

카드사의 효과적인 분석 CRM을 위한 회원정보 정제 및 보강 방안 사례 연구

Customer Data Cleansing & Enrichment for Analytical CRM : A Credit Card Company Case

신숙원 · 배성민 · 김태호 · 장남식

Shin, Sookwon · Bae, Sungmin · Kim, Taeho · Chang, Namsik

Abstract

Since the late 1990s, many corporations have implemented customer relationship management(CRM) to establish marketing and customer response strategies based on customer analysis, but have realized few benefits for their efforts. This has largely been the result of corporations focusing on implementing CRM solutions while neglecting the conditions necessary for effective utilization. In particular, the quality of customer data gathered has not been a priority in the data gathering, analysis and storage process. Data quality is essential for Analytical CRM, where strategies are developed based on the results of data analysis. Socio-economic data of customers is both easily gathered and widely used for establishing marketing strategies, customer segmentation, and demand forecasting. As a consequence, data enrichment can significantly improve the accuracy of data analysis. This study examines two credit card companies where customer data plays a critical role and proposes methods for standardization and enrichment of address-based data.

Keywords : Analytical CRM, Data Quality, Data Enrichment, Data Cleansing, Credit Card Company

1. 서론

1.1 연구의 배경 및 목적

1990년대 초부터 등장한 인터넷과 정보기술의 급격한 발전은 사람들의 일상생활뿐만 아니라, 기업의 운영방법 및 마케팅방법, 나아가 기업간의 경쟁 규칙을 바꾸어 놓았다. 소비자들은 인터넷을 통해 서로의 정보를 공유함으로써 기업들이 제공하는 상품과 서비스에 대해 더 많은 지식을 가질 수 있게 되었으며, 이를 바탕으로 기업에 영향력을 행사할 수 있게 되었다. 반대로 기업들은 소비자에 대한 더 많은 정보를 수집할 수 있게 되었으며 대용량의 데이터들을 신속하고 정확하게 처리하고 분석할 수 있게 되었다.

처음에는 고객들의 신상정보가 정보수집의 주 대상

이었으나, SFA(Sales Force Automation) 등을 통해 고객들의 거래데이터를 파악할 수 있게 되었으며, 여러 설문조사들을 통해 고객의 성향정보까지도 파악할 수 있게 되었다. 이러한 대용량의 고객 관련 정보들을 분석하여, 기업의 대 고객 전략 수립 및 마케팅에 활용하려는 시도가 CRM(Customer Relationship Management) 이라고 볼 수 있다.

분석 CRM(Analytical CRM)은 고객 관련 데이터의 분석을 통해 마케팅 전략 수립을 위한 고객 선정, 고객군 파악, 이벤트 효과 분석 등에 사용되고 있다. 특히, 기업에서 많은 비용을 소비하는 마케팅 또는 이벤트의 성공여부는 대상 고객군 선정의 정확성 정도에 따라 달라질 수 있기 때문에 고객 분석을 위한 양질의 데이터를 확보하는 것은 매우 중요한 일이다.

하지만, 기업에서의 고객 데이터 확보는 주로 회원 가입 시 고객의 기재를 통해 이루어지게 되며, 고객의 신상변화 또한, 고객이 스스로 업데이트 하지 않는 한 고객이 가입한 시점의 정보로 고정되므로 정확성이 떨어진다고 할 수 있다. 데이터의 불확실성은 분석 결

신숙원 : (주)LG전자, 업무혁신팀, 과장
배성민 : 한밭대학교, 산업공학과, 교수
김태호 : 연세대학교, 정보대학원, 석사과정
장남식 : 서울시립대학교, 경영학부, 교수

과의 신뢰도를 떨어뜨려 기업은 특성에 알맞은 고객들을 선별해 낼 수 없게 되며 잘못 선정된 고객들에 대한 마케팅은 오히려 역효과를 낼 가능성이 크다.

고객정보에 대한 데이터 품질(Data Quality)을 높이기 위한 방법은 여러 가지가 있을 수 있으나, 가장 널리 이용되는 방법은 고객정보를 항상 최신의 정보와 일치시키기 위해 고객정보 수정을 유도하는 이벤트를 수행하는 것이다. 또, 기존 데이터에서 오염된 부분을 제거하거나 또는 외부 데이터를 이용하여 기존 데이터를 보강하여 기존의 데이터들로부터 더 많은 데이터를 추출해낼 수 있도록 하는 방법이다.

특히, 카드사는 다른 업계보다 더 정확한 고객의 인구 통계학적/사회 통계학적 정보를 필요로 하게 된다. 그러므로 카드업계에서의 데이터 품질 향상 및 보강을 위한 방법론을 제시하는 것은 효과적인 CRM 분석을 위해서 매우 필요하며, 해당 분야뿐만 아니라 금융업 전반으로 확산되어 적용될 수 있을 것이다.

1.2 연구의 범위와 방법

본 연구는 2장에서 데이터의 정제, 보강을 포괄하는 개념인 데이터 품질 관리를 소개하고 이 데이터를 이용한 분석형 CRM에 대해 살펴본 후 3장에서는 카드사의 회원정보 정제 및 데이터 보강 방안에 대하여 연구하고, 4장에서는 이를 실제 적용한 카드사의 사례에 대하여 알아보겠다. 그리고 마지막 5장에서는 사례를 통해 시사점을 도출하였다.

2. 문헌연구 및 사례연구

2.1 Data Quality Management의 개념 및 전망

Gartner Group 리포트에 따르면 2007년까지 80% 이상의 기업들이 데이터 품질 문제에 직면하게 될 것이라고 한다[3]. 이러한 데이터 품질의 문제는 대부분 조직의 의사결정이 KPI(Key Performance Index)에 의해 이루어지는데, 시장 점유율, 매출 예측, 고객군 특성 등의 KPI들은 대부분이 기업이 보유하고 있는 데이터를 기준으로 하여 만들어진다. 그러므로, 양질의 데이터를 가지고 있지 않다면 그로부터 얻어지는 KPI도 신뢰할 수 없으며 이는 잘못된 의사결정을 하도록 만들 수도 있다.

이러한 관점에서 데이터 품질 관리(Data Quality Management)는 정보기술의 발전과 급변하는 현대사회에서 축적되는 정보들을 효율적이고 효과적으로 관리하기 위해 등장한 개념이다. 국내에서는 90년대 말부

터 데이터 품질의 중요성이 부각되면서 이에 대한 연구가 진행되었다. 특히 국내 기업들이 데이터웨어하우스를 구축하고 CRM을 구현하면서 국내에서도 지속적인 개선을 위한 데이터 품질에 관한 연구가 활발하게 진행되고 있다. 데이터 품질에 대한 기존의 연구를 살펴보면 Redman은 데이터 품질은 데이터 사용자의 요구 사항을 충족하는 정도 즉, 사용자가 특정 작업을 수행할 때 계속적으로 기대되는 데이터를 제공해 줄 수 있는 것이라고 말했다[11]. Wang과 Strong은 데이터 사용자에게 의해 사용되기 적합한 데이터라고 설명하였고[12], Kelly는 사용자에게 의해서 정의되는 것이라 하였다[7]. Orr는 정보시스템의 역할은 실제세계를 제대로 반영하는 시스템이어야 하며, 이때 조직의 구성원들은 제품을 생산할 수 있고, 의사결정을 할 수 있다. 이러한 관점에서 데이터의 품질은 정보시스템과 실제세계의 동일한 일치성을 측정하는 것이라고 정의하였다[9]. Brackett은 기업의 현재와 미래의 비즈니스 정보 요구에 어떻게 잘 지원하는가를 측정하는 것이라고 정의하고 있다[2].

