KU LEUVEN

FACULTY OF SCIENCES

Visualizing and analyzing of wine reviews

Serhat Erdogan

Supervisor: Prof. dr. K. Verbert

Assessors:
ir. M. Millecamp
Dr. E. Sansone

Mentor: Dr. N. N. Htun

Master thesis submitted in fulfillment of the requirements for the degree in Master of Science in Applied Informatics

Academic year 2020-2021

Preface

Ik wil beginnen met mijn begeleider Nyi-Nyi Htun te bedanken voor alle hulp en inzichtrijke discussies gedurende de hele periode van deze studie. Aan professor K. Verbert voor het interessante en uitdagende onderwerp. Alle test die deelnamen aan de evaluatie van het systeem. Tot slot wil ik mijn ouders bedanken voor mij de mogelijkheid te geven om mijn studies te voltooien.

I want to start by thanking my supervisor Nyi-Nyi Htun for all the help and insightful discussions during the entire period of this study. To professor K. Verbert for the interesting and challenging subject. To all participants that partook in the evaluation of the system. Finally, I want to thank my parents for giving me the opportunity to complete my studies.

Contents

Preface....i
Contents....ii
Abstract....iv
List of Figures....vi
List of Tables....viii
Introduction....1
1.1 Context....1
1.2 Research questions....2
1.3 Approach....2
1.4 Contributions....3
1.5 What is Vivino?....3
Literature review....5
2.1 Data preparation and analysis....5
2.1.1 Data extraction and refinement....6
2.1.2 Data preparation and analysis....6
2.1.3 Interpretation and discussion....8
2.2 Content-based analysis....8
2.2.1 TF-IDF....8
2.2.2 Content-based Emotion Lexicon....10
2.2.3 Sentiment analysis....11
2.3 Content-based visualization....13
2.3.1 Radar chart....13
2.3.2 Bar chart....14
2.3.3 Sankey diagram....14
2.3.4 Word clouds....16
2.4 Visual analytics methods....17
2.4.1 Visual Summary Reports....17
2.4.2 Cluster Analysis....18
2.4.3 Circular Correlation Map....19
2.5 Color combinations....19
2.5.1 Deuteranopia....19
2.6 Research gap....21

2.7 Conclusion 22
Evaluation methods..24
3.1 Think aloud study..24
3.2 User study..24
3.3 Custom questionnaire 26
3.4 Thematic analysis 27
3.5 Flow of the procedure..27
Implementation..30
4.1 Software stack..30
4.1.1 Ploty..30
4.1.2 Qualtric..31
4.1.3 Matplotlib..32
4.2 Data 32
4.2.1 Data collection and preparation..32
4.2.2 Data description 33
4.2.3 Data cleaning 34
4.3 Final version..35
Evaluation results..39
5.1 First evaluation: Think aloud..39
5.1.1 Results and changes..39
5.2 Second evaluation: User study..40
5.2.1 Information participants 40
5.2.2 System usability score..42
5.2.3 Results custom questionnaire 43
Discussion..58
6.1 Feedback and possible improvements 59
Conclusion and future work 61
Appendices 63
A.1 System Usability Scale (SUS) questionnaire 63
A.2 Custom questionnaire: Information participants 64
A.2 Custom questionnaire: end questions..65
Bibliography 67

Abstract

This thesis starts by giving a clear overview of the existing visualizations that were used to visualize text data. Then, a concise list of techniques to analyze and visualize this data is presented. It gives an idea of the various ways text data can be visualized.

Upon inspection of academic examples, several visualizations were used to visualize product characteristics. By taking a look at all these examples a clear research gap is identified. The need to know which visualizations would be best fitting to visualize product characteristics from positive, negative and neutral reviews, is the focus of this thesis. To accomplish this four existing visualizations are used to visualize the product characteristics.

The data was provided by Vivino with permission to use it for research purposes. This dataset contains user reviews of wines on the Vivino website. The four visualizations were chosen as they are good methods to represent certain attributes of objects, in this case, wine. Simple visualizations were chosen since user interpretation and comprehension are one focus of this research.

In the end, a thorough evaluation of the implemented system is presented. The fact that the implementation was web-based leads to a higher number of possible participants and an easier evaluation procedure. The implementation is screened on results of questionnaires data. In the end the results are discussed together with some possible improvements and future work.

Deze masterproef begint met het geven van een duidelijk overzicht van de bestaande visualisaties die werden gebruikt om tekstgegevens te visualiseren. Vervolgens wordt een beknopte lijst met technieken gepresenteerd om deze gegevens te analyseren en te visualiseren. Het geeft een idee van de verschillende manieren waarop tekstgegevens kunnen worden gevisualiseerd.

Bij inspectie van academische voorbeelden werden verschillende visualisaties gebruikt om product eigenschappen te visualiseren. Door al deze voorbeelden te bekijken, wordt een duidelijke onderzoekskloof geïdentificeerd. De noodzaak om te weten welke visualisaties het beste passen om productkenmerken van positieve, negatieve en neutrale recensies te visualiseren, is de focus van deze masterproef. Om dit te bereiken worden vier bestaande visualisaties gebruikt om de producteigenschappen te visualiseren.

De gegevens zijn door Vivino verstrekt met toestemming om deze voor onderzoeksdoeleinden te gebruiken. Deze dataset bevat gebruikersrecensies van wijnen op de Vivino-website. De vier visualisaties zijn gekozen omdat ze goede methoden zijn om bepaalde eigenschappen van producten weer te geven, in dit geval wijn. Er is gekozen voor eenvoudige visualisaties omdat interpretatie en begrip van de gebruiker centraal staan in dit onderzoek.

Ten slotte wordt een grondige evaluatie van het geïmplementeerde systeem gepresenteerd. Het feit dat de implementatie webgebaseerd was, leidt tot een groter aantal mogelijke deelnemers en een eenvoudigere evaluatieprocedure. De implementatie wordt gescreend op resultaten van vragenlijstgegevens. Uiteindelijk worden de resultaten besproken samen met enkele mogelijke verbeteringen en toekomstig werk.

List of Figures

Figure 1.1: Vivino website 3
Figure 2.1: Overall framework data analyzing [9] 5
Figure 2.2: Procedure of review-data extraction [9] 6
Figure 2.3: Pseudocode K-means algorithm [15] 7
Figure 2.4: Inverse document frequency [16] 8
Figure 2.5:TF-IDF process [14] 9
Figure 2.6: Distribution of emotion of six most popular product reviews of Amazon users [21] 10
Figure 2.7: sentiment analysis process on product reviews [22] 11
Figure 2.8: Sentiment classification techniques [22] 12
Figure 2.9: Rader chart of wine characteristics [32] 13
Figure 2.11: Data based on review categories [33] 14
Figure 2.10: Data based on product categories[33] 14
Figure 2.12: Example visualizations using Sankey diagrams [5] 15
Figure 2.13: Reviews that are displayed as a separate layer on top of the Sankey visualization [5] 15
Figure 2.14: ReCloud of a Yogurtland store near UC Berkeley campus [29] 16
Figure 2.15: Summary Report of printers [2] 17
Figure 2.16: Scatterplot of customer reviews on printers [2] 18
Figure 2.18: Analyzing the score 1 comments: Service is one of the attributes that is often mentioned negatively [2] 19
Figure 2.17: All Customer Comments. Most comments have an overall positive tendency [2]. 19
Figure 2.19: Subset of comments on service. Not all the customers are dissatisfied with the service. But this was a hot topic in the score 1 comments [2]. 19
Figure 2.20: Colors as they appear to readers with normal vision and deuteranopia [37] 20
Figure 2.21: Point classes typical of a dot map distinguished by saturation, hue and shape [37] 20
Figure 2.22: plotly interface 21
Figure 3.1: counterbalanced design, four conditions [44] 25
Figure 3.2: Different visualizations used to visualize wine characteristics 26
Figure 3.3: Guided sequence of pages 28
Figure 4.1: Example radar chart [45] 30
Figure 4.2: Example Sankey diagram [45] 30
Figure 4.3: Qualtric survey's 31
Figure 4.4: Qualtric individual answers 31
Figure 4.5: Pseudo code, creating word cloud using Matplotlib with Plotly 32
Figure 4.6: Review example Vivino website 33
Figure 4.7: Review example json form 33
Figure 4.8: Python transform_data.py 34
Figure 4.9: Python json_to_csv.py 34
Figure 4.10: CSV data file of user reviews 34
Figure 4.11: Page one, visualizations for one specific wine 35

Figure 4.12: Page two, comparing wine....35
Figure 4.13: page three, filtering wine attributes....35
Figure 4.14: Final implementation....36
Figure 4.15: Final implementation using word cloud....37
Figure 4.16: Final implementation using Sankey diagram....37
Figure 4.17: Final implementation using bar chart....37
Figure 5.1: Bar chart question: “What's your gender?"....40
Figure 5.2: Bar chart question: “What's your age group?"....40
Figure 5.3: Bar chart question: “How often do you buy wine on average?"....41
Figure 5.4: Bar chart question: “Have you ever searched/browsed for wine online before?"....41
Figure 5.5: Bar chart question: "Do you read the reviews when you are buying wine?"....41
Figure 5.6: Bar chart question: "Do you prefer to read the negative reviews, the positive reviews or both?"....42
Figure 5.7: Bar chart question: “When do you consider a review negative?”....42
Figure 5.8: Score formula SUS-questionnaire (SQx= score of question x)....42
Figure 5.9: Box-plots: score of SUS-questionnaire....43
Figure 5.10: Bar chart Q1 results....44
Figure 5.11: Bar chart occurrence codes for Table 5.3....45
Figure 5.12: Bar chart Q2 results....46
Figure 5.13: Bar chart occurrence codes for Table 5.5....47
Figure 5.14: Bar chart Q3 results....48
Figure 5.15: Bar chart occurrence codes for Table 5.7....49
Figure 5.16: Bar chart Q4 results....50
Figure 5.17: Bar chart occurrence codes for Table 5.9....52
Figure 5.18: Bar chart Q5 results....52
Figure 5.19: Bar chart occurrence codes for Table 5.11....54
Figure 5.20: Bar chart Q6 results....54
Figure 5.21: Bar chart occurrence codes for Table 5.13....56
Figure 5.22: Bar chart Q7 results....56

List of Tables

Table 5.1: SUS-questionnaire results 43
Table 5.2: Results Q1 custom questionnaire 44
Table 5.3: Result thematic analysis Q1 for bar chart visualization 44
Table 5.4: Results Q2 custom questionnaire 46
Table 5.5: Result thematic analysis Q2 for Sankey diagram visualization.............................46
Table 5.6: Results Q3 custom questionnaire 48
Table 5.7: Result thematic analysis Q3 for bar chart visualization 48
Table 5.8: Results Q4 custom questionnaire 50
Table 5.9: Result thematic analysis Q4 for Sankey diagram visualization.............................51
Table 5.10: Results Q5 custom questionnaire 52
Table 5.11: Result thematic analysis Q5 for bar chart visualization 53
Table 5.12: Results Q6 custom questionnaire 54
Table 5.13: Result thematic analysis Q6 for bar chart visualization 55

Chapter 1

Introduction

1.1 Context

The internet provides not only a retail channel for purchase but also the development of a vast repository of customer and user reviews. Existing research indicates that 81% of Internet users have searched for a product online at least once. Customers are also more willing to buy products with a rating of 5 stars than a rating of 4 stars. This means that customers reviews can significantly impact a business's profit margin on each product. Competitive organizations now collect large amounts of review data from their customers and can potentially use these customer reviews to improve the quality of their products. Customers reviews are a valuable source of information [1].

Not only are customer reviews beneficial to organizations, customers themselves also find that reading other customers opinion about the product provides an invaluable source of sometimes more detailed and more objective information than the company publishes. However, customer reviews are unstructured by nature and far too many customer comments to read them all sequentially [2]. For example, "Mellanni Bed Sheet Set" is the most-reviewed product on Amazon.com with some 62,500 reviews [3]. Basic filters allow the reviews to be selected based on rating or other numerical features. This limits the amount of reviews that a customer has to read but also limits the amount and kind of information they gather. The great quantity of reviews provides so much valuable input but the sheer amount makes it difficult for an individual customer to actually access all of it. Visualization tools can help summarize reviews and highlight popular sentiment so a customer can more easily draw a conclusion about the characteristics of the product.

Visualizing and ordering the product characteristics mentioned in customer reviews can make it easier for potential customers and the company to have general idea what the positive and negative aspects are of the product. By distinguishing the positive and negative reviews and visualizing them, such a tool can save customers time, prevent them from becoming overwhelmed, and help them understand the product better. The goal of this research is to identify which visualization techniques are most suitable for comparing wine characteristics. This will give the customer better insight about various products and help companies understand what improvements can be made to suit their customer demand.

1.2 Research questions

Previous research has focused on improving visualization techniques and visualization content in different ways. Usability and user comprehension of various visualizations have not been studied. In this research, we aim to answer the following research questions:

Research question 1: *How can we better visualize the positive, negative, and neutral characteristics of wine from review text?*

Research question 2: *How can we make positive, negative, and neutral reviews easily understandable?*

1.3 Approach

As the intention is to assess different visualization techniques to provide understandable information summarized from customer reviews, an appropriate selection of techniques must be chosen. Table Based on the data provided by Vivino, certain techniques were identified as possible candidates for the visualizations of the customer review.

Out of the existing visualizations that have been used in previous research to visualize product characteristics, bar chart, Sankey diagram, word cloud, and radar chart were selected for their ability to display multiple characteristics at once. These techniques also display the relative importance/weight of the characteristics [4] [5] [6]. Simple visualizations were chosen since user interpretation and comprehension are one focus of this research. To answer all the research questions, a web-based application will be constructed, where users will have the ability to search for a specific wine and have an visualization that display the characteristics with their relative importance/weight. This web-based application has several advantages:

- **Control**: it's possible to monitor and evaluate every move the user does. This offers a detailed description of what the individual has done while using the tool, which can be compared with questionnaire results.

- **Ease of evaluation**: using an online tool eliminates the need to do an in-person evaluation. It is also possible to include the questionnaires in the platform itself and automatically calculate the responses. Therefore, the evaluation is more efficient and the results are more reliable.

- **Audience**: it’s easier to reach a higher amount of participants because of the web-based approach. However, there would also be an increase in the number of incomplete or invalid answers. We can filter these invalid responses by locking the finish button for a certain amount of time, so that the participants can’t rush to finish the task. Responses can be required in order for the participant to continue with the study.

1.4 Contributions

The research questions aim to bridge the gaps in the research identified earlier leading to two interesting contributions:

- Concluding which of the chosen visualization is the better to use for the end user when visualizing wine characteristics from customers reviews. It offers a way for the user to have a general idea what people's opinions are about the product.

- Comparison of four different visualization (radar chart, bar chart, word cloud, and Sankey diagram). All four will be compared based on the different groups of test-users. Each group will see the visualizations in different order so that the order won't affect the result.

1.5 What is Vivino?

Vivino is a Danish-based company that is currently the most used wine application used by over 26 million consumers worldwide. The database makes up the world's most comprehensive wine library with millions of wines included. Vivino has built a community and developed an app that is really perfect for those of us who enjoy red wine, but are not professional sommeliers. The Vivino app encourages customers to take a snapshot of the label of a bottle, and the app can automatically gather details about the wine, its ranking score, feedback, and much more [7].

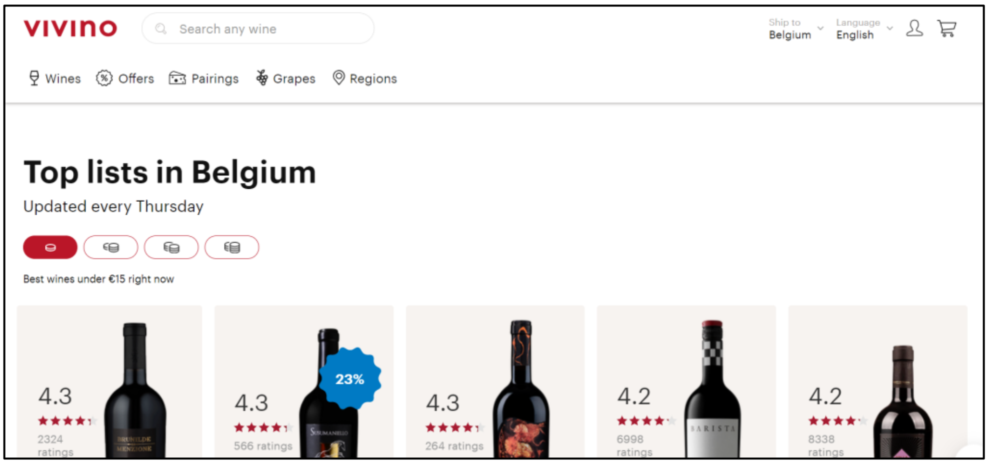

Figure 1.1: Vivino website

Chapter 2

Literature review

In the following chapter, a series of different approaches to prepare, analyze and visualize the data are discussed. For example, sentiment analysis using TF-IDF is a technique to score the reviews based on the content. For each of these approaches, some concrete examples are given that implement said approaches. This should give a clear view of in which cases each approach can be applied and what types of visualizations can be used.

2.1 Data preparation and analysis

The overall research architecture of review data analyses is shown on Figure 2.1. The research framework that is shown on Figure 2.1 is an improvement of the designers' previous work [8]. Figure 2.1 shows the entire from data extraction and analysis to visualizing the data and drawing an conclusion out of it [9]. Looking at the architecture in Figure 2.1, we can see three main sections:

1. Data extraction and refinement
2. Data preparation and analysis
3. Conclusion

The flow in Figure 2.1 is mostly followed throughout this study. The following subsections provide detailed descriptions and explanation of the three main sections of this architecture.

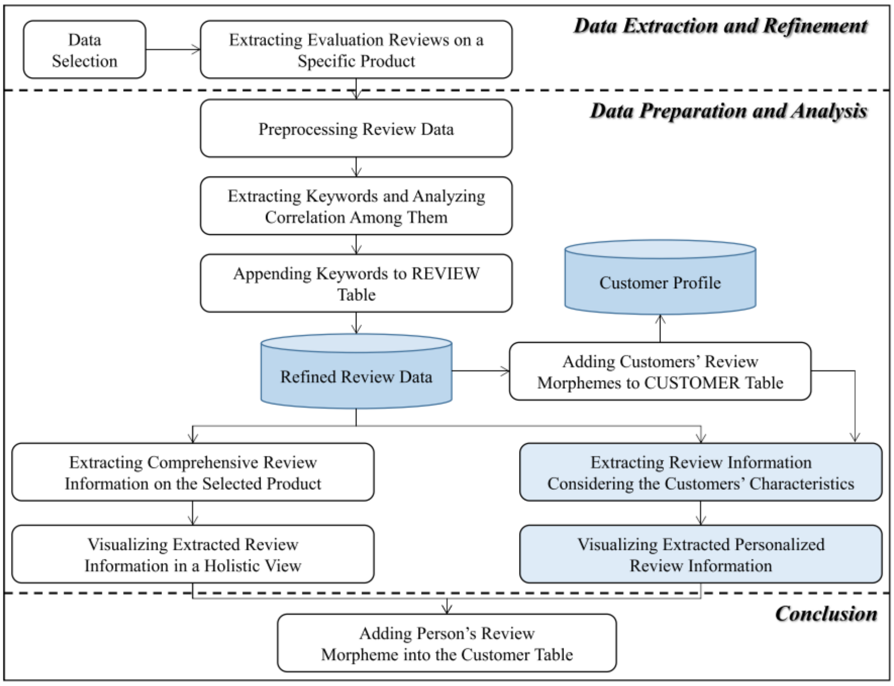

Figure 2.1: Overall framework data analyzing [9]

2.1.1 Data extraction and refinement

Current methods of keyword extraction and refinement were used for testing to find an appropriate method for evaluating and visualizing detailed and customized online product reviews. Figure 2.2 shows the process of data extraction. We can extract free-form review data on e-commerce sites as a text file using web crawling methods. “Web crawling, a process of collecting web pages in an automated manner, is the primary and ubiquitous operation used by a large number of web systems and agents starting from a simple program for website backup to a major web search engine” [10]. We can use this text file to create review data table as shown in Figure 2.2 [9].

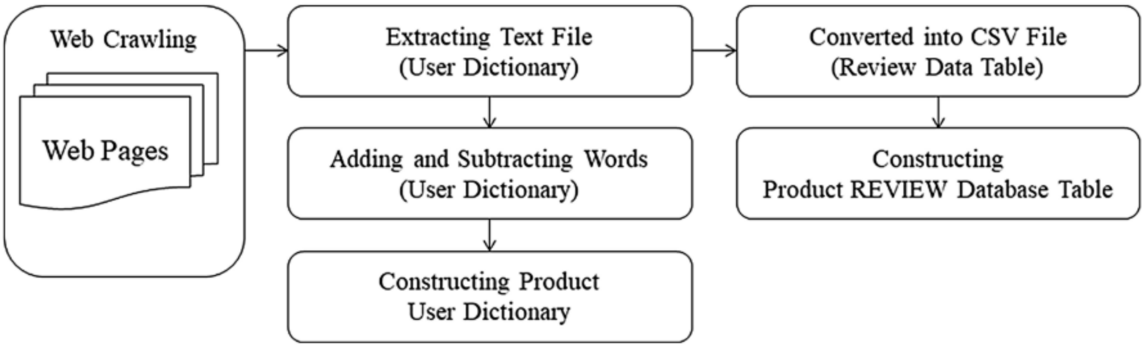

Figure 2.2: Procedure of review-data extraction [9]

2.1.2 Data preparation and analysis

“Review text mining is a mining technique that finds meaningful and useful information from unstructured review text data”. Because all of the data that is going to be analyzed is made up of a sequence of characters, there will be differing characteristics of languages. This means that there is no well-established formal analysis method [9]. We can categorize text mining as [11]:

- Text categorization
- Text clustering
- Association rule extraction
- Text visualization

Some text mining techniques are discusses in the following subsections.

