Master Thesis Data Science & Marketing Analytics

"What makes Instagram content engaging?
Understanding visual content and its interestingness"

Abstract

This research aims to explore the relationship between visual content and engagement on Instagram, focusing on capturing the effect of intangible concepts known as visual interestingness. Visual interestingness captures to what extent people find an image or a video interesting. In this research, interestingness is explored through four key concepts: novelty, complexity, valence, and emotions. Visual content elements and their groupings are constructed through extensive text analysis and Natural Language Processing models. Their relationships are then explored through regression analysis. The findings indicate significant positive effects of humans and hedonistic activities such as summer travel, leisure, cozy breakfast, etc., on engagement.

Furthermore, significant effects are observed for two concepts relating to visual interestingness: novelty and complexity. These effects suggest that familiar images (not novel) and images that have dissimilar objects positively influence engagement. Valence and emotions are also significant in this research, such as anger, fear, joy, and positive and negative valence. Based on these results, this paper ends with marketing recommendations and points for further research.

The views stated in this thesis are those of the author and not necessarily those of the supervisor, second assessor, Erasmus School of Economics or Erasmus University Rotterdam.

Student name: Isidora Stojanac
Student ID number: 557536

Supervisor: A.C.D. Donkers
Second assessor: O. Vicil

Date hand-in: 21/04/2022

ERASMUS UNIVERSITEIT ROTTERDAM

Table of Contents

1 Introduction

Instagram has become a powerful tool for social media marketing and a favorite for targeting millennials. With around 1 billion users worldwide, it ranks as the fifth most popular social network platform and the most popular photo-sharing application. A favorite among social influencers, the volume of brand interactions exceeds the other two big social platforms (Facebook and Twitter), with 77% of brand interactions happening on Instagram. Moreover, the worldwide expenditure on Instagram marketing is expected to double in 2020 vs. 2018. Additionally, the total number of promoted posts is expected to double, surpassing 6 billion in 2020. (Statista, 2021)

On Instagram, it is all about the content. Pleasant, engaging pictures are shared every day, and photography is made available to anyone. One only needs a smartphone and one of the famous Instagram filters to make the picture trendy. The picture's appeal is measured by the level of engagement it generates. The most common key performance indicators used by influencers and brands alike are likes and comments received by post. Engagement is defined as all interactions users make with a person's profile or picture, including likes, comments, shares, views, direct messages, link click rates, story reactions, acts of following, etc.

An essential role in a picture appeal is played by the observer itself, their personal preferences, personality, and cultural background. Computer vision has allowed researchers to go beyond objects in a picture and look at more subjective concepts such as visual interestingness. Interest is classified as a defining factor for human motivation and behavior (Berlyne 1949). As such, it is of interest in this thesis. In the computational area, the concept of interestingness is projected in two perspectives: (1) visual interestingness, which is related to human emotions, motivation, and situational interest; and (2) social interestingness, which is related to social media concepts like popularity and virality (Constantin et al., 2019).

Visual interestingness and visual content have not been previously explored regarding influencer marketing on the social media platform Instagram. As this platform is the most prominent visual platform, this research can help understand what kind of visual content influencers generate leads to more engagement. This can help both influencers and businesses in improving their marketing decisions. For this reason, this thesis aims to answer the main research question:

"*What type of visual content can help influencers create Instagram posts that maximize engagement?*"

This paper presents scientific ways to detect and represent in-picture visual features using text as image descriptors to answer the research question. Additionally, this paper focuses on quantification and measurement of the visual interestingness of an image in a theory sound manner. Finally, the author examines the relationship between these features and aspects of the post engagement. Thus, the paper answers the following list of sub research questions:

1. *RQ1: "What type of visual features can be found in images?"*
2. *RQ2: "Which theme groups can be extracted from the images?"*
3. *RQ3: "What is the extent of concepts describing visual interestingness of an image posted by influencers on Instagram?"*
4. *RQ4: What relationships can be found between features of and concepts in posts and engagement with those posts on Instagram?"*

2 Literature review

A picture is worth a thousand words is a famous quote. This phenomenon is known as the "Picture superiority effect" in psychology. It has shown that pictures are more accessible to remember in conceptual and perceptual memory (Stenberg, 2006) and can change a viewer's perception using framing theory (Goffman, 1974). The visual rhetorical theory has been used to explain, among others, how different elements of a picture communicate together and form one entity. This entails observing the overall design using individual features and interactions (Foss, 2004). Observing the problem in this way is supported by the marketing theory for visual rhetoric established by Scott (1994), which states that when it comes to advertising, visual rhetoric plays a central role in how consumers process the image.

On the other hand, the area of visual interestingness explores the intangible concepts related to visual media and tries to explain what people perceive as interesting and how they react to that perception. This has been a popular topic for many researchers, especially humanities and psychology. Recently, due to the enormous increase of visual content streamed online and the rise of social media, researchers from fields such as multimedia and computer vision have started to dive deeper into what makes an image interesting. Thus, this paper focuses on and builds the theoretical framework around visual rhetoric and visual interestingness theories to define and quantify tangible and intangible visual features of images and model them to predict engagement.

2.1 Customer engagement

Harmeling et al. (2017) define customer engagement marketing as a "*firm's deliberate effort to motivate, empower, and measure customer contributions to marketing functions*." On social media, customer contributions that define engagement can vary from comments and likes to clicking on links, direct messaging, resharing, answering polls, etc. However, brands primarily focus on likes and comments as a measure of engagement. They are easily accessible, promptly available, and often used as direct feedback for a specific post. In their study, Mochon et al. (2017) conclude that page likes on Facebook positively affect customer behavior offline, mainly stemming from user exposure to the message. Therefore, this paper uses likes to measure engagement an Instagram image has received and considers this quantification appropriate in answering the main research question.

2.2 Extracting visual features

Bulmer & Buchanan-Oliver (2006) state that visual features images can be represented in forms of objects or personas, use of color, background, use of lighting, etc. These can be summarized together to represent the respective image and used to describe or quantify it. Highfield & Leaver (2016) elaborate further and, in their report, show that focusing on elements of an image maximizes the recognition of the sentimental value of an image, creating a marketing impact. According to Unnava & Burnkrant (1991), this way of looking at an image is close to how people themselves process the visual content. They state that visual content stimulates labeling in consumers' memory and requires more imaginary power than understanding text content. This entails that the visual feature of images has a decisive role in considering and understanding the image as a whole.

Therefore, this paper focuses on these individual features as potential drivers of engagement and relies on such descriptions of images generated by the software Clarifai. It is a third-party tool that uses computer vision and image recognition through neural networks to recognize and generate a textual description of an image. Image object tags and their metadata represent this description. These tags include objects (woman, man, car, etc.), feelings (happy, fun, pretty, etc.), and ideas (leisure, love, the act of giving, etc.) (Clarifai, n.d.; Jaakonmäki et al., 2017). An example of an image from the available dataset and it is corresponding Clarifai description is given in Figure 1 below:

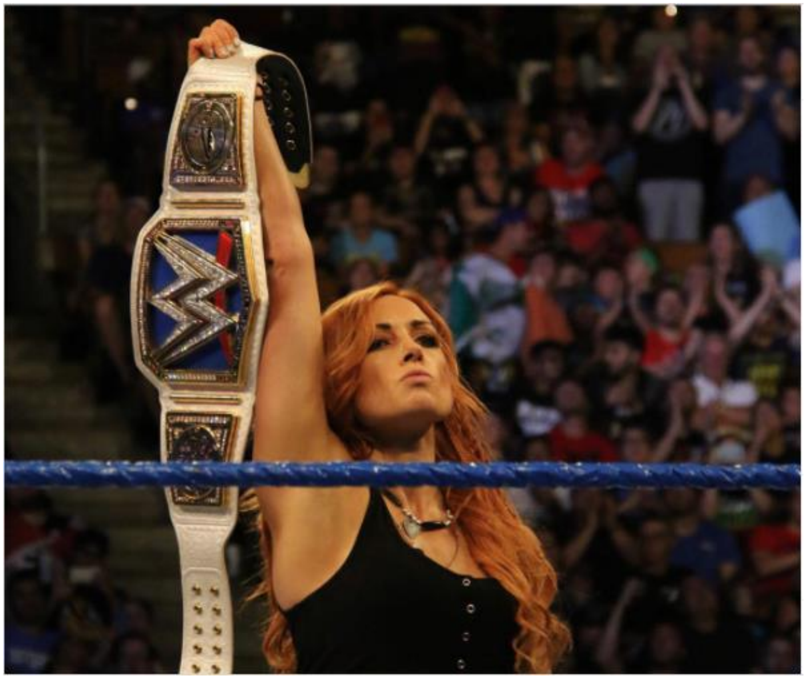


Figure 1: Image example from the dataset and Clarifai objects tags attached: competition, athlete, sports, equipment, people, adult, strength, victory, many, championship, stadium, strong, effort, jewelry, band, one

2.2.1 Visual content and engagement

The use of individual features as drivers of engagement was used in other papers before. They serve as a solid foundation and reasoning for using individual features in this paper. First, Hu et al., 2014 defined eight distinct categories of Instagram images using an extensive cluster and classification analysis of User Generated Content (UGC). Nearly half of the photos contain faces (friends and selfies). Also, these types of photos are 38% more likely to receive likes from the users (Bakhshi et al., 2014). The study of Jaakonmäki et al., 2017 supported this idea. It confirmed the positive effect of human faces in an Instagram environment.

2.3 Understanding interestingness

Visual interestingness captures to what extent people find an image or a video interesting (Silvia, 2005). In the Oxford English language dictionary (Stevenson, 2010), interest represents how someone, or something can hold or catch someone's attention. Nowadays, due to the worldwide popularity of social media and an enormous number of generated images and videos daily, the attention span of users is considerably reduced per content (Romero et al., 2011), making it hard for businesses to stay relevant. Thus, recommender systems need to recognize and suggest exciting items to users. Furthermore, suppose the content itself is relevant and interesting. In that case, users will spend more time on the platform, engaging with its content. Thus, understanding and predicting visual interestingness is crucial for engagement and user retention, which links to company profit.

Capturing and understanding visual interestingness is a complex endeavor due to the complexity and ambiguity of the topic and the limitation of available data. Traditionally, visual interestingness was explored in relationship with human perception and emotions. However, with the rise of social media, research of interestingness in the area of computer vision is projected in two perspectives (1) visual interestingness, which is related to human emotions, motivation, and situational interest; and (2) social interestingness, which is related to social media concepts like popularity and virality (Constantin et al., 2019). While virality is mostly defined as a probability of an image being reshared, popularity is seen as the probability for an image to receive likes. This paper makes an effort to model social interestingness through tangible and intangible concepts of visual content such as visual features and abstract concepts of visual interestingness. Furthermore, the social interestingness of an Instagram image in this paper is expressed through the number of likes it receives, capturing engagement as explained in section 2.1 Customer Engagement.

2.3.1 Importance of visual interestingness

Constantin et al., (2019) summarized the problem of modeling interestingness in three points. First, they mention that interestingness depends on a person's subjective perception, making it complex for understanding and quantification. Second, they mention that the data needed to understand the topic is not widely available and sometimes requires particular methodologies for data collection. Thirdly, the feature extraction or model creation for a rating of interestingness is complex and, in a way, paradoxical task for the algorithm. E.g., when recognizing objects, the algorithm aims to minimize the differences. However, when searching for interestingness, the aim is to enhance them. Modeling interestingness is, therefore, non-trivial.

Nonetheless, previous studies have shown that visual interestingness correlates to abstract and subjective concepts such as complexity or unexpectedness, but this area is still unexplored. Furthermore, to the author's knowledge, linking visuals to social interestingness in the form of likes or shares has been done twice in the literature. A brief literature overview is given in Table 1 below:

Table 1: Overview of abstract concepts previously linked to visual interestingness

Positively correlated	**Negatively correlated**
Valence (Gygli et al., 2013)	Valence (Turner and Silvia, 2006)
Arousal (Soleymani, 2015)	Virality (Deza and Parikh, 2015)
Novelty (Gygli et al., 2013)	Popularity (Hsieh et al., 2014)
Unusualness (Zhao et al., 2011)	
Unexpectedness (Padmanabhan and Tuzhilin, 1999)	
Complexity (Silvia, 2005)	
Popularity (Gygli and Soleymani, 2016)	

2.3.1.1 *Complexity & Novelty*

Berlyne (1960) identified four factors that create or influence visual interest: novelty, complexity, uncertainty, and conflict. He found that new, unexpected, and complex events generate interest in his work. This is further explored by Silvia (2005), who defined two phases that lead to interest (1) a high novelty-complexity appraisal, and (2) a high coping-potential appraisal, where coping-potential is defined as a vast potential of the object.

Complexity can be defined as “the amount of variety and diversity in a stimulus pattern.” (Berlyne, 1960) Additionally, Berlyne (1960) defined important characteristics of complexity: if the number of elements in an object increases, the complexity also increases; Similarly, if the number of elements is held constant, the increase in their dissimilarity increases the complexity of the object. Thus, the complexity decreases if the objects can be grouped in a standard unit. The opposite concept is simplicity. Yu & Winkler (2013) gave an additional definition of complexity: “The complexity of an object or a system measures the inherent difficulty of performing the tasks associated with it.”

Novelty represents something new, something rarely seen. Maher (2020) defines novelty as “a measure of a distance from other artifacts in space.” Novelty is related to originality, unexpectedness, and unusuality (Constantin et al., 2019). Thus, in the effort to quantify novelty, this paper reaches for literature defining the focal word and its synonyms or related words.

Zhao et al. (2011) focused on detecting anomalies to recognize unusualness. Padmanabhan and Tuzhilin (1999) defined interestingness as an intensely subjective concept and developed ways to detect unexpected patterns leveraging apriori management knowledge. Gygli et al. (2013) measured novelty through The Local Outlier Detector (LOF) algorithm, which worked well on images with solid context but could not use apriori knowledge. All three papers have found a positive correlation between these concepts and with visual interestingness of an image.

2.3.1.2 Valence and Arousal

Emotional reactions can be summarized into two groups, arousal and valence (Holbrook and Batra, 1987). Arousal describes how we feel stimulated, excited, alert, or active. On the other hand, valence describes how we feel happy, joyful, etc., or the opposite of those emotions (Russell and Mehrabian, 1974). Moreover, a study by Schwarz (2000) concludes that consumers are driven by emotions in the absence of a specific goal, which is often the case when browsing social media content. Thus, emotional content could potentially drive high engagement on social media.

In their paper, Gygli et al. (2013) research what arouses human interest. They argue that even though there is a strong correlation between interestingness and aesthetics, one does not imply another. They found that certain emotions and valence strongly correlate with interestingness, such as *pleasant, exciting, makes happy, makes sad, and arousing*. Additionally, emotion-inducing content was found to be more critical than informative content in driving the engagement, with high positive and negative values of arousal being the main drivers (Rietveld et al., 2020).

2.3.2 Importance of social interestingness

As previously mentioned, social interestingness is related to the popularity and virality of an object. In a study by Hsieh et al. (2014), visual and social interestingness was found to be very weakly correlated, and usually, the correlation is negative. The popular content on social media is not necessarily visually interesting (Hsieh et al., 2014). The authors worked with still images from Pinterest and used colors, edge information, saliency, and texture information as features. Social interestingness was determined by social media ranking and visual interestingness through crowdsourcing.

On the other hand, Gygli and Soleymani (2016) found the opposite effect with different media types. They focused on Graphics Interchange Format (GIF) images. They found that although the visual interest was correlated with the number of likes the GIF receives, this was not the case for reshares and the social popularity of the user.

