

Horizon 2020



Understanding Europe's Fashion Data Universe

Requirement Analysis Document

Deliverable number: D2.2

Version 4.0



Funded by the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 732328

Project Acronym: FashionBrain
Project Full Title: Understanding Europe's Fashion Data Universe
Call: H2020-ICT-2016-1
Topic: ICT-14-2016-2017, Big Data PPP: Cross-sectorial and cross-lingual data integration and experimentation
Project URL: <https://fashionbrain-project.eu>

| | |
|----------------------------|--|
| Deliverable type | Report (R) |
| Dissemination level | Public (PU) |
| Contractual Delivery Date | 30 June 2018 |
| Resubmission Delivery Date | 27 February 2019 |
| Number of pages | 23, the last one being no. 17 |
| Authors | Ines Arous, Mourad Khayati - UNIFR |
| Peer review | Alessandro Checco, Jennifer Dick - USFD Jennie Zhang - MDBS |

Change Log

| Version | Date | Status | Partner | Remarks |
|---------|------------|-------------------|----------------------|---------------------|
| 0.1 | 01/08/2017 | Draft | UNIFR | |
| 1.0 | 08/09/2017 | Final | UNIFR, USFD, Zalando | Rejected 15/03/2018 |
| 2.0 | 20/04/2018 | Resubmitted Final | UNIFR, USFD | Rejected 26/04/2018 |
| 3.0 | 21/05/2018 | Resubmitted Final | UNIFR, USFD | Rejected 15/10/2018 |
| 3.1 | 15/02/2018 | Revised Draft | UNIFR, USFD, Zalando | |
| 4.0 | 27/02/2019 | Resubmitted Final | UNIFR, USFD, Zalando | |

Deliverable Description

This deliverable will consist of an extended requirement analysis document detailing expert sourcing. It will collect requirements based on the data sources identified in WP1 (T1.1) and it will feed into WP3.

Abstract

In the context of the FashionBrain project, several datasets have been provided by our partners which will be used in order to perform different tasks such as extract key features from text and/or images, capture customers' preferences or predict new fashion trends as they emerge on social media and the blogosphere. In this deliverable, we describe these datasets (size, format, properties, etc.) and we explain how to identify and expand the sources of data relevant to the domain of fashion by means of crowdsourcing. We conducted a survey in order to assess the importance of social media in the online shopping experience. As a result, we are able to extract the most important social media platforms as well as the most important features in online shopping.

Table of Contents

| | |
|---|-----------|
| List of Figures | v |
| List of Tables | v |
| List of Acronyms and Abbreviations | vi |
| 1 Introduction | 1 |
| 1.1 Motivation of the Deliverable | 1 |
| 2 Data Description | 2 |
| 2.1 Fashion Retailer Data | 3 |
| 2.1.1 Datasets Analysis | 3 |
| 2.1.2 Relational Schema | 6 |
| 2.2 Fashion-tech startup | 6 |
| 3 Fashion Survey | 9 |
| 3.1 Presentation | 9 |
| 3.2 Results and Analysis | 9 |
| 3.2.1 Shopping Online/ Social Media Habits | 9 |
| 3.2.2 Influence of Bloggers and Social Media on Their Viewers . . . | 11 |
| 3.2.3 Comments/Reviews | 13 |
| 3.2.4 Fashion Content | 13 |
| 3.2.5 Other Forms of Content | 14 |
| 4 Conclusions | 16 |
| Bibliography | 17 |

List of Figures

| | | |
|-----|--|----|
| 2.1 | Relational model. | 6 |
| 2.2 | Fashion items classifier. | 8 |
| 3.1 | Platform preference by gender. | 10 |
| 3.2 | Platform preference for discovering clothing trend by age. | 11 |
| 3.3 | Frequency of purchased items discovered through Social Media. | 12 |
| 3.4 | Percentage of participants answering: When are you more likely to review or comment about a fashion product? | 13 |
| 3.5 | Interaction with fashion brands. | 14 |
| 3.6 | Content type preference. | 14 |
| 3.7 | Focus preference. | 15 |

List of Tables

| | | |
|-----|--|---|
| 2.1 | Datasets provided by Zalando. | 2 |
| 2.2 | Datasets provided by Fashwell. | 3 |

List of Acronyms and Abbreviations

| | |
|----------------|---|
| AR | Augmented Reality |
| CF | Crowdfunder (www.crowdfunder.com), micro-task crowdsourcing platform |
| CSV | Comma Separated Values |
| ID | Identifier |
| FB | Facebook (www.facebook.com) |
| JSON | JavaScript Object Notation |
| NLP | Natural Language Processing |
| OEM | Original Equipment Manufacturer |
| R&D | Research and Development |
| SKU | Stock Keeping Unit |
| SM | Social Media |
| TF-IDF | Term Frequency-Inverse Document Frequency |
| VR | Virtual Reality |
| WP | Work Package |
| YT | YouTube (www.youtube.com) |

