

Gap Analysis among Customers and e-Customers in Retail Sector using Data Mining

Pradeep Singh Guleria¹, Dr. Harsh Sadawarti²

¹RIMT--Institute of Engineering and Technology, Punjab Technical University, Sirhind Side, Punjab, India
pradeepguleria91@gmail.com

²Director at RIMT- Institute of Engineering and Technology, PTU, Sirhind Side, Punjab, India
harshsada@yahoo.com

Abstract

We have been living in an era of rapid integration of IT technologies into business processes and people's day to day lives. Large volumes of functional and non functional data is being produced everywhere which companies desire to make use of. Data mining techniques and data analytics have become essential in making useful sense of this sea of data for any business to prosper. Taking an example of retail sector, internet users in India have been becoming increasingly aware and receptive towards online retail. Noticing this traditional retailers have been focusing on increasing their online presence and adapting to new marketing strategies thereby utilizing this medium. In this study we attempted to sense the changing needs of consumers while trying to identify the factors influential in affecting consumer perception and become causes for their satisfaction or dissatisfaction

Key Words: Data mining, online retailing, e-shopping, association

INTRODUCTION:

Data mining and data analytics deal with extraction of value-able information from large data sets. They are widely used together and most helpful in supporting decision making activities of a firm. The knowledge derived after analysis of data can be used to predict future trends, thus enabling firms to make more focused, knowledge-driven decisions. Data mining has found a lot of applications in the current business scenario due to its ability to process summarize and derive conclusions from large volumes of data.

These days many companies have been choosing to integrate data mining and analytics into their business processes to gain competitive advantage. A steady rise in growth rate of online retail sector in India can be seen in the past several years. Data mining helps formulate marketing strategies based on previous records to predict who will respond to new marketing campaigns such as online campaigns etc. Through this knowledge marketing personnel may have a clearly defined approach to sell profitable products to targeted consumers while providing them with a greater degree of satisfaction

1. RECENT ADVANCEMENTS IN RETAIL SECTOR:

A significant part of today's market scenario is influenced by the advancements in IT industry and increasing usage of internet as a medium for connecting with the consumer. Though India is an emerging economy on the world scale still the numbers of online shoppers in India has been rapidly increasing.

These days' companies have been diligently working on increasing their online presence and improving quality of customer relationships. Progress in IT technologies and the growing popularity of internet in India have led to computer based systems taking over a significant portion of retail world although they have not completely substituted the importance of physical stores in the market.

The lifestyle of Indian consumer has been changing steadily changing with time. Due to long working hours, comfort seeking lifestyle, and increasing usage of devices and gadgets utilizing IT technologies the consumer may avoid the effort and time required to go and buy at a physical store. Factors such as earnings, educational background and awareness also determine buying behavior of a customer. These days online shopping offers a wide variety of products to compare and choose from with the click of a button, the products are less expensive and they generally offer discounts and lucrative deals to the customer. This has not only decreased overall effort of the customer in a fast paced world but also given him/her an opportunity to choose from a wide variety of products and buy at lesser expense than the physical market while being in the comfort of his home.

It has been found that the numbers of users opting for online shopping has increased considerably compared to increase in number of internet users in recent years. Most of these consumers prefer to buy some selected products online because they will get heavy discounts in comparison to store purchases. Trust may also be a

significant factor in affecting customer behavior. It has been seen that sometimes the customer is not satisfied with the product which reaches his doorstep and finds unacceptable quality difference in the look and feel of the product. This affects his judgment of the brand value and may also affect his overall trust on online shopping as a reliable medium for purchase. If the customer has many such experiences or hears about similar experiences happening in his social group it's highly likely that he will not choose to shop online in future. The user may often be required to give his personal information while completing a transaction during shopping online. The misuse or misplacement of such information may result in the user losing his trust in online shopping medium.

Although the count of online shoppers has been rising, still physical stores have also maintained their stronghold in the market. An old Consumer -retailer relationship may amount for a certain degree of trust and might be one of the many factors for this fixation as many users are suspicious by nature and prefer to remain in the comfort zone of their buying behaviors. Unawareness among users and illiteracy as regarding the usage of internet may also be an influential factor.

Also when shopping at physical stores the customer does not have to go through time lags and delay issues between date of purchase and delivery. In many instances when the consumer needs to acquire the product immediately the consumer has no choice but to buy from a physical store. This factor of time delay between purchase and acquiring the object is not there when one shops directly from a store.

The customer experiences ease of use and flexibility while using this medium for purchasing. Seeing the increased inclination of customers towards online shopping the online retailers have also modified many of their business policies in attempts to improve quality of customer relationship and build trust. Implementation of some of these policies, such as "Cash on delivery" and "Return if not satisfied" are in an effort to strengthen trust relationship with consumers.

2. OBJECTIVES:

- 1) To recognize the core reasons for the users with high inclination towards online or personal shopping of products.
- 2) To recognize the causes of satisfaction or dissatisfaction amongst the users with respect to online shopping of products.

3) To establish the factors that influences the conception of the consumers in relation to online shopping of products.

4) To establish the pros and cons aspects of online shopping.

3. RESULTS AND DISCUSSIONS:

Various responses from e-customers have been collected according the questionnaire framed. The response collected is mined and analyzed that is presented below.