Additionally, other factors have been found to impact social interest next to the visual interest and the content of an image. Khosla et al. (2014) have found that image popularity largely depends on social ties within the network, such as using the number of followers. In addition, visual content and context, user context, tags, and text have been used as features in predicting popularity by other researchers (Gelli et al. 2015, McParlane et al. 2014, Aloufi et al. 2017). They have proven to correlate to some extent. Finally, the aspects that define visual interestingness are influenced by subjective experiences and feelings that are almost impossible to measure, which introduces more complexity to the process.

3 Methodology

3.1 Research approach

This paper aims to answer the main research question "*What type of visual content influences engagement on Instagram*" by utilizing the research framework shown in Figure 2 below:

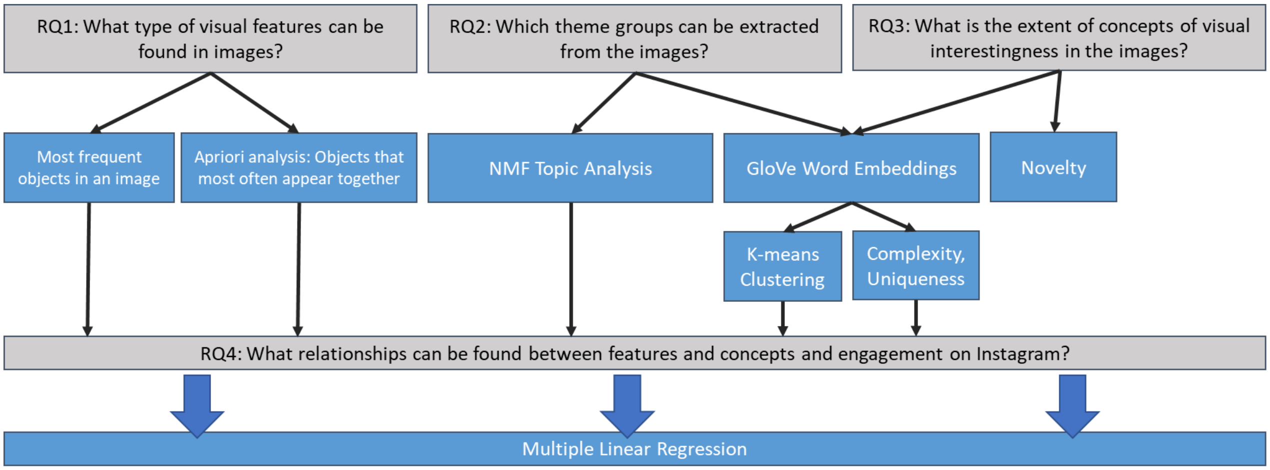

RQ1: What type of visual features can be found in images?
RQ2: Which theme groups can be extracted from the images?
RQ3: What is the extent of concepts of visual interestingness in the images?
Most frequent objects in an image
Apriori analysis: Objects that most often appear together
NMF Topic Analysis
GloVe Word Embeddings
Novelty
K-means Clustering
Complexity, Uniqueness
RQ4: What relationships can be found between features and concepts and engagement on Instagram?
Multiple Linear Regression

Figure 2: Research framework

The approach entails the use of both supervised and unsupervised modeling techniques. First, for feature engineering, an unsupervised approach is taken. To extract visual features, the author used text and basket analysis in the form of the Apriori algorithm. This approach answers the first sub-research question: *“Which visual features can be extracted from the image content of influencer promotional posts on Instagram?”.*

Word embeddings are computed using the global vector model (GloVe) to answer the second and third research questions. Next, K-means cluster analysis is applied to identify underlying themes and groups in images based on word representations previously extracted from GloVe. Additionally, NMF is used to extract underlying topics in the data. Both NMF and K-means aim to answer the second research sub-question: *“Which theme groups can be extracted from the image content of influencer promotional posts on Instagram?”.*

After feature extraction and topic identification, an effort is made to quantify visual interestingness. The quantification is strongly guided by literature, and the formulas and the exact approach are defined later in this chapter. This answers the third research sub-question: *“What is the extent of intangible concepts of visual interestingness present in the image content of influencer promotional posts on Instagram?”.*

Finally, a supervised model is used to answer the fourth research sub-question: “What relationships can be found between features and concepts and engagement of influencers’ promotional posts on Instagram?”. Features, clusters, and intangible concepts of visual interestingness are used in this step. The predictive analysis is employed using linear regression.

3.1.1 Global Vectors for Word Representation

Word vector models strive to learn the meaning of words by representing them as a real-valued vector (embeddings), where similar words are clustered together in a vector space. Global Vectors (GloVe) is a semantic vector space model. Word representation is created using local context information of words and word co-occurrence matrix, capturing local and global corpus statistics. Local context is derived through the neighborhood approach on the word level using a focal window. A focal window consists of a focus word and neighborhood words in a specific size window. The global context is derived using conditional probabilities of words appearing together in the global word corpus. Unlike its popular alternative, Word2Vec, which relies solely on local context, and thus the word Global in the name. GloVe builds on the advantages of using local and global statistics where the use of local context allows for performing analogy tasks. In contrast, the global aspect enables it to handle word ambiguity.

GloVe is an unlikely choice for this paper since the words generated by Clarifai software are an isolated, unstructured set of words without a local context of text semantics. On the other hand, GloVe helps establish the meaning of an object in a photo by taking into consideration the neighborhood objects and, in such, capturing the synergy and contextual information of these objects. Moreover, utilizing the global context ensures the quality of word representations by analyzing them on a deeper level than local context models. This ensures a higher probability of recognizing essential words and is why GloVe proved to outperform other word vector models (Pennington et al., 2014). Additionally, GloVe focuses on interpreting the meaning of keywords, where frequent stop words, pronouns, prepositions, and adverbs ("I", "the", "a", etc.) might not add valuable information to the model, especially in the context of this paper. Since these words are absent in Clarifai tags, it is not an issue.

The resulting word embeddings serve as input for cluster analysis in uncovering clusters of content in images to answer sub-research question 3. We expect that this approach will provide higher granularity and, as such, result in clusters with specific context that can be more valuable for interpretation than the more robust information provided by other solutions like topic modeling.

Next, to understand the technical details behind the algorithm, we look at the paper published by Pennington et al., 2014. The authors start by building a co-occurrence matrix X with dimensions V x V, where V represents the number of words in a corpus. This matrix contains all words from the corpus, while its entries X_{ij} represent how often does the word j appear in the context of word i. Next, co-occurrence probability ratios are calculated, taking into consideration k number of probe words:

$$\frac{P_{ik}}{P_{jk}}, where\ P_{ik} = \frac{X_{ik}}{X_i},$$

For words k related to i but not j, ratio $\frac{P_{ik}}{P_{jk}}$ will be large, and vice versa. If the words k are related or unrelated to both i and j, ratio is close to one. This ratio is better than using raw probabilities because it can better distinguishing relevant from irrelevant words and can better discriminate between two relevant words.

Next, since some function F should capture information present in the above-mentioned ratio in a word vector space, the function F can be restricted to depend on the differences of two target words:

$$F(w_i - w_j, u_k) = \frac{P_{ik}}{P_{jk}}, where \tag{1}$$

$w \in R^d$ are word vectors and $u_k \in R^d$ are context word vectors.

Next, since left side of equation (1) are vectors and right side is a scalar, to prevent F of mixing vector dimensions, left side is represented as a dot product:

$$F\left((w_i - w_j)^T u_k\right) = \frac{P_{ik}}{P_{jk}}. \tag{2}$$

Since in the context of word co-occurrence matrix we are free to exchange the roles of words and its context words, a symmetry and homomorphism between groups must be maintained, which is lost in equation (2). First, homomorphism is ensured by:

$$F\left((w_i - w_j)^T u_k\right) = \frac{F(w_i^T u_k)}{F(w_j^T u_k)}, which\ by\ equation\ (2)\ is\ solved\ by: \tag{3}$$

$$F\left(w_i^T u_k\right) = P_{ik} = \frac{X_{ik}}{X_i}. \tag{4}$$

Assuming that the solution to equation (4) is F = exp, the authors transform the entries to logarithms in order to scale down on large non-zero entries in the matrix:

$$w_i^T u_k = \log(P_{ik}) = \log(X_{ik}) - \log\,(X_i). \tag{5}$$

Next, to ensure symmetry $\log\,(X_i)$ is absorbed in a bias b_i for w_i, and additional bias b_k for u_k is added.

$$w_i^T u_k + b_i + b_k = \log(X_{ik}). \tag{6}$$

To ensure not all co-occurrences are weighted equally since less frequent ones contain less information and can introduce noise compared to the frequent ones, the authors observe equation (6) as a least squares problem and introduce a weighting function $f(X_{ij})$ into the cost function. This approach provides the following model:

$$J = \sum_{i,j=1}^{V} f(X_{ij})(w_i^T u_k + b_i + b_k - \log(X_{ij}))^2\,, where \tag{7}$$

V is the number of words in a corpus. Finally, the weighting function should be zero in case of zero values ($f(0) = 0$), to ensure the convergence of $\log\,(x_{ij})$; it should be non-decreasing to prevent overweighting of rare co-occurrences and it should be small for very large co-occurrences to prevent overweighting them. This function is defined as:

$$f(x) = \begin{cases} (x/x_{\max})^{\alpha}, if\ x < x_{max} \\ 1, \quad otherwise. \end{cases} \tag{8}$$

3.1.2 Non-Negative Matrix Factorization

Topic models focus on global context to establish topics in the text. The global context is here referred to as document-level context, which is image-level in the case of this paper. Topic models strive to identify which topics are dominant in which documents through the bag-of-word approach, meaning that the location and order of words are not relevant. This fits well with the tags provided by Clarifai that do not have a specific location and are generated in the order of probabilities that the object actually reflects what is depicted in the picture. This approach essentially summarizes the extensive collection of individual objects in the form of more meaningful and distinctive topics for interpretation and modeling. For example, suppose images contain words such as “chair”, “desk”, “sofa”, “vase”, “decoration”, “rug”, and “light”. In that case, they can be summarized as a topic “interior design”. These topics might provide more information than looking at objects individually (“interior design” versus “chair”).

The model used in this thesis from the family of Matrix Factorization models is Non-Negative Matrix Factorization. It is introduced by Lee and Seung, 2001. Like other matrix factorization models, it works by decomposing a matrix **V** to two approximate matrices **W** and **H** ($V \approx WH$), but in this case matrix **V** is non-negative. It achieves this by setting all negative values in the matrix W and H to zero.

In this paper, matrix V is a document-term matrix with $\boldsymbol{n} \times \boldsymbol{m}$ dimensions, where $\boldsymbol{n}$ stands for the number of unique terms or object tags in this case, and $\boldsymbol{m}$ stands for the number of observations. This is then approximated to matrix W (basis matrix) with dimensions $\boldsymbol{n} \times \boldsymbol{r}$ and matrix H (document matrix) with dimensions $\boldsymbol{r} \times \boldsymbol{m}$ where $\boldsymbol{r}$ stands for the number of dimensions (topics) and is chosen by the researcher. In general, the number chosen should be lower than m or n: $r < \min{(m, n)}$. This reduction in dimensionality implies that the basis matrix W can capture some latent structures in the data, which are then interpreted as topics.

Cost function used to measure the quality of approximation is Euclidean distance. The aim is to minimize the distance between the input matrix V and it's two approximate matrices W and H:

$$\min_{W,H \geq 0} \parallel V - WH \parallel^2.$$

This is a non-negative least squares problem that is solved in iterations. The algorithm fixes one factor W or H while optimizing the other and then alternates. The resulting topics are interpreted based on their high scoring terms.

3.1.3 Apriori

Association rules are widely used in business and are based on "if-then" rules. These rules consist of two metrics that express the support and confidence of the rule found in the dataset. This method originates from basket analysis and is now one of the most popular techniques for data mining. One of the first and most popular algorithms in this area is the Apriori. In this thesis, Apriori is used to uncover rules in the dataset that yield two terms that often appear together. Since the object tags are an unordered set, they can be observed as items in a basket analysis, where n-grams techniques would not be appropriate. A famous example of one such found rule is that people that buy diapers also buy beer.

Every rule is given with accompanying metrics of support and confidence. Support gives information on what fraction of transactions contains the specific itemset. For this thesis, transactions are referred to as images and item sets as object sets. Given a set of images D, support for object set (O) is therefore calculated as:

$$support(O) = \frac{Number\ of\ images\ containing\ O}{Total\ number\ of\ images}$$

Confidence is a measure that gives information on how often an object Y appears in an image that contains object X:

$$confidence(X \rightarrow Y) = \frac{Number\ of\ images\ containing\ X\ and\ Y}{Number\ of\ transactions\ containing\ X}$$

The problem being solved by association rules mining is to generate object sets (rules) that have a higher support and confidence than the minimum support (*minsup*) and minimum confidence (*minconf*) specified by the user. This problem is decomposed into two parts by (Agrawal and Srikant, 1994) the authors of the Apriori algorithm:

First, the goal is to find sets of objects that have image support above the *minsup*. Support for every object set is calculated with the formula mentioned above. These object sets are referred to as *large* object sets, and all others are referred to as *small* object sets. The algorithms pass the data multiple times in order to discover these sets. First, the support of individual items is calculated, and those with higher support than specified *minsup* are classified as large. In every next pass, these objects classified as large are used to generate new candidate object sets. The objects classified as small are therefore not taken into consideration anymore. The support is again calculated and measured against the minimum. Candidate sets that satisfy this condition are labeled as large. The object sets labeled as small are not considered in the next pass. The process continues until no more large sets are found in the data.

Second, the resulting large object set is used to generate association rules. If XYZW and XY are large object sets, then we can determine if the rule XY -> ZW satisfies our minimum threshold (*minconf*) by computing the ratio $conf = support(XYZW)/support(XY)$. If $conf \geq minconf$, then the rule stands. The improvement of Apriori algorithm over the older ones is that it allows for generation of multiple objects in consequent. For example, it is possible to find a rule that in X% of cases, object woman appears together with objects man, skirt and coffee.

3.1.4 K-means

K-means is an algorithm belonging to unsupervised models' family, used as a clustering technique, to uncover clusters of similar objects in the data. The goal is to cluster in such a way, where observations in one cluster are similar, while observations between clusters are dissimilar. These similarities are expressed through distances in data points. In this thesis, K-means is used to uncover groups of images that can be used as visual features input to predict engagement. These groups are valuable because they summarize similar objects or ideas, opposed to only using individual objects as input. For example, chair, desk, indoor, bed, sofa, decoration can be possibly summarized as interior design and as such provides more wholesome information than observing chair or desk individually.

The distance measure used is Euclidian distance, which is a classical measure used in this algorithm. The general idea is to find such clusters where total within-cluster variation is minimized. The algorithm used is by (Hartigan and Wong, 1979) and total within-cluster variation is given as the sum of squared Euclidian distances between items and it's centroid:

$$W(C_k) = \sum_{x_i \epsilon C_k} (x_i - \mu_i)^2$$

Where x_i is an observation belonging to the cluster C_k; and μ_k is the mean value of observations belonging to the cluster C_k.

The algorithm works as following: after indicating the number of desired cluster (*k*), the algorithm randomly selects *k* number of images as initial centroids. Then, for each of the remaining images, k-means calculates their distance (Euclidian) from the centroids and assigns them to the closest one. After the clusters have been assigned, a mean is calculated for each cluster, and these serve as new centroids. After the centroids have been updated, the distance calculation from each image to each cluster is calculated again and images are assigned to the closest cluster. Then, the cluster mean is calculated again, and centroids adjusted. These iterations are repeated until there are no more changes in clusters (Hartigan and Wong, 1979).

3.1.5 Custom measures: Novelty, Uniqueness & Complexity

In an effort to capture intangible aspects of a picture such as its complexity and novelty, custom measures were constructed.

