



HAL
open science

Text Mining Amazon Mobile Phone Reviews: Insight from an Exploratory Data Analysis

S. Suhasini, Vallidevi Krishnamurthy

► **To cite this version:**

S. Suhasini, Vallidevi Krishnamurthy. Text Mining Amazon Mobile Phone Reviews: Insight from an Exploratory Data Analysis. 5th International Conference on Computational Intelligence in Data Science (ICCIDS), Mar 2022, Virtual, India. pp.345-355, 10.1007/978-3-031-16364-7_27 . hal-04381290

HAL Id: hal-04381290

<https://inria.hal.science/hal-04381290v1>

Submitted on 9 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Text Mining Amazon Mobile Phone Reviews: Insight from an Exploratory Data Analysis

Suhasini S.^[1], Vallidevi Krishnamurthy^[2],
Research Scholar^[1], Associate Professor^[2],
Department of Computer Science and Engineering^{[1][2]},
Sri Sivasubramaniya Nadar College of Engineering^{[1][2]}.
suhasinis@ssn.edu.in^[1], vallidevik@ssn.edu.in^[2]

Abstract:

Online shopping is used widely by all the people across the world. Only due to that the sites of e-commerce is receiving more review about the products which includes positive, negative and neutral reviews about the product purchased by the people. Those reviews would help the peer customer to know about the quality of the product and to suggest the best one to purchase. Similarly, this paper is focused on analyzing the reviews of the mobile phone purchased on the amazon website and gives a graphical analysis of all reviews given after each purchase of the mobile phone and this analysis could help the customers to easily identify the best mobile phone. In future such analysis would be helpful for the amazon mobile phone buyers to make a better purchase. The exploratory data analysis is made with amazon review dataset which visualizes the review data and analysis of star rating, features rating of each mobile phones. The overall review of the product is also mapped using EDA.

Keywords: Exploratory Data Analysis (EDA), Online Reviews, Sentiment Analysis, Star Rating.

1. Introduction

Online reviews have become useful for people to make their purchase in online. As people are showing more interest in online shopping, to make a useful online shopping the reviews are collected from every customers for each product which would be make the purchase increase or decrease depending upon the review given by the previous customer. In fact, clients frequently simply demand a modest number of positive ratings. Some internet retailers have systems in place to detect which evaluations are the most beneficial to customers.

Although online reviews have aided consumers in determining the benefits and drawbacks of various products, ultimately assisting in the selection of the best product for an individual's needs, they also present a challenge in terms of analyzing this massive amount of data. The review data is rapidly growing in size day by day. Some users have begin to include images of the product in their reviews in order to make them more appealing and user-friendly. As a result, the review dataset could be viewed as a big data analytics challenge. Businesses can be interested in digging into the review data to learn more about their product.

To handle huge amount of review data, the prediction analysis techniques is used which would directly fetch the over analysis of the reviews given by the prior customers[8][10]. The analysis which gives the overall positivity and negativity of the product instead of reading each review. This prediction analysis method will save the computation time for finding the review of the product.

Customers can grade or score their customer experience and happiness with a product on most ecommerce platforms. One of the key aspects that determines online customers' purchasing behavior is product ratings and reviews. According to research, not all online consumer evaluations are helpful and have an impact on purchasing decisions. Consumers who perceive online reviews to be useful have a bigger influence on purchasing decisions than those who do not. Previous research on online consumer reviews had focused on quantitative aspects such as frequent purchase and helpfulness of each brand[2][7][10].

First and foremost, consumers receive firsthand information from their peers about a product, allowing them to make an informed judgement about the product or service. However, due to the large number of reviews, they are overburdened with knowledge on the reviews. It is impossible for any customer to read all of the reviews before making a decision.