Text clustering

“Text clustering is based on the cluster hypothesis which proposes that relevant documents must have more similarities with one another than the non-relevant ones” [12]. This clustering technique is used for analyzing large amount of data and is known as a trustworthy technique. Research has been done on text clustering which was found to be a very effective technique for thematic analysis [11] [13].

K-Means algorithms

K-means algorithm works as follows: first the data is divided into k clusters. Each cluster is defined by the average of points, called the centroid. Step one of the algorithm is to assign every point to the closest centroid. Step two is re-evaluating the centroids based on the new groupings. When the cluster centroid arrives at a constant value, the process is finished [11].

The steps of the k-means algorithm [14]:

1. In the first step we specify the number of clusters.
2. The second step involves initialization of centroids by shuffling the dataset first and then picking K data points for the centroids randomly without substitution.
3. Step three, keep iterating until the centroids are not altered. That is, the distribution of data points to clusters does not change.
4. Step four, we have to calculate the sum of the squared distance between data points and all centroids.
5. Step five, assign each point of data to the closest cluster (centroid) .
6. In the last step we take the average of the all data points that belong to each cluster and then calculate the centroids for the clusters.

```
Initialize K means with random values


For a given number of iterations:
    Iterate through items:
        Find the mean closest to the item
        Assign item to mean
        Update mean
```

Figure 2.3: Pseudocode K-means algorithm [15]

2.1.3 Interpretation and discussion

The final step is drawing a conclusion from our results. As mentioned before, the steps in Figure 2.1 are mostly followed in this thesis to be able to answer the research questions. Steps in this architecture can be used to decide which visualization approaches can better express the meaning of data [11].

2.2 Content-based analysis

It's possible to analyze the review based on different factors. As the name suggests, we can analyze the reviews based on the actual content instead of basing it on the review rating itself. It's becoming very difficult to process structured or semi-structured data in organizations because data has been increasing tremendously [16] [17] [18]. There are several techniques or algorithms that can be used to process this data. One of the technique is called tf-idf or TFIDF, short for term frequency–inverse document frequency.

2.2.1 TF-IDF

"TF-IDF is a numerical statistic that shows the relevance of keywords to some specific documents or it can be said that, it provides those keywords, using which some specific documents can be identified or categorized" [16].

Example

An example of how TF-IDF can be used is as follows. Let's consider that we have a document, this document contains 50 words wherein the word 'dog' appears 3 times. The term frequency for 'dog' is then (3/50) = 0,06. Now, let's assume we have a 1 million documents and the word 'dog' appears in one thousand documents. Then, the inverse document frequency is calculated as log(1.000.000 / 1.000) = 3. Thus, the TF-IDF weight is the product of these quantities: 0.06 * 3 = 0.18 [19].

This technique can be used to compute the importance of characteristics in reviews. Which tells us the characteristics that are mentioned the most. The TF-IDF process is shown in Figure 2.4. In the next section I discuss some limitations of TF-IDF.

$$idf(t_i) = \log \frac{N}{n_i}$$

Figure 2.4: Inverse document frequency [16]

Limitations

There are several limitations to using a TF-IDF. The algorithm will treat words with a slight change in it's tense as different words. To give an example, The algorithm will identify "mark" and "marking", "play" and "playing", "year" and "years" as different words. Because TF-IDF is not able to check the semantic of the text in documents, it is only useful until the lexical level and it also unable to check the co-occurrences of words [16].

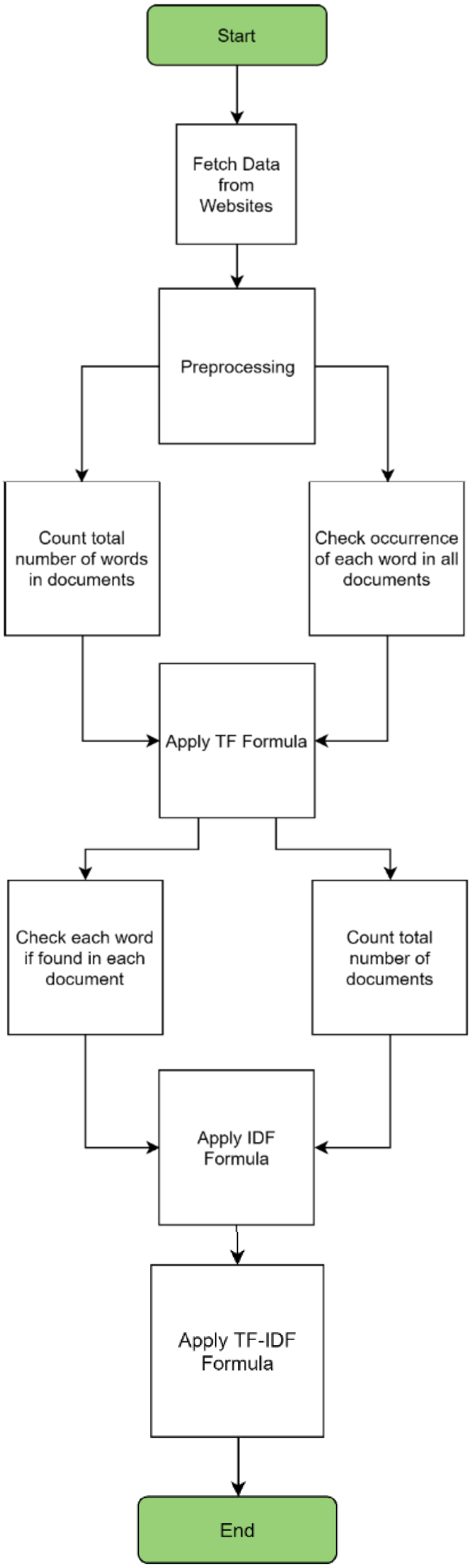

Figure 2.5:TF-IDF process [14]

2.2.2 Content-based Emotion Lexicon

We are able to analyze a product review and associate words in that review with emotions. For example, "*happy*" and "*excited*" are indications that you are feeling a joyful emotion, "*lonely*" and "*cry*" are indications of sadness. We first need to identify the set of basic emotions before we can categorize reviews to basic emotions. There have been a number of theories that propose which emotions are basic. "Ekman (1992) argues that there are six basic emotions: joy, sadness, anger, fear, disgust, and surprise". "Plutchik (1962, 1980, 1994) proposes eight basic emotions. These include Ekman's six as well as trust and anticipation" [20]. The large word–emotion association lexicon that has been created by this research, has decided to annotate words with Plutchik's eight basic emotions [20].

So this NRC emotion lexicon is able to categorize customers' reviews into eight emotions and two sentiments i.e positive or negative. In this example, the dataset that was used was collected from Kaggle which consists of reviews of fine foods from Amazon from October 1999 to October 2012 [21]. This dataset includes:

- 568,454 reviews
- 74,258 products
- 256,059 users
- 260 users with more than 50 reviews.

For analyzing the product review two approaches were used. A user-centric approach, they found potential reviewer who posted their own opinion to the m number of products on the Amazon. The results are visualized in Figure 2.6 using an bar chart visualization.

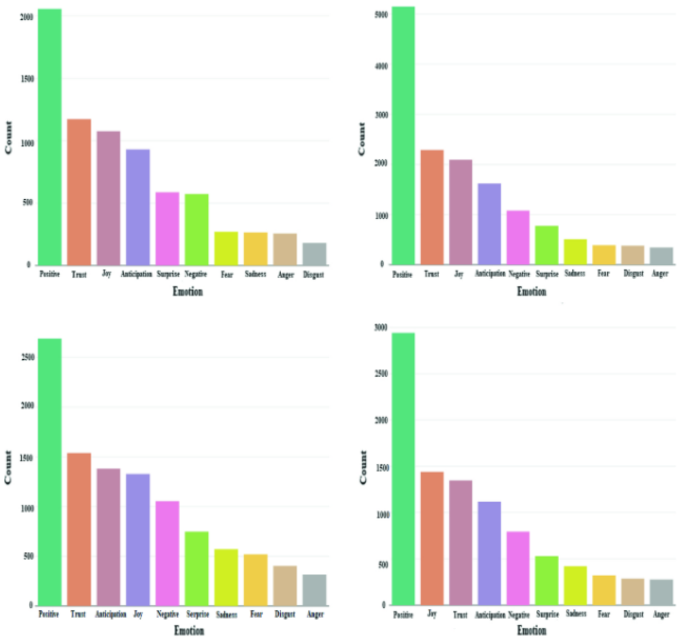

Figure 2.6: Distribution of emotion of six most popular product reviews of Amazon users [21]

2.2.3 Sentiment analysis

Sentiment analysis or opinion mining is a computer study of the beliefs, behaviors and feelings of individuals towards an entity. Individuals, activities or subjects may reflect the entity. Reviews are most likely to discuss these topics. Sentiment analyses or opinion mining are synonymous. They convey a sense which is mutual. Some scholars however, have reported that opinion mining and sentiment analysis have somewhat different notions. Opinion mining captures and analyzes the opinion of people regarding an entity, while sentiment analysis detects and analyzes the sentiment expressed in a text. The aim of sentiment analysis is to locate opinions, classify the feelings they express, and then categorize their polarity as seen in Figure 2.7 [22].

The main sources of data that is used in sentiment analysis are from the product reviews [22]. For company owners, these reviews are relevant since they may take business decisions based on the empirical findings of the feedback of consumers regarding their goods. Sentiment analysis can also be used on stock markets [23] [24], news articles [25] or political debates [26].

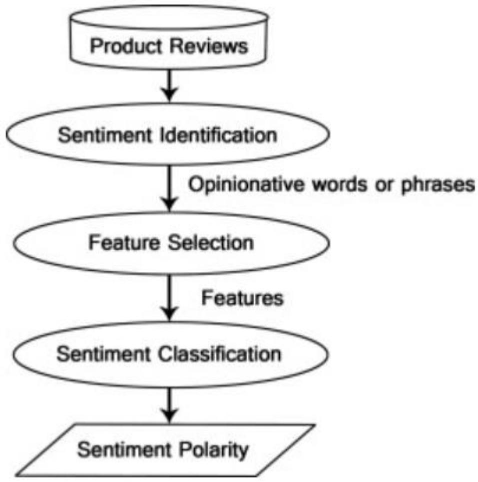

Figure 2.7: sentiment analysis process on product reviews [22]

As shown in Fig 2.7, sentiment analysis may be called a classification method. The three main levels of classification are [22]:

1. Document-level
2. Sentence-level
3. Aspect-level.

Document-level sentiment analysis tries to categories text as expressing, opinion or feeling that is positive, negative or neutral. It considers the whole text as a basic information unit. The purpose of sentence-level sentiment analysis is to identify the emotions conveyed in each sentence. The first step is to determine whether the sentence is subjective or objective. If it's the case that the sentence is subjective, then the sentence-level sentiment analysis will decide whether the sentence expresses positive or negative opinions. In order to obtain this information, classifying text at the document level or at the sentence level does not have the requisite of comprehensive opinions on all facets of the organization required in certain applications; we need to go to the aspect level. “The aim of aspect-level sentiment analysis is to classify the sentiment with respect to the specific aspects of entities” [22].

In sentiment analysis algorithms, there are many applications and enhancements that in the last few years were proposed. In this study [22], the authors have collected fifty-four articles which presented important enhancements to the sentiment analysis field lately. They are categorized according to the purpose of the paper explaining the algorithms and data used in their work. A more detail explanation of feature selection and sentiment classifications as shown in Figure 2.7, is given in the next 2 subsections [22].

Feature selection

The first step in the sentiment classification problem is extracting and selecting text features. Some of the latest features are [27] [22]:

- **Terms presence and frequency**: these characteristics are actual words or n-grams of the word and their frequency counts, explained in section 2.2.1.
- **Parts of speech**: since they are important markers of opinions, seeking adjectives.
- **Opinion words and phrases**: there are terms widely used to express views, such as positive or bad, like or dislike.
- **Negations:** the presence of negative words will modify the orientation of thought as not good is equal to bad.

Sentiment classification techniques

We can roughly divide sentiment classification techniques into lexicon based approach, hybrid approach and machine learning approach [28]. The machine learning approach uses the familiar ML algorithms and linguistic characteristics. The lexicon-based approach, explained in Section 2.2.2, relies on a sentiment lexicon, a set of known and precompiled sentiment terms. Approaches that use statistical or semantic methods to find sentiment polarity are dictionary-based approach and corpus-based approach. The hybrid approach blends both methods and is very common in most strategies, with sentiment lexicons playing a key role. The various approaches and the most popular algorithms of sentiment classification are shown in Figure 2.8 [22].

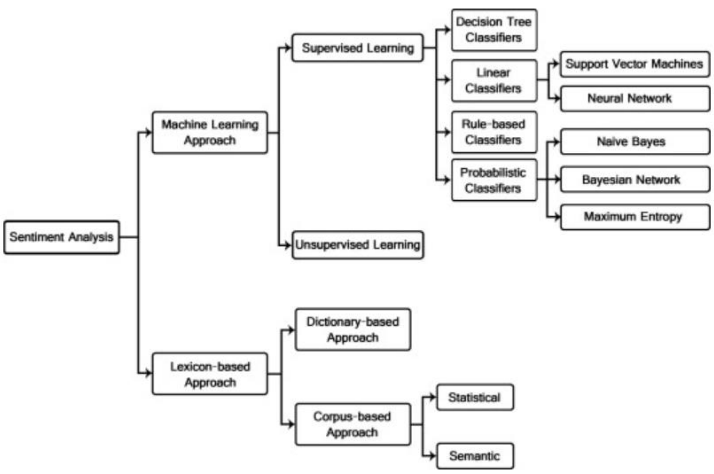

Figure 2.8: Sentiment classification techniques [22]

2.3 Content-based visualization

Customer reviews, like those found on Amazon, Yelp, Coolblue, etc. have become an important guide for decision making in everyday life. For example, in shopping, dining and entertainment. However, the reading process is tedious with large volumes of feedback available. Visualization of online reviews can be categorized into two types [29].

First, visualization of the measurable features of reviews are regularly used to display customer price level, ratings, and other numerical measurable features of a product. Wu et al., for example, introduced a method focused on quantitative analysis characteristics to display hotel customer reviews. In reviews however, many products or services cannot easily be defined with quantitative values; for consumers to make informed choices, finer perspectives into individual review content are required [29].

Second, visualizations of the textual content of reviews can provide a deeper view. By taking a look at the examples that were used in previous research [30] [31] [5], several types or styles of visualizations can be used to visualize and explain reviews. So we are able to combine these visualizations with the content-based model. In the next sections, four types of content-based visualizations that have been used in previous studies and that have been applied in this user study will be discussed [29].

The focus here is not the data presentation but the different possible visualizations that can be used to visualize explain text. The reason for taking a look at the way explanations are visualized is because providing fitting explanations can be beneficial for customers who are searching for wine or other products and want to be have a summary of the reviews.

2.3.1 Radar chart

One of the types of visualization that has been used for content-based recommendation methods is the radar chart [32]. Figure 2.9 is an example of a radar chart, where wine characteristics are used as features. When visualizing reviews of wine, we can visualize the features of each wine that are mentioned in reviews. This makes it possible to visualize multiple characteristics at once and how many times a characteristic occurs.

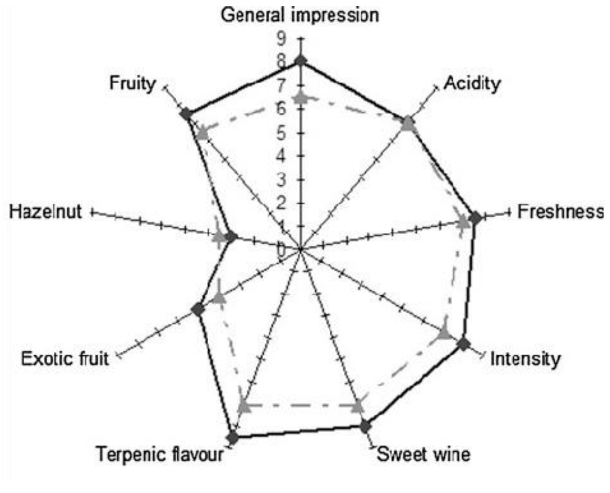

Figure 2.9: Rader chart of wine characteristics [32]

2.3.2 Bar chart

Another concrete example of a visualization that has been used to visualize content is the bar chart, as it is capable of representing multiple features or categories of an item at once. The graphs in Figure 2.10 and Figure 2.11 are the results of a past study where sentiment analyses was used on product review [33]. These two graphs show us a correlation between:

- Product categories with amount of reviews
- Review categories with amount of reviews

For this thesis, bar chat is used with a correlation between product characteristics and the amount of occurrence of the product characteristic. The advantages of using a bar chart are that they are visually strong, familiar and they allow the comparison of multiple datasets [4]. Disadvantages of using a bar chart are that you can only use discrete data and graph categories can be reordered to emphasize certain effects [4].

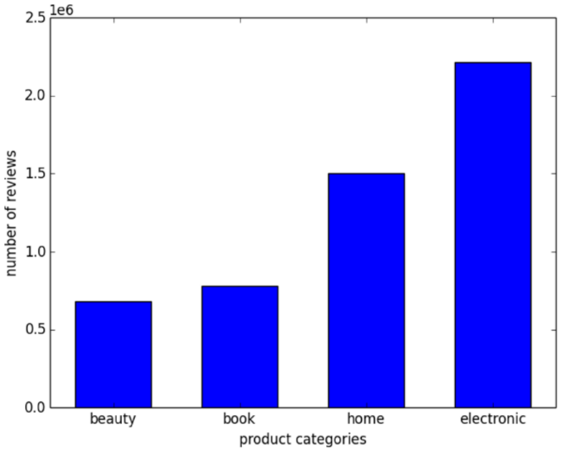

Figure 2.11: Data based on product categories[33]

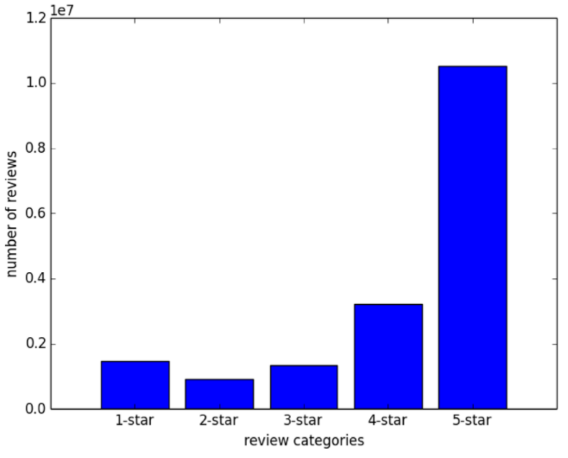

Figure 2.10: Data based on review categories [33]

2.3.3 Sankey diagram

Another concrete example of a visualization that has been used to visualize content is the Sankey diagram, as it is capable of representing multiple features or categories of an item at once. Also Sankey diagrams appear to be more readable. This is due to their flow structure, which generally follows a left-to-right orientation that can easily be understood by users. The diagram in Figure 2.12 are the results of a previous study interactive Visualizations of personalized review data for a hotel recommender system were designed. In Figure 2.12, Various colors (red and green) and symbols ("-" and "+"), respectively, are used for denoting positive and negative references. These graphical elements were used to help users perceive quickly the prevailing user sentiment on a given hotel characteristic. Figure 2.12 is explained as following [5]:

- In diagram (a) in Figure 2.12, characteristics of the hotel are mentioned by two users in their reviews about Hotel A.
- In diagram (b) in Figure 2.12, characteristics of the hotel are mentioned by a user in her hotel reviews.
- In diagram (c) in Figure 2.12, opinions regarding the location of two hotels have been aggregated based on users' travel category.
- In diagram (c) in Figure 2.12, subset of characteristic mentioned by a group of users who reviewed Hotel A.

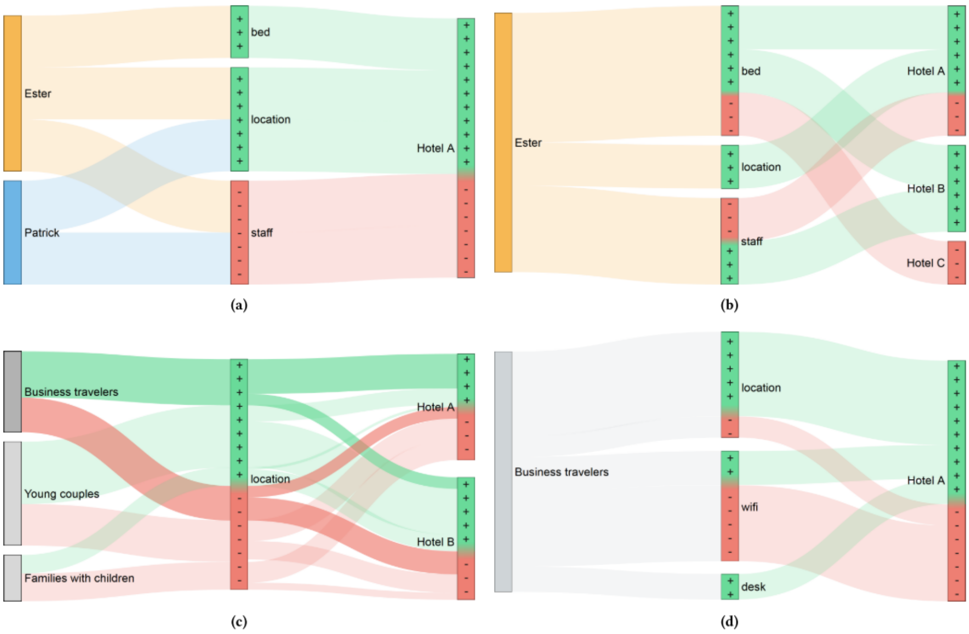

Figure 2.12: Example visualizations using Sankey diagrams [5]

In Figure 2.13, we can see the reviews being displayed as a separate layer on top of the Sankey visualization. Parts of the sentences in the reviews are highlighted that mention characteristic of the hotel. Green highlight indicates a positive mention of the hotel. Red highlight indicates a negative mention of the hotel. The same principle is used in thematic analysis. Where we highlight parts of sentences that describe certain characteristic of the product [34] [35]. This can be very beneficial for the user because he can easily access this information on demand.

