International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 5, 2025
https://theaspd.com/index.php

Sentiment Signals: Decoding Online Consumer Emotions In India's E-Commerce Sector

Sachin Mahendru[1], **Dr. Rachin Suri**[2]
[1]PhD Research Scholar, MM Institute of Management, MM (Deemed to be) University, Mullana, Ambala, Haryana, India, sachinmahendrummu@gmail.com
[2]Associate Professor, MM Institute of Management, MM (Deemed to be) University, Mullana, Ambala, Haryana, India, rachin.suri@mmumullana.org

Abstract
The aim of the research is to comprehend and examine the sentiments of the consumers in form of app-based review in the leading Indian e-retailers namely Amazon, Flipkart and Snapdeal, the study. This paper is a study based on the basis of text mining and sentiment analysis that investigates how the customers tend to emotionally respond to their treatment in such sites. The data encapsulation and processing thereof using Python and RStudio and analysis of the data warehouse which has a data set of over 25,000 user reviews collected on Google Play Store were conducted. It employed a sentiment categorizing method based on lexicon in which a positive, negative sentiment and neutral sentiments along with a more detailed sentiment emotion of trust, joy, anticipation and anger are identified using NRC Emotion Lexicon. Based on thematic analyses of the sentiment trend on each of the three platforms, visualization in the form of word clouds, comparison clouds, and polarity boxplots methodology detailed the content.
This finding reveals that most of the positive messages were received by Flipkart as compared to Amazon and Snapdeal. Ordinary words were used such as service, order and shopping which pointed to the areas of interest to the consumer and words such as good, nice and worst pointed to various emotional reactions. The interpretation reveals that the consumer sentiment is strongly influenced by the how easy the app was to use, the rate at which the delivery was received and fastest, the promotional activity, and the response given by the customer service desk.
Such a study would come in handy in contributing to the generalizing new body of the literature regarding the digital consumer behavior and providing realities of practices at the e-commerce level in the determination of effectiveness on user engagement. It also shows how sentiment analytics is important in competitive benchmarking and brand positioning.
Keywords: *Sentiment analysis, Text mining, E-Commerce, Consumer reviews, Lexicon based classification, Flipkart, Amazon, Snapdeal, RStudio.*

1. INTRODUCTION & BACKGROUND

Digital revolution has not only changed the face of commerce but it has changed the relationship between the consumer and the brands. The new force of the consumer voice, the e-commerce sites has been developed by the 21st century, there it is the place where the satisfaction, disappointment, trust, and criticism are shared through the use of star ratings, reviews, and app feedback. Sentiment is the heart beat of the world and in the event that it is driven by the consumer data as is being done today is literally a direct reflection of the customer perspective that a business can base their strategies on.

Through the proliferation of the internet connection, the use of smartphones, digital literacy levels, and ease of transacting online, the e-commerce industry in India has seen records of its whopping growth. India e-commerce is expected to increase to US350 billion in 2030 which is a trend in some of the fastest-growing markets in the world (IBEF, 2023). The change in pattern of online purchases that was catalyzed due to the COVID-19 outbreak has led to entry of such e-commerce giants as Amazon, Flipkart or Snapdeal in everyday lives of consumers. However with maturity of the industry the demands of the customers have become complex and nuanced. In order to retrieve such expectations, a mere quantitative rating is not sufficient to retrieve them all, additional attempts are required to retrieve the qualitative feedback itself, as well as the background emotion, and this is where sentiment analysis is already required (Pang & Lee, 2008).

Conventional ways of getting feedback through the use of surveys and questionnaires is slowly being done away with as text mining and sentiment analysis provide a scalable, efficient, and real-time view of

consumer sentiment. There is a lot of written data available on the sites of application review and electronic commerce, but it is untructured. To interpret it, one or more tools detecting the context, polarity (positive, negative or neutral), and emotional states, including trust, joy, anger, anticipation, or sadness are required (Liu, 2012). Consequently, sentiment analysis has emerged as an important tool in capturing practical insight on textual feedback turning the consumer voice into an empirical brand intelligence (Medhat et al., 2014).

The main aim of the proposed research work is to carry out text analysis and sentiment analysis of app based reviews of three prominent Indian e-commerce portals Amazon, flipkart and Snapdeal. It is not to measure simply the mood of consumer reviews but also to gauge customer perception of such sites with regard to the quality of services, reliability, user-friendliness, and customer satisfaction. This study offers an analytic framework by which businesses have access to know what the consumer is feeling and why he or she feels so, far above the ordinary business intelligence terms.

In a theoretical aspect, the study is based on the Consumer Behavior Theory with the approval of Natural Language Processing (NLP) techniques. The interdisciplinary technique of text mining appertains to the fields of linguistics, computer science and data analytics to perform preprocessing, cleaning, and transformation by which raw user-generated content is transformed into analyzable forms (Feldman & Sanger, 2007). After that, the lexicon-based methods of sentiment classification are applied, especially the NRC Emotion Lexicon that divides the textual material into such simple emotions as trust, anger, fear, surprise and joy, among the others (Mohammad & Turney, 2013).

A number of researches have pointed to the growing influence of electronic word-of-mouth (eWOM) as a factor in consumer purchasing behaviour. EWOM is contagious, and a wider digital audience gets affected by it as compared with traditional WOM, which is narrow in scope and reach. The research by Jimenez and Mendoza (2013) revealed that eWOM has great influence on consumer trust, brand perception, and purchase intention. Chatterjee (2019) even states that reviews are supposed to be valued just second to friends and family at the provision of product information. This highlights the fact that customer reviews should be interpreted not only as individual thoughts but emotional stories that talk about stronger brand-customer interactions.

In India, reviews provided by customers on such websites as Flipkart and Amazon are likely to offer the product satisfaction rating as well as the review regarding the delivery, the refund system, application interface, and customer service. They combine multi-dimensional feelings that should be understood as a whole. Another example would be a review that says, Great product but terrible customer service, which would be two-fold message and would never be represented with star points. Polarity scoring and emotion detection hence, becomes a necessity (Thelwall et al., 2010).

Besides, the importance of sentiment analysis in the Indian e-commerce environment is highlighted by studies included in the research pool, namely, by those conducted by Shivaprasad and Shetty (2017) and Sarin et al. (2021). They highlight that the customer attitude does not just depend on the product, but on the whole environment of purchase- price, package, delivery, so far the app responsiveness and after purchasing communication. Such facts are particularly important in price-sensitive and culturally diverse market as India.

On a methodological side, this study uses the methods of text preprocessing including its tokenization, removal of stop-words, lower-casing, and filtering of punctuation by applying libraries like NLTK and pandas in Python. After cleaning, the reviews are run in RStudio where sentiment lexicon tools such as syuzhet, meanr and sentimentr (along with text mining tools) help in analysing the pattern in terms of frequency and also the polarity of emotion. The research also represents the sentiment trends using different tools including word clouds, commonality clouds as well as polarity boxplots to provide a comparative view between the three sites.

It is interesting to note the early results. One can take the example that when viewed, Flipkart is widely referred to as a local favourite but when it comes to the points of positive sentiment regarding the use of the application and delivery time of arrival of the order is high and Amazon is being given credit regarding the quality of goods and trust to the company. The sentiment in Snapdeal on the other hand is quite conflicting in nature where its users raise their complains to the question of authenticity of the products and also the inconsistency of their services. The managers of platforms, marketers, and UI/UX designers

can have a strategic sense of these differences, which were depicted on sentiment polarity charts and the top-frequency words.

In addition, this article presents the decent case regarding sentiment analytics as a benchmarking tool of the competition. Kumar (2015) notes that the ideology of the Indian e-commerce websites ought to be based on an emotional and perception appeal that the customers feel to stick and hold on the shares as the world of competition is increasing in the form of such businesses. Sentiment mapping would enable companies to help them know what red flags to watch out of at early stages (e.g. the rise of negative sentiment when it comes to being late with deliveries or returning policies). This not only assists in improving the customer satisfaction level but also assists in solidifying the brands in the unpredictable online markets. The suggested study can bridge the existence of such a vital gap in research as sentiment analysis tools of a higher sophistication compared to the ones used by Indian e-commerce setting will be employed. It doe s not look at customer sentiment as an issue of the periphery but as a part of their resource. Through new forms of data mining, categorizing emotion, effects of comparative visual analytics, it will make available real-time information to decision-making abilities on e-commerce sites that will ultimately be able to make decisions based off of the emotion.