이러한 데이터 품질은 바라보는 시각에 따라 다양하게 정의되지만 이를 정리하면 <표 1>과 같다.

표 1. 고객 데이터 품질 관리 속성[4]

속성	내용
정확성 (Accuracy)	기록된 값이 실제 값과 일치하는 상태
보전성 (Integrity)	개체와 속성간의 관계를 일관적으로 유지하고 있는 상태
개념적 일관성 (Consistency)	시기적으로 필요한 데이터 요소들의 개념이 반영되어 있는 상태
타당성 (Validity)	관측치들이 사전 정의 된 개념에서 발생 가능한 범위에 존재 하고 있는 상태
시기성 (Timeliness)	기록된 값이 시간적으로 실제 값과 일치하는 상태
접근성 (Accessibility)	사용자가 데이터를 쉽게 접근하고 이해하고 사용할 수 있는 상태

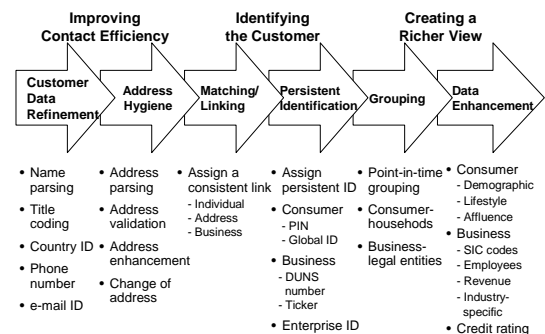


그림 1. 고객 데이터 획득의 6단계

Gartner Group도 몇 년 전부터 데이터 품질 관리의 필요성을 강조하고 있으며, 조만간 데이터 품질 관리가 보편화 될 것으로 전망하고 있다[6].

Radcliffe에 따르면 양질의 고객 데이터를 획득하기 위해서는 <그림 1>에서 제시된 바와 같이 6단계를 거쳐야 한다고 하였다[10].

처음 두 단계(Step 1, Step 2)는 고객과의 효율적인 접근을 위한 고객 정보의 정제 및 수정 부분이다. 이 부분에서는 이름의 형식을 통일하고, 전화번호, e-mail 주소들을 형식에 맞도록 통일하는 일을 하게 된다. 영어권 국가와 달리 우리나라의 이름은 일반적으로 2~4 글자로 구성되며, 성과 이름을 그대로 붙여쓰기 때문에 큰 문제가 되지는 않는다. 하지만, 주소의 경우는 그렇지 못하다. 외국의 경우처럼 우편번호와 정확히 일치하거나 혹은 주소가 어떠한 규칙에 의해 일관적으로 부여되지 못하였기 때문에 주소를 통일하는 것은 기업의 입장에서 큰 부담이 된다. 특히, 우리나라의 경우 같은 주소에 사용할 수 있는 우편번호의 개수는 2~3개가 있기 때문에 어떤 것이 고객의 특성을 잘 나타내어 주는 것인지를 구분하는 것은 매우 중요한 일이며, 다음 장에서 어떻게 이 문제를 해결할 것인지에 대해 제시할 것이다.

다음 두 단계(Step 3, Step 4)는 고객의 정보와 기존에 기업이 가지고 있던 정보를 서로 일치시키는 작업을 수행하는 것을 말한다. 고객의 정보는 여러 부서로부터 얻을 수 있으며 그러한 정보들을 각 고객 개인의 통합된 정보로 인식할 수 있어야 하며, 이를 고객 정보 통합(CDI : Customer Data Integration)이라고 한다. 이 작업은 단순 작업으로 여겨지기 쉬우나, 데이터 품질에 가장 큰 영향을 미치는 부분이므로 전문가에 의해 수행되어야 한다.

마지막 단계인 데이터 보강 단계에서는 기업이 일반적으로 획득하기 힘든 데이터들을 기존 고객 데이터와 일치시키는 것을 의미한다. 예를 들어, 고객들에 대한 만족도 조사, 라이프스타일에 대한 설문조사 등의 결과들을 고객 개개인의 데이터에 보강함으로써 추후 마케팅에 도움을 줄 수 있는 부가 정보들을 추가해 줄 수 있다. 다음 장에서 언급될 것처럼, 개인의 주소의 특성에 대한 데이터를 보강하는 것도 데이터 품질 향상에 긍정적인 영향을 줄 수 있을 것이다.

그러나 아직까지 국내기업들은 데이터 품질 관리의 필요성을 인정하면서도 그에 대한 투자나 개선은 적극적이지 못하다. 데이터 자체의 결함이나 문제로 인한 비용 잠식 규모 등을 파악하지 못하고 있어 투자에 대한 효과를 자신하지 못하기 때문이다. 하지만 훼손된 데이터를 근간으로 한 비즈니스는 실패할 수 밖에

없다는 점에서 데이터 품질 관리는 갈수록 그 필요성과 중요도가 높아질 것으로 보인다.

2.2 분석 CRM에 관한 연구

CRM의 개념은 많은 연구를 통해 어느 정도 정립되어 있으나, 이에 대해서 한마디로 정의하기는 쉽지 않다. CRM이란 용어 자체에 내포된 의미를 살펴보면 CRM은 Customer Relationship Management의 약자로, 말 그대로 ‘고객관계관리’ 혹은 ‘고객관계경영’으로 이해할 수 있다. 여기서 고객이란 막연한 대중 집단이 아니라 ‘확인된 개별 고객’, 혹은 ‘매우 작은 단위로 세분화된 고객 집단’을 말한다. 또한 관계는 기업과 고객이 서로 가지게 되는 우호적인 관계를 의미하는데, 쌍방향적인 상호작용(대화)을 지속적으로 전개함으로써 구축될 수 있다. 그리고 ‘M’은 ‘Marketing’이 아니라 ‘Management’란 용어를 사용하고 있음을 볼 때, 일반적인 마케팅 활동보다는 상위적인 개념의 경영전략적 수준을 의미하는 것으로 이해된다. 즉, 기업내의 다양한 부서와 수단, 자원 및 역량을 통합적으로 활동할 수 있는 ‘관리적인 능력’도 요구된다는 것이다[1]. 이러한 CRM을 <그림 2>에서 보는 바와 같이 Meta Group에서는 운영 CRM과 분석 CRM, 협업 CRM의 세 가지로 분류하였는데[8], 본 논문에서는 세 분류 중 분석 CRM의 효과적인 운영 방안에 대해 알아보겠다.

