INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN 2278 – 0211 (Online)

A Semantic Analysis of the User-Generated Content from Social Network Sites of Cashbuild Limited: Application of Text Mining, Machine Learning and Big Data Analytics toward the Development of a Customer-Centric Marketing Strategy

Dugmore Simosezwe Simphiwe Lushaba
Chartered Director, Corporate Governance Specialist,
Talent Growth Partners, South Africa

Abstract:
Social Network Sites (SNS) have grown rapidly over the past 15 years, which is in line with the growth of smartmobile phones. A key phenomenon of SNSs is User-Generated Content (UGC). It enables social engagements between users and their followers. UGC shares various perspectives on news, entertainment, and fun, as well as their perspectives on brands like Customer Experience (CX) and Customer Satisfaction (CS). UGC stimulates the Brand Attitudes (BA) of the readers if the user is trusted and credible. BAs influence the Purchase Intentions (PI) of the readers of UGC. Marketers like that UGC influences PI. They learn about CX and CS, which users share on the UGC. This semantic study uses Text Mining (TM) and Big Data Analysis (BDA) to extract keywords from the UGC of a building materials retail company, Cashbuild Limited (CBL) from the Republic of South Africa (RSA). The UGC was analysed using Power BI (PBI) to identify its sources and sentiments. The majority of the UGC in this study was obtained from Facebook (94.51%), followed by X.com (3.95%). The rest of the UGC was obtained from Blogs (0.79%) and Instagram (0.76%). 65.63% of the UGC was neutral, 29.74% positive, and 4.46% negative. UGC influences PI by mediating hedonic and utilitarian Brand Equity. The generation of a high-quality UGC was encouraged by CS - an outcome of CX. CX comes from the touchpoints across the CJ. Marketers want to ensure a good customer experience (CX) to achieve customer satisfaction (CS). Ultimately, CS influences future customer expectations (CE) when shared through UGC. 84 top keywords were extracted out of 16 328 hit sentences (HS) from the UGC of CBL using the MELTWATER program. The keywords were ranked to 51 frequency positions. CONCOR analysis was used to determine the connectivity and centrality of the keywords using Freeman's coefficient and the Eigenvector value for each keyword. The results of the CONCOR analysis were used to identify 41 significant keywords based on their higher connectivity and centrality. The 41 significant keywords were matched with the 16 328 HSs. 14 keywords that matched less than 100 HSs were excluded. The remaining 27 keywords were subjected to exploratory factor analysis EFA on Statistical Program for Social Sciences (SPSS). 21 keywords were loaded on 8 factors, and the model explained 68.363% of the variance with a Kaizer Meyer Olkin (KMO) of 0.736 above 0.6. The 8 categories were compared to input from a panel of industry experts. 7 categories proposed by experts matched 7 out of 8 categories from the statistical model. 7 key themes proposed to be used in the study to develop Customer-centric Marketing Strategies (CCMS) in the building materials retail industry (BMRI) are DIY Kings, House Proud, Project Master, Renovators, Interior Image, Building Materials Inventory, and Promotions. All 5 objectives of the study were met. The study adds value to the development of CCMS learning from the UGC.

Keywords: *Brand attitude, big data analytics, brand equity, business intelligence, building materials retail industry, cashbuild limited, customer-centric marketing strategy, customer expectations, customer experience, customer journey, customer orientation, customer perspective, customer satisfaction, social network sites, purchase intention, text mining and user-generated content*

1. Introduction

1.1. Background

The influence of user-generated content (UGC) on purchase intention (PI) is important to brands. The S-O-R Theory create an understanding of the influence, showing that the UGC creates a stimulus (S) with pictures, visuals, and text that entertains, delights, and pulls users (Kunja, Kumar, and Rao, 2021). Putra et al. (2021) posit that the impact of UGC on PI is influenced by brand equity (BE), and BE is impacted by customer experience (CX) and customer satisfaction

(CS). A large quantity of UGC is communicating CX and comments (Fu et al., 2022). This study investigated the semantics of the UGC in the building materials retail industry (BMRI).

Cashbuild Limited (CBL) is the largest BMRI in the Republic of South Africa (RSA), operating 320 stores in 6 countries (RSA, Lesotho, Botswana, Namibia, Eswatini, and Malawi) in southern Africa. It also operates an e-Commerce platform that ranks number 3 out of similar outlets on Google search for BMRI in RSA. It is the brand where the study was conducted. CBL is active on the web and has social network sites (SNSs), mainly on Facebook, Instagram, X.com (formerly called Twitter), and Blogs. CBL operates standardized shops. It does not segment its market by income levels, country, or any other criteria except for its VIP customers who qualify for discounts. CBL advertises extensively on SNSs, radio, TV, and print media. In terms of its promotions, CBL conducts promotional campaigns. It sponsors a popular program on TV (SABC) – called the Reno Race. It showcases house renovations using building materials from CBL. CBL also has an online shop where customers can place orders and get supplies from the nearest CBL stores.

CBL users between 1st of September 2022 and 31st of December 2023 posted 261 million impressions (i.e., count of instances that UGC was shown), 43 million reaches (i.e., the count of viewers who witnessed CBL posts), and 8.4 million impressions (i.e., count of viewers who appreciated, expressed, or responded) of UGC on the SNSs (Cashbuild, 2023). Publicly available UGC from the SNSs linked to CBL is mined and analyzed to identify semantics between the 01st of September 2022 and the 31st of December 2023. The study aims to gain valuable insights from CBL customers and strengthen its CCMS.

The topic is chosen based on the interest of the researcher in the analysis and application of UGC and how it develops a customer-centric marketing strategy (CCMS). A literature search did not identify any similar study from the BMRI globally. 16328 hit sentences (HSs) were extracted from the UGC. 84 top keywords were extracted using the MELTWATER program and reduced to 41 using the Frequency analysis, Freeman's coefficients, and the Eigenvector values. These determined the frequency, connectivity, and centrality of the top keywords. The 41 significant keywords were reduced to 27 after 14 failed to match 100 or more HSs. The study used exploratory factor analysis (EFA) to derive 21 keywords that loaded on 8 factors. 7 categories were developed in consultation with a panel of experts based on a review of the statistical model. All the objectives of the study were met, and it contributed to the body of knowledge and development of CCMS.

2. Review of the Literature

2.1. Literature Search

A search for journal articles 8 years or later was carried out on ProQuest and Google Scholar online libraries using the search criteria below:

- ("e-WOM OR UGC AND PURCHASE INTENTIONS"),
- ("UGC INFLUENCE PURCHASE INTENTIONS"),
- ("UGC IMPACT ON PURCHASE INTENTIONS"),
- ("SOCIAL MEDIA OR UGC AND PURCHASE INTENTIONS")
- ("SOCIAL MEDIA")
- ("SOCIAL NETWORK SITES")
- ("BUSINESS INTELLIGENCE AND POWER BI")
- ("CUSTOMER EXPERIENCE AND CUSTOMER SATISFACTION")
- ("CUSTOMER EXPECTATIONS AND CUSTOMER PERSPECTIVES")

80 articles were considered, as shown in table 1, and the selection criteria that were used to exclude or include articles led to 51 references that are included in the List of References.

Outcome	ProQuest	Google Scholar	Total
Identified	31	49	**80**
Duplicates	-	4	**4**
Excluded	10	15	**25**
Included	21	30	**51**

Table 1: Literature Search

Four duplicate articles were excluded. Twenty-five publications that contained the selection elements but did not study them were left out. Four other older articles that covered other supporting topics in this research, like statistical analysis and research philosophy, were included. The Cashbuild Integrated Report, a Cashbuild presentation on its social media, and one older article on the benefits of UGC were reviewed.

2.2. Social Networks Sites

Mobile communication is a developing area for marketers. The utilization of social network sites has grown fast throughout the last 15 years. The development of technology and the use of mobile units offer users easier access to information and ease of communication. Globally, many users use the top ten SNSs, including more than 90% of young adults (Borrero, 2023, citing Statista, 2022; Kaplan & Haenlein, 2010; and Pew Research Centre, 2019). SNSs provide a communication platform that impacts the marketing of brands and creates a new and deeper understanding of customer perspectives (CPs). Linked to the growing use of mobile devices is the growing use of smartphones.

Smartphones drove the use of SNSs, which changed consumer communication behaviour. The emergence of Web 2.0 coupled SNSs, sites where videos are shared like Instagram, TikTok, and YouTube, ushered in a new form of online communication (Yoo, Katsumata and Ichikohji, 2019 citing Lamberton and Stephen, 2016). Facebook is the biggest SNS, with a growth rate of 234% between 2009 and 2016 (Mayrhofer et al., 2020). A key advantage of SNSs is their ability to enable users to communicate directly. They improved the ease of communication and encouraged users to connect and disseminate information – a development that brought brands and their customers closer. There are many other SNSs over and above Facebook, as discussed above, and companies have caught on to this growing trend.

Companies use Twitter (renamed X.com) to communicate information to their customers. It transmits more than 500 million tweets daily and has 436 million users. Twitter data analytics help marketers develop CCMSs. It is more exposed to the daily life experiences of users than Facebook and YouTube. 78% of companies use Twitter, 74% use LinkedIn, and 44% use Facebook to communicate with users (Sanchez, 2023; citing Statista, 2022; Smith, Fischer & Yongjian, 2012; Go & You, 2016). According to Mishra (2022, citing Statista 2020 and Digital Marketing Institute, 2019). 40% of Twitter users confirmed making purchase decisions based on a tweet. In a study on banks from Nigeria, Stanley and Chinelo (2017) found that it was difficult to separate consumers from their mobile phones and recommended that banks combine their digital marketing with their mobile marketing. This trend is in line with the growing use of mobile phones and the growth of SNSs discussed above. This high utilization of SNS to transmit messages among users has improved its usefulness and relevance. This development attracted marketers and researchers to the huge amounts of data on SNSs.

This research pursues the value of SNSs to establish how data collected from customers can be used to develop CCMSs. Researchers mine huge amounts of data from SNSs to derive patterns and trends and infer insights, or they use SNA to inform CCMSs. Twitter data word count dominates this research approach followed by content analysis, sentiment analysis, machine learning (ML) and text analysis methods (Sanchez, 2023 citing Congosto, Basanta-Val & Sanchez Fernandes, 2017; Crisci et al., 2018; Liere-Netheler et al., 2019; Angelopoulos & Merali, 2017; Vidal et al., 2015; Alaparthi & Mishra, 2021; Chakraborty et al., 2020; Ibrahim, Wang & Bourne, 2017; Mostafa, 2019; Recuero-Virto, Valilla- Arrospide, 2021; Monero-Sandoval et al., 2018; Sings, Shuklab & Mishrac, 2018). SNSs, through these developments, grew and took centre stage in market research. The reviewed literature agrees that data from SNS became valuable to market researchers. It is based on the real-life experience of users. It is provided to communicate with their followers. It is not filtered and can be considered to provide users with feelings and experiences that are as authentic as possible. This semantic study uses a similarly authentic UGC that was posted on the SNSs of CBL to identify keywords and recommend themes to be considered in developing CCMSs.

2.3. Business Intelligence

This study analyses big data that was extracted from the SNSs. It is important to understand the sources of data for the study, such as which SNSs it was extracted from, which locations the data was posted from, and what sentiments were expressed in the UGC. One of the useful types of statistics for BI is descriptive statistics. It responds to questions that occur at the beginning of data analysis, like what is going on? (Sharma & Sarkar, 2022). This helps researchers understand the data, its sources, and what they can potentially do with it. Managers use BI tools to base their decisions (Goncalves, Goncalves, and Campante, 2023). BI creates the ability to control the data when it is better understood (Necochea-Chamorro, Larrea-Goyochea, 2023). Understanding the sources of data and the sentiments of messages from which it was extracted will help with its validity, repeatability, and comparability to other studies. To gain this understanding of the UGC from the SNSs, this study used Microsoft Power BI due to its wide use in big data analytics (BDA).

Power BI is widely used by businesses to monitor performance, visualize data, and support decision-making. Companies like Rolls Royce, Heathrow, Hewlett-Packard, Meijer, and Aston Martin use Power BI to analyze data and make decisions (Nabil et al., 2023, citing Iliashanko et al., 2019). Power BI can integrate data and create visuals for the information that improves decision–making (Gonclaves & Campante, 2023). The ability to support business decisions is a key advantage for Power BI, hence its early adoption. It was an appropriate tool for this study due to the advantages cited in the literature.

There are numerous advantages of Power BI cited in the literature. According to Necochea-Chamorro and Larrea-Goyochea (2023), it facilitates appropriate positioning in the workspace to identify challenges related to the rendering of services. It can be used to support strategic decisions that affect human talent, the needs of consumers, and their preferences, leading to improved customer experience (CX). Power BI leads to improved performance. Goncalves, Goncalves, and Campante (2023) seem to agree when they cited the ability of Power BI to create insights that improve performance. From raw data, it builds visuals and dashboards that improve understanding and support innovation. This study uses Power BI to analyze data sources by SNSs and geographic locations and sentiments of UGC.

