



**SCHOOL OF STATISTICS**  
UNIVERSITY OF THE PHILIPPINES DILIMAN



## WORKING PAPER SERIES

**Identifying Influencers of Consumer Activity**  
by

*Angela D. Nalica<sup>1</sup>, Erniel B. Barrios<sup>2</sup>,*  
and  
*Joseph Ryan G. Lansangan<sup>3</sup>*

UPSS Working Paper No. 2010-15  
September 2010

School of Statistics  
Ramon Magsaysay Avenue  
U.P. Diliman, Quezon City  
Telefax: 928-08-81  
Email: [updstat@yahoo.com](mailto:updstat@yahoo.com)

---

<sup>1</sup> Assistant Professor, School of Statistics, University of the Philippines Diliman

<sup>2</sup> Professor, School of Statistics, University of the Philippines Diliman

<sup>3</sup> Assistant Professor, School of Statistics, University of the Philippines Diliman

Marketing activation usually entails a universal blast of information to all consumers. Oftentimes, only a small proportion of the consumers react positively to such activation, resulting to waste in marketing expenses. If a circle of influencers can be identified for certain events or phenomena, then such activities can be focused into a group of factors or individuals, thus, optimizing the outcomes. With the identification of such group of influencers, resources for strategic optimization of outcomes can be allocated efficiently. A usage database is used to identify consumers who could initiate or influence the complex dynamics of consumer behavior. The data mining process of clustering, sampling, aggregation, modeling, and validation are used to mine such information from the database.

**Keywords:** logistic regression, segmentation, influencers, consumer behaviour, customer relationship management

## 1. Introduction

The consumer products and service industry changed from a rapidly growing market into a state of intense competition and saturation. Companies are now becoming more and more aware of the crucial importance of their existing clients database as inputs for marketing strategies. Predictions of behavior, customer value, customer satisfaction and customer loyalty are examples of some of the information that can be extracted from the data that should already be stored within a company's database. Examples of such data are those from scanned information in the point of sales (POS) machines, automatic teller machines (ATM), credit card transactions, and other records of customer/consumer interactions.

There are many marketing activities that are being bombarded to the consumers. However, not all consumers are responsive to these types of activities. It is thus a waste of resource and time to include the entire customer base to these types of activities.

The major motivation for this study is the *Tipping Point Hypothesis*, (Gladwell, 2002) which states the following: (a) a circle of influencers can be identified for certain events or phenomena, and (b) that the search for explanations for the occurrence or onset of certain events or phenomena can be zoomed into a group of factors or individuals. This will facilitate problem solving, mitigation of certain negative effects or simply optimize the outcomes. In the service sector or recurring purchasable products/services, the hypothesis is translated into the following:

- There are circles or segments of customers responsible (or influence or are catalysts) of some activities like acquisition/initiation, churn, patronage, or loyalty.
- A group of influencers can be identified so that resources for strategic optimization of outcomes of such events can be allocated efficiently. There is no need to spread thin the resources to the whole population of customers, but rather focus on segments that can yield multiplier effect in fast-tracking acquisition, improvement of patronage and loyalty and mitigation of churns. This will simplify marketing programs and activities focused only on selected few customers.
- It is possible to identify from the existing database key customers who could initiate or influence the complex dynamics of consumer behavior in general.

This study desires to derive some empirical evidence on the tipping point hypotheses above, specifically, on churn and loyalty of customers. Also, if segments or circles of influencers exist, the next step is to identify or determine their profiles. Contingent on the existence of influencers, a roadmap on how a continuous consumer tracking system or customer

relationship management (CRM) can be planned and implemented from the results of this study.