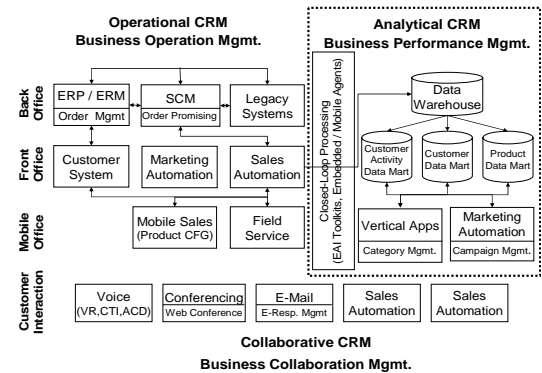


그림 2. CRM 분류

분석 CRM은 고객 데이터를 저장하고 데이터를 분석하여 의미있는 결과를 도출하는 데 필요한 시스템을 말한다. 이는 CRM 운영이나 협업 환경에서 생성되는 데이터를 분석하는 것으로 데이터웨어하우스(Data Warehouse)와 데이터마트(Data Mart)와 연계되는 시스템을 의미하기도 한다.

이러한 분석 CRM의 첫 단계는 분석하고자 하는

데이터를 수집(Collection)하는 것이다. 1990년대 후반에 여러 기업들이 앞다투어 도입하였던 CRM이 그리 큰 효과를 내지 못했던 가장 큰 이유도 데이터를 수집하는데 있어서 기존 시스템에 저장되어 있던 데이터를 그대로 사용하였을 뿐, 그 데이터들에 대한 정제(Cleansing), 맞춤(Matching), 연결(Linking), 식별(Identifying) 등의 작업이 제대로 수행되지 않고, 저장된 상태 그대로 이용되었기 때문이다.

데이터 수집 및 정제 작업을 통해 기업이 얻을 수 있는 장점은 크게 네 가지로 요약해 볼 수 있다[5].

첫째, 전사적 관점에서 개개인의 고객을 식별해내고, 중복된 고객들에 대한 데이터를 제거함으로써 추후 마케팅 비용을 절감할 수 있으며, 둘째로 자사의 신규가입자를 기존 고객과 명확히 구분해 내어 CRM의 효과를 확대할 수 있다. 특히, 카드업계는 고객과의 관계가 오래될수록 기업이 얻게 되는 이윤이 커지기 때문에 신규 고객을 얼마나 오랫동안 유지(Retention)시킬 수 있느냐가 기업의 이윤 창출에 큰 영향을 미치게 된다. 셋째로 기업이 얻을 수 있는 장점은 양질의 데이터로 고객군을 구분하고 고객 개개인의 특성을 파악함으로써 고객들의 만족도(Customer Satisfaction)와 충성도(Loyalty)를 높이는 데 도움을 줄 수 있다. 마지막으로 기업이 데이터 품질을 높이기 위해 끊임없이 노력한다면 궁극적으로 전사적 데이터 통합의 바탕이 될 수 있을 것이다.

3. 카드사의 회원정보 정제 및 데이터 보강 방안

3.1 회원정보 정제 및 데이터 보강의 필요성

3.1.1 회원정보의 활용성 부족

카드사는 상대적으로 회원정보가 풍부한 곳이다. 회원 보유 기준 상위 3사의 각 회원 수는 1000만을 초과하고 있다. 카드사 회원정보의 특징은 청구지 주소가 비교적 정확하나 활용성이 미흡하다는 것이다. 회원의 신상정보는 누락되거나 오류 데이터가 많이 존재하며, 또한 한번 수집된 회원정보에 대하여 업데이트 및 검증 등의 데이터의 객관성이 부족하다는 것이다. 따라서, 회원 신상정보의 활용도를 높이기 위해서는 회원 신상정보의 정제 및 표준화를 추진하고 다양한 외부정보의 수집 및 체계 구축이 되어야 한다.

3.1.2 외부 데이터 보강

고객의 잠재 구매력 및 소비패턴 등 회원성향 파악

이 가능한 외부 데이터의 수집 및 데이터 베이스화를 통한 지역별 속성 정보를 확보하고, 지역 속성 정보와 내부 데이터를 활용하여 회원성향 분석을 제공하기 위해서 외부 데이터를 보강한다. 외부 데이터는 고객 분석 모델을 정교화하고 지역별 고객 성향에 근거한 마케팅 역량 기반을 강화하기 위한 목적으로 수행한다.

3.1.3 내부 데이터 통합

카드사는 회원 속성 정보 및 거래 정보 등의 내부 정보를 이용하여 분석 CRM 및 마케팅 전략 수립/실행 등에 활용할 수 있다. 회원정보가 주소를 핵심 변수로 하여 표준화 및 정제되면 내부의 데이터를 회원 정보로 통합할 수 있다. 여기에 포함될 수 있는 데이터는 각 카드사마다 차이가 있겠으나 독립 카드사의 경우에는 카드 총 이용 금액, 업종별 카드 실적, 복수 카드 구매 금액/서비스 금액, 가맹점 매출 금액 등이 될 수 있다. 또한 은행과 데이터를 공유할 수 있는 카드사의 경우에는 평균 월평균잔액, 급여 이체 정보, 연소득 금액, 부동산 금액, 유동성 수신 평균잔액, 총 대출 정보 등이다.

3.2 회원정보 정제 및 데이터 보강 방안

3.2.1 회원정보 정제 및 표준화 방안

1) 개념 및 수행 방안

회원정보의 정제 및 표준화는 회원의 주소를 핵심 변수(Key Variable)로 하여 정제 및 표준화를 수행한다. 주소 표준화는 고객이 혼용하고 있는 주소 체계 및 단지 명칭을 하나의 기준으로 통일하는 것이다(<표 3> 참조). 주소 표준화 작업은 그 자체에도 의미가 있으나 내·외부 데이터를 보강하기 위한 선행 작업으로 볼 수 있다. 현재 우리나라의 주소 관리 체계는 법정동과 행정동 그리고 우편동으로 관리된다. 실제 회원들은 가입 서류의 주소란에 적은 본인의 주소가 법정동 기준인지 행정동 기준인지 알지 못하며, 회원의 입장에서는 그러한 구분이 무의미하다. 그러나 외부 데이터를 보강해야 하는 측면에서는 주소를 매칭키(Matching Key)로 외부 데이터를 보강하기 때문에 우선적으로 고객정보의 정제 및 표준화가 우선되어야 한다.

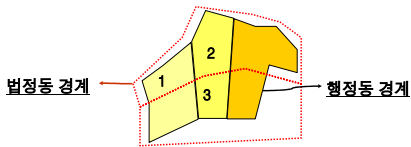
회원 주소의 정제 및 표준화는 크게 두 가지 단계로 이루어진다. 첫 번째 단계는 주소 체계를 정비하고 통일하는 일이다. 법정동/행정동/우편동이 혼재되어 있는 주소 체계를 정비하는 것이다. 예를 들자면 행정동

으로 신사동은 법정동으로는 압구정동이며, 행정동으로 압구정1동은 법정동으로는 신사동이다(<그림 3> 참조). 법정동은 예로부터 전래되어 온 동(洞) 명칭으로 개인의 권리, 의무 및 법률 행위시 사용된다. 모든 법정동 내 지번은 유일하며 주민등록증 등 정부 발행 서류에 표기되는 것이다. 행정동은 행정상 관리를 위해 인구 및 지역규모를 기준으로 부여된 명칭으로 동사무소, 우체국 등의 관리 단위이다. 국가 통계자료 등은 행정동 기준으로 조사 발표된다. 우편등은 우편 업무의 편의를 위해 우체국에서 관리하는 동 명칭이다.

