

## Supermarket Analysis Based On Product Discount and Statistics

Komal Kumawat\*

\*(Department of Computer Science and Engineering, Walchand Institute of Technology, Solapur University, Solapur, India)

### ABSTRACT

E-commerce has been growing rapidly. Its domain can provide all the right ingredients for successful data mining and it is a significant domain of data mining. E commerce refers to buying and selling of products or services over electronic systems such as internet. Various e commerce systems give discount on product and allow user to buy product online. The basic idea used here is to predict the product sale based on discount applied to the product. Our analysis concentrates on how customer behaves when discount is allotted to him. We have developed a model which finds the customer behaviour when discount is applied to the product. This paper elaborates upon how a different technique like session, click stream is used to collect user data online based on discount applied to the product and how statistics is applied to data set to see the variation in the data.

**Keywords** - —Customer Session, Data Mining, Product discount, standard deviation

### I. INTRODUCTION

Availability of huge data and need of converting such data into useful information and knowledge motivated the concept of Data Mining. Data Mining refers to extracting the useful information from large amount of available data. This retrieved information & knowledge is used for application ranging from market analysis, fraud detection, medical systems etc to production control and science exploration. [1] As Electronic commerce is growing fast, and with this growth companies are willing to spend more on improving the online experience, our research work helps the executive management to take valuable decisions related to particular product depending upon the customer's feedback. The research work involves designing and developing software which gives an edge over present e-commerce existing systems. We have created a system that learns the response provided by a customer to the stimulus giving to him as a discount. The system helps to understand the behavioral pattern of a customer so that the executive management can take expert right time steps to see that business continues in a straight forward manner without ups and downs. Further we also applied statistics to our data set to see how the data spread out from the mean. The algorithm required for this work is given in pseudo code

Step 1: Take username and password for User Login

Step 2: Initialize User session

Step 3: While user session is on

- Allow user to browse and purchase products
- Add products to customer shopping cart
- View final price
- Check Delivery Information

- Check Payment method
- Make Confirmation from User
- Store whole information into database
- Done

Step 4: Analyze stored data.

Step 5: Calculate number of products sold based on Discount

Step 6: Display results based on discount and based on Standard Deviation

After analyzing the results achieved, we see that if discount increases number of product sold also increases but at some point when product discount increases too much, product sale automatically decreases as it creates doubt in the customer mind regarding the quality of product. The statistical method, Standard deviation is also applied for the same dataset.

### II. LITERATURE REVIEW

Alexandru M. Degeratu, Arvind Rangaswamy, Jianan Wu, thought out that Are brand names more valuable online or in traditional supermarkets? Does the increasing availability of comparative price information online make consumers more price-sensitive? They address these and related questions by first conceptualizing how different store environments (online and traditional stores) can differentially affect consumer choices. They have used the liquid detergent; soft margarine spread, and paper towel categories to test their hypotheses [2].

Shahriar Ansari Chaharsoughi and Tahmores Hasangholipor Yasory strove to understand the impact of sales promotions on consumers' behavior. Sales promotions are action-focused marketing events whose purpose is to have a

direct impact on the behavior of the firm's consumers. Consumer behavior focuses on how consumers decide what to buy, why to buy, when to buy, where to buy and how often they buy, how frequently they use it, "how they evaluate it after the purchase and the impact of such evaluations on future purchases, and how they dispose it off". They developed a framework known as A-B-C-D paradigm to study and understand consumer behavior. The acronym A-B-C-D stands for the four stages of the paradigm namely access, buying behavior, consumption characteristics and disposal [3].

Sriram Thirumalai, Kingshuk K. Sinha focuses on the proverbial "last mile" of the retail supply chain – i.e., delivering products to the end-customer –and highlights the need for recognizing product type differences in configuring order fulfilment processes in electronic business-to-customer (B2C) transactions. They analyze the need to examine how the underlying dimensions of physical distribution service – availability, timeliness, and quality – should be integrated into the overall customer service package to best meet customer's expectations and needs [4].

