

# CRM 고객데이터 분석을 통한 이탈고객 연구\*

## A Study of Customer Churn by Analysing CRM Customer Data

김 상 용(Kim, Sang Yong)\*\*

송 지 연(Song, Ji Yeon)\*\*\*

이 기 순(Lee, Gi Soon)\*\*\*\*

고객관계관리(customer relationship management: 이하 CRM)는 고객에 대한 정보를 수집하고 수집된 정보를 효과적으로 활용하여 신규고객획득, 우수고객 유지, 고객가치 증진, 잠재고객 활성화, 평생 고객화의 순환을 통하여 고객을 적극적으로 관리하고 유지하며 고객의 가치를 극대화시키기 위한 기업 마케팅 전략의 일환이다. 특히 경쟁 환경이 급변하고 치열해 짐에 따라 기업의 수익 극대화를 위한 고객가치 증대 및 고객과의 관계 형성을 위한 CRM활동 중 고객의 이탈방지를 통한 유지관리의 중요성이 점차 커지고 있으며, 이러한 움직임은 고객 세분화를 통한 이탈고객 관리분석으로 주로 금융시장에서 다루어져왔다. 한편, 금융시장뿐만 아니라 모든 사업 분야에서 고객 유지 및 이탈방지를 위한 분석의 필요성은 높아지고 있다. 그 이유는 자사가 보유하고 있는 고객의 특성을 파악함으로써 기존의 고객을 효과적으로 유지·관리하여 고객이탈을 막는 것이 고객관리에서 점차 그 중요성을 더하기 때문이다. 그러나 아직까지 필요성만 대두될 뿐 어떠한 속성을 보유하고 있는 고객이 쉽게 이탈하는지를 판별할 수 있는 이탈고객에 대한 체계적인 연구가 진행되지 않았다는데 한계점이 있다. 이에 본 연구에서는 TV 홈쇼핑사의 실제 고객자료를 통하여 고객의 유지 및 이탈방지를 위한 CRM 전개방안, 이탈고객과 유지고객간의 인구통계적 속성 및 거래 행동의 특성 차이를 분석, 이탈에 미치는 영향력이 높은 변수를 밝혀내고 이탈고객예측 모형을 통하여 개별고객의 이탈확률을 예측하고자 했다. 더 나아가 실증 분석 결과를 바탕으로 이탈예측고객을 대상으로 고객 이탈을 방지하고 거래유지 및 활성화를 위한 CRM전개 방안을 도출, 이를 바탕으로 TV 홈쇼핑사가 수립해야할 마케팅 전략을 제시한다.

핵심개념: CRM, VIP 고객유지, 이탈고객

\* 본 연구는 고려대학교 특별 연구비에 의하여 수행되었음.

\*\* 고려대학교 경영대학 교수(sangkim@korea.ac.kr)

\*\*\* SK(주) 컨설턴트(ziofl@hanmail.net)

\*\*\*\* 고려대학교 대학원 경영학과 박사과정(min4504@korea.ac.kr)

## I. 서론

오늘날의 기업은 급속한 환경변화와 치열한 경쟁 하에서 신속한 전략수립으로 대표되는 새로운 패러다임의 비즈니스에 대한 대응이 요구되고 있다. 이러한 변화와 경쟁의 상황에서 고객관리의 문제는 중요성이 더욱 부각되고 있다. 따라서 기업이 고객을 올바르게 이해하고 행동을 예측하여 그들이 요구하는 제품과 서비스를 제공함으로써 기업은 성공적인 비즈니스를 영위하고 높은 성장을 이룰 수 있을 것이다(연순희, 최희수, 유시정 2003).

환경변화와 경쟁에 대응하고 기업이 수익 가치를 높이기 위해서는 한 번 획득한 고객을 적극적으로 관리하여 지속적으로 유지함으로써 기업에게 높은 수익을 가져다주는 충성 고객을 확보해야 한다. 이를 위해서 각 기업들은 고객관계관리라는 CRM 경영 기법을 도입하여 운영하고 있다. CRM은 고객에 대한 정보를 수집하고 수집된 정보를 분석하고 효과적으로 활용하여 '신규고객의 획득→기존 고객의 유지→휴면 고객 활성화→이탈고객 방지→우수고객 증대→평생 고객화'와 같은 순환과정을 통하여 고객의 가치를 극대화시킴으로써 기업의 수익성을 증대하기 위한 경영 전략이다. CRM경영기법을 적극적으로 활용하기 위해서는 무엇보다 획득한 고객 데이터 분석과정이 중요하다.

추적된 거래 데이터를 바탕으로 기존 고객과 잠재 고객의 요구를 잘 파악하고 그에 따른 제품과 서비스를 고객 접점을 통해서 제공하기 위한 프로세스인 CRM의 과정 중에서 기업 입장에서 가장 중요한 부분은 이미 획득한 고객

의 이탈을 방지하고 관계를 지속적으로 유지시켜서 수익 가치가 높은 평생 고객을 늘려가는 것이다. 고객이탈 방지 분석의 궁극적인 목적은 고객 이탈원인을 파악하고 이를 이용하여 이탈예상 집단에 대한 관리활동을 강화하여 이탈고객을 최소화시키는데 있다고 할 수 있다. 이탈고객의 정보를 잘 활용하면 기업 활동에 있어 잘못된 점을 파악할 수 있고, 이 문제점을 적절히 개선한다면 고객이탈은 더 이상 기업에게 있어 위협이 아닌 더 견고한 토대 위에서 고객과의 관계를 쌓아갈 수 있는 좋은 기회가 될 수 있다. 이는 신규고객을 확보하는데 더 많은 비용이 소요되고, 고객의 재구매가 이어지지 않으면 지속적 수익을 창출하기 어렵다는 점에서 뿐만 아니라, 고객과의 관계를 돈독히 함으로써 생애 가치가 큰 우수 고객을 확보할 수 있으므로 수익증대라는 기업의 목표달성 차원에서 중요하다. 그런데 현대 기업들의 문제는 대부분 고객의 속성 및 거래를 통해 얻어진 데이터를 보유하고 있지만, 이러한 데이터를 실질적으로 마케팅 목적에 맞게 적극적으로 활용하고 있지 못하다는데 있다.

특히 CATV홈쇼핑의 경우 대부분의 기존연구는 홈쇼핑을 이용하는 고객의 인구통계학적 분석, 라이프스타일, 만족 등의 홈쇼핑구매자들의 태도와 행동에 초점이 맞추어졌다. 미국의 홈쇼핑 시장의 경우 주부의 취업상태, 자녀의 수, 연령, 교육수준, 가계소득, 성, 직업 등이 홈쇼핑을 하는 소비자들의 특성을 설명하고 있으며(Berkowitz, Walton, and Walker 1979; Cox and Rich 1964; Cunningham and Cunningham 1973; Darian 1987; Gillette 1970), 국내의 경우 성별, 교육수준, 소득, 결혼, 직업과 같은 구

매자의 인구통계학적인 특성들과 주로 이용하는 시간대, 결제수단, 구매제품, 구매경험수준 등 CATV홈쇼핑을 이용하고 있는 소비자들의 이용실태에 관한 연구가 있다(김영숙, 심미영 2002). 이러한 CATV홈쇼핑을 이용하는 소비자들의 특성을 규명하는 것도 중요하지만, 경쟁이 치열해 지고 있는 CATV홈쇼핑에서의 CRM을 통한 고객데이터 분석의 필요성도 증가하고 있다. 집에서 채널변경을 통해 경쟁사의 상품정보를 쉽게 얻을 수 있고, 구매절차가 편리하기 때문에 고객이탈이 쉽다는 CATV홈쇼핑의 특징을 고객데이터 분석을 통해 고객유지 및 이탈방지를 전략적으로 관리할 수 있기 때문이다.

따라서 본 논문에서는 국내 CATV홈쇼핑 A사의 데이터에 기초하여 기존 고객의 이탈 방지 및 유지를 위한 CRM 방안에 대해서 연구해보고자 한다. 특히 기업이 보유하고 있는 고객의 속성 및 거래 데이터의 실증 분석을 통해 이탈고객과 유지고객간의 속성 및 거래 특성간에 어떠한 차이가 있는지를 알아낸다. 또한 고객이탈예측 모형을 개발해내어 이탈행동에 영향력이 높은 변수를 발견하고 이탈예측고객을 찾아내서 궁극적으로 고객 이탈로 이어지는 것을 막고 거래를 안정화시킬 수 있는 결정적인 거래 안정화 분기점을 찾아보고자 한다. 실증 분석 결과를 바탕으로 이탈가망고객을 대상으로 사전에 이탈을 방지하고 장기간 관계를 유지시켜서 충성도 높은 고객을 만들기 위한 CRM 전개 방안을 제시하며 더 나아가 TV홈쇼핑 업계 외에도 일반 기업이 실제로 고객의 유지 및 이탈 관리 전략을 수립하고자 할 때 본 논문의 연구 결과가 실질적으로 적용

될 수 있도록 도움을 주는 것이 본 연구의 의의이다.