표 3. 주소 체계 표준화 방향

주소 체계 표준화 방안	장점	단점
행정동으로 통일	통계청 등 정부발표 정보 활용시 정보의 손실 없이 활용가능	법정동 기준으로 관리되는 정보활용시 행:법이 1:N인 지역에서는 정보의 손실
법정동으로 통일	건물 및 토지관련 정보 활용시 정보의 손실 없이 활용가능	행정동 기준으로 관리되는 정보활용시 행:법이 N:1인 지역에서는 정보의 손실
행:법이 1:N 이면 법정동 행:법이 N:1 이면 행정동	법정동 행정동 기준으로 관리되는 자료 활용시 정보의 손실 없이 활용가능	행정동과 법정동이 혼용되어 지역별 분석이 용이하지 않음
행정동, 법정동을 따로 관리	법정동, 행정동으로 관리되는 자료 활용시 정보의 손실 없이 활용가능	행정동과 법정동을 각각 관리하는 데 따른 관리비용 발생

..... 표준화 방안



구역	행정동	법정동	고객 표기 주소
1	신사동	압구정동	서울 강남구 압구정동 미성아파트 **동 ***호
2	압구정1동	압구정동	서울 강남구 압구정동 현대아파트 **동 ***호
3	압구정1동	신사동	서울 강남구 압구정동 ***번지

그림 3. 동 체계 혼용(행정/법정/우편 체계의 차이) 현황

두 번째 단계는 텍스트 기반의 주소를 분할(Parsing)하여 숫자 코드화하는 것이다. 분할의 기본 개념은 회원 주소 데이터의 별도 변경없이 기존 회원 주소뿐만 아니라 신규 입력 주소까지도 회원 주소를 분석하여 정비할 수 있도록 구성 요소 별로 자동적으로 세분화한다. 분할된 텍스트를 숫자로 코드화하고, 행정동 체계 중심과 법정동 체계 중심의 외부 데이터 보강에 적합하게 데이터 원천(Source)별 주소 체계 및 코드 테이블

을 정비를 수행한다.



그림 4. 코드화 작업 흐름

2) 표준화 및 정제 작업 범위

회원정보의 표준화 및 정제 범위는 회원정보 주소 표준화, 전화번호 표준화, 그 결과를 활용한 세대화가 있다.

주소 표준화

주소 표준화 작업은 주소 체계의 통일을 위해 행정동, 법정동 별로 관리하고 공동주택의 경우 번지수를 추가한다. 또한 유사 명칭 통일, 분류 및 코드화 작업을 수행한다(<표 4> 참조).

표 4. 표준화 작업 범위

동 체계 통일	행정동, 법정동 및 고객이 오기입한 주소로 혼재되어 있는 주소 체계를 번지 행정동, 법정동으로 각각 Dual관리 예 : 서울 광진구 자양동 644번지 로얄동아아파트 1011803 → 서울 광진구 자양2동 644번지 로얄동아아파트 1011803 행정동 코드 1121583000 법정동 코드 1121510500
주소요소 추가 및 수정	공동주택의 미기입 및 오기입번지를 추가 및 수정 예 : 서울 강남구 압구정동 현대아파트 17-904 → 서울 강남구 압구정동 433번지 현대아파트 17-904호
오타 및 명칭통일	용어상의 다양한 표현을 하나로 통일 (apt, @, APT → 아파트 / BD, 벨딩 → 빌딩) 예 : 서울 강남구 신사동 507번지 강남벨딩 2층 → 서울 강남구 신사동 507번지 강남빌딩 2층
단지 명칭 통일	아파트 단지 명칭(약명칭, 유사 명칭)을 통일화된 하나의 명칭으로 변환 예 : 신시가지아파트, 북동아파트 → 목동3단지아파트
주소요소 분할 및 코드화	주소 세부 항목별 분할(Parsing) 및 코드화(주소, 단지명 등) 예 : 서울 노원구 하계2동 354번지 학여울청구 아파트 103-504 → 11 35 612 804 354 103 504

주소 요소 분할(Parsing)의 기본 개념은 기존 고객 주소 및 신규 입력 주소에 대하여 자동적으로 세분화하는 작업이다. 코드화는 주소 분할 후 세분화 된 자료에 숫자를 부여하는 것으로 간결성, 질서성, 식별성, 공용성, 확장성 등을 갖추어야 한다.

고객정보 Input Format

주민번호	입회날짜	청구지 구분	지역주소 1	지역주소 2	직장주소 1	직장주소 2
------	------	--------	--------	--------	--------	--------

고객정보 Output Format (지역, 직장주소 동일)

주민번호	입회날짜	청구지 구분	주소 1	주소 2	표준화주소	파싱주소 1	파싱주소 10	법정동 코드	결과 코드
------	------	--------	------	------	-------	--------	---------	--------	-------



파싱주소 1~5는 행정동 코드, 파싱주소 11은 법정동 코드
 행정동 코드는 시도(2), 시군구(2), 구(1), 읍면동(3), 리(2) 코드로 분할관리
 법정동 코드는 하나(10)로 붙여서 관리함

그림 5. 주소 표준화 데이터 포맷

전화번호 표준화

구(舊) 지역번호 및 구 국번을 수정해주며, 일률적 번호나 자릿수 오류 등의 외형적인 오류 번호를 찾아준다(<표 5> 참조). 고객이 오기입 또는 미기입한 번호(Garbage)를 회원의 최신번호로 보정 및 갱신해주는 것은 아니며, 수행 방안은 주소 표준화와 마찬가지로 전화번호를 지역번호, 국번, 실번호로 분할한 후 코드화 하는 것이다.

표 5. 전화번호 표준화 작업 범위

구 국번 수정	2자리 국번 및 구 국번 수정 예 : 02 297 5340 → 02 2297 5340
구 지역번호수정	구 지역번호(DDD) 갱신 예 : 0347 792 7751 → 031 792 7751
일률적 오류체크	입력시 코딩자에 의한 일률적 오류 번호 체크 예 : 02-123-4567, 031-111-2222 → 오류
분할 및 코드화	지역번호(DDD), 국번, 번호 별로 분할(Parsing) 예 : 02-573-2305 → 02 573 2305

가구화

표준화된 회원의 주소 정보를 활용하여 고객을 Grouping한 뒤 세대 코드를 부여한다. 이때 그룹핑 키(Grouping Key) 변수로는 표준화 된 주소와 전화번호를 사용하여 일반주택과 공동주택 별로 수행한다. 우체국, 기숙사, 합숙소와 같은 집단 가구 등에 속하는 회원들은 세대화 대상에서 제외하고 세대화 예외 처리 테이블을 만들어 별도 코드로 관리한다(<그림 6> 참조).