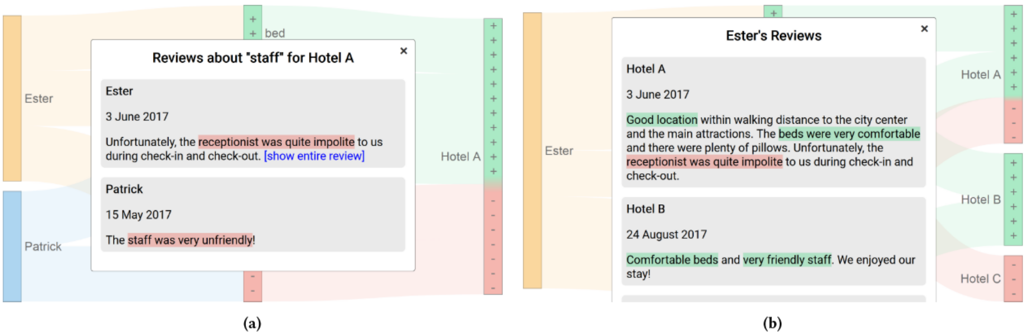

Figure 2.13: Reviews that are displayed as a separate layer on top of the Sankey visualization [5]

The authors concluded that these visualizations could provide a better understanding of the reviews usually written by the guests. This could help monitor and focus on areas that require improvement, i.e. characteristics with a lot of negative mentions [5].

2.3.4 Word clouds

Another concrete example of a visualization that has been used to visualize content are the word clouds, as they are capable of representing multiple features or categories of an item at once. Word clouds are common methods for visually displaying large volumes of text, presenting the material with the font size and color of words mapped to the word frequency, popularity or significance in a space-filling, informative and esthetically pleasing way [29].

Most word clouds, however, randomly organize the words. The randomness of word layout does not provide a meaningful representation of the data, even though they are useful and insightful tools. There are two main disadvantages to using word clouds [29]:

1. “It requires significant mental demand for users to understand the review content, because users need to scan the entire visualization to gain an overview or to find specific keywords of interest” [29].
2. “It only provides one dimension of information, such as word frequency, without semantic relationships among keywords, which is critical for understanding the review content” [29].

In this study an attempt was made to make word cloud visualization of user reviews that arranges semantically related words as spatially proximal, called ReCloud. The steps for their implementations is as follow [29]:

1. First, to create a semantic graph of review content, then use a natural language processing technique called grammatical dependency parsing.
2. Second, add the semantic graph to a force-directed layout, which produces a clustered word layout by decreasing the energy mode.

This will result in giving users more context of the review content and insight about the semantics. Also providing users an important additional dimension of information, so they can better comprehend reviews in a quick and easy manner [29].

Figure 2.14: ReCloud of a Yogurtland store near UC Berkeley campus [29]

2.4 Visual analytics methods

In order to assess the positive and negative views shared by consumers, we incorporated several techniques to interactively evaluate consumer feedback and reviews. D. Oelke et al [2] introduced the next three visual analytics techniques. Summary reports provide a quick overview of the customer reviews without reading them. A visualization shows the clusters of reviews in which similar opinions are expressed. Circular correlation maps uncover relationships between the user given score and various attributes. In the next sections, we briefly discuss these three techniques [2].

2.4.1 Visual Summary Reports

A quick analysis of the consumer feedback data collection is provided by visual summary reports. They show whether it belongs to a category of attributes with a positive tendency (in this case the color blue represents positive) or a category with a negative tendency for each attribute extracted by our automated algorithm (in this case the color red represents negative). Figure 2.15 is an example of a summary report of printers. The scale of the inner rectangles is calculated by the percentage of feedback that commented on the attribute, indicating from its assessment the value that the analyst can assign to this attribute. The proportion of positive or negative opinions, respectively, is mapped to a color. The measuring of the average percentage of positive feedback per attribute is done using an automated review process, and it is used as a threshold [2].

Compared to the other attributes, the attributes whose number of favorable feedback is above that threshold indicate a positive tendency (blue). The ones that are below the threshold reflect a negative tendency (red). The darker the color value gets, the greater the positive or negative tendency. The amounts of the set of positive or negative attributes decide the intervals for the four shades of red/blue tones [2].

In Figure 2.15 a visual summary report of reviews from Amazon.com is shown on three different printers, in total there were 1876 reviews analyzed. Furthermore for printer 1, the data is displayed separately for two different styles of printers. This makes it possible to examine the strengths and limitations of particular printer families in depth. It can be shown that there are certain characteristics that clients are normally happy with. This is the case for the attribute "printer", but also for other "photo, print, copy or quality attributes, as shown in Figure 2.15 [2].

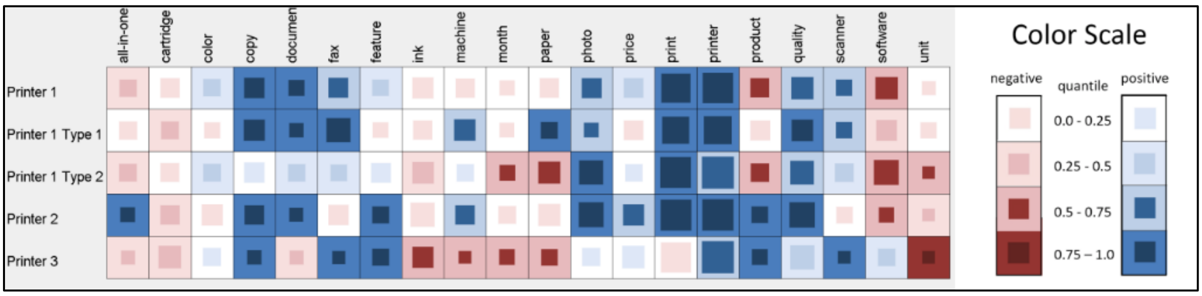

Figure 2.15: Summary Report of printers [2]

2.4.2 Cluster Analysis

The visualization in Figure 2.16 shows groups of customers with comparable views. The authors wrote: “this kind of visualization are important for companies that would like to learn about different groups of customers” [2].

In previous research [2], a hierarchical clustering algorithm is used to find the different groups of customer opinions. Each cluster representative is projected in 2D space using multidimensional scaling as a dimensionality reduction method. Each cluster is then visualized using a thumbnail image that displays the percentage of reviews in the cluster commenting on it, separated into negative and positive feedback for each attribute. The number of feedback found in the cluster is mapped (non-linear) to the scale of the thumbnails and the resulting Voronoi cell's grey color. Figure 2.16 illustrates an example in which consumer clusters of Printer 1 is shown. Figure 2.16 is interpreted as follow [2]:

- The biggest cluster is the one that summarizes all feedback that did not condemn any characteristics but were pleased with the printer.
- For cluster 2, you can see that there is a group of customers that don’t like most attributes.
- Cluster 3 summarizes the reviews that had an overwhelmingly positive sound, but still noted certain critical points.
- Clusters 4-6 provide consumer groups with a distinctly differentiated view of the product..

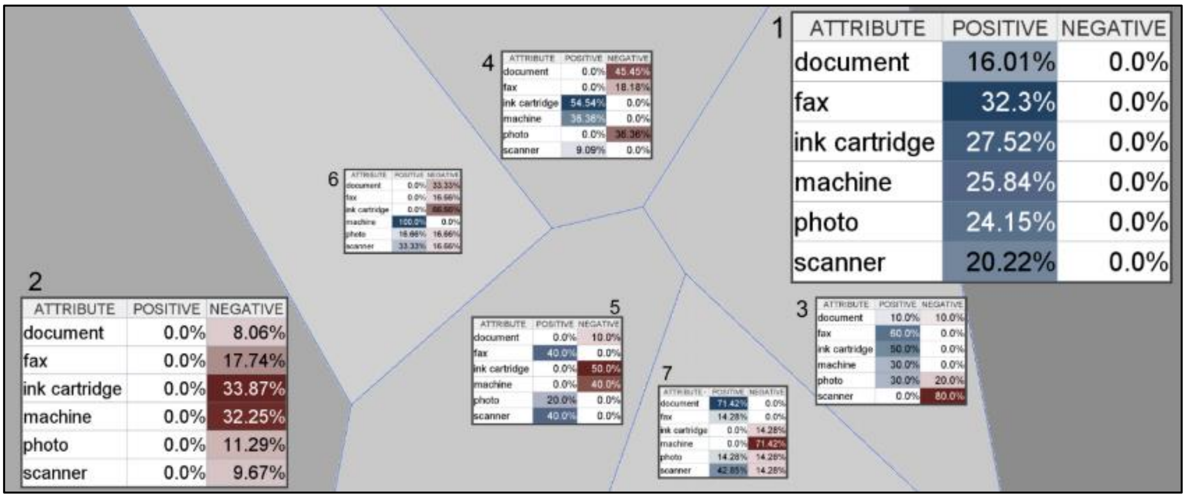

Figure 2.16: Scatterplot of customer reviews on printers [2]

2.4.3 Circular Correlation Map

The Circular Correlation Map gives a comprehensive view of the data. It was implemented in [36] and can be used to identify similarities between the various elements of the data collection [2]. An implementation example in which input was analyzed from consumers who purchased a notebook in an online store as seen in Figure 2.17-2.19. Figure 2.17-2.19 can be interpreted as follows [2]:

- For each attribute commented on a line by the client, it is taken from the position in the right semicircle of the document ID to the corresponding score value in the center of the analysis and from the score to the attribute location in the left semi-circle.
- The color of a line is determined by the opinion that was expressed on the attribute.
- The percentage of positive comments is determined if several lines are on top of each other and the lines are colored accordingly (color = blue).
- The line width indicates the number of lines on top of each other.

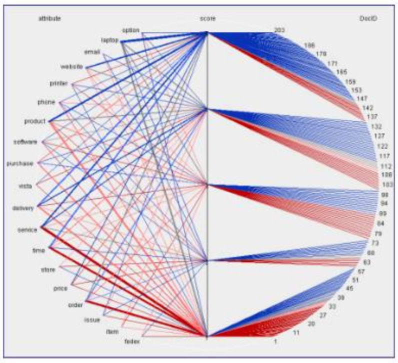

Figure 2.18: All Customer Comments. Most comments have an overall positive tendency [2].

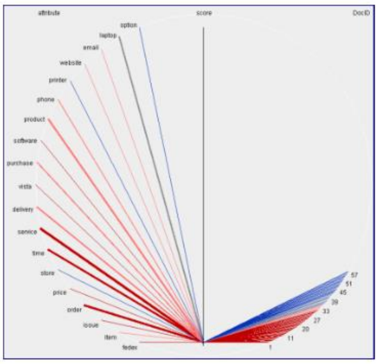

Figure 2.17: Analyzing the score 1 comments: Service is one of the attributes that is often mentioned negatively [2].

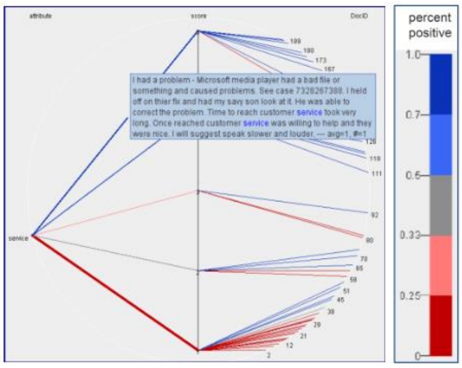

Figure 2.19: Subset of comments on service. Not all the customers are dissatisfied with the service. But this was a hot topic in the score 1 comments [2].

2.5 Color combinations

One out of every twelve men sees color differently than the majority of the population. Affected by a condition sometimes misrepresented as colorblindness, these men misinterpret those colors that can be distinguished by other people. Their definition of color, saturation, and brightness vary. This means when choosing color for our visualizations, we have to take the colorblindness and similar conditions into account [37].

2.5.1 Deuteranopia

Deuteranopia is a type of red-green color blindness characterized by the inability to distinguish red and green pigments [38]. Figure 2.20 illustrates how readers with normal color vision see colors and readers with deuteranopia interpret color. The right side shows how red-green impaired readers confuse these colors. That's why in previous study they used blue and red color combinations to visualize negative and positive reviews. Figure 2.21 show color combinations that are bad and good for people with deuteranopia.

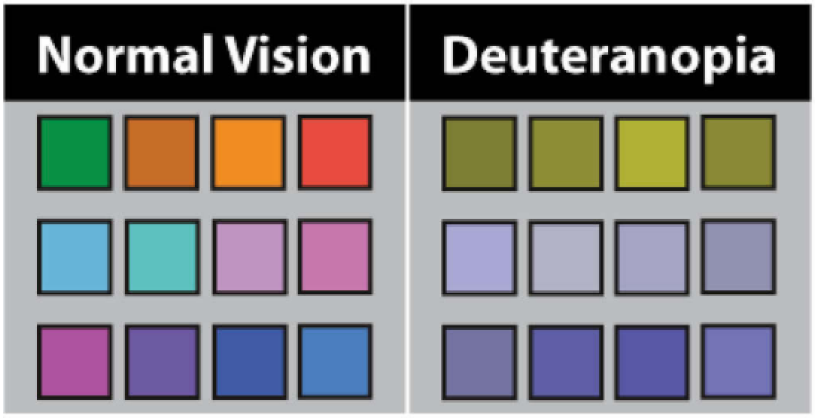

Figure 2.20: Colors as they appear to readers with normal vision and deuteranopia [37]

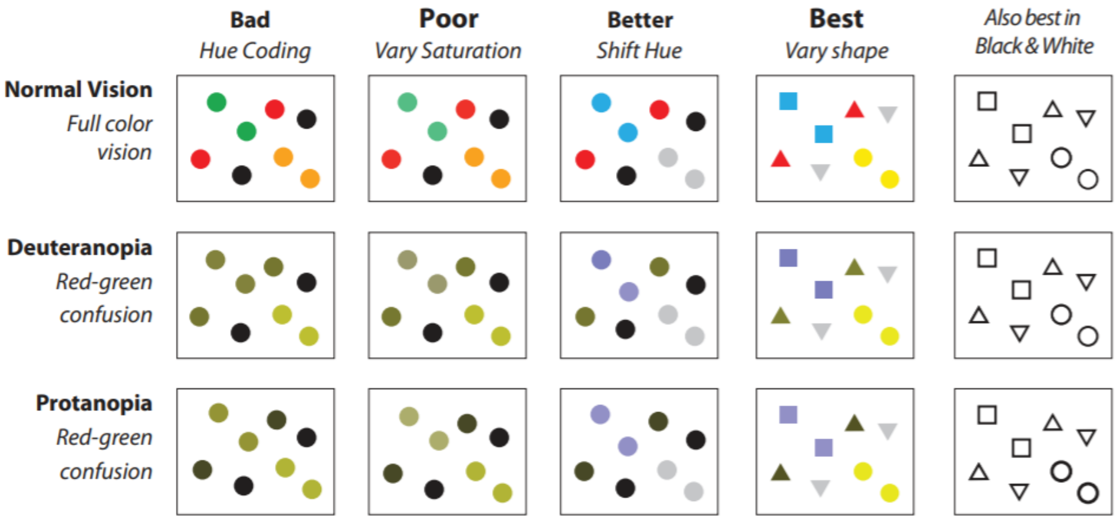

Figure 2.21: Point classes typical of a dot map distinguished by saturation, hue and shape [37]

2.6 Research gap

Previous work in content visualization mainly focuses on visualization of reviews and different techniques to visualize the results [2] [29] [5] [32]. In most applications, like the ones mentioned above, the end user was able to see the reviews being visualized in different types of visualizations.

So we can conclude that there exist various manners to visualize customers reviews. Knowing which visualizations would be the best for the end user to visualize these reviews is where the gap lies in existing literature. Identifying which visualization would be best-suited to display positive and negative reviews for the end user and which would be the most understandable are the focus of this thesis. The aim is to develop an interface with different visualizations to visualize reviews and to ask questions to the end user in order to conclude which visualization is the best. As a starting point we decided to go for simple visualizations because we believe that it's easier for user to interpreted.

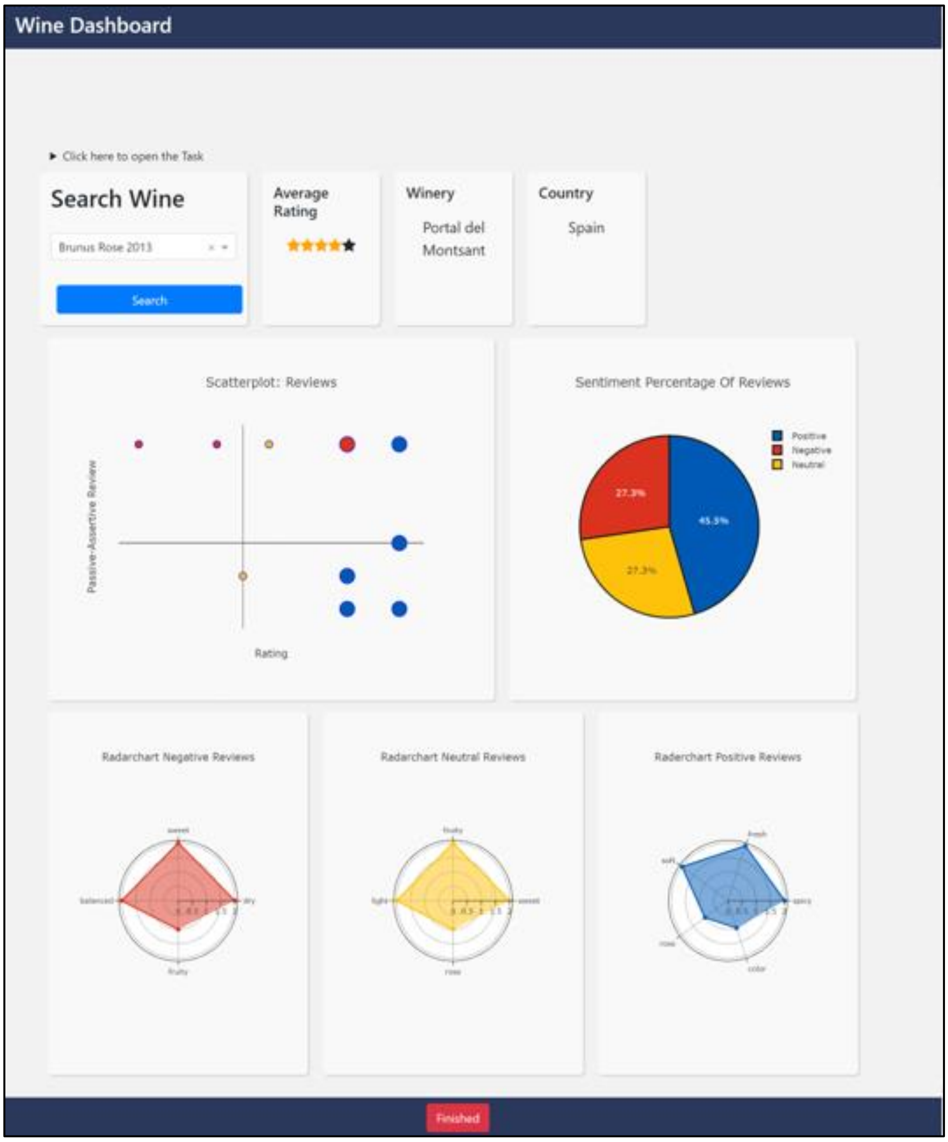

Figure 2.22: plotly interface

2.7 Conclusion

During this chapter, it became clear that there exists a broad spectrum of different visualizations methods/techniques that offer the ability to visualize the customer reviews, each in a distinct manner. These approaches vary from more simple visualizations like bar charts [33] to complex visualization like a circular correlation map or a heatmap [2].

As seen, each type of the mentioned methods has been used in numerous applications that each have varying visualizations. For example the sentiment analysis method has been used in almost all the applications above to distinguish positive, negative, and neutral reviews from each other.

Future work should focus on improving the interfaces with more complex visualizations and conduct a future user study. This should offer a better understanding of which visualization would be best-suited to display positive and negative reviews for the end user

Chapter 3

Evaluation methods

The evaluation of the visualization will happen by conducting a user study. First, participants will be asked to provide some personal information. After providing this information, participants will be asked to use the interface and then extract information from the visualization. After each interface, participants will be asked to fill in a System Usability Scale (SUS) questionnaire [39]. This is a 10-Question questionnaire that offers a quick, cost-effective yet accurate way to evaluate the usability of a website. In the end, a customer questionnaire will be shown to form a final evaluation.

