). This increases engagement and this way, the post is more likely to get on the feed of their followers. Some influencers even go further and buy followers and likes to keep their engagement at an attractive level for brands (Osman, 2017).

To the best of our knowledge, no research has been done on the impact of the influencer's engagement ratio on brand engagement. However, a study of E. L. De Vries (2019) has examined the consequences of likes-to-followers ratios on the credibility of influencers on Instagram. Their conclusions state that too high as well as too low likes-to-followers ratios have a negative influence on the perceived trustworthiness of influencers. This is because these influencers are perceived as people who buy likes or followers. It is therefore important that brands focus on influencers with a decent engagement ratio. This research investigates whether these conclusions also apply to brand engagement on Instagram.

> **RQ3**: *Does the engagement ratio of an influencer have an impact on brand engagement on Instagram?*

Activity

Besides the number of followers and the engagement ratio, literature has shown that the frequency of posting of the influencer is also an important KPI for brands. Evidence indicates that frequent activity is considered as one of the most valued characteristics of an influencer (Brorsson & Plotnikova, 2017) and is important to gain appeal (Ledbetter & Redd, 2016). Most of the respondents in the study of Brorsson and Plotnikova (2017) support this by stating that rich content is important to perceive an influencer as interesting. Furthermore, Stephen et al. (2017) have found that posts from active Twitter users generate the most engagement. Their posts are seen as the most up-to-date, leading them to be perceived as most attractive. In addition, the number of Facebook posts is positively correlated to the number of likes of the influencers (Hughes et al., 2019).

On the other hand, Y. Zhang et al. (2018a) have concluded that frequently advertising products can be counterproductive on the effectiveness of influencer marketing on Instagram. Likewise, the study of Bakhshi et al. (2014) has come to the conclusion that the level of activity is also negatively correlated with likes and comments on Instagram. Furthermore, on Twitter, active posters are not necessarily the most influential (Bodrunova et al., 2016). Given these conflicting findings and the fact that the impact of an influencer's activity on brand engagement has not been investigated, the following research question is formulated.

> **RQ4**: *Does the activity of an influencer have an impact on brand engagement on Instagram?*

2.2 Post Specifics Influencer

Caption Sidedness

Although brands usually consider the photo or video to be the most important part of advertisement, the caption of the post also plays a major role. While photos and videos are particularly useful in customer attraction and influencing behavioral intentions, the caption is more powerful in transferring information (Decrop, 2007).

It seems logical that brands expect their influencers to be as positive as possible about their product. However, in past literature this has not always been advised as the best option. Including unfavorable information in the message can lead to a higher perceived trustworthiness of the influencer (Braatz, 2017). However, it is claimed that while the credibility of the source increases with the amount of negative information, the behavioral intentions decrease (Eisend, 2006). This trade off leads to a preference for two-sided messages instead of purely negative and positive messages. Similarly, the study by Wee et al. (1995) has concluded that negative messages have a more significant negative impact on behavioral intentions compared to two-sided and positive messages. Additionally, two-sided messages have been found to have a more positive impact than positive messages.

There has been a lot of research about the effects of message sidedness. However, the results are not conclusive. Moreover, the specific impact on brand engagement has not been investigated yet. Therefore, the following research question is constructed.

> **RQ5**: *Does the caption sidedness of the influencer have an impact on brand engagement on Instagram?*

Discount Code

The use of discount codes is often used in the context of influencer marketing on Instagram. Influencers who collaborate with a particular brand sometimes receive a personalized discount code, for example "LAURA20". This represents the code offered by influencer Laura[3] which corresponds to a 20% discount on the purchase for that specific brand. Discount codes offer an extra advantage for brands. As these codes provide a direct link between the influencer and the people who actually use the code, brands can keep track of how many times this code is used. The brand can thus determine whether the collaboration with the influencer is beneficial or not. According to Korotina and Jargalsaikhan (2016), these discount codes are recognized as the most favorable influencer marketing tool in terms of purchase intentions.

This study investigates whether the positive impact of the use of discount codes on purchase intentions on Instagram can also be found in case of brand engagement.

> **RQ6**: *Does the use of a discount code by the influencer have an impact on brand engagement on Instagram?*

Ad Disclosure

The main strength of influencer marketing compared to traditional marketing is the fact that the relationship between the sponsoring brand and the influencer may be unclear. This could give the impression that the sponsored post is the honest opinion of the influencer and not the result of compensation (Woods, 2016). However, this practice of influencer marketing has been widely criticised. As a result, the global regulation implies that commercial and editorial content must be separated (EASA, 2018; FTC, 2017). Fines can range from €80K in Belgium to €5M in Italy. These are only theoretical figures as no such fines have actually been issued in practice (Osborne Clarke, 2017). Some in-depth interviews with Belgian influencers have revealed that many influencers are not even aware of these strict rules. Others have even admitted that they circumvent these rules. For example, they edit their caption after a few weeks by adding something like "#ad" or "#sponsored". They even have revealed that some brands explicitly ask not to mention that the post in question is an advertisement (Van den Abeele, 2019).

Some research has already been done on the impact of the use of disclosure language by influencers on Instagram. Most of the research agrees that the use of disclosure language in the caption has a

[3]Laura Gardin is an influencer with 8500 followers from Belgium

positive effect on advertising awareness (Evans et al., 2017; De Veirman & Hudders, 2020; Mathisen & Stangeby, 2017; Delrue & Sinigaglia, 2017) and the credibility of the influencer (De Veirman & Hudders, 2020; Delrue & Sinigaglia, 2017). However, some of these studies have found that advertising awareness has a negative impact on brand attitude (Evans et al., 2017; De Veirman & Hudders, 2020; Delrue & Sinigaglia, 2017) and purchase intentions (Delrue & Sinigaglia, 2017). Contrarily, other studies have found that people are in fact more motivated to interact with a post (Boerman, 2020) and to buy the product (Dhanesh & Duthler, 2019) when it is recognized as an advertisement. The study of Johnson et al. (2019) claims that there is no significant impact on brand attitude and purchase intentions.

This aspect of influencer marketing is important, confirmed by the vast amount of research in this area. However, surprisingly, no research has been conducted on brand engagement. This fact, together with the inconsistency of the prior results, led us to construct the following research question.

> **RQ7**: *Does explicitly disclosing that the influencer's post concerns an advertisement have an impact on brand engagement on Instagram?*

2.3 Post Specifics Brand

Regram

Regram refers to the act of publishing someone else's post on your own feed. As displayed in Figure 1 , a regram can be recognized by the acknowledgement of the source Instagram account (Highfield & Leaver, 2015) by means of a tool. However, this tool is not provided by Instagram itself but by other applications (Chen, 2019). Therefore, there is always the possibility that brands do not use those applications and just put the print screen of the photo on their feed with a link to the post-maker and a corresponding hashtag like *#Regram* or *#Repost* in the caption.

Figure 1: Example of a regram making use of a tool (Source: Instagram)

A study of Utz et al. (2012) examines the effects of online customer reviews on trust by conducting two experiments. In both experiments, customer reviews have proved to be a strong predictor of trust. Trust is perceived to be a prelude to repeated purchase intentions (Kassim & Abdullah, 2010).

Regrams can basically be considered customer reviews by users visiting the brand page. These customer reviews can be from paid influencers as well as satisfied customers. By sharing these alleged customer reviews, the brand builds trust to its followers (Stouthuysen et al., 2018). Another reason for brands to make use of regrams is to value and maintain a relationship with its customers (Tresna & Wijaya, 2015). The study of Tresna and Wijaya (2015) concludes that regrams from customers are a valid variable to investigate the relationship between a user and the brand on Instagram. This relationship with the brand has a major influence on brand loyalty.

Vignisdóttir (2017) has analyzed the top five beauty brands on Instagram in 2016. The purpose of the study was to investigate whether there is a difference in the level of customer engagement on posts created by the brand or regrammed posts created by users on Instagram brand pages. One of their conclusions is that UGC is positively related to more likes, but the opposite is true for comments.

This study investigates whether this positive impact on engagement of beauty brands can also be found in the categories fashion and food.

RQ8: *Does brands posting regrams of influencers have an impact on brand engagement* *on Instagram?*

2.4 Posts General

Interaction

In the past few years, research has already shown that being interactive can significantly increase engagement. Ha and James (1998) define interactivity as "the extent to which the communicator and the audience respond to each other's communication needs" (p.461). Interaction is an umbrella term for several aspects. Table 2 displays the different aspects used in this study, each with an example and their sources.

Interaction	Example	Sources
Hashtags	This *#sunscreen* of *#Nivea* helps you against the *#sun* *#beyourself* *#worthit*	Schultz (2017) Suh et al. (2010)
Question	This sunscreen of Nivea is amazing! *Have you already tried them?*	L. De Vries et al. (2012) Lei et al. (2017) Schultz (2017) Vignisdóttir (2017)
Call to action	I really love the smell of the sunscreen of Nivea!! If you agree, *like* this! Certainly *use* my code to *get* 10% discount!	Lei et al. (2017) Schultz (2017) L. De Vries et al. (2012)
Second pronouns	*You* really need to try the new sunscreen of Nivea!! *You* will not regret this	Cruz et al. (2017)
Mention	Love the sunscreen of *@nivea_be*!	Schultz (2017) L. De Vries et al. (2012)

Table 2: Interaction aspects

Firstly, hashtags are used because a larger reach can be achieved. Schultz (2017) supports this by showing that the use of hashtags is positively related to brand engagement with respect to likes and comments on Facebook. A study by Suh et al. (2010) analyzed the factors that influence retweets[4] on Twitter. They have concluded that a tweet using hashtags is more likely to get retweeted.

A second way to interact is by asking questions. In the past, research has shown that asking a question in your post is beneficial for the number of comments on Facebook (L. De Vries et al., 2012; Schultz, 2017) and Instagram (Vignisdóttir, 2017). While Schultz (2017) has also found a positive impact on the number of likes, L. De Vries et al. (2012) claim that including question

[4] Retweeting is sharing a tweet on Twitter and can therefore be seen as engagement

marks reduces the number of likes. On the other hand, Lei et al. (2017) have established that asking questions has no significant impact on engagement.

A next interaction element is the use of action words like "get", "go", "like", "use"... Earlier research on Facebook has discovered that interactivity, in the sense of including action words, has a positive influence on brand engagement in terms of likes (Lei et al., 2017; Schultz, 2017) and comments (Schultz, 2017). Contrarily, L. De Vries et al. (2012) have not found a significant relation between a call for action and brand engagement in terms of likes and comments.

The fourth aspect is the inclusion of second pronouns such as "you" or "yours" in the brand message. This creates a sense of personalisation because the customer perceives it as a message directed straight towards him (Vesanen, 2007). Cruz et al. (2017) have researched this interactivity dimension and concluded that it has a positive impact on brand engagement.

Finally, adding a link to a website is also perceived as a type of interactivity, because people are invited to click on it. However, Schultz (2017) has discovered that including a link does not enhance brand engagement, while L. De Vries et al. (2012) have even found a negative effect on the number of comments on Facebook. On Instagram, it is common practice to mention other accounts in your caption ("@username") instead of including links. It will therefore be investigated whether this has a more positive impact than including links. In Table 2, this aspect is referred to as "Mention", as this is the most commonly used term to refer to other Instagram pages.

The study of Tafesse (2015) in the automotive industry has discovered that the probability of receiving likes declines as a higher level of interactivity is added to the brand post. Consequently, they advise the brand to keep their interactivity level within limits. However, this study took external links, hashtags, page/person tags, questions, call to actions and contests together as one variable interactivity.

As can be seen, research has recognized the importance of being interactive on Facebook in the case of brands. In this study, we want to investigate whether these conclusions can also be applied to Instagram. Moreover, while Brorsson and Plotnikova (2017) have shown that the openness and integrity of influencers is an important characteristic, quantitative research is needed to determine whether influencer's interactivity also has a positive effect on brand engagement. Therefore, the following two research questions are stated.

RQ9a: *Does the interaction level of the influencer's post have an impact on brand en-*

gagement on Instagram?

RQ9b: *Does the interaction level of the brand's post have an impact on brand engagement on Instagram?*

Giveaway

A giveaway is a campaign incentive designed to elicit responses and thus engagement from consumers (Hughes et al., 2019). In fact, a giveaway contest can also be seen as a component of interactivity. However, this is investigated separately, as this is a very common practice in influencer marketing on Instagram.

A giveaway on Instagram is a type of contest, initiated by the influencer or the brand, in which participants can win a product or service of that specific brand. In exchange for liking, sharing, commenting, tagging others, re-posting a photo and/or following the poster and maybe some other influencers, the participant gets a chance to get picked as a winner. Depending on these conditions, giveaways can 1) increase the number of followers, 2) increase brand awareness and 3) increase engagement because the post will get extra exposure. Followers become a network of advertorial channels by duplicating, amplifying and multiplying the influencer's content to their own circle of followers and personal friends. Therefore, giveaways are considered as an extremely important advertising tool (West, 2019).

Sometimes giveaways are even organized by the influencer without collaboration with a brand. This is just to increase their engagement ratio and get the advantages of the new Instagram algorithm (Van den Abeele, 2019). On the other hand, the interviews from the study of Abidin (2016) have shown that many Instagram users are afraid to be considered "spammers" by their friends and followers. They find it embarrassing to actively participate in Instagram contests. They, therefore, participate when they only have to like or comment. Some interviewees have admitted that they have separate "contest accounts", where they can share and post influencer's content without being noticed by their friends and followers.

To our knowledge, no quantitative assessment of the effects of giveaways has yet been made on brand engagement on Instagram. However, a quantitative travel study of Lei et al. (2017) on Facebook, supports that doing giveaways, initiated by the brand, has a positive impact on brand engagement. According to Schultz (2017), the number of comments significantly increases when performing a Facebook contest in the food sector, but not in the fashion sector. The number of likes decreases in both sectors. Surprisingly, L. De Vries et al. (2012) have found that doing contests positively influences the number of likes of the brand whereas the number of comments remains unchanged. The reason for this may be that this study has been carried out using the Facebook platform in 2012 when tagging other people was not a prerequisite yet. Furthermore, some studies have been done on giveaways as an influencer strategy on SNSs in general. They have concluded that campaign giveaways are classified as one of the key drivers of engagement on the SNS of the influencer (Hughes et al., 2019) and that people tend to generate 20% more WOM when they get a product for free (Berger & Schwartz, 2011).

It has been proved in the past that giveaways have a positive impact on influencer engagement when initiated by the influencer and on brand engagement when initiated by the brand. However, the following research questions are examined, because no quantitative research has been done on Instagram. Furthermore, it is questionable whether a giveaway initiated by influencers also has a positive impact on brand engagement.

> **RQ10a**: *Does a giveaway contest initiated by the influencer have an impact on brand engagement on Instagram?*
> **RQ10b**: *Does a giveaway contest initiated by the brand have an impact on brand engagement on Instagram?*

Video

Instagram advertisements mainly consist of photos or videos. Different media types, such as photos or videos can be seen as the vividness of a post (L. De Vries et al., 2012). Vividness can be defined as "the representational richness of a mediated environment as defined by its formal features; that is, the way in which an environment presents information to the senses" (Steuer, 1992, p.11). Video and audio, both identified as rich media tools, enhance the vividness (Coyle & Thorson, 2001), whereas images tend to evoke a lower number of sensory dimensions (Steuer, 1992).

Research shows that an increase in vividness is considered as feeling a more positive attitude towards the brand's website (Coyle & Thorson, 2001). Moreover, L. De Vries et al. (2012); Cvijikj and Michahelles (2013) and Sabate et al. (2014) agree on the fact that rich vividness, i.e. videos, has a positive relationship with the number of likes of the brand post on Facebook. While L. De Vries et al. (2012) claim that there is no significant relationship between vivid post features and the number of comments of a brand post, Cvijikj and Michahelles (2013) and Sabate et al. (2014) have found a negative relation. In addition, the studies of Kwok and Yu (2013) and Schultz (2017) have found this negative relation with respect to both likes and comments on Facebook. Moreover, an Instagram study on vividness has revealed that high vividness posts were negatively related to the number of likes, but positively related to the number of comments (Vignisdóttir, 2017).

It is remarkable that many studies contradict each other. One reason can be that the effectiveness of vividness depends on the type of product represented. The study of Swani and Milne (2017) confirms this by stating that vividness generates more likes and comments for goods rather than for services. In the study of Hellberg et al. (2015) most of the interviewees claim that the usefulness of using videos depends on their purpose, which can be considered as a second reason. Some interviewees do not appreciate videos at all, regardless of their purpose, because it interrupts the flow of scrolling.

Again, the impact of influencers using videos has not yet been investigated on brand engagement. This gap together with the contradictory conclusions of prior research result in the following research questions.

> **RQ11a**: *Does posting videos in case of the influencer have an impact on brand engagement on Instagram?*
> **RQ11b**: *Does posting videos in case of the brand have an impact on brand engagement on Instagram?*

Multiple Post

It is already mentioned that Instagram is a fast-growing platform. In addition, it often releases new features and possibilities. From 2017, Instagram users were able to combine multiple photos and videos into a single post. To this day, no research has been conducted into the impact of using so-called multiple posts on brand engagement. Nevertheless, it can be assumed that the results regarding personas from the study of Salminen et al. (2018) can be extended to the use of multiple posts in advertisements. The findings show that, on the one hand, using more photos can substantially increase the knowledge acquired by users of a profile character. On the other hand, using different photos can generate ambiguity which reduces the informativeness.

For both brands and influencers, it is examined whether the use of multiple messages has the same impact on brand engagement regarding Instagram.

RQ12a: *Does using multiple photos/videos in one post in case of the influencer have an impact on brand engagement on Instagram?*

RQ12b: *Does using multiple photos/videos in one post in case of the brand have an impact on brand engagement on Instagram?*

Image Content

As Instagram is a social platform for sharing photos and videos, it is crucial to investigate what drives engagement in the photo. The most common types of photos that are placed on SNSs are photos of human faces (Bakhshi et al., 2014).

In the same study of Bakhshi et al. about Instagram, they conclude that the presence of a face in a photo significantly increases engagement. That is, including a face increases the likelihood of receiving likes with 38% and comments with 32%. They further conclude that more than one face in a photo does not significantly affect engagement. Further research of Z. Zhang et al. (2018b) on predicting Instagram popularity reveals that a human face gets a relatively high popularity score. In addition, after clustering the images, they have found that photos in classes such as "group photo", "small group photo", and "selfie" receive more likes. Furthermore, the findings of Ding et al. (2019) show that pretty faces, lovely kids and cute animals lead to high predictions of image popularity.

However, a study of Vignisdóttir (2017) in the beauty sector comes to the opposite conclusion and states that Instagram posts containing a face are negatively related with customer engagement. The structured in-depth interviews conducted by Hellberg et al. (2015) have uncovered conflicting insights. Most of the interviewees admit that they want to see people wearing the clothes in order to get inspired. However, some also emphasize the fact that they want to see clearly the products in which they are interested.

What is perceived beneficial for the engagement in the photo is vital for brands in imposing rules on the influencer. Therefore, it is investigated whether or not posting a photo with a person on it is important for the engagement.

RQ13a: *Does using a post including a person on the influencer's post have an impact on brand engagement on Instagram?*

RQ13b: *Does using a post including a person on the brand's post have an impact on brand engagement on Instagram?*

2.5 Brand Categories

Given the many advantages of influencer marketing, it is not surprising that not only fashion brands use this advertising tool. Although fashion is considered as the most important part of influencer marketing, literally all types of brands use influencers. Travel, electronics, food, beauty, lifestyle etc. are only some of the many examples (Statista, 2016). According to Zeren and Gökdağlı (2020), fashion is the most commonly used category for endorsement, followed by beauty and food. Therefore, these three brand categories are taken into account.

Research exhibits that social media content (Cvijikj & Michahelles, 2013) and brand engagement (Erkan, 2015) vary across different categories and sectors. In the beauty category, for example, it can be assumed that many influencers use video tutorials on how to apply make up. This may explain the conflicting findings about the use of videos. In contrast, fashion influencers use more photos than videos (Buryan, 2016). Besides, Kim et al. (2015) conclude that photos receive more likes and comments in this category. The study of Vignisdóttir (2017) about regrams considered only beauty brands for which it can be assumed that regrams are prominently useful for beauty brands. The study of Shen and Bissell (2013) concludes that there are no differences in brand engagement by comparing discount codes and giveaways in the beauty sector. The question then arises whether this also applies to other brand categories. Another difference between categories is how influencers should be selected. The report of Launchmetrics (2019) shows that beauty brands (50%) are more likely to choose influencers based on their engagement ratio than fashion brands (29,7%). The study by Schultz (2017) compares the food and fashion sector with regard to the impact of the use of contests, the use of videos and interaction aspects of brands on Facebook engagement. They have found no marked differences, but they did not take influencer characteristics into account.

A number of assumptions can be made based on previous research into influencer marketing features across different categories. It is necessary to carry out a comprehensive study, as these are mainly hypotheses considering only a few features. Therefore, in this research it is examined whether the conclusions on the previously mentioned research questions differ across the categories fashion, beauty and food.

> **RQ14**: *Is there a difference in importance of the features and their relations with brand engagement across the categories fashion, beauty and food on Instagram?*

2.6 Research Framework

In light of the aforementioned research questions concerning brand engagement in terms of influencer characteristics, post influencer characteristics and post brand characteristics in the brand categories, the conceptual framework in Figure 2 is proposed. The features on the left side of the figure are possible drivers for brand engagement in terms of likes and comments based on the described literature on Instagram and other SNSs. The goal is to find what matters most for brand engagement across the categories fashion, beauty and food on Instagram. Moreover, the brand engagement is controlled for several brand characteristics and time variables which will be explained in more detail in Section 3 Data.

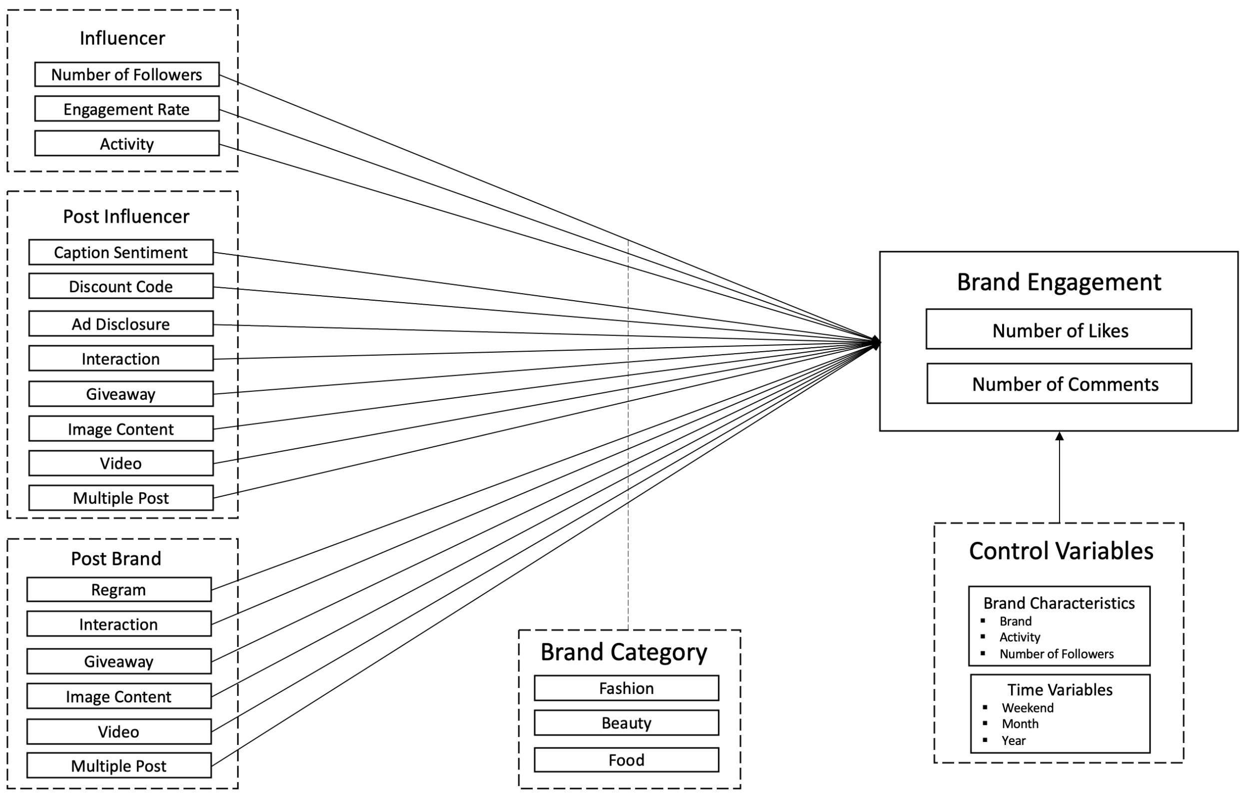


Figure 2: Conceptual Framework

3 Data

For this research, data from 83 brands and 4758 influencers are scraped from the platform Instagram. For each brand category, the list of considered brands with their corresponding number of followers can be consulted in Appendix A. Table 3 displays the number of brands and the number of observations per category. Fashion brands post more frequently and hence, they have more observations.

	Number of Brands	**Observations**
Fashion	17	12030
Beauty	26	4683
Food	46	2243

Table 3: Number of brands and observations per category

3.1 Time Window

The data of the brands and influencers must match in terms of time. Posts from both brands and influencers are examined between 01/01/2018 and 31/08/2019. This period of 20 months already gives us an enormous load of data because Instagram is a very active platform. Moreover, every year new features are added or removed in a way that a time span of 20 months makes the data more comparable. The end date is 31/08/2019 because that is when the first brands and influencers started entering the database.

3.2 Influencers

For each brand, it is necessary to look for the corresponding influencers. Therefore, two types of tools are used. First, influencers are sought using *influence.co*, a site where you can find which influencers worked or featured for which brand. However, since influencers need to have an account on *influence.co* in order to get selected by us, this site is not complete at all. This is where the second tool comes in: manually scrolling through the tagged photos of each brand between 01/01/2018 and 31/08/2019. The influencers are selected when they meet some implicit conditions. First, the photo or video requires a considerable amount of likes. Secondly, it is usually not difficult to distinguish influencers from ordinary users. For example, by the specific editing of the pictures, or the way of posing. Thirdly, when a post is selected fulfilling the two conditions mentioned above, the page of the influencer is examined. If the user belongs to one of the categories in Table 4, which means he or she has a significant number of followers, the user can be considered an influencer. In addition, it is often mentioned in the biography to contact the influencer for collaborations. Finally, it is investigated whether the feed itself has some similarities with a typical influencer's feed.

	Number of Followers	**Engagement ratio benchmark**	
		Underbound	*Upperbound*
Nano-influencers	1K - 10K	0,8%	76%
Micro-influencers	10K-100K	0,4%	43%
Medium influencers	100K-1M	0,33%	32%
Mega-influencers	>1M	0,066%	19%

Table 4: Bounds for detecting fake influencers for each category

This way of looking for influencers is labour-intensive but feasible for the food and beauty categories. It gives us a complete list of influencers collaborating with the considered brands. However, it is unrealistic to carry out these practices for the fashion brands. Every day, thousands of posts are posted with hashtags of different brands by influencers and non-influencers who want to show off their outfit, for example with the hashtag #OOTD[5]. For this reason, we look at the metadata that contains the posts in which the brand is tagged. We then include the influencers in the analysis if their post has more than 400 likes. Unfortunately, not all the influencers are caught using this method because when comparing to Instagram, these tagged users are incomplete. In any case, there is still a huge amount of influencers added to the analysis.

However, any Instagram user can pretend to be an influencer, these are so-called "fake influencers". For example, some influencers admit that they initially pretended to work with a brand in order to get real contracts in the future (Van den Abeele, 2019). In this respect, *influence.co* is a great opportunity for these people because any person can create an account and thus appears to be an influencer, even if they only have 100 followers. Moreover, *influence.co* is completely built by Instagram users themselves. Hence, nobody checks whether the collaborations are false or not. The same study of Van den Abeele (2019) also reveals that Instagram users buy likes or followers in order to be perceived as more attractive. This topic of influencer fraud has gained a lot of interest in research, especially since the launch of the Instagram algorithm that has forced users to artificially increase engagement in order to get noticed (Schröder, 2019; Williamson, 2019; Anand et al., 2019).

It is clear to see that Instagram users with fake followers or likes are not beneficial for brands.

[5] "Outfit Of The Day"

To solve this issue, some rules of thumb regarding the number of likes and comments on posts, the number of followers and the engagement ratio are used in order to eliminate as many fake influencers as possible.

The first two rules of thumb are the most evident. It is quite straightforward that an influencer must have at least 1000 followers, since none of the categories displayed in Table 4 include accounts with less than 1000 followers. Therefore, those accounts are considered as non-influencers and are omitted from the research. Subsequently, the median number of likes on photos of the influencers is checked. If this amount is lower than 150, the account is eliminated.

However, the largest group of non-influencers is found using the engagement ratio. The engagement ratio is calculated as the ratio of the median number of likes and comments to the number of followers. If this ratio is considered extremely low, this can be a signal that the account has bought its followers. On the other hand, if this ratio is extremely high, it is likely that the account has bought likes. Research has shown that the higher the number of followers, the lower the expected engagement ratio is (Launchmetrics, 2019). One reason for this difference could be that mega-influencers have a lot of followers just because they are famous and not because the audience is interested in what they share. Moreover, nano-influencers tend to have more time to engage with their limited number of followers. That is why different benchmarks have been set for the different categories (Influencer Marketing Hub, 2019). The different benchmarks used to detect fake influencers with the associated category are also shown in Table 4. The under bound of the benchmarks is one third of the benchmarks a decent influencer in that category should have according to the report of Influencer Marketing Hub (2020b). It is obvious that when the engagement ratio is lower than one third of the pre-defined benchmarks, the influencer is perceived as a fake influencer. Similarly, an engagement ratio higher than the upper bound benchmark is also considered infeasible and therefore omitted. Here it is decided to multiply the pre-defined upper bound benchmark with 10, to make sure no extremely loved influencers are deleted. After deleting the fake influencers, the database is reduced from 4758 to 4162 influencers.