공동주택	파싱주소1	...	파싱주소5	파싱주소6	파싱주소7	파싱주소8	파싱주소9	파싱주소10	변경동 코드
	시도 코드		리 코드	대표 번지	가지 번지	아파트 코드	동	호	

공동주택 거주고객 세대 ID 부여 : A + 법정동 코드+대표번지+가지번지+동+호

일반주택	시도 코드	리 코드	대표 번지	가지 번지	아파트 코드	동	호	
	지역번호	국번	번호					

일반주택 거주고객 세대 ID 부여 : H + 법정동 코드+대표 번지+가지 번지+국번+번호

가구화 예외 사항	법정동 코드	본번	부번	비고
4127310800	955	2		대덕전자기숙사
2650010500	60	4		부산세관직원합숙소
2714013500	551	5		대구공산우체국

Ex) 회원 주소가 '경기 안산시 단원구 원곡동 955-2 번지'이고
 전화번호는 031-491-****인 경우 세대화 코드

'G412731080095500204912078'

그림 6. 가구화 데이터 포맷

3.2.2 데이터 보강 방안

1) 개념 및 수행 방안

어떤 외부 데이터를 보강할 지를 결정하기 위해서는 다음과 같은 사항을 고려해야 한다.

첫째, 외부 데이터가 향후 마케팅 활용을 위한 연관성이 있어야 한다. 둘째, 가급적 개인 단위에 접근해야 한다. 셋째, 고객 정보와의 결합이 가능해야 한다. 넷째, 주기적으로 갱신이 가능해야 한다. 다섯째, 포괄 범위가 넓어야 한다. 그리고 외부 데이터를 수집 및 사용하는 데 있어서 법적인 문제가 없어야 한다. 개인의 정보를 사용하는 것은 법적인 제약이 따르며 비합법적인 정보 취득 및 사용은 큰 문제를 일으킬 수 있다.

주요 외부 데이터로서는 공동주택 관련 데이터, 지가 데이터, 지역 속성 관련 데이터 등이 있다.

2) 데이터 보강 작업 범위

외부 데이터를 수집한 후에 고객 신상정보 데이터 베이스와 결합할 수 있는 매칭키(Matching Key)를 정의한다. 외부 데이터 보강을 위해 동 단위 지역 정보 데이터 베이스는 광역시도, 시/군/구, 구, 읍/면/동 필드로 매칭하고, 공동주택 데이터 베이스는 아파트 코드, 아파트 동, 아파트 호 필드로 매칭한다. 공시 지가 데이터 베이스는 광역시도, 시/군/구, 구, 읍/면/동, 리가 통합된 행자부 법정동 코드, 번지 구분, 대표 번지, 가지 번지 필드로 매칭한다. 그리고 변수 변환을 통한 추가 변수를 확보하며, 결측치를 보완/생성하는 방법으로 데이터 보강을 수행한다.

표 6. 공동주택 데이터 포함 내용

아파트 단지기본 정보	아파트 단지별 특성정보	시도/시군구/읍면동 코드
		단지 명, 단지면적, 준공일자, 세대수
		아파트 구조, 난방방식, 방향, 평형 교통, 교육, 주거환경
평형 정보	아파트 평형별 특성정보	분양가, 유자 금액, 관리비
		구조 : 방, 거실, 주방, 현관구조 평형타입, 형별 면적, 평형
지하철 / 철도정보	주변교통 환경정보	지하철 코드, 노선 코드, 노선 명
		버스거리, 도보거리
시세 정보	평수별매매 임대시세정보	조사일자, 평형, 단지 코드
		매매상한가, 임대 상한가
동 정보	아파트동별 특성 정보	단지 코드, 동, 방향
		현관구조, 평형, 세대수
		층, 호, 형태, 특징
동호 정보	가구 단위로 정보를 결합 하기 위한 필수 정보	단지 코드
		동, 호 (층+호)
		평형, 평형타입

공동주택 데이터

공동주택 데이터는 일반 정보 및 시세 정보를 포함하고 있다. 아파트의 경우는 모든 단지에 대한 정보가 제공되는 것이 원칙이고, 연립주택은 아파트와 같이 단지를 구성하는 건에 대한 정보가 포함된다. 공동주택 데이터 베이스에는 단지 정보, 평형 정보, 시세 정보, 동 정보, 동호 정보 등 공동 주택에 대한 세분화된 정보를 포함하고 있다(<표 6> 참조).

지가 데이터

표 7. 지가 데이터 포함 내용

범주	내용
행정구역 관련정보	광역시도
	시, 군/구
	읍/면/동
필지 구분	리
	부여된 코드에 의해 필지가 나누어짐
번지 관련 정보	본번 (대표 번지라고도 함)
	부번 (가지 번지라고도 함)
부가 관련 정보	지목
	면적 (단위 : m ²)
	지가 (단위 : 원)

공시지가 데이터는 전국 필지에 대한 지가, 면적, 지목 등의 정보를 포함한다. 필지란, 하나의 지번이 붙는 토지 등록 단위를 말한다. 지목이란, 토지 분류의 유형으로 토지의 주된 목적 또는 용도에 따라 토지의 종류로 구분하여 표시한 것이며, 전/답/대지 등 총 24종으로 분류된다. 지가 데이터의 지가는 표준 공시지가가 아니라 개별 공시지가에 해당된다(<표 7> 참조).

지역 속성 데이터

인구 주택 총 조사는 5년 주기로 통계청에서 실시하며 그 범위는 대한민국의 행정권이 미치는 영토 내에 상주하는 모든 한국인과 외국인 및 이들이 거주하는 거처에 대한 특성을 조사한다. 가장 최근에 조사한 것은 2005년도 통계청에서 실시하였다.

지역속성 데이터는 이러한 개인별, 주택별 조사 결과를 행정동별로 재집계하고 분석한 자료이다. 인구수 관련 정보, 가구수 관련 정보, 주택 관련 정보 등 인구주택에 대한 전수 조사를 실시하는 것이다. 이러한 자료를 바탕으로 지역별 특성을 Clustering 분석 기법으로 세분화하여 유용한 정보로 활용한다.

4. 카드사 회원정보 정제 및 데이터 보강 사례

본 장에서 Entru Consulting Partners가 수행한 A 카드사와 B 카드사의 회원정보 정제(표준화) 및 데이터 보강 작업을 통하여 분석 CRM 기반을 구축한 사례를 살펴 본다. A 카드사와 B 카드사의 전체 회원 중 10% 샘플을 활용하여 회원정보 정제 및 표준화와 데이터 보강 작업을 수행한 결과를 분석하였다.

