Ecommerce: Analysis of Customer Behaviour to Enhance Sales

Shrikant Samarth

CRM Project Report

Msc in Data Analytics

Student Id: X18129137

National College of Ireland

Guided by Prof. Vikas Sahni

Abstract - **E-commerce alludes to the utilization of the Internet and the Web to execute business. Moreover, E-commerce business is about carefully empowered business exchanges between and among organizations and people. Nowadays, people are moving towards this digital shopping and many retail sellers are now selling goods online. But to grab the attention of a customer, it is important to understand the trend, preference and customer behaviour to target sales. This study clarifies an examination of one of the such ecommerce order transactions between various sellers and the customers from 2016 to 2018 made at multiple marketplaces in Brazil. This dataset can be further used to improvise the seller service by checking, the relationship between the timestamp and the customer ratings for the service. This examination attempts to build up a connection between the customer satisfaction level with respect to the delivery service, payment methods, timestamp using different methods for example, cross table, binning, graphical portrayals of the connections of the variable utilizing R and Tableau programming. The result of the study portrays that the customer satisfaction is one of the important factors to retain and get bulk orders. By offering various services such as payment options, free and on time deliveries, the user satisfaction can be maintained which would help the business to fetch excellent and quality reviews and would in turn attract new customers towards the service.**

Keywords: E-commerce, Brazilian Market, Customer Satisfaction, Analysis, Order sales, Online Sales

I. INTRODUCTION:

In any industry it is important to meet the customers demand in order to meet and maintain the customer relationship. With the fast-growing users on the internet, not only the retailers but also many entrepreneurs have started business to sell goods online. In 2014, according to [1] the growth rate for ecommerce industry was growing at a pace of 12% annually than the 3.5% of the retail industry. Within just few years, E-commerce has expanded from desktop to mobile devices. Though most of the sales takes place through the traditional system, E-commerce continues to grow rapidly and has created a way for many companies to come into this business in order the meet the market demand. But, in this vast market foreseeing the customers demand is the major issue. To convert an effective sale, it is important to give luring offers and gain

a trust of a customer. Once a relationship has built up a company need to check the customer behaviour in order to retain their customer base. This study tries to focus on how the customer can be retained and check what factors could be responsible for the customers satisfaction level. A Brazilian E-Commerce customer order dataset over the two years is used as the base for this study. Such industry level dataset would be helpful to study customer behaviour.

In Brazilian industry, boleto is a different payment option which is more popular. As market like brazil, which is heavily cash based society over, 40% of population does it isn't even have a bank. A boleto option allowed customer to pay cash at ATMs branch facilities or Internet banking of any Bank, Post Office, Lottery Agent and some supermarkets within the due date. After the due date it must be paid at the guarantor bank offices. Boleto must be gathered by an approved Collector Agent in the Brazilian domain [2]. Thus, we will look into the analysis of payment type transaction in this study

This analysis will not only benefit the users to get good customer service but also to the sellers to make a proper strategy and improve their services which will benefit them in the long run. The prime focus of this study is to check, which are the top customer preferred product categories to ensure which products needs to be stocked and avoid blowouts. What are the payment preferences, is it related to the product price? If number of payment instalments has any role is related to selling more products? These kinds of analysis can help online sellers to understand the customer behaviour and can help them in building campaign, strategies and provide promotions if necessary, to improve sales.

II. HYPOTHESIS:

The null hypothesis of this study is "Payment methods and the duration of the delivery do not account to the customer satisfaction on an ecommerce platform"

III. LITERATURE REVIEW:

As mentioned in the previous section, there are many studies done to analyse the main factors which affects the customer satisfaction rate.

Some studies have been done in the ecommerce sector to find out the which factors significantly affects the customer satisfaction, in order to convert sales more and to retain customers base. A previous study work done by [3] to check the influence of experimental marketing on customer satisfaction based on the quantity relationship. The study was done on the ticketing online through survey. Analysis was done using descriptive analysis and statistical methods such as simple linear regression resulted as a positive influence between customer satisfaction and relationship quantity. Another study done by [4] on rating prediction of consumer products and fusion of Egg (electroencephalogram) response. Natural Language Processing was used for the reviews posted by the global viewers. Further, random forest regression technique was used to model Egg data which was helpful to build rating prediction framework which was considered as local rating. Egg dataset was for the 40 participants were recorded which was for 42 different E-commerce website products.