Shu-hsien Liao, Pei-hui Chu, Yin-ju Chen, Chia-Chen Chang, have used online group buying. It is an effective marketing method. By using online group buying, customers get unbelievable discounts on premium products and services. This not only meets customer demand, but also helps sellers to find new ways to sell products sales and open up new business models, all parties benefit in these transactions. this study proposes a data mining approach for exploring online group buying behavior in Taiwan. Thus, this study uses the Apriori algorithm as an association rules approach, and clustering analysis for data mining, which is implemented for mining customer knowledge among online group buying customers in Taiwan. The results of knowledge extraction from data mining are illustrated as knowledge patterns, rules, and knowledge maps in order to propose suggestions and solutions to online group buying firms for future development [5].

Ralph-C bayer and Changxia Ke implement a simple two-shop search model in the laboratory with the aim to investigate if consumers behave differently in equivalent situations, where prices are displayed either as net prices or as gross prices with discounts. They compared two types of experimental treatments (in which the price in either of the shops was presented as a gross price with a discount) to their corresponding baseline treatments (where prices in both shops were given as net prices) [6].

### III. METHODOLOGY

We developed a model which contains authenticated administrative module and

user/customer module. Login module checks whether user has entered username and password. If without entering username or password user tries to submit then system does not allow him/her to enter into the system. When user provides correct username and password then only login is successful. The developed model is based on web mining and click stream analysis, along with this it also uses customer session and standard deviation.

**3.1 Customer Session:** As HTTP (Hyper Text Transfer Protocol) is a stateless protocol, it becomes very difficult to maintain the state [like when you login, the site remembers you until you logout]. So there are two ways to do it. Either store cookies at client side which stores the user information and other state related information. Session serves the same purpose but it is better than cookies as a user can disable cookies. For our research it is very important to keep a track of the customer once he/she gets logged in.

**3.2 Standard Deviation:** Standard deviation is the measure of spread over data, most commonly used in statistical practice when the mean is used to calculate a central tendency. Thus, it measures the spread around the mean. Because of its close links with the mean, standard deviation can be greatly affected if the mean gives a poor measure of central tendency.

**3.3 Data Collection:** Data collection is carried on through historical sources to set threshold points and online user data is used to generate the behavior of the customer. The data collection is part of every customer login. Customer logs into the system and his behavior is recorded by using click stream technique. Data is collected in user module, according to the data collection reports from online sources and historical data collected from super markets and media results are generated in the Admin module. The data collected from historical sources pinpoints the discount on various products and the online data is used to identify the customer behavior.

## IV. EXPERIMENTAL WORK

We implemented this idea using PHP/MYSQL software and above mentioned techniques. An administrative module has privileges to add product details into the database and he can also apply different discounts on different products.

### 4.1 ADMINISTRATIVE MODULE

Administrator has to enter username and password. After login administrator has different privileges like, view all category, add category, modify or delete category, view all manufacture, add manufacture, modify/delete manufacturer, add

product, and modify/delete product. Administrator can also apply discount to the products. He can see the order status whether it is delivered or in process. Reports are also generated in administrative module based on discount applied and standard deviation. The change password facility is also given to him/her. Using logout option administrator can leave the system. All privileges of administrator uses web page and each page is validated using JavaScript. These privileges are shown in figure 1. This figures also shows that one new category is added successfully, and it also shows all available categories along with modify/delete options.

4.1.1 Manufacturer: After adding details of category, manufacturer details can be added.



Fig 1 Administrator Module with privileges

As manufacturers are categorized based on category administrator has to select first category then list of manufacturer belong to that category will be displayed. He can add, modify and delete manufacturer based on particular category.