## **2. Customer Relationship Management**

There are various definitions of customer relationship management. (Chalmeta, 2006) defined CRM as a set of business, marketing and communication strategies and technological infrastructure designed with the aim of building a lasting relationship with customers, which involves identifying, understanding and meeting their needs. Although most companies are using a customer-focused strategy which demands the use of CRM solutions, many of its implementation failed. This is mainly due to the inadequacy of existing methodologies used to approach a CRM project since they do not satisfactorily integrate and complement the strategic and technological aspects of CRM.

(Coltman, 2007) noted that a superior CRM capability can create positional advantage and subsequent improved business performance. However, CRM suffers when it is not properly understood and implemented. Thus, in order to be successful, CRM should focus on the underlying or latent customer needs that emphasize a proactive market orientation. (Coltman, 2007) further demonstrates that by integrating three schools of thought – capabilities, market orientation and conversion feasibility – CRM programs can be successful.

There are also some models proposed to model consumer acceptance probabilities. (Thomas et al, 2006) introduced techniques that can be used to build models of the probabilities that a particular consumer will accept different variants of a generic borrowing product such as credit card. It was emphasized that the probability acceptance models will become increasingly vital as the consumer lending markets mature and the market becomes more consumer-centered. The models also satisfy the marketing credo of tailoring the product to the customer.

Because of saturated markets and fierce competition, the focus is now being shifted from building a large customer base to keeping the existing ones – i.e. prevent customers from churning. It is much cheaper to retain customers than to gain new ones. Thus, attention has been directed on coming up with churn models.

(Glady et al, 2009) propose a new framework for the churning detection process using the earnings a customer brings to the company. They use customer lifetime value (CLV) defined as the discounted value of future marginal earnings, based on the customer's activity. They assess the performance of several classifiers for churn prediction such as logistic regression, decision tree and neural networks together with cost-sensitive classifiers such as AdaCost and cost-sensitive decision tree. (Glady et al, 2009) further emphasized that besides good overall classification, it is also important to correctly classify potentially profitable churners.

(Hadden et al, 2005) reviewed the main trends for developing models to predict consumer behavior specifically those used in the development of customer churn management platform. These include traditional methods such as decision trees, regression analysis, Markov models, and genetic algorithms (GA), among others. Most modeling efforts focused mainly on feature extraction rather than feature selection problem.

(Popović and Bašić, 2009) illustrated how data mining methods based on the fuzzy logic could be successfully applied in the retail banking analysis and, moreover, that the fuzzy c-means clustering performed better than the classical clustering algorithms in the problem of churn prediction. Canonical discriminant analysis was used to reveal variables that provide maximal separation between clusters of churners and non-churners.

(Padmanabhan and Tuzhilin, 2003) further emphasized that automated CRM problems of customer analysis, customer interactions, and optimization performance methods can be better analyzed with the simultaneous use of data mining and optimization methods.

### **3. Methodology**

The data used in verifying the hypothesis was derived from the database of customers of a service company with a nationwide coverage of the services offered. Information about detailed transactions is stored in a data warehouse. Data on the usage and frequency of use of certain services are used in this study. At the customer level, transactional weekly time series data for 1 year are used. Aggregated macro level indicators for the same reference year are likewise used.

Cluster analysis was employed to generate simple usage groupings based on usage profile and revenue indicators. Once groupings have been identified, a sample of customers by cluster was obtained and information on detailed activities was gathered. This helps in the identification of possible influencers.

To generate company-level (macro-level) benchmark of various activities, several indicators were aggregated on a weekly basis over a period of 53 weeks. Together with some customer/account level variables, these were used in coming up with transfer function models (TFM). A TFM with macro-level activities (number of churns and number of continuing accounts) as outputs and consumer activities as inputs were fitted for each of the potential influencer from the segmentation of customers.

