



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

碩士學位論文

데이터 마이닝기법을 이용한  
외국인전용카지노 고객 가치도

濟州大學校 大學院

電算統計學科

高 權 瑚

2016年 12月

# 데이터 마이닝기법을 이용한 외국인전용카지노 고객 가치도 분석

指導教授 金鐵洙

高權瑚

이 論文을 理學 碩士學位 論文으로 提出함

2016年 12月

高權瑚 理學 碩士學位 論文으로 提出함

審査委員長 金鐵洙 

委員 宋竣模 

委員 吳昇喜 

濟州大學校 大學院

2016年 12月



**Analysis of client values of foreign exclusive casino through  
Datamining technology**

Kwonho Ko

(Supervised by professor Chulsoo Kim)

A thesis submitted in partial fulfillment of the requirement for the degree of  
Master of Science

2016. 12.

Department of Computer Science and Statistics

GRADUATE SCHOOL

JEJU NATIONAL UNIVERSITY

# 데이터 마이닝기법을 이용한 외국인전용카지노 고객 가치도 분석

## Abstract

1. 서론	
1-1. 연구 배경 및 동기.....	10
1-2. 연구 방법 및 범위.....	11
2. 이론적 배경	
2-1. 카지노의 정의와 의의.....	12
2-2. 국내 카지노 현황.....	14
2-3. 데이터 마이닝 기법 적용 필요성.....	17
3. 데이터 마이닝 적용 방안	
3-1. 주요변수 결정 및 관련 데이터 생성.....	18
4. 분석 결과	
- 접목 및 결과 도출 적합성 여부 검토.....	23
5. 결론 및 시사점.....	41

## List of Tables

[표 1] 우리나라 카지노 현황.....	15
[표 2] RFM 기준표.....	20
[표 3] 다중회귀 분석 모형 요약 (고객별 매출액) .....	23
[표 4] 다중 회귀모형의 분산분석 (고객별 매출액) .....	23
[표 5] 다중 회귀모형의 계수 (고객별 매출액) .....	23
[표 6] 다중회귀 분석 모형 요약 (기대수익) .....	24
[표 7] 다중 회귀모형의 분산분석 (기대수익) .....	24
[표 8] 다중 회귀모형의 계수 (기대수익) .....	24
[표 9] 다중회귀 분석 모형 요약 (RFM) .....	25
[표 10] 다중 회귀모형의 분산분석 (RFM) .....	25
[표 11] 다중 회귀모형의 계수 (RFM) .....	25
[표 12] 다중회귀 분석 모형 요약 (고객생애가치) .....	25
[표 13] 다중 회귀모형의 분산분석 (고객생애가치) .....	26
[표 14] 다중 회귀모형의 계수 (고객생애가치) .....	26
[표 15] 상관분석 결과(고객별매출액, 기대수익, RFM스코어, 고객생애가치).....	28
[표 16] 분석에 사용된 변수(의사결정나무) .....	28
[표 17] 의사결정나무 분석의 오분류 테이블(고객별 매출액).....	31
[표 18] 의사결정나무 분석의 오분류 테이블(기대수익).....	32
[표 19] 의사결정나무 분석의 오분류 테이블(RFM스코어).....	33
[표 20] 의사결정나무 분석의 오분류 테이블(고객생애가치).....	33
[표 21] Hosmer와 Lemeshow 검정(고객별 매출액).....	34
[표 22] 방정식에 포함된 변수(고객별 매출액).....	34
[표 23] 로지스틱 모형의 오분류 테이블(고객별 매출액).....	34
[표 24] Hosmer와 Lemeshow 검정(기대수익).....	35
[표 25] 방정식에 포함된 변수(기대수익).....	35
[표 26] 로지스틱 모형의 오분류 테이블(기대수익).....	35
[표 27] 가치고객 가능성 점수 샘플 데이터(기대수익).....	36
[표 28] 가치, 비가치 고객 비율 정리 (기대수익).....	36
[표 29] Hosmer와 Lemeshow 검정(RFM).....	36
[표 30] 방정식에 포함된 변수(RFM).....	36
[표 31] 로지스틱 모형의 오분류 테이블(RFM).....	37
[표 32] 가치고객 가능성 점수 샘플 데이터(RFM).....	37
[표 33] 가치, 비가치 고객 비율 정리 (RFM).....	37
[표 34] Hosmer와 Lemeshow 검정(고객생애가치).....	38
[표 35] 방정식에 포함된 변수(고객생애가치).....	38
[표 36] 로지스틱 모형의 오분류 테이블(고객생애가치).....	38
[표 37] 가치고객 가능성 점수 샘플 데이터(고객생애가치).....	39
[표 38] 가치, 비가치 고객 비율 정리 (고객생애가치) .....	39
[표 39] 각 분석별 결과 비교표.....	40

## List of Figures

[그림1]	연구 방법 및 범위에 관한 도표	11
[그림2]	연도별 카지노 매출 추이	16
[그림3]	데이터 생성 관련 프로세스(기대수익)	19
[그림4]	SQL (기대수익)	19
[그림5]	샘플데이터 (기대수익)	20
[그림6]	데이터 생성 관련 프로세스(RFM)	20
[그림7]	SQL (RFM)	21
[그림8]	샘플데이터 (RFM)	21
[그림9]	데이터 생성 관련 프로세스(고객생애가치)	21
[그림10]	SQL (고객생애가치)	22
[그림11]	샘플데이터 (고객생애가치)	22
[그림12]	독립변수별 회귀모형 그래프(고객별 매출액)_평균배팅액,참여율,게임일수,게임시간,핸드수	26
[그림13]	독립변수별 회귀모형 그래프(기대수익)_평균배팅액,참여율,게임일수,게임시간,핸드수	26
[그림14]	독립변수별 회귀모형 그래프(RFM스코어)_평균배팅액,참여율,게임일수,게임시간,핸드수	27
[그림15]	독립변수별 회귀모형 그래프(고객생애가치)_평균배팅액,참여율,게임일수,게임시간,핸드수	27
[그림16]	고객별 매출액(y) 과 기대수익(x) 기준 고객 분류	29
[그림17]	고객별 매출액(y)과 RFM스코어(x) 기준 고객 분류	29
[그림18]	고객별 매출액(y)과 고객생애가치(x) 기준 고객 분류	29
[그림19]	고객별매출 구분에 대한 의사결정나무	30
[그림20]	기대수익 구분에 대한 의사결정나무	31
[그림21]	RFM 구분에 대한 의사결정나무	31
[그림22]	고객생애가치 구분에 대한 의사결정나무	32
[그림23]	로지스틱 회귀모델에 대한 ROC곡선	39

# 데이터 마이닝기법을 이용한 외국인전용카지노 고객 가치도 분석

## Abstract

대한민국의 카지노는 총 17 곳이 있다. 내국인 출입이 허용되는 강원랜드 1 곳 과 외국인 전용 카지노 16 곳으로 서울에 3 개, 부산, 강원 각 2 개, 인천, 대구 각 1 개, 제주에 8 개로 구성되어 있다. 강원랜드를 제외한 나머지 16 개 외국인전용 카지노는 보통 VIP 고객의 매출이 80% 이상을 차지하고 있는데 중국 고객이 대다수이며 각 업체별로 활발한 마케팅 활동으로 인한 고객 유치를 위해 치열한 경쟁을 벌이고 있다. 특히 제주 시장은 항공, 인프라, 쇼핑 등의 모든 하드웨어시설이 부족할 뿐만 아니라 8 개 업체가 밀집되어 있어 더욱더 치열한 경쟁이 벌어지고 있으며 그 결과 과다 경쟁으로 인한 원가 상승, 이익율 저하, 누적되는 영업 손실로 인한 휴업, 소규모 영세 카지노 업체의 잦은 소유권 이전 등의 악순환이 반복되고 있으며 심각한 경영난을 겪고 있다. 때문에 고객 가치도 분석을 통한 효율적인 고객 관리의 중요성이 그 어느 때 보다 더욱 절실해지고 있는 시점이다. 더욱이 외국인 전용 카지노와 같은 VIP 고객위주의 서비스업은 고객 한명, 한명의 세부 관리가 더욱더 중요하다고 할 수 있다. 하지만 안타깝게도 현재 제주도내 카지노는 시스템에 기반을 둔, 데이터를 근거로 한 CRM(고객관계관리) 측면 역시 많이 부족한 실정이다. 고객 관리를 통한 고객별 타겟 마케팅 및 프로모션 제공은 효과를 극대화 시킬 수 있음을 상기해야 할 것이다. 따라서 본 논문에서는 현재 카지노의 고객 관리에 대한 문제점을 짚어보고 보다 데이터에 근거를 둔 체계적인 분석을 함으로써 실제 데이터에 적용 및 시뮬레이션을 통해 적합성 여부를 검토해보고 한다. 먼저 P사의 데이터 베이스 내에서 샘플 데이터를 토대로 고객 가치도를 추려 낼 수 있는 주요지표 (기대수익, RFM, 고객생애가치) 등을 생성하여 통계적 분석에 접목 시킬 예정이다. 먼저 각 통계방법에 대한 이론을 살펴 본 후, **(1)회귀 분석(regression analysis)** 을 통한 고객별 매출(종속변수)에 영향을 미치는 요인(설명변수) 등에 대해 분석을 하고, **(2)상관 분석(correlationanalysis)**을 통하여 변수가 연관성 또는 이질성에 대해서 분석할 예정이다. 또한 **(3)의사결정 트리(Decision Tree)**를 이용하여 가치 고객군 과 비가치 고객군을 분류 분석할 예정이다. 그리고 매출의 수익과 손실이 발생할 수 있는 특징을 갖고 있는 카지노의 실제 상황과 얼마나 적합한지에 대한 검증은 하고자 한다.



# Analysis of client values of foreign exclusive casino through Datamining technology

## Abstract

There are a total of 17 casinos in Korea including 1 in Gangwon Land that allows entry of domestic persons and 16 foreign exclusive casinos composed of 3 in Seoul, 2 each in Busan and Gangwon, 1 each in Incheon and Daegu, and 8 in Jeju. At the 16 foreigner exclusive casinos with the exception of the Gangwon Land, the revenue generated by the VIP clients generally account for more than 80% of the total revenues. Majority of such VIP clients are Chinese clients. These casinos are waging severe marketing competitions to secure and entice clients into their casinos. In particular, there is even severer competition in the Jeju market due to the concentration of 8 casinos as well as lack of all the hardware facilities including air flight, infrastructure and shopping, etc. As the result of the excessive competition, vicious cycle of temporary closure and frequent transfer of the ownership of the small casinos due to the increase in cost, drop in profit rate and accumulated operating deficits is being repeated with the casinos confronted with severe financial crisis. Therefore, the need for efficient management of clients through the analysis of the level of the value of the clients is becoming even more important at the moment. Furthermore, detailed management of each and every one of the clients is even more important for VIP client-centered service businesses such as foreigner exclusive casinos. However, casinos in Jeju Island unfortunately are also lacking substantially in the aspect of system-based and data-based customer relationship management (CRM). It must be remembered that the effects of target marketing and provision of promotion for each of the clients can be maximized through customer management. Accordingly, this thesis will examine the current problems in the customer management by casinos. Appropriateness of their management practices will be reviewed through application to actual data and simulation by executing data-based systematic analysis. Firstly, key indices that can screen out the level of the value of clients (anticipated profit, RFM and client's life value) will be generated on the basis of sample data within the database of the company P and grafted to the statistical analysis. Firstly, after having examined the theories of each of the statistical methods, **(1)** factors (explanatory variables) that impart effects on the revenues for each of the clients (dependent variables) will be analyzed through **regression analysis**, and **(2)** the correlation or heterogeneity of

the variables will be analyzed through **correlation analysis**. In addition, **(3)** valuable client and non-valuable client groups will be categorized and analyzed by using **Decision Tree**. Moreover, the actual circumstances of the casinos with the features that can result in the profits and losses of revenues as well as the extent of their appropriateness will be verified.

