

화장품 고객 정보를 이용한 마이크로 마케팅[†]

석경하¹ · 조대현² · 김병수³ · 이종언⁴ · 백승훈⁵ · 전유중⁵ · 이영배⁵ · 김재길⁵

¹²³⁵인제대학교 데이터정보학과

⁴미켄즈 (주)

접수 2010년 4월 17일, 수정 2010년 5월 20일, 게재확정 2010년 5월 23일

요약

고객 정보를 활용하는 방법에는 고객의 구매액을 활용한 마일리지 방법과 구매 횟수에 따라 등급을 나누어 활용하는 방법 등이 있다. 본 연구에서는 회사 매출에 직결되는 고객의 재구매 여부에 초점을 맞추어 고객정보와 구매정보를 이용하여 로지스틱 회귀분석을 통한 재구매 예측 모형을 만들었다. 예측 모형 평가 측도로는 하이드게 점수를 사용하였으며 하이드게 점수를 최대로 하는 점수를 기준으로 분계점을 선택하였다. 재구매 예측모형을 이용하여 재구매 지수를 만들어 고객을 등급화하여 보다 효율적인 고객 관리가 가능하게 하였다.

주요용어: 로지스틱 회귀분석, 분계점, 재구매 모형, 재구매지수, 하이드게 점수

1. 서론

과거에는 전체 시장의 구성원들이 원하는 바가 균집화 패턴을 보이지 않거나 시장세분화가 불가능 할 때 전체 시장에 대하여 한 가지 마케팅 믹스를 제공하는 전략을 사용하였다. 이를 대량마케팅 (Mass Marketing)이라고 한다. 그러나 현대사회에서는 눈부신 컴퓨팅 기술의 발전을 바탕으로 엄청난 고객의 정보를 저장 가공할 수 있어 전체 시장 구성원들이 원하는 바의 차이를 인식하여 시장을 보다 세분화하고 세분화된 시장별로 최적의 마케팅 믹스를 개발하여 제공하는 것이 가능하게 되어 보다 고객의 필요를 만족시킬 수 있게 되었다 (김이태, 2005; 지혜영과 조완현, 2009; Cho 와 Park, 2007).

이러한 대량마케팅에서 CRM (Customer Relation Management)으로의 마케팅 기법의 발전에서 기본적인 가정은 개별 고객을 보다 잘 이해하고 성향과 욕구를 파악하는데 있다. 그러므로 현재 기업들은 전략적 마케팅과 전략적 경영을 위하여 크게 세 가지 사항을 체크해야 한다. 첫째, “우리 회사의 고객의 특성에 대하여 상세히 알고 있는가?”이다. 이를 위해 이상적인 고객 집단의 인구학적 특성과 원하는 제품이나 서비스, 선호도, 주요 소비자 행동특성에 대해 조사하고 분석한다. 둘째, 기업 이익에 직결되는 사항으로 “우리 회사의 이상적인 고객은 누구인가?”이다. 고객 중에 수익률이 높고 충성도가 높은 고객은 누구인가를 찾아야한다. 이는 타겟 마케팅을 실시하기 위해 필수적인 것이다. 셋째, “현재 고객의 특성이 이상적인 표적 고객의 특성에 적합/일치/부합 되는가?”이다. 기업이 가지고 있는 고객 데이

[†] 본 연구는 산학협동재단 2009년 학술연구비 지원에 의해 이루어졌음.

¹ (621-749) 경남 김해시 어방동 607, 인제대학교 데이터정보학과 통계정보연구소, 교수.

² 교신저자: (621-749) 경남 김해시 어방동 607, 인제대학교 데이터정보학과 통계정보연구소, 교수.
E-mail: statcho@inje.ac.kr

³ (621-749) 경남 김해시 어방동 607, 인제대학교 데이터정보학과 통계정보연구소, 조교수.

⁴ (135-010) 서울시 강남구 논현동 206 일당빌딩 7층, 미켄즈 (주), 연구소장.

⁵ (621-749) 경남 김해시 어방동 607, 인제대학교 데이터정보학과, 석사과정.

터를 활용하여 고객의 특성을 조사, 분석 하여 특성에 부합되는지를 평가하고, 부합되는 경우 유지 강화 방안과, 부합되지 않는 경우 유인책을 강구한다 (강현철 등, 2003; 고봉성 등, 2009; 조대현 등, 2009).