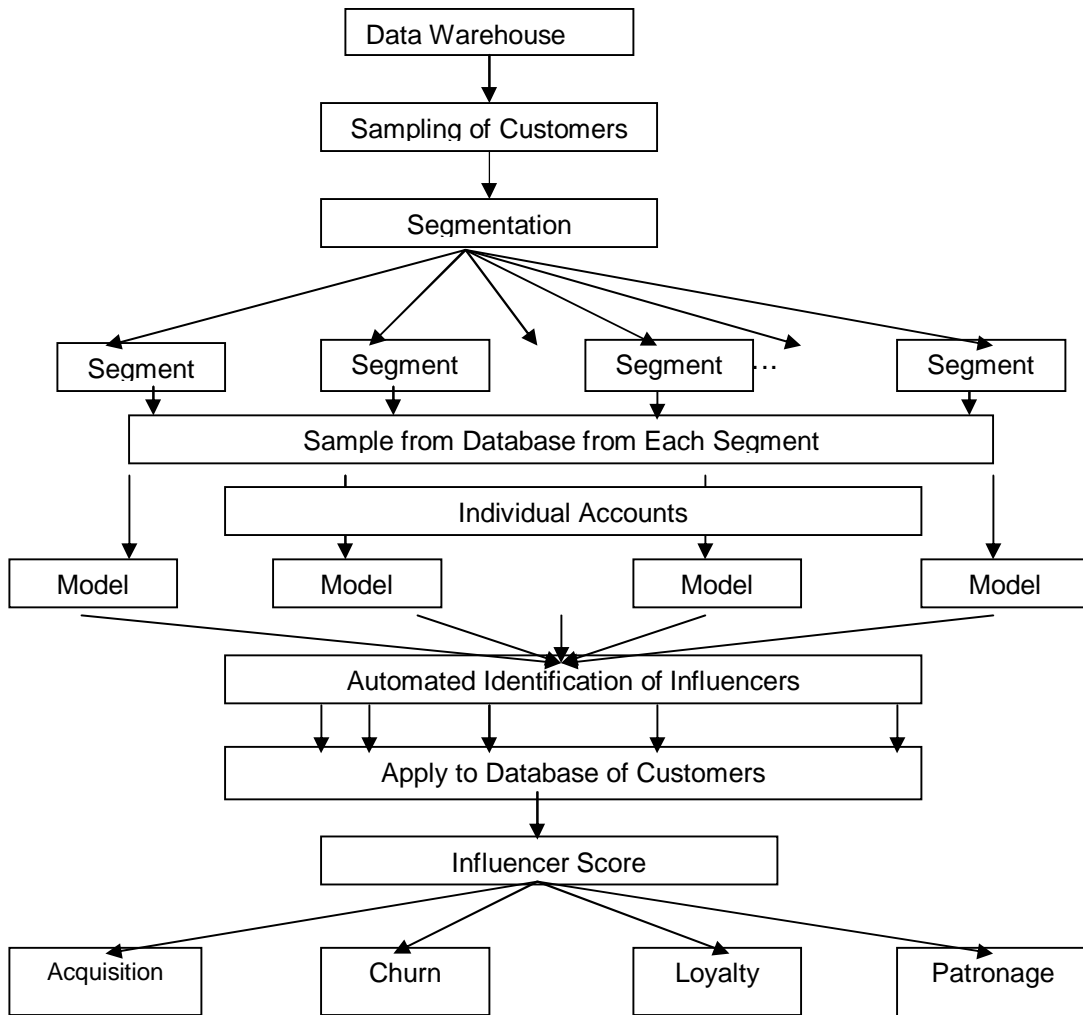
For churn, a customer that exhibit significant cross-correlation between the number of churns and the occurrence of its own churn is tagged as influencer and the delay (in weeks) is identified from the same cross-correlations. This results to a dichotomy that will either confirm or negate the postulation that the customer is an influencer of churn.

For loyalty, a customer that exhibit significant cross-correlation between the number of continuing accounts and the number of distinct transactions made by the customer is tagged as influencer and the delay (in weeks) is identified from the same cross-correlations. This results to a dichotomy that will either confirm or negate the postulate that the customer is an influencer of loyalty.