4.1 표준화 실행 결과

4.1.1 각 카드사별 데이터 현황 분석

1) 건물 유형별 현황

표 8.A, B 카드사 건물 유형별 현황

구분	상세 구분	자택주소		직장주소	
		A	B	A	B
공동 주택 건물	아파트/아파트예약	36.24%	41.82%	1.41%	2.16%
	빌라/맨션	5.94%	6.26%	0.20%	0.33%
	연립/주택	2.16%	2.34%	0.22%	0.25%
	기타(타운/파크 등)	0.63%	0.54%	0.85%	0.83%
	소계	44.98%	50.98%	2.68%	3.56%
목적성 건물	학교/관공서/부대	0.09%	0.07%	0.86%	0.95%
	상가/백화점	0.08%	0.07%	0.97%	1.05%
	기타(공단/은행 등)	0.01%	0.01%	0.42%	0.17%
	소계	0.18%	0.16%	2.24%	2.18%
빌딩 건물	빌딩	0.35%	0.29%	10.80%	9.89%
기타 건물	일반주택/ 일반주택 예약	54.29%	48.57%	84.28%	84.37%
	전체 합계	100.00%	100.00%	100.00%	100.00%

공동주택 건물은 단지명과 동호를 통해 공동주택이라고 파악되는 주소이며 아파트, 빌라, 연립, 오피스텔 등이 해당된다. 목적성 건물은 주소 내 건물명 이름에

목적이 뚜렷하게 나타난 주소이며 학교, 공단, 은행, 호텔 등이 해당된다. 빌딩 건물은 빌딩명이 주소에 포함되어 있거나 주소 레퍼런스 데이터 베이스에 참조하여 확인되는 경우이다. 나머지 주소는 기타 건물로 분류한다. A 카드사와 B 카드사의 경우 건물 유형 현황은 <표 8>과 같다.

2) 주소 기입 현황

A 카드사의 자택주소에서 주소 필수 정보 기입률은 71.23%이고, 직장주소에서 필수 정보 기입률은 61.00%이다. 이에 비하여 B 카드사의 자택주소에서 주소 필수 정보 기입률은 70.73%이고, 직장주소의 필수 정보 기입률은 72.30%이다(<표 9> 참조).

표 9. A, B 카드사 주소 기입 현황

상세구분	자택주소		직장주소	
	A	B	A	B
주소, 필수 정보 모두 기입	71.23%	70.73%	61.00%	72.30%
필수 주소 정보 미 기 입	번지미기입(공동주택)	18.37%	24.50%	0.42%
	번지미기입(비공동주택)	0.72%	0.02%	1.08%
	공동주택 동 미기입	0.24%	0.17%	0.36%
	공동주택 호 미기입	0.05%	0.03%	0.45%
	빌딩층 or 호 미기입	0.02%	0.01%	1.44%
	주소항목 틀이상 미기입	2.45%	0.53%	5.83%
	리 미기입	0.72%	0.27%	0.68%
	소계	22.57%	25.53%	10.26%
Missing or	주소 이외의 내용	6.20%	3.74%	28.74%
전체	합계	100.00%	100.00%	100.00%

3) 전화번호 기입 현황

표 10. A, B 카드사 전화번호 기입 현황

구분	전화번호 기입	
	A	B
자택전화	기입	94.01%
	미기입	5.99%
	합계	100.00%
직장전화	기입	88.33%
	미기입	11.67%
	합계	100.00%
이동전화	기입	82.59%
	미기입	17.41%
	합계	100.00%

A 카드사와 B 카드사의 전화번호 기입률은 각각 94.01%, 90.72%이며, 직장 전화번호 기입률은 각각 88.33%와 70.93%이다(<표 10> 참조).

4.1.2 주소, 전화번호 표준화

1) 결과 유형

주소, 전화번호 정제 및 표준화의 결과, 유형은 크게 표준화 대상과 표준화 제외 대상으로 구분할 수 있으며, 다시 표준화 대상 데이터는 표준화 성공과 실패로 구분할 수 있다. 이 가운데 표준화 제외 대상은 주소 표준화의 경우는 공동주택의 동/호수 정보가 없거나, 일반주택의 읍/면/동 이하 번지가 없거나, 주소를 아예 기입하지 않은 경우이다. 전화번호 표준화 제외 대상은 지역 국번이 누락되거나, 자릿수 부족 또는 초과, 일률적/인위적 번호(공단, 숫자 '0'으로만 표시된 경우) 기입 등이다(<표 11> 참조).

표 11. 표준화 작업 결과 유형

분류	주소	전화번호	
표준화 대상	표준화 성공	<ul style="list-style-type: none"> 행정동, 법정동 코드 생성 공동주택인 경우 법인 수정 및 추가, 단지명 통일 	<ul style="list-style-type: none"> DDD 변경 서울지역 국번 변경 (2xx → 22xx)
	표준화 실패	<ul style="list-style-type: none"> 단지명(아파트) 레퍼런스에 존재하지 않아 정상 주소를 확인 못한 경우 공동주택 번지 수정 및 추가하지 못한 경우 	<ul style="list-style-type: none"> DDD 존재 & 2자리 국번 변경 불가인 경우 DDD 변경 & 2자리 국번 변경 불가인 경우
표준화 제외 대상	가비지 (Garbage)	<ul style="list-style-type: none"> 주소 필수 항목 부족 공동주택 동, 호 정보 없거나 중복 일반주택 읍/면/동 이하 번지 없음 주소공백 	<ul style="list-style-type: none"> 전화번호 필수 항목 부족 DDD 누락인 경우 필드 오류인 경우 (자릿수 부족 또는 오버) 일률적/인위적 번호, 공단, 숫자, '0', 구분자로만 되어진 경우

2) 주소, 전화번호 표준화 결과

A 카드사의 자택주소 표준화 성공은 90.59%, 직장주소 표준화 성공은 99.99%, 전화번호 표준화 성공은 100%이다. 반면에 B 카드사의 자택주소 표준화 성공은 98.54%, 직장주소 표준화 성공은 99.92%, 전화번호 표준화 성공은 100%이다(<표 12>, <표 13>, <표 14> 참조).

표 12. A, B 카드사 자택주소 표준화 결과

표준화 결과분류		공동주택 건물		빌딩/목적성건물		기타건물	
		A	B	A	B	A	B
표준화 성공	정상	11.59%	22.37%	0.32%	0.43%	47.58%	48.62%
	수정 후 정상	6.70%	4.64%	0.1%	0.06%	6.26%	2.86%
	추가 후 정상	18.02%	19.54%	0.01%	0.02%	0.00%	0.00%
	소계	36.31%	46.55%	0.43%	0.51%	53.85%	51.48%
표준화 실패	레퍼런스 없음 (번지없음)	1.37%	1.38%	0.00%	0.00%	0.00%	0.00%
	레퍼런스 없음 (리없음)	8.05%	0.08%	0.00%	0.00%	0.00%	0.00%
	소계	9.42%	1.46%	0.00%	0.00%	0.00%	0.00%
	성공/실패합계 (Garbage 제외)	45.73%	48.01%	0.43%	0.51%	53.85%	51.48%
Garbage	주소필수항목 부족/오류	0.00%	14.3%	0.00%	0.38%	14.29%	28.16%
	주소공백	22.1%	0.00%	0.00%	0.00%	62.73%	57.16%
	소계	22.1%	14.3%	0.00%	0.38%	77.01%	85.32%