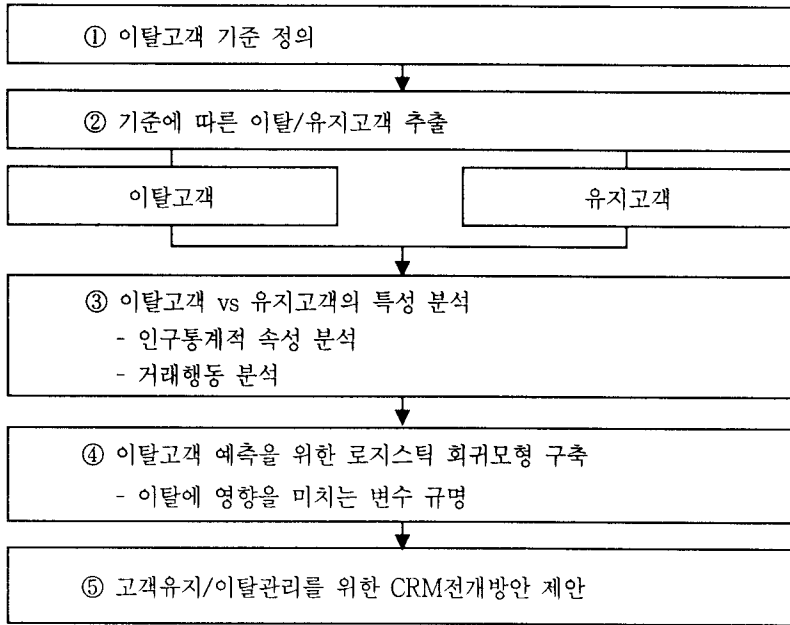
## II. 연구방법 및 실증분석 결과

### 2.1 연구방법

본 연구는 실제 고객 데이터의 실증 분석을 통하여 이탈고객과 유지고객간의 속성 및 거래 행동상의 차이를 분석하고, 고객 이탈에 영향을 미치는 변수를 발견하여 고객의 이탈을 결정적으로 방지할 수 있는 거래상의 안정화 분기점(key retention point)을 찾고자 하는 것이다. 이러한 분석 결과를 바탕으로 예측된 이탈가망고객을 대상으로, 고객 이탈 방지 및 유지를 위한 전략적 고객 관리 방안을 제시하고자 한다.

본 연구에서는 TV홈쇼핑, 인터넷 쇼핑몰과 카탈로그 상품판매를 동시에 운영하는 국내 A홈쇼핑사의 거래 고객 데이터를 자료로 하여 분석한다. 구체적으로는 이하에서 간략하게 서술한바와 같이 크게 세 가지를 그 내용으로 한다. <그림 2-1>의 분석단계에서 제시하듯이, 우선적으로 A홈쇼핑사의 이탈고객의 기준을 정의하고, 그 기준에 따라서 이탈고객 또는 유지고객으로 구분하여 고객 데이터를 추출한다. 그리고 이탈고객과 유지고객의 특성간의 차이점을 분석한다. 즉, 인구통계학적 속성에서는 어떤 차이가 있는지, 구매 거래 행동 패턴에서는 어떤 차이를 보이는지를 각 변수별로 이탈고객과 유지고객의 구성 비율을 분석함으로써 두 집단간의 차별적 특성을 구분해내고자 한다.

〈그림 2-1〉 분석 단계



다음으로는 로지스틱 회귀모형<sup>1)</sup>을 통하여 고객이탈에 영향을 주는 변수를 찾아내어 고객이탈행동이 주로 어떤 변수에 의해 좌우되는지 규명해보고, 이탈예측 모형 수식을 설정하고자 한다. 이러한 모형에 A홈쇼핑사의 개별고객을 대입해보면 이탈 예측 집단과 유지 예측 집단으로 분류해낼 수 있으며, 이탈 확률이 높은 고객을 찾아낼 수 있다. 고객 이탈에 영향을 주는 관련변수와 예측이탈확률은 주로 데이터마이닝 분야에서 다루어져 왔다. 그러나 본 연구는 마케팅 분야에서 CRM의 부각을 통한 이탈고객을 다루는데 초점을 맞추고 있으므로 스코어링 모델과 관련된 데이터마이닝은 문헌연구에서

다루지 않았다.

셋째로, 신규 진입 고객이 궁극적으로 이탈고객이 될 것인가 아니면 거래를 지속적으로 유지하는 안정화 고객이 될 것인가를 결정해주는 거래상의 분기점(turning point)을 발견하고자 한다. 다시 말해서, 이탈로 이어지지 않고 거래를 안정화시킬 수 있는 결정적인 거래 안정화 기준점(key retention point)을 찾아내고자 하는 것이다. 거래 안정화 기준점을 넘어섰을 경우에는 고객은 유지될 가능성이 커지는 것이고, 거래 안정화 기준점을 넘지 못했을 경우에는 이탈이 예상된다고 할 수 있다. 따라서 개별고객마다 거래상의 접점에서 거래 현황을 관리하여,

1) 고객이탈을 모형은 이탈확률이 높은 고객을 분류하여 그 고객층을 대상으로 데이터베이스 마케팅 프로그램을 실행하여 이탈율을 줄이는 것이 목적이며, 로지스틱 회귀모형을 통해서 고객이탈에 영향을 주는 관련변수 또는 통계가 가능한 변수를 밝혀낼 수 있고 개인별 이탈확률을 계산하여 예측이탈확률이 높은 고객들을 찾아내어 그들의 특징을 알아낼 수 있다(성용현 2001).

거래 안정화 포인트를 넘지 않아서 이탈이 예측되는 고객에게 이탈이 일어나기 전에 이탈을 막을 수 있는 고객관리 방안을 제시할 수 있을 것이다.

다시 요약하면, 본 연구는 실증 분석 결과를 통해서 A홈쇼핑사의 이탈고객의 인구통계적/거래행태적 특성을 파악할 수 있고, 이탈 가망 고객을 예측하여 이탈방지 및 유지를 위한 CRM 전개 방안을 제시한다.

분석에 사용된 데이터는 2001년 9월부터 2003

년 12월까지의 A홈쇼핑사의 TV홈쇼핑, 인터넷 쇼핑, 카탈로그에서 구매를 한번이상 일으킨 고객으로, 사용된 변수는 인구통계적 속성 정보에서는 성별, 연령대, 직업, 거주지역 변수, 구매와 관련된 거래정보에서는 매출, 이용 채널, 구매 상품군, 반품/취소와 관련된 변수이며, 최종 분석에 이용된 변수는 <표 2-1>에 제시되어 있다. 분석도구는 SAS 8.0 version을 사용했다.

<표 2-1> 분석에 사용된 고객 데이터 변수

| 변수 구분 | 변 수              |          | 설 명                                  |
|-------|------------------|----------|--------------------------------------|
| 속성 정보 | 성별*              | SEX      | 남성, 여성                               |
|       | 연령대*             | AGE      | 연령대분류(10세단위)                         |
|       | 직업*              | JOB      | 전문직, 회사원, 주부 등의 직업구분                 |
|       | 거주지역*            | ADDR     | 서울/수도권, 광역시, 기타시도(대분류)               |
| 거래 정보 | 구매의 최근성          | R        | 최근 구매정도에 따른 점수 (1~6점, 점수 높을수록 최근 구매) |
|       | 구매횟수             | CNT      | 실제 구매를 일으킨 횟수                        |
|       | 구매금액             | AMT      | 실제 구매한 상품의 금액                        |
|       | 주이용 채널*          | CH       | TV, 인터넷, 카탈로그 채널 중 가장 많이 이용한 채널      |
|       | 이용채널 수           | CH_N     | TV, 인터넷, 카탈로그 채널 중 구매시 이용해본 모든 채널의 수 |
|       | 최초 구매 채널*        | FST_CH   | 최초 구매시 이용한 채널                        |
|       | 주이용 상품군*         | GD       | 가장 많이 구매한 상품군(대분류)                   |
|       | 이용상품군 수          | GD_N     | 구매 상품이 속한 상품군의 개수                    |
|       | 최초구매상품군*         | FST_GD   | 첫 구매상품의 상품군                          |
|       | 반품/취소 횟수         | CNCL     | 구매상품 중 반품 또는 취소한 횟수                  |
|       | 최초 구매의 반품/취소 여부* | FST_CNCL | 첫 구매상품의 반품/취소여부                      |

\* 범주형 변수

## 2.2 실증분석 결과

### 2.2.1 이탈기준 정의

이탈고객 관리를 위해서는 가장 먼저 해야 할 일은 이탈의 기준을 설정하는 것이다. 즉, 고객의 어떤 상태를 이탈로 볼 것인가에 대한 명확한 정의를 내려야 한다. CRM업계에 따르면 고객이탈 캠페인 도구를 도입한 업체 등은 대부분 '이탈고객'에 대한 정확한 정의를 내리지 못하고 있어 캠페인 타겟이 불분명해지는 어려움을 겪고 있다(김응렬 2002). 하지만 은행 거래의 경우 고객들은 거래를 끊기 전에 신용카드 결제 계좌를 바꾼다거나 정기적금, 정기에금을 해약한다거나 또는 거래빈도가 준다거나 하는 독특한 행동변화를 보일 수 있다. 또는 유통업의 경우 주거래 품목의 구매 금액이 준다거나 또는 주거래 품목만 사고 나머지는 사지 않는 다거나 전체적으로 구매 빈도가 떨어진다는 등의 변화가 나타날 수 있다. 일반적으로 업계현황을 보면 금융기관에서는 고객 이탈의 기준을 증권계좌, 예금계좌, 보험, 신용카드 등을 계약을 해지하는 시점으로 삼고 있다(이건창, 권순재, 신경식 2002).

인터넷 쇼핑몰의 경우 이탈고객의 정의를 천용호, 임현택, 배준성(2002) 연구에서 디지털제품을 구매한 사용자가 재구매 의도를 보이지 않을 경우 이탈고객이라 간주하여 재구매 의도를 목표변수로 하고 성별, 응답자 직업, 가족관계, 연령의 사회 환경적 변수와 인터넷 구매경험, 최근 구매사이트 유형, 구매시 결제방법의 구매유형 변수를 예측변수로 설정하여 이에 대한 이탈고객의 속성을 파악하였다. 하지만 이

러한 방법은 단순한 재구매 의도 측정이 고객 이 이탈하였다고 보는데 한계가 있으므로 이탈의 정의를 규정하는데 어려움이 있다.