Moreover, [5] research work proceeds this research by proposing a relation between brand value and customer satisfaction in an E-commerce business. Artificial intelligence with the help of vosag technology was used on 50 consumers on creating a brand value and customer satisfaction in E-commerce business to make their purchase more flexible and more reliable. The result shows how buying through online could be automated by using AI according to customer preference and expectation of consumers. Also, this technique will help motivate customers to be loyal to a brand because of their goods and better services.

In the process achieving customer satisfaction it is important to strengthen the delivery service. One such study by [6] based on Big data platform, express delivery method was proposed for the E-commerce. The time range distribution was analysed to obtain penalty function using the time window concept. In this study, path optimization method for optimal route was suggested by using the collaborative delivery system. This experiment shows how minimum and optimized time can be achieved for delivering products on time.

In another research work done by [7], who tries to explore the dependent fields of customer satisfaction. In an attempt to measure customer satisfaction various measures such as combining MIS field and the marketing/customer behaviour field. The importance about this study is that the researchers found out that payment methods need to be reviewed in any online platform which could be another way to attain the customer satisfaction and increase the sale. This method is also useful for any seller to build a marketing strategy and develop the business.

Thus, after assessment of previous studies, the project is implemented by linear regression and can be found out whether payment instalments, price and delivery time significantly affect the customer satisfaction.

IV. BACKGROUND OF THE DATASET:

The E-commerce order dataset which was used for the analysis named as Brazilian E-commerce public dataset. This dataset has been provided by Olist, which offers platforms for multiple sellers to sell products across different marketplaces. The Master file dataset contains both the customer and the seller product information such as product id, customer id, payments methods, ratings, customers geolocation, timestamp (ordered date & delivery date), customer satisfaction ratings and classifications. The dataset contains 22 columns and 32952 rows. Each customer and seller have a unique Id and details regarding the products is mentioned in the file. Detailed information and attributes of dataset is given in the below table:

Dataset**:** https://www.kaggle.com/olistbr/brazilian-ecommerce

TABLE 1: ATTRIBUTES OF E-COMMERCE ORDER DATASET

Attributes	Description
product_id	Every product has its Unique code which is represented as product_id
product_category_name	This column gives information about the Product's main category.
product_name_length	It gives product's title length which was displayed on the online listing.
product_description_length	It gives information about the total description words count.
product_photos_qty	Total number of photos displayed on the listing is mentioned.
product_weight_g	Product weight is mentioned.
seller_id	It gives information about the seller who sold the product
order id	It gives information about the order unique code
order_purchase_timestamp	Date of the order purchased in yy-mm-dd hh:mm
order_delivered_customer_date	Order deliverey date in yy-mm-dd hh:mm
payment_type	Payment information e.g. Credit, debit, boleto etc.
payment_installments	Number of instalments taken for the purchase.
payment_value	Amount of the transaction with shipping amount is mentioned.
price	Original price of product
review_score	Ratings given for the service. (1 star – 5 star)
geolocation_lat	Latitude of customer delivery address.
geolocation_lng	longitude of customer delivery address.

Below section explains regarding the data pre-processing and techniques.

V. METHODS AND TECHNIQUES:

A. Data Pre-processing:

In general, a real-world data contains a lot of issues such as inconsistent or incomplete information, could require some transformation in the data etc. To perform any type of analysis, it is important to resolve such issues. R studio has been used for cleaning and merging different files. For data pre-processing following steps are taken:

1) Originally there were nine different data files related to the customer, seller, payment information, customer geolocation details etc. All these files were merged with the reference of customer and seller ids, and a Master file is created for our project.

2) After creating a Master file, out of 22 columns 6 columns were dropped as those columns information which was not useful for our analysis.

3) In R studio .na() function was used to take care of missing values in the dataset. As the dataset contains different data types which was corrected with the corrected data types for example: categorical values which was in string was changed to factors using as.factor() function in R studio.

B. Proposed methodology:

The main goal of this project is to find the customer behaviour pattern with respect to the payment type and the order purchase. To understand this, we need to find the correlation between the variables first and need to check what is the significance between the variables. For this we have used SPSS multiple regression technique to check the correlation between the variables as studied in the previous literature review section. The multiple regression technique is used to check the relationship between the dependent variable and multiple independent variable [8]. Multiple regression allows to decide the general fit (variance) of the model and the overall contribution of each of the indicators to the total variance.

Further we have used data visualization techniques to find the pattern in the customer order dataset in order to see which are the factors that could affect the customer satisfaction rate. As there are a good amount of information in our dataset, data visualization is the best strategy to discover the trend , pattern and correlation between the variable in a visual context [9]. Hence, Tableau, SPSS and R studio were used for this project.