The results of the TFM were then used in logistic regression modeling that will help automate the determination of influencers. Outputs from logistic regression together with principal components analysis were then used to come up with influencer scoring equation. For each activity (churn and loyalty), customers were considered as possible influencer or non-influencer based on some pre-defined characteristics. For example, in modeling churn, the number of distinct relations (for different services) was counted.

It is important to determine the soundness and validity of the results. Hence, additional samples of about 10% of the training set are used as validation data set.

The Flow Chart of the activities done can be summarized as follows:



#### 4. Results and Discussion

For purposes of identification of influencers from the history of usage, a vital assumption is that influence takes effect when the customer uses the services provided by the company. The segments then should be able to aptly represent various dynamics of customer usage of services provided by the company.

The results showed that 7% or a total of 91 sample customers are potential influencers and the remaining 93% or 1,214 sample customers are non-influencers which exhibited low usage on practically all services (i.e. low linkages to other customers thereby reducing the opportunity to influence acquisition, churn, loyalty or patronage).

#### 4.1 Churn Model

A sample of 205 churners was stratified by week of churning from week 14 to week 40. About 72 or 35% of the sample churners are not influencers and more than one-half influence churning on the same week they churned. More than half of the potential influencer of churn actual influenced churn counts within the same week the account churned (contemporaneous), over 3% influences churning one week after they churned, 8% influences 2 weeks after, and a very few would influence churners after 3-8 weeks (see Table 1).

**Table 1. Churn Influencers By Number of Weeks After They Churned**

Weeks After Influence Is Measured	Frequency	Percent
0	103	50.49
1	7	3.43
2	16	7.84
3	1	0.49
4	1	0.49
6	1	0.49
7	2	0.98
8	1	0.49
Not Significant	72	35.29

Among the non-influencers, 95% do not have any activity in the last 6 months. While among those who can influence, 41% still have activities in the last 6 months prior to churning.

There is a problem however on the use of logistic regression to automate identification of influencers since there is minimal usage months prior to churning. A possible solution is to use churn prediction model among those who are probable churners but still have activities and can be considered probable influencers.

The aggregate extent of influence is impossible to compute because the week of churning was a variable used as a basis in assessing the impact of influencers. There are several accounts that could churn in that same week as other accounts churn. This is an empirical assessment of causation between the week they churned and the number of churners. Also, once the account churned, no more activity is observed, thus it will be impossible to verify the extent of influence.

The best way to validate churning influence is to look at the activities before churning, whether the account relationships/circles (i.e. linkages to other customers within the same service/s) include accounts that also churned. Problem though is that 6 months prior to churning, the accounts hardly had any activities, and the data warehouse keeps data only for 13 months. It is impossible then to trace back many months before churning to understand the usage dynamics of churners. One possible solution here is to retrieve data from archives, but at the time of study, such effort takes longer period.

#### 4.2 Loyalty Model

The study also aimed to identify influencers of continuing accounts, i.e. model on loyalty. From a sample of 1,305 loyal accounts, the number of distinct transactions/activities was determined per account. It was assumed that the more distinct transactions made, there will be more loyal subscribers associated with this account. Results show that 172 or 13.4% are

significant influencers based on transaction A, 10 or 0.78% are influencers based on Transaction B. Overall, 177 or 13.8% are influencers based on the distinct Transactions A/B. This supports the earlier observation that influence is exhibited more on Transaction A and less on Transaction B.

Influencers can be identified based on quality of transaction (higher revenue) together with the quantity (no. of transactions). To model the probability of being an influencer of continuing accounts, logistic regression model is used and results show that factors (mostly quantity measures) are significant. It shows that the odds of being influencer are higher when the proportion of incoming transactions from another account from a competitor is high. Although some indicators that are included in the model are not significant, they are not removed from the model since they contribute towards the stability of the model.

