Anubin Binoy
CHRIST, Deemed to be
University, Bengaluru,
Karnataka, India

Fathimath Safna
CHRIST, Deemed to be
University, Bengaluru,
Karnataka, India

Maria David
CHRIST, Deemed to be
University, Bengaluru,
Karnataka, India

Factors influencing online shopping behaviour: An empirical study in Bangalore

Anubin Binoy, Fathimath Safna and Maria David

Abstract
Online Shopping is growing rapidly in India, predominantly driven by tremendous and substantial divulgatory activities among millennial consumers. In this era where competition is considered, marketing has become the keystone for competition. Online shopping is progressing and is becoming more popular and is attracting great attention because it has great potential for both consumers and vendors. The convenience of online shopping makes it more successful and makes it an emerging trend among consumers. When all the companies are striving against one another, certain factors influence the behavior of customers. There are significant associations within these factors. This paper analyses the relationship between the important independent variables, including consumer behavior, cultural, social, personal, psychological, and marketing mix factors. The results revealed that the influence of Brand as a factor had positively influenced the customer's decisions in shopping online and evaluates the customer's level of satisfaction with online shopping. Results provided in this research could be employed as reference information for Ecommerce app builders and marketers regarding such issues in the city.

Keywords: marketing, e commerce, online shopping, consumers

1. Introduction
Today a large number of people around the world are using online shopping. Over the last few years, e-commerce has become an indispensable part of global retailing, and it has also undergone a number of changes and will keep changing according to the trends. And This research is based on the factors that influence the online shopping behavior of consumers. Online shopping has been rapidly increasing in India. There are over 75 million online buyers in India, and the number of online shoppers is over 2.05 billion, which makes 26.28% of the 7.8 billion in the world. But this study contains the online shopping behavior of consumers from Bangalore, Karnataka. Online shoppers have been growing at a massive rate, and the annual shopping of consumers has been increasing. But there are advantages and disadvantages for the same, and there are several factors that are being considered by the consumers while doing online shopping. This study mainly focuses on these factors such as Brand, quality, price, review, etc. (Ngo Tan Vu Khanh, Gwangyong Gim, 2014.) [5], (Chayapa Katawetawaraks, Cheng Lu Wang, 2011) [4], (Anurag Pandey and Jitesh Parmar, January 2019.) [11]. The disadvantage is that consumers may stay online and shop but do not purchase any product, which in turn results in wastage of time. It can also be agreed upon that online shopping is not possible without a gadget and internet services (Jozef Bucko, 2018.) [3], and only those who can enjoy this who has the privilege of both of these. This has increased because of the detailed product information and improved services attract more and more people to change their consumer behavior from the traditional mode to rely on online shopping. More, companies have realized that the factors affecting consumer behavior are unavoidable, and they also have to keep changing their marketing strategy. The consumer can purchase products online 24X7, and it also provides the minimum possible price with various cash backs and discounts. It can also be agreed upon that online shopping is not possible without a gadget and internet services, and only those who can enjoy this who has the privilege of both of these. Sometimes faulty products are delivered, and it consumes time for exchanging the product too. There are delivery charges, and the products arrive in greatly preserved packages, so one cannot touch and feel them (Gunashekharan, 2017.) [8].

Corresponding Author:
Anubin Binoy
CHRIST, Deemed to be
University, Bengaluru,
Karnataka, India

Many factors are affecting the online shopping behavior of consumers, and this study is mainly focused on this. This study endows a deeper understanding of the effect of different factors on consumer buying behavior. This study revolves around customer behavior, specifically customers of online shopping. To analyze and evaluate the factors influencing online shopping behavior, that includes understanding the usage of different apps for online shopping. Studying the customer's level of satisfaction about online shopping and analyzing the dependability of age groups in online shopping during the Covid19 pandemic.

2. Methodology
Data employed in this study were obtained from a structured questionnaire design. The questionnaire was distributed among Bengalureans. The questionnaire was divided into two sections. The first part includes socioeconomic and demographic variables. The second part presents the factor that leads to customer shopping online. The first five questions were the demographic information of the respondents. Then each of the remaining questions presents the factor that leads to customer shopping online. All statements were formulated on a 5-point Likert-type scale ranging from "strongly disagree (1)" to "strongly agree (5)" and "Yes" or "No".