The Research makes many contributions. In terms of academics it further contributes to the growing completion of literature on consumer sentiment in online medias in particular in the emergent economies. In practice, it has the potential to give to the brand which is interested in using the power of customer voice as an improvement of its service and competitive advantage a replicable example. The companies will be able to re-code their actions based on the more customers-oriented, responsive, and empathic understanding of the reviews of the emotions using the means of translating the understanding of the reviews of the emotions.

1.1 Theoretical Background:

Over the past years, the growth of the digital platform has transformed the consumer relation to a drastic pace with businesses. E-commerce has emerged to be the primary soft interface, which is being utilized by consumers to visualize the brands and share the experiences as well as influencing the brand preferences within peer groups. The change has not only introduced the possibilities of the new prospects to the businesses, but also has presented theoretical dilemma in the accomplishment of the perception, satisfaction and behavior of the online consumers. It has necessitated the use of various theoretical models such as use of information systems, marketing and psychology models in the examination of the digital consumer sentiment. The study conducted by the author is informed by such main theories as Electronic Word of Mouth (eWOM), Technology Acceptance Model (TAM), Sentiment Analysis Theory and Consumer Behavior Theory that, together, form the conceptualization of user-generated content in Indian e-commerce study as a research objective.

1.1.2 Electronic Word-of-Mouth (eWOM) Theory

In eWOM, the word electronic indicates the involvements of internet and the word Word-of-Mouth signifies any affirmative or negative spread of a statement by prospective, real and previous purchasers of a commodity or organization across a large number of people and organizations (Hennig-Thurau et al., 2004). Unlike the traditional word-of-mouth method, eWOM is broader in reach and will never diminish, thereby greatly influencing the decision making process of the consumers. It is possible to place customer reviews, the reviews that are accompanied by the star rating, or the reviews that include the remarks of any app that praises or criticizes as the good examples of eWOM in the e-commerce environment.

The empirical observation suggests that eWOM has considerable influence on brand image, trust, and intention of purchase (Cheung & Thadani, 2012). To illustrate, Chatterjee (2019) explains that eWOM has the ability to affect the trust levels and determination of the quality of services in online Indian retail markets. The internet reviews are increasingly viewed in terms of accepted user generated messages of praise or extortion that define the mass consumption behavior. This theory is used to expound on the pertinence of extraction of the consumer sentiment of the bulk of data that is in the form of text in the e-commerce websites.

1.1.2 Technology acceptance model code (TAM)

It was Davis (1989) who fully elaborated the Technology Acceptance Model (TAM) and it describes the factors which influence the acceptance of information technology on its users. The model poses the

assumption that the attitude held by the users of the technology correlates to two key beliefs namely Perceived Usefulness (PU) and Perceived Ease of Us (PEOU) which eventually influence the behavioral intention to use the technology. TAM has been developed over the years as it has added other constructs like that of Subjective Norms, Image, Job Relevance and Output Quality to be changed to set to TAM2 and TAM3 (Venkatesh & Davis, 2000).

The TAM in the current context of the research paper will have relevance because the technological devices of the e-commerce applications (Amazon, Flipkart, Snapdeal) are concerned. The way the users will employ their feelings and action will depend on how they perceive these apps (workability, responsiveness, and integration of service). App attitude towards technology is generally based upon the reviews of the users as they always make it a point to leave a review such as, easy to use, confusing layout, very helpful to make returns in other words similar to the modified TAM.

1.1.3 Theory of sentiment analysis

A sentiment analysis is type of natural language processing which was being used to calculate the feeling behind or attitudes of a piece of written material, also known as opinion mining. The subjective/sentiments analysis theory is on the assumption that any piece of writing has a subjective meaning and this subjective meaning may be termed as having either the positive, negative sentiments or as a sentiment of neither (Liu, 2012). The more advanced models even identify specific emotions such as trust, anticipation, anger or joy.

In this research, the sentiment analysis is performed on the basis of lexicon, which in this example is the NRC Emotion Lexicon (Mohammad & Turney, 2013) that defines words in the framework of emotional types and sentiment (polarized). The theoretical strength of sentiment analysis is that it can capture subtle customer response on unstructured data and therefore, this procedure can become the source of insight on brand perception, service quality and user experience (Thelwall et al., 2010).

Text mining and sentiment analysis are not just computational instruments but also a behavioral decoder by which the researcher could unfurl more convoluted state of existence in man or woman that are represented using digital methods. Sentiment theory also allows conducting such a study because it allows taking a coherent approach to the quantification of subjective information to achieve the goal of producing business intelligence.

1.1.4 Theory of Consumer Behavior

Guided primarily by the understanding of psychology, sociology and economics, the consumer behavior theory aims at theorizing the consumer action in determining how to spend resource on shopping related products (Schiffman & Kanuk, 2007). The core concepts include motivation, perception, learning, beliefs, attitudes, and also decision-making. The theory when applied in e-commerce case expands to include digital interactions, level of perceived risk, searching of information, behavior and evaluation after purchase.

The theory is relevant to prove the reasons why the customers write reviews, what stimulates factors produce the desire to share their emotions and how these emotions impact the continuation of the buying process. It is also somewhat connected to the emotional branding and user experience design, which always stand at the center of customer loyalty in the digital market. The expectations of the consumers in the Indian market place are highly situational and it is after conducting a behaviour study that we are able to find out such useful information by way of market segmentation and brand association (Kumar, 2015).

1.5.5 Unified Theory of Acceptance and Use of Technology (UTAUT)

Unified Theory of Acceptance and Use of Technology (Venkatesh et al., 2003) develops TAM by incorporating some new variables into it, such as the Social Influence, the Facilitating Conditions, and the Performance Expectancy. The surrounding environment and the convenience in which the technology must fit in their day to day lives still affects the users according to the model.

The UTAUT has an indirect mention in consumer reviews since when users discuss the peer effect of product and service (My friend has recommended Flipkart) support (Snapdeal offers poor customer care) or infrastructure (Amazon App goes unresponsive at awkward times) they had UTAUT in their minds. All these elements refer to the massive ecosystem, which preconditions the context of user experience and feeling.

1.6 The Role of Trust and Perceived Risk

The online consumer behavior models are very sensitive to the antecedent measures of the trust and perceived risk (Gefen et al., 2003; Featherman & Pavlou, 2003). Other than the mailing address, users are involved in evaluating reliability grouped on the quality of services, the accuracy of the product and how fast the product was attended to. The negative review that states delays, counterfeits and poor customer service are all directly related with increase in perceived risk which then attracts negative feeling. There are also positive relations between satisfaction and loyalty and trust. Ratings offered by a user with the line tagged on the product Flipkart never disappoints with high level of trust and positive emotional involvement. The affinity of this lens will have the implication such that the affective component of trust in digital commerce will be decomposed. Altogether, these theoretical foundations constitute a strong ground were trends in sentiments of Indian e-commerce environment can be investigated. This study would show a clear image of consumer behavior according to the eWOM, TAM, Sentiment Analysis, Consumer Behavior Theory, and UTAUT, and this aspect would assist us towards appreciating consumer behavior as a whole in relation to the interaction and assessment of e-commerce markets by the consumers. The amalgamation of behavioral and computational theories ensures not that the research be used to mine data but to interpret this into the meaningful strategic and customer-based results.

2. Objectives of the Study:

So, representing the above issue and in line with the above information this study tries to highlight the following objectives:

Objective 1: To determine the emotions and sentiments expressed by the users through the application review.

Objective 2: To establish the various themes generated by the e-commerce websites and qualitative analysis of the user generated content.

2.1 Hypothesis Development

Responses by consumers to e-commerce businesses at an emotional level ought to be determined in the context of a digital economy. As the online reviews on the better or worse quality of the product as well as the experience with the services grow, the textual content of those feedbacks becomes a fine source of the insight into the behavior of the consumers. Grouping the main theoretical assumptions to Electronic Word-of-Mouth (eWOM), Technology Acceptance Model (TAM), or Sentiment Analysis Theory, there is a set of hypothesis stated in the paper in the form of testing the perceptions of the individuals with the most common and popular Indian online resellers, which are Amazon, Flipkart, and Snapdeal.