Novelty

This measure was based on the frequency of appearance of objects in the dataset. The analogy is that if objects present in an image are not so frequently found in the dataset, those objects are likely novel.

The quantification approach was as follows: frequency of each object in the data was counted, then frequencies of objects per image are summed and divided by the number of objects in that image. For a dataset of images I with the size of n, and m number of unique objects, we look at object k ($O_{k,i}$) in image I_i, where $k = \{1, \dots, m\}\ and\ i = \{1, \dots, n\}$. If object k is found in an image then $O_k = 1$, otherwise $O_k = 0$:

$$Freq(O_k) = \sum_{i=1}^{n} O_{k,i} \tag{1}$$

Then, the frequencies of objects are summed per image, based on the objects in that image and divided by their count:

$$Novelty(I_i) = \frac{\sum_{k=1}^{p} Freq(O_k)}{p} \tag{2}$$

Where p is the total number of objects found in image I_i.

To make the measure more intuitive for interpretation, it is reversed scored in order for high values to reflect novel images, and vice versa:

$$Rev. Novelty(I_i) = \frac{1}{Novelty(I_i)} \tag{3}$$

Reversed novelty score indicates how novel the image is on average, taking into consideration all objects found in it. Higher score indicates more novel image consisting of more uncommon object(s) and lower score indicates a less novel image with common objects.

Uniqueness

As Novelty accounts for the average novelty of the whole image, a measure accounting for a single unique object in the image itself might also be interesting. This comes from the premise that seeing an object in an unique, unusual environment is something rarely seen, which can be defined as novel or unique.

The quantification of this measure is based on Maher, 2020 definition of novelty: "*a measure of a distance from other artifacts in space*", and as such uses GloVe similarity matrix as its base. If an object is more dissimilar with other objects in a picture, we can say that it does not belong to that environment and is considered novel. Thus, the unique object is quantified as an object with maximum cosine distance over all associated objects in a single image. Given the symmetrical similarity matrix A with $m \times m$ dimensions, where m is the total number of distinct objects in the dataset, the most unique object per image is the one least similar with all other objects in that image:

$$Unique\ object(I_i) = MIN\left(\sum_{k=1}^{p} a_{jk}\right) \tag{1}$$

Where a_{jk} ($j, k = \{1, \dots, p\}$) are the elements (words) in similarity matrix A, and p is the number of objects found in image I_i.

To make the formula reflect the measure more intuitively, it is reverse scored in order for high values to indicate more unique object (higher distance from the other objects) and vice versa:

$$Rev. Unique\ object(I_i) = \frac{1}{Unique\ object(I_i)}. \tag{2}$$

This measure gives the score of the most unique object per image and when compared across all dataset can indicate a range where the lower value means it’s more unique and upper value means it’s less unique than average score.

Complexity

This measure was constructed by referring to the definition of Berlyne, 1960. In the context of Instagram, complexity by his definition refers to the difference in objects found in the image, as well as their number. The more different they are, and the more objects can be found, complexity increases. Opposite is simplicity which indicates the objects are similar to each other and/or there are not many objects present. The data used in this report does not contain full list of objects in the image, but rather only the ones with high confidence threshold provided by the object recognition software used. Because of this, every image has around 20-21 object tags. Due to the low variance, this paper focuses only on the first part of Berlyne definition, related to dissimilarity of objects.

This paper uses GloVe model to calculate the similarity matrix between objects in an image, which then serves as a base for the quantification of Complexity. Since the input is a similarity matrix, first formula is giving us a measure of simplicity which is opposite of complexity. This formula is then reverse scored to have a more intuitive interpretation in the model. After the GloVe similarity matrix A is constructed, a function iterates through all observations filtering the similarity matrix for only the terms appearing in that specific observation. Since this similarity sub-matrix is symmetric, the formula takes the mean of the upper triangle without the diagonal. This represents the average simplicity of the image per post.

Given the dataset of images I with the size of n, and the similarity matrix A with the size $m \times m$, a sum of matrix values a_{jk} is calculated for a single image based on p number of objects in that image. Next, the diagonal is subtracted and since it is a diagonal of ones (1), its value is equal to the number objects in the image p. Next, the value is divided by 2 to account for only one triangle, and divided by the number of objects in the image:

$$Simplicity(I_i) = \frac{(\sum_{j=1}^{p} \sum_{k=1}^{p} a_{jk}) - p}{2p}. \tag{1}$$

Next, we reverse score the measure to make the measure reflect complexity as it is the focus of this paper:

$$Complexity(I_i) = \frac{1}{Simplicity(I_i)}. \tag{2}$$

4 Data

4.1 Data description

The dataset is provided by a Dutch influencer platform TheCirqle. The company bridges the gap between companies and influencers by providing supervised marketing campaigns in fashion, lifestyle, and travel. The platform is open for registration to anyone; thus, the data might contain posts from people with a small number of followers and, as such, still not recognized as influencers. The dataset contains influencers of over 50 nationalities, and the scope of their work is global. All of the posts from the dataset are made on the social media platform Instagram. Moreover, most of the posts are commercial, e.g., brand or product promotion, and mostly there is one post per influencer.

Table 2 Data description

Variable	Description
ID	ID of the influencer
Content	Caption (text accompanying the post underneath
Reach	Number of followers of the influencer
Hashtags	Tags written in the caption of the picture
No_comments	Number of comments on the picture
No_Likes	Number of likes on the picture
General	Text description of the picture in forms of object and themes tags generated by Clarifai software

4.2 Data pre-processing

Before diving into text cleanup, the dataset was cleaned from all observations that contained missing values in the column General instead of object tags. Since this variable is the main one used in this paper, observations that did not have a text description were unusable. Second, duplicated observations based on influencer ID and image description were also removed. Finally, the images containing the word “giveaway” in the variable caption were also removed. This is because giveaway posts typically ask for followers and a broader audience to like, comment, tag others and reshare content posted to be eligible for a giveaway. This leads to posts that have artificially increased engagement and are not the focus of this paper. The author’s opinion is that these posts are biased and thus are removed. These steps lowered the number of observations from 355 188 to 228 317.

The primary variable used in this paper is named General. This variable gives a text description of the image in the form of objects and theme tags. The object tags are separated by non-meaningful tag IDs, classes, numbers, and interpunctions. Thus, the first step was to remove all the above-mentioned nonsensical text. The object tag was nested in the following pattern: *name: “object tag “, value:,* thus recognizing and extracting the meaningful tags was straightforward. Moreover, since object tags generated by Clarifai software can contain multiple words, these were concatenated together by eliminating the space between (e.g., sports equipment -> sportsequipment). After this, the column general consisted only of object tags separated by space. Finally, all object tags were set to lowercase since the software sometimes generated the same words with capital letters and sometimes without (e.g., “Christmas” and “christmas”).

Stemming was not applied because there was no need for it. Most of the object tags are generated in the forms of nouns, and there are very few adjectives if any. Additionally, for the process of sentiment, unstemmed words are required due to the use of dictionaries. Moreover, Clarifai software does not generate stop words. In any exceptional cases where they would get generated, they would be meaningful. Thus, no stop words were removed. The same goes for negations; they are rarely generated. However, there are instances such as the object tag “no person” most probably indicating an absence of humans in the image. Therefore, it is essential to retain this information by concatenating it to “noperson” as mentioned in the previous paragraph.

After cleaning the dataset and extracting the object tags, 4 573 218 occurrences of objects were counted. In contrast, there were only 4 944 unique ones in those occurrences. The author decided to remove terms appearing less than 0.1%, resulting in about 80% of terms being removed from the dataset. In typical cases, the rule of the thumb is removing about 1% of infrequent words. In the case of this dataset, the author decided it was too much, as infrequent words are one of the focuses of this paper in quantifying image interestingness.

4.3 Data exploration

This section gives a high-level overview of the dataset, exploring the main variables containing image objects and supporting variables.

The most frequent objects found in images in this dataset are mainly associated with the presence of people. Terms such as “people”, “woman”, “adult”, “portrait”, “one”, “girl”, etc., are objects with the highest frequencies, occurring more than 50 000 times. This corresponds with finding of Hu et al., (2014) that the biggest group of photos on Instagram contain people. On the other hand, terms such as “angle”, “anemone”, “anchovy”, “anatomical”, “alternativeEnergy”, “abascus”, etc., are objects with the least frequency, appearing only once in the whole dataset. Figure 3 and Figure 4 below show the top 20 most and least frequent objects, respectively. As least frequent objects are not helpful since their sample size is too small, they are removed in the pre-processing step before feeding it to the model.

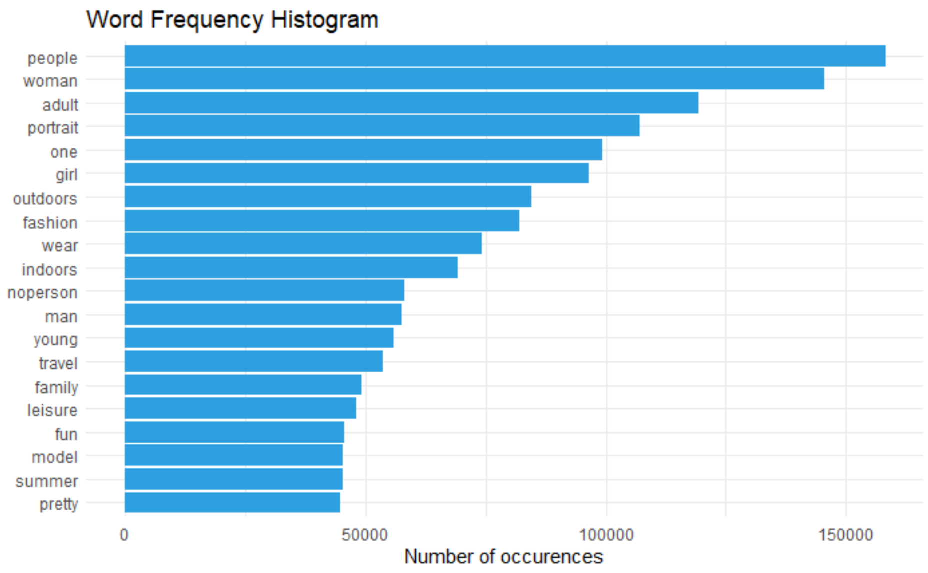


Figure 3 Most frequent objects

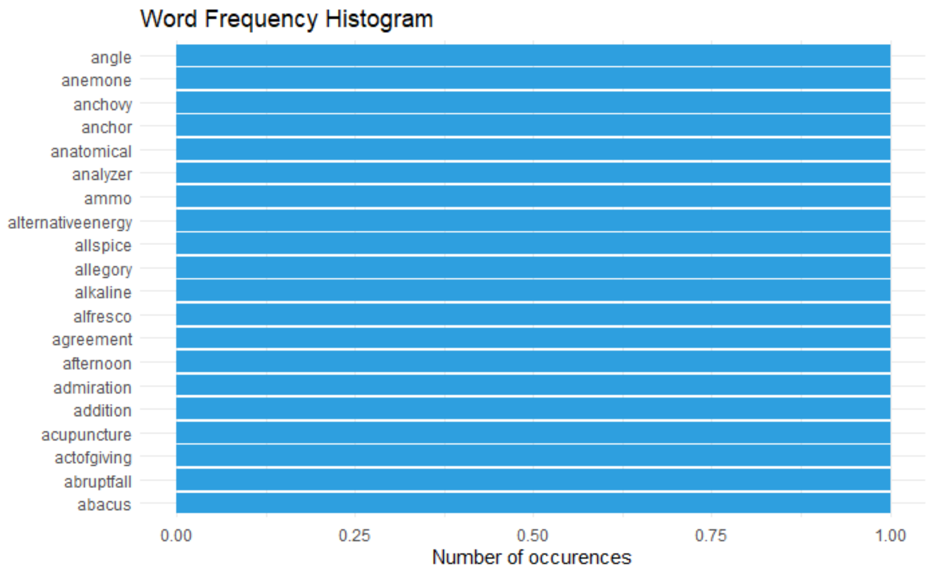


Figure 4 Least frequent objects

Next, multidimensional scaling was applied (MDS) to uncover the topics appearing in the data. This gives a high overview of topics found in images and anticipates groups that are later uncovered by unsupervised modeling techniques. The MDS graph is plotted in Figure 5. We can identify 8 theme groups: summer, city, fitness/sport, fashion, education, web design, medicine/health, food and lifestyle. These groups seem intuitive and similar to groups found on Instagram by other researchers (Bakhshi et al., 2014).

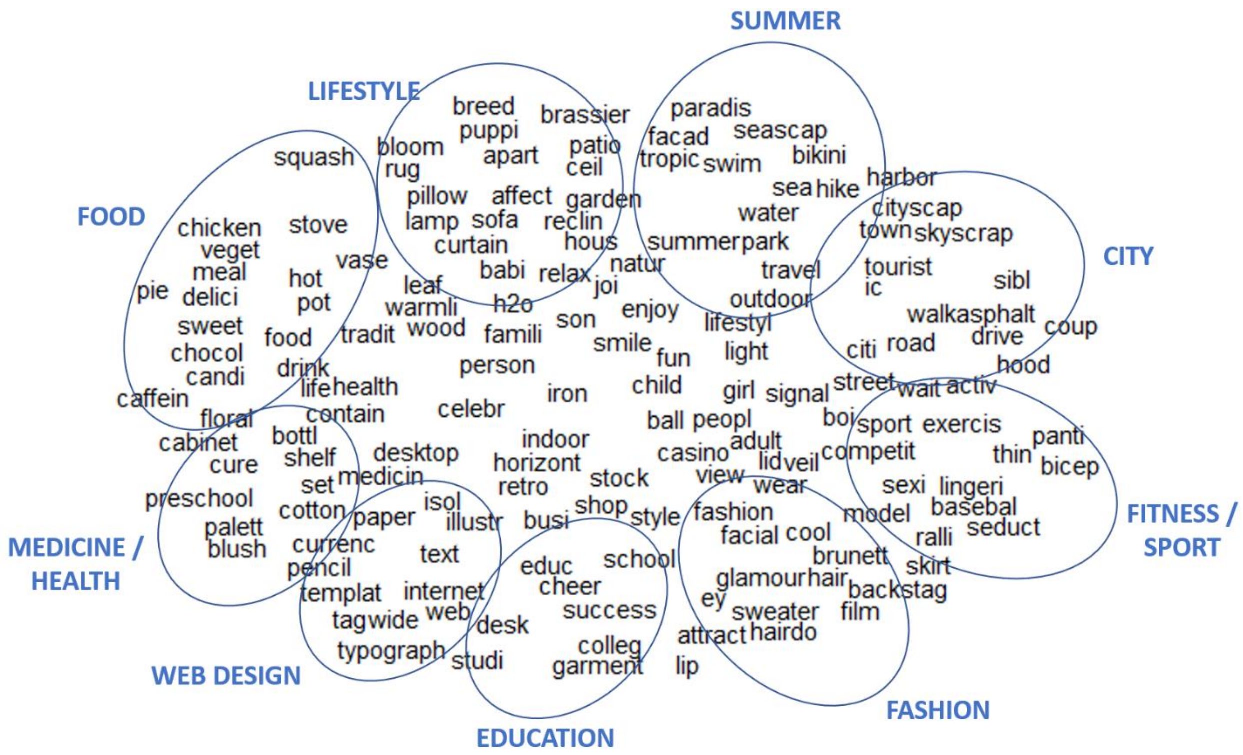


Figure 5 MDS plot

Next, looking at the quantitative variables, the dataset represents a wide range of images from least popular with only zero likes to very popular with half a million likes. An even higher spread can be seen in the variable reach, representing the number of followers the post's author had at the moment of scraping the data. Even though this dataset is acquired from the company that employs influencers, their popularity status differs significantly, going from only one follower to more than 3 million. Moreover, the number of comments has the smallest variation as comments are more rarely occurring than likes.