본 연구에서는 이러한 전략적 마케팅과 전략적 경영을 위한 세 가치를 통계적 방법론을 적용하여 점검하는 법을 제공하고자한다. 첫 번째 질문에 대하여 고객들의 과거 구매기록을 탐색적으로 분석하여 재구매에 영향을 주는 변수를 중심으로 로지스틱 회귀분석을 이용하여 재구매 확률을 예측하고 이를 통해 기업에 수익을 내는 재구매 고객들을 분류해낼 것이다. 두 번째 질문에 대해서는 선택된 변수를 통하여 이상적인 고객들의 특성을 파악할 것이다. 마지막 질문에 대하여는 효율적인 마케팅에 도움을 줄 수 있도록 이상적인 표적고객을 위하여 앞에서 구한 확률을 바탕으로 재구매 지수를 개발하고자 한다.

2장에서는 구매 자료에 대한 분석을 통해 모형화에 필요한 변수를 생성하고 생성된 변수를 이용하여 재구매 여부에 따른 변수들을 탐색하고, 중요한 변수를 중심으로 이상적인 고객 특성을 파악한다. 3장에서는 로지스틱 회귀분석을 이용하여 각 제품별 재구매 예측 모형을 개발하여 재구매 확률과 재구매 여부를 파악한다. 4장에서는 앞에서 구한 재구매 확률을 통하여 재구매 지수를 개발하고 지수에 따른 고객의 구매성향을 파악한다. 5장에서는 결론 및 향후 과제에 대하여 기술한다.

2. 자료 분석

2.1. 원자료

원자료는 우리나라의 SKII 화장품에 대한 2005년 1월부터 2008년 12월 까지 총 48개월간의 구매고객들의 구매 정보이다. 원자료는 Essence를 포함하여 10가지 제품으로 구성되어 있고 총고객의 수는 498,748명이고 전체 거래 수는 1,557,054건으로 고객 1인당 평균거래 수는 3.1건으로 나타났다.

2.2. 분석 자료와 변수 생성

분석할 자료는 10가지의 제품에 대한 거래자료이다. 제품에 따른 고객들을 구매성향을 분석하기 위하여 제품별로 나누었다. 분석은 10개 제품군의 자료 중 고객의 수가 많고 구매간격이 뚜렷하게 나타나는 제품인 Essence를 분석한다. 2005년 1월부터 2006년 12월까지 총 2년의 자료를 이용하여 총 구매금액, 총 구매횟수, 총 구매수량, 나이, 평균구매간격, 기준시점과 마지막 거래일자와의 차 등의 변수를 생성한다.

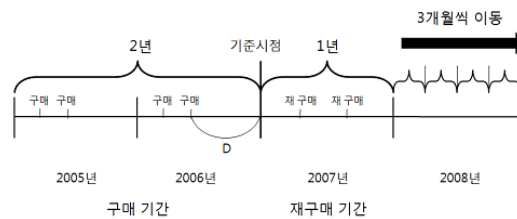


그림 2.1 변수 생성

다음으로 그림 2.1과 같이 기준시점 이후 1년인 2007년 1월부터 2007년 12월까지의 고객의 구매여부를 이용하여 재구매 변수를 생성한다. 또한 계절적인 영향을 제거하기 위하여 기준시점을 3개월 단위로 이동하면서 분석자료 집합을 구성하였다.

분석에 사용된 구매에 관련된 변수는 표 2.1과 같다. 종속변수는 재구매 여부 이고, 총 구매횟수는 그림과 같이 기준 시점 이전에 구매를 한 고객의 구매한 횟수이다. 평균구매간격은 구매기간 내에 고객이

표 2.1 분석에 사용한 변수 설명

변수	변수명	설명
id	customer_num	고객코드
독립변수	age	고객연령
	payment	총 구매금액
	qty	총 구매수량
	totalcount	총 구매횟수
	mcy	평균 구매간격
	D	마지막 거래일자 와의 차이
종속변수	re	고객의 재구매 여부

첫 번째 구매가 일어난 시점에서 마지막 구매 시점까지의 기간을 총 구매횟수에서 1을 뺀 값으로 나눈 변수이다. 마지막 거래일자 와의 차이는 그림에서 D에 해당되는 값이다.