C. Results and Analysis:

To start with the experiment, first we need to see the descriptive overall statistics for the selected variables. Fig 1 shows mean average score which is our dependent variable is 4.04; the independent variables price, payment value are 145.54 and 187.41 respectively. Whereas the mean payment instalments in the order history is 3.15 which can be considered as 3. That means average customer prefers 3 payment instalments for their transaction.

Descriptive Statistics

|	Mean	Std. Deviation	N
review_score	4.04	1.396	32951
price	145.5468271	248.1431084	32951
payment_value	187.4174511	285.8980260	32951
payment_installments	3.15	2.910	32951

Figure 1: Descriptive Statistics

The below correlation table shows the significance between the dependent and the independent variables.

Correlations

		review_score	price	payment_value	payment_installments
Pearson Correlation	review_score	1.000	-.008	-.050	-.043
	price	-.008	1.000	.891	.292
	payment_value	-.050	.891	1.000	.310
	payment_installments	-.043	.292	.310	1.000
Sig. (1-tailed)	review_score	.	.085	.000	.000
	price	.085	.	.000	.000
	payment_value	.000	.000	.	.000
	payment_installments	.000	.000	.000	.

**Figure 2: Multiple Regression - Correlation table**

Fig. 2 shows the correlation matrix between the dependent variable i.e. Review score whereas the independent variables i.e. price, payment value and payment instalments. From this table the correlation between the payment value and the review score is more significant with respect to the other variables because the significance value is less than .05 and linear relationship is medium strength, as the persona correlation 'r' value (-.05 is nearly equidistance between 0 and -1 [10]. Similarly, the significance can be seen amongst the other variables except between price and review score. Thus, we can proceed with our further analysis as we can see the correlation between the variables.

Now to understand the customer satisfaction pattern, we need to see which cluster of people are more satisfied with their transaction provided in terms of their purchase in different categories. This study will also guide sellers to see how this analysis would be used to for the improvement services which will ultimately help to achieve the customer satisfaction.
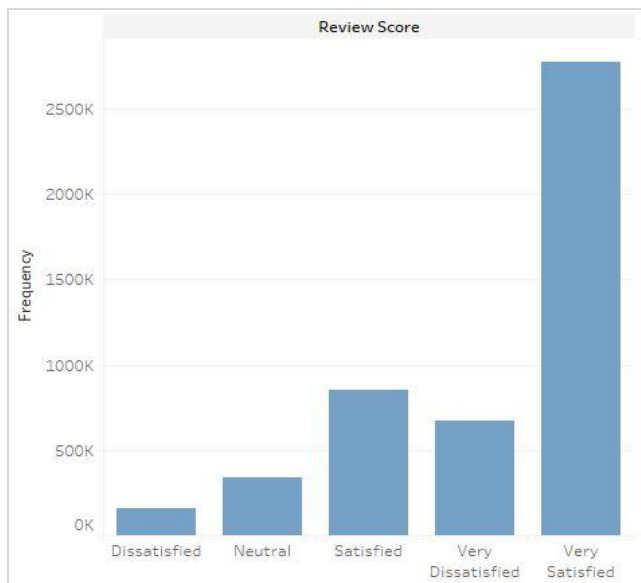
Figure 3: Overall Customers Satisfaction

It seems from the overall dataset that majority of the customers are in the very satisfied and satisfied categories which shows the customers are happy with their purchase. But there are customers who are very dissatisfied. As we know in the ecommerce industry maintaining a performance matrix for any seller is important as it decides if the future customers would trust the seller for the service. The performance matrix is the overall score which is given by the customers which also tracks the on-time delivery, deficit rate, stock maintenance. All these things are important but secondary. From the above graph, it shows 25% of the customers are below neutral category, which would rotten the overall score. Hence, to dig further we will investigate which are those categories where customers are satisfied i.e. which is above the neutral.
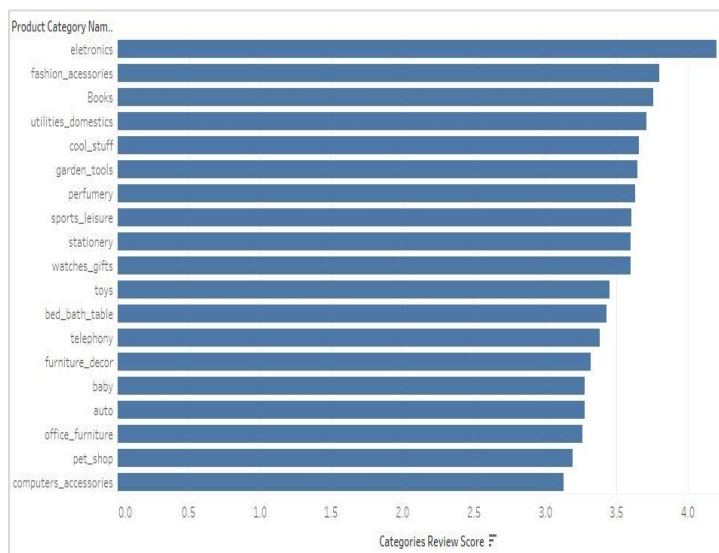