3.3 Basetable Set-up

For each brand and each influencer a lot of metadata is scraped of Instagram. The initial variables that are kept are the username, postid, date, caption, number of photos in the post, whether the post is a video or not, number of likes, number of comments, tagged accounts and image description.

The steps in constructing the basetables are shown schematically in Figure 3.

If an influencer sponsors one of the 83 brands, the brand is mentioned somewhere in the post. Therefore, the tagged accounts and the caption are investigated. An additional variable *Brand* is created with a list of which of the 83 brands are sponsored in the particular influencer's post. An extra row per brand is then created for each brand mentioned in this list. This way, the post can later be merged with a brand basetable (BB) based on the profile name. The influencer's posts with an empty list are omitted in a way that only sponsored posts remain. Thereafter, the individual basetables of the fake influencers are eliminated, as explained in the previous subsection. Next, all the sponsored posts of the influencers are combined into one complete influencer basetable (IB). On the other hand, all the brand posts are kept. Moreover, the brands have their own basetable, i.e. the BBs and are not yet joined together. The next step is to combine each BB with the

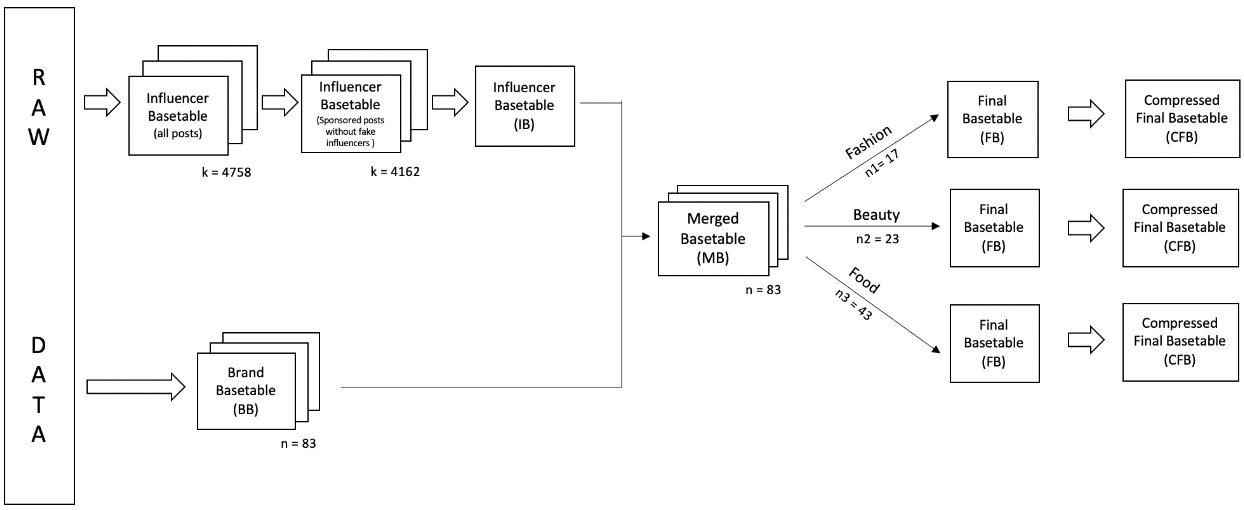


Figure 3: Basetable set up

corresponding influencers of the IB. It is obvious that the brand's post must be posted after the influencer's post in order to assess the impact of the influencer. However, the most crucial decision is to determine the maximum number of days over which an influencer's post could have an impact on the brand's post. Since most Instagram users log in daily (Pew Research Center, 2019), it is assumed that a post circulates not longer than three days on a news feed. Therefore, it is decided to merge the influencer's post with the brand's post when (1) the first is posted before the latter and (2) the brand's post is posted maximum three days after the influencer's post. When more than one post of a particular influencer can be linked to the same brand post, the closest influencer post is chosen. This process is done for each brand and results in 83 merged basetables (MBs).

These MBs are stacked per category in a final basetable (FB). Hence, there are three final basetables. Since brands often collaborate with multiple influencers over a time frame of three days, more influencers are merged with a single post of a brand. This way, multiple observations of influencers belong to one brand post and thus to one dependent variable. Therefore, the influencer posts in the FB are compressed into a kind of summary in a way that each row corresponds to one dependent variable. This results in a compressed final basetable (CFB) for each category. This practice of going from the FB to the CFB is explained in more detail in the next subsection.

3.4 Independent Variables

Once the FBs have been made, the various important features must be obtained. While some features are already present in the basetable, most need to be created. This feature engineering step involves 26 variables which have been established by reviewing the literature. This section discusses how these variables are extracted. Moreover, for the (post) influencer characteristics, it is explained how we get from the FB to the CFB. An overview of the final independent variables, including control variables, can be found in Table 5. This table displays the variables and how they will be used in further analysis.

Independent Variables CFB	**Description**
Influencer Characteristics	
Number of Followers	
Category 1	Numerical variable : Number of influencers in category 1
Category 2	Numerical variable : Number of influencers in category 2
Category 3	Numerical variable : Number of influencers in category 3
Engagement Ratio	Numerical variable : Number of engaged influencers
Activity	Numerical variable : Number of active influencers
Post Influencer Characteristics	
Caption Sideness	Numerical variable : Number of influencer posts that contains a two-sided caption
Discount Code	Numerical variable : Number of influencer posts that contains a discount code
Giveaway	Numerical variable : Number of influencer posts that contains a giveaway
Ad Disclosure	Numerical variable : Number of influencer posts that contains an ad disclosure.
Interaction	
Hashtags	Numerical variable : Number of influencer posts that contains one or more hashtags
Question	Numerical variable : Number of influencer posts that contains one or more questions
Action	Numerical variable : Number of influencer posts that contains one or more action words
Second pronouns	Numerical variable : Number of influencer posts that contains one or more second pronouns
Mentions	Numerical variable : Number of influencer posts that contains a mention of the brand
Image Content (Person)	Numerical variable : Number of influencer posts that contains one or more people
Video	Numerical variable : Number of influencer posts concerning a video
Multiple Post	Numerical variable : Number of influencer posts that is a multiple post
Post Brand Characteristics	
Regram	Dummy = 1 if the post of the brand is a regram, 0 otherwise
Interaction	
Hashtags	Dummy = 1 if the post of the brand contains one or more hashtags, 0 otherwise
Question	Dummy = 1 if the post of the brand contains one or more questions, 0 otherwise
Action	Dummy = 1 if the post of the brand contains one or more action words, 0 otherwise
Second pronouns	Dummy = 1 if the post of the brand contains one or more second pronouns, 0 otherwise
Giveaway	Dummy = 1 if the post of the brand contains a giveaway, 0 otherwise
Image Content (Person)	Dummy = 1 if the post of the brand contains one or more people, 0 otherwise
Video	Dummy = 1 if the post of the brand concerns a video, 0 otherwise
Multiple Post	Dummy = 1 if the post of the brand is a multiple post, 0 otherwise
Control Variables	
Profile Name Brand ID	Categorical variable : Unique ID of the brand
Activity	Numerical variable : Median time between consecutive posts of the brand in days
Number of Followers	Numerical variable : Number of followers of the brand
Weekend	Dummy = 1 if the post of the brand is posted in the weekend, 0 otherwise
Month	Categorical variable : The month in which the post of the brand is published
Year	Numerical variable : The year in which the post of the brand is published

Table 5: Description of independent and control variables

Influencer Characteristics

All five influencer variables need to be created. Table 6 gives an overview of the descriptive statistics of these variables. For each brand category, the percentage of influencers that occurs in each category is displayed. The median engagement ratio and the median activity of influencers of the influencers are shown for each brand category and influencer category. The numbers in bold show remarkable differences across the categories. In the remainder of this subsection, the extraction of these variables is explained.

	Fashion	**Beauty**	**Food**
Number of Followers			
Category 1	46,41%	40,57%	**30,46%**
Category 2	34,35%	36,46%	**43,21%**
Category 3	19,24%	22,96%	**26,32%**
Median Engagement ratio			
Category 1	0,0238	0,0213	**0,0180**
Category 2	0,0108	0,0115	**0,0097**
Category 3	0,0085	0,0091	**0,0075**
Median Activity	1d 5h	1d 3h	1d 1h

Table 6: Descriptive analysis - Influencer Characteristics

Number of Followers: The most obvious parameter of an influencer is his or her number of followers. There are many different ways to categorize the influencers according to their number of followers. In order to detect fake influencers, the division of Table 4 on page 23 is used. However, the number of nano- and micro-influencers is much larger than the number of mega-influencers. This can be deduced from the histogram in Appendix B. Therefore, in this study the division is done as in Table 7. For each category, the number of influencers can be consulted. Usually more than one influencer

	Number of Followers	**Number of Influencers**
Category 1	1K-20K	1688
Category 2	20K-200K	1587
Category 3	>200K	887

Table 7: Three categories of influencers based on the number of followers

is involved with one particular brand post in the FB. To summarise this, for each category the sum of the number of influencers in that category is collected. Three variables are thus created in the CFB. For example, when the variable *Category 3* has the value 5, it means that five influencers in category 3 are linked to that particular brand post. Table 6 shows for each brand category, what percentage of the influencers belong to Category 1, Category 2 and Category 3 respectively. Food brands focus more on influencers with a larger number of followers compared to fashion and beauty brands.

Engagement: The engagement ratio is considered to be the most powerful ratio to determine the efficiency of an influencer. This formula puts likes, comments and followers in a more reliable perspective. There is no standard formula to calculate this ratio and therefore there are several ways to measure the engagement ratio in order to better match the objectives of social media. In this study, the engagement ratio is calculated using the following formula:

$$Engagement\ ratio = \frac{(0,4\ likes + 0,6\ comments)}{followers}$$

Both parameters "likes" and "comments" are calculated by taking the median number of likes and number of comments of all posts of the influencer dating from 01/01/2018 to 31/08/2019. The combination of these two is then calculated by taking 40% of the likes and 60% of the comments. The reason for this uneven division is because a like is easier, and therefore faster, to give than a comment (Yew et al., 2018). This number is then divided by the number of followers of the influencer. In this way, an engagement ratio between 0 to 1 is obtained for each influencer.

Since every influencer has one engagement ratio, these multiple engagement ratios must be reduced to one measure. Therefore, it is first decided to make a dummy of the engagement ratio. The engagement ratio of every influencer is divided by the median engagement ratio of the category to which he or she belongs, because engagement ratios are expected to decrease with an increasing number of followers. If this result is higher than 1, the influencer is considered engaged, 0 otherwise. Hence, for each brand post, the sum of the number of influencers with a high engagement ratio is obtained. The median engagement ratios are expressed in Table 6 for each category. We notice that food influencers have a lower engagement ratio compared to the other categories.

Activity: The variable activity is created over all posts of the influencer in the relevant time frame. First, the difference in time between two consecutive posts is taken. Next, the median of these differences is calculated. Hence, this number indicates how frequently an influencer post. After creating the median activity per influencer, a dummy is created where influencers with a higher activity rate than the median are considered as active and inactive otherwise. Again, the activity variable has to be compressed into one observation. As with the previous variables, the sum of active influencers is used. Table 6 shows the median activity level for each brand category.

Post Influencer Characteristics

A total of 12 post influencer variables need to be created and compressed. Table 8 provides an overview of related descriptive statistics. For every variable, the percentage of occurrence is given per brand category. The percentages in bold indicate notable differences across the categories. The explanation of the extraction of these variables is explained below the table.

	Fashion	**Beauty**	**Food**
Caption Sidedness	49,88%	**29,46%**	42,03%
Discount Code	**2,55%**	5,71%	5,73%
Ad Disclosure	13,17%	14,23%	**20,84%**
Giveaway	9,78%	**21,61%**	13,28%
Interaction			
Hashtags	76,25%	76,76%	**84,00%**
Question	20,90%	24,10%	24,62%
Action	20,64%	**43,25%**	29,51%
Second Pronoun	**30,35%**	48,09%	45,10%
Mentions	**29,86%**	57,75%	48,41%
Person	97,48%	95,21%	90,26%
Video	3,07%	**10,22%**	4,13%
Multiple Post	14,41%	**22,04%**	17,65%

Table 8: Descriptive analysis - Post Influencer Characteristics

Caption Sidedness: To determine whether an influencer's caption is two-sided or not, non-English languages are first translated into English via the *googletrans* package. This package can detect 109 languages, so it is assumed that no languages are left out. Afterwards, the Python package *TextBlob* is applied to the caption. This package gives a polarity score for a sentence between -1 and 1. It is assumed that an influencer will never be negative about a product he or she sponsors. Hence, the caption is considered to be two-sided when it has a polarity score lower or equal to zero. For the brand post, the variable *Caption Sidedness* refers to the number of two-sided captions. Table 8 displays the percentages of influencers using sided messages per brand category. It is remarkable that two-sided messages are more common in fashion and food brands.

Discount Code: To find out whether an influencer post concerns a discount code, the caption is searched for discount related words. Since influencers are considered over the whole world, the translated captions are used again. A search for words like "discount", "promo", "code", "reduction" and "offer" is done. This is decided by looking up influencer posts including discount codes. If one of these words is found, we assume a discount code is offered by the influencer and the dummy variable is set to 1. Again, the sum is taken to get the number of influencers using a discount code. Table 8 gives an idea of the percentage of the influencer posts including a discount code. We can notice that discount codes are used less in the fashion industry.

Ad disclosure: In the same way as the discount code, the translated caption is searched for the words like "#ad", "sponsor", "publication", "cooperation","collab"... If the translated caption of the influencer contains one of these terms, the dummy *Ad Disclosure* is set to 1. The CFB contains the number of influencer posts that contains an ad disclosure. Table 8 shows for each brand category the occurrences of influencer posts containing ad disclosure. We can deduce from the table that ad disclosure occurs more in the food category.

Giveaway: The same procedure is followed for giveaways. A dummy is created by checking the translated caption for the occurrence of "giveaway", "contest", "gift", "win","competition"... After this, the dummy is summed up for each brand post in the CFB. Table 8 shows what percentage of the influencer posts concern a giveaway for each brand category. We see that more giveaways are organized in the beauty category compared to the other categories.

Interaction: Research has made it clear that interaction is an overarching term. Therefore, different variables are created with respect to interaction. These different variables are:

- Hashtags: a dummy variable whether the caption includes one or more hashtags ("#") or not
- Question: a dummy variable whether the caption includes one or more questions ("?") or not
- Action: a dummy variable whether the translated caption includes an action word like "get", "use", "go", "start", "like", "comment", "share"... or not
- Second pronoun: a dummy variable whether the translated caption includes "you" or not
- Mentions: a dummy variable whether the caption includes a mention of the brand ("@brandname") or not

All five variables are represented in the FB. When making the CFB, these variables are also aggregated per post brand. In Table 8, the percentage of occurrence in influencer posts is given for each of these features, for each brand category. It is notable that in the food category, hashtags are more used, while in the beauty category, actionable verbs are more common in the beauty category. Furthermore, in the fashion category, second pronoun words and mentions are more used.

Image Content: The Instagram metadata itself contains an accessibility caption with a list of objects that appears in the photo for each post. Examples of these objects are things like "sky", "outdoor", "one people or more", "sea"... Since the previous literature emphasized the importance of people in photographs, it is checked for each post whether or not "person" or "people" appear in that list. This is the only variable where missing values occur. No accessibility caption is available for videos. It is assumed that these videos contain a person. For the influencer's photos where no objects have been found, it is also hypothesized that a person occurs in the photo since this is most common. Hence, for each post of the brand, the number of influencer posts containing a person are summarized in the CFB. Table 8 shows how many influencer posts contain a person in the photo.

Video: The dummy variable *Video* is already available in the metadata and therefore did not need to be created. However, in order to reduce the multiple influencer posts to one observation, the sum is taken here as well. Table 8 shows the percentage of influencer posts containing a video. We see that beauty brands makes more use of videos.

Multiple Post: Similarly, the dummy variable *Multiple post* did not need to be created, only the sum has to be taken. Table 8 shows the percentage of influencer posts concerning a multiple post. Multiple posts are more used in the beauty category compared to the fashion and food category.

Post Brand Characteristics

The post brand variables are created in the same way as the post influencer variables. The explanation of these variables can be found in Table 5 on page 27. One difference is that dummy variables do not need to be summarized. For the missing values of image content, it is assumed that there is no person in the image because we have noticed that brands mainly post their products. For the sake of simplicity, only the brand specific variable *Regram* is explained. Table 8 displays the percentage of occurrence of each brand variable per category. Regrams and giveaways are more applied in the food category. In the beauty category, the interaction level is most used. It is noteworthy that a brand posting a photo with people in it are more common in the fashion industry while videos are more common in the beauty industry.

	Fashion	**Beauty**	**Food**
Regram	12,85%	11,98%	**4,82%**
Giveaway	7,04%	8,24%	**15,43%**
Interaction			
Hashtags	80,00%	**69,81%**	80,34%
Question	**13,33%**	26,33%	25,86%
Action	15,35%	**21,27%**	15,74%
Second Pronoun	31,88%	**51,29%**	40,88%
Person	**77,86%**	30,69%	48,15%
Video	8,79%	**17,00%**	4,13%
Multiple Post	8,32%	5,79%	**17,65%**

Table 9: Descriptive analysis - Post Brand Characteristics

Regram: There are two ways to verify whether or not the post of a brand concerns a regram of an associated influencer. Firstly, when “regram” or “repost” appears in the caption, the post is considered to be a regram. Secondly, this is also true when one of our influencers is mentioned or tagged in the post. This variable occurs as a dummy: 1 when the brand post is a regram, 0 otherwise.

3.5 Control Variables

It is necessary to add variables to our research that may influence brand engagement, although we are in fact not interested in the behaviour of those variables. By adding these so-called control variables, this research tries to isolate the causal effect between the brand engagement and other variables. In this study, control variables related to time and brand characteristics are investigated, because these variables already appeared to have a relationship with engagement in previous research.

The timing of a post is often used as a control variable. Cvijikj and Michahelles (2013) have examined the relationship between timing of the post and engagement. They conclude that posts on workdays increase the number of comments. Sabate et al. (2014) have discovered that posts published during business hours are more likely to get comments. Therefore, the dummy variable *Weekend* is created. Additionally, the categorical variables *Month* and *Year* are added to the model as indicated in Table 5.

Since several brands are combined in one basetable, it is important to add some brand characteristics to eliminate these effects. In our study, three brand characteristics variables are included as control variables. The first one is the brand itself. Secondly, the number of followers is included in the model because prior research has concluded that the number of followers has an impact on brand engagement (Sabate et al., 2014; Bakhshi et al., 2014). The last control variable is the activity of the brand page. Several studies have indicated the effect of volume of posts on engagement (Stephen et al., 2017; Hughes et al., 2019; Meire et al., 2019).

3.6 Dependent Variable

The goal of this research is to investigate which of the aforementioned influencer characteristics, post influencer characteristics over a period of three days and post brand characteristics have an impact on the brand engagement on Instagram. Brand engagement can be seen as both the number of likes and the number of comments on posts of the brand in question. That is why two seperate models are made: one for the likes and one for the comments.

The basetable of a category comprises several brands, each having a different number of followers and engagement. It would, thus, be inappropriate to simply take the absolute number of likes and comments as the dependent variable. One solution could be to normalize this by dividing the number of likes and comments by the number of followers. However, brands with a high number of followers are not expected to have the same engagement ratios as brands with fewer followers. The approach in this study is to divide the likes of each post by the median likes of the particular brand. Hence, a post with a measure lower than 1 indicates a post that is less popular than normal whereas a post with a measure higher than 1 is better than normal. The same approach is applied to the number of comments. This way, it is possible to analyse which factors have an impact on the success of a post for the three different categories.

4 Methodology

The previous section has described the extraction of the features and the creation of the CFB. The aim of this methodology section is to explain how the different research questions of this study are answered. Firstly, the different models used to answer **RQ1** are explained. Secondly, the classification algorithm, i.e. Random Forest, that is used for these models is clarified. This concerns the approach of successfully classifying brand posts into engaged (1) and unengaged (0) brand posts. Afterwards, the variable importance plots and partial dependence plots are discussed in detail because these are essential in order to answer **RQ1-RQ14**. Finally, the methods for evaluating the models are specified.

4.1 Different Models

To answer **RQ1** on page 5, different models are taken into account as displayed in Table 10. Model 1 relates only to the post characteristics of the brand. Model 2 also adds post influencer characteristics while Model 3 adds influencer characteristics. Finally, Model 4 is the overarching model that takes all three aspects into consideration. By comparing the performance of these different models, we can check whether or not the (post) influencer characteristics have an additional impact on brand engagement. The performance measure is explained in detail in Subsection 4.3. The difference in performance is statistically tested with the Delong (1988) test. This is an empirical (non-parametric) method that compares the Receiver Operating Characteristic (ROC) curves. The R package *pROC* is used to employ the Delong test.

	Influencer Characteristics	**Post Influencer Characteristics**	**Post Brand Characteristics**
Model 1			✓
Model 2		✓	✓
Model 3	✓		✓
Model 4	✓	✓	✓

Table 10: Different Models

This approach is implemented for the three categories and for both likes and comments. Hence, 24 models are assessed in total. For each category and each engagement measure, it is checked which model performs best. In this way, **RQ1** can be partially answered because the variable importance plots, explained in the next subsection, give us additional information to complete the research question. The other research questions (**RQ2 - RQ14**) are investigated using the overarching model, Model 4.

4.2 Classification Algorithm

For this research, the correct response variable is available and thus, supervised machine learning techniques can be used. The technique used in this study is based on probabilistic classifiers. This means that the resulting predictions can be interpreted as the probability of belonging to a certain class (Burez & Van den Poel, 2009). In our study, this classifier gives us the probability of having an engaged brand post. By choosing an appropriate threshold, these probabilities can be turned into binary outcomes, i.e. engaged (1) if it is higher than the threshold and unengaged (0) otherwise. This study uses Random Forest as a classification algorithm. In the remainder of this section, it is clarified why this algorithm is used, together with the purpose of the corresponding variable importances and the partial dependence plots.

4.2.1 Random Forest

There has been a lot of interest in ensemble techniques, i.e. methods that generate many classifiers and aggregate their results (Liaw et al., 2002). In other words, ensembles generate predictions by averaging the predictions made by the single classifiers (Dietterich, 2000). While single classifiers suffer from a high variance, ensemble algorithms follow the principle of averaging a set of classifiers which reduces variances. Regarding tree-based methods, the single classifiers are decision trees.

An example of a tree-based technique is Bagging. This algorithm is about building multiple models on different bootstrap samples to obtain a final prediction. A bootstrap sample is the result from resampling with replacement from the original dataset. In addition, each model is created based on recursive binary splitting (Merkle & Shaffer, 2011). This means that for each split, all the predictors and their corresponding possible values and cutpoints are considered. The split which results in the lowest error rate is then selected. This process is repeated within the resulting regions until a stop criterion is reached. In case of classification, the majority vote is then taken, which means that the overall prediction is the most common class among the predictions. Liaw et al. (2002) define Bagging as “successive trees that do not depend on earlier trees — each is independently constructed using a bootstrap sample of the data set” (p.18).

Random Forest is another tree-based method ensemble built on the principles of Bagging. As with Bagging, the Random Forest algorithm constructs each tree using a different bootstrap sample of data and uses binary recursive splitting when building the trees. However, it adds an additional layer of randomness to Bagging (Liaw et al., 2002). Instead of searching for the best split among all variables, a random sample of m predictors is chosen as split candidates from the full set of p predictors.

Only considering a subset of the predictors can be seen as an improvement over bagged trees, because the resulting trees will be less correlated. Suppose there is one very strong predictor in the dataset. In the case of Bagging trees, those predictors will appear at the top of each tree. Hence, Bagging trees have a high probability of being correlated, which leads to a smaller decrease in variance. Random Forests are known to be very robust, consistent and do not overfit (Breiman, 2001).

With regard to Bagging and Random Forest, it is necessary to decide how many trees need to be created. Since increasing the number of trees does not lead to overfitting, it is advised to set this number high enough (Breiman, 2001). In this research, the number of trees is therefore set to 1000. Besides the number of trees, Random Forest has an additional parameter, i.e. the number of variables considered at each split (m). In this research, a common used rule of thumb is applied: this parameter is set to the square root of the total number of predictors. To implement the Random Forest Classification algorithm in the analysis, the *sklearn.ensemble.RandomForestClassifier* package for Python is employed.

4.2.2 Variable Importances and Partial Dependence Plots

In practice, the best split is determined by selecting the feature with the least node impurity as the root node (Berk, 2008). The Gini Index is often used as a measure for this node impurity. It is calculated as follows for a classification task of K classes where p_k stands for the probability that an object will be classified to class k.

$$\text{Gini Index} = \sum_{k=1}^{K} p_k(1 - p_k)$$

Since this study considers only two classes, this simply becomes $p(1-p)$ with p representing the probability of having an engaged post. The Gini Index is, thus, used to evaluate the quality of each specific split.

Based on the Gini Index, the variable importances can be calculated. This is an additional advantage of Random Forest and Bagging. This is done based on the Mean Decrease in Gini Index (MDGI). The MDGI of a variable represents the average decrease of node impurity across all trees in the forest when that variable is omitted. The MDGI of a variable is computed by averaging the decreases in the Gini Index of all the splits in the forest formed on the basis of that variable (Archer & Kimes, 2008). Variables with a high MDGI are considered to be the most important because the node impurity decreases significantly when these variables are omitted. These measures will help us identify which predictors have the greatest influence on brand engagement. The variable importance plots are discussed for the most complete model, i.e. Model 4, for all three categories and for both likes and comments. Beside the use of the different models explained in the previous subsection, the variable importance plots provide additional information to answer **RQ1**. These plots could tell us that if for example mainly influencer characteristics appear at the top of the plot, the influencers have an additional impact on brand engagement. To answer **RQ2-RQ13**, our study assumes that the first 12 variables of the variable importance plot are important and, thus, have an impact on brand engagement.

The MDGI can also help to create the so-called partial dependence plots (PDPs). A PDP shows the marginal effect a feature has on the predicted outcome of a machine learning model (Friedman, 2001). This is in our case the probability of having an engaged post. In this way, the PDPs can show whether the relationship between the target and a feature is linear, monotonic or more complex. This is very effective because unlike e.g. logistic regression, in Random Forest no coefficients can be retrieved for the predictors. In this research, the partial dependence plots are constructed for all the models related to the most complete model, i.e. Model 4. For the numerical variables, line charts are used while for the dummy variables bar charts are employed. Since PDPs do not show the feature distribution, it is indicated for each data point how many observations correspond to that value of the feature. In this way, it is avoided that regions with almost no data are not over-interpreted. For the top 12 features of the variable importance plot, the PDPs make it possible to answer **RQ2-RQ13** in more detail by determining the relationships.

This approach and analysis is done separately for the categories fashion, beauty and food. In this way, differences between the categories in terms of important variables and their relations with brand engagement are determined. With this information, **RQ14** can be answered.

4.3 Performance Measures

The most commonly used performance measure in classification approaches is the Area Under the ROC Curve (AUC). This metric is preferred because it does not depend on the selected threshold compared to cut-off dependent metrics such as accuracy, sensitivity and specificity. This is important because it is possible to have very good results for the latter, while having a completely inappropriate cut-off value. The calculation of AUC is based on a comparison between the predicted state of an event and the actual state of an event for all possible cut-off measures between 0 and 1 (Larivière & Van den Poel, 2005). The AUC can be calculated as follows:

$$AUC = \int_0^1 \frac{TP}{(TP + FN)} d \frac{FP}{(FP + TN)} = \int_0^1 \frac{TP}{P} d \frac{FP}{N}$$

With TP: True Positives, FN: False Negatives, FP: False Positives, TN: True Negatives, P: Positives and N: Negatives. The values of the AUC can range from 0 to 1. A value of 0,5 indicates that the model does not outperform a random classifier. On the other hand, an AUC equal to 1 suggests a perfect predictive model.

Using the Receiver Operating Curve (ROC), the AUC can be represented graphically. This is done by plotting the TPR (sensitivity) against the FPR (1-specificity) (Bogaert et al., 2018). In this study, sensitivity refers to the proportion of engaged posts that are effectively classified as engaged. The specificity is then the proportion of unengaged posts that are correctly classified as unengaged. Furthermore, the 1-specificity is the proportion of engaged posts that are incorrectly classified as unengaged. The surface under the ROC curve should be as large as possible.

For each of the 24 models, 5x2 fold cross validation is applied. This means that the data is randomly split into two mutually exclusive subsets or folds of more or less equal size. The model is trained and tested on those two subsamples. This is done five times for each model. Furthermore, the average ROC curve is constructed for each model.

5 Results

The results of this research are divided into three parts. The first part concerns assessing the performance of the four models discussed in the previous section. This is done for each category and each engagement measure. In this part, **RQ1** can be partially answered. In the second part of this section, the features of the most complete model are analyzed in more detail using variable importance plots and partial dependence plots. For each feature it is investigated whether it has an impact on brand engagement and if so, whether this impact is positive or negative. The research questions **RQ2-RQ13** are answered in this part. The last part of the results provides a summary of the responses to **RQ1** and **RQ2-RQ13** for each category. Based on this summary, differences between the different categories are observed and an answer to **RQ14** is given.

5.1 Models

The four models of Table 10 on page 35 will be investigated in order to determine whether influencers have an additional impact on brand engagement. First, an overview of the AUCs is given for each model. Thereafter, the AUCs are checked for significant differences by making use of the Delong test.

5.1.1 Performance of the Models

The AUCs of the four models regarding likes and comments are represented in Table 11 for each category. Consequently, 24 models are performed. For all 24 models, the detailed results of the 5x2 fold cross validation can be found in Appendix D. Moreover, Figure 4, Figure 5 and Figure 6 show the average 5x2 fold cross validation ROC curves for each of the 24 models.