3.1 Think aloud study

First a think aloud study was conducted on the first version of the implementation. This provides a clear way of seeing what information or graph is used during problem solving. All participants will be asked to verbally clarify their thought process while using the application. The main advantage of this approach is that in this process only a small group needs to be evaluated as the aim is to gather rich, in-depth data.

The participants were mostly hand-selected in order to get participants with diverse academic backgrounds, for example economics, computer science, physiotherapist, etc. They have no previous knowledge about the functionality or subject of the study and certainly no previous experience. Once they have participated in the think aloud study, they of course can't participate in the next study. The goal of this phase is to find large flaws due to bad visualizations, flaws against the expected logic/flow of the application, measure the duration of the study and evaluate the clarity of the content.

3.2 User study

These participants were also selected intentionally to include participants with different academic backgrounds. They have no previous knowledge about the functionality or subject of the study and certainly no previous experience. To know which type of subject is using the system and undergoing this instance of the evaluation, a custom questionnaire is presented to retrieve some personal demographic data. This questionnaire contains some basic questions, e.g.:

- What's your gender?
- How often do you buy wine on average?
- Do you read the reviews when you are buying wine?
- …

In appendix A.2 one can find all the question asked to the participants. Only the essential questions are asked because participants are often concerned about their privacy. This custom questionnaire is also asked because it is important to know if the participants have ever read reviews, or have ever searched for a wine before. While using the system the participants will be presented by a total of 5 questionnaires. Because there are 4 interfaces, after each interface the participant is presented with a SUS-questionnaire [39]. After completing the 4 SUS-questionnaires, in the end one custom questionnaire is presented to get a final conclusion of the participant which visualization he or she thought was the best.

The SUS-questionnaire [39] is used as an indication of the usability of the system/interface. A score of 100 is determined using the content of the questionnaire. SUS also considers many other factors to determine usability, such as the environment of the program and the users that communicate with it. [39] This is necessary as ISO 9241-11:2018 defines that usability should also cover other factors like effectiveness, efficiency, and satisfaction [40] [41]. In appendix A.1 one can find the presented questionnaire concerning.

In comparison with the SUS-questionnaire, the custom questionnaire gives a more detailed result describing which visualizations the end user found the most fitting to visualize the characteristics in reviews. To avoid participants that do not comply with the goals of the study, results that do not surpass a certain time threshold or answer the questionnaire in a contradictory fashion are filtered out. The participants can only navigate to the next interface after a certain amount of time. This is implemented to make sure that the participant doesn't rush through the study just to get it over with. This approach of having the participants evaluating all the interfaces has some downsides:

- **Fatigue**: the evaluation of all 4 interface and total of 6 questionnaire can lead to participant being fatigue [42]. This could add to the scenario that the participant finishes the experiment in an unwanted hurry.
- **Experience**: as mentioned before, the user does not have previous knowledge about the functionality or subject of the study and certainly no previous experience. When different versions of the system are presented sequentially the result would be influenced by a higher (unwanted) experience factor [43]. This problem will be solved by using counterbalanced design.

As mentioned before, all the participants will have to evaluate each interface. To reduce the chances of the order of treatment or other factors adversely influencing the results, counterbalanced design is used [44]. In Figure 3.1 is used as an example to display in which order the interfaces are presented to each group of participants.

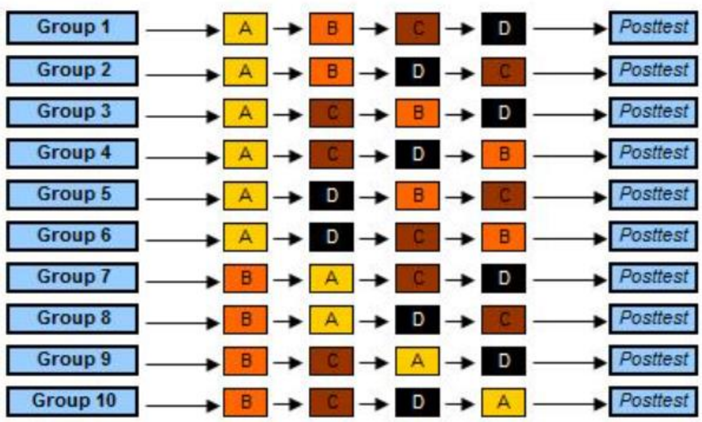

Figure 3.1: counterbalanced design, four conditions [44]

3.3 Custom questionnaire

As mentioned at the end of the experiment, the participants will be presented a custom questionnaire. The goal of the questionnaire is to get a final conclusion of the participant which visualization he or she thought was the best, the worst and the most understandable. For every answer, respondents were asked why they came to that conclusion. A thematic analysis is used on their reasoning to get a final result [34]. The result are discussed in chapter 5. Figure 3.2 was shown in the custom questionnaire and the following questions were asked:

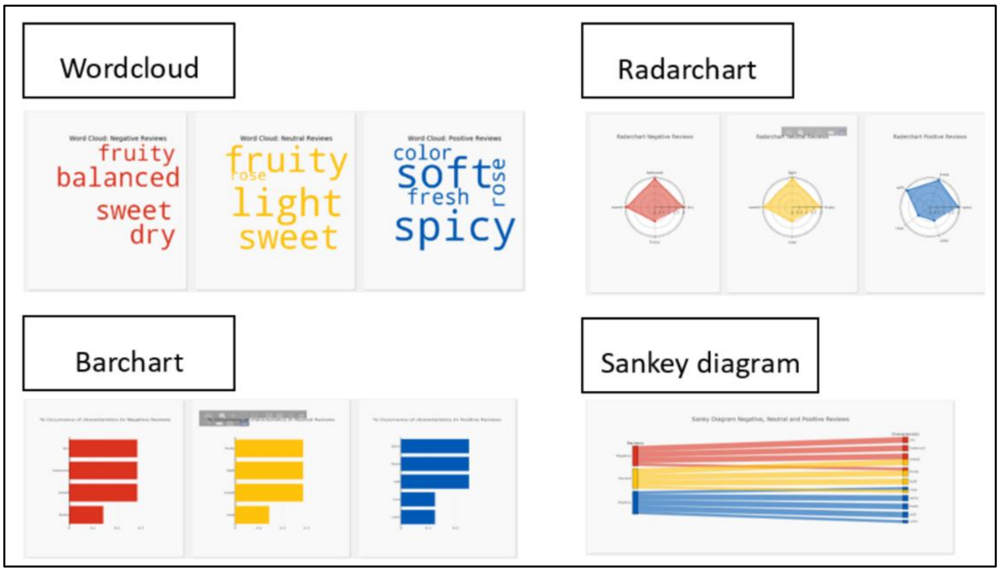

Figure 3.2: Different visualizations used to visualize wine characteristics

Q1) Which of the four visualizations helped you to find the wine characteristics the fastest? Please explain your answer.

Q2) Which of the four visualizations took you the longest to interpret the wine characteristics? Please explain your answer.

Q3) Which of the four visualizations did you find the best to visualize the wine characteristics from positive, negative and neutral reviews ? Please explain your answer.

Q4) Which of the four visualizations did you find the worst to visualize the wine characteristics from positive, negative and neutral reviews? Please explain your answer.

Q5) Which of the four visualizations would you like to see when you are buying wine online? Please explain your answer.

Q6) Which of the four visualizations did you understand the most? Please explain your answer.

Q7) Do you find it useful to visualize the wine characteristics from positive, negative and neutral reviews? Please explain your answer.

Q8) Do you have further suggestions for improvement on this page and please be critical?

3.4 Thematic analysis

Thematic analysis is a method of analyzing qualitative data. It is typically applied to a collection of texts, such as interview transcripts or in this case reviews. The researcher carefully analyses the data to recognize common themes – subjects, concepts and trends of interpretation that come up frequently. A thematic analysis can be performed in the following six-step process [34]:

1. **Familiarization**: the first step is to get to know our data. Before we start analyzing individual items, it is necessary to get a thorough summary of all the data we gathered.
2. **Coding:** coding involves highlighting parts of our reviews, usually sentences or phrases, and describing their content with shorthand labels or "codes".
3. **Generating themes:** looking through the codes that we have created, finding patterns within them, and starting to come up with themes.
4. **Reviewing themes:** making sure that our themes are useful and accurate representations of the data.
5. **Defining and naming themes:** it's time to name and describe each of them now that you have a final list of themes. Defining themes includes formulating precisely what each theme means and figuring out how it allows one to understand the data.
6. **Writing up:** finally, we will write down our analysis of the data.

Thematic analysis is a good technique to find out something about people's views, opinions, knowledge, experiences or values from a set of qualitative data – for example, product reviews. [34].

3.5 Flow of the procedure

Figure 3.3 shows a strictly guided sequence that was included in the application, to be able to test the application properly. The user cannot deviate from the sequence shown in Figure 3.3. The flow starts with a short information page that states the goals and subject, which also is a consent-form in which the subjects agree to participate. By doing this he/she gives permission to use all the data that is gathered from using the system. After agreeing to the consent-form, the tutorial is shown. This tutorial aims to guide the participant through the subsequent steps.

After the tutorial video, the participants will be directed to a custom questionnaire where some personal information is asked. This is done because it is important to know if the participants have ever read reviews, or have ever searched for a wine before. Completing these first steps, the participants then start using the implementation. The participant is asked to complete a task in each interface and after completing the task, the participant has to fill in a SUS-questionnaire. After going through all the interfaces, a custom questionnaire is asked to get a final conclusion.

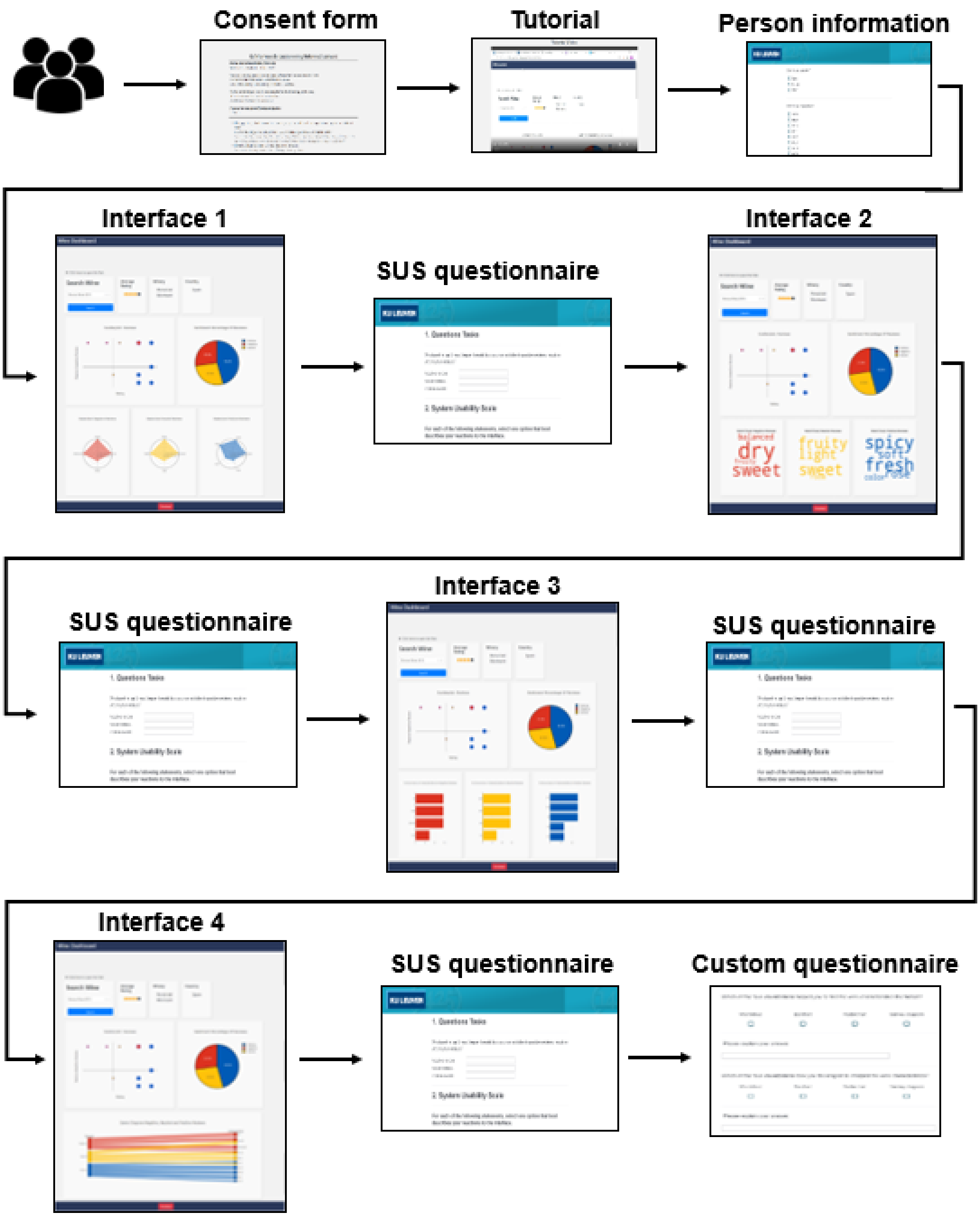

Figure 3.3: Guided sequence of pages

Chapter 4

Implementation

This chapter explains the implementation of the application in detail. Justification for each technology that was used for the application is provided and including a short description of the technology itself.

4.1 Software stack

4.1.1 Ploty

Plotly is a platform to create data visualization. Python, R, Julia, and a few other languages can be used for it. It's very user-friendly, has detailed documentation, and helps you to create very powerful graphs and maps in a few lines of code. Interactive, publication-quality diagrams are made by Plotly's Python graphing library. Some examples are error bars, box plots, line plots, scatter plots, histograms, area charts, bar charts, heatmaps, subplots, polar charts, multiple-axes and bubble charts. Figure 4.1 and 4.2 are examples of a radar chart and Sankey diagram using Plotly [45].

One strength of Plotly is that Dash renders web applications as a "single-page app". This implies that when the user navigates the application, the application does not fully refresh, making browsing very easy. Another strength and the main reason why Plotly was used is the ability to use Dash to deploy graphs on the internet as web applications. This allowed easy deployment of the application [45].

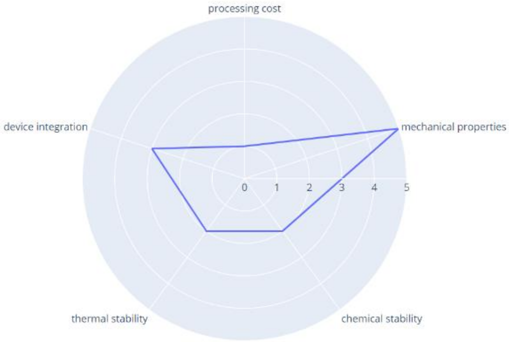

Figure 4.1: Example radar chart [45]

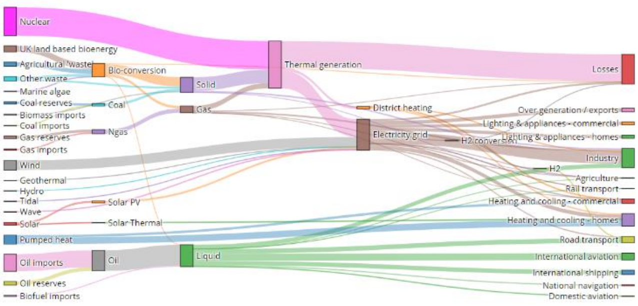

Figure 4.2: Example Sankey diagram [45]

4.1.2 Qualtric

Qualtric is a web-based survey tool that is easy to use to perform survey studies, evaluations and other data collection activities. This research suite can be used by anybody to develop surveys, submit surveys and review responses - all from any online location, everywhere you want to [46]. Qualtric is used in combination with Plotly, to analyze and save all the data. The main reason why Qualtric was used is because it's easy to use, free, can be used as an database, and is web based.

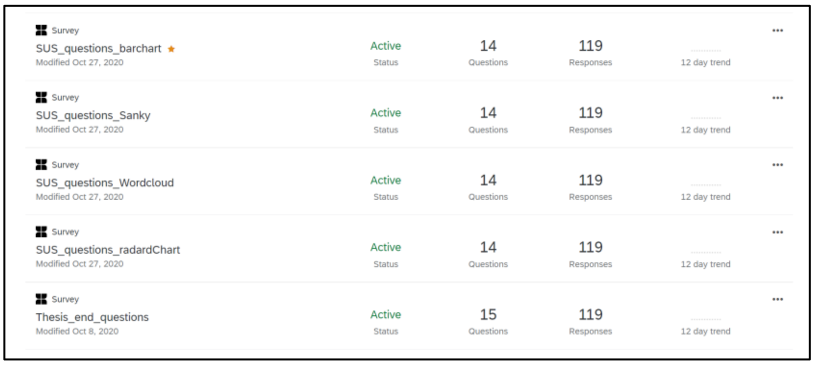

Figure 4.3: Qualtric survey's

The Qualtrics survey tool is entirely web based and offers the following features [46]:

- It's free
- No need to install software
- Easy to use point and click interface - Process is simple and fast, and results are easy to access!
- Anyone can create surveys with graphics, complex branching and randomization
- Many question and survey templates to choose
- Tools for basic and advanced users and surveys
- Multiple surveys can be posted at the same time and in 48 different languages
- Ability to export data directly to SPSS, CSV, PDF, Word, Excel, and PowerPoint
- ….

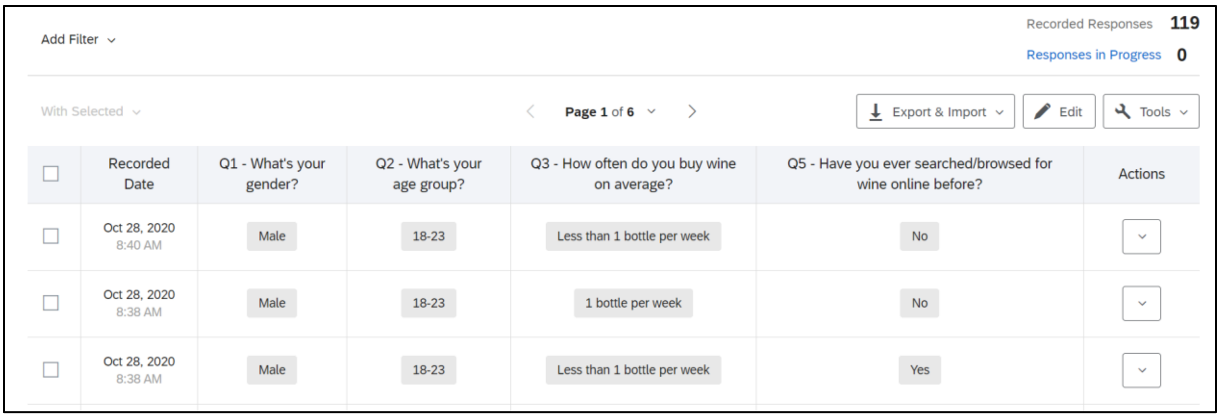

Figure 4.4: Qualtric individual answers

4.1.3 Matplotlib

Matplotlib is a Python library used for data visualization. Matplotlib is a 2D plot library that produces good quality Figures. It consists of several plots like line, bar, scatter and histogram e.t.c. Matplotlib uses an object oriented API to embed plots in Python applications. The main reason why Matplotlib was used in combination with plotly is because plotly doesn't support word clouds. For the implementation of the word clouds Matplotlib was used. Matplotlib was also used to generate graphs for the result of the data, which will be discussed in Chapter 5.

```
def create_wordcloud(n_clicks, input_value):
  if input_value is not None:
    specific_wine = search the reviews for the specific wine
    positive_reviews = filter only reviews where the sentiment is positive
    if (len(positive_reviews) != 0):
      characteristics = filter all the irrelevant words and only keep the characteristics
      characteristics_occurrence = count occurence characteristics
      if (len(characteristics_occurrence) != 0):
        create wordcloud
      else:
        create wordcloud wich says "no positive characteristics found"
```

Figure 4.5: Pseudo code, creating word cloud using Matplotlib with Plotly

4.2 Data

4.2.1 Data collection and preparation

The Vivino Review Dataset was provided by Vivino with permission to use it for research purposes. This dataset contains user reviews of wines on the Vivino website. This data set consists of 3 parts. The first part contains about approximately 381 000 reviews spanning 13 years of product review data. It contains information related to the user providing the review, the time related information and the characteristics of the product. The second part consists of about 4 million products which consists of product description, brand name etc. In order to obtain conclusive results, a subset of this enormous dataset was used for analysis. The third part was a user dataset but this dataset wasn't used to analyze the reviews.

4.2.2 Data description

In order to understand the structure and schema of the dataset, let us look at a sample Vivino Review. As seen in Fig. 1.2, an Vivino's User Review consists of four important aspects:

- Review text : The actual content of the review.
- Rating : User rating of the product on a scale of 1 to 5.
- Date : The time when the review was placed.