Figure 4: Customers Satisfaction in Categories

From the fig. 4 shows customer satisfaction category-wise. Once we understood the overall satisfaction of customer, category-wise ratings give an idea which are the top preferred categories and what is the average rating score. Moreover, this analysis also shows which are the topmost preferred categories. Electronics which is proven to be the best category in the industry has best ratings which shows whereas categories such as pet products, office furniture and the computer accessories categories has the moderate ratings. First, it is important to understand these customer needs to be contacted first; as these are the easy targets who can convert into the good score by giving a feedback call. By doing so, sellers can easily improve their overall score and can also investigate into which area needs some improvement. For the ratings below neutral, shows the customer clearly unhappy with the transaction. It can be either with the service he received or there could be an issue with the product he received. These are the customers who needs more attention and are the top priority customers because, in order to maintain a good relationship and build a trust with the customer a feedback from the seller is important in such cases.

In order to achieve the customer satisfaction level, we need to understand the pattern from the order date to the time required to deliver the product.
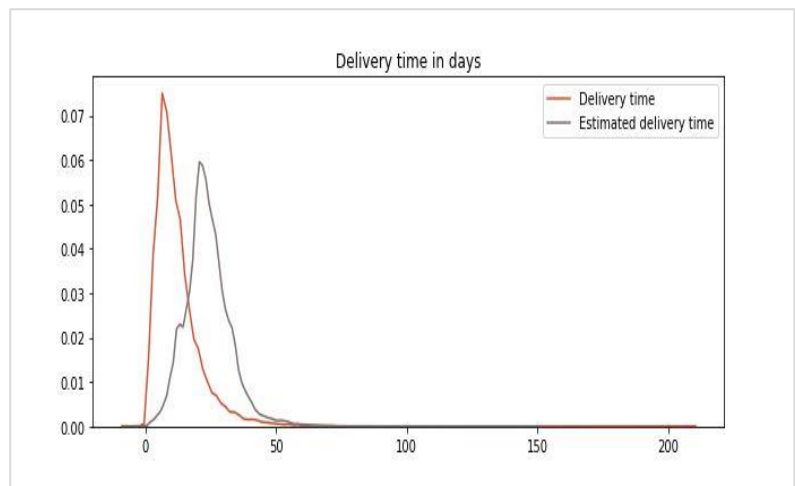
Figure 5: Delivery time in Days

From fig. 5 clearly most of the order are delivered before time got good ratings whereas there are some outliers in the delivery time. Hence, we can say that the customer satisfaction is most probably related with the time taken to deliver the product to the customer.

Another factor from where seller could get more sales is the payment type and number of instalments offered and needs to see if these factors are important to achieve the customer satisfaction score. Below fig. shows the for overall orders which are the most preferred payment type and how many instalments customers took for the transaction.

**Figure 6: Orders and Payment total value vs payment type**

From Fig. 6 pie chart we can see that most of the orders are done from credit card followed by the boleto payment method famous in Brazilian industry (explained in the introduction section). From the second pie chart customers gives clarity about the preference to pay with boleto (as explained in the Introduction section) for any single order then with the credit card.
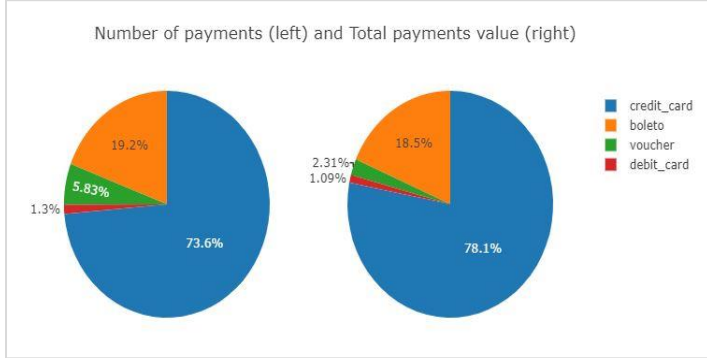