A. Demographic Profile

The distribution of respondents according to various socio-economic characteristics is described below:

Gender

Majority of e-shoppers are females as compared to males as listed in Table 1.

Table 1: Gender Demographic Profile

Demographic Profile	E-Shoppers (N=100)
Gender	Frequency
Male	38
Female	62

Age

A perusal of Table 2 below shows that highest proportion of e-shoppers belongs to age group of <25 years, followed by 25-60 years. The lowest proportion was of age >60 years. The mean age in case of males and females comes out to be 25 and 29 years respectively. This shows that during the student life e-shopping may be more as compared to other age groups. It may be due to the fact that in student life social interactions with friends increase shopping awareness and laidback lifestyle influences buying behavior.

Table 2: Age Demographic Profile

Demographic Profile	E-Shoppers (N=100)	
Age (years)	Male	Female
<25	25	36
25-60	12	20
>60	1	6
Mean	24.97	29.03

Marital Status

As illustrated in the Table 3 below, majority of people under consideration are single.

Table 3 Marital Status Demographic Profile

Demographic Profile	E-Shoppers (N=100)
Marital Status	Frequency
Single	61
Married	39

Occupation

The examination of 1 below shows that majority of the female shoppers are housewives or unemployed as in the place where the study has been conducted due to tradition unemployed females are mainly housewives.

Males who do service e-shop more as compared to females who do service. Businessmen and retired e-shoppers are approximately the same in males and females.

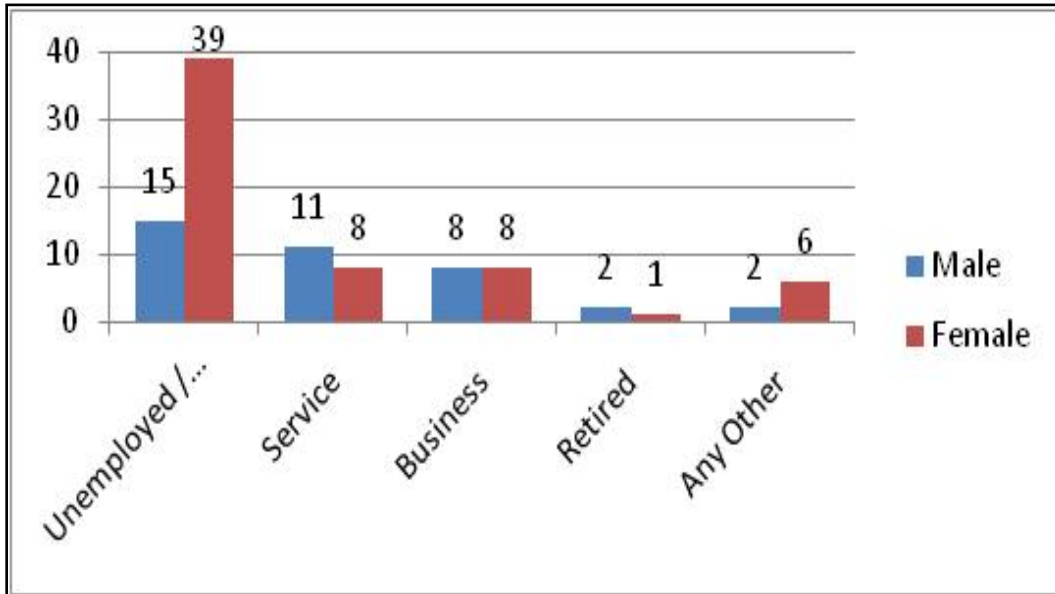


Figure 1 Occupation Demographic Profile

Annual Household Income

As depicted in the Figure 2 below, the highest portion of e-shoppers belongs to group with income in the range of 3-5 lacs, followed by the range 5-7 lacs then by 1-3 lacs whereas lowest is >7 lacs followed by <1 lac. The same proportion is followed among males and females.