	AUC					
	Likes			**Comments**		
	Fashion	*Beauty*	*Food*	*Fashion*	*Beauty*	*Food*
Model 1	0,6656	0,6743	0,6055	0,6011	0,6264	0,5487
Model 2 (Post Influencer)	0,6838	0,6507	0,6287	0,6051	0,6121	0,5510
Model 3 (Influencer)	0,6801	0,6773	0,6161	0,6066	0,6175	0,5454
Model 4 (Both)	0,6784	0,6520	0,6148	0,6046	0,6102	0,5510

Table 11: Median 5x2fcv AUC for each model

One can see that all AUCs are higher than 0,50. This indicates that all algorithms perform better than random. The AUCs range from 0,5454 to 0,6838. Besides the food model regarding comments, all the AUCs are higher than 0,60. Given the fact that there are probably other relevant factors influencing brand engagement that are not included in this study, these measures are quite good. Moreover, the goal of this study is to indicate relationships between features and engagement and their differences across different categories rather than purely predicting whether a post will be engaged or not.

We find that the AUCs for the models of likes are higher than those of comments. This can be explained by the fact that likes are given faster because less effort is needed (Yew et al., 2018). There are a lot of brand posts with hardly any comments. The reason for this is because users find it unnatural to communicate directly with a brand (Hellberg et al., 2015). Thus, when a post has only two comments, it can quickly be perceived as "engaged" as the median of this brand would be very low. Hence, the probability of having an engaged post with respect to comments, is less stable than an engaged post with respect to likes. Furthermore, one can observe that fashion and beauty models have a higher performance than food models. This can be explained by the fact that beauty and fashion have a higher number of observations than food.

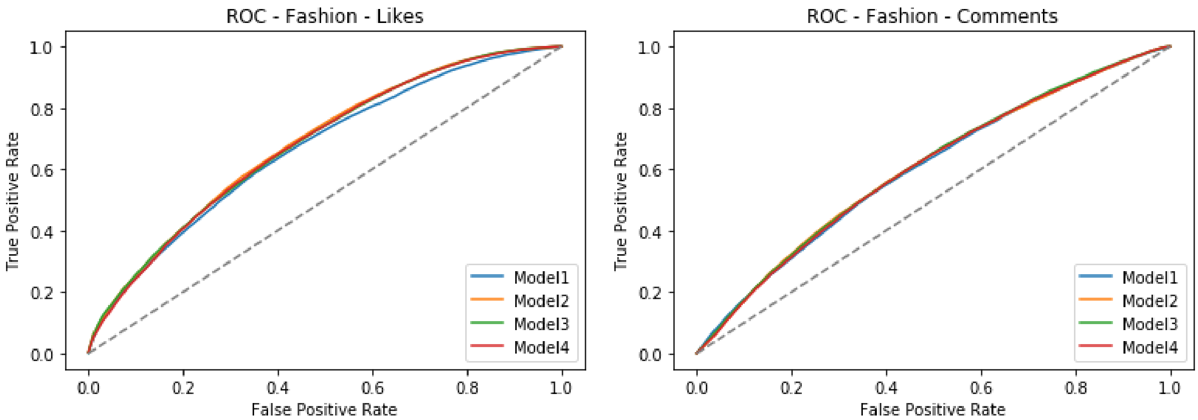
We find for the fashion category that Model 2 (0,6838) performs best and Model 1 (0,6656) performs worst in terms of likes. Model 3 (0,6801) and Model 4 (0,6784) look similar. Moreover, their AUCs differ less with Model 2 than with Model 1. The ROC curves in Figure 4a confirm this by showing that Model 1, Model 2 and Model 3 are better than Model 4 because they lay above them. With respect to the comments, there are hardly any differences between the models. Likewise, the different lines of the ROC curves in Figure 4b are indistinguishable.

As for beauty, we see that Model 3 (0,6773) performs best with respect to the likes. However, the difference with Model 1 (0,6743) is only minor. Both models perform better than Model 2 (0,6507) and Model 4 (0,6520). Likewise, the ROC curves in Figure 5a show that Model 1 and Model 3 are placed above Model 2 and Model 4. Concerning the comments, we find less differences in the models. However, Figure 5b indicates that Model 1 (0,6262) performs slightly better than the other models.

As for the category food, we find that Model 2 (0,6287) performs the best and Model 4 (0,6055) performs the worst for the likes. Model 3 (0,6161) and Model 4 (0,6148) are somewhere in between. From the ROC curves in Figure 6a, it is noticeable that Model 2 outperforms the rest. The low

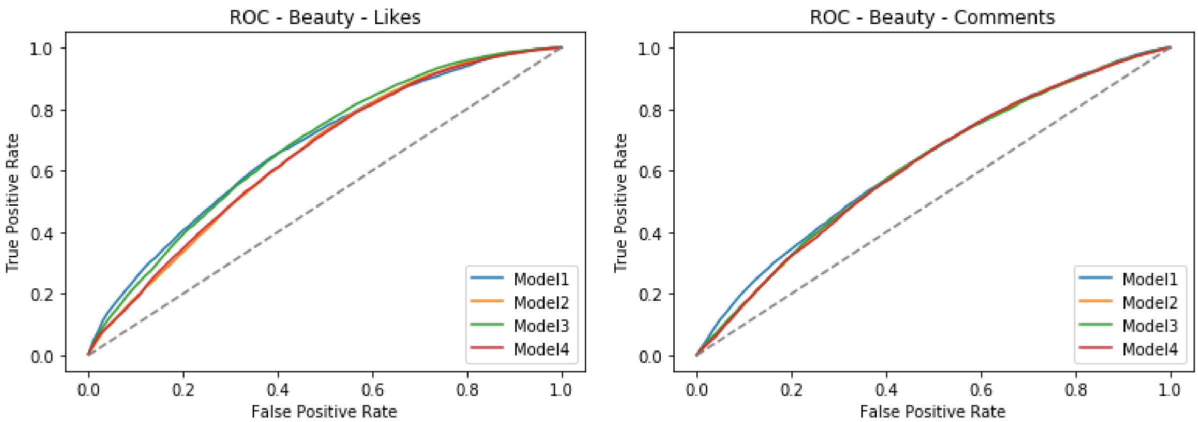
AUC values for the comments in case of food, show little differences. The same conclusions can be drawn when looking at the ROC curves in Figure 6b.

The next subsection examines whether the aforementioned performance measures differ significantly.



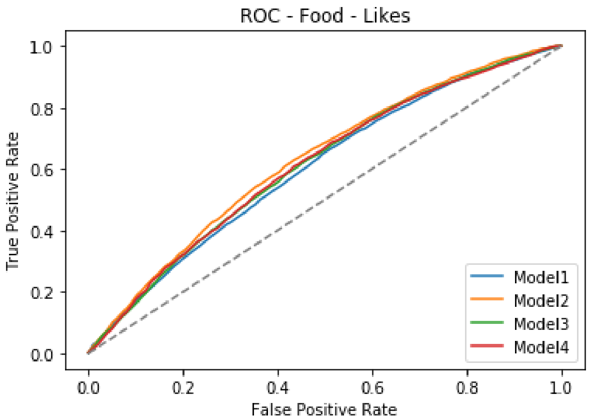
(a) ROC - Likes as dependent (b) ROC - Comments as dependent

Figure 4: Averaged 5x2fcv ROC Curves for category Fashion

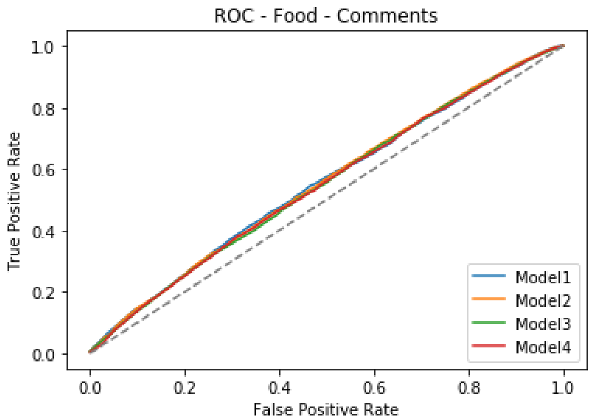


(a) ROC - Likes as dependent (b) ROC - Comments as dependent

Figure 5: Averaged 5x2fcv ROC Curves for category Beauty



(a) ROC - Likes as dependent



(b) ROC - Comments as dependent

Figure 6: Averaged 5x2fcv ROC Curves for category Food

5.1.2 Delong Test

The Delong test is conducted to determine for which models the AUCs show significant differences. This is done based on the estimated probabilities of the full dataset that generated the median AUC. For each possible combination of models within the same category and engagement measure, a pairwise comparison is performed. The results in terms of p-values are displayed in Table 12. We find that there are only significant differences between the models regarding likes.

	Delong Test					
	Likes			**Comments**		
	Fashion	*Beauty*	*Food*	*Fashion*	*Beauty*	*Food*
Model 1 - Model 2	0,0002***	0,0162**	0,0182**	0,3234	0,1870	0,9504
Model 1 - Model 3	0,01493**	0,4199	0,3839	0,2762	0,3861	0,8026
Model 1 - Model 4	0,01209**	0,0154**	0,5177	0,4522	0,1408	0,8673
Model 2 - Model 3	0,1004	0,0005***	0,0957*	0,9530	0,5639	0,8462
Model 2 - Model 4	0,1210	0,9631	0,0435**	0,7677	0,8169	0,8196
Model 3 - Model 4	0,8091	0,0002**	0,8608	0,7340	0,4268	0,6808

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 12: Delong Test - P-values

For fashion, we conclude that Model 2, Model 3 and Model 4 perform significantly better than Model 1. As expected for the beauty category, Model 1 and Model 3 perform significantly better than Model 2 and 4. There are no significant differences between Model 1 and Model 3 and between Model 2 and Model 4. For the food category, Model 2 performs significantly better than the other models. There are no significant differences between Model 1, Model 3 and Model 4.

Different answers can be given to **RQ1**, depending on the type of category. It has to be said that the inclusion of influencer characteristics and post influencer characteristics does not add extra value if we want to predict brand engagement using comments. However, it does not harm the model either. On the other hand, for the likes, including both influencer characteristics and post influencer characteristics significantly add value to the fashion model. For beauty, we can conclude that influencer characteristics (Model 3) add extra value in explaining the brand engagement. However, this added value is not significant. Finally, for food brands, post influencer characteristics (Model 2) have a significant additional impact on the number of likes. However, **RQ1** is only partially explained because the variable importance plots discussed in the next subsection, give us additional information.

5.2 Features

In order to answer research questions **RQ2-RQ13**, Model 4 is examined in more detail. For each of the three categories fashion, beauty and food the results regarding influencer characteristics, post influencer characteristics and post brand characteristics are investigated in more detail.

For each category, the variable importance plot of all the features is given for the model regarding likes and the model regarding comments. We assume that the first 12 characteristics of the plots are important. The other variables are assumed to have no impact on brand engagement.

Afterwards, based on PDPs, the relationships of the 12 most important characteristics are discussed regarding brand engagement. The PDPs can be found in Appendix F. For each point, we have added the number of observations because we need to be careful in interpreting areas where nearly observations occur. Moreover, it happens sometimes that outliers appear. Therefore, when we explain the relations, we only focus on the first part of the plots where a significant amount of observations occur.

5.2.1 Fashion

Likes

It is clear from the variable importance plot, shown in Figure 7, that the influencer characteristics and post influencer characteristics are mainly present in the fashion industry. This confirms our prior findings that Model 2, Model 3 and Model 4 perform significantly better than Model 1.

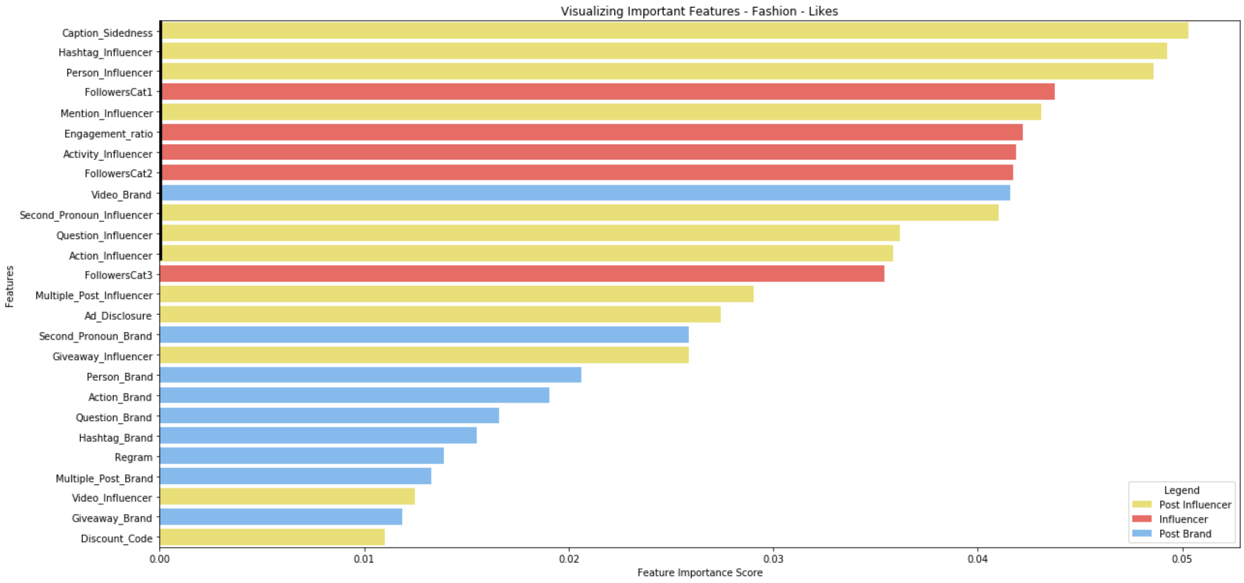


Figure 7: Variable Importances Fashion - Likes as dependent

In the top of the most important variables only one post brand characteristic is present. The PDP indicates that the variable *Video_Brand* has a negative relation with the likes of the brand's post (**RQ11b**).

Seven post influencer characteristics are present in top of the variables importance plot. The most important variable is *Caption_Sidedness*. The use of influencers with two-sided captions about the fashion brand seems to have a negative impact on the number of likes of the brand's post (**RQ5**). The next most important variable is the influencer's interaction level. All five interaction variables occur in the top 12. The use of hashtags (*Hashtag_Influencer*), second pronouns (*Second_Pronoun_Influencer*) and questions (*Question_Influencer*) has a negative relation with the number of likes. On the other hand, mentioning the brand in the influencer's caption (*Mention_Influencer*) has a positive relation with the number of likes. No clear relationship is found for *Action_Influencer*. We can conclude that the influencer's interaction level has a clear impact on the number of likes of the brand's post (**RQ9a**). The last influencer characteristic variable is *Person_Influencer*. The PDP shows that adding a person in the post has a positive effect on the number of likes of the brand's post (**RQ13a**).

Finally, almost all influencer characteristics appear to have an important impact on the number of likes of the brand post. The variable *FollowersCat1* and *FollowersCat2* occur in this plot, but *FollowersCat3* is just outside the top 12. Therefore, influencers of category 3 seem less important. Influencers of category 1 have a rather mixed relation, i.e. first a positive relation is observed, but after more influencers of this category are present, the relation with the number of likes becomes negative. Influencers of category 2, however, seem to have a positive relationship with the number of likes of a brand's post. We can conclude that the number of followers have an impact on the likes (**RQ2**). The *Engagement_ratio* of an influencer is also considered important. The PDP indicates that engaged influencers negatively affect the number of likes. *Activity_Influencer* has the same negative relation with the number of likes. Hence, both the engagement ratio and the activity of an influencer have a negative impact on the number of likes of the brand (**RQ3**, **RQ4**).

Comments

In Table 12 on page 43, no significant differences were found between the models. However, it is noteworthy that all post brand characteristics are at the bottom of the variable importance plot (Figure 8). Hence, these results suggest that (post) influencer characteristics are more important to enhance the number of comments than brand characteristics.

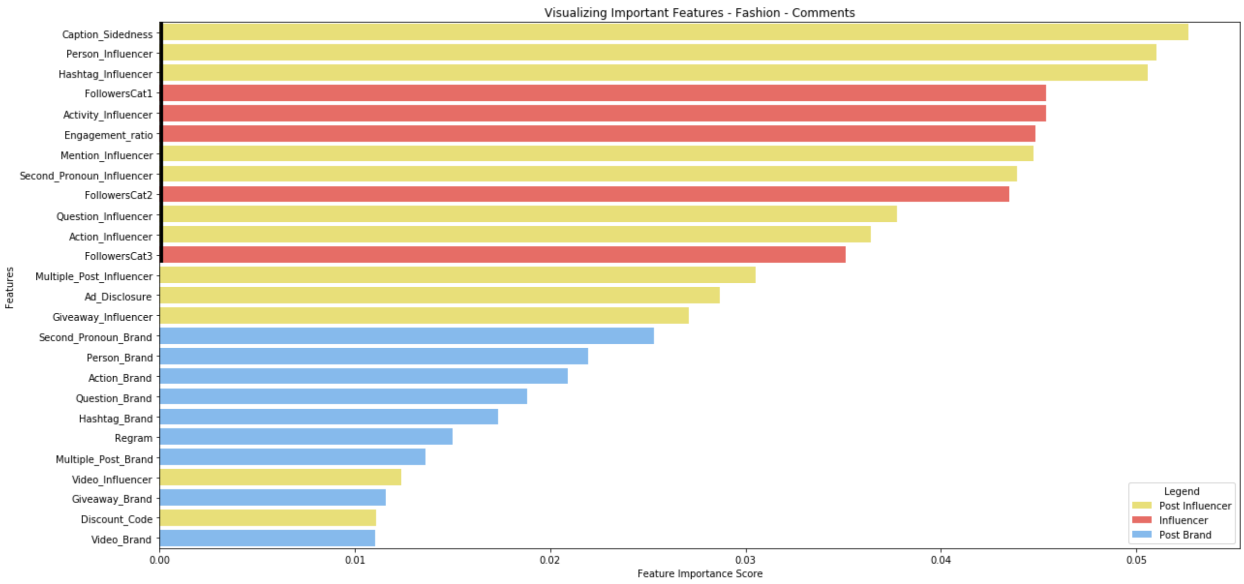


Figure 8: Variable Importances Fashion - Comments as dependent

As with the model with likes, a two-sided caption has the most influence on the number of comments of the brand's post. The PDP suggests that *Caption_Sidedness* has a positive influence on the number of comments (**RQ5**). The opposite is true for the likes. The second most important variable is whether or not there is a person in the photo or video. Again, the variable *Person_Influencer* has a positive impact (**RQ13a**). The third most important variable is the interaction level of the caption of the influencer. Again, all five interaction variables are present in the top most important variables. It is furthermore remarkable that all the interaction variables have a positive impact on the number of comments. Thus, the use of hashtags (*Hashtag_Influencer*), mentions (*Mention_Influencer*), second pronouns (*Second_Pronoun_Influencer*), questions (*Question_Influencer*) and actionable verbs (*Action_Influencer*) enhances the number of comments of the brand's post (**RQ9a**). It is striking that all post influencer characteristics have a positive influence on the number of comments of the brand's post.

In Figure 8, it is clear that all influencer characteristics are again present in the top most important variables with respect to the number of comments. *FollowersCat1*, *FollowersCat2* and *FollowersCat3* all have a positive relation with the number of brand comments. Our results suggest that influencers of category 1 have the biggest impact on the number of likes, followed by influencers of category 2 and 3 (**RQ2**). The next key influencer characteristics are *Engagement_ratio* and *Activity_Influencer*. Similar to the likes model, engaged and active influencers seem to have a negative impact on the number of comments of the brand (**RQ3**, **RQ4**).

5.2.2 Beauty

Likes

Although Model 1 and Model 3 perform significantly better than Model 2 and Model 4, we see that in the beauty sector mainly influencer characteristics and post influencer characteristics occur at the top of the variable importance plot. This can be found in Figure 9.

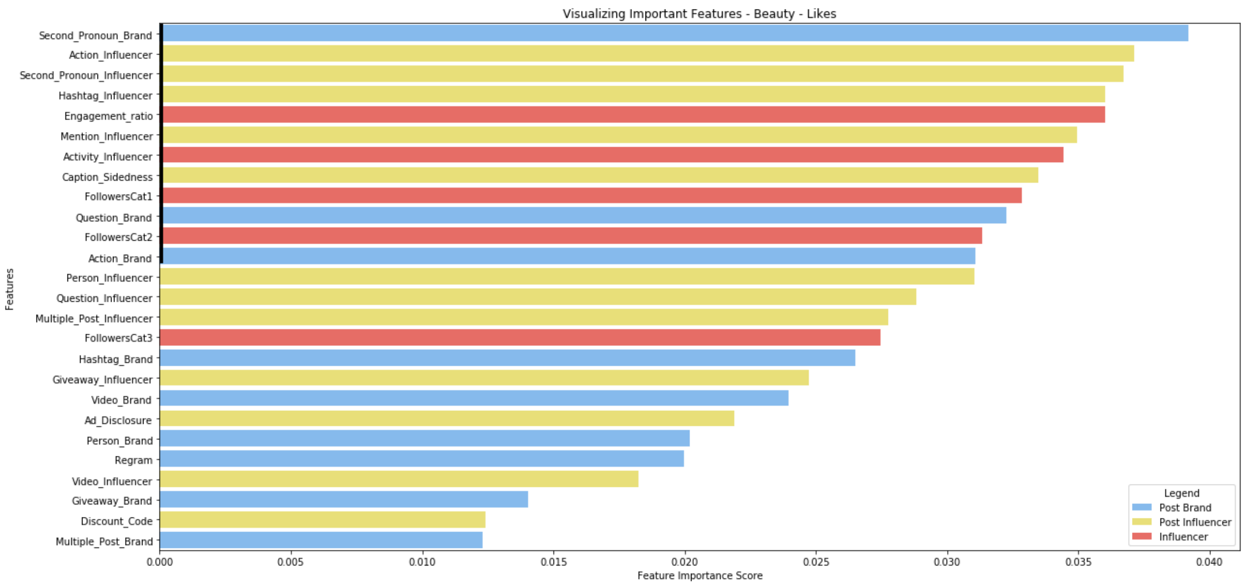


Figure 9: Variable Importances Beauty - Likes as dependent

The interaction variable *Second_Pronoun_Brand* is the most important feature. However, the next brand interaction variables, *Question_Brand* and *Action_Brand*, only occur at the 10th and 12th place. Nevertheless, these results show that interaction has an impact on the likes of a brand post (**RQ9b**). Furthermore, if we look at the PDPs of these variables, we see that for *Second_Pronoun_Brand* no differences occur. For *Question_Brand* there is a a positive impact, while *Action_Brand* has a rather negative impact on likes.

Figure 9 tells us that all interaction variables, except for *Question_Influencer*, are important with regard to influencers. The variables *Action_Influencer* and *Hashtag_Influencer* have a positive relation with the likes of the brand, while the variables *Second_Pronoun_Influencer* and *Mention_Influencer* have a mixed relation. More influencers using second pronouns or mentions results in increasing likes. However, there should not be too many influencers who use these aspects, because this could lead to a decrease in the number of likes. Our results suggest that the interaction of an influencer post is important and more or less positively related to the likes of the brand (**RQ9a**). A second post influencer characteristic that is important is the *Caption_Sidedness*. The PDP shows that the more influencers use caption sidedness, the higher the likes of the brand post. We can conclude that there is a positive relation between caption sidedness of the influencers and the likes of the brand post (**RQ5**).

Four out of the five influencer characteristics occur in the top 12 of the plot in Figure 9. The most important variable here is the *Engagement_ratio*. There exists a positive relation between engaged influencers and the likes of the brand post (**RQ3**). Secondly, *Activity_Influencer* appears shortly after the engagement ratio in the variable importance plot. Our results also indicate a positive relationship (**RQ4**). Finally, influencers of *FollowersCat1* have a larger impact on the brand post's likes compared to influencers of *FollowersCat2*. Remarkably, *FollowersCat3* is not among the most important variables. The PDPs show that there is a negative relation between *FollowersCat1* and the likes of a brand post, while a rather positive relation is detected by the *FollowersCat2* (**RQ2**).

Comments

No significant differences were found between the Models 1 , 2 , 3 and 4 considering the comments as dependent variable in the beauty sector. It is remarkable, however, that the top of the variable importance plot shows mainly post influencer characteristics, followed by influencer characteristics. This is displayed in Figure 10.

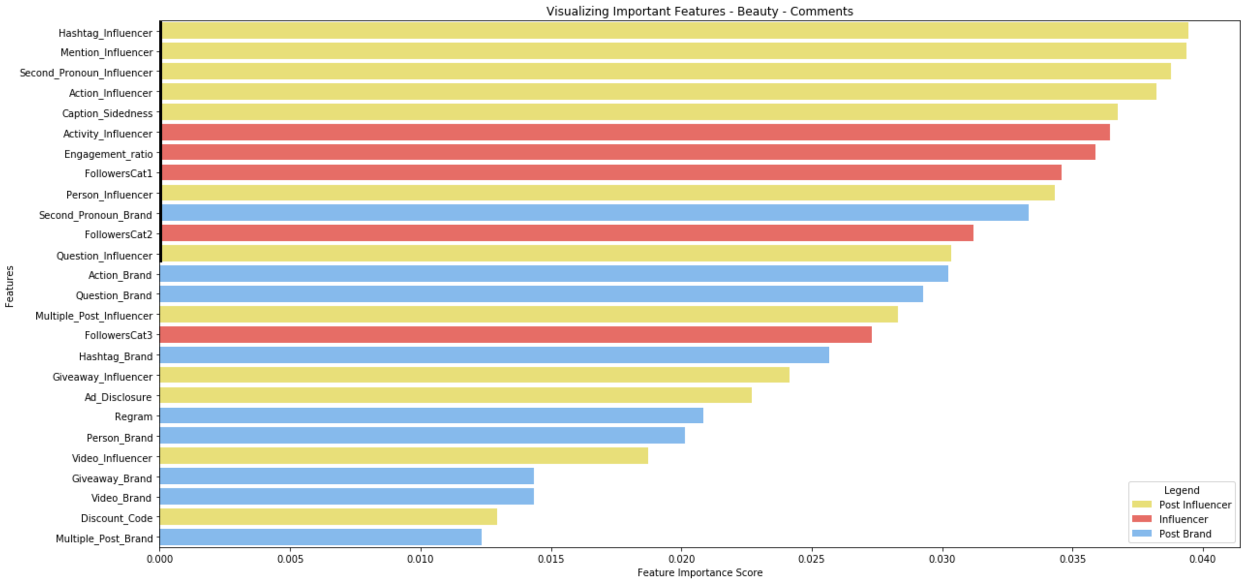


Figure 10: Variable Importances Beauty - Comments as dependent

Once again, *Second_Pronoun_Brand* is one of the important features. This is the only important brand feature present in the top 12. In the model with likes this feature comes first, while in the model with comments it comes tenth. The PDP indicates clearly that the use of second pronouns has a positive effect on the number of comments of the brand post. Therefore, we can conclude that *Second_Pronoun_Brand* has an impact on brand engagement (**RQ9b**). However, the four other interaction variables are not present in the top 12 variables of Figure 10.

As with the likes model, the interaction post influencer characteristics and the *Caption_Sidedness* variable appear at the top of the variable importance plot. *Hashtag_Influencer*, *Mention_Influencer*, *Second_Pronoun_Influencer* and *Action_influencer* appear consecutively in the plot. In contrast to the comments model, *Question_influencer* is here in the top 12. The use of hashtags and second pronouns has a rather negative influence on the number of comments of the brand's post, while bringing the followers into action , asking questions and providing a link to the brand has a positive influence on the number of comments of the brand's post. We can conclude that the interaction of influencers has an impact on the number of comments, because all interaction variables are at the top of the plot (**RQ9a**). The second most important variable of post influencer characteristics is the *Caption_Sidedness*. Our results suggest a positive impact of influencers using caption sidedness on the number of comments of the brand post (**RQ5**). In the engagement model with respect to comments, a third post influencer characteristic is important, i.e. *Person_Influencer*. Having a person in the influencer's post has a negative effect on the number of comments of the brand's post

(**RQ13b**).

The important influencer characteristics variables are similar to those in the likes model. However, *Activity_Influencer* is more important than *Engagement_ratio* which is the opposite in the model with the likes. The same conclusions concerning the relations can be drawn here. More active and engaged influencers result in an increase in the number of comments (**RQ3**, **RQ4**). Similar to the model with likes, the influencers in *FollowersCat1* are more important than influencers in *FollowersCat2*. Again, influencers in *FollowersCat3* are considered less important. The PDP of *FollowersCat1* indicates a positive relation on the number of comments whereas there is a negative relation on the number of likes (**RQ2**). The PDP of *FollowersCat2* indicates a mixed relation. First, there is a negative trend when adding more influencers from category 2 but after a certain number of influencers in this category, the probability of having an engaged post increases. However, the probability of having an engaged post is not higher than having no influencers of category 2. Therefore, we can conclude that there is a negative relationship with the number of comments (**RQ2**).

5.2.3 Food

Likes

A mix of (post) influencer characteristics and brand characteristics can be found in the top of the variable importance plot shown in Figure 11, while we have found that Model 2 is significantly better than the other models.