더욱이 홈쇼핑고객의 경우 이탈에 대한 기준을 명확하게 하기보다는 주관적인 판단에 의존할 수밖에 없는 것이 현실이다. 따라서 본 연구에서는 이탈고객의 정의를 실질적으로 홈쇼핑업계에서 이탈의 기준으로 삼는 고객의 구매거래 패턴을 바탕으로 기준으로 삼았다. 즉 새롭게 구매거래를 시작한 이후, 거래가 발생하지 않는 기간이 일정기간 이상 경과되었을 때 거래에서 이탈되었다고 보고, 이러한 고객을 이탈고객이라 정의하였다. 이와 같은 개념적 정의를 바탕으로, 조작적 정의를 하였다. 구매거래가 일어나지 않는 일정기간(=이탈기간)의 설정은 A홈쇼핑 고객의 구매거래 패턴을 분석하여 설정하였다. 일반적으로 고객은 이탈하기 전에는 휴면기간을 거치게 되고 휴면 후 재활성될 수도 있다. 따라서 휴면 후 재활성된 고객을 제외하기 위해서 휴면기간이 연속되었을 때를 이탈되었다고 정의하였다.

구체적으로 본 연구에서는 휴면기간을 정의하기 위해서 2회 이상 구매 고객 대상으로 평균구매주기 분석을 실시하였다. 평균구매주기가 지나도 구매가 일어나지 않을 때를 휴면화되었다고 볼 수 있으므로, 평균구매주기를 휴면기간으로 설정하고자 한다. 평균구매주기의 사분위수(quartile)가 <표 2-2>와 같으며, 전체 고객의 평균구매주기를 작은 값부터 배열하면 50%(Q2)에 해당하는 평균구매주기가 약 3개월(93일)이고, 75%(Q3)에 해당하는 평균구매주기가 약 6개월(174일)이다.

평균구매주기를 휴면기간으로 설정하기로 했

〈표 2-2〉 A홈쇼핑사 고객의 평균구매주기의 사분위수(Quartile)

| Min | Q1(제1사분위수, 25%) | Q2(제2사분위수, 중앙값) | Q3(제3사분위수, 75%) | Max  |
|-----|-----------------|-----------------|-----------------|------|
| 1일  | 47일             | 93일             | 174일            | 896일 |

〈표 2-3〉 A홈쇼핑사의 유지/휴면/이탈고객 정의

| 구분   | 정의                           |
|------|------------------------------|
| 유지고객 | 1년간 매 6개월마다 1회이상 구매실적이 있는 고객 |
| 휴면고객 | 6개월간 구매실적이 없는 고객(휴면기간: 6개월)  |
| 이탈고객 | 1년간 구매실적이 없는 고객(이탈기간: 1년)    |

으므로, 만일 휴면기간을 3개월로 설정하고 2003년 12월 기준으로 A홈쇼핑사 고객에 적용시켜 보면, 34%의 고객만이 거래가 유지되는 고객이고 휴면이거나 이탈된 것으로 보이는 고객이 나머지 66%로 비교적 많은 편이다. 이것은 홈쇼핑사의 활성고객이 1/3에 불과하다는 것을 뜻하고 고객 구성이 비정상적이므로 휴면기간 3개월은 홈쇼핑업계나 본 연구자에게 짧다고 판단된다. 반면에 휴면기간을 6개월로 설정한다면, 2003년 12월 기준으로 거래가 유지된 고객이 44%, 휴면 또는 이탈된 고객이 56%로 그 비율측면에서 적정한 고객 구성이라고 보이므로 본 연구에서는 휴면기간은 '6개월'로 설정하였다. 따라서 이탈기간은 휴면기간이 연속되는 '1년'으로 정의하여, A홈쇼핑 고객이 일정 월을 기준으로 이전 6개월간 구매가 없었으면 휴면고객으로, 1년간 구매가 없었으면 이탈고객으로 정의하고, 1년간 휴면기간 없이 매 6개월마다 1회 이상 구매가 있었던 고객은 유지고객이라고 정의하였다.

## 2.2.2 이탈고객과 유지고객의 특성 분석

기존 고객이탈의 영향요인에 관한 연구들은 주로 다양한 데이터 마이닝기법을 이용하여 금융산업에서 다루어져왔다. 그러나 대부분의 연구들이 CRM을 기반으로하여 고객이탈의 영향요인을 밝히고 이를 전략적으로 다루는 것이 아니라 고객을 정확하게 세분화할 수 있는 방법 즉, 고객세분화(이탈/유지고객)에 초점을 두고 여러 가지 데이터마이닝기법(이건창, 권순재, 신경식 2001; 천용호, 임현택, 배준성 2002; 이용구, 이현정 2001)을 적용하였다. 이진창, 권순재, 신경식(2001)은 C은행의 실제 고객자료를 통하여 클레멘타인(Clementine 5.2)를 활용하여 고객을 세분화 하였다. 그 결과 고객이 가계성 예금을 해지하고 다른 은행으로 이탈하는 집단의 특성이 통장의 개수와 자동이체건수, 일정액 이상의 카드사용액 및 월평균 잔액 등에 영향을 받은 것으로 분석되었다. 특히 이들은 인공지능망과 클레멘타인 분석의 결과를 비교함으로써 고객 유지, 해지에 영향을 미치는 변수가 유사하다고 주장하였다. 의사결정나무(decision

tree analysis)를 이용하여 인터넷쇼핑몰의 이탈 고객을 분석한 천용호, 임현택, 배준성(2002)의 연구에서는 남성이 뚜렷한 직업을 가지고 있으면서 신용카드나 자동이체로 결제한 경우에 이탈고객들이 보이는 특성으로 나타났다. 이용구, 이현정(2001) 연구에서도 보험회사를 대상으로 클레멘타인 분석을 통해 이탈고객을 세분화하였다.

그러나 마케팅 분야에서는 고객데이터 분석을 통한 이탈고객에 대한 관련된 연구가 CATV홈쇼핑업계를 포함하여 대부분 영역에서 다루어지지 않았다는데 문제점이 있다. 따라서 본 연구에서는 이탈고객과 유지고객간의 특성 분석을 인구통계적 속성 및 구매거래 행동에서 서로 다른 특성을 보일 것이라는 가정(Imhoff, Gentry 2000; Raaen 2000)하에, CATV 홈쇼핑을 이용하는 고객이 어떤 변수에서 무슨 차이를 나타내는지 밝혔다. 분석대상 고객은 2001년 12월 또는 2002년 3월 또는 2002년 6월에 각각 최초 구매를 일으킨 고객 중에서 최초구매월로부터 6개월이 지난 시점부터 이후 1년간 매 6개월마다 거래가 유지된 '유지고객'과 거래가 발생하지 않은 '이탈고객'을 각각 5만명씩 랜덤추출법(random sampling)으로 추출하였다.

유지고객과 이탈고객의 거래상의 특성 차이는 추출된 '유지고객'과 '이탈고객'의 최초구매월로부터 6개월간의 구매거래 데이터를 비교 분석하였다. 최초구매월을 특정월로 한정하지 않고, 계절별로 다른 월의 고객을 골고루 추출함으로써, 구매시기로 인한 거래 행태의 특이성이 반영되지 않도록 했다.

#### 1) 인구통계적 속성변수 분석

이탈고객과 유지고객간의 인구통계적 속성 차이에 대한 분석은 성별, 연령, 직업, 거주지역 변수에 대해서 실시하였다. 범주형 변수인 성별, 직업, 거주지역은 카이제곱검정, 연속형 변수인 연령은 T검정 한 결과, 유의수준 1%에서 이탈고객집단과 유지고객집단은 성별, 직업, 거주지역 및 연령변수에서 서로 다른 특성 차이를 보이는 것으로 나타났다( $p < .01$ ). 이에 대한 자세한 결과는 <표 2-5>에 제시되어 있다.

먼저 성별의 경우 유지고객과 이탈고객이 구분변수(남성, 여성)에서 각각 비율 차이가 있는지 여부를 검정하기 위해서 표준정규분포로 유의성 검정한 결과, 유의수준 1%에서 유지고객과 이탈고객의 비율에 차이가 나는 것으로 나타났다( $p < .01$ ). 즉, 남성은 이탈고객 비율이

<표 2-4> 분석 표본집단 선정

| 구분      | 내역  |         |
|---------|---|---------|
| 분석 표본집단 | 2001년 12월 또는 2002년 3월 또는 2002년 6월에 A홈쇼핑에서 최초로 구매를 일으킨 신규고객 중 최초구매월로부터 6개월이 지난 시점부터, 이후 1년간 매 6개월마다 거래가 유지된 '유지고객'과 거래가 없었던 '이탈고객' |         |
| 추출법     | 랜덤추출법(random sampling)  |         |
| 대상 모수   | 유지고객  | 50,000명 |
|         | 이탈고객  | 50,000명 |



69.9%로 더 높고, 여성은 유지고객 비율이 54.7%로 더 높아서 남성이 여성보다 이탈될 확률이 더 높고, 여성은 유지될 확률이 더 높다고 할 수 있다.