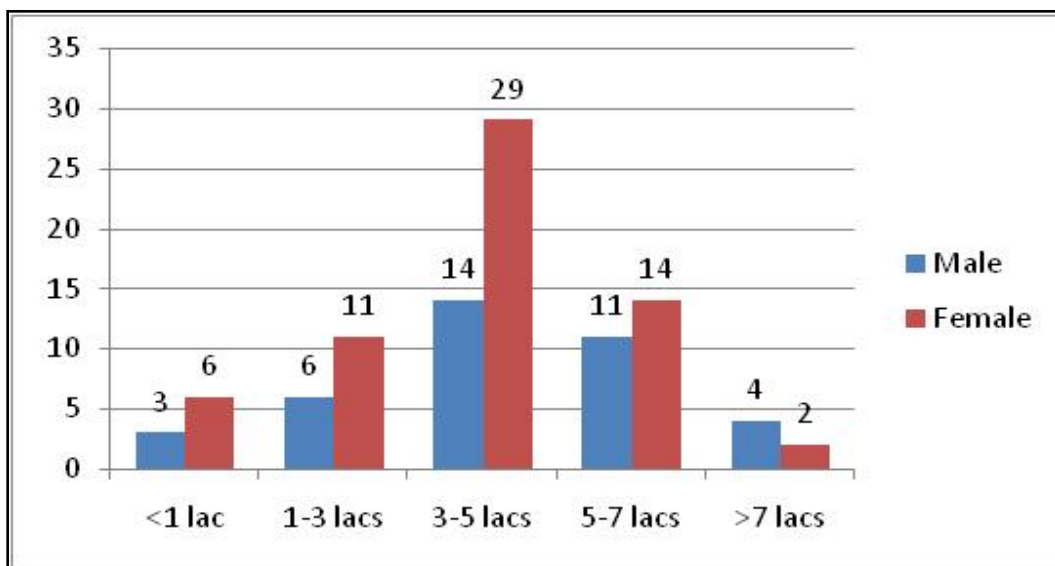


Figure 2 Annual Household Incomes Demographic Profile

Education

As illustrated in the Figure 3 below, the highest proportion of e-shoppers belongs to higher secondary group, followed by graduate and post graduate group whereas lowest proportion belongs to any other group followed by illiterate. The same proportion is followed among males and females.

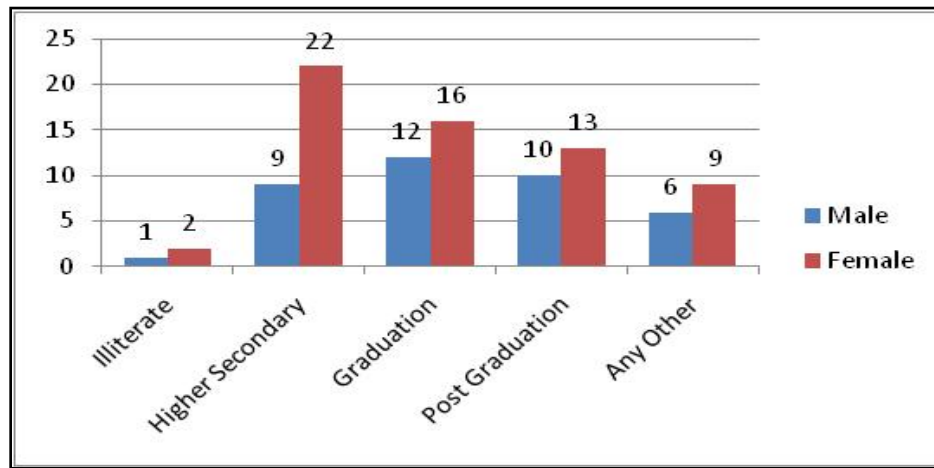


Figure 3: Educations Demographic Profile

B. Perception Male and Female E-Shoppers

It is clear from below Table 4 that there is a significant difference during e-shopping among males and females. According to females, parameters such as cost effectiveness, convenient replace and return policy, easy refund process, easy product selection, customer support service are highly significant in contrast to males as indicated by “***”.

Other parameters such as accurate/relevant information, better quality than retails, reliability of products, promotional schemes, variety of products, timely delivery of goods, judging the quality of products were found to be significant as indicated by “*”. This indicates that female perception is more significant as compared to males.

Other factors were found at par both for males and females.

Table 4 Perception of E-Shoppers

PARAMETERS	E-SHOPPERS		
	MALE	FEMALE	T-TEST
Website easy to navigate and operate	3.97	3.77	1.087
Website is clear with attractive design	4.11	4.21	0.638
Accurate/Relevant information	3.68	4.05	2.106*
Cost effectiveness	3.47	4.13	3.511**
Sales and discounts	4.11	3.89	1.292
Better quality than retails	3.74	4.15	2.597*
Reliability of products	3.53	3.92	2.478*
Branded products	4.16	4.27	0.702
Easy comparison of products	4.32	4.35	0.32
Promotional schemes	3.79	4.21	2.552*
Saves time	4.24	4.29	0.345
Ease of ordering process	4.05	3.92	0.823
Variety of products	3.76	4.26	2.68*
Timely delivery of goods	3.87	4.27	2.637*
Payment security	4.34	4.29	0.354
Secure personal information	4.13	3.82	1.865
Convenient replace and return policy	4.24	3.56	3.655**
Easy refund process	4.24	3.66	3.548**
Multiple payment options	4.37	4.35	0.091
Easy product selection	3.83	4.29	2.731**
Judging the quality of products	3.37	3.77	2.212*
Automated emails of new products	4.45	4.47	0.153
Customer support service	3.47	4.19	3.99**
Order tracking system	3.97	3.92	0.304
Feedback option	3.58	3.89	1.953

All the significant values have been represented in Figure 4. The values that were non-significant have been removed from the figure. It gives the clear picture of the perception.