1 Introduction

1.1 Motivation of the Deliverable

The fashion domain involves multiple industry players e.g., fashion designers, production chains, retailers, marketing companies, value-added services, Social Media (SM) outlets etc. Each player produces important pieces of information that can be used to provide a better end-user experience, improve individual processes and drive more profit. Fashion data has recently gained a lot of attention in various fashion applications e.g., clothing recommendation, clothing recognition, fashion trend prediction, etc. Yet, the relationship between datasets emanating from retailers and social media is rarely explored. One major impediment is the distribution of the data sources containing relevant products information. Once we uncover these datasources, we are still faced with the challenge of integrating them into a federated data repository. For instance, two online retailers might deal with common brands yielding redundant items in their datasets, thus, the need of eliminating redundant data. In the context of the FashionBrain project, the datasets provided by our partners will be used in order to i) automatically process text and image data and extract key features such as fashion items, ii) capture customers' preferences, iii) build a fashion-based recommendation engine and iv) predict new fashion trends as they emerge on SM and the blogosphere.

2 Data Description

This section depicts the datasets provided by the project's industrial partners. We give an overview on a retailer and on fashion-tech startup potential datasets. We will also describe the datasets content and focus on their future use in the context of FashionBrain. This section will be briefly extended by introducing the data integration technique that will be deployed. Table 2.1 summarizes the main properties of the datasets provided by Zalando.

| File name | Format | Size | Brief description | Date |
|--------------------------------------|--------|----------|--|------------|
| region1_items region2_items | csv | 22.2 MB | Items classification in Zalando anonymized tree with the amount and # of items sold | 31.03.2017 |
| region1_orders region2_orders | csv | 42 KB | Items' orders with the gross amount | 31.03.2017 |
| zalando-images-attributes-FB-release | csv | 1.01 GB | A detailed identification of products from different brands and their specialized features | 24.04.2017 |
| zalando-reviews-FB-release | csv | 355.2 MB | Users' comments on different products | 24.04.2017 |
| OEM_translations | csv | 5.2 MB | Mapping of item attributes rendered in different languages on the web site | 31.03.2017 |

Table 2.1: Datasets provided by Zalando.

Table 2.2 summarizes the main properties of the datasets provided by Fashwell.

| File name | Format | Size | Brief description | Date |
|--------------------|--------|--------|--|------------|
| dump-example-macys | json | 3.6 MB | A detailed identification of products from “Macys” with their specialized features | 30.03.2017 |
| dump-example-gap | json | 4.3 MB | A detailed identification of products from “GAP” with their specialized features | 30.03.2017 |
| products-example | json | 512 KB | an example of the json description of a given product | 30.03.2017 |
| fashion_tree_dump | csv | 21 KB | Fashwell Fashion tree | 04.05.2017 |
| products-schema | json | 3 KB | A schema explaining the content of the json files | 30.03.2017 |

Table 2.2: Datasets provided by Fashwell.

2.1 Fashion Retailer Data

This section provides a detailed description of the datasets provided by Zalando and that will be used in order to offer exclusive data services to their customers as they will be able to anticipate customer needs, and predict short and long term trends or frauds. The datasets are in the form of csv files containing each a table of attributes. This section will be divided into two parts. In the first part, the attributes of each file will be presented by providing their type and description. Next, we will analyse how these datasets can be used in terms of research. In the second part, we will summarize the relationship between the files in a relational schema.

2.1.1 Datasets Analysis

OEM_Translation.csv : In order to keep a tight relation with its clients, Zalando offers the possibility for users to see product attributes in their own language depending on the localization of the website. The OEM_Translation file includes items’ attributes rendered in different languages on the website. This file is used to map multilingual description to one fashion vocabulary.

- **Attribute_type (String)**: It can be the item’s color, color family or fabric. It can also give an indication about the appropriate season for the item.
- **Attribute_id (String)**: It gives a unique attribute identifier up to translation i.e. there are multiple instances of each attribute, but their translations have similar meaning in different languages.
- **Language (String)**: It can be French, English, German, Italian, etc.

- Translation (String): It is the translation of the attribute in the indicated Language.

Region*_item.csv (* is either 1 or 2) : The items dataset splits sales into categories, subcategories and sub-subcategories as well as the item's season (category-level 1-3).

- Date (String): The date when an item was sold. The date ranges from March 2013 to March 2015.
- Category_level 1,2,3 (Integer): This is the categorization of items in the Zalando category hierarchy. Example: A long dress could be under clothing, women, dress category. The mapping for "clothing" could be found in the first category with an id=0 for instance. The mapping for "women" could be found in the second category with an id=1 and "dress" could be found under the third category with an id =11. In this case, the long dress would be under the 0,1,11 category. The real values are anonymized and they provided values are given for the sake of illustration.
- Season (String): This attribute represents the clothing season for the item e.g., spring/ summer, autumn/ winter, no-season.
- Gross_amount (Float): This attribute represents the amount spent in total normalized by the mean over the column (anonymised revenue value)
- No_items (Float) No of items bought normalized by the mean over the column no_items could be less or larger than the revenue. It actually depends on the item itself. In fact, selling a lot of socks can generate less revenue than selling a much smaller quantity of suits.