표 13. A, B 카드사 직장주소 표준화 결과

표준화 결과분류		공동주택 건물		빌딩/목적성건물		기타 건물	
		A	B	A	B	A	B
표준화 성공	정상	1.68%	1.37%	12.34%	14.48%	70.37%	74.78%
	수정 후 정상	0.33%	0.19%	2.77%	2.37%	6.04%	5.11%
	추가 후 정상	1.64%	0.63%	4.73%	1.00%	0.00%	0.00%
	소계	3.65%	2.19%	19.84%	17.84%	76.41%	79.89%
표준화 실패	레퍼런스 없음 (번지없음)	0.10%	0.07%	0.00%	0.00%	0.00%	0.00%
	레퍼런스 없음 (리없음)	0.01%	0.01%	0.00%	0.00%	0.00%	0.00%
	소계	0.11%	0.08%	0.00%	0.00%	0.00%	0.00%
	성공/실패합계 (Garbage 제외)	3.76%	2.26%	19.84%	17.84%	76.41%	79.89%
Garbage	주소필수항목 부족/오류	10.84%	3.91%	20.33%	4.91%	30.80%	16.19%
	주소 공백	0.00%	0.00%	0.00%	0.00%	38.03%	75.68%
	소계	10.84%	3.31%	20.33%	4.91%	68.83%	91.78%

표 16. 가구화 결과 예시

고객 번호	표준화 후 자택주소	주택 파싱 번호			가구화 코드	
***	서울 종로구 청운동 15-**번지 (6/1)	02	738	****	H11110101000015005007384608	
***	서울 종로구 청운동 15-**번지 (6/1)	02	738	****	H11110101000015005007384608	
***	광주 동구 계림2동 100-1번지 금호아파트 101동 ***호	062	514	****	A2911010900010000010101050305147509	
***	광주 동구 계림2동 100-1번지 금호아파트 101동 ***호	062	514	****	A2911010900010000010101050305147509	
***	인천 중구 항동7가 27-107번지 라이프아파트 1동 ***호	032	883	****	A2811011800002701070001011008830673	
***	인천 중구 항동7가 27-107번지 라이프아파트 1동 ***호	032	883	****	A2811011800002701070001011008830673	
***	부산 중구 영주1동 16-**번지 (10/3)	051	469	****	H26110101000016001904699524	
***	부산 중구 영주1동 16-**번지 (10/3)	051	469	****	H26110101000016001904699524	
***	서울 중구 신당6동 840번지 현대아파트 8동 ***호	02	6325	****	A111401620008400000008020263252589	
***	서울 중구 신당6동 840번지 현대아파트 8동 ***호	02	6325	****	A111401620008400000008020263252589	

표 14. A, B 카드사 전화번호 표준화 결과

표준화 결과분류		공동주택건물		빌딩/목적성건물		기타건물	
		A	B	A	B	A	B
표준화 성공	정상 번호	99.03%	97.99%	96.76%	98.37%	100.00%	100.00%
	수정 후 정상	0.19%	1.72%	0.23%	1.42%	0.00%	0.00%
	소계	99.21%	99.71%	96.99%	99.79%	100.00%	100.00%
표준화 실패	000 존재 & 2자리 국번 변경불가	0.77%	0.29%	0.99%	0.21%	0.00%	0.00%
	000 변경 & 2자리 국번 변경 불가	0.02%	0.00%	0.02%	0.00%	0.00%	0.00%
	소계	0.79%	0.29%	1.01%	0.21%	0.00%	0.00%
	성공/실패합계 (Garbage 제외)	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Garbage	자릿수 부족 /오류	18.99%	6.51%	16.99%	7.94%	3.90%	0.46%
	공란 또는 'D' 또는 구분자	81.01%	93.49%	81.01%	92.06%	96.10%	99.54%
	소계	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

4.1.3 가구화 결과

카드 회원 가운데 주소와 전화번호 표준화를 성공하고 그 가운데 집합건물 주소 대상을 제외한 나머지 회원에 대하여 가구화를 수행하였다. 가구화는 표준화된 자택주소 및 전화번호에 기반하여 동일 주소, 동일 전화번호를 가진 회원을 가구로 묶은 것이며 혈연관계인 가구를 정의한 것은 아니다. 각 카드사의 가구화 결과는 <표 15>과 같으며, 결과 예시는 <표 16>과 같다.

표 15. A, B 카드사 가구화 결과

가구별 고객 수	가구 비율		고객 비율	
	A	B	A	B
1인	85.2%	76.4%	73.1%	59.8%
2인	13.3%	20.1%	22.7%	31.4%
3인	1.3%	2.8%	3.5%	6.8%
4인이상	0.2%	0.6%	0.7%	2.0%

4.2 데이터 보강

4.2.1 결과 유형

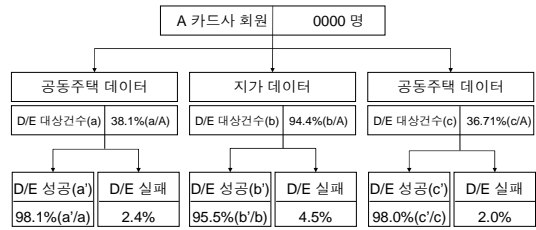
데이터 보강 결과 유형은 크게 데이터 보강 대상과 데이터 보강 제외 유형으로 구분되며, 데이터 보강 대상은 다시 데이터 보강 성공과 실패 유형으로 구분된다(<표 17> 참조). 데이터 보강은 표준화된 자택주소 중에서 공동주택, 일반주택 주소를 대상으로 하고 빌딩 및 목적성 건물과 표준화 결과 Garbage건은 데이터 보강 대상에서 제외한다.

표 17. 데이터 보강 결과 유형

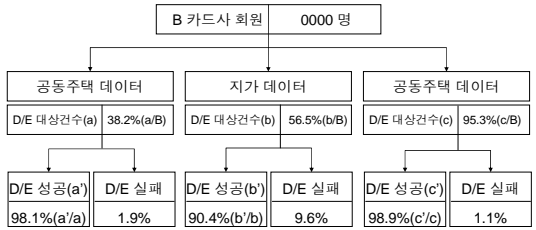
분류		공동주택 데이터	지가 데이터	동 단위 데이터
데이터 보강 대상	성공	<ul style="list-style-type: none"> 공동주택 평형/시세 결합 공동주택 평형결합 공동주택 임대여부 결합 	지가결합 지번(본번)에 대한 평균 지가 결합	동 단위 지역정보 결합
	실패	<ul style="list-style-type: none"> 유사 동 명칭이 레퍼런스에 존재하지 않은 경우(가동, A동) 공동주택 데이터 동 or 호 구간이 벗어난 경우 	지가 레퍼런스에 존재하지 않은 경우	행정동 변환 실패로 인해 행정동 코드가 미부여 된 경우
데이터 보강 제외대상		<ul style="list-style-type: none"> 공동주택으로 분류되었으나 아파트 코드 미부여건 공동주택 표준화 실패건 주소 표준화 결과가 Garbage이거나 일반주택/빌딩/목적성 건물일 경우 	주소 표준화 결과 실패건 주소 표준화 결과가 Garbage이거나 일반주택/빌딩/목적성 건물일 경우	주소 표준화 결과가 Garbage이거나 일반주택 / 빌딩 / 목적성 건물일 경우

4.2.2 데이터 보강 결과

데이터 보강의 수행은 공동주택 데이터는 표준화 결과 공동주택 레퍼런스에 포함된 아파트 코드가 부여된 건을 대상으로 한다. 지가 데이터는 행정구역 및 번지가 존재하는 건을 대상으로 한다. 동 단위 데이터는 표준화 결과 읍/면/동 이상 행정구역이 식별되는 건을 대상으로 한다. 데이터 보강 결과는 <그림 7>과 같다.