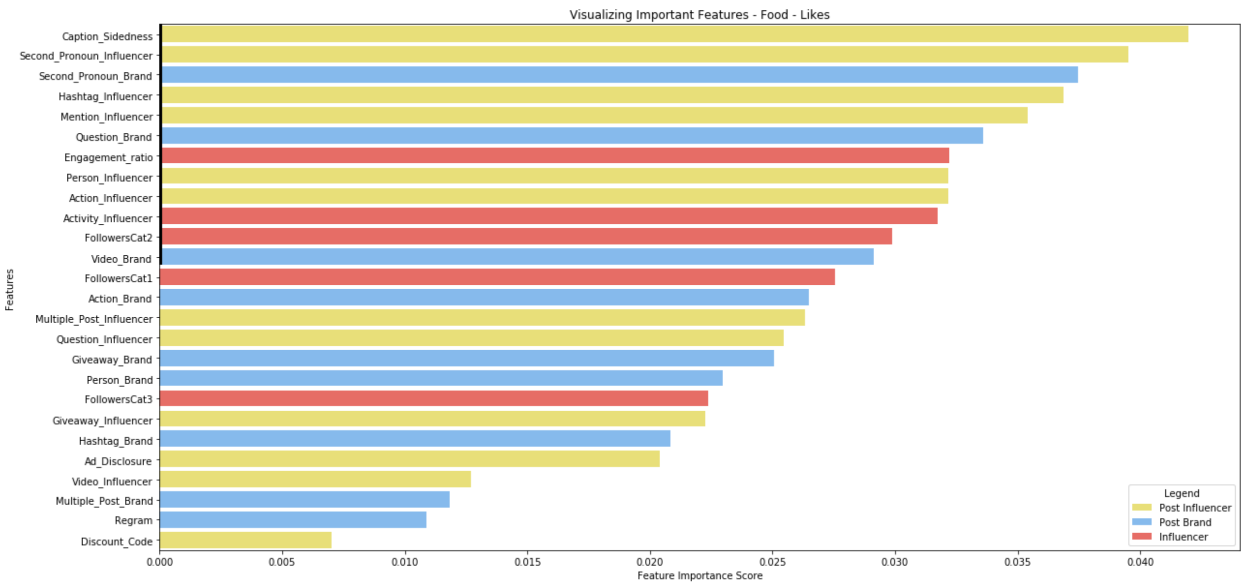


Figure 11: Variable Importances Food - Likes as dependent

Only three post brand characteristics seem to be important for improving the number of likes. The *Second_Pronoun_Brand* variable is the third most important variable here. The PDP shows a slightly negative trend. Therefore, we can conclude that using second pronouns in the caption of the brand rather has a negative effect on the likes of that post. Asking questions in the caption of the brand (*Question_Brand*) or posting a video (*Video_Brand*) seems to negatively influence the likes. Hence, we can conclude that posting videos has a negative impact on brand engagement with regard to likes (**RQ11b**). As only *Question_Brand* and *Second_Pronoun_Brand* of the five interaction terms occur in the top, we can conclude that the interaction of the brand has a semi-impact on the number of likes (**RQ9b**).

The post influencer characteristics are prominently present in the top 12 important features. The most important variable related to the number of likes, is the *Caption_Sidedness*. The PDP of this variable indicates that the trend first increases and afterwards decreases. However, this increase is rather small compared to the decrease. Therefore, we conclude that the caption's sidedness of an influencer has a negative impact on the number of likes of the brand's post (**RQ5**). Secondly, the interaction aspect of the post of an influencer is a feature that needs attention. Four out of five interaction variables occur in the most important variables. The most important feature among them is *Second_Pronoun_Influencer*. The use of second pronouns in the caption of an influencer, increases the likes of the brand. Likewise, the use of hashtags (*Hashtag_Influencer*) or the use of mentions (*Mention_Influencer*) in the caption enhances the total number of likes of the brand post. On the other hand, bringing followers into action (*Action_Influencer*) has a negative effect. Therefore, we can conclude that interaction has nearly a positive impact, except for *Action_Influencer* and *Question_Influencer*, on brand engagement with regard to the likes (**RQ9a**). The last post influencer characteristic to be discussed is the *Person_Influencer*. The PDP indicates that showing a person in the post of an influencer has a positive impact on brand engagement with respect to likes (**RQ13a**).

When more active and engaged influencers are used, the likes of the brand are more likely to increase. The influencer characteristic variables *Engagement_ratio* and *Activity_Influencer* have a positive impact on the likes of the brand's post (**RQ3**,**RQ4**). The *FollowersCat2* variable has a negative impact on brand engagement (**RQ2**).

Comments

Although all the comments models in the food sector do not differ significantly, the (post) influencer characteristics appear to be the most important variables to enhance the number of comments. This is displayed in Figure 12.

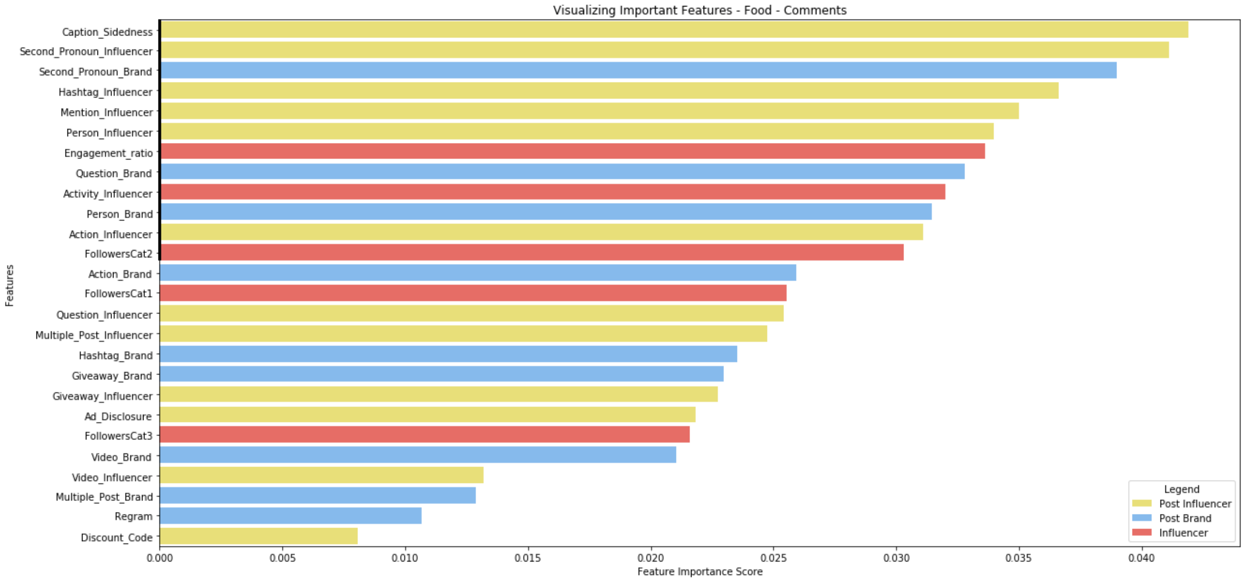


Figure 12: Variable Importances Food - Comments as dependent

The post brand variables are less common in the top 12. As with the model of likes, *Second_Pronoun_Brand* and *Question_Brand* occur in the top important variables. However, in contrast with the model on likes, adding second pronouns or questions in the caption of the brand has a positive effect on the number of comments (**RQ9b**). The last important post brand variable is *Person_Brand*. The PDP suggests that publishing a brand post with a person in it tends to increase the number of comments. We can therefore conclude that including person in the brand post has a positive impact on brand engagement with respect to the comments (**RQ13b**).

The same post influencer variables as in the model with likes occur at the top of the plot. However, the variables come in a different sequence and the relations are different. The most important variable is *Caption_Sidedness*. The PDP shows a positive relation and hence, there is a greater chance that the brand post gets a comment when more influencers use caption sidedness (**RQ5**). Secondly, the interaction terms *Second_Pronoun_Influencer* and *Hashtag_Influencer* belong to the most important variables. Our results suggest that adding second pronouns or hashtags in the caption of the influencer has a negative impact on the number of comments. The other mentioned interaction terms, *Mention_Influencer* and *Action_Influencer* have a positive influence on the number of comments. We can conclude that the interaction level has an impact on the number of comments of the brand post (**RQ9a**). The last post influencer variable is *Person_Influencer*. The PDP indicates that when an influencer posts a photo or video with a person in it, it has a negative impact on the number of comments (**RQ13a**).

Finally, similar to the model with likes, the influencer characteristics *Engagement_ratio* and *Activity_Influencer* have a positive impact on the number of comments of the brand's post. The last variable in the top is *FollowersCat1*. Influencers from category 1 seem to have a negative influence on the number of comments of the brand post (**RQ2**).

5.2.4 Control Variables

The variable importance plots shown above, are without the control variables. The reason for this is because these variables are not the focus of our research. However, the interested reader can find the variable importance plots with control variables in Appendix E and their PDPs in Appendix F. It is remarkable that for all three categories, the month in which a post is published seems to be very important. We find that the most likes and comments are received in January and February. Next, we find that more active brands receive more likes, whereas for comments the relation is less clear. Furthermore, brands in the category food are less active compared to the other categories. Fashion brands post on a daily basis while some food brands do not post anything for a week. Finally, if a post is published in a weekday instead of the weekend, the post is more likely to receive comments for all categories. This is confirmed in the study of Cvijikj and Michahelles (2013). With respect to the likes, more likes are received in the weekend in the fashion category. For the other categories, no relationship is found.

5.3 Summary

This subsection provides a summary of the answers to all research questions. First, an explanation is given to **RQ1**. Secondly, Table 13 displays an overview of **RQ2-RQ13**. In this table, the differences across the categories are analysed and hence, an answer to **RQ14** is established.

Concerning the likes of brands in the fashion category, the models with (post) influencer characteristics perform significantly better than the model with only post brand characteristics. Therefore, it is clear that influencers do have an impact on the likes in the fashion industry. This is confirmed by Table 13 as only one post brand characteristic has an impact on the likes. Furthermore, we find no significant differences between the models with regard to the number of comments. However, all important variables in Table 13 are influencer variables. Hence, we can conclude that influencers still have an influence on the number of comments of a brand's post in the fashion industry.

		Likes			Comments		
		Fashion	*Beauty*	*Food*	*Fashion*	*Beauty*	*Food*
Influencer Characteristics							
RQ 2	Number of Followers						
	Category 1	+/-	-	•	+	+	-
	Category 2	+	+	-	+	-	•
	Category 3	•	•	•	+	•	•
RQ 3	Engagement Ratio	-	+	+	-	+	+
RQ 4	Activity	-	+	+	-	+	+
Post Influencer Characteristics							
RQ 5	Caption Sidedness	-	+	+/-	+	+	+
RQ 6	Discount Code	•	•	•	•	•	•
RQ 7	Ad Disclosure	•	•	•	•	•	•
RQ 9a	Interaction						
	Hashtags	-	+	+	+	+	-
	Action	=	+	-	+	+	+
	Question	-	•	•	+	+	•
	Second Pronouns	-	+/-	+	+	-	-
	Mention	+	+/-	+	+	+	+
RQ 10a	Giveaway	•	•	•	•	•	•
RQ 13a	Image Content (Person)	+	•	+	+	-	-
RQ 11a	Video	•	•	•	•	•	•
RQ 12a	Multiple Post	•	•	•	•	•	•
Post Brand Characteristics							
RQ 8	Regram	•	•	•	•	•	•
RQ 9b	Interaction						
	Hashtags	•	•	•	•	•	•
	Action	•	-	•	•	•	•
	Question	•	+	-	•	•	+
	Second Pronouns	•	=	-	•	•	+
RQ 10b	Giveaway	•	•	•	•	•	•
RQ 13b	Image Content (Person)	•	•	•	•	•	+
RQ 11b	Video	-	•	-	•	•	•
RQ 12b	Multiple Post	•	•	•	•	•	•

+ : positive relation, - : negative relation, +/- : mixed relation, • : not important

Table 13: Summary of results

We find that for beauty, the model with additional influencer characteristics and the model with only post brand characteristics perform equally well with respect to the likes. However, Table 13 shows that post influencer characteristics are also important. The results of the Delong test suggest further that influencers do not have a significant impact on the number of comments. However, the opposite is true for the variable importance plots. In Table 13, it can be observed that only (post) influencer characteristics are considered important. We can therefore conclude that influencers have an impact on both likes and comments for beauty brands. As for food, the model with post influencer characteristics outperform the other models with respect to the likes. No significant differences are found regarding the number of comments. However, the variable importance plots regarding likes and comments again show that both influencer and post influencer characteristics are important. Therefore, we can conclude that influencers have an impact on food as well. Hence, a clear conclusion can be drawn with respect to **RQ1**. It is noteworthy that the influencer characteristics and post influencer characteristics are prominently present in every category compared to post brand characteristics. In all three categories, the influencers and their posts have a clear impact on brand engagement.

With respect to **RQ2 - RQ13**, the summary of the impact and relations of variables can be found in Table 13 for each category and dependent variable. In what follows, the differences across the categories are analysed and a formulation on **RQ14** is constructed.

Surprisingly, the impact of the number of followers of an influencer is different across the categories. It is clear that influencers from category 3 are only important for the number of comments in the fashion sector. Moreover, influencers from category 2 have a positive impact on the number of comments in fashion and a negative impact in beauty. The same can be observed for influencers from category 1 with respect to the likes. Furthermore, these influencers seem to have no impact on the likes of a food brand page and a negative effect on the number of comments. Another remarkable finding is that engaged and active influencers have a negative influence on both the number of likes and comments in fashion, whereas a clear positive relation is found for beauty and food.

We have seen that the caption sidedness is often the most important feature with respect to likes and comments. The positive impact of two-sided captions of an influencer on the number of comments is steady across the three categories. However, with respect to the likes, two-sided messages have a negative influence for fashion brands, whereas for beauty a positive relation exists. With regard to the discount code and ad disclosure, we can draw a coherent conclusion that they do not affect both the likes and comments for all categories.

We can notice the importance of the influencer's interaction level and the many differences between the categories. The use of hashtags has a negative influence on the number of likes for a fashion brand, whereas a positive influence is found for the other categories. Moreover, with respect to the number of comments, the use of hashtags has a negative impact for food brands whereas a positive relation occurs for the other categories. Next, adding actionable verbs in the caption is beneficial for the likes of beauty brands but not for those of food brands. The likes of a fashion brand are negatively affected by influencers asking questions, whereas no impact was found for the beauty and food category. Influencers using second pronouns also differs in impact across the categories. As for the likes of a fashion brand, this has a negative impact whereas for beauty or food brands the impact is rather positive. The opposite is true for the number of comments. Including a person in the photo or video of the influencer also appears to have different effects on brand engagement. This has a positive impact for the likes of the fashion and food category but no impact was found for the beauty category. Moreover, this has a negative impact on the number of comments for both beauty and food brands whereas a positive influence is found for fashion brands. No substantial impact was further found for giveaways, videos and multiple posts of the influencer on brand engagement for all categories.

We find that post brand characteristics have less impact on the brand engagement compared to influencer characteristics. A regram, giveaway and multiple post of the brand seem not to have an impact on brand engagement in all three categories. The interaction level of the brand only has an impact on brand engagement for beauty and food brands. Adding actionable words in the brand's caption only has a negative impact on likes for beauty brands whereas adding questions and second pronouns only has a positive impact on the number of comments for food brands. With respect to likes, asking questions has a positive impact for beauty brands whereas a negative relation is found for food brands. Including a person in the photo or video of the brand only has a positive impact on number of comments for food brands. Finally, whether the post is a video or not solely has a negative impact on the likes of fashion and food brands.

6 Discussion and Managerial Implications

In this study, we have found that influencers do have an impact on brand engagement. In addition, the results of the analysis show that influencer characteristics, post influencer characteristics and post brand characteristics do not have the same effect on the number of comments and likes. Moreover, there are some differences between the categories fashion, beauty and food. In what follows, a more practical explanation is given on how marketeers and managers of a certain category can use this research as a guideline to increase their brand engagement on Instagram.

6.1 Fashion

If the focus of managers is to increase likes, it is important not to post a lot of videos on the brand page. The reason can be retrieved from the in-depth interviews of Hellberg et al. (2015), in which respondents argue that videos interrupt the flow of scrolling. This is in contrast with the study of Vignisdóttir (2017) that states that videos increase likes. This difference can be explained by the fact that the latter did a study in the beauty sector.

For fashion managers, it is important to collaborate especially with influencers with less than 200K followers in order to optimize brand engagement. This is consistent with the experiment of De Veirman et al. (2017) that found that influencers with too high a number of followers have a negative impact on brand attitude. Surprisingly, working with many engaged and active influencers can have a detrimental effect on brand engagement. This is in contrast to previous Instagram studies. This contrast can be explained by the fact that the engagement ratio (E. L. De Vries, 2019; Henderickx & De Wolf, 2019) and the activity level (Brorsson & Plotnikova, 2017) have only been investigated with respect to the influencer's credibility based on questionnaires on Instagram.

Concerning the influencer's own page, the fashion company should advise the influencer to be positive in the caption, post photos of themselves and set certain rules regarding interaction. To increase the likes on the brand page, a fashion influencer has to be extremely positive about the brand. This is in line with the general study of Eisend (2006) that found that behavioral intentions decrease with a higher amount of negative information. This is an important guideline for fashion managers because nowadays almost 50% of their influencer posts are two-sided, which can be retrieved from Table 8. The fact that the influencer should be in the photo, is in line with the studies of Bakhshi et al. (2014); Z. Zhang et al. (2018b); Ding et al. (2019) that found that photos of people are more popular. Our study extends this by stating that the influencer posting himself/herself also has a positive impact on brand engagement. An explanation can be found in the in-depth interviews conducted by Hellberg et al. (2015) where some respondents claim that in the case of fashion, they find it important to clearly see how the outfit fits.

Regarding the level of interaction, the guidelines depend on whether the emphasis is on comments or likes. An influencer should include hashtags, questions and second pronouns in the caption if the focus is on comments. However, this is at the expense of the likes. Mentioning the brand in the caption, and thus offering a direct link to the brand, is positive for both likes and comments. While Brorsson and Plotnikova (2017) have emphasized the fact that interactivity increases influencer's credibility, no research has been done into the quantitative impact of influencer's interaction level on brand engagement. Several Facebook studies, including Schultz (2017); Lei et al. (2017); Cruz et al. (2017), have found that brand interaction characteristics are important for brand engagement. However, this study finds that primarily the interaction of influencers is important on Instagram. This can be due to the fact that less influencers are active on the Facebook platform (Launchmetrics, 2019). This is also the case for the beauty and food sector.

6.2 Beauty

Beauty managers are advised to take interaction aspects into account when managing their brand page. On the one hand, they should ask questions, because this increases the likes. On the other hand, they should not be too obtrusive, so the caption should not contain many actionable verbs. Previous studies were inconclusive about the impact of questions and actionable verbs on brand engagement on Facebook. Our results were the first for beauty brands on Instagram. However, these results are consistent with the Facebook study of Schultz (2017).

Smaller influencers with followers between 1K and 200K have a greater impact on engagement of beauty brands. This is consistent with the study of De Veirman et al. (2017). This study states that it is more important to choose smaller influencers with similar interests and the same target audience as the brand, than influencers with an overall larger reach. If the brand wishes to enhance the comments, a preference is given to influencers with 1K-20K followers. On the other hand, to enhance the likes, influencers with 20K-200K should be chosen. Beauty managers should also take the engagement ratio and activity of the influencers into account as they enhance the brand engagement in general. The studies of Henderickx and De Wolf (2019) and Brorsson and

Plotnikova (2017) on Instagram have concluded that the engagement ratio and activity ratio have a positive impact on the influencer's credibility and trust. Our results extend this by stating that this credibility and trust also relates to brand engagement in the beauty industry.

Managers should advise their influencers to be honest in their posts in order to increase the brand engagement. This means that influencers should also mention the negative points of a product. This way, an influencer is perceived as reliable, as concluded in the study by Braatz (2017). We find that this reliability evokes likes and comments on the beauty brand page. In addition, if a manager wishes to enhance the comments, no person should be included in the post. This is confirmed by the study of Vignisdóttir (2017), which is also an Instagram study concerning the beauty industry. The interaction of an influencer's post appears to have an impact on the brand engagement. It is therefore advisable to impose rules on the influencer's caption. Adding hashtags, actionable words and a mention of the brand, enhances the brand engagement in general. We find that adding words like "you" and "yours" do not have a clear relation with the likes but is harmful for the number of comments. Therefore, we advise not to use second pronouns in the influencer's caption.

6.3 Food

Food managers should manage their own Instagram page carefully since brand posts are considered more important for the food category in comparison with the other categories. If the aim is to enhance likes, it is recommended to avoid asking questions or using second pronouns like "you" or "your" in the caption. On the other hand, if the goal is to improve the comments, the opposite is true. The study of Schultz (2017) has found a positive impact for both likes and comments. The difference can be explained by the fact that the latter is a Facebook research incorporating only 700 observations. As with the fashion industry, videos are detrimental for the likes. This is consistent with the study of Vignisdóttir (2017). However, this study analyzed Instagram posts of the beauty sector. Strangely enough, we find no impact of using videos on engagement of beauty brands. This can be explained by the fact that we take a lot more data into account. Finally, if the manager wants to enhance the number of comments, we advise to include a person in the photo or video. This is confirmed by the general Instagram study of Bakhshi et al. (2014).

In the food sector, managers should focus on the engagement ratio and the activity of an influencer rather than on the absolute number of followers of the influencer. The importance of high engagement ratios and high activity levels are in line with the prior studies of Henderickx and De Wolf

(2019) and Brorsson and Plotnikova (2017). However, the fact that the number of followers does not have an impact on brand engagement is in contrast with the study of Jin and Phua (2014). The latter study has found that influencers with higher number of followers lead to purchase intentions in the food industry. The differences between the studies can be explained by the fact that the study of Jin and Phua (2014) was an experiment and focused on purchase intentions. These findings are particularly important for food managers because we notice that in Table 6, nowadays, the engagement ratios of food influencers are lower compared to fashion and beauty influencers. In addition, food managers invest the most in influencers with a large number of followers. This is money thrown away as the number of followers is less important to them.

We advise managers to ask the influencers to post pictures of themselves in order to increase likes. However, this is at the expense of the comments. The study of Bakhshi et al. (2014) finds a positive impact for both likes and comments. This study has conducted a general study examining Instagram photos while our study focuses on the impact of influencers using people in posts on brand engagement. This could explain the differences in outcome. It is further recommended to ask food influencers to be honest. This honesty increases the influencer credibility (Braatz, 2017) and therefore has a positive impact on the engagement of food brands. Finally, the interaction level of the influencer is an important factor to improve brand engagement. The influencers should include appropriate hashtags, use second pronouns such as "you" and "yours" but omit the use of actionable words in order to enhance likes. The opposite is true if the food managers want to increase comments. Again, the influencers should mention the brand page in their caption. This is a consistent trend applicable to fashion, beauty and food influencers for both comments and likes. This is in contrast with previous Facebook studies by Schultz (2017) and L. De Vries et al. (2012). These inconsistent findings can be clarified by the fact that these two studies use links to the brand website instead of the Facebook brand page.

7 Conclusion

This study sheds light on the underinvestigated area of influencer marketing in relation to brand engagement on Instagram. Previous research has often omitted the fact that influencers can have an additional impact on brand engagement. On an active influencer platform like Instagram, this statement is not plausible. Moreover, in contrast to prior literature, this study examines differences across the three most used influencer product categories: fashion, beauty and food. A quantitative analysis of 83 brands and 4162 influencers shows that both influencer characteristics and post influencer characteristics have an impact on brand engagement besides the traditional brand post characteristics. Moreover, there exist many differences in these characteristics across the various product categories. The findings of this study are of great importance to managers because the brand engagement can be optimized by formulating appropriate guidelines for influencers.

We find that the fashion industry should focus on finding influencers with less than 200K followers, whereas food influencers only require a high engagement ratio and activity level. As for the beauty sector, all these three influencer characteristics are important. The findings further show that fashion influencers should be positive while beauty and food influencers need to be honest to enhance brand engagement. In addition, interaction - defined as the use of second pronouns, questions, actionable verbs, hashtags and the mention of the brand in the influencer's caption - has a considerable impact on brand engagement. For the fashion industry, the interaction level is only important in the influencer's posts, while for beauty and food it is important for both the influencer's and brand's posts. The relationships of the various interaction aspects are different regarding the impact on likes and comments across the categories. The only interaction aspect that is positively related to both likes and comments for all categories is mentioning the brand in the influencer's caption.

8 Limitations and Further Research

Since this is the first comprehensive research making the bridge between influencer marketing and brand engagement on Instagram, there are some limitations. Hence, a lot of opportunities for further research are possible. This work can be seen as a starting point for future research combining influencer marketing and brand engagement on Instagram.

First, we need to mention that due to time limitations and the tremendous loads of data, not all the influencers of the brands have been implemented in the research. However, this should not have a big impact on the results if we assume that a random sample of the influencers is omitted. Further research could also examine whether our assumption of merging brands and influencers over a time period of three days should be changed. Moreover, further research could examine other ways to summarize the influencers into one row corresponding to one dependent variable. While this study works with absolute numbers and takes the sum, another option is to work with proportions.

An interesting option for future research is to also implement the use of Instagram stories. In contrast to posts on the feed, those stories are temporarily which means that they only stay on the user's profile for 24 hours. By means of the swipe up feature, the followers of an influencer can be directly taken to the URL included in the story (Nardello, 2017). According to the study of the company Klear (2018), around one third of the sponsored content is shared along stories instead of traditional posts. Hence, the overall impact of influencers on brand engagement is not considered in this study.

Another limitation in this research is that less focus is done on the image content itself. An option for future research is to add more focus on the different aspects of images since this is after all the core of Instagram. It could for example be examined which colors have the most impact or what kind of compositions are most favorable.

The primary focus of the analysis of this study is the posts characteristics omitting the content of comments. In further research, the comments may be examined in search of interesting features. For instance, the impact of an influencer reacting on questions of followers could be examined. In a qualitative study of Hellberg et al. (2015), it is mentioned that if a user leaves a comment, he/she would feel disappointed if the brand or influencer did not answer.

Besides the Instagram stories and a deeper analysis of the image and comments, there will probably be other important variables that are not taken into account in this research. It is obvious that the behaviour of a person in the sense of liking or commenting on a brand post, depends on more factors (e.g. social pressure, mood of the person or even the weather outside). Furthermore, there could be cultural differences and differences between male brands and female brands in influencer marketing. Moreover, further research can also investigate whether there are important interaction effects which are not examined here.

Finally, attention should be given to other platforms. Although Instagram is nowadays seen as the best platform regarding influencers, the music app TikTok has increased in importance and a lot of influencers are also active there (Influencer Marketing Hub, 2020a). This app is known for being able to achieve a huge reach for certain videos compared to other SNSs (Kumar, 2020).

Although our research has some limitations, we are convinced that this study is valuable to the literature combining influencer marketing and brand engagement.

References

Abidin, C. (2016). Visibility labour: Engaging with influencers' fashion brands and #ootd advertorial campaigns on instagram. *Media International Australia*, *161*(1), 86–100.

Alalwan, A. A., Rana, N. P., Dwivedi, Y. K., & Algharabat, R. (2017). Social media in marketing: A review and analysis of the existing literature. *Telematics and Informatics*, *34*(7), 1177–1190.

Anand, A., Dutta, S., & Mukherjee, P. (2019). Influencer marketing with fake followers. *IIM Bangalore Research Paper*(580).

Archer, K. J., & Kimes, R. V. (2008). Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis*, *52*(4), 2249–2260.

Bakhshi, S., Shamma, D. A., & Gilbert, E. (2014). Faces engage us: Photos with faces attract more likes and comments on instagram. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 965–974).

Berger, J., & Schwartz, E. M. (2011). What drives immediate and ongoing word of mouth? *Journal of Marketing Research*, *48*(5), 869–880.

Berk, R. A. (2008). *Statistical learning from a regression perspective* (Vol. 14). Springer.

Bodrunova, S. S., Blekanov, I. S., & Maksimov, A. (2016). Measuring influencers in twitter ad-hoc discussions: active users vs. internal networks in the discourse on biryuliovo bashings in 2013. In *2016 ieee artificial intelligence and natural language conference (ainl)* (p. 1-10).

Boerman, S. C. (2020). The effects of the standardized instagram disclosure for micro-and meso-influencers. *Computers in Human Behavior*, *103*, 199–207.

Bogaert, M., Ballings, M., & Van den Poel, D. (2018). Evaluating the importance of different communication types in romantic tie prediction on social media. *Annals of Operations Research*, *263*(1-2), 501–527.

Booth, N., & Matic, J. A. (2011). Mapping and leveraging influencers in social media to shape corporate brand perceptions. *Corporate Communications: An International Journal*, *16*(3), 184–191.

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication*, *13*(1), 210–230.

Braatz, L. (2017). *Influencer marketing on instagram, consumer responses towards promotional posts: the effects of message sidedness* (Unpublished master's thesis). University of Twente, the Netherlands.

Brandwatch. (2019). *50 incredible instagram stats.* Retrieved from `https://www.brandwatch.com/blog/instagram-stats/`

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Brorsson, A., & Plotnikova, V. (2017). *Choosing the right social media influencer.: A quasi-experiment to explore the impact of influencers' different characteristics.* (Unpublished master's thesis). Linnaeus University, Sweden.

Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, *36*(3), 4626–4636.

Buryan, M. (2016). *Why fashion brands are thriving on instagram.* Retrieved from `https://www.socialbakers.com/blog/2626-why-fashion-brands-are-thriving-on-instagram`

Chen, J. (2019). *How to regram on instagram.* Retrieved from `https://sproutsocial.com/insights/how-to-regram/`

Coyle, J. R., & Thorson, E. (2001). The effects of progressive levels of interactivity and vividness in web marketing sites. *Journal of advertising*, *30*(3), 65–77.

Cruz, R. E., Leonhardt, J. M., & Pezzuti, T. (2017). Second person pronouns enhance consumer involvement and brand attitude. *Journal of Interactive Marketing*, *39*, 104 - 116.

Cvijikj, I. P., & Michahelles, F. (2013). Online engagement factors on facebook brand pages. *Social network analysis and mining*, *3*(4), 843–861.

DataReportal. (2020). *Digital 2020 - global digital overview.* Retrieved from `https://datareportal.com/global-digital-overview`

Decrop, A. (2007). The influence of message format on the effectiveness of print advertisements for tourism destinations. *International Journal of Advertising*, *26*(4), 505-525.

Delrue, L., & Sinigaglia, N. (2017). What is the impact of instagram's social influencers on consumer attitude and purchase behaviour of lifestyle products of young belgian women? *Belgium: Louvain School.*

De Veirman, M., Cauberghe, V., & Hudders, L. (2017). Marketing through instagram influencers: the impact of number of followers and product divergence on brand attitude. *International Journal of Advertising*, *36*(5), 798–828.

De Veirman, M., & Hudders, L. (2020, Jan 2). Disclosing sponsored Instagram posts: the role of material connection with the brand and message-sidedness when disclosing covert advertising. *International Journal of Advertising*, *39*(1, SI), 94-130.