Reading all the reviews points is not possible and it may take more time and also we should consider all the review commands to select the best product. To overcome this complexity, an data analysis technique can be done with the amazon dataset which will process all the review commands and gives an overall statistical representation of product, with that representation the best product can be easily identified. Based on the different features of the mobile phones the analysis can be made. This paper focuses on the Exploratory Data Analysis(EDA) for the statistical prediction.

2. Literature Review:

Few works has conducted research into both customer and seller feedback. Customers research a variety of factors before making a purchase, including customer reviews, seller feedback, pricing comparisons with other marketplaces, and so on. All of these factors were employed in this study to estimate customers' willingness to pay using a conceptual model based on consumer uncertainty[1]. Unlike previous researchers who have studied online shopping reviews, and few people conducted study on online travel reviews. Consumer produced reviews were found to have a greater influence on travel decisions such as where to go, what to eat, where to stay, and what to see than ideas offered by the travel and tourism agency. They looked into the relationship between the review text's qualitative qualities, review helpfulness, and the impact of review helpfulness on the review score. They used a dataset of 37,221 reviews acquired from Amazon UK to validate their model. They also used intergroup comparisons to look into the interrelationships between extremely helpful and unhelpful reviews, as well as utterly favorable and negative reviews[3]. Instead of making these comparison we can have a pictorial representation which includes comparison as well as it gives an highly purchased product. Similarly, the content-based and source-based review elements that were examined have a direct impact on the usefulness of product reviews. Customer written reviews with less content abstractness and good comprehension were also shown to be more useful[2]. One work looked into the relationship between the review text's qualitative qualities, review helpfulness, and the impact of review helpfulness on the review score. They used a dataset of 37,221 reviews acquired from Amazon to validate their model. They also used intergroup comparisons to look into the interrelationships between extremely helpful and unhelpful reviews, as well as utterly favorable and negative reviews[4].Some research focused on only the limited volume of evaluations, the business's reputation may shift too frequently in the early years, according to the findings. For a longer period of time, there may be a modest change in a company's reputation. When opposed to a business that has an unstable start, a business with a steady start

is more likely to preserve and strengthen its reputation later on[5]. They made the effect of the emotional intensity index on the final review ranking which is not substantial, according to a comparison analysis. The emotional intensity index can be selected based on the business environment. The sensitivity analysis of the parameters reveals that the loss is not only the attenuation coefficient, it also ensure that the parameters are correct, but it can verify that the parameters are correct loss avoidance psychology of the decision maker. But ensure that the review ordering is somewhat stable in a when the parameters vary, a specific range[6].

They addressed the problem on handling large amount review which confuses the frequent buyers who tend to look at the previous experience. This is handled by ranking the reviews provided in websites in order to improve the experience of the customers. The process is done by creating a CNN-GRU model[7]. Handling of large amount of review can be made easier by using EDA, which shows all the reviews in a graphical format.

3. Dataset Description:

The goal of this study is to discover the factors that influence the usefulness of internet reviews. The dataset was collected from the amazon website using web mining technique. During August 2020, Amazon.in collected 73,151 online reviews for various smart phone brands such as Oneplus, Apple, Vivo, Mi, Redmi, Honor, Samsung and Oppo, with only 12,389 reviews being valid. It is because of the large number of zero counts in the whole sample, the analysis was limited to 12,389 online reviews.

The collected dataset was cleaned through the pre-processing tasks using the python programming library Pandas. The reviews of the respective mobile phone brands were processed and stored in a worksheet. The length of the review is also calculated with help of WordNet which count the number of words used in the review post.

Using the dataset, the sentiment analysis can also be made with three categories of sentiment as positive, negative and neutral.