Figure 4.6: Review example Vivino website

These aspects will help us understand and analyze the reviews in order to derive insights. Figure 4.7 shows an example of the data shown in json format, its show clearly which keys are available to use in this study. The most keys that where mostly used throughout the study are the following:

- Rating: the rating of the review.
- Note: the comment of the user.
- Created_at: the date when the review is created.


```
{
   "rating": 4.0,
   "language": "en",
   "vintage": {...},
   "created_at": "2014-08-08T23:01:45.000Z",
   "note": "Very good. Well balanced and lots of berries and cherry, smoothly accompanied by
   "flavor_word_matches": [{...}],
   "user": {...},
   "activity": {...},
   "id": 14014016
}, {
   "rating": 3.5,
   "language": "en",
   "vintage": {...},
   "created_at": "2018-02-25T02:27:30.000Z",
   "note": "Really impressed by this one. Vibrant and flavorful with formidable structure an
   "user": {...},
   "activity": {...},
   "id": 89016423
},
```


Figure 4.7: Review example json form

4.2.3 Data cleaning

The provided Vivino user review dataset was originally in JSON format. There was no standard schema for this dataset. For example some of the review texts were more than 5 kb (~2 pages of text) while some other reviews were much shorter. Pandas library wasn't able to read the provided JSON file because the JSON objects were not valid. So the first challenge was to make the JSON objects valid. The code below was used to make the JSON objects valid.

```
from itertools import islice
import re


with open("wine_input.txt", encoding="utf8") as wine_file:
        head = list(islice(wine_file, 1000))


outF = open("wine_output.txt", "w", encoding="utf8")
for line in head:
        line = re.sub(r''"'"'(?=[^"]*"[^"]*(?:"[^"]*"[^"]*)*$)', '', line)
        outF.write(line)
outF.close()
```

Figure 4.8: Python transform_data.py

After transforming the data, the JSON file was converted to a CSV file to make it easier for analyzing and viewing the data.

```
with open('json_reviews.json', encoding='utf-8-sig') as f_input:
        df = pd.read_json(f_input)

result = pd.json_normalize(df["Allreviews"])
result.to_csv('wineReviews.csv', encoding='utf-8', index=False)
```

Figure 4.9: Python json_to_csv.py

The result was a CSV file where all the names of the name/value pairs were column names and every JSON object is a row were every name/value pair is written in the corresponding cell.

	vintage.wine.name	rating	language	created_at	note
0	Reserve Zinfandel	4.0	en	2018-02-28T21:19:47.000Z	Ws: a lot of oak smells, sweet spices, cherries. Dry, but sweet fruit. Mediur
1	Reserve Zinfandel	3.5	en	2018-03-02T16:06:01.000Z	Medium ruby color. Medium+ nose intensity. Red berries, oaky elements w
2	Zinfandel	4.0	en	2017-11-10T13:34:40.000Z	Violet bouquet. Blackcurrant, black cherry and blueberry with cocoa flavou
3	Zinfandel	4.0	en	2014-08-08T23:01:45.000Z	Very good. Well balanced and lots of berries and cherry, smoothly accomp
4	Zinfandel	3.5	en	2018-02-25T02:27:30.000Z	Really impressed by this one. Vibrant and flavorful with formidable structu
5	Napa Valley Cabernet Sauvignon	4.5	en	2016-08-06T02:01:25.000Z	Yummy .....$30 bucks local store was steal
6	Napa Valley Cabernet Sauvignon	5.0	un	2018-05-27T18:05:59.000Z	Good
7	Brunus Rose9	4.0	en	2018-04-29T10:29:24.000Z	A marvelous soft, inviting cherry colored rose that shifts gears on the palat
8	Brunus Rose9	4.0	en	2015-08-24T20:15:25.000Z	Smooth and nice

Figure 4.10: CSV data file of user reviews

4.3 Final version

In Chapter 2 there were several options to visualize characteristics of a product. The first version of the implementation was spread out over 3 different pages. Figure 4.11-4.13 shows the first version. During the think aloud study it was determined that this implementation was over complicated and took too long to go through all the tasks. This is why it was decided to choose simple visualizations for this study and recommend more complex visualizations for future study.

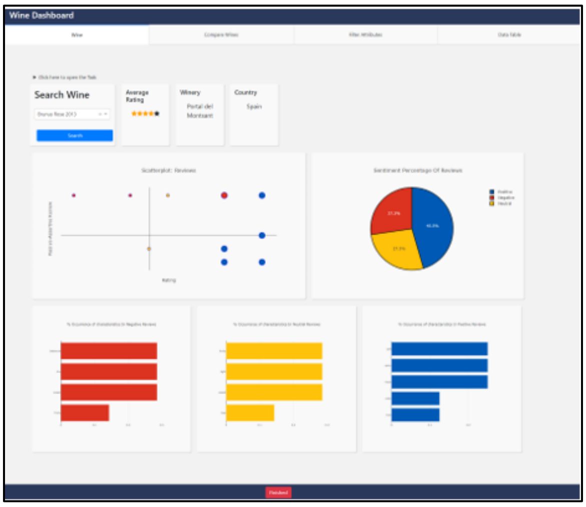

Figure 4.11: Page one, visualizations for one specific wine

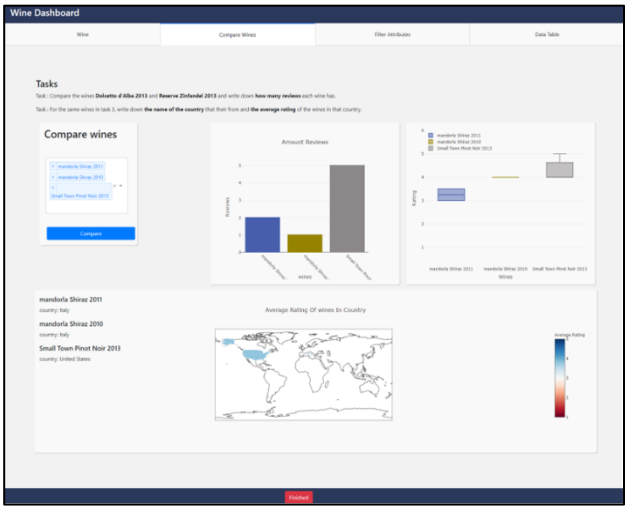

Figure 4.12: Page two, comparing wine

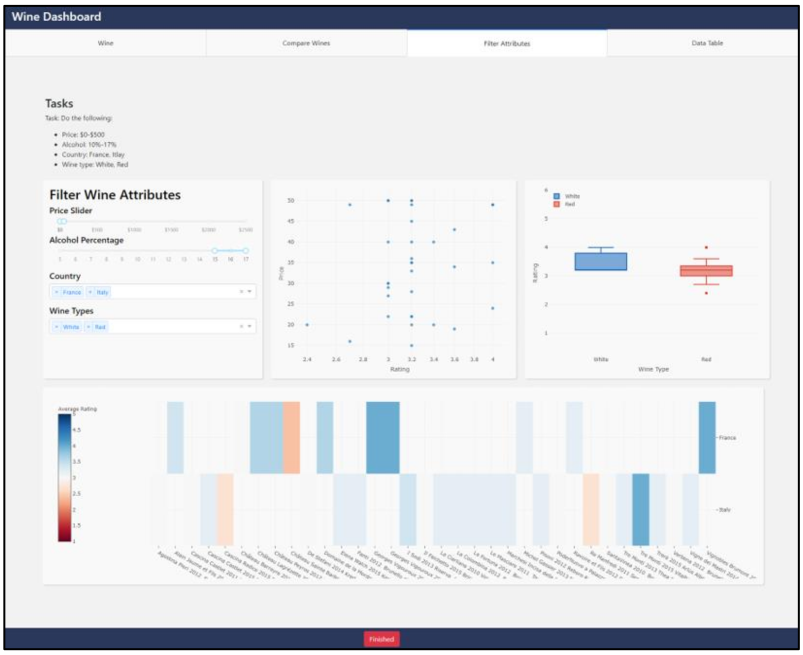

Figure 4.13: page three, filtering wine attributes

The final visualization was designed to keep it as simple as possible as over-complication is detrimental for the trust and thus usage as seen in the think aloud study [47].

The final implementation was made out of 3 sections (each row is considered one section). Figure 4.8 shows all the sections. The first section lets you search for a specific wine and gives you general information about the wine. In section 2, the results of the searched wine are displayed in a scatterplot and a pie chart. This is a clear way of visualizing multiple reviews at the same time and also displaying the percentage of the amount of positive, negative and neutral reviews. The text review will be displayed if you hover over the scatterplot as seen in Figure 4.8. In section 3, the results are presented in radar chart, bar char, Sankey diagram or word cloud. As mentioned in Chapter 2, these visualizations display multiple features of data at the same time in a clear way.

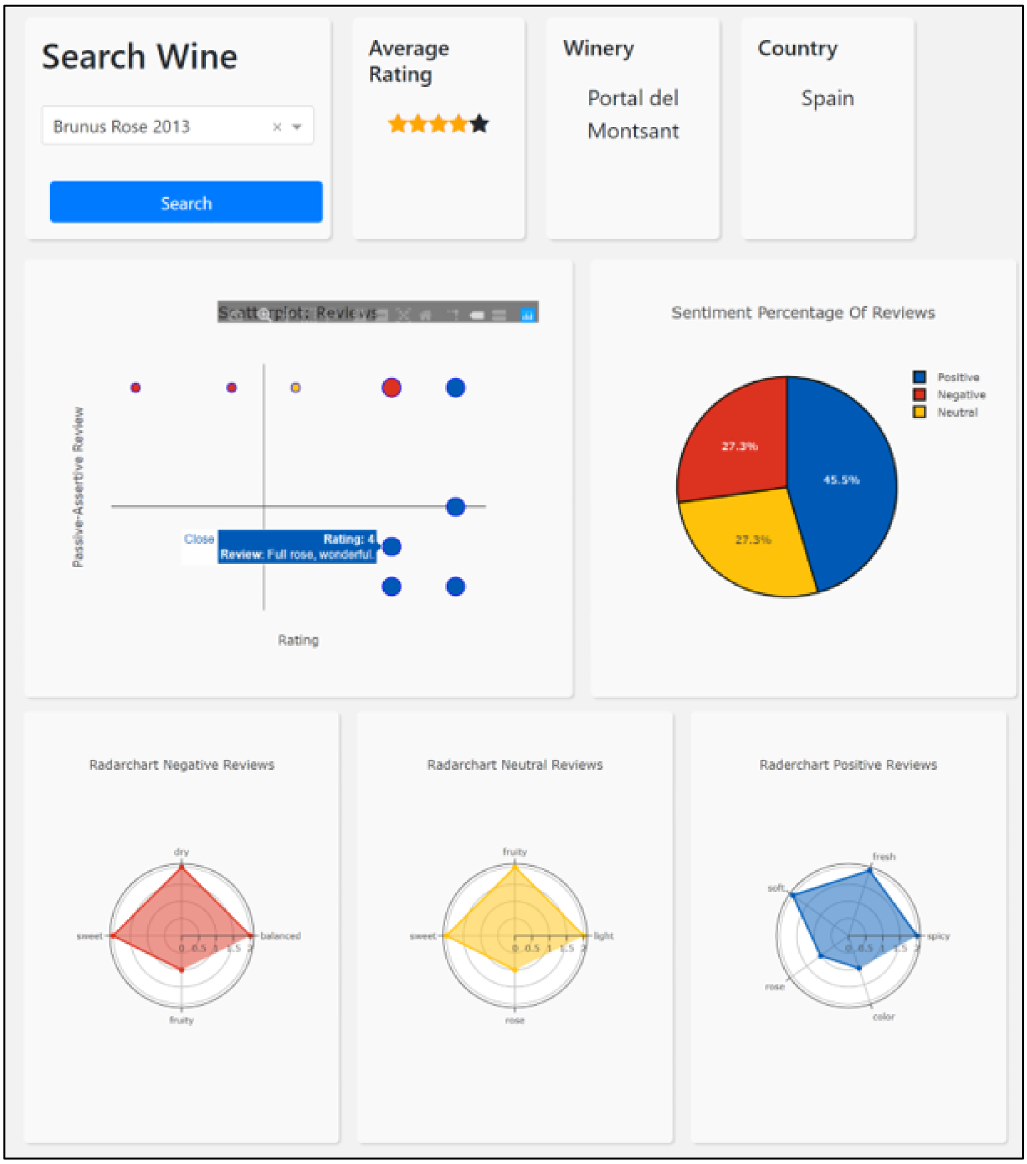

Figure 4.14: Final implementation

Figure 4.9 – 4.11 shows the other variants using the bar chart, word cloud and Sankey diagram in a similar way as the radar chart but differs based on the visualization used. The aim is to consider another visualization which is known to be able to represent attributes of a given item, similarly to how radar graphs are able to.

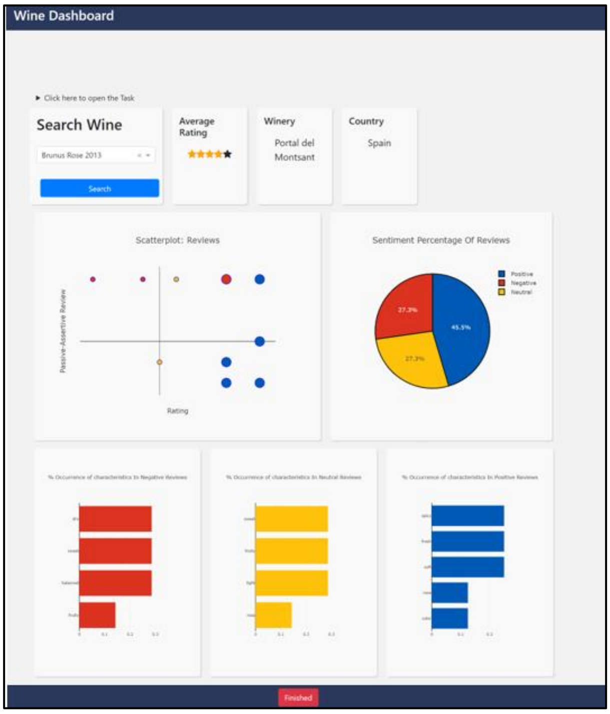

Figure 4.17: Final implementation using bar chart

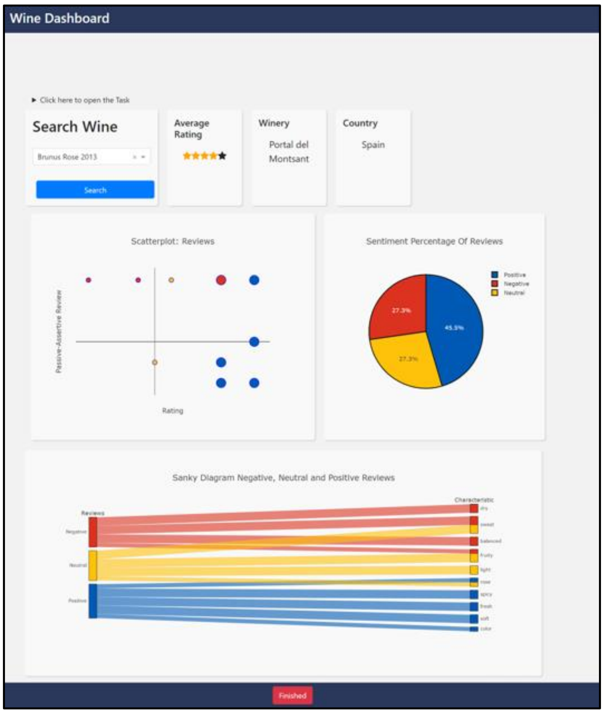

Figure 4.16: Final implementation using Sankey diagram

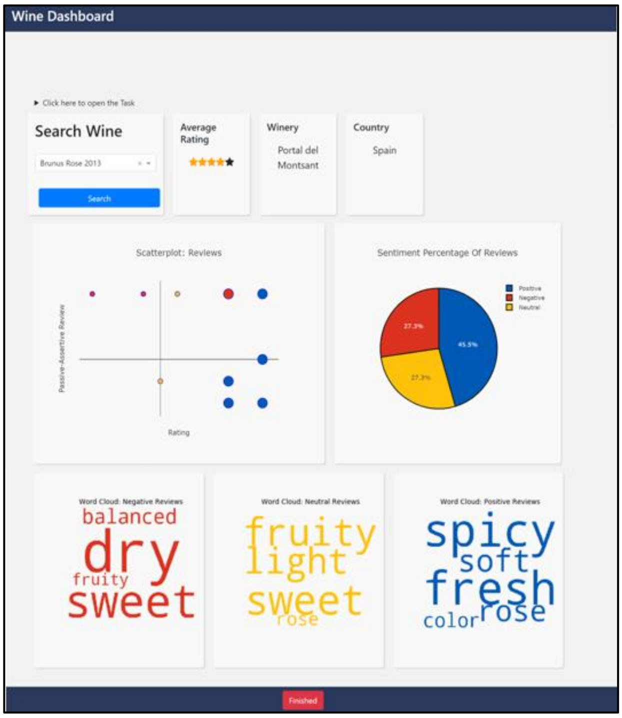

Figure 4.15: Final implementation using word cloud

Chapter 5

Evaluation results

5.1 First evaluation: Think aloud

During this phase, a small group of people, which were hand-picked and were asked to evaluate the first prototype. A total of 15 people participated in the think aloud study. The participants were selected to include participants with diverse backgrounds. They have no previous knowledge about the functionality or subject of the study and certainly no previous experience. Once they have participated in the think aloud study, they of course can’t participate in the next study. The goal of this phase is to find large flaws due to bad visualizations, flaws against the expected logic/flow of the application, measure the duration of the study and evaluate the clarity of the content.

5.1.1 Results and changes

The consent form and the tutorial page were generally well received. Only some small notes regarding grammar mistakes. The first page was mostly very well received, because of the use of simple visualizations. One participant gave the comment: “it’s like that the task is part of the page”. This was solved by using a dropdown for displaying the task. Page one is shown in Figure 4.11.

The second page shown in Figure 4.12, was not very clear for the majority of the participants. This was because of the boxplot visualization, which most of the participants didn’t quite understand. This was also mentioned during the feedback presentation that this type visualization is very complicated for people without a technical background. Therefore it was decided to remove this page because it took too long the understand this page, the duration of the study was too long and mainly because it had minimal contributions to the research questions.

The third page shown in Figure 4.13, was also not very clear for the majority of the participants. This was because of the boxplot visualization, which most of the participants didn’t quite understand. The heatmap which was also hard to comprehend for some participants because they were not familiar with the visualization. Therefore it was decided to remove this page because it took too long the understand this page and the duration of the study was too long. This choice of removing these complex visualization led to the decision that this study will mainly focus on comparing simple visualizations as a starting point because we believe that it's easier for user to interpret.

Some questions contained words that were unclear in the SUS and custom questionnaire, such as 'cumbersome'. For some participants with no technical background found that some more technical words were unclear. The cause of this problem is probably due to the fact that English is not the native language of the participants. The content of the questionnaires cannot, however, be altered to preserve meaning.

The conclusion from all this is that the application has to be smaller by removing some pages and visualizations. This will help to keep the duration of the study to max 25 to 30 minutes, otherwise it was almost 1 hour until the person understood and finished all the task. Here and there was a need to resolve a small bug in the interface, but nothing major came up. The content of the questionnaire was kept the same to preserve meaning and some additional information should clear up any uncertainties.

5.2 Second evaluation: User study

During this phase, a much larger group was sourced to evaluate the system. The total amount of participants for this study was 119. Meaning, each participants evaluated each version of the implementation, see the entire ‘flow’ on Figure 3.3. The only requirements to be eligible to test the system was being of age (18+).

5.2.1 Information participants

Figure 5.1 shows the amount of men, woman and other that have participated in this study. When looking at Figure 5.1, out of the 119 participants 93 were men, 25 woman and 1 other.

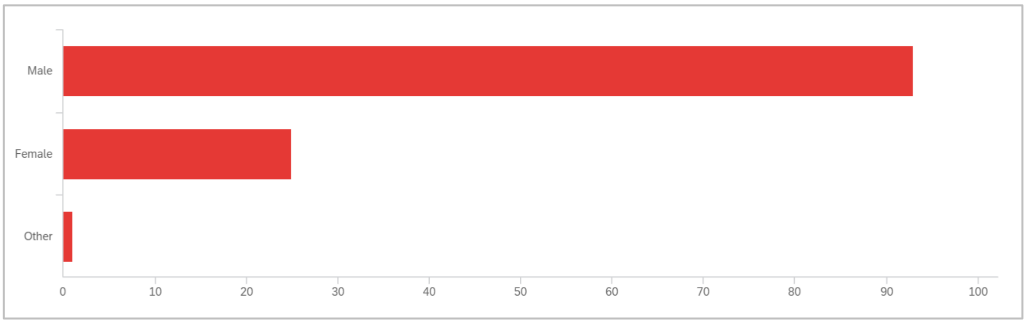

Figure 5.1: Bar chart question: “What's your gender?"

Figure 5.2 shows the amount of users in different age groups that have participated in this study. Between the ages 18 and 23, there were 102 participants. Between the ages 24 and 29, there were 9 participants. Between the ages 30 and 35, there were 3 participants. Between the ages 36 and 41, there were 4 participants. Between the ages 48 and 53, there was 1 participant.

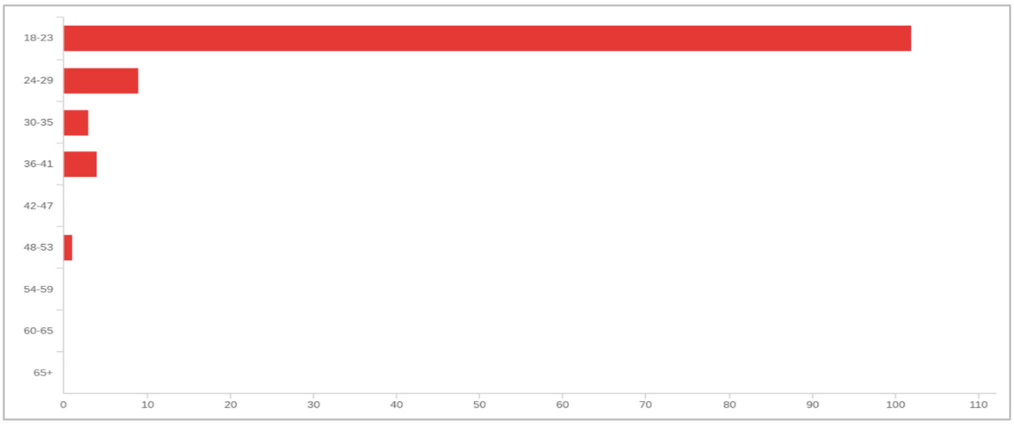

Figure 5.2: Bar chart question: “What's your age group?"

