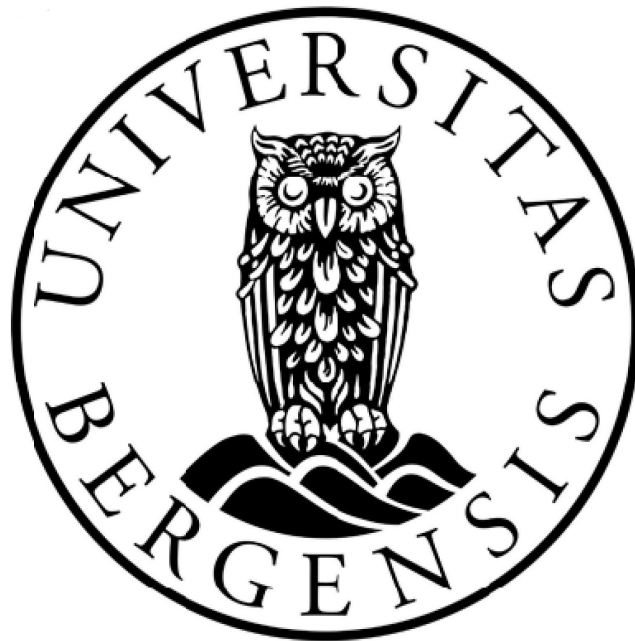# Web Based Brand Equity

Web mining as an alternative approach to Customer Based Brand Equity measurement

Emil Knudsen Pamer

Master thesis
Department of Information Science and Media Studies

UNIVERSITETET I BERGEN

23.05.2017

2017

Web Based Brand Equity

Emil Knudsen Pamer

https://bora.uib.no/

Abstract

Brand equity denotes the value that is transferred from a brand to its products, and which translates into higher sales volumes and higher profit margins. Consumer based brand equity (CBBE), a perspective of brand equity as a set of associations held in consumers' minds, was conceptualized in the early 1990's. CBBE measurement has always been conducted by the use of surveys, administered as personal interviews, telephone calls and the like. Surveys are expensive and time-consuming, which means that alternative approaches for CBBE measurement should be investigated.

The internet offers multiple types of information which may relate to CBBE in various ways. This research investigated whether publicly available online resources can be collected and analyzed with regard to CBBE. In order to do so, the author re-conceptualized CBBE as applicable for measurement using online resources. The online conceptualization was then operationalized in terms of relevant information available online, resulting in an online CBBE scale consisting of 11 metrics. The metrics were gathered from four domains: Google search results, Google search statistics, social media, and online reviews.

Web mining is the process of automatic web information retrieval. In order to automate web retrieval procedures within the four domains, the author designed a computer application, intended to work across product categories. The web mining approach was performed on the product category of headphones, featuring 62 headphone models from 18 different brands. Product information was supplied by leading Norwegian electronics retail chain Spaceworld Soundgarden, and included all headphone models sold in-store or online in 2016. The study revealed that several online measures of CBBE are correlated with sales. One of the key findings of the multiple regression analysis was a general shift of explanatory power, when comparing expensive models with budget models. In the high-end segment, product-specific measures, such as professional review scores, best explained differences in sales, while several brand-specific measures were negatively correlated. In the low-end segment, the result was the opposite.

The web mining approach is shown to be able to provide insights on CBBE, but generalizability across product categories is limited, due to reliance on social media and reviews. Recommendations for future research are proposed, such as utilizing nationwide sales figures, and analyzing products that are in direct competition. Expanding the set of algorithms, such as including sentiment analysis, may also increase the performance of the web mining approach to CBBE measurement. The research concludes with the argument that the web based CBBE may, with refinement, become a valid alternative to survey research.

Preface

This thesis was written in the spring of 2017, for the Department of Information Science and Media Studies at the University of Bergen, in collaboration with consumer electronics retail chain Spaceworld Soundgarden.

Table of contents

List of figures and tables

1: Introduction

1.1: Definitions

Brand equity: "A brand's power derived from the goodwill and name recognition that it has earned over time, which translates into higher sales volume and higher profit margins against competing brands" (Businessdictionary.com).

Customer based brand equity: "The differential effect that brand knowledge has on consumer response to the marketing of that brand" (K.L. Keller, 1993).

Web mining: "The process of using data mining techniques and algorithms to extract information directly from the Web by extracting it from Web documents and services" (techopedia.com).

1.2 Customer Based Brand equity measurement

Customer based brand equity (CBBE) was defined in the beginning of the 1990's, when it was conceptualized by Keller (1993). CBBE is a perspective that views brand equity as a set of associations towards a brand, held in the minds of consumers. Because of its intangible nature, brand equity has always been hard to measure directly. However, the work performed by Keller and Aaker (1991, 1996) on building conceptual models attempting to describe brand equity, provided a guiding light for CBBE measurement studies in the following years. The conceptual models have been validated in dozens of studies during the last two decades, and have been shown to provide accurate insights about this intangible phenomenon. However, the research on the topic of CBBE measurement has always employed the survey approach; people studies in which surveys are administered as personal interviews, telephone calls, and later, email and online surveys. This type of research is time-consuming and expensive, and suffers from challenges inherent in many studies involving people, such as confirmation bias, the use of student samples and so on.

As research on CBBE sprung into existence, so did the internet. After its conception in the mid 1990's, the world wide web now contains a wealth of information that may relate to brand equity. This includes companies' presence on social media platforms such as Facebook and Twitter, online professional and amateur reviews, professional 'best of' web sites and so on. However, the author could find no previous attempt to harness this information to gain insights on brand equity.

1.3 Brand Equity measurement using internet resources

Although no research exists that examines the explanatory power of social media, reviews and so on, with regard to brand equity, many studies have examined the explanatory power of these domains individually, with regard to sales figures. Several studies have reported that both professional and amateur reviews are significant factors in explaining sales in domains such as cinema movies, wine, computer games and more (Zhu & Zhang, 2010). Social media elements such as 'likes' on Facebook have been shown to correlate with sales in other domains (Barnes, 2014). Since sales figures represents one of two ways of measuring CBBE (the other being price premium), this suggests that measuring CBBE using online resources may be a worthwhile approach.

Web mining is a set of techniques that can be used to programmatically retrieve information from the web. In recent years, much effort has been made in order to make web retrieval more available to developers. This has usually taken the form of API's (application programming interface); online services that provide public access to company data or functionality. By building computer programs that consume API's, developers can retrieve large quantities of structured, accurate data automatically, and much faster than any manual methods.

In this thesis, the author designed such a computer program. The main goal of the thesis is to investigate whether traditional conceptualizations of CBBE can be adapted to the internet domain, and measured using modern tools for automatic web retrieval. If an approach based on web mining should prove to perform as well as traditional survey methods, then web mining as an approach to CBBE measurement would constitute a faster, cheaper, and more efficient way of doing research on this topic.

2: Brand equity

Brand equity is the conceptualization of the value of a brand which is transferred to the brand's associated products, resulting in increased sales volumes and enabling businesses to command higher margins for their products and services. Research on brand equity has been conducted in two fields: Informational economics and cognitive psychology, the latter being the focus of this thesis.

2.1: The financial perspective

Because of brand equity's obvious ties with financial performance, much research has been conducted in the field of informational economics. This research has mainly been focused on the company level, treating brand equity as a quantifiable asset - the value of the brand itself. This is useful in many cases, such as acquisitions, mergers and so on. In one case, an estimated 90% of the total sum of an acquisition deal was attributed to brand assets (Lassar, Mittal and Sharma, 1995). However, it was clear that company level financial brand equity would always be an indirect measurement of something else: Increased revenue. And if increased revenue could be attributed to the value of the company's brand(s), then those gains had to come from either increased sales volumes, increased margins, or both. Thus, research in financial economics started focusing on the product level, with the basic idea being "How much of the sales volume can be attributed to brand equity?". This perspective manifested itself in many research papers concerning the relationship between sales of branded products, and generic (unbranded), or fictitiously branded products. This type of research yielded, for the first time, a scientific measure of the value that is transferred from brands to related products.

2.2: Customer based brand equity

Research on the product level revealed that some quantifiable value is indeed transferred from brands to products. The next logical step was to ask the questions of "why" and "how". This shifted the focus of research from statistics and mathematical analysis, to cognitive psychology.

It was obvious that brand equity depended on the minds of consumers, and their experience with a brand. Cognitive psychology studies on brand equity continued the work on unraveling consumers' response to a brand, and started developing models on the individual level, intended to explain and predict consumer behavior. This perspective on brand equity has been the main source of research on the subject since the 1990's, and is now called Customer Based Brand Equity, or CBBE.

2.3: Conceptualizations and measurement

Although the general idea of brand equity has been studied for decades, the first dedicated conceptual models were developed in the 1990's, by authors Kevin L. Keller (1993) and David A. Aaker (1991). Common to both approaches is the perspective of CBBE, as well as the assumption that measurement will involve people studies, in the form of surveys such as interviews.

2.3.1: Keller's Model

Keller formalized the first CBBE model, proposing that brand equity consists of two central components: Brand awareness and brand image. Brand awareness represents the consumers' previous exposure to, or experience with, the brand, and is a prerequisite for brand image. Brand image is defined by Keller as "perceptions about a brand as reflected by the brand associations held in consumers' memory".

FIGURE 1
Dimensions of Brand Knowledge
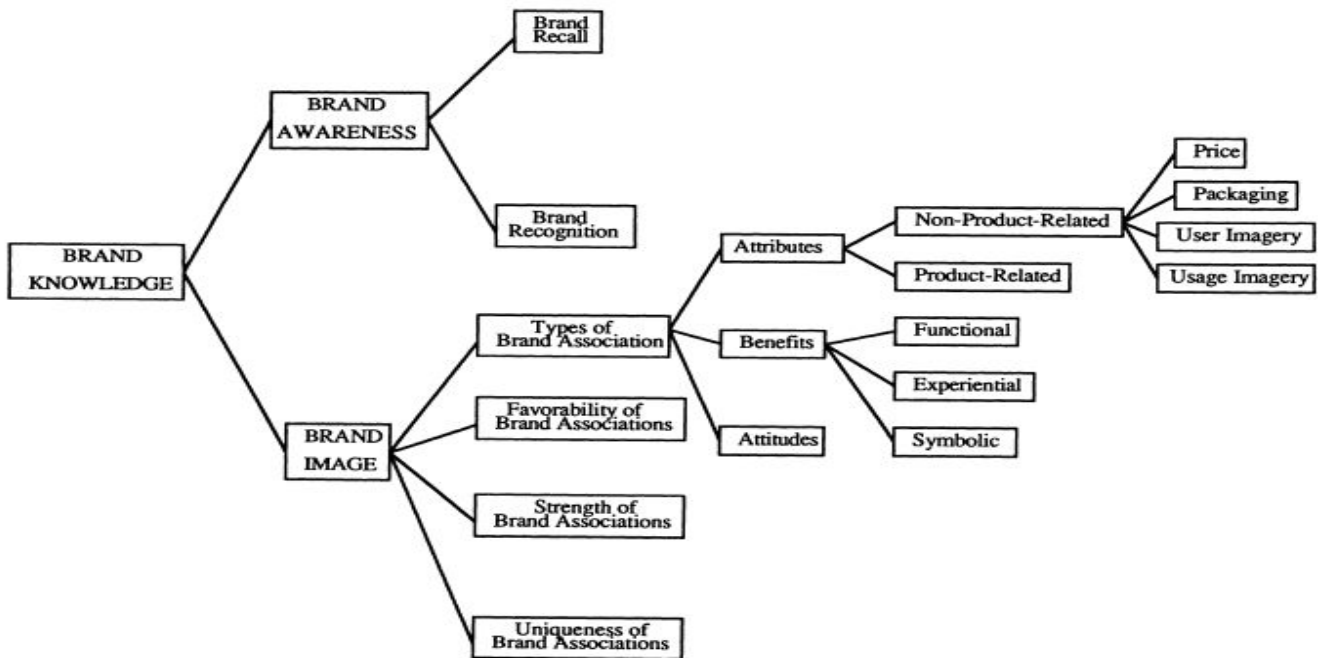
Figure 1: Keller's CBBE conceptualization

Brand awareness

Brand awareness consists of two basic metrics: Brand recognition and brand recall. Brand recognition studies involve asking research participants whether they can remember seeing or knowing about the brand, given the brand as a que. The que may be the brand's name, logo or some other direct reminder. Brand recall is measured by asking research participants whether they can identify the brand, given some other indirect que, such as the product category.

Brand image

Brand image is the set of brand associations held in a consumer's mind. This definition was influenced by associative memory models presented in studies from cognitive psychology. Associations are categorized by type, favorability, strength and uniqueness.

The types of brand associations are differentiated by scope, and are classified as attributes, benefits, and attitudes. Attributes are broken down into product-related and non-product-related attributes. Product-related attributes may be the size, shape, expected performance and so on. Non-product-related attributes are price, packaging, user imagery (i.e what type of person uses

this type of product), and usage imagery (i.e where, and in what type of situation the product is used).

Benefits are "the personal value consumers attach to a product or service attributes". Benefits have a widened scope compared to attributes, and may be functional, experiential, and symbolic.

Functional benefits are the benefits intrinsic in the use of the product or service, and usually involves some practical or psychological need. For example, purchasing a TV because of its ability to show programmes. *Experiential benefits* are also often linked to the product's attributes, and are the sort of feelings or experiences one can expect from the consumption of the product. A TV's ability to immerse the watcher in the story of dramatic movies, is an experiential benefit. *Symbolic benefits*, however, are not linked to the actual performance of the product. A certain product may resonate with one's personality, and the benefit of owning or consuming the product is based on what it says about you. As such, symbolic benefits may often be based on needs for social approval, personal expression and so on. Symbolic benefits may be especially relevant for visibly branded products, such as clothes with a brand's logo printed on them.

The *favorability* of an association is not necessarily linked to the specific product or brand, and may reflect attitudes about the product category, the situations in which the product is used and so on. For example, one may have positive associations towards cabriolets, because they are associated with nice weather. The *uniqueness* of an association is the degree to which the association is not shared with competing products or brands.