4.1.2 Product: Products are classified according to category and manufacturer, to add particular product administrator has to select first category name, then manufacturer name and then he/she can add product details. The Administrator has rights to modify or delete a product. The whole information related to product will be shown to administrator after selecting category and manufacturer name and he/she can change it and submit it. To delete a particular product he should select delete option then confirmation message is displayed to the administrator, if he clicks on "OK" then only product will be deleted. Modify/delete product is shown in figure 2.

4.1.3 Discount: The Administrator can apply discount on a particular product, as this thesis is based on discount applied to the products.

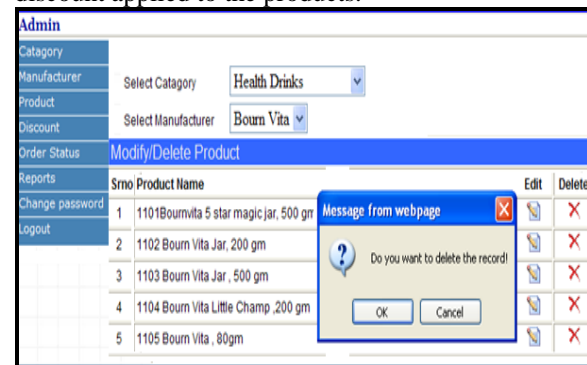


Fig 2 Modify/ Delete Product

The Administrator has to select a product name and discount. After clicking on submit button discount will be applied to the product.

Order Status: The Administrator can check order status whether it is delivered or in process. To see these reports of the administrator he has to select date duration. Accordingly date reports are generated.

## 4.2 User Module

When the customer visits a site, the first page displayed to him is login module. If the customer is already existing user then he/she can give username and password and if username, password is correct then he/she can login to the system. If the customer is new to the site then he/she should submit first form by clicking on "New User? Click Here to Create an Account" link. After clicking on this link the form will be shown to customer. The given form is validated for all fields. The form takes first name, last name, gender, birthdates, occupation is optional, correct email address, residential address including street1, street2, city, state etc, correct phone number, password and confirm password. After filling whole details, user has to click on "Please Sign In" link to login into system. Login module will be displayed to him/her. Then he/she should enter an email address as a username and his/her own password given while filling the form. These username and password are authenticated by system; if they are correct then privileges of user/customer are visible to customer otherwise system will not allow to login as it is unauthenticated user. User/customer privileges are Browse Products, Cart Contents, My account, My Order, Contact Us, Change password and Logout. All privileges of user are shown in figure 3 along with Browse Products.

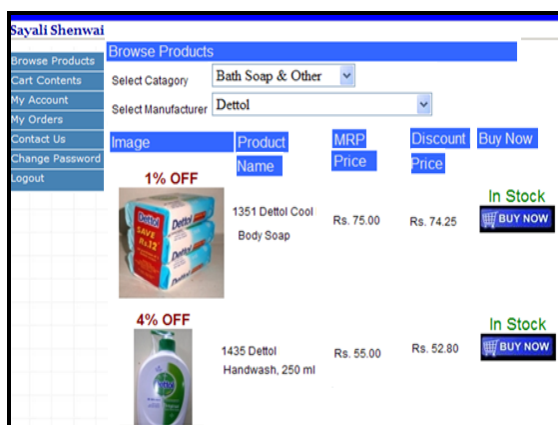


Fig 3 Browse Product

4.2.1 Browse Products: Users can browse any product based on category and manufacturer. Then he/she can select or click on the particular product to see the product details. After reading the details user can click on “Buy Now” button. After selecting “Buy Now” the user has to enter quantity for the selected product. Without entering quantity system does not allow checkout to user. Quantity should be only integer. After entering a correct quantity the user has to click on “Checkout” option. It navigates to the cart content page. From this page a user can “Select another Product” and can browse different products. All these products get added into cart content of the user. Cart Contents are shown in figure 4.