De Vries, E. L. (2019). When more likes is not better: the consequences of high and low likes-to-followers ratios for perceived account credibility and social media marketing effectiveness. *Marketing Letters*, *30*(3-4), 275–291.

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.

Dhanesh, G. S., & Duthler, G. (2019). Relationship management through social media influencers: Effects of followers' awareness of paid endorsement. *Public Relations Review*, *45*(3).

Dietterich, T. G. (2000, June). Ensemble methods in machine learning. In *International workshop on multiple classifier systems* (pp. 1–15).

Ding, K., Ma, K., & Wang, S. (2019). Intrinsic image popularity assessment. In *Proceedings of the 27th acm international conference on multimedia* (pp. 1979–1987).

Djafarova, E., & Rushworth, C. (2017). Exploring the credibility of online celebrities' instagram profiles in influencing the purchase decisions of young female users. *Computers in Human Behavior*, *68*, 1–7.

Doh, S.-J., & Hwang, J.-S. (2009). How consumers evaluate ewom (electronic word-of-mouth) messages. *CyberPsychology & Behavior*, *12*(2), 193–197.

Domingues Aguiar, T., van Reijmersdal, E., et al. (2018). *Influencer marketing* (Vol. 76). Amsterdam: Stichting Wetenschappelijk Onderzoek Commerciële Communicatie, SWOCC.

Dwivedi, Y. K., Kapoor, K. K., & Chen, H. (2015). Social media marketing and advertising. *The Marketing Review*, *15*(3), 289–309.

EASA. (2018). *Easa best practices and recommendations on influencer marketing.* Retrieved from `https://www.easa-alliance.org/about-ad-standards/ad-standards-and-trust`

Eisend, M. (2006, Jun). Two-sided advertising: A meta-analysis. *International journal of research in marketing*, *23*(2), 187-198.

Erkan, I. (2015). Electronic word of mouth on instagram: customers' engagements with brands in different sectors. *International Journal of Management, Accounting and Economics*, *2*(12), 1435–1444.

Evans, N. J., Phua, J., Lim, J., & Jun, H. (2017). Disclosing instagram influencer advertising: The effects of disclosure language on advertising recognition, attitudes, and behavioral intent. *Journal of Interactive Advertising*, *17*(2), 138–149.

Filieri, R. (2015). What makes online reviews helpful? a diagnosticity-adoption framework to explain informational and normative influences in e-wom. *Journal of Business Research*, *68*(6), 1261–1270.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.

FTC. (2017). *The ftc's endorsement guides: what people are asking.* Retrieved from `https://www.ftc.gov/tips-advice/business-center/guidance/ftcs-endorsement-guides-what-people-are-asking`

Ha, L., & James, E. L. (1998). Interactivity reexamined: A baseline analysis of early business web sites. *Journal of broadcasting & electronic media*, *42*(4), 457–474.

Harrison. (2017). *Top 10 trends that will transform digital marketing in 2017.* Retrieved from `https://www.forbes.com/sites/kateharrison/2017/01/09/top-10-trendsthat-will-transform-digital-marketing-in-2017/#7e6d507d3bf5`

Hellberg, M., et al. (2015). *Visual brand communication on instagram: A study on consumer engagement.* (Unpublished master's thesis). Hanken School of Economics, Finland.

Henderickx, A., & De Wolf, R. (2019). Social influencers, instapods en het nieuwsoverzicht-algoritme van instagram. *Tijdschrift voor Communicatiewetenschap*, *47*(3-4), 239–254.

Highfield, T., & Leaver, T. (2015). A methodology for mapping instagram hashtags. *First Monday*, *20*(1), 1–11.

Huang, J., Zhang, J., Li, Y., & Lv, Z. (2014). Business value of enterprise micro-blogging: Empirical study from weibo. com in sina. *Journal of Global Information Management (JGIM)*, *22*(3), 32–56.

Hughes, C., Swaminathan, V., & Brooks, G. (2019). Driving brand engagement through online social influencers: An empirical investigation of sponsored blogging campaigns. *Journal of Marketing*, *83*(5), 78–96.

Influencer Marketing Hub. (2019). *Instagram influencers' performance benchmarks and fraud report.* Retrieved from `https://influencermarketinghub.com/instagram-influencers-performance-benchmarks-and-fraud-report/`

Influencer Marketing Hub. (2020a). *11 tiktok influencer marketing case studies to inspire you.* Retrieved from `https://influencermarketinghub.com/tiktok-influencer-marketing-case-studies/`

Influencer Marketing Hub. (2020b). *The state of influencer marketing 2020: Benchmark report.* Retrieved from `https://influencermarketinghub.com/influencer-marketing-benchmark`

-report-2020/

Jin, S.-A. A., & Phua, J. (2014). Following celebrities' tweets about brands: The impact of twitter-based electronic word-of-mouth on consumers' source credibility perception, buying intention, and social identification with celebrities. *Journal of Advertising*, *43*(2), 181-195.

Johnson, B. K., Potocki, B., & Veldhuis, J. (2019). Is that my friend or an advert? the effectiveness of instagram native advertisements posing as social posts. *Journal of Computer-Mediated Communication*, *24*(3), 108–125.

Kassim, N., & Abdullah, N. A. (2010). The effect of perceived service quality dimensions on customer satisfaction, trust, and loyalty in e-commerce settings. *Asia pacific journal of marketing and logistics*.

Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? get serious! understanding the functional building blocks of social media. *Business horizons*, *54*(3), 241–251.

Kim, D.-H., Spiller, L., & Hettche, M. (2015). Analyzing media types and content orientations in facebook for global brands. *Journal of Research in Interactive Marketing*.

Klear. (2018). *Instagram stories trends.* Retrieved from `https://klear.com/Instagram-stories-report.pdf`

Korotina, A., & Jargalsaikhan, T. (2016). *Attitude towards instagram micro-celebrities and their influence on consumers' purchasing decisions.*

Kumar, G. (2020). Tik tok advertising policies in india: An overview. *Studies in Indian Place Names*, *40*(69), 297–305.

Kwok, L., & Yu, B. (2013). Spreading social media messages on facebook: An analysis of restaurant business-to-consumer communications. *Cornell Hospitality Quarterly*, *54*(1), 84–94.

Larivière, B., & Van den Poel, D. (2005). Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems with Applications*, *29*(2), 472–484.

Launchmetrics. (2019). *State of influencer marketing 2019.* Retrieved from `https://www.launchmetrics.com/landing/influencer-marketing-report-2019`

Ledbetter, A. M., & Redd, S. M. (2016). Celebrity credibility on social media: A conditional process analysis of online self-disclosure attitude as a moderator of posting frequency and parasocial interaction. *Western Journal of Communication*, *80*(5), 601-618.

Lee, E., Lee, J.-A., Moon, J. H., & Sung, Y. (2015). Pictures speak louder than words: Motivations for using instagram. *Cyberpsychology, behavior, and social networking*, *18*(9), 552–556.

Lei, S. S. I., Pratt, S., & Wang, D. (2017). Factors influencing customer engagement with branded content in the social network sites of integrated resorts. *Asia Pacific Journal of Tourism Research*, *22*(3), 316–328.

Liaw, A., Wiener, M., et al. (2002). Classification and regression by randomforest. *R news*, *2*(3), 18–22.

Lou, C., & Yuan, S. (2019). Influencer marketing: how message value and credibility affect consumer trust of branded content on social media. *Journal of Interactive Advertising*, *19*(1), 58–73.

Luarn, P., Lin, Y.-F., & Chiu, Y.-P. (2015). Influence of facebook brand-page posts on online engagement. *Online Information Review*.

Mathisen, A., & Stangeby, M. F. (2017). *Factors influencing advertising effectiveness and purchase intention on instagram* (Unpublished master's thesis). BI Norwegian Business School, Norway.

Mediakix. (2019). *Influencer marketing 2019 industry benchmarks.* Retrieved from `https://mediakix.com/influencer-marketing-resources/influencer-marketing-industry-statistics-survey-benchmarks/`

Meire, M., Hewett, K., Ballings, M., Kumar, V., & Van den Poel, D. (2019). The role of marketer-generated content in customer engagement marketing. *Journal of Marketing*, *83*(6), 21–42.

Merkle, E. C., & Shaffer, V. A. (2011). Binary recursive partitioning: Background, methods, and application to psychology. *British Journal of Mathematical and Statistical Psychology*, *64*(1), 161–181.

Müller, M. (2020). Coronavirus: A magic moment for social media? Retrieved from `https://www.dw.com/en/coronavirus-a-magic-moment-for-social-media/a-53180172`

Nardello, E. (2017). Best practices for producing stories on instagram. *Journal of Digital & Social Media Marketing*, *5*(4), 332–340.

Naumanen, E., et al. (2017). *Celebrities of instagram-what type of content influences followers' purchase intentions and engagement rate?* (Unpublished master's thesis). Aalto University School of Business, Finland.

Oliveira, M., Barbosa, R., & Sousa, A. (2020). The use of influencers in social media marketing. In *Marketing and smart technologies* (pp. 112–124). Springer.

Osborne Clarke. (2017). *Influencer marketing: a guide to regulation across europe.* Retrieved from `https://www.osborneclarke.com/insights/influencer-marketing`

-a-guide-to-regulation-across-europe/

Osman, M. (2017). *Here's why you should never buy instagram likes.* Retrieved from https://www.scrunch.com/blog/what-is-a-good-engagement-rate-on-instagram

Patel. (2016). *Top 8 marketing trends that will define 2017.* Retrieved from https://www.entrepreneur.com/article/284525

Pew Research Center. (2019). *How often americans are using social media.* Retrieved from https://www.pewresearch.org/internet/chart/how-often-americans-are-using-social-media/

Pittman, M., & Reich, B. (2016). Social media and loneliness: Why an instagram picture may be worth more than a thousand twitter words. *Computers in Human Behavior*, *62*, 155–167.

Richter, D., Riemer, K., & vom Brocke, J. (2011). Internet social networking: Stand der forschung und konsequenzen für enterprise 2.0. *Wirtschaftsinformatik*, *53*(2), 89–103.

Sabate, F., Berbegal-Mirabent, J., Cañabate, A., & Lebherz, P. R. (2014). Factors influencing popularity of branded content in facebook fan pages. *European Management Journal*, *32*(6), 1001–1011.

Salminen, J., Nielsen, L., Jung, S.-g., An, J., Kwak, H., & Jansen, B. J. (2018). "is more better?" impact of multiple photos on perception of persona profiles. In *Proceedings of the 2018 chi conference on human factors in computing systems* (pp. 1–13).

Schouten, A., Janssen, L., & Croes, E. (2019). Influencer advertising on instagram: Product-endorser fit and number of followers affect influencer and product evaluations via credibility and identification. Language, Communication and Cognition.

Schröder, J. (2019). *Influencer fraud on instagram* (Unpublished master's thesis). University of Mannheim, Germany.

Schultz, C. D. (2017). Proposing to your fans: Which brand post characteristics drive consumer engagement activities on social media brand pages? *Electronic Commerce Research and Applications*, *26*, 23–34.

Shen, B., & Bissell, K. (2013). Social media, social me: A content analysis of beauty companies' use of facebook in marketing and branding. *Journal of Promotion Management*, *19*(5), 629–651.

Solutions, N. C. (2016). *Sales effect study: Influencer marketing.* Retrieved from https://cdn2.hubspot.net/hubfs/1882019/TapInfluence/Resources/1009%20-%20Nielsen_Study_Case_Study.pdf

Statista. (2016). *Relevance of influencers on purchase decisions by product category in the* *u.s. 2017.* Retrieved from https://www.statista.com/forecasts/805883/relevance-of-influencers-on-purchase-decisions-by-product-category-in-the-us

Statista. (2019). *Average amount paid per influencer marketing post in the u.s.-2014-2019.* Retrieved from https://www.statista.com/statistics/1092754/cost-per-influencer-post-usa-gender/

Statista. (2020a). *Media consumption increase due to the coronavirus worldwide 2020, by country.* Retrieved from https://www.statista.com/statistics/1106766/media-consumption-growth-coronavirus-worldwide-by-country/

Statista. (2020b). *Most popular social networks worldwide as of january 2020, ranked by number of active users.* Retrieved from https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

Stephen, A. T., Dover, Y., Muchnik, L., & Goldenberg, J. (2017). Pump it out! the effect of transmitter activity on content propagation in social media. *Saïd Business School WP*, *1*.

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of communication*, *42*(4), 73–93.

Stouthuysen, K., Teunis, I., Reusen, E., & Slabbinck, H. (2018). Initial trust and intentions to buy: The effect of vendor-specific guarantees, customer reviews and the role of online shopping experience. *Electronic Commerce Research and Applications*, *27*, 23–38.

Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010). Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *2010 ieee second international conference on social computing* (pp. 177–184).

Swani, K., & Milne, G. R. (2017). Evaluating facebook brand content popularity for service versus goods offerings. *Journal of Business Research*, *79*, 123–133.

Tafesse, W. (2015). Content strategies and audience response on facebook brand pages. *Marketing Intelligence & Planning*.

Talaverna. (2015). *10 reasons why influencer marketing is the next big thing.* Retrieved from http://www.adweek.com/digital/10-reasons-why-influencermarketingis-the-next-big-thing/

Time. (2020). *Will the coronavirus ever go away? here's what one of the who's top experts thinks.* Retrieved from https://time.com/5805368/will-coronavirus-go-away-world-health-organization/

Tresna, L., & Wijaya, J. C. (2015). The impact of social media towards brand equity: An empirical study of mall x. *iBuss Management*, *3*(2).

Utz, S., Kerkhof, P., & Van Den Bos, J. (2012). Consumers rule: How consumer reviews influence perceived trustworthiness of online stores. *Electronic Commerce Research and Applications*, *11*(1), 49–58.

Van den Abeele, E. (2019). *De influencer aan het woord: Kwalitatief onderzoek naar de perceptie van influencers over hun rol in influencer marketing* (Unpublished master's thesis). University of Ghent, Belgium.

Vesanen, J. (2007). What is personalization? a conceptual framework. *European Journal of Marketing*.

Vignisdóttir, A. (2017). *Customer engagement on instagram brand pages in the make-up cosmetics industry* (Unpublished doctoral dissertation). Reykjavik University, Iceland.

Wee, C.-H., Lim, S., & Lwin, M. (1995, 12). Word-of-mouth communication in singapore:with focus on effects of message-sidedness, source and user-type. *Asia Pacific Journal of Marketing and Logistics*, *7*, 5-36. doi: 10.1108/eb010260

West, C. (2019). *How to create successful social media contests that convert.* Retrieved from `https://sproutsocial.com/insights/social-media-contests/`

Williamson, D. A. (2019). *What to know about spending, stories, fraud and microinfluencers.* Retrieved from `https://iabperu.com/wp-content/uploads/2019/08/Global_Influencer_Marketing_2019_eMarketer.pdf`

Woods, S. (2016). *# sponsored: The emergence of influencer marketing* (Unpublished master's thesis). University of Tennessee, Knoxville, United States of America.

Yew, R. L. H., Suhaidi, S. B., Seewoochurn, P., & Sevamalai, V. K. (2018). Social network influencers' engagement rate algorithm using instagram data. In *2018 fourth international conference on advances in computing, communication & automation (icacca)* (pp. 1–8).

Zeren, D., & Gökdağlı, N. (2020). Influencer versus celebrity endorser performance on instagram. In A. Kavoura, E. Kefallonitis, & P. Theodoridis (Eds.), *Strategic innovative marketing and tourism* (pp. 695–704). Cham: Springer International Publishing.

Zhang, Y., Lin, Y., & Goh, K. H. (2018a). Impact of online influencer endorsement on product sales: Quantifying value of online influencer. In *Pacis* (p. 201).

Zhang, Z., Chen, T., Zhou, Z., Li, J., & Luo, J. (2018b). How to become instagram famous: Post popularity prediction with dual-attention. In *2018 ieee international conference on big data (big data)* (pp. 2383–2392).

Appendices

A Brands with their followers on Instagram (retrieved on 24 April 2020)

A.1 Fashion Brands

Instagram Brandname	Number of Followers
batashoes	104.000
crocs	780.000
emperorapparel	14.600
farfetch	2.000.000
hunkemoller	647.000
intimissimiofficial	2.900.000
jcrewmens	253.000
jjillstyle	82.300
levis_nl	67.500
moncler	3.100.000
primark.man	175.000
soliverfashion	141.000
sylviap	75.000
teva	501.000
thefryecompany	215.000
thetiebar	186.000
underarmourdach	94.600

A.2 Beauty Brands

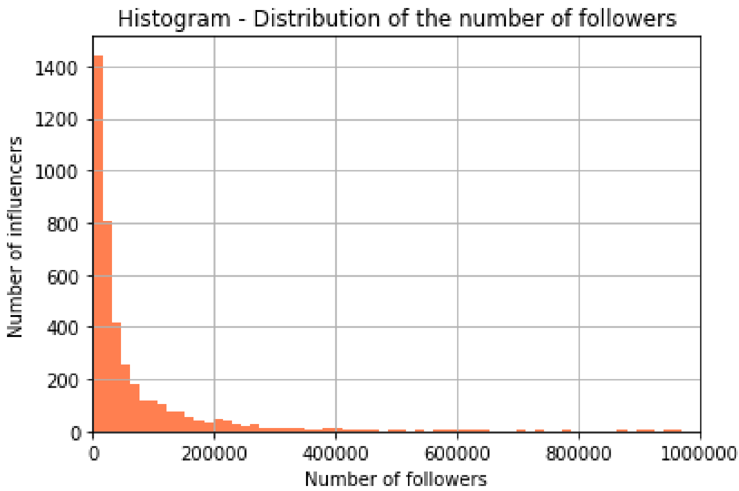
Instagram Brandname	**Number of Followers**
carboncocoau	760.000
caudalieus	44.300
cliniqueperu	58.400
dollarshaveclub	240.000
evahairnyc	202.000
garnieruk	157.000
givenchybeauty	1.700.000
guhl	6.000
headandshouldersusa	18.600
hismileteeth	1.300.000
johnsonsbabyuk	8.300
kiss_aesthetics	226.000
lovebeautyandplanet	140.000
myclarinsofficial	38.500
nivea_be	13.700
placentaplus	24.100
ranahairextensions	3.800
revlon_uk	255.000
sanctuaryspa	40.200
swisseusa	5.000
theritualofholi	48.700
vaselineuk	17.800

A.3 Food Brands

Instagram Brandname	**Number of Followers**
acqua_rocchetta	3.700
acquapannausa	2.800
almondbreeze	30.400
aperolspriz_be	78.600
arnottsbiscuits	12.100
belvederevodkaitalia	42.300
benjerrysnl	21.200
bibigousa	14.300
blacktaplv	43.900
bluepyrenees	2.300
cadburyaust	23.900
cheerios	58.100
cocacolabelgium	30.800
communitycoffee	18.800
eatalyflatiron	186.000
ferrerorocherfr	21.100
greenandblacks	36.100
hummkombucha	35.300
indelight	28.600
juicyjuiceusa	5.900
kevitadrinks	54.600
kyrodistillery	25.300
lennyandlarrys	186.000
lipton.nl	1.500
maltesers	41.300
martinmillersgin	29.200
mccainuk	3.800
moavodka	1.400
moetchandon	634.000
naluenergydrink	5.900
nespresso.belux	11.900
nutterpuffs	5.000
ozarkaspringwtr	2.800
perrierusa	15.600
polandspringorigin	2.200
popchips	26.300
redbullbe	43.200
riondoprosecco	2.400
schweppes	11.400
seedsofchange	5.700
starburst	131.000
theskinnypop	24.400
v8juice	9.600

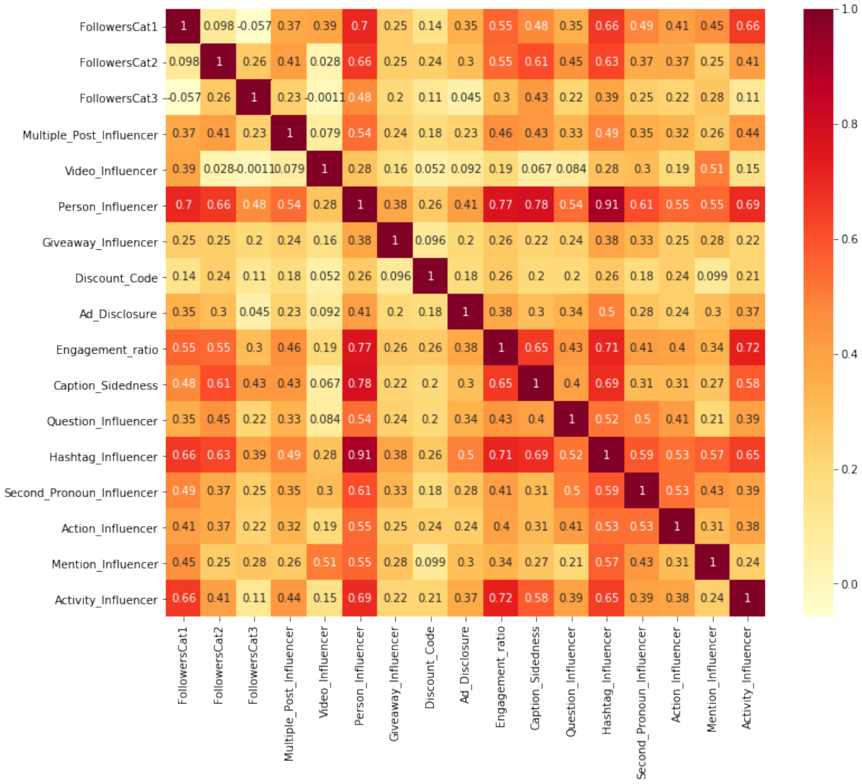
B Histogram Number of Followers

The x-axis is limited to 1.000.000 instead of the maximum number of followers because only 73 influencers have more followers. This way, a more readable table could be provided.

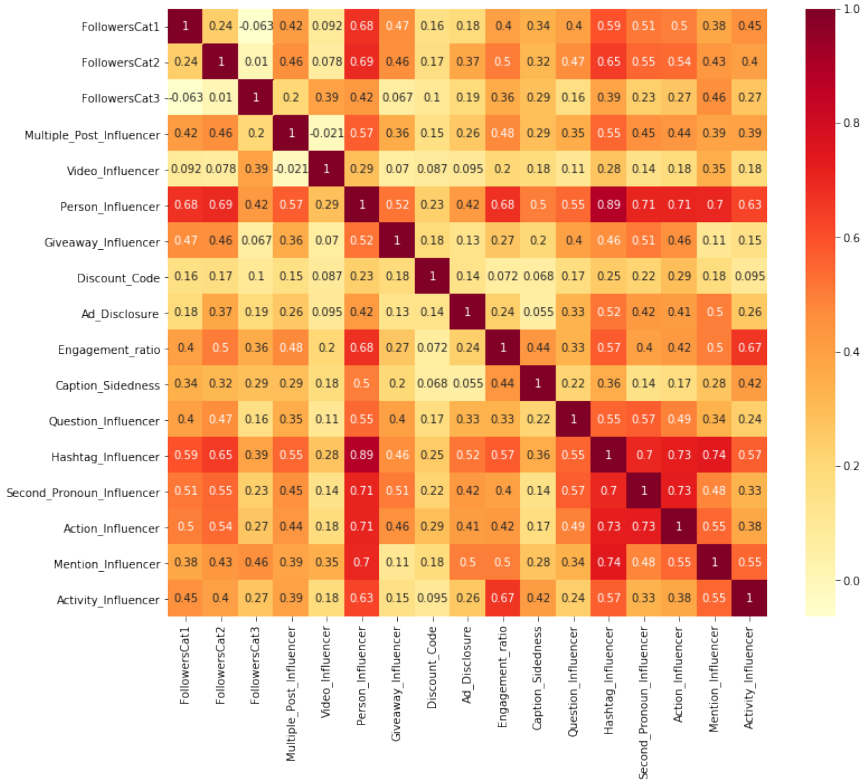


C Correlation Matrix

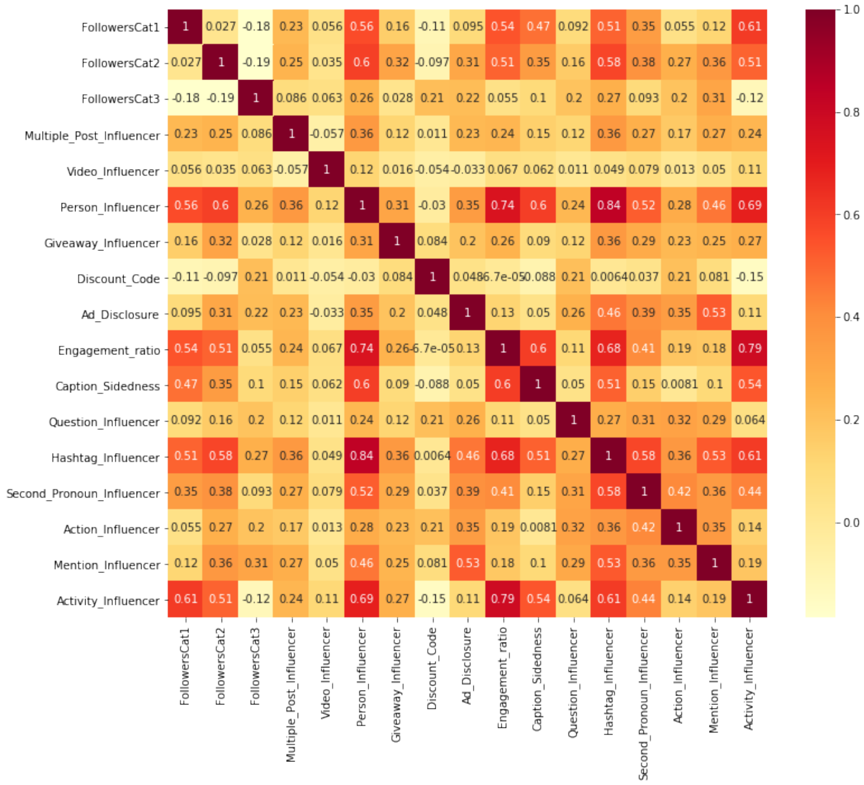
C.1 Fashion



C.2 Beauty



C.3 Food



D Model Cross Validation Evaluation Results

D.1 Fashion: 5X2 Cross Validation AUC per model

		Likes				Comments			
Time	Fold	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
1	1	0,6640	0,6821	0,6734	0,6829	0,6042	0,6011	0,6055	0,6004
1	2	0,6723	0,6887	0,6803	0,6796	0,6084	0,6072	0,6065	0,6057
2	1	0,6664	0,6935	0,6800	0,6804	0,5950	0,6020	0,6079	0,5926
2	2	0,6703	0,6868	0,6775	0,6791	0,5994	0,6077	0,6007	0,5967
3	1	0,6682	0,6856	0,6819	0,6778	0,6047	0,5942	0,6068	0,6055
3	2	0,6649	0,6893	0,6815	0,6847	0,5940	0,6030	0,6101	0,6016
4	1	0,6685	0,6756	0,6787	0,6746	0,6053	0,6028	0,6020	0,6060
4	2	0,6633	0,6785	0,6720	0,6747	0,6012	0,6084	0,6035	0,6096
5	1	0,6599	0,6739	0,6841	0,6735	0,5949	0,6074	0,6100	0,6044
5	1	0,6641	0,6817	0,6906	0,6754	0,6009	0,6130	0,6108	0,6047
Median		**0,6656**	**0,6838**	**0,6801**	**0,6784**	**0,6011**	**0,6051**	**0,6066**	**0,6046**

D.2 Beauty: 5X2 Cross Validation AUC per model

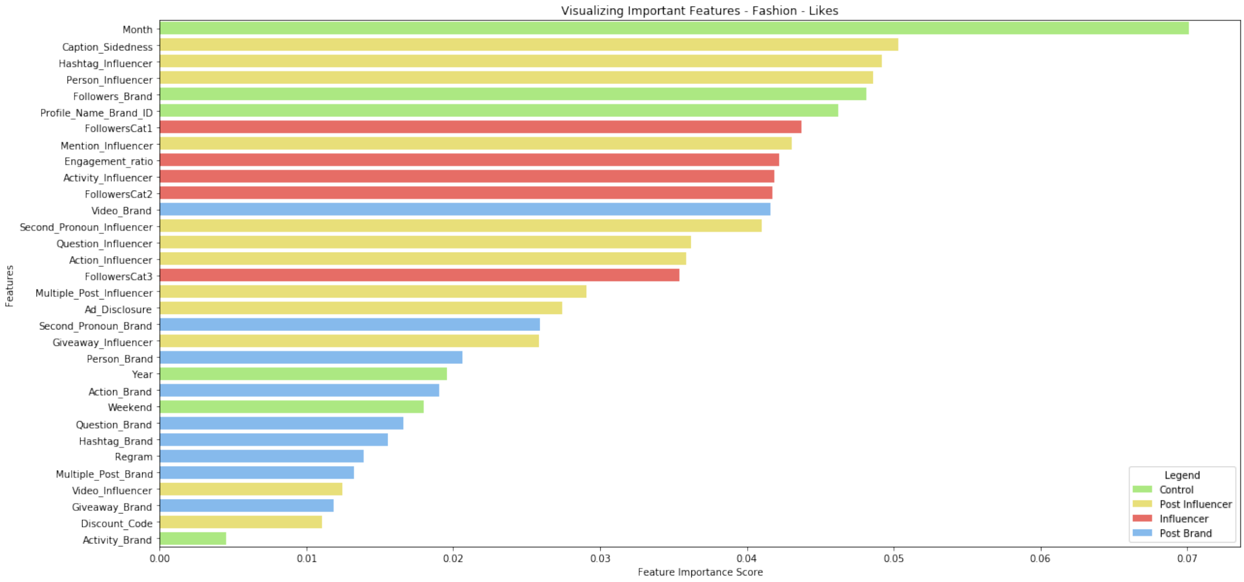
		Likes				Comments			
Time	Fold	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
1	1	0,6676	0,6473	0,6858	0,6548	0,6295	0,6064	0,6070	0,6192
1	2	0,6695	0,6394	0,6885	0,6447	0,6401	0,6172	0,6012	0,6111
2	1	0,6806	0,6478	0,6685	0,6576	0,6151	0,6124	0,6148	0,6101
2	2	0,6786	0,6611	0,6752	0,6486	0,6233	0,6021	0,6185	0,6089
3	1	0,6667	0,6394	0,6848	0,6568	0,6164	0,6340	0,6193	0,6011
3	2	0,6784	0,6671	0,6768	0,6508	0,6298	0,6359	0,6194	0,6062
4	1	0,6947	0,6563	0,6680	0,6618	0,5972	0,6188	0,6033	0,6081
4	2	0,6903	0,6493	0,6756	0,6425	0,6302	0,5986	0,6166	0,6103
5	1	0,6623	0,6624	0,6777	0,6532	0,6375	0,6118	0,6185	0,6311
5	1	0,6701	0,6521	0,6821	0,6485	0,6221	0,6118	0,6243	0,6259
Median		**0,6743**	**0,6507**	**0,6773**	**0.6520**	**0,6264**	**0,6121**	**0,6175**	**0,6102**