연령에 대해서 T검정을 실시한 결과, 유의수준 1%에서 유지고객집단과 이탈고객집단의 연령의 평균이 다르다고 나타났다(p-값<.01). 즉 유지고객(평균 41세)의 연령이 이탈고객(평균 39세)의 연령보다 높은 편이며, 특히 20대는 70.1%의 고객이 이탈되었으므로 이탈 확률이 매우 높고, 40대 이상은 유지될 확률이 높다. 성별과 연령을 동시에 분석해보면, 20대 남성의

이탈고객 비율(87.6%)이 가장 높고, 다음으로 30대 남성(74.2%)의 이탈고객 비율이 높다. 유지고객 비율은 40대 여성(62.1%)이 가장 높으며, 다음으로 50대 여성(59.6%), 30대 여성(54.1%)의 순으로 유지고객 비율이 높다.

직업의 경우는 직업의 구분변수별로 유지고객 비율과 이탈고객 비율이 차이가 나는 것으로 검정되었지만(p-값<.01), 각 구분변수별 비율 차이가 그다지 크지 않은 편이다. 다만, 학생과 회사원은 이탈율이 높고, 주부는 유지율이 높은 편이라고 할 수 있다. 이는 상기의 성별, 연령대 변수의 분석 결과를 고려해볼 때 20~30대

〈표 2-5〉 인구통계적 속성변수 분석결과

| 구분    |       | 비율(%)* |       | 유의성 검정결과**          |   |
|-------|-------|--------|-------|---------------------|---|
|       |       | 유지고객   | 이탈고객  | 구분 변수               | 변수(직업)  |
| 성별    | 남성    | 30.1%  | 69.9% | z=59.7 p-value=0.00 | x <sup>2</sup> =3,712.5<br>df=1, p-value=0.00 |
|       | 여성    | 54.7%  | 45.3% | z=26.7 p-value=0.00 |   |
| 연령    | 20대   | 29.9%  | 70.1% | z=48.5 p-value=0.00 | t=34.6 p-value=0.00                           |
|       | 30대   | 49.3%  | 50.7% | z=2.7 p-value=0.00  |   |
|       | 40대   | 57.3%  | 42.7% | z=27.3 p-value=0.00 |   |
|       | 50대이상 | 51.8%  | 48.2% | z=4.4 p-value=0.00  |   |
| 직업    | 전문직   | 46.3%  | 53.7% | z=2.6 p-value=0.00  | x <sup>2</sup> =434.2<br>df=6, p-value=0.00   |
|       | 공무원   | 45.5%  | 54.5% | z=2.4 p-value=0.00  |   |
|       | 회사원   | 40.2%  | 59.8% | z=10.2 p-value=0.00 |   |
|       | 자영업   | 44.4%  | 55.6% | z=5.2 p-value=0.00  |   |
|       | 주부    | 58.6%  | 41.4% | z=11.2 p-value=0.00 |   |
|       | 학생    | 22.8%  | 77.2% | z=11.9 p-value=0.00 |   |
|       | 기타    | 49.7%  | 50.3% | z=1.5 p-value=0.06  |   |
| 거주 지역 | 서울/경기 | 47.3%  | 52.7% | z=8.7 p-value=0.00  | x <sup>2</sup> =714.9<br>df=2, p-value=0.00   |
|       | 광역시   | 51.7%  | 48.3% | z=5.1 p-value=0.00  |   |
|       | 기타    | 49.2%  | 50.8% | z=2.0 p-value=0.02  |   |

\* 비율(%): 전체 표본고객 중 구분변수에 속하는 고객 중에서 유지고객과 이탈고객의 구성비율

\*\* x<sup>2</sup>: chi-square 검정통계량값, z: 표준정규분포의 검정통계량값, p-value<.01

(학생, 회사원)가 이탈율이 높고, 40대 이상의 여성(주부)이 유지율이 높은 것과 동일한 맥락의 결과임을 알 수 있다.

마지막으로 거주지역의 경우 서울/수도권이 지방보다는 이탈고객비율이 약간 높은편인데, 이는 홈쇼핑 외에도 백화점이나 대형할인점 등의 쇼핑 대체제가 가까이 많기 때문으로 보인다.

## 2) 구매거래 관련 변수 분석

### ① 구매횟수, 구매 최근성, 구매금액

구매거래와 관련된 변수들은 연속형 변수인 구매횟수, 구매금액, 구매의 최근성, 이용상품군 종류수, 이용채널 수, 반품/취소횟수는 T검정으로, 범주형 변수인 주이용상품군, 주이용채널은

카이제곱검정으로 유의성 검정을 했으며, 그 결과 모든 변수에서 유지고객과 이탈고객의 거래 특성이 통계적으로 유의한 차이가 나는 것으로 나타났다(p-value<.01). 분석에 사용된 데이터는 <표 2-4>의 표본고객 10만명이 최초 구매일로부터 6개월간 일으킨 구매거래 데이터(단, 정상 거래 데이터에 한함)를 분석한 것이고, 구매거래를 일으킨 표본고객이 이후 1년간 구매실적에 따라서 유지 또는 이탈고객으로 분류되었을 때의 구성 비율의 크기 비교를 통해서 분석 결과를 해석했다.

구매횟수가 많을수록 유지될 확률이 높아지고 있다. 6개월간 1회만 구매한 고객은 이탈고객 비율이 유지고객보다 약 2배 이상 높으나, 2회 이상 구매를 일으킨 고객은 대체로 유지고객

<표 2-6> 구매횟수, 구매의 최근성

| 구분       |      | 비율(%) |       | 유의성 검정결과             |                         |
|----------|------|-------|-------|----------------------|-------------------------|
|          |      | 유지고객  | 이탈고객  | 구분 변수                | 변수                      |
| 구매 횟수*   | 1회   | 29.1% | 70.9% | z=103.1 p-value=0.00 | t=121.8<br>p-value=0.00 |
|          | 2회   | 56.0% | 44.0% | z=17.0 p-value=0.00  |                         |
|          | 3회   | 72.2% | 27.8% | z=48.9 p-value=0.00  |                         |
|          | 4회   | 82.3% | 17.7% | z=61.1 p-value=0.00  |                         |
|          | 5회   | 87.5% | 12.5% | z=63.4 p-value=0.00  |                         |
|          | 6회   | 89.7% | 10.3% | z=58.4 p-value=0.00  |                         |
|          | 7회이상 | 94.4% | 5.6%  | z=129.6 p-value=0.00 |                         |
| 구매 최근성** | 1점   | 27.7% | 72.3% | z=111.2 p-value=0.00 | t=128.7<br>p-value=0.00 |
|          | 2점   | 53.1% | 46.9% | z=5.1 p-value=0.00   |                         |
|          | 3점   | 61.7% | 38.3% | z=19.1 p-value=0.00  |                         |
|          | 4점   | 69.1% | 30.9% | z=35.7 p-value=0.00  |                         |
|          | 5점   | 75.5% | 24.5% | z=59.4 p-value=0.00  |                         |
|          | 6점   | 83.1% | 16.9% | z=104.6 p-value=0.00 |                         |

\* 유지고객: 이탈고객 평균구매횟수 = 2.5회: 1.4회(6개월 기준)

\*\* 유지고객: 이탈고객 평균 점수 = 3.2점: 1.7점(6개월 기준)

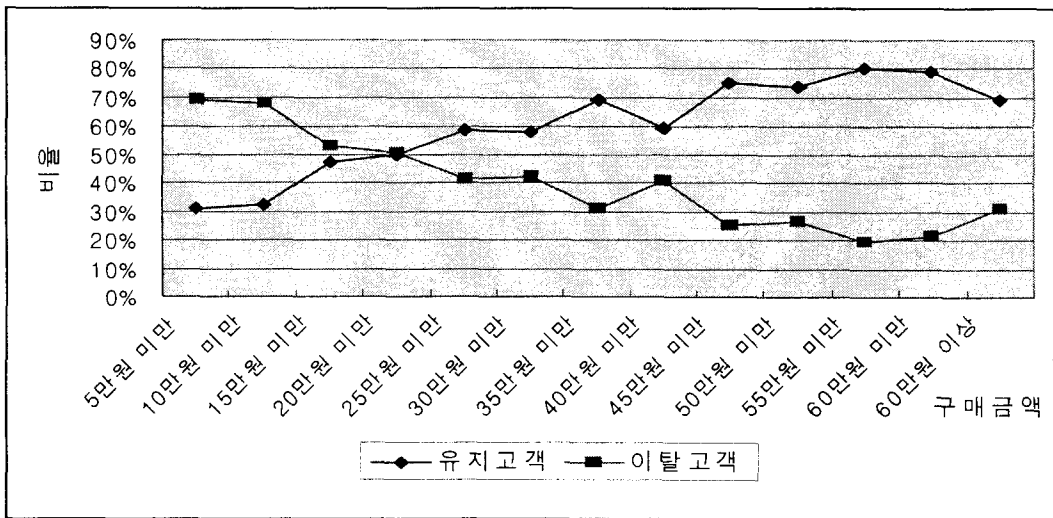
비율이 월등히 높다. 또한, 구매횟수가 1회에서 2회로 늘어났을 때, 유지고객 비율의 증가폭이 가장 크고 이탈고객 비율은 약 1/2로 감소하므로, 1회만 구매한 고객을 1회 이상 추가 구매가 일어나도록 유도하는 것이 고객 유지율을 높이기 위한 방법 중의 하나임을 알 수 있다. 뿐만 아니라 유지고객의 평균 구매횟수가 2.5회, 이탈고객의 평균 구매횟수가 1.4회이므로 구매횟수가 최소한 2회는 되어야 고객유지의 비율이 높아짐을 알 수 있다. 따라서 구매 거래의 안정화를 위한 거래상의 분기점은 구매횟수 1회를 넘어서서 구매횟수 2회 시점이라고 볼 수 있다. A홈쇼핑사의 경우는 6개월간 1회 거래 고객이 표본고객의 과반수가 넘는 56%를 차지하고 있기 때문에, 고객 유지 차원에서 이 고객들의 관리가 필수적이라 판단된다.