2.4. User-Generated Content

UGC is based on the customer experience (CX) of a brand. It is communicated through testimonials, tweets, text messages, images, blogs, sound, and visuals. It is any form of communication that is generated by users online. There is a strong relationship between phone-based content behaviour and content usage behaviour (Yoo, Katsumata & Ichkohji, 2019). The increased use of smartphones to generate and share content, as discussed above, facilitated an increase in the usage of content. This relationship is beneficial to marketers when the content relates to brands.

It is on SNSs that users share their CX through UGC. UGC is at the centre of the exploding SNS phenomenon (dos Santos, M.L.B., 2021; Halliday, 2016, citing Chaffey, 2008). Due to the widespread availability of information, users became well-informed. They can seek information by themselves, create content, and share it (Polanco-Diges & Debasa, 2020).

Users became powerful publishers that influence the PI of their followers on SNSs. This has become a significant development for customer-centric marketing strategies. It brought customers to the centre of the marketing of brands. Marketers use UGC to learn about customer sentiments, preferences, and experiences.

UGC is shared freely and willingly amongst associates. It is a covert marketing tool that blends into editorial social media content. UGC is cheaper yet more effective for brands. It is recommended by Adweek and MIT Sloan Management Review (Mayrhofer et al., 2020). Brands become part of conversations between friends and associates. Users use UGC for engaging friends, online self-image, collecting and publishing knowledge, social engagements, communicating their views, and sharing CX, impressions, fun, and entertainment (Sanchez, 2023; Mayrhofer et al., 2020; Mohammad et al., 2020, citing Bakshy et al., 2012; Kim & Lee, 2017; Shim & Lee,2009; Daugherty et al., 2008; Blackshaw & Nazzaro, 2006; and Boyd and Ellison, 2008). UGC is a form of electronic word-of-mouth (eWOM). Its content is developed by users on SNS (Tariyal et al., 2022). It is beneficial to the brands that readers learn about whilst consuming information on matters of interest. Its ability to feature brands in the social engagements of friends is a smart approach that pushed it to prominence with marketers. It is a sublime tool that positions brands in conversations in an unthreatening manner.

Brands encourage UGC among their customers. They use strategies like real-world tie-ins and contests to encourage users to generate UGC that benefits the brand. Alcohol brands use launch events series, sports events, hashtags, photos, and video competitions where users link themselves to the brand image and share this with their networks. Coca-Cola printed cans and bottles with popular names of people and a prompt for customers to share their unique pictures with the hashtag ShareACoke. UGC is affordable covert advertising for brands (Mayrhofer et al., 2020, citing Nicholls, 2012; Pinsky et al., 2017; and Lobstein et al., 2016). UGC is a good marketing avenue for advertising brand images through credible users. It is evident from the literature that brands are developing innovative strategies to influence and encourage beneficial UGC. It is affordable and effective for brands compared to traditional advertising forms.

Brands do not pay for the advertising benefits of UGC. Companies share product attributes, advertise, publish knowledge, and CS, whereas users post content for many other reasons, like personal identity, acceptance and integration into their networks, and conversations with their friends (Mayrhofer et al., 2020). This ability of UGC to transcend advertising and integrate other objectives of social interaction is valuable to brands. Brands, however, may not always benefit from UGC.

Brands do not have control over what the user states in UGC. It may be negative toward the brand. To counter this, brands pay celebrities who are popular on SNSs (called Influencers) to share their CX (Karunanayake & Madubashini, 2019; Nuseirat et al., 2021 citing Ngah et al., 2021). Despite the risks that are cited in the literature, UGC offers merits that brands cannot afford to ignore. It weaves itself into influential engagements that brands cannot afford to miss. Therefore, keeping a brand away from UGC is not an option for marketers. Caring about friends is at the heart of UGC.

Users generate UGC because they care about others, a phenomenon linked to CO that influences the quality of UGC. Out of the 5 motivations for users to create content, researchers found that the value function motivated the creation of UGC. Utilitarian and social functions secure customer-oriented users that produce high-quality and -quantity content (Yoo, Katsumata & Ichikohji, 2019). When users acquire an experience of a product or service, they become keen to share such CX with others that they care about. It is the arousal of interest for users to share information about brands that marketers seek to influence. The creation of high-quality UGC-based CX facilitates the sharing of UGC and the fulfilment of their CO. This attracts marketers to the value of UGC.

UGC attracts marketers and market researchers. Users post information like their CX and CS, from which marketers can extract and derive insights. UGC is a discussion between associates and friends; unlike research questionnaires, it contains unfiltered input from users for research purposes. For retailers, UGC improves understanding of product attributes as seen by customers, service characteristics that impress or concern their customers, and consumer behaviour (Kitsios et al., 2021). The usefulness of UGC is in its ability to deeply connect brands with customers. It is authentic and original and more useful in providing insight into the brands, helping marketers formulate effective strategies.

This study explores the potential to learn strategic insights from UGC in the BMRI. Such insights can be used by management, along with its experience, to develop CCMSs. UGC and data analytics help to develop CCMSs (Sanchez, 2023; Yoo, Katsumata & Ichikohji, 2019). There are five (5) benefits that brands enjoy from using UGC and the analysis of data. They gain customer insights, improve loyalty, build awareness, grow sales, and reduce costs (Shea, 2008). Literature shows that UGC is more beneficial than detrimental to brands. There are many other reasons cited in the literature on how brands benefit from UGC.

Some authors present monetary reasons (Utilitarian), knowledge, values like expressing oneself, self-image or belonging, and social interaction as reasons for generating UGC. Researchers also use UGC to understand customer needs (Artem & Hauser, 2019), to identify brand positioning, to understand the competitive landscapes of hotel chains in China, and to develop brand positioning strategies (Hu & Trivedi, 2020). Companies learn from UGC reasons why customers support their products (Shridhar, 2023). This helps marketers to position their brands competitively. It is a source of valuable information for brands to gain customer insight and develop CCMSs. This study makes use of this benefit from UGC to contribute toward the development of CCMSs. There are other forms of UGC over and above the text that can also be used.

UGC is not only shared through text messages. Posting of visuals and images is also referred to as VSMM (Al-Gasawneh, et al., 2023, citing Gretzel, 2017). Connection to images of the brand also motivates users to include these in their content for personal identity (Mayrhofer et al., 2020). Readers who associate with the user who posts images may be attracted to the brand. This association of the brand image with the user that the reader associates with benefits the brand. The social standing of the user benefits from association with the brand image, whereas the brand gains influence on the followers. Such mutual benefits created the attraction of UGC with marketers. The users share their knowledge of the brand with their networks. UGC is a valuable source of influence.

By influencing the knowledge shared on UGC, marketers get more value out of the users. Literature highlights two (2) types of knowledge that users share. These are (1) familiarity - the exposure that the user had to the product, and (2) expertise knowledge - their ability to use the product (Yoo, Katsumata & Ichkohji, 2019; Kaosiri et al., 2019 citing Nezakati et al., 2015). Due to the value of knowledge to potential customers, they develop an interest in the product. UGC educates potential customers, which is a powerful advantage for brands and marketers. The potential impact of UGC and its shaping of brand attitude (BA) responses on the PI rely on knowledge that makes it important to marketers.

2.5. Customer Purchase Intentions

Research published by Mohammad et al. (2020) demonstrated a strong link between the strength of the UGC and functional and emotional brand attitudes. They argued that simple UGC that is easy to understand, popular, offers technological quality, and is what users want to hear will attract and enhance their hedonic and utilitarian attitudes, resulting in their improved interest and customer expectations (CE) at the level of thinking and feelings about the brand. Putra et al. (2021) found that brand equity (BE) influenced the impact of UGC on PI, and it was influenced by customer experience (CX) and customer satisfaction (CS). For brands to gain insight from large UGC data on their SNSs, they mine the text that contains such UGC. It gives them access to customer perspectives (CPs) that help sharpen their marketing effort.

Quality UGC, like pictures, visuals, or text that is playful or believed, invites the users. It encourages brand attitudes to have a functional or emotional impact on the recipients. According to the Theory of Planned Behaviour, values impact conduct or direct it. Behaviour results from specific attitudes (Kunja, Kumar & Rao, 2021). The stature of the source impacts purchase intention (PI) through values in relation to the UGC (Muda & Hamzah, 2021). Yu-Jin (2019, citing Stalcka's Consumer Content Report, 2018) agreed when he found that as high as 87% of customers consent that UGC influences their purchase decisions. The research was done in the beauty industry (Nusairat et al., 2021), property industry (Al-Gasawneh et al., 2023), smart mobile phones, store purchases, and e-products (Kunja, Kumar & Rao, 2021 citing To and Ho, 2014; Erkan & Evans, 2016; Kudesha & Kumar, 2017; Wang et al., 2018), like environmentally friendly products and clothing all established that UGC strongly impacted PI. In the intoxicating beverages sector, the link between the application of SNSs and the utilization of such beverages was high at 93.10% (Alhabash et al., 2022). Davcik et al. (2021) found an affirmative impact of UGC on Instagram users' aim to participate in the SNS. The impact of UGC on PI is well-documented in the literature. It is a relationship that hugely benefits marketers.

UGC shapes purchase intentions, leading to purchase decisions. PI is a conduct that leads to purchase decisions. PI is the likelihood of a customer who intends to buy a product executing their intention (Long & Nga, 2020). Customers use UGC to make purchase decisions. In China, for instance, studies have found that 77% of users on the internet search for information on products before buying them. They want to see what others share as their customer experience with the product or service (Wang et al., 2023). This is important to marketers in their pursuit of customer purchase decisions. UGC, therefore, attracts interest from marketers by influencing PI. This relationship draws from consumer psychology, and the theories and models that explain the relationships between UGC and PI are discussed below.

2.6. S-O-R Theory

Understanding the process of how UGC influences PI is crucial for marketers. The S-O-R Theory was developed by Mchrabian and Russel in 1974 (Mohammad et al., 2020). It creates a link between the lived experience and the value response of people. It derives from the stimulus (S) from the lived experience, which influences functions like thinking and feeling awareness, made from the organism (O) stage. It impacts either practical BAs or emotional BAs between the stimulus and the customers' response (R) (Mohammad et al., 2020, citing Fiore & Kim, 2007; Kunja, Kumar & Rao, 2021). This impact on recipients of UGC is crucial to influence behaviour. It has the potential to shape attitudes that may benefit the brand.

Brands want customers to be positively disposed toward their products. Affirmative behaviour encourages the customer's buying intentions (Muda, M. & Hamzah, MI, 2021, citing Ajzen, 1991; Mosavi & Ghaedi, 2012). The positive influence of UGC on the behavioural intentions of the readers improves PIs. Its permeation of social engagements without antagonizing the readers attracts marketers to UGC. It is more accepted than other forms of advertising. UGC does not come across as an advertisement. This mitigates the resistance from readers.

2.7. The Persuasion Knowledge Model (PKM)

Customers resist advertising communication. In terms of PKM customers that establish that the UGC is an advert, their defence is alerted, and they resist the message (Mayrhofer, 2020, citing van Reijmersdal et al., 2016; De Jans, Cauberghe and Hudders, 2018; Evans et al., 2017; Friedstad & Wright, 1994). Such resistance works against the successful marketing of brands. UGC does not trigger negative affect (PKM). It leads to higher PI than advertising and producer-generated content (PGC) (Mayrhofer et al., 2020). Trust influences the relationship between the users and their followers. UGC is a powerful tool for marketing brands due to its high levels of acceptance by recipients, even when it is brand-oriented. Its acceptance is also influenced by other social factors.

Culture affects engagements on SNS. Strong and enduring relations, like guanxi in China, impact PI (Bilal et al., 2021, citing Shaalah et al., 2013; Dimoka et al., 2012; Samaha et al., 2014). They are established on the strong trust and credibility of those who relate. Such trust and credibility also apply to UGC; hence, it is less affected by PKM. Users can share their customer experience and customer satisfaction without resistance from the readers.

2.8. UGC, Customer Satisfaction and Customer Experience

Users share their customer experience and customer satisfaction with their followers on UGC. UGC influences the purchase intention of the followers and shapes their expectations. It influences the customer satisfaction by shaping their expectations. Information about the experience of others is valued by customers when they purchase in tourism (Kaosiri et al., 2019), the automotive industry (Karunanayake & Madubashini, 2019), and hotels (Fu et al., 2022). This is in line with the findings of a study by Alwan and Alshurideh (2022, citing Wijaya et al., 2020) that digital marketing positively impacts both value creation and customer satisfaction. Customers, therefore, find reference points from what others experienced from the brand on UGC. This experience of others, if assuring, leads to higher customer expectations.