3. 재구매 분석

3.1. 재구매 변수 분석

자료 861,468개중 랜덤하게 40%를 골라 훈련용 자료로, 30%를 검증용 자료로, 나머지 30%를 평가용 자료로 사용하였고, 각각 자료에 따른 재구매자 비율은 표 3.1과 같다. 전체 자료의 재구매자 비율은 19.70%이고, 훈련용 자료의 재구매자 비율은 19.70%, 검증용 자료의 재구매자 비율은 19.67%, 평가용 자료의 재구매자 비율은 19.72%로 나왔다.

표 3.1 재구매자 비율표; 백분율 (%)

재구매	훈련용 자료	검증용 자료	평가용 자료	총 합
0	276708 (80.30)	207592 (80.33)	207483 (80.28)	691783 (80.30)
1	67879 (19.70)	50848 (19.67)	50958 (19.72)	169685 (19.70)
총 합	344587	258440	258441	861468

표 3.2와 그림 3.1은 재구매 여부에 따른 고객들의 평균구매 간격에 대한 기술통계량과 상자도표이다. 재구매를 하지 않았을 때 평균 구매간격의 평균은 484.23일, 재구매를 했을 때 구매간격의 평균은 266.49일로 나타났다.

표 3.2 재구매 여부에 따른 구매간격의 기술통계량

변수	재구매 유무	N	평균	표준편차	최소값	최대값
평균구매간격	0	691783	484.23	146.70	1.00	709.00
	1	169685	266.49	191.54	1.00	1093.00

표 3.3과 그림 3.2은 재구매 여부에 따른 총 구매횟수에 대한 기술통계량과 상자도표이다. 재구매를 하지 않은 경우의 총 구매횟수의 평균은 1.20회, 재구매를 한 경우의 총 구매횟수의 평균은 3.38회로 나타났다.

표 3.4와 그림 3.3에서 나타난 것과 같이 재구매에 따른 고객의 기준시점과 마지막 거래와 차이 평균을 나타낸다. 재구매를 하지 않았을 때 마지막 거래와의 차이의 평균은 303.91일, 재구매를 했을 때 마지막 거래와의 차이의 평균은 181.50일로 나타났다.

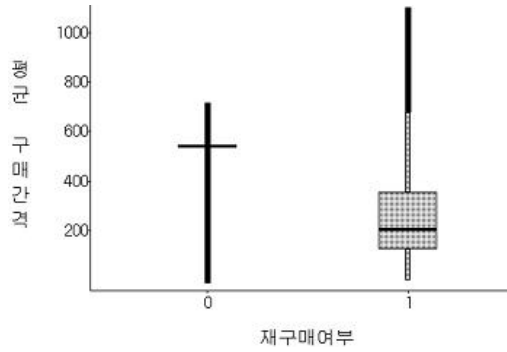


그림 3.1 재구매 여부에 따른 구매간격

표 3.3 재구매 여부에 따른 구매횟수의 기술통계량

변수	재구매 유무	N	평균	표준편차	최소값	최대값
총 구매횟수	0	691783	1.20	0.63	1.00	18.00
	1	169685	3.83	3.02	2.00	87.00

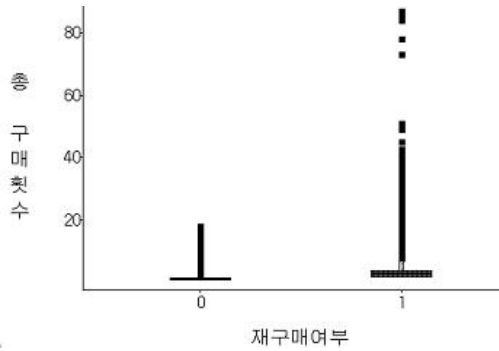


그림 3.2 재구매 여부에 따른 총 구매횟수

표 3.4 재구매 여부에 따른 마지막거래일자의 기술통계량

변수	재구매 유무	N	평균	표준편차	최소값	최대값
마지막거래일자와의간격	0	401297	303.91	203.07	1.00	730.00
	1	169685	181.50	166.49	1.00	730.00

총 구매횟수에 따른 평균구매간격을 알아 본 결과 표 3.5와 같은 결과를 얻었다. 구매횟수 전체의 평균 구매 간격은 214.69일이며, 두 번째 구매를 했을 때 272.57일로 나타났다. 두 번째 구매를 했을 때의 평균 구매간격은 신규구매가 일어난 다음에 다음 구매가 일어나기 까지 평균적으로 걸리는 시간으로 신규고객의 마케팅을 위해 유용한 정보가 될 수 있을 것이다. 구매횟수가 증가함에 따라 구매간격의 평균은 줄어드는 것을 알 수 있다.