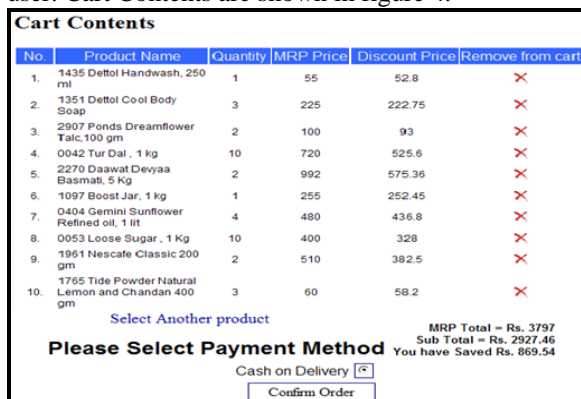


Fig 4 cart contents

4.2.2 Cart Contents: It shows the customer basket. If a customer wants to remove any product from cart he/she can do so by clicking on remove button shown with “Red Cross”. It asks user whether he/she really wants to delete a record? If the user answers positively then that product is removed from the cart. After selecting all products from a site a user has to select payment method and then he/she can click on “Confirm order” option. It navigates the user to final confirmation page. On this page the order information like shipping address, payment method and contact details are displayed for the user. The user can change shipping address by clicking on Edit

option. Then it shows form to change address with existing details present into the database. After clicking on Submit button “Address Updated successfully” message is shown to user. By clicking on “Get Product Order Information” cart contents will be displayed to the user. In the same way a customer can change contact information also and finally customer can click on “Confirm order” then it shows “Thank you” message to customer.

### 4.3 Results

Results are generated in administrative module. They are generated based on customer shopping in customer module. We have collected data from 472 customers and we have total 6,147 entries for 182 products. Based on these data results are generated. Results are of two types that are 1) Discount Based Results, 2) Standard Deviation Based Results.

4.3.1 Discount Based Results: As shown in figure 5 customer behaviour is examined. We observed that if discount increases, number of product sold also increases but at some point when product discount increases too much, product sale automatically decreases as it creates doubt in the customer mind regarding the product. Figure 6 shows graphical representation of the result.

Sr.No	Discount Applied	No. of Product Sold
1	1%	68
2	2%	71
3	4%	113
4	5%	114
5	8%	118
6	10%	121
7	12%	127
8	16%	130
9	18%	159
10	20%	162
11	22%	168
12	25%	189
13	27%	234
14	30%	276
15	32%	24

Fig 5 Discount Based Result

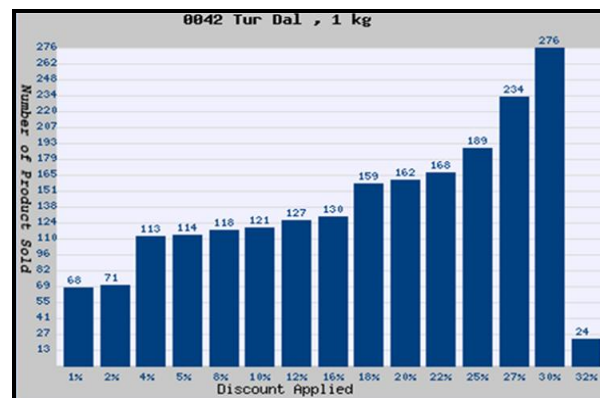


Fig 6 Graphical Representation of Discount Based Result

Sr.No	Discount Applied	No. of Product Sold
1	1%	83
2	2%	82
3	3%	82
4	4%	84

Sr.No	Discount Applied	No. of Product Sold
1	1%	78
2	2%	78
3	3%	76
4	4%	78
5	5%	79

Sr.No	Discount Applied	No. of Product Sold
1	1%	88
2	2%	64
3	3%	65
4	4%	64

Sr.No	Discount Applied	No. of Product Sold
1	1%	52
2	2%	53
3	3%	55
4	4%	57

Fig 7 People Likelihood regarding product

Figure 7 shows people likelihood of product, here same discount is applied for all four products. By looking at figure we can say that people likes Dove most when we compare it with other products.