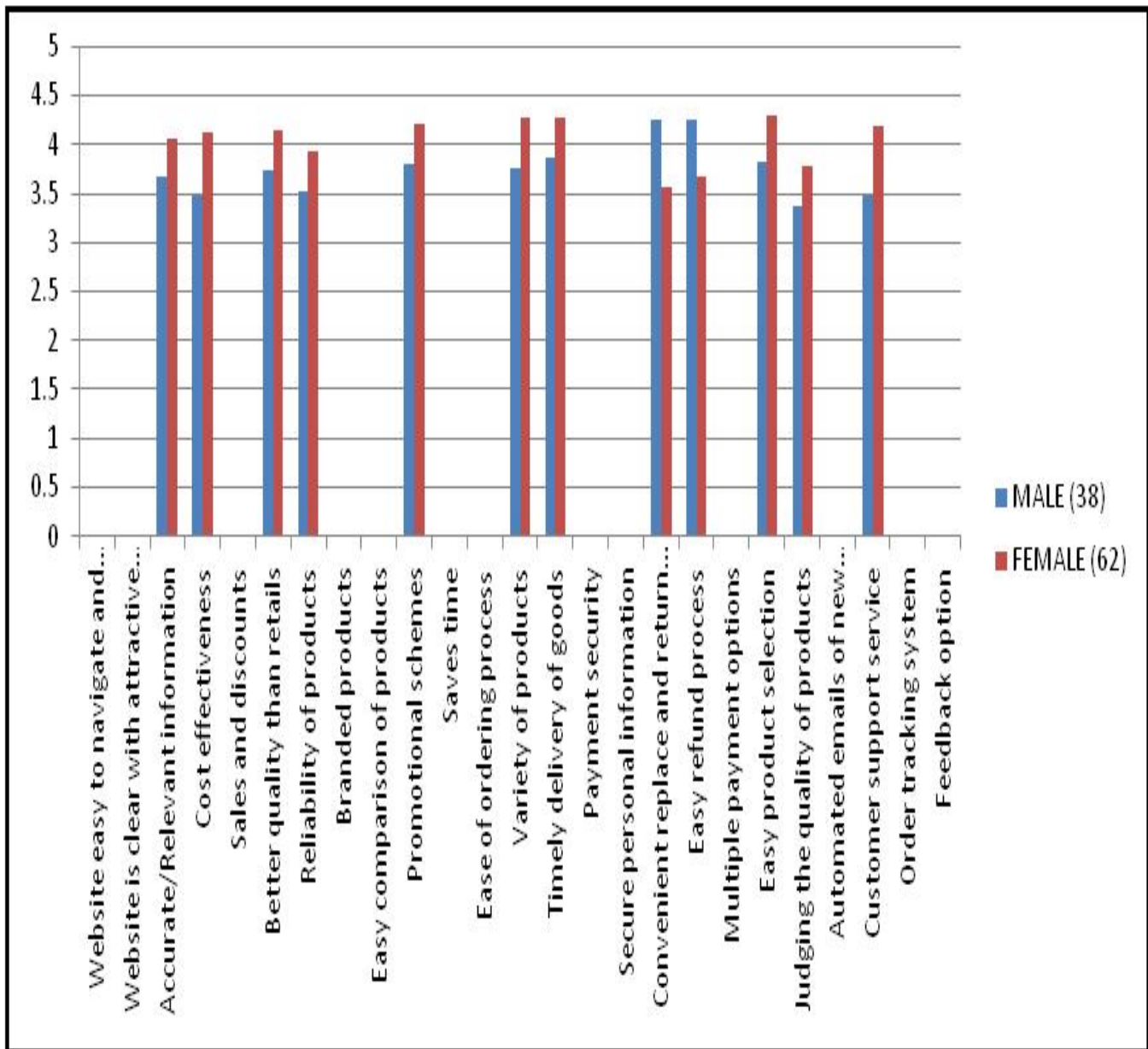


Figure 4: Perception of E-Shoppers

C. Perception of Singles and Married E-Shoppers

It is clear from Table 5 below that there is a significant difference according to the marital status.

According to married customers parameters such as accurate/relevant information, sales and discounts, better quality the retails, branded products, variety of products, secure personal information, convenient replace and return policy, easy product selection, judging

the quality of products are highly significant in contrast to singles as indicated by "***".

Other parameters such as cost effectiveness, reliability of products, ease of ordering products were found to be significant as indicated by "*". This indicates that married customers' perception is more influential as compared to singles customers.

Other factors were found at par both for single and married customers.

Table 5 Perception of Single and married e-shoppers

PARAMETERS	E-SHOPPERS		
	MALE	FEMALE	T-TEST
Website easy to navigate and operate	4.07	4.03	0.24
Website is clear with attractive design	3.95	3.77	1.134
Accurate/Relevant information	3.56	4.23	3.591**
Cost effectiveness	3.77	4.13	2.285*
Sales and discounts	3.64	4.23	3.489**
Better quality than retails	3.48	4.03	3.044**
Reliability of products	4.16	3.77	2.611*
Branded products	3.89	4.44	3.498**
Easy comparison of products	3.89	4.03	0.897
Promotional schemes	4.05	3.85	1.226
Saves time	4.31	4.33	0.147
Ease of ordering process	3.92	4.26	2.042*
Variety of products	3.77	4.23	3.091**
Timely delivery of goods	4.11	3.87	1.438
Payment security	3.95	4.13	1.153
Secure personal information	3.77	4.23	3.033**
Convenient replace and return policy	3.72	4.18	3.128**
Easy refund process	3.79	4.13	2.273
Multiple payment options	4.23	4.31	0.528
Easy product selection	3.28	3.97	3.873**
Judging the quality of products	4.07	3.46	3.404**
Automated emails of new products	4.33	4.41	0.569
Customer support service	3.57	4.21	3.381
Order tracking system	3.87	4.03	0.947
Feedback option	3.62	3.77	0.861

All the significant values have been represented in Figure 5. The values that were non-significant have been removed from the figure. It gives the clear picture of the perception of single and married people.