2.1.1 Distinctions of sentiments in polarity across the platforms

In the case of e-commerce, consumer perception is also very different depending on how e-commerce is viewed on the various sites in terms of user friendliness, shipping and customer care services. In past cases, it has been mentioned that platform sentiment is different among the customer due to brand perception and application design and app that satisfies the request to the greatest extent (Shivaprasad & Shetty, 2017). These distinctions find their expression in the differentiation of user generated content in the eWOM theory which is quantifiable and can be compared across brands (Hennig-Thurau et al., 2004).

In India Flipkart is somehow assumed to be a local brand that is regional in uniformity but Amazon is an international brand that has some differences in the service output to an extent (Kumar, 2015). Snapdeal used to be one of the forces to reckon with until recently when because of the lack of confidence in the companies at the consumer level, the sentiment polarity might be influenced. Accordingly,

H1a: There is significant difference between the sentiment polarity (positive, negative and neutral) of the three of the major e-commerce websites i.e., Amazon, Flipkart and Snapdeal.

H 1b: More positive sentiments of Flipkart are conferred to Flipkart than to Amazon and people confess that they are using Snapdeal.

H1c: Snapdeal has a greater expression of negative sentiments as compared to Amazon and Flipkart.

2.1.2 Cross Platform Differences on a basis of emotion

In further research, the emotional aspect of the reviews in terms of trust, joy, anticipation, anger, and fear, provides more psychological explanation on what the consumer is feeling (Mohammad & Turney, 2013). The analysis that uses NRC Emotion Lexicon in e-commerce literature has revealed that trust and

joy are the most common emotions in the positive reviews, and anger and fear are the most prevailing emotions when relating to negative user experiences (Thelwall et al., 2010; Medhat et al., 2014).
Due to the focus of Flipkart on the regional offers and promotional campaigns, it is expected to create a feeling of anticipation and surprise within the user. Conversely, Snapdeal can be more frustrating as it does not care about its records of service delivery.
Based on this, we propose the following hypotheses:
H2a**:** Trust and joy are the most frequently expressed emotions across e-commerce app reviews.
H2b**:** Negative emotions such as anger and fear are significantly higher in Snapdeal reviews compared to Amazon and Flipkart.
H2c: Anticipation and surprise are more strongly associated with Flipkart, suggesting better customer engagement.

2.1.3 Relationship Between App Ratings and Sentiment Polarity

The association of review ratings (star values) and textual sentiment is enormous (Liu, 2012; Feldman & Sanger, 2007). The low rating star values normally imply that it is linked with negative sentiment and emotions like disappointment that is mostly characterized by anger or sadness.
This association is used to provide congruency to polarity scores determined in computer models, and provides an interface between quantitative (rating) and qualitative (text) comments.
In such a way, we suggest:
H3a: 5-star ratings would have positive associations with positive sentiment scores.
H3b: 1-star ratings are highly related to having a negative polarity and negative emotions of anger, sadness or disgust.

2.1.4 The position of Service Quality/App usability in Sentiment

This group of hypotheses, which is based on the Technology Acceptance Model (TAM), ties the perception of usefulness and ease of use with the emotional reaction of the users to the application (Davis, 1989). The delivery speed, product authenticity and navigability of apps are some of the elements that may appear frequently in reviews and determine the resultant sentiment (Venkatesh & Davis, 2000).
Negative polarity reviews may point out the existence of failed deliveries, bad packaging, or even the rage of refunds. Conversely, sentiment words such as easy-to-use, fluid interface, fast refund are usually the words that were included in positively graded emotions (Cheung & Thadani, 2012; Sarin et al., 2021).
We, therefore, hypothesize:
H4a: Negative sentiment is more likely to be present in the reviews that refer to the word delivery, customer service, or refund.
H4b: Reviews with the mention of usability, interface, or easy returns are more prone to a positive sentiment and trust.

2.1.5 Linguistic Features and Sentiment Correlation

The theory of sentiment analysis holds that there are some keywords which have tendency to appear together with a particular sentiment class or emotion category. Such is the essence of lexicon-based styles through which lexical counts and emotional enumerations are the foundation blocks of classification (Liu, 2012; Mohammad & Turney, 2013).
This study hopes to obtain insights on what words are most effective in determining the emotion of the consumers after analyzing term frequency and sentiment labels.
It would hence propose the following hypotheses:
H5a: The trust and the joy of sentiment are correlated positively with the frequency of such keywords as good, nice and easy.
H5b: The frequency of such keywords as worst, late and fake has an important positive correlation with negative sentiment polarity.
These group of hypotheses include theoretical logic including behavioral sciences and computational linguistics to investigate the digital consumer sentiment. The fact that both polarity and emotion-based hypotheses are included means that users perceptions will be assessed in a multi-dimensional way. Moreover, it will be linked to the features of platforms, and language patterns, as well as user ratings, which will help explain the wisdom-of-crowds effect more thoroughly. In the following section, the

research methodology and tools will be described through which these hypotheses are going to be empirically tested.

Hypothesis Code	Statement	Construct / Variable	Supporting Theory / Citation
H1a	There is a significant difference in sentiment polarity across Amazon, Flipkart, and Snapdeal.	Platform-wise sentiment polarity (Positive/Negative/Neutral)	Hennig-Thurau et al., 2004; Shivaprasad & Shetty, 2017
H1b	Flipkart receives more positive sentiment expressions compared to Amazon and Snapdeal.	Platform reputation and localized user experience	Kumar, 2015; Chatterjee, 2019
H1c	Snapdeal receives higher negative sentiment expressions compared to Amazon and Flipkart.	Brand trust and perceived risk	Gefen et al., 2003; Featherman & Pavlou, 2003
H2a	Trust and joy are the most frequently expressed emotions across e-commerce app reviews.	Emotion frequency (NRC categories)	Mohammad & Turney, 2013; Thelwall et al., 2010
H2b	Anger and fear are significantly higher in Snapdeal reviews than Amazon and Flipkart.	Negative emotional intensity	Medhat et al., 2014; Liu, 2012
H2c	Anticipation and surprise are more strongly associated with Flipkart.	Customer engagement and satisfaction	Sarin et al., 2021; Chatterjee, 2019
H3a	5-star ratings are positively correlated with positive sentiment polarity.	Ratings vs. polarity	Feldman & Sanger, 2007; Liu, 2012
H3b	1-star ratings are associated with negative sentiment and emotions like anger or sadness.	Rating-emotion mapping	Pang & Lee, 2008; Thelwall et al., 2010
H4a	Reviews mentioning delivery, refund, or customer service are more negative in sentiment.	Service failure keywords vs. sentiment	Cheung & Thadani, 2012; Shivaprasad & Shetty, 2017
H4b	Reviews with terms like interface, usability, or smooth return express more positive sentiment.	Usability keywords vs. sentiment	Davis, 1989; Venkatesh & Davis, 2000
H5a	Keywords like “good,” “nice,” and “easy” correlate with joy and trust sentiments.	Positive lexicon mapping	Mohammad & Turney,

			2013; Liu, 2012
H5b	Keywords like "worst," "late," and "fake" correlate with negative sentiment.	Negative lexicon mapping	Medhat et al., 2014; Thelwall et al., 2010

3 RESEARCH METHODOLOGY

This paper aims at looking into consumer sentiment as found in the app based reviews on the major Indian e-commerce websites of Amazon, Flipkart, and Snapdeal through text mining and sentiment analysis algorithms. In this section, the methodological framework that has been used to conduct the study is outlined, the research design, the data sources, tools and procedures of analysis are outlined, and justification of every method has been provided.

3.1 The Research Design

The study is quantitative, exploratory in nature with a design of computational linguistic analysis of large scale audio-text data of consumer reviews. The main purpose is to examine and contrast the sentiment polarity (positive, negative and neutral) and emotional features (e.g. trust, joy, anger, fear) in the context of three leading e-commerce platforms working in India. Since the form of online reviews cannot be described as structured, text mining and lexicon-based sentiment analysis serve as the main analytical techniques to be applied further.