Keller's conceptualization of CBBE was the first to be formulated as a hierarchical structure. Since its proposal in 1993, it has been validated in dosens of research papers on CBBE, and remains one of two central foundations for the research that followed. The other, proposed by David A. Aaker two years earlier, does not focus on the structure of associations. Rather, 'The Brand Equity Ten' was simply intended to be a minimal set of measures, while still including everything thought to contribute to brand equity. Aaker's goal for future research was a continuous reduction of the set, resulting in the definition of a single, quantifiable measure of brand equity.

2.3.2: David Aaker - The Brand Equity Ten

Aaker (1991) presented brand equity as consisting of four
dimensions: loyalty, perceived quality, associations, and
awareness. He developed a model consisting of ten measures,
which he termed The Brand Equity Ten. The ten measures were
grouped into five categories, with the first four categories
representing the four dimensions, and the fifth category
representing market behavior measures not obtained directly from
customers.

Loyalty Measures

Aaker placed special emphasis on the loyalty measures, as a loyal
customer base "represents a barrier to entry, a basis for a price
premium, time to respond to competitor innovations, and a
bulwark against deleterious price competition". He further stated
that, because of loyalty's central position in brand equity as a
whole, other measures such as perceived quality could be
measured by their ability to influence loyalty. Aaker's loyalty
category includes the measures price premium and
satisfaction/loyalty.

TABLE I. The Brand Equity Ten

Loyalty Measures

- Price Premium
- Satisfaction/Loyalty

Perceived Quality/ Leadership Measures

- Perceived Quality
- Leadership

Associations/ Differentiation Measures

- Perceived Value
- Brand Personality
- Organizational Associations

Awareness Measures

- Brand Awareness

Market Behavior Measures

- Market Share
- Price and Distribution Indices

Figure 2:The brand equity ten (Aaker, 1996)

Price premium, as defined by Aaker, is "the amount a customer
will pay for the brand in comparison with another brand (or set of comparison brands) offering
similar benefits". Aaker further highlighted price premium as perhaps the single best indicator of
brand equity, as all drivers of brand equity should have an impact on this measure.

Satisfaction/Loyalty of existing customers can be easily measured directly via interviews, or by
monitoring recurring purchases of a service or product. Satisfaction and loyalty may be
influenced by several factors within the customers' minds, such as inflated expectations, or
lacking performance of the product or service. Customer satisfaction can also be influenced by
competitors in categories where the benefits may change over time, such as introduction of new

features, objectively improved performance and so on. As such, a computer owner may report high customer satisfaction shortly after the purchase, but the satisfaction may steadily decline over time, when the customer notices newer and better products being introduced by competing brands. This is generally not a problem in product categories where benefits stay the same, such as with commodities like soap. These product categories generally enjoy high levels of customer loyalty, with many individuals becoming lifetime consumers of the product.

Perceived Quality/Leadership Measures

Perceived Quality has been shown to be associated with other components of brand equity, such as price premiums, brand usage and others. Aaker submits that perceived quality is applicable across product categories, making it one of the most revealing and safe measures of brand equity. Examples of perceived quality associations include statements like "Volvo makes the safest cars", and "LG TVs have the most natural image".

Leadership, or popularity, is defined in terms of current market share. The general observation is that people tend to buy products which are popular, or considered the leading product in its category. This may be because people consider a very popular product to be safe, in terms of quality. Three relevant elements of leadership are identified by Aaker, in order to measure this phenomenon: Is the brand the leading brand, or one of the leading brands in the category? Is the brand growing in quality? Innovative, first with advances in product or service?

Despite its proposed influence on brand equity, leadership measures have not received the same attention in research as other measures, such as perceived quality, loyalty and so on.

Associations/Differentiation measures

Aaker distinguishes between three image dimensions of brand associations: The brand as product (value), the brand as person (brand personality) and brand as organization (organizational associations).

The *Value* proposition usually involves a functional benefit, and is designed to work across product classes. It suggests that brand equity can be measured by a) whether the brand represents good value for money and b) whether there are reasons to buy this brand over competitors.

Brand Personality can provide a link between customer/brand relationships and differentiation, as well as forming the basis for the brand's emotional and self-expressive benefits. Aaker proposes that this dimension is especially important for brands which are visible in social settings, enabling the brand to make "a visible statement about the consumer". As such, Aaker's definition of brand personality is analogous to what Keller describes as a brand's symbolic benefit. Ways of investigating the existence and nature of brand personality include testing statements like a) this brand has personality, b) this brand is interesting, c) is is easy to envision a specific type of user of this brand.

The last statement is connected to user imagery in Keller's model, and Aaker(1996) claims that user imagery can be a driver of brand personality, i.e. the personality of a typical user of the brand becomes part of the brand personality. Aaker warns that brand personality may be very stable, and its measurement can be poorly suited as a measurement of brand equity in dynamic markets.

Organizational Associations is the brand-as-organization dimension, and it considers the people, values and programs that lie behind the brand. This dimension can play an important role in demonstrating that the brand is about more than just the products or services that it provides. Apple's founder Steve Jobs probably played an important factor in establishing Apple as an innovative and exciting company in people's minds.

Awareness measures

Brand Awareness is the measure of whether people are aware of the brand, as well as their level of awareness. As in Keller's model, Aaker distinguishes between brand recognition and brand recall. In addition, Aaker appends four descriptors of brand awareness, in the form of *top-of-mind* (analogous to Keller's description of some brands as *prototypical* in a certain product category), *brand dominance* (the only brand recalled), *brand knowledge* (knowing what the brand stands for), and *brand opinion* (a user's opinion about the brand). For less known brands, recognition may be the most important aspect. For well known brands, recall and top-of-mind are more relevant marketing goals. Aaker specifies that measurement using only the brand name is insufficient. Recall and recognition of logos, slogans and visual imagery are necessary elements when attempting to gather the full extent of people's awareness of a brand.

Market behavior - brand performance measures

The performance of a brand as measured by *Market Share* can be a valid measure of a brand's standing with its customers. A brand which is considered the biggest brand in the product category, should see its sales increase, rather than decrease. Similarly, if a company's brand equity improves, increased market share should follow (Aaker, 1996). In contrast with the other four categories in the Brand Equity Ten, market share measures are available and accurate. However, market share can be increased by promotions, offers and the introduction of lower-priced products into the market. These are activities that can decrease the perceived quality of a brand, and is also by definition the opposite of price premium. Thus, increased market share in the short term may decrease brand equity in the long term.

Price and Distribution Indices

Since market share increases as a result of promotions and offers, it is important to compensate by adjusting the price points according to the relative market price at which the brand is being sold. The distribution coverage also impacts market share, and differences in distribution coverage may obscure the reality which data analysis attempts to convey. For example, a brand whose products sell very well, but is only available in 80% of the total market, may seem like it performs worse than a company whose products are sold in the entire market. Aaker suggests two measures of distribution coverage which may serve to reduce this issue: a) the percentage of stores carrying the brand and b) the percentage of people who have access to it.

Aaker admits that the measurement of brand equity using so many measures may become unwieldy, and that the end goal of brand equity research should culminate in an agreed single value of brand equity. This value would be built by combining the measures with the highest diagnostic value. Then, marketers would be able to create specific surveys and monitor this single value as it changes with short term marketing efforts, as well as long term growth. The question, then, is which measures should form the basis of such a model, and whether a truly general model in fact can be built and used effectively across product classes.

2.4: Brand equity measurement using survey methods

Brand equity was an established concept long before Keller and Aaker proposed their CBBE models. The measurement of brand equity (which was always the end goal of conceptualizing it in the first place) has an equally long history of research. However, until the mid 1990's, research on brand equity measurement was still dominated by the company-and product based perspective of brand equity as a financial asset. This was made clear by Lassar, Mittal and Sharma (1995): "In spite of the increasing importance of the brand equity concept, an instrument to measure brand equity from a customer perspective has been lacking. Because the source of brand equity is customer perceptions… it is important for managers to be able to measure and track it at the customer level". Following the work of Keller and Aaker on conceptualizing CBBE, research on brand equity measurement gradually moved towards CBBE as the dominant perspective in brand equity research.

CBBE measurement research in the early and mid 1990's immediately reported interesting findings, based on this new perspective. For example, Dacin and Smith (1994) found that consumers accept brand extensions more readily when the brands' existing product lines do not vary greatly in quality. The findings suggested that the release of a single sub-standard product could negatively impact the performance of other, often unrelated products from the same brand.

Lassar, Mittal and Sharma (1995) attempted to operationalize the conceptual models, in order to diminish the gap between concept and measurement. They found it necessary, among other things, to limit the 'brand image' dimension to social settings, defining it as "the consumer's perception of the esteem in which the consumer's social group holds the brand". Similar adjustments were made to all the relevant parts of CBBE, culminating in a collection of five central measurement dimensions: Performance, social image, price/value (based on consumers' perceived balance of the price of a product and all its utilities), trustworthiness and identification/attachment (analogous with brand loyalty). They refined their scale through three pilot tests, and evaluated their final measurement model by administering a questionnaire to 113 consumers, in which they tested two product categories. They found that their newly developed CBBE measurement model correlated significantly with an overall measure of brand equity,

reporting that "we found that prices reflected the equity associated with the brand". The study also found that consumers demonstrate a halo effect across dimensions, stating that "if consumers evaluate a brand to perform well, consumers also expect the brand to have high levels of value, or be more trustworthy". Figure 3 depicts the final questionnaire.

Performance
P1 From this brand of television, I can expect superior performance
P2 During use, this brand of television is highly unlikely to be defective
P3 This brand of television is made so as to work trouble free
P4 This brand will work very well

Social image
I1 This brand of television fits my personality
I2 I would be proud to own a television of this brand
I3 This brand of television will be well regarded by my friends
I4 In its status and style, this brand matches my personality

Value
V1 This brand is well priced
V2 Considering what I would pay for this brand of television, I will get much more than my money's worth
V3 I consider this brand of television to be a bargain because of the benefits I receive

Trustworthiness
T1 I consider the company and people who stand behind these televisions to be very trustworthy
T2 In regard to consumer interests, this company seems to be very caring
T3 I believe that this company does not take advantage of consumers

Attachment
A1 After watching this brand of television, I am very likely to grow fond of it
A2 For this brand of television, I have positive personal feelings
A3 With time, I will develop a warm feeling toward this brand of television

Figure 3: The CBBE questionnaire (Lassar, Mittal and Sharma, 1995)

CBBE measurement methodology

Classical CBBE measurement has always been performed by administering a survey. Surveys in the literature have been administered in the form of telephone calls, personal interviews, mail questionnaires and, later, email and online questionnaires (Aaker, 1996). Research participants have been, to a large degree, students (Poppu, Quester & Cooksey, 2005), but research targeting potential customers have also been performed.

CBBE measurement by survey administration is challenging for a number of reasons:
- Surveys are expensive, especially if one is targeting several markets

- Large scale surveys are time consuming
- Responses may be difficult to interpret
- It is hard to ensure external validity, for multiple reasons:
    - Many studies involve student participants instead of potential customers.
    - Greatly varying markets in terms of culture, average consumers' purchasing power etc.
    - Surveys in the literature revolve around brands in a limited set of product categories, making it difficult to draw generalized conclusions based on the results.

Thus, attempts to investigate alternative methods for conducting CBBE research should be welcomed.

2.5: CBBE research using internet resources

The internet contains much information about brands, products and consumers that could be viewed as representing aspects of CBBE. Many brands now have social media profiles, which may be analyzed in order to to gain information about their popularity, image, and so on. There are also a wealth of web sites that write and publish professional reviews. Furthermore, many e-commerce web sites let consumers write and publish reviews of their own. Many web sites also offer 'best of' articles; professional summaries of the best products in a given product category. It is possible that much of this information can be collected and analyzed in order to gain insights in terms of brand equity.

Advances in the field of computer science has led to easier access to web information retrieval techniques, in the form of APIs (application programming interface) and other related tools. These tools enable computer scientists to collect and analyze web based content, and may provide a basis for a new approach in CBBE measurement. If an approach based on web retrieval can be shown to yield results that are as accurate as survey methods, then it should be considered a superior approach. This is because web retrieval can be automated, reducing time spent on data collection from weeks or months, to minutes. A web mining approach also avoids many of the challenges inherent in people studies, such as confirmation bias, student samples, ethical considerations, and so on.

Research questions

The goal of this research is to investigate whether web mining constitutes a worthwhile alternative approach with regard to CBBE measurement. In order to do this, four research questions were devised:

Research Question 1: Can classical consumer based brand equity be re-conceptualized into dimensions that are appropriate for web information retrieval?

Research Question 2: Can these dimensions be operationalized in terms of freely available online data?

Research Question 3: Is it possible to create a web mining application which accurately measures CBBE across product classes?

Research Question 4: Do these measurements of web based brand equity correlate with sales figures?

Project description

In order to answer the research questions, several activities has to be performed. First, a re-conceptualization must take place, in which CBBE is redefined in terms of the internet domain. The re-conceptualization should stay as true to the models proposed by Keller(1993) and Aaker(1996) as possible, as these models have been thoroughly examined and validated over the last 20 years. The online conceptualization should contain dimensions that both reflect the established CBBE models, as well as the online information that is available for web retrieval. Then, the dimensions must be operationalized into metrics that are possible to measure directly, using web retrieval. These metrics will constitute a scale for online CBBE.

Data collection will be performed by a computer application, designed by the author. The application should work across product categories. In order to do so, the application must be agnostic about the product category, and hold no previous knowledge about the brands or products being tested. The application will perform all web retrieval tasks automatically, based only on information about the product category, brands and products which must be supplied by the user. When all web retrieval algorithms are completed, the application should output the results on a format that is suitable for analysis. The CBBE metrics will then be analyzed with regard to sales figures, in order to assess the explanatory power of the approach.

3: Related work

The means for performing CBBE measurement studies using online resources have been readily available for some time. For example, e-commerce businesses can, and do, track the ip-address of customers, in order to categorize them based on whether they are new or recurring. A high percentage of recurring customers could be interpreted as high levels of customer loyalty. Additionally, most brands today have an online presence on social media sites like Facebook and Twitter. As this presence does not lead to sales directly, it could be viewed as an attempt, conscious or unconscious, to improve their brand's equity along dimensions like brand awareness and brand personality. Because this data is publicly available, it is possible to collect and analyze this data using web mining. However, web mining has not yet been utilized in any full-fledged CBBE measurement study.