# 1. 서론

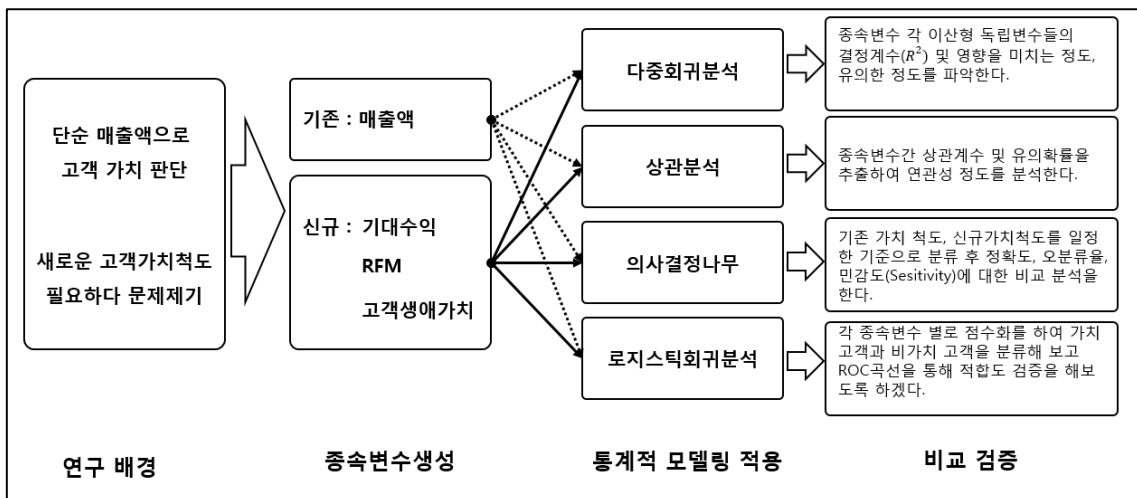
## 1-1. 연구 배경 및 동기

관광 산업은 현재 다른 산업과 달리 두드러진 발전을 거듭하고 있으며, 주요 연계산업과 병행하여 눈부신 도약을 준비 중에 있다. 특히 제주도의 산업은 현재 1차(농·임업 등), 2차(제조업 등) 등의 약세의 효과로 인해 3차(관광산업)의 역할이 그 어느 때 보다 더 중요해졌는데 호텔, 면세점, 관광지, 골프, 그 외 요식업 등이 주요산업으로 각광 받고 있으며 그 중에서도 특히 카지노산업의 중요성이 무엇보다도 이슈화 되고 있다. 카지노산업은 어떤 관광 산업보다도 고 부가 가치 산업이며, 고용창출 효과 및 외국인을 통한 외화 획득에 가장 큰 역할을 한다고 할 수 있다. 또한 사행성 산업 특성상 매출액 대비 낮은 원가율로 인해 영업 이익율이 높아 모두가 선호하는 산업이다. 하지만 '카지노는 문만 열어도 돈을 번다'는 낭설은 이미 옛날 이야기이고 현재는 업체들 간 고객 유치를 위한 마케팅 전쟁이 한창 뜨겁게 달아 오르고 있다. 특히 제주도는 이미 도내 8개 카지노가 있으며, 대형 복합 리조트(Integrated Resort)내 카지노가 허가 신청서를 제출해 있는 상태이다. 이렇듯 제주 카지노는 무한경쟁 시대에 접어 들기에 임박해 있으며 업체들간 과열 경쟁은 결국 원가율을 높이고 영업 이익율을 떨어뜨리는 악영향을 초래하기 마련이다. 더욱이 최근 불거진 중국 정부와의 외교적인 문제는 앞으로 다가올 카지노와 관광산업의 미래를 한층 더 어둡게만 하고 있는 것도 사실이다. 이런 경우 매출의 증대와 원가의 효율적인 절감을 위하여 무엇보다 중요한 것이 바로 '고객 관계 관리(CRM)'에 있다. 기존 고객의 재방문 유도 및 신규 고객 발굴, 고객의 서비스를 제공함에 있어 체계적인 관리 등 CRM에 근거한 고객 관계 관리가 중요한데 이러한 모든 부분이 '데이터 마이닝'을 통한 고객 분석이 전제되어야만 한다. 고객관계관리(CRM)의 목적은 고객을 파악하고 고객의 구매(실적) 패턴을 이해하고 예측하며 적절한 마케팅 도구를 결정한 뒤 고객에게 개인화된 방법을 통해 직접 전달하는 것이다. 따라서 고객관계관리는 차별적인 프로모션 활동의 전제인 고객 집단을 표적화하는 마케팅 행위의 전제 조건으로 여겨지고 있다.(Rossi et al., 1996) 특히 카지노 처럼 전체 고객 군 중 20%의 VIP고객이 전체 매출에 80%를 차지하는 [파레토 법칙]의 이론이 적용되는 산업에는 더욱 중요하다고 할 수 있다. 다만 본 논문에서는 개인정보 보호법 강화에 따른 민감 정보, 개인 사생활 정보(업종, 소득수준, 가족정보, 자금 출처)등의 정보 수집이 어려워 기본적인 범주형 데이터(성별, 연령대, 주거지역 등)에 의존할 수 밖에 없는 한계점도 있었다.(카지노 VIP는 정치계, 연예계 등의 민감 업종 종사자도 있어 개인 사생활 정보 공유를 극히 꺼려한다.) 때문에 활용할 수 있었던 정보는 극히 제한적이긴 하였지만 연속형 데이터(배팅액, 참여율, 게임일수, 시간, 회전수 등)를 위주로 새로운 종속변수(기대수익, RFM, 고객생애가치)를 생성할 수 있었다.

새로운 종속 변수는 기존 종속변수와 연관성 및 이질성, 또한 비교를 통한 적합성 등을 본 논문에서 연구 해보고 검토 해보기로 한다.

## 1-2. 연구 방법 및 범위

본 연구는 현재 국내에서 영업중인 P사 카지노의 데이터를 활용하여 행동 및 구매(실적), 마케팅 전략을 예측하기 위해 먼저 데이터 수집을 통한 데이터 웨어 하우스(Data Ware House)를 구축하고, 분석을 위한 표본 샘플을 추출 한다. 그리고 변수를 설정 및 통계적 모델 적용하고, 그 후 구매(실적), 마케팅 전략 및 고객 방문 예측 및 결과 도출 후 성과 분석 을 하고자 한다. 마지막으로 결과에 대한 통계적 검정을 통해 적합성 판단을 한다.



<그림1> 연구 방법 및 범위에 관한 도표

<그림 1>에서 보여지듯 먼저 다중회귀 분석을 통하여 종속변수가 가장 영향력 있는 설명변수를 선별한다. 그리고 의사결정나무를 통해서 가치 고객군과 비가치 고객군을 구분짓는데 가장 영향을 미치는 변수를 기준으로 샘플을 추출 후 로지스틱 회귀분석을 통하여 가치 고객군과 비가치 고객군에 대한 예측 확률을 추출, 점수화 하여 예측 가치 고객군과 예측 비가치 고객군을 나누어 분석 하여 보고자 한다. 그리고 상관분석을 이용하여 신규 종속변수가 기존 종속변수인 매출액에 과하여 얼마나 상관성이 있는지에 대해서 분석하여 신규 종속변수가 매출 증대에 있어 크게 벗어나지 않는지에 대해서 알아보고자 한다.

## 2. 이론적 배경

### 2-1. 카지노의 정의

카지노의 정의를 아래와 같이 크게 4가지 분류로 나누어 정리해 보았다.

#### ① 법률적 개념

카지노는 관광사업의 한 종류로서 '전용영업장을 갖추고 주사위, 트럼, 슬롯머신, 등 특정한 기구 등을 이용하여 우연의 결과에 따라 특정인에게 재산상의 이익을 주고 다른 참가자들에게 손실을 주는 행위 등을 하는 업'으로 규정(관광진흥법 제3조 1항 5호)

#### ② 일반적 개념

카지노(Casino)란 도박, 음악, 쇼, 댄스 등 여러가지의 오락시설을 갖춘 연회장이라는 의미 이태리어의 카자(Casa)가 어원으로 르네상스 시대의 귀족이 소유하고 있었던 사교.오락용의 별관을 뜻하였음. (Wykes, 1994) 지금은 해변, 온천,휴양지 등에 있는 일반 실내 도박장을 의미한다. 한편 웹스터 사전에 의하면 카지노란 모임(Meeting), 춤(Dancing) 그리고 전문갬블링(Professional Gambling)을 위해 사용되는 건물이나 넓은 장소라고 정의 되어 있음. (정은영, 2011)

#### ③ 관광산업적 개념

카지노는 주요 관광산업들과 크게 연관되어 있으며 특히 특급 관광 호텔 내에 위치하여 관광객에게 게임, 오락, 유흥을 제공하여 체류시간을 연장하고 관광객의 지출을 증대시키는 관광산업의 주요한 사업종의 하나로 규정. (문화체육관광부, 2015)

#### ④ 실질적 개념

게임장에 입장하여 게임을 즐기고자 하는 다수의 고객들에게 게임중재를 하고 게임중재의 대가로 일정 수수료를 지급받는 행위를 통해 수익을 창출 하는 사업

아직까지는 일반적으로 카지노를 도박 또는 사행성 산업으로 생각하는 것이 현실이다. 가끔씩 나오는 언론 보도 역시 카지노의 이미지에 좋지 않은 영향을 미치고 있는 것도 사실이다. 도박 중독, 도박 폐해, 비자금 조성, 재산 탕진, 자해, 자살 등의 언론 보도 등으로 다른 종류의 도박으로 인한 문제 역시 대부분의 카지노의 탓으로 비춰지는 경향이 많다. 허나 관광산업적인 개념으로 볼 때 마카오의 예를 들자면, 1990년대 카지노 산업에서 들어오는 세수는 마카오 정부 재정수입의 50% 이상을 충당했다. 2007년에도 카지노 산업이 재정수입의 70% 이상을 충당했다. 2007년 재정수입은 총 51억 미국달러였는데, 이 중

카지노 산업의 재정수입액은 36억 미국 달러였다. 2002년에는 4월, 카지노 영업권 독점화 종식 및 2003년 중국인 개인관광객 허용으로 카지노 산업이 급속 발전했다. 2007년 기준, 중국은 49개 도시에서 마카오의 자유 관광이 허용되었다. 기존의 카지노 사업 독점권자였던 스탠리 호(Stanley Ho)의 STDM 외에도, 미국 라스베이거스에 기반을 둔 윈 리조트(Wynn Resorts)에도 라이선스를 부여했다. 또 미국계 베네시안(Venetian)과 홍콩 및 마카오 기업인들 간의 합작업체인 갤럭시 카지노(Galaxy Casino) 등 5개 업체에 추가로 라이선스를 부여했다. 2007년에만 메머드급 카지노인 마카오계의 '그랜드 리스보아(Grand Lisboa)'와 미국계인 '베네시안 호텔 카지노(Venetian Hotel Casino)'가 개장했다. 뿐만 아니라, 마카오-오스트레일리아계인 '크라운 마카오(Crown Macau)', 마카오-미국계인 'MGM 그랜드 마카오(MGM Grand Macau)'가 개장함에 따라 마카오는 세계 카지노 업계의 각축장으로 부상했다.

마카오 카지노 산업은 2006년 69억 달러의 매출액을 기록해, 65억 달러를 기록한 라스베이거스를 능가하는 카지노 도시로 성장했다. 이어 2007년에는 카지노 수입이 100억 달러를 초과하는 등 급성장하고 있다. 마카오의 카지노 세금은 순매출의 40%를 징수하며, 이를 통해 마카오 국민들에 환원하여 복지 및 운영에 큰 힘이 되고 있으며, 자국민을 대상으로 직원 채용을 우선시 하기 때문에 고용창출 효과에도 막대한 영향을 끼치고 있다.

**(홍콩·마카오 개황, 2009. 8., 외교부)**

우리나라 역시 현재 인천 영종도에 대형 복합 리조트 카지노를 운영하기 위해 현재 공사 중에 있으며, 2017년 4월부터 순차적으로 오픈 할 예정에 있다. 이 부분 역시 우리나라 관광산업에 막대한 영향을 미칠 것으로 예상하고 있으며, 현재의 청년실업 문제에도 긍정적인 영향을 미칠 것으로 보인다. 물론 카지노의 부작용을 무시할 순 없으나 긍정적인 영향 역시 가볍게 볼 순 없다. 부가가치가 큰 사업인 만큼 그 부작용에 대한 보안책 마련을 잘해 나간다면 현재 어려움에 빠진 국내 관광의 큰 견인차 역할을 할 것이라 예상해 본다.

## 2-2.국내 카지노 현황

우리나라 카지노 산업 발전과정을 살펴 보면 1961년 11월에 카지노 설립의 법적 근거가 최초로 법률 제정되었고(복표발행현상기타사행행위단속법), 1962년 9월에 동법 개정을 통해 외국인을 상대로 하는 오락시설로 외화획득에 기여 할 수 있음이 인정될 때 이를 허가 할 수 있다고 개정 되었다. 1967년에 대한민국 최초로 인천 올림포스호텔 카지노가 개설되었고, 그 다음해 주한 외국인 및 외국인 관광객 전용인 파라다이스 워커힐 호텔 카지노가 개장 하였다. 그 후 1970년대 주요 관광지에 확산되어 속리산 관공호텔 카지노(1971년), 제주칼호텔 카지노(1975년), 파라다이스비치호텔 카지노(1978년), 코오롱 관광호텔 카지노(1979년)이 개장하였으며, 1980년대에는 설악파크호텔 카지노(1980년)가 강원도에 최초 개설, 제주하얏트 호텔 카지노(1985년), 제주그랜드호텔 카지노(1990년), 제주남서울호텔 카지노, 제주서귀포칼 호텔 카지노, 제주오리엔탈호텔 카지노, 제주신라호텔 카지노 순으로 개장하였다. 그 후 1994년 8월 '관광진흥법'이 개정되어 카지노 산업을 관광산업으로 규정하였고, 주무부서가 교통부(현 건설교통부)에서 문화체육부로 이관, 문화체육부 관광국에서 허가 및 운영, 감독 을 맡게 되었다. 대한민국에선 최초로 내국인 출입허용 카지노인 강원랜드가 2000년 10월 에 소규모로 2003년 3월엔 메인카지노를 개장하여 현재 명실상부 대한민국 최고의 매출을 올리고 있으며 스키장, 콘도, 위락시설등을 갖춰 다양한 복합 리조트식 카지노를 운영 중에 있다. 그리고 2005년 1월에는 한국관광공사 자회사인 (주)그랜드코리아레저의 세븐럭 카지노 가 서울 2곳, 부산 1곳을 개장하여 현재 17개 (내.외국인 카지노) 카지노가 운영 중 이며 절반 가까운 카지노가 제주도(8개)에 밀집되어 있는 실정이다. <표1>를 보면 서울권 카지노(3 곳)가 16개업체(강원랜드 제외) 중 매출의 61.7% 차지하며 제주권 카지노는 16.8%의 매출 점유율을 차지하고 있다.이렇듯 제주는 현재 시장규모가 작은 반면 8개 업체가 밀집되어 있어 경쟁 구도가 심각한 상황이다. 때문에 과열 경쟁이 심화되어 그 관리 감독 및 제주 카지노를 발전 방안을 모색하기 위한 제주특별자치도 내 카지노감독과 가 2016년 7월에 출범하였다. 허나 이는 관광진흥 기금 및 감독체제의 강화로 인하여 제주 카지노의 운영을 더욱 힘들게 하고 있다는 적지 않은 의견도 반발하고 있다.