구매의 최근성은 구매가 얼마나 최근에 일어

났는가에 따라 부여한 점수(1~6점)<sup>2)</sup>로, 신규고객이 최초구매월로부터 6개월사이에 최종구매를 일으킨 월이 최근월일수록 점수가 높다. 구매의 최근성 점수가 클수록 유지고객 비율이 높아지므로, 최근에 구매한 고객일수록 유지될 확률은 높아지고 이탈될 확률은 낮아진다. 따라서 최종 구매 시점에서 시간이 어느 정도 경과된 고객은 이탈 가능성이 점차 높아지므로, 빠른 시일 내에 방문을 촉진하여 재구매를 유도하는 것이 필요하겠다.

〈그림 2-2〉는 6개월간 구매금액이 많을수록 유지고객 비율은 점차 증대하고 이탈고객 비율은 점차 감소하는 추세를 보여주고 있다. 유지고객과 이탈고객의 6개월간 평균구매금액은 각각 342,026원과 202,511원으로 유지고객이 139,515원이 더 높다. 구매횟수와 더불어 구매금액도 고객의 유지, 이탈 여부에 영향을 미치므로 구

〈그림 2-2〉 구매금액별 유지고객/이탈고객의 비율



2) A홈쇼핑사 내부 점수

매금액을 높일 수 있는 방안이 고려되어야 할 것이다.

② 주이용 상품군, 이용상품군 종류수

표본고객 중 구매횟수가 2회 이상인 고객 기준으로 주이용상품군별 유지, 이탈고객 비율을 분석한 결과, 유지고객 비율이 70%이상을 차지하는 상품군은 의류(78.6%)>보석/장신구(75.2%)>식품(72.6%)>패션잡화(70.0%)의 순이었다. 유

지고객 비율이 높은 상품군이 주로 여성과 관련된 것은 여성의 유지비율이 높다는 상기 결과와 연관지어 해석할 수 있다. 이 외에 침구/침장, 주방용품도 유지고객 비율이 비교적 높았으나, 전기전자, 스포츠/레저용품 구매자는 이탈고객 비율이 타 상품군에 비해서 상대적으로 높았다.

이용상품군 종류수가 많아질수록 유지고객 비율이 높아지고 있다. 즉, 여러 종류의 상품군에서 골고루 상품을 구매하는 고객이 유지율이

〈표 2-7〉 주이용상품군, 이용상품군 종류수

| 구분                |        | 비율(%) |                      | 유의성 검정결과             |  |
|-------------------|--------|-------|----------------------|----------------------|--|
|                   |        | 유지고객  | 이탈고객                 | 구분 변수                | 변수   |
| 주<br>이용<br>상품군    | 가구     | 59.9% | 40.1%                | z=4.9, p-value=0.00  | x <sup>2</sup> =4,311.4<br>df=14<br>p-value=0.00 |
|                   | 문화/취미  | 60.2% | 39.8%                | z=2.7, p-value=0.00  |  |
|                   | 보석/장신구 | 75.2% | 24.8%                | z=25.9, p-value=0.00 |  |
|                   | 생활 잡화  | 64.3% | 35.7%                | z=13.5, p-value=0.00 |  |
|                   | 속옷     | 60.7% | 39.3%                | z=9.7, p-value=0.00  |  |
|                   | 스포츠/레저 | 59.2% | 40.8%                | z=4.4, p-value=0.00  |  |
|                   | 식품     | 72.6% | 27.4%                | z=34.6, p-value=0.00 |  |
|                   | 유아동용품  | 62.7% | 37.3%                | z=6.0, p-value=0.00  |  |
|                   | 의류     | 78.6% | 21.4%                | z=89.2, p-value=0.00 |  |
|                   | 이미용품   | 63.0% | 37.0%                | z=12.9, p-value=0.00 |  |
|                   | 자동차    | 53.9% | 46.1%                | z=1.0, p-value=0.18  |  |
|                   | 전기전자   | 55.6% | 44.4%                | z=5.9, p-value=0.00  |  |
|                   | 주방     | 64.0% | 36.0%                | z=16.4, p-value=0.00 |  |
|                   | 침구/침장  | 65.5% | 34.5%                | z=21.6, p-value=0.00 |  |
| 패션잡화              | 70.0%  | 30.0% | z=21.4, p-value=0.00 |                      |  |
| 이용<br>상품군<br>종류수* | 1종류    | 32.3% | 67.7%                | z=90.4 p-value=0.00  | t=126.3<br>p-value=0.01                          |
|                   | 2종류    | 63.7% | 36.3%                | z=41.3 p-value=0.00  |  |
|                   | 3종류    | 80.6% | 19.4%                | z=73.8 p-value=0.00  |  |
|                   | 4종류    | 88.4% | 11.6%                | z=75.4 p-value=0.00  |  |
|                   | 5종류이상  | 93.8% | 6.2%                 | z=102.7 p-value=0.00 |  |

\* 유지고객: 이탈고객 평균 종류수=2.2종류: 1.3종류(6개월 기준)

높다고 할 수 있다. 이용상품군 종류가 한 가지 인 경우는 이탈고객 비율이 높고, 2가지 이상이 될수록 유지 가능성이 크므로 고객들에게 다양한 상품군의 상품을 구매할 수 있도록 교차 판매를 유도하는 것이 필요해 보인다.

### ③ 이용채널, 이용채널 수

신규고객이 최초 구매시 이용한 채널 여부가 이후 거래의 유지와 이탈에 영향을 미치는 것으로 나타났다. A홈쇼핑사는 TV홈쇼핑과 인터넷쇼핑몰, 카탈로그를 동시에 운영하고 있는데, 첫 구매시, TV홈쇼핑 채널을 이용할 때의 유지고객 비율이 인터넷쇼핑이나 카탈로그를 통해 구매시보다 높았다. 특히, TV홈쇼핑의 ARS

채널<sup>3)</sup>로 구매한 고객이 상담원 전화로 구매한 고객보다 유지고객 비율이 높았다. 이는 ARS 채널로 구매시, 일정을 할인혜택을 주기 때문으로 보인다. 카탈로그는 이탈고객 비율이 70.2%에 달해 채널경쟁력이 크게 떨어진다.

2회 이상 구매자의 주 이용채널에 따른 유지 및 이탈고객 비율을 분석해본 결과, 최초구매채널의 영향을 그대로 받고 있다. 2회이상 구매자 대상의 분석이므로 모든 채널의 유지고객 비율이 이탈고객 비율보다 높긴 하나, 최초구매채널의 경우와 동일하게 TV홈쇼핑의 ARS채널을 주로 이용하는 구매자의 유지고객의 비율이 상대적으로 가장 높고, 인터넷, 카탈로그의 유지고객 비율이 상대적으로 낮음을 알 수 있다.

〈표 2-8〉 주이용채널, 이용채널수

| 구분                    |         | 비율(%) |       | 유의성 검정결과              |  |
|-----------------------|---------|-------|-------|-----------------------|--|
|                       |         | 유지고객  | 이탈고객  | 개별 변수                 | 변수                                       |
| 주 이용 채널<br>[최초 구매채널]  | TV(ARS) | 53.9% | 46.1% | $z=14.0$ p-value=0.00 | $\chi^2 = 772.1$<br>df=3<br>p-value=0.00 |
|                       | TV(전화)  | 49.3% | 50.7% | $z=3.6$ p-value=0.00  |  |
|                       | 인터넷     | 32.5% | 67.5% | $z=22.4$ p-value=0.00 |  |
|                       | 카탈로그    | 29.8% | 70.2% | $z=12.2$ p-value=0.00 |  |
| 주 이용 채널<br>[2회이상 구매자] | TV(ARS) | 77.0% | 23.0% | $z=98.0$ p-value=0.00 | $\chi^2 = 3,825$<br>df=3<br>p-value=0.00 |
|                       | TV(전화)  | 64.1% | 35.9% | $z=36.9$ p-value=0.00 |  |
|                       | 인터넷     | 61.7% | 38.3% | $z=9.9$ p-value=0.00  |  |
|                       | 카탈로그    | 58.1% | 41.9% | $z=9.4$ p-value=0.00  |  |
| 이용 채널수*               | 1개      | 38.9% | 61.1% | $z=60.1$ p-value=0.00 | $t_{값}=100.7$<br>p-value=0.00            |
|                       | 2개      | 73.9% | 26.1% | $z=81.5$ p-value=0.00 |  |
|                       | 3개      | 86.1% | 13.9% | $z=50.7$ p-value=0.00 |  |
|                       | 4개      | 90.0% | 10.0% | $z=10.3$ p-value=0.00 |  |

\* 유지고객: 이탈고객 평균 이용채널수 = 1.5개: 1.1개(6개월 기준)

3) TV홈쇼핑에서 상품을 주문할 때 사용하는 자동주문전화 채널

④ 반품/취소 횟수

반품/취소 횟수대별 고객 비율 분석에 들어가기 전에, 반품/취소 횟수가 많을수록 홈쇼핑사의 상품이나 서비스에 대한 실망으로 인해 이탈고객 비율이 높아질 것이라고 예상했었으나, 실제로 결과는 반품/취소 횟수가 늘어날수록 오히려 유지고객 비율이 높아졌다. 이는 반품/취소를 하는 고객일수록 유지도 잘 된다는 의미인데, 유지되는 고객일수록 상품을 구매하는 횟수가 많으므로 그 만큼 상품을 반품/취소할 경우의 수도 많아지기 때문에 이와 같은 결과가 나온 것이라 생각된다.