Since the response variable 'brand' is binary, the study employed Logistic Regression in predicting the influence of Brand in online shopping by developing the relationship between the factors/predictors like gender, monthly income level, education level, age group of participants.

Let π denote the probability of influence of Brand on online shopping when p predictor variables are given, and the relationship between the probability π and p predictors is represented in the form of a logistic model i.e.

$$\pi = Pr(Y = 1 | X_1 = x_1, \ldots, X_p = x_p)$$

$$= \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p}} \tag{1}$$

The function given in equation (1) is the logistic regression function. Among the non-linear regression coefficients β_0, $\beta_1 \ldots \beta_p$, it is linearised by the logit transformation, i.e., if the probability of Brand influencing online shopping is π then the resulting ratio $\frac{\pi}{1-\pi}$ is the log odds of Brand influencing online shopping.

Here,

$$1 - \pi = Pr(Y = 0 | X_1 = x, \ldots, X_p = x_p)$$

$$= \frac{1}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p}} \tag{2}$$

Then,

$$\frac{\pi}{1-\pi} = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p} \tag{3}$$

Take natural log on both sides,

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) \tag{4}$$

$$= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

In equation (4), logit (π) is a linear function in terms of the regression parameters called the logit function and the range of π in equation (1) is between 0 to 1, and the range of values of $\log\left(\frac{\pi}{1-\pi}\right)$ is between $-\infty$ and ∞, which makes the logits quite suitable for linear regression model and the error term 'e' satisfies all of the basic assumptions of ordinary least squares.

From the survey responses of the factors influencing online shopping as shown in Fig 1., the percentage of respondents on each row who agree with the statement is shown to the right of the zero lines; the counts (or percentages)who disagree are shown to the left.
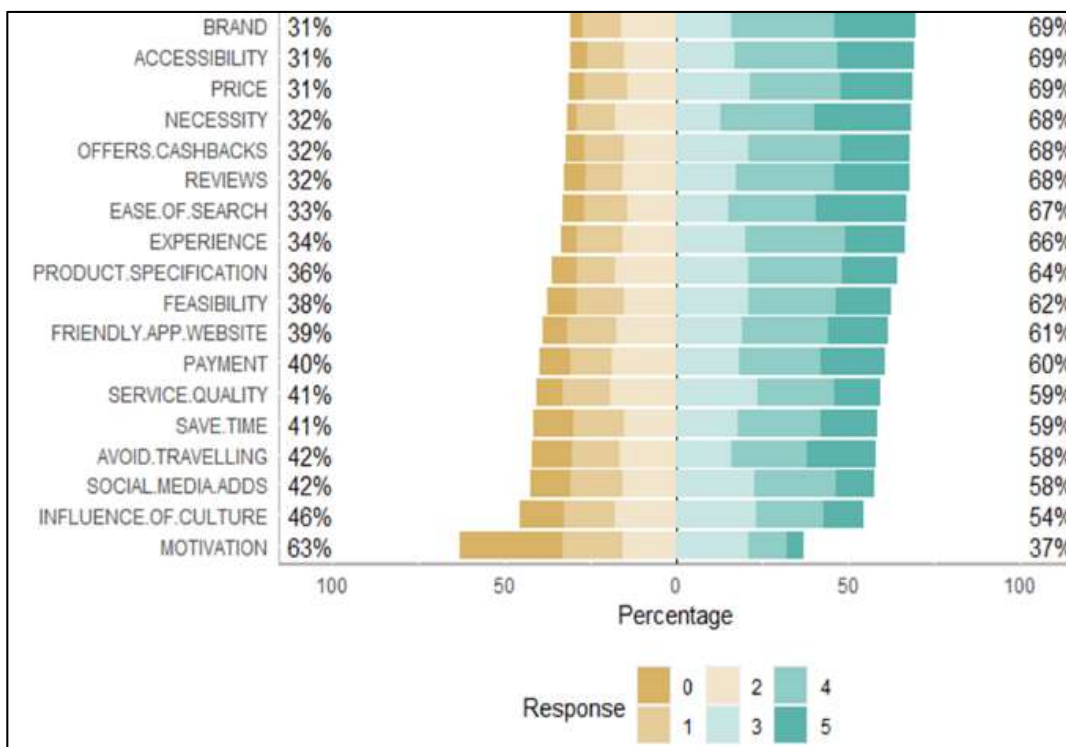
Fig 1