(단위 : 명, 백만원, m<sup>2</sup>)

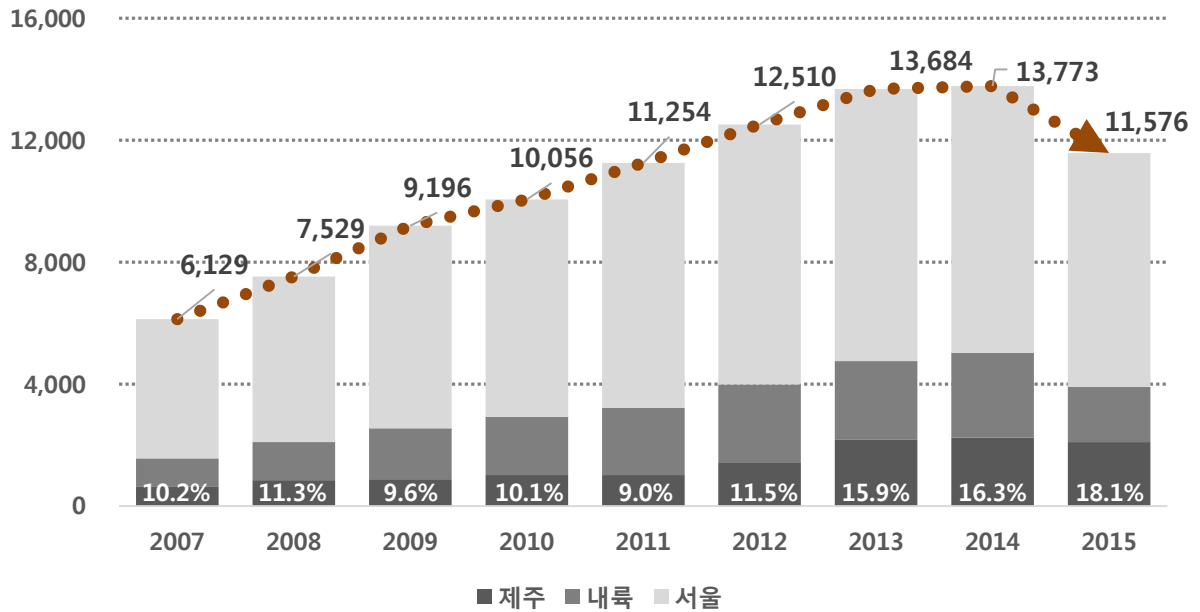
시.도	업소명 (법인명)	허가일	운영형태 (등급)	직원 수	2015 매출액	2015 입장객	허가 면적
서울	워커힐카지노 【(주)파라다이스】	68.03.05	임대 (특1)	923	355,454	572,959	3,970.97
	세븐럭카지노 서울강남코엑스점 【그랜드코리아레저(주)】	05.01.28	임대 (컨벤션)	809	215,275	470,735	6,093.57
	세븐럭카지노서울강북힐튼점 【그랜드코리아레저(주)】	05.01.28	임대 (특1)	493	196,215	919,799	2,811.94
부산	세븐럭카지노 부산롯데점 【그랜드코리아레저(주)】	05.01.28	임대 (특1)	320	85,157	210,276	2,554.50
	부산카지노지점 【(주)파라다이스】	78.10.29	임대 (특1)	352	73,151	90,098	2,283.50
인천	인천카지노 【(주)파라다이스세가사미】	67.08.10	임대 (특1)	412	9,510	48,526	1,703.57
강원	알펜시아카지노 【(주)지바스】	80.12.09	임대 (특1)	8	185	3,893	689.51
대구	인터볼고대구카지노 【(주)골든크라운】	79.04.11	임대 (특1)	171	13,141	69,483	3,473.37
제주	더케이제주호텔카지노 【(주)엔에스디영상】	75.10.15	임대 (특1)	158	17,805	30,734	2,359.10
	제주카지노지점 【(주)파라다이스】	90.09.01	임대 (특1)	252	49,502	45,702	2,756.76
	마제스타카지노 【(주)마제스타】	91.07.31	임대 (특1)	264	40,175	36,597	2,886.89
	로얄팔레스카지노 【(주)건하】	90.11.06	임대 (특1)	199	16,305	18,362	1,353.18
	파라다이스카지노 제주롯데【(주)두성】	85.04.11	임대 (특1)	189	36,748	33,461	1,205.41
	제주썬카지노 【(주)지앤엘】	90.09.01	직영 (특1)	217	4,127	19,559	2,802.09
	랜딩카지노 【그랜드익스프레스코리아(주)】	90.09.01	임대 (특1)	230	32,204	15,474	803.30
	골든비치카지노 【(주)골든비치】	95.12.28	임대 (특1)	157	12,713	27,962	1,528.58
<b>16개 업체(외국인 대상)</b>			<b>직영 : 1 임대 : 15</b>	<b>5,154</b>	<b>1,243,268</b>	<b>2,613,620</b>	<b>39,276.24</b>
강원	강원랜드 카지노 【(주)강원랜드】	00.10.12	직영 (특1)	3,672	1,561,145	3,133,391	12,792.95
<b>17개 업체(내.외국인 대상)</b>			<b>직영 : 2 임대 : 15</b>	<b>8,826</b>	<b>2,804,413</b>	<b>5,747,011</b>	<b>52,069.19</b>

<표1> 우리나라 카지노 현황 자료 : 문화체육관광부, 관광동향에 관한 연차 보고서 (2015년)

<표1>에서도 보여지듯이 제주 카지노 8곳을 매출을 합쳐도 전체(강원랜드 제외)카지노의 약 17%, 서울의 약 27% 수준에 그친다. 이렇게 시장규모가 작은 반면 8개가 카지노가 밀집되어 있어 제주 카지노 시장은 예년부터 많은 어려움을 겪고 있다.



국내 카지노 매출 추이\_강원랜드 제외 (단위 :억원)



<그림2> 연도별 카지노 매출 추이 : 문화체육관광부, 관광동향에 관한 연차 보고서 (2015년)

제주는 적은 매출 시장 규모(점유율: 약15%) 대비 많은 업체(8개)가 밀집되어 있어 업체간 과열 경쟁이 점점 심화되어 가고 있다. 2014년 이후부터 매출이 하락세로 접어들었는데 중국 정부의 현지 카지노 마케팅 단속을 강화하였기 때문이며 이는 2015년 메르스 여파 등으로 인해 현재까지도 하락세를 면치 못하고 있다. 더욱이 업체들간의 과다 경쟁은 원가 상승으로 이어지고 있어 영업이익률은 더욱 떨어질 것으로 전망하고 있다. 저 매출 고 비용의 경영은 소규모 영세 카지노 업체들로 이루어진 제주 지역 카지노에 커다란 악영향을 끼칠 뿐이다. 따라서 고객관리를 통한 특화된 타겟 마케팅 및 비용 절감을 위한 효율적인 관리가 무엇보다도 필요한 시기이며, 이를 위해서 데이터 분석 및 관리, 운영에 보다 능동적이고 적극적일 필요가 있다. 일환으로 데이터 마이닝을 이용한 CRM분석을 통해 보다 효율적인 매출 및 고객 관리를 통하여 수익성 향상을 도모해야 할 것이다.

## 2-3. 데이터마이닝기법적용필요성

데이터 마이닝이란 데이터 베이스 내에서 어떠한 방법(순차 패턴, 유사성 등)에 의해 관심 있는 지식을 찾아내는 과정. 데이터 마이닝은 대용량의 데이터 속에서 유용한 정보를 발견하는 과정이며, 기대했던 정보뿐만 아니라 기대하지 못했던 정보를 찾을 수 있는 기술을 의미한다. 데이터 마이닝을 통해 정보의 연관성을 파악함으로써 가치있는 정보를 만들어 의사 결정에 적용함으로써 이익을 극대화시킬 수 있다. 기업이 보유하고 있는 일일 거래 데이터, 고객 데이터, 상품 데이터 혹은 각종 마케팅 활동의 고객 반응 데이터 등과 이외의 기타 외부 데이터를 포함하는 모든 사용 가능한 근원 데이터를 기반으로 감춰진 지식, 기대하지 못했던 경향 또는 새로운 규칙 등을 발견하고, 이를 실제 비즈니스 의사 결정 등을 위한 정보로 활용하고자 하는 것. 데이터 마이닝의 적용 분야로 가장 대표적인 것은 데이터 베이스 마케팅이다. (컴퓨터인터넷IT용어대사전, 2011)

### 분석에 사용된 데이터 마이닝

- 다중회귀분석[multiple regression analysis]  
종속변수(고객별매출액, 기대수익, RFM, 고객생애가치)에 각 이산형 독립변수(평균배팅액, 참여율, 핸드수, 게임시간, 게임일수 등)들의 결정계수( $R^2$ ) 및 영향을 미치는 정도, 유의한 정도를 파악한다.
- 상관분석[correlation analysis]  
종속변수(고객별 매출액, 기대수익, RFM, 고객생애가치)간 상관계수 및 유의확률을 추출하여 연관성 정도를 분석한다.
- 의사결정나무[decision Tree]  
CART알고리즘을 이용하여 기존 가치 종속 변수(고객별 매출액), 신규 종속 변수(기대수익, RFM, 고객생애가치)를 특정 변수 기준으로 분류 후 정확도(Accuracy), 오분류율(Error rate), 민감도(Sensitivity)에 대한 비교 분석을 한다.
- 로지스틱 회귀분석[logistic Regression Analysis]  
의사결정나무에 의해 분류된 노드중에서 표본 샘플을 추출 하고, 각 변수별로 로지스틱 회귀분석 후 각 종속변수 별로 점수화를 하여 가치 고객과 비가치 고객을 분류해 보고 ROC곡선을 통해 적합도 검증을 해보도록 하겠다.

### 3.데이터 마이닝 적용 방안

#### 3-1. 주요변수 결정 및 관련 데이터 생성

P사의 데이터베이스에 있는 데이터를 대상으로 분석에 필요한 변수를 산출해 내는 데이터 웨어 하우스를 구축하여 고객 분석을 위한 데이터마이닝에 이용한다. 분석에 필요한 변수로는 '기대수익', 'RFM', '고객생애가치(CLV)'등을 생성하고 그 외 지역, 국가 등은 데이터베이스 상에 기본정보를 응용한다.

- ① 기대수익 : 고객의 게임내용을 토대로 기대 할 수 있는 수익. 다음과 같은 공식으로 데이터를 산정한다. (카지노의 특성상 이익과 손실이 발생할 수 있기 때문에 고객의 매출(구매도)는 기대수익으로 대체한다.)

기대수익 = 평균 베팅액 × 게임참여율 × 게임 시간 × 핸드수 × House Edge

※용어 설명 : House Edge - 이론적으로 알려진 종목별 카지노의 승률

- ② RFM : 카지노 고객 대상으로 얼마나 최근에 방문했으며(Recency), 얼마나 자주 방문했으며(Frequency), 얼마나 많은 게임을 했는지(Monetary) 를 중심으로 고객 별 점수를 산정한다. 중요도 별로 비율을 배정한다. (R=10%, F=10%, M=80%)
- ③ 고객생애가치 : 단순 매출액으로 평가하는 현재 문제점을 보완하여 향후 잠재 미래가치를 평가 한다. 고객생애가치 판단을 위해 **공헌마진, 활동확률, 할인율**을 이용하여 산정한다.

#### ③-① 공헌마진

고객이 처음 거래를 시작한 시점부터 현재까지 그 고객이 기여한 총가치

$$\text{공헌마진 측정 모델 : } CM_i = \sum_{n=1}^n GC_{in} \times (1+r)^n$$

i = i번째 고객, r = 이자율, n = 고객의 구매시점부터 현재시점까지의 기간 수

$GC_{in}$  = i번째 고객이 n번째 시점에 구매한 거래의 총 기여도(Gross Contribution)

#### ③-② 활동확률

특정시점 t에 고객이 활동할 확률 -> P(Active)

$$\text{활동확률 측정 모델 : } P(\text{Active}) = 1 - \text{expondist}(LPL, \lambda, TRUE)$$

expondist = 지수분포확률값 함수, LPL = 최종구매경과일,  $\lambda$  = 단위 시간당 발생한 구매횟수(1회기준)

③-③ 할인율

미래에 발생하게 될 고객가치를 현재 가치로 환산하기 위해 필요한 비율

$$\text{할인율의 모델} : \left( \frac{1}{1+r} \right)^n : \text{이자율의 역수}$$

③-④ 고객 생애가치 측정 모델

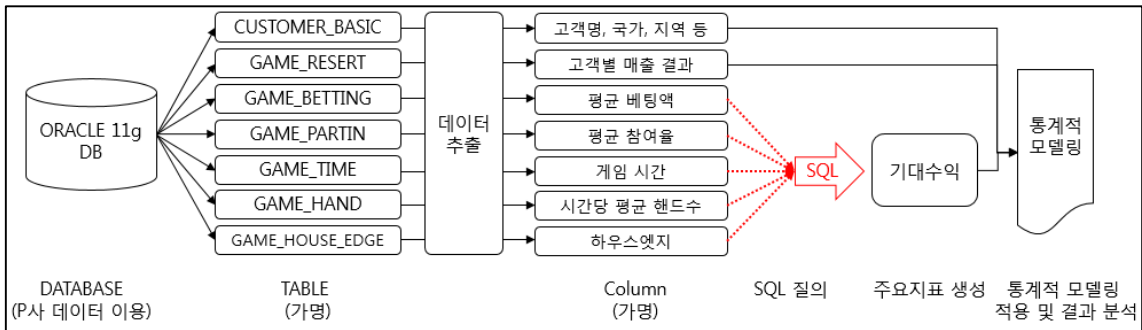
$$CLV_i = \sum_{t=1}^t CM_t \times P_t \times \alpha_t$$

CM : 공헌마진, P : 활동확률, α : 할인율, i : i번째 고객, t : 시간당위

(김주성, 박유라, 2011)

1) 기대수익

① 먼저 주요 지표 데이터 생성을 위한 세부 계획을 수립한다.