3) 고객이탈확률 예측을 위한 로지스틱 회귀 모형  
고객이탈에 영향을 주는 변수 및 그 영향 정도

도를 밝혀내는 동시에, 개별 고객의 이탈예측확률을 구하여, 이탈확률이 높은 고객을 선별하고 관리하기 위해서 로지스틱 회귀모형을 통한 이탈고객예측 모형을 구축해보겠다. 분석 표본집단은 <표 2-4>에서 추출된 고객데이터와 동일하며, 이분화 종속변수 및 모형에 사용될 독립변수는 아래의 <표 2-10>과 같다.

① 최량부분선택법에 의한 최종모형 결정  
로지스틱 회귀모형을 설정 할 때 일반적으로 어떤 사건이 발생할 확률에 영향을 미칠 수 있는 가능한 모든 독립변수를 고려하게 된다. 그 중에서 특정한 사건이 발생할 확률에 유의한 영향을 미치는 변수는 선택하고 그렇지 못한 변수는 제외시켜야 한다. 본 연구에서는 설정된 모형에 포함된 많은 변수들 중에서 유의한 변

<표 2-9> 반품/취소 횟수

| 구분    | 비율(%) |       | 유의성 검정결과              |                         |
|-------|-------|-------|-----------------------|-------------------------|
|       | 유지고객  | 이탈고객  | 개별 변수                 | 변수                      |
| 0회    | 36.8% | 63.2% | $z=67.8$ p-value=0.00 | t값=71.7<br>p-value=0.00 |
| 1회    | 63.0% | 37.0% | $z=39.1$ p-value=0.00 |                         |
| 2회    | 74.0% | 26.0% | $z=49.3$ p-value=0.00 |                         |
| 3회    | 81.0% | 19.0% | $z=49.0$ p-value=0.00 |                         |
| 4회    | 83.5% | 16.5% | $z=39.9$ p-value=0.00 |                         |
| 5회 이상 | 87.9% | 12.1% | $z=74.2$ p-value=0.00 |                         |

\* 유지고객: 이탈고객 평균 반품/취소 횟수 = 1.4회; 0.4회(6개월 기준)

<표 2-10> 로지스틱 회귀모형의 종속변수/독립변수

|         |  |
|---------|--|
| 종속변수(Y) | 고객이탈=1, 고객유지=0   |
| 독립변수(X) | 성별(SEX), 연령대(AGE), 직업(JOB), 거주지역(ADDR), 구매횟수(CNT), 구매금액(AMT), 구매의 최근성(R), 주이용채널(CH), 이용채널수(CH_N), 최초구매상품군(FST_GD), 주이용상품군(GD), 이용상품군 종류수(GD_N) |

수를 선택하기 위한 변수선택방법으로 '최량부분선택법(best subject selection)<sup>4)</sup>'을 사용했고, 이를 통해서 적합한 최종모형을 결정했다. 변수선택과정에서 서로 다른 모형의 비교를 위해서 스코어 카이제곱 통계량을 사용했으며, 모형의 적합도를 평가하기 위해서 통계량 Pseudo-R<sup>2</sup> 과 정분류율을, 모형의 예측력 정확도를 평가하기 위해서는 c통계량을 사용했다.

최량부분선택법에 의해 모형에 포함될 독립변수를 선택하기 위해서 독립변수의 개수에 따라 가능한 모든 모형을 적합시킨 후, 독립변수의 개수별 스코어 카이제곱값이 가장 큰 하나의 모형만을 출력했다. 출력된 결과는 <표 2-11>과 같으며, 모형에 포함된 변수의 개수가 증가함에 따라 스코어 카이제곱값은 증가하고 있고, 정분류율, Pseudo-R<sup>2</sup>, c통계량의 수치도 커지고 있

<표 2-11> Regression Models Selected by Score Criterion

| 변수개수 | Score Chi-Square | Variables Included in Model   | 정분류율         | c통계량         | Pseudo-R <sup>2</sup> |
|------|------------------|---|--------------|--------------|-----------------------|
| 1    | 19,210           | R   | 68.8%        | 0.687        | 0.1861                |
| 2    | 20,743           | SEX_M R   | 69.4%        | 0.718        | 0.2145                |
| 3    | 22,082           | SEX_M R GD_N  | 72.0%        | 0.774        | 0.3069                |
| 4    | 23,249           | SEX_M AGE_20 R GD_N   | 72.6%        | 0.783        | 0.3233                |
| 5    | 23,865           | SEX_M AGE_20 R FAVOR_GD_10 GD_N   | 72.8%        | 0.790        | 0.3326                |
| 6    | <b>24,339</b>    | SEX_M AGE_20 R FAVOR_GD_10 FAVOR_CH_01 GD_N   | <b>72.9%</b> | <b>0.794</b> | 0.3390                |
| 7    | <b>24,513</b>    | SEX_M AGE_20 AGE_30 R FAVOR_GD_10 FAVOR_CH_01 GD_N                                  | <b>72.9%</b> | <b>0.795</b> | 0.3416                |
| 8    | <b>24,596</b>    | SEX_M AGE_20 AGE_30 R FAVOR_GD_08 FAVOR_GD_10 FAVOR_CH_01 GD_N                      | <b>73.0%</b> | <b>0.796</b> | 0.3427                |
| 9    | <b>24,699</b>    | SEX_M AGE_20 AGE_30 R AMT_LOG FAVOR_GD_08 FAVOR_GD_10 FAVOR_CH_01 GD_N              | <b>73.0%</b> | <b>0.796</b> | 0.3437                |
| 10   | <b>24,783</b>    | SEX_M AGE_20 AGE_30 R AMT_LOG FST_GOODS_13 FAVOR_GD_08 FAVOR_GD_10 FAVOR_CH_01 GD_N | <b>73.1%</b> | <b>0.796</b> | 0.3447                |

(변수 10개 이상의 경우 생략함)

4) 회귀모형에서 독립변수를 어떠한 방법으로 선택하는가에 따라 변수추가법, 변수제거법, 최량부분 선택법이 있는데, 최량부분선택법은 변수가 없는 상태에서 변수추가법을 이용하여 한개 변수를 추가한 후, 변수 제거법에 의하여 제거될 변수가 있는지를 조사하는 방법을 말한다(이군희 2001).

다. 정분류율, Pseudo-R<sup>2</sup>, c통계량은 그 수치가 클수록 모형의 적합도와 정확도는 높은 것을 의미한다.

그런데 독립변수의 수가 6개 이상부터는 스코어 카이제곱값의 증가가 둔화되어서 차이가 거의 없게 나타나고 있고, 정분류율과 c통계량도 변수 개수가 6개 이상부터는 증가율이 둔화되면서 거의 같게 나타남에 따라, 6개 이상의 변수를 포함하더라도 모형의 적합도와 예측력의 차이는 없다고 판단된다. 따라서 가장 적합한 최종모형으로 6개 독립변수(SEX\_M, AGE\_20, R, FAVOR\_GD\_10, FAVOR\_CH\_01, GD\_N)를 포함하는 것으로 결정하였다.

## ② 로지스틱 회귀 모형 수식

최종모형에 포함되는 6개 독립변수와 각 변수의 계수추정치 및 계수추정치에 대한 카이제곱검정 결과는 아래의 <표 2-12>와 같다.

6개의 변수는 성별(남성 여부), 연령(20대 여부), 구매의 최근성, 주이용상품군(의류여부), 주이용채널(ARS여부), 이용상품군 종류수로 결정됐다. 유의수준 1%에서 계수추정치의 유의성을 검정한 결과, 통계적으로 유의하다(p-value <.01).

6개의 변수 중에서 이탈율을 높이는데 가장 영향력이 큰 변수는 '연령 20대'이고 다음으로 '성별 남성'으로 나타났다. 변수의 승비의 값이 클수록 이탈율을 높이는데 영향도가 큰 변수이다. 20대고객은 타 연령대 고객보다 이탈할 확률이 2.477배(승비)가 높고, 남성은 여성보다 2.233배(승비)가 높다.

이 외에 승비의 값이 큰 순서대로 구매의 최근성>주이용채널(ARS여부)>이용상품군 종류수>주이용상품군(의류여부)의 순으로 이탈확률을 낮추는데 영향을 미치는 것으로 나타났다. 연령과 성별은 계수추정치가 정(positive)이므로, 20대이거나 남성이면 이탈확률이 커지고, 나머지 변수는 계수추정치가 부(negative)이므로, 구매의 최근성 점수가 높을수록, 이용상품군 종류수가 많아질수록, 주이용채널이 ARS이고, 주이용상품군이 의류일 때 이탈확률이 적어지고 유지확률이 높아진다.

고객이탈확률 예측을 위한 로지스틱 회귀모형에서 개별고객의 이탈확률은 계수추정치를 사용한 모형수식으로 계산된다. 모형수식은 다음과 같이 표현된다(강병서, 김계수 2001).