Region*_orders (* is either 1 or 2) : A fashion retailer keeps track of item's orders that it can have a better handle on stock levels. This data was provided in the form of two files Region*_item and Region*_orders where * could be 1 or 2 corresponding to two regions where deliveries were made from 2013 until 2015. This data will be used in the prediction package (WP5) where we propose to help retailers to estimate their needed stock. More specifically, we will use the data for the following tasks: i) Prediction of the category with the largest revenue and ii) Prediction of the orders amount on a specific category.

- No_items (Float) Number of items bought normalized by the mean over the column
- No_orders (Float): Number of orders placed normalized by the mean over the column
- Gross_amount (Float): Amount spent in total normalized by the mean over the column

- Date (Float): Date when the order was made. It varies between March 2013 and March 2015

zalando-reviews-FB-release.csv : Comments and reviews are increasingly important part of the purchase journey for online consumers. Fashion retailers need to take full advantage of this important source of feedback and adjust their offers according to customer's desires. The file "zalando-reviews-FB-release.csv" contains user's feedback on purchased items through Zalando website. The data extracted from this file will be part of the input of the Aspect-Oriented Opinion Mining and Entity Linkage packages. For example, some items are very well appreciated by users while in other cases, some items may not match their taste or their idea about them. Thus, research in this area could lead to a deep understanding of consumer's preferences and subsequently improve the consumer's online shopping experience.

- Title (String): This could be interpreted as the summary of the user's comment.
- Text (String): This is the user's comment in its original language.
- sku (String): This is the stock keeping unit, it enables the identification of one specific item.
- Language_Code (String): Here we have two-letter language code which is the original language of the comment.

zalando-images-attributes-FB-release.csv : This file includes a detailed identification of products from different brands with their specialized features. This file is the one that will be used as a reference to identify the items and extract their specialized features.

- As_sku (String): This is the stock keeping unit, it enables the identification of one specific item.
- M_path (String): Image paths, separated by spaces where each image path needs to be appended to the URL: <https://mosaic01.ztat.net/vgs/media/pdp-gallery/>
- Article_name (String): Name of the product
- Brand_name (String): Name of the brand
- Pool_name (String): Name of the category
- Subpool_name (String): Name of the subcategory
- Main_color (String): Name of the main color
- Additional_attributes (JSON key-value pairs): Additional attributes as JSON key-value pairs (each article may have any number of additional attributes)

2.1.2 Relational Schema

The relational schema of Figure 2.1 describes the relationships between the files above-mentioned, and provides a comprehensive view of our database. Note that Region*_items and Region*_orders are related since the item specification in terms of categorization, season and gross amount is found in the first file while in the second one, we find the number of ordered items per date. Each review is identified by the product SKU which can be recognized with the image-attributes file. Some specific features named in the latter (color, material, etc.) can be translated into other languages with the OEM translation file.

