13.

Customer Stratification as a Means For Targeted Audits

M. Maphaka, W. Terblanche, D. John and J. Pillai

ENERGY TRADING ESKOM and EON Consulting,
South Africa
E-mail:johnd@eskom.co.za, pillayja@eskom.co.za

Abstract — **Customer Stratification looks at the proactive identification of potential energy loss customers. The adopted approach involves segmenting the customer base by making use of customer characteristics that drive consumption. Specific customer segment consumptions are then distributed to identify possible energy loss customers. By targeting loss customers in this manner it is possible to avoid the often ineffective and expensive blanket audits of customer areas.**

Keywords — **Customer stratification, Customer Care and Billing (CC&B), non-technical losses, audits.**

1. Introduction

Electrical utilities around the world are constantly faced with the pressures of reducing operational costs and improving the quality and reliability of supply. Energy losses constitute a considerable burden in terms of operating costs to the utility and hence have serious implications towards future network planning and design strategies. These losses which translate into commercial losses for the utility come from a variety of sources, all of which have in common that energy was delivered but not accounted for. Losses incurred in electrical power systems are generally made up of two components: technical losses and non-technical losses. Technical losses in power delivery systems are caused by the physical properties of the components of the power system which can be computed and brought under control. Non-technical losses are caused by actions external to the power system, or are caused by loads and conditions that the technical loss computation failed to take into account. In general, non-technical losses come from a variety of sources, all of which have in common that energy was delivered but not accounted for as sales. Adequate measuring, metering and the monitoring of various flows of energy within the various networks are essential for reducing energy losses. This should be complimented by proper energy audits, strategies that are put in place to effectively determine the losses in different customer segments and target those loss producing customers.

Technical losses can be further divided into copper losses (I^2R) and iron losses due to the core magnetization in transformers. In general, the copper loss component accounts for the majority of technical losses incurred on the network. The technical losses are generally calculated based on load flow simulations using appropriately modeled distribution network and load models. The effective flow of energy through the network must also be available through adequate installation of stats meters across the network which can enable the recording of data going in and out of the network.

Non-technical losses on the other hand can be defined as the component of energy loss that is not related to the physical characteristics and functions of the electrical system. Non-technical losses are almost impossible to measure using traditional power system analysis tools and can only be determined as the residual loss after subtracting technical loss from total losses. The causes for these types of losses are numerous and varied in nature. Studies have shown that of the numerous forms or causes for non-technical losses, the dominant component of non-technical losses are electricity theft and non-payment [1]. This can be broadly defined as a conscious attempt by a person to reduce or eliminate the amount of money owed to the utility for electric energy consumed. It has been identified that the theft of utility services is an international problem and has become an area of focus for the majority of utilities as experience shows that it is a difficult problem to deal with or attempt to isolate from the total losses incurred [2]. Efforts to combat non-technical losses have to be put in place and then be effectively enforced in order for the utilities to try and reduce overall losses.

Eskom Distribution currently has a methodology in place for the calculation of technical loss on the sub-transmission side and use a 10% assumption on the reticulation networks across the regions. This is due to the fact that there is limited statistical metering coverage at the MV and LV levels of the network. The method currently used within Distribution is carrying out blanket audit procedures over large areas where non-technical loss type issues have been suspected. This reduces the effectiveness of audits in that there is a lower hit rate in audits carried out and is hence not an efficient means of resource allocation (audit personnel & audit expenditure).

Amongst numerous methods proposed to reduce non-technical losses, the Customer Stratification methodology stands out with significant prominence over other methods, whereby using this methodology would ensure targeting of issues related to non-technical losses to be executed on an individual customer basis. This novel method within the Eskom Distribution environment could provide huge strides in terms of combating non-technical losses both from an audit effectiveness as well as cost perspective. The research aims to prove or disprove the hypothesis that possible loss customers within Distribution can be identified through the stratification approach. The hypotheses states that customers can be segmented such that outliers in their consumption distribution can be targeted as causing energy losses. This methodology if proven successful will provide a huge advantage over the current blanket audit procedure followed in Eskom Distribution, as it targets individual premises and drastically reduces audit costs required to do blanket audits. The data considered for the research is ESKOM Distribution’s Large Power User (LPU) and Small Power User (SPU) consumption for the period September 2003 – July 2006.