The study is cross-sectional, and the data to be taken within one week to obtain a standard picture in temporal dynamics of consumer sentiment. The (qualitative) analysis is descriptive and comparative, with the emphasis placed on the difference between the platforms through the visual analytics (word clouds, polarity graph).

4 Data Source

The research is based on publicly collected user-reviews downloaded in the Google Play Store of:

- Flipkart
- Amazon India
- Snapdeal

Reviews posted by consumers at Google Play store present an immense warehouse of unedited natural thoughts. They contain star rating, written reviews, time stamps and, sometimes, emojis and other sentiment carrying markers. This qualifies them to be very appropriate towards sentiment analysis (Liu, 2012; Sarin et al., 2021).

4.1 Period and Size of Sampling

Information was gathered that offered a view of one week. The remaining sample sizes were:

- Flipkart: 20,000 reviews
- Amazon; 4.7 thousand reviews
- Snapdeal: 1041 reviews

The reviews have been scraped with the help of the Google-Play-Scraper Python library, which ensures real-time and an automatic way of grabbing the data. Reviews in any other language other than English was not included to ensure standardization, compatibility with sentiment lexicon.

4.2 Data Preprocessing

The raw review test turned out to be noisy and unstructured and thus preprocessing was important. The procedure adopted is as under:

1. Capitalizing all the text to have a uniformity
2. Cleaning of stop words, punctuations / numbers and special characters i.e. NLTK (Natural Language Toolkit)
3. Tokenization- splitting words up in text.
4. Normalization of word forms by Lemmatization
5. Editing out of blank rows, emojis and HTML characters
6. Text mining in RStudio conversion into corpus format

Data was preprocessed and written in structured format in .csv files to be further analyzed.

4.3 Tools and Technologies Used

• Python (for scraping, cleaning): Pandas, numpy, nltk, google-play-scraper and re libraries were utilised.
• RStudio (text mining and sentiment analysis): Packages to use were tm, syuzhet, wordcloud, sentimentr, ggplot2 and NRClex.

Additional visualizations and tables were created by using Excel and Tableau.

4.4 Sentiment Analysis Procedure

4.4.1 Approach

The proposed study applies lexicon-based sentiment analyses, where one compares words present in the reviews, with a pre-defined dictionary containing sentiment-marked words. The NRC Emotion Lexicon (Mohammad & Turney, 2013) was chosen because it can represent not only the most basic sentiment polarity (positive, negative, neutral) but also the classes of emotions (trust, joy, anger, fear and so on).

4.4.2 Classification of Sentiments

Process was such that each cleaned review was run through the NRC lexicon, categorizing words according to the following 10 emotions:

- Positive
- Negative
- Joy
- Trust
- Anticipation
- Fear
- Anger
- Disgust
- Sadness
- Surprise

This enabled cross-platform emotion distribution mapping enabling deep counterintuitive features beyond binary sentiment labelling.

4.4.3 Aggregation and Seeing

On both platforms the following were created:
Word clouds (first 200 words)

- Comparison clouds (to demonstrate sentiment-heavy terms platforms)
- Common Platter Clouds (communality clouds)
- Polarity boxplots (between the star-ratings)

All these visualizations help to test hypotheses and clear interpretation of sentiment patterns.

5 Stability and reliability

5.1 Reliability of Lexicon

NRC Emotion Lexicon applied in the current research has been tested during various scholarly works (Thelwall et al., 2010; Mohammad & Turney, 2013). The high rate of adoption in digital sentiment analysis ensures good internal consistency of results.

5.2 The process of triangulation

Analysis was done with the help of cross-verification of sentiment trends with the help of multiple tools (Python + RStudio + Tableau). Finding triangulation by statistical and visual analysis enhances the construct validity of the study.

5.3 Justification of Methodology

The sentiment analysis has been selected because first, it has a proven ability to provide the customer opinions at the scale and in real-time, which cannot be duplicated even by traditional methods of conducting surveys or interviews (Medhat et al., 2014).

In addition, publicly available, naturally produced consumer data provide ecological validity, or a guarantee that observed results reflect actual behaviour instead of those distorted by survey effects.

5.4 Ethics Ethics is a factor to consider when building CSI.

The research uses only open user reviews, and no personally identifiably information is collected. The aggregate and anonymized data is analysed, which can be regarded as compliant with the ethical standards of conducting a study with open digital information (Townsend & Wallace, 2016).

5.5 Constraints of the Technique

Only English language reviews were analysed, and this fact can create bias based on language.

• Lexicon based analysis might not be able to understand sarcasm, irony, or slang hence some percentage of the misclassification of sentiment.

The data analyzed takes only one week of the review and can be not representative of seasonal and event-related changes in sentiment.

6 Analysis and Outcomes of Data

In this section, it shows the results of the sentiment analysis of the review of consumers on the mobile apps of Flipkart, Amazon India and Snapdeal. The sentiment polarity, emotional tone, and linguistic frequencies patterns analysis is done followed by hypothesis testing with the help of visualizations.

6.1 Dataset description

The preprocessed data gave the final structure containing:

Flipkart: 20 000 reviews

Amazon: 4700 reviews

Snapdeal: 1041 reviews

Metadatas of each review included star rankings (1-5), text, and the name of the app. The Python (NLTK) was used to clean the reviews, and the RStudio packages were applied: tm, syuzhet, wordcloud, and ggplot2.

Exhaustive text analytic and sentiment analysis of the application reviews to the three E- commerce websites have been undertaken and besides lexicon-based dictionary was also utilised in sentiment analysis which was done with the help of RStudio. Citizens in India are showing interest in e-commerce websites more frequently and it is in light of this, that a social media competitive research of the three most popular e-commerce websites appears on top the agenda.

the case of selection of these businesses was that they not only have a large market share but are also in fashion these days.

6.2 Text Mining:

Text mining is meant to sort out and make sense of rudimentary unvarnished reviews and ratings that can be found on the site. The satisfaction of customers is an entry point to repeated sales and generation of revenue. The raw data approach to interpretation of customer feedback might be too tedious to be undertaken by organizations. The objectives are thus fulfilled through the newest procedures such as text mining. The technique that is applied to know the emotions presented by the customers once the products and services have been accessed is through sentiment analysis. The analysis of the texts and data conducted in this study has adopted R-studio software. R-studio offers domain in an analytical computing, video game interface, pictorial formation, analyze data through star clusters, develop API, convert web application (Simplystats, 2019). R is freely available under GNU general public license; due to the strength of groups working on it, various extensions, which are referred to as packages, have been created over the period, in addition to robust documentation. With this flexibility and capability, R has been continuously popular in the implementation of data and text miner in a variety of fields and it does consist of effective text miner tools (Meyer, D. and Hornik, K. and Feinerer, I., 2008). Various kinds of analysis are also used in the present study to include text analysis, sentiment analysis, comparison and commonality of three files of the data. Sentiment Analysis It is an undertaking that establishes the reaction, opinions, revelations, thinking and feelings of the customers, in short what do the customers feel after buying the product were they happy or sad. What kind of emotions were theirs, did they possess negative, clumsy and alarming emotions or they were neutral. In this sentiment analysis, various emotions of the users of all the four companies were derived and studied through using RStudio software and inbuilt packages. Depending on the results of the sentiment analysis in the RStudio tool, the text was divided into ten categories of emotions listed above. The sentiment analysis is carried out solely to analyse the information retrieved (in our case it is the reviews of the customers of the two different websites (Flipkart, Amazon) and to determine the reaction, attitude, context and emotion of the customers. When I wanted to work on sentiment analysis, there are basically three ways and those are:- working on individual file, working on group of files, working on corpus or term document matrix. To operate individual file meanr RSentiment and sentiment package is needed and when multiple files then you will have to use syuzhet

package and to operate corpus you will need Sentiment Analysis package. In order to get the results of the basic score of positive, negative wordcount we used (meanr) package of RStudio. The results of combined files which is the customer reviews in e-commerce websites (Flipkart, Amazon, snapdeal) is used to determine the emotions the customer is feeling using the method of sentiment analysis. It can be said that it involves classifying the text into three forms of sentiments i.e., positive, negative or neutral. Textual data is usually used to perform sentiment analysis. It also aids the businesses to review the defects on the product and review the needs of the consumers (Shivaprasad, T. K., & Shetty, J., 2017). The analysts are able to analyze the public feelings regarding the products and services. It has superseded the traditional approach of surveys that were prevalently done by the organizations to determine the views and articulations of the products and services that the customer portrays. It assists people to have a close overview of the feelings of previous customers before the buying.