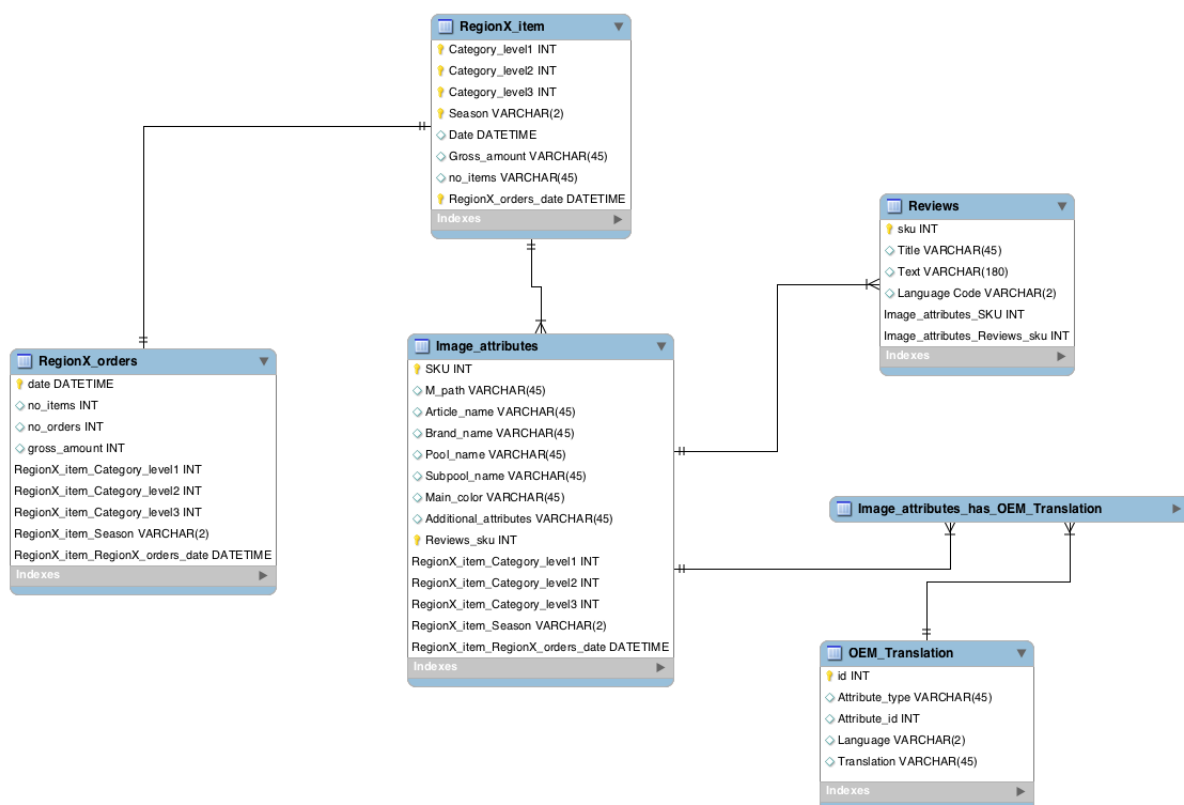


Figure 2.1: Relational model.

2.2 Fashion-tech startup

This section provides an overview of the datasets provided so far by Fashwell. The datasets are in the form of json and csv files. These files are a sample of the data that will be provided by Fashwell. The json files contain brand's catalogue while the csv files contain a fashion taxonomy. This taxonomy is a fine grained tree of fashion items depending on the gender. We propose to analyse the provided files in order

to extract their potential use in terms of research. For the json files, we describe the main attributes and their type. The fashion taxonomy file contains a table of attributes. We describe the most relevant attributes with their respective types. We begin by describing two example json files. Each file describes respectively GAP and MACYS online store's catalogue joined with a schema and an example. These attributes help to extract the product specification, e.g., the categorization of the items will be used in order to build a taxonomy in (WP1).

dump-example-*.json (* can be GAP or MACYS) :

- Brand (String): Name of Brand
- Category (Array of Strings): A categorization hierarchy of the item
- Gender ("enum": ["unisex", "female", "male", "boy", "girl", "infant", "unknown"]): It could be unisex, female, male, boy, girl, infant or unknown
- Name (String) Name of the item
- InternalId (String): Internal Order Number
- Product_url (String): Link to the product in the website
- Images_url (Array of Strings): Link to the product image
- Desc (String) Gives a description of a product

Below we give an example of the json description of one product.

```
{ "name": "Nike SB Stefan Janoski Max",
  "desc": "The Nike SB Stefan Janoski Max is designed with ...",
  "price": 140.00,
  "msrp": 210.00,
  "currency": "EUR",
  "sizes": ["34", "38", "40"],
  "category": ["men", "accessories", "glasses"],
  "images_url": ["http://www.foo.com/bar0.jpg", "http://www.foo.com/bar0.jpg"],
  "gender": "male", "brand": "Nike",
  "product_url": "https://www.foo.com/bar/baz",
  "ean": 123456789,
  "internal_id": "999111",
  "features": "110% cotton",
  "color": "tomato"
}
```

Fashwell has different sources of data. In order to have one unique dataset with non redundant elements, they integrate them all together to get one unique fashion tree using the model described in Figure 2.2.

We provide in what follows a brief description of the model used by Fashwell. This model takes as an input an image and a text associated with one product. For

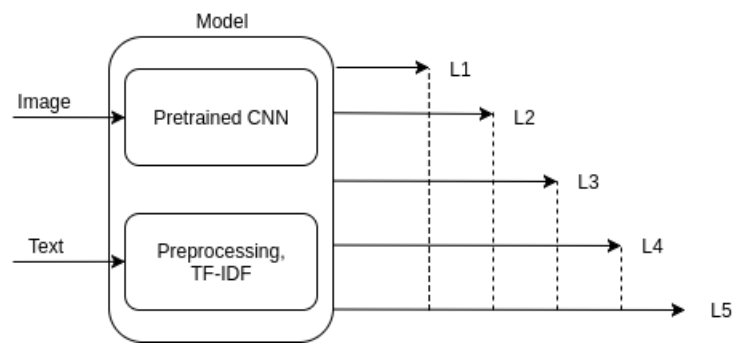


Figure 2.2: Fashion items classifier.

the vision part of the model, a convolutional neural network is used. the model is pre-trained on ImageNet dataset and then fine-tuned to fashion products taxonomy task. The text part of the model uses some NLP techniques for preprocessing, like stemming and stop-word removal, and then uses TF-IDF for feature extraction from the text. In the end, these two parts of the model (vision and text) are combined in a single neural network model which can predict a taxonomy. Interesting finding was that, in most cases, text was much more informative for inferring taxonomy correctly than an image. When a product has multiple images, the classifier is applied on all images with the text describing the product. In a later step, predictions average is made, which usually leads to better results. If there are multiple images per product, the vision part of the model improves in prediction accuracy. This model learns 5 tasks in joint - for each level of taxonomy. The higher the level, the more fine-grained the classification is. It is also taken into consideration the information from the lower levels of predictions to enhance the predictions of higher levels, but still the higher the level, the more imprecise the predictions are. Products with already-labeled taxonomies were used to train the classifier.