Marketers are interested in what influences customer satisfaction; in addition to developing customer loyalty, it can also influence expectations and, subsequently, the customer satisfaction of others. In a study by Bakri, Zamli, and Azman (2012) to identify factors that influence customer satisfaction, they found that customer satisfaction is influenced by reliability (ability to meet promise), tangibility (design information, given), assurance (ability to communicate trust and confidence), responsiveness (ability to respond to questions quickly), and empathy (ability to be approachable). Assurance was found to be the most influential factor in determining customer satisfaction. When marketers crawl through UGC, they gain a better understanding of these factors. It is the ultimate fulfilment that brands wish to achieve toward customer loyalty.

Customer satisfaction is a crucial stage in a customer journey. To achieve this, brands should try to meet the expectations of their customers (Nirmalasari et al., 2022). Brands are, therefore, keen to learn about the experiences that shape customer expectations. Customer satisfaction is a valuable platform that stimulates the generation of UGC and future customer satisfaction. Since UGC can communicate these customer experiences, it plays a significant role in shaping customer satisfaction. UGC is a repository for valuable insights into customer preferences and triggers for their sought-after customer satisfaction.

The customer experience determines customer satisfaction. Superior customer experience is very important for success. Its nature is multidimensional and includes cognitive, affective, social (behavioural), and physical (sensorial) responses to the retailer. Marketers want to find these responses at different points of the customer journey. Customer experience involves every touchpoint, from initial browsing of the brand on the internet, to post-purchase interaction (Susiang et al., 2023). A strong customer experience is a competitive differentiator for brands. Customer experience gives brands a competitive edge beyond product quality (Verhoef, 2021, citing Lemon and Verhoef, 2016). The literature agrees that brands that deliver exceptional customer experience attract high levels of customer trust. They enjoy higher levels of repeat business in a highly competitive Indonesian e-commerce market (Susiang et al., 2023). Improving customer experience, therefore, needs multiple responses, including messages that shape customer expectations to develop an interest in the brand. Customer sentiments developed from customer experience influence association with the brand, stimulate interest to generate UGC, and make purchase decisions (Mishra, 2022). Customer experience in retail plays a crucial role in the generation of UGC. It influences purchase intentions. The nature and form of UGC are, therefore, crucial to marketers.

Companies encourage high-quality UGC to increase customer expectations and create PIs. Marketing strategies employed by companies try to grow the customer experience. The UGC campaign, also known as viral strategy or buzz marketing, when compared to affiliate marketing, search engine (SE) campaigns, SNS marketing, and corporate blogs, was found to be the strategy that enhanced customer experience the most. It was able to maximize customer experience (Nuseir et al., 2023). UGC interacts with customer satisfaction and customer experience and benefits the brands. Brands must understand UGC that interests their customers.

Brands encourage UGC to influence the purchase intentions. Satisfaction is an outcome of cognitive and affective processes (Kaosiri et al., 2019, citing Castaneda, Frias, and Rodriguez, 2007; Oliver, 1993; Tao and Kim, 2019, citing Bigne, Andreu & Gnoth, 2005). It derives from both utilitarian (utility) and hedonic (feeling) BAs of customers. In passenger cruise shipping, both CX and CS are known to influence repeat purchases, customer loyalty, and favourable UGC (Tao & Kim, 2019; Rane, Achari & Choudhary, 2023). Brands, therefore, benefit from UGC influencing these attitudes when communicated by credible and trusted users.

2.9. Trust and Credibility of UGC

Trustworthiness and believability of the user impact on acceptance of UGC. According to Yu-Jin (2019, citing Bright Local 2018), when users crawl the internet, what they engage with is determined by trust. He states that 85% of customers trust online messages as much as they trust recommendations from friends and family, as agreed by Kaosiri et al. (2019). Customers believe in UGC more than they believe in producer-generated content for users (Nosita, Lestari, 2019). UGC videos are more believed by customers than PGC videos (Israfilzade & Baghirova, 2022). A study by Zinko et al. (2020) in travel agrees when it argues that UGC that contains pictures is liked and thought to provide higher quality content than UGC of messages only. Marketers should think carefully when developing strategies that they employ when encouraging UGC.

Influencers are also liked and believed by their followers. This creates quasi-social links (Kanwar & Huang, 2022). Huang and Copeland (2020, citing Rogers and Bhowmik, 1970), mentioned that stature on SNS is derived from the believability and know-how possessed by the user. Kunja, Kumar and Rao (2021, citing Chu & Choi, 2011; Wallace et al., 2009) agree that the UGC is the most referenced source by customers. Trust and credibility of users improve the reliability of UGC. Marketers should consider the opposite and unintended outcomes of UGC.

There is a limit to how much UGC can be used to persuade customers. Studies show that the quality of the sources, whether they are liked or not, and their credibility influence the acceptance of their persuasive advertisement. If the user lacks credibility or is not liked, readers may choose to become critics of their UGC (Mayhofer et al., 2020, citing Ertimur & Gilly, 2012; Steyn et al., 2011; Thompson & Malaviya, 2013). Criticism reduces interest in the brand associated with the UGC. The trustworthiness and credibility of the source of UGC is crucial for its success. When UGC is credible and trusted, the purchase intention is impacted. Methods that researchers use to access and analyze large quantities of UGC data to gain insights are fast developing. We consider some of these methods below.

2.10. Big Data Analytics and Text Mining

UGC occurs in large amounts of data posted on digital platforms like websites, social media, and blogs. In this form, the data is referred to as big data. Such data can be text, videos, and images. It is produced on the internet through clickstream, mobile communication, user-generated content, and social network sites, and UGC is deliberately sourced through sensor networks, business transactions, and other functional platforms like biological informatics. Big data largely increased the volume, speed, and diversity of data for brands. It provides feelings, attitudes, and views of users that are based on their customer experience. Big data is the UGC coupled with the individual users' footprints and their behaviours (Tao & Kim, 2019, citing Davenport, Barth & Bean, 2012; George, Haas & Pentland, 2014). This boom of data provides a wealth of information to brands. It provides information about consumers, such as their experiences, preferences, and attitudes.

The rapid growth of UGC volume and TM has improved qualitative and quantitative research methods (Mastrogiacomo et al., 2021). It provides opportunities for statistical inference and large analysis that reveal valuable insights (Tao and Kim, 2019, citing Talon-Bellestero, Gonzales-Serrano, Soguero-Ruiz, Munoz-Rumero, Rojo-Alvarez, 2018). It ushered in an era of big data analytics (BDA), which enabled a deeper understanding of customer sentiments. It starts with large but low-quality data. Raw UGC is unstructured and massive with low density. It needs techniques to extract insights, knowledge, and meaning.

This study derives its approach from text-mining semantic network analysis with visual analytics and exploratory factor analysis (EFA). There are similarities to methodologies used in sentiment analysis like obtaining information that is relevant to a specific topic, pre-processing such information like tokenizing such data into keywords, and obtaining relevant information from them like their connectivity to other words in large amounts of UGC, and identify sentiment from such information relating to the brand (Tao and Kim 2019 citing Schmunk, Hopken, Fuchs and Lexhagen, 2013). TM enables researchers to access large amounts of data to gain insight. It is a technique that unlocks the information potential of big data and grows the quantity and quality of knowledge that brands acquire from big data like UGC.

There are different text mining methods and techniques that are used. Text mining is the approach of obtaining good trends from unshaped data. It fundamentally arranges the UGC through tools like text parsing, vectorization, linguistic feature extraction, topic modelling, and sentiment analysis drawn from machine learning, statistics, and computational linguistics. Typical methods include algorithms and models. Methods used in literature include statistical representation of items like words and sentences to machine learning algorithms and deep learning models that vectorize UGC, derive important embedded features, and establish UGC understanding functions like similarity recognition, semantic analysis, and opinion mining (Li et al., 2022). A researcher must choose an analysis method that suits their study. The approach should, therefore, be well considered, from text mining methods to data analysis methods.

UGC analyses of topics, keywords, sentiments, and opinions are enabled by data mining and other applications. To gain access to insights, researchers mine the UGC on SNSs and other digital platforms. Semantic and sentiment TM analysis are techniques that help researchers uncover meaningful information (Li et al., 2022). UGC helps brands understand consumers and improve their business intelligence. There is a growing interest in the business in getting high-quality UGC, accessing it and extracting useful insights. This is giving rise to a deeper understanding of customer perspectives.

The broad process followed by researchers involves data input, text mining, and the use of business applications for the analysis of data. Data input is market data shared by the users on social network sites, user reviews, brand or news websites, and product-related UGC (Li et al., 2022). This study broadly follows a similar process to access the UGC. It extracts keywords from big data and analyses these to influence strategic thinking and hopefully inspire customer-centric marketing strategies. For this study, we conduct semantic analysis since it best enables the analysis of keywords from the UGC, including their frequency, connectivity, and centrality.

Semantic analysis involves evaluating and representing text and analyzing its meaning and interpretations. To understand text semantics, a more complex semantic analysis method, such as text representation using vectors in high-dimensional space, is applied. Semantic analysis methods are used to carry out quantitative text analysis and enable a strong methodical and theoretical basis to outline their semantic structure. There are also statistics-based methods like vectorizing text by counting the frequency of words used in this study. The vector space of the text can be determined by using the n-gram model. The significance of keywords may be determined using word importance and frequency. Word frequency and vector analysis methods will be used in this study, as discussed in the methodology section below (Li et al., 2022). The frequency of occurrence of words indicates their importance in the text. The connectivity and centrality of such frequently occurring keywords demonstrate the structure of the semantic network (Tao and Kim, 2019, citing Carley, 1997; Popping, 2000; and Xiang, Gretzel, and Fesenmaier, 2009). Semantic analysis creates a deeper understanding of the semantic structure of UGC. While this study applies semantic analysis, other methods, such as sentiment analysis, can be used in UGC analysis.

Although sentiment analysis is not used in this study, it is worth mentioning briefly. Sentiment analysis methods include neural-network-based methods, transformer-based models, conditional random field models, frequency-based feature extraction, and rule-based feature extraction that are usually applied. Lexicon-based methods and machine-learning techniques are key to sentiment analysis (Li et al., 2022). As discussed above, some of the methods used in this study, like the pre-processing of data and its tokenization, are similar to methods used in sentiment analysis.

3. Methodology

3.1. Research Method

A UGC analysis was carried out using a combination of qualitative and quantitative data analysis techniques. This study used big data analytics like semantic analysis studies in casino hotels and cruise shipping industries (Fu et al., 2022; Tao & Kim, 2019). Figure 1 shows the research approach that was followed and how UGC was collected, processed and analysed in this study.

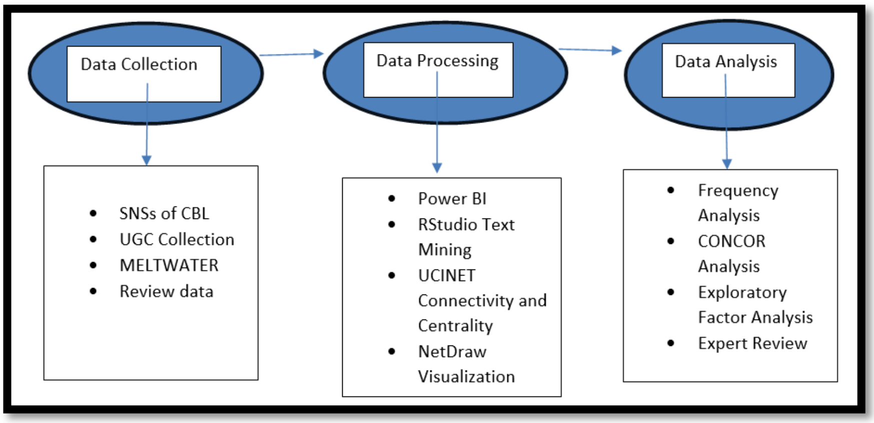

Figure 1: Data Collection, Data Processing and Data Analysis (Fu Et Al. 2022)

3.2. Data Collection

Secondary data was collected from the social network sites (SNSs) (Saura & Bennett, 2019). The period of UGC was from the 01st of September 2022 to the 31st of December 2023. This UGC was extracted from the above sources using a social media crawler written in the MELTWATER program. The following search was carried out on the MELTWATER program:

- Title of the search: Cashbuild/Social Search>Cashbuild Keywords Michael
- Filter Set: Source Type (10):
- Included> ("Facebook" AND "Instagram" AND "X.com" AND "Reddit" AND "Pinterest" AND "YouTube" AND "TikTok" AND "LinkedIn" AND "Blogs" AND "Twitch")
- Not included> "NOT (handle "Cashbuild" OR source name: "Cashbuild" from "Cashbuild")
- Search queries> (Cashbuild OR @Cashbuild OR hashtag: Cashbuild) AND ("service" OR "deliveries" OR "stock" OR "cement" OR "Afrisam cement" OR "PPC cement" "bricks" OR "timber" OR "geysers" OR "roof" OR "glass" OR "glass-cutting" OR "tiles" OR "floor tiles" OR "paving" OR "plumbing" OR taps" OR "lashers" OR" stores" OR "store locations" OR "deliveries" OR "product quality" OR "product prices" OR "renovations" OR "RenoRace" OR "Kitchen" OR "kitchens" OR "bathroom" OR "bathrooms" OR "ceiling" OR "doors" OR "door frames" OR "patty" OR "windows" OR "window frames" OR "roof" OR "roofs" OR "roofing" OR "roof trusses" OR "nails" OR "screws" OR "electrical" OR "plugs" OR "switches" OR "security cameras" OR "patio" OR "pool" OR "hobb" OR "building" OR "tubing" OR "curtain rails" OR "fireplace" OR "chimney" OR "sink" OR "bath" OR "grout" OR "stove" OR "oven" OR "cabinets"
- Location: Include> ("Botswana" AND "Eswatini" AND "Lesotho" AND "Malawi" AND "Namibia" AND "South Africa"
- Excluded searches were producer-generated content of Cashbuild Limited (CBL). The search queries are based on industry knowledge derived from boardroom discussions, board meeting packs, and general industry knowledge. The six locations were selected because Cashbuild Limited operates stores in those countries.