6.3 Comparison Cloud:

Comparison cloud is a unique mapping of the intertwining of both the database texts. We did this by the RStudio. The creation of the comparison clouds followed the process as mentioned below. The text was scraped in all of the three of the e- commerce sites i.e., Amazon, Flipkart and Snapdeal and the same was converted into a pdf file using the pdf convertors and the corpus was generated to the same. To begin with, the corpus got purified of the stop words, emoticons, numeric digits, punctuations and other of such kind. Then a cleaned up corpus again was produced and on the basis of this term document matrix was produced. The cloud indicates that the customers possess a good and nice intention when using the flipkart application as opposed to others.

Fig 6.1: Word Cloud

6.4 Commonality Cloud:

Commonality cloud Commonality cloud is used to summarize the most frequently appearing words in all the three of the text files of Amazon, Flipkart and snapdeal. The commonality cloud provides the pair of identical words in corpus. From Error! Ancillary material is missing. it can be determined that the most frequently used words are the one of the words, namely, service, order, and shopping in both situations. It also displays words such as worst which are the similar words in the data retrieved. Thus, it displays the ambivalent responses.

	WORD	FREQ
1	good	6076
2	nice	3222
3	app	2691
4	flipka	1664
5	super	1240
6	product	953
7	best	909
8	service	900
9	delivery	818
10	customer	611
11	order	608
12	shopping	606
13	worst	568
14	bad	498
15	time	453

Fig 6.2: Word Cloud (Commonality Cloud)

The top 15 words of Flipkart are indicated in the above image and the words like; good, nice and super show the positiveness. It also comes out that the customers are satisfied after consuming this specific product.

	WORD	FREQ
1	app	1224
2	good	1030
3	amazon	807
4	nice	526
5	service	347
6	delivery	324
7	worst	308
8	product	270
9	pay	261
10	best	256
11	bad	243
12	customer	243
13	order	238
14	time	216
15	shopping	192

Fig 6.3: Word Cloud-Amazon

The above figure shows the word clouds of amazon application and the top 200 words are displayed in it. The words such as “good”, “nice” depicts the positiveness.

The above image shows the top 15 words of Amazon and the words such as “good”, “nice”, “super” depict the positiveness. The word “worst” depicts the negative emotions. It also reveals that the consumers are happy and might be disappointed as well after using this this application.

	WORD	FREQ
1	good	364
2	app	293
3	nice	145
4	product	126
5	snapdeal	126
6	best	95
7	delivery	73
8	service	70
9	quality	69
10	products	68
11	shopping	67
12	order	51
13	customer	47
14	experience	45
15	return	44

Fig 6.4: Word Cloud- Flipkart

The above figure shows the word clouds of Flipkart application and the top 200 words are displayed in it. The words such as “good”, “nice” depicts the positiveness. The above image shows the top 15 words of snapdeal and the words such as “good”, “nice”, “super” depict the positiveness. The customer might had great experience after using this application.

	WORD	FREQ
1	app	1224
2	good	1030
3	amazon	807
4	nice	526
5	service	347
6	delivery	324
7	worst	308
8	product	270
9	pay	261
10	best	256
11	bad	243
12	customer	243
13	order	238
14	time	216
15	shopping	192

Fig 6.5: Word Cloud Snapdeal

The above figure shows the word clouds of Snapdeal application and the top 200 words are displayed in it. The words such as “good”, “nice” depicts the positiveness.

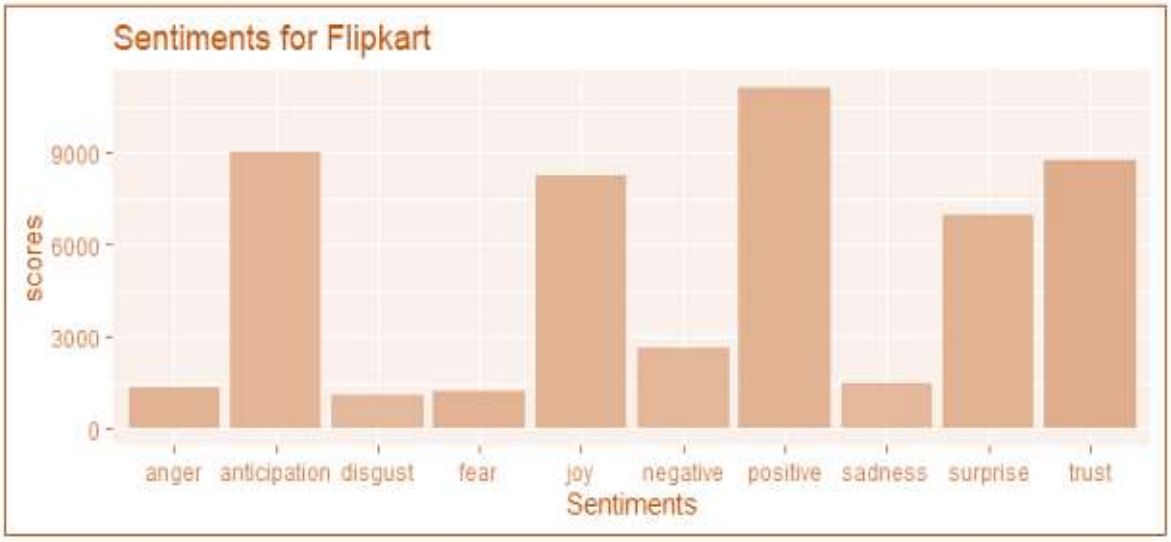

Fig 6.6: Sentiments for Flipkart

The figure 6.6 justifies the sentiment analysis of the app reviews that were produced through the play store on the Flipkart application. The feeling expressed in figure is used to describe a positive expression and expectation as is also portrayed by the shoppers of the online sites. The positive emotions are the ones that register the highest sentiment points but in this order, the positive emotions are anticipation and then the trust and Joy. The facial expressions shown by the customers in the reviews are pleasant ones expressed by the users. The emotions have been sorted out through NRC emotion lexicon based.

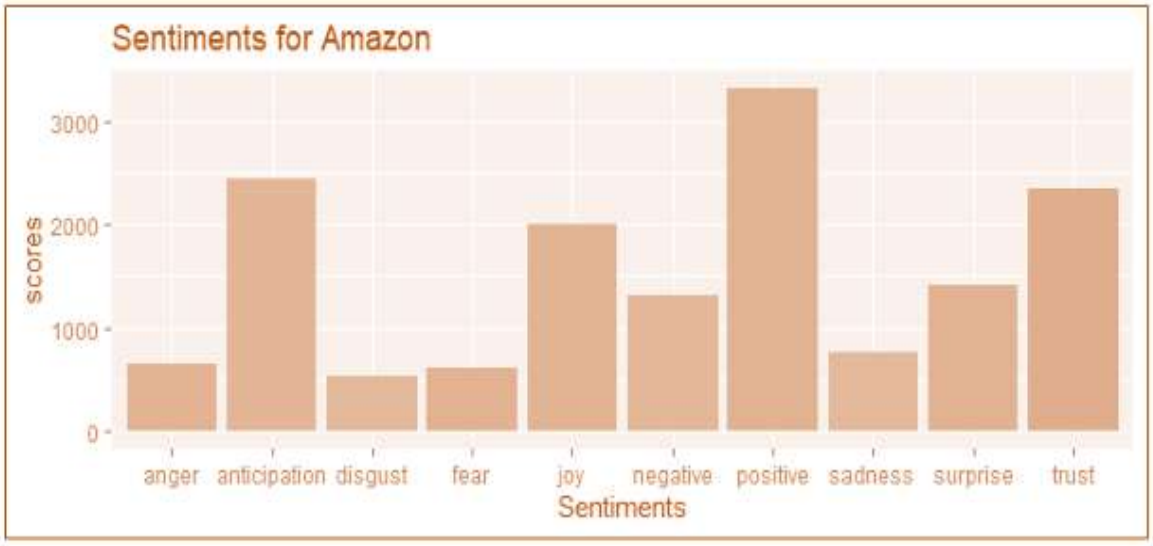

Fig 6.7: Sentiments for Amazon

The figure 6.7 describes the sentiment analysis of the app reviews that are produced by the play store in regard to the Amazon application. The feelings in figure describe a good expression and expectancy as also exhibited by the clients of the e-trading websites. The sentiments with the most score is the positive ones, anticipation and then the trust and Joy. The responses exhibited by the customers in the reviews present an enjoyable facial expression by the users. The sentiments have been categorized through the

application of NRC based emotion lexicon.