Table 3 Summary of quantitative variables in the dataset

Variable	Min	1st Quantile	Median	Mean	3rd Quantile	Max	SD
No_likes	0	17.8	611	2142.7	2 057	675 198	6006.182
No_comments	0	1	15	49.74	53	114 694	303.4747
Reach	1	11.743	33 194	82 473	101 451	3 397 141	147481.6

5 Analysis & Results

5.1 Global Vectors for Word Representation: Results

Global Vectors for Word Representation (GloVe) was conducted as one of the methods of extracting visual features and calculating similarities between image tags. These similarities are then used to construct the custom measure representing complexity. Visual features take the form of word embeddings as an output of the model. At the same time, the similarities are computed with cosine distance. The package used is text2vec (Selivanov et al., 2020, p. 2).

Data used as input was cleaned data, whereby cleaning is explained in the data preparation section of the paper. As unordered data is unconventional to use as input, relying on default values might not be the best approach. Thus, the parameters were tuned in a combination of using manual tuning and rules of thumb. Tuning was done in reference to analogies “travel” – “vacation” and “celebration” + “groom”. Additionally, the author looked at similarities of the following words: “woman”, “man”, “travel”, “fashion”, “sport”, “concert”, and “vehicle”. The final model parameters are as follows: the context window was set to 8, meaning that the model looks at a total of 16 words around the focal word to derive its local context. Here, the focal word is not referred to as a focal tag in an image but as a focal word from the model methodology, explained in chapter 3.1.1. The dimensionality of the word vectors is set to 50. Additionally, this value provided satisfactory results in word analogies and similarities tests. The results of word analogies and similarities can be observed in Table 4 and Table 5 below:

Table 4 Top 10 most similar and dissimilar words for analogies "travel" - "vacation" & “celebration” + “groom”

“travel” – “vacation”			**“celebration” + “groom”**	
traffic	0.4589441		bride	0.7943048
subwaysystem	0.4574653		groom	0.7886823
road	0.4415924		ceremony	0.7823068

offense	0.4161271		wedding	0.7818548
pavement	0.3954203		marriage	0.7522444
street	0.3849828		celebration	0.7425671
production	0.3770684		bridal	0.6842078
bus	0.3745714		romance	0.6446242
outdoors	0.3686357		bouquet	0.6369261
accident	0.3650254		romantic	0.6363675
swimsuit	-0.3445989		machinery	-0.2817537
vacation	-0.3447774		order	-0.2887586
tropical	-0.3535199		abandoned	-0.3090480
island	-0.3736584		analogue	-0.3359703
idyllic	-0.3839817		navigation	-0.3392994
paradise	-0.4084024		electricity	-0.3628470
exotic	-0.4113117		waterfront	-0.3662583
resort	-0.4198008		device	-0.3776794
turquoise	-0.4508548		control	-0.3780110
balloon	-0.4635274		grinder	-0.3795991

The analogy presented in Table 4 above shows that GloVe learned analogies between words surprisingly well. It recognizes that when we take out “vacation” from “travel” it relates to vehicles and transportation systems as most similar words, which follows the human understanding and logic. The second analogy, “celebration” + “groom” also shows logical connections. The resulting words are related to the wedding and romance, while the most dissimilar words are far from being associated with it by common sense.

Table 5 Word similarities

	woman	man	travel	fashion	sport	vehicle
woman	1.0000000	0.9170865	0.6474519	0.7884652	0.3721537	0.5041284
man	0.9170865	1.0000000	0.6152984	0.7116686	0.4366892	0.5301425
travel	0.6474519	0.6152984	1.0000000	0.4474354	0.1836958	0.6232033
fashion	0.7884652	0.7116686	0.4474354	1.0000000	0.3919008	0.4085469
sport	0.3721537	0.4366892	0.1836958	0.3919008	1.0000000	0.3229582

vehicle	0.5041284	0.5301425	0.6232033	0.4085469	0.3229582	1.0000000

Word similarities from Table 5 above also show logical connections: “woman” is more similar than “man” to concepts that would generally be associated more with females, such as “fashion” and “travel”; “man” is more similar to “sport” and “vehicle” which would generally be associated more with males than females. In addition, logical groups are also observed, as “woman” is very similar to “man”, and “travel” to “vehicle”.

The author, therefore, deems this model satisfactory and proper to be used for further computations of complexity. The embeddings themselves will not be discussed in this paper, as they are very hard to interpret and therefore are outside of the scope of this thesis.

Finally, to use the word vectors matrix in K-means to obtain clusters of images, the word vectors must be summarized to image embeddings. This was done using the mean of word vectors. Other approaches are also possible but were not tested in this paper.

5.2 Non-Negative Matrix Factorization: Results

Non-Negative Matrix Factorization (NMF) was performed to identify groups of images presented as topics. These groups would serve as visual features used in the final model to predict engagement. The package used was *NMF* (Gaujoux and Seoighe, 2020).

Due to the computational restraints, the number of topics (*k*) was chosen by looking at the results of a Singular Value Decomposition (SVD) matrix. This is a widely used alternative to cross-validation to establish the optimal number of topics. This fit is explored by Qiao, (2015). It was concluded that results are similar to those obtained using cross-validation of NMF itself. Based on the graph shown in Appendix, the number of topics was set to *k=25* in the final model. Because of the same computational restraints, the author used a deterministic method to initialize values instead of a random seed. Nonnegative Double Singular Value Decomposition (*nndsvd*) is a popular method that allows for intelligent and efficient one-time initialization of results (Gaujoux and Seoighe, 2020).

The interpretation of the resulting topics is given in Table 6 below. Graphs depicting all the topics can be found in the Appendix.

Table 6 Interpretation of NMF topics

	Topic name	Topic description
1	Portrait photography	portrait related words: one, portrait, adult, people
2	Architecture travel	architecture related words: architecture, building, sky, outdoors, city, town travel related words: travel, tourism
3	People indoors	people related words: people, family, woman, facial expression indoor related words: indoors, room, vertical, horizontal
4	Street photography	street related words: street, city, urban, outdoors, pavement, road
5	Indoor decoration without people	no people related word: no person indoor related words: wood, desktop decoration related words: color, luxury, food, decoration, paper, traditional
6	Digital art	digital related words: desktop, abstract, vector, graphic, symbol art related words: illustration, design, art, abstract, text, decoration
7	Children images	children related words: child, fun, cute, little, girl, people, baby, family, joy, son
8	Food images	food related words: food, delicious, meal, dinner, lunch, cooking, plate, vegetable, no person, dish
9	Summer travel	summer related words: water, beach, sea, ocean, seashore, summer, sand, sun travel related words: travel, vacation
10	Female fashion models (adult)	fashion related words: fashion, dress, fashionable, skirt female model related words: girl, woman, model, people, adult, portrait
11	Female fashion models (young)	fashion related words: fashion, casual, contemporary young related words: young, adolescent, cute female model related words: pretty, woman, looking
12	Glamour female models	glamour related words: sexy, glamour, pretty, beautiful, elegant female model related words: model, hair, fashion, woman, girl
13	Sport athletes competing	sport athletes related words: man, adult, woman, child, athlete competition related words: people, group, two, competition, group together
14	Recreation & leisure	related words: recreation, leisure, competition, exercise, lifestyle, athlete, fitness, sport

15	E-commerce	commerce related words: shopping, stock, shop, commerce, market, sale, option
16	Interior design	related words: furniture, room, house, seat, table, chair, window, interior design, inside
17	Female lifestyle models	female related words: beautiful, young, person, girl, adult, woman, people lifestyle related words: lifestyle, happiness, smile
18	Music festivals	related words: music, festival, musician, performance, singer, group, celebration, concert, movie, people
19	Career building	career related words: business, technology, internet, office, computer, communication career building related words: education, paper, text, achievement
20	Females enjoying summer	summer related words: summer, fair weather enjoyment related words: leisure, relaxation, enjoyment, fun, lifestyle female related words: woman, people, girl
21	Wear portraits (fashion)	wear related words: wear, fashion, lid, pants, fashionable, veil portrait related words: facial expression, portrait, adult
22	Couples in love	couple related words: togetherness, two, family, portrait, friendship, facial expression love related words: love, happiness, affection, enjoyment
23	Outdoor nature	outdoors related words: outdoors, summer, fall, landscape, fair weather nature related words: nature, park, tree, grass, leaf
24	Indoors seated people	indoor seat related words: sit, room, furniture, coffee, seat, sitting people related words: family, adult, people, woman
25	Transportation vehicles	related words: vehicle, transportation system, road, car, travel, wheel, drive, traffic, driver

Overall, the topics look well defined and can be interpreted. Specifically, NMF produced a distinction between adult and young female fashion models and between architecture and street photography. Furthermore, there is a clear distinction between different types of female influencers' content represented by topics 10, 12, 17, and 20 depicting female influencers in fashion, glamour, general lifestyle, and summer photography.

5.3 K-means: Results

K-means clustering was performed with a similar intention as NMF – to obtain groups of images presented as clusters. These groups would be further used as visual features. The input for K-means is the

GloVe similarity matrix constructed using the Cosine similarity measure. The similarity is calculated as an angle between two-word vectors. This way, the measure considers the direction of the vectors but not their magnitude. The idea is that if the GloVe produced correct and logical associations, clusters in K-means would also be logical and interpretable. This is also another way to check if the GloVe word similarities are suitable for constructing complexity.

The number of clusters (*k*) was chosen based on the elbow method and based on NMF topics, as these two are expected to yield similar clusters (topics) and thus, can be compared. The number was set to *k = 25*. This produced apparent, interpretable clusters described in Table 7 below. Wordcloud graphical representation can be found in Appendix, along with cluster sizes. This graph was also used to describe the clusters. To ensure the independence of random initialization, the number of centroid initializations (*nstart*) was set to *10*.

Table 7 Clusters produced by K-means based on GloVe similarity matrix

	Cluster name	Cluster description
1	Urban / city photography	urban, street, city, pavement, graffiti, luggage, step, umbrella
2	Music events	music, festival, man, group, people, musician, singer, many, event, performance, concert, stage
3	Healthcare	health, treatment, healthcare, medicine, bottle, care, soap, plastic, medical
4	Glamour female models	sexy, pretty, glamour, fashion, hair, woman, beautiful, young, look, nude, skin, attractive
5	Animals	animal, mammal, pet, cute, little, dog, puppy, fur, canine, sit, cat, breed
6	Shopping	shopping, stock, market, shop, commerce, sale, option, sell, foot, shelf, bag, footwear
7	Bedtime	people, woman, adult, family, indoors, room portrait, bed, pajamas, reclining
8	Landscapes	landscape, sky, scenic, sunset, daylight, mountain, river, dusk, dawn
9	Flora	nature, leaf, flora, flower, color, tree, wood, season, garden, bright, color, rose
10	Career motivation	technology, computer, office, telephone, laptop, vertical, school, serious, desk
11	Couples	adult, facial expression, couple
12	Digital art	business, illustration, text, internet, symbol, graphic, paper, facts, abstract, education, sign, vector

13	Fashion female models	model, girl, portrait, one, person, fashionable, dress, knitwear, scarf, skirt, sweater
14	Gifts and wraps	desktop, design, decoration, art, pattern, box, gift, retro, gold, ornate, focus, craft, card
15	Sport events	athlete, exercise, competition, sport, fitness, active, game, stadium, tennis, ball, soccer
16	Cozy breakfast	breakfast, noperson, chocolate, drink, fruit, cup, sugar, cake, coffee, cream, candy, pastry, baking, homemade
17	Summer leisure	carefree, freedom, woman, outdoors, summer, water
18	Summer beach vacation	water, vacation, beach, summer, ocean, travel, sea, sun, sand, tropical, leisure, bikini, exotic, island, coconut
19	Weddings	bride, wedding, groom, veil, painting, gown, hand, mask, marriage, human, bridal, engagement, jewelryband
20	Children	child, fun, joy, baby, love, son, togetherness, happiness, enjoyment, boy, affection, offspring, two, three
21	Architecture travel	architecture, building, old, tourism, tourist, tower, church, cityscape, monument, historic, museum, temple
22	Interior design	furniture, room, house, indoors, interior design, family, table, rug, stove, shower, wall
23	Road vehicles	vehicle, transportation system, car, drive, fast, wheel, automotive, hurry, driver, asphalt, sedan, race, hood, bike
24	Sitting in park	outdoors, park, lifestyle, smile, outside, bench
25	Food	meal, food, dinner, delicious, lunch, plate, dish, meat, nutrition, restaurant, healthy, tasty, nutrition, rice, beef

Clusters are very well distinguished and can be interpreted. Although there are similarities with topics produced by NMF, clusters produced by K-means based on the GloVe similarity matrix give surprisingly specific and distinct groups of images. There is a clear distinction between urban (street) and architecture photography; music and sports events; cozy breakfast food images and general food images; career motivation and digital art; flora and nature landscapes, etc.

5.4 Apriori: Results

Another method for identifying objects that often appear together was Apriori. Although this was not the focus of this research, it was interesting to see these rules applied to a case in point. A common problem with the Apriori algorithm is finding generally interesting, not straightforward rules or something that would be common sense. The top 10 rules produced by the method can be found in Table 8 below:

Table 8 Top 10 Apriori rules

Top 10 Apriori rules
infancy_innocence
performance_concert
performance_music
symbol_illustration
medicine_healthcare
scenic_landscape
scenic_outdoors
guitar_musician
narrow_travel
aerial_travel

Although these rules do not give an unexpected relationship, they provide more context than observing one term on its own. E.g., performance could relate to both musical or athletic performance, but relating it to music or concert gives it a more specific meaning.

5.5 Custom measures: Results

To better understand the custom measures, it is helpful to look at what kind of images score high/low on them. As the images in this dataset are not available and are very hard or impossible to find online, we look at their text description in the form of objects used in this paper.

Novelty

Since this measure is based on the frequency of objects in an image, it is expected to have low Novelty scores for images already identified as belonging to the biggest clusters in the sample. Object tags that are most frequently found in images that score high or low on novelty are presented below. From this table, we can see that images that score low on novelty are the ones that contain human objects, as well as popular themes such as fashion. On the other hand, images that score low on novelty are often without people, indicated by the tag “noperson”. Additionally, they are images of less popular categories on Instagram, such as food, interior design, etc.

Table 9 Words found in images that score high/low Novelty

Measure	High	Low
Novelty	Noperson, food, desktop, business, design	Woman, man, people, girl, portrait, adult, fashion

Unique object in an image

Table 10 with a couple of examples directly from the dataset helps illustrate this measure and better understand why some objects are unique in each image. Unique objects are bolded.

Table 10 Example images that scored high on Unique object in an image measure

Example 1	"food pizza cheese meal slice **people** cooking restaurant refreshment pie fruit tasty pepper hot cuisine epicure knife meat vegetable"
Example 2	"abstract **mouth** desktop art noperson color woman love texture reflection painting celebration decoration wear shape people symbol image"
Example 3	"people hand market woman adult creativity wall business man art desktop person chalkout money street symbol knife **pizza** face"
Example 4	"noperson garden nature bird outdoors grass food flower leaf poultry summer family stone rural farm color outside house **chicken**"

Example 1 shows “people” as an outlier in the rest of the tags, as they all stand for food, indicating that food is in focus. Thus, people are less expected. Example 2 shows tags hint at an image that does not have a human object in its focus, at least not a portrait. A “mouth” would be unexpected to see, but it could be a part of the painting in this case. Example 3 illustrates an image with tags related to people and careers, and “pizza” does seem as an odd one out.