In this paper, an explanation of the Customer Stratification approach used to target possible loss customers within various segment groups in the customer base is discussed in detail. Comparison of results performed through actual field audits and those obtained through stratification are provided. Section 2 briefly reviews the segmentation criteria available for various systems. In section 3, we present the Customer Segmentation approach and highlight the various steps involved in the design of the methodology and segmentation criterion used. In section 4, the analysis performed to determine the diagnostic variables to be used for the segmentation and the target percentiles to be used are explained. Section 5, compares the results obtained from the sample of field audits and the stratification approach for the same sample. Finally, some concluding remarks are provided in section 6.

2. Segmentation

It is a well known fact that the entire customer base cannot be identified as identical in terms of importance or value to the business. It is usual that a certain small percentage of the consumer base constitutes almost ninety percent of the revenue recovered. Any approach that provides greater resolution about the types of customers served is a good start. Segmentation of customer base is one such method in this regard. Segmentation can also be used to better understand and respond to the common needs of the customers represented within each segment. Best-in-class segmentation systems evaluate customers across multiple categories, including financial metrics, growth, innovation and fit with the business and its strategy. The most innovative customer segmentation processes link segmentation results directly to the strategic plan for the business. In doing so, businesses can ensure they place investment bets on their most innovative customers, ultimately setting themselves up for success years into the future.

A segmentation system can be defined as a system that is created by performing a process of clustering, also known as cluster analysis, where similar objects are grouped into homogenous clusters (segments) [3]. Care should be taken so as to ensure that the clusters are as different from each other as possible. Segmentation systems provide synthetic information about objects by focusing on differences in demography, lifestyles, consumer behavior, etc. Some typical examples of segmentation systems are a Geodemographic segmentation system where systems are usually based on such variables as age, housing, family status, income, education, ethnicity, mobility, labor force, religion and other census variables. Such systems assume rather homogenous characteristics for all customers within a given geographical neighborhood. The second type of systems are Lifestyles segmentation systems, which include additional data derived from various surveys on preferences and lifestyle behavior, journals/magazines, TV, leisure, food and drinks, credit cards, etc. The third being Customer segmentation systems which are systems that refer more to individual customers than to geographic neighborhoods and are usually based on proprietary survey data about customers.

Segmentation in the Eskom Distribution environment will mainly fall into the third category. The segmentation of Eskom Distribution customers are based on data like customer type, voltage levels, maximum demand levels etc. From an Eskom Distribution perspective the segmentation methods and processes become even more complex than in most other industries due to the fact that Distribution are not in a position to select their customer base based on the best segmentations. Segmentation in the Eskom Distribution environment and from a Losses Reduction perspective must function as a management tool to effectively identify and address possible problem areas amongst their customer base. The fundamentals of segmentation in this environment is more a method of identifying problematic scenarios amongst the customer base than filtering for the best group of customers as in the sales and marketing environment.

Irrespective of the type of segmentation system approach used, the systems can be created using a similar procedure or methodology. It involves the selection of numerous input variables from various possible data sources which then eventually gets narrowed down into a final set of variables that are determined to be sufficient enough to define the segmentation system. These final sets of variables are known as ‘diagnostic’ variables. A general approach to this kind of segmentation can be summarized into the following basic steps as shown below [4],

1. Identification and selection of input variables
2. Reduction of the total number of variables
3. Deducing the total number of segments available
4. Agglomerating objects (customers, markets, geographic neighborhood) into segments
5. Characterization of segments and determination of key objects that are most typical for each segment
6. Mapping segments and linking them to other data sets.

As stated above the final sets of variables are known as 'diagnostic' variables. These variables should be relevant to the given type of business, they should be available at the desired geographic level of aggregation and additionally, they should not be interrelated. Diagnostic variables should also be geographically differentiated and relatively stable in time.

There are thousands of initial input variables used in commercially available segmentation systems. However, the final list of diagnostic variables should be much smaller to ensure the optimization between its application value versus complexity. An issue that arises with the use of too many potential input variables can lead to what is known in statistics as "overfeeding the model," and this situation should be avoided. If two variables are highly (significantly) correlated, their explanatory value is very similar and using both of them induces or introduces redundancy. Multiple linear regression analysis can help to select the set of variables (predictors) that are the most suitable for predicting another single variable. These tasks to be performed sometimes require mathematically intensive and enormous amounts of data to be processed, and hence need the use of special statistical software in order to assist in performing some of the steps mentioned above. Some examples of software packages available in the market that are in use are SPSS, SAS and S-PLUS or Clementine.