Figure 7: Number of instalments and Total payment value Orders and Payment total value vs payment type

From fig. 7, for multiple payment customers prefer Credit card and the total cost of the transaction is also more with the credit card. Thus, we can say that the for big expensive products people prefer credit card over other payment options.
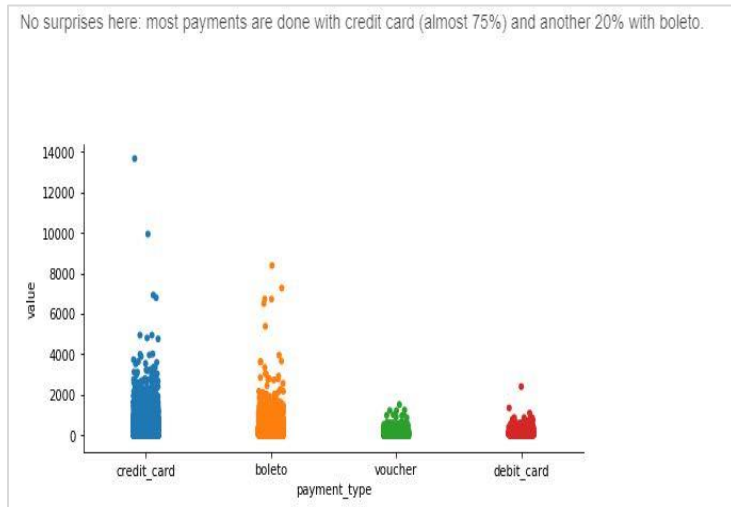
Figure 8: Payment Methods

Fig. 8 gives clearer picture in terms the choice of the payment method. That clearly shows for an expensive product and the choice of payment instalments customer prefer credit card and boleto over other payment methods.

.
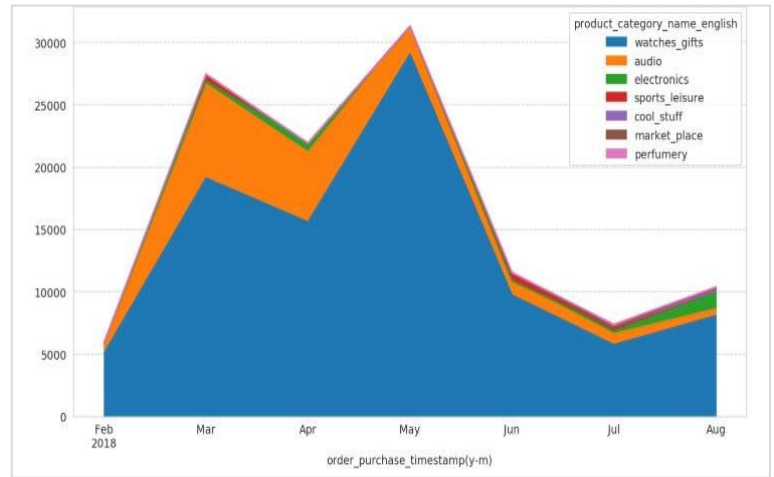
Figure 9: Monthly Revenue by Top product Category

From the above fig. 9, which is top category by revenue over the year shows that most of the revenue is coming from the luxury items followed by the electronics categories. We can see the spike in March and May which are the months that contributed more in terms of revenue. Hence, we can conclude from this that, sellers can give certain offers focus on these months and can convert more sales if given variety of products to the customer.

Figure 10: Sales per month

Fig. 10 gives an idea abouth customer ordering pattern and have more buying tendancy. Christmas and year end festivals are the season month for ecommerce industry. Hence sellers should be prepared for huge amount of order flow considering the pressure of good customer service that leads to achieve the customer satisfaction target.

VI. VALUE

The goal of this study is to analyse what are the various factors which needs to be taken care of, to understand the customer behaviour and ensure the customer retention. As discussed in this paper customer satisfaction is one of the factors which plays a major role in retaining the customer, to ensure a long-term relationship. For a industry like E-commerce, it is important to improvise the service continuously according to the customer needs in order to make an impact on the customer which will ultimately help in maintaining and increase a customer base.

