Analysis of e-commerce behavior in Multi-Category Store

Sanya Sachdeva[1] and Supriya Raheja[2]

[1,2] Amity University, Noida, India, 201313
supriya.raheja@gmail.com

Abstract. Availability of multiple brands for a product leads a competitive environment among the manufacturers as well as sellers. This situation raises a requirement to understand their customers better and how their choices differ from one another. As in this competitive environment, it is very vital to keep customers close and happy. So, every manufacturer wants to correctly analyze their customer choices and behavior. Keeping this in mind, this work presents as analysis on the human behavior towards purchase as it helps in growth and earning profit which is very crucial for any manufacturer.

Keywords: Data analysis, e-commerce, consumer behavior, data mining

1 Introduction

Due to the mass usage of Internet by the consumers, enormous data are produced daily. Most of the data available on Web is in unstructured form which cannot be directly used by any filed. Here, the role of data mining comes which transforms it into useful data and creates large business for the mass [1].

Data science focusses on utilizing the modern mass data produced for estimation, assessment, insight, involvement, and analysis [2]. The collection of data itself is very intricate and tricky but once done appropriately, it is used to produce large business profits [3].

This paper focuses on analyzing eCommerce behavior data from multi category store. The current work presents how the products and users are connected to each other in an irresistible way and analysis of how it affects the business. Human behaviors can also be predicted through digitized texts where the way they interact with a person is clearly depicted and henceforth prediction and analysis are accurate and relevant. These data can easily help us understand the major chunk of our society and finding patterns to derive theories and transform them into something more solid and undeniable.

2 Literature Review

Uyoyo Zino Edosio [8] proposed idea of using the biggest assets like the needs of customers, predicting tendencies in customer's conduct, democratizing of advertisement for consumers and also making sure in meeting customers' needs. In another study, author extracted important information from bigger datasets where data gets produced at more speed, different range, and elevated volumes [9]. One of the study of Big Data Analytics in E-Commerce discussed on compilation of enormous data-sets that cannot be handled using conservative computing practices and accomplishing by executing appropriate schemes for attainment of precise and deep understanding of the information obtained by analyzing the data [10]. In another study on Big data analytics in E-commerce, authors discussed about how rarely-explored the ecommerce sector is and presents a framework based on definitional facets, distinct traits, kinds, business values and challenges of Data Analytics in the ecommerce [11].

Amarnadh and Akhilla [12] discussed on how online buying keeps increasing and how efficient it is to understand how to grow it even more in future. Weihs and Ickstadt, [13] showed that figures are one of the most crucial fields to offer tools and techniques to find structures and offer richer understanding into data and how to analyze it. Huseynov and Yıldırım discussed on organizing and categorizing the collected information on B2C e-commerce to find out less-explored regions and offer future study paths [14]. In one of the studies, authors discussed on breakdown of online consumers into various types to form a better interpretation and depiction of buying behavior in electronic business market. Sigurdsson et al. aimed on field built on broad trials with the aim of uncovering legitimate behavioral practices appropriate for consumer investigation [16]. Authors discussed on using consumer's information from several product categories prior to deciding, how these purchase situations create dependencies in choice outcomes across categories and change it into profit [17]. In one of the studies, authors presented several surveys on consumer buying behaviors and presented them and used in genuine problems using methods for more efficient customer behavior analysis [18].

3 Methodology

This section discusses the methodology adopted for the present work. The flow of work is illustrated through fig. 1.

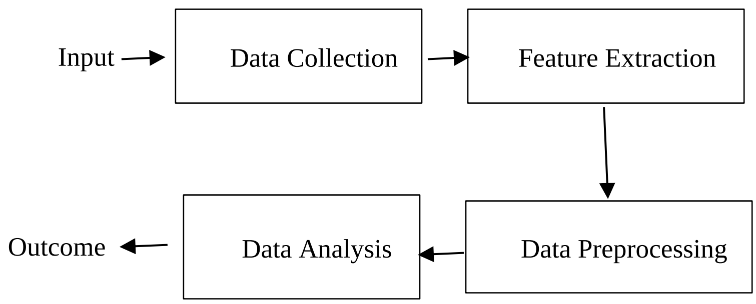

Fig. 1 Flow of Proposed work

3.1 Data Set Used:

The dataset used in this work is taken from the Kaggle.com [12]. It is a valid repository of datasets. The data is for the month of November 2019. There are 4635837 rows and 9 columns in the respective dataset. Each row in the data signifies an event which are related to products and users. There are different types of events according to the customers.

Different fields in the dataset are given below:

- event_time
- Event type
- product_id
- category_id
- category_code
- Brand
- Price
- user_id

Event_time describes when the event took place, event_type takes one entry out of the list [view, cart, remove_from_cart, purchase], product_id is the ID of the purchased product, category_id, is the category of the product, category_code; each categrory has a code which is represented by this field, Brand represents the brand name of the product, Price is the price of product and user_id is the id of the user which is generated at login. It is a unique value.