3.3. Data Processing

Data pre-processing was conducted using morphological analysis. Useless information in unstructured data was removed during pre-processing, which improved the accuracy of sentiment feature extraction by properly representing the key features. Invalid information like stop words and symbols was removed. Stop words like "however," "is," and "are" were removed as they do not indicate specific semantics (Li et al., 2022). Duplicate content was removed.

<u>3.3.1. Determining Data Sources and Sentiment</u>

The .csv files from MELTWATER with source data for UGC contained information about social network sites, source data like IP addresses, countries, provinces, and cities where posts were sent from, hit sentences, keywords, etc. The data was loaded on Power BI for processing and visualisation. Power BI analysis was conducted to determine the sources of UGC and the sentiments.

<u>3.3.2. Identification of the Keywords</u>

16328 hit sentences and 84 keywords were extracted from the UGC on the social network sites of Cashbuild Limited using the MELTWATER program.

3.3.3. Determining the Frequency of Occurrence of Keywords

Extracted keywords were grouped around categories using the 5 Ps and service (Saura & Bennett, 2019). RStudio program and Ucinet 6.0 were packaged with the visualisation tool NetDraw, which was used in data processing. The unstructured data was developed into structured data. The frequency of 84 keywords was determined and ranked with the occurring word as number 1. 84 top keywords were retained for further analysis. The NetDraw program in UCINET 6.0 was utilised to display patterns of the keywords (Fu et al., 2022).

3.4. Data Analysis

3.4.1. Determining Connectivity and Centrality of Keywords

The CONCOR Analysis was used to establish and show the connectivity and centrality of top keywords. The connectivity of central keywords represents the linkage and importance of keyword/s that make up a node for the cluster. Freeman's degree of centrality is calculated by the connectivity obtained by a node. The words that had the highest connectivity were the most central. The eigenvector values were computed to determine the most influential nodes (Fu et al., 2022).

Measures and patterns among connecting words were discovered. Connecting words were determined as 41 significant words by considering their ranking on Eigenvector and Freeman's degree of centrality. The meaning of the words, relevance, and duplications were also considered. In big data analytics, systematic data reduction is important (Tao & Kim, 2019, citing Boyd and Grawford, 2012). Significant keywords that had similar characteristics were grouped. Significant keywords were used to determine correlation coefficients and group similar words into clusters.

3.4.2. Determining Categories of Keywords

Data was analysed using SPSS. 41 significant keywords identified above were matched to the 16328 hit sentences. 27 words that matched 100 hit sentences and more were central to the UGC. Exploratory factor analysis (EFA) was chosen over the Principal Components Analysis because components are created by variables, and factors create variables (Pallant, 2016 citing Tabachnick and Fidell 2001), so EFA is a better analysis to consider.

27 words were processed using EFA to determine factors that create variables in the data (Pallant, 2016, citing Tabachnick & Fidell, 2001). The EFA was conducted to determine factor loads for the keywords. EFA is a data reduction technique. It takes a large set of data and reduces it to smaller factors by identifying clusters among intercorrelations of related data (Pallant, 2016). 21 keywords loaded in 8 factors. Finally, a panel of experts was convened consisting of management and a marketing expert to independently provide input on 8 factors and keywords based on their industry and business insights.

3.5. Conceptual Framework

Figure 2 shows a conceptual framework that demonstrates how UGC influences the PI, leading to purchase decisions.

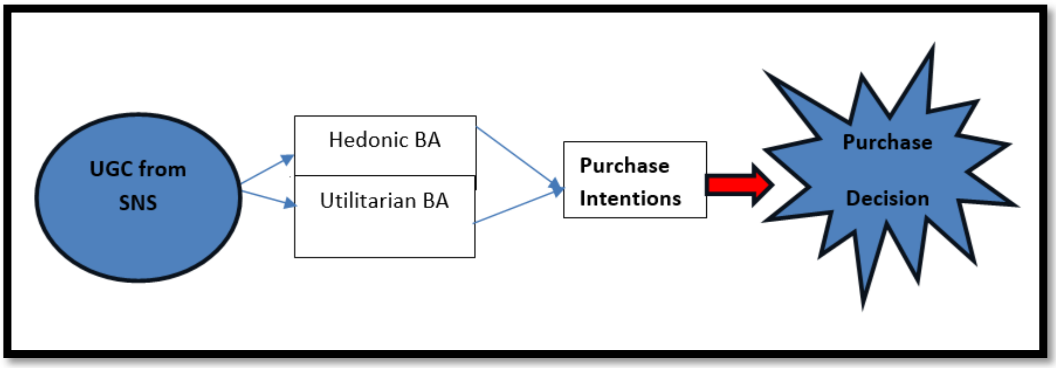

Figure 2: Conceptual Model

Figure 2 above shows the UGC from SNSs stimulating (S) the hedonic and utilitarian BAs. The organism (O) in the BAs influences a response (R), which is the purchase intention (PI). The PI leads to purchase decisions.

3.6. Ethical Considerations

Accurate recognition and referencing of authors were important in this research. The analysis was carried out objectively. Permission to access UGC was obtained from CBL through the Chairman of the Board of Directors, the CEO, and the Commercial Director. There was no infringement of privacy for the users. Ethics challenges in UGC analysis include the use of UGC to segregate against consumers by practising big data segregation pricing in violation of consumer rights and interests (Li et al., 2022). This study does not segregate customers by type, country, sex, race, or any other segmentation, and therefore, results cannot be used to discriminate against any customers.

4. Results

The results provide the top keywords and the categories based on their connectivity and centrality in the UGC.

4.1. Understanding the Social Network Sites

Figure 3 shows the results of the social network sites that the UGC was extracted from.

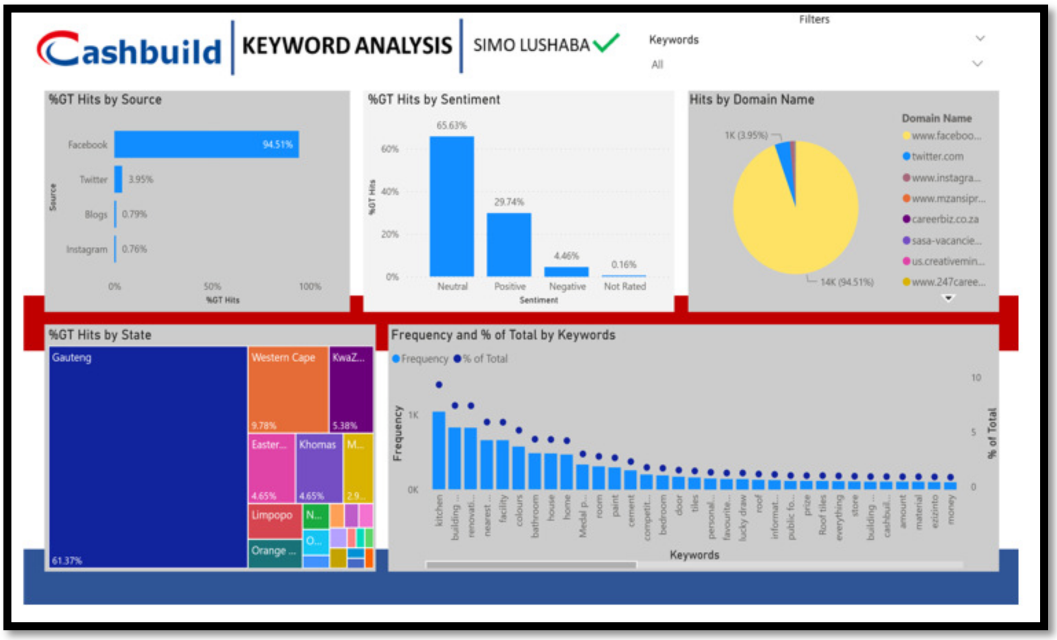

Figure 3: Hits by Sources and Sentiments

Figure 4 shows locations where the UGC was posted.

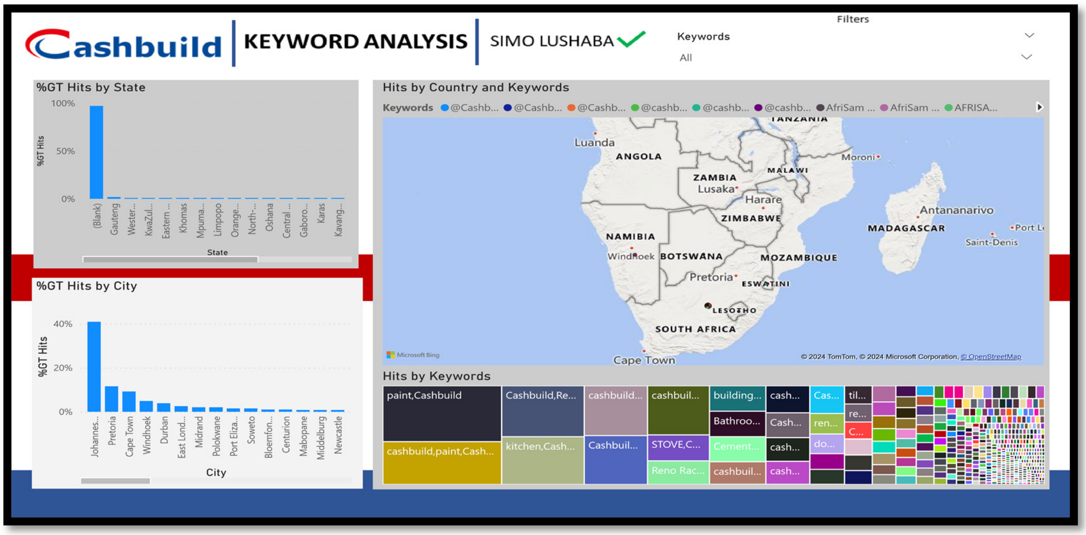

Figure 4: Hits by Location

4.2. Identification and Frequency of Top Keywords

Table 2 presents 84 of the most frequently used keywords in the UGC. The keywords are ranked using the frequency of the highest occurring keywords. “Kitchen” in the UGC was ranked first on the list.