4. Methodology:

A statistical method is identified to measure the overall review of the product. Using the exploratory data analytics and data visualization method. EDA (Exploratory Data Analysis) is a data analysis approach that uses visual methods to summarize data. The primary focus of EDA is on examining what data can tell us in addition to formal models or the hypotheses are being tested. Initial Data is not the same as EDA. IDA is a type of analysis that focuses on confirming assumptions. It is crucial for the model and hypothesis evaluation, as well as resolving missing values and changing variables in the suitable manner where, IDA is included in EDA. The Bag-of-Words model is a technique for extracting textual features. It's a textual representation of the situation. The occurrence of words in a corpus of text There are two of things: a lexicon of well-known terms and the measuring of the presence of words that are well-known. It is called a "bag" because it contains everything. There is no information on the arrangement of the words or their meaning ignored.

5. Implementation

By applying the EDA, amazon review dataset was analyzed using data visualization libraries in python. The various visualizations for the amazon review dataset are shown below.

5.1 Trustworthiness of online purchase of mobile phones:

Initially, the overall rating of the online purchase of a mobile phone is visualized. **Figure 1:** illustrates the distribution of rating given by the different customers who have purchased mobile phones through Amazon. It is inferred from the Figure 1 that the online purchase of the mobile phone is recommended, as, more than 6000 customers have given the star rating as 5 with their highest satisfaction of online purchase. Further through the sentiment analysis the positivity level of the review comments given for the online purchase of mobile phones is identified as 66.73%. Similarly, the negativity level and the neutral levels for the online purchase of mobile phones is identified as 8.08% and 25.07% respectively.