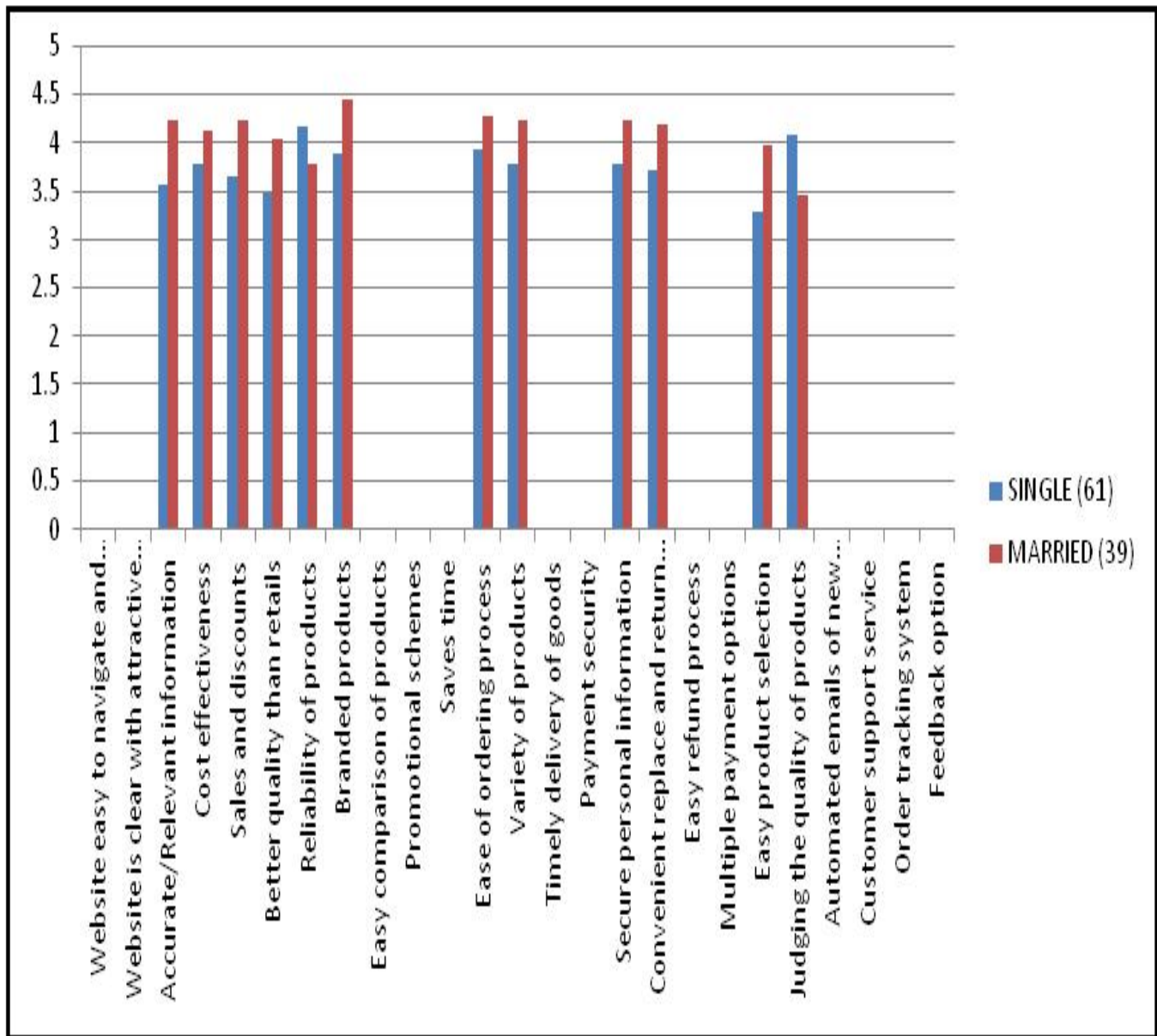


Figure 5 Perception of single and married customers

ASSOCIATION RULE MINING:

Association rule mining is one of the important problems of data mining. The goal of the Association rule mining is to detect relationships or associations between specific values of categorical variables in large data sets. This is a common task in many data mining projects.

Suppose I is a set of items, D is a set of transactions, an association rule is an implication of the form $X \Rightarrow Y$, where X, Y are subsets of I , and X, Y do not intersect. Each rule has two measures, support and confidence. Association rule mining was originally proposed in the domain of market basket data. The association rule mining on Market "Basket Data" is Boolean Association Rule Mining in which only Boolean attributes are considered.

We are using association rule mining on demographics data obtained through data collection to identify attributes which are associated with each other and how strongly do they hold in this scenario. By identifying these attributes we will be able to identify the primary features which influence buying behavior. By varying value of support and confidence while mining we will be able to narrow down our vision and identify how strong a rule holds in our observation.

We applied Apriori algorithm on the dataset to find associations while varying support and confidence values to our needs and found the following results.