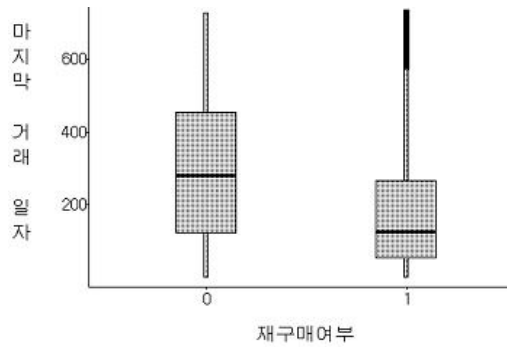


그림 3.3 재구매 여부에 따른 마지막 거래 일자와의 거리

표 3.5 구매횟수에 따른 평균구매간격

구매횟수	평균구매간격	
	평균	표준편차
총 구매횟수1 (전체)	214.69	178.36
총 구매횟수=2	272.57	217.21
총 구매횟수=3	188.93	103.08
총 구매횟수=4	151.75	66.00
총 구매횟수=5	129.63	49.01

3.2. 재구매에 대한 예측 모형

본 연구는 자료에 포함된 변수들을 이용하여 고객에 대한 재구매를 예측하는 모형을 개발한다. 예측 모형을 위해 로지스틱회귀분석을 사용한다. 전체자료를 훈련용 40%, 검증용 30%, 평가용 30%로 나누어 모형을 생성하고 평가하여 통계적 예측 모형을 생성한다.

예측모형이 얼마나 잘 예측하는가를 알아보기 위하여 정분류율을 구하여 예측 모형을 평가한다. 본 연구에서는 하이드게 점수 (Heidke skill score, Heidke, 1926)를 사용하여 분계점 (threshold)을 정한다. 하이드게 점수를 사용하는 방법이란 순수하게 예측 모형에 의해 분류한 경우에만 점수를 부여하여 우연에 의한 예측의 확률을 배제한 모형의 예측성을 평가할 수 있게 해준다. 표 3.6에 나타난 수치를 이용하여 와 같이 PCM (예측 모형에 의해 예측한 부분)과 PCR (임의로 예측된 부분)을 구하여 최종적으로 하이드게 점수를 얻을 수 있다. 하이드게 점수가 클수록 순수하게 예측에 의한 부분이 높아지게 된다.

$$PCM = \frac{A + D}{N}, \quad PCR = \frac{O_1 F_1 + O_2 F_2}{N^2}, \quad HSS = \frac{PCM - PCR}{1 - PCR}$$

본 연구에서는 하이드게 점수가 최대인 분계점을 선택하여 재구매 여부에 대한 예측성을 평가하였다.

4. 로지스틱 회귀분석

일반적으로 회귀분석이란 연속형 변수들에 대해 설명변수인 독립변수와 반응변수인 종속변수사이의 인과관계에 따른 수학적인 모형을 구하여 변수간의 관련성을 규명하는 분석기법이다. 또한 분석결과로

표 3.6 하이드게 점수를 최대화 하는 분계점 계산표

		Observations		Total
		YES	NO	
Forecasts	YES	A (hit)	B (false alarm)	$F_1 = A + B$
	NO	C (miss)	D (correct negative)	$F_2 = C + D$
Total		$O_1 = A + C$	$O_2 = B + D$	$N = A + B + C + D$

나온 회귀계수는 어떤 독립변수가 주어졌을 때 반응변수의 값을 예측하기 위해서 활용된다. 반응변수가 두 개 혹은 두 개 이상의 범주를 가지는 범주형 변수인 경우 로지스틱회귀 모형이 적용된다 (성웅현, 2001).