Keyword	Frequency	Category	Percentage	Rank	Keyword	Frequency	Category	Percentage	Rank
Kitchen	1038	Product	9,32866	1	Cashbuild stock	54	Product	0,485306012	34
Building materials	826	Product	7,4233486	2	Wide range	52	Product	0,467331716	35
Renovations	824	Service	7,4054103	3	Area of responsibility	41	Service	0,368473083	36
Nearest Cashbuild store	658	Place	5,9135436	4	Good condition	39	Service	0,350498787	37
Facility	656	Service	5,8955693	5	Bank	31	Service	0,2786016	38
Colours	574	Product	5,1586232	6	Cheaper quotation	31	Service	0,2786016	38
Bathroom	483	Product	4,3407927	7	Interests	31	Service	0,2786016	38
House	480	Product	4,3138312	8	Building materials equivalent	25	Product	0,22468709	39
Home	468	Product	4,2059854	9	Small small sigo dha gwana	25	Service	0,22468709	39
Medal paint	333	Product	2,9927204	10	Tokala to tulako	25	Service	0,22468709	39
Room	308	Product	2,7680417	11	Tools and accessories	24	Products	0,215691561	40
Paint	294	Product	2,6422216	12	Bleak picture	20	Service	0,179742968	41
Cement	255	Product	2,2917228	13	Front door	19	Product	0,170755819	42
Competition	196	Promotion	1,77614811	14	Transactions	19	Service	0,170755819	42
Bedroom	188	Product	1,68955839	15	SA's biggest building materials	18	Product	0,161768671	43
Door	168	Product	1,5098409	16	Bakkie builders	18	People	0,161768671	43
Tiles	159	Product	1,4289566	17	Ceiling	18	Product	0,161768671	43
Personal Information	146	Service	1,3121237	18	Cement purchase	18	Product	0,161768671	43
Favourite room	139	Product	1.2492136	19	Economy at ground level	18	Service	0,161768671	43
Lucky draw	139	Promotion	1,2492136	19	Home renovations and building work	18	Service	0,161768671	43
Roof	128	Product	1.150355	20	Low-income earners	18	Service	0,161768671	43
Information	124	Service	1,1144064	21	Paint sponsor	17	Service	0,152781522	44
Public forum	113	Service	1,10155478	22	Tools	17	Products	0,152781522	44
Roof tiles	112	Product	1,0065606	23	Accordance	15	Service	0,134807226	45
Prize	112	Promotion	1,0065606	23	Activation	15	Promotion	0,134807226	45
Everything	109	Service	0,9795992	24	Ages	15	People	0,134807226	45
Store	106	Place	0,95263377	25	Bags	15	Products	0,134807226	45
Building project	102	Product	0,9166891	26	Second winner	15	Promotion	0,134807226	45
Cashbuild voucher	102	Promotion	0,9166891	26	Responsibility	13	Service	0,116832929	46
Amount	101	Service	0,907702	27	Stock	13	Product	0,116832929	46
Material	101	Product	0,907702	27	Zinc Roof	13	Product	0,116832929	46
ezizinto	99	Product	0,8997277	28	Available in 101 locations	12	Place	0,017845781	47
Money	96	Service	0,8627622	29	Building or DIY project	12	Service	0,017845781	47
Card	94	Service	0,8447919	30	Holidays	12	Service	0,017845781	47
Acrylic roof paint	77	Product	0,6920104	31	Product	12	Product	0,017845781	47
Building and plumbing materials	65	Product	0,5841646	32	Aisle	11	Place	0,098858632	48
Cashbuild increase	57	Service	0,51226675	33	Bulk deposit	9	Service	0,080884335	49
Fixed deposit account	57	Service	051226675	33	Comprehensive customer service	9	Service	0080884335	49
Inflation	57	Service	0,5122675	33	Customers	7	People	0,062910039	50
Issues	57	Service	0,5122675	33	Customer satisfaction	6	Service	0,05392289	51
Price meaning	57	Price	0,5122675	33	Experience	6	Service	0,05392289	51
Zero interest	57	Service	0,5122675	33	Good	6	product	0,05392289	51

Table 2: Top Frequently Used Keywords

The keywords were displayed as presented in figure 5, where the top keywords identified in table 2 were in white labels, with the dimensions of the blue blocks next to the keywords showing their frequency. The black lines represent the connectivity of the keywords.

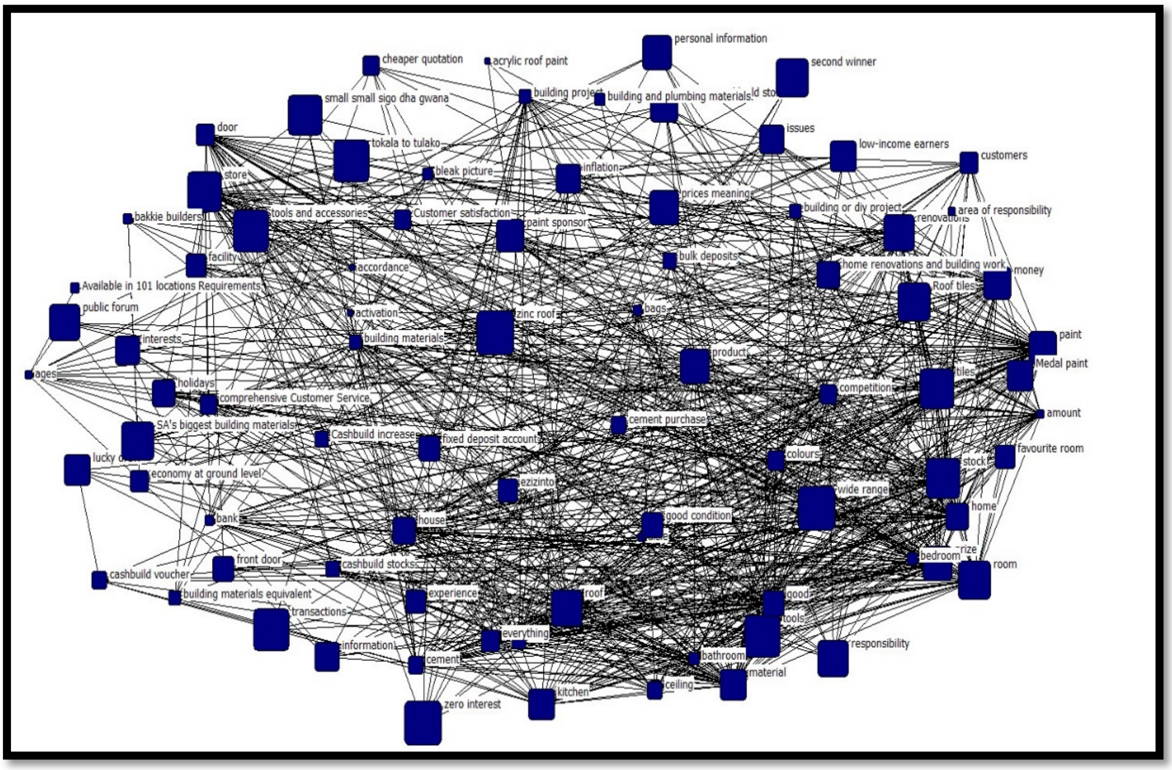

Figure 5: Visualisation of Top Keywords

Table 3 shows the results of the frequency of occurrence from the frequency analysis and the Freeman's Degree and Eigenvector, both of which determine the centrality of the keywords.

	Keyword		Frequency of Occurrence		Freeman's Degree Centrality		Eigenvector Centrality	
Keyword	**Frequency**	**Percentage**	**Rank**	**Degree**	**nDegree**	**Rank**	**Eigenvector**	**Rank**
Kitchen	1038	9,328660016	1	3496,000	0,019	14	0,207	6
building materials	826	7,42338456	2	1072,000	0,006	23	0,022	25
Renovations	824	7,405410263	3	4901,000	0,026	6	0,085	13
nearest cashbuild store	658	5,913543633	4	3923,000	0,021	13	0,042	22
Facility	656	5,895569336	5	3927,000	0,021	12	0,042	21
Colours	574	5,158623169	6	1346,000	0,007	21	0,096	10
Bathroom	483	4,340792666	7	2938,000	0,016	15	0,176	7
House	480	4,313831221	8	1752,000	0,009	20	0,084	14
Home	468	4,205985441	9	8347,000	0,044	3	0,411	3
Medal paint	333	2,99272041	10	2062,000	0,011	18	0,131	9
Room	308	2,7680417	11	7998,000	0,043	4	0,400	4
Paint	294	2,642221623	12	10169,000	0,054	2	0,436	2
Cement	255	2,291722836	13	1848,000	0,010	19	0,093	11
Competition	196	1,761481082	14	10170,000	0,054	1	0,472	1
Bedroom	188	1,689583895	15	4300,000	0,023	10	0,268	5
Door	168	1,509840927	16	1069,000	0,006	24	0,053	18
Tiles	159	1,428956592	17	2240,000	0,012	17	0,092	12
personal information	146	1,312123663	18	40,000	0,000	77	0,001	64
favourite room	139	1,249213625	19	379,000	0,002	38	0,020	26
lucky draw	139	1,249213625	19	192,000	0,001	48	0,008	33
Roof	128	1,150354992	20	2333,000	0,012	16	0,075	16
Information	124	1,114406399	21	86,000	0,000	65	0,002	50
public forum	113	1,015547767	22	40,000	0,000	78	0,001	65
Prize	112	1,006560618	23	4064,000	0,022	11	0,150	8
Roof tiles	112	1,006560618	23	642,000	0,003	28	0,013	30

	Keyword		Frequency of Occurrence		Freeman's Degree Centrality		Eigenvector Centrality	
Keyword	**Frequency**	**Percentage**	**Rank**	**Degree**	**nDegree**	**Rank**	**Eigenvector**	**Rank**
Everything	109	0,979599173	24	490,000	0,003	37	0,022	24
Store	106	0,952637728	25	4795,000	0,026	8	0,053	19
building project	102	0,916689135	26	843,000	0,004	25	0,003	47
cashbuild voucher	102	0,916689135	26	313,000	0,002	41	0,017	28
Amount	101	0,907701986	27	813,000	0,004	26	0,003	48
Material	101	0,907701986	27	6049,000	0,032	5	0,083	15
Ezizinto	99	0,889727689	28	304,000	0,002	42	0,004	40
Money	96	0,862766244	29	1266,000	0,007	22	0,017	27
Card	94	0,844791948	30	4804,000	0,026	7	0,045	20
acrylic roof paint	77	0,692010425	31	258,000	0,001	43	0,010	31
building and plumbing materials	65	0,584164645	32	71,000	0,000	67	0,001	55
Cashbuild increase	57	0,512267458	33	513,000	0,003	33	0,001	67
fixed deposit account	57	0,512267458	33	513,000	0,003	34	0,001	68
Inflation	57	0,512267458	33	524,000	0,003	31	0,001	60
Issues	57	0,512267458	33	516,000	0,003	32	0,001	66
prices meaning	57	0,512267458	33	513,000	0,003	35	0,001	69
zero interest	57	0,512267458	33	513,000	0,003	36	0,001	70
cashbuild stocks	54	0,485306012	34	358,000	0,002	39	0,005	37
wide range	52	0,467331716	35	352,000	0,002	40	0,005	38
area of responsibility	41	0,368473083	36	54,000	0,000	75	0,000	76
good condition	39	0,350498787	37	143,000	0,001	61	0,005	39
Bank	31	0,2786016	38	228,000	0,001	46	0,002	52
cheaper quotation	31	0,2786016	38	186,000	0,001	49	0,000	72
Interests	31	0,2786016	38	186,000	0,001	50	0,000	73
building materials equivalent	25	0,224678709	39	175,000	0,001	52	0,001	56
small small sigo dha gwana	25	0,224678709	39	175,000	0,001	53	0,001	57
tokala to tulako	25	0,224678709	39	175,000	0,001	54	0,001	58
tools and accessories	24	0,215691561	40	208,000	0,001	47	0,001	63
bleak picture	20	0,179742968	41	180,000	0,001	51	0,003	41
front door	19	0,170755819	42	68,000	0,000	68	0,002	51
Transactions	19	0,170755819	42	20,000	0,000	81	0,000	79
bakkie builders	18	0,161768671	43	174,000	0,001	56	0,003	42
Ceiling	18	0,161768671	43	699,000	0,004	27	0,033	23
cement purchase	18	0,161768671	43	66,000	0,000	70	0,000	75
economy at ground level	18	0,161768671	43	174,000	0,001	57	0,003	43
home renovations and building work	18	0,161768671	43	174,000	0,001	58	0,003	44

	Keyword		Frequency of Occurrence		Freeman's Degree Centrality		Eigenvector Centrality	
Keyword	**Frequency**	**Percentage**	**Rank**	**Degree**	**nDegree**	**Rank**	**Eigenvector**	**Rank**
low-income earners	18	0,161768671	43	174,000	0,001	59	0,003	45
SA's biggest building materials	18	0,161768671	43	166,000	0,001	60	0,003	46
paint sponsor	17	0,152781522	44	132,000	0,001	62	0,007	36
Tools	17	0,152781522	44	564,000	0,003	30	0,014	29
Accordance	15	0,134807226	45	25,000	0,000	79	0,000	81
Activation	15	0,134807226	45	60,000	0,000	72	0,000	77
Ages	15	0,134807226	45	238,000	0,001	45	0,009	32
Bags	15	0,134807226	45	129,000	0,001	63	0,002	49
second winner	15	0,134807226	45	60,000	0,000	73	0,000	78
Responsibility	13	0,116832929	46	78,000	0,000	66	0,001	59
Stock	13	0,116832929	46	575,000	0,003	29	0,008	34
zinc roof	13	0,116832929	46	50,000	0,000	76	0,001	61
Available in 101 locations Requirements	12	0,107845781	47	24,000	0,000	80	0,000	82
building of DIY project	12	0,107845781	47	175,000	0,001	55	0,001	71
Holidays	12	0,107845781	47	59,000	0,000	74	0,002	54
Product	12	0,107845781	47	244,000	0,001	44	0,007	35
Aisle	11	0,098858632	48	5,000	0,000	83	0,000	83
bulk deposits	9	0,080884335	49	5,000	0,000	84	0,000	84
comprehensive Customer Service	9	0,080884335	49	67,000	0,000	69	0,000	74
Customers	7	0,062910039	50	109,000	0,001	64	0,002	53
Customer satisfaction	6	0,05392289	51	14,000	0,000	82	0,000	80
Experience	6	0,05392289	51	63,000	0,000	71	0,001	62
Good	6	0,05392289	51	4449,000	0,024	9	0,063	17

Table 3: Comparison of Keyword's Frequency and Centrality

4.3. Determining Connectivity and Centrality of Keywords

To determine the connectivity of the keywords, we conducted a CONCOR analysis. It clustered the top keywords according to their context. The display of CONCOR analysis is illustrated in figure 6. Nodes for these clusters were identified using the CONCOR analysis.