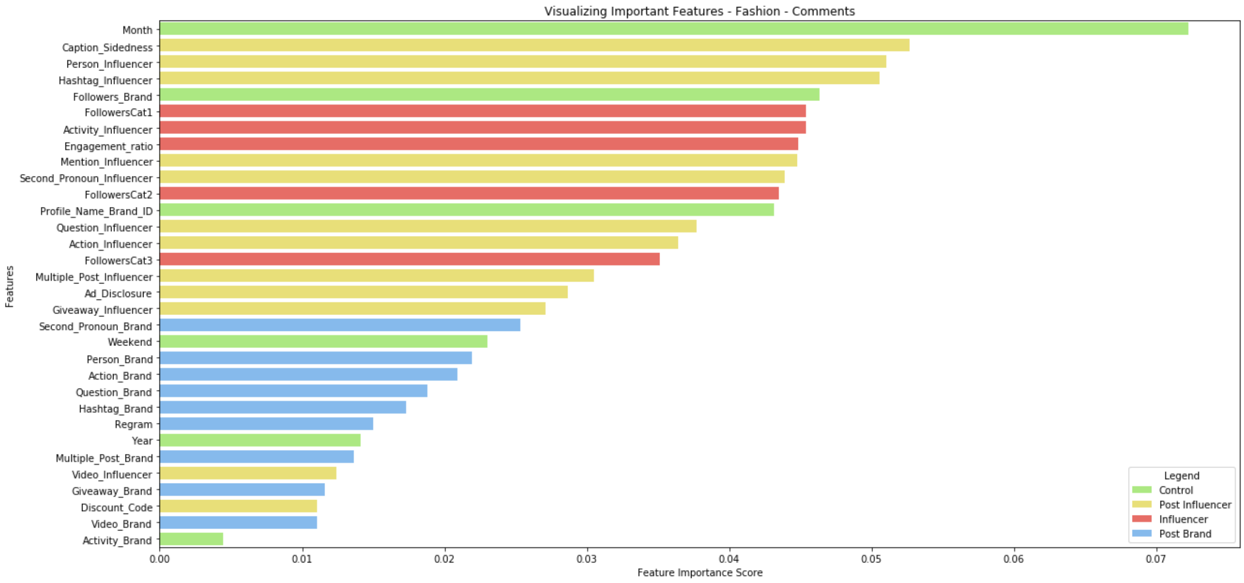
D.3 Food: 5X2 Cross Validation AUC per model

		Likes				Comments			
Time	Fold	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
1	1	0,6126	0,6242	0,5907	0,6303	0,5429	0,5395	0,5574	0,5586
1	2	0,6038	0,6122	0,5938	0,6156	0,5569	0,5536	0,5586	0,5422
2	1	0,5919	0,6385	0,6097	0,5947	0,5353	0,5313	0,5495	0,5522
2	2	0,6072	0,6277	0,6276	0,6079	0,5420	0,5484	0,5360	0,5454
3	1	0,5853	0,6492	0,6329	0,6163	0,5703	0,5814	0,5376	0,5534
3	2	0,6017	0,6345	0,6239	0,6281	0,5789	0,5683	0,5364	0,5594
4	1	0,6153	0,6367	0,6225	0,6051	0,5482	0,5613	0,5408	0,5296
4	2	0,6108	0,6297	0,6028	0,6178	0,5525	0,5481	0,5670	0,5284
5	1	0,5994	0,6138	0,6269	0,6140	0,5483	0,5455	0,5425	0,5498
5	1	0,6084	0,6138	0,5993	0,6088	0,5491	0,5561	0,5483	0,5544
Median		**0,6055**	**0,6287**	**0,6161**	**0,6148**	**0,5487**	**0,5510**	**0,5454**	**0,5510**

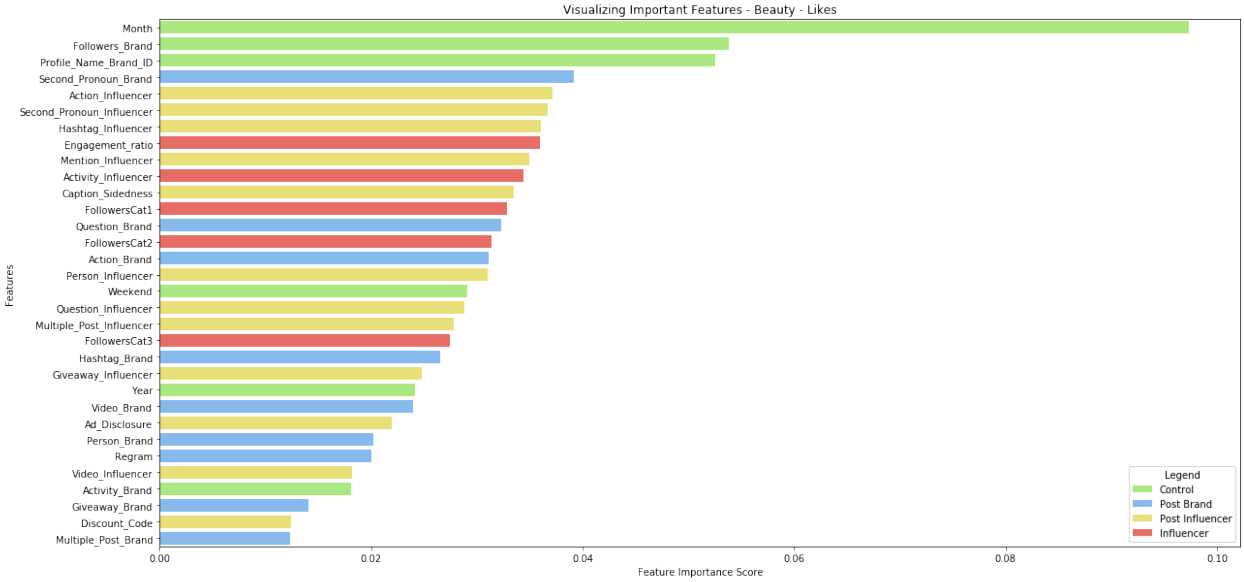
E Variable Importance Plots with Control Variables

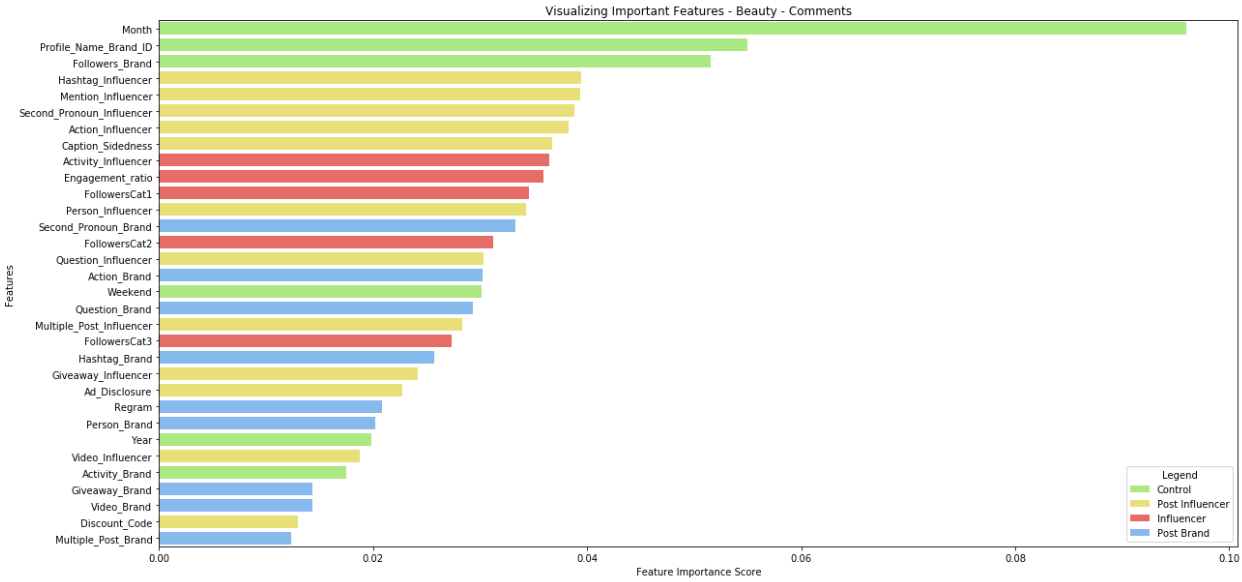
E.1 Fashion



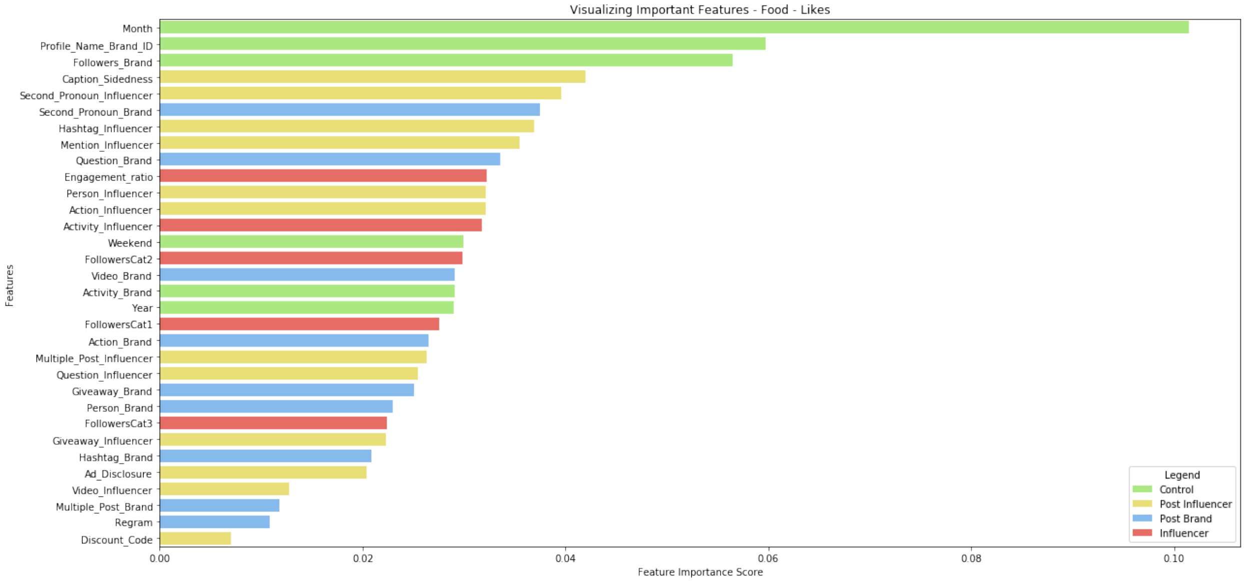


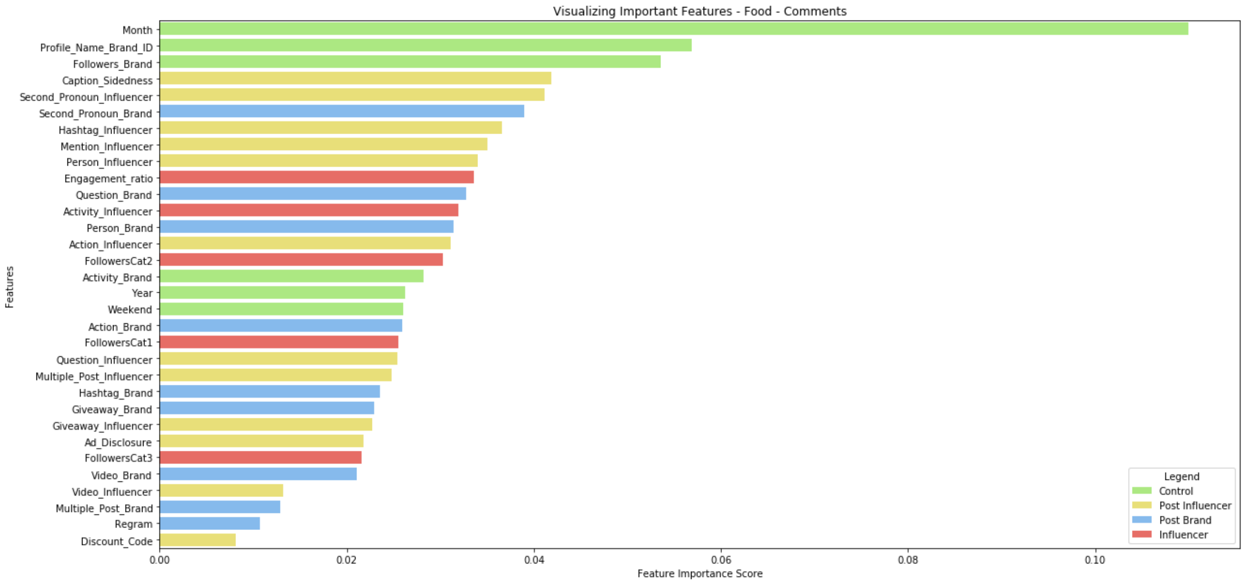
E.2 Beauty


Visualizing Important Features - Beauty - Likes
Month
Followers_Brand
Profile_Name_Brand_ID
Second_Pronoun_Brand
Action_Influencer
Second_Pronoun_Influencer
Hashtag_Influencer
Engagement_ratio
Mention_Influencer
Activity_Influencer
Caption_Sidedness
FollowersCat1
Question_Brand
FollowersCat2
Action_Brand
Person_Influencer
Weekend
Question_Influencer
Multiple_Post_Influencer
FollowersCat3
Hashtag_Brand
Giveaway_Influencer
Year
Video_Brand
Ad_Disclosure
Person_Brand
Regram
Video_Influencer
Activity_Brand
Giveaway_Brand
Discount_Code
Multiple_Post_Brand
Features
0.00
0.02
0.04
0.06
0.08
0.10
Feature Importance Score
Legend
Control
Post Brand
Post Influencer
Influencer


Visualizing Important Features - Beauty - Comments
Month
Profile_Name_Brand_ID
Followers_Brand
Hashtag_Influencer
Mention_Influencer
Second_Pronoun_Influencer
Action_Influencer
Caption_Sidedness
Activity_Influencer
Engagement_ratio
FollowersCat1
Person_Influencer
Second_Pronoun_Brand
FollowersCat2
Question_Influencer
Action_Brand
Weekend
Question_Brand
Multiple_Post_Influencer
FollowersCat3
Hashtag_Brand
Giveaway_Influencer
Ad_Disclosure
Regram
Person_Brand
Year
Video_Influencer
Activity_Brand
Giveaway_Brand
Video_Brand
Discount_Code
Multiple_Post_Brand
Features
0.00
0.02
0.04
0.06
0.08
0.10
Feature Importance Score
Legend
Control
Post Influencer
Influencer
Post Brand

E.3 Food


Visualizing Important Features - Food - Likes
Month
Profile_Name_Brand_ID
Followers_Brand
Caption_Sidedness
Second_Pronoun_Influencer
Second_Pronoun_Brand
Hashtag_Influencer
Mention_Influencer
Question_Brand
Engagement_ratio
Person_Influencer
Action_Influencer
Activity_Influencer
Weekend
FollowersCat2
Video_Brand
Activity_Brand
Year
FollowersCat1
Action_Brand
Multiple_Post_Influencer
Question_Influencer
Giveaway_Brand
Person_Brand
FollowersCat3
Giveaway_Influencer
Hashtag_Brand
Ad_Disclosure
Video_Influencer
Multiple_Post_Brand
Regram
Discount_Code
Features
Legend
Control
Post Influencer
Post Brand
Influencer
0.00
0.02
0.04
0.06
0.08
0.10
Feature Importance Score


Visualizing Important Features - Food - Comments
Month
Profile_Name_Brand_ID
Followers_Brand
Caption_Sidedness
Second_Pronoun_Influencer
Second_Pronoun_Brand
Hashtag_Influencer
Mention_Influencer
Person_Influencer
Engagement_ratio
Question_Brand
Activity_Influencer
Person_Brand
Action_Influencer
FollowersCat2
Activity_Brand
Year
Weekend
Action_Brand
FollowersCat1
Question_Influencer
Multiple_Post_Influencer
Hashtag_Brand
Giveaway_Brand
Giveaway_Influencer
Ad_Disclosure
FollowersCat3
Video_Brand
Video_Influencer
Multiple_Post_Brand
Regram
Discount_Code
Features
Legend
Control
Post Influencer
Post Brand
Influencer
0.00
0.02
0.04
0.06
0.08
0.10
Feature Importance Score

F Partial Dependence Plots - Model 4

<u>*Legend colors:*</u>

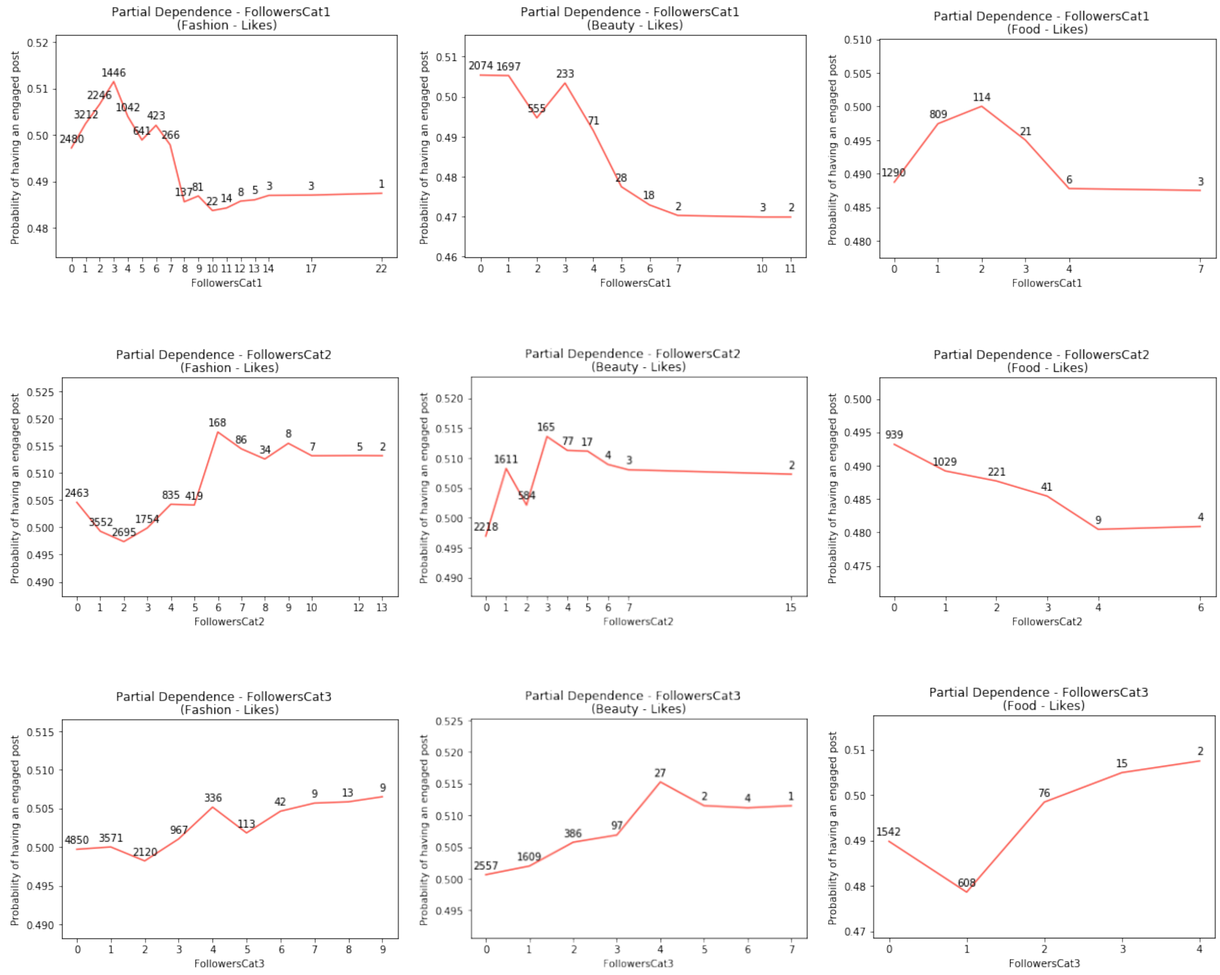
Red = Influencer Characteristics

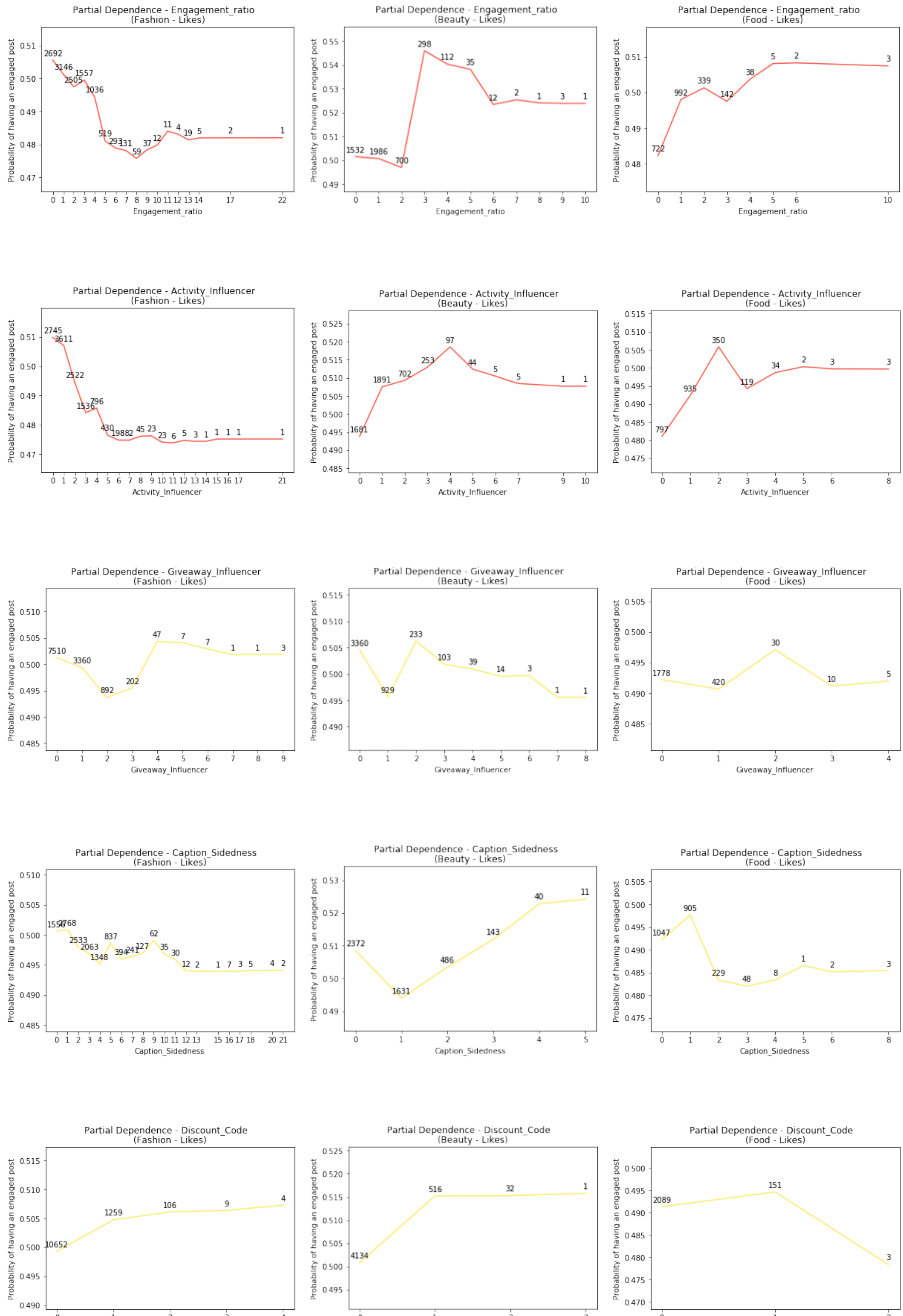
Yellow = Post Influencer Characteristics

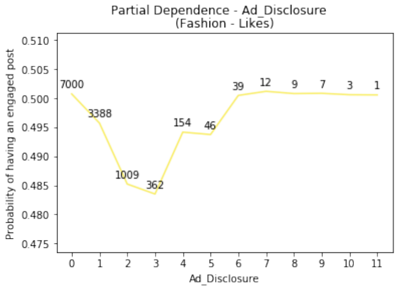
Blue = Post Brands Characteristics

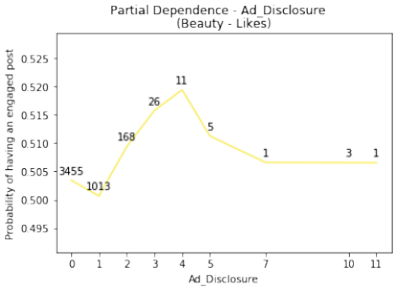
Green = Control Variables

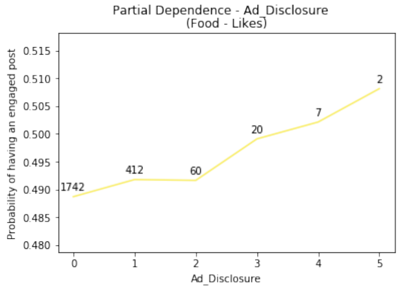
F.1 Likes

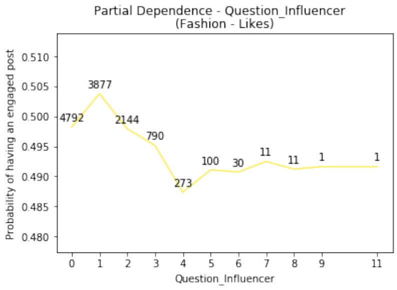


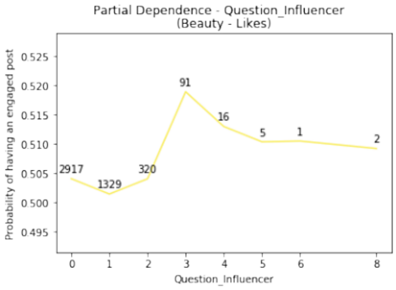

Partial Dependence - Engagement_ratio (Fashion - Likes)
Partial Dependence - Engagement_ratio (Beauty - Likes)
Partial Dependence - Engagement_ratio (Food - Likes)
Engagement_ratio
Partial Dependence - Activity_Influencer (Fashion - Likes)
Partial Dependence - Activity_Influencer (Beauty - Likes)
Partial Dependence - Activity_Influencer (Food - Likes)
Activity_Influencer
Partial Dependence - Giveaway_Influencer (Fashion - Likes)
Partial Dependence - Giveaway_Influencer (Beauty - Likes)
Partial Dependence - Giveaway_Influencer (Food - Likes)
Giveaway_Influencer
Partial Dependence - Caption_Sidedness (Fashion - Likes)
Partial Dependence - Caption_Sidedness (Beauty - Likes)
Partial Dependence - Caption_Sidedness (Food - Likes)
Caption_Sidedness
Partial Dependence - Discount_Code (Fashion - Likes)
Partial Dependence - Discount_Code (Beauty - Likes)
Partial Dependence - Discount_Code (Food - Likes)
Discount_Code
Probability of having an engaged post

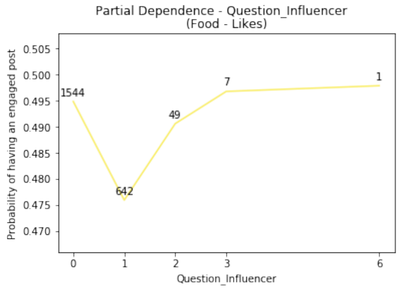

Partial Dependence - Ad_Disclosure
(Fashion - Likes)
Probability of having an engaged post
Ad_Disclosure

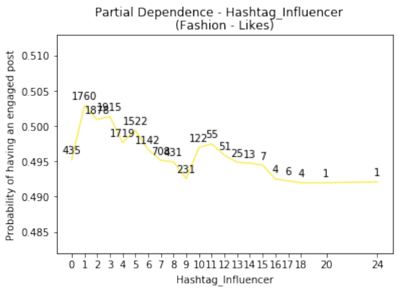

Partial Dependence - Ad_Disclosure
(Beauty - Likes)
Probability of having an engaged post
Ad_Disclosure

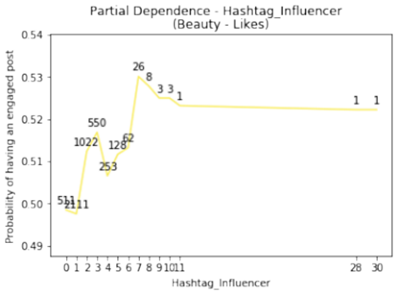

Partial Dependence - Ad_Disclosure
(Food - Likes)
Probability of having an engaged post
Ad_Disclosure