두 개의 범주를 가지는 이진형 변수가 반응변수인 경우 첫 번째 범주가 나올 확률을 $P(Y = 1) = p$ 로 나타내고 두 번째 범주가 나올 확률은 $P(Y = 0) = 1 - p$ 로 나타낼 수 있다. 또한 이진형 변수의 확률분포에서 평균은 $E(Y) = p$, 분산은 $Var(Y) = p(1 - p)$ 이다. 확률 p_i 가 x_i 에 대해서 선형관계를 가질 때 $p_i = \beta_0 + \beta_1 x_i$ 와 같은 함수식으로 표현 될 수 있다. 하지만 p_i 는 0과 1사이의 값을 가지므로 적절하지 않다. 이에 오즈비에 로그를 취한 로짓모형을 이용하여 회귀분석을 실시할 수 있다 (배희수 등, 2004).

$$\log\left(\frac{p_i}{1 - p_i}\right) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}$$

재구매 유무를 나타내는 변수를 반응변수로 하여 이진형 로지스틱회귀 모형을 한 결과 표 4.1과 같은 결과를 얻었으며 회귀식은 식 (4.1)과 같다. 변수선택 기준은 오분류율을 사용하였고 변수선택 방법은 단계적 선택방법으로 하였다. 모형의 다중공선성을 보기 위하여 VIF (variance inflation factor) 값을 구한 결과 가장 높은 값이 4.2370으로 나와 다중공선성은 무시하였다.

표 4.1 자료의 로지스틱회귀모형 회귀계수의 추정치

variable	DF	Estimate	s.e	Wald- χ^2	Pr-Chisq
Intercept	1	-6.2038	0.0409	23040.93	<.0001
age	1	0.0079	0.0009	79.94	<.0001
cy	1	0.0473	0.0002	40569.25	<.0001
payment	1	-9.23E-08	3.49E-08	6.99	0.0082
totalcount	1	0.6701	0.0077	7653.67	<.0001
D	1	-0.0002	7.2E-05	6.91	0.0085

$$\begin{aligned} \text{Logit}(p_{\text{Essence}}) &= \log\left(\frac{p}{1 - p}\right) \\ &= -6.2038 + 0.0079 \times \text{age} + 0.0473 \times \text{cy} - 9.23\text{E} - 08 \times \text{payment} \\ &\quad + 0.6701 \times \text{totalcount} - 0.0002 \times \text{D} \end{aligned} \tag{4.1}$$

표 4.2와 그림 4.1은 훈련용 자료에서 재구매 확률에 대한 기술통계량과 상자그림이다. \hat{p} 의 평균은 0.2970이고, 최소값은 0.0001보다 작은 값을 가지고, 최대값은 1.0000을 가진다. 그림을 보면 \hat{p} 이 낮은 쪽에 몰려 있음을 알 수 있다.

표 4.2 혼련용 자료에서 얻은 재구매 확률의 기술통계량

변수	N	평균	표준편차	최소값	최대값
\hat{p}	344587	0.2970	0.2947	0.0000	1.0000

표 4.3과 그림 4.2는 검증용 자료에서 재구매 확률에 대한 기술통계량과 상자그림이다. \hat{p} 의 평균은 0.2970이고, 최소값은 0.0001보다 작은 값을 가지고, 최대값은 1.0000을 가진다. 그림을 보면 \hat{p} 이 낮은 쪽에 몰려 있는 것을 알 수 있다.

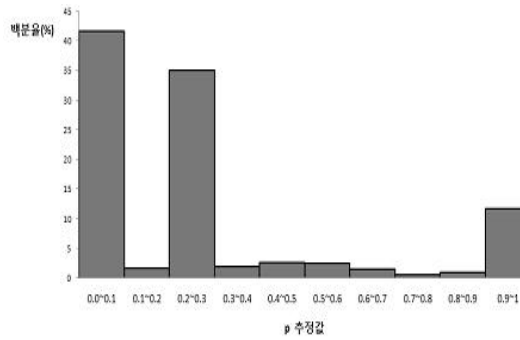


그림 4.1 혼련용 자료에서 얻은 재구매 확률의 분포

표 4.3 검증용 자료에서 얻은 재구매 확률의 기초통계량

변수	N	평균	표준편차	최소값	최대값
p	258440	0.2970	0.2946	0.0000	1.0000

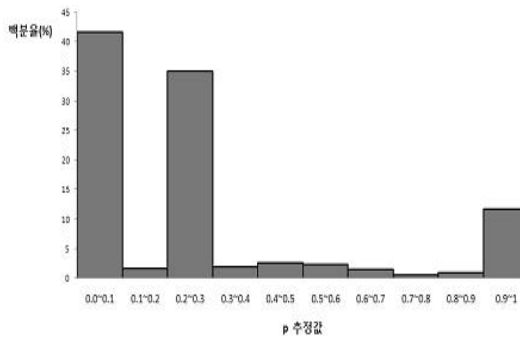


그림 4.2 검증용 자료에서 얻은 재구매 확률의 분포

표 4.4와 그림 4.3은 평가용 자료에서 재구매 확률에 대한 기술통계량과 상자그림이다. \hat{p} 의 평균은 0.2972이고, 최소값은 0.0001보다 작은 값을 가지고, 최대값은 1.0000을 가진다. 그림을 보면 \hat{p} 이 낮은 쪽에 몰려 있는 것을 알 수 있다.