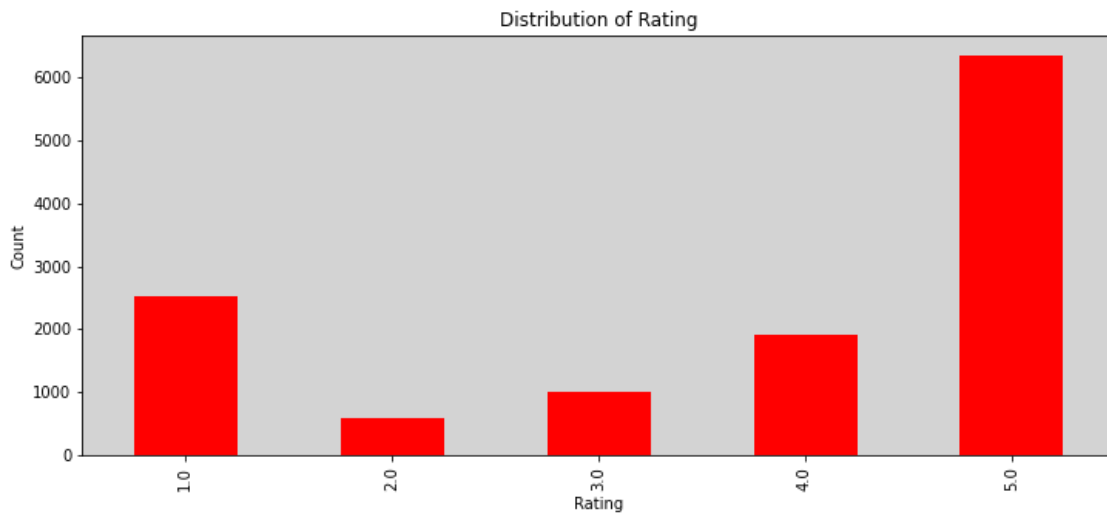


Figure 1: Overall Distribution of Star Rating for Online Mobile Phone Purchase

5.2 Reviews available for the different brands of mobile phones:

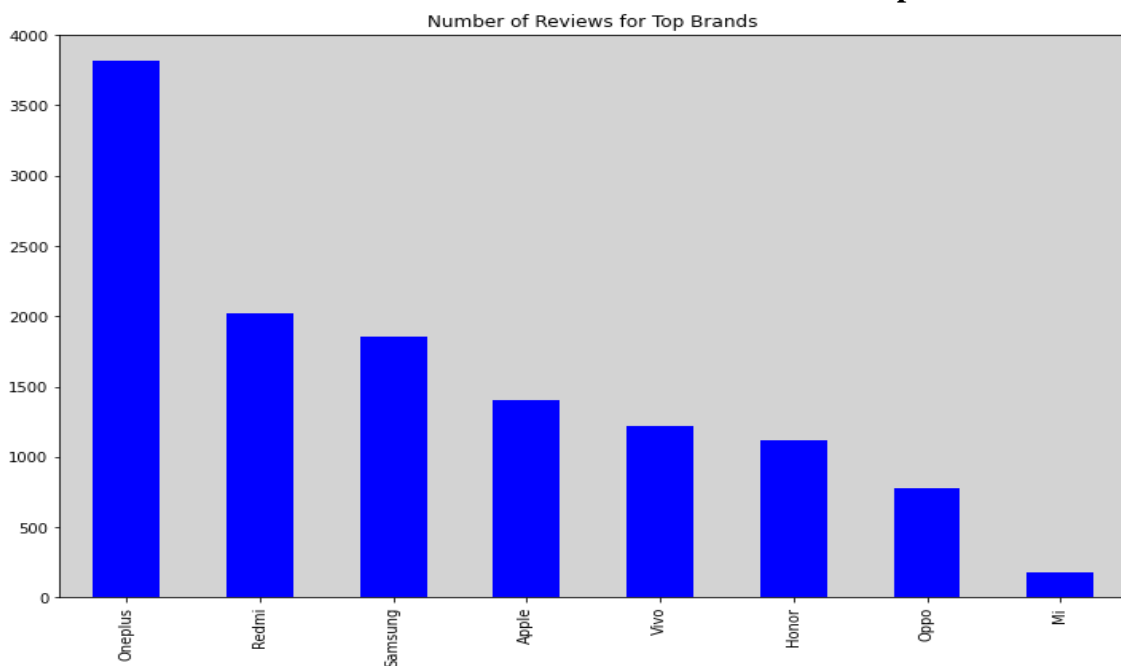


Figure 2: Distribution of Frequent purchase of Mobile Phone Brand

In **Figure 2**, it visualizes the number of reviews given to each mobile phone brands by the customers irrespective of positive, negative, and neutral reviews. Based on that it is observed that the frequently purchased mobile is Oneplus with more than 3500 reviews by the customers. **Table 1**, represents the ranking for each mobile phone brand based on the no of review comments given by customers.

Table 1: Ranking for each mobile phone brands

Brands	Ranking
Oneplus	Rank 1 (Reviews were given by more than 3500 customers)
Redmi	Rank 2 (Reviews were given by more than 1800 customers)
Samsung	Rank 3 (Reviews were given by more than 1500 customers)
Apple	Rank 4 (Reviews were given by more than 1300 customers)
Vivo	Rank 5 (Reviews were given by more than 1200 customers)
Honor	Rank 6 (Reviews were given by more than 1000 customers)
Oppo	Rank 7 (Reviews were given by more than 500 customers)
Mi	Rank 8 (Reviews were given by more than 50 customers)

5.3 Comparison of different features of various brands:

Apart from the above analysis, few customers were interested in good camera quality, and few may concentrate only on battery life, etc. To satisfy those customers, the below graphs were visualized by comparing with different features of various smart phone brands. This visual analysis would help the customer for easy purchase based on unique requirement. **Figure 3** represents the analysis of battery life for each mobile phone brand and the analysis is made based on the star rating given by the previous buyers. Similarly, **Figure 4, 5 & 6** represents the star rating analysis of picture quality, sound quality and camera quality with respect to the various mobile brands. This star rating analysis would help the customers to buy the mobile phone with their unique feature requirement.

