A Study to Improve the Response in Email Campaigning by Comparing Data Mining Segmentation Approaches in Aditi Technologies

1*P. Theerthaana, 2S. Sharad

1 Department of Marketing, Anna University, Tamil Nadu, India

2 Aditi Technologies, Marketing, Tamil Nadu, India

ABSTRACT:
Email marketing is increasingly recognized as an effective Internet marketing tool. In this study, a questionnaire is constructed and distributed to a sample of 146 prospects of Aditi Technologies to find the factors associated with higher response rates. The collected data is analyzed using Factor Analysis and the 11 factors, From Line, Subject Line, Personalization of the subject line, Timings for sending mails, Frequency of mailing, Length of the Emails, Incentives to respond, Pre-existing Business Relationship, Permission based emails, Links and Image are extracted and it explains 78.363% of variance. These 11 factors is analyzed using Multiple Linear Regression and the .922 R square value indicates that 9 independent variables, Permission based emails, Length of Email, Timings, From Line, Frequency of mailing, Preexisting Business, Personalization, Incentives to respond, Subject Line contributes to higher response rate.
This study also investigates marketing campaigns of Aditi Technologies using RFM, CHAID, and logistic regression segmentation methods. One-way ANOVA is used to analyze the data and it is found that there exists no difference between the three approaches.
The study concludes that RFM is the most commonly used segmentation approach, however RFM may focus too much attention on transaction information (recency, frequency, and monetary value) and ignore individual difference information (e.g., values, motivations, lifestyles) that may help a firm to better market to their customers. This consideration would favor analytical techniques such as CHAID and logistic regression that can accommodate a variety of personality and individual difference information.

Keywords: *Data mining segmentation, RFM (Recency, Frequency, and Monetary value), CHAID, Logistic regression, Email, Marketing campaigns, Response rate*

INTRODUCTION

E-mail marketing is a popular marketing communications tool. Over 80% of US marketers (Forrester, 2005) and 90% of Canadian marketers (Inbox Marketer, 2005) are doing some form of e-mail marketing. But few companies are deploying high volumes.

E-mail marketing has the highest ROI effectiveness rating of any direct marketing medium. Most marketers agree that e-mail has excellent cost efficiency and on a cost per response basis, e-mail marketing is ranked number one. But achieving a higher response rate in email campaigning is a big challenge for the marketers. This study attempts to determine

*Corresponding Author, Email: *sahanatheerthee@gmail.com*

effective ways to improve the response rate in email campaigning. This study is done in Aditi Technologies which is an IT services company partnered with Microsoft, selling out Microsoft cloud services and Microsoft SharePoint services in US, UK and INDIA. They use email-marketing method only for their business development purposes.

Mass emailing is one among their marketing operations using an online tool called HubSpot. Email Campaigns is used for:
✓ Promotional Messages
✓ Newsletters
✓ Official Announcement
✓ Event invitations, Pre and Post follow up of the Event

Where the above said is solely used only for business development. But currently they are able to achieve only 10% of HTML opens or Number of opens. This led to a poor Click Through or Unique -Click rate, which reduced their business development efficiency by 45 %. So this study aims at proposing a model to achieve a high rate of HTML Opens or Number of Opens and thereby increase the Click Through or Unique -Click rate.
The objective of the study is

✓ To determine the major factors that affect the response rates in Email Campaigning and this information could be utilized to achieve higher response rate in email campaigning.
✓ To investigate RFM, CHAID, and Logistic Regression as analytical methods for email marketing segmentation and to determine the most efficient data mining segmentation method for Aditi Technologies.

Literature Review
Analytical Segmentation Methods in Data Mining

Segmentation in Email Campaigning is a means of dividing the email list based on interest categories, purchasing behavior, demographics and more for the purpose of targeting specific email campaigns to the audience most likely to respond to your messaging or offer. This list segmentation and targeting efforts pay off in higher open and click-through rates (Keegan, 2012). Segmentation in direct marketing has become more efficient in recent years because of the development of database marketing techniques. These data-mining approaches provide direct marketers with better ways to segment their current customers and develop marketing strategies tailored to particular segments and/or individuals. Over the recent years, database-marketing techniques have evolved from simple RFM models (models involving recency of customer purchases, frequency of their purchases, and the amount of money they have spent with the firm) to statistical techniques such as chi- square automatic interaction detection (CHAID) and logistic regression. More recently, neural network models are employed in the database-marketing arena (Yang, 2004).

A study suggests that various data mining techniques can be useful for efficient customer segmentation and targeted marketing. Variants of RFM-based predictive models are constructed and compared to classical data mining techniques of logistic regression, decision trees, and neural networks. RFM is found to be a better statistical practice and Logistic regression can include many variables (Olson et al., 2012). In spite of recent statistical advances in data mining, marketers continue to employ RFM models. A study by Verhoef et al. (2002) shows that RFM is the second most common method used by direct marketers, after cross tabulations, in spite of the availability of more statistically sophisticated methods. There are a couple of related reasons for the popularity of RFM. As Kahan (1998) notes, RFM is easy to use and can generally be implemented very quickly. Furthermore, it is a method that managers and decision makers can understand (Marcus, 1998). This is an important consideration in that a successful technique for a direct marketer is one that differentiates likely responders to a particular mailing from those who are unlikely to respond, yet does so in a way that is easy to explain to decision makers. However, it has been argued that the simplicity of RFM has been overemphasized, but its ability to differentiate, relative to statistical techniques, has not been considered to the extent that it should be (Yang, 2004).

RFM can help identify valuable customers and develop effective marketing strategy for not only profits organizations but also non-profit organizations and government agencies. Through the application of RFM model, decision

makers would gain insights on RFM and would be able to apply RFM more effectively to resolve the problems encountered in daily activities and develop effective strategy to satisfy a wide variety of customer needs (Wei et al., 2010). Asllani and Chattanooga Diane Halstead conducted a study and found that the proposed linear programming model identifies the customer segments based on RFM profile, which should be targeted in order to maximize profitability. The proposed model also identifies the RFM segments, which are not worthy of pursuing either due to unprofitability or due to an insufficient campaign budget (Asllani and Halstead, 2011). The RFM model, which is used to identify the customer behavior using a recency, frequency and monetary together with customer life time value (LTV) model can more effectively segment and target valuable customers than random selection (Chan, 2008). RFM captures the effects of past marketing activities and the original marketing impact, which is represented by temporal changes from the purchase process and also there exists a relationship between RFM and marketing (Reimer and Albers, 2011). CHAID-based approach is useful in detecting classification accuracy heterogeneity across segments and also gives a better insight into factors influencing customer behavior (Antipov and Pokryshevskaya, 2009). A study conducted in children's dental clinic indicates that RFM (recency, frequency, and monetary) model along with self-organizing maps is used to segment dental patients of a children's dental clinic in Taiwan and also suggests that one cluster with both R and F values greater than the overall average R and F values can be viewed as loyal patients (Wei et al., 2011). RFM is less reliable than CHAID when the response rate is low and when the mailing is relatively to a small portion of the database. Alternatively, when the response rate is relatively high or the database marketer desires to mail to a relatively large portion of the file, RFM may provide results similar to CHAID and logistic regression (McCarty and Hastak, 2007).