The main attributes of the fashion tree are depicted as follows:

fashion_tree_dump.csv :

- id (Integer): Category id in the Fashwell tree
- name (String): Name of the category
- parent_id (Integer): The parents id of a particular category.
- gender (enum: ["unisex", "female", "male", "boy", "girl", "infant", "unknown"]): The gender could be any element of the enum set

In the following section, we describe the results of the survey on how social media influences the shopping experience of their viewers. This survey aims to help us discovering the platforms and the fashion blogs that are used as fashion drivers.

3 Fashion Survey

Fashion bloggers and social media in a more general way are becoming an influential force within the fashion industry. Fashion related applications pulls data from Instagram and other social media platform in order to enlarge their datasets and consolidate it. The purpose of this survey is to extract more features and discover platforms that are used to spot new clothing trend. In a first part, we present the survey and in the second part, we describe the survey's analysis and results.

3.1 Presentation

This survey was performed on Crowdfunder (www.crowdfunder.com) (CF) asking 800 men and women from 16 European countries questions about their online shopping habits. The choice of the 16 European countries was based on Zalando's delivery destinations. When the participants accepts the task (or accesses the system web page), they will be asked to read the Information Sheet and Consent Form and indicate their agreement and this is the first page of the system. Then, the participants are given an entry questionnaire to gather demographic information such as age, gender, nationality, etc. Next, each participant will be presented with a series of questions about his/her social media habits, his/her behaviour towards fashion blogs and preferred content presentation in both retailer websites and social media platforms. The results of the survey are described in the following section.

3.2 Results and Analysis

3.2.1 Shopping Online/ Social Media Habits

Online shopping is becoming increasingly popular for multiple reasons. People find it more convenient because it saves time, it is available 24/7 and the comparison between products can be made in few clicks. Clothes shopping is not an exception, among the 800 men and women we asked, 94.7% of them shopped clothes online at least once. These digital buyers turn to social media for a variety of reasons like reading comments and discovering new trends. For instance, 75% of the participants who check their social media account several times a day, use them to follow fashion trend. While a clear majority use social media to follow new clothing trends, the platform choice differs slightly depending on the age and the gender. Figure 3.1 shows the top platforms depending on the gender. We can see that Facebook (www.facebook.com) (FB) is the most popular platform for both men and women,

Instagram and YouTube (www.youtube.com) (YT) are ranked just behind, then Pinterest who is more popular among women more than men.

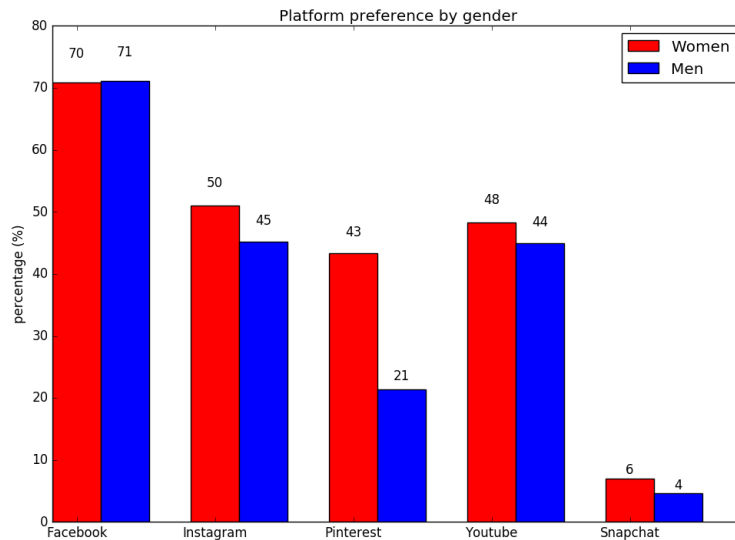


Figure 3.1: Platform preference by gender.

The preference for the social media platform depending on the age range is shown in Figure 3.2. FB has a uniform and large user base (2 billion monthly active users in 2017) which explains why 70% of each age range chooses FB to discover Fashion trend. Instagram is the second choice with more popularity among the younger generation (under 30). Around half of each age range uses YT, then we have Pinterest and Snapchat ranked just behind.

The results of the survey show that the majority of the participants, i.e., 71.8% of them, uses FB platform to discover new fashion trends. When studying the demographic of this majority, around 70% of each age range uses this platform to discover new trends. This result was expected since the user base of FB is uniform. Besides reaching a large community, a FB post has more opportunity for getting feedback of all kinds and is more likely to get a deeper level of interaction since the comments can contain text, emoji and images unlike other platforms where only text/emoji are allowed.