3.2 Feature Extraction and Data Preprocessing

Some of the features are irrelevant for current work [19]. We have extracted only the useful features such as Event type, product_id, category_id, brand, price. Some of the fields are having null values. For proper analysis of data, there is a need of cleaning the data. Cleaning is basically preprocessing, in which we are removing all the null entries from the dataset. We have also removed the duplicate entries from the dataset. One more column added to the dataset i.e. 'purchased'. Using the new added column, we have categorized the rows into 1 or 0 if the purchase took place or not simultaneously by comparing the value of event_type with 'purchase' then it gives 1 to purchase column and if any other value like view, cart or remove_from_cart in there then it will give the value 0 in purchase column[20].

4 Result & Analysis

To analyze the customer behavior, we have plotted a chart with respect to event type that is shown in fig. 2 and fig. 3. It can be observed that the ratio of viewing the product is more as compare to other events. Figure 3 clearly illustrates the minimum ratio in terms of purchasing the product by customer. Through this we can analyze that the customers usually just viewed the product and left it there to make up their mind or to explore somewhere else after which a lesser amount of people put the good in cart and then removed it from cart. The least amount of people actually bought the product which proves that the customers are wise enough to understand the importance of that product in their life hence they must have explored it somewhere else before putting into cart after which they must have had doubts if they really want the product

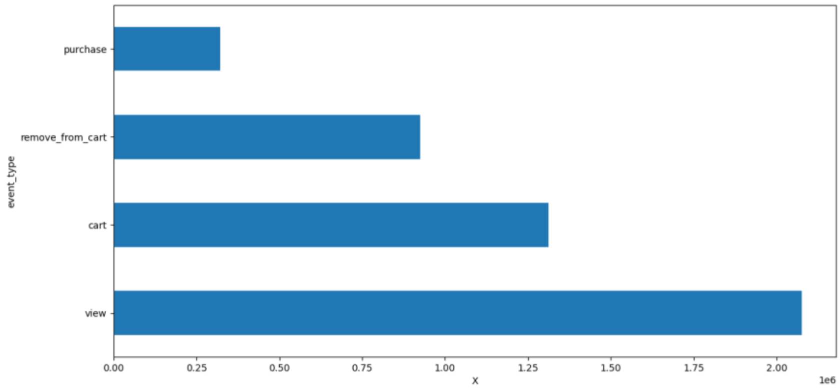

Fig. 2. Relation between event_type and x (le6)

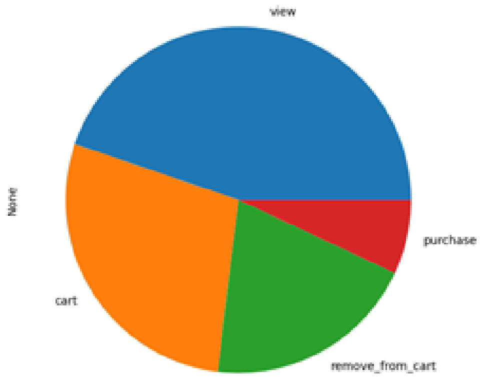

Fig. 3. Different values for event_type

nevertheless, of the less price and if it all matched then only they purchased it hence the amount of people doing that was significantly very less [21].

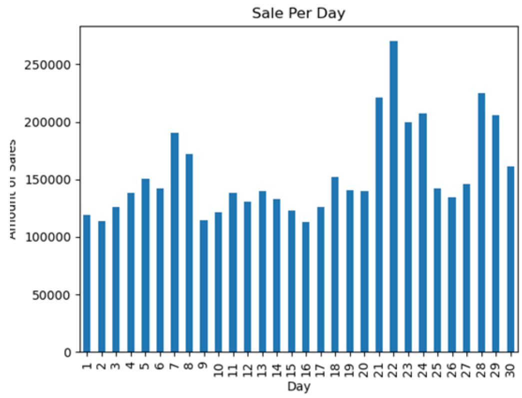

Fig. 4. Amount of sales per day

Figure 4 illustrates the relation between amount of sales per day where it was observed that they are usually average but on 22nd it was the highest.

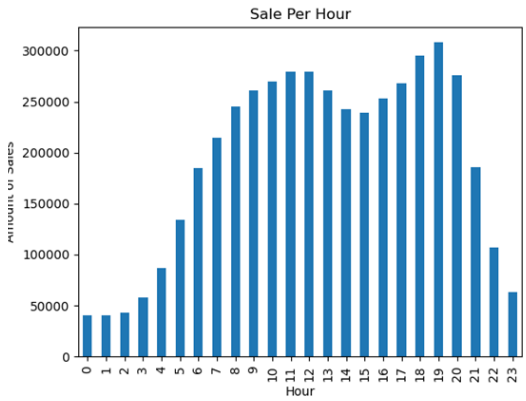

Fig. 5. Amount of sales per hour

Figure 5 depicted the amount of sales per hour where it clearly showed that people bought goods in early morning or mid night really less, the sales grew as the time reached between 6 am to 1 pm in afternoon after which there was again a dip in sales as it came near to lunch time and rose again as evening set in around 4 pm and kept increasing till 7 pm, this hour had the maximum sales in whole day and after this the sales dropped again in night exponentially as dinner time came and it goes on for the next day as well. Hence this proved that the maximum sales take place when people are getting ready for their work in the morning or when they get free in the afternoon, rest all time it is low and unsure [22]. Thus, it is important that in that time there is no technical error or shortage of products as great loss will be incurred [23].