Although the efficiency of RFM has been questioned, little research documents its ability relative to newer statistical techniques. This paucity of research is partly because RFM refers to a general approach to data mining; there are a variety of ways of applying the use of recency, frequency, and monetary value. Research that has been conducted on the efficacy of RFM generally focuses on proprietary or judgmental models of RFM (Magidson, 1988; Levin and Zavari, 2001) and not on empirically based RFM models. More recently, research has moved away from RFM and has focused instead on newer, more sophisticated approaches to data mining (Deichmann et al., 2002; Linder et al., 2004). The current study evaluates one popular, empirically based (as opposed to judgmental) approach to RFM. This RFM approach is compared to CHAID and logistic regression, in an effort to understand its capabilities as a database marketing analytical tool.

RFM Analysis

Recency, frequency, and monetary (RFM) analysis have been used in direct marketing for a number of decades (Baier et al., 2002). This analytical technique grew out of an informal recognition by catalog marketers that three variables seem particularly related to the likelihood those customers in their house data files would respond to specific offers. Customers who recently purchased from a marketer (recency), those who purchase many times from a marketer (frequency), and those who spend more money with a marketer (monetary value) typically represent the best prospects for new offerings.

As noted, RFM analysis is utilized in many ways by practitioners, therefore, RFM analysis can mean different things to different people. One common approach to RFM analysis is what is known as hard coding (Drozdenko and Drake, 2002). Hard coding RFM is a matter of assigning a weight to each of the variables recency, frequency, and monetary value, then creating a weighted score for each person in the database. The assignment of weights is generally a function of the judgment of the database marketers with a particular database; for example, past experience may tell a marketer that recency should weigh twice as much as frequency and monetary value. Therefore, this application of RFM is often referred to as judgment based RFM. The weightings could also vary as a function of the particular mailing (Baier et al., 2002). The weights can, of course, be empirically derived based on offerings mailed

to database members in the past, thus relying on previous data rather than judgments.

Regardless of the way that RFM is utilized; there are two common characteristics of RFM procedures. First, RFM is used to segment a house file (i.e., a company's current customers) using information related to recency, frequency, and monetary value. RFM is not applicable to the prospecting for new customers because a marketer would not have transaction information for prospects. Second, RFM analysis generally focuses on the three behavioral variables of recency, frequency, and monetary value. Although these variables are considered powerful predictors of future behavior, traditional RFM is limited to these three things.

A well known, empirically based RFM method is a procedure advocated by Arthur Hughes (2000). Hughes' approach is applicable in instances when a marketer intends to send a mailing to customers in its database and would like to find those in the database who are the most likely to respond to the specific mailing. Hughes recommends a test mailing to a sample of customers in the file; then the selection of the members of the rest of the file is made as a function of the results of the test. Thus, compared with hard coding RFM, Hughes' method is not arbitrary with respect to the weighting of recency, frequency, and monetary value. The importance of each of these is determined by the test mailing for the particular offer.

The first step in the method is for the marketer to sort the customer file according to how recently customers have purchased from the firm. The database is then divided into equal quintiles and these quintiles are assigned the numbers 5 to 1. Therefore, the 20% of the customers who most recently purchased from the company are assigned the number 5; the next 20% are assigned the number 4, and so on. The next step involves sorting the customers within each recency quintile by how frequently they purchase from the marketer. For each of these sorts, the customers are divided into equal quintiles and assigned a number of 5 to 1 for frequency. Each of these groups (25 groups) is sorted according to how much money the customers have spent with the company. These sorts are divided into quintiles and assigned numbers 5 to 1. Therefore, the database is divided into 125 roughly equal groups (cells) according to recency, frequency, and monetary value.

Hughes recommends conducting a test mailing to a randomly sampled subset of each cell (e.g., 10%). After the responses of the test mailing are received, the proportion of respondents in each cell can be calculated. The cells can then be ordered as a function of response percent. The marketer can then elect to mail to a certain portion of the remaining file (e.g., the top 20% of the cells). Alternatively, the marketer can elect to mail to the cells that are above a break-even percent, given the cost of the mailing and the expected revenue for each return. For example, if a mailing costs $1.50 and the revenue received is $50.00 per order, the break-even percentage would be 3%. Thus, for the 90% of the file that is left after the test mailing, the direct marketer would mail to the RFM cells that the test mailing predicted a 3% or better return.

It is important to note that Hughes' method does not assume a monotonic relationship between the dependent variable (responded/did not respond) with the variables of recency, frequency, and monetary value. Each cell is a discreet group that is considered individually in terms of its performance. Thus, if middle levels of one of the independent variables (e.g., frequency) are more related to response compared with higher or lower levels of this variable, then the procedure can accommodate the non-monotonic nature of the relationship.

CHAID

Chi Square Automatic Interaction Detector (CHAID) (Sargeant and McKenzie, 1999) is a method of database segmentation that has been used for a number of years. Research has shown that CHAID is superior to judgment based RFM with respect to the identification of likely responders (Magidson, 1988; Levin and Zavari, 2001). CHAID is similar to the RFM approach of Hughes because it creates groupings (nodes) of database members. The main difference is that these groupings are not created a priori as is the case with RFM. Rather, the file is split according to a statistical algorithm after a test mailing is conducted. After the returns of the test mailings are received, the procedure starts with a node that includes everyone in the test file. The

procedure then searches for the independent variable (e.g., number of times purchased) that best discriminates among the file members with respect to a dichotomous variable (i.e., purchased/did not purchase on current mailing). It splits the original node on this independent variable into as many subgroups as are significantly different with respect to the dichotomous variable. The procedure then splits these new nodes according to the variables that discriminate each of them. The procedure continues until no other splits are significant. CHAID analysis is often called tree analysis because a trunk (original node) is split into branches, then more branches, etc. The terminal nodes are those that cannot be split any further.

The analysis is similar to RFM because the terminal nodes can be evaluated according to which ones break even with respect to expected profit and mailing costs. The direct marketer can then use the rules that define the terminal nodes in the test mailing (i.e., levels of the independent variables that define each terminal node) to select the groups of people left in the file after the test that should receive the mailing. It is also similar to RFM in that CHAID can accommodate relationships between the dependent variable and the predictor variables that are non- monotonic. For example, if the number of times purchased relates to the dependent variable, CHAID may divide the file members into three nodes: those who purchase 1 to 3 times, those who purchase 4 to 8 times, and a third node of those who purchase 9 or more times. These three nodes represent discreet groupings.

An important difference between CHAID and RFM is that CHAID can accommodate a variety of independent variables. The independent variables could include recency, frequency, and monetary value, but could also include other transaction variables (e.g., used a credit card or not), as well as individual difference variables such as demographic and psychographic variables.