Eventually segmenting Eskom Distribution customers into groups according to their needs has a number of advantages such as,

- identify Eskom Distribution's most and least profitable customers
- focus the Eskom Distribution effort on the customers where they can achieve the biggest saving in the shortest possible time frame and hence avoid wasting time on customers or segments where the reduction in losses is minimal or not an issue at all
- use Eskom Distribution resources wisely and effectively
- improve service to meet customer needs
- increase profit potential by keeping losses down
- group Eskom Distribution customers by factors such as geographical location, size, type, Voltage level and maybe later attitudes and behaviour

3. Customer Stratification Approach

Amongst the various possible approaches that might be considered as feasible in accurately identifying possible loss customers, the approach of employing a segmentation criteria stands out with significant promise in tackling the issue at hand. It would require understanding whether the use of a detailed cluster based segmentation which involves segmenting of customers based on certain criteria over a specified sample of time, could actually shed light or give some indication of a possible loss customer. It is possible then to formulate a hypothesis which would either prove or disprove the hypothesis that “energy loss customers can be proactively identified by segmenting customers into customer classes of similar consumption behaviour”

Data Sourced

Eskom Distribution currently stores all the information pertaining to individual customers and their monthly consumption figures to relatively fair amount of accuracy in a database system called Customer Care and Billing (CC&B) information database. The system contains millions of records of information regarding the customer’s billing information, consumption figures, geographic regional locations, specific premise locations, account details, service agreement details, customer classification details, and a variety of other information associated with each and every customer. This information is stored in a number of different tables in CC&B depending on the various functions being performed on the data on a regular basis.

The data stored in the CC&B database consist of a number of different tables containing information regarding the customer base which contain specific customer characteristic details which in turn can be used to define customer segmentation criteria. Various sets of information regarding the characteristic details of the customers were sourced from the different tables and later combined into a single table, which was then be manipulated effectively. Consumption details and characteristic information such as customer types, voltage ratings, geographic locations, service agreement details, NMD details, urban/rural identifications, termination status flags, SIC codes and descriptions, LPU/SPU/Pre-paid indications and service point details are some the vital information required to make decisions in classifying the customers into different segments of similar consumption patterns.

Adopted Approach

The approach adopted involved the following steps:

- Segment Customers into groups of same type and expected consumption
- Determine the average consumption per segment
- Plot the distribution of the consumption for that segment
- Identify abnormal consumption customers
- Conduct site visits and resolve factors causing abnormal consumption

The end result of the approach adopted is to ensure that it is possible to identify a possible loss customer on an individual premise basis and then send out a resource, which would then go and audit that specific premise. This approach, if a success would then be able to save the business time and money as it provides a huge advantage over the blanket audits that is the current practice followed in Eskom Distribution.

Segmentation Criteria

Customer segmentation is an ideal way to deal with energy losses because each customer read can be compared to other customers with the same profile (i.e. same tariff, customer type, meter read type, etc.) located within the same area [6]. Keeping this target in mind it is necessary to identify and analyze the factors that might influence the average consumption of a specific segment. This will be dealt with in greater detail in section 4. Once the critical factors or basis variables are determined, the rest of the variables are discarded and the segmentation process is carried out using only the identified critical variables.

The different factors initially assumed to contribute to the consumption patterns of the customers are listed below,

1) Customer Type
2) Customer Service Area
3) Notified Maximum Demand (NMD) value
4) Voltage Level
5) Service Point (SP) count
6) Urban/Rural (geographical split)
7) Standard International Code (SIC)

Once these basis variables are chosen, the next step in segmentation is to use the identified variables and proceed to group the customers into the different segments based on this criteria. The more the number of variables considered, the greater will be the total number of possible segments. Hence it is critical to determine the significance of each of the initially chosen factors through some form of significance analysis on the basis variables. Regression analysis of the variables selected enables us to determine the critical variables returned and the impact each variable has on the average consumption of the customers. An elaborate look at the methodology used for the regression analysis and numerical analyses performed on the data can be found in section 4. On deciding the segmentation criteria to be used it was also observed that the consumption data for certain customers have intermittent periods where the consumption data has not been populated. This could possibly be due to the customer being billed for three or four consecutive months all together in a single billing or lack of sufficient billing information for those months. In order to smoothen out such discrepancies, a twelve month moving average (12 MMA) was applied to the consumption data for all the customers. The consumption data obtained was for the period September 2003 – July 2006 (35 months).

Consumption Distributions per Segment

For the customer consumptions considered the average of twelve months data was taken into consideration also taking into consideration the fact as to whether the customer was active or not. In order to target the loss customers, the approach used is to identify the average consumption of the customers within a segment and then identify the mean, standard deviation, median, minimum value, maximum value, 25th and 95th percentile values from the distribution. The values of the mean and standard deviation give the user an indication of the nature of the distribution of the segment considered. Once the distribution has been plotted, the customers to be targeted will be any consuming below the lowest 25th percentile within the selected segment. This will be the primary criteria in targeting loss customers across all the different segments.