The percentage of respondents who neither agree nor disagree are split down the middle and are shown in a neutral color. Here, we can see that Brand, accessibility, and price are the most influencing factors, with 69% of the respondents agreeing to this. The factors Necessity, Offers/Cashbacks, and reviews have 68% of respondents. The factor that had the highest percentage (63%) of disapproval was Motivation, which is followed by the influence of culture (46%). Therefore, the top five factors influencing online shopping are Brand, accessibility, price, necessity, and offers/cashback. The least five factors that influence online shopping from our study are found to be 'motivation', 'influence of culture', 'social media ads', 'avoid traveling' and 'save time' factors.

From the Polychoric correlations of the factors influencing online shopping as given in Fig 2., the dark blue shaded values indicate a high correlation (> 0.72), and a lighter shade of blue indicates a moderate correlation between the factors.
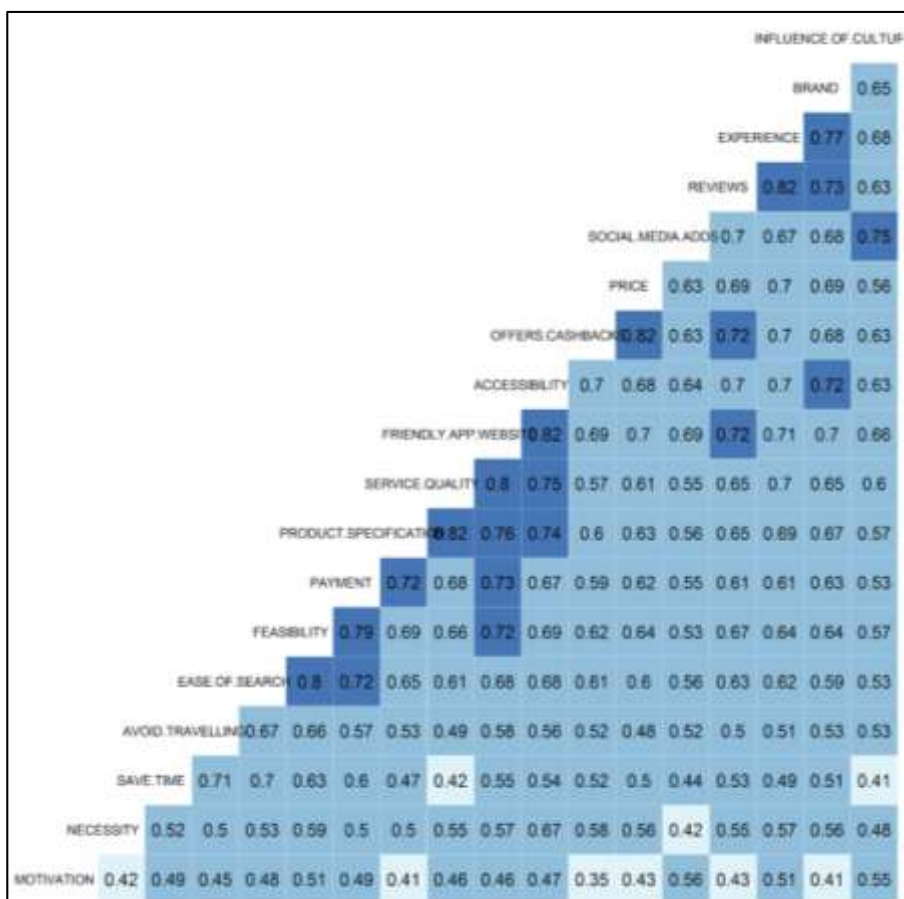

Fig 2

The lightest shade of blue indicates a weak correlation (<0.45) between the factors. From the responses of our study, we found that there is a high correlation (>0.82) between the factors- product specification and service quality, friendly app website, and Accessibility, Offer cashback and price, Reviews, and experience.

The Chi-square test of independence was done with the x variable(did you rely more on shopping during covid 19) and the variable y(Age) to find if there is a relationship between these two variables.

Chi-square test of independence is defined as,

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

In the survey conducted, We had asked our respondents, "Did you rely more on shopping during the Covid-19 pandemic". This question was answered with either Yes or No. The age of the respondents was also recorded in the variable "Age". The variable Age had 6 groups 18-24,25-31,32-38,39-45,46-52 and above 52.