Logistic Regression

Logistic regression is a modeling procedure where a set of independent variables is used to model a dichotomous criterion variable.

Therefore, it is appropriate for direct marketers who would like to model the dichotomous variable of respond/don't respond to a mailing. Logistic regression is particularly useful in these circumstances in that the actual criterion variable is dichotomous; however, the predicted variable is the response probability, which varies from zero to one. Therefore, the model can provide a probability of response for everyone in the file, given the estimated parameters for a set of predictor variables.

After a test mailing similar to CHAID, logistic regression can be used to analyze the response variable as a function of several independent variables (e.g., number of times purchased) and provide an equation that can calculate the response probability for the entire house file. The marketer can then mail to everyone left in the file (excluding those in the test) that has a probability higher than the break even percent. Similar to CHAID, the independent variables are not restricted to recency, frequency, and monetary value.

Logistic regression differs from both RFM and CHAID in two important ways. First, logistic regression provides a response probability for individual members of the dataset rather than creating discreet groups of people. Therefore, in theory, each person in the dataset may have a different response probability. In practice, however, if few independent variables are used to construct the logistic function and each has a small number of different possible values, then there would be a relatively small number of different response probabilities across the people in the file. Second, for continuous predictor variables, logistic regression model relationships of the independent variables with the dichotomous dependent variable that are monotonic; both RFM and CHAID are distribution free. This has implications for the performance of logistic regression in instances where the relationship between a predictor variable and the response variable is neither continuously increasing nor decreasing. For example, when the relationship between recency of previous purchases and purchase on the test mailing is curvilinear, logistic regression may not be able to capture the relationship in ways similar to that for RFM or CHAID.

Factors Affecting the Response Rate in Email Campaigning

A study conducted by Lisa Chittenden and Ruth suggests that Permission based E -Mails positively affects the response rate. The response rate model suggests that there are three stages in effective e-mail marketing: getting recipient to open the email, holding their interest and persuading them to respond, hence response rate should depend upon email header, email contents and recipients. There is a significant correlation between response rate and subject line, e-mail length, incentives and number of images (Chittenden and Rettie, 2002) and also Permission marketing, personalization, brand equity influenced response rates of email marketing (Tezinde, et al., 2002). The permission email messages can be linked to retention of the customers (Jolley, et al., 2012). A study also indicates that there exists a positive correlation between the Internet marketing and business performance (Saeedi et al., 2012).

Incentive to open the mails or WIIFM or "What's In It For Me?" is a question at the forefront of every email recipient's mind when making a decision to open, read and take action on the email sent and thus influences the response rate of the Email campaigning (Keegan, 2012). Links (text links, hyperlinks, graphics or images) when clicked or when pasted into a browser, send the prospect to another online location (e.g., a landing page or other pages of a website). A link in an email is a call-to-action. To be most effective in motivating action, links should be visible, clear and compelling (Keegan, 2012).

A Preexisting Business Relationship that is the recipient of the email has made a purchase, requested information, responded to a questionnaire or a survey, or had offline contact with the firm running the campaign influences the response rate of the Email campaigning (Keegan, 2012).

RESEARCH METHOD
Research Design

The research design adapted in the study is *"Descriptive Research"*, as it aims at exploring the causes for low response rate in email campaigning and proposing a model to enhance the response rate in email campaigning.

Data Collection
Primary Data

The data, which was gathered to determine the major factors affecting the response rate in email campaigning, was primary data, as it was collected for the first time. Survey Method and more specifically E-questionnaire method was used to collect data from a sample of respondents, who were prospects of Aditi Technologies.

Secondary Data

The data, which was gathered to determine the most effective data mining segmentation method for Aditi Technologies, was secondary data, as it was not collected for the first time. The secondary data was from the tool called HubSpot (internal sources), which was used by Aditi Technologies to send out campaigns.

Sampling Method

A sample of 146 prospects was chosen based on non – probability sampling method, more specifically convenient sampling method from Aditi Technologies to determine the major factors affecting the response rate in email campaigning.

And a sample of 120 datasets was collected from the tool called HubSpot using a non – probability sampling method, more specifically Convenient sampling method to determine the most effective data mining segmentation approaches for Aditi Technologies.

Tools for Data Collection

A structured Questionnaire was used as a research instrument or a tool for collecting data and administered to the prospects of Aditi Technologies to determine the factors affecting the response rate in email campaigning.

In order to construct the questionnaire first a list of variables that was affecting the response rate in email campaigning is collected from empirical studies and thirty-one questions are written to assess these indices. The questions are all close ended and respondents are asked to answer these close ended questions on a 7-point Likert-type scale - strongly disagree, moderately disagree , slightly disagree, neutral, slightly agree, moderately agree, strongly agree.

The questionnaire used in this study is divided into 2 parts. In Part 1, the respondents were asked general demographic questions such

as gender, age, and designation. Part 2 of the questionnaire contains items measuring various dimensions of factors affecting the response rate in email campaigning, namely From Line, Subject Line, Personalization of the subject line, Timings for sending mails, Frequency of mailing, Length of the Emails, Incentives to respond, Pre-existing Business Relationship, Permission based emails, Links and Image.

The Questionnaire is then distributed to prospects of Aditi Technologies to know what factors led to their non-response for the mail sent by the company. Depending on the content of the question, answers were later converted to 7-point "favorability" scores based on the response indicated (1 = strongly disagree through to 7 = strongly agree).

Validity and Reliability
Validity

The validity of the instrument was established by taking email-marketing experts' opinion on the questions framed and asking a test sample if they could comprehend the questions in the questionnaire. For this mean, the

questionnaires were given to the experts in management and email marketing, and after their modifications were being used and they confirmed it, the Questionnaires were given to the participants.

Reliability Test

Reliability of the questionnaire is established using a pilot test by collecting data from 18 people chosen randomly from the samples. Data collected from pilot test is analyzed using SPSS (Statistical Package for Social Sciences).

In order to prove the internal reliability of the questionnaires 'Cronbach Alpha technique' was performed on the responses obtained from the sample of respondents (n=18). The 'Cronbach Alpha' values were calculated for all the variables and it was found that the reliability results were more than reasonable threshold (0.7) and hence reliability of questionnaire was confirmed. Table 1 shows the Cronbach Alpha values for all the dimensions identified and specifies that the items pertaining to each dimension are internally consistent and it is measuring the same dimension that it intends to.

Table 1: Reliability statistics

S. No	Variables	N of Items	Cronbach's Alpha
1	From Line	4	0.90
2	Subject Line	4	0.88
3	Personalization of the subject line	3	0.86
4	Timings for sending mails	3	0.80
5	Frequency of mailing	2	0.83
6	Length of the Emails	2	0.81
7	Incentives to respond	3	0.86
8	Pre-existing Business Relationship	1	NA
9	Permission based emails	2	0.81
10	Links	3	0.82
11	Image	3	0.80
Overall Cronbach's Alpha: 0.87			

Tools Used for Analysis

Factor analysis method was conducted for the responses using principal component method and varimax rotation for rotation of the axis. A factor analysis is a data reduction technique to summarize a number of original variables into a smaller set of composite dimensions, or factors. In this study it is employed to explore the underlying factors associated with 31 items.