On the other hand, example 4 shows the tag “chicken” as most dissimilar, but that is not obvious when looking at all tags. The image possibly represents a farm, with a house and yard and chicken would not be unexpected to see. However, chicken has a very low frequency in the dataset and can have multiple meanings, such as a living chicken or a chicken prepared as food.

Complexity

Complexity as a measure in this paper is represented as having dissimilar objects in one image. Thus, looking at the images as a whole helps give a better picture of what is considered complex in this dataset. Below are two tables representing examples of images that score high and low on this measure.

Table 11 Example of images that score high on Complexity

Example 1	"game square leisure chocolate victory dark competition intelligence"
Example 2	"flag fun noperson color art motley child mouth stripe sportsfan pride abstract desktop bridge bright"
Example 3	"time guidance clock precision noperson luck game number leisure achievement gambling business sport"
Example 4	“fence noperson iron steel chemistry gold web indoors winter desktop security art light design museum”
Example 5	“diagram chalkout noperson chalk chalkboard electricity graphicdesign business wine graph cooking symbol conceptual achievement balloon”

Table 12 Example of images that score low on Complexity

Example 1	“woman fashion people dress girl outdoors one summer wear portrait young pretty adult leisure beautiful lifestyle happiness enjoyment fun model”
Example 2	“adult people recreation girl lifestyle two woman man leisure happiness wear child boy outdoors group one facialexpression fun enjoyment portrait”
Example 3	“time clock watch precision analogue”

Looking at these examples of high complexity images, the meaning of object tags is so dissimilar that it is hard to understand what is happening in that scene as it appears to have multiple themes. Additionally, sometimes these images seem to tell a story more clearly, like in Example 4, where a person who took the picture was possibly visiting a museum. On the other hand, images that score low on complexity seem to portray one focal theme. All objects point to it, like in example 1 depicting fashion or in example 3 depicting a watch.

5.6 Assessing the effect on engagement: Multiple Linear Regression

A multiple linear regression model was used to identify important visual features and their effect on engagement.

The dependent variable represents engagement expressed by the number of likes divided by reach. Independent variables are first explored in clusters, grouped according to the theoretical literature. Then a variable selection is performed to construct the final model. The conclusions and recommendations are then written based on this. Independent variables are standardized to ensure that their effects can be compared. A logarithm of the dependent variable is taken to ensure the linear regressions' assumption on the normal distribution of the variables has been met. Additionally, to avoid logs of zero, the distribution is shifted for plus one.

5.6.1 Visual Interestingness

A summary of the model with features measuring visual interestingness is given below. Reach is used as a control variable. The theory states that an image might receive a high number of likes simply due to the high number of followers an influencer has. It also represents the popularity of that influencer that is strongly linked to social interestingness.

Table 13 Summary of coefficients for Multiple Linear Regression with features capturing Visual Interestingness

Feature	**Coefficient on likes per 1 follower (log)**
Intercept	-0.009*
log(reach + 1)	0.014*
Novelty	-0.012*
Complexity	0.004*
Uniqueness of an object	0.001*
Anger	0.0002
Anticipation	0.004*
Disgust	0.002*
Fear	-0.003*
Joy	-0.006*
Sadness	-0.001
Surprise[1]	-0.002*
Trust	0.002*
Negative	-0.0004
Positive	0.008*
Adjusted R^2	**0.024**

[1] * 5% significance level or lower

The coefficients in Table 13 indicate an increase or decrease in engagement, and due to the transformation of variables and standardization, their interpretation is challenging. However, the coefficients can still be compared to each other. This still gives valuable insights and allows for connection with literature.

All three custom measures of visual interestingness have been labeled as important. This corresponds with the literature review that states that intangible visual concepts such as uniqueness, novelty, and complexity are essential factors influencing visual interestingness. To reiterate, the novelty in this paper is observing the image as a whole and how it fairs against other images when it comes to depicting distinctive elements. Novel images are defined as images with objects that are rarely seen with respect to the global corpus; the uniqueness of a single object in an image is trying to capture an object that does not generally belong with other objects in that image (is not commonly found together), and as such is focused on the something rarely seen in an image itself. Lastly, complexity is capturing how dissimilar objects in a single image are with respect to their local and global corpus.

Novelty has a negative effect on engagement. Opposite to expectations from literature, the negative impact of novelty implies that the less familiar objects in an image are, the less engagement that image gets. Additionally, the novelty has the second strongest effect (after reach). Since this measure is highly correlated with words depicting humans (e.g., woman, girl, man, family, people), the measure may be capturing their positive impact on engagement more than it does capture the overall novelty of an image.

On the other hand, unique objects and complexity positively affect engagement, which aligns with the theoretical expectations for these two measures. Suppose an image has a unique object compared to other objects in that same image. In that case, it is more likely to get engagement, as it is seen as something interesting. Since this is an opposite effect of novelty that focuses on the whole image, novelty may be further researched in a more local context: the rarity of an object depends on its environment. Next, if an image has multiple dissimilar objects present, meaning that it is complex in its composition, it is likely to get more engagement as those images are found to be more interesting.

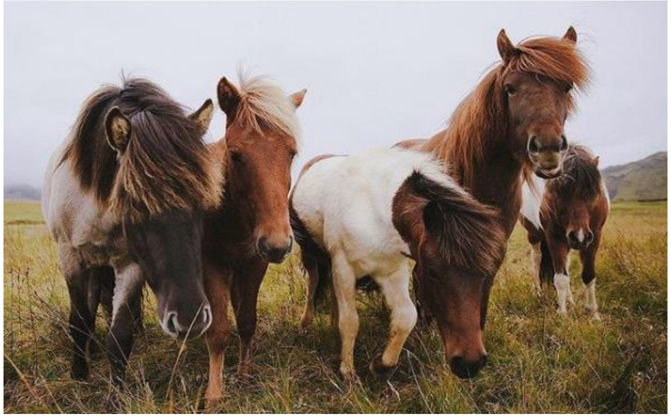


Figure 6 mammal, animal, cavalry, pasture, farm, grass, domestic, hayfield, field

Figure 6 represents an image taken as an example from the dataset. This image represents wild ponies captured up close in a meadow. In the Instagram environment of the dataset, images with people represent a significant group. Thus, this type of image can be considered novel in the global environment.

Most of these features are deemed significant by the model regarding valence and emotions. Interestingly, disgust drives positive engagement. Images that score high on disgust are mainly because of the keyword “dirty”. Although, in general, this term might be tied to a feeling of disgust, this is not necessarily the case in the Instagram environment. For example, a subject might have gotten dirty because of some activity, hinting that some exciting story might be happening in the image. An example of one such image that scores high on disgust is Figure 7.

Figure 7 nature, dirty, old, tree, outdoors, color, summer, art, retro, abstract, travel, soil, vintage, antique, wood, leaf, pictureframe

Joy and surprise appear as significant negative factors, but this is due to the inclusion of valence, which already contains their sentiment. When valence is excluded, both joy and surprise are found insignificant.

5.6.2 Clusters

A summary of the model with features representing themes or the picture's context is given in Table 14 below. The reference cluster for this model is Cluster 11 and was excluded to avoid multicollinearity. This cluster is the biggest, and it represents female fashion models.

Table 14 Summary of coefficients for Multiple Linear Regression with features capturing Visual Interestingness

Feature	**Coefficient on likes per 1 follower (log)**
Intercept	-0.012*
log(reach + 1)[2]	0.011*
Gifts and wraps	-0.009*
Interior design	-0.008*
Career motivation	-0.007*
Healthcare	-0.006*

[2] * 5% significance level or lower

Music events	-0.005*
Food	-0.005*
Cozy breakfast	-0.004*
Road vehicles	-0.004*
Urban / city photography	0.003*
Weddings	-0.003*
Flora	-0.003*
Local parks	0.002*
Summer leisure	0.002*
Summer beach vacation	0.002*
Shopping	-0.002*
Bedtime	-0.002*
Landscapes	-0.002*
Digital art	-0.013*
Glamour female models	0.001*
Animals	0.001*
Couples	-0.001*
Sport events	0.001*
Children	-0.001*
Architecture travel	-0.0002
Adjusted $\boldsymbol{R^2}$	**0.03**

Since cluster variables are binary, their effects should be interpreted regarding the reference cluster left out. E.g., Urban/city photography on average, has a higher engagement score per follower than the fashion female models cluster. Table 14 shows all clusters fall into the 5% significance category except for images depicting architecture. As expected, due to the dominance of the reference category, many clusters have negative coefficients except for: Urban/city photography, Glamour female models, Animals, Sports events, Summer leisure, Summer beach travel, and images depicting happy people in parks. Out of these, on average, Urban / city photography has the highest engagement score per follower, followed by 2 clusters related to summertime.

On the other hand, images from the category, gifts, interior design, career motivation, and healthcare are some with the strongest negative effects. They all score low on novelty compared to other clusters, indicating that these images could be more niche types in the sample corpus. Additionally, they appear to indicate the absence of humans, and as such, they could be not well received in an environment where human images have the highest engagement.

5.6.3 Topics

A summary of the model with topics as features is given in the Table 15 below.

Table 15 Summary of coefficients for Multiple Linear Regression with features capturing Topics

Feature	**Coefficient on likes per 1 follower (log)**
Intercept	-0.0099*
log(reach + 1)	0.011*
Architecture travel	0.005*
Glamour female models	0.005*
Career building	-0.005*
Digital art	-0.005*
Female fashion models (adult)	0.004*
Recreation & leisure	0.003*
Shopping	0.003*
Females enjoying summer	0.003*
Summer travel	0.003*
Female lifestyle models	0.002*
Indoors seated people	0.002*
Children images	0.002*
Sport athletes competing	0.002*
Street photography	0.002*
Interior design	-0.002*
Indoor decoration without people	-0.002*
Couples in love	0.001*
Portrait photography	-0.001*
People indoors	-0.001*
Food images	-0.001*
Outdoor nature	0.0005
Music festivals[3]	0.0003
Wear portraits (fashion)	-0.0003
Female fashion models (young)	-0.0002
Transportation vehicles	-0.0002
Adjusted R^2	**0.035**

[3] * 5% significance level or lower

The defined topics are like clusters and represent a theme or context of the picture. Although small, the explanatory power is still 14% higher with included topics versus clusters. On the other hand, observed topics seem to be less specific than clusters obtained through K-means.

Almost all topics depicting female models/influencers are significant with a positive effect on engagement except for Female fashion models (young). The topic with the strongest positive effect is Architecture travel, while the strongest negative effects come from the topics of Digital art and Career building.

5.6.4 Unigrams

A summary of the model with 20 unigrams representing objects in a picture is given in the Table 16 below. The objects were chosen based on the maximum frequency of their appearance in the dataset.

Table 16 Summary of coefficients for Multiple Linear Regression with features representing objects in a picture

Feature	**Coefficient on likes per 1 follower (log)**
Intercept	-0.009*
log(reach + 1)	0.012 *
Travel	0.007 *
Noperson	-0.005*
People	0.004*
Model	0.003*
Adult	-0.003*
Family	0.002*
Pretty	0.002*
Girl	0.002*
Woman	0.002*
Indoors	-0.002*
Young	-0.002*
Outdoors	0.001
Fashion	0.001*
Wear	0.001
Man	0.001*
Leisure	0.001*
Fun	0.001*
Summer	0.001*
Portrait	-0.001*
One	-0.0002

Adjusted R^2	**0.026**

People, woman, girl, man, family, model and pretty suggest the presence of people in pictures, and all have a significant positive effect on the engagement rate. On the other hand, noperson indicates the absence of people in the photo. This suggests that although the photos with people are not novel or unique, they lead to an increased engagement rate on the posts versus the photos without people. Next, photos taken indoors have a significant negative effect versus the outdoors, which has a positive, but insignificant effect. This insignificance might be due to the other variables that suggest they were taken outdoors, such as summer and travel. Terms with significant negative effects “adult”, “portrait”, “one” all suggest to refer to either “woman” or “man” and thus, their correlation might be the reason behind the effect.

5.6.5 Bigrams

Table 17 shows a summary of the multiple linear regression model with object bigrams. These bigrams are generated with the Apriori algorithm and top 10 rules were selected and are presented here.

Table 17 Summary of coefficients for Multiple Linear Regression with features representing combinations of objects as bigrams

Feature	Coefficient on likes per 1 follower (log)
Intercept	-0.012*
log(reach + 1)	0.014*
symbol_illustration	-0.01*
medicine_healthcare	-0.002*
narrow_travel	0.001*
aerial_travel	0.001
infancy_innocence	-0.001*
performance_concert	-0.001*
performance_music	-0.001*
scenic_outdoors[4]	0.0005
scenic_landscape	-0.0005
guitar_musician	0.0003
Adjusted R^2	**0.01**

[4] * 5% significance level or lower

Interestingly, almost all unigrams with significant coefficients have a negative effect, except for narrow_travel. This bigram suggests the narrow field of vision or narrow roads, places, etc., that might be interesting to people due to the perspective and depth of field these images create. On the other hand, although the literature has found that cuteness, defined by the words cute, child, innocence, etc., has a positive impact on social interestingness, this model suggests otherwise. Since the dataset comes from the pool of influencers hired by a single company, not many of them may post pictures of children or babies. Thus, these images are not very popular in the data sample used in this paper. Other negative bigrams with significant effects seem like they do not have people depicted, at least not in a straightforward way as having an object, woman, or man. This could further support the conclusion that people in pictures drive engagement.

5.7 Multiple Linear Regression: Variable selection

Multiple combinations of features were tested, and their performance was measured. This section gives an overview of the variable selection process for the final model.

First, the starting point was a multiple linear regression model including features representing objects in images. Previous literature has found that these types of features influence Instagram engagement and thus are selected as a starting point. Since this thesis focuses on visual and social interestingness, we want to see if the explanatory power increases significantly from using only objects in pictures. Next, clusters and bigrams are added, and their performance is measured. The overview of all models can be seen in Table 18 below.

Table 18 Experiments with varied multiple linear regression models

	Variables included	Number of variables	*Adj* R^2	Increase vs. previous
1	Individual objects (terms)	21	0.02633	
2	1 + Visual interestingness	32	0.0292	**9.83%**
3	2 + clusters	57	0.03589	**18.64%**

4	3 + bigrams	67	0.03646	**1.56%**

5.7.1 Final Model

Variable selection is performed based on the found literature and the interpretation power of the features themselves. Unigrams describe essential objects in the picture with significant positive or negative effects. The influence of individual objects has been confirmed many times in the previous literature and is thus included in the final model. Since visual interestingness is one of the main focuses of this paper, all features reflecting this are selected. Next, the context or theme of the picture is described by features derived from K-means based on GloVe word similarities and NMF topics. Since clusters and topics derived this way describe similar themes, only one group is needed. Clusters from K-means have more specific themes, and the model explanatory power increased significantly by adding them. Thus, they are selected for the final model. Even though NMF Adjusted R^2 is slightly higher, the goal of this paper is not prediction but interpretation. Finally, bigrams did not seem to portray exciting relations. They were mostly constructed of two very similar terms describing the same concept. Since their inclusion also did not lead to a significant increase in explanatory power, they are excluded from the final model.