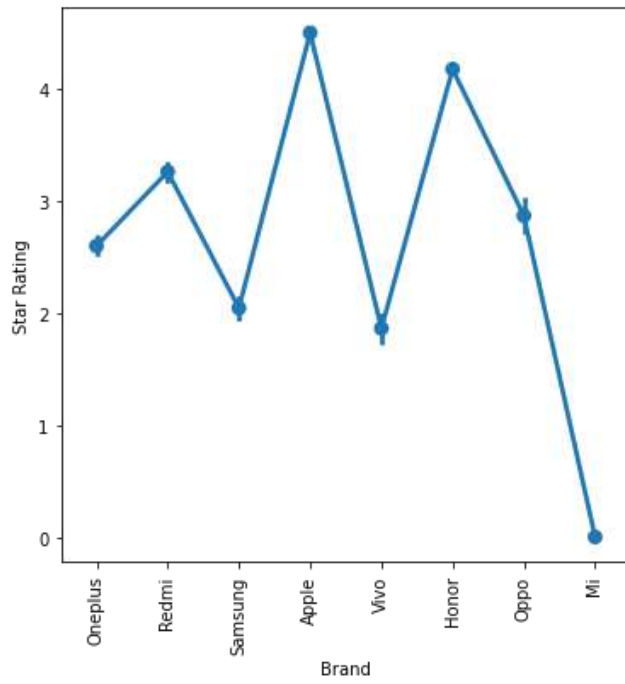


Figure 3: Analysis of battery capacity of various bands

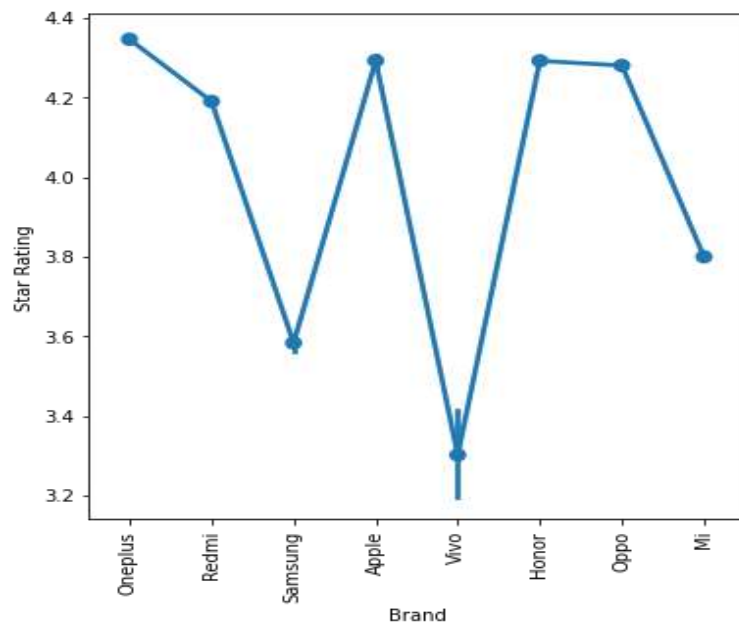


Figure 4: Analysis of Picture quality of various brands

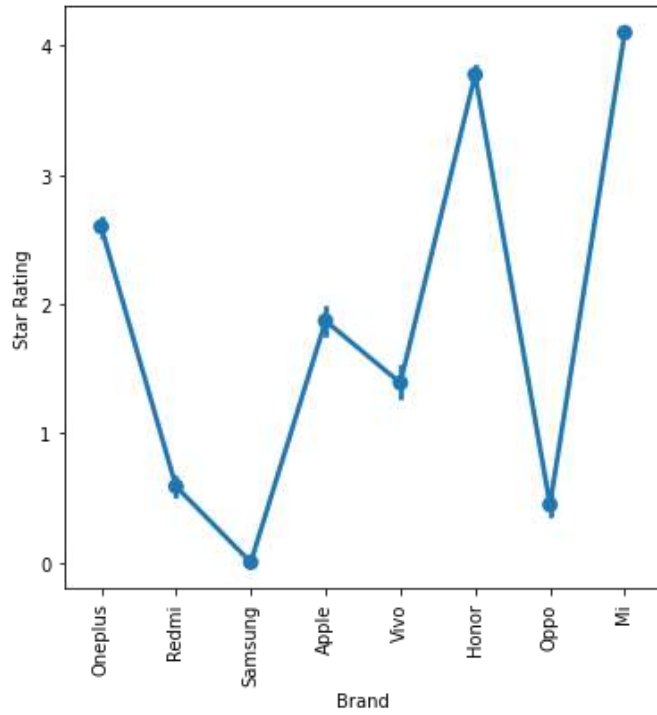


Figure 5: Analysis of Sound quality of various brands