4.3.2 Standard Deviation Based Results: Standard deviation is used to measure spread or dispersion around the mean of a data set. Standard deviation measures how concentrated the data are around the mean. A small standard deviation means that the values in the data set are close to the mean, on average, and a large standard deviation means that the values in the data set are farther away from the mean, on average. Without standard deviation, we can't get handle on whether the data are close to the average or whether the data are spread out over a wide range. In this dissertation work we considered discrete data that is different discount applied to the products, then the variance for the discrete variable made up of n observations is defined as ,

$$S^2 = \frac{\sum (x - \bar{x})^2}{n}$$

Where,

- x is data set of different discount applied to a particular product,
- n is number of observations
- $\bar{x}$  is mean of data set calculated as,
- $\bar{x} = \text{sum of } x / n$
- $S^2$  is the variance

Then, Standard Deviation S is calculated as follows,

$$S = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

In frequency table, frequency f is considered as number of product sold when discount is applied to the product. The formulas for variance and

standard deviation change slightly if observations are grouped into a frequency table. Squared deviations are multiplied by each frequency's value, and then the total of these results is calculated.

In frequency table, the variance for a discrete variable is defined as

$$S^2 = \frac{\sum (x - \bar{x})^2 f}{n} \quad \text{where: } n = \sum f$$

Where,

f is data set of number of product sold,  
The standard deviation for a discrete variable is defined as

$$S = \sqrt{\frac{\sum (x - \bar{x})^2 f}{n}}$$

As shown in figure 8 we can say that 68 % of data lies between 4.43 and 24.57. As shown in figure we can see that standard deviation calculated using discrete data, where X is discount applied to the product is close to standard deviation calculated using frequency table, where f is number of product sold.

## V. CONCLUSION

We created a system that learns the response provided by a customer to the stimulus giving to him as a discount. Our primary purpose was to show customer behaviour when discount is applied to the product. Our narrow focus is on click stream data.

Standard Deviation Result for 0042 Tur Dal, 1 kg						
N	Discount Applied (X)	Frequency (f) (No. of Product Sold)	$F_x = (F \cdot X)$	$X \cdot \bar{x}$	$(X - \bar{x})^2$	$F(X - \bar{x})^2$
1	1	68	68	-13.5	182.25	12393
2	2	71	142	-12.5	156.25	11093.75
3	4	113	452	-10.5	110.25	12458.25
4	5	114	570	-9.5	90.25	10288.5
5	8	118	944	-6.5	42.25	4985.5
6	10	121	1210	-4.5	20.25	2450.25
7	12	127	1524	-2.5	6.25	793.75
8	16	130	2080	1.5	2.25	292.5
9	18	159	2862	3.5	12.25	1947.75
10	20	162	3240	5.5	30.25	4900.5
11	22	168	3696	7.5	56.25	9450
12	25	189	4725	10.5	110.25	20837.25
13	27	234	6318	12.5	156.25	36562.5
14	30	276	8280	15.5	240.25	66309
15	32	24	768	17.5	306.25	7350
N= 15	Sum of X= 232	Sum of F (n)= 2074	sum of $F_x = 368$			Sum of $F(X - \bar{x})^2 = 202112.5$
Mean ( $\bar{x}$ )= Sum of X / N $\bar{x} = 232 / 15 = 14.5$						
Sum of $(X - \bar{x})^2 = 1521.75$						