표 4.4 평가용 자료에서 얻은 재구매 확률의 기초통계량

변수	N	평균	표준편차	최소값	최대값
p	258441	0.2972	0.2944	0.0000	1.0000

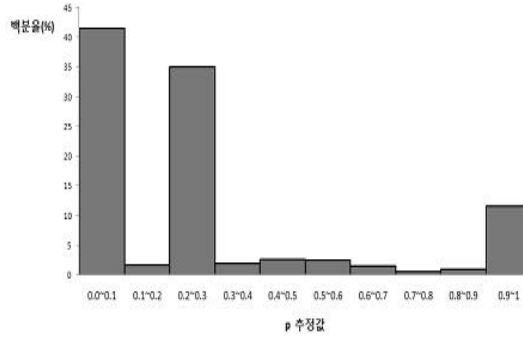


그림 4.3 평가용 자료에서 얻은 재구매 확률의 분포

다음으로 Essence 자료를 이용하여 생성된 예측 모형에 대한 훈련용, 검증용, 평가용 자료에 따른 정오분류표를 작성한다. 이는 HSS값이 가장 높은 분계점을 가지고 생성된 결과모형을 이용한 분류결과이다. 표 4.5를 보면 훈련용 자료와 검증용자료의 분계점은 모두 0.30이었다.

표 4.5 제품에 따른 정분류율의 비교

제품	자료	분계점	정분류율
Essence	훈련용	0.30	96.85
	검증용	0.30	96.86
	평가용	0.30	96.82

표 4.6 로지스틱 회귀모형을 이용한 검증용 자료의 정오분류표

Frequency(percent)	Observations		Total	
	0	1		
Forecasts	0	200554 (77.60)	7038 (2.72)	207592 (80.32)
	1	1065 (0.42)	49783 (19.26)	50848 (19.68)
Total	201619 (78.02)	56821 (21.98)	258440 (100.00)	

표 4.6은 정분류율이 가장 좋게 나타난 검증용자료의 정오분류표이다.

5. 재구매 지수의 생성 및 해석

5.1. 재구매 지수의 생성 및 분석

재구매 지수는 고객의 구매정보를 이용하여 재구매할 가능성을 지수화한 것으로 재구매 지수를 계산하는 과정은 로지스틱회귀분석 결과로 얻어진 회귀모형 (4.1)로부터 재구매확률의 추정값인 \hat{p} 을 이용한

다. 위에서 구한 재구매확률의 예측값 (\hat{p})에 100을 곱하여 재구매 지수 (RPI: repurchase index)로 사용한다. 즉, $RPI=100 \times \hat{p}$ 이다. RPI의 기초통계량은 표 5.1과 같다. 재구매지수의 평균은 29.7070으로 나타났다.

표 5.1 재구매지수 기초통계량

변수	N	평균	표준편차	최소값	최대값
재구매지수	861468	29.7070	29.4567	0.0000	100.0000

재구매지수를 이용하여 고객을 6개월 이내의 재구매 확률을 통하여 5등급으로 분류하였다. 분류결과에 대한 고객들에 대한 등급별 해석과 고객들의 비율은 표 5.2와 같다.

표 5.2 고객의 등급에 따른 재구매 지수의 범위와 해석

고객의 등급	RPI		해석	고객비율
	상한	하한		
1	100	86	재구매 지수가 매우 높은 고객군으로 6개월 이내 재 구매 확률이 90% 이상이다.	4.71%
2	86	63	재구매 지수가 높은 고객군으로 6개월 이내에 재구매 확률이 70% 이상이다.	13.55%
3	63	40	재구매 지수가 잠재된 고객군으로 6개월 이내 재구매 확률이 50% 정도이다.	44.10%
4	40	22	재구매 지수가 낮게 잠재된 고객군으로 1년 이내 재구매 확률이 40% 정도이다.	12.79%
5	22	0	재구매 지수가 아주 낮게 잠재된 고객군으로 1년 이내 재구매 확률이 20% 미만이다.	24.85%