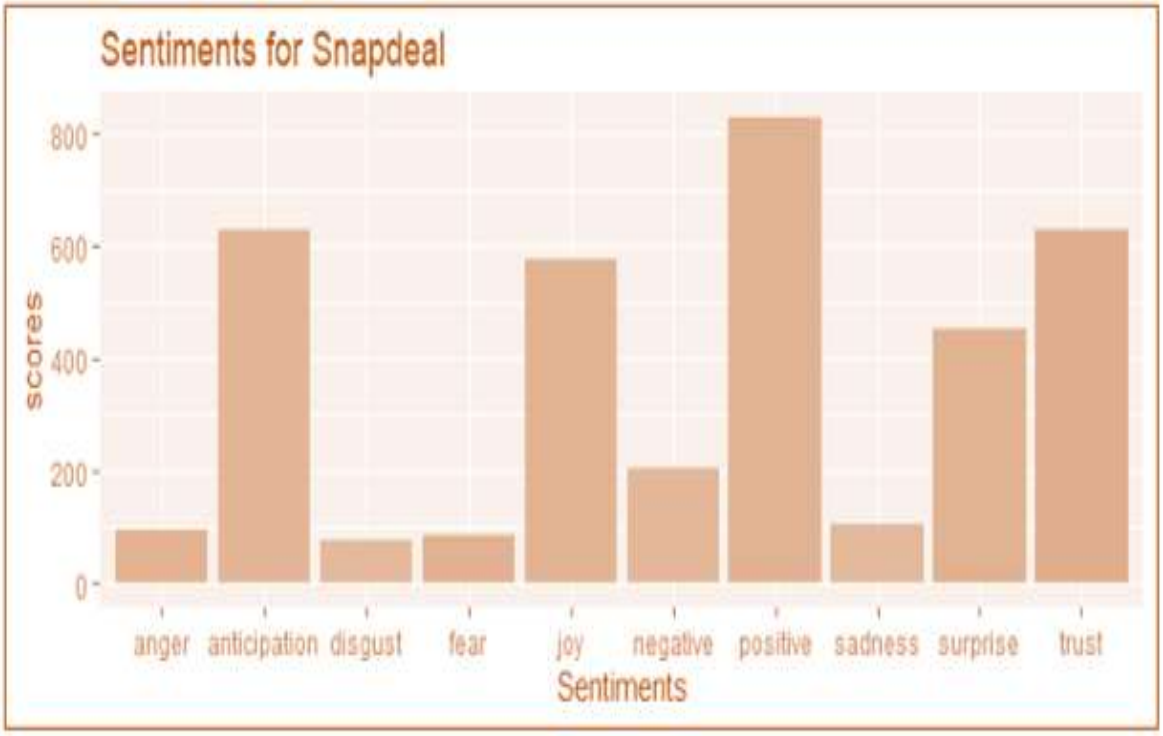

Fig 6.8: Sentiments for Snapdeal

The figure above explains the sentiment analysis of the app reviews generated by the play store of the Snapdeal application. The feeling that is portrayed in figure tells us about a feeling of positivity in expression which is also portrayed by the customers of the e-commerce websites as far as eagerness to anticipate is concerned. Emotions that exhibit the highest sentiment scores are individualistic and they are positive, anticipation, the trust and Joy respectively. The personality that is exhibited by customers on the reviews presents a good face by the users. Using NRC emotion lexicon based sentiments have been classified.

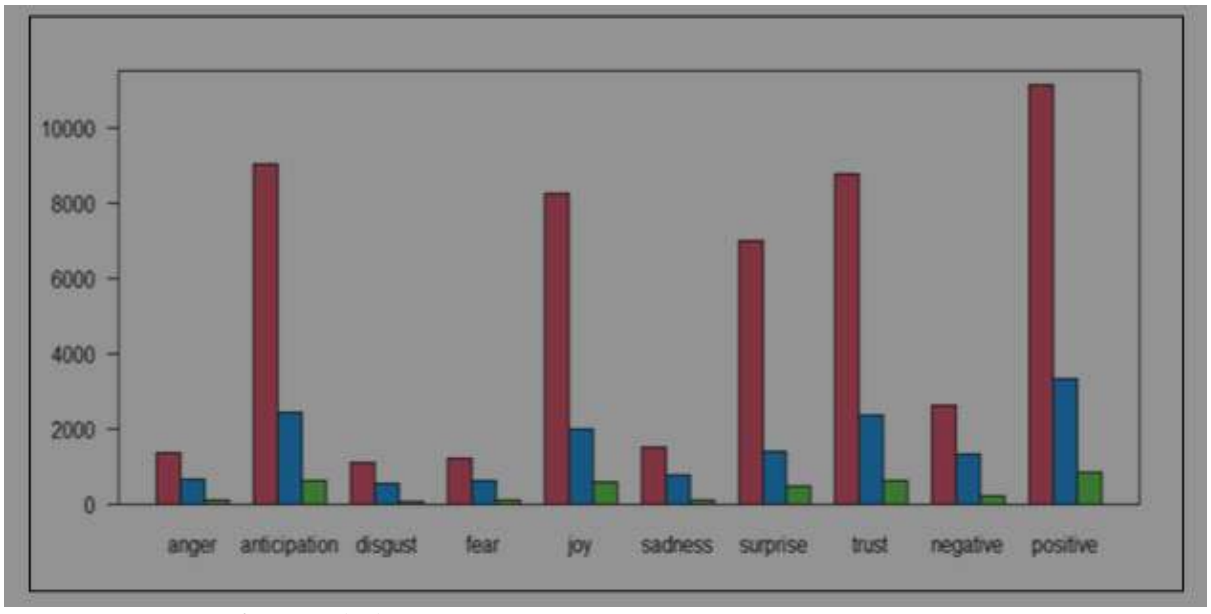

Fig 6.9: Sentiments for Snapdeal

The figure 6.9 denotes the collective emotions of all the food delivery apps. Through these graphs, it can be seen that the customers have exhibited positive emotions especially compared to the negative emotions. Sentiment analysis will aid in explaining the sentiments of what the customers feel concerning the

product. It also provides a good picture of what is in the heads of the customers. The figure above represents the unambiguous reflection of the mood of all the three e-commerce websites. Upon looking at the above image, it is immediate that Flipkart has the greatest numbers of positive sentiments, as compared to others. Among these two the users that used Amazon had superior sentiments as compared to snapdeal. In comparison with others, the negative sentiments of the snapdeal were the least and flipkart has the utmost positive sentiments. Out of these two the consumers who were utilizing flipkart expressed better sentiments than what they would have used against amazon. The negative sentiments of the snapdeal were the least.

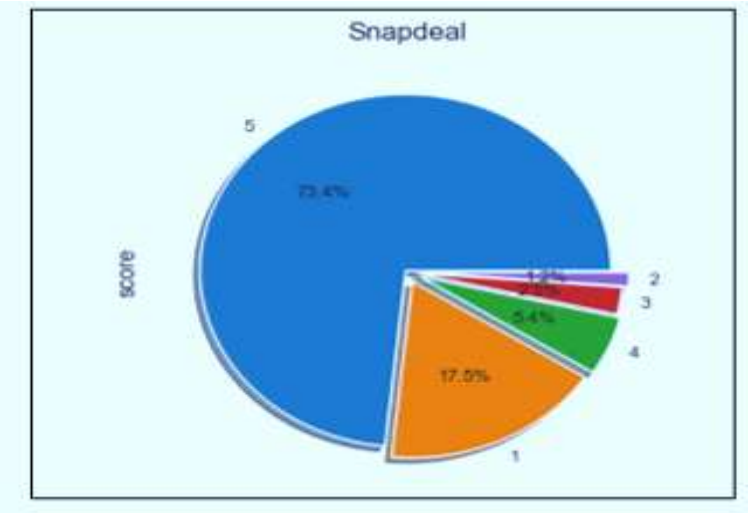

Fig 6.10: Pie Chart for Snapdeal

The above pie chart displays the pictorial representation of snapdeal and the maximum rating of 5 star is given by most of the people i.e. 73.4% people and 1.2% of the people have given 2 star rating.