5 Conclusion and Future Work

The work discussed the need of analyzing the e-customer behavior while purchasing the products from the multi-category store. The results of analyzes can help the sellers as well as the manufacturers to increase their selling rate, gain maximum profit and benefit the customers as well [24]. It is concluded that most of the time e-customer is just viewing the product and very often, customer is exactly purchasing the item. It is also concluded that the maximum sales are in the evening time by the e-customer and at peak around 7pm. Further the work can be extended by analyzing the e-customer behavior during discount time and with respect to festival time [25].

References

1. Paffenroth, R., & Kong, X. (2015). Python in Data Science Research and Education.
2. Blei, D. M., & Smyth, P. (2017). Science and data science. Proceedings of the National Academy of Sciences, 114(33), 8689-8692.
3. Galeano, P., & Peña, D. (2019). Data science, big data and statistics. TEST, 28(2), 289-329.
4. Odegua, R. (2019). DataSist: A Python-based library for easy data analysis, visualization and modeling. arXiv preprint arXiv:1911.03655.
5. Crowe, J., & Candlish, J. R. (2013, September). Data Analytics: The next big thing in information. In Fourteenth International Conference on Grey Literature (p. 139).
6. Russell, G. J., Ratneshwar, S., Shocker, A. D., Bell, D., Bodapati, A., Degeratu, A., ... & Shankar, V. H. (1999). Multiple-category decision-making: Review and synthesis. Marketing Letters, 10(3), 319-332.
7. https://www.kaggle.com/danofer/ecommerce-cosmetic-predict-purchases-data-prep/notebook#Target:
8. Edosio, U. Z. (2014). Big Data Analytics and its Application in E-Commerce. Proceedings E-Commerce Technologies. University of Bradford.
9. Avinash, B. M., & Akarsha, B. M. (2007). Big data analytics for e-commerce–its impact on value creation. International Journal of Advanced Research in Computer and Communication Engineering ISO, 3297.
10. Pavithra, B., Niranjanmurthy, M., Kamal Shaker, J., & Martien Sylvester Mani, F. (2016). The Study of Big Data Analytics in E-Commerce. International Journal of Advanced Research in Computer and Communication Engineering, 5(2), 126-131.
11. Akter, S., & Wamba, S. F. (2016). Big data analytics in E-commerce: a systematic review and agenda for future research. Electronic Markets, 26(2), 173-194.
12. Amarnadh, V., & Akhila, M. (2019, May). Big Data Analytics in E-Commerce User Interest Patterns. In Journal of Physics: Conference Series (Vol. 1228, No. 1, p. 012052). IOP Publishing.
13. Weihs, C., & Ickstadt, K. (2018). Data Science: the impact of statistics. International Journal of Data Science and Analytics, 6(3), 189-194.
14. Huseynov, F., & Yıldırım, S. Ö. (2016). Behavioral Issues in B2C E-commerce: The-state-of-the-art. Information Development, 32(5), 1343-1358.

15. Wu, R. S., & Chou, P. H. (2011). Customer segmentation of multiple category data in e-commerce using a soft-clustering approach. Electronic Commerce Research and Applications, 10(3), 331-341.
16. Sigurdsson, V., Larsen, N. M., & Fagerstrøm, A. (2016). Behavior analysis of in-store consumer behavior. The Routledge companion to consumer behavior analysis, 40-50.
17. Russell, G. J., Ratneshwar, S., Shocker, A. D., Bell, D., Bodapati, A., Degeratu, A., ... & Shankar, V. H. (1999). Multiple-category decision-making: Review and synthesis. Marketing Letters, 10(3), 319-332.
18. Raorane, A., & Kulkarni, R. V. (2011). Data mining techniques: A source for consumer behavior analysis. arXiv preprint arXiv:1109.1202.
19. Silambarasan, G., & Raj, P. S. Consumer Behavior Marketing Analysis Using Data Mining Apriori Algorithm.
20. Park, C. H. (2012). Essays On Shopping Dynamics In Customer Base Analysis.
21. Liu, J., Tang, T., Wang, W., Xu, B., Kong, X., & Xia, F. (2018). A survey of scholarly data visualization. IEEE Access, 6, 19205-19221.
22. Krum, R. (2013). Cool infographics: Effective communication with data visualization and design. John Wiley & Sons.
23. Zheng, J. G. (2014). Data visualization.
24. Fares, A. (2019). The Effects of In-Store Category Adjacencies on Consumer Purchase Behavior. Available at SSRN 3454126.
25. Assael, H. (1995). Consumer behavior and marketing action.