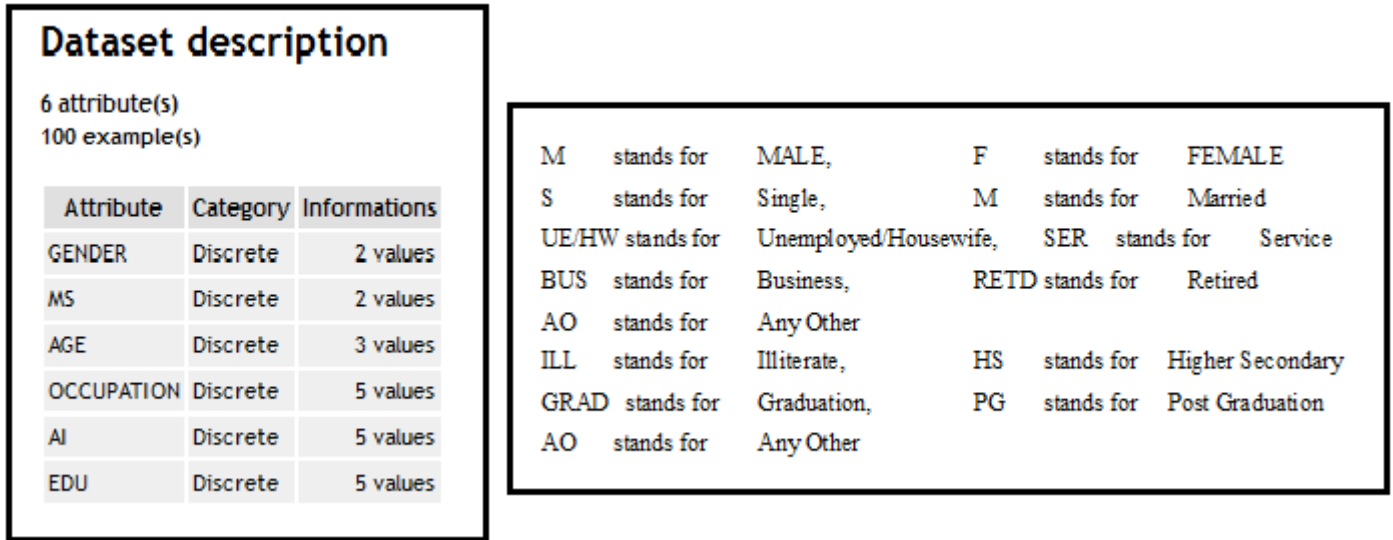
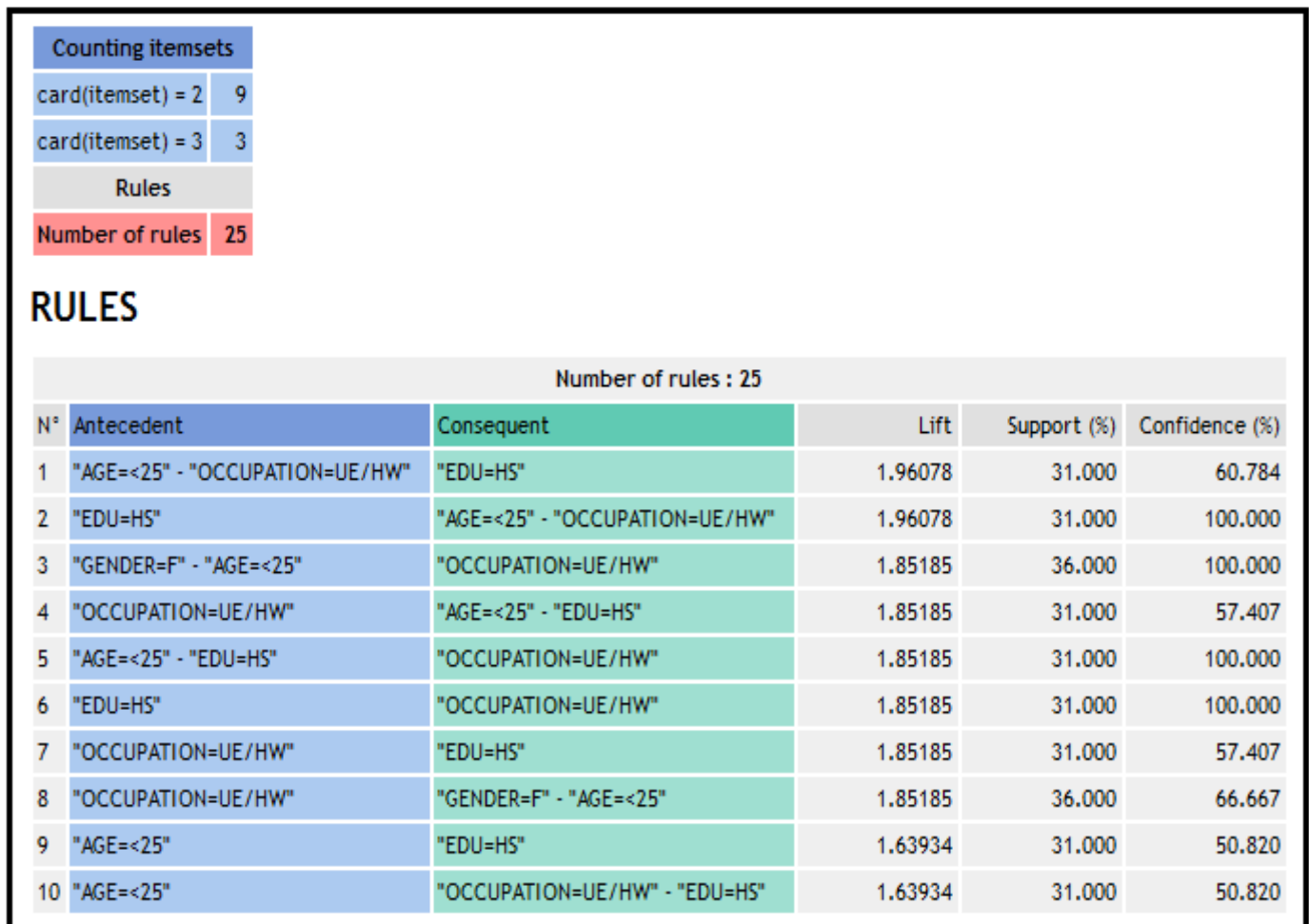