Multiple linear regression method was employed to determine the major factors affecting the response rate in email campaigning, and the coefficient of determination score was used to find the correlation relationship among From Line, Subject line, Personalized subject line, Frequency of mailings, Timing of mailings, Length of the email, Incentive to respond, Permission based email, Images, Links, Preexisting Business Relationship and response rate of the email campaigning

Independent Sample T – Test is used to determine whether the male or female respondents have different responses to the 11 factors (obtained from factor analysis) which are From Line, Subject line, Personalized subject line, Frequency of mailings, Timing of mailings, Length of the email, Incentive to respond,

Permission based email, Images, Links, Preexisting Business Relationship.

Chi-Square Goodness-of Fit is used to determine whether the observed frequencies of gender, designation and age group of the respondents are significantly different from what it is expected to get by chance that is the gender, designation and age group of the respondents are not equally likely to be chosen.

One-way ANOVA is used to determine whether there exists difference in effectiveness of RFM, CHAID and Logistic Regression segmentation method on response rate of the customers.

RESULTS AND DISCUSSION
Results of Segmentation
RFM Segmentation

RFM segmentation is run using SPSS 19.0 and the prospects that are having high RFM scores are most likely to respond to the solicitation. Hence, the prospects are segmented using the RFM score generated by the SPSS tool and their response rate at various levels of depth (20% to 50%) are calculated and tabulated as in table 2.

Table 2: Percent of total responses for various levels of depth of total file

Data Mining Technique		
RFM (%)	CHAID (%)	Logistic (%)
20% depth of file		
Test Sample — 33.3	41.67	50
Hold Sample — 16.67	33.3	33.3
Difference — 16.63	8.37	16.67
30% depth of file		
Test Sample — 27.7	33.33	33.33
Hold Sample — 22.2	38.46	22.2
Difference — 5.5	5.13	11.13
40% depth of file		
Test Sample — 20.8	25	25
Hold Sample — 16.67	33.33	29.17
Difference — 4.13	8.33	4.17
50% depth of file		
Test Sample — 23.3	20	20
Hold Sample — 20	26.67	26.67
Difference — 3.3	6.67	6.67

CHAID Segmentation

CHAID is a technique of decision tree or regression tree, and is the best tool used to discover the relationship between variables. CHAID analysis determines how the variables best combine to explain the outcome in given dependent variables. CHAID segmentation is run using SPSS 19.0 and non-binary classification tree is generated as in figure 1. This is where more than two branches may go from the dependent node (Open_Response). In the CHAID technique, we can visually see the relationship between the dependent variable (Open_Response) and the associated related factor with a tree (Last Order Date). The Tree visually explains that Last Order Date is a variable that best discriminates the respondents and non-respondents. Hence, the prospects are segmented using the relevant variable Last Order Date and their response rate at various levels of depth (20% to 50%) are calculated and tabulated as in table 2.

Logistic Regression

The prospects were segmented using the equation derived by running the test file on the SPSS 19.0. And this equation was primarily used to derive the probability for the likely respondents. The logistic regression equation is:

logit (Response Rate) = -346.009 + 0.028 (Number of Transactions)
This equation is derived from table 3.

Open_Response

Node 0	
Mean	0.150
Std. Dev.	0.358
n	120
%	100.0
Predicted	0.150

Last Order Date
Adj. P-value=0.006, F=12.340, df1=1, df2=118

<=13-Jul-2010

Node 1	
Mean	0.057
Std. Dev.	0.234
n	70
%	58.3
Predicted	0.057

>13-Jul-2010

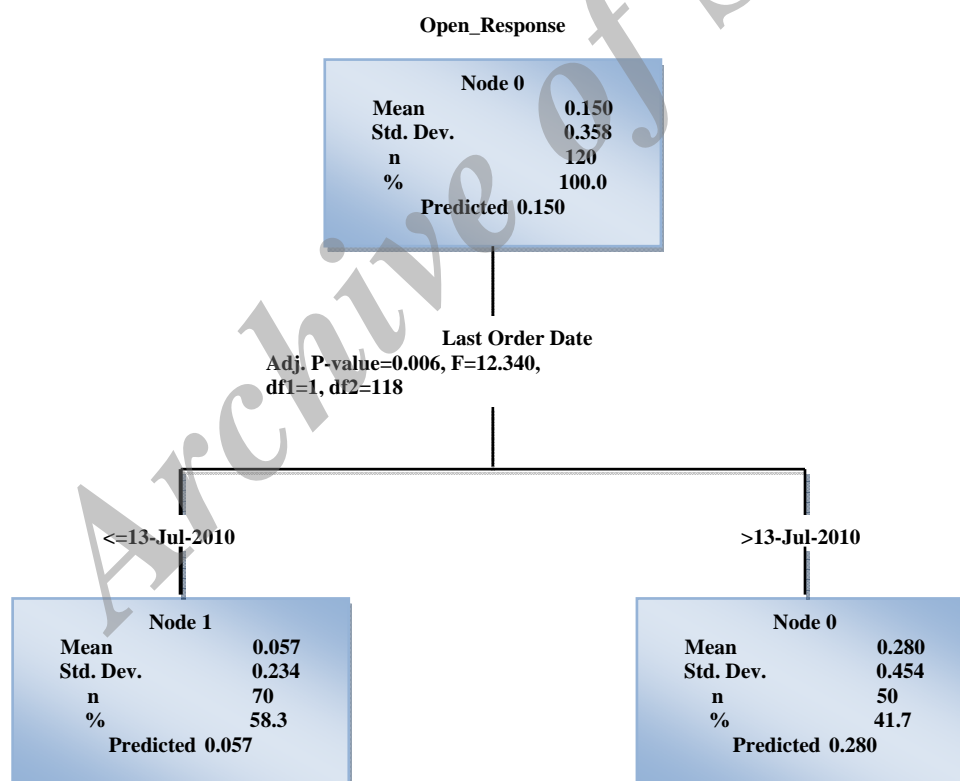| Node 0	
Mean	0.280
Std. Dev.	0.454
n	50
%	41.7
Predicted	0.280

Figure 1: CHAID segmentation tree structure

281

Table 3: Segmentation using logistic regression - variables in the equation

	B	S.E.	Wald	df	Sig.	Exp (B)
Last Order Date	0	0	8.539	1	0.003	1
Number of Transactions	0.028	0.023	1.416	1	0.234	1.028
Total Revenue	0	0	0.22	1	0.639	1
Constant	-346.01	117.804	8.627	1	0.003	0

Table 4: Test of homogeneity of variances

Levene Statistic	df1	df2	Sig.
0.619	2	21	0.548

Table 5: Summary table of ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	362.032	2	181.016	3.166	0.063
Within Groups	1200.690	21	57.176		
Total	1562.721	23			

Hypothesis Testing

One-way ANOVA was employed for testing the hypothesis of this research. The data was analyzed using One-way ANOVA and the summary is presented in table 4.