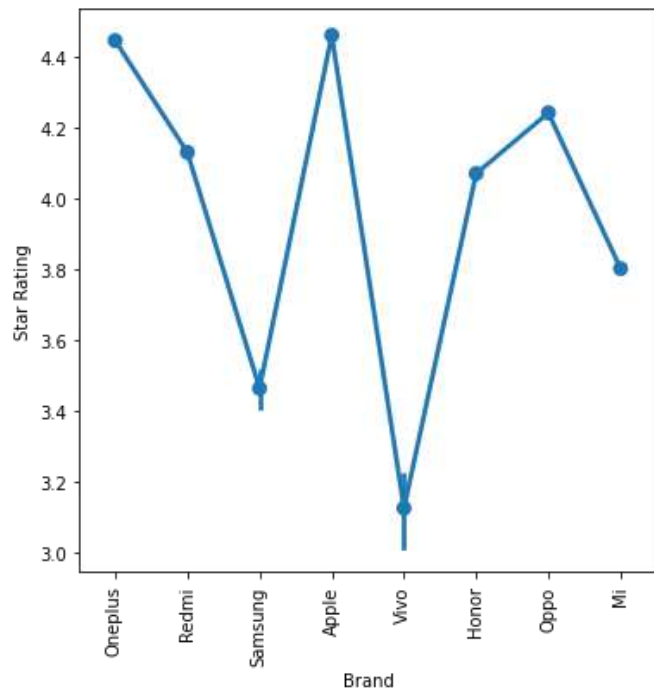


Figure 6: Analysis of Camera quality of various brands

Figure 7, illustrates the distribution of rating by various customers after purchase of the respective brands of their interest. It visualizes the number of customers who have given 1 star rating, 2 star rating, for each brand, and so on.