<표 2-12> 변수의 계수추정치와 검정 결과

| Variables Included in Model | 계수추정치   | 승비    | x <sup>2</sup> | df | Pr>Chisq |
|-----------------------------|---------|-------|----------------|----|----------|
| Intercept                   | 2.5143  |       | 11185.8        | 1  | <.0001   |
| 성별_남성여부(SEX_M)              | 0.8032  | 2.233 | 1568.0         | 1  | <.0001   |
| 연령_20대여부(AGE_20)            | 0.9072  | 2.477 | 1334.8         | 1  | <.0001   |
| 구매의 최근성(R)                  | -0.3111 | 0.733 | 3450.3         | 1  | <.0001   |
| 주이용상품군=의류여부(FAVOR_GD_10)    | -0.5063 | 0.603 | 881.4          | 1  | <.0001   |
| 주이용채널=ARS여부(FAVOR_CH_01)    | -0.3839 | 0.681 | 603.8          | 1  | <.0001   |
| 이용상품군 종류수(GD_N)             | -0.4986 | 0.607 | 1826.7         | 1  | <.0001   |



$$P_i = \frac{\exp(Q_i)}{1+\exp(Q_i)}$$

$$Q_i = 2.5143 + (0.8032 \times \text{SEX\_M}_i) + (0.9072 \times \text{AGE\_20}_i) - (0.5063 \times \text{FAVOR\_GD\_01}_i) - (0.3111 \times \text{R}_i) - (0.3839 \times \text{FAVOR\_CH\_01}_i) - (0.4986 \times \text{GD\_N}_i)$$

$P_i$ 는 고객 $i$ 의 이탈확률이며, 고객 $i$ 의 독립변수 ( $X_i$ )의 값(범주형은 1 or 0, 연속형은 해당값)을 수식에 대입했을 때,  $P_i > 0.5$ 이면 이탈예측 고객이고,  $P_i < 0.5$ 이면 유지예측 고객이다(Prob level=0.5). 이렇게 모형을 통해서 개별고객마다 이탈예측집단 또는 유지예측집단으로 분류할 수 있으며, 이탈확률값( $P_i$ )에 따라 이탈 관리가 요구되는 고객을 선정하여 적절한 고객 관리 활동을 할 수 있다.

### 2.3 분석결과 요약

지금까지 연구모형에 따른 A홈쇼핑사 고객 데이터의 실증 분석 결과를 요약하면 다음과 같다. 유지고객과 이탈고객은 인구통계적 및 거래 행태 변수에 따라 서로 다른 특성을 가지는 것으로 나타났다. 이는 <표 2-13>과 같이 차이를 보였다.

인구통계적 속성에서는 남성이 여성보다 이탈

<표 2-13> 유지고객과 이탈고객의 변수별 특성 차이

| 구분       | 변수         | 유지비율 높은 집단             | 이탈비율 높은 집단      |
|----------|------------|------------------------|-----------------|
| 인구통계적 속성 | 성별         | 여성                     | 남성              |
|          | 연령         | 40대 이상                 | 20대~30대         |
|          | 직업         | 주부                     | 회사원, 학생         |
|          | 거주지역       | 광역시                    | 서울/수도권          |
| 거래행동적 속성 | 구매횟수       | 2회이상                   | 1회              |
|          | 구매금액       | 높을수록                   | 낮을수록            |
|          | 구매의 최근성    | 2점~6점(최근구매일이 5개월~1개월전) | 1점(최근구매일이 6개월전) |
|          | 주이용 상품군    | 의류, 보석/장신구, 식품, 패션잡화   | 전기전자, 스포츠/레저    |
|          | 이용 상품군 종류수 | 2종류 이상                 | 1종류             |
|          | 최초구매채널     | ARS, 전화                | 인터넷, 카탈로그       |
|          | 주이용채널      | ARS, 전화                | 인터넷, 카탈로그       |
|          | 이용채널수      | 2개 이상                  | 1개              |
| 반품/취소 횟수 | 1회 이상      | 0회                     |                 |

\* 거래 행동적 속성 변수는 6개월 기준.  
 \* 각 변수별 유지 또는 이탈고객 비율이 상대적으로 높은 구분변수를 기입함.

가능성이 크고, 연령대가 낮아질수록 이탈가능성이 크다. 즉, 20대 남성의 이탈비율이 가장 높고, 40대 이상 여성은 유지비율이 높다. 직업은 학생, 회사원의 이탈비율이 높고, 주부의 유지비율이 높다. 지역별로는 서울/수도권의 이탈비율이 높은 편이다. 구매 거래상에서는 6개월간의 구매횟수가 1회 일 때 이탈가능성이 가장 크고 2회 이상으로 갈수록 이탈가능성은 작아진다. 구매금액이 많아질수록, 최종 구매월이 최근에 가까울수록 이탈가능성은 줄어든다. 이용상품군 종류가 한 종류이거나 이용채널수가 하나일 때 이탈가능성이 가장 높으나, 이용상품군 종류수와 이용채널수가 2개 이상으로 늘어날수록 이탈가능성은 낮아진다. 주이용 상품군이 의류, 보석/장신구, 식품, 패션잡화이고, 구매 채널도 주로 ARS나 전화를 이용한 고객이 그렇지 않은 고객보다 이탈할 확률이 낮아진다. 반품/취소 횟수는 예상과 달리, 반품/취소 횟수가 늘어날수록 이탈 가능성은 줄어드는 것으로 나타났다.

유지고객과 이탈고객간의 특성차이 분석 결과에 따라, 고객의 거래를 안정화시켜서 이탈을 방지하고 유지고객화 시킬 수 있는 거래상의 안정화 분기점은 구매횟수 2회(6개월간)라고 판단하였으며, 단, 2회 구매시에 이전과 다른 상품군과 다른 채널로 전환이 이루어졌을 때 거래 안정화된다고 판단된다. 이는 이용채널수가 1개, 이용상품군 종류수가 1개일 때 이탈고객비율이 가장 높았으나, 2개이상으로 늘어나면서 유지고객비율이 높아지는 것과 연관된다. 따라서 구매횟수 2회 이상이면서 이전과 다른 상품군, 다른 채널을 이용할 경우가 거래 안정화를 위한 핵심 기준이라 할 수 있다. 이 때 채널

전환은 가급적 TV홈쇼핑 채널로 전환되는 것이 거래 안정화 확률을 높일 수 있다.

마지막으로 고객이탈확률을 예측하기 위해 가장 적합한 로지스틱 회귀모형을 통하여 이탈에 유의한 변수를 찾아냈으며, 로지스틱 회귀 모형 수식을 통해서 개별고객의 이탈여부 및 이탈확률을 예측할 수 있게 되었다. 이탈율( $P_i$ )을 높이는데 가장 영향력이 큰 변수는 '연령 20대'이고 다음으로 '성별 남성'이다. 이 밖에 구매의 최근성, 주이용채널(ARS여부), 이용상품군 종류수, 주이용상품군(의류여부)의 순서대로 이탈확률( $P_i$ )을 낮추는데 유의한 영향을 주는 변수이다. 구매의 최근성 점수가 낮을수록, 주이용채널이 ARS가 아니고, 주이용상품군이 의류가 아닐 때 그리고 이용상품군 종류수가 적을수록 이탈확률( $P_i$ )은 높아진다.

### Ⅲ. 고객 유지 및 이탈 관리를 위한 CRM전개방안

앞서 분석 결과를 통하여 유지고객과 이탈고객의 인구통계적 및 거래 행태적으로 서로 다른 특성이 무엇인지 알게 되었고, 이탈에 영향력이 큰 변수에 대해서도 살펴보았다. 이에 이탈고객의 특성과 유사한 특성을 보이는 고객들을 대상으로 이탈방지 및 고객 재활성화를 위한 CRM프로그램을 전개해야 한다. 따라서 로지스틱 회귀식에 의해 이탈 확률 수치가  $P_i > 0.5$  이상인 고객을 대상으로 하되, 개별고객의 특성별로 해당 고객을 세분화하여 세분화된 고객집단에 맞게 전개하는 것이 효과적일 것이다.

### 3.1 인구통계적 특성에 따른 이탈방지

로지스틱 회귀모형에서 성별과 연령은 이탈 또는 유지 여부에 영향을 미치는 중요한 변수였으며, 남성일수록, 연령대가 20대에 가까울수록 이탈확률이 높고, 여성일수록, 연령대가 40대 이상일수록 유지확률이 높았다. 이는 이탈방지를 위해서 관리가 필요한 고객군은 이탈확률이 상대적으로 높은 20대~30대 초반 남성 및 20대 여성고객으로 판단된다. A홈쇼핑사는 이들 세분 고객군의 직업과 구입 선호 상품, 평균 구매금액, 이용 채널 등의 구매 거래 행태를 사전에 분석하여, 고객군의 연령과 거래 특성에 맞는 이탈방지관리 프로그램을 기획해야 할 것이다. 예를 들어서 이들 고객군이 구입하기에 적절한 가격대의 구입 선호 상품들로 구성된 기획전을 시행하고, 상품 구입시 이들에게 효과적인 인센티브를 주는 프로그램을 실행할 수 있다. 이 때 이들 고객군에게만 E-Mail이나 DM을 통해서 할인이나 적립 쿠폰을 사전에 발송하여, 구매시 즉시 혜택을 받을 수 있도록 하여 구매를 촉진할 수 있을 것이다.

### 3.2 구매거래 특성에 따른 이탈방지

분석 결과에서 이탈을 방지하고 구매 거래의 안정화를 위한 거래상의 분기점은 구매횟수 1회(6개월간)를 넘기는 것이며, 이용채널과 이용상품군도 1개를 넘겨야 한다. 따라서, 구매횟수가 1회인 고객들을 별도로 분류하여 구매횟수를 1번 더 늘리고 이용상품군과 채널을 전환시키는 프로그램을 시행하여 이탈을 방지해야 한다.