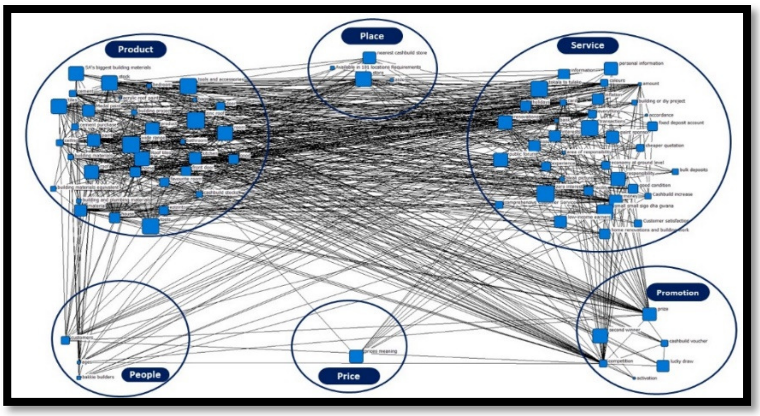

Figure 6: Visualisation of CONCOR Analysis

Table 4 shows how the groups were illustrated in the results. 41 significant keywords were identified based on their centrality ranking on Eigenvector value and Freema's Coefficient. Duplicate words were removed, and words that carried no meaning, like "everything", were excluded.

Category	Extracted Words	Significant Words
Products	kitchen/building materials/colours/bathroom/house/home/medal paint/room/paint/cement/bedroom/door/tiles/favourite room/roof/roof tiles/building project/material/ezizinto/acrylic roof paint/building plumbing materials/cashbuild stocks/wide range/building materials equivalent/tools and accessories/front door/ceiling/cement purchase/SA's biggest building materials/tools/bags/stock/zinc roof/product/good	paint/home/room/bedroom/kitchen/ bathroom/colours/cement/tiles/ house/material/roof/good/door/ ceiling/tools/stock/product/ building project
Service	renovations/facility/personal information/information/public forum/everything/amount/money/card/cashbuild increase/fixed deposit account/inflation/issues/zero interest/area of responsibility/good condition/bank/cheaper quotation/interests/small small sigo dha gwana/tokala to tukalo/bleak picture/transactions/economy at ground level/home renovations and building work/low income earners/paint sponsor/accordance/responsibility/building or DIY project/holidays/bulk deposits/comprehensive customer service/customer satisfaction/experience	renovations/card/facility/money/ paint sponsor/information/bank/ issues/cashbuild increase/fixed deposit account/cheaper quotation/ interests/transactions/bulk deposits
Promotions	competition/lucky draw/prize/cashbuild voucher/activation/second winner/	competition/prize/cashbuild voucher/lucky draw/activation
People	bakkie builders/ages/customers/	bakkie builder
Place	nearest cashbuild store/store/available in 101 locations requirements/aisle/	store/aisle

Table 4: CONCOR Analysis

4.4. Determining Categories of Significant Top Keywords

41 significant keywords were matched to the 16328 hit sentences, and 14 significant keywords shown in table 5 that matched less than 100 hit sentences were excluded.

No.	Significant Keyword	Unmatched HS	Matched HS	Application (Inclusion/Exclusion)
1	Activation	16 313	15	exclude
2	Aisle	16 327	1	exclude
3	bakkie_builder	16 310	18	exclude
4	Bank	16 275	53	exclude
5	Bathroom	15 238	1 090	include
6	Bedroom	15 131	1 197	include
7	building_project	16 185	143	include
8	bulk_deposit	16 323	5	exclude
9	Card	15 479	849	include
10	cashbuild_increase	16 271	57	exclude
11	cashbuild_voucher	16 191	137	include
12	Ceiling	16 054	274	include
13	Cement	15 058	1 270	include
14	cheaper quotation	16 297	31	exclude
15	Colours	15 564	764	include
16	Competition	10 372	5 956	include
17	Door	15 670	658	include
18	Facility	15 666	662	include
19	fixed_deposit_account	16 271	57	exclude
20	Good	15 416	912	include
21	Home	12 694	3 634	include
22	House	15 538	790	include
23	Information	16 301	27	exclude

No.	Significant Keyword	Unmatched HS	Matched HS	Application (Inclusion/Exclusion)
24	Interests	16 297	31	exclude
25	Issues	16 270	58	exclude
26	Kitchen	14 014	2 314	include
27	lucky draw	16 238	90	exclude
28	Material	14 757	1 571	include
29	Money	16 001	327	include
30	Paint	11 175	5 153	include
31	paint sponsor	16 303	25	exclude
32	Prize	14 362	1 966	include
33	Product	16 224	104	include
34	Renovations	15 215	1 113	include
35	Roof	15 328	1 000	include
36	Room	13 419	2 909	include
37	Stock	16 158	170	include
38	Store	15 301	1 027	include
39	Tiles	15 247	1 081	include
40	Tools	16 070	258	include
41	Transactions	16 321	7	exclude

Table 5: Significant Keywords Match to Hit Sentences

EFA is presented in table 6, showing the results of the factor loading and the Kaiser-Meyer-Olkin (KMO) test. 21 significant keywords loaded on 8 components with distinct amounts of variance. These groups (components) were named into categories according to relevant features of the keywords represented in each cluster.

Category	Words	Factor Loading	Eigen Value	Variance (%)
DIY Renovations	Facility Card Store Good Renovations Material	0.954 0.866 0.854 0.841 0.809 0.728	4.511	21.480
House Plans	Room Bathroom Bedroom	0.916 0.721 0.706	2.125	10.121
Building Project Planning	Building Project Money	0.859 0.807	1.567	7.462
Home Image	Kitchen Home	0.807 0.744	1.463	6.967
Makeovers	Paint Colours	0.788 0.735	1.334	6.351
Promotions	Prize Competition	0.786 0.776	1.211	5.767
Building Materials Inventory	Stock Tools	0.786 0.744	1.091	5.196
House Interior Finishes	House Ceiling	0.731 0.714	1.054	5.018
Total Variance (%) = 68.363				
KMO (Kaiser-Meyer-Olkin) = 0.736				
Bartlett's Test of Sphericity Chi-Squared = 125480.351 (p<0.001)				

Table 6: Results of Exploratory Factor Analysis (EFA)

Figure 7 shows a Scree Plot of the components where the eight components above in table 6 were selected based on the amount of variance that they created as shown from the Scree Plot.

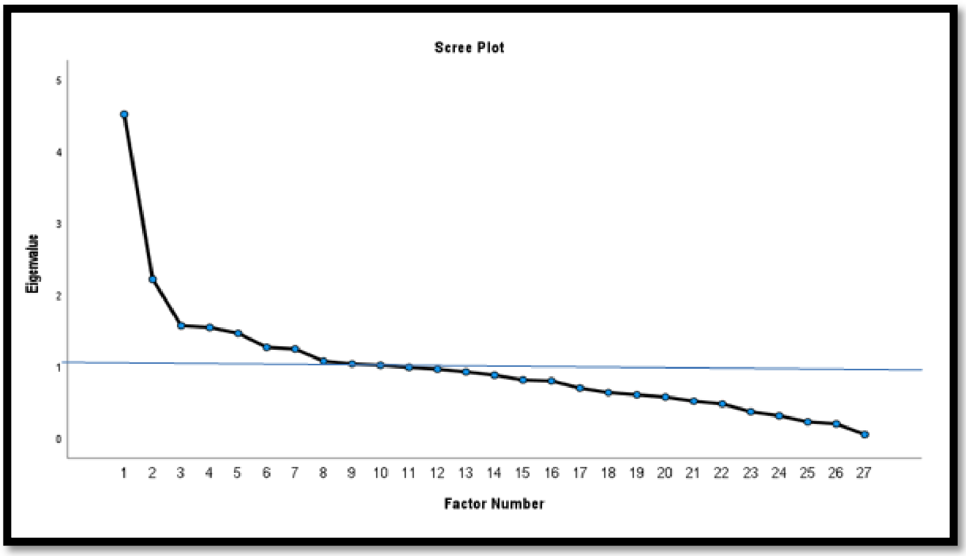

Figure 7: Scree Plot

4.4.1. Expert Review

A panel of experts from CBL management and Ogilvy, the marketers for CBL, developed the categories and keywords shown in table 7 after considering the categories and keywords resulting in table 6.

Category	Keywords
DIY Kings	Good/Store/Card/Room
Houseproud	Room/Home/House
Bob the Builder/Project master	Facility/Money/Building Project
Renovators	Renovations/ Paint/Colour
Interior Image	Ceiling/Bathroom/Bedroom/Kitchen
Building Materials Inventory	Material/Stock/ Tools
Promotions	Prize/ Competition

Table 7: Experts Review Results

Table 8 compares the alternative categories that were proposed by the panel of experts outlined in table 7 to the categories from the EFA presented in table 6.

New Expert Category	Factor Analysis Categories
DIY Kings	DIY Renovators
House Proud	House Interior Finishes & House Image
Bob the Builder/ Project Master	Building Project Planning & DIY Renovators
Renovators	Makeovers & DIY Renovators
Interior Image	House Plans & Home Image
Promotions	Promotions
Building Materials Inventory	Building Materials Inventory
N/A	House Interior Finishes

Table 8: Comparison of Categories

5. Discussion

This study was designed to establish top keywords from social networking sites and analyse them to determine the most significant keywords.

5.1. Determining the Sources and Sentiments of UGC

The UGC from the SNSs is analysed using Microsoft Power BI. Results are shown in figure 3. Facebook extracted the highest hits at 94.51%, followed by a distant Twitter at 3.95%, blogs at 0.79%, and Instagram at 0.76%. This shows that most customers post their views on Facebook, followed by Twitter. There is very little engagement on Blogs and Instagram. Other social network sites like TikTok and YouTube did not yield any significant results. CBL should, therefore, focus on Facebook and Twitter to communicate with its customers and track their perspectives. This is in line with the literature that was reviewed, which showed that Facebook was a leading social network site. Twitter, according to the literature, is the largest social network site used by brands to communicate with their customers.

Figure 3 shows that 65.63% of the UGC is neutral, 29.74% is positive, and 4.46% is negative. This shows that the majority (95.37%) of the hits are neutral to positive. This is good for Cashbuild Limited in that it may use the results of this study to reinforce its strategies, understanding that the sentiment from which the keywords were extracted was largely neutral to positive.

Figure 3 also shows that 61.37% of the UGC was posted from Gauteng, the largest economic province of the Republic of South Africa (RSA), where Cashbuild is headquartered, followed by the Western Cape at 9.78%, KwaZulu-Natal at 5.38%, and the Eastern Cape and a province from Namibia called Khomas at 4.65%. RSA locations dominate the UGC for Cashbuild, but figure 4 shows that there are posts from other countries like Khomas, Oshana, Karas, and Kavango East provinces in Namibia and Gaborone in Botswana. The results of this study emulate the economic activity of the region and the location of Cashbuild stores. The first objective of this study, "To determine sources and sentiment of the UGC of Cashbuild posted between 1st of September 2022 and 31st of December 2023," was met because the sources and sentiments were determined.

5.2. Identification of Top Keywords

84 top keywords were extracted from 16 328 hit sentences. This is in line with other similar previous studies where a study that was conducted for casino hotels in Busan, Fu et al. (2022) extracted 70 top keywords, and a study conducted for cruise lines in South Korea extracted 99 top keywords (Tao & Kim, 2019). The keywords represent the interaction of the users on "Products" like "kitchen, building materials, bathroom, cement, roof, etc.", "Services" like "renovations, facility, good condition, bank, cheaper quotation, etc.", "Promotion" like "Cashbuild voucher, lucky draw, competition, prize, activation, etc.", "Price" like "prices meaning", "Place" like "aisle, store, nearest Cashbuild store, etc.", "People" like "bakkie builder, ages, customers, etc." This showed that users use words from all 5 Ps of marketing in the UGC. The second objective, "To identify the top 80 to 90 keywords used in the UGC that were posted on social network sites of Cashbuild between 1st of September 2022 and 31st of December 2023", was met because 84 top keywords as outlined above and listed on table 2 above were identified.

5.3. Determining Frequency of Occurrence of Top Keywords

The top keywords identified in 5.2 occurred at varying frequencies. They are ranked 51st on the frequency of occurrence, as shown in table 2. Some shared similar frequencies. Figure 5 shows the keywords in white blocks, and the size of the blue blocks next to each word shows their frequency. The larger the blue bloc, the more frequently the keyword occurs in the UGC.