According to the survey results, Instagram is the second platform to be used in order to spot new clothing trends, i.e. 49.1% of the participants. A majority of these Instagram users are under 30 as a matter of fact, this platform appeals more to the younger generation at the opposite of Pinterest that is slightly more popular in older generations. Almost half of the participants use YT to discover fashion. As a matter of fact, in June 2016, beauty-related content generated more than 5 billion views per month. Popular types of YT beauty content include tutorials, reviews,

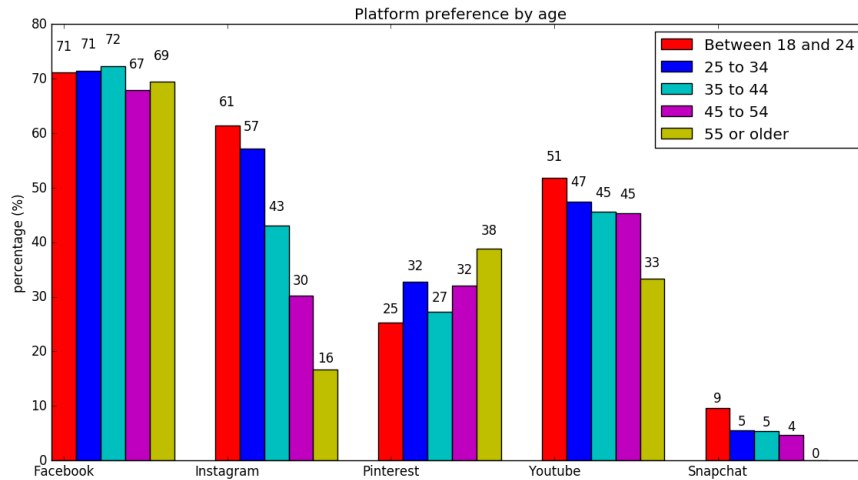


Figure 3.2: Platform preference for discovering clothing trend by age.

and videos produced by beauty vloggers [1].

Since FB and Instagram are the most popular platforms to discover fashion according to the survey results, we will briefly describe the difference between the two platforms. FB is focused more on text and is largely informational while, Instagram is more about “capturing the moment”. It is also worth mentioning that FB content is so diverse that one can easily get distracted, while Instagram focuses on image/video sharing only. Users are more able to be engaged with the visual content shared on Instagram which explains why when asking participants to name fashion influencers they know, we count only 9 FB pages while 87 fashion instagram accounts. The difference between the number of FB pages and the number of instagram accounts mentioned bring us to conclude that while FB has the largest user base, Instagram audience is more engaged. This last result suggests that social media platforms have a key role in creating, advertising and spreading clothing trends. Hence, social media data will play a pivotal role in the rest of our project. Since social media data often comes in an unstructured format, the ensuing data integration task will consist of recognizing key entities to be linked with our fashion catalog (Fashion Knowledge Base), this will be done either via text or image processing.

This question revealed some other platforms used to discover fashion trend like Reddit 21 buttons, google+, Soup.io, lookbook, bantoa, elcorteingles, etc.

3.2.2 Influence of Bloggers and Social Media on Their Viewers

In this section, we gather the results related to how participants use social media to discover fashion trends and who are the influencers they follow to spot new trends.

A common way to stay updated with the latest trend is to follow brands on social media. For instance, 63.6% of the participants follow their favourite brands on social media. They also find the best way to get inspiration about what clothing to purchase is going through online clothing shops (75.5%) and by following them on social media (48.6%). These methods beat by far the traditional way of looking into magazines (27.5%) and visual merchandising in shops (34.2%). Therefore, social media platforms are not only used to discover trend but also have an influence on people's decision to purchase items. As shown in the chart of Figure 3.3, 44.8% of the participants frequently purchase items discovered through social media.

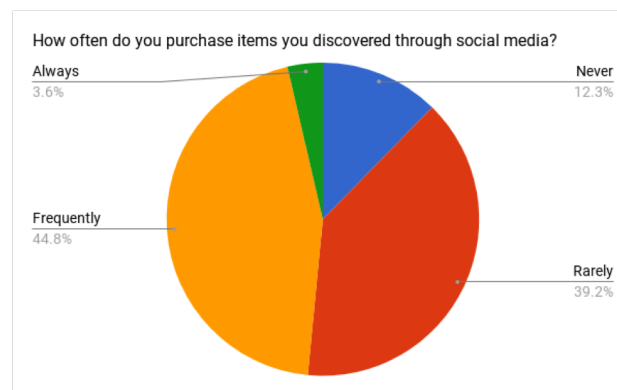


Figure 3.3: Frequency of purchased items discovered through Social Media.