3.1: Brand equity measurement using online resources

The author could find only one CBBE study in the online domain, conducted by Christodoulides et.al. (2006). The study revolved around the conceptualization and measurement of brand equity for businesses in online retail service (ORS). The researchers re-conceptualized brand equity as defined by Keller and Aaker, in order to make the models applicable for measurement utilizing internet resources. Their re-conceptualization was informed by 'experience interviews', which are in-depth interviews with domain experts (marketing in this case). It resulted in a scale of five dimensions: Emotional connection, online experience, responsive service nature, trust, and fulfillment.

While the dimensions of emotional connection and trust are applicable to brands that are not primarily internet-based, the other dimensions are not. Furthermore, since the scale focused solely on ORS brands, it is of limited usefulness for brand equity measurement for brands that are not primarily internet-based.

3.2: Correlation studies on online metrics and sales

Despite the lack of pure CBBE studies utilizing internet resources, much research has been conducted measuring would-be elements of brand equity, utilizing web mining and other types of information-gathering in the internet domain. Among these are professional and amateur reviews, as well as social media information.

3.2.1: Professional reviews

Several studies have measured the explanatory power of online reviews on sales. Brand equity is sometimes defined as the sum of past and present marketing efforts (Aaker, 1996). Brands often deliver new products to domain experts in order to get them reviewed. This strategy should be considered part of the marketing effort for the product, and a long history of glowing product reviews should communicate to the consumer that the brand consistently delivers high-quality products. Conversely, a history of mixed or negative reviews may decrease consumer confidence in the brand, drawing on the findings of Dacin and Smith (1994) that inconsistent quality of past products negatively impact brand equity.

Litman (1983) studied the impact of critics' ratings on cinema movies' box office revenue, over a period of six years (1972-1976). He reported that critics' ratings are significant factors in explaining box office revenue. Mahajan, Muller, and Kerin (1984) conducted a study using diffusion models, and found that word of mouth (WOM) was a significant predictor of movies' attendance figures.

Eliashberg and Shugan (1997) found that professional reviews are indeed correlated with sales, but stressed that their findings suggested that reviews do not impact sales directly, and are useful only for their predictive or explanatory power. Some later studies have supported this finding, while others, such as Friberg and Grönquist (2012), found clear indications that favorable and neutral reviews on wine significantly influenced demand in the following weeks. Boatwright, Kamakura, and Basuroy (2007) also reported a positive correlation between expert reviews and sales. Furthermore, they found that some critics are especially influential, suggesting that further research should consider adding weights to the individual critics.

3.2.2: Consumer reviews / online word of mouth

Many ORS websites offer consumers the option to write a review of the products they have purchased. These amateur reviews may have an impact similar to that of expert reviews, on consumer associations towards the brand. The literature supports this idea; Zhu & Zhang (2010) argues that consumer reviews work as a proxy for overall online word of mouth, while Bambauer-Sachse and Mangold (2010) found that brand equity is diluted through negative online word of mouth.

The reported findings on the correlation between consumer reviews and word of mouth on sales have been mixed. Chen, Wu, and Yoon (2004) found that consumer reviews on books are not correlated with sales on the online retail service Amazon.com. Liu (2006) found that word of mouth information offers significant explanatory power for box office revenue in the domain of cinema movies, while Duan, Gu, and Whinston (2008) reported that consumer reviews have no effect on box office revenue.

3.2.3: Social media

Social media platforms offer many benefits for businesses when compared to traditional media;
- Publicity of campaigns are often cheaper and can be updated and adjusted frequently,
- Company news and updates may spread virally at a speed unmatched by traditional media,
- Companies can gain insight into consumer discussions and opinions in situations where users are at leisure (increasing the confidence in them being honest about their opinions),
- They provide brand personality building opportunities
- They offer built-in customer relationship management (CRM).

Several studies have been conducted on businesses' adoption of social media and related tools for analysis, as brands attempt to harness social media in order to increase revenue and build stronger consumer connections.

Barnes (2014) studied the 500 biggest companies in the world (Forbes Magazine's 'Fortune 500') and their activity of social media platforms. Analysis showed that 77% of the companies

had active corporate Twitter accounts. In the paper, Barnes discussed what has been coined *social commerce*, defined by Yahoo! as "a set of online collaborative shopping tools such as shared pick lists, user ratings and other user-generated content sharing of online product information and advice". She observed that large companies have started to actively leverage online engagement with customers on social media platforms, in order to boost sales, and build brand awareness and brand personality using these platforms. One of the central conclusions is that "one can certainly assume that online discussions (eWOM) can, and do, impact sales, reputations and brands".

Chen, De and Hu (2011) analyzed the size of music artists' network of 'friends' with regard to sales. They reported that "for the artists with many friends, broadcasting activities on MySpace have a significant impact on music sales". Li and Wu (2012) investigated the connection between Facebook 'likes' and twitter 'tweets' on posts about voucher deals from the online retail service GroupOn. They found that a single Facebook 'like' corresponded to an additional 4.5 voucher sales. With regard to Twitter, they reported that "we do not find consistent evidence of Twitter-mediated WOM having an effect on sales". The study also found that average ratings of the voucher deals on Yelp/Citysearch act as a complement to Facebook 'likes' when ratings are moderate.

These studies show that research not intended to cover brand equity, still conducted measurements, and reported findings, that are interesting to CBBE. Furthermore, they show that freely available online information has explanatory power when it comes to sales volumes. Thus, the related literature on reviews and social media indicate that these resources are natural inclusions in online CBBE measurement.

4: Methodology and research design

4.1: Adapting CBBE for measurement by web mining

Research Question 1: Can classical consumer based brand equity be re-conceptualized into dimensions that are appropriate for web information retrieval?

Research Question 2: Can these dimensions be operationalized in terms of freely available online data?

4.1.1: Re-conceptualization

The goal of this study is to investigate the merits of web mining as a tool for measuring consumer based brand equity. As existing general models of CBBE are conceptualized as applicable for measurement using surveys, CBBE was re-conceptualized as applicable for measurement using online resources. This part of the research was exploratory, and so it had to be informed by existing literature in order to stay true to the original, verified concepts of CBBE. To ensure the validity of the conceptualization, the process followed three guiding requirements:

1. The conceptualization should, as far as possible, be a direct translation of the existing and verified models of CBBE as proposed by Keller (1993) and Aaker (1996), introducing no extra elements, and keeping as many of the existing elements as possible.

2. The conceptualization should only contain elements which are possible to measure via web mining.

3. The data gathering by web mining should be easily replicated by manual search, in order to verify the results.

4.1.2: Operationalization - the online CBBE scale

The online CBBE scale took form according to the following activities:

1. A list of brand-and product related metrics which are possible to measure using web mining was created.

2. The list of metrics was reduced by consulting the existing CBBE models, and removing the metrics which did not have a justifiable corresponding element in the models.

3. All remaining metrics were tested against the third requirement of easy replication of their measurement.

The resulting scale consisted of 11 metrics, categorized into four dimensions:
Brand Awareness, perceived quality / leadership, Loyalty / price premium, and Online word of mouth / share of voice. The scale included five product-related metrics, in order to compare the impact of the brand metrics with the impact of the individual product metrics, with regard to explanatory and predictive power:

- The number of professional and consumer reviews
- The average score of professional and consumer reviews respectively,
- The number of product mentions in the corpus (explained below)

making for a total of 11 metrics.

All of the metrics are defined in such a way that higher numbers should denote higher levels of brand equity, and thus all metrics are expected to be positively correlated with sales figures.

The online CBBE scale

	Brand Awareness	Perceived Quality / Leadership	Loyalty / price premium	Online word of mouth / share of voice
Search results				
Brand mentions in the corpus*	✓	✓		✓
Product mentions in the corpus*	✓	✓		✓
Search statistics				
Total number of reviews for a brand's products	✓	✓		✓
Total number of reviews for a product	✓	✓		✓
Brand connection to product category	✓	✓		
Social media				
Facebook profile likes	✓		✓	✓
Twitter profile likes	✓		✓	✓
Twitter profile followers	✓		✓	✓
Reviews				
Professional review scores (average)		✓		
Consumer review scores (average)	✓	✓		✓
Number of consumer reviews of a product	✓			✓

* The corpus is described in detail below

Table 1: The online CBBE scale

Search results - the corpus

The corpus is a collection of 100 web sites, and is unique to each product category. In order to identify the prominent brands and products within a product category, a Google search based on the product category is performed. The search is a string of text of this form: "Best [category]". So, if the product category being tested was, for instance, cars, the search string would read "best cars", and the corpus would consist of the first 100 web sites in the result.

The quotation marks here are important, as the Google search engine interprets terms inside double quotes as a literal string, whereas without the quotes, the order of words is ignored.
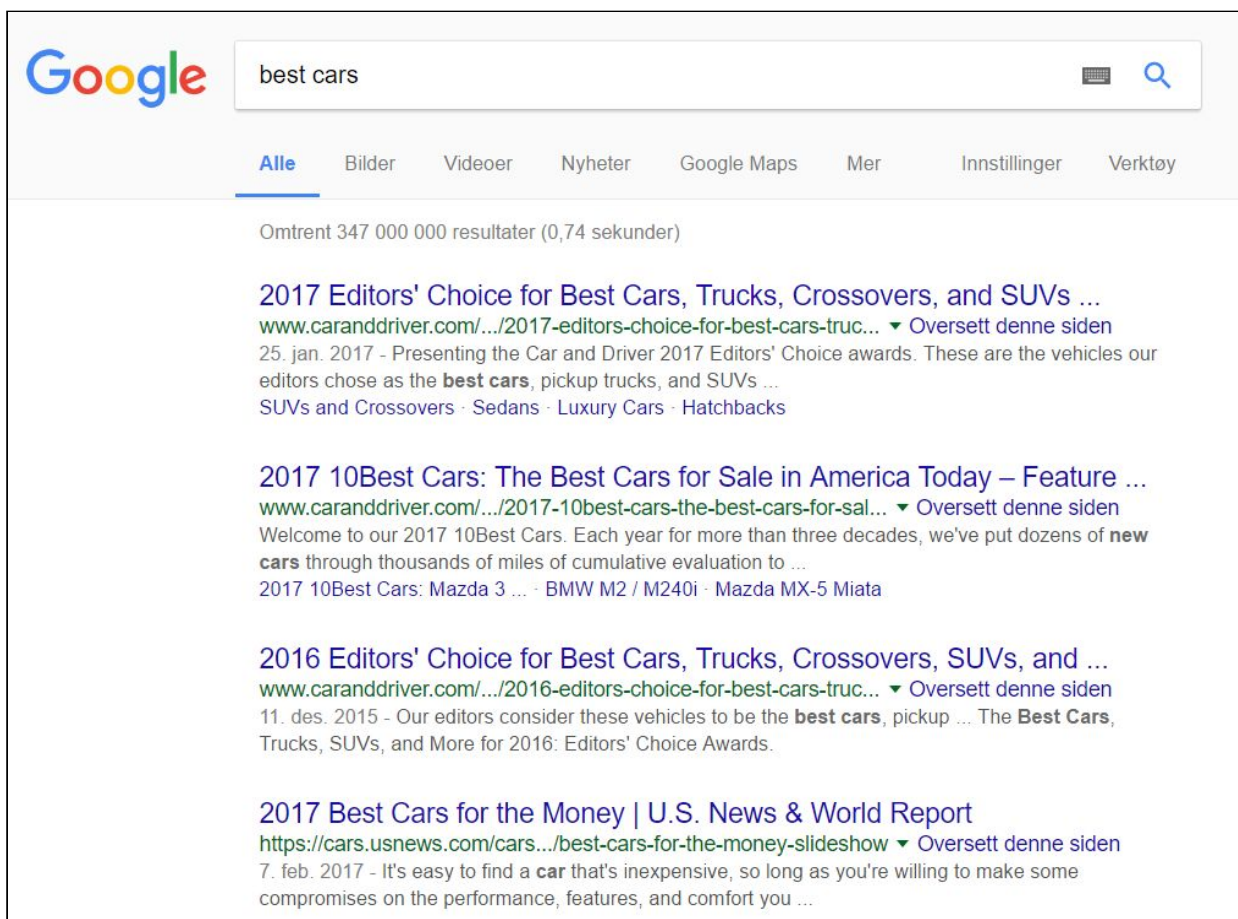

Figure 4: The search performed when collecting web sites for the corpus, done manually.

The two metrics that are collected from the corpus are brand and product mentions. The manual procedure is as follows:

1) Visit each of the 100 web sites.
2) Write down all the brands and products which are mentioned on each page.

3) For each brand and product, count the number of websites on which the brand or product was mentioned.

The reason for searching "best cars" instead of just "cars", is that a search for "best cars" will include (in many product categories) a high number of 'best-of'-lists. A search for "cars", however, will lead to a corpus that includes many web sites that explain what a car is, where you can buy one, and so on. Another reason for prefixing the term "best", is the assumption that potential buyers are more likely to do the same when using the web, in an attempt to make an informed purchase decision.

This means that a high number of corpus mentions for a brand means that the brand has produced many products that are often included in such 'best-of' web sites, which suggests that the brand is well known, and is associated with quality products. Similarly, a high number of product corpus mentions suggests that the product is well known, and of high quality. Thus, in terms of the online CBBE scale, these metrics are assumed to be associated with the dimensions brand awareness, perceived quality/ leadership, and online word of mouth/ share of voice.

Search statistics - total hits

Similar to the corpus procedure, the search statistics metrics are collected by performing a search using Google's search engine. However, instead of visiting the web sites in the result, the total number of web sites in the result is collected. For instance, the search "best cars" in figure 4 returned 37 million web sites. The search statistics category comprises three CBBE metrics: The total number of reviews for a brand's products, the total number of reviews for a product, and a brand's connection to the product category. These are measured by conducting three individual searches, in which the total number of hits are collected.

Collecting the number of reviews of any product by a particular brand, follows the same logic as the counting procedure performed on the corpus; if a brand has an outstanding number of reviews of its products, then that serves as an indication that a lot of people have some degree of awareness of that brand. Furthermore, it could be an indication that the brand is considered by many to be a leader in one or more product categories, and it also contributes to the knowledge about how much attention this brand receives by reviewers. This metric does not, however, say

anything about whether the reviews are favorable. Thus, the number of reviews for any of a brand's products is considered to have a positive relationship with brand awareness, perceived quality/ leadership, and online word of mouth/ share of voice.