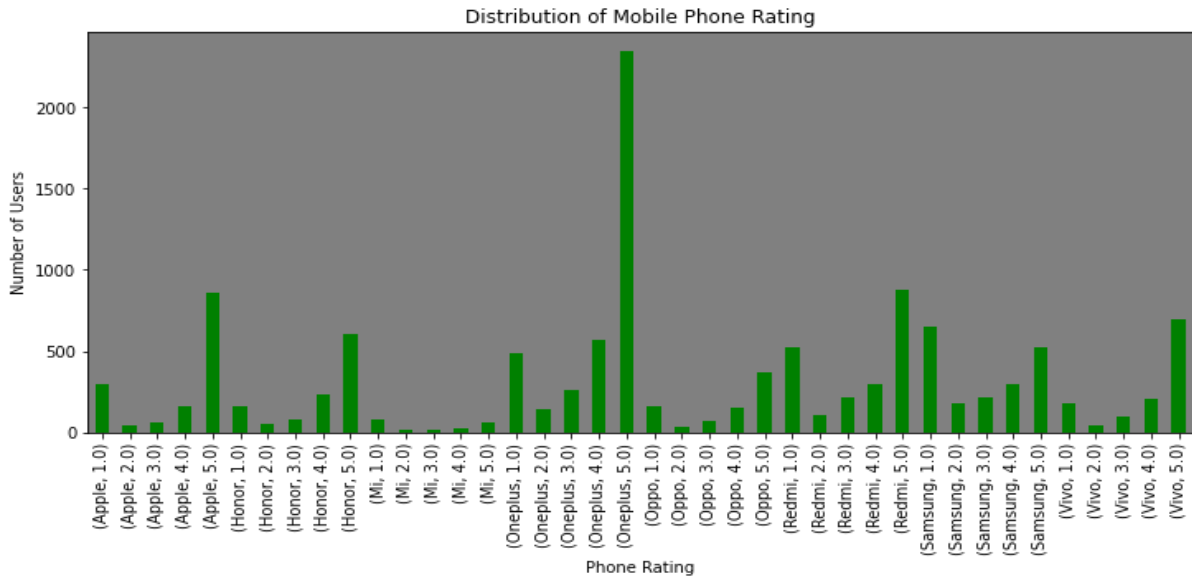


Figure 7: General distribution of star rating for each mobile phone brand

From the **Figure 8**, the overall distribution of rating score for each brand ranging from 1 to 5 is visualized. Oneplus mobile phone brand is having average of 4.1 star rating out of 5. Similarly, the average star rating for other mobile phone brands is also visualized and the same is summarized in **Table 2**.

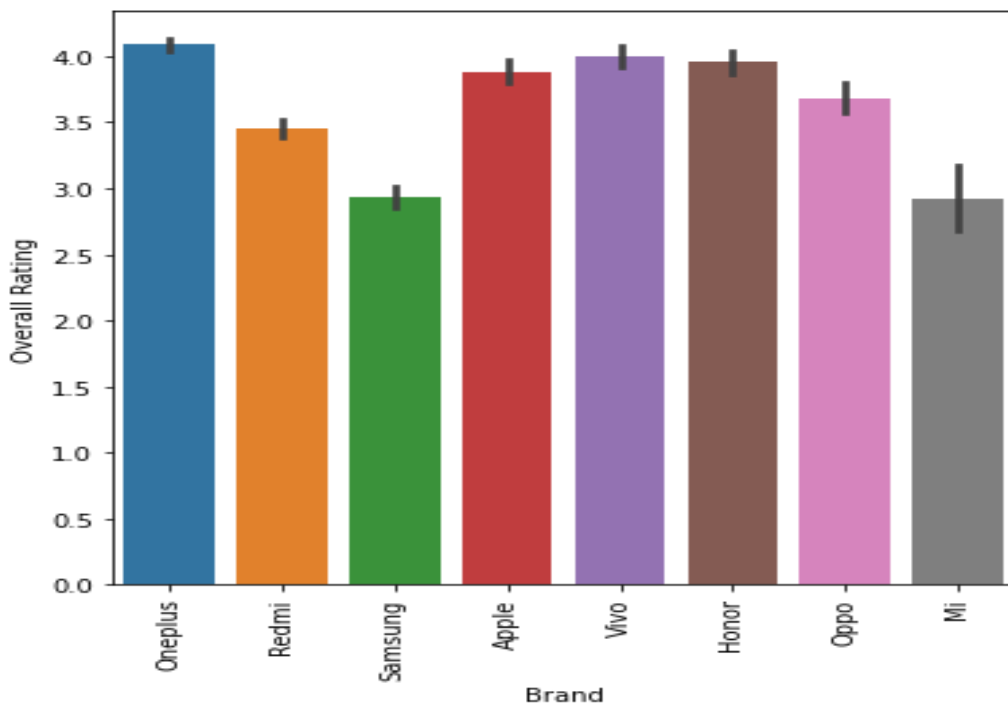


Figure 8: Distribution of rating score for each mobile phone brand

Table 2: Rating score for each mobile phone brand

Mobile Phone Brand	Rating Score
Oneplus	4.1
Redmi	3.4
Samsung	2.9
Apple	3.8
Vivo	4.0
Honor	3.9
Oppo	3.6
Mi	2.9

Conclusion and Future Work:

Data visualization for the customer reviews of the following are explained in this paper:

1. Trustworthiness of online purchase of mobile phones: From the analysis it is clearly observed that online purchase of smart phones has positive review.
2. Reviews available for the different brands of mobile phones: From the analysis it is inferred that the filtering of mobile phone brand for online purchase could be made easier than checking the review comments of the various brands.
3. Comparison of various features of the different brands of the mobile phones were analysed in this paper with the help of the amazon review dataset. This comparative analysis would help the future buyers to find the best phone which is more frequently brought through online.