Figure 5.3 shows how often the participants buy wine on average. In all, 43 participants buy less than 1 bottle per week, 12 participants buy 1 bottle per week, 1 participant buys more than one bottle per week, and 63 participants don't drink wine. It's worth noting that the majority of the participants doesn't drink wine.

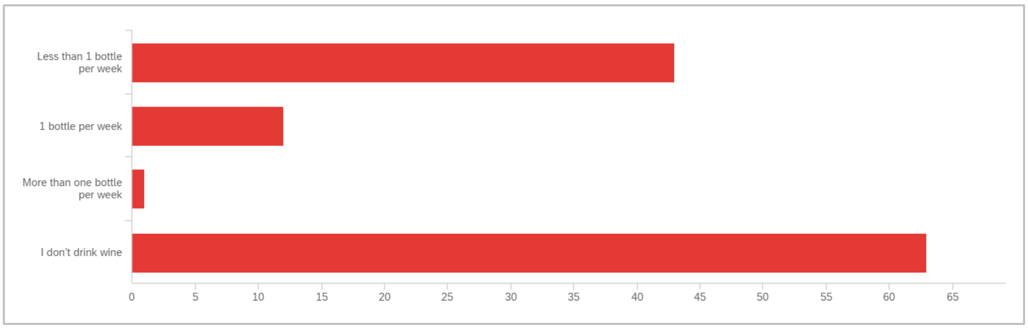

Figure 5.3: Bar chart question: "How often do you buy wine on average?"

Figure 5.4 shows how many participants have ever searched/browsed for wine before. Figure 5.4 shows that, 35 participants have searched for wine before and 84 participants have never searched for wine before online.

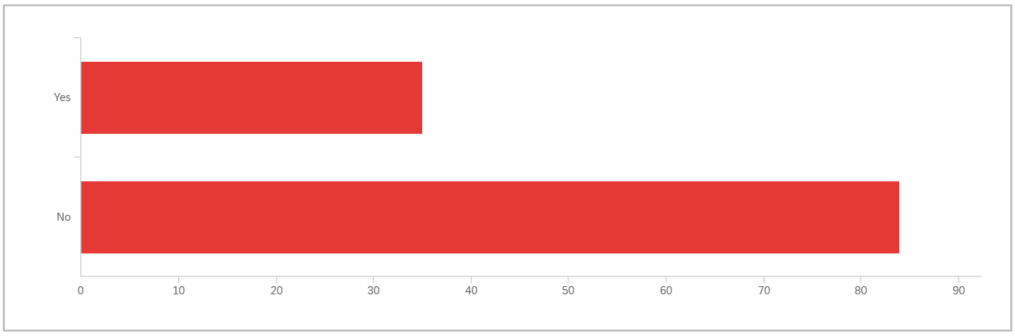

Figure 5.4: Bar chart question: "Have you ever searched/browsed for wine online before?"

Figure 5.5 shows how many participants read the reviews when they are buying wine. This question is only shown for the participants if they answered "yes" to the question, "Have you ever searched/browsed for wine online before?". Figure 5.5 shows that, 14 participants read reviews when they are buying wine, 11 participants don't read reviews when they are buying wine, and 10 participants sometime read the reviews.

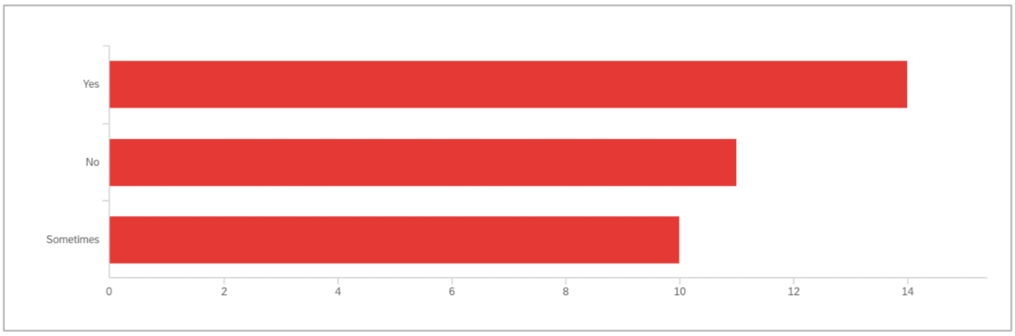

Figure 5.5: Bar chart question: "Do you read the reviews when you are buying wine?"

Figure 5.6 shows how many participants prefer to read negative, positive, or both kind of reviews. This question is only shown for the participants if they answered “yes” to the question, “Have you ever searched/browsed for wine online before?”. Figure 5.6 shows that, 3 participants prefer to read negative reviews, 2 participants prefer to read positive reviews and 30 participants prefer to read both negative and positive reviews.

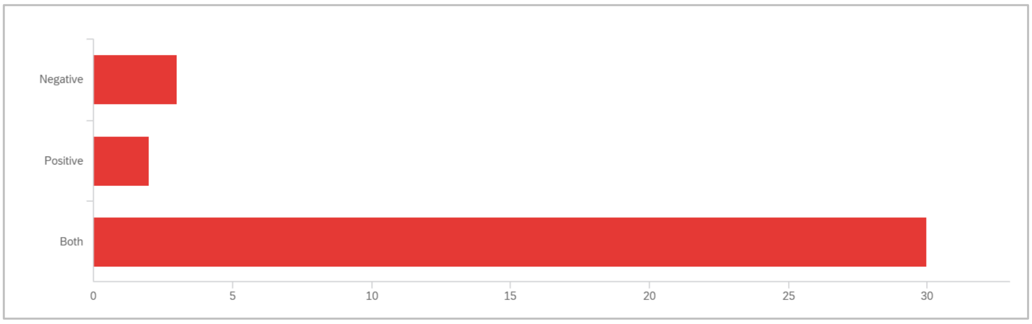

Figure 5.6: Bar chart question: "Do you prefer to read the negative reviews, the positive reviews or both?"

Figure 5.7 shows when participants consider a review negative. Figure 5.7 shows that, 11 participants consider a review negative when it has only one rating, 75 participants consider a review negative when it has a 2 star rating and 33 participants consider a review negative when it has a 3 star rating.

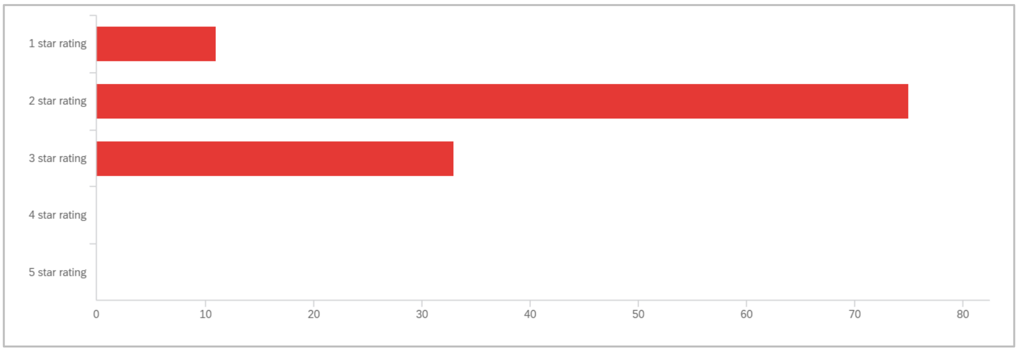

Figure 5.7: Bar chart question: “When do you consider a review negative?”

5.2.2 System usability score

After each interface, each participant had to fill out a SUS-questionnaire. This gives an insight into the usability of the system. The score was calculated with the formula in Figure 5.8. A score under 68 means that there is room for improvement and anything above that means the system is accepted [39].

$$((SQ_1 - 1) + (5 - SQ_2) + (SQ_3 - 1) + (5 - SQ_4) + (SQ_5 - 1) + (5 - SQ_6) + (SQ_7 - 1) + (5 - SQ_8) + (SQ_9 - 1) + (5 - SQ_{10})) * 2.5$$

Figure 5.8: Score formula SUS-questionnaire (SQ_x= score of question x)

Results SUS questionnaire

The median of the scores with the standard deviation are shown in the Table below and are visualized in Figure 5.9. By comparing the four visualizations, the bar chart score notably better than the other visualizations and is the more 'usable' approach. The Sankey diagram scores the lowest of all four visualizations. The reasoning why one visualization has a high score or a low score will be discussed in the next section.

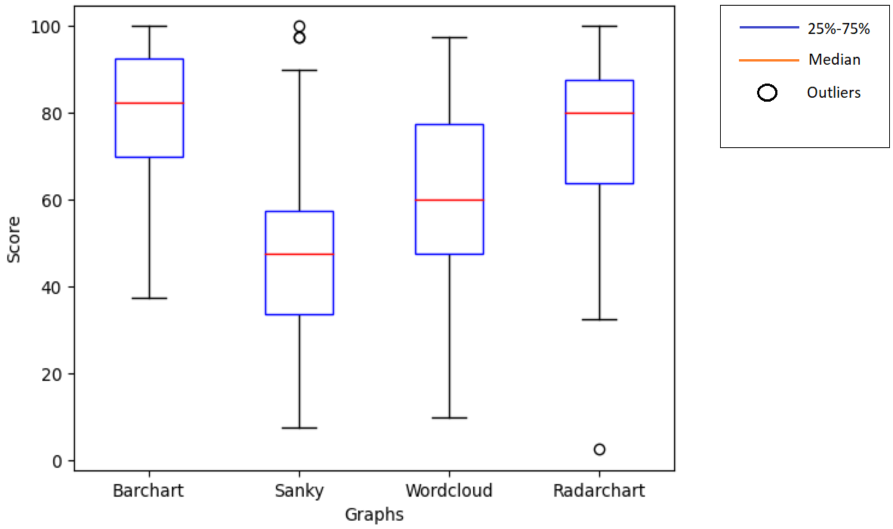

Figure 5.9: Box-plots: score of SUS-questionnaire

Visualization	Median	Standard deviation
Bar chart	82,5	16,62
Sankey diagram	47,5	21,05
Word cloud	60	18,63
Radar chart	80	17

Table 5.1: SUS-questionnaire results

5.2.3 Results custom questionnaire

As explained in Chapter 3, at the end of the study each participant is shown a custom questionnaire where they are asked to choose which visualization was the best or the worst overall. In this section a detailed description of the results for each question in the custom questionnaire will be discussed. As mentioned in Chapter 3, a thematic analysis is used for each explanation for their answer.

Identification of characteristics

***Q1)** Which of the four visualizations helped you to find the wine characteristics the fastest? Please explain your answer.*

Figure 5.10 shows that the majority of the participants voted for bar chart. Meaning that the bar chart is the visualization that helped participants find the wine characteristics the fastest.

In Table 5.2, we can see that 78 participants out of the 119 voted for bar chart. The participants were able to choose more than one option, this is why the total of votes are more than 119.

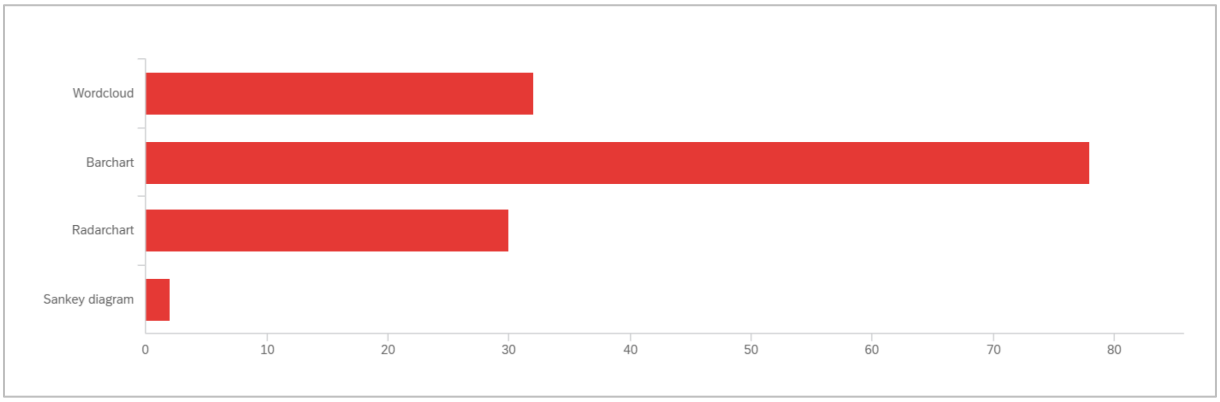

Figure 5.10: Bar chart Q1 results

Visualization	Count
Word cloud	32
Bar chart	78
Radar chart	30
Sankey diagram	2

Table 5.2: Results Q1 custom questionnaire

As mentioned before, to find out why the majority of the participants choose bar chart a thematic analyses is conducted on the explanation of the participants. The following themes were defined for the bar chart visualization:

Codes	Theme
• Understandable • Straight forward • Obvious	Understandability
• Usability • Simple to comparing the characteristics	Practicality
• Easy to read • Clear and simple	Simplicity
• Visually pleasing	Aesthetically pleasing
• Familiar	Familiarity

Table 5.3: Result thematic analysis Q1 for bar chart visualization

Description of the themes:

- **Understandability**: this covers everything that has to do with why participants found the visualization easy to understand. This will help to understand if understandability of the visualization helped the find wine characteristics the fastest.
- **Practicality**: this covers everything that has to do with why participants found the visualization easy to use. This entails finding and comparing data out of the visualization. This will help to understand if the practicality of the visualization helped the find wine characteristics the fastest.
- **Simplicity**: this covers everything that has to do with why participants found the visualization simple. This will help to understand if the simplicity of the visualization helped the find wine characteristics the fastest.
- **Aesthetically pleasing:** This covers everything that has to do with participants finding the visualization aesthetically pleasing. This will help to understand if "aesthetically pleasing" is the reason that made the majority of the participant choose this visualization. As the visualizations that helped them to find the wine characteristics the fastest.
- **Familiarity:** This covers everything that has to do with participants finding the visualization familiar. This will help to understand if the familiarity of the visualization helped the find wine characteristics the fastest.

In Figure 5.11, the codes "simple to comparing the characteristics", "easy to read" and "clear and simple" were mentioned the most times compare to other codes. This results indicates that it was the simplicity and the practicality for the majority of the participants that made them prefer the bar chart.

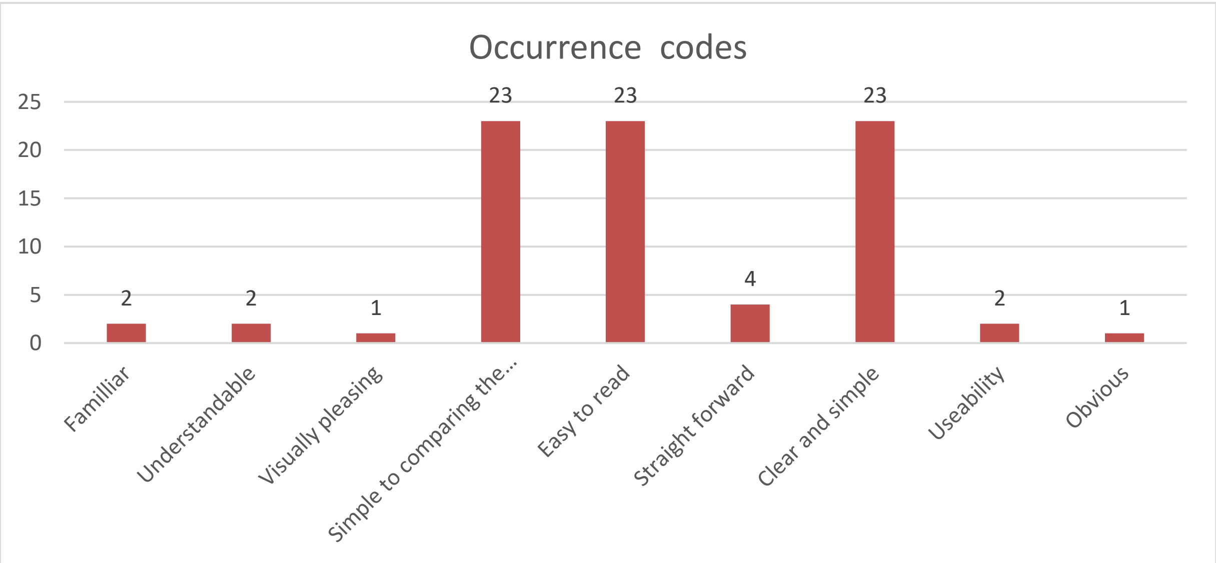

Figure 5.11: Bar chart occurrence codes for Table 5.3

Q2) *Which of the four visualizations took you the longest to interpret the wine characteristics? Please explain your answer.*

Figure 5.12 shows that the majority of the participants voted for Sankey diagram. Meaning that the Sankey diagram is the visualizations that took the longest time to find the wine characteristics. In Table 5.4, we can see that 75 participants out of the 119 voted for Sankey diagram. The participants were able to choose more than one option, this is why the total of votes are more than 119.

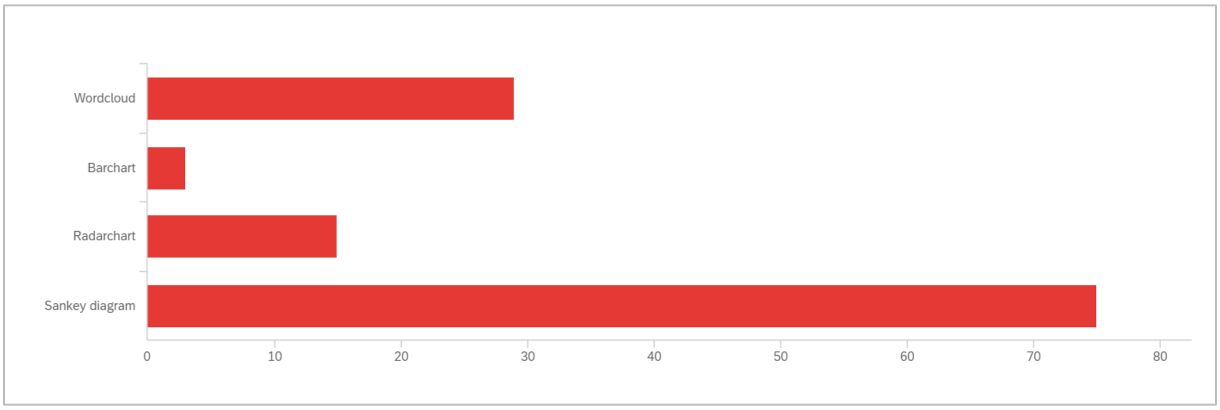

Figure 5.12: Bar chart Q2 results

Visualization	Count
Word cloud	29
Bar chart	3
Radar chart	15
Sankey diagram	75

Table 5.4: Results Q2 custom questionnaire

Table 5.5 is the result of conducting a thematic analyses on the explanation of question 2. The following themes were defined for the Sankey diagram visualization:

Codes	Theme
• Not understandable • Difficult to see at first glance • Difficult to read • Hard to interpret	Not understandable
• Unclear • Complex • Confusing	Complex
• Difficult to compare characteristics	Difficult to compare values

Table 5.5: Result thematic analysis Q2 for Sankey diagram visualization

Description of the themes:

- **Not understandability**: this covers everything that has to do with why participants found the visualization difficult to understand. This will help to understand if the lack of understandability of the visualization was the reason that made the majority of the participants choose this visualization. As the visualization that took the longest to interpret the wine characteristics.
- **Complex**: this covers everything that has to do with why participants found the visualization complex. This will help to understand if the complexity of the visualization was the reason that made the majority of the participants choose this visualization. As the visualization that took the longest to interpret the wine characteristics.
- **Difficult to compare values**: this covers everything that has to do with why participants found it difficult to compare values in the visualization. This will help to understand if the usability was the reason of that made the majority of the participants choose this visualization. As the visualization that took the longest to interpret the wine characteristics.

In Figure 5.13, the codes "Not understandable", "unclear" and "complex" where mentioned the most times compared to other codes. These results indicate that it was the lack of understandability and complexity for the majority of the participants that made them choose the Sankey diagram.