The hypothesis was constructed to test if respondents relying more on shopping online during covid-19 had a significant difference between the age groups.

Null Hypothesis: Respondents relying on online shopping during the Covid-19 pandemic were independent of the age groups of the respondents.

Alternate Hypothesis: Respondents relying on online shopping during the Covid-19 pandemic were dependent on the age groups of the respondents.

Table 1: Chi-Squared test

	DF	Chi-Squared Value	P-value	LOS
Pearson's Chi-Squared Test	5	15.15	0.009739	0.05

From contingency Table 1, the p-value (0.009739) obtained is less than 0.05, which indicates that there is a relationship between the variables x and y. Therefore, we reject Ho(Null hypothesis) and fail to reject H1(Alternate hypothesis). From the chi-square test, Respondents relying on online shopping during the Covid-19 pandemic is dependent on the age groups of respondents.
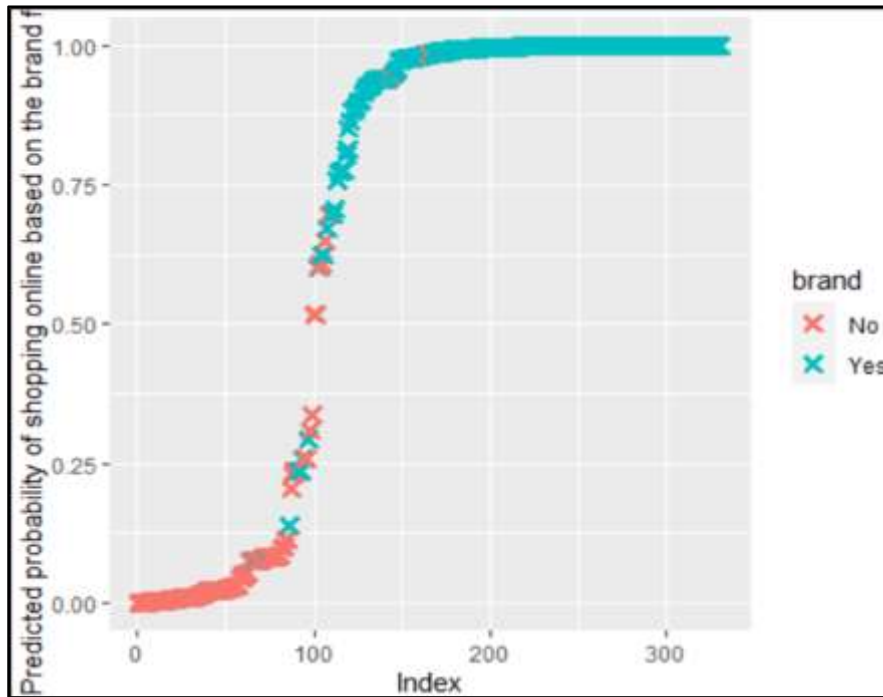
Fig 3

From the Survey responses to the missing factors in online shopping, the word cloud in Fig 3 answers to "What is missing in online shopping?"

The words "nothing", "products", "quality", "delivery" are the words that are highly common from the responses. The word with the highest frequency is "nothing", which signifies that there is nothing missing in online shopping and are satisfied with the current shopping experience. The word with the second-highest frequency is found to be "quality", this indicates that respondents are concerned about the quality of the products, and that is missing in online shopping. "Products"

And "product" also appears frequently in the responses, which indicates that they are less satisfied with the products available in online shopping.

Fig 4

The bar graph in Fig 4. Shows us the top ten frequent words that appeared in the responses to the survey question, "What are the changes to be made in online shopping?"

The bar graph clearly shows us that "products", "nothing", "quality", "delivery", "online", "time" are the words with the highest frequencies. The changes expected by the respondents are mainly dealt with the products, quality of the products, and the delivery of the products. Whereas, the word with the second-highest frequency is "nothing" which implies that most of the respondents do not expect any changes in online shopping and are satisfied with the present online shopping experience.

This study employed Logistic Regression in predicting the factors influencing the online shopping behavior of the Consumers. The model was created based on the response variable Brand, a factor that influences online shopping, and eight independent variables education, time, necessity, experience, accessibility, the influence of culture, gender, and social media.