The total number of reviews for a specific product is expected to act as a proxy for the overall hype surrounding a product, and possibly its position as a leader or an underdog, compared to other products. Thus, this metric is also expected to contribute positively to the online CBBE dimensions of brand awareness, perceived quality/ leadership, and online word of mouth/ share of voice.

The argument for collecting information about a brand's connection to the product category, is that it investigates whether people are more inclined to purchase a product from a company that is well known on a general basis, or that they prefer to purchase a product from a company that specializes in that particular category. For example, Samsung is one of the largest and most well known electronics companies in the world, and is one of the leaders in product categories like TVs and smartphones. Does this entail that consumers will be inclined to purchase a vacuum cleaner from Samsung, despite the fact that this product category is dominated by other brands, such as Miele and Dyson? A brand's connection to the product category is expected to contribute to the brand awareness and perceived quality/ leadership dimensions of online CBBE, as a high number of website hits will reveal that the brand is strongly associated with, and well known in, the product category.

Social media

Facebook and Twitter are the two largest social media sites on the web, as of april 2017 (Lifewire.com). Both platforms publicly supply quantitative information about company pages/profiles: Facebook presents information about how many 'likes' a certain page has received, while Twitter presents several statistics, including how many people 'follow' a certain page (a subscription that notifies a user whenever there is activity on that page), and the number of people who 'like' the page. Thus, every brand's Facebook like count, and its Twitter likes and number of followers were collected.

The number of Facebook and Twitter likes and followers are assumed to be related to brand awareness, brand loyalty and online word of mouth. However, enjoying high levels of likes and followers serve an additional purpose for brands: Whenever a user has 'liked' a company's Facebook page, the user is automatically presented with updates from that company. This means that Facebook 'likes' give the brands an audience which already has positive associations towards the brand, and the means to broadcast information to them directly. Twitter followers work in the same way. This level of targeting is unmatched by conventional marketing channels like TV or radio commercials, and it gives rise to the assumption that all brands actively try to increase their score on these metrics.

Reviews

Both professional and consumer reviews are expected to act as proxies for perceived quality/ leadership in the online CBBE scale. Consumer reviews have also been shown in the literature to relate to brand awareness and online word of mouth, and is expected to have a similar contribution in this research.

Professional reviews - average score

Professional review scores are extracted in much the same way as the corpus web sites: By issuing a search. In the case of reviews for a product, a search of the type "[brand] + [product]", is performed, including the keyword 'review' in order to weed out irrelevant results. Then, the scores of the first 20 online reviews are collected, and the average score is calculated. At some point in recent years, Google made it possible for web sites to display the score of the review directly on the results page. This feature has been implemented by many professional review websites, and it simplifies the procedure of collecting review scores.

Figure 5: Review scores in the Google results when searching manually

Consumer reviews - Average score and number of reviews

Consumer reviews differ from professional reviews in that they are offered by an online retail service. By giving consumers the choice to review a product, the service provider hopes that consumers will recommend the product to other potential buyers. Customers may place more trust in a complete stranger than in the company that makes the product, as they expect other consumers to be unbiased and honest about flaws and drawbacks. In terms of collecting data about consumer reviews, this means that data collection would take place by targeting online retail services that offer this option.

Amazon.com is a leading ORS website which has offered consumer product reviews for years. Starting with online book sales, Amazon.com now offers millions of products in a wide range of categories, making it a natural starting point for collecting consumer reviews. Amazon.com is Google-enabled, making it possible to gather review information about products in the same way as with professional reviews.

4.2: Web retrieval using developer tools

Research Question 3*: Is it possible to create a web mining application which accurately measures CBBE across product classes?*

Manual web retrieval such as the procedures which were described in the preceding section is less time consuming than conventional CBBE research that utilizes survey methods with real people. However, when using Google.com manually, the results that are returned by the search engine are heavily influenced by the information that Google has about the user, such as demographical and geographical information. Thus, the corpus, search statistics and review scores will vary according to who and where the user is, which is not ideal. Furthermore, the procedures as described above must be done manually, by people. While faster than conventional methods, the procedures would be more efficient still if they were automated. Thus, the procedures that were implemented in this research, utilized developer tools that lend themselves well to automation via computer programs.

4.2.1: Google Custom Web Search API

Google's 'Custom Web Search API' (application programming interface) lets developers use Google's search engine without a graphical user interface (GUI) such as a browser. Furthermore, it returns results which are independent of Google's personalization algorithms. This means that the results from the API are not tailored for specific user demographics (age, sex) or geographical information, and it makes this strategy more accurate than performing the same searches using Google's search engine manually in a browser. Lastly, the Google Custom Web Search API is easily integrated in computer programs, which makes it possible to automate search activity. The Google Custom Web Search API was used to collect data in three of the four categories in the online CBBE scale: Search results (the corpus), search statistics, and reviews.

Search results - the corpus

The search "best [category]" was configured to return the first 100 results (web page URLs). As all URLs point to a specific HTML document (the actual web page), this makes it possible to

download and store the contents of the web pages. The 100 documents were downloaded, constituting the corpus. The corpus represents an excellent basis for many types of text analysis. In this research, the goal was counting the number of websites on which brands and products were mentioned. The procedure for extracting these numbers is explained by this pseudo-code:

for each brand and product:

      for each web site in the corpus:

            check if the brand or product is mentioned on the web site.

Search statistics

In addition to the actual web sites, the Custom Web Search API also includes metadata about the search results in the response. Specifically, the metadata includes the number of total hits (web page URLs found) when performing the search. The number of total hits is easily extracted from the response, and stored. This strategy was used to gather information on three metrics on the CBBE scale: The total number of reviews for a brand's products, the total number of reviews for a specific product, and the brand's connection to the product category. Naturally, each of the metrics have their own search string, but the collection procedure is the same for all three, differing only in that the number of reviews for a specific product is not performed for brands, and vice versa:

for each brand (or product):

      perform search using specific search string

      extract the number of hits from the metadata in the response

Reviews

Google-enabled reviews simplify the manual labour of collecting and calculating average review scores online. However, for the purposes of this research, Google-enabled reviews provide an additional advantage: The review scores of Google-enabled reviews are included in the metadata returned by the Google Custom Web Search API. This means that also this metric can be made available for automatic collection and processing by a computer program. Professional review scores were collected by issuing a search on this form: "[Brand name] + [product name]" + "review". The procedure continues by scanning the metadata of the first 20 results for each search, and collecting the review scores as it finds them (not all of the web sites in the results are necessarily Google-enabled, and may not even actually be a review. This is all down to the

performance of the Google search engine). Lastly, the average review score for the product is calculated, and the number of consumer reviews is collected as well.

Web retrieval procedures using Google's Custom Web Search API

Online CBBE metric	Search string formula	Example search string	Collection procedure
Search results			
Brand mentions in the corpus	"Best [category]"	"Best cars"	100 first results are downloaded, and brand mentions are counted
Product mentions in the corpus	Same as above	Same as above	100 first results are downloaded, and product mentions are counted
Search statistics			
Total number of reviews for a brand's products	"[Brand name] + review"	"Ferrari review"	The search is performed, and the number of hits is extracted
Total number of reviews for a product	"[Brand name] + [Product name]" + "review"	"Ferrari F430" "review"	Same as above
Brand connection to product category	"[Brand name] + [Category]"	"Ferrari cars"	Same as above
Reviews			
Professional review scores (average)	"[Brand name] + [Product name]" + "review"	"Ferrari F430" "review"	20 first results are collected, and all google-enabled review scores are used to calculate the average score.
Consumer review scores (average)	"Amazon.com" + "[Brand name] + [Product name]"	"Amazon.com" "Ferrari F430"	The first result is collected, and the google-enabled review score is extracted
Number of consumer reviews of a product	Same as above	Same as above	The first result is collected, and the number of reviews is extracted

Table 2: Web retrieval procedures using the Google Custom Web Search API

4.2.2: Direct web retrieval

Collecting the number of likes and followers on Twitter and Facebook is merely a matter of visiting the pages of the brands in question. However, this information is also available through web retrieval. In contrast to search results, search statistics and reviews, this information can not be retrieved via the Google Web Search API. This means that, in order to extract this information, a computer program needs to interact with the actual brand profiles on these social media platforms.

Conveniently, both Facebook and Twitter append the page name as a suffix in the URL. This means, for example, that the URL to Apple's Facebook page is www.facebook.com/apple. This makes it possible for a computer program to target the web site URLs directly, based only on the brand name. The URLs point to an HTML document, which can be downloaded just like any other publicly available online document. HTML code comprises everything that is visible on a web site, and is organized as nodes. Thus, simply searching through the document for keywords such as 'likes' or 'followers' is bound to fail, as there could be multiple places in a web page where these terms are mentioned. Thus, the very structure of these web pages needs to be considered. This is commonly done with HTML parsers; purpose-built functionality which identifies the node hierarchy within a web page, and makes it possible to target specific nodes. Thus, the structure of Facebook and Twitter pages are analyzed, and the relevant nodes for likes and followers are identified. After that, collecting the number of likes and followers is done by simply reading the value of the respective nodes. This is a procedure which is, naturally, more challenging and time-consuming for a person to perform, but a computer program can retrieve data in this way in milliseconds.

Procedures using direct URL based web retrieval

Online CBBE metric	URL formula	Example URL	Collection procedure
Social Media			
Facebook Likes	Facebook.com/ [Brand Name]	Facebook.com/ Sony	The HTML document is downloaded and parsed. Then, the 'likes' node is identified and its value is collected
Twitter Likes	Twitter.com/ [Brand Name]	Twitter.com/ Sony	The HTML document is downloaded and parsed. Then, the 'favorites' node is identified and its value is collected
Twitter Followers	Same as above	Same as above	The HTML document is downloaded and parsed. Then, the 'followers' node is identified and its value is collected

Table 3: Procedures using direct URL based web retrieval

4.2.3: The application

In order to automatically perform the procedures, the author designed a computer application. The application was designed to work across product categories, and should be able to retrieve accurate and relevant brand and product data from the web, without any prior 'knowledge'. This meant that the application had to be designed in a way which enabled the program to receive information about which category to investigate, as well as some information about brands and products. In order to allow this, the application was designed to accept data input from the user, prior to running the algorithms. The data input must be in the form of a .csv (comma seperated variables) file. This is a format that most popular databases can export. The file must contain product data such as brand name, product name and so on, and the first line in the file declares

the product category. This allowed the application to be completely agnostic about product categories, brands and products.
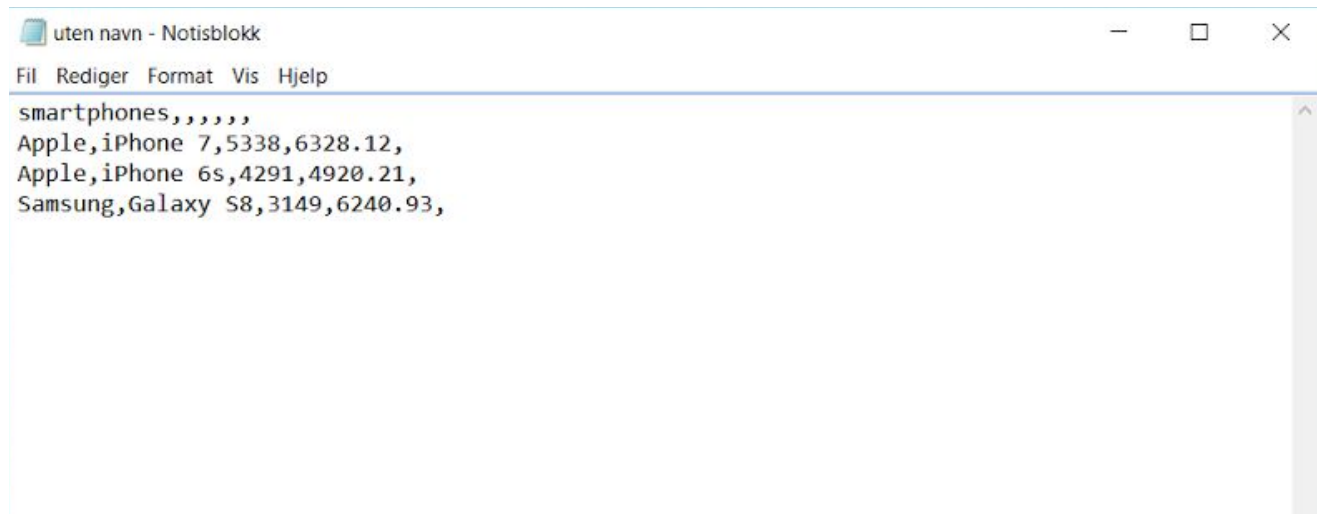
Figure 6: Example of a .csv input file

The application uses the input data in order to automatically perform the previously described searches on the internet, based on the product information. The application then collects the relevant information through web retrieval, and populates the metrics with the results, through a series of purpose-built algorithms. When all algorithms have been performed, and all metrics have been populated with the retrieved information, the application outputs a new .csv file. This file contains all the original data, as well as the new data which has been collected from the web.

4.3: Analysis

4.3.1: Multiple regression

"Multiple regression is a statistical tool used to derive the value of a criterion from several other independent, or predictor, variables. It is the simultaneous combination of multiple factors to assess how and to what extent they affect a certain outcome" (Techopedia.com).

With regard to this research, multiple regression is used to assess the impact on sales, of all metrics, individually. Multiple regression analysis provides information about the direction of

correlations (positive or negative), ranging from 1 to minus 1, and also returns the levels of statistical significance for each independent variable (the online CBBE metrics).

The application's output file consists of brand and product data submitted by the user, as well as one column for each of the web mining metrics. This file was loaded into R studio, a computational statistics program. Then, multiple regression analysis was performed on the complete set of headphones.

According to Keller (1993), "The favorability, strength, and uniqueness of brand associations are the dimensions distinguishing brand knowledge that play an important role in determining the differential response that makes up brand equity, especially in high involvement decision settings". Thus, there is reason to believe that brand equity's impact on sales could increase as price goes up. In order to investigate this, three additional analyses were conducted. The complete set was partitioned into three subsets based on price, with cut-offs at NOK 1000 and NOK 2500. These subsets were then analyzed individually.