Table 19 Summary of the final model

	Feature	**Coefficients**
	(Intercept)	-0.013*
	log(reach + 1)	0.09*
Visual interestingness	novelty	-0.02*
	complexity	0.004*
	unique object	-0.0001
Emotion & Valence	joy	-0.005*
	trust	0.004*
	fear	-0.003*
	positive	0.002*
	surprise	0.002*
	anger	0.002*
	negative	-0.001*

	anticipation	-0.001
	disgust	-0.0005
	sadness	-0.0001
Unigrams	noperson	-0.01*
	model	0.004*
	travel	0.004*
	adult	-0.004*
	outdoors	-0.003*
	wear	-0.003*
	indoors	-0.003*
	fun	0.002*
	people	0.002*
	man	0.002*
	fashion	0.002*
	young	-0.002*
	summer	-0.002*
	portrait	-0.002*
	pretty	0.001*
	girl	0.001*
	family	0.001*
	leisure	0.001
	woman	-0.001
	one	0.0001
Clusters	Architecture travel	0.005*
	Summer beach vacation	0.005*
	Digital art	-0.005*
	Summer leisure	0.004*
	Animals	0.004*
	Urban / city photography	0.004*
	Local parks	0.004*
	Career motivation	-0.004*
	Cozy breakfast	0.003*
	Food	0.003*
	Sport events	0.003*
	Landscapes	0.002*
	Flora	0.002*
	Children	0.002*
	Music events	-0.002*
	Glamour female models	0.001*
	Couples	0.001*
	Interior design	-0.001*
	Gifts and wraps	-0.001
	Bedtime	0.0004
	Weddings	-0.0002
	Healthcare	-0.0001

	Road vehicles	-0.00002
	Shopping	0.00001
[5]	**Adj R^2**	**0.036**

Visual interestingness: custom measures

Novelty is still significant and with a negative effect. This is contrary to the expected effect theorized by (Berlyne, 1960). This could be because Instagram is a setting where people choose what content to follow. They are more likely to opt for more familiar people and content they generate based on familiarity theory. In addition, this measure mainly identifies images that are more often represented in the dataset, as it is based on the frequency of object tags in a global context. Capturing novelty in a local context (e.g., inside clusters) could yield different results. **The uniqueness** of an object in an image, measured by the highest distance of an object from others, is no longer found significant in the final model. Lastly, **complexity** expressed through the average dissimilarity of objects in an image is found to be significant and with a positive effect. It suggests that the more the objects in an image are dissimilar, the more likely the engagement will be higher for that image. This finding aligns with the idea theorized by (Berlyne, 1960) & (Silvia, 2005).

Valence & Emotions

Anger shows a significant positive effect on likes per follower. As anger is a high arousal emotion that drives people to action, this could be the reason for its positive influence. It is possible that in the context of Instagram, the images scoring high on anger have a cause that drives people's interest, as well as likes. It is not uncommon for influencers on Instagram to call on their followers for justice with images portraying protests of some sort. Terms found in images with high scores on this emotion are "man", "offense", "police", "accident", "vehicle", "calamity", and "street," which supports this explanation. Fear

[5] * 5% significance level or lower

is found to have a significant but negative effect on likes per follower. As anger is an emotion that drives action and confrontation, fear is an emotion that does the opposite, flight and avoiding. As such, it seems intuitive to negatively influence the engagement. Terms found in images that score high on fear are “people”, “man”, “gun”, “emergency”, “vehicle”, “weapon”, and “battle”. As can be observed, although these terms might describe similar scenes as ones that score high on anger, the distinction is that the images with fear show possible repercussions to individuals in the forms of words indicating weapons and fights.

Emotions of surprise and trust have significant positive effects. Images scoring high on surprise are mainly associated with “gift” and “celebration”, showcasing an event or occasion that induces positive emotions for the author and possibly their viewers. Positive and negative valence have significant positive and negative influences, respectively. This is intuitive as it is expected for positive content to stimulate positive emotions and negative the opposite (Berger and Milkman, 2012).

Unigrams

“People”, “girl”, “man”, “family”, “model”, “pretty” all have significant positive effects and relate to the presence of people, the main women in images. “Noperson” indicates that an image does not contain people and has the strongest significant negative effect out of all features. This suggests that people, mainly women, are significant subjects depicted in pictures. Their presence indeed leads to increased likes per follower, while their absence does the opposite. This is similar to the findings by (Bakhshi et al., 2014) that suggested presence of people's faces leads to more likes and comments.

Clusters

Overall, images with a hedonistic atmosphere depicting pleasure, leisure, indulgence, and enjoyment in life and nature are, on average, more popular than other, more action-oriented ones (e.g., career goals, web design). Image clusters depicting **travel (Urban / city photography, Summer leisure,**

Summer beach vacation, Architecture travel), **influencers (Glamour female models, Couples, Children)**, **nature (Landscapes, Local parks, Flora)**, **cozy breakfast, food** all have significant comparative positive coefficient against the biggest, reference cluster depicting female fashion models. Furthermore, image clusters such as **career motivation**, **digital art**, **interior design, and music events** are significant negative ones. Additionally, the clusters with positive effects also suggest people's presence. In contrast, those with negative ones do not, or at least imply that people are not the focus of the image. This further confirms the finding by Bakhshi et al. (2014) mentioned in unigrams. This could also be explained by the social interestingness theory, where the authors' popularity can be the success behind the visual content itself. Next, clusters depicting **children** and **animals** have a significant positive effect compared to the reference cluster. This might be found in the aesthetic appeal of cuteness, often described with terms such as "baby face." According to (Lorenz and Martin, 1971), this appeal stimulates a desire in others to take care of the subject in question. This desire might be potentially linked to visual interestingness but is, to this date, still unexplored (Constantin et al., 2019). Finally, clusters depicting summer travel have the strongest comparative positive effect, suggesting that this type of content is most popular on Instagram.

6 Conclusion and Recommendations

The results in the previous section suggest that the presence of humans in images drives higher engagement, and these types of images are most found on the platform. This is shown in features such as "girl", "man", and novelty. Novelty is highly correlated with the most common elements in the picture (woman, girl, man, family, people). Its negative effect on engagement also points to the conclusion that the presence of humans increases engagement. This is further confirmed by the element "noperson" which is found to have a negative effect on engagement and suggests the absence of people. However, more people do not necessarily mean more likes. Complexity is found significant with positive effect, indicating that presenting multiple objects with dissimilar meaning also helps increase engagement. The recommendation for marketeers is to pursue images with people but should introduce a story to the image to catch people's attention.

Moreover, the number of followers has been a significant variable in determining engagement. It shows that social interestingness plays a significant role in influencer marketing. Some posts could get many likes purely from the considerable audience size that follows the content creator. Marketers should do an ROI analysis to determine which influencers to work with, and keep in mind that the higher following does imply higher exposure that could, in the end, lead to higher engagement as well.

The biggest category of images posted on Instagram is female models/portraits. This corresponds with results from other research papers (Hu et al., 2014). Furthermore, enjoyment is a concept widely presented on Instagram. Images that depict its actors indulging in relaxing activities such as summer travel, couples, landscapes, cozy breakfast, and food increase engagement. Travel seems to be very popular, as a group of images depicting both urban travel photography and summer/landscape photography has the most substantial positive effect. Categories that depict less enjoyment and are more action-oriented activities, such as career goals, are not so popular. Recommendation for marketeers is to pursue images with influencers depicted indulging in relaxing activities, showcasing enjoyment and positive sentiment.

Additionally, images depicting children and animals have a strong positive impact on engagement. This could be explained by the appeal of cuteness that is often linked to both subjects. This appeal has been noted by Constantin et al., 2019 as a possible additional aspect of visual interestingness but is still unexplored.

Positive emotions and valence mostly positively impact engagement and vice versa. Interestingly, images depicting anger are found to impact engagement positively. Instagram is a platform where people often stand up for their beliefs with impactful images with strong messaging. It is not rare to find images that induce anger in a viewer to call to action. Often these images get high engagement through the number of likes and comments. However, content creators should be careful in choosing their images and caption words, as posts that emit fear negatively impact engagement.

Novelty shows that people prefer images that have more familiar objects in them. This can be explained with the well-known familiarity social theory, which was repeatedly proven true. It can hold even more in a controlled setting such as Instagram, where people can choose whom to follow and are more likely to follow content related to them. Since the dataset used in this research is provided by a company that hires influences, it is possible that novelty cannot be identified as a result of sample bias. Moreover, defining novelty in the future with respect to the local context might show better results.

Furthermore, trying to capture the presence of unique objects in an image did not show as significant in the final analysis. However, it is still worth exploring and finding other ways to capture this characteristic. Lastly, complexity showed that images with different and dissimilar objects tend to get more engagement. Thus, content creators are advised to be creative and showcase complexity in their images to attract higher engagement.

To summarize, this paper made an effort to quantify intangible concepts found in the literature to identify what makes an Instagram image interesting. Findings indicate that familiar pictures, mainly containing people drive more engagement than their counterparts. The power of people in images is seen even in different content themes, where the ones without human presence have lower engagement. However, introducing complexity to the content makes an image more interesting and helps increase engagement. This is likely due to the effect of storytelling transmitted through these images, where different objects have different meanings but create a story together. Additionally, images with positive sentiment and emotions are generally better received. However, powerful emotions that call to action can also resonate well with the audience.

7 Limitations and Further Work

The first limitation is the explanatory power of the model used, limiting internal validity. Low R^2 produced in this research is very likely due to the failure of including other relevant variables and the complexity of the topic itself. Unfortunately, there is no similar model found in the literature that could be used as a baseline. This implies that even though some of the variables have been found significant in earlier studies, some of their effects could be affected due to omitting some control variables. Thus, for further research, it is recommended to include a wider variety of variables, as this could potentially increase the model's explanatory power.

Moreover, the measures of interestingness in this paper are simple quantifications of very complex theoretical terms. Relying only on them is a good start. However, more quantifications should be explored and constructed to capture more information in the future. Novelty has been shown to have the opposite effect of the one expected by the literature review, but this could be due to the limitation of how the measure was constructed and/or the sample used. Exploring this measure in a local context, e.g., novel images in the category of summer travel or focusing more on the unique object in an image itself could yield different results.

This research was performed on the sample taken from a company that provides influencer marketing services. Although the data analysis showed influencers vary in their number of followers and likes on their post, there could be selection bias, as they come from the same company and might have similar behaviors. Additionally, the research did not have access to any information on the influencers themselves, such as gender, age, where they come from, etc. This kind of information can influence results, as it can also impact social interestingness talked about in this research.

Next, the object tags in this data originated from Clarifai software. Even with high reported accuracy, a margin of error can produce an incorrect description of images. Moreover, the software provided tags indicating sentiment, which relies highly on how the algorithm processes facial expressions, a very complex topic in computer vision.

Bibliography

Agrawal, R., Srikant, R., 1994. Fast Algorithms for Mining Association Rules. pp. 487–499.

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, CHI '14. Association for Computing Machinery, New York, NY, USA, pp. 965–974. https://doi.org/10.1145/2556288.2557403

Berger, J., Milkman, K.L., 2012. What Makes Online Content Viral? J. Mark. Res. 49, 192–205. https://doi.org/10.1509/jmr.10.0353

Berlyne, D.E., 1960. Conflict, arousal, and curiosity, Conflict, arousal, and curiosity. McGraw-Hill Book Company, New York, NY, US. https://doi.org/10.1037/11164-000

Bulmer, S., Buchanan-Oliver, D.M., 2006. Visual Rhetoric and Global Advertising Imagery. J. Mark. Commun. 12, 49–61. https://doi.org/10.1080/13527260500289142

Butterfield, D., Fake, C., Henderson-Begg, C., Mourachov, S., 2006. Interestingness ranking of media objects. US20060242139A1.

Casaló, L.V., Flavián, C., Ibáñez-Sánchez, S., 2017. Understanding Consumer Interaction on Instagram: The Role of Satisfaction, Hedonism, and Content Characteristics. Cyberpsychology Behav. Soc. Netw. 20, 369–375. https://doi.org/10.1089/cyber.2016.0360

Clarifai, n.d. Clarifai Computer Vision, NLP & Machine Learning Platform [WWW Document]. URL https://www.clarifai.com (accessed 6.1.21).

Computational Understanding of Visual Interestingness Beyond Semantics: Literature Survey and Analysis of Covariates: ACM Computing Surveys: Vol 52, No 2 [WWW Document], n.d. URL https://dl.acm.org/doi/abs/10.1145/3301299 (accessed 5.19.21).

Constantin, M.G., Redi, M., Zen, G., Ionescu, B., 2019. Computational Understanding of Visual Interestingness Beyond Semantics: Literature Survey and Analysis of Covariates. ACM Comput. Surv. 52, 25:1-25:37. https://doi.org/10.1145/3301299

Foss, S.K., 2004. Theory of Visual Rhetoric, in: Handbook of Visual Communication. Routledge.

Gaujoux, R., Seoighe, C., 2020. NMF: Algorithms and Framework for Nonnegative Matrix Factorization (NMF).

Gygli, M., Grabner, H., Riemenschneider, H., Nater, F., Van Gool, L., 2013. The Interestingness of Images. Presented at the Proceedings of the IEEE International Conference on Computer Vision, pp. 1633–1640.

Gygli, M., Soleymani, M., 2016. Analyzing and Predicting GIF Interestingness, in: Proceedings of the 24th ACM International Conference on Multimedia. Presented at the MM ’16: ACM Multimedia Conference, ACM, Amsterdam The Netherlands, pp. 122–126. https://doi.org/10.1145/2964284.2967195

Harmeling, C.M., Moffett, J.W., Arnold, M.J., Carlson, B.D., 2017. Toward a theory of customer engagement marketing. J. Acad. Mark. Sci. 45, 312–335. https://doi.org/10.1007/s11747-016-0509-2

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A K-Means Clustering Algorithm. J. R. Stat. Soc. Ser. C Appl. Stat. 28, 100–108. https://doi.org/10.2307/2346830

Highfield, T., Leaver, T., 2016. Instagrammatics and digital methods: studying visual social media, from selfies and GIFs to memes and emoji. Commun. Res. Pract. 2, 47–62. https://doi.org/10.1080/22041451.2016.1155332

Holbrook, M.B., Batra, R., 1987. Assessing the Role of Emotions as Mediators of Consumer Responses to Advertising. J. Consum. Res. 14, 404–420. https://doi.org/10.1086/209123

Hsieh, L.-C., Hsu, W.H., Wang, H.-C., 2014. Investigating and predicting social and visual image interestingness on social media by crowdsourcing, in: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Presented at the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4309–4313. https://doi.org/10.1109/ICASSP.2014.6854415

Hu, Y., Manikonda, L., Kambhampati, S., 2014. What We Instagram: A First Analysis of Instagram Photo Content and User Types, in: Eighth International AAAI Conference on Weblogs and Social Media. Presented at the Eighth International AAAI Conference on Weblogs and Social Media.

Jaakonmäki, R., Müller, O., vom Brocke, J., 2017. The Impact of Content, Context, and Creator on User Engagement in Social Media Marketing. Presented at the Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2017.136

Lab, P.W., n.d. Visual Rhetoric: Overview // Purdue Writing Lab [WWW Document]. Purdue Writ. Lab. URL https://owl.purdue.edu/owl/general_writing/visual_rhetoric/visual_rhetoric/index.html (accessed 6.9.21).

Lee, D., Seung, H., 2001. Algorithms for Non-negative Matrix Factorization. Adv Neural Inf. Process Syst 13.

Lee, R.K.-W., Hoang, T.-A., Lim, E.-P., 2017. On Analyzing User Topic-Specific Platform Preferences Across Multiple Social Media Sites, in: Proceedings of the 26th International Conference on World Wide Web, WWW ’17. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, pp. 1351–1359. https://doi.org/10.1145/3038912.3052614

Lee_Seung_NMF.pdf, n.d.

Lorenz, K., Martin, R., 1971. Studies in Animal and Human Behaviour. Br. J. Philos. Sci. 22, 81–82.

Maher, M.L., n.d. Evaluating creativity in humans, computers, and collectively intelligent systems 7.