4. Data Analysis

The adoption of a segmentation criterion involving a systematic and logical approach becomes crucial, prior to arriving at the final segments. In accordance with this approach, it is necessary to identify and analyze the factors that might influence the variables considered, which in this case should be the average consumption. Care should also be taken to keep the total number of segments within acceptable limits, and this can only be done by ensuring that only the most critical variables are used in the final segmentation process.

A regression analysis was carried out on the basis variables to determine the significant segmentation variables. Regression analysis is a statistical analysis method that is used to describe and evaluate the relationship between a given variable (usually called dependant variable) and one or more other variables (usually called independent variable). The technique involves developing a mathematical equation or model that accurately describes the nature of the relationship that exists between the dependant variable and the independent variable. The aim of the whole process is to ascertain the relationship, or the causal effect between the dependant and independent variables [7].

Regression analysis with a single explanatory variable is termed as simple regression. Multiple regression analysis is a technique that allows additional factors to enter the analysis separately so that the effect of each can be estimated. This technique is a valuable tool for quantifying the impact of an arbitrarily large number of explanatory variables. This technique is of particular interest in the stratification process as there are a number of independent variables that are to be evaluated and predictors or explanatory variables with the smallest significance in the regression test can be eliminated. This decision will be made during the regression analysis using the coefficient of determination (R^2) values which acts as a measure of the strength of the linear relationship between the independent and dependent variable. The aim is to obtain a R^2 value as close as possible to the adjusted R^2 value to have confidence in the outputs of the regression analysis.

Regression Analysis for Customer Stratification

The magnitude of Eskom's customer data is such that it requires powerful software tools to carry out the regression analysis. SAS Enterprise Guide [5] was used to carry out the regression analysis for customer stratification (SAS Enterprise Guide is a very powerful software tool that has the ability to handle large amounts of information and perform a number of data manipulations, analytical testing, statistical analyses, and host of other mathematical operations with ease).

A random sample (10% of population size - approx 60,000) was considered for the analysis. Average consumption was selected as the dependent variable. The following variables assumed to influence the average consumption were selected as the independent variables.

- Customer Type
- NMD Range
- Voltage Level
- Urban/Rural
- SP Count

Tests for normality and heteroscedasticty were carried out. Tests for multi co-linearity amongst the considered independent variables to avoid any redundancy of information were also performed. Once all these tests were performed, the regression analysis was carried out and the variables that have the most significant contribution towards a change in the consumption were determined. The variables of significance that were returned are;

- NMD range; &
- Customer Type

The coefficient of determination R^2 value obtained was 82.3 and adjusted R^2 value obtained was 82.3 which shows a very high level of confidence in the output returned. It also showed that NMD range accounted for the majority of the contribution along with Customer type coming in the next highest. Hence NMD range and Customer type will be used as the main independent variables along with Customer service area in the final customer segmentation process. The independent variables were hence reduced to three main variables after the regression analysis process.

5. Results

The Customer Stratification methodology was tested on a sample set of feeders provided by the SPU audit team in North West Region. Field audits carried out on a sample set of predominantly agricultural feeders within the region during May 2006 were compared to the customers targeted through the stratification methodology for the same sample set. The results have shown that on an average a hit rate of almost 52% is achieved using the stratification method on the selected sample. Taking into consideration that Eskom Distribution's current approach of using blanket audits has a much lower hit-rate, the magnitude of the impact that the stratification methodology can provide is immense.

An example of the PAS feeder's consumption distribution is shown in Fig. 2 indicating the lowest, average and 25th percentile consumptions as an example.