<그림3> 데이터 생성 관련 프로세스 : 통계적 모델 적용을 위한 변수를 생성함.(기대수익)

② SQL 작성 총베팅액 & 기대수익은 상단에 명시한 기준에 맞게 질의한다.

```

SELECT A.CUSTOMER_NUM AS '고객 번호',
       A.CUSTOMER_NAME AS '고객 이름',
       A.NATION_CODE AS '국가코드',
       A.REGION_CODE AS '지역코드',
       B.SALE_RESERT AS '고객별 매출',
       C.BETTING_AMOUNT AS '평균베팅액',
       D.PARTIN_RATE AS '평균 참여율',
       E.GAMESTART_TIME AS '게임 시작시간',
       F.GAMEEND_TIME AS '게임 종료시간',
       F.HAND_COUNT AS '핸드수',
       G.HOUSE_EDGE AS '하우스엣지',
       ((NVL(C.BETTING_AMOUNT,0) * (NVL(C.PARTIN_RATE,0)/100) * ((E.GAMEEND_TIME - E.GAMESTART_TIME)*24) * NVL(F.HAND_COUNT,0)) AS '총정 총베팅액',
       ((NVL(C.BETTING_AMOUNT,0) * (NVL(C.PARTIN_RATE,0)/100) * ((E.GAMEEND_TIME - E.GAMESTART_TIME)*24) * NVL(F.HAND_COUNT,0) * (HOUSE_EDGE/100)) AS '기대수익'
FROM   CUSTOMER_BASIC A,
       GAME_RESERT B,
       GAME_BETTING C,
       GAME_PARTIN D,
       GAME_TIME E,
       GAME_HAND F,
       GAME_HOUSE_EDGE G
WHERE  A.CUSTOMER_NUM = B.CUSTOMER_NUM
AND    A.CUSTOMER_NUM = C.CUSTOMER_NUM
AND    A.CUSTOMER_NUM = D.CUSTOMER_NUM
AND    A.CUSTOMER_NUM = E.CUSTOMER_NUM
AND    A.CUSTOMER_NUM = F.CUSTOMER_NUM
AND    A.TABLE_CODE = G.TABLE_CODE
AND    B.GAME_DATE = C.GAME_DATE
AND    B.GAME_DATE = D.GAME_DATE
AND    B.GAME_DATE = E.GAME_DATE
AND    B.GAME_DATE = F.GAME_DATE
AND    B.SEQUENCE = C.SEQUENCE
AND    B.SEQUENCE = D.SEQUENCE
AND    B.SEQUENCE = E.SEQUENCE
AND    B.SEQUENCE = F.SEQUENCE
AND    A.VALIDATION_YN='Y'
AND    B.VALIDATION_YN='Y'
AND    C.VALIDATION_YN='Y'
AND    D.VALIDATION_YN='Y'
AND    E.VALIDATION_YN='Y'
AND    F.VALIDATION_YN='Y'
AND    G.VALIDATION_YN='Y';
    
```

<그림4> SQL: SQL쿼리를 통해 P사의 데이터베이스 내에서 필요한 데이터를 추출함 (테이블 명과 컬럼명은 보안을 위한 가명 사용)

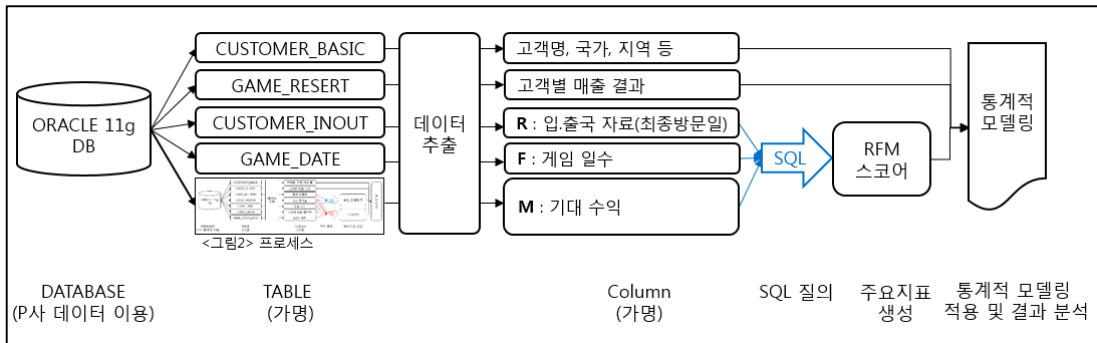
③ 통계적 모델링 적용을 위한 기초 데이터 생성 및 데이터에 대한 오류 검증을 한다.

고객번호	고객명	국가코드	지역코드	고객별매출	평균베팅액	평균참여율	게임시간	평균핸드수	하루스렛지	총베팅액	기대수익
0496037	ZI NIKG	중국	국제	3000000	3000000	50	0.08	7	1.23	10500000	193725
0150856	WIKAVV	중국	화동	10000000	10000000	50	0.04	4	1.23	20000000	246000
0071129	BWIA	중국	화북	1000000	1000000	90	0.08	2	1.23	1800000	22140
0525153	YHWAJI	일본	동경	100000	100000	90	0.07	1	1.23	90000	1107
0491047	MSHCII	일본	기타	-40000	20000	95		61	1.28	1159000	3115
0145346	YIITHO	일본	동경	500000	200000			61	1.28	10980000	27973
0071129	BWIA	중국	화북	1800000	500000		0.04	2	1.23	800000	9840
0489876	WFIQU	중국	기타	200000	50000	80	0.06	30	2.01	3600000	6512
0519551	LIIFZ	중국	기타		30000	95	0.07	30	2.01	855000	6162
0132285	MHUURU	일본	기타	950000	200000	90	0.14	61	1.28	10980000	19676
0322618	LIZNHG	중국	기타	-50000	100000	100	0.07	30	2.01	3000000	29264
0513623	LINI	중국	기타	150000	0	0	0	0	0	0	0
0119712	PKIC	한국	국제	5000000	1000000	70	0.17	17	1.23	11900000	146370
0514495	HIGYC	브라질	기타	60000	50000	50	0.07	1	1.23	25000	461
0492980	LIJNHU	중국	기타	80000	50000	30	0.03	1	1.23	15000	277

<그림5> 샘플데이터 : 변수로 사용될 데이터 추출. (총베팅액, 기대수익)

## 2) RFM

① 먼저 주요 지표 데이터 생성을 위한 세부 계획을 수립한다.



<그림6> 데이터 생성 관련 프로세스 : 통계적 모델 적용을 위한 변수를 생성함.(RFM)

② RFM 스코어를 추출함에 있어 다음과 같은 차트에 의한 기본 점수로 계산한다.

점수 기준은 각항복면 상위 10% 단위 비율로 계산된 자료이다.

R (최근 방문일)		F (게임 일수)		M (기대수익)	
90일 이내	10	70일 이상	10	5 억원 이상	10
180일 이내	9.6	60일 이상	9.6	1 억원 이상	9.6
270일 이내	8.2	45일 이상	8.2	8천만원 이상	8.2
360일 이내	6.7	30일 이상	6.7	6천만원 이상	6.7
450일 이내	5.3	20일 이상	5.3	4천만원 이상	5.3
540일 이내	3.9	10일 이상	3.9	2천만원 이상	3.9
630일 이내	2.4	5일 이상	2.4	1천만원 이상	2.4
그 외	1	그 외	1	그 외	1
<b>10%</b>		<b>10%</b>		<b>80%</b>	

<표2> RFM기준표 : 최근 2년간 데이터 기준. 세가지 항목을 값을 더하여 고객별 RFM 가치도를 산정.



② SQL 작성 고객생애가치산출에 필요한 항목별로 질의 한다.

```

기본지표 {
    SELECT AA.CUSTOMER_NUM AS '고객 번호',
           AA.CUSTOMER_NAME AS '고객 이름',
           AA.NATION_CODE AS '국가코드',
           AA.REGION_CODE AS '지역코드',
           SUM(NVL(BB.SALE_RESEPT, 0)) AS '고객별 매출',
           COUNT(DD.GAME_DATE) AS '게임일수',
           SUM(EE.EXPECTATION_PROFIT) AS '기대수익합계',
           SUM(EE.EXPECTATION_PROFIT) * SUM(POWER(1+(1-평균 이자율/12), 12)) AS '공헌마진',
           SUM(EE.EXPECTATION_PROFIT) * SUM(POWER(1+(1-평균 이자율/12), 12)) AS '활인율',
           (SELECT POWER(1/(1+(1-평균 이자율/12), 12) FROM DUAL) AS '활인율',
           '공헌마진' * '활동확률' * '활인율' AS '고객생애가치'
    }
    
```

<그림10> SQL: SQL쿼리를 통해 P사의 데이터베이스 내에서 필요한 데이터를 추출함. Table이하 절은 생략 (테이블 명과 컬럼명은 보안을 위한 가명 사용)

④ 통계적 모델링 적용을 위한 기초 데이터 생성 및 데이터에 대한 오류 검증을 한다.

고객번호	고객명	지역	국가	고객별매출	매출순위	기대수익	기대수익 순위	공헌마진	활동확률	활인율	고객생애가치	CLV 순위
0078372	HL JI JG	후쿠오카	JPN	6,034,700,000	4	6,472,629,000	2	6,648,049,724	26.290%	95.711%	1,672,827,545	1
0230009	YU V AI IV	화북	CHN	3,643,320,000	9	4,602,629,114	3	4,664,664	36.447%	95.711%	1,646,288,483	2
0154457	LIL JH J JH	화북	CHN	3,309,120,000	11	2,962,390,687	11	2,962,390,687	53.857%	95.711%	1,581,706,842	3
0298497	JH JH ZI	상동북	CHN	-412,100,000	10833	2,336,144,278	11	2,365,955,675	63.863%	95.711%	1,446,149,792	4
0157234	WI JI BI	산동	CHN	4,345,358,000	5	4,751,041,180	9	4,751,041,180	77.508%	95.711%	1,257,978,306	5
0123911	MI IA S	화북	CHN	3,751,540,000	7	3,751,540,000	5	2,686,260,673	48.867%	95.711%	1,256,395,974	6
0075668	SA JI AS KI	동경	JPN	1,842,500,000	15	1,751,041,180	7	1,786,695,599	72.140%	95.711%	1,233,635,190	7
0260280	YA AI A' J JRI	오사카	JPN	1,496,700,880	15	1,496,700,880	12	1,519,229,443	62.543%	95.711%	909,417,822	8
0094809	ZH V ZE U V	하동북	CHN	1,496,700,880	7	1,261,883,719	17	1,298,889,513	71.827%	95.711%	892,937,014	9
0081123	LI EI E	화북	CHN	397,683,000	141	1,048,457,888	23	1,080,233,021	80.353%	95.711%	830,771,787	10
0207434	HC IE AI JI	후쿠오카	JPN	1,378,094,900	26	1,333,536,904	13	1,357,139,415	62.047%	95.711%	805,950,111	11
0469440	KA S UL	국제	CHN	777,150,000	54	979,403,490	25	1,000,039,543	53.431%	95.711%	511,411,409	12
0270903	IN IA A AI HISA	나고야	JPN	417,806,600	128	664,524,018	45	684,578,185	74.979%	95.711%	491,270,355	13
0132898	KA AM . C N	동경	JPN	-294,837,300	10821	591,191,392	53	604,862,062	78.895%	95.711%	456,737,011	14
0154074	PIA J NH H J	하동북	CHN	1,216,820,000	31	1,624,524,372	10	1,674,326,234	26.536%	95.711%	425,245,351	15
0277741	PIA J NH H J	하동북	CHN	1,000,052,000	42	1,017,202,832	24	1,044,638,128	40.389%	95.711%	403,819,303	16
0266451	ZH J J J J G AI	상동북	CHN	-1,884,600,000	10844	761,112,087	36	783,565,761	52.221%	95.711%	391,636,315	17

<그림11> 샘플데이터 : 변수로 사용될 데이터 추출. (고객생애가치)

## 4.분석 결과

### 접목 및 결과 도출 적합성 여부 검토

#### 4-1 다중 회귀분석을 이용한 '고객별 매출액'의 분석

종속변수를 기존의 가치 척도인 고객별 매출액, 독립 변수는 평균배팅액, 참여율, 게임시간, 게임일수, 핸드수로 설정 후 회귀분석을 실행한다.

모형	R	R제곱	수정된 R제곱	추정값의 표준오차
1	.567	.321	.321	107158697.2

[표 3] 다중회귀 분석 모형 요약 (고객별 매출액)

[표 3]에서 보여지듯 R제곱의 값이 .321 이므로 설명변수는 약 32.1%를 설명한다고 할 수 있다.

모형	제곱합	자유도	평균제곱	F	유의확률
회귀모형	5.138E+19	5	1.028E+19	894.823	.000
잔차	1.084E+20	9444	1.148E16		
합계	1.598E+20	9449			

[표 4] 다중 회귀모형의 분산분석 (고객별 매출액)

F값에 대한 유의확률(p값)은 .000 이므로 0.05보다 작기 때문에 회귀모형이 적합하다고 할 수 있다.

종속변수	독립변수	회귀계수	표준화 회귀계수	t	p	공선성 통계량
매출액	평균배팅액	24.831	.494	58.244	.000	1.003
	참여율	208930.845	.106	4.357	.000	8.201
	핸드수	5648.380	.141	6.886	.000	5.827

[표 5] 다중 회귀모형의 계수 (고객별 매출액)

공선성 통계량값은 세변수 모두 10미만이므로 다중공선성에는 문제가 없고, 유의 확률은 모두 0.05보다 작기 때문에 종속변수에 유의한 영향을 미친다고 볼 수 있다. 베타값은 모두 양의 값이기 때문에 매출액에 플러스 영향을 미치며, 특히 평균배팅액이 가장 큰 영향을 미친다고 할 수 있다.



## 4-2 다중 회귀분석을 이용한 '기대수익'의 분석

종속변수를 신규 가치 척도인 기대수익, 독립변수는 평균배팅액, 참여율, 게임시간, 게임일수, 핸드수로 설정 후 회귀분석을 실행한다.

모형	R	R제곱	수정된 R제곱	추정값의 표준오차
1	.735	.541	.540	67646943.28

[표 6] 다중회귀 분석 모형 요약 (기대수익)

[표6]에서 보여지듯 R제곱의 값이 .541 이므로 설명변수는 약 54.1%를 설명한다고 할 수 있다.