The "Products" category dominates the top 10 most frequently used words, featuring 7 top frequently used words, followed by 2 in the "Service" category and only one from "Place". Users frequently discuss products of Cashbuild Limited. The third objective, "To determine the frequency of the top 80 to 90 keywords in the UGC from the social network sites of Cashbuild between 1st of September 2022 and 31st of December 2023," was met because 84 top keywords identified in 5.2 were ranked by frequency.

5.4. Determining Connectivity and Centrality of Top Keywords

The 84 top keywords that are extracted are ranked based on their frequency of occurrence and further analysed for their connectivity using the CONCOR analysis. Freeman's coefficient, which measures how many keywords are connected to the word, was used. The ranking of the keywords changed when considering their connections to other keywords, as shown in table 3. The word "Competition" ranked the highest (compared to 14th on frequency ranking), followed by "Paint" (compared to 12th on frequency ranking), followed by "Home" (compared to 9 on frequency ranking). All the top 5 keywords on ranking using the Freeman's Coefficient are ranked higher than in frequency. This showed that even though these words occur less frequently, they connect to other keywords, thus demonstrating their significance.

The Eigenvector is used to determine the most influential nodes (Fu et al., 2022). Again, the keywords are ranked using their eigenvector values, as shown in table 3. Like the ranking using Freeman's coefficient, the words "Competition," "Paint," "Home," and "Room" all ranked in the top four keywords, respectively, using their eigenvector values. This confirmed the importance of these words using their connectivity.

Figure 5 shows connections in black lines between the keywords. Figure 6 shows the keywords in the categories that they were arranged into using the 5Ps and service as discussed above. The black lines also show connections between the keywords in the categories that they are arranged into. It can be seen in figure 6 that there is significant connectivity between words that were arranged into "Products" and "Services", and keywords that were arranged into "Promotions" also show a high number of connections. This seems to agree with the results from Freeman's Coefficient and the Eigenvector values that showed the highest connectivity for words like "Competition", which is arranged into "Promotions", "Paint", and "Home", which were arranged into "Products". Keywords that were arranged into "Place", "Price", and "People" had very little connectivity to other keywords. This indicates that the users significantly share their experience of products, services, and promotions.

41 significant keywords were determined out of 84 top keywords by considering their Eigenvector value from highest to lowest, Freeman's Coefficient, Frequency of occurrence, the meaning of the word, the relevance of the word, and repetition. Words like "Competition", "Paint", "Home", etc., that had high eigenvector values and Freeman's coefficients are placed first. Words like "Medal paint", even though it was ranked 9th by eigenvector value and 5th by Freeman's coefficient, are not included because the word "Paint" is already included in the list of significant keywords. Other keywords that are excluded due to repetition are "nearest Cashbuild store", "building materials", "favourite room", "Acrylic roof paint", "Cashbuild stock", "good condition", "home renovations and building", "SA's biggest building materials", "amount", "front door" "building materials plumbing", "Zinc roof", "Prices meaning", "building or DIY Project" and "Cement Purchases". Words also not included due to their lack of relevance like "Ages", "Wide Range", "Low-Income Earners",

"Holidays", "Responsibility", "Inflation", "Experience", "Personal Information", "Public Forum", "Zero Interest", "Area of Responsibility", and "Second Winner". Other words that were not included due to their lack of meaning were "Everything", "ezizinto", and "accordance". 41 significant words that were selected are shown in table 4. 41 significant keywords are central to the content on UGC that was extracted. The fourth objective, "To determine connectivity and centrality of the top 80 to 90 keywords in the UGC of Cashbuild that were posted on social network sites between 1st of September 2022 and 31st of December 2023," is met because the above analyses determined the connectivity and centrality of the top keywords.

5.5. Determining Categories of Significant Top Keywords

The 41 significant keywords are matched with the 16 328 hit sentences. 14 Keywords that have less than 100 matches to the hit sentences are excluded, as shown in table 5. Exploratory Factor Analysis (EFA) is conducted to determine the factor loadings of the remaining 27 significant keywords by explaining as much variance as possible between the factor groups.

The EFA is applied with spatial adjustment of each factor. As a result, 21 of the keywords are down to 8 factors by applying the Varimax rotation process. Common factorial criteria are utilised in computing the factors, selecting keywords that loaded above 0.4 are selected and used in the final model. Four keywords: "Products", "Cement", "Tiles", and "Ceiling", featured in more factor groups than one with factor loadings of less than 0.4, suggesting that they are not creating a sufficient variance. These keywords were, therefore, excluded in the final EFA. The 8 distinct factor groups are also shown in the Scree Plot in figure 7, above the red line with Eigenvalues of above 1.

Keywords that loaded above 0.4 were also obtained, considering that this varies from EFA using metric units. This research uses factors that are obtained from the semantic spaces in the UGC, and loadings from semantics are usually lower. The Eigenvalue of the variables was more than 1.0 to explain a substantial percentage of the total variance. Previously, 27 factors were derived from 41 significant keywords.

Finally, 21 words from the original 41 significant keywords are included in the 8 final factors with the most variance, commentating 68.363% of all variances. Table 6 shows the results of the EFA with the KMO of 0.736, which is more than 0.6, signifying it is primarily based on the recommender value. The 8 factors are named by the themes in each factor, as shown in table 6.

The 8 categories and keywords in table 6 were given to a panel of to comment based on their market and business insights. Table 7 shows the expert's input that recommends 7 categories with some of the keywords rearranged. In the first category, which had the biggest variance, the panel of experts retained 3 keywords, "Good", "Store", and "Card", and added a new word "Room", which is the second category of the EFA model, and they retained that category. All the 18 keywords that the panel of experts categorised were included in the model from the EFA. Only 'Ceiling", "Stock", and "Tools" were not used in table 7. Only "Room" was used twice by the panel of experts to explain the categories that they proposed. Table 8 compares categories proposed by the panel of experts to those derived from the model that was developed from the EFA. Only one category, "Building Materials Inventory", was not considered a fit by the panel of experts. The remaining 7 categories from the EFA compared favourably to the categories that were suggested by the panel of experts, as shown in table 8 above. This shows a very good fit, and the categories proposed by experts have been accepted. The fifth objective: "To determine categories of keywords from the UGC of Cashbuild, collect expert feedback regarding the proposed classification derived from statistical analysis, and investigate alternative categories based on the experts' industry and business insights," is met because the categories of keywords are determined using the EFA and alternative categories are investigated based on input from a panel of experts.

6. Conclusion and Recommendations

The study contributes to the customer-centric marketing strategy (CCMS) of Cashbuild Limited. The study aimed to conduct a semantic analysis of the UGC. The study uses secondary data from the UGC from 1st of September 2022 to 31st of December 2023. The objectives of the research are to determine the sources and sentiments of UGC, determine the keywords from users, determine the rank of keywords by frequency, establish connectivity and centrality of keywords, determine factors from significant keywords and investigate categories considering input from a panel of experts.

The data was mined using the MELTWATER program to extract keywords from the UGC. The extracted data was processed first using Power BI to understand its sources and sentiments. These showed that the majority of the UGC was posted on Facebook (94.51%), followed by Twitter (3.95%), and to a negligible extent, Blogs and Instagram. The majority of the UGC was posted from RSA Gauteng (61.37%) province and the City of Johannesburg (+40%). This is in line with the economic landscape of the region of southern Africa where Cashbuild Limited operates. The majority of the UGC had neutral (65.63%) sentiment followed by positive (29.74%) sentiment. More than 95% of the UGC had a neutral to positive sentiment, making it useful for considering a customer-centric marketing strategy.

RStudio program and Ucinet 6.0 were used to establish the connectivity and the centrality of keywords. UGC was enhanced from unshaped data to shaped data, and the frequency of keywords was visualised using NetDraw. The Social Program for Social Sciences (SPSS) was used to do the CONCOR analysis and the EFA. 85 top keywords were reduced to 41 significant keywords based on connectivity and centrality. 27 significant keywords that achieved more than 100 matches to the HS were retained for the EFA on SPSS. 8-factor that showed significant variances were modelled using the EFA. The 8 factors comprising 21 keywords were named. These categories were compared to categories that were developed by a panel of experts using industry and business insights. Only 3 out of 21 keywords were excluded by experts, and one word, "Room", was included in two categories. Only one category from the statistical model, "building materials inventory," did not fit the categories by experts, and one category, "Promotions", was retained by experts. The fact that 18 keywords in the statistical model were retained by experts and 7 out of 8 categories fitted expert opinion demonstrates that the statistical model is largely acceptable and may be considered with some changes in developing a customer-centric marketing strategy (CCMS). The key themes that are recommended for consideration in the formulation of a CCMS, namely DIY Kings, Houseproud, Project Master, Renovators, Interior Image, Building Materials Inventory, and Promotions, are illustrated on a theoretical model, as shown in figure 8.

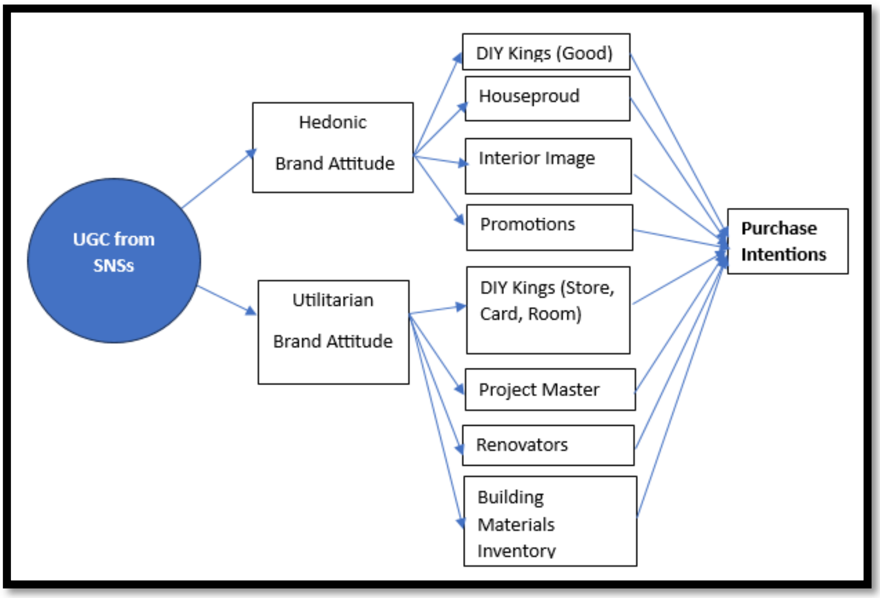

Figure 8: Theoretical Model

The categories position Cashbuild Limited closer to experiential marketing than the traditional market mix of 5Ps. The study is the first of its kind to be conducted in the building materials retail industry (BMRI). The study added to the knowledge in this discipline. It also contributed to the CCMS by providing results that analysed UGC from customers and determined categories of words that are at the centre of customer engagements on social network sites. Cashbuild Limited gained a better understanding of its UGC. They learned customer perspectives of offerings that satisfy customers, encouraging them to publish their customer experience on social network sites. Cashbuild learned the frequency, connectivity, and centrality of the keywords from UGC. It gained a deeper understanding of customers and their experiences of the brand.

6.1. Practical Implications

The study demonstrates that the UGC can contribute toward the development of a customer-centric marketing strategy. It proposes categories that are central in UGC, like “DIY Kings,” “Houseproud,” “Renovators,” “Project Master,” “Building Materials Inventory,” “Interior Image” and “Promotions”. These categories and keywords can be used by management to advertise Cashbuild Limited. The connectivity of keywords also showed that there was a high link between the use of keywords in the “Products” and the “Services” categories. There are also significant links between these two categories and “Promotions”. It is recommended that management consider increasing its use of Product-Services Systems (PSS) supported by intense promotional campaigns. This can be achieved through promotions that package Cashbuild Limited products and services, such as offering extra loyalty points to customers who purchase products like kitchens and bathrooms using partner bank facilities. This can be achieved through promotions that package Cashbuild Limited products and services, such as offering free deliveries and/or extra customer loyalty points to customers who purchase products like kitchens and bathrooms using partner bank facilities. It is recommended that Cashbuild Limited continue with UGC analysis to gain more insights.

Other retailers in southern Africa and globally may consider the results of this study to create customer-centric marketing strategies.

6.2. Limitations and Future Research Studies

This study had limitations regarding the collection of data and data analysis. Collected data is limited to social network sites. Future research should consider sourcing data from social network sites of other building materials industry (BMRI) players globally.

Data from the UGC of BMRI should be considered for improved generalisation of the results. There was no segmentation of the market based on their Living Standards Measures (LSM), social identity, gender, age, race, or any other customer segmentation criteria.

The study was limited to UGC in the English language, even though the extraction of keywords did pick up some words like “ezizinto” (meaning these things) from vernacular languages. However, a better understanding of CPs could be obtained by including UGC, which uses other languages.

The analysis of data was limited to frequency, connectivity, and centrality of keywords. More sentiment analysis should enhance understanding of UGC.