H_0: There is no significant difference in effectiveness of RFM, CHAID and Logistic Regression segmentation method on response rate of the email campaigning.

H_1: There is a significant difference in effectiveness of RFM, CHAID and Logistic Regression segmentation method on response rate of the email campaigning.

The Levene's test of variance is used to check the homogeneity of variance that is to check if the performance is uniformly distributed or not. Here, the value of "Levene Statistics ANOVA" is greater than .05. There the response rate is uniformly distributed.

In the ANOVA table 5, we can see that the value of the significance is greater than .05 as in the output. So, we accept H_0 that is there is no

significant difference created by the three segmentation methods on the response rate.

Measure of Sampling Adequacy

The Kaiser-Meyer-Olkin measure of sampling adequacy tests whether the partial correlations among variables are small. High values (close to 1.0) generally indicate that a factor analysis may be useful with data. Bartlett's test of sphericity tests the hypothesis that correlation matrix is an identity matrix, which would indicate that variables are unrelated. Small values (less than 0.05) of the significance level indicate that a factor analysis may be useful with data. Table 6 indicates that in the present test The Kaiser-Meyer-Olkin (KMO) measure was 0.6 which is greater than the threshold value of 0.5 and hence the sample chosen is adequate. The Bartlett's sphericity test also indicates that Chi-Square = 126.952, df = 55 with a significance of 0.000 and hence the variables are unrelated.

Table 6: KMO and Bartlett's test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.6
Bartlett's Test of Sphericity	Approx. Chi-Square	126.952
	Df	55
	Sig.	0.000

Factor Analysis
Extraction Method

Extraction communalities are estimates of the variance in each variable accounted for by the components. The communalities are ranging from .526 to .880, which indicates that the extracted components represent the variables well.

Principle Component Analysis

✓ *Factor* - The initial number of factors is the same as the number of variables used in the factor analysis. However, not all 31 factors will be retained. In this study, only the first 11 factors were retained.

✓ *Initial Eigenvalues* - Eigenvalues are the variances of the factors. Because factor analysis is conducted on the correlation matrix, the variables are standardized, which means that the each variable has a variance of 1, and the total variance is equal to the number of variables used in the analysis, in this case, 31.

✓ *Total* - This column contains the eigenvalues. The first factor will always account for the most variance (and hence have the highest eigenvalue), and the next factor will account for as much of the left over variance as it can, and so on. Hence, each successive factor will account for less and less variance.

✓ *% of Variance* - This column contains the percent of total variance accounted for by each factor.

✓ *Cumulative %* - This column contains the cumulative percentage of variance accounted for by the current and all preceding factors. In this study, the first 11 factors together account for 78.363% of the total variance.

✓ *Extraction Sums of Squared Loadings* - The number of rows in this panel of the table correspond to the number of factors retained. Here, 11 factors were retained, so there are 11 rows, one for each retained factor. The values in this panel of the table are calculated in the same way as the values in the left panel, except that here the values are based on the common variance. The values in this panel of the table will always be lower than the values in the left panel of the table, because they are based on the common variance, which is always smaller than the total variance.

Interpretation

Table 7 reveals that the total variance explained is 78.363% and the rotated component matrix shows 11 factors. All 11 factors are having high value loadings and it ranges from .721 to .904 as shown in the table 4.10, which indicates that the extracted components represent the variables well.

Factor-1 loading about 14.273%, Factor-2 loading 12.015%, Factor -3 loading 9.279%, Factor- 4 loading 7.670%, Factor- 5 loading 6.886%, Factor- 6 loading 6.489%, Factor- 7 loading 5.860%, Factor- 8 loading 4.857%, Factor- 9 loading 4.913%, Factor- 10 loading 3.668%, Factor- 11 loading 3.444%. All eleven factors explain nearly 78.363% of the variability; it means only a 21.637% loss of information. According to Kenova and Jonasson (2006) and Garson, (2002) 60% is arbitrary level for good factor loadings in likert scale cases and hence the extracted factors show good factor loadings.

Table 7: Principal component analysis - total variance explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.425	14.273	14.273	4.425	14.273	14.273
2	3.724	12.015	26.288	3.724	12.015	26.288
3	2.876	9.279	35.567	2.876	9.279	35.567
4	2.378	7.670	43.236	2.378	7.670	43.236
5	2.135	6.886	50.123	2.135	6.886	50.123
6	2.012	6.489	56.612	2.012	6.489	56.612
7	1.817	5.860	62.472	1.817	5.860	62.472
8	1.422	4.587	67.058	1.422	4.587	67.058
9	1.300	4.193	71.251	1.300	4.193	71.251
10	1.137	3.668	74.919	1.137	3.668	74.919
11	1.068	3.444	78.363	1.068	3.444	78.363
12	0.868	2.798	81.161			
13	0.677	2.183	83.345			
14	0.643	2.076	85.420			
15	0.568	1.832	87.252			
16	0.453	1.461	88.713			
17	0.401	1.294	90.008			
18	0.399	1.288	91.295			
19	0.333	1.073	92.368			
20	0.314	1.014	93.382			
21	0.282	0.908	94.290			
22	0.262	0.844	95.134			
23	0.254	0.819	95.953			
24	0.235	0.759	96.712			
25	0.210	0.678	97.390			
26	0.197	0.636	98.026			
27	0.181	0.582	98.608			
28	0.164	0.531	99.139			
29	0.155	0.501	99.639			
30	0.112	0.361	100.000			
31	1.801E-17	5.808E-17	100.000			
Extraction Method: Principal Component Analysis.						

Rotated Component Matrix

Rotated Factor contains the rotated factor loadings (factor pattern matrix), which represent both how the variables are weighted for each factor and also the correlation between the variables and the factor. Because these are correlations, possible values range from -1 to +1. In the factor analysis option pane, the option suppress small coefficient less than (.30) is resorted to and hence any of the correlations that are .3 or less will not be shown. And they were sorted by size as displayed in table 8. This makes the output easier to read by removing the clutter of low correlations that are probably not meaningful.

Table 8: Rotated component matrix

	Component										
	1	2	3	4	5	6	7	8	9	10	11
QuF1	0.854										
QuF2	0.892										
QuF3	0.866										
QuF4	0.837										
QuS5		0.893									
QuS6		0.841									
QuS7		0.893									
QuS8									-0.316		-0.451
QuS9		0.721									
QuP10				0.904							
QuP11				0.897							
QuP12				0.796							
QuT13						0.697					
QuT14						0.871					
QuT15						0.897					
QuF16								0.880			
QuF17								0.886			
QuL18									0.864		
QuL19									0.866		
QuI20		0.884									
QuI21		0.858									
QuI22		0.821									
QuR23											0.806
QuPB24										0.855	
QuPB25										0.873	
QuLi26					0.859						
QuLi27					0.880						
QuLi28					0.813						
QuIm29							0.875				
QuIm30							0.786				
QuIm31							0.838				
Variance	10.412	9.694	7.753	7.719	7.415	7.216	7.154	5.810	5.750	5.592	3.848
Cumulative Variance	10.412	20.106	27.859	35.578	42.992	50.208	57.362	63.172	68.922	74.515	78.363

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Interpretation

The items were collectively renamed for component 1 to component 11 as From Line, Subject line, Incentive to respond, Personalized subject line, Links, Timing of mailings, Images, Frequency of mailings, Permission based email, Preexisting Business Relationship respectively

Scree Plot

The scree plot graphs the eigenvalue against the factor number. These values can be found in the first two columns of the Extraction Table.