모형	제곱합	자유도	평균제곱	F	유의확률
회귀모형	5.085E+19	5	1.017E+19	2222.374	.000
잔차	4.322E+19	9444	4.576E+15		
합계	9.407E+19	9449			

[표 7] 다중 회귀모형의 분산분석 (기대수익)

[표 7]에 보여지듯 F값에 대한 유의확률(p값)은 .000 이므로 0.05보다 작기 때문에 회귀모형은 적합하다고 할 수 있다.

종속변수	독립변수	회귀계수	표준화 회귀계수	t	p	공선성 통계량
기대수익	평균배팅액	24.465	.635	90.905	.000	1.003
	참여율	328946.371	.217	10.866	.000	8.201
	핸드수	7663.589	.249	14.800	.000	5.827

[표 8] 다중 회귀모형의 계수 (기대수익)

공선성 통계량값은 세변수 모두 10미만이므로 다중공선성에는 문제가 없으며 유의 확률은 모두 0.05보다 작기 때문에 종속변수에 유의한 영향을 미친다. 베타값은 모두 양의 값이기 때문에 기대수익에 플러스 영향을 미치며, 기대수익 또한 평균배팅액이 가장 큰 영향을 미치고 있음을 알 수 있다.

### 4-3 다중 회귀분석을 이용한 'RFM'의 분석

종속변수를 가치 척도인 RFM 스코어, 독립변수는 평균배팅액, 참여율, 게임시간, 게임일수, 핸드수로 설정 후 회귀분석을 실행한다.

모형	R	R제곱	수정된 R제곱	추정값의 표준오차
1	.692	.479	.478	11.843

[표 9] 다중회귀 분석 모형 요약 (RFM)

[표 9]에서 보여지듯 R제곱의 값이 .479 이므로 설명변수는 약 47.9%를 설명한다고 할 수 있다.

모형	제곱합	자유도	평균제곱	F	유의확률
회귀모형	1215645.482	5	243129.096	1733.596	.000
잔차	1324478.502	9444	140.246		
합계	2540123.985	9449			

[표 10] 다중 회귀모형의 분산분석 (RFM)

[표 10]에 보여지듯 F값에 대한 유의확률(p값)은 .000 이므로 0.05보다 작기 때문에 회귀모형은 적합하다고 할 수 있다.

종속변수	독립변수	회귀계수	표준화 회귀계수	t	p	공선성 통계량
RFM	평균배팅액	2.074E-006	.328	44.014	.000	1.003
	참여율	.057	.229	10.777	.000	8.201
	핸드수	.000	.089	4.981	.000	5.827

[표 11] 다중 회귀모형의 계수 (RFM)

공선성 통계량값은 세변수 모두 다중공선성에는 문제가 없으며 유의 확률(p값)은 모두 0.05 보다 작기 때문에 종속변수에 유의한 영향을 미친다. 베타 값은 모두 양의 값을 갖으며 RFM스코어 에 플러스 영향을 미치고 평균배팅액과 참여율이 좋은 영향을 미치고 핸드수 역시 플러스 영향을 미치나 다른 종속 변수에 비해서 RFM 스코어에 큰 영향을 못미치고 있음을 알 수 있다.

### 4-4 다중 회귀분석을 이용한 '고객생애가치'의 분석

종속변수를 신규 가치 척도인 고객생애가치, 독립변수는 평균배팅액, 참여율, 게임시간, 게임일수, 핸드수로 설정 후 회귀분석을 실행한다.

모형	R	R제곱	수정된 R제곱	추정값의 표준오차
1	.729	.531	.531	58614777.96

[표 12] 다중회귀 분석 모형 요약 (고객생애가치)

[표 12]에서 보여지듯 R제곱의 값이 .531 이므로 설명변수는 약 53.1%를 설명한다고 할 수 있다.

모형	제공합	자유도	평균제공	F	유의확률
회귀모형	3.670E+19	5	7.340E+18	2136.430	.000
잔차	3.245E+19	9444	3.436E+15		
합계	6.915E+19	9449			

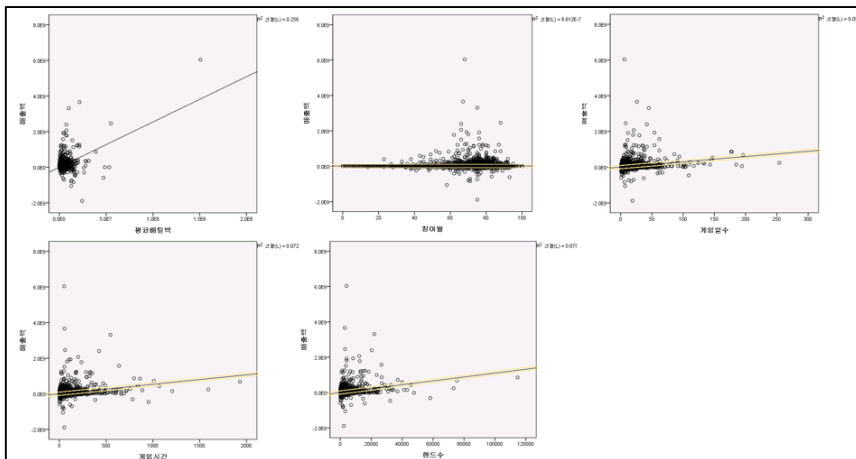
[표 13] 다중 회귀모형의 분산분석 (고객생애가치)

[표 13]에 보여지듯 F값에 대한 유의확률(p값)은 .000 이므로 0.05보다 작기 때문에 회귀모형은 적합하다고 할 수 있다.

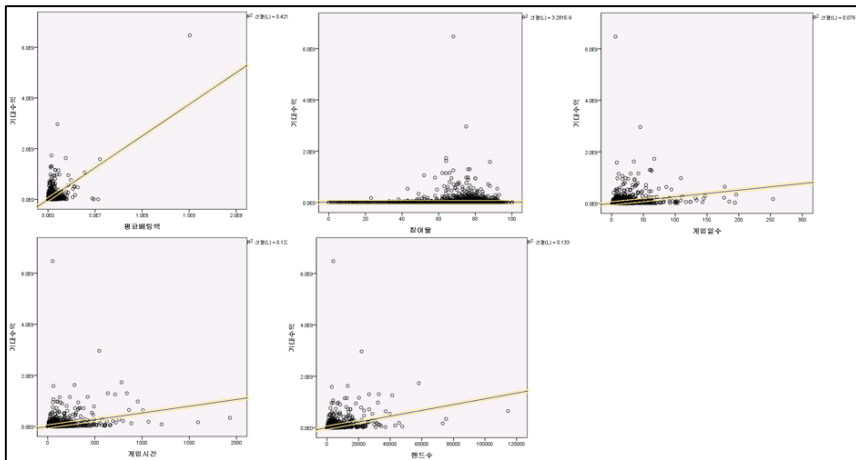
종속변수	독립변수	회귀계수	표준화 회귀계수	t	p	공선성 통계량
고객생애 가치	평균배팅액	19.912	.603	85.386	.000	1.003
	참여율	301650.960	.232	11.500	.000	8.201
	핸드수	7455.835	.283	16.617	.000	5.827

[표 14] 다중 회귀모형의 계수 (고객생애가치)

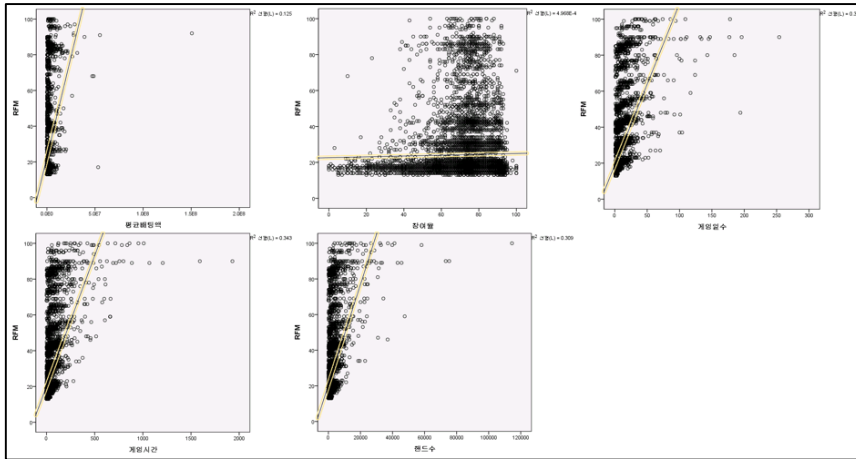
공선성통계량값은 세변수 모두 다중공선성에는 문제가 없으며 유의 확률(p값)은 모두 0.05 보다 작기 때문에 종속변수에 유의한 영향을 미친다. 베타 값은 모두 양의 값을 갖으며 종속변수(고객생애가치)에 플러스 영향을 미치고 평균배팅액이 가장 좋은 영향을 미치고 있다.



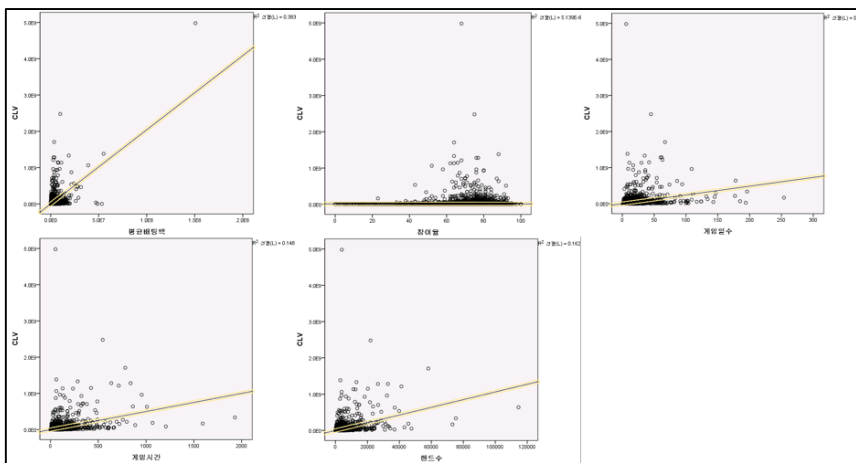
[그림 12] 독립변수별 회귀모형 그래프(고객별 매출액)\_평균배팅액, 참여율, 게임일수, 게임시간, 핸드수



[그림 13] 독립변수별 회귀모형 그래프(기대수익)\_평균배팅액, 참여율, 게임일수, 게임시간, 핸드수



[그림 14] 독립변수별 회귀모형 그래프(RFM스코어)\_평균배팅액, 참여율, 게임일수, 게임시간, 핸드수



[그림 15] 독립변수별 회귀모형 그래프(고객생애가치)\_평균배팅액, 참여율, 게임일수, 게임시간, 핸드수

다중회귀분석의 결과로 보아 기존 종속 변수 인 고객별 매출액과 신규 종속 변수인 기대수익, RFM, 고객 생애 가치 모두 독립변수 평균배팅액, 참여율, 핸드수의 영향을 많이 받으며 모든 독립변수의 유의수준(P-value)이 0.05 이하이므로 종속변수에 유의함을 알 수 있다. 다만 결정계수(R제곱)을 비교하면 고객별 매출액은 .321(32.1%), 기대수익은 .541(54.1%), RFM은 .479(47.9%), 고객생애가치는 .531(53.1%)로서 신규 종속 변수 3가지는 모두 고객별 매출액보다 회귀 직선이 실제 Data를 보다 잘 설명하고 있어 더 적합한 모형이라 할 수 있다.

#### 4-5 상관분석을 이용한 종속 변수 별 연관성 분석

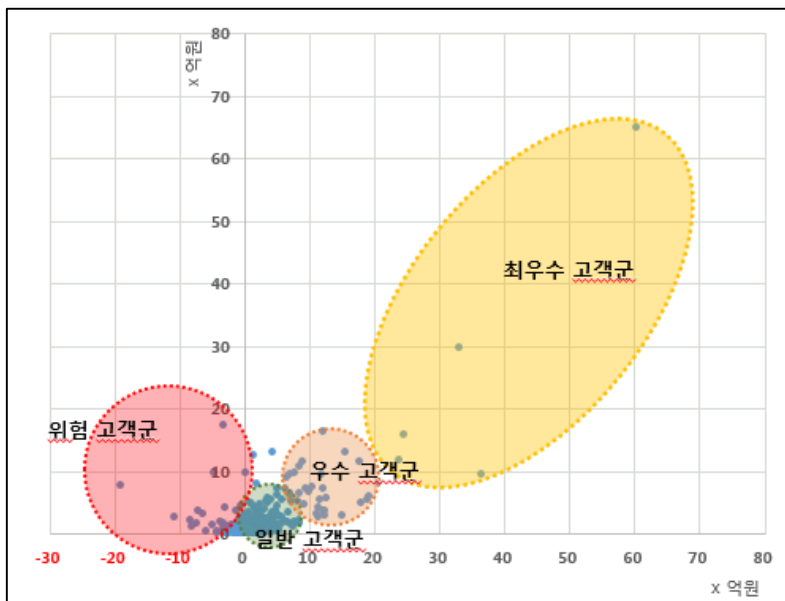
기존 종속 변수인 '고객별 매출액' 과 신규 종속 변수인 '기대수익','RFM스코어','고객생애가치'에 대해서 상관분석을 하고 서로간에 연관성 및 유의한 영향이 있는지에 대해 알아보고자 한다.