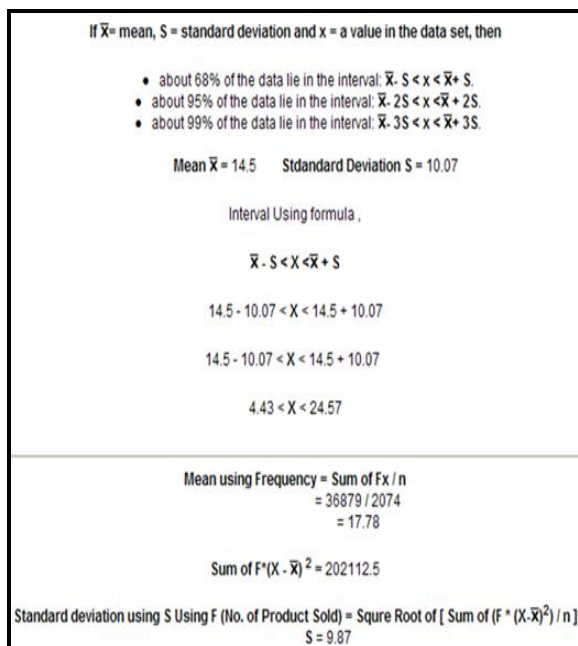


Fig 8 Standard Deviation Based result

Based on the click stream technique collect the information and store it into database. By using this information generated results shows customer behavior on the product purchase when discount is applied to it.

Standard deviation measures how concentrated the data are around the mean. The standard deviation can be difficult to interpret as a single number on its own. The findings of this study are, first for the same discount, a particular product gains more sales. As we compared the results of different bath soap like, Dove, Pears Blue, Pears Orange and Pears green. As shown in the result people likes Dove most as compare to other soap. This also shows the customer liking of product. As discount increases it is found that product sale also increases. We can also state that, for branded products discount never increases to large extent. Second is, irrespective of the discount applied to some products the sale remains constant. We analyze that, the product sale of ponds dream flower talc remains constant though discount increases. This also shows the people belief regarding particular product. Third is that more comfortable product in terms of cost and weight getting sold more easily and to a great extent? As we compared the product sale of Bournvita 80gm and Bournvita jar 500 gm, we observed that the product sale of Bournvita 80 gm is more than Bournvita jar 500 gm.

## VI. FUTURE SCOPE

This research work is related to the prediction of product sale based on discount applied to the product. As discount increases product sale

also increases but if product sale increases to large extent then automatically product sale goes down. This research can be extended to compare the product sale based on discount and without discount. We can analyze the product sale when discount is not applied to any product i.e. it's actual MRP price and also analyze the product sale when discount is applied to it.

Thus we can compare the results of both and our prediction can be determined that if discount is applied then product sale automatically increases as compare to it without discount.

## REFERENCES

### Books:

- [1] Han, I, & Kamber M, *Data mining: Concepts & Techniques* (2nd Edition, Morgan Kaufmann, 2006)

### Journal Papers:

- [2] Alexandru M. Degeratu, Arvind Rangaswamy, Jianan Wu, "Consumer choice behaviour in online and traditional supermarkets: The effects of brand name, price, and other search attributes", *International Journal of Research in Marketing, Volume 17, Issue 1*, 31 March 2000, Pages 55-78
- [3] Shahriar Ansari Chaharsoughi and Tahmores Hasangholipor Yasory, "Effect of sales promotion on consumer behavior based on culture", *African Journal of Business Management, Vol. 6(1)*, pp. 98-102, 11 January, 2012 Also Available online at <http://www.academicjournals.org/AJBM> DOI: 10.5897/AJBM11.739 ISSN 1993-8233 ©2012 Academic Journals.
- [4] Sriram Thirumalai, Kingshuk K. Sinha, "Customer satisfaction with order fulfillment in retail supply chains: implications of product type in electronic B2C transactions", *Journal of Operations Management* 23 (2005) 291-303, doi:10.1016/j.jom.2004.10.015
- [5] Shu-hsien Liao, Pei-hui Chu, Yin-ju Chen, Chia-Chen Chang, "Mining customer knowledge for exploring online group buying behaviour", *Expert Systems with Applications, Volume 39, Issue 3*, 15 February 2012, Pages 3708-3716
- [6] Ralph-C bayer and Changxia Ke, "Discounts and Consumer Search Behaviour: The Role of Framing", *The University of Adelaide Australia school of Economics, Research Paper No-2010-21*