4.3.2: Significance level

The significance level was defined as $p < 0.1$, which is slightly higher than the norm ($p < 0.05$). This is justified by the fact that this research is explorative, and based upon a novel conceptualization which is itself based on a conceptualization of the highly intangible phenomenon that brand equity is. Because of this, drawing definitive conclusions based on the analysis results of individual metrics would be prone to statistical errors. Thus, the interpretation of the results will assume a bird's eye view, focusing more on general trends than specific results. Setting the significance level at $p < 0.1$ will ensure that all aspects of general trends will be captured, while sacrificing some of the confidence that can be placed in individual results.

5: Data collection

The application's web mining approach was tested on 62 products from 18 brands in the category of headphones. Headphones constitute single sale products, enabling the results to be compared with previous studies in the literature, most of which also revolves around single sale products such as movies, books and video games.

Product data was made available by Spaceworld Soundgarden, a leading norwegian consumer electronics retail chain. The data was gathered directly from Spaceworld Soundgardens product database, and included the following: Brand name, product name, number of sold items and price, as well as boolean values (true/false) for whether the headphone had bluetooth and/or noise-cancelling features. The average price was calculated and used instead of retail price, to avoid the effects of situations where some products are sold at a reduced price, such as promotions. The product data contained information about all headphones sold in a physical store or online in 2016.

5.1: The application - web mining

5.1.1: The corpus

The application issued the search "Best headphones" which returned the first 100 URLs of the results from Google's custom web search API. Of the 100 URLs, 97 web sites were downloaded. The application analyzed the 97 websites once for each product and once for each brand. The most prominent brand, Bose, was mentioned in 49 of the web sites, slightly over 50 percent of the corpus. Sennheiser and Sony followed with 43 and 38, respectively. Two brands (Tiny Audio and XTZ) were not mentioned on any of the 97 web sites. The most prominent products were Bose QuietComfort 35 (22 mentions) and Sony MDR-1000 (16 mentions), with three other products receiving 15 mentions. 25 products were not mentioned at all, and a further 15 were mentioned in less than 5 of the web pages.

5.1.2: Search statistics - total hits

Search statistics were collected by issuing a number of searches using Google's custom web search API, and collecting the total number of results (search hits). Information about the total number of results were collected on on three metics. An overview of the search strings and results can be found in table 4 below.

	Example search string	Highest	Lowest
Total number of reviews for a brand's products	"Bose review"	60 million (Pioneer)	14.200 (Tiny Audio)
Total number of reviews for a product	"Bose QuietComfort 35" "review"	6.7 million (Beats Solo 2)	11 (Koss MyOwn)
Brand connection to product category	"Bose Headphones"	17.7 million (Sony)	25.000 (Tiny Audio)

Table 4: Search statistics strings and results

5.1.3: Social media - Likes and followers

Social media metrics were collected by letting the application build the various URLs directly, using the brand names as described above. Of 18 brands, only 12 of the returned web sites correctly corresponded to the actual company's Facebook page. The application performed better on Twitter, but still missed 3 brand pages. After investigating the issue, it became clear that some of the pages did not follow the convention of adding the brand name as a suffix. Some brands have several Facebook pages, and did not have a single main Facebook page, instead using suffixes denoting geographical information such as appending USA at the end. Additionally, some brands used suffixes to denote their various divisions, such as 'audio' or 'electronics'. This lead to Facebook URLs like Sennheiser's being on the form: www.facebook.com/SennheiserUSA, and it caused the application to fail in retrieving the brand information. The same was true for Twitter, but in fewer cases. Thus, additional methods in the application were implemented in the case that the retrieved social media page was missing. These methods utilize the Google web search API, and searches for the social media pages by issuing requests on this form: "Facebook.com" + "[Brand name]". The top result which was

returned was the correct social media page in all of the faulty cases, enabling the application to collect the correct data from all the brands in the set.

After the correct social media page was downloaded, the application employed an HTML parser in order to retrieve the relevant data about likes and followers. In the beginning of 2017, Twitter altered the structure of their HTML code in such a way that the application no longer found the Twitter scores for likes and followers. Thus, the part of the application's logic that was responsible for parsing and extracting those metrics from the download Twitter web pages had to be remade. This sort of changes in the resources used by the application may happen any time, and without notice, making this strategy for collecting social media data unfit for production software. A safer alternative would be to utilize Twitter's API, thus outsourcing the technical data gathering to Twitter itself.

The scores in social media showed a significant span between the headphone brands in the set. The highest performer on Facebook was Oppo with over 20 million likes, followed by Beats, Philips and Sony, all having between 7.5 and 10 million likes. The brands with the fewest likes were Tiny Audio and XTZ, both with less than five thousand likes, and the only ones with less than one hundred thousand likes.

Twitter likes was led by Beats, with over 26.000 likes. Beats was followed by Sony, Bose and Monster, all with over 10.000 likes. Tiny Audio and XTZ had zero likes, while Panasonic, interestingly, had only 14 likes on their Twitter page. Sony performed the best in terms of Twitter followers with over 4 million, while Tiny Audio (2), XTZ (43) and Koss (142) had the lowest amount of Twitter followers.

An overview showing the highest and lowest scores is presented in table 5.

	Highest	Lowest
Facebook Likes	Over 21 million (Oppo)	751 (Tiny Audio)
Twitter Likes	Over 26.000 (Beats)	0 (Tiny Audio, XTZ)
Twitter Followers	Over 4 million (Sony)	2 (Tiny Audio)

Table 5: Social media results

5.1.4: Reviews

Professional reviews

Professional review scores were gathered using Google's Custom Web Search API. The application found between 5 and 9 review scores for each product, and calculated the average score. Taking for granted that all the retrieved web pages were indeed reviews of the product, this means that between 55% and 75% of the relevant web pages were not Google-enabled. All products apart from one outlier (Koss Myown, with an average score of 20%) had received average scores between 67% and 95%. Only 12 products had received an average score of less than 80%, while 6 products scored 90% or over. The brands with the biggest variance in terms of percentage points were Koss (60), AKG (18), Sennheiser (15) and Sony (14). Oppo had the lowest variance in average scores, with 89% and 91% for its two headphone models.

Consumer reviews

Consumer reviews were also gathered using Google's Custom Web Search API, but targering only results coming from Amazon.com. The way Amazon.com is Google-enabled, is that it displays the average score directly in Google's results, rendering mathematical calculations unnecessary. Amazon.com also displays the number of reviews for each product directly in the search results, making this one of the simplest procedures of gathering product data by web retrieval.

Figure 7: Average score and the number of reviews on Amazon.com

The main challenge of using only Amazon.com for the purpose of collecting consumer reviews, is that the number of reviews for each product varies greatly. This means that the confidence that can be placed in each average score may differ from product to product. The number of reviews on Amazon.com varied from zero (two products), to about 8.500 (Sony MDR-100). The average score varied from 46% (Philips M2BT) to 100% (Sony MDR-XB450 and Oppo PM1). However, these extreme scores were all supported by less than ten reviews. It would be natural to assume that the scores would approach normality given a higher number of consumer reviews. A comparison with the professional review scores supports this; the biggest differences were found in the products who had received the fewest Amazon.com reviews, with differences in average scores between 11 and 34 percentage points for the ten products with the fewest Amazon.com reviews. For all products who had received 10 or more Amazon.com reviews, the differences ranged from 0 to 15. This suggests that niche products with few Amazon.com reviews may yield average scores that are artificially high or low.

6: Analysis

The set of headphones was analyzed utilizing multiple linear regression on all the product data gathered from Spaceworld Soundgarden's database, together with all the data about each product and brand that was collected by the application. The set was also partitioned into three subsets based on price range, which were analyzed individually. In addition to the product data and the application data, one additional column were added to the analysis: Reviews vs Category. This column was calculated manually after the application had populated the data through web mining. Reviews vs Category was calculated by subtracting the brand connection to category hits from the total number of a brand's product reviews. This measure was intended to distinguish between brands who are well known in general, from the brands who are mainly associated with the product category.

The low-price category included headphones from up to NOK 999, the mid-price category between NOK 1000 and NOK 2499, while the high-price category contained headphones whose average price was above NOK 2500.

7: Results

7.1: The complete set

Multiple linear regression on all 62 headphones revealed four statistically significant results. Noise cancelling (product feature), product corpus mentions, the number of Amazon.com reviews, as well as the calculated reviews vs category measure were all statistically significant. All of the significant measures were in the positive direction, meaning that higher numbers in these categories are correlated with higher sales figures. One interesting observation is that half of the metrics were negatively correlated, although none of these correlations were statistically significant.

```
Whole set

Coefficients:
                      Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)          6.256e+02   7.821e+02    0.800    0.4278
Price               -7.759e-02   4.654e-02   -1.667    0.1021
BTTRUE              -1.065e+02   2.030e+02   -0.524    0.6025
NCTRUE               5.304e+02   2.742e+02    1.935    0.0591 .
AvgReviewScore      -7.985e+00   9.645e+00   -0.828    0.4119
ProductReviewsCount -1.653e-04   1.322e-04   -1.250    0.2176
BrandCategoryHits   -3.045e-05   4.640e-05   -0.656    0.5149
BrandReviewHits      1.377e-06   3.618e-05    0.038    0.9714
FacebookLikes       -1.920e-05   2.610e-05   -0.735    0.4657
TwitterFollowers    -2.073e-04   1.263e-04   -1.642    0.1072
TwitterLikes         4.185e-02   4.020e-02    1.041    0.3031
BrandCorpusMentions -4.352e+00   6.996e+00   -0.622    0.5369
ProductCorpusMentions 5.672e+01  2.166e+01    2.618    0.0119 *
AmazonReviewsCount   1.527e-01   7.663e-02    1.993    0.0521 .
avgAmazonReviewScore 3.286e+00   5.365e+00    0.613    0.5431
ReviewsVsCategory    2.130e-05   1.061e-05    2.007    0.0505 .

Residual standard error: 647.4 on 47 degrees of freedom
Multiple R-squared:  0.4207,    Adjusted R-squared:  0.2481
F-statistic: 2.438 on 14 and 47 DF,  p-value: 0.01146
```

Figure 8: Results of multiple regression on the complete set

The product data metrics gathered from Spaceworld Soundgarden's database consisted of price, and bluetooth and noise cancelling features (boolean). Both price and bluetooth functionality seem to be negatively correlated with sales, while noise cancelling is positively correlated. Product corpus mentions and brand corpus mentions have opposite directions; while product mentions on web sites is positively correlated with sales, brand corpus mentions is negatively correlated.

Neither of the search statistics metrics (product reviews count, brand category hits, and brand review hits) were statistically significant. Brand review hits (the total number of online reviews for that brand's products) was the only search statistics measure to have a positive correlation with sales, while the other two were negatively correlated with sales.

The social media metrics comprised Twitter likes and followers, and Facebook likes. None of the metrics had statistically significant correlations with sales. Twitter likes was positively correlated with sales, while Twitter followers and Facebook likes were negatively correlated. The analysis results of consumer review measures gathered from Amazon.com suggests that both the score and number of reviews for a product is positively correlated with sales, with the number of reviews being a statistically significant result. The average score of professional reviews was not a significant measure, and its correlation with sales was in the negative direction.

7.2: Price segments

7.2.1: High-end (NOK 2500 and over)

The high-end segment consisted of the 17 most expensive headphones in the total set, with prices spanning from NOK 2.508 to NOK 13135. 8 metrics were statistically significant in this segment.

```
Over 2500

Coefficients:
                      Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)          -3.316e+03  1.206e+03   -2.749  0.05140 .
Price                -2.204e-02  9.272e-03   -2.378  0.07619 .
BTTRUE                2.235e+02  1.436e+02    1.556  0.19470
NCTRUE                2.835e+02  1.198e+02    2.366  0.07714 .
AvgReviewScore        1.789e+01  7.717e+00    2.318  0.08134 .
ProductReviewsCount  -4.306e-03  1.219e-03   -3.531  0.02420 *
BrandCategoryHits     1.888e-04  4.000e-05    4.720  0.00917 **
BrandReviewHits       2.141e-05  2.137e-05    0.022  0.85162
FacebookLikes         3.619e-06  8.739e-06    0.414  0.70002
TwitterFollowers     -7.224e-04  8.641e-05   -8.360  0.00112 **
TwitterLikes          2.160e-03  5.675e-02    0.038  0.97146
BrandCorpusMentions   2.245e+01  2.875e+01    1.480  0.56590
ProductCorpusMentions 1.310e+01  6.549e+00    2.000  0.11612
AmazonReviewsCount    3.883e-01  1.106e-01    3.512  0.02464 *
avgAmazonReviewScore  1.718e+01  1.070e+01    1.605  0.18379
ReviewsVsCategory    -1.524e-03  3.984e-03   -0.383  0.70137

Residual standard error: 81.95 on 4 degrees of freedom
Multiple R-squared:  0.9985,    Adjusted R-squared:  0.994
F-statistic: 222.2 on 12 and 4 DF,  p-value: 4.688e-05
```

Figure 9: Results of multiple regression on the high-end set

The high-end segment revealed several differences when compared to the analysis on the whole set of headphones. Here, the composite linear model (marked as intercept in figure 9) was significant in the negative direction.

Price was still negatively correlated in this segment, and the result was now statistically significant. Noise cancelling was still positively correlated with sales, and still a significant result. Bluetooth was still not statistically significant, but had changed direction and was now positively correlated with sales.

Brand mentions in the corpus was positively correlated in this segment, which was not the case in the complete set. Product mentions was still positively correlated, but unlike in the complete set, it was not statistically significant in the high-end segment.

Two of the search statistics metrics were statistically significant: The number of product reviews was negatively correlated, while brand category hits were positively correlated. Brand review hits were not a significant result.

The social media results were quite different in the high-end segment, with Facebook likes being positively correlated, and Twitter followers a statistically significant correlation in the negative direction.

The average professional review score had changed direction, now being positively correlated with sales, and a significant result. The number of reviews on Amazon.com remained a statistically significant positive correlation.

In the high-end segment, reviews vs category was negatively correlated with sales, contrary to the complete set.