		고객별매출액	기대수익	RFM스코어	고객생애가치
고객별매출액	P상관계수	1	.751	.748	.395
	유의확률		.000	.000	.000
기대수익	P상관계수	.751	1	.992	.441
	유의확률	.000		.000	.000
RFM스코어	P상관계수	.748	.992	1	.465
	유의확률	.000	.000		.000
고객생애가치	P상관계수	.395	.441	.465	1
	유의확률	.000	.000	.000	

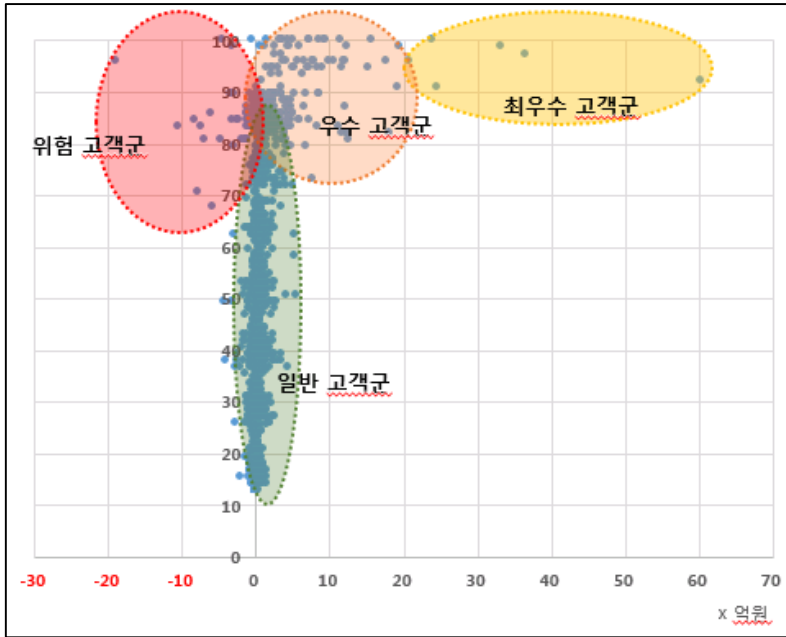
[표 15] 상관분석 결과 (고객별매출액, 기대수익, RFM스코어, 고객생애가치)

[표15]에서 보여지듯 고객별 매출액과 기대수익간 상관계수는 .751, 고객별 매출액과 RFM스코어는간 상관계수는 .748 , 고객별 매출액과 고객생애가치 상관계수는 .395 이다. 또한 기대수익과 RFM스코어 간 상관계수는 .992, 기대수익과 고객생애가치는 .441 이다. 그리고 RFM스코어 고객 생애가치 간의 상관계수는 .465로 고객생애가치와의 상관성이 서로 약하긴 하나 모든 변수들이 비교적 양호한 상관계수를 갖고 있음을 알 수 있다. 이는 이 종속 변수들 간에 연관성이 양호하다고 할 수 있다.

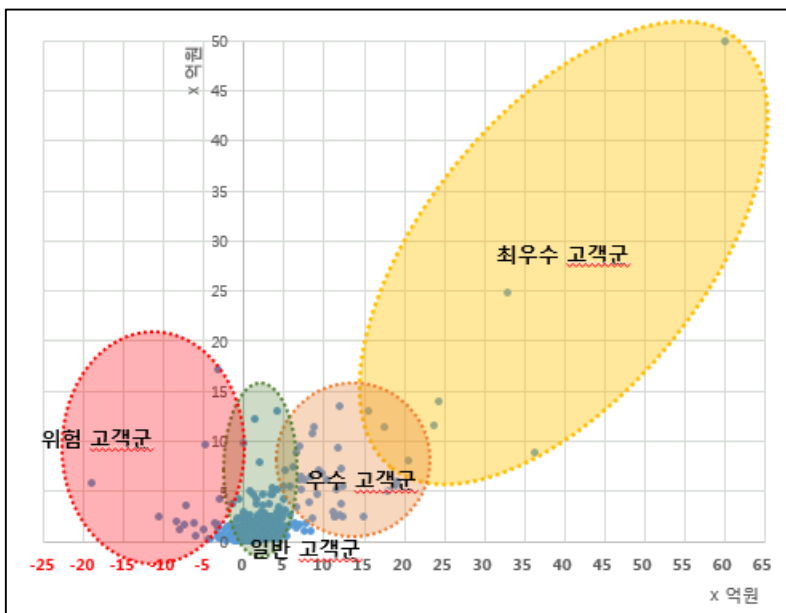
위와 같은 결과를 고객별 매출액을 기준으로 신규 종속 변수를 대입하여 산점도 그래프화를 하고 고객군을 분류 해보도록 하겠다.



[그림 16] 고객별 매출액(y) 과 기대수익(x) 기준 고객 분류



[그림 17] 고객별 매출액(y)과 RFM스코어(x) 기준 고객 분류



[그림 18] 고객별 매출액(y)과 고객생애가치(x) 기준 고객 분류

최우수 고객군과 우수 고객군은 매출 공헌 과 게임 내용 및 방문이 모두 우수하여 지속적인 고객관리가 필요한 고객군이며 일반 고객, 일반 고객군은 매출에 대한 기여도는 떨어 지나 잠재 가능성이 있는 고객군으로 게임에 좀 더 충실할 수 있도록 유도 하여 관리 하여야 하는 고객군이다. 위험 고객군은 오랜기간동안 매출을 올리기 힘든 위험 겜블러 이므로 선별 하여 게임자체를 하는 것 외에 VIP로서 서비스(COMP) 까지 지급하여 관리 할지에 대해서 신중히 생각하여 관리하여야 하는 고객군이다.

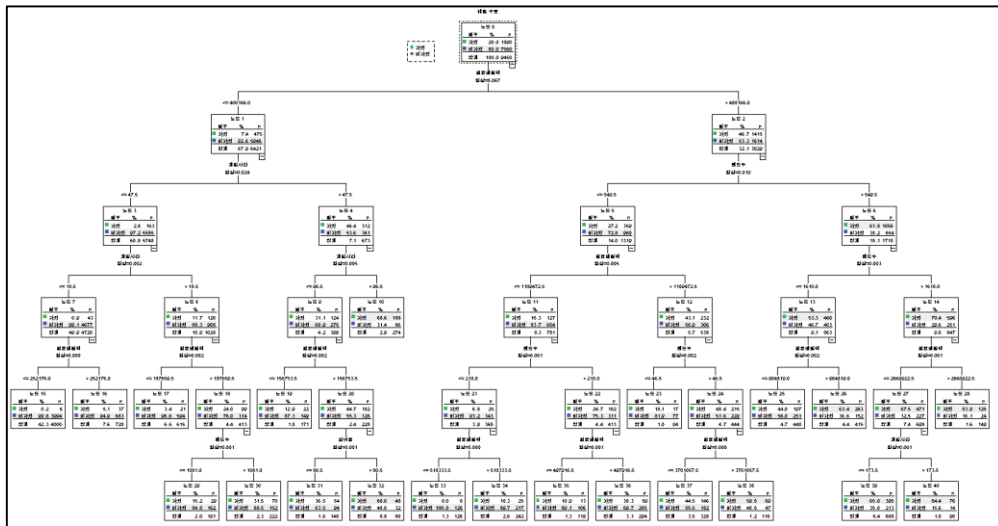
#### 4-6 의사결정나무를 이용한 고객 분류

의사결정나무를 통해 분류 분석 후 기존 종속 변수인 '고객별 매출액'과 신규 종속 변수인 '기대수익', 'RFM', '고객생애가치'의 정확도(Accuracy), 오분류율(Error rate) 및 민감도(Sensitivity)에 대한 비교 분석을 해보고자 한다. 먼저 종속 변수에 대해 상위 30%를 '가치고객군', 그 나머지를 '비가치고객군'으로 분류 후 의사결정트리에 적용 시켰다. 분류 기준은 CART알고리즘을 사용하였다.

변수명	유형	내용
고객 구분	범주형	가치 / 비가치
성별	범주형	남 / 여
연령대	범주형	20대 / 30대 / 40대 / 50대 / 60대 / 70대 / 80대
평균배팅액	연속형	0 ~ 1억5천
참여율	연속형	0 ~ 100 %
게임일수	연속형	1 ~ 254 일
게임시간	연속형	0.01 ~ 1930 시간
핸드수	연속형	1 ~ 114,545 회
미방문일	연속형	29 ~ 394 일
활동 확률	연속형	0 ~ 100 %

[표 16] 분석에 사용된 변수(의사결정나무)

먼저 기존 종속 변수인 고객별 매출에 대한 분류를 해보았다.



[그림 19] 고객별매출 구분에 대한 의사결정나무

분류 결과 총 노드 41개, 터미널 노드 21개, 깊이 5인 나무구조가 형성되었다. 고객별 매출로 인한 분류를 결정짓는 변수로는 평균배팅액, 핸드수, 게임시간, 참여율 이며 특히 평균배팅액이 구분 짓는데 가장 큰 영향을 미치는 것으로 보여진다. 또한 오분류 테이블에 의해 얻어진 결과는 다음과 같다.

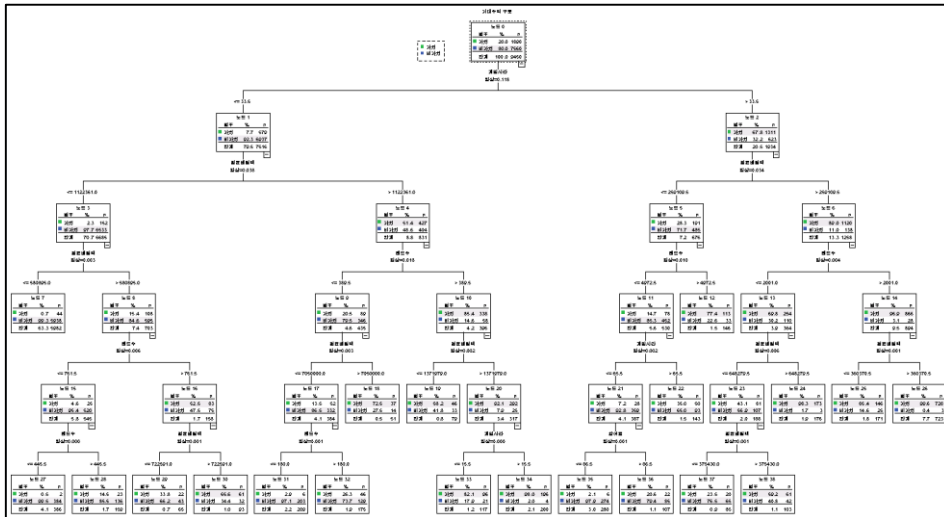
고객별매출액 오분류테이블		예측		계
		가치	비가치	
실제	가치	1164 12.32%	726 7.68%	1890
	비가치	568 6.01%	6992 73.99%	
계		1732	7718	9450

Error Rate = 13.7%      Accuracy = 86.3%  
Sensitivity = 61.6%      Specificity = 92.5%

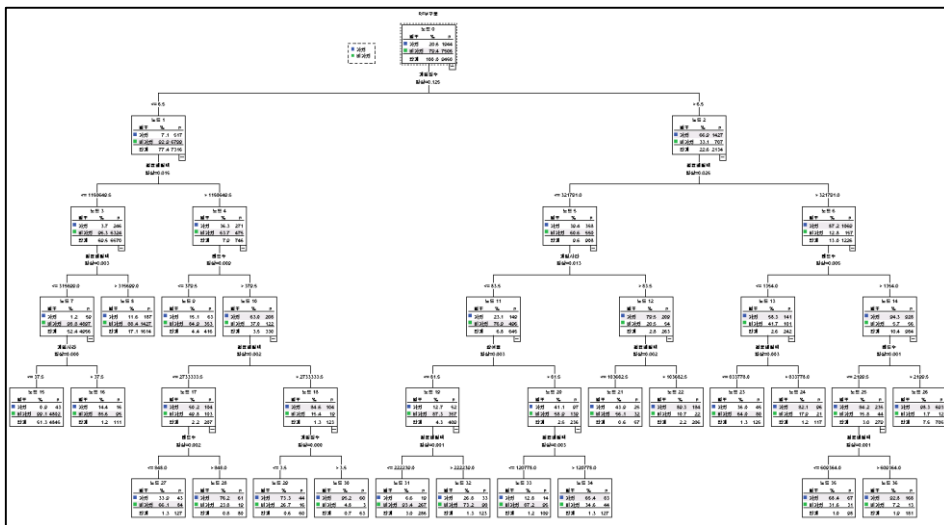
[표 17] 의사결정나무 분석의 오분류 테이블(고객별 매출액)

분석결과 오분류율(Error rate) 과 정확도(Accuracy)는 각각 13.7%, 86.3%으로 비교적 잘 분류되었다. 결과적으로 고객별 매출은 카지노게임의 내용(배팅, 참여, 시간등 )에 영향을 받는것을 알 수 있다. 다음으론 신규 종속 변수인 기대수익, RFM, 고객생애가치 순으로 분류해보도록 하겠다.

신규 종속 변수(기대수익, RFM, 고객생애가치) 에 대한 분류 결과는 다음과 같다.

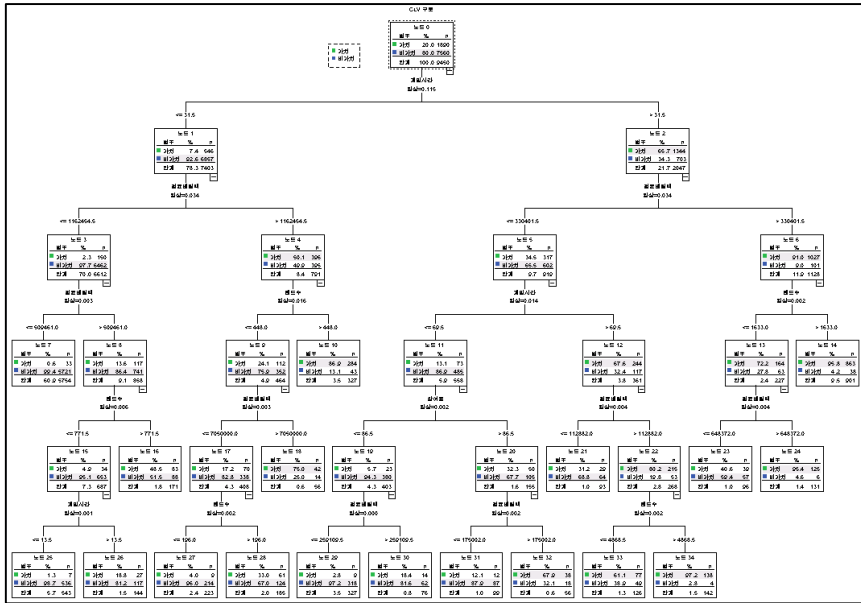


[그림 20] 기대수익 구분에 대한 의사결정나무



[그림 21] RFM 구분에 대한 의사결정나무





[그림 22] 고객생애가치 구분에 대한 의사결정나무

각 분류 결과 기대수익에 의한 의사결정나무는 총 노드 39개, 터미널 노드 20개, 깊이 5인 나무구조가 형성되었고, RFM에 의한 의사결정나무는 총 노드 37개, 터미널 노드 19개, 깊이 5인 나무구조가 형성되었으며 고객생애가치에 의한 의사결정나무는 총 노드 35개, 터미널 노드 18개, 깊이 5인 나무구조가 형성되었다. 세가지 종속 변수 모두가 나무 분류를 결정짓는 변수로는 평균배팅액, 핸드수, 게임시간, 참여율 이며 특히 평균배팅액이 구분 짓는데 가장 큰 영향을 미치는 것으로 보여진다. 고객별 매출액도 마찬가지로 엮지만 성별이나 연령대 보다는 게임의 내용(배팅, 참여도, 시간, 회전수 등)에 의해 가치도가 분류되고 있음을 알 수 있다. 또한 각각의 오분류 테이블에 의해 얻어진 결과는 다음과 같다.