고객의 재구매지수 등급에 따른 1개월 간의 미구매시와 1회 구매 시 재구매지수의 변화와 고객 등급 이탈율을 알아본 결과는 표 5.3과 같다. 동일한 등급에서 다른 변수들은 동일하나 최종구매일이 30 (pm 2)일 차이가 나는 고객들 간의 구매지수 비교 결과와 다른 변수들은 동일하나 총 구매횟수만 1회 차이가 나는 고객들 간의 구매지수 비교결과가 각각 2열과 3열에 나와 있다. 고객 이탈율은 등급별로 6개월 경과 후 등급에서 이탈하는 비율이 고객 이탈율이다. 등급이 낮을수록 1회 구매에 따른 재구매 지수의 변화가 심함을 알 수 있다. 또한 등급이 높을수록 고객 이탈율이 낮음을 알 수 있다.

표 5.3 재구매 지수의 이탈율

등급	재구매 지수 평균 감소율 (1개월 미구매시)	재구매 지수 평균 상승률 (1회 구매시)	고객이탈율
1등급	-3.85%p	3.47%p	5.52%
2등급	-6.51%p	15.24%p	15.82%
3등급	-6.77%p	20.18%p	33.11%
4등급	-4.09%p	24.17%p	59.31%
5등급	-2.13%p	30.22%p	71.44%

5.2. 재구매 지수별 매출 변화

재구매 지수에 의한 등급별 총매출이 전체에서 차지하는 비율이 표 5.4에 나와 있다. 1, 2, 3 등급 층성고객의 전체매출에서 차지하는 매출비율이 높으며 1등급의 1인당 매출액이 다른 등급에 비해 상대적으로 아주 높음을 알 수 있다. 이탈율이 높은 4, 5등급의 고객보다는 1, 2 등급의 고객들의 효율적 관리 방안이 매출액을 높이는데 더욱 도움이 될 것임을 알 수 있다.

6. 결론 및 향후과제

기업의 목적은 이익의 창출에 있다. 이익을 높이기 위해서는 고객의 구매 행태를 파악하여 고객의 재구매가 일어날 수 있도록 해야 한다. 본 연구 목적은 로지스틱회귀 분석을 이용하여 고객의 향후 재구매

표 5.4 재구매 지수의 등급별 매출 변화

등급	N	매출 (5년)	매출비율 (%)	1인당 매출액 (5년)
1	9601	81,708,609,000	32.52	8,510,427
2	27601	57,747,394,000	22.98	2,092,221
3	89883	62,837,547,000	25.01	699,104
4	26074	22,366,786,000	8.90	857,820
5	50653	26,595,748,000	10.59	525,058

여부와 재구매 시점을 파악하여 고객의 정보를 제공함으로써 마케팅 프로모션 시점을 찾아 보다 효율적인 마케팅을 통하여 매출을 극대화 하는데 많은 도움을 주고자 하는 데 있다.

본 연구에서는 2005년 1월부터 2008년 12월까지 총 48개월간의 구매를 한 고객들의 구매 정보를 이용하였다. 원자료에서의 구매 자료의 고객 수는 498,748명이고 거래 수는 1,557,054건이며 고객 1인당 평균거래 수는 3.1건이었다. 원자료 중에서 Essence 제품에 대한 구매자료 861,468건 고객 수 203,812명의 자료만을 분석에 사용하였다. 이 자료를 이용하여 새로운 변수를 생성하고 생성된 변수를 이용하여 재구매 예측 모형을 만들었다.

재구매 모형의 예측율을 평가하기 위해 정오분류표를 사용하였으며, 가장 높은 예측력을 가지는 분 계점을 찾는 과정에서 하이드게 점수를 최대로 하는 값을 이용하였다. 본 연구에서는 훈련용 자료 96.93%, 검증용 자료 96.97%, 평가용 자료 96.92%로 아주 높은 예측율을 보였다.

로지스틱 회귀모형에서 얻은 고객의 재구매 확률을 이용하여 재구매 지수를 개발하고 그 특징을 알아 보았다. 재구매 지수의 분포를 보면 50~60%p 대의 고객이 많았다. 이는 제품을 한번 구매하고 이탈하는 사람보다는 재구매를 적어도 1번 이상 하는 사람들이 많음을 나타낸다.