Interpretation

In the Scree plot it can be noted that from the eleventh factor the line is almost flat, meaning the each successive factor is accounting for smaller and smaller amounts of the total variance.

Multiple Linear Regression
Multiple Linear Regression before Elimination

From table 9, it is evident that Links and Image variables are not significant as the sig>0.05.

Testing of Hypothesis

H_0: $\beta=0$. From Line, Subject line, Personalized subject line, Frequency of mailings, Timing of mailings, Length of the email, Incentive to respond, Permission based email, Images, Links, Preexisting Business Relationship are not good predictors of email response in email campaigning.

H_1: $\beta\neq0$. From Line, Subject line, Personalized subject line, Frequency of mailings, Timing of mailings, Length of the email, Incentive to respond, Permission based email, Images, Links, Preexisting Business Relationship are good predictors of email response in email campaigning.
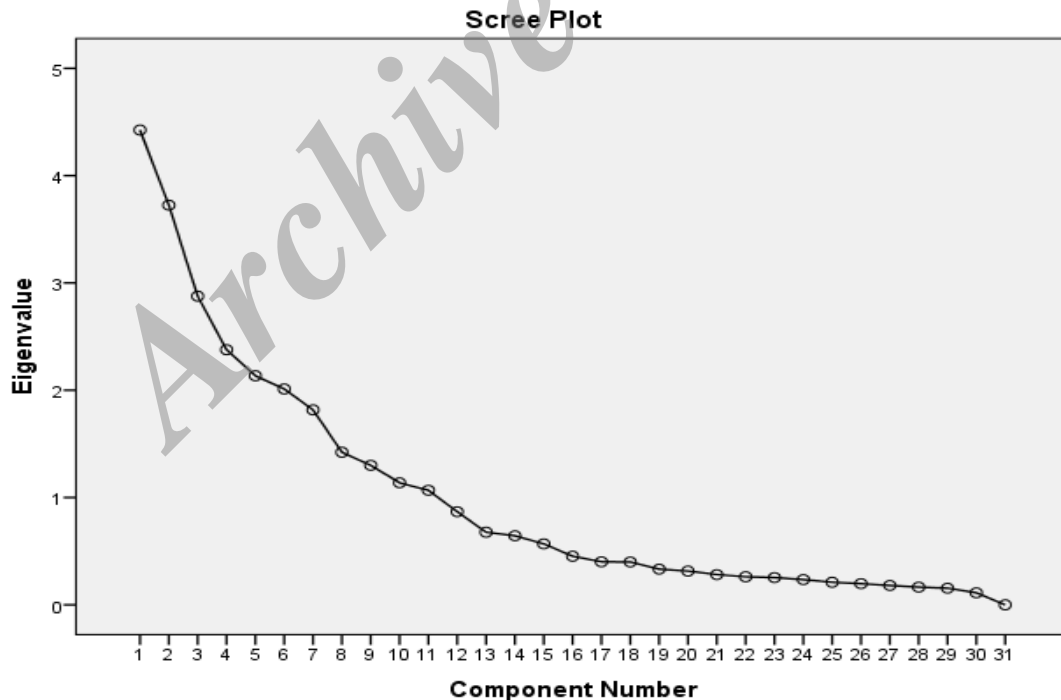

Figure 2: Scree Plot

Table 9: Multiple linear regression before elimination

Models	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Null Hypothesis
	B	**Std. Error**	**Beta**			
(Constant)	-0.649	1.150		-0.564	0.573	Reject
From_Line	0.040	0.068	0.021	0.583	0.561	Reject
Sub_Line	0.000	0.118	0.000	0.001	0.999	Reject
Personalization	0.038	0.075	0.018	0.508	0.612	Reject
Timings	-0.072	0.042	-0.060	-1.735	0.085	Reject
Frequency_of_Mailing	0.044	0.057	0.029	0.773	0.441	Reject
Length_of_Email	1.019	0.041	0.922	24.982	0.000	Reject
Incentives_to_Respond	0.262	0.173	0.097	1.509	0.134	Reject
Preexisting_Business	-0.021	0.080	-0.009	-0.267	0.790	Reject
Permission_based_Emails	0.035	0.036	037	0.986	0.326	Reject
Links	-0.035	0.055	-0.022	-0.630	0.530	Reject
Image	0.188	0.124	-0.093	-1.510	0.133	Reject
a. Dependent Variable: Response in email campaigning						

Table 10 indicates that 9 independent variables, *Permission based emails, Length of Email, Timings, From Line, Frequency of mailing, Preexisting Business, Personalization, Incentives to respond, Subject Line* explains 92.2% of variance of the Response rate (dependent variable). The summary table indicates that *Permission based emails, Length of Email, Timings, From Line, Frequency of mailing, Preexisting Business, Personalization, Incentives to respond, Subject Line* were good predictors of response rate in email campaigning because R Square value 0.922.

From table 11 the following regression equation can be derived:

Response rate in email campaigning= -.649+ (0.040) From Line + (0.00) Subject Line + (0.038) Personalization - (0.072) Timings + (0.044) Frequency of mailing + (1.019) Length of Email+ (0.262) Incentives to respond - (0.021) Preexisting Business + (0.035) Permission based emails.

From the Table 11 the F value can be calculated as:

$F (11,134) = 65.055, p < .000$

Testing the Hypothesis Using Independent Sample T- Test

H_0: $\mu_{male} = \mu_{female}$ There is no significant differences in responses given by males and females of the population on the 11 factors namely, From Line, Subject line, Personalized subject line, Frequency of mailings, Timing of mailings, Length of the email, Incentive to respond, Permission based email, Images, Links, Preexisting Business Relationship.

H_1: $\mu_{male} \neq \mu_{female}$ There is no significant differences in responses given by males and females of the population on the 11 factors namely, From Line, Subject line, Personalized subject line, Frequency of mailings, Timing of mailings, Length of the email, Incentive to respond, Permission based email, Images, Links, Preexisting Business Relationship

An independent-sample t-test was conducted to examine whether there was a significant differences in responses given by males and females of the population on the 11 factors namely, From Line, Subject line, Personalized subject line, Frequency of mailings, Timing of mailings, Length of the email, Incentive to

287

respond, Permission based email, Images, Links, Preexisting Business Relationship.