7.2.2: Mid-range (NOK 1000 - NOK 2500)

The mid-range segment consisted of headphones between NOK 1000 to NOK 2500, and comprised 27 headphones, making this the largest of the price-partitioned sets. The mid-range segment showed no statistically significant correlations in any of the metrics.

```
Between 1000 og 2500

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          5.530e+02  7.073e+02   0.782    0.449
Price               -1.525e-01  1.197e-01  -1.275    0.227
BTTRUE               1.180e+02  1.088e+02   1.085    0.299
NCTRUE               2.154e+02  1.526e+02   1.412    0.183
AvgReviewScore      -2.562e+00  9.093e+00  -0.282    0.783
ProductReviewsCount  1.725e-05  4.970e-05   0.347    0.735
BrandCategoryHits    2.388e-05  2.994e-05   0.798    0.441
BrandReviewHits      4.370e-07  5.287e-07   0.012    0.673
FacebookLikes       -4.363e-05  5.115e-05  -0.853    0.410
TwitterFollowers     9.026e-05  1.758e-04   0.513    0.617
TwitterLikes        -7.500e-03  1.969e-02  -0.381    0.710
BrandCorpusMentions -3.541e+00  4.890e+00  -0.724    0.483
ProductCorpusMentions 1.491e+01 1.228e+01   1.214    0.248
AmazonReviewsCount  -2.331e-02  7.910e-02  -0.295    0.773
avgAmazonReviewScore 1.141e+00  2.116e+00   0.539    0.600
ReviewsVsCategory   -3.547e-07  1.318e-05  -0.027    0.979

Residual standard error: 196.2 on 12 degrees of freedom
Multiple R-squared:  0.5517,    Adjusted R-squared:  0.02876
F-statistic: 1.055 on 14 and 12 DF,  p-value: 0.4682
```

Figure 10: Results of multiple regression on the mid-range set

Average professional review score assumed a negative direction (as in the complete set, but opposite from the high-end segment), while bluetooth was positively correlated (as in the

high-end segment, but opposite from the complete set). All of the search statistics metrics were now positively correlated with sales. Twitter followers were positively correlated in this segment, contrary to both the complete set and the high-end segment, with both Facebook and Twitter likes being negatively correlated. The number of consumer reviews on Amazon.com was negatively correlated with sales in this segment, unlike both the complete set and the high-end segment, while the average consumer review score was positively correlated, as in the complete set and the high-end segment. Reviews vs Category was negatively correlated in this segment (as in the high-end segment, but unlike the complete set).

7.2.3: Low-end (NOK 1000 and below)

The low-end segment consisted of 18 headphones priced NOK 1000 or lower.
There were three statistically significant correlations in the low-end segment: Brand category hits was negatively correlated with sales, while the number of Amazon.com reviews remained positive, as in all of the analyzed sets. Twitter followers were also positively correlated with sales in this segment, opposite from the results in the high-end segment.

```
Under 1000

Coefficients: (1 not defined because of singularities)
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)          -1.322e+02  1.610e+03  -0.082   0.9385
Price                -8.551e-01  8.014e-01  -1.067   0.3461
BTTRUE                7.739e+02  3.993e+02   1.938   0.1247
NCTRUE                      NA         NA      NA       NA
AvgReviewScore       -1.912e+01  9.524e+00  -2.008   0.1151
ProductReviewsCount   1.164e-02  9.773e-03   1.191   0.2996
BrandCategoryHits    -1.995e-04  5.562e-05  -3.587   0.0230 *
BrandReviewHits       2.212e-03  2.890e-04   1.104   0.5327
FacebookLikes        -2.666e-05  4.711e-04  -0.057   0.9576
TwitterFollowers      1.375e-03  6.327e-04   2.174   0.0954 .
TwitterLikes         -2.323e-01  2.896e-01  -0.802   0.4674
BrandCorpusMentions   1.913e+01  1.787e+01   1.070   0.3448
ProductCorpusMentions -1.243e+02 1.213e+02  -1.025   0.3633
AmazonReviewsCount    7.111e-01  2.183e-01   3.258   0.0311 *
avgAmazonReviewScore  2.935e+01  2.359e+01   1.244   0.2813
ReviewsVsCategory     1.028e-05  9.499e-06   1.082   0.3402

Residual standard error: 416.8 on 4 degrees of freedom
Multiple R-squared:  0.9468,    Adjusted R-squared:  0.774
F-statistic: 5.477 on 13 and 4 DF,  p-value: 0.05666
```

Figure 11: Results of multiple regression on the low-end set

Price remained negatively correlated with sales also in this segment. Having bluetooth was also positively correlated. None of the headphone models in this segment had noise cancelling features. The two corpus metrics are opposite from the complete set, with brand mentions being positively correlated and product mentions being negatively correlated with sales.

The number of online reviews for individual products, as well as the total number of reviews for a brand's products, were both positively correlated with sales in the low-end segment. Both Facebook and Twitter likes were negatively correlated with sales. Professional reviews scores were negatively correlated with sales, while consumer review numbers and scores on Amazon.com were positively correlated. Reviews vs category was positively correlated with sales. A complete overview over results are is summarized in table 6 below. Green cells indicate positive correlations, while red cells indicate negative correlations. Statistically significant results are also highlighted.

Metric	Complete set	High-end	Mid-range	Low-end
Combined linear model		Significant		
Price		Significant		
BTTrue				
NCTrue	Significant	Significant		N/A
AvgReviewScore		Significant		
ProductReviewsCount		Significant		
BrandCategoryHits		Significant		Significant
BrandReviewHits				
FacebookLikes				
TwitterFollowers		Significant		Significant
TwitterLikes				
BrandCorpusMentions				
ProductCorpusMentions	Significant			
AmazonReviewsCount	Significant	Significant		Significant
AvgAmazonReviewScore				
ReviewsVsCategory	Significant			

Table 6: Summary of Results

8: Discussion

This section is divided into two main subsections; first a discussion about the results of the four analyses, followed by a discussion about the validity of the web mining approach as a means for measuring consumer based brand equity.

8.1: Analysis results

8.1.1: The complete set headphone models

The analysis of the complete set of 62 headphones yielded several interesting observations. Firstly, almost half of the metrics were negatively correlated with sales. This is curious, given that all of the included metrics were expected to contribute positively to brand equity. This seems to suggest that many of the collected metrics have less explanatory power when it comes to sales volumes, than results from previous research in the literature has suggested. However, the composite linear model itself, although not statistically significant, was positively correlated with sales. This means that despite the high number of negatively correlated metrics, the combination of all the brand equity metrics suggests that higher scores should translate to higher sales volumes.

Secondly, none of the social media statistics were statistically significant, and of the three, only likes on Twitter was positively correlated with sales.

Thirdly, the average score of professional reviews were negatively correlated with sales, while consumer review scores were positively correlated. Furthermore, the number of consumer reviews on Amazon.com had a statistically significant positive correlation.

Fourthly, brand mentions in the corpus was negatively correlated with sales, while product mentions in the corpus was positively correlated.

Search results - the corpus

The corpus comprised the first 97 (of 100 attempted) downloaded websites when issuing the search "Best + [category]". The measures 'BrandCorpusMentions' and 'ProductCorpusMentions' denote the number of web sites on which a brand or product is

mentioned. Product corpus mentions was positively correlated with sales, a statistically significant result. This indicates that products that figure often in 'best-of'-lists on prominent web sites in that category, sell better than products who are mentioned rarely or never on these web sites. It is, however, difficult to interpret the direction of the correlation. Does the product sell more because people visit these sites (or searches similar terms on Google) to get advice before the purchase, or does the product figure in many web sites because it is a popular product with significant distribution in the market? The brand mentions in the corpus was not a statistically significant correlation. However, it is interesting that, contrary to the product mentions, the brand mentions in the corpus were negatively correlated with sales figures. It could be that the readers of such online articles place more interest in individual products than in brands, but this interpretation presupposes the direction of the correlation to go from product corpus mentions to sales, instead of the other way around. Excluding the possibility that both product corpus mentions and sales are affected by a third variable, namely the product's actual quality, could lead to a type three error, as both sales and 'best-of' lists are likely to be affected by the performance of the product.

Search statistics - Google hits

The Google search statistics were gathered by performing searches via Google's Web Search API, and counting the total number of hits in the result. Three of the metrics were collected in this way by the application: Product reviews count (number of reviews for a given product), brand category hits (a brand's connection to the product category), and brand review hits (total number of product reviews for a brand). None of these measures were significantly correlated, indicating that search statistics are not predictors of product sales.

The reviews vs. category metric was calculated as the difference between brand review hits and brand category hits, in order to seperate specialists from brands that are generally well known. A high score on this metric would indicate that being a big brand in general is more important than being a specialist. Reviews vs. category was positively correlated with sales, and also a statistically significant result. This adds support to the notion that being a big brand is more important than being a headphone specialist.

Social media

Despite the high adoption of social media platforms, the results suggest that the efforts do not translate to higher sales volumes, as neither of the three metrics are statistically significant. Furthermore, likes on Facebook and followers on Twitter are negatively correlated with sales.

Reviews

As consumer reviews are shown in the literature to act as proxies for online word of mouth, the results suggest that online word of mouth plays a bigger part in explaining sales than professional reviews. This result supports the idea that consumers tend to place more trust in the opinions of other consumers, rather than professionals. Furthermore, it lends support to the notion that brands who are successful in creating activity among their customers, also enjoy higher sales figures.

Complete set: Summary of interpretations

The results of the analysis of the complete set suggests that a headphone model will have higher sales figures if its brand

   a) is a big brand in general, rather than a specialist in the product category.
   b) enjoys high levels of online word of mouth (many and favorable consumer reviews), rather than many and favorable professional reviews.
   c) creates class-leading products that are included in 'best-of' articles online.
   d) creates products that are not too expensive.

8.1.2: The high-end segment

The high-end segment consisted of 17 headphone models ranging in price from NOK 2.508 to NOK 13.135. In this segment, the multiple linear regression model was statistically significant in the negative direction, indicating that a high score on many of the metrics are correlated with lower sales figures. The high-end segment revealed eight statistically significant correlations; four negative and four positive. In contrast to the complete set, bluetooth functionality is positively correlated with sales in this segment. This is most likely explained by the fact that most of the models in this segment that have bluetooth functionality, also have noise cancelling

functionality (six of the models have both, one has only bluetooth and one has only noise cancelling), and so the metrics are almost completely connected in this segment.

Search results - the corpus

Neither of the corpus metrics are statistically significant in this segment. If operating on the assumption that consumers consult 'best-of' web sites before a purchase, this is an interesting result, especially given that product corpus mentions was a statistically significant result when analyzing the complete set. More than in the other segments, one would expect consumers to do research about the expensive headphone models, and that this would lead to a statistically significant correlation between product mentions and sales. Another difference between the high-end segment and the complete set, is that brand corpus mentions was positively correlated in this set. However, without statistical significance and a plausible theory about causation, it is difficult to draw conclusions based on this result.

Search statistics - Google hits

The number of product reviews was negatively correlated with sales, a statistically significant result in the high-end segment. This result suggests that significant hype surrounding a product does not contribute to higher sales figures. However, the high-end segment contains headphone models that are priced well above what most people are prepared to pay. Thus, even if the truly world-class models (priced over NOK 10.000) had received a lot of publicity in terms of professional reviews, they still would not sell as many units as the more reasonably priced headphones. This could contribute to the negative direction of this correlation.

The brand's connection to the category is positively correlated with sales in this segment, contrary to the result from the complete set. This result was statistically significant. This suggests that, when purchasing an expensive headphone model, consumers may prefer to buy from a headphone specialist. The number of reviews of a brand's products is also positively correlated, albeit not a statistically significant measure. This supports the idea that brands with a long history of product reviews in any product category tend to sell more, also when it comes to more expensive models.

Social media

One of the social media metrics stand out in the high-end segment: Twitter followers. In this segment, the number of twitter followers is negatively correlated with sales, and is a statistically significant result. This result may indicate that brands who are prominent on social media, and thus perhaps quite popular, may be considered 'mainstream' by the customers in this segment. It could also mean that, in combination with the positive correlation of professional average review score, buyers in this segment are more inclined to be influenced by product quality than by brand popularity. There exists a different possible explanation as well: If the brands who produce the most expensive headphones (NOK 10.000 and above) have many Twitter followers because they create exciting, state-of-the-art products, then that could explain the negative correlation with sales, as most people can not afford the most expensive headphones in the world, and thus even companies held in high esteem may not boast impressive sales figures of all their headphone models. This view would also explain why Twitter followers is negatively correlated, while both Facebook and Twitter likes are positively correlated, although neither of those metrics are statistically significant.

Reviews

The high-end segment is the only segment where the average score of professional reviews are positively correlated with sales, and the only segment where the correlation is statistically significant. This result supports the idea that buyers of expensive headphones rely on information on the web. Alternatively, it could be explained by the notion that buyers of expensive headphones listen to several models within their price range, and so they end up with a headphone model which delivers great value for money on the same terms as reviewers use, such as sound quality, build quality, features and so on. The most plausible explanation is that the result is based on a combination of these factors. As for consumer reviews on Amazon.com, the number of reviews per product is positively correlated with sales, with a higher significance level than in the complete set. The average consumer review score is also positively correlated with sales, but is not a statistically significant result.
This makes the high-end segment the only segment in which professional reviews have more explanatory power than consumer reviews

High-end: Summary of interpretations

The results of the analysis of the high-end segment suggests that a product will have higher sales figures if it

a) is made by a specialist in that product category, rather than a big brand in general.

b) receives high scores on professional reviews, but also enjoys high activity in terms of consumer reviews

c) is on the lower end of the price range, rather than being a top-of-the-line product

d) is made by a brand which is not considered 'mainstream', in terms of few followers on Twitter

8.1.3: The mid-range segment

The mid-range segment consists of 27 headphone models from 13 different brands, ranging in average price from NOK 1.056 to NOK 2.436. Ten of the models offer bluetooth functionality, while three models offer noise cancelling (one product offers both). The most striking result of the mid-range segment, is that none of the metrics are statistically significant.

The lack of statistically significant results in the mid-range segment, could be explained by the fact that this segment is by far the smallest in terms of product sales: 4423 units sold in this segment, with 9838 and 8484 in the low-and high-end segment, respectively.

The mid-range segment is the only set in which all of the search statistics metrics are positively correlated with sales. In general, these results seem to represent a transition from the trends of the high-end segment to the trends of the low-end segment. This is evident in the brand connection to the product category, which goes from being positively correlated with sales in the high-end segment to being negatively correlated in the low-end segment, both results being statistically significant. Also in social media, the results seem to constitute a transition from high-end results to low-end results, with Twitter followers being a statistically significant positive correlation in the low-end segment.