New SVD based initialization strategy for non-negative matrix factorization - ScienceDirect [WWW Document], n.d. URL https://www.sciencedirect.com/science/article/pii/S0167865515001762?casa_token=NHlzxhWfhhsAAAAA:q1somZTDHQnff8ZWiccxwdN2N9DWrRXr8PV9gRRGB94koH6e64UOdlMJrlGIAXQ-UkqPfd0MO20 (accessed 9.5.21).

NMF.pdf, n.d.

Padmanabhan, B., Tuzhilin, A., 1999. Unexpectedness as a measure of interestingness in knowledge discovery. Decis. Support Syst. 27, 303–318. https://doi.org/10.1016/S0167-9236(99)00053-6

Pennington, J., Socher, R., Manning, C., 2014. Glove: Global Vectors for Word Representation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Presented at the Proceedings of the 2014 Conference on Empirical Methods in Natural

Language Processing (EMNLP), Association for Computational Linguistics, Doha, Qatar, pp. 1532–1543. https://doi.org/10.3115/v1/D14-1162

Qiao, H., 2015. New SVD based initialization strategy for non-negative matrix factorization. Pattern Recognit. Lett. 63, 71–77. https://doi.org/10.1016/j.patrec.2015.05.019

Rietveld, R., van Dolen, W., Mazloom, M., Worring, M., 2020. What You Feel, Is What You Like Influence of Message Appeals on Customer Engagement on Instagram. J. Interact. Mark. 49, 20–53. https://doi.org/10.1016/j.intmar.2019.06.003

Romero, D.M., Galuba, W., Asur, S., Huberman, B.A., 2011. Influence and Passivity in Social Media, in: Gunopulos, D., Hofmann, T., Malerba, D., Vazirgiannis, M. (Eds.), Machine Learning and Knowledge Discovery in Databases, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 18–33. https://doi.org/10.1007/978-3-642-23808-6_2

Russell, J.A., Mehrabian, A., 1974. Distinguishing anger and anxiety in terms of emotional response factors. J. Consult. Clin. Psychol. 42, 79–83. https://doi.org/10.1037/h0035915

Schmidhuber, J., 2009. Driven by Compression Progress: A Simple Principle Explains Essential Aspects of Subjective Beauty, Novelty, Surprise, Interestingness, Attention, Curiosity, Creativity, Art, Science, Music, Jokes, in: Pezzulo, G., Butz, M.V., Sigaud, O., Baldassarre, G. (Eds.), Anticipatory Behavior in Adaptive Learning Systems, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 48–76. https://doi.org/10.1007/978-3-642-02565-5_4

Schwarz, N., 2000. Emotion, cognition, and decision making. Cogn. Emot. 14, 433–440. https://doi.org/10.1080/026999300402745

Scott, L.M., 1994. Images in Advertising: The Need for a Theory of Visual Rhetoric. J. Consum. Res. 21, 252–273. https://doi.org/10.1086/209396

Selivanov, D., models), M.B. (Coherence measures for topic, code), Q.W. (Author of the W.C., 2020. text2vec: Modern Text Mining Framework for R.

Siedlecka, E., Denson, T.F., 2019. Experimental Methods for Inducing Basic Emotions: A Qualitative Review. Emot. Rev. 11, 87–97. https://doi.org/10.1177/1754073917749016

Silvia, P.J., 2005. What Is Interesting? Exploring the Appraisal Structure of Interest. Emotion 5, 89–102. https://doi.org/10.1037/1528-3542.5.1.89

Stevenson, A., 2010. Oxford Dictionary of English. OUP Oxford.

text2vec.pdf, n.d.

Unnava, H.R., Burnkrant, R.E., 1991. An Imagery-Processing View of the Role of Pictures in Print Advertisements. J. Mark. Res. 28, 226–231. https://doi.org/10.1177/002224379102800210

What Are Likes Worth? A Facebook Page Field Experiment - Daniel Mochon, Karen Johnson, Janet Schwartz, Dan Ariely, 2017 [WWW Document], n.d. URL https://journals.sagepub.com/doi/full/10.1509/jmr.15.0409?casa_token=AnFiGd5Ln28AAAAA%3A-ul8Ybp35JdH44DrWLHNDNrYyghuA50DgOFZWUF5zkTcENhhpy3JFE68ODVW8X2cZpf1hdw9E3372A (accessed 6.14.21).

What We Instagram: A First Analysis of Instagram Photo Content and User Types | Proceedings of the International AAAI Conference on Web and Social Media [WWW Document], n.d. URL https://ojs.aaai.org/index.php/ICWSM/article/view/14578 (accessed 6.1.21).

White, L., Togneri, R., Liu, W., Bennamoun, M., 2015. How Well Sentence Embeddings Capture Meaning, in: Proceedings of the 20th Australasian Document Computing Symposium. Presented at the ADCS ’15: The 20th Australasian Document Computing Symposium, ACM, Parramatta NSW Australia, pp. 1–8. https://doi.org/10.1145/2838931.2838932

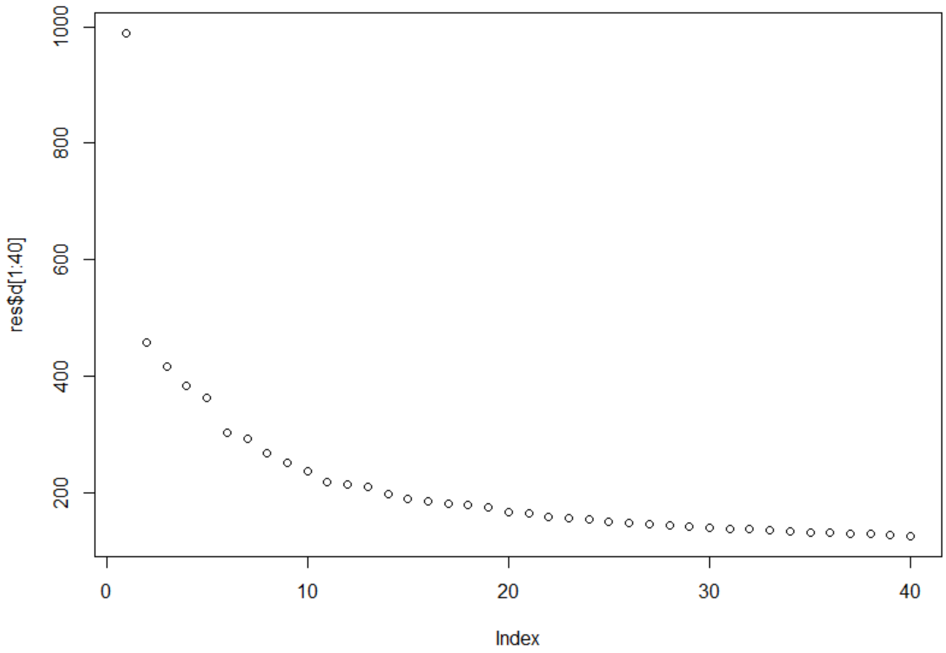
Yu, H., Winkler, S., 2013. Image complexity and spatial information, in: 2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX). Presented at the 2013 Fifth International

Workshop on Quality of Multimedia Experience (QoMEX), pp. 12–17. https://doi.org/10.1109/QoMEX.2013.6603194

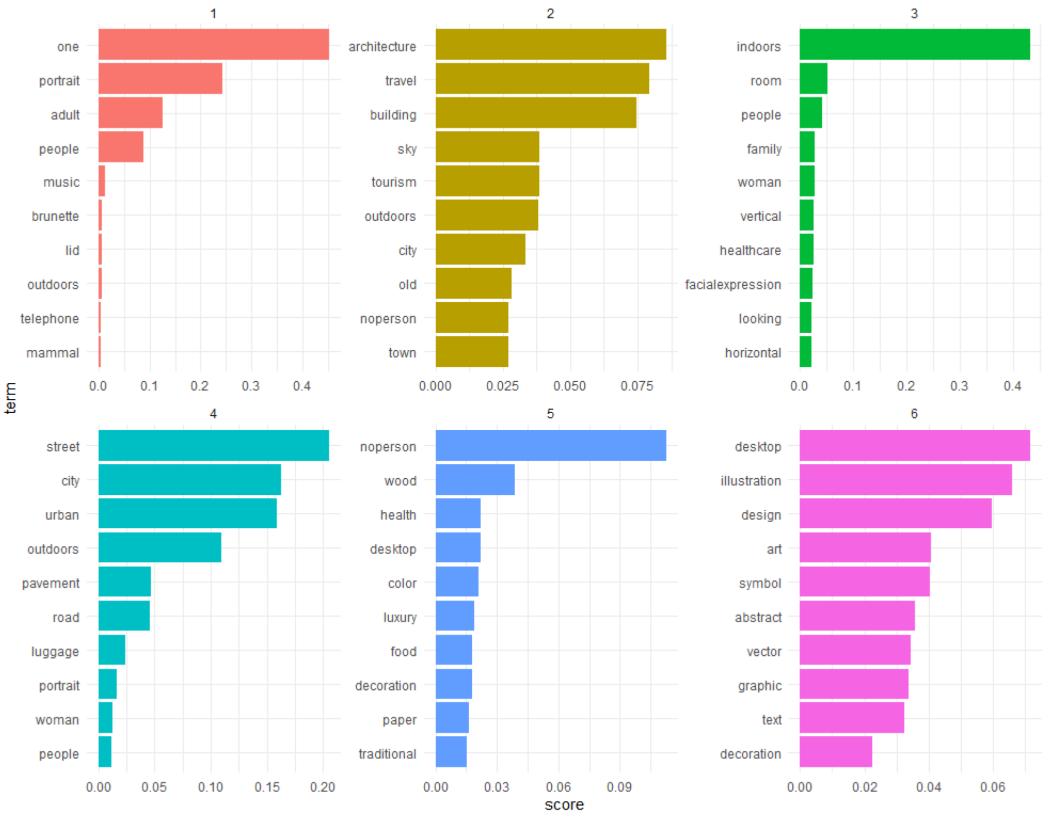
Zhao, B., Fei-Fei, L., Xing, E.P., 2011. Online detection of unusual events in videos via dynamic sparse coding, in: CVPR 2011. Presented at the CVPR 2011, pp. 3313–3320. https://doi.org/10.1109/CVPR.2011.5995524

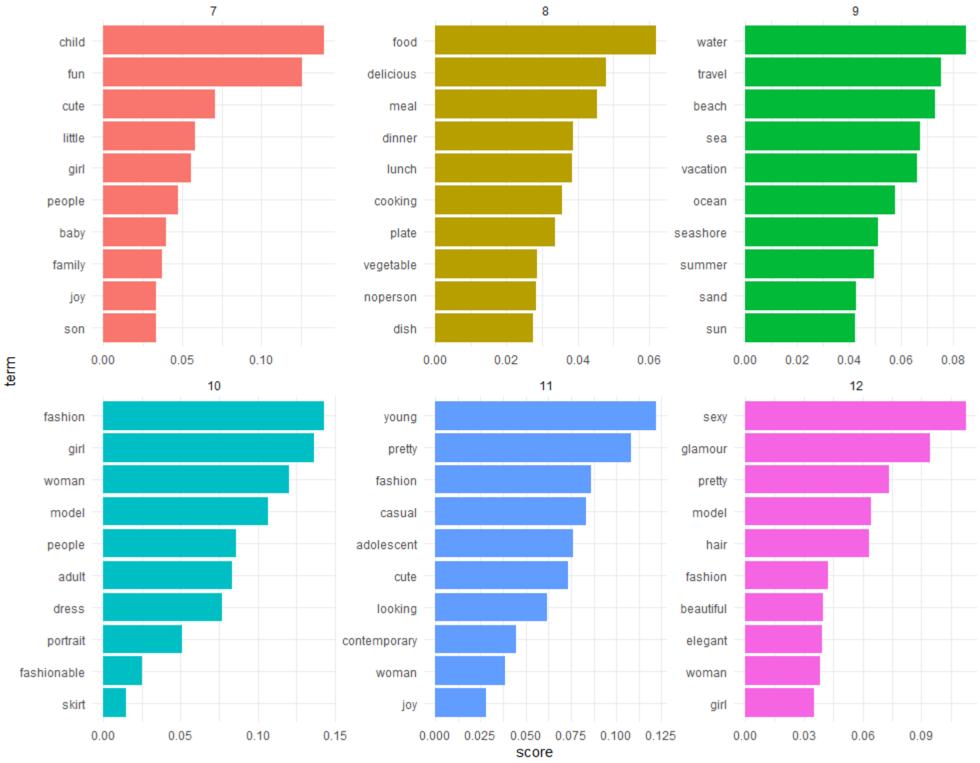
Appendix

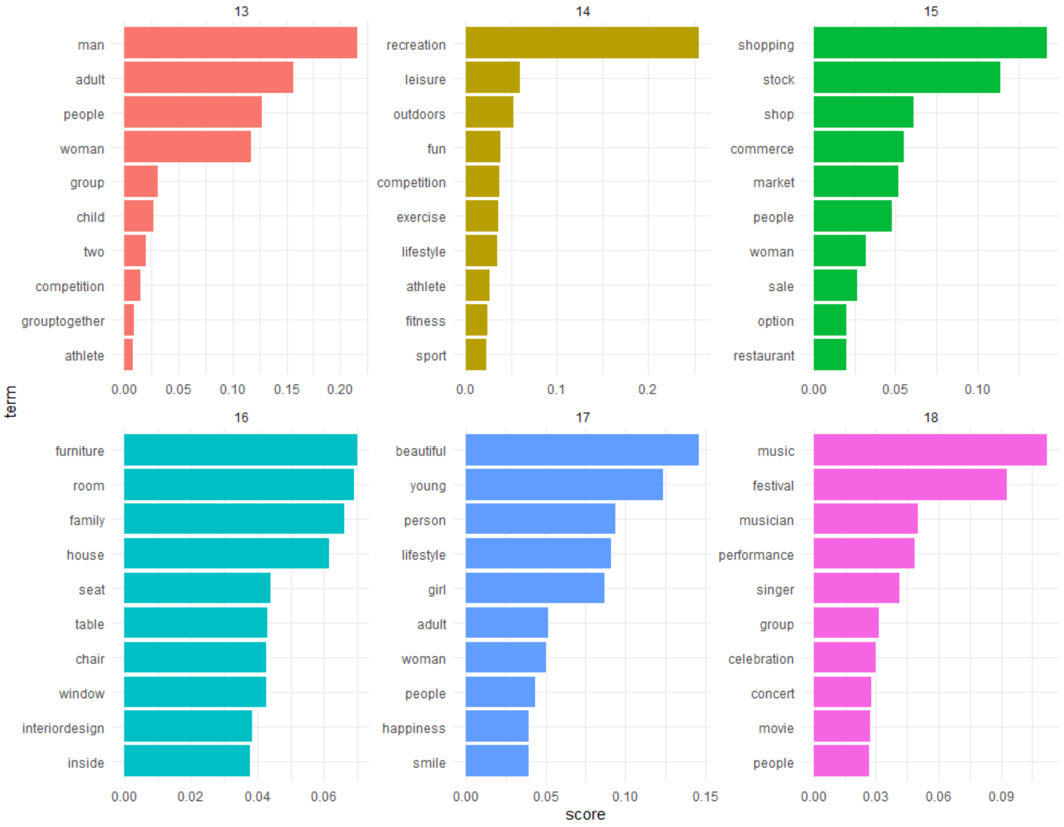
Appendix AP1: SVD: NMF Topic number selection