If Levene's test for equality of variances is significant, the statistics for the row equal variances not assumed would be reported. In this study only for the sub scale From Line Levene's

test for equality of variances was significant and the row equal variances not assumed was considered. For all other subscales the Levene's test for equality of variances was not significant and hence the row equal variances assumed were considered as in table 12.

Table 10: Multiple linear regression – coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Null Hypothesis
		B	Std. Error	Beta			
	(Constant)	-0.649	1.150		-0.564	0.573	Reject
	From_Line	0.040	0.068	0.021	0.583	0.561	Reject
	Sub_Line	0.000	0.118	0.000	0.001	0.999	Reject
	Personalization	0.038	0.075	0.018	0.508	0.612	Reject
1	Timings	-0.072	0.042	-0.060	-1.735	0.085	Reject
	Frequency_of_Mailing	0.044	0.057	0.029	0.773	0.441	Reject
	Length_of_Email	1.019	0.041	0.922	24.982	0.000	Reject
	Incentives_to_Respond	0.262	0.173	0.097	1.509	0.134	Reject
	Preexisting_Business	-0.021	0.080	-0.009	-0.267	0.790	Reject
	Permission_based_Emails	0.035	0.036	037	0.986	0.326	Reject

a. Dependent Variable: Response rate in email campaigning

Table 11: Model summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.922[a]	0.850	0.838	0.583

a. Predictors: (Constant), Permission_based_emails, Length_of_Email, Timings, From_Line, Frequency_of_mailing, Preexisting_Business, Personalization, Incentives_to_respond, Sub_Line.

	Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Null Hypothesis
		B	Std. Error	Beta			
	(Constant)	-0.649	1.150		0.662	0.508	Reject
	From_Line	0.040	0.068	0.021	2.93	0.004	Reject
	Sub_Line	0.000	0.118	0.000	12.67	0	Reject
	Personalization	0.038	0.075	0.018	3.969	0	Reject
1	Timings	-0.072	0.042	0.053	-0.060	0	Reject
	Frequency_of_mailing	0.044	0.057	0.05	0.029	0.002	Reject
	Length_of_Email	1.019	0.041	0.023	0.922	0.004	Reject
	Incentives_to_respond	0.262	0.173	0.141	0.097	0.007	Reject
	Preexisting_Business	-0.021	0.080	0.212	-0.009	0	Reject
	Permission_based_emails	0.035	0.036	0.02	0.037	0.016	Reject

a. Dependent Variable: Response rate in email campaigning

Table 12: Independent sample T-Test

		Levenes test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Internal of the Difference	
		F	Sig.	t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
From_Line	Equal variance assumed	4.947	0.028	-0.158	144	0.875	-0.02356	0.14903	-0.31813	0.27102
	Equal variance not assumed			**-0.191**	**93.150**	**0.8449**	**-0.02356**	**0.12319**	**-0.26818**	**0.22107**
Sub_Line	Equal variance assumed	0.059	0.808	**-1.070**	**144.0**	**0.286**	**-0.09526**	**0.089**	**-0.27118**	**0.08088**
	Equal variance not assumed			-1.116	67.106	0.286	-0.09526	0.08538	-0.28587	0.07514
Personalization	Equal variance assumed	0.062	0.82	**-0.501**	**144.0**	**0.617**	**-0.06646**	**0.13276**	**-0.32887**	**0.19594**
	Equal variance not assumed			-0.512	64.817	0.61	-0.06646	0.12973	-0.32557	0.19285
Timings	Equal variance assumed	0.41	0.523	**0.582**	**144.0**	**0.582**	**0.13278**	**0.22819**	**-0.31826**	**0.58382**
	Equal variance not assumed			0.557	57.797	0.58	0.13278	0.23851	-0.34468	0.61024
Fequency_of_ Mailing	Equal variance assumed	0.519	0.473	**1.086**	**144.0**	**0.279**	**0.19688**	**0.18127**	**-0.16143**	**0.55518**
	Equal variance not assumed			1.058	59.533	0.294	0.19688	0.18602	-0.17529	0.58904
Length_of_ Email	Equal variance assumed	0.287	0.593	**1.678**	**144.0**	**0.095**	**0.41594**	**0.24782**	**-0.07388**	**0.90577**
	Equal variance not assumed			1.621	58.654	0.11	0.41594	0.25884	-0.09785	0.92954
Incentives_to_ Respond	Equal variance assumed	0.45	0.503	**-0.671**	**144.0**	**0.503**	**-0.08846**	**0.10198**	**-0.27002**	**0.13311**
	Equal variance not assumed			-0.624	55.340	0.535	-0.08846	0.10966	-0.28818	0.15127
Preexisting_ Business	Equal variance assumed	0	0.994	**0.078**	**144.0**	**0.939**	**0.00893**	**0.11704**	**-0.22242**	**0.24027**
	Equal variance not assumed			0.077	63.228	0.939	0.00893	0.11596	-0.22278	0.24083

Permission_based_Emails	Equal variance assumed	0.361	0.549	**2.013**	**144.0**	**0.046**	**0.5755**	**0.28592**	**0.01038**	**1.141085**
	Equal variance not assumed			1.958	59.341	0.055	0.5755	0.29399	-0.0127	1.1837
Links	Equal variance assumed	1.343	0.249	**-0.132**	**144.0**	**0.895**	**-0.02379**	**0.17989**	**-0.37895**	**0.33138**
	Equal variance not assumed			-0.128	57.441	0.9	-0.02379	0.18855	-0.40129	0.35371
Image	Equal variance assumed	3.696	0.057	0.500	144.0	0.618	0.06874	0.13739	-0.20283	0.3403
	Equal variance not assumed			**0.419**	**48.271**	**0.677**	**0.06874**	**0.16388**	**-0.26072**	**0.39819**

Table 13: Group statistics

Gender		N	Mean	Std. Deviation	Std. Error Mean
From_Line	Male	109	5.9427	0.84343	0.08079
	Female	37	5.9662	0.56569	0.09300
Sub_Line	Male	109	6.4128	0.47709	0.04570
	Female	37	6.5081	0.43867	0.07212
Personalization	Male	109	6.1314	0.70562	0.06759
	Female	37	6.1978	0.67359	0.11074
Timings	Male	109	5.3393	1.17109	0.11217
	Female	37	5.2065	1.28034	0.21049
Frequency_of_mailing	Male	109	5.9266	0.93992	0.09003
	Female	37	5.7297	0.99019	0.16279
Length_of_Email	Male	109	3.8349	1.27848	0.12246
	Female	37	3.4189	1.37191	0.22554
Incentives_to_Respond	Male	109	6.2107	0.51444	0.04927
	Female	37	6.2792	0.59589	0.09796
Preexisting_Business	Male	109	6.4954	0.61800	0.05919
	Female	37	6.4865	0.60652	0.09971
Permission_based_Emails	Male	109	4.4404	1.48094	0.14185
	Female	37	3.8649	1.56635	0.25751
Links	Male	109	5.0397	0.92005	0.08813
	Female	37	5.0635	1.01391	0.16669
Image	Male	109	6.1498	0.64067	0.06137
	Female	37	6.0811	0.92432	0.15196

Interpretation

The test revealed a statistically no significant difference between males (M=5.9, SD=.84) and females (M=5.9, SD=.57) for Permission based emails (t=1.91, df=144, p<0.001). The test also revealed a statistically no significant difference between males (M=6.49, SD=.62) and females (M=6.49, SD=.61) for Preexisting Business (t=.076, df=144, p<0.001) (table 13).