This study gathered information on only few mobile phone brands. Future research could include other brands, and additional features of the various brands for comparative analysis.

Reference:

- [1]. Wu, J., Wu, Y., Sun, J., & Yang, Z. (2013). User reviews and uncertainty assessment: A two stage model of consumers' willingness-to-pay in online markets. *Decision Support Systems*, 55(1), 175-185.
- [2]. Li, M., Huang, L., Tan, C. H., & Wei, K. K. (2013). Helpfulness of online product reviews as seen by consumers: Source and content features. *International Journal of Electronic Commerce*, 17(4), 101-136.
- [3]. Sparks, B. A., Perkins, H. E., & Buckley, R. (2013). Online travel reviews as persuasive communication: The effects of content type, source, and certification logos on consumer behavior. *Tourism Management*, 39, 1-9. Spool J. (2009), The Magic behind Amazon's 2.7 Billion Dollar Question available online at, <http://www.uie.com/articles/magicbehindamazon/2009>.
- [4]. Korfiatis, N., García-Bariocanal, E., & Sánchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications*, 11(3), 205-217.

- [5]. Bilal, Muhammad, et al. "Profiling users' behavior, and identifying important features of review "helpfulness"." *IEEE Access* 8 (2020): 77227-77244.
- [6]. Dong, H., Hou, Y., Hao, M., Wang, J., & Li, S. (2020). Method for Ranking the Helpfulness of Online Reviews Based on SO-ILES TODIM. *IEEE Access*, 9, 1723-1736.
- [7]. Basiri, M.E. and Habibi, S., 2020, April. Review helpfulness prediction using convolutional neural networks and gated recurrent units. In *2020 6th International Conference on Web Research (ICWR)* (pp. 191-196). IEEE.
- [8]. Kong, L., Li, C., Ge, J., Ng, V. and Luo, B., 2020. Predicting Product Review Helpfulness A Hybrid Method. *IEEE Transactions on Services Computing*.
- [9]. Benlahbib, A., 2020. Aggregating customer review attributes for online reputation generation. *IEEE Access*, 8, pp.96550-96564.
- [10]. Fouladfar, F., Dehkordi, M.N. and Basiri, M.E., 2020. Predicting the helpfulness score of product reviews using an evidential score fusion method. *IEEE Access*, 8, pp.82662-82687.
- [11]. S. Urologin, "Sentiment analysis, visualization and classification of summarized news articles: A novel approach," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 8, pp. 616–625, 2018.
- [12]. Khan, S.A. and Velan, S.S., 2020, June. Application of exploratory data analysis to generate inferences on the occurrence of breast cancer using a sample dataset. In *2020 International Conference on Intelligent Engineering and Management (ICIEM)* (pp. 449-454). IEEE.
- [13]. Alamoudi, E.S. and Al Azwari, S., 2021, March. Exploratory Data Analysis and Data Mining on Yelp Restaurant Review. In *2021 National Computing Colleges Conference (NCCC)* (pp. 1-6). IEEE.
- [14]. E. S. Alamoudi and N. S. Alghamdi, "Sentiment classification and aspect-based sentiment analysis on yelp reviews using deep learning and word embeddings," *J. Decis. Syst.*, vol. 00, no. 00, pp. 1– 23, 2021.
- [15]. Gupta, R., Sameer, S., Muppavarapu, H., Enduri, M.K. and Anamalamudi, S., 2021, September. Sentiment Analysis on Zomato Reviews. In *2021 13th International Conference on Computational Intelligence and Communication Networks (CICN)* (pp. 34-38). IEEE.