Applying Statistical Techniques To Analyze Discount Based Data Of Shopping Malls

Komal V.Kumawat

ME CSE Student

Walchand Institute of Technology, Solapur

Prof. L.M.R.J.Lobo

Associate Professor

Walchand Institute of Technology, Solapur

Abstract

In recent years, people's life and living styles have been extremely inclined by the Internet, which enables electronic commerce for companies and varies small scale fields to perform electronic transaction. Computers and software play an important role in working of such kind of business. Large amount of data is generated with the use of software and these dataset need to be analyzed for useful information which will give benefit to the organisation. Data mining is the approach that fulfils this requirement. It is the process of applying mathematical, statistical technique on large quantity of data. We 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.

1. Introduction

There are different types of electronic commerce such as business to business, business to consumer, business to employee, consumer to consumer and consumer to business. Our research work is based on businesses to consumer model. The idea basically is to create a system that learns the response provided by a customer to the stimulus giving to him as a discount. The system will help to understand the behavioural 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.

The methodology used is based on monitoring the user activities by creating user login and initializing the user session. While user is logged on to the system a user can select the product based on discount applied to the product. We need to save all details regarding a user and his purchase of product into the database. By analyzing these details we generate the graph based results for each product based on discount. 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 analysing 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. This analytical work will definitely be useful to executives of supermarkets who can take decisions on the product retention.

2. 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 [1].

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 [2].

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 [3].

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 [4].

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) [5].

3. Methodology Adopted

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.

4. Data Collection

Data collection is carried on through historical sources to set threshold points and online user data is used to generate the behaviour of the customer. The data collection is part of every customer login. Customer logs into the system and his behaviour 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 behaviour.

5. Implementation

We implemented this idea using above mention techniques. An administrative module has privileges to add product details into the database and he can also apply different discounts on different products. The login module for administrator is shown in figure 1.
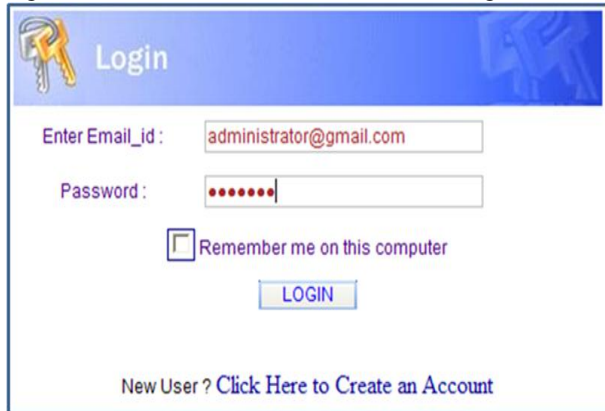

Figure 1. Login Module for administrator

After login administrator has different privileges as shown in figure 2.
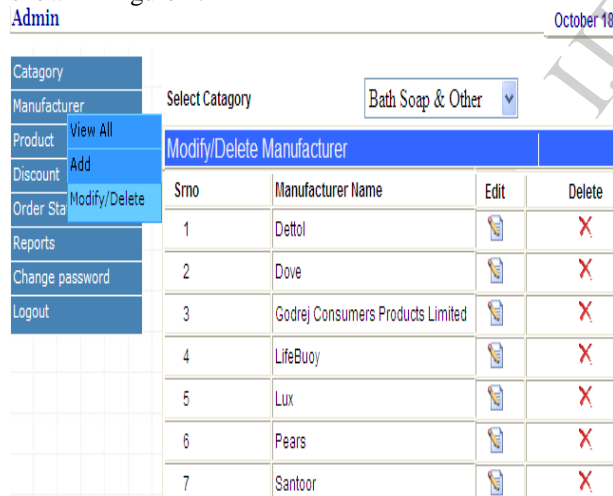

Figure 2. Administrator Privileges

As products are classified based on category and manufacturer, administrator has rights to add, modify or delete category, manufacturer, and product. He can also apply discounts to the products. The changes made by administrator are visible to the customer in customer module. Based on discount applied to the product customer purchase the product and confirms the order. The whole data of customer module is saved into database. Based on this data results are generated

in administrator module. 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.