(a) A 카드사 데이터 보강 결과



(b) B 카드사 데이터 보강 결과

그림 7. 카드사별 데이터 보강 결과

5. 결론

5.1 사례의 시사점

A 카드사와 B 카드사의 사례를 통하여 데이터 표준화 및 보강은 마케팅 조직이나 분석 CRM을 수행하는데 있어서 의미 있는 과정임이 입증되었다. 표준화 및 외부 데이터의 보강으로 회원 데이터의 분석 및 전략적 활용을 위한 기반이 구축되었다.

CRM 영역에서의 데이터 표준화와 보강을 통한 데이터 품질 향상은 다음과 같은 기대 효과를 얻을 수 있다.

첫째, 우수 고객 획득 및 우수 잠재 고객 개발을 통한 타겟 마케팅이 가능하다. 보강된 데이터를 바탕으로 지역 성향, 객관적 소득/자산, 직업 등을 기준으로 잠재 우수 고객 그룹을 추출하고 타겟 마케팅을 실시한다.

둘째, 활성/비활성/휴면 고객 등 활성화도에 따라 고객 대상의 마케팅 전략을 개발할 수 있다. 고객의 지역 특성, 라이프 스타일 등을 고려한 차별적 마케팅을 수행한다.

셋째, 신상, 구매 행동, 지번 지역정보, 세대화 정보, 가구당 카드 수, 주거 정보, 주거 형태, 주거지 시가, 소득 정보 등 고객 자산 척도(Wealth Indicator)를 표준화 및 외부 데이터 보강을 통한 데이터 품질의 향상으로 고객별, 지역별로 교차 판매, 공동 마케팅 등에 활용할 수 있다.

넷째, 지역별 회원/비회원을 파악을 통한 지역 거점별 영업 지원이 가능하다. 지역별 점유(침투)율을 통해 영업 실행 전략을 수행할 수 있다.

what data quality means to data consumers”, *Journal of MIS*, Vol.12(4), 1996, pp.5-34.

5.2 향후 연구방향

분석 CRM을 수행하는 근본 목적은 고객에 대한 이해를 기반으로 마케팅/CRM 실행 전략을 수립하는데 있다. 본 연구는 주소기반의 데이터 표준화 및 보강 방안을 통한 고객이해의 접근을 수행하였다. 향후 주소를 매칭키(Matching Key)로 하는 데이터뿐 아니라, 다른 요소를 매칭키로 하는 외부 데이터를 추가적으로 발견하고 분석하여 다양한 관점에서 고객에 대한 이해를 증진 시키고 데이터 품질을 향상시키는 노력이 필요하다.

참고문헌

- [1] 이재문, “고객관계관리(CRM)의 이해 및 활용”, *Marketing Communication Review*, Vol.10(겨울호), 2000, p.4.
- [2] Bracket, M., “The concept of data resource data”, *Data Resource Management*, Vol.4, 2002, pp.149-150.
- [3] Desisto, R. P., *Bad Sales Data is worse than No data at all*, Gartner, 2004.11.
- [4] Eckerson, W. “Excerpt from TDWI’s Research Report Data Quality and the Bottom Line”, *Business Intelligence Journal*, Vol.8(1), 2003.
- [5] Friedman, T., Nelson, S. D. and Radcliffe, J., *CRM demands Data Cleansing*, Gartner, 2004.12.
- [6] Friedman, T. and Bitterer, A., *Market Scope Update : Data Quality Technology*, Gartner, 2005.6.
- [7] Kelly, S., *Data Warehousing in Action*, 1st edition, John Wiley & Sons, Ltd. 1997.
- [8] META Group, *The CRM Ecosystem*, META Group, 2000.
- [9] Orr, K., “Data quality and systems theory”, *Communication of the ACM*, Vol.41(2), 1998, pp.66-71.
- [10] Radcliffe, J., *Six Steps to Creating Accurate, Rich Customer Information*, Gartner, 2004. 11.
- [11] Redman, T. C., “The Impact of poor data quality on the typical enterprise”, *Communication of the ACM*, Vol.41(2), 1998, pp.79-82.
- [12] Wang, R. Y. and Strong, D. M., “Beyond accuracy :

저자소개



신숙원
(Shin, Sookwon)

연세대학교 금속공학과를 졸업하고 포항공과대학교에서 자성재료를 전공하였다. Entrue Consulting Partens의 금융그룹에서 은행과 카드산업의 BPR, 업무연속성계획(BCP), 전사적 아키텍처(EA)수립 등의 다양한 프로젝트를 수행하였다. 현재 LG전자 업무혁신팀, 과장으로 근무 중이며 주요 관심분야는 변화관리(Change Management), EA 등이다.

E-mail : paul_shin@lge.com

Tel : +82-2-3777-5474



김태호
(Kim, Taeho)

충북대학교 경영정보학과를 졸업하고 현재 연세대학교 정보대학원에서 디지털 비즈니스 전공으로 석사 학위 과정에 있으며 주요 연구 관심분야로는 CRM, IT Governance 등이다.

E-mail : taeho@yonsei.ac.kr



배성민
(Bae, Sungmin)

한국과학기술원(KAIST) 산업경영학과를 졸업하고, 한국과학기술원 산업공학과에서 석사, 박사를 취득하였다. 현재 한밭대학교 산업경영공학과 조교수로 재직 중이다. 관심분야는 데이터마이닝(Data Mining), 고객관계관리(CRM), 마케팅 분석, 서비스 TQM, e-Manufacturing 이며, IEEE Intelligent Systems, Expert Systems with Applications, Computers & Industrial Engineering 등의 저널에 논문을 발표하였다.

E-mail : loveiris@hanbat.ac.kr

Tel : +82-42-821-1756



장남식
(Chang, Namsik)

University of Missouri에서 경영학 석사, 그리고 University of Arizona에서 경영정보시스템을 전공하여 경영학 박사학위를 취득하였으며, LG-EDS Systems(현 LG CNS)에서 IT 컨설턴트로 근무한 바 있다. 현재 서울시립대학교 경상대학 경영학부에 재직 중이다. 주요 관심분야는 데이터웨어하우스 시스템 설계 및 구축, 데이터마이닝을 이용한 각종 산업별 데이터 분석 및 활용, 그리고 이들과 CRM의 효과적인 접목 방안 연구 등이다.

E-mail : nchang@uos.ac.kr

Tel : +82-2-2210-2169