Below some answers to the question “Describe how you use social media to discover fashion trends”: “I visit profiles of people who publish photos and reviews of clothes from new collections of well-known brands. I also visit the profiles with ideas how to join various pieces of clothing to make a great and fashionable outlook.” “FB and Twitter are great for discovering new trends, and Pinterest is wonderful for a visual experience.” These answers reveal how fashion bloggers are gaining a lot of popularity and have a clear influence on their followers as a matter of fact 40% of the participants use fashion blogs to have an idea about their next shopping. The participants provided us with 168 distinct blogs, YT channels and Instagram accounts (The list is kept in our shared repository). We name here those with the highest number of followers:

- @chiaraferragni
- @songofstyle
- @galagonzalez
- @thefashionguitar

From this list we can hypothesis that the list of influencers is large and growing which makes it difficult to keep up manually. The recommendation that we issue for (WP3) is the development of an interface for the continuous collection and curation of an influencer list by means of crowdsourcing.

3.2.3 Comments/Reviews

Users share their feelings and opinions on products mostly to clarify some doubts before purchasing (48.8%) or to give some subjective feedback afterwards (41%) as depicted in Figure 3.4. The most important aspects participants look for in a review are: (i) the item's general features specially the quality of the product and (ii) whether is it similar to the picture and description.

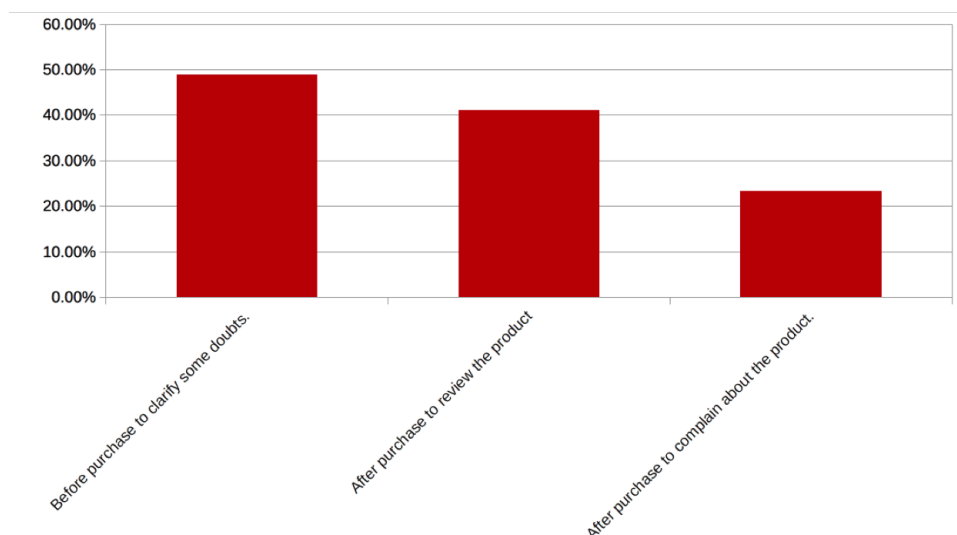


Figure 3.4: Percentage of participants answering: When are you more likely to review or comment about a fashion product?

With social media, regular consumers have a direct access to fashion designers, clothing and shoes firms. When fashion companies do social media marketing, they are more able to reach and engage a wider audience. When asking the participants “how do you interact with fashion brands on social media?”, we found as depicted in Figure 3.5, 43% of them answering by “liking posts/publications” and 25% of them by “commenting on posts/publications”. Therefore, it is possible to measure the popularity of a specific trend or style with the number of likes and comments. This same approach was used in different domain such as prediction of the new popular models [3] and the box-office performance of newly-released blockbuster movies [2].

3.2.4 Fashion Content

In this section, we describe the preference of data type according to the survey participants (image, text, videos) and how they prefer items to be displayed.

Figure 3.6 shows that image is the preferred content users want to see on fashion social media then comes videos and text consecutively. Even though videos seem to have less importance compared to image, they are an important source of

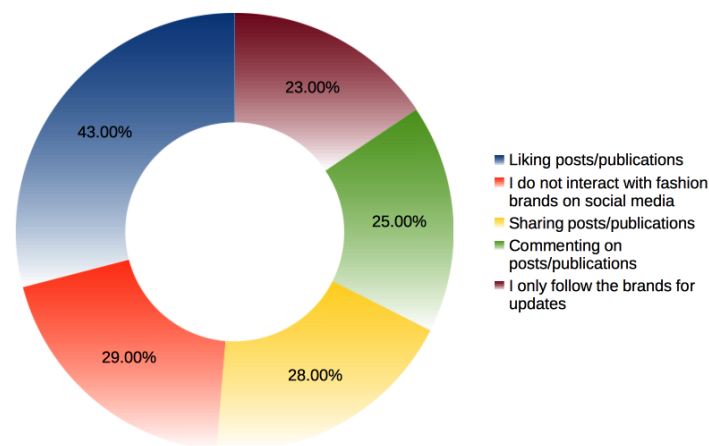


Figure 3.5: Interaction with fashion brands.

information to spot new trends. For example, the participants provided us with around 30 distinct YT channels dedicated to fashion. Such a high number of fashion videos shows that the latter are fashion drivers.