기대수익 오분류테이블		예측		계
		가치	비가치	
실제	가치	1649	241	1890
	비가치	210	7350	
계		1859	7591	9450
Error Rate = 4.8%		Accuracy = 95.2%		
Sensitivity = 87.2%		Specificity = 97.2%		

[표 18] 의사결정나무 분석의 오분류 테이블(고객별 매출액)

기대수익에 의해 분류된 의사결정나무의 오분류 테이블을 분석 결과 오분류율(Error rate)과 정확도(Accuracy)는 각각 4.8%, 95.2%로 고객별매출액에 의한 결과 보다 오분류율은 감소하였고 정확도는 증가하였음을 알 수 있다. 민감도 역시 87.2% 아주 높은 수치를 보여준다.

RFM 오분류테이블		예측		계
		가치	비가치	
실제	가치	1456 15.41%	488 5.16%	1944
	비가치	181 1.92%	7325 77.51%	
계		1637	7813	9450
Error Rate = 7.1%		Accuracy = 92.9%		
Sensitivity = 74.9%		Specificity = 97.6%		

[표 19] 의사결정나무 분석의 오분류 테이블(RFM 스코어)

RFM에 의해 분류된 의사결정나무의 오분류 테이블도 마찬가지로 분석 결과 오분류율(Error rate) 과 정확도(Accuracy)는 각각 7.1%, 92.9%로 기존 종속 변수인 고객별매출액에 의한 결과 보다 오분류율은 감소하였고 정확도는 증가하였음을 알 수 있다. 민감도 역시 74.9%로 전체 적으로 기대수익에 의한 수치보단 근소하게 낮지만 고객별 매출액보다는 높은 수치를 보여준다.

고객생애가치 오분류테이블		예측		계
		가치	비가치	
실제	가치	1567 16.58%	323 3.42%	1890
	비가치	172 1.82%	7388 78.18%	
계		1739	7711	9450
Error Rate = 5.2%		Accuracy = 94.8%		
Sensitivity = 82.9%		Specificity = 97.7%		

[표 20] 의사결정나무 분석의 오분류 테이블(고객생애가치)

고객생애가치에 의해 분류된 의사결정나무의 오분류 테이블 역시 분석 결과 오분류율(Error rate) 과 정확도(Accuracy)는 각각 5.2%, 94.8%로 기존 종속 변수인 고객별매출액에 의한 결과 보다 오분류율은 감소하였고 정확도는 증가하였음을 알 수 있다. 민감도 역시 82.9%로 전체 적으로 고객별 매출액보다는 높은 수치를 보여준다. 전체적으로 기대수익, 고객생애가치, RFM스코어, 고객별 매출액 순으로 높은 적합도를 보이고 있으며, 신규 종속 변수 3가지가 기존 종속 변수인 고객별 매출액 보다 월등히 높은 적합도를 보여 주고 있다.

## 4-7 로지스틱 회귀분석을 이용한 가치 고객 가능성 점수화

[표 16]의 분석에 필요한 변수들에 의해 분류된 각각의 의사결정나무들 중에서 표본 샘플을 추출 하고, 각 가치도 별로 로지스틱 회귀분석 후 각 종속 변수별로 점수화를 하고 ROC곡선을 통해 적합도 검증을 해보도록 하겠다.

### 4-7-1. 고객별 매출에 대한 로지스틱 회귀분석

먼저 고객별 매출에 대해서는 1)평균배팅액이 400,166원을 초과하는 노드 중에서 핸드수가 550회 미만인 노드 와 2)평균배팅액이 400,166원 이하의 노드 중에서 게임시간이 47.5시간을 초과하고 96.5시간 이하의 노드에서의 자료(표본 약 1750개)만을 이용하여 로지스틱 회귀 모형을 구축한다. 분석에 사용할 변수는 의사결정나무 분류시 직접적으로 연관있는 평균배팅액, 핸드수, 게임시간, 참여율을 선택하도록 하겠다. 고객별 매출액에 대한 로지스틱 회귀분석 결과는 다음과 같다.

단계	카이 제곱	자유도	유의확률
1	11.506	8	.175

[표 21] Hosmer와 Lemeshow 검정(고객별 매출액)

종속변수	독립변수	P(유의확률)	Exp(B)
고객별 매출액	평균배팅액	.000	1.000
	참여율	.000	10.708
	핸드수	.190	1.000

[표 22] 방정식에 포함된 변수(고객별 매출액)

분석결과 Hosmer와Lemeshow적합도 검정에서 유의확률이 .175로서 0.05보다 크므로 모형은 적합하다고 할 수 있다. 그리고 각 독립변수별 유의 확률은 평균배팅액과 참여율은 유의 확률 0.05 보다 낮아 고객별 매출에 대해 유의한 영향을 미치나 핸드수는 그렇지 못하고, Exp(B) 값은 독립변수 모두 1을 초과 하여 양의 영향을 미친다. 특히 참여율 같은 경우는 10.708로 가장 큰 영향을 미치는 것으로 분석된다. 또한 고객별 매출액 로지스틱 모형에 의한 오분류 테이블은 다음과 같다.

고객별매출액 오분류테이블		예측		계
		가치	비가치	
실제	가치	25 1.43%	460 26.23%	485
	비가치	20 1.14%	1249 71.21%	
계		45	1709	1754
Error Rate = 27.4%		Accuracy = 72.6%		
Sensitivity = 5.2%		Specificity = 98.4%		

[표 23] 로지스틱 모형의 오분류 테이블(고객별 매출액)

[표 23]에서와 같이 고객별 매출액 로지스틱 모형에 의한 오분류테이블 분석결과 전체적인

적합도는 72.6%로 양호한 편이긴 하나 민감도는 5.2%로 상당히 낮기 때문에 고객별 매출에 대한 가치고객 예측이 비효율적이다. 여기서는 고객별 매출액에 대한 가치 고객 가능성 점수화는 하지 않고, 신규 종속 변수인 기대수익, RFM, 고객생애가치에 대한 가치 고객 가능성 점수화를 하도록 하겠다.

- 가치고객가능성 점수 =  $P(Y = \text{가치}) \times 100$

#### 4-7-2. 기대수익에 대한 로지스틱 회귀 분석

기대수익에 대해서는 1)게임시간이 33.5시간을 초과하는 노드 중에서 평균배팅액이 260,000원 이하의 노드 와 2)게임시간이 33.5시간 이하의 노드 중에서 평균배팅액이 1,120,000원 이하 이고 580,000원 초과하는 노드에서의 자료(표본 약 1,380개)만을 이용하여 로지스틱 회귀 모형을 구축한다. 분석에 사용할 변수는 의사결정나무 분류 시 직접적으로 연관있는 평균배팅액, 핸드수, 게임시간, 참여율을 선택하도록 하겠다. 기대수익에 대한 로지스틱 회귀분석 결과는 다음과 같다.

단계	카이 제곱	자유도	유의확률
1	477.978	8	.552

[표 24] Hosmer와 Lemeshow 검정(기대수익)

종속변수	독립변수	P(유의확률)	Exp(B)
기대수익	평균배팅액	.000	1.000
	참여율	.000	463.414
	핸드수	.009	1.000

[표 25] 방정식에 포함된 변수(기대수익)

분석결과 Hosmer와Lemeshow적합도 검정에서 유의확률이 .552로서 0.05보다 크므로 모형은 적합하다고 할 수 있다. 그리고 각 독립변수별 유의 확률은 평균배팅액과 참여율, 핸드수 모두 유의확률 0.05 보다 낮아 기대수익 대해 유의한 영향을 미치고, Exp(B) 값은 독립변수 모두 1을 초과 하여 양의 영향을 미친다. 특히 참여율 같은 경우는 463.414로 가장 큰 영향을 미치는 것으로 분석된다. 또한 기대수익 로지스틱 모형에 의한 오분류 테이블은 다음과 같다.

기대수익 오분류테이블		예측		계
		가치	비가치	
실제	가치	188 13.62%	88 6.38%	276
	비가치	32 2.32%	1072 77.68%	
계		220	1160	1380
Error Rate = 8.7%		Accuracy = 91.3%		
Sensitivity = 68.1%		Specificity = 97.1%		

[표 26] 로지스틱 모형의 오분류 테이블(기대수익)

[표 26]에서와 같이 기대수익 로지스틱 모형에 의한 오분류테이블 분석결과 전체적인 적합도는 91.3%로 매우 높고, 민감도는 68.1%로 양호하므로 고객별 매출에 대한 가치고객 예측에 비해 상당히 효율적임을 알 수 있다. 로지스틱 회귀모형에서 추정된 확률값을 이용하여 가치고객 가능성에 대한 점수화를 실시한다.

성별	연령대	평균 배팅액	참여율	게임 일수	게임 시간	핸드수	기대수익 구분	기대수익 가치 점수	기대수익 가치 예측
남	60	1,287,000	88	178	862	114,545	가치	100	가치
남	60	3,032,000	79	11	40	1,663	가치	90.06	가치
남	20	4,571,000	80	3	8	408	가치	85.68	가치
남	60	657,000	77	13	74	2,982	가치	78.62	가치
남	40	491,000	75	7	69	2,216	가치	57.38	가치
남	20	972,000	79	7	54	2,592	가치	57.26	가치
남	30	2,253,000	74	9	27	913	가치	56.96	가치
남	40	337,000	90	12	56	1,680	가치	51.85	가치
남	30	584,000	88	9	47	2,552	가치	45.49	비가치
남	50	292,000	44	23	56	2,338	비가치	45.49	비가치
남	50	222,000	86	6	51	4,148	비가치	39.59	비가치
남	60	177,000	90	9	48	3,543	비가치	39.57	비가치

[표 27] 가치고객 가능성 점수 샘플 데이터(기대수익)

비율	점수	인원	기대수익	고객별매출액
가치	50 점▲	15.9%	86.8%	81.2%
비가치	50 점▽	84.1%	13.2%	18.8%

[표 28] 가치, 비가치 고객 비율 정리 (기대수익)

### 4-7-3. RFM에 대한 로지스틱 회귀 분석

RFM에 대해서는 1)게임일수가 6.5일을 초과하는 노드 중에서 평균배팅액이 321,000원 이하의 노드 하위에 게임시간이 83.5시간 이하의 노드 와 2)게임일수가 6.5일 이하의 노드 중에서 평균배팅액이 115,000원을 초과하고 핸드수가 380회 이하의 노드 에서의 자료(표본 약 1,060개)만을 이용하여 로지스틱 회귀 모형을 구축한다. 분석에 사용할 변수는 의사결정나무 분류 시 직접적으로 연관있는 평균배팅액, 핸드수, 게임시간, 참여율을 선택하도록 하겠다. RFM에 대한 로지스틱 회귀분석 결과는 다음과 같다.

단계	카이 제곱	자유도	유의확률
1	245.766	8	.162

[표 29] Hosmer와 Lemeshow 검정(RFM)

종속변수	독립변수	P(유의확률)	Exp(B)
RFM	평균배팅액	.000	1.000
	참여율	.000	154.560
	핸드수	.030	1.000

[표 30] 방정식에 포함된 변수(RFM)

분석결과 Hosmer와Lemeshow적합도 검정에서 유의확률이 .162로서 0.05보다 크므로 모형은 적합하다고 할 수 있다. 그리고 각 독립변수별 유의 확률은 평균배팅액과 참여율, 핸드수 모두 유의확률 0.05 보다 낮아 RFM 대해 유의한 영향을 미치고, Exp(B) 값은 독립변수 모두 1을 초과 하여 양의 영향을 미친다. 기대수익과 마찬가지로 RFM도 참여율의 경우가 154.560으로 가장 큰 영향을 미치는 것으로 분석된다. RFM 로지스틱 모형에 의한 오분류 테이블은 다음과 같다.

RFM 오분류테이블		예측		계
		가치	비가치	
실제	가치	137 12.91%	81 7.63%	218
	비가치	27 2.54%	816 76.91%	843
계		164	897	1061
Error Rate = 10.2%		Accuracy = 89.8%		
Sensitivity = 62.8%		Specificity = 96.9%		

[표 31] 로지스틱 모형의 오분류 테이블(RFM)

[표 31]에서와 같이 RFM 로지스틱 모형에 의한 오분류테이블 분석결과 전체적인 적합도는 89.8%로 매우 높고, 민감도는 62.8%로 양호하므로 RFM도 마찬가지로 고객별 매출에 대한 가치 고객 예측에 비해 상당히 적합함을 알 수 있다. 로지스틱 회귀모형에서 추정된 확률값을 이용하여 가치고객 가능성에 대한 점수화를 실시한 결과는 다음과 같다.