Quality control checks by making sure all of the factor levels are represented by people shopping online on the response variable Brand and otherwise. For example, as listed a few in Table 2:

Table 2: Cross Tabs for Quality Control Check

Brand	Education			Necessity		Accessibility	
	Other	Postgraduate	Undergraduate	No	Yes	No	Yes
No	13	14	75	30	72	87	15
Yes	28	23	178	77	152	16	213

The model is employed to predict whether consumers purchase only branded products online concerning factors influencing the decision. We can fit a generalized linear model as the variables are of a categorical type.

The model provides us with the respective z-values, the null deviance, and residual deviance by which we found the *Mc Fadden's Pseudo R^2*.

McFadden's R-square is defined as:

$$R^2_{MF} = 1 - \frac{LL_{full}}{LL_0}$$

where, LLfull is the full log-likelihood model and LL0 is the log likelihood function of the model with the intercept only. The value of McFadden's R-square is found to be approximately equal to 79%. And this depicts that the model is a good fit. We can plot the predicted probabilities for each sample having shopped online based on Brand and color by whether or not they shopped online.

Logistic Regression yields deep predictions and obtains the best model yielding high and accurate results with an accuracy of 79%

Fig 5

In Fig 5. We found that more than 200 people have shopped online based on the factor brand. A c-statistic is found using the model and is 0.9869 or 99%, which indicates that the model is almost perfect. Further, we will work on the Hosmer-Lemeshow statistic.

To compute HL test, the data is divided into a number of subgroups, and the observed and expected frequencies in each of the group are calculated, and then the Chi-squared statistic calculated gives the Hosmer-Lemeshow statistic as

$$\chi^2_{HL} = \sum_{g=1}^{G} \frac{(O_g - E_g)^2}{E_g(1 - E_g/n_g)}$$

Where O_g is the observed events, E_g is the expected events and n_g signify the number of observations for the g' th group, and the G is the number of groups. The data is grouped in 10 groups, and the Hosmer and Lemeshow (HL) goodness of fit test is obtained as χ^2 HL = 2.9718, df = 8, p−value = 0.9361, from this it can be seen that the p-value is greater than 0.05 at 95% confidence level and it shows that the model is a good fit.

Table 3: Hosmer and Lemeshow Goodness of fit

Hosmer and Lemeshow goodness of fit (GOF)		
data: logistic$y, fitted(logistic)		
X-squared = 2.9718	df = 8	p-value = 0.9361

Classification models are models that predict a categorical label. We will build our model on the training dataset and evaluate its performance on the test dataset. This is called the holdout-validation approach to evaluating model performance. The data is split into two partitions, 70% for training and 30% for tests. The majority class of the target variable has a baseline accuracy of 69 percent. The predictions on the training set and Confusion matrix on training data are found divided by the number of rows of training data to get the accuracy of 95.66%, and predictions

on the test set are found to get an accuracy of 94%. The baseline accuracy for the data was 69 percent, while the accuracy on the training and test data was 96 percent and 94 percent, respectively. Overall, the logistic regression model is beating the baseline accuracy by a big margin on both the train and test datasets, and the results are very good.
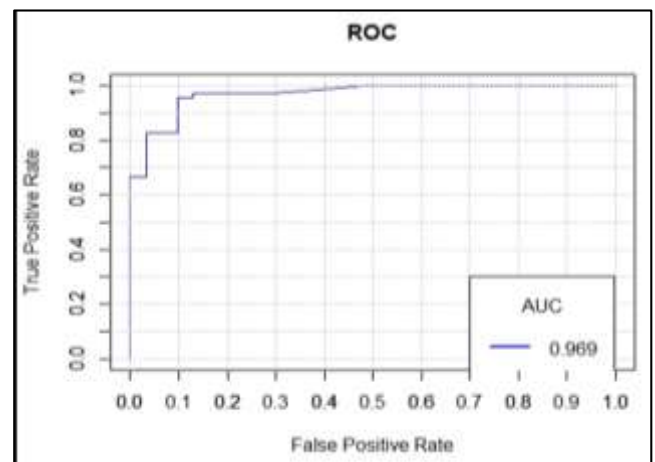

Fig 6

The ROC Curve (Fig 6.) demonstrates the amount of area covered by the predictive model graphically. It can be seen in Fig 6. how the true positive rate (specificity) is plotted against the false positive rate (1-specificity).The curve for the model associated with Brand as the predictor variable is above baseline and is close to the upper left corner, which the classifier performance approximately equals 97%.