Figure 6 It shows dataset with number of rows, columns, attributes and neighbouring figure shows abbreviations used.



Support=0.3, Confidence= 0.5, Itemsets=6

Figure 7 It shows values of the parameters and results in 25 rules.

11	"EDU=HS"	"AGE=<25"	1.63934	31.000	100.000
12	"OCCUPATION=UE/HW" - "EDU=HS"	"AGE=<25"	1.63934	31.000	100.000
13	"OCCUPATION=UE/HW"	"AGE=<25" - "MS=5"	1.63399	30.000	55.556
14	"AGE=<25" - "MS=5"	"OCCUPATION=UE/HW"	1.63399	30.000	88.235
15	"AGE=<25"	"OCCUPATION=UE/HW"	1.54827	51.000	83.607
16	"OCCUPATION=UE/HW"	"AGE=<25"	1.54827	51.000	94.444
17	"MS=5" - "OCCUPATION=UE/HW"	"AGE=<25"	1.53689	30.000	93.750
18	"AGE=<25"	"GENDER=F" - "OCCUPATION=UE/HW"	1.51324	36.000	59.016
19	"GENDER=F" - "OCCUPATION=UE/HW"	"AGE=<25"	1.51324	36.000	92.308
20	"AGE=<25"	"AI=3-5"	1.25810	33.000	54.098
21	"AI=3-5"	"AGE=<25"	1.25810	33.000	76.744
22	"GENDER=F"	"OCCUPATION=UE/HW"	1.16487	39.000	62.903
23	"OCCUPATION=UE/HW"	"GENDER=F"	1.16487	39.000	72.222
24	"GENDER=F"	"AGE=<25" - "OCCUPATION=UE/HW"	1.13852	36.000	58.065
25	"AGE=<25" - "OCCUPATION=UE/HW"	"GENDER=F"	1.13852	36.000	70.588

Figure 8 Continuation of 25 rules

It shows various rules such as Age <25 and Education is HS gives support as 0.31 and confidence as 0.508 this may point towards the younger generation.

Similarly Gender as F, Age <25 and Occupation as UE/HW gives support as 0.36 and confidence as 1.000.

Income in the range of 3-5 lacks and age<25 gives 0.33 support and 0.76 confidence values.

Occupation=UE Gender =F gives 0.39 support and 0.72 confidence showing that females are frequent shoppers.

Marriage status = single, Occupation = UE/HW gives 0.3 support and 0.93 confidence this maybe due to attraction towards sales and discounts availability on online shopping medium.

Though 25 rules can be analyzed to identify primary factors but they are to be narrowed down significantly so that the rules which hold more strength and hence more significant in deciding the perception are identified.

Now we set **Support = 0.3, Confidence= 0.6, Itemsets=6**

RULES					
Number of rules : 17					
N°	Antecedent	Consequent	Lift	Support (%)	Confidence (%)
1	"EDU=HS"	"AGE=<25" - "OCCUPATION=UE/HW"	1.96078	31.000	100.000
2	"AGE=<25" - "OCCUPATION=UE/HW"	"EDU=HS"	1.96078	31.000	60.784
3	"AGE=<25" - "EDU=HS"	"OCCUPATION=UE/HW"	1.85185	31.000	100.000
4	"EDU=HS"	"OCCUPATION=UE/HW"	1.85185	31.000	100.000
5	"GENDER=F" - "AGE=<25"	"OCCUPATION=UE/HW"	1.85185	36.000	100.000
6	"OCCUPATION=UE/HW"	"GENDER=F" - "AGE=<25"	1.85185	36.000	66.667
7	"EDU=HS"	"AGE=<25"	1.63934	31.000	100.000
8	"OCCUPATION=UE/HW" - "EDU=HS"	"AGE=<25"	1.63934	31.000	100.000
9	"AGE=<25" - "MS=5"	"OCCUPATION=UE/HW"	1.63399	30.000	88.235
10	"AGE=<25"	"OCCUPATION=UE/HW"	1.54827	51.000	83.607
11	"OCCUPATION=UE/HW"	"AGE=<25"	1.54827	51.000	94.444
12	"MS=5" - "OCCUPATION=UE/HW"	"AGE=<25"	1.53689	30.000	93.750
13	"GENDER=F" - "OCCUPATION=UE/HW"	"AGE=<25"	1.51324	36.000	92.308
14	"AI=3-5"	"AGE=<25"	1.25810	33.000	76.744
15	"OCCUPATION=UE/HW"	"GENDER=F"	1.16487	39.000	72.222
16	"GENDER=F"	"OCCUPATION=UE/HW"	1.16487	39.000	62.903
17	"AGE=<25" - "OCCUPATION=UE/HW"	"GENDER=F"	1.13852	36.000	70.588