In this study the p value (Sig. 2-tailed) for all 11 factors is greater than or equal to 0.05. This implies that there is not sufficient evidence to conclude that male or female respondents have different responses to the 11 factors which are From Line, Subject line, Personalized subject line, Frequency of mailings, Timing of mailings, Length of the email, Incentive to respond, Permission based email, Images, Links, Preexisting Business Relationship.

Chi-Square Goodness-of-Fit Test for Demographic Profile of the Respondents

Table 14 provides the observed frequencies (Observed N) for each of the demographic profile of the respondents, namely, Gender, Designation and Age group, as well as the expected frequencies (Expected N), which are the frequencies expected if the null hypothesis is true. The difference between the observed and expected frequencies is provided in the Residual column.

Table 15 provides the actual result of the chi-square goodness-of-fit test. From this table it can be noted that the test statistic is statistically significant: $\chi2(2) = 35.507$, $p < .0005$. Therefore, we can reject the null hypothesis and conclude that our observed frequencies of Gender, Designation and Age Group of the respondents are significantly different from what it is expected to get by chance that is the Gender, Designation and Age Group of the respondents are not equally likely to be chosen.

Table 14: Chi-square goodness-of-fit test for demographic profile of the respondents

		Observed N	Expected N	Residual
Gender	Female	37	73	-36
	Male	109	73	36
Designation	CIO	78	20.9	57.1
	CTO	26	20.9	5.1
	VP of IT	6	20.9	-14.9
	Director of IT	7	20.9	-13.9
	VP of Engineering	12	20.9	-8.9
	Director of software development	8	20.9	-12.9
	Director of product management	9	20.9	-11.9
Age Group	21 and under	81	48.7	32.3
	22 to 34	60	48.7	11.3
	35 to 44	5	48.7	-43.7

291

Table 15: Test statistics for demographic profile of the respondents

Demographic Profile		Frequency
Gender	Chi-Square	35.507[a]
	Df	1
	Asymp. Sig.	0
Designation	Chi-Square	196.041[a]
	Df	6
	Asymp. Sig.	0
Age group	Chi-Square	63.301[a]
	Df	2
	Asymp. Sig.	0

CONCLUSION

The current study attempted to examine a contribution of various factors in improving the response rate in email campaigning. However, a result of principle component analysis indicates that, Subject Line, From Line, Incentives to respond and Personalization of the Subject Line are important factors in enhancing the response rate in email campaigning as it explains 43.24 per cent of variance. The Percentage Analysis also confirms the same. Therefore, the email marketer should think over these factors and make possible changes in the email campaigning. This will be help Aditi Technologies to enhance the response rate and would help them in getting leads.

The study also attempted to examine the most efficient segmentation approaches for Aditi Technologies. The result of One-way ANOVA indicates that there are no significant differences in the three segmentation approaches (RFM, CHAID and Logistic Regression) on the response rate in email campaigning. However, CHAID and logistic regression are not constrained with respect to the variables of recency, frequency, and monetary value as in RFM method. Response to a mailing can be modeled with a variety of variables using CHAID and Logistic Regression. One would assume that more precise modeling could be achieved using other variables. A consideration of relational data such as information about the motivations, attitudes, values, and lifestyles is taking more of a marketing approach to customers. Although these variables may be less useful than transaction information in their ability to predict a response to an immediate marketing activity (i.e., a mailing), they may be enormously useful in understanding the underlying tendencies in customers. This consideration would favor analytical techniques such as CHAID and logistic regression that can accommodate a variety of personality and individual difference information.

REFERENCES

Asllani, A. and Halstead, D. (2011). Using RFM Data to Optimize Direct Marketing Campaigns: A Linear Programming Approach. *Academy of Marketing Studies Journal*, 15 (1), pp. 59-75.

Chan, C. C. H. (2008). Intelligent Value-Based Customer Segmentation Method for Campaign Management: A Case Study of Automobile Retailer. *Expert Systems with Applications*, 34 (4), pp. 2754–2762.

Chittenden, L. and Rettie, R. (2002). An Evaluation of Email Marketing and Factors Affecting Response. *Journal of Targeting, measurement and Analysis for Marketing*, 11 (3), pp. 203–217.

Chittenden, L. and Rettie, R. (2003). An Evaluation of E-Mail Marketing and Factors Affecting Response. *Journal of Targeting, Measurement and Analysis for Marketing*, 11 (3), p. 203-217.

Chiu, Ch.-Y., Lin, Z.-P., Chen, P.-Ch. and Kuo, I.-T. (2009). Applying RFM Model to Evaluate the

E-Loyalty for Information-Based Website. *International Journal of Electronic Business Management*, 7 (4), pp. 278-285.

Coussement, K., Van den Bossche, F. A. M. and De Bock, K. W. (2012). Data Accuracy's Impact on Segmentation Performance: Benchmarking RFM Analysis, Logistic Regression, and Decision Trees. *Journal of Business Research*, 67 (1), pp. 2751-2758.

Dibb, S. and Simkin, L. (1997). A program for Implementing Market Segmentation. *Journal of Business and Industrial Marketing*, 12 (1), pp. 51 -65.

Fader, P., Hardie, B. G. S. and Lee, K. (2005). RFM and CLV: Using Iso-Value Curves for Customer Base Analysis. Journal of Marketing Research, 42 (4), pp. 415-430.

Golmah, V. and Mirhashemi, G. (2012). Implementing a Data Mining Solution to Customer Segmentation for Decayable Products – A Case Study for a Textile Firm. *International Journal of Database Theory and Application*, 5 (3), p. 73.

McCarty, J. A. and Hastak, M. (2007). Segmentation Approaches in Data Mining: A Comparison of RFM, CHAID, and Logistic Regression. *Journal of Business Research*, 60 (6), pp. 656–662.

Miglautsch, J. R. (2000). Thoughts on RFM Scoring. *Journal of Database Marketing*, 8, pp. 67-72.

Rajagopal, S. (2011). Customer Data Clustering Using Data Mining Technique. *International Journal of Database Management Systems*, 3 (4), p. 11.

Saeedi, N., Askari Masouleh, S., Abdolah, S. K., Mousavian, S. I. and Zendehbad, S. (2012). Impact of Internet Marketing on Business Performance. *American Journal of Scientific Research*, 71, pp. 39-47.

Wei, J.-T., Lin, Sh.-Y. and Wu, H.-H. (2010). A Review of the Application of RFM Model. *African Journal of Business Management*, 4 (19), pp. 4199-4206.