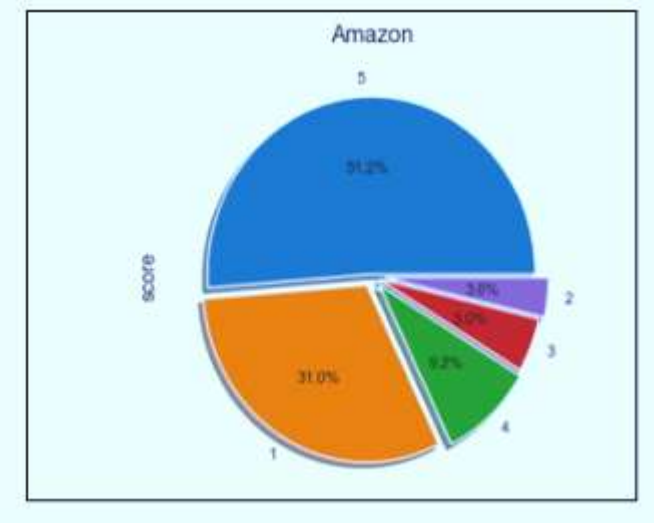

Fig 6.10: Pie Chart for Amazon

The above pie chart displays the pictorial representation of amazon as an e commerce website and the maximum rating of 5 star is given by 51.2% people and 1 star rating is given by approximately 31% people. The minimum rating is given 2 star by 3.6% people.

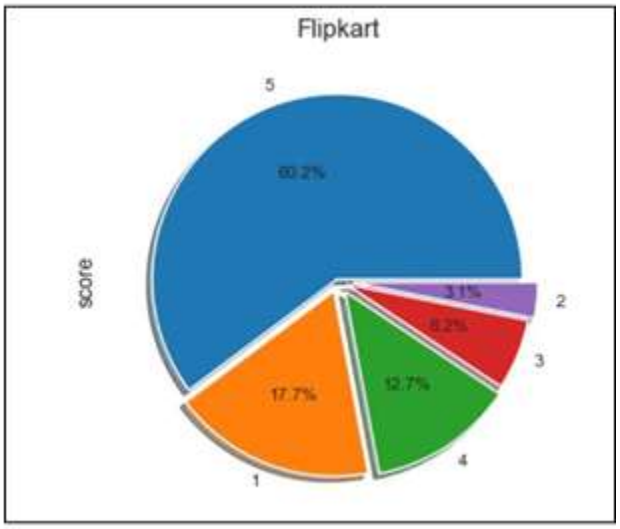

Fig 6.12: Pie Chart for Amazon

The above pie chart displays the pictorial representation of flipkart as an e commerce website and the maximum rating of 5 star is given by 60.2% people and 1 star rating is given by approximately 17.7% people. The minimum rating is given 2 star by 3.1% people.

Below is the table shown which reveals the ratings ranging from 5 star to 1 star.

Name of the organisation	5 Star	4 Star	3 Star	2 Star	1 Star
Flipkart	60.2	12.7	6.2	3.1	17.7
Amazon	51.2	3.6	5.0	9.2	31.0
Snapdeal	73.4	1.2	2.5	5.4	17.5

Table 6.1: Ratings

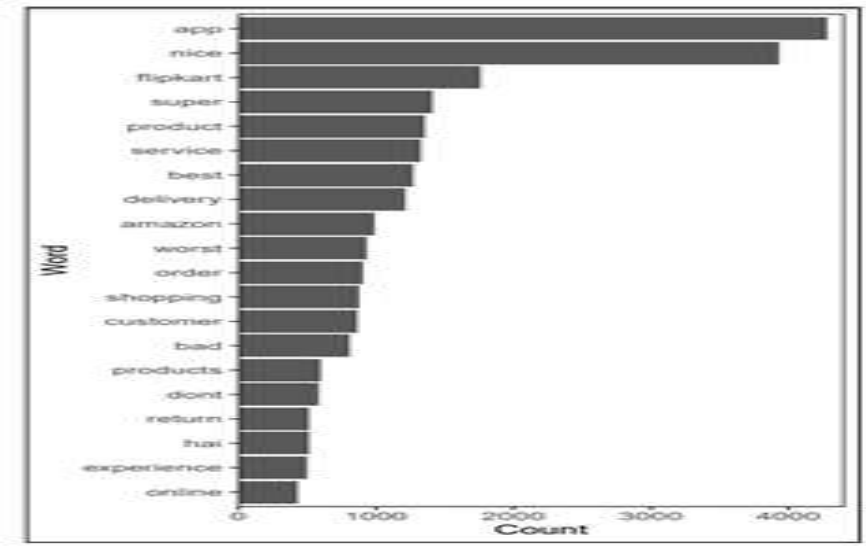

Fig 6.11: Word Count

These were the words with maximum count and the word that had the maximum count was the word app that was approximately 4000. Compiling the list of the occurrence of the most common phrases or words may give a foresight of an overt subject. These terms have been discovered in reviews of the three e-commerce websites and are the most common ones. Based on the above figure, it is apparent that word like app or app has the largest number of how many times they are used in review given.

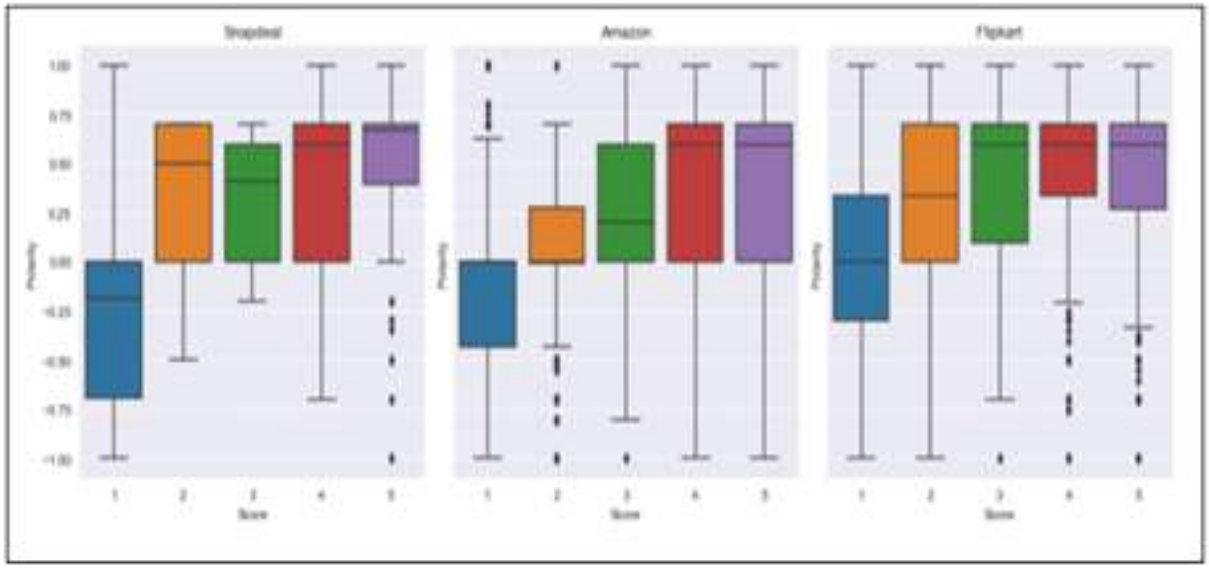

Fig 6.12: Box Plot

The box plots shown above, make it apparent that there is a greater degree of negativity of polarity in snapdeal with 1 star rating, and that the 1 star rating is the most positive with flipkart application. However, it has plenty of outliers in the 5-star rating of flipkart and snapdeal in comparison to the other and there are no outliers in case of amazon.