When customer/user first time visits to the site he has to fill registration form and then his username and password will be created. Then after confirmation user can login to the system. After correct login customer is able to browse and purchase different products as shown in figure 3..
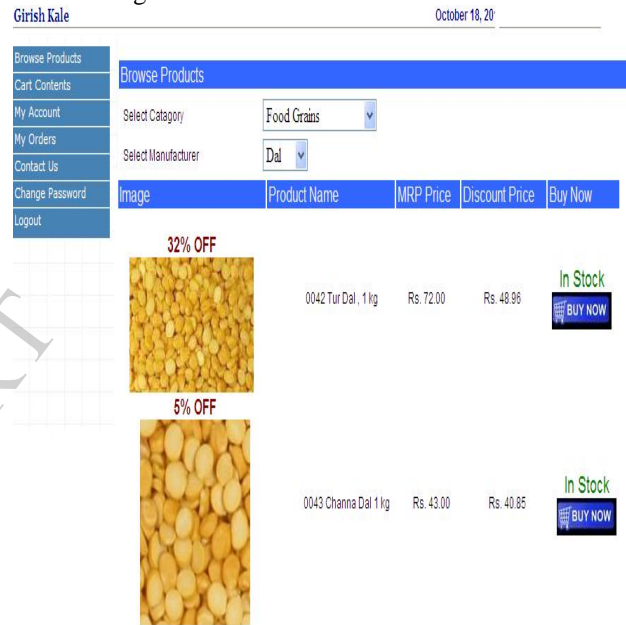

Figure 3. Customer Privileges along with Browse Products

After clicking on "Bye Now" option customer has to enter quantity and submit then the product gets added into the shopping cart and user can select more products. All the products get added into the shopping cart of the customer, finally he has to confirm order as shown in figure 4. After confirming order the whole data get saved into the database. User also has privilege to see his/her account information by clicking on "My Account". He/she can see his/her order details by clicking on "My order" privilege. They can send their query by clicking on "contact us" option. They can also change their password. By clicking on logout option they can leave the system.

Cart Contents

No.	Product Name	Quantity	MRP Price	Discount Price	Remove
1.	0044 Kabuli Channa Big, 1kg	1	65	62.4	X
2.	0056 Massor Dal(Black) 1 kg	1	65	63.05	X
3.	0046 Moong Dal Chilka, 1 Kg	1	114	109.44	X
4.	0193 Hira Besan, 250 Gm	2	22	21.56	X
5.	0063 Ramdev Haldi Powder, 100gm	1	21	20.79	X

MRP Total = Rs. 287
Sub Total = Rs. 277.24
You have Saved Rs. 9.76

Select Another product

Please Select Payment Method

Cash on Delivery ○

Confirm Order

Figure 4. Cart Contents

5.1 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 6,196 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. The developed software is generalized. Results are generated for all 182 products. Discount based result for product "Loose Sugar" is shown in tabular form in figure 5. Its graphical representation is shown in figure 6. As shown in figure we can 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 product. 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 of the data set, 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.

Figure 7 shows standard deviation result for "loose

Admin

| Catagory |
| Manufacturer |
| Product |
| Discount |
| Order Status |
| Reports |
| Change password |
| Logout |

Report

Select Product 0053 Loose Sugar , 1 Kg ▾

Submit   Reset

Standerd Deviation Result
Show Graph Based on Disount Applied

Sr.No.	Discount Applied	No. of Product Sold
1	1 %	73
2	2 %	90
3	3 %	95
4	4 %	103
5	5 %	105
6	6 %	109
7	7 %	112
8	8 %	127
9	9 %	129
10	10 %	133
11	12 %	144
12	14 %	166
13	16 %	181
14	18 %	200
15	20 %	224
16	22 %	84