Mid-range: Summary of interpretations

Without any statistically significant results, it is difficult to draw conclusions about relationships between the metrics of the CBBE scale and product sales. There are signs that point towards the mid-range segment representing a transition from traits of high-end products, to traits of low-end products, notably in the metrics from social media and search statistics. As a whole, the results suggest that the CBBE scale performs poorly when it comes to explaining sales in this segment. However, the comparative lack of unit sales in this segment is likely to influence these results.

8.1.4: The low-end segment

The low-end segment consists of 18 headphone models from 10 different brands, priced from NOK 58 to NOK 977. None of the headphone models offer noise cancelling functionality, but five of the models offer bluetooth functionality. There were three statistically significant metrics in this segment; one was negatively correlated and two were positively correlated with sales figures.

Search results - the corpus

The brand and product corpus mention metrics are opposite from that of the mid-range segment, as well as the complete set, in that brand mentions are positively correlated while product mentions are negatively correlated with sales. This result supports the idea that brands are more important than individual models in this segment, and also suggests that not much research is performed by the customer before the purchase decision is made. However, not much confidence can be placed in these interpretations, as the results are not statistically significant.

Search statistics - Google hits

The total number of reviews received by an individual product is positively correlated with sales, but not a statistically significant result. It still marks a distinction between the high-end and low-end segments, as this metric had a statistically significant negative correlation with sales in the high-end segment. The total number of product reviews for a given brand is the same as in all other sets: Positively correlated but not a statistically significant result. The interesting metric within this web mining strategy is the brand connection to product category. Being strongly associated with the product category was a positive correlation in the high-end segment, while it

was negatively correlated with sales in the low-end segment. Both results were statistically significant, meaning that there are real differences between segments in this regard. Both brand review hits and reviews vs. category were positively correlated with sales in this segment, which, in combination with the category connection metric, indicates that buyers in the low-end segment prefers products from brands that are generally well known, and not mainly associated with headphones. This is a clear distinction between the high-end and low-end segments.

Social media

Another interesting result is that both Facebook and Twitter likes were negatively correlated with sales in the low-end segment, while they were positively correlated with sales in the high-end segment. However, neither Facebook likes nor Twitter likes have been statistically significant in any of the analyzed partitions, and so it is difficult to assess their explanatory power when it comes to sales volumes. The number of Twitter followers were positively correlated with sales figures in this segment, a statistically significant result. This marks the second clear distinction between the high-end and low-end segments, as Twitter followers were negatively correlated with sales in the high-end segment, also a statistically significant result. Drawing on the idea that buyers of high-end products respond negatively to 'mainstream' brands, the low-end segment embraces brands who perform well in terms of followers on Twitter. Viewed alongside the negative correlation of a brand's connection to the product category, it supports the idea that brands that are generally well known and popular are more attractive to buyers in the low-end segment, than brands whose associations are limited to the product category.

Reviews

Professional reviews are not a statistically significant metric in the low-end segment, and is also negatively correlated. Following the discussion of the opposite result in the high-end segment, this could mean that

a) Buyers in this segment do not consult online professional reviews in order to make an informed purchasing decision, or

b) Buyers in this segment do not test several models in order to find a headphone which represents good value for money in terms of relevant performance dimensions such as

sound quality, build quality and so on. Thus, they do not purchase products that, coincidentally, score well on professional reviews.

Both points support the idea that brand equity becomes more important in purchasing decisions that command high involvement, such that buyers in this price segment are more likely to purchase a product that has the right features, appealing packaging, or that it is made by a brand that the customer knows and towards which the customer has positive associations.

The results of amateur reviews on Amazon.com are identical to that of the high-end segment, with both metrics being positively correlated with sales, and the number of reviews being a statistically significant result. This suggests that products that enjoy high levels of consumer activity in the form of positive reviews are attractive to buyers in both the high-end and low-end segment. Viewed in combination with the Twitter followers result, it may seem that the general popularity of both the brand and the product positively contributes to sales volumes in the low-end segment. However, it is likely that many of the sales in this segment happen without the customer consulting the web before the purchase decision is made.

Low-end: Summary of interpretations

The results of the analysis of the low-end segment suggests that a product will have higher sales figures if it

   a) is made by a well known, popular brand rather than a brand that is mainly associated with the product category
   b) is popular, in that it enjoys high online activity in terms of consumer reviews
   c) is appealing for other reasons than just product performance, such as having desirable features, appealing packaging, nice design and so on.

8.1.5: Summary of interpretations

The results highlight a general shift from product-and quality orientation in the high-end segment, to brand-and popularity orientation in the low-end segment, with the mid-range segment representing mainly a transition between the two. The segmentation by price provides more insight than the complete set of headphones, which suggests that further segmentation by price might reveal more fine-grained differences. Partitioning based on other metrics than price could provide additional insights as well, but that is beyond the scope of this research.

Complete set	In order to sell many headphones, brands should strive to be well known in general, be able to stimulate users into writing many and favourable online reviews, and produce class-leading headphones without venturing into price points that are beyond most buyers' reach. Brands should not be too focused on professional reviews or social media, as higher performance in these areas will not contribute to increased sales.
High-end	In order to sell many headphones, brands in the high-end segment should strive to be perceived mainly as a headphone specialist, and not as a mainstream brand that attracts many followers. Brands should focus on producing headphones that impress professional and amateur reviewers alike, while avoiding the very highest price points.
Mid-range	The mid-range segment seems to represent a transition between traits of the high-end segment and traits of the low-end segment, and there exists no clear relationship between sales and the online CBBE scale in this segment.
Low-end	In order to sell many headphones, brands in the low-end segment should strive to be well known in general, and to produce well known products, as buyers in this segment seem to be influenced mostly by popularity. This suggests that buyers in this segment could be more easily persuaded by factors other than product performance, such as appealing design and desirable features.

Table 7: Summary of interpretations

8.2: The web mining approach to CBBE measurement

The results showed that the CBBE model is not immediately suitable for analysis based on any set of products. This is evident in the results from the complete set, which provided limited insight when compared to the high-and low-end segments. The fact that some of the operationalized variables were statistically significant, shows that the web mining approach is able to uncover brand-related online attributes' impact on sales. However, many of the operationalized variables were insignificant in all four sets. This suggests that a) the online CBBE model may contain metrics that do not have explanatory power on brand equity, or b) the product data and analyses were insufficient, in terms of uncovering insights from all metrics.

	Complete set	High-end	Mid-range	Low-end
Search results				
Brand mentions in the corpus				
Product mentions in the corpus	Significant			
Search statistics				
Brand connection to product category		Significant		Significant
Total number of reviews for a brand's products				
Total number of reviews for a product		Significant		
Social media				
Facebook profile likes				
Twitter profile followers		Significant		Significant
Twitter profile likes				
Reviews				
Professional review scores (average)		Significant		
Number of consumer reviews of a product	Significant	Significant		Significant
Consumer review scores (average)				

Table 8 - The online CBBE model in terms of significant results

The following section describes key challenges and limitations of utilizing web mining as an approach to CBBE measurement, grouped into four categories: Conceptual transitions, conceptual challenges, practical challenges, and data limitations. Then follows a discussion about whether web mining should be considered a worthwile approach to CBBE measurement, and recommendations for future research are proposed.

8.2.1: Conceptual transitions

The main challenge for doing CBBE research in this way, is the number of conceptual transitions required. The process began with the concept of brand equity itself; the value which is transferred from a brand to its associated products, that translates into increased sales volumes. It is natural to assume that some elements of brand equity were lost in translation when Keller and Aaker defined and conceptualized CBBE. However, the conceptualizations were validated through research in the following years, and were shown to accurately describe brand equity from the perspective of the consumer, and so this first step from BE to CBBE is assumed to preserve much of the essence of the original concept.
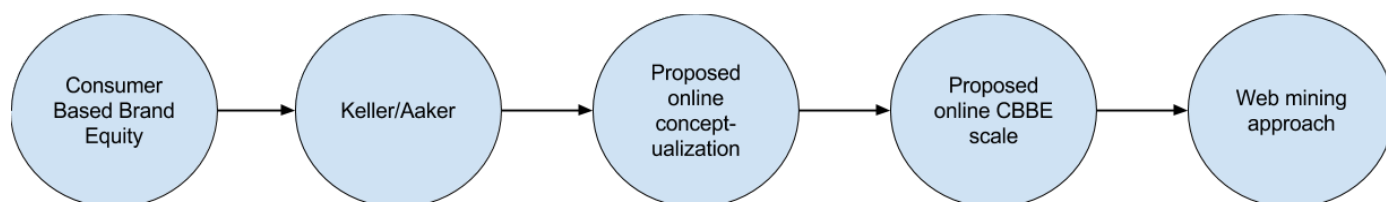

Figure 12: Conceptual transitions involved in CBBE measurement using web mining

In order to use web mining as a tool for measuring CBBE, a new re-conceptualization took place, which adapted the CBBE concept to the internet domain. It is natural to assume that some elements that were included in the original conceptualizations, are lost in the online conceptualization. Similarly, the online conceptualization may contain elements of CBBE that were not captured by the original conceptualizations.

Next in the process was the operationalization of the online conceptualization into metrics suitable for online measurement. Also in this step, parts of the original definition may have been lost, and new parts added. Thus, both the online conceptualization, and the online CBBE scale

that was formed on its basis, introduce uncertainty about the internal validity of the study and its results.

The last step of the process is the set of web retrieval algorithms. The uncertainty introduced in this step is limited by the fact that every measurement was possible to replicate manually. Furthermore, the data that is retrieved by the web mining algorithms corresponds quite directly to the metrics of the online CBBE scale, and so the conceptual transition is less dramatic than in the two preceding steps.

Thus, the conceptual challenges of performing web mining for the purpose of measuring CBBE, seem to lie in the conceptualization and operationalization steps.

8.2.2: Conceptual challenges

In addition to the challenges in the conceptual transitions, there are some specific observations that can be made about the nature of the online conceptualization of CBBE, and the online CBBE scale:

1) They favor brands with a long history. This may not be a significant deviation from reality, as brand equity as a concept also favors brands with a long history of marketing efforts. However, a young brand that in reality enjoys a sharp rise in brand equity, will still struggle to compete with established brands in terms of number of professional and amateur reviews, the number of reviews of a brand's products, likes and followers on social media and so on. Thus, in cases where a young brand enjoys a surge in equity, the brand's equity according to the online CBBE scale will struggle to keep up with reality. The same is true for brands whose equity is quickly diminishing; social media metrics, search statistics and review metrics will still reflect the brand's equity at its top.

2) They are not applicable to every product category. The reliance on information about professional and amateur reviews, means that the online CBBE scale will perform worse with products or services that are not typically reviewed by external web sites. It also excludes many service providers, such as stores, retail chains, public transportation companies and so on. The same is true for product categories where brands are generally not on social media platforms.

3) They do not consider product design and packaging. Keller(1993) includes packaging and design in his conceptualization of CBBE, as attributes of brand associations. This shortcoming of the online conceptualization is natural, given the fact that product and packaging aesthetics are not suited for measurement by computer algorithms.

8.2.3: Practical challenges

When designing the application, several practical challenges were revealed. These challenges are direct threats to the accuracy of measurement, and some of them constitute limitations of generality to the approach as a whole.

1) Social media profile naming conventions are not followed by all brands. The application builds social media URLs on the basis that the brand name will be added as a suffix. This was accurate for most of the brands in the headphone set, but not all. For example, JBL's twitter page is called "JBL Audio". This means that www.twitter.com/JBL, which is built by the application, does not correctly identify the profile. Furthermore, many of the biggest brands have several profiles on the same social media platforms. Some brands have dedicated profiles based on geographical areas. For example, Sennheiser has dedicated Facebook profiles for the United States and Europe, instead of one main profile. This issue is somewhat solvable, and was overcome in this research by finding the correct profile using the Google API, as a plan b in the cases where the URL-building algorithm returned empty results. This procedure succeeded in identifying the correct profiles of all brands in the test set, but is generally not a bulletproof procedure, as the Google API may return the wrong profile in other test sets.

2) Product lines with the same name, but no clear naming convention. Sennheiser's Momentum range proved difficult to deal with in terms of search statistics and reviews. The Momentum range was introduced in 2012, with three models: One around-ear design, one on-ear design, and one in-ear design (not included in the set, as these are earbuds and not headphones). In 2014, the range was renewed as Momentum 2.0, now with four models: On-ear and around-ear as before, and also two versions that were wireless via bluetooth. This lead to challenges in automatically performing searches based on product names, as it lead to searches on the form "Sennheiser momentum 2.0 on ear wireless". The quotations could not be removed, as the results would then

include any of the momentum models. This search would exclude web sites that did not contain the exact terms, in that order. So, in order to limit the results to only this model, some web sites that did contain reviews for this product, but did not contain that exact sentence, were left out of the results. No search strategy was identified that could include all relevant results for that product, and at the same time exclude all results for the other models in that product line.

3) Brands and products whose names are common words. This proved to be a challenge for the accuracy of several of the algorithms, regarding the corpus and the search statistics.
One of the brands in the headphone set was Monster, and one of Jabra's headphone models was called 'Move'. This lead to artificially high numbers in terms of corpus mentions, and also in terms of the search statistics regarding reviews. The corpus mentions issues were fixed by adding application logic that tested for capitalization, thus removing those corpus mentions that were caused by sentences like "The headphone stays on your head when you move around" (the word 'move' being previously counted as a corpus mention). The search statistics issue was fixed by refining the use of quotation marks in such a way that the brand name and product name were grouped together, on the form of "Jabra Move" instead of "Jabra" "Move".

4) Changes in the structure of HTML-parsed web sites. The discovery of this problem happened by coincidence: During the testing phase of developing the application (in january 2017), Twitter altered the HTML-structure of their website in such a way that likes and followers were no longer found by the application. Changes to the HTML-structure of web sites can happen at any time, and thus render parts of the application useless. This challenge can be overcome by utilizing Twitter's own API, as Twitter would then be responsible for ensuring that changes in their code do not impact the results of web retrieval. However, in cases where web sites do not provide API's, URL-based web retrieval may fail at any time due to changes in HTML structure.