Social media is receiving attention by many sectors these days. It will also help to comprehend how social media data may be applied in making the significant strategic decisions on the industry level. There are a number of companies, which used social media information of large companies and performed competitive analysis. In this study, the researcher will rely on the competitive analysis, by using text mining of the data generated by the consumers on Appstore and among three e-commerce websites namely:- Amazon, flipkart, Snapdeal. This study suggests that even customer perception can be different to various people and using of e-commerce websites can also be according to the wish of the customers who prefer using the e-commerce websites due to the promotional campaigns and offerings of the applications.

Other than this the consumers also utilize e-commerce websites because the primary reason to utilize it is the fact that e-commerce has an easy and quick delivery and this can be achieved at a fair price. The e-commerce site should also be characterized by the good service support or the customer care.

• The other fundamental functionality that a e-commerce application should have is the seasonal offers particularly during some festivals like Diwali, Holi and some special thigs they should offer in big numbers.

Another finding in this study is that a client would prefer the method according to which he or she receives the highest savings, e.g. cash on delivery, net banking. They only buy the ones that fit their budget and normally they have ordered once every 30 days but there exist some clients who desire to order regularly.

• On the three apps, a combined sentiment analysis was conducted utilising two distinct Lexicon-based sentiment classification algorithms, and diverse sentiments were obtained.

• According to the comparative analysis, the companies' entire performance was geared toward users' good emotions. Flipkart, on the other hand, received far more positive feedback than the other two companies.

• The findings show that all three organizations were using social media to engage with their customers. For flipkart there were 28 Million reviews and the average rating was of 4.3Star, and 100 million plus was downloads was there. For Snapdeal there were 100. million plus downloads and the average rating was 4.5 star, but the total reviews for this was 2 Milllion only. For Amazon as an e-commerce website there were 100 million plus downloads and the average rating was of 4.2 star and there were 8 million reviews in total.

7. CONCLUSION AND DISCUSSIONS:

Social media is catching the attention of many sectors now. It is important to realize the application of social media data in putting forward strategic decisions at large levels of the industry. Competitive analysis on social media has been done using the data available in large companies by a couple of studies. In this research, the data that is produced by the consumers on Appstore and on three E-commerce websites:- Amazon, flipkart, Snapdeal has been used through text mining in terms of competitive analysis. As suggested in this study, customer attitude towards the e-commerce website differs among different individuals, and mostly the consumers are willing to use e-commerce websites due to promotions and campaigns made by the applications.

Other than this, the consumers also used e-commerce websites as the primary reason was easy and fast delivery and moreover at reasonable rate. E-commerce website should also have an attribute of good service support or the customer care.

• Another minimum requirement that any e-commerce application should also have is the seasonal deals especially at few occasions like Diwali, Holi and there should be great offers on few special things.

• Another finding that is revealed by this study is that clients prefer the one that offers him the best savings e.g. cash on delivery, net banking etc. They buy the entities that do not exceed their budget and they order once in a month but there are some clients who wish to order regularly.

• A cumulative sentiment analysis was performed on the three apps by use of two separate Lexicon-based sentiment classification algorithms and different sentiments were derived.

As compared in the analysis, the total performance of the companies was based on the good emotions of users. Compared to the two other companies, however, Flipkart was also the target of much more positive feedback.

• Based on their findings, it shows that all the three organizations were deployed to use social media to reach out to their customers. In flipkart 28 Million other reviews were there and the average rating there is of 4.3Star and 100 million or more was there in the number of downloads. In the case of Snapdeal, it had 100. million plus downloads, and the average rating was 4.5 star, however, the overall ratings had only 2 Milllion. In the case of Amazon which is an e-commerce site had a download of more than 100 million and the average rating is 4.2 star and there are 8 million reviews altogether.

8. RECOMMENDATIONS and LIMITATION '

Future studies could also take a place in different seasons or years in order to assess broader perceptions of food delivery applications.

Second, the social media analytics tools like text mining, frequency analysis of words, and sentiment analysis were used to investigate some of the most interesting trends in social media communications (Bruns, 2012; Pang and Lee, 2008).

Future studies may apply other methods to figure out the most powerful Twitter accounts, including network mapping analysis; regression analysis may be potentially used to locate antecedent features that permit the dissemination of information. Trivedi, S. K., Singh, A., 2021).

• The other aspect that is possible to work on this study is the absence of consideration of the location of study. Social media may be location-specific in the future, and consequences may depend on the location of data.

• This study only considered reviews carried out in English. Future studies may however analyze other languages.

• Make use of opinion bearers such as emoticons, emojis as well as slangs in enhancing the material.

These other social media may also be used like Instagram, twitter, facebook and twitter analysis may also be carried. Explain how a social media competitive analysis is important and illustrate the value and competitiveness of text mining and sentiment analysis success to a business.

The suggestions, 37 business, and research implications can be used by organisations in helping them develop a social media competitive analysis plan.

REFERENCES:

1. Chatterjee, S. (2019). The impact of eWOM on consumer purchase intention. *Journal of Retailing and Consumer Services*, 50, 356–364.
2. Cheung, C. M., & Thadani, D. R. (2012). The impact of electronic word-of-mouth. *Internet Research*, 22(4), 461–480.

3. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340.
4. Feldman, R., & Sanger, J. (2007). *The text mining handbook*. Cambridge University Press.
5. Featherman, M. S., & Pavlou, P. A. (2003). Predicting e-services adoption. *Decision Sciences*, 34(2), 371–410.
6. Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping. *MIS Quarterly*, 27(1), 51–90.
7. Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth. *Journal of Interactive Marketing*, 18(1), 38–52.
8. Kumar, N. (2015). E-commerce trends in India. *International Journal of Management*, 6(1), 45–52.
9. Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.
10. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113.
11. Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3), 436–465.
12. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135.
13. Sarin, P., Kar, A. K., & Ilavarasan, V. P. (2021). Understanding app-based consumer reviews through sentiment analysis. *Information Systems Frontiers*, 23, 1457–1473.
14. Shivaprasad, T. K., & Shetty, J. (2017). Review-based sentiment analysis for e-commerce platforms. *International Journal of Advanced Research in Computer Science*, 8(5), 95–99.
15. Thelwall, M., Buckley, K., & Paltoglou, G. (2010). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 61(12), 2544–2558.
16. Venkatesh, V., & Davis, F. D. (2000). Theoretical extension of the Technology Acceptance Model. *Management Science*, 46(2), 186–204.
17. Chatterjee, S. (2019). The impact of eWOM on consumer purchase intention: A study on Indian online retail sector. *Journal of Retailing and Consumer Services*, *50*, 356–364. https://doi.org/10.1016/j.jretconser.2018.07.008
18. Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge University Press.
19. India Brand Equity Foundation (IBEF). (2023). *E-commerce industry report*. https://www.ibef.org/industry/ecommerce
20. Jiménez, F. R., & Mendoza, N. A. (2013). Too popular to ignore: The influence of online reviews on product choice. *Journal of Interactive Marketing*, *27*(3), 226–235.
21. Khalid, W., Ali, R., Khan, M. A., & Hameed, I. (2021). Determinants of sustainable energy technology adoption: A behavioral reasoning theory approach. *Energy Reports*, 7, 799–812.
22. Kumar, N. (2015). E-commerce trends in India: Current scenario and future predictions. *International Journal of Management*, 6(1), 45–52.
23. Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.
24. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113.
25. Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, *29*(3), 436–465.
26. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*(1–2), 1–135.
27. Sarin, P., Kar, A. K., & Ilavarasan, V. P. (2021). Understanding app-based consumer reviews through sentiment analysis. *Information Systems Frontiers*, *23*, 1457–1473.
28. Shivaprasad, T. K., & Shetty, J. (2017). Review-based sentiment analysis for e-commerce platforms. *International Journal of Advanced Research in Computer Science*, *8*(5), 95–99.
29. Thelwall, M., Buckley, K., & Paltoglou, G. (2010). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, *61*(12), 2544–2558.