What kind of content do you prefer to see on Fashion's brand social media on a scale from 1 to 3 (1 is the lowest interest)

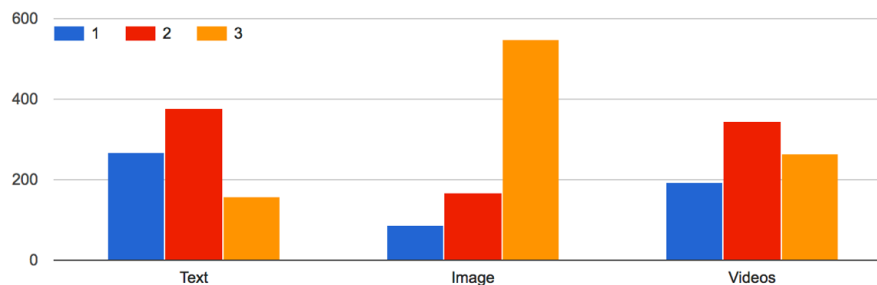


Figure 3.6: Content type preference.

Items on models and showing a combination of items is preferred over focusing on showing items solely as shown in Figure 3.7. Thus, for the image processing tasks, we recommend to take into account overall style recommendation, i.e., combination of multiple items instead of focusing on showing different items separately.

3.2.5 Other Forms of Content

VR and AR are still in early development stages. However as technology becomes mainstream, with companies like FB and Snapchat investing millions in R&D, these

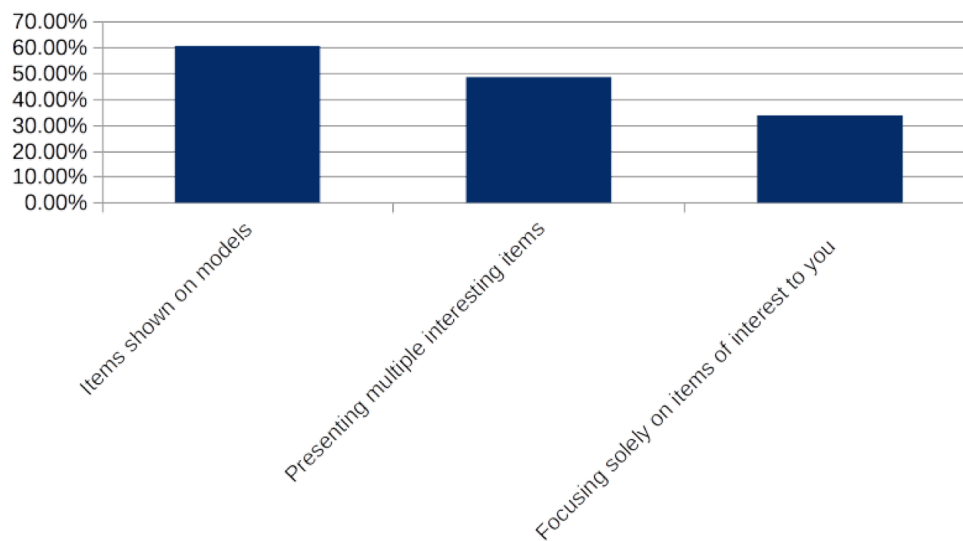


Figure 3.7: Focus preference.

technologies are expected to be integrated in the social media experience quite soon. Fashion related applications are already launched and even though they were not mentioned by the 800 participants, they will soon become a powerful channel for brand-consumer interaction :

- Uniqlo's Magic Mirror: Uniqlo offered shoppers the unique ability to imagine themselves in a range of color choices of a single silhouette without the hassle of having to remove the garment.
- The Sampler mobile app from Converse: A mobile application that gives people the ability to imagine what a Converse product would look like on-foot without having to actually try on a pair. One has only to point his iPhone camera at his leg and subsequently getting a visual reference on how it would look.

We add as a target in WP3 the tracking of the adoption of new technologies.

4 Conclusions

In this deliverable, we have described and analyzed all the internal datasets provided by the FashionBrain partners. This dataset description will help to identify the datasets needed for each task in the project as well as the connections between the provided datasets. In addition to the dataset description, we have also performed a survey on CF to evaluate the importance of SM platforms in online shopping. One of the main results of this survey shows that more than 44% of the social network users frequently purchase items discovered through SM, which confirms the increasing importance of SM in the fashion industry. The results of this survey will help us identifying the social platform to focus on for the task of identifying fashion trends using social media (related to D5.5).

Bibliography

- [1] Beauty-related content views on youtube. <https://www.statista.com/statistics/294655/youtube-monthly-beauty-content-views/>, 2018. Accessed: 2018-12-15.
- [2] S. Asur, B. A. Huberman, G. Szabo, and C. Wang. Trends in social media: Persistence and decay. *5th International AAAI Conference on Weblogs and Social Media*, 2011. URL <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2815/3205>.
- [3] Jaehyuk Park, Giovanni Luca Ciampaglia, and Emilio Ferrara. Style in the age of instagram: Predicting success within the fashion industry using social media. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, pages 64–73, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-3592-8. doi: 10.1145/2818048.2820065. URL <http://doi.acm.org/10.1145/2818048.2820065>.