5) Not all review web sites are google-enabled. The application was originally configured to collect the first ten reviews from the Google API results, and extract the google score from the metadata. However, the results for some products contained only one or two google-enabled reviews. Thus, the number of web sites retrieved was increased to 20. The percentage of review sites that offer Google-enabled review scores is expected to rise in the future, as more web sites take advantage of the functionality. In the meantime, alternative approaches should be found

when examining product categories in which reviews occur less frequently than in the headphones category.

8.2.4: Data limitations

The biggest limitation in terms of data is the size and completeness of the data set. The brand and product data was provided by Spaceworld Soundgarden, a sizeable electronics retail chain with stores in most parts of Norway. Still, there are some notable sources of potential threats to the internal validity of the research:

1) Distribution. Not all headphone models are sold in all stores, which means that some headphone models in the set have had limited distribution, compared to others. This may affect the analysis results, as some of the headphone models with limited distribution may come from a brand that enjoys high levels of brand equity. As these models are likely to have sold less than if they were equally distributed across all stores, the linear model may place too little weight on the performance of the brands producing these products, in terms of the online CBBE scale.

2) Products sold for a limited time. Some products are purchased by the retail chain in bulk, intended for marketing campaigns during a limited time, such as a summer or christmas sale. Furthermore, some of the headphone models was released during the year, and thus had less time to sell, compared to models that were released in 2015 or older. This point is distinguished from the previous point, by that the headphone models affested by this may well have been sold in all stores. However, the potential impact on analysis results is the same: Headphone models affested by this have had less time to sell, and thus are likely to have sold less units than if they had been sold year-round. This could have a similar impact on the results as distribution differences.

3) Not all products are in direct competition. This is mainly because of price. The analysis of the complete set of headphones revealed that price was negatively correlated with sales. However, some of the products are so expensive that most people can not afford to buy them. When Oppo released the headphone model PM1, at a price of over NOK 13.000, they probably did not expect to sell thousands of units in one retail chain alone. In terms of the online CBBE scale, Oppo may score well on many brand equity metrics, and because of the relatively low sales figures, this may cause the linear model to add to little weight to those metrics, as they had low

explanatory power in terms of sales. This suggests that it would be a good idea to further partition the data set on the basis of similar price and features, so that analysis is based on products that are in actual competition.

Challenges inherent in the online conceptualization and operalization of CBBE, as well as the practical challenges listed above, cast doubts about whether web mining is a worhwile approach to CBBE measurement. However, this does not mean that the approach is without merit.

8.2.5: Merits of web mining as an approach to CBBE measurement

Throughout this research, some challenges to the web mining approach to CBBE measurement has been identified. Two main limitations have been presented; the first being the re-conceptualization and operationalization phases, where it is argued that some aspects of CBBE are lost in translation. The second limitation is the applicability across product categories. Here, it is argued that product categories where brands are generally not on social media platforms, and where products are not typically reviewed, are not suitable for measurement by web mining. These limitations are considered the main shortcomings of the research approach. However, the web mining approach to CBBE measurement is not without merit, as the approach shows promise in several key areas:

1) Web mining as a tool for measurement is effective and reliable. This was confirmed in this research by manual reproduction of the results of the web mining algorithms.

2) Related work supports the approach. Research in the literature suggests that both reviews and social media have a direct or indirect impact on sales, and should thus be able to provide insights on dimensions of CBBE. Some of the research in the literature claim that reviews and social media metrics have predictive power over sales. However, for the purposes of this research, it is sufficient that they have explanatory power, predictive or otherwise.

3) The analysis results suggest that real insights can be gained through this approach. The main finding was the shift from product-and quality-orientation in the high-end segment, to brand-and popularity-orientation in the low-end segment. The apparent shift was supported by several statistically significant metrics, suggesting that the online CBBE scale is not without merit.

4) Web mining is cheap and efficient. This constitutes the main advantage of the web mining approach over conventional research using survey methods, as they are expensive and time consuming.

5) The web mining approach is applicable to many product categories. Although this research has highlighted some challenges in terms of applicability across product categories, the approach is applicable to many interesting categories such as cars, TVs, smartphones, furniture, tools, and more. In general, as long as the brands in the category have a presence on social media, and the products are reviewed, the web mining approach may be applied. Furthermore, it is likely that the number of eligible product categories will increase over time, as more brands adopt social media strategies, and product reviews in additional categories emerge.

This research constitutes the first attempt to measure CBBE by retrieving and analyzing publicly available online resources. Several key challenges have been identified, some of which are not immediately solvable. However, the prospects of a much cheaper and more efficient approach compared to conventional CBBE research, makes web mining an interesting alternative. Furthermore, the analysis results found in related literature, and also in this research, suggest that the approach may prove to be a valid alternative to surveys. However, before the approach can be considered trustworthy, it needs to be validated through further study.

8.3: Future work

There are a number of factors related to this research that requires further study. One of the key aspects is the need for validating the re-conceptualization and the online CBBE scale. This validation should consider the aspects of CBBE that are lost in translation in the re-conceptualization and operationalization phases of this study. It is only when these phases are shown to preserve enough of the original concepts, that the results of web mining can be compared with confidence to the results of existing literature using survey methods.

Another potential area of further research is the expansion of the CBBE scale, along various dimensions. One potential addition is sentiment analysis on online content, such as social media

and reviews. In social media, consumer discussions about brands and products may provide a more fine-grained measurement than likes and followers alone. In terms of reviews, the scores may be tied to individual product aspects such as build quality, performance and so on. Further textual analysis of this type may even provide information about design and packaging, product-related aspects of brand equity that are beyond the scope of the current online CBBE scale.

Some of the limitations related to data should also be improved upon in future research. Challenges of this research in terms of data, included the analysis of headphones with limited distribution, as well as headphones that were available during limited times during the year. A satisfactory solution to this issue would be to collect nation-wide data about headphone sales, and removing products that were unavailable during large parts of the year.

Analysis could be improved as well. This research analyzed the complete set with regard to sales figures, as well as three subsets based on price. One possible improvement on this would be to analyze products that are, in reality, in direct competition. The results of this research indicates that multiple regression analysis performed on groups of competing products, in terms of price and features, would likely be a better way of validating the online CBBE scale. This would perhaps require expert knowledge, and could be done in combination with experience interviews.

Lastly, the web mining approach should be applied to other product categories. Throughout this research, domain-specific challenges like product naming conventions emerged. Applying the web mining approach to different product categories may reveal other domain-specific challenges, and would also constitute a natural step in validating the generality of the online CBBE scale.

9: Conclusions

The goal of this research was to investigate whether web mining constitutes a worthwile alternative approach to CBBE measurement. The idea was motivated by the prospects of efficiency and cost reduction, and was supported by results from related literature on social media and reviews. The research consisted of several activities necessary for adapting existing conceptualizations for applicability of measurement by online resources. These activities included re-conceptualization based on existing work, operationalization based on available online resources, and the production of a computer application intended to perform web mining.

The operationalization phase resulted in an online CBBE scale, which was validated through a multiple regression analysis. The analysis based on web mining showed that the online CBBE scale as a whole was not significantly correlated with sales, but individual metrics were able to highlight differences between expensive headphone models and budget models.

Throughout the research, challenges to the validity of the approach has been presented, some of which are not easily solvable. The approach is promising in that it constitutes a cheap, efficient alternative to existing methods based on surveys. Furthermore, the approach has been shown in this research to provide insights on the relationship between online resources and CBBE. Recommendations for future work was proposed, highlighting the need for improved data, the application of the approach to different product categories, and further work on the re-conceptualization and operationalization phases, as well as limiting analysis to competing products.

Four research questions were devised:

RQ 1: Can classical consumer based brand equity be re-conceptualized into dimensions that are appropriate for web information retrieval?
RQ 2: Can these dimensions be operationalized in terms of freely available online data?

RQ 3: Is it possible to create a web mining application which accurately measures CBBE across product classes?

RQ 4: Do these measurements of web based brand equity correlate with sales figures?

RQ1: Yes, but not completely. Some aspects that are covered in the conceptual models of Keller (1993) and Aaker (1996), such as design and packaging, were excluded in the online conceptualization. Other aspects may have been excluded, and some aspects may have been added in the process. The extent of the correlation between the online conceptualization and the original ones, needs to be examined through further work.

RQ2: Yes. This part of the research contains somewhat less uncertainty than the conceptualization phase, and the author has presented supporting arguments for the inclusion of each of the metrics in the online CBBE scale. However, it is plausible that some aspects of the dimensions have limited coverage in the online CBBE scale. This could be improved by expanding the scale to include, for example, sentiment analysis.

RQ3: Yes, but not all product classes. During the research, it became apparent that web mining is an accurate tool for measurement of online resources. Thus, any lack of confidence in the approach as a whole, lies in the activities that must be performed in order to make web mining applicable. These activities are also responsible for the limitation in terms of generality across product classes; the re-conceptualization and operationalization phases lead to a CBBE definition that excludes many product classes. This limitation is due to the online CBBE scale's reliance on social media and reviews.

RQ4: This question is not clearly answered in this research. This is down to a number of factors:
1) This research gathers information that is used to measure CBBE indirectly, as opposed to people studies.
2) There are many external variables that are not accounted for, such as actual product quality, and design.
3) The significance level was set at a higher level than is normal.
4) Data and analysis needs improvement, by incorporating nation-wide datasets and performing analysis on groups of products that are in actual competition.

These factors limit the confidence that can be placed on the analysis results. This research showed that general trends may be identified, but the level of uncertainty in terms of the factors listed above, means that not much confidence should be placed in each individual metric. However, through future research, the last two factors may be erradicated. Then, web mining will have taken the first steps of becoming a valid alternative to surveys, when it comes to explaining sales figures by CBBE measurement.

In the end, this research showed that performing web mining in order to measure CBBE using online resources, is possible. Key challenges, as well as the merits of the approach compared to conventional methods, have been identified and discussed. The research was also able to provide satisfactory answers to three of the four research questions. As such, this research represents a successful first probe into this previously unexplored CBBE measurement approach.

References

1. Keller, K.L. (1993). Conceptualizing, measuring, and managing customer-based brand equity, Journal of Marketing; Jan 1993; 57, 1; ABI/INFORM Global

2. Aaker, D.A. (1996). Measuring Brand Equity Across Products and Markets; spring 1996; 38, 3; California management review

3. April, B. M., Pappu, R., Quester, P. G., Cooksey, R. W., & Pappu, R. (2005). Consumer-based brand equity : improving the measurement – empirical evidence, (April). http://doi.org/10.1108/10610420510601012

4. Bambauer-sachse, S., & Mangold, S. (2011). Journal of Retailing and Consumer Services Brand equity dilution through negative online word-of-mouth communication. Journal of Retailing and Consumer Services, 18(1), 38–45. http://doi.org/10.1016/j.jretconser.2010.09.003

5. Barnes, N. G. (2014). Social Commerce Emerges As Big Brands Position Themselves to Turn " Follows ", " Likes " and " Pins " into Sales, 14(4), 11–18.

6. Chen, H., De, P., Hu, Y. J., & Lafayette, W. (2011). IT-Enabled Broadcasting in Social Media : An Empirical Study of Artists ' Activities and Music Sales.

7. Christodoulides, G., & Furrer, O. (2006). Conceptualising and Measuring the Equity of Online Brands, (August). http://doi.org/10.1362/026725706778612149

8. Lassar, W., Mittal, B., & Sharma, A. (2012). Measuring customer-based brand equity.

9. Li, X., & Wu, L. (2012). Measuring Effects of Observational Learning and Social-Network Word-of-Mouth (WOM) on the Sales of Daily-Deal Vouchers.

10. Zhu, F., & Zhang, F. (2010). Impact of Online Consumer Reviews on Sales : The Moderating Role of Product and Consumer, 74(March), 133–148.

11. Friberg and Grönquist (2012) - Do expert reviews affect the demand for wine? Available online at https://www.jstor.org/stable/41419429?seq=1#page_scan_tab_contents

12. Dacin, Peter A. and Daniel C. Smith (1994). "The Effect of Brand Portfolio Characteristics on Consumer Evaluations of Brand Extensions," Journal of Marketing Research, 31 (May), 229-242.

13. Litman, Barry R. (1983), "Predicting the Success of Theatrical Movies: An Empirical Study," Journal of Popular Culture, 16(4), 159–75.

14. Mahajan, Vijay, Eitan Muller, and Roger A. Kerin (1984), "Introduction Strategy for New Products with Positive and Negative Word-of-Mouth," Management Science, 30 (12), 1389–1404.

15. Eliashberg, Jehoshua and Steven M. Shugan (1997), "Film Critics: Influencers or Predictors?" Journal of Marketing, 61 (April), 68–78.

16. Boatwright, Peter H., Wagner Kamakura, and Suman Basuroy (2007), "Reviewing the Reviewers: The Impact of Individual Film Critics on Box Office Performance," Quantitative Marketing and Economics, 5 (4), 401–425.

17. Chen, Pei-Yu, Shin-Yi Wu, and Jungsun Yoon (2004), "The Impact of Online Recommendations and Consumer Feedback on Sales," in Proceedings of the International Conference on Information Systems, ICIS 2004. Seattle: Association for Information Systems, 711–24.

18. Liu, Yong (2006), "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," Journal of Marketing, 70 (July), 74–89.

19. Duan, Wenjing, Bin Gu, and Andrew B. Whinston (2008), "Do Online Reviews Matter? An Empirical Investigation of Panel Data," Decision Support Systems, 45 (4), 1007–1016.

20. Lifewire.com, Top social networking sites people are using. Retrieved December 08, 2016 from https://www.lifewire.com/top-social-networking-sites-people-are-using-3486554

21. Coupons in the news. "Generic cola". Retrieved December 05, 2016, from http://couponsinthenews.com/wp-content/uploads/2014/03/Generic-cola-1024x625.jpg

22. What is Web Mining? - Definition from Techopedia. (n.d.). Retrieved November 04, 2016, from https://www.techopedia.com/definition/15634/web-mining definition

23. What is Multiple Regression? - Definition from Techopedia. (n.d.). Retrieved November 04, 2016, from https://www.techopedia.com/definition/27369/multiple-regression

24. What is brand equity? definition and meaning. (n.d.). Retrieved November 04, 2016, from http://www.businessdictionary.com/definition/brand-equity.html