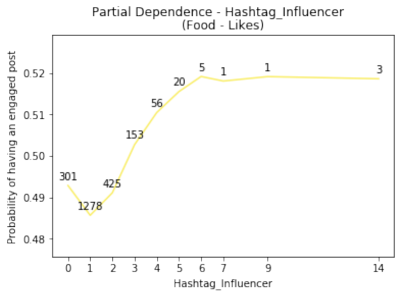

Partial Dependence - Question_Influencer
(Fashion - Likes)
Probability of having an engaged post
Question_Influencer

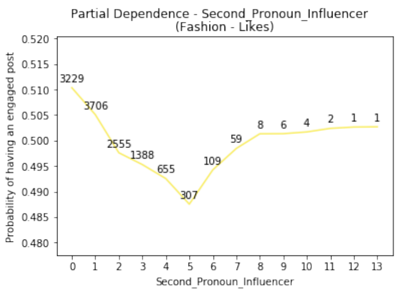

Partial Dependence - Question_Influencer
(Beauty - Likes)
Probability of having an engaged post
Question_Influencer

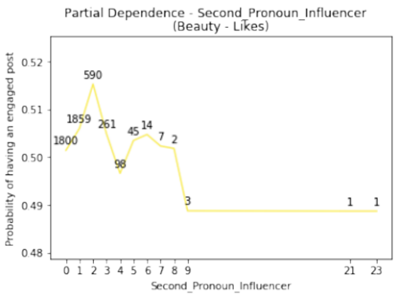

Partial Dependence - Question_Influencer
(Food - Likes)
Probability of having an engaged post
Question_Influencer

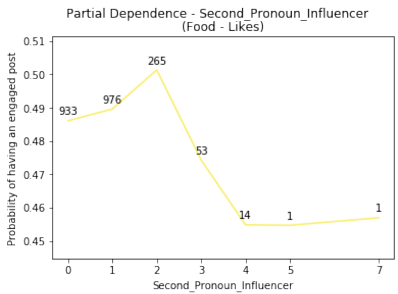

Partial Dependence - Hashtag_Influencer
(Fashion - Likes)
Probability of having an engaged post
Hashtag_Influencer

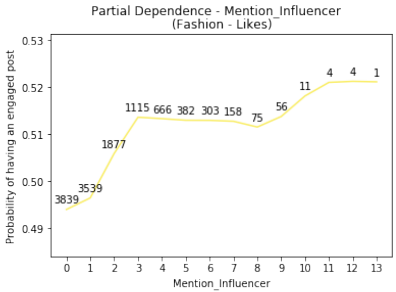

Partial Dependence - Hashtag_Influencer
(Beauty - Likes)
Probability of having an engaged post
Hashtag_Influencer

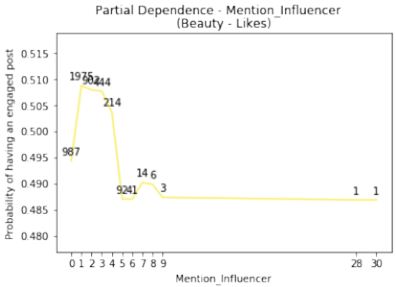

Partial Dependence - Hashtag_Influencer
(Food - Likes)
Probability of having an engaged post
Hashtag_Influencer

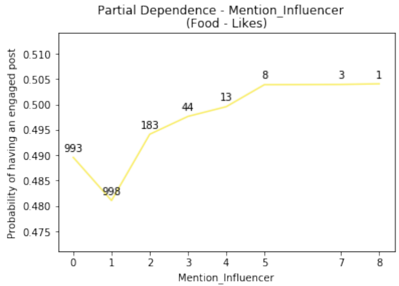

Partial Dependence - Second_Pronoun_Influencer
(Fashion - Likes)
Probability of having an engaged post
Second_Pronoun_Influencer

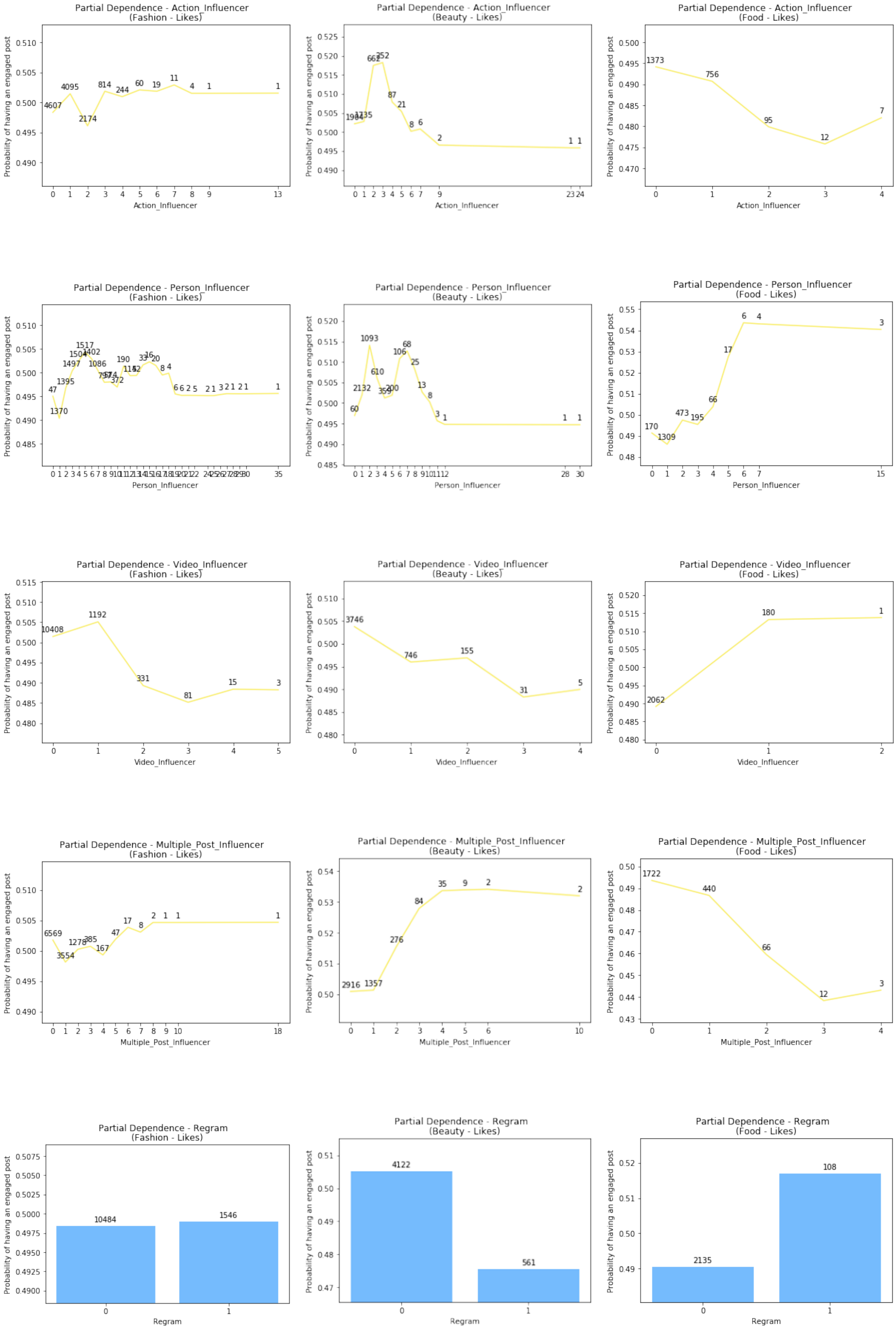

Partial Dependence - Second_Pronoun_Influencer
(Beauty - Likes)
Probability of having an engaged post
Second_Pronoun_Influencer

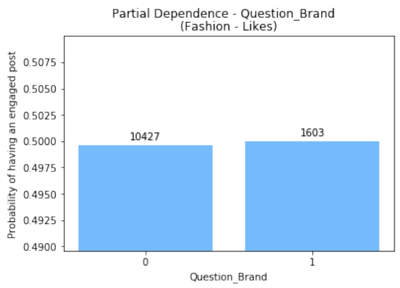

Partial Dependence - Second_Pronoun_Influencer
(Food - Likes)
Probability of having an engaged post
Second_Pronoun_Influencer

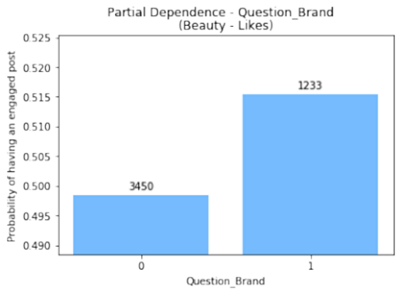

Partial Dependence - Mention_Influencer
(Fashion - Likes)
Probability of having an engaged post
Mention_Influencer

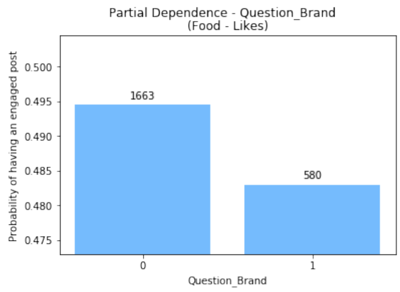

Partial Dependence - Mention_Influencer
(Beauty - Likes)
Probability of having an engaged post
Mention_Influencer

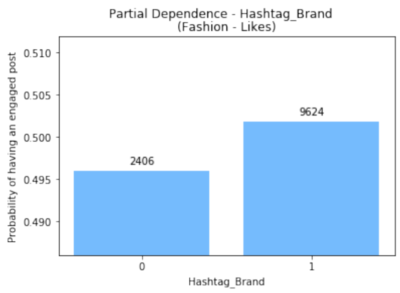

Partial Dependence - Mention_Influencer
(Food - Likes)
Probability of having an engaged post
Mention_Influencer

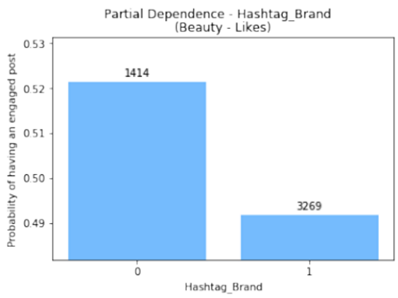

Partial Dependence - Action_Influencer (Fashion - Likes)
Partial Dependence - Action_Influencer (Beauty - Likes)
Partial Dependence - Action_Influencer (Food - Likes)
Partial Dependence - Person_Influencer (Fashion - Likes)
Partial Dependence - Person_Influencer (Beauty - Likes)
Partial Dependence - Person_Influencer (Food - Likes)
Partial Dependence - Video_Influencer (Fashion - Likes)
Partial Dependence - Video_Influencer (Beauty - Likes)
Partial Dependence - Video_Influencer (Food - Likes)
Partial Dependence - Multiple_Post_Influencer (Fashion - Likes)
Partial Dependence - Multiple_Post_Influencer (Beauty - Likes)
Partial Dependence - Multiple_Post_Influencer (Food - Likes)
Partial Dependence - Regram (Fashion - Likes)
Partial Dependence - Regram (Beauty - Likes)
Partial Dependence - Regram (Food - Likes)
Probability of having an engaged post
Action_Influencer
Person_Influencer
Video_Influencer
Multiple_Post_Influencer
Regram

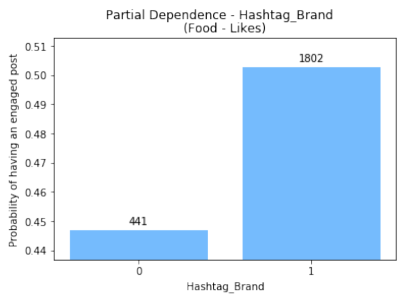

Partial Dependence - Question_Brand
(Fashion - Likes)
Probability of having an engaged post
10427
1603
Question_Brand

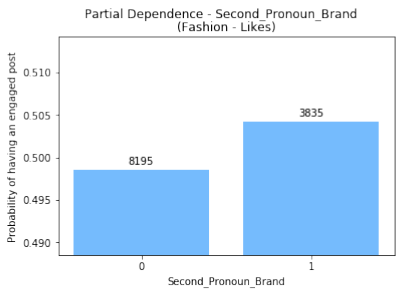

Partial Dependence - Question_Brand
(Beauty - Likes)
Probability of having an engaged post
3450
1233
Question_Brand

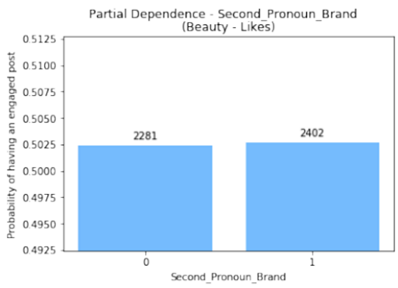

Partial Dependence - Question_Brand
(Food - Likes)
Probability of having an engaged post
1663
580
Question_Brand

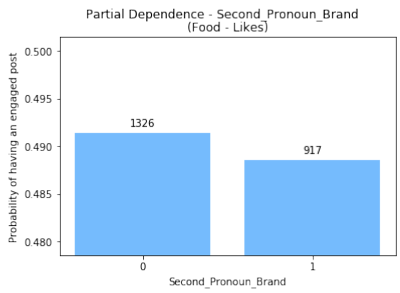

Partial Dependence - Hashtag_Brand
(Fashion - Likes)
Probability of having an engaged post
2406
9624
Hashtag_Brand

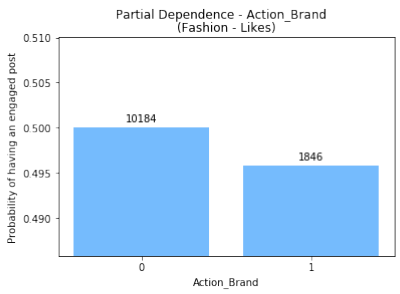

Partial Dependence - Hashtag_Brand
(Beauty - Likes)
Probability of having an engaged post
1414
3269
Hashtag_Brand

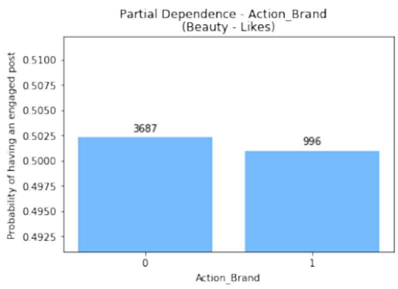

Partial Dependence - Hashtag_Brand
(Food - Likes)
Probability of having an engaged post
441
1802
Hashtag_Brand

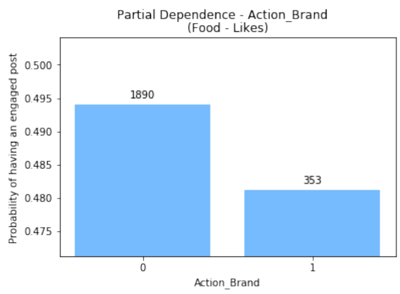

Partial Dependence - Second_Pronoun_Brand
(Fashion - Likes)
Probability of having an engaged post
8195
3835
Second_Pronoun_Brand

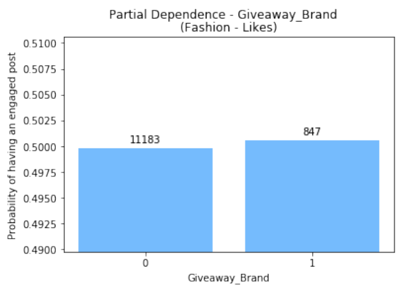

Partial Dependence - Second_Pronoun_Brand
(Beauty - Likes)
Probability of having an engaged post
2281
2402
Second_Pronoun_Brand

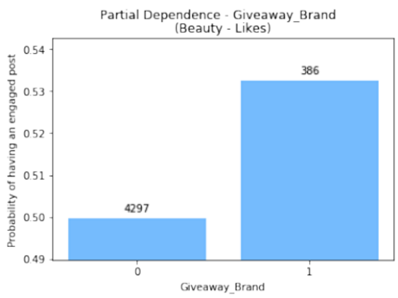

Partial Dependence - Second_Pronoun_Brand
(Food - Likes)
Probability of having an engaged post
1326
917
Second_Pronoun_Brand

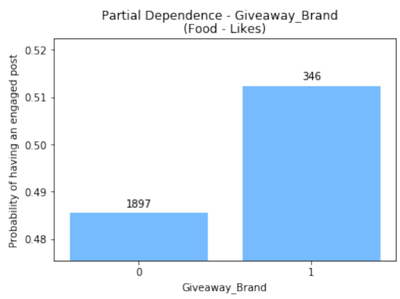

Partial Dependence - Action_Brand
(Fashion - Likes)
Probability of having an engaged post
10184
1846
Action_Brand

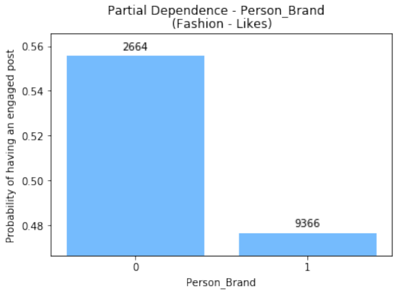
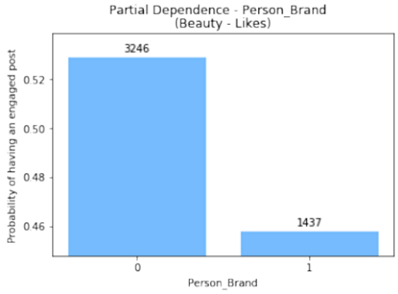
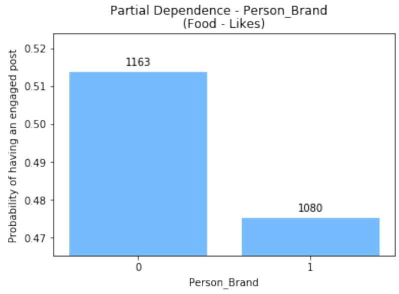
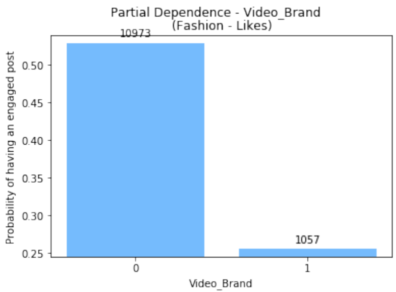
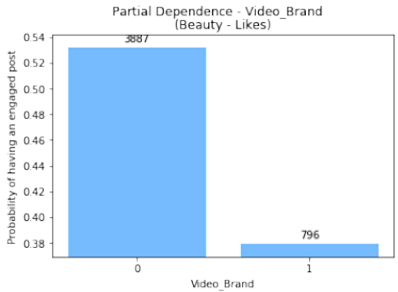
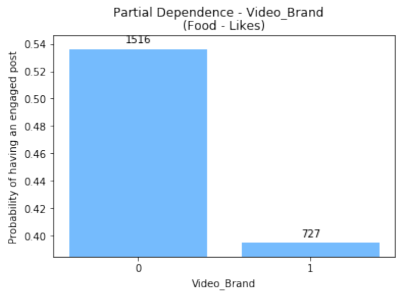
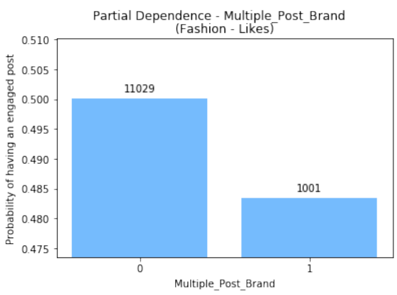
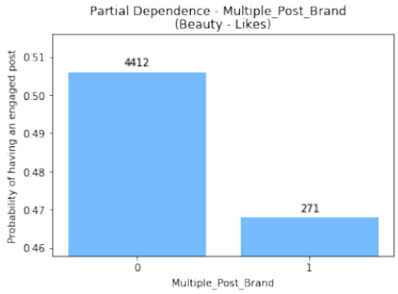
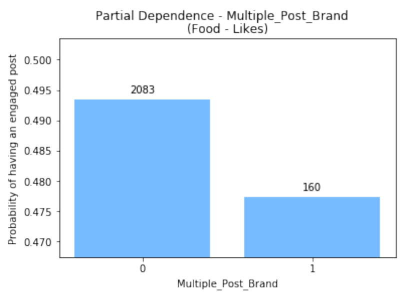
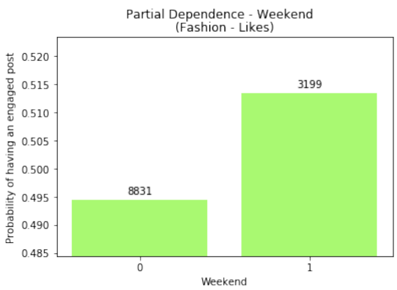
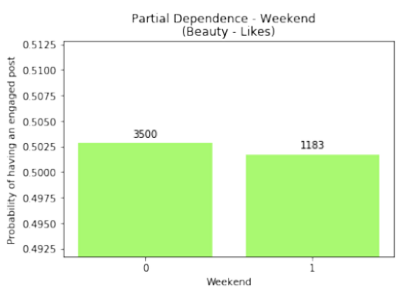
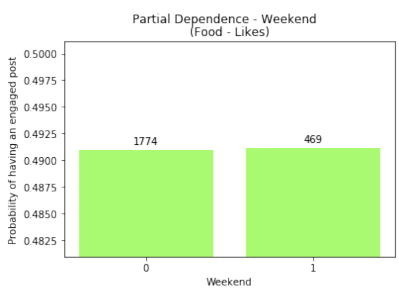
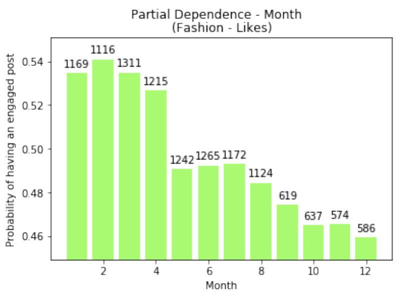
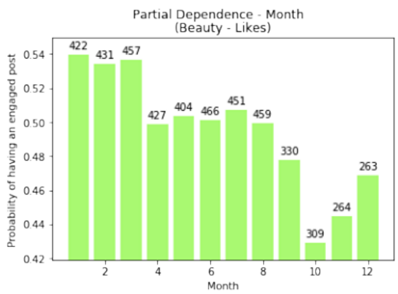
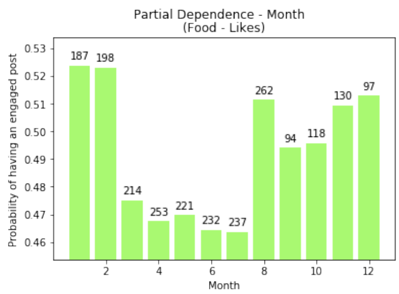

Partial Dependence - Action_Brand
(Beauty - Likes)
Probability of having an engaged post
3687
996
Action_Brand

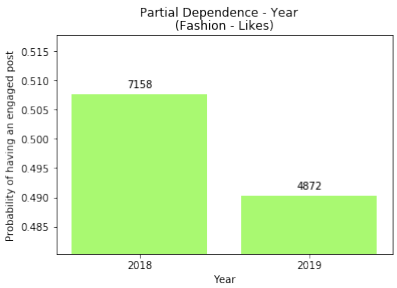

Partial Dependence - Action_Brand
(Food - Likes)
Probability of having an engaged post
1890
353
Action_Brand

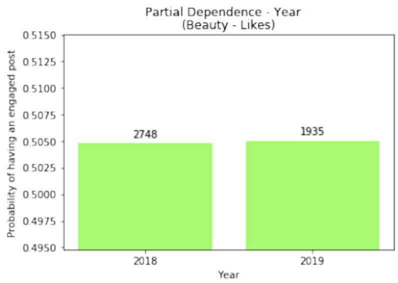

Partial Dependence - Giveaway_Brand
(Fashion - Likes)
Probability of having an engaged post
11183
847
Giveaway_Brand

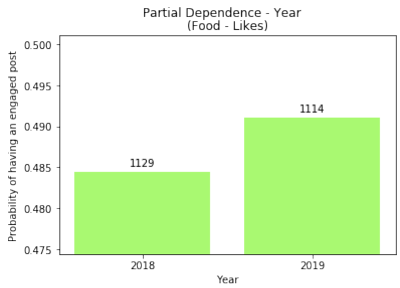

Partial Dependence - Giveaway_Brand
(Beauty - Likes)
Probability of having an engaged post
4297
386
Giveaway_Brand

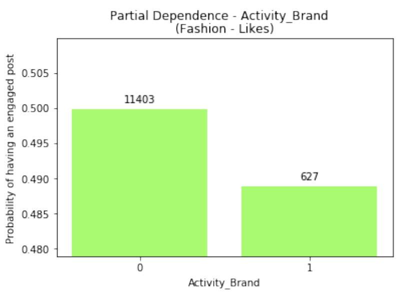

Partial Dependence - Giveaway_Brand
(Food - Likes)
Probability of having an engaged post
1897
346
Giveaway_Brand

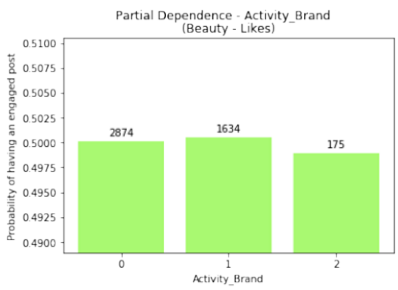

Partial Dependence - Person_Brand
(Fashion - Likes)
Probability of having an engaged post
2664
9366
Person_Brand

Partial Dependence - Person_Brand
(Beauty - Likes)
Probability of having an engaged post
3246
1437
Person_Brand

Partial Dependence - Person_Brand
(Food - Likes)
Probability of having an engaged post
1163
1080
Person_Brand

Partial Dependence - Video_Brand
(Fashion - Likes)
Probability of having an engaged post
10973
1057
Video_Brand

Partial Dependence - Video_Brand
(Beauty - Likes)
Probability of having an engaged post
3887
796
Video_Brand

Partial Dependence - Video_Brand
(Food - Likes)
Probability of having an engaged post
1516
727
Video_Brand

Partial Dependence - Multiple_Post_Brand
(Fashion - Likes)
Probability of having an engaged post
11029
1001
Multiple_Post_Brand

Partial Dependence - Multiple_Post_Brand
(Beauty - Likes)
Probability of having an engaged post
4412
271
Multiple_Post_Brand

Partial Dependence - Multiple_Post_Brand
(Food - Likes)
Probability of having an engaged post
2083
160
Multiple_Post_Brand

Partial Dependence - Weekend
(Fashion - Likes)
Probability of having an engaged post
8831
3199
Weekend

Partial Dependence - Weekend
(Beauty - Likes)
Probability of having an engaged post
3500
1183
Weekend

Partial Dependence - Weekend
(Food - Likes)
Probability of having an engaged post
1774
469
Weekend

Partial Dependence - Month
(Fashion - Likes)
Probability of having an engaged post
1169 1116 1311 1215 1242 1265 1172 1124 619 637 574 586
Month

Partial Dependence - Month
(Beauty - Likes)
Probability of having an engaged post
422 431 457 427 404 466 451 459 330 309 264 263
Month

Partial Dependence - Month
(Food - Likes)
Probability of having an engaged post
187 198 214 253 221 232 237 262 94 118 130 97
Month

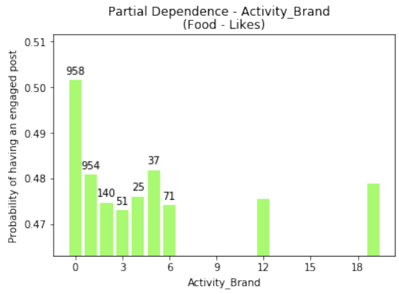

Partial Dependence - Year
(Fashion - Likes)
Probability of having an engaged post
7158
4872
2018
2019
Year

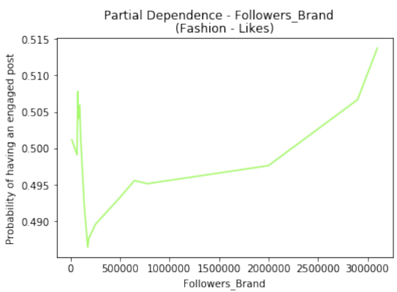

Partial Dependence - Year
(Beauty - Likes)
Probability of having an engaged post
2748
1935
2018
2019
Year

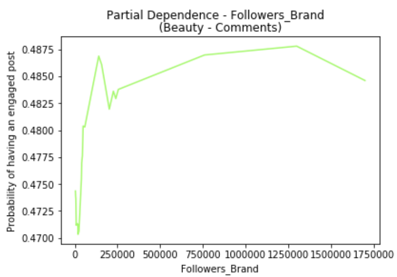

Partial Dependence - Year
(Food - Likes)
Probability of having an engaged post
1129
1114
2018
2019
Year

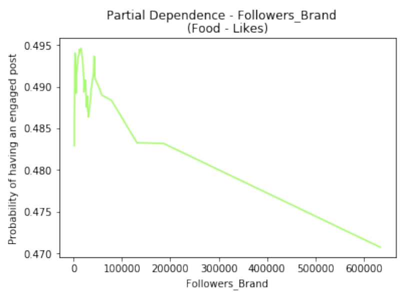

Partial Dependence - Activity_Brand
(Fashion - Likes)
Probability of having an engaged post
11403
627
Activity_Brand


Partial Dependence - Activity_Brand
(Beauty - Likes)
Probability of having an engaged post
2874
1634
175
Activity_Brand

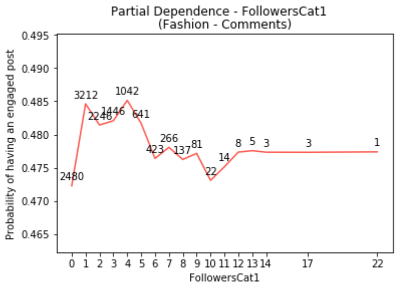

Partial Dependence - Activity_Brand
(Food - Likes)
Probability of having an engaged post
958
954
140
51
25
37
71
Activity_Brand


Partial Dependence - Followers_Brand
(Fashion - Likes)
Probability of having an engaged post
Followers_Brand