분석 결과 구매횟수가 1회에서 2회로 증가되

었을 때 유지고객 비율이 약 2배로, 이탈고객 비율이 약 1/2로 감소하며, 이용채널과 이용상품군 종류수도 1개에서 2개로 증가될 때부터는 유지고객 비율이 높아졌으므로, 6개월내에 1회만 구매한 고객이 1회 이상 추가 구매가 일어나도록 유도하는 2번째 구매(second trial) 프로그램을 개발하는 것이 고객이탈방지를 위한 방법 중의 하나일 것이다. 더구나, A홈쇼핑사의 경우는 6개월간 1회 거래 고객이 표본 고객의 과반수가 넘는 56%를 차지하고 있으므로, 이들 고객군에 집중적 관리가 요구 된다. Outbound TM을 통해서 고객에게 직접 전화를 하여 구매를 독려할 수도 있고, DM이나 E-Mail을 통해서 2번째 구매시의 특별 인센티브를 제공함으로써 구매를 유도할 수도 있다. 인센티브는 가격 할인부터 적립금 제공, 경품/사은품 행사 등 고객 특성에 맞게 효과적인 것을 선택할 수 있다.

또한, 1회 구매자가 2회째 구매를 일으킬 때는 타 채널로 전환되어 이용채널수가 늘어날 수 있도록 유도해야 이탈을 줄일 수 있다. 특히 이탈율이 높은 인터넷쇼핑몰이나 카탈로그 채널 이용자의 경우는, TV홈쇼핑에서 ARS나 전화로 상품을 구매해보는 경험을 할 수 있도록 체험 프로그램 제공이 필요하다. TV홈쇼핑의 채널 번호와 주문 방법을 알려주고, ARS나 전화 주문시 할인을 받을 수 있는 쿠폰을 제공하여 이용채널이 늘어날 수 있도록 해야 한다.

## IV. 결론

2003년부터 TV홈쇼핑 업계는 케이블 TV의

가시청 가구수가 포화상태에 이르러, 신규고객을 통한 양적 팽창보다는 기존고객의 유지 및 활성화를 통한 질적 성장이 어느 때보다 요구되고 있다. 이러한 때에 고객이탈방지를 위한 CRM활동의 중요성은 점차 커지고 있다. 사실상 기업이 고객유지율을 높이고 고객 이탈방지를 위해서 가장 기본적으로 해야 할 일은 기업이 보유한 고객데이터를 활용하여 이탈고객의 특성을 분석하는 것이다. 그러나 대부분의 기업은 실제 마케팅 현장에서 이론대로 고객데이터를 제대로 활용하고 있지 못한 실정이다.

본 연구에서는 실제 고객 데이터의 실증 분석을 통해서 이탈고객과 유지고객의 속성 및 거래상의 특성 차이를 분석하고, 이탈에 영향을 미치는 변수를 파악하였다는데 의의가 있다. 또한 이탈예측모형을 통해서 이탈 확률을 예측하여 실증분석 결과를 바탕으로 고객유지 및 이탈방지를 위한 CRM전개방안을 제시하였다. 그리고 통계적 분석 과정을 거쳐서 CRM실행 방안을 제안함으로써, TV홈쇼핑사나 인터넷 쇼핑물업체에게 데이터를 활용한 마케팅 활동에 대한 방향성을 제시했다고 생각한다. 따라서 다른 업종의 경우도 본 논문의 연구 결과를 토대로 자사가 보유한 고객 데이터의 실증적 분석을 통해서 자사 고객 특성을 제대로 파악하고, 고객반응을 예측할 수 있는 모형을 개발함으로써, 고객 유지 및 이탈관리를 위한 구체적이고 효과적인 CRM전개 방안을 수립할 수 있다는 점에서 실천적 의미를 가질 것이다.

그러나 TV홈쇼핑사의 고객의 이탈 및 유지에 영향을 미치는 요소를 고객 데이터의 인구통계적 속성 변수와 구매거래 행동 변수에 따른 실증 분석과 모델링을 통해서만 찾아내었다

는데 본 연구의 한계가 있다. 고객이 거래를 이탈하고 유지하는 요인에는 통계적 고객 분석 결과 외에도 정성적, 경쟁적, 사회적 요인이 있을 수 있으며, 홈쇼핑의 경우 예를 들어 C/S불만, 배송 지연, 가격 불만족, 이용 불편, 상품 불만, 주변사람들의 권유와 구전 등의 정성적 요인과 브랜드 선호도 약화, 경쟁 쇼핑물의 변화, 오프라인 할인점의 확대 등의 경쟁적 요인에 의해서도 좌우될 수 있다. 또한 경기 침체, 소비 심리 위축 등의 사회적 요인도 영향을 미칠 것이다. 따라서 향후 연구에서는 이러한 요소들까지도 복합적으로 분석하여 고객의 이탈 및 유지에의 영향 요소를 결정해야 한다.

〈논문 접수일: 2004. 11. 30〉

〈게재 확정일: 2005. 02. 11〉

## 참고문헌

- 강병서, 김계수(2001), **사회과학 통계분석**, SPSS 아카데미.
- 강혜림(1995), "CATV 홈쇼핑 이용자의 구매의사 결정과정에 관한 실험적 연구," 연세대학교 신문방송학과 석사학위 논문.
- 김영숙, 심미영(2002), "Cable TV 홈쇼핑이용 소비자의 특성 및 소비자 특성별 상품 구매 결정요인 분석," **대한 가정학회지**, 제40권 제4호, 85-103.
- 김응렬(2002), "CRM핵심모듈 고객이탈방지 툴 기업전략 부재 효과 못 거둬," **디지털타임즈**, 2002. 08. 07.
- 박찬욱(1996), **데이터베이스 마케팅**, 연암사.

- 성응현(2001), **응용 로지스틱 회귀분석(이론, 방법론, SAS활용)**, 탐진.
- 연순희, 최희수, 유시정(2003), “금융기관의 CRM 전략과 성공요인에 관한 연구,” **서비스경영학회지**, 제4권 제3호, 89-110.
- 이건창, 권순재, 신경식(2001), “은행고객 세분화를 통한 이탈고객 관리 분석-가계성예금을 중심으로-,” **한국지능정보시스템학회 학술대회논문집**, 261-281.
- 이근희(2001), **사회과학연구방법론**, 법문사.
- 이용구, 이현정(2001), “데이터마이닝을 이용한 보험회사 고객 이탈분석에 관한 연구,” **수학·통계논문집**, 제8집, 37-57.
- 천용호, 임현택, 배준성(2002), “의사결정나무를 이용한 이탈고객의 방지에 관한 연구,” **과학기술연구**, 제13집, 1235-1247.
- 한국통계학회 편(1994), **통계용어사전**, 자유아카데미.
- 허명희(1997), **통계적 개념 방법 응용**, 자유아카데미.
- Berkowitz, Eric N., John R. Walton, and Orville C. Walker, Jr.(1979), “In-Home Shoppers: The Market of Innovative Distribution Systems,” *Journal of Retailing*, 55(Summer), 15-33.
- Cox, Donald F. and Stuart U. Rich.(1964), “Perceived Risk, and Consumer Decision-Making: The Case of Telephone Shopping,” *Journal of Marketing Research*, 1(November), 32-49.
- Cunningham, Isabella C. M. and William H. Cunningham(1973), “The Urban In-Home Shopper: Socioeconomic and Attitudinal Characteristics,” *Journal of Retailing*, 49(Fall), 42-50.
- Darian, Jean C.(1987), “In-Home Shopping: Are There Consumer Segments,” *Journal of Retailing*, 63(Summer), 163-186.
- Gillette, Peter L.(1970), “A Profile of Urban In-Home Shoppers,” *Journal of Marketing*, 34(July), 40-45.
- Imhoff, C. and Gentry, J.(2000), “CRM Building an Environment for The Future,” *Technology Guide*, 20, 178-185.
- Marsh, J.(1998), *Managing Financial Service Marketing*, London: Pitman Publishing.
- Raaen, D. A(2000), “A Strategic Approach to CRM,” America's Network.

# A Study of Customer Churn by Analysing CRM Customer Data

Sang Yong Kim\*  
Ji Yeon Song\*\*  
Gi Soon Lee\*\*\*

## Abstract

Customer Relationship Management (CRM) is a corporate marketing strategy maintaining and managing customers. And with CRM companies maximize the customer's value through a series of processes of new customer retention, VIP customer retention, customer value increase, potential customer activation, and customers for lifetime by collecting the customer information and taking advantage of it effectively. In particular, as the competitive environment is changing rapidly and getting more intense, maintaining the customer retention through customer churn management becomes more important in order to increase the customer value for maximizing the company's profit and to build up the relationship with customers. For example, the financial industry has managed the customer churn with the concept of customer segmentation. Recently the customer retention and churn management is becoming increasingly important in all business fields as well as financial industry since the companies expect the effect of preventing the customer churn by identifying characteristics of customers. However, despite the increasing interest and importance of the management of the customer churn, not many of studies are systematically executed by analyzing the data of customer churn. In this study we analyze the actual data of CRM activities for the customer retention, specifically the data of TV home-shopping. By doing so, we hope to identify the differences of demographic attributes and transaction specific characteristics in consumer behaviors between the churning customer and the retained customers. In addition, we try to find out the variables which can impact the churning of the customers and to predict the churn rate of individual customer through our proposed model of customer churn. In the end, based on our findings we suggest the possible marketing strategies for TV home-shopping companies.

Keywords: CRM, VIP customer retention, customer churn

---

\* Associate Professor of Business School, Korea University

\*\* CRM Consultant, SK Corporation

\*\*\* Ph. D. Student of Business School, Korea University