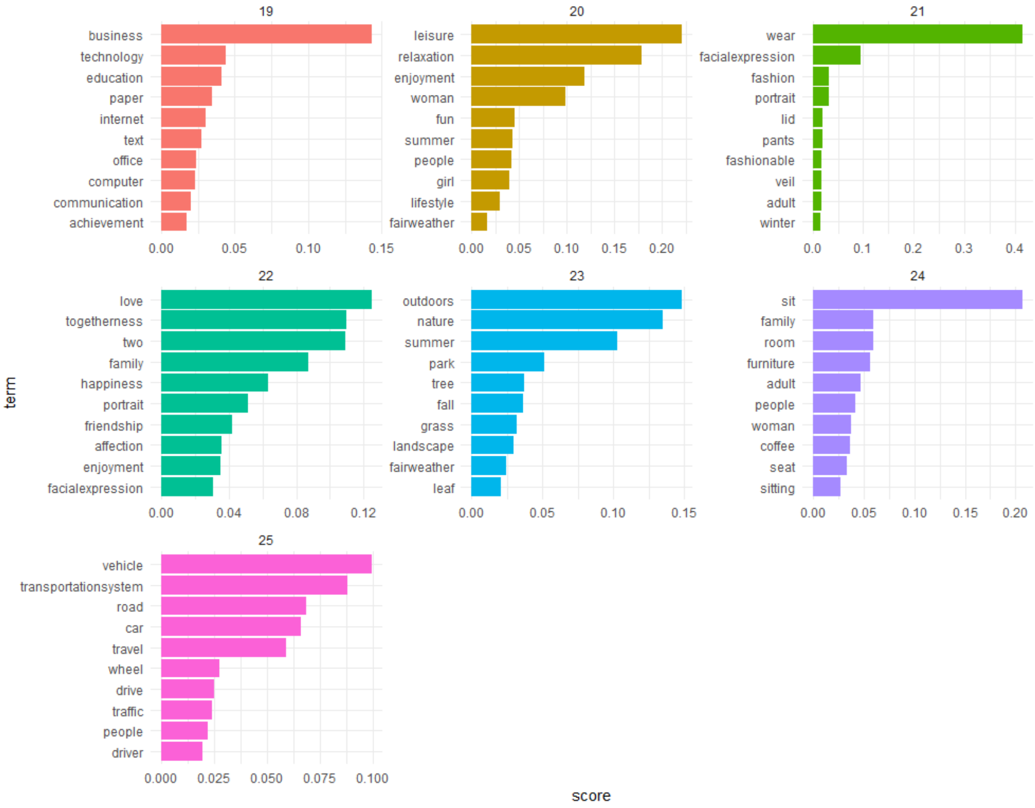


Appendix AP2: NMF Topics

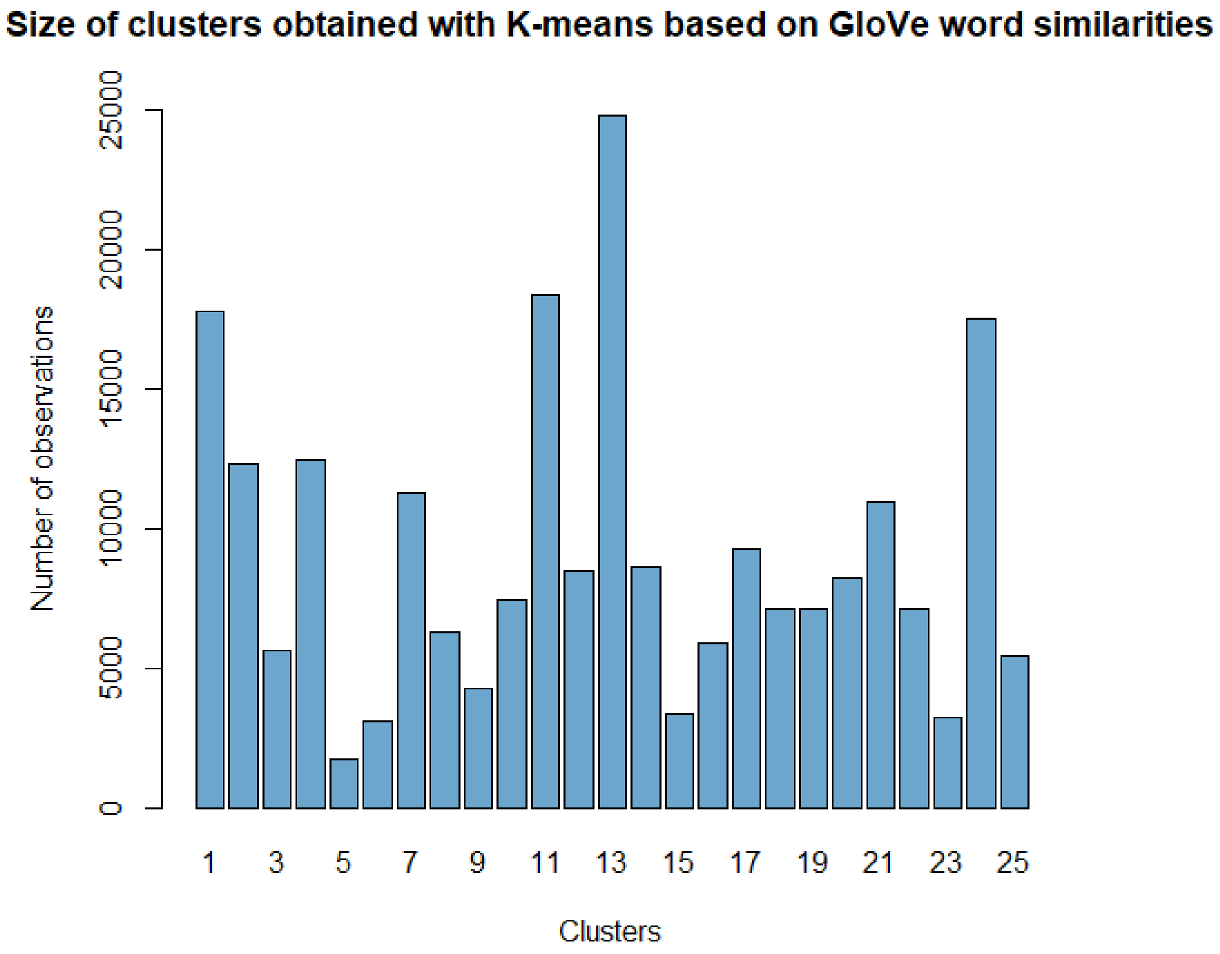




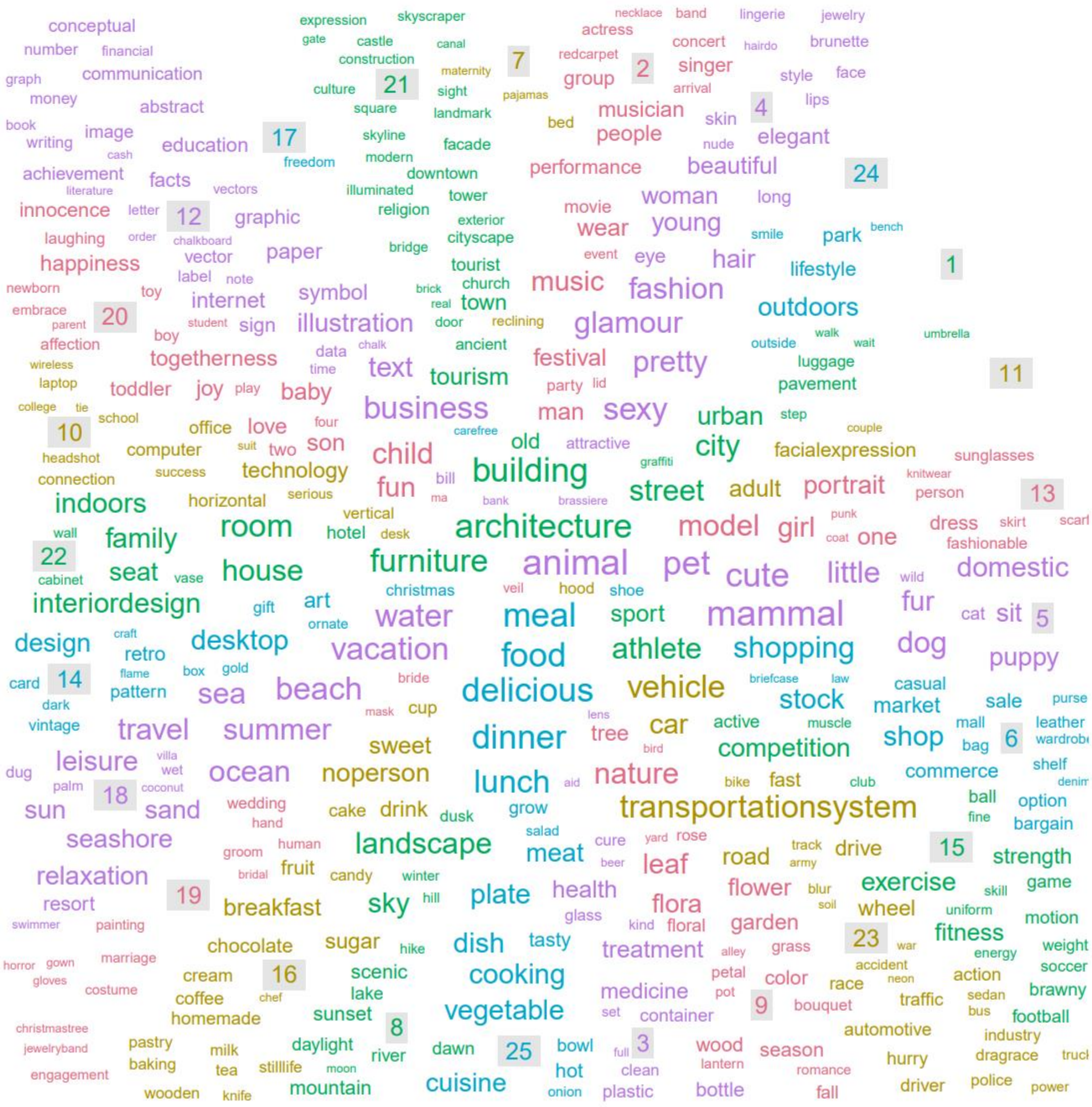

13
man
adult
people
woman
group
child
two
competition
grouptogether
athlete
0.00
0.05
0.10
0.15
0.20
14
recreation
leisure
outdoors
fun
competition
exercise
lifestyle
athlete
fitness
sport
0.0
0.1
0.2
15
shopping
stock
shop
commerce
market
people
woman
sale
option
restaurant
0.00
0.05
0.10
term
16
furniture
room
family
house
seat
table
chair
window
interiordesign
inside
0.00
0.02
0.04
0.06
17
beautiful
young
person
lifestyle
girl
adult
woman
people
happiness
smile
0.00
0.05
0.10
0.15
18
music
festival
musician
performance
singer
group
celebration
concert
movie
people
0.00
0.03
0.06
0.09
score


19
business
technology
education
paper
internet
text
office
computer
communication
achievement
0.00
0.05
0.10
0.15
20
leisure
relaxation
enjoyment
woman
fun
summer
people
girl
lifestyle
fairweather
0.00
0.05
0.10
0.15
0.20
21
wear
facialexpression
fashion
portrait
lid
pants
fashionable
veil
adult
winter
0.0
0.1
0.2
0.3
0.4
term
22
love
togetherness
two
family
happiness
portrait
friendship
affection
enjoyment
facialexpression
0.00
0.04
0.08
0.12
23
outdoors
nature
summer
park
tree
fall
grass
landscape
fairweather
leaf
0.00
0.05
0.10
0.15
24
sit
family
room
furniture
adult
people
woman
coffee
seat
sitting
0.00
0.05
0.10
0.15
0.20
25
vehicle
transportationsystem
road
car
travel
wheel
drive
traffic
people
driver
0.000
0.025
0.050
0.075
0.100
score

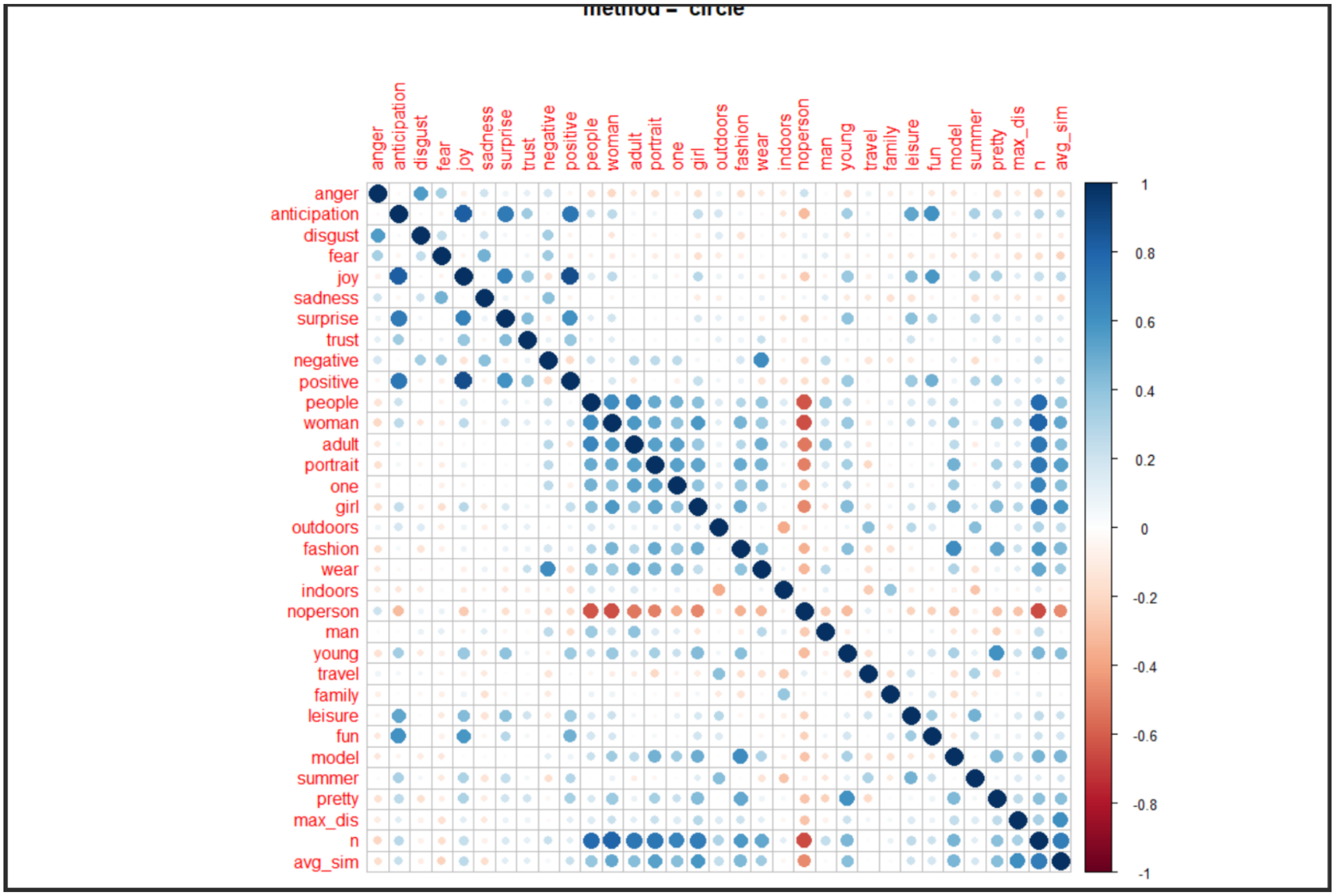
Appendix AP3: Cluster sizes



Appendix AP4: Wordcloud: K-means



Appendix AP5: Features correlation plot


method = "circle"
anger
anticipation
disgust
fear
joy
sadness
surprise
trust
negative
positive
people
woman
adult
portrait
one
girl
outdoors
fashion
wear
indoors
noperson
man
young
travel
family
leisure
fun
model
summer
pretty
max_dis
n
avg_sim