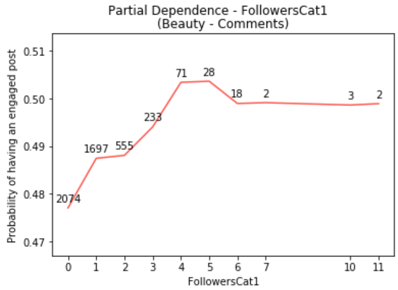

Partial Dependence - Followers_Brand
(Beauty - Comments)
Probability of having an engaged post
Followers_Brand


Partial Dependence - Followers_Brand
(Food - Likes)
Probability of having an engaged post
Followers_Brand

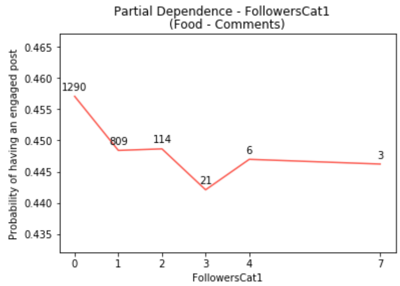
F.2 Comments



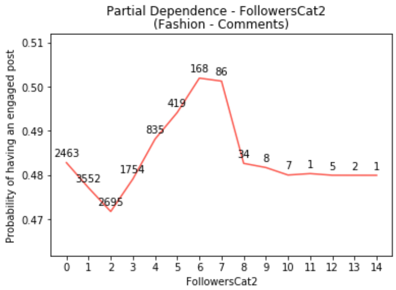
Partial Dependence - FollowersCat1
(Fashion - Comments)
Probability of having an engaged post
FollowersCat1



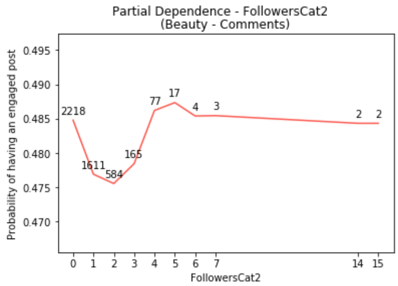
Partial Dependence - FollowersCat1
(Beauty - Comments)
Probability of having an engaged post
FollowersCat1



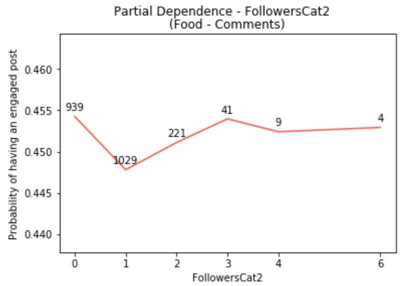
Partial Dependence - FollowersCat1
(Food - Comments)
Probability of having an engaged post
FollowersCat1



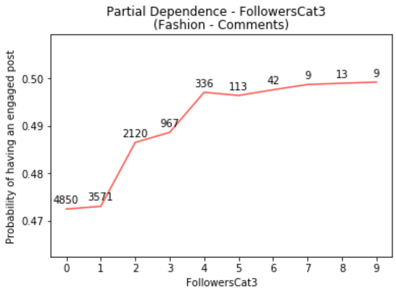
Partial Dependence - FollowersCat2
(Fashion - Comments)
Probability of having an engaged post
FollowersCat2



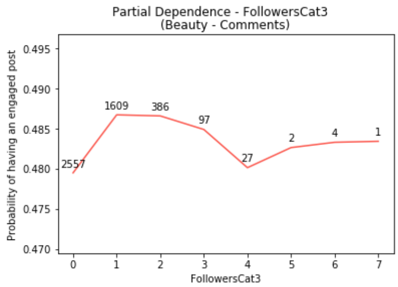
Partial Dependence - FollowersCat2
(Beauty - Comments)
Probability of having an engaged post
FollowersCat2



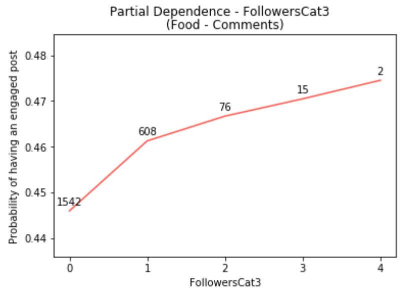
Partial Dependence - FollowersCat2
(Food - Comments)
Probability of having an engaged post
FollowersCat2



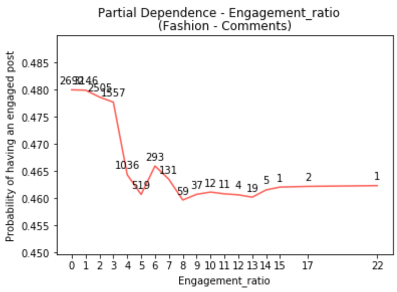
Partial Dependence - FollowersCat3
(Fashion - Comments)
Probability of having an engaged post
FollowersCat3



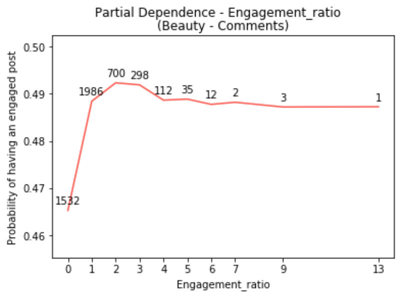
Partial Dependence - FollowersCat3
(Beauty - Comments)
Probability of having an engaged post
FollowersCat3



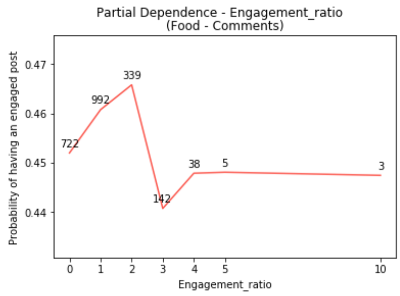
Partial Dependence - FollowersCat3
(Food - Comments)
Probability of having an engaged post
FollowersCat3



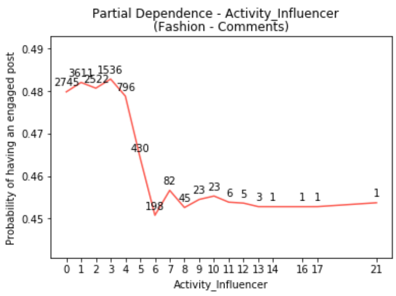
Partial Dependence - Engagement_ratio
(Fashion - Comments)
Probability of having an engaged post
Engagement_ratio



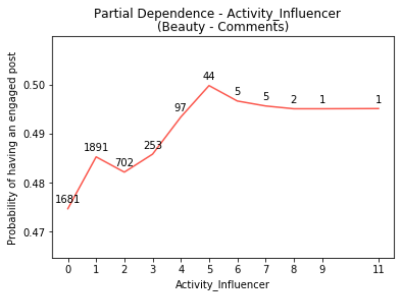
Partial Dependence - Engagement_ratio
(Beauty - Comments)
Probability of having an engaged post
Engagement_ratio



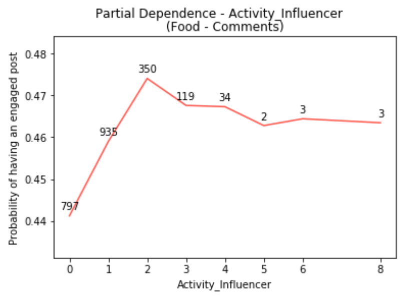
Partial Dependence - Engagement_ratio
(Food - Comments)
Probability of having an engaged post
Engagement_ratio



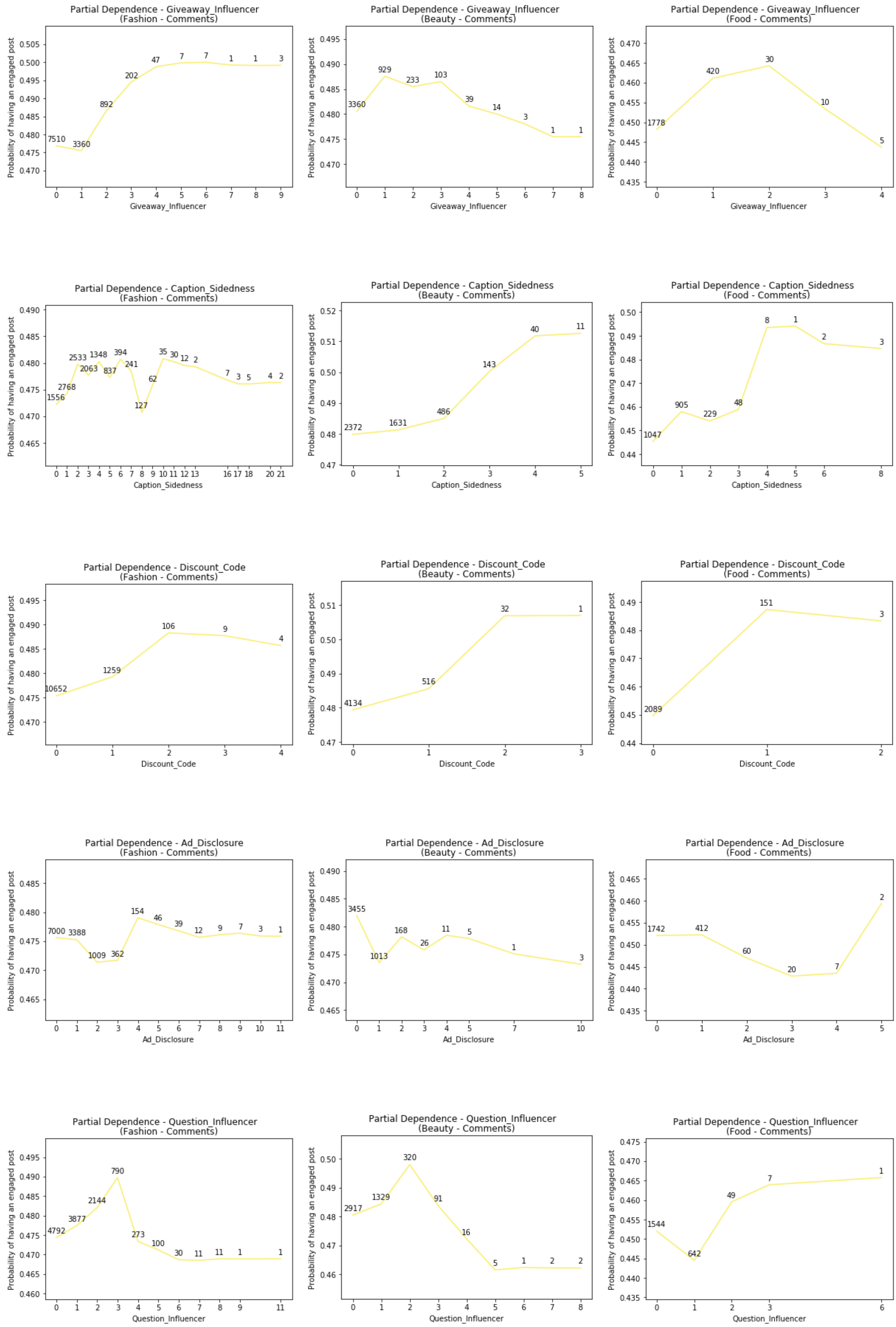
Partial Dependence - Activity_Influencer
(Fashion - Comments)
Probability of having an engaged post
Activity_Influencer

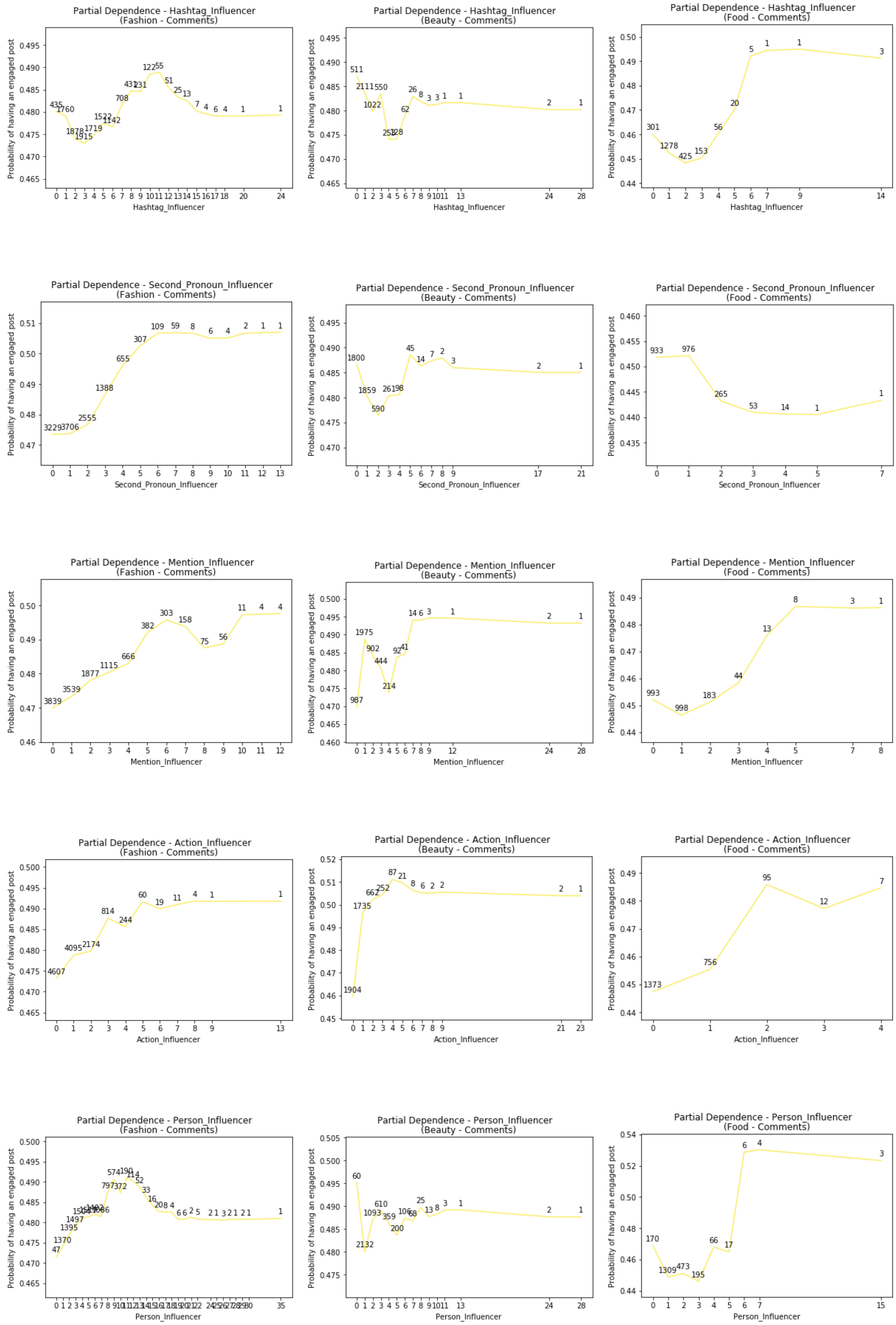


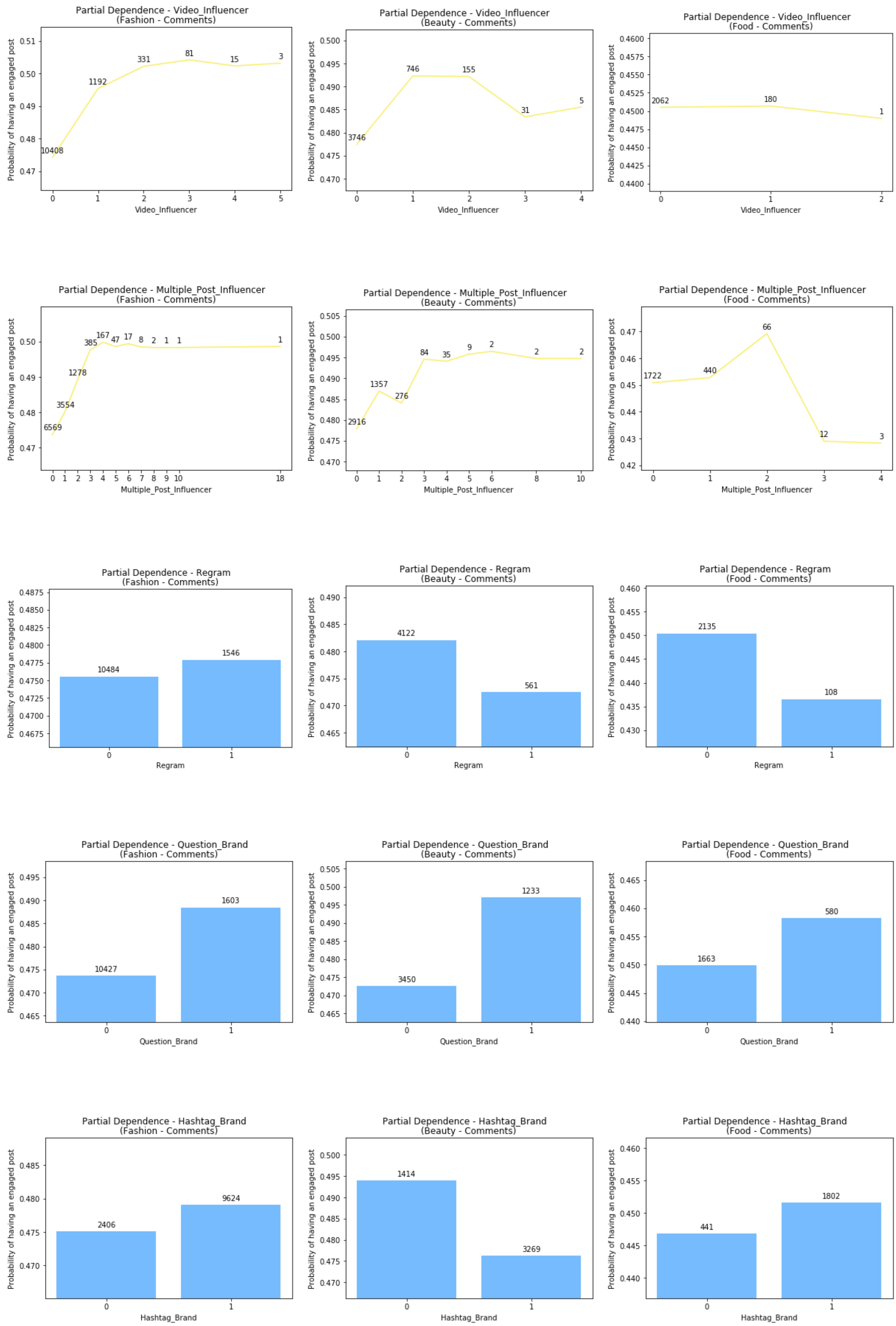
Partial Dependence - Activity_Influencer
(Beauty - Comments)
Probability of having an engaged post
Activity_Influencer

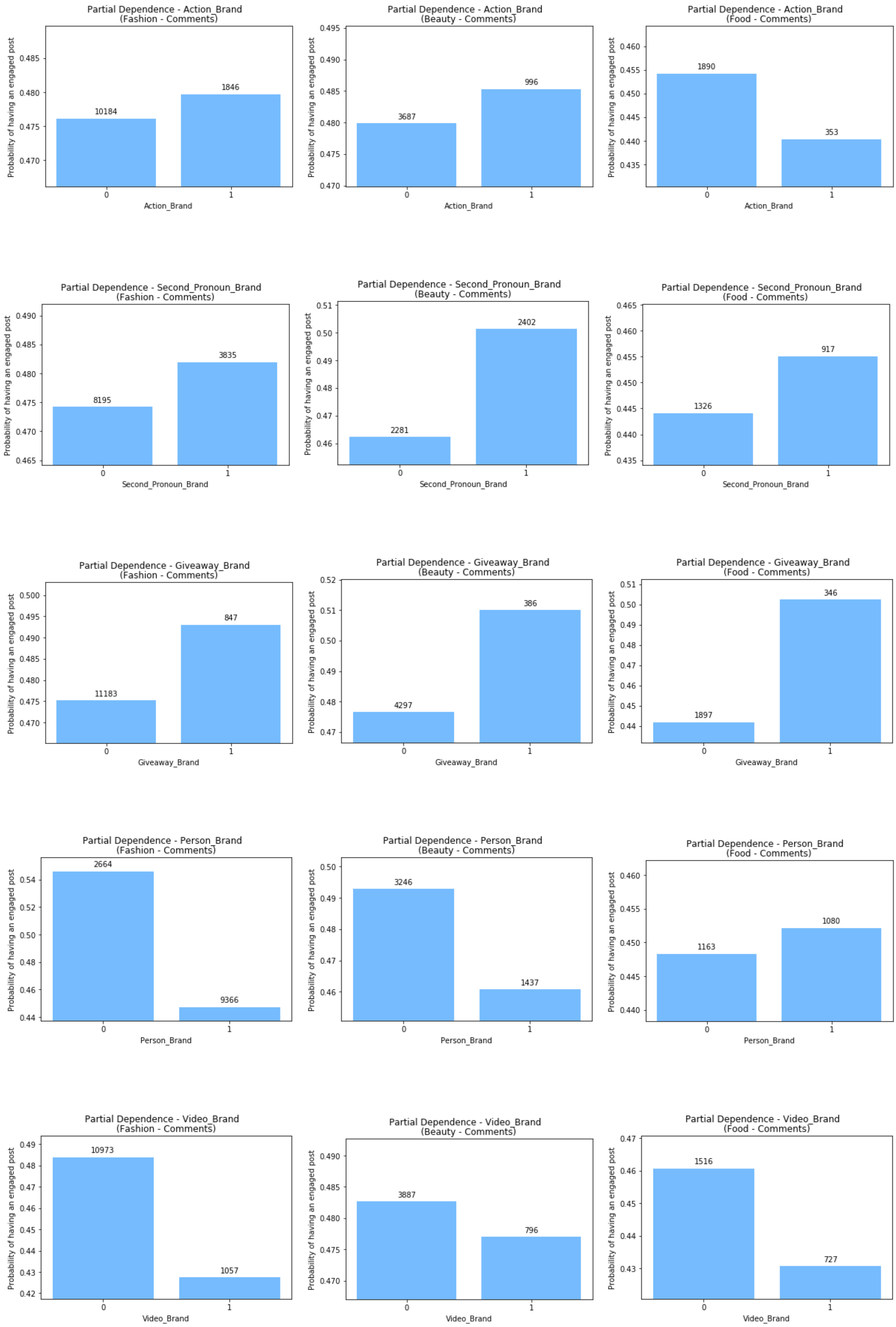


Partial Dependence - Activity_Influencer
(Food - Comments)
Probability of having an engaged post
Activity_Influencer

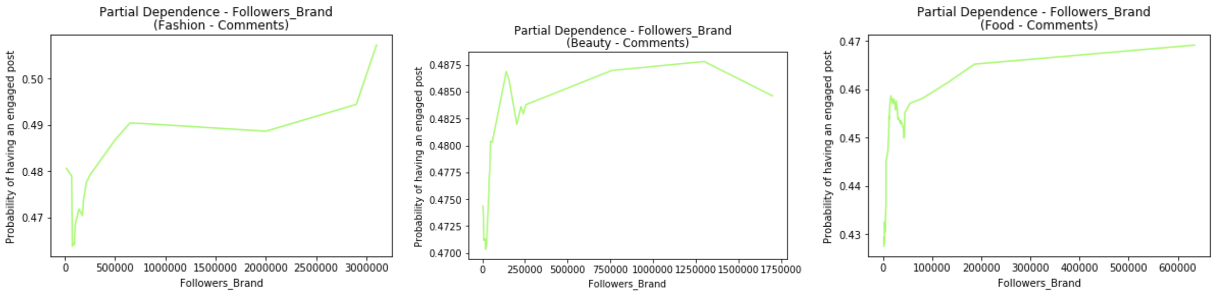

Partial Dependence - Giveaway_Influencer
(Fashion - Comments)
Partial Dependence - Giveaway_Influencer
(Beauty - Comments)
Partial Dependence - Giveaway_Influencer
(Food - Comments)
Giveaway_Influencer
Partial Dependence - Caption_Sidedness
(Fashion - Comments)
Partial Dependence - Caption_Sidedness
(Beauty - Comments)
Partial Dependence - Caption_Sidedness
(Food - Comments)
Caption_Sidedness
Partial Dependence - Discount_Code
(Fashion - Comments)
Partial Dependence - Discount_Code
(Beauty - Comments)
Partial Dependence - Discount_Code
(Food - Comments)
Discount_Code
Partial Dependence - Ad_Disclosure
(Fashion - Comments)
Partial Dependence - Ad_Disclosure
(Beauty - Comments)
Partial Dependence - Ad_Disclosure
(Food - Comments)
Ad_Disclosure
Partial Dependence - Question_Influencer
(Fashion - Comments)
Partial Dependence - Question_Influencer
(Beauty - Comments)
Partial Dependence - Question_Influencer
(Food - Comments)
Question_Influencer
Probability of having an engaged post


Partial Dependence - Hashtag_Influencer (Fashion - Comments)
Partial Dependence - Hashtag_Influencer (Beauty - Comments)
Partial Dependence - Hashtag_Influencer (Food - Comments)
Hashtag_Influencer
Partial Dependence - Second_Pronoun_Influencer (Fashion - Comments)
Partial Dependence - Second_Pronoun_Influencer (Beauty - Comments)
Partial Dependence - Second_Pronoun_Influencer (Food - Comments)
Second_Pronoun_Influencer
Partial Dependence - Mention_Influencer (Fashion - Comments)
Partial Dependence - Mention_Influencer (Beauty - Comments)
Partial Dependence - Mention_Influencer (Food - Comments)
Mention_Influencer
Partial Dependence - Action_Influencer (Fashion - Comments)
Partial Dependence - Action_Influencer (Beauty - Comments)
Partial Dependence - Action_Influencer (Food - Comments)
Action_Influencer
Partial Dependence - Person_Influencer (Fashion - Comments)
Partial Dependence - Person_Influencer (Beauty - Comments)
Partial Dependence - Person_Influencer (Food - Comments)
Person_Influencer
Probability of having an engaged post


Partial Dependence - Video_Influencer (Fashion - Comments)
Partial Dependence - Video_Influencer (Beauty - Comments)
Partial Dependence - Video_Influencer (Food - Comments)
Video_Influencer
Partial Dependence - Multiple_Post_Influencer (Fashion - Comments)
Partial Dependence - Multiple_Post_Influencer (Beauty - Comments)
Partial Dependence - Multiple_Post_Influencer (Food - Comments)
Multiple_Post_Influencer
Partial Dependence - Regram (Fashion - Comments)
Partial Dependence - Regram (Beauty - Comments)
Partial Dependence - Regram (Food - Comments)
Regram
Partial Dependence - Question_Brand (Fashion - Comments)
Partial Dependence - Question_Brand (Beauty - Comments)
Partial Dependence - Question_Brand (Food - Comments)
Question_Brand
Partial Dependence - Hashtag_Brand (Fashion - Comments)
Partial Dependence - Hashtag_Brand (Beauty - Comments)
Partial Dependence - Hashtag_Brand (Food - Comments)
Hashtag_Brand
Probability of having an engaged post


Partial Dependence - Action_Brand
(Fashion - Comments)
10184
1846
Action_Brand
Partial Dependence - Action_Brand
(Beauty - Comments)
3687
996
Action_Brand
Partial Dependence - Action_Brand
(Food - Comments)
1890
353
Action_Brand
Partial Dependence - Second_Pronoun_Brand
(Fashion - Comments)
8195
3835
Second_Pronoun_Brand
Partial Dependence - Second_Pronoun_Brand
(Beauty - Comments)
2281
2402
Second_Pronoun_Brand
Partial Dependence - Second_Pronoun_Brand
(Food - Comments)
1326
917
Second_Pronoun_Brand
Partial Dependence - Giveaway_Brand
(Fashion - Comments)
11183
847
Giveaway_Brand
Partial Dependence - Giveaway_Brand
(Beauty - Comments)
4297
386
Giveaway_Brand
Partial Dependence - Giveaway_Brand
(Food - Comments)
1897
346
Giveaway_Brand
Partial Dependence - Person_Brand
(Fashion - Comments)
2664
9366
Person_Brand
Partial Dependence - Person_Brand
(Beauty - Comments)
3246
1437
Person_Brand
Partial Dependence - Person_Brand
(Food - Comments)
1163
1080
Person_Brand
Partial Dependence - Video_Brand
(Fashion - Comments)
10973
1057
Video_Brand
Partial Dependence - Video_Brand
(Beauty - Comments)
3887
796
Video_Brand
Partial Dependence - Video_Brand
(Food - Comments)
1516
727
Video_Brand
Probability of having an engaged post


Partial Dependence - Multiple_Post_Brand (Fashion - Comments)
11029
1001
Multiple_Post_Brand
Probability of having an engaged post
Partial Dependence - Multiple_Post_Brand (Beauty - Comments)
4412
271
Partial Dependence - Multiple_Post_Brand (Food - Comments)
2083
160
Partial Dependence - Weekend (Fashion - Comments)
8831
3199
Weekend
Partial Dependence - Weekend (Beauty - Comments)
3500
1183
Partial Dependence - Weekend (Food - Comments)
1774
469
Partial Dependence - Month (Fashion - Comments)
1169 1116 1311 1215 1242 1265 1172 1124 619 637 574 586
Month
Partial Dependence - Month (Beauty - Comments)
422 431 457 427 404 466 451 459 330 309 264 263
Partial Dependence - Month (Food - Comments)
187 198 214 253 221 232 237 262 94 118 130 97
Partial Dependence - Year (Fashion - Comments)
7158
4872
Year
Partial Dependence - Year (Beauty - Comments)
2748
1935
Partial Dependence - Year (Food - Comments)
1129
1114
Partial Dependence - Activity_Brand (Fashion - Comments)
11403
627
Activity_Brand
Partial Dependence - Activity_Brand (Beauty - Comments)
2874
1634
175
Partial Dependence - Activity_Brand (Food - Comments)
958 954 140 51 25 37 71


Partial Dependence - Followers_Brand
(Fashion - Comments)
Probability of having an engaged post
Followers_Brand
Partial Dependence - Followers_Brand
(Beauty - Comments)
Probability of having an engaged post
Followers_Brand
Partial Dependence - Followers_Brand
(Food - Comments)
Probability of having an engaged post
Followers_Brand