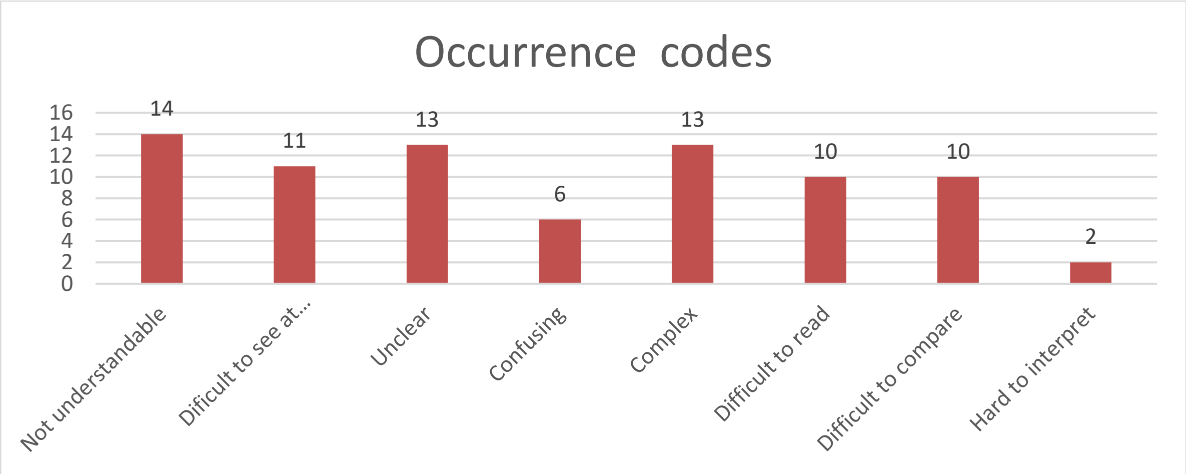

Figure 5.13: Bar chart occurrence codes for Table 5.5

Usability

***Q3)** Which of the four visualizations did you find the best to visualize the wine characteristics from positive, negative and neutral reviews ? Please explain your answer.*

Figure 5.14 shows that the majority of the participants voted for bar chart and radar chart. Meaning that those two visualizations are the best to visualize the wine characteristics from positive, negative and neutral reviews. In Table 5.6, we can see that 56 participants out of the 119 voted for bar chart and 49 for radar chart. The participants were able to choose more than one option, this is why the total of votes are more than 119.

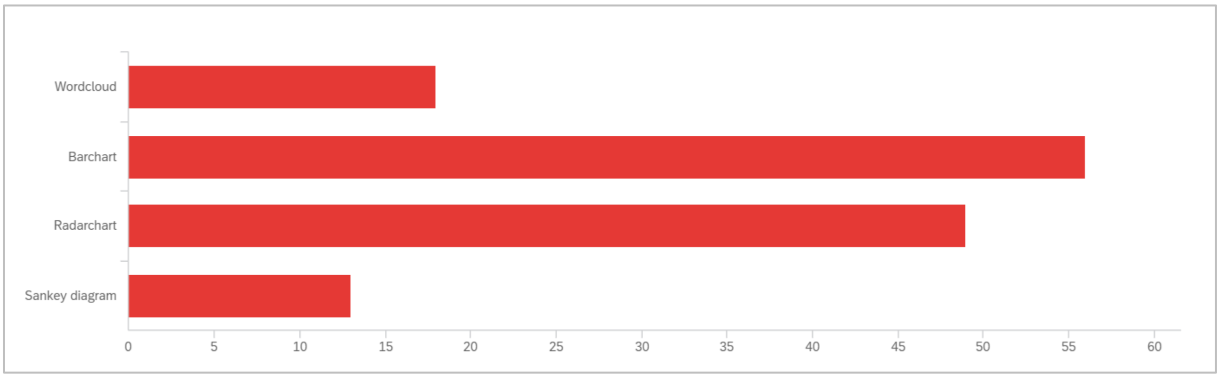

Figure 5.14: Bar chart Q3 results

Visualization	Count
Word cloud	18
Bar chart	56
Radar chart	49
Sankey diagram	13

Table 5.6: Results Q3 custom questionnaire

Table 5.7 is the result of conducting a thematic analyses on the explanation of question 3. The following themes were defined for the bar chart visualization:

Codes	**Theme**
• Understandable • Straight forward	Understandability
• Simple to comparing the characteristics	Practicability
• Easy to read • Clear and simple	Simplicity
• Visually pleasing • Organized	Aesthetically pleasing
• Familiar	Familiarity

Table 5.7: Result thematic analysis Q3 for bar chart visualization

Description of the themes:

- **Understandability**: this covers everything that has to do with why participants found the visualization easy to understand. This will help to understand if understandability of the visualization is the reason that made the majority of the participant choose this visualization. As the best visualization to visualize the wine characteristics from positive, negative and neutral reviews.
- **Practicality**: this covers everything that has to do with why participants found the visualization easy to use. This entails finding and comparing data out of the visualization. This will help to understand if practicality of the visualization is the reason that made the majority of the participant choose this visualization. As the best visualization to visualize the wine characteristics from positive, negative and neutral reviews.
- **Simplicity**: this covers everything that has to do with participants found the visualization simple. This will help to understand if simplicity of the visualization is the reason that made the majority of the participant choose this visualization. As the best visualization to visualize the wine characteristics from positive, negative and neutral reviews.
- **Aesthetically pleasing:** This covers everything that has to do with participants finding the visualization aesthetically pleasing. This will help to understand if “aesthetically pleasing” is the reason that made the majority of the participant choose this visualization. As the best visualization to visualize the wine characteristics from positive, negative and neutral reviews.
- **Familiarity:** This covers everything that has to do with participants finding the visualization familiar. This will help to understand if “familiarity” is the reason that made the majority of the participant choose this visualization. As the best visualization to visualize the wine characteristics from positive, negative and neutral reviews.

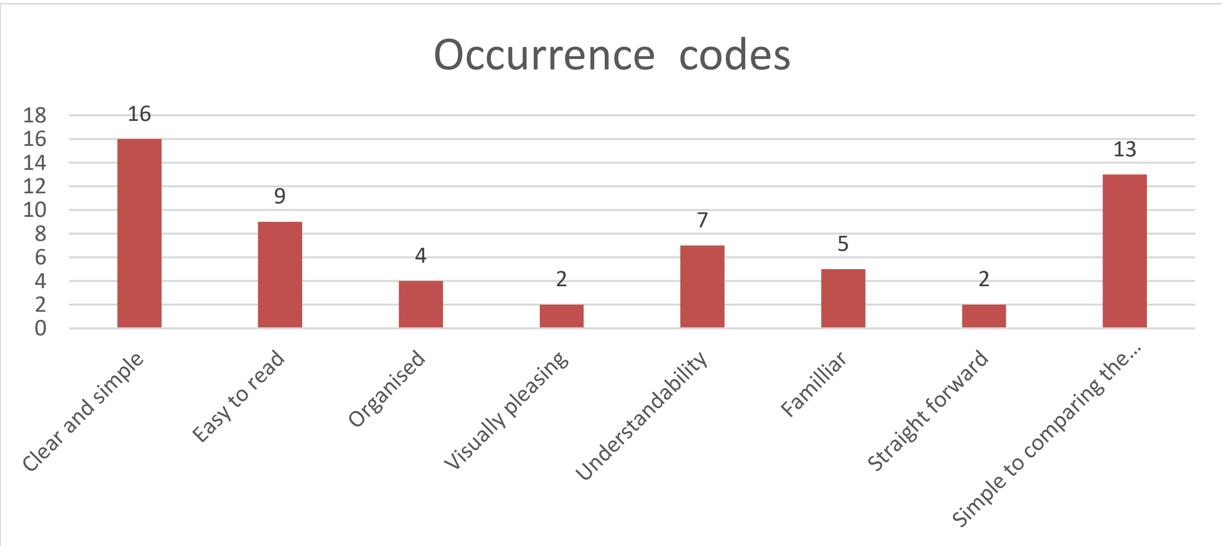

Figure 5.15: Bar chart occurrence codes for Table 5.7

In Figure 5.15, the codes “simple to comparing the characteristics”, “easy to read” and “clear and simple” were mentioned the most times compare to other codes. These results indicate that it was the simplicity and the practicality for the majority of the participants that made them prefer the bar chart.

Q4) *Which of the four visualizations did you find the worst to visualize the wine characteristics from positive, negative and neutral reviews? Please explain your answer.*

Figure 5.16 shows that the majority of the participants voted for bar chart and word cloud. Meaning that these two visualizations are the worst to visualize the wine characteristics from positive, negative and neutral reviews. In Table 5.8, we can see that 67 participants out of the 119 voted for Sankey diagram and 35 voted for word cloud. The participants were able to choose more than one option, this is why the total of votes are more than 119.

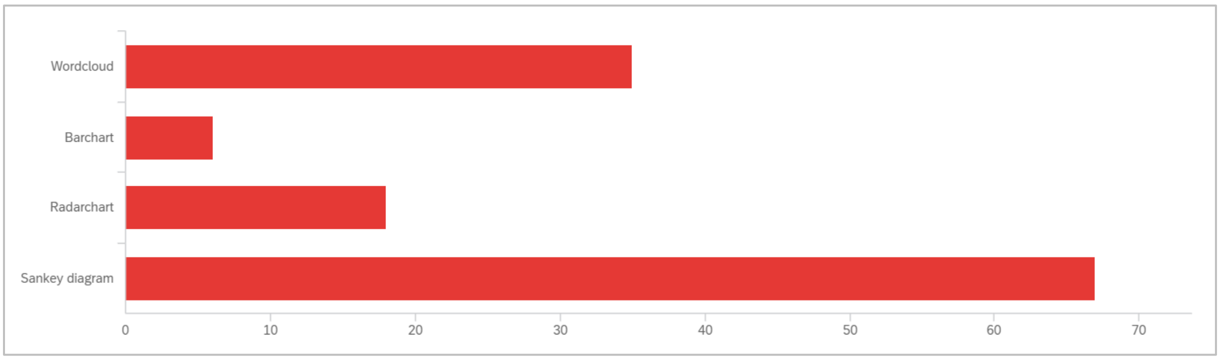

Figure 5.16: Bar chart Q4 results

Visualization	Count
Word cloud	35
Bar chart	6
Radar chart	18
Sankey diagram	67

Table 5.8: Results Q4 custom questionnaire

Table 5.9 is the result of conducting a thematic analyses on the explanation of question 4. The following themes were defined for the Sankey diagram visualization:

Codes	Theme
• Not understandable • Difficult to see at first glance • Difficult to read • Hard to interpret	Not understandable
• Unclear • Complex • Confusing	Complex

• Difficult to compare	Difficult to compare values
• Unpleasing visualization	Aesthetically unpleasing
• Uncommon	Unfamiliarity

Table 5.9: Result thematic analysis Q4 for Sankey diagram visualization

Description of the themes:

- **Not understandability**: this covers everything that has to do with why participants found the visualization difficult to understand. This will help to understand if the lack of understandability of the visualization was the reason that made the majority of the participants choose this visualization. As worst visualization to visualize the wine characteristics from positive, negative and neutral reviews.
- **Complex**: this covers everything that has to do with why participants found the visualization complex. This will help to understand if the complexity of the visualization was the reason that made the majority of the participants choose this visualization. As worst visualization to visualize the wine characteristics from positive, negative and neutral reviews.
- **Difficult to compare values**: this covers everything that has to do with why participants found it difficult to compare values in the visualization. This will help to understand if the this was the reason of that made the majority participants choose this visualization. As worst visualization to visualize the wine characteristics from positive, negative and neutral reviews.
- **Aesthetically unpleasing:** This covers everything that has to do with participants finding the visualization aesthetically unpleasing. This will help to understand if “aesthetically unpleasing” is the reason that made the majority of the participant choose this visualization. As worst visualization to visualize the wine characteristics from positive, negative and neutral reviews.
- **Unfamiliarity:** This covers everything that has to do with participants finding the visualization unfamiliar. This will help to understand if “unfamiliarity” is the reason that made the majority of the participant choose this visualization. As worst visualization to visualize the wine characteristics from positive, negative and neutral reviews.

In Figure 5.17, the codes “Not understandable”, “unclear” and “complex” were mentioned the most times compared to other codes. These results indicate that it was the lack of understandability and complexity for the majority of the participants that made them choose the Sankey diagram.

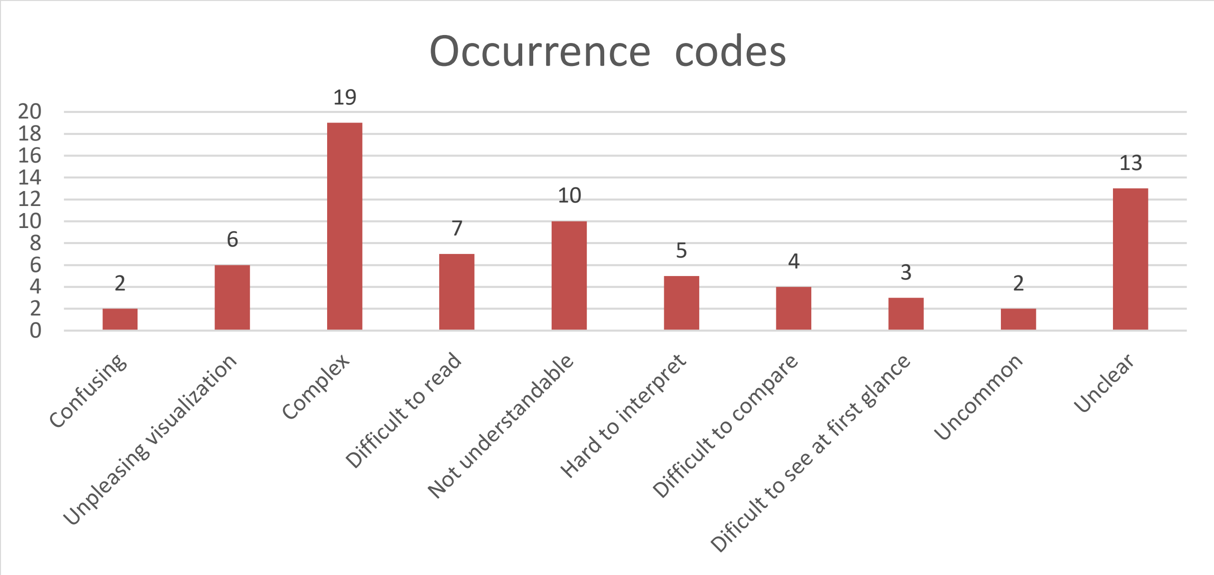

Figure 5.17: Bar chart occurrence codes for Table 5.9

Preference of visualization

Q5) *Which of the four visualizations would you like to see when you are buying wine online? Please explain your answer.*

Figure 5.18 shows that the majority of the participants voted for bar chart. Meaning that the bar chart is the visualization that the participants prefer to see when they are buying wine. In Table 5.10, we can see that 65 participants out of the 119 voted for bar chart and 45 for radar chart. The participants were able to choose more than one option, this is why the total of votes are more than 119.

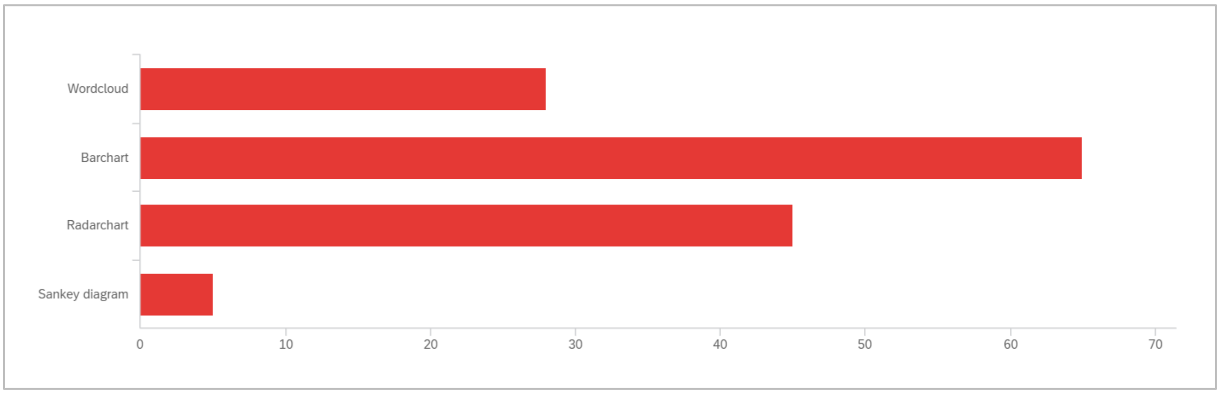

Figure 5.18: Bar chart Q5 results

Visualization	Count
Word cloud	28
Bar chart	65
Radar chart	45
Sankey diagram	5

Table 5.10: Results Q5 custom questionnaire

Table 5.11 is the result of conducting a thematic analyses on the explanation of question 5. The following themes were defined for the bar chart visualization:

Codes	Theme
• Understandable • Straight forward • Easy to interpret	Understandability
• Simple to comparing the characteristics • User friendly	Practicability
• Easy to read • Clear and simple	Simplicity
• Visually pleasing	Aesthetically pleasing
• Familiar	familiarity

Table 5.11: Result thematic analysis Q5 for bar chart visualization

Description of the themes:

- **Understandability**: this covers everything that has to do with why participants found the visualization easy to understand. This will help to understand if understandability of the visualization is the reason that made the majority of the participant choose this visualization. As the visualization that they like to see when they are buying wine.
- **Practicality**: this covers everything that has to do with why participants found the visualization easy to use. This entails finding and comparing data out of the visualization. This will help to understand if practicality of the visualization is the reason that made the majority of the participant choose this visualization. As the visualization that they like to see when they are buying wine.
- **Simplicity**: this covers everything that has to do with participants found the visualization simple. This will help to understand if simplicity of the visualization is the reason that made the majority of the participant choose this visualization. As the visualization that they like to see when they are buying wine.
- **Aesthetically pleasing:** This covers everything that has to do with participants finding the visualization aesthetically pleasing. This will help to understand if “aesthetically pleasing” is the reason that made the majority of the participant choose this visualization. As the visualization that they like to see when they are buying wine.
- **Familiarity:** This covers everything that has to do with participants finding the visualization familiar. This will help to understand if “familiarity” is the reason that made the majority of the participant choose this visualization. As the visualization that they like to see when they are buying wine.

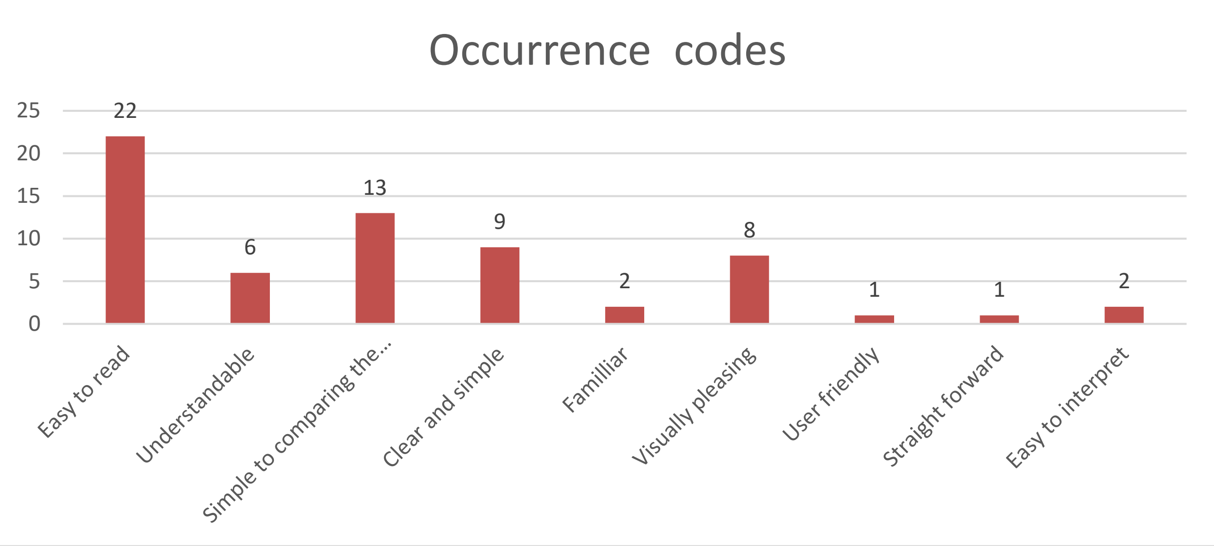

Figure 5.19: Bar chart occurrence codes for Table 5.11

In Figure 5.19, the codes “simple to comparing the characteristics”, “easy to read” and “visually pleasing” where mentioned the most times compare to other codes. These results indicate that it was the simplicity, practicality and pleasing aesthetics that made the majority of the participants prefer the bar chart.

Understandability

***Q6)** Which of the four visualizations did you understand the most? Please explain your answer.*

Figure 5.20 shows that the majority of the participants voted for bar chart, meaning that the participants understand the bar chart visualization the best out of the four visualizations. In Table 5.12, we can see that 77 participants out of the 119 voted for bar chart. The participants were able to choose more than one option, this is why the total of votes are more than 119.

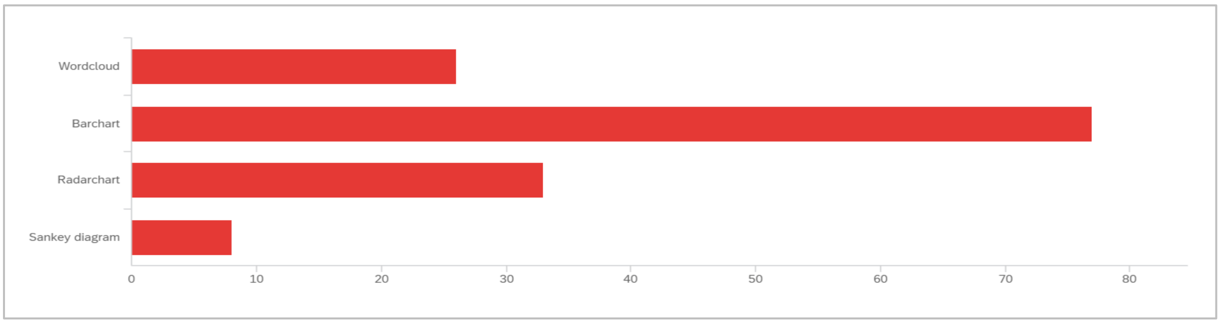

Figure 5.20: Bar chart Q6 results

Visualization	Count
Word cloud	26
Bar chart	77
Radar chart	33
Sankey diagram	8

Table 5.12: Results Q6 custom questionnaire

Table 5.13 is the result of conducting a thematic analyses on the explanation of question 6. The following themes were defined for the bar chart visualization:

Codes	Theme
• Understandable • Straight forward	Understandability
• Simple to comparing the characteristics	Practicability
• Easy to read • Clear and simple	Simplicity
• Familiar	Familiarity

Table 5.13: Result thematic analysis Q6 for bar chart visualization

Description of the themes:

- **Understandability**: this covers everything that has to do with why participants found the visualization easy to understand. This will help to understand if understandability of the visualization is the reason that made the majority of the participant choose this visualization. As the visualization that they like to see when they are buying wine.
- **Practicality**: this covers everything that has to do with why participants found the visualization easy to use. This entails finding and comparing data out of the visualization. This will help to understand if practicality of the visualization is the reason that made the majority of the participant choose this visualization. As the visualization that they understood the most.
- **Simplicity**: this covers everything that has to do with participants found the visualization simple. This will help to understand if simplicity of the visualization is the reason that made the majority of the participant choose this visualization. As the visualization that they understood the most.
- **Familiarity:** This covers everything that has to do with participants finding the visualization familiar. This will help to understand if “familiarity” is the reason that made the majority of the participant choose this visualization. As the visualization that they understood the most.