Internal consistency describes the degree to which items in a dataset will measure a comparable variable and will link to the affinity of the rest of the items in that dataset. Usually, a Cronbach's alpha (α) value of more than 0.70 and closer to 1 is deemed to be a reliable score (acceptable). The internal consistency of the usage of different Online applications for shopping is considered (Table 4.), and Cronbach's alpha (α= 0.80) value for this dataset used in this paper demonstrating high reliability of the data collected.

Table 4: Cronbach's Reliability Test results

Results of reliability test (Cronbach alpha) (n = 331)		
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	Number of Items
0.8	0.8	6

The Cronbach's alpha for the subscales identified were all above 0.70. Cronbach's alpha values for the subscales are given in Table 5.

Table 5: Cronbach's Alpha test for subscales

Cronbach's alpha test for the subscales.	
Factor(Usage of Different Apps)	Cronbach Alpha
Amazon	0.77
Myntra	0.76
Flipkart	0.8
Zomato	0.76
Big Basket	0.76
Medlife	0.74

3. Conclusion

Nowadays, online shopping has become a regular part of an individual's life; optimizing e-commerce stores is important to have a better experience anticipated by potential customers. The findings of this research will help the e-commerce industry to create a positive experience that can result in higher revenues; the negative experiences might result in permanent loss of customers. The main objective of this paper was to determine factors that influence the consumers' disposition to purchase products from online stores. Based on the theoretical study, we conducted a survey among Bengalureans and analyzed the perception of the users, which were influenced by the factors when purchasing products online. Amid the COVID-19 spread in 2020, Online consumption habits are growing rapidly in cities like Bangalore, and the satisfaction level of using online products by customers is increasing each day because of the innovative technologies being used by the online shopping platforms. In this study, it is observed that many people do trust online shopping platforms, which is a plus point for such online platforms. This is one of the main factors influencing online shopping, that, whether to or not shop online.

4. References

1. A Study of Factors Affecting Online Buying Behavior: A Conceptual Model, Dr. Ujwala Dange, Prof. Vinay Kumar 2012.
2. Factors Influencing Online Shopping Behavior of Urban Consumers in India, Nittala Rajya Lakshmi 2019.
3. Online shopping: Factors that affect consumer purchasing behavior, Jozef Bucko, 2018.
4. Online Shopper Behavior: Influences of Online Shopping Decision, Chayapa Katawetawaraks, Cheng Lu Wang 2011.
5. Factors Affecting the Online Shopping Behavior: An Empirical Investigation in Vietnam, Ngo Tan Vu Khanh, Gwangyong Gim, 2014.
6. Exploring the Influential Factors of Online Purchase Intention in Finland, Heather Abrafi Agyapong, 2018.
7. Factors Influencing online shopping experience– A conceptual model and implications, M. Swapna, C. Padmavathy, 2017.
8. An Empirical Study on the Factors Affecting Online Shopping Behavior of Millennial Consumers, A sivakumar, A Gunashekharan 2017.
9. Logistic regression modelling for consumer purchase behavior of passenger cars by Balakrishnan Menon and Dr. Jagathy Raj V.P, AIMA Journal of Management & Research, August 2012, 6.
10. Determination of Factors Influencing the Decision on Purchasing Organic Food by Walailak Atthirawong 2017.
11. Factors Affecting Consumer's Online Shopping Buying Behavior by Anurag Pandey and Jitesh Parmar, 2019.
12. A review of factors affecting online consumer search behavior from an information value perspective, Robert Grant, Rodney J.Clarke, Elias Kyriazis, University of Wollongong, Australia, 2007.
13. An Empirical Study of Online Shopping Behaviour: A Factor Analysis Approach, Bikram Jit Rishi, Journal of Marketing & Communication, 2008.
14. Internet shopping and its impact on consumer behavior, Fayu Zheng, 2006.
15. Consumer behavior towards e-marketing: a study of jaipur consumers, Dr. Sanjay Hooda, Mr. Sandeep Aggarwal, 2012.
16. Rezai G, Mohamed Z, Shamsudin MN, Phuah KT. "Demographic and Attitudinal Variables Associated with Consumers' Intention to Purchase Green Produced Foods in Malaysia 2011.