Some of the categories in the EFA only had 2 words, so conducting a Cronbach's Alpha value to determine internal consistency would have been meaningless on only 2 - 6 keywords. Cronbach Alpha Values are sensitive to the number of items in the data. When there are fewer than 10 items, Cronbach's Alpha tends to be low (Pallant, 2016).

Since CBL no longer conducts regular customer satisfaction surveys, it is not possible to use the results of this study to determine the relationship between the 8 categories of the EFA results and customer satisfaction from the survey.

Cashbuild may consider conducting a customer satisfaction survey and carrying out a regression analysis, where the average customer satisfaction rating will be a dependent variable and 7 factors from the study will be independent variables. Cashbuild Limited may also consider conducting an ongoing customer satisfaction survey for its online shop customers.

Cashbuild must also consider conducting a sentiment analysis on its UGC to gain insight into the sentiments of the UGC. Lastly, Cashbuild may consider using computer vision for brand marketing and pre-trained, ready-to-use computer vision models like YOLOV2, Google Cloud Vision, and Clarifai to analyse UGC images like videos and pictures to further enhance its analyses of UGC.

7. References

i. Al-Gasawneh, J. A., Hasan, M., Joudeh, J. M. M., Nusairat, N. M., Ahmad, A. M. K., & Ngah, A. H. (2023). Mediating role of e-Word of Mouth on the Relationship between Visual Social Media Marketing and Customer Purchase Intention in Jordanian Real Estate Companies. *Quality Access to Success, 24*(193).

ii. Alhabash, S., Park, S., Smith, S., Hendriks, H., & Dong, Y. (2022). Social Media Use and Alcohol Consumption: A 10-Year Systematic Review. *International Journal of Environmental Research and Public Health, 19*.

iii. Alharahsheh, H. H., & Pius, A. (2020). A Review of key paradigms: positivism vs interpretivism. *Global Academic Journal of Humanities and Social Sciences, 2*(3), 39–43.

iv. Artem, T., & Hauser, J. R. (2019). Identifying Customer Needs from User-Generated Content. *Institute for Operations and the Research Management Sciences*.

v. Bakri, M. H., Zamli, N. S., & Azma, H. (2012). The Effectiveness of Advertising in Digital Marketing towards Customer Satisfaction. *Journal of Technology Management and Technopreneurship*.

vi. Bilal, M., Akram, U., Rasool, H., Yang, X., & Tanveer, Y. (2022). Social Commerce isn't the Cherry on the Cake, it's the New Cake! How Consumers' Attitudes and eWOM Influence Online Purchase Intention in China. *International Journal of Quality and Services Sciences, 14*(2), 180–196.

vii. Cashbuild. (2022). *Integrated Report*.

viii. Cashbuild. (2023). *DM Presentation*.

ix. Davcik, N. S., Langaro, D., Jevons, C., & Nascimento, R. (2022). Non-sponsored brand-related user-generated content: Effects and mechanisms of consumer engagement. *Journal of Product and Brand Management, 31*(1).

x. Dos Santos, M. L. B. (2021). The "so-called" UGC: An updated definition of user-generated content in the age of social media.

xi. Fu, W., Wei, S., Wang, J., & Kim, H.-S. (2022). Understanding the Customer Experience and Satisfaction of Casino Hotels in Busan through Online User-generated Content. *Sustainability, 14*(5846).

xii. Goncalves, C. T., Goncalves, M. J. A., & Campante, M. I. (2023). Developing Integrated Performance Dashboards Visualisations Using Power BI. *Information, 14*(614).

xiii. Halliday, S. V. (2016). User-Generated Content about Brands: Understanding the Creators and Consumers. *Journal of Business Research, 69*(1), 137–144.

xiv. Hu, F., & Trivedi, R. (2020). Mapping Hotel brand positioning and competitive landscapes by text mining user-generated content. *International Journal of Hospitality Management, 84*(102317).

xv. Huang, O., & Copeland, L. (2020). Gen Z, Instagram Influencers, and Hashtags' Influence on Purchase Intention of Apparel. *Academy of Marketing Studies Journal, 24*(3).

xvi. Israfilzade, K., & Baghirova, S. (2022). Comparing the Impact of Brand and User-Generated Video Content on Online Shoppers' Purchasing Intention. *Sciendo*.

xvii. Kanwar, A., & Huang, Y.-C. (2022). Exploring the impact of social media influencers on customer' purchase intention: A sequential mediation model in Taiwan context. *Entrepreneurial Business and Economics Review, 10*(3).

xviii. Kaosiri, Y. N., Fiol, L. J. C., Tena, M. A. M., Artola, R. M. R., & Garcia J. S. (2019). User-Generated Content Sources in Social Media: A New Approach to Explore Tourists Satisfaction. *Journal of Travel Research, 58*(2), 253–265.

xix. Karunanayake, T., & Madubashini, C. (2019). The Influence of User-Generated Content on Purchase Intention of Automobiles in Sri Lanka. *Asian Social Science, 15*(6).

xx. Kitsios, F., Kamariotou, M., Karanikolas, P., & Grigoroudis, E. (2021). Digital Marketing Platforms and Customer Satisfaction: Identifying eWoM using Big Data and Text Mining. *Applied Science*.

xxi. Kroeze, J. H. (2011). Interpretivism and Information Systems: A Postmodern Epistemology? *Sprouts: Working Papers on Information Systems, 11*(17).

xxii. Kunja, S. R., Kumar, A., & Rao, B. (2021). Mediating role of hedonic and utilitarian brand attitude between eWOM and Purchase Intentions: A context of brand fan pages in Facebook. *Young Consumers, 23*(1).

xxiii. Li, S., Liu, F., Zhang, Y., Zhu, B., Zhu, H., & Yu, Z. (2022). Text mining of User-Generated Content (UGC) for Business Applications in E-Commerce: A Systematic Review. *Mathematics, 10*.

xxiv. Long, P. D., & Nga, D. Q. (2020). Factors of Facebook Advertising affecting the Purchase Intention – A Case of Psychological Consulting Service.

xxv. Mastrogiacomo, L., Barravicchia, F., Franceschini, F., & Marimon, F. (2021). Mining quality determinants of product-service systems from user-generated content. *Quality Engineering*.

xxvi. Mayrhofer, M., Matthes, J., Einwiller, S., & Naderer, B. (2020). User-generated content presenting brands on social media increases young adults' purchase intention. *International Journal of Advertising, 39*(1), 166–186.

xxvii. Mishra, M. (2022). Customer Experience: Extracting Topics from Tweets. *International Journal of Market Research, 64*(3), 334–353.

xxviii. Mohammad, J., Quoquab, F., Thurasamy, R., & Alolayyan, M. N. (2020). The Effect of User-Generated Content Quality on Brand Engagement: The Mediating Role of Functional and Emotional Values. *Journal of Electronic Commerce Research, 21*(1).

xxix. Muda, M., & Hamzah, M. I. (2021). Should I suggest this YouTube clip? The Impact of UGC Source Credibility on eWOM and Purchase Intention. *Journal of Research in Interactive Marketing, 15*(3), 441–459.

xxx. Necochea-Chamorro, J. I., & Larrea-Goyochea, L. (2023). Business Intelligence Applied in the Corporate Sector: A Systematic Review. *TEM Journal, 12*(4), 2225–2234.

xxxi. Nirmalasari, L., Alwiyah, Sunarya, P. A., & Panjaitan, A. S. (2022). A Digital Marketing Strategy Based on Applications to Rising Customer Satisfaction. *International Journal of Cyber and IT Service Management (IJCITSM), 2*(2), 139–145.

xxxii. Nosita, F., & Lestari, T. (2019). The Influence of User-Generated Content and Purchase Intention on Beauty Products. *Journal of Management and Marketing Review, 4*(3), 171–183.

xxxiii. Nusairat, N. M., Alroale, M. A. Qeed, M. A., Al-Gasawneh, J. A., Hammouri, Q., Ahmad, A., Sumaya, P., & Abdellatif, H. (2021). User-Generated Content – Consumer Buying Intentions Nexus: The Mediating Role of Brand Image. *Academy of Strategic Management Journal, 20*(4).

xxxiv. Nuseir, M. T., El Rafae, G. A., Aljumah, A., Alshurideh, M., Urabi, S., & Al Kurdi, A. (2023). Digital Marketing Strategies and the Impact on Customer Experience: A Systematic Review.

xxxv. Pallant, J. (2016). *SPSS Survival Manual – A step-by-step guide to data analysis using IBM SPSS*. McGraw Hill Education.

xxxvi. Panopoulos, A., Poulis, A., Theodoridis, P., & Kalampakas, A. (2023). Influencing Green Purchase Intention through Eco Labels and User-Generated Content. *Sustainability, 15*(764).

xxxvii. Polanco-Diges, L., & Debasa, F. (2020). The Use of Digital Marketing Strategies in the Sharing Economy: A Literature Review. *Journal of Spatial and Organizational Dynamics, VIII* (3), 217–229.

xxxviii. Putra, A. H. P. K., Nurani, N., Ilyas, G. B., Samiha, Y. T., & Lestari, S. D. (2021). Configure the Symmetrical and Asymmetrical Paths of Brand Equity and the Relationship between Firm Created Content and User Generated Content as Antecedents. *Jurnal Manajemen Bisnis, 8*(1).

xxxix. Rane, N. L., Achari, A., & Choudhary, S. P. (2023). Enhancing Customer Loyalty through Quality of Service: Effective Strategies to Improve Customer Satisfaction, Experience, Relationship, and Engagement. *International Research Journal of Modernization in Engineering and Technology and Science, 5*(5).

xl. Sanchez, B. J. D. (2023). Analysis of tweets from food retailers operating in Spain and the UK: How user-generated content on Twitter can help agrifood cooperatives build better relationships with their customers. *REVESCO, Revista de Estudios Cooperativos, 1*(143), e85557.

xli. Saura, J. R., & Bennett, D. (2019). A Three-Stage Method for Data Text Mining: Using UGC in Business Intelligence Analysis. *Symmetry, 11*(519).

xlii. Scotland, J. (2012). Exploring the Philosophical Underpinnings of Research: Relating Ontology and Epistemology to the Methodology and Methods of the Scientific, Interpretive, and Critical Research Paradigms. *Canadian Center of Science and Education, 5*(9).

xliii. Sharma, N., & Sarkar, D. (2022). Healthcare Data Analytics Using Power BI. *International Journal of Software Innovation, 10*(1).

xliv. Shridhar, R. (2023). Digital Marketing and Customer Satisfaction: A Comparative Study. *Social Science Journal. Res Militaris, 13*(1).

xlv. Stanely, O. A., MacAttama, & Chinelo, A. (2017). Improving Customer Satisfaction through Digital Marketing in the Nigerian Deposit Money Banks. *Open Access International Journal of Science and Engineering, 2*(7).

xlvi. Susiang, M. I. N., Suriyaningrum, D. A., Masliardi, A., Setiyawan, E., & Abdillah, F. (2023). Enhancing Customer Experience through Effective Marketing Strategies: The Context of Online Shopping, *6*(2), 437–447.

xlvii. Tao, S., & Kim, H.-S. (2019). Cruising in Asia: What can we dig from online cruiser reviews to understand their satisfaction and experience? *Asia Pacific Journal of Tourism, 24*(6), 514–528.

xlviii. Tariyal, A., Bisht, S., Rana, V., Roy, S., & Pratap, S. (2022). Utilitarian and Hedonic Values of eWOM Media and Online Booking Decisions for Tourist Destinations in India. *Journal of Open Innovations: Technology, Market, and Complexity, 8*(137).

xlix. Vazquez, D., Cheung, J., Nguyen, B., Dennis, C., & Kent, A. (2020). Examining the influence of user-generated content on the fashion consumer online experience. *Journal of Fashion Marketing and Management, 25*(3).

l. Verhoef, P. C. (2021). Customer experience creation in today's digital world. In B. B. Schlegelmilch, & R. S. Winer (Eds.), *The Routledge Companion to Strategic Marketing* (107–122). Routledge.

li. Wang, S., Lin, Y., Yan, Y., & Zhu, G. (2023). Social media user-generated content, online search traffic and offline car sales. *Kybernetes, 52*(11).

lii. Yoo, B., Katsumata, S., & Ichikohji, T. (2019). The Impact of Customer-orientation on the quantity and quality of User-Generated Content: A multi-country case study of mobile applications. *Asia-Pacific Journal of Marketing and Logistics, 31*(2), 516–540).

liii. Yu-Jin, K. (2019). The Relative Impact of User- and Market-Generated Content on Consumer Purchase Intentions: The Case of the Social Media Marketing Platform, 0.8L. *International Association of Societies of Design Research Conference*.

liv. Zinko, R., Furner, C. P., de Burgh-Woodman, H., Johnson, P., & Sluhan, A. (2020). The Addition of Images to eWOM in the Travel Industry: An Examination of Hotels, Cruise Ships and Fast-Food Reviews. *Journal of Theoretical and Applied Electronic Commerce Research, 16*, 525–54.