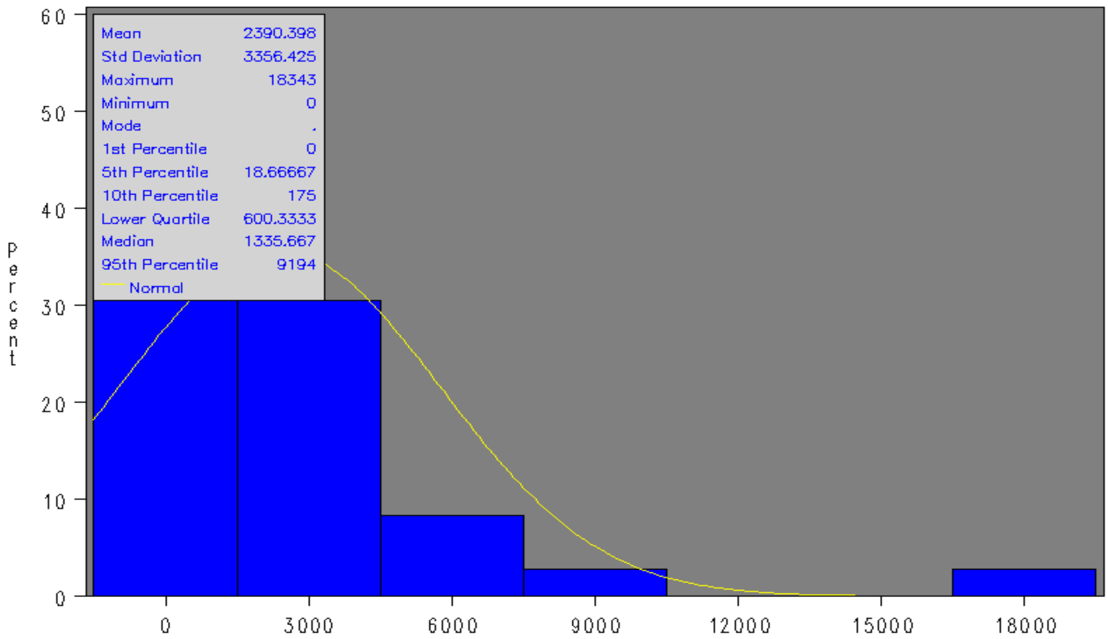

Fig. 2 PAS feeder example

Table 1 shows the results obtained on the sample set of six feeders considered for the study. The representation of the results obtained for the various segments per feeder are also highlighted.

FEEDER NAME	CUSTOMER TYPE	NMD-RANGE	25TH PERCENTILE	PREMISES <= 25TH PERCENTILE	MATCH WITH AUDIT	HIT RATE
PAL	AGRICULTURAL	0-25	363	8	7	87.50%
PAL	AGRICULTURAL	26-50	484	9	5	55.56%
PAL	AGRICULTURAL	51-100	300	9	9	100.00%
HKD	AGRICULTURAL	0-25	508	13	4	30.77%
HKD	AGRICULTURAL	26-50	404	11	3	27.27%
PAS	AGRICULTURAL	0-25	472	8	4	50.00%
PAS	AGRICULTURAL	25-50	212	13	9	69.23%
PAS	AGRICULTURAL	50-100	752	8	5	62.50%
GAAN	AGRICULTURAL	0-25	196.5	15	10	66.67%
GAAN	AGRICULTURAL	25-50	1030	14	8	57.14%
BLBE	AGRICULTURAL	0-25	190.5	14	3	21.43%
BLBE	AGRICULTURAL	25-50	645	8	6	75.00%
BLBE	AGRICULTURAL	51-100	1060	7	4	57.14%

GATC	AGRICULTURAL	0-25	255	26	7	26.92%
GATC	AGRICULTURAL	25-50	560	34	12	35.29%
GATC	AGRICULTURAL	51-100	869	15	3	20.00%
GATC	RESIDENTIAL	0-25	463	9	3	33.33%
					AVERAGE	51.52%

Table 1 Results comparison with North West Region's Audits

Further comparison tests are to be conducted in the other regions for other typical networks with different customer groupings. Customization of target percentiles for the lower cut-off values are to be finalized. Tests for the 95th percentile consumption are also to be tested to target customers who might be allocated in incorrect segments.

6. Conclusions

The Customer Stratification methodology has shown to provide a means of effectively targeting possible loss customers with a relatively high hit rate in comparison to the current blanket audit approach within Distribution. This approach allows for a cost effective approach for utilities to curb their non-technical losses and provides a mechanism to target individual customers across the broad spectrum of customer types and geographic locations. Further improvements to the methodology are to be incorporated based on further tests that are to be performed in order to customize the methodology to be more effective in its targeting capabilities.

References

[1] D. Suriyamongkol, "Non-Technical Losses in Power Systems", thesis submitted at *The University of Ohio*, as part of Master of Science 2002.

[2] A Strategy for the management of energy losses in a local electricity distribution network. Johannes Wilhelmus Fourie – University of Pretoria

[3] K. Dramowicz, "A Primer on How to Create a Customized Segmentation System", November 2004.

[4] AMA Core Marketing – Segmentation –American Marketing Association

[5] www.sas.com

[6] "Customer Segmentation, Key Questions Your Peers are Asking," http://www.forum.executiveboard.com/BLF/1,3204,0-0-Public_Display - 116011,00.html

[7] G. Keller, "Statistics for Management and Economics", Thomson Brooks / Cole Publications, ISBN 0 – 495 – 01339 – 0.