Figure 5 Discount Based Result for Loose Sugar
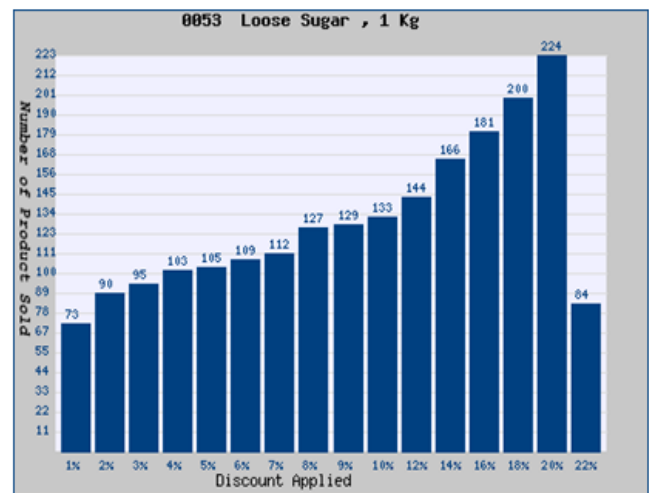
Figure 6 Graphical representation of figure 5

Sugar" product. We can say that 68 % of data lies between 2.84 and 15.83. 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.

	Standard Deviation Result for		0053 Loose Sugar , 1 Kg			
N	Discount Applied (X)	Frequency (F) [No. of Product Sold]	F_x =(F*X)	$X-\overline{X}$	$(X-\overline{X})^2$	$F(X-\overline{X})^2$
1	1	73	73	-8.24	67.82	4950.87
2	2	90	180	-7.24	52.35	4711.45
3	3	95	285	-6.24	38.88	3693.49
4	4	103	412	-5.24	27.41	2823.06
5	5	105	525	-4.24	17.94	1883.46
6	6	109	654	-3.24	10.47	1140.92
7	7	112	784	-2.24	5	559.61
8	8	127	1016	-1.24	1.53	193.8
9	9	129	1161	-0.24	0.06	7.14
10	10	133	1330	0.76	0.58	77.78
11	12	144	1728	2.76	7.64	1100.68
12	14	166	2324	4.76	22.7	3768.6
13	16	181	2896	6.76	45.76	8282.79
14	18	200	3600	8.76	76.82	15364.01
15	20	224	4480	10.76	115.88	25956.87
16	22	84	1848	12.76	162.94	13686.77
N = 16	Sum of X = 157	Sum of F (n)= 2075	sum of F_x = 23296			Sum of $F(X-\overline{X})^2$ = 88201.29

Mean ($\overline{X}$)= Sum of X / N

$\overline{X}$ = 157 / 16 = 9.24

Sum of $(X-\overline{X})^2$ = 653.77

Standard Deviation S Using X(Discount Applied)= Sqaure root of [(Sum of $(X-\overline{X})^2$ / N]

S = 6.39

If $\overline{X}$= mean, S = standard deviation and x = a value in the data set, then

- about 68% of the data lie in the interval: $\overline{X}$- S < x < $\overline{X}$+ S.
- about 95% of the data lie in the interval: $\overline{X}$- 2S < x < $\overline{X}$ + 2S.
- about 99% of the data lie in the interval: $\overline{X}$- 3S < x < $\overline{X}$+ 3S.

Mean $\overline{X}$ = 9.24   Stdandard Deviation S = 6.39

Interval Using formula ,

$\overline{X}$ - S < X < $\overline{X}$ + S

9.24 - 6.39 < X < 9.24 + 6.39

9.24 - 6.39 < X < 9.24 + 6.39

2.84 < X < 15.63

Mean using Frequency = Sum of Fx / n

= 23296 / 2075

= 11.23

Sum of $F*(X-\overline{X})^2$ = 88201.29

Standard deviation using S Using F (No. of Product Sold) = Squre Root of [Sum of $(F*(X-\overline{X})^2)$ / n]

S = 6.52

Figure 7 Standard Deviation Based Result

6. 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 behavior when discount is applied to the product. Our narrow focus is on click stream data. 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.

7. 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.

We have developed a module which gives only "Cash on Delivery" facility to customers. This work can be extended to allow user to do transaction online using PayPal or any other online transaction facility.

8. References

[1] 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

[2] 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.

[3] 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

[4] 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

[5] 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