In Figure 5.19, the codes “understandable”, “clear and simple”, “familiar” were mentioned the most times compared to other codes. These result indicate that it was the simplicity and understandability for the majority of the participants that made them prefer the bar chart.

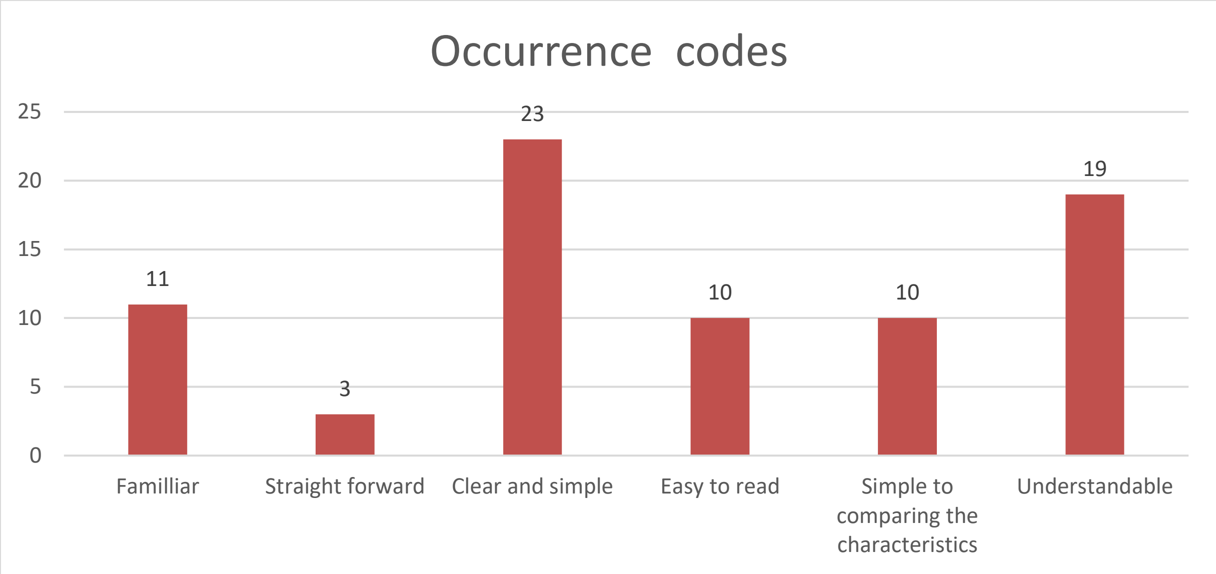

Figure 5.21: Bar chart occurrence codes for Table 5.13

Usefulness of visualizing characteristics

Q7) *Do you find it useful to visualize the wine characteristics from positive, negative and neutral reviews? Please explain your answer.*

Figure 5.20 shows that the majority of the participants find it useful to visualize the wine characteristics from positive, negative and neutral reviews.

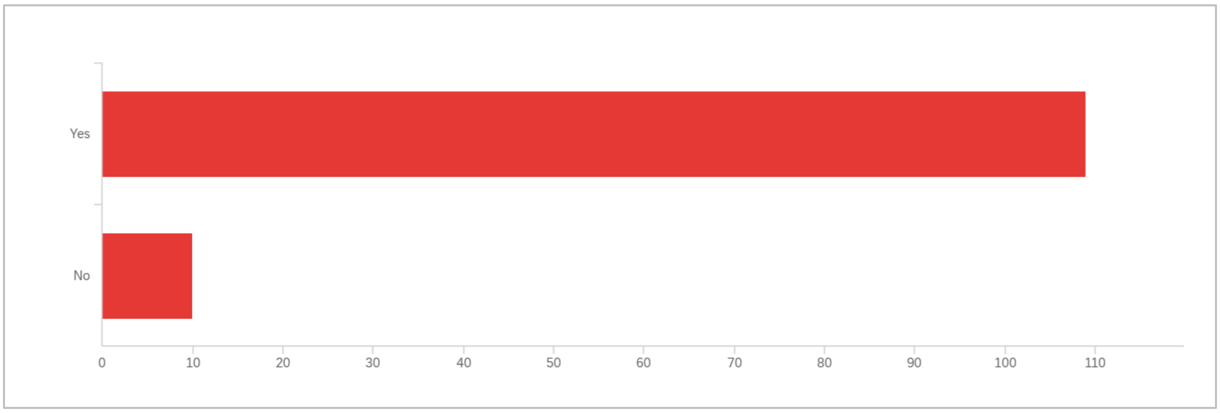

Figure 5.22: Bar chart Q7 results

Chapter 6

Discussion

This chapter gives an in-depth description of the results from the previous one and discusses them. Also, any critical feedback received that is strongly grounded will be covered. The goal of this study was to find out which visualization is the most fitting to visualize the wine characteristics with the help of different visualization techniques. For the first research question, there was the need for a system that included different type of visualizations to visualize wine characteristics that where mentioned in reviews. Therefore four variants using a bar chart, radar chart, word cloud, and Sankey diagram were constructed.

By looking at the results of the SUS-questionnaire the usability of all four variants can be evaluated. These results indicate that the version using the bar chart and radar chart were more usable in comparison to the other variant. Both the radar chart and bar chart scored higher than 68, meaning the system is accepted and the other variants scored under 68 which means that there is room for improvement [39]. A more detailed explanation for these scores might come up by looking at the in-depth custom questionnaire. These results might also give an indication of which parts could be improved.

Questions 1, 3, and 5 will give the result of which visualization can better visualize the characteristics of wine from review text. The results are visualized in Figures 5.10, 5.14, 5.18 and 5.20. The results of the thematic analyses are shown in Tables 5.3, 5.7, 5.11 and 5.13, and visualized in Figures 5.11, 5.15, 5.19 and 5.21.

In Figure 5.11, the codes “simple to comparing the characteristics”, “easy to read” and “clear and simple” were mentioned the most times compare to other codes. These results indicate that it was the simplicity and the practicability for the majority of the participants that made them prefer the bar chart. This means that the bar char is the visualization that helped to find the wine characteristics the fastest.

The same result was given for Question 3, which was “Which of the four visualizations took you the longest to interpret the wine characteristics?”. The reason was also mostly because of the simplicity and practicability, looking at Figure 5.15 and Table 5.7.

For the Question 5, which was: “which of the four visualizations would you like to see when you are buying wine online?“, as shown in Figure 5.19 and Table 5.11, it was mostly because of the simplicity, practicability, and pleasing aesthetics for the majority of the participants.

The results of Questions 2 and 4 gives a more in-depth answer to why the participants weren’t fond of the Sankey diagram. The results are visualized in Figures 5.12 and 5.16. The results of the thematic analyses are shown in Tables 5.5 and 5.9, and visualized in Figures 5.13 and 5.17.

For Question 2 which was: “which of the four visualizations took you the longest to interpret the wine characteristics?”, as shown in Figure 5.12 and Table 5.13, it was mostly because of the lack of understandability and complexity for the majority of the participants that made them choose the Sankey diagram.

The same result was given for Question 4 which was: “Which of the four visualizations did you find the worst to visualize the wine characteristics from positive, negative and neutral reviews?”. The reason was also mostly because of the lack of understandability and complexity, looking at Figure 5.17 and Table 5.13.

For the second research question, Question 6 was asked which was: “which of the four visualizations did you understand the most?”. Looking at Figure 5.21 and Table 5.13, it was mostly because of the simplicity and understandability for the majority of the participants that made them prefer the bar chart.

The results in section 5.2.3 are in line with the results of the SUS-questionnaire. This indicates that the bar chart is the combination of all these results clarifies the reason why bar chart variant is preferred and answers the research questions.

6.1 Feedback and possible improvements

After using the system there was the possibility to provide some feedback in an open text field. From all the replies the most concrete and rightfully grounded ones are described. These state some specific points of improvement for the application that did not come up from the recorded data.

There were a few users who gave some constructive feedback on the system. One user indicated that the text boxes on the chart are not very readable and in general that the section could use some improvement. Comments on the task question showed that it was not clear what was expected of the participants.

There were also few features that were requested, like seeing an image of the wine and a link to buy the wine. This so that they could easily navigate to the website to buy the wine. Then there was one user that indicated to maybe arrange the word cloud from most to least answered. This so that they could easily find the characteristics that occurred the most.

Finally, the interface should include more information containing the explanation of different attributes. This could for example help in understanding the purpose of the visualizations better.

Chapter 7

Conclusion and future work

To close the research gap, a system was constructed that visualized the wine characteristics mentioned in text. This allowed us to analyze which simple visualization is the best to visualize wine characteristics for the end user.

The result for research question one was that using a bar chart visualization is the best way to visualize the positive, negative, and neutral characteristics of wine from review text. This mainly because of the simplicity, understandability, and practicality.

The result for research question two was that using a bar chart visualization is also the best way to make positive, negative, and neutral reviews easily understandable. This is mainly because of the simplicity and understandability.

To conclude it shows that the development of this system was a step in the right direction and definitely promising as the average scores from the questionnaires indicated. There is, however, still some room for improvement as was just mentioned in the previous chapter. Therefore, future work for the application should focus even more to make clear at all times what is happening in the system and to focus on more complex visualizations.

Appendices

A.1 System Usability Scale (SUS) questionnaire

Each of these questions had to be answered on a 5 point Likert scale.

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use

1. this system.

5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

A.2 Custom questionnaire: Information participants

1. What's your gender?
2. What's your age group?
3. How often do you buy wine on average?
4. Which wine type do you prefer? (display this question depending on answer for question 3).
5. Have you ever searched/browsed for wine online before?
6. Do you read the reviews when you are buying wine? (display this question depending on answer for question 5).
7. Do you prefer to read the negative reviews, the positive reviews or both? (display this question depending on answer for question 6).
8. When do you consider a review negative?

A.2 Custom questionnaire: end questions

1. Which of the four visualizations helped you to find the wine characteristics the fastest? Please explain your answer.
2. Which of the four visualizations took you the longest to interpret the wine characteristics? Please explain your answer.
3. Which of the four visualizations did you find the best to visualize the wine characteristics from positive, negative and neutral reviews ? Please explain your answer.
4. Which of the four visualizations did you find the worst to visualize the wine characteristics from positive, negative and neutral reviews? Please explain your answer.
5. Which of the four visualizations would you like to see when you are buying wine online? Please explain your answer.
6. Which of the four visualizations did you understand the most? Please explain your answer.
7. Do you find it useful to visualize the wine characteristics from positive, negative and neutral reviews? Please explain your answer.
8. Do you have further suggestions for improvement on this page and please be critical?

Bibliography

[1] R. K. Bakshi, N. Kaur, R. Kaur, and G. Kaur, "Opinion mining and sentiment analysis," in *Proceedings of the 10th INDIACom; 2016 3rd International Conference on Computing for Sustainable Global Development, INDIACom 2016*, Oct. 2016, vol. 2, no. 1–2, pp. 452–455, doi: 10.1561/1500000011.

[2] D. Oelke *et al.*, "Visual opinion analysis of customer feedback data," *VAST 09 - IEEE Symp. Vis. Anal. Sci. Technol. Proc.*, pp. 187–194, 2009, doi: 10.1109/VAST.2009.5333919.

[3] "10 of the Most Surprising Top-Reviewed Products on Amazon." https://www.entrepreneur.com/slideshow/342345 (accessed Dec. 14, 2020).

[4] D. Boduszek, "Using Graphs to Display Data," *Wisconsin Hosp. Assoc.*, no. Spss 19, 2018, doi: 10.1007/7854.

[5] C. M. Barbu and J. Ziegler, "Designing interactive visualizations of personalized review data for a hotel recommender system," *CEUR Workshop Proc.*, vol. 2222, pp. 7–12, 2018.

[6] J. Wang, J. Zhao, S. Guo, C. North, and N. Ramakrishnan, "ReCloud: Semantics-Based Word Cloud Visualization of User Reviews."

[7] "Web Scraping and Analysis of Wines on Vivino.com | NYC Data Science Academy Blog." https://nycdatascience.com/blog/student-works/web-scraping-analysis-wines-vivino-com/ (accessed Mar. 01, 2020).

[8] "ScholarWorks@Soongsil University: A method for extracting organized information from online product reviews based on text mining." https://scholarworks.bwise.kr/ssu/handle/2018.sw.ssu/5681 (accessed Mar. 10, 2020).

[9] J. Kim and D. Kim, "Analyzing and visualizing comprehensive and personalized online product reviews," *Cluster Comput.*, vol. 22, no. s1, pp. 2115–2128, 2019, doi: 10.1007/s10586-018-2645-6.

[10] D. Shestakov, "Current challenges in web crawling," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, vol. 7977 LNCS, pp. 518–521, doi: 10.1007/978-3-642-39200-9_49.

[11] S. A. Salloum, M. Al-Emran, A. A. Monem, and K. Shaalan, "Using text mining techniques for extracting information from r," in *Studies in Computational Intelligence*, vol. 740, Springer Verlag, 2018, pp. 373–397.

[12] A. Huang, "Similarity Measures for Text Document Clustering."

[13] C. Clifton and R. Cooley, "TopCat: Data mining for topic identification in a text corpus," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1999, vol. 1704, pp. 174–183, doi: 10.1007/978-3-540-48247-5_19.

[14] "K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks | by Imad Dabbura | Towards Data Science." https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a (accessed Nov. 23,

2020).

[15] "K means Clustering - Introduction - GeeksforGeeks." https://www.geeksforgeeks.org/k-means-clustering-introduction/ (accessed Jan. 03, 2021).

[16] S. Qaiser and R. Ali, "Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents," *Int. J. Comput. Appl.*, vol. 181, no. 1, pp. 25–29, 2018, doi: 10.5120/ijca2018917395.

[17] P. Bafna, D. Pramod, and A. Vaidya, "Document clustering: TF-IDF approach," in *International Conference on Electrical, Electronics, and Optimization Techniques, ICEEOT 2016*, Nov. 2016, pp. 61–66, doi: 10.1109/ICEEOT.2016.7754750.

[18] B. Trstenjak, S. Mikac, and D. Donko, "KNN with TF-IDF based framework for text categorization," in *Procedia Engineering*, Jan. 2014, vol. 69, pp. 1356–1364, doi: 10.1016/j.proeng.2014.03.129.

[19] "TF-IDF for NLP | Data Science - Machine Learning - Deep Learning." http://www.ashukumar27.io/tfidf/ (accessed Dec. 14, 2020).

[20] S. M. Mohammad and P. D. Turney, "NRC Emotion Lexicon. NRC Technical Report," pp. 1–234, 2013, [Online]. Available: http://www.mturk.com/mturk/welcome.

[21] R. Bose, R. K. Dey, S. Roy, and D. Sarddar, "Sentiment Analysis on Online Product Reviews," *Adv. Intell. Syst. Comput.*, vol. 933, no. August, pp. 559–569, 2020, doi: 10.1007/978-981-13-7166-0_56.

[22] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, Dec. 2014, doi: 10.1016/j.asej.2014.04.011.

[23] L. C. Yu, J. L. Wu, P. C. Chang, and H. S. Chu, "Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news," *Knowledge-Based Syst.*, vol. 41, pp. 89–97, Mar. 2013, doi: 10.1016/j.knosys.2013.01.001.

[24] M. Hagenau, M. Liebmann, and D. Neumann, "Automated news reading: Stock price prediction based on financial news using context-capturing features," *Decis. Support Syst.*, vol. 55, no. 3, pp. 685–697, Jun. 2013, doi: 10.1016/j.dss.2013.02.006.

[25] "Scopus preview - Scopus - Document details." https://www.scopus.com/record/display.uri?eid=2-s2.0-85019139911&origin=inward&txGid=eb886dba7e67d45ff8954e8058058bad (accessed Dec. 10, 2020).

[26] I. Maks and P. Vossen, "A lexicon model for deep sentiment analysis and opinion mining applications," in *Decision Support Systems*, Nov. 2012, vol. 53, no. 4, pp. 680–688, doi: 10.1016/j.dss.2012.05.025.

[27] J. M. Chenlo, A. Hogenboom, and D. E. Losada, "Sentiment-based ranking of blog posts using rhetorical structure theory," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, vol. 7934 LNCS, pp. 13–24, doi: 10.1007/978-3-642-38824-8_2.

[28] D. Maynard and A. Funk, "Automatic detection of political opinions in tweets," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and*

Lecture Notes in Bioinformatics), 2012, vol. 7117 LNCS, pp. 88–99, doi: 10.1007/978-3-642-25953-1_8.

[29] J. Wang, J. Zhao, S. Guo, C. North, and N. Ramakrishnan, “ReCloud: Semantics-based word cloud visualization of user reviews,” *Proc. - Graph. Interface*, pp. 151–158, 2014.

[30] Y. Zhang and X. Chen, “Explainable Recommendation: A Survey and New Perspectives,” Apr. 2018, Accessed: Mar. 01, 2020. [Online]. Available: http://arxiv.org/abs/1804.11192.

[31] “Predicting Polarity of User Reviews - Towards Data Science.” https://towardsdatascience.com/predicting-polarity-of-user-reviews-8774c2a83dd3 (accessed Mar. 01, 2020).

[32] A. Calabretti *et al.*, “Characterization of volatile fraction of typical Irpinian wines fermented with a new starter yeast,” *World J. Microbiol. Biotechnol.*, vol. 28, no. 4, pp. 1433–1442, Apr. 2012, doi: 10.1007/s11274-011-0943-8.

[33] X. Fang and J. Zhan, “Sentiment analysis using product review data,” *J. Big Data*, vol. 2, no. 1, pp. 1–14, Dec. 2015, doi: 10.1186/s40537-015-0015-2.

[34] V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qual. Res. Psychol.*, vol. 3, no. 2, pp. 77–101, 2006, doi: 10.1191/1478088706qp063oa.

[35] “How to Do Thematic Analysis | A Step-by-Step Guide & Examples.” https://www.scribbr.com/methodology/thematic-analysis/ (accessed Dec. 03, 2020).

[36] M. C. Hao, D. A. Keim, U. Dayal, and J. Schneidewind, “Business Process Impact Visualization and Anomaly Detection,” *Inf. Vis.*, vol. 5, no. 1, pp. 15–27, Mar. 2006, doi: 10.1057/palgrave.ivs.9500115.

[37] B. Jenny and N. V. Kelso, “Color design for the color vision impaired,” *Cartogr. Perspect.*, no. 58, pp. 61–67, 2007, doi: 10.14714/CP58.270.

[38] “Deuteranopia: Red-Green Color Blindness.” https://www.healthline.com/health/deuteranopia (accessed Dec. 13, 2020).

[39] J. Brooke, “SUS-A quick and dirty usability scale.”

[40] “ISO 9241-11:2018(en), Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts.” https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en (accessed Dec. 15, 2020).

[41] “Usability Evaluation In Industry - 1st Edition - Patrick W. Jordan -.” https://www.routledge.com/Usability-Evaluation-In-Industry/Jordan-Thomas-McClelland-Weerdmeester/p/book/9780748404605 (accessed Dec. 15, 2020).

[42] Ö. Gülen Ertosun, O. Erdil, N. Deniz, and L. Alpkan, “Positive Psychological Capital Development: A Field Study by the Solomon Four Group Design,” *Int. Bus. Res.*, vol. 8, no. 10, Sep. 2015, doi: 10.5539/ibr.v8n10p102.

[43] A. Vanspauwen, “Flexible interactions with recommendation systems,” 2019.

[44] “Counterbalanced Measures Design - Counterbalancing Test Groups.” https://explorable.com/counterbalanced-measures-design (accessed Dec. 15, 2020).

[45] “Plotly Python Graphing Library | Python | Plotly.” https://plotly.com/python/ (accessed Dec. 17, 2020).

[46] K. Duong, “Research Guides: Qualtrics: What is Qualtrics?,” Accessed: Dec. 17, 2020. [Online]. Available: https://csulb.libguides.com/qualtrics/about.

[47] M. Z. Al-Taie and S. Kadry, “Visualization of Explanations in Recommender Systems,” *J. Adv. Manag. Sci.*, vol. 2, no. 1, pp. 140–144, 2014, doi: 10.12720/joams.2.2.140-144.

AFDELING
Straat nr bus 0000
3000 LEUVEN, BELGIË
tel. + 32 16 00 00 00
fax + 32 16 00 00 00
www.kuleuven.be