Figure 9 As the confidence value is increased, only 17 rules left

Next we further increase confidence and change **Support = 0.35, Confidence= 0.7, Itemsets=6**

Number of rules : 6					
N°	Antecedent	Consequent	Lift	Support (%)	Confidence (%)
1	"GENDER=F" - "AGE=<25"	"OCCUPATION=UE/HW"	1.85185	36.000	100.000
2	"AGE=<25"	"OCCUPATION=UE/HW"	1.54827	51.000	83.607
3	"OCCUPATION=UE/HW"	"AGE=<25"	1.54827	51.000	94.444
4	"GENDER=F" - "OCCUPATION=UE/HW"	"AGE=<25"	1.51324	36.000	92.308
5	"OCCUPATION=UE/HW"	"GENDER=F"	1.16487	39.000	72.222
6	"AGE=<25" - "OCCUPATION=UE/HW"	"GENDER=F"	1.13852	36.000	70.588

Figure 10: It shows only 6 rules are left.

5. CONCLUSION AND FUTURE SCOPE:

Association predicts which group of customers used to e-shop more than other. According to the data collection, it is predicted that females do more e-shopping than males. Maximum e-shoppers belong to age group <25 years which depicts that e-shopping is more during student life as they have a more active online presence

and are more aware. Housewives maybe frequent e-shoppers as they may not have time to visit retail stores or in some cases contrary to this they may be active online. Cost effectiveness, convenient replace & return policy, easy refund process and product diversity were found to be the influential factors for satisfaction among e shoppers. The general populace seems to be attracted

by discounts and easy product selection especially married people. In addition to this in married e shoppers Website security concerns and trust issues were found to be influential. Awareness of the internet medium is also a negative influential factor affecting the choices of aged people perhaps due to their lack of familiarity with the medium. By setting the values of support and confidence, every marketer can check the factors influencing consumer satisfaction rate and can hence take future decisions for betterment of services and growth.

The responses gathered from e-shoppers are confined to the specific region. It can be extended with wider area such as state or different states. Difference between various websites or states can be done. More powerful tools such as neural networks etc. or other data mining techniques can be applied for refined results. Sample size can be taken more in order to gather more response which may improve and further narrow down the results.

6. REFERENCES:

1. Gurleen, K. (2012), 'Consumer's perception towards online shopping- the case of Punjab', International Journal of Management & Information Technology, 1, (1), pp.1-6.
2. Kamaladevi, Vanithamani M. Exploration of E-tail business with reference to E-Customer Experience (2014).
3. Irene S.Y. Kwan*, Joseph Fong and H.K. Wong (2005) An e-Customer Behavior Model with Online Analytical Mining for Internet Marketing Planning.
4. Ki-Han Chung, Jae-Ik Shin (2008) The Relationship among e-Retailing Attributes, e-Satisfaction and e-Loyalty.
5. Lee, J. and Know, K. (2003), 'Concerns about payment security of Internet purchases: a perspective on current on-line shoppers', Clothing and Textiles Research Journal, 21, (4), pp.174-184.
6. Suresh, A. M. and Shashikala, R. (2011), 'Identifying Factors of Consumer Perceived Risk towards Online Shopping in India', Journal of Information and Financial Engineering, 12, pp.336-338.
7. Ramayah, T. (n.d.), 'Impact of Perceived usefulness, Perceived ease of use and Perceived Enjoyment on Intention to shop online', Available at: <http://ramayah.com/journalarticlespdf/impactpeu.pdf>.
8. Constantinides, E., (2004), 'Influencing the online consumer's behavior: the Web experience', Journal of internet research, 14, (2), pp.111-126.
9. Haq, U. Z. (n.d.), 'Perception towards online shopping: an empirical study of Indian consumers', International Journal of Electronic Commerce, 1, (8), pp. 9-20.
10. Hoffman, D.L., Novak, T. P. and Peralta, M., (1999), 'Building consumer trust online', Communications of the ACM, Vol. 42, (4), pp. 80-85.
11. Iftikhar M, Hussain M, Kahn Z, Liyas S. 'Social Class is myth or reality in buying behavior.' African Journal of Business Management Vol. 7(9), pp. 713-718, 7 March, 2013.
12. Lee, J, Y. (n.d.), 'Exploring the Influence of Online Consumers' Perception on Purchase Intention as Exemplified with an Online Bookstore', Available at: <http://www.jgbm.org/page/3%20Yu-Je%20Lee%20.pdf>, accessed on 09/05/2013.
13. Rastogi, K. A. (2010), 'A study of Indian online consumers & their buying behavior', Internet Research Journal, 1, (10), pp. 80-81.
14. P. Beatty, I. Reay, S. Dick, J. Miller, "Consumer Trust in E-Commerce Websites: A Meta-Study" ACM Computing Surveys, v. 43 no.3, 2011, pp.14:1-14:46.
- 15.