### **4.3 Business Implications**

#### **Churns**

Results showed that some churners can potentially influence other churners. Generally, churners have minimal usage details 6 months prior to churning, making it difficult to validate influencers of churn. However, there is an indication that potential influencers still exhibit usage/transactions a few months prior to churning. As churners can be classified as either voluntary or involuntary, it is thus recommended that the churn prediction model be used in predicting churners. Then, among those with high chance of churning, current activities can be assessed, and those that exhibit relatively higher usage level can be considered as potential influencers.

#### **Loyalty**

The output from the transfer function model shows the extent of the influence of each account on the number of continuing accounts are computed. The results show an average impact of influencers. Some accounts influence immediately, while others may have lags up to about 4 weeks in actual increase of the number of continuing accounts. If incentives are given for 1,071 accounts (<1% of actual base, on the assumption that they are the only influencers) every time their account relationships/circles increase, they can encourage from 98-180 weekly loyal accounts. In the most optimistic case, they can bring in almost 1 loyal account each on a weekly basis.

### **5. Conclusions and Recommendations**

It is verified that influencers of certain activities such as churn and loyalty can be identified from the customer database. Expenditures aimed to activate certain activities need not be spread across all customers. By focusing on some fewer influencers only, activation can become more efficient and higher returns can be expected.

Further validation is recommended by doing the following:

- Conducting an FGD on the possible influencers.
- Conducting a survey that will serve as a baseline of customer preferences and conditions or situations that will motivate them to influence other customers to do certain activities (with use of qualitative questions).
- The survey shall also provide input data in profiling the influencers. This will facilitate the development of strategies that will encourage them to influence other customers to engage into a specific target activity.

This study identified the influencers through a purposive selection of samples from possible influencers and non-influencers. Purposive selection is intended to increase the likelihood of selecting influencers (hypothesized to be smaller in number compared to the total customer database). The individual transaction patterns were analyzed to confirm the customer's being an influencer or not. Identification of influencers is automated through a logistic regression with monthly aggregate measures as determinants.

- The hypothesized influencers, once identified, can be subjected to a continuous customer tracking (CCT).
- Activation or promotions can be focused on certain influencers and their reaction in terms of usage and transactions with internal and/or external circles can be analyzed.
- Such data can be used not only to validate a customer being influencer, but it can also be used in studying effectiveness and efficiency of certain activation strategies. The number of new customers per peso spent on activating influencers of acquisition can be derived from a CCT data.
- CCT data can be generated from usage history and can be augmented by response to one or two questions (related to marketing activity and/or strategy) pushed to the customers

#### Reference:

- Chalmeta, R., 2006, Methodology for Customer Relationship Management, *The Journal of Systems and Software*, 79:1015-1024.
- Coltman, T., 2007, Why Build a Customer Relationship Management Capability?, *Journal of Strategic Information Systems*, 16: 301-320.
- Gladwell, M., 2002, **The Tipping Point: How Little Things Can Make a Big Difference**, New York:Little Brown and Company.
- Glady, N., Baesens, B., Croux, C., 2009, Modeling Churn Using Customer Lifetime Value, *European Journal of Operational Research*, 197: 402-411.
- Hadden, J., Tiwari, A., Roy, R., and Ruta, R., 2005, Computer Assisted Customer Churn Management: State-of-the-art and Future Trends, *Computers and Operations Research*, 34:2902-2917.
- Padmanabhan, B. and Tuzhilin, A., 2003, On the Use of Optimization for Data Mining: Theoretical Interactions and eCRM Opportunities, *Management Science*, 49: 1327-1343.
- Popović, D. and Bašić, B., 2009, Churn Prediction Model in Retail Banking Using Fuzzy C-Means, *Informatica*, 33:243-247.
- Thomas, L., Jung, K., Thomas, S., and Wu, Y., 2006, Modeling Consumer Acceptance Probabilities, *Expert Systems with Applications*, 30:499-406.