1) The first and the foremost important thing for any seller is to find the right and wide range of products to attract the customers. Sellers needs to understand

the customer needs and accordingly they must maintain a range of products. Also, products should be selected based on the common preference, geographic location, product category and specific group of customer base. Additionally, seller must consider the product to be within competitive in price to attract more customers.

2) It comes out from the overall survey that, there are more satisfied customer than the unsatisfied ones. Though comparatively, there are less disappointed customers it is important to convert these customers ratings into positive, which can be done by reaching out to the customers and asking them about where the sellers needs to improvise. If the customer is unhappy with service or the product which he received; offer him a free replacement, addons, giveaway products and try to resolve the issue as soon as possible. Any delay in the service or getting the feedback from the customer only worsen the situation. So, situation should be handled on time by the sellers. In this customer driven industry, it is very important for any seller to make sure a smooth and good experience for its customer.

3) It can be seen from the pie chart (fig.) that for most of the customer has purchased product with boleto payment type order value. As market like brazil, which is heavily cash based society over, 40% of population does it isn't even have a bank. But when it comes to instalments, it can be seen that customer prefer credit card for their payments as with the instalments they can buy more and expensive products. So, in order to reach out to the customer, seller should give flexible options to the customers which could surely increase the sale and help customer to develop more trust in the brand.

VII. CONCLUSION AND FUTURE WORK:

In conclusion from the above analysis, we can conclude that customer can be retained if the product delivered in time and if there is a delay in the product delivery, it is a duty of a seller to inform customer for the same. From the above analysis it can also be seen that usually for the heavy and slightly more expensive products it takes more time for the delivery, in such cases customer should be promptly informed for the same. It will help in maintaining a good relationship as well as it will help seller to get good ratings. Moreover, payment method proven to be the important parameter to enhance the sales over a period of time. As it increases the purchasing power of customer. Hence above analysis suggests on time delivery, flexibility in payment method and good customer service would help seller to gain customer trust which would help them to concert more sales. Feedback call also help seller in retaining the customer trust.

Customer retention can be further analysed using the artificial intelligence, neural network NLP for predicting customer reviews as a future work. This analysis could be helpful in building rating prediction framework and understanding how buying through online could be automated to build a brand name.

REFERENCES

[1] K. C. a. L. J. P. Laudon, "E-Commerce: Digital Markets, Digital Goods," in *Management Information Systems: Managing the Digital Firm Plus MyMISLab with Pearson eText--Access Card Package*, Prentice Hall Press, 2015, pp. 415-418.

[2] Wikipedia, "Boleto," Wikimedia Foundation, Inc, [Online]. Available: https://en.wikipedia.org/wiki/Boleto. [Accessed 21 jul 2019].

[3] A. a. R. P. Hidayat, "The Impact of Relationship Quality on Experiental Marketing and Customer Satisfaction in the Case of Ticketing Online," *2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA),* pp. 796--799, 2019.

[4] S. a. Y. M. a. R. P. P. Kumar, "Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction," *Information Fusion,* vol. 52, pp. 41--52, 2019.

[5] S. a. o. Sha, "Creating a Brand Value and Consumer Satisfaction in E-Commerce Business Using Artificial Intelligence," *Rajeswari, Creating a Brand Value and Consumer Satisfaction in E-Commerce Business Using Artificial Intelligence,* 2019.

[6] Z. a. Y. P. Chu, "Innovation of E-Commerce Terminal Express Cooperative Distribution based on Big Data Platform," *International Journal of Performability Engineering,* vol. 15, p. 3, 2019.

[7] H.-R. Kim, "Developing an index of online customer satisfaction," *Journal of financial services marketing,* vol. 10, pp. 49--64, 2005.

[8] W. Foundation, "Linear regression," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Linear_regression. [Accessed 21 July 2019].

[9] F. a. S. S. L. a. X. D. a. C. T. Yang, "Improved correlation analysis and visualization of industrial alarm data," *ISA transactions,* pp. 499--506, 2012.

[10] B. College, "Regressions and Correlations: Multiple Regressions," Barnard, [Online]. Available: https://erc.barnard.edu/spss/multiple_regressions. [Accessed 21 jul 2019].

[11] F. C. Y. &. C. L. (. Hou, "Analysis on the quality factors of b2c cross-border ecommerce service," 2019.

[12] P. &. L. P. (. Aungkulanon, "Relationship management of customer demand and production planning on e-business of thai natural cosmetics," in *ACM International Conference Proceeding ,* 2019.