얻어진 재구매지수를 이용하여 고객들을 5등급으로 나누어 등급별 고객군의 특성을 제시 하여 이를 해석하였다. 에센스 고객군의 특징은 1등급 고객들의 개인 매출액이 다른 고객 군에 비해 월등히 높은 데 있다. 그리고 제품에 대한 충성도 또한 높아 1,2,3 등급에 고르게 높게 분포 되어 있다는 것이다. 이는 기존 고객에 대한 유치에 힘쓰고 신규고객의 충성화를 마케팅의 우선 목표로 정해야 함을 보여준다고 할 수 있다.

본 연구에서 얻어진 모형을 이용하면 고객에 대한 재구매에 대한 예측력을 가지고 고객들의 등급상태를 파악할 수 있다. 그러나 이러한 등급화를 이용하여 마케팅 프로모션을 한 경우 고객들의 등급변화와 등급에 대한 비율 변동을 알 수는 없다. 앞으로 이러한 이번 연구결과를 활용하여 나타나는 변화들을 연구할 필요가 있다. 마케팅 프로모션 결과로 얻어지는 자료들을 이용한 이러한 연구들을 통하여 고객들에 대한 좀 더 나은 이해와 구매패턴 파악을 위한 예측력이 높은 모형을 만들어야 할 것이다.

참고문헌

- 김이태 (2005). <CRM 고객관계관리>, 대경출판사.
 배회수, 조대현, 석경하, 최국렬, 김병수 (2004). < SAS Enterprise Miner를 이용한 데이터마이닝>, 교우사.
 성용현 (2001). <응용로지스틱 회귀분석>, 탐진.
 Cho, D., Kim, B., Seok, K.H., Lee, J., Kim, J. and Kim, S. (2009). A study on the behavior of cosmetic customers. *Journal of the Korean Data & Information Science Society*, **20**, 615-627.
 Cho, K. and Park, H. (2007). A study of association rule mining by clustering through data fusion. *Journal of the Korean Data & Information Science Society*, **18**, 927-935.
 Heidke, P. (1926). Berechnung des Erfolges und der Gute der Windstar kevorhersagen im Sturmwarnungsdienst. *Geografiska Annaler*, **8**, 301-349.
 Ji, H. and Cho, W. (2009). Predictive model plan of customer using purchasing items in internet shopping mall. *Journal of the Korean Data & Information Science Society*, **20**, 25-37.

- Kang, H., Han, S. and Shin, H. (2003). On the development of customer lifetime value evaluation model by using mining technique. *Journal of the Korean DataAnalysis Society*, **5**, 927-936
- Koh, B., Lee, S. and Heo, J. (2009). An empirical study on telemarketing efficiency at life insurance. *Journal of the Korean Data & Information Science Society*, **20**, 673-684.

Micro marketing using a cosmetic transaction data[†]

Kyung Ha Seok¹ · Dae Hyeon Cho² · Byung Soo Kim³ · Jong Un Lee⁴ · Seung Hun Paek⁵
Young Bae Lee⁵ · Yu Joong Jeon⁵ · Jae Gil Kim⁵

¹²³⁵Department of Data Science, Inje University

⁴Head of research center, Micans

Received 17 April 2010, revised 20 May 2010, accepted 23 May 2010

Abstract

There are two methods in grouping customers for micro marketing promotion. The one is based on how much they paid and the other is based on how many times they purchased. In this study we are interested in the repurchase probability of customers. By analysing the customer's transaction data and demographic data, we develop a forecasting model of repurchase and make repurchase indexes of them. As a modeling tool we use the logistic regression model. Finally we categorize the customers into five groups in according to their repurchase indexes so that we can control customers effectively and get higher profit.

Keywords: Logistic regression model, micro marketing promotion, repurchase index, repurchase probability, transaction data.

[†] This work was supported by the 2009 research grant from the Korea Sanhak Foundation.

¹ Professor, Department of Data Science Institute of statistical Information, Inje University, Kimhae 621-749, Korea.

² Corresponding author: Professor, Department of Data Science Institute of statistical Information, Inje University, Kimhae 621-749, Korea.
E-mail: statcho@inje.ac.kr

³ Assistant professor, Department of Data Science Institute of statistical Information, Inje University, Kimhae 621-749, Korea.

⁴ Head of research center, micans, Ilyang Bulding 7, Nonhyun Dong, Kangnam Gu, Seoul 135-010, Korea.

⁵ Graduate student, Department of Data Science, Inje University, Kimhae 621-749, Korea.