성별	연령대	평균 배팅액	참여율	게임 일수	게임 시간	핸드수	RFM 구분	RFM 가치 점수	RFM 가치 예측
남	40	48,580,000	10	1	0	6	가치	100	가치
남	40	1,510,000	82	26	155	6039	가치	99.91	가치
남	50	68,000	68	15	93	3256	비가치	85.67	가치
여	30	4,950,000	38	9	31	1671	가치	77.65	가치
남	40	2,514,000	67	4	64	4882	가치	77.61	가치
남	60	1,041,000	63	8	83	3498	가치	77.46	가치
남	30	196,000	69	15	62	2247	비가치	62.51	가치
여	40	3,277,000	82	7	31	1139	가치	62.43	가치
남	40	4,297,000	84	5	14	529	가치	55.97	가치
남	40	212,000	76	17	42	1820	비가치	49.49	비가치
남	70	276,000	91	6	53	5673	가치	49.39	비가치
남	60	194,000	83	15	46	1440	비가치	49.33	비가치

[표 32] 가치고객 가능성 점수 샘플 데이터(RFM)

비율	점수	인원	RFM	고객별매출액
가치	50 점▲	15.4%	72.6%	79.4%
비가치	50 점▽	84.6%	27.4%	20.6%

[표 33] 가치, 비가치 고객 비율 정리 (RFM)

#### 4-7-4. 고객생애가치에 대한 로지스틱 회귀 분석

고객생애가치에 대해서는 1)게임시간이 31.5시간을 초과하는 노드 중에서 평균배팅액이 330,000원 이하의 노드와 2) 게임시간이 31.5시간 이하의 노드 중에서 평균배팅액이 116,000원이하 이고 51,000원 이상의 노드 에서의 자료(표본 약 1,780개)만을 이용하여 로지스틱 회귀 모형을 구축한다. 분석에 사용할 변수는 기대수익, RFM 분석때와 마찬가지로 사결정나무 분류 시 직접적으로 연관있는 평균배팅액, 핸드수, 게임시간, 참여율을 선택하도록 하겠다. 고객생애가치에 대한 로지스틱 회귀분석 결과는 다음과 같다.

단계	카이 제곱	자유도	유의확률
1	488.921	8	.239

[표 34] Hosmer와 Lemeshow 검정(고객생애가치)

종속변수	독립변수	P(유의확률)	Exp(B)
RFM	평균배팅액	.000	1.000
	참여율	.000	253.530
	핸드수	.001	1.000

[표 35] 방정식에 포함된 변수(고객생애가치)

분석결과 Hosmer와Lemeshow적합도 검정에서 유의확률이 .239로서 로지스틱회귀모형은 적합하고 각 독립 변수별 유의 확률은 평균배팅액과 참여율, 핸드수 모두 유의확률 0.05 보다 낮아 고객생애가치 대해 유의한 영향을 미친다. Exp(B) 값은 독립변수 모두 1이 넘어 양(+)의 영향을 미치고 있음을 알 수 있다. 다른 신규 종속 변수(기대수익, RFM)와 마찬가지로 고객생애가치 역시 참여율의 경우가 253.530 으로 가장 큰 영향을 미치는 것으로 분석된다. 고객생애가치 로지스틱 모형에 의한 오분류 테이블은 다음과 같다.

고객생애가치 오분류테이블		예측		계
		가치	비가치	
실제	가치	237 13.31%	119 6.69%	356
	비가치	42 2.36%	1382 77.64%	
계		279	1501	1780
Error Rate = 9.1%		Accuracy = 90.9%		
Sensitivity = 66.5%		Specificity = 97.1%		

[표 36] 로지스틱 모형의 오분류 테이블(고객생애가치)

[표 36]에서와 같이 고객생애가치 로지스틱 모형에 의한 오분류 테이블 분석결과 전체적인 적합도는 90.9%로 매우 높고, 민감도는 66.5%로 양호하다. 고객생애가치도 마찬가지로 고객별 매출에 대한 가치 고객 예측에 비해 상당히 유효함을 알 수 있다. 로지스틱 회귀모형에서 추정된 확률값을 이용하여 가치고객 가능성에 대한 점수화를 실시한 결과는 다음과 같다.

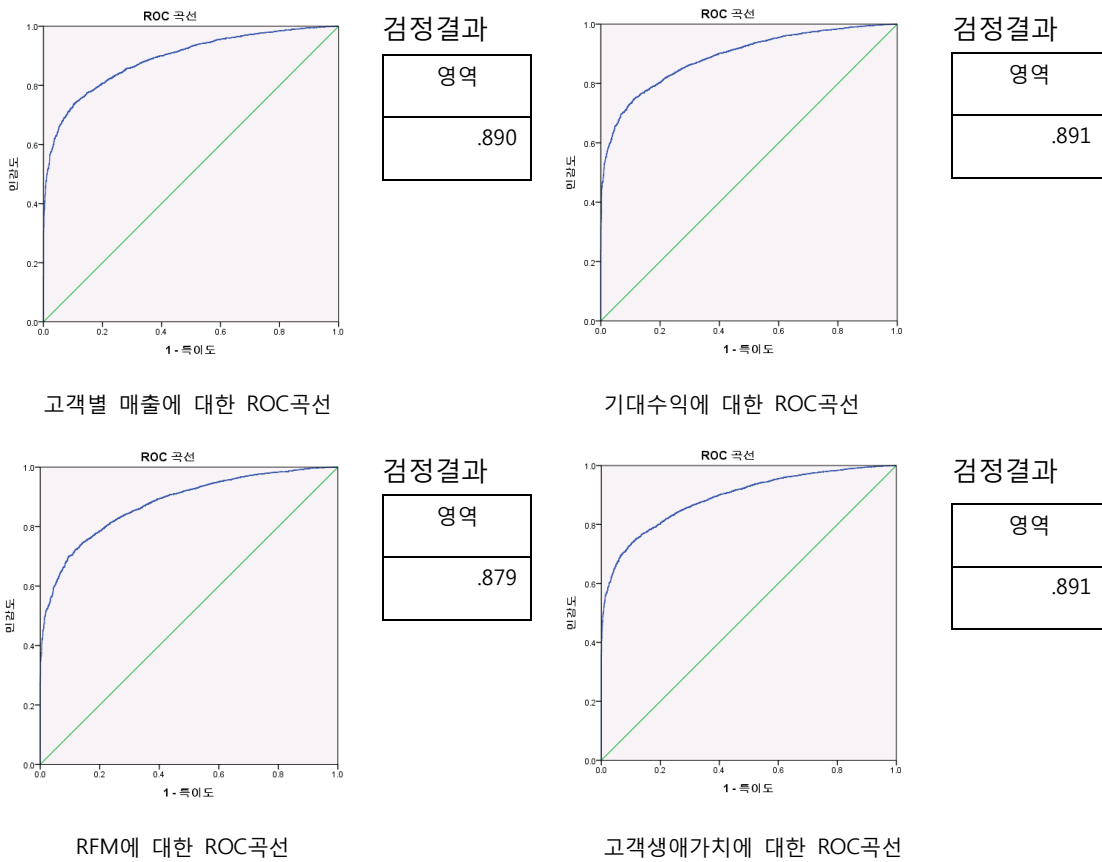
평균배팅액	참여율	게임일수	게임시간	핸드수	고객생애가치 구분	고객생애가치 가치 점수	고객생애가치 가치 예측
4,077,000	74	59	268	14841	가치	100	가치
99,720,000	84	4	14	590	가치	99.9606	가치
389,000	69	18	112	3696	가치	94.8522	가치
441,000	71	15	95	3066	가치	87.2841	가치
58,000	88	18	65	2801	비가치	65.5701	가치
2,864,000	70	4	20	786	가치	51.9928	가치
3,303,000	70	3	10	465	가치	49.8585	비가치
3,060,000	73	2	6	175	비가치	37.1777	비가치
570,000	83	9	47	1671	비가치	37.1627	비가치
216,000	90	10	45	3213	가치	37.1093	비가치
553,000	69	5	55	2680	가치	37.1036	비가치

[표 37] 가치고객 가능성 점수 샘플 데이터(고객생애가치)

비율	점수	인원	고객생애가치	고객별매출액
가치	50 점▲	15.6%	87.2%	80.9%
비가치	50 점▽	84.4%	12.8%	19.1%

[표 38] 가치, 비가치 고객 비율 정리 (고객생애가치)

기존 종속 변수인 고객별매출액과 신규 종속 변수인 기대수익, RFM, 고객생애가치의 로지스틱 회귀분석결과를 ROC곡선으로 표현을 하면 다음과 같다.



[그림 23] 로지스틱 회귀모형에 대한 ROC곡선



## 4.8 분석 비교 결과

각 통계적 모델 비교 분석 결과는 다음과 같다.

		고객별매출액	기대수익	RFM	고객생애가치
다중회귀분석	결정계수( $R^2$ )	.321	.541	.479	.531
	유의확률	.000	.000	.000	.000
상관분석	고객별매출액	1	.751	.748	.395
	기대수익		1	.992	.441
	RFM			1	.465
	고객생애가치				1
	오분류	13.7%	4.8%	7.1%	5.2%
의사결정나무	적합도	86.3%	95.2%	92.9%	94.8%
	민감도	61.6%	87.2%	74.9	82.9%
	유의확률	.175	.552	.162	.239
로지스틱	오분류	27.4%	8.7%	10.2%	9.1%
회귀분석	적합도	72.6%	91.3%	89.8%	90.9%
	민감도	5.2%	68.1%	62.8%	66.5%

[표 39] 각 분석별 결과 비교표

서론에서 제시한 바와 같이 기존 단순 매출액만으로 가치도를 평가하는 방법에서 보다 고객의 게임 내용에 비중을 두어 신규 종속 변수인 기대수익, RFM, 고객생애가치 라는 종속 변수를 생성하여 여러 가지 통계적 모델링을 적용하여 분석하여 보았다. 비교 분석결과 기존 종속 변수 인 고객별 매출액 보다 신규 종속 변수인 기대수익, RFM, 고객 생애가치가 결정 계수 및 적합도 민감도 등 모든 부분에서 더 적합한 모델링이 되는 것을 확인할 수 있으며 신규 종속 변수에서도 미세하지만 기대수익, 고객생애가치, RFM 순으로 통계적 모델링이 더 적합함을 알 수 있다. 또한 상관 분석을 통해 기존 종속변수와 신규 종속변수간에 연관성이 있는것으로 보아 물론 예외의 결과도 있지만 전체적으로 신규 종속변수(기대수익, RFM, 고객생애가치)가 높으면 기존종속변수(매출액)에도 영향을 미침을 알 수 있다.

## 5. 결론 및 시사점

본 논문에서는 외국인 전용 VIP카지노의 고객 관리의 중요성에 대해서 강조하며, 기존 카지노 고객별 매출액 (WinLoss) 으로만 판별하던 고객 가치도 측정의 한계점을 짚어 보고, 보다 좀 더 세부적으로 고객의 게임내용과 방문 주기, 활동내용 등에 중점을 두어 분석을 해 보았다. 기존 카지노에는 다루지 않았던 변수인 '기대수익', 'RFM', '고객생애가치' 등을 추출하였고, 회귀분석을 통해 이 종속 변수에 직접 영향을 주는 변수에 대해서 분석해 보았고 상관분석을 통해 종속 변수들 간의 연관성 및 이질성이 있는지에 대한 분석을 파악해보았으며 또한 의사결정 나무를 통한 우수 고객군과 일반 고객군 그리고 위험 고객군을 분류 후 적합성 검토 및 로지스틱 회귀분석을 통한 적합성 분석 고객 가치에 대한 예측 파악 후 점수화를 해보았다. 본 논문을 위한 데이터는 국내 외국인 전용 대표 카지노인 P사의 데이터를 이용하였고, 보안상 데이터에 대한 명칭은 대부분 가칭을 사용하였다. 물론 본 논문의 분석 결과 고객 가치도가 높다 하여도 카지노의 특성상 손실을 입을 수도 있으며, 분석 결과 보다 훨씬 더 많은 금액의 이익을 취할 수도 있다. 자주 방문하며 게임을 많이 하는 고객이라 하더라도 카지노에 많은 손실을 입힐 경우 Black List 가 되어 카지노 Protection을 위해 출입 금지가 될 수 있는 것이 현 카지노의 현실이다. 그렇기 때문에 카지노 VIP고객에 대한 분석은 보다 더 현실성에 가깝고 더욱 세밀하게 파고 들여다 볼 필요가 있다. 본 논문에서 제시하는 시사점은 기존 종속 변수인 고객별 매출이 아닌 게임의 내용에 중점을 둔 '기대수익', 게임 내용 및 게임 일수에 초점을 맞춘 'RFM', 게임 내용, 방문 패턴, 미래의 가치성에 중점을 둔 '고객생애가치' 는 모두 고객별 매출액과 상관성이 깊은 수치를 나타낸다는 것이다. 물론 분석 결과와 크게 빛나가는 고객도 존재 하나 대부분이 비슷한 패턴을 보이는 것으로 보아 카지노 역시 자주 방문 하며 게임에 충실히 하는 고객일수록 카지노에 공헌하는 실 수익이 크다는 것을 알 수 있다. 또한 지역간에도 카지노의 실적은 차이가 보여지고 있음을 알 수 있다. 특정지역은 관측수가 높은 반면 매출의 합계나 평균이 현저히 떨어 지는 반면 관측 수는 낮은 지역이 매출액의 합계나 평균이 월등히 높게 분석 되는 등 지역별로 매출액 및 게임성향이 다르게 나타나고 있으며 주요 지표(기대수익, RFM, 고객생애가치)는 모두 고객별 매출액 과 상관성은 깊으나, 분산 분석 결과 서로간에 상관성은 적어 서로 차별화된 수치를 나타내고 있음을 알 수 있다. 이로 인해 의사결정 나무(Decision Tree) 분류 분석을 통해 세 지표간 각 다른 고객 군 분포 결과를 얻었으며 이를 토대로 적합성 검토 후 기존 종속 변수(고객별 매출액)보다는 적합도가 더 높음을 알 수 있었다. 본 논문을 마무리 지으며 카지노는 특성상 일반 호텔, 백화점, 대형 마트와는 차별화된 고객 관리가 필요하다는 결론을 내릴 수 있었으며, 또한 현재 카지노의 고객 분석에 대한 다양성 부족과 동시에 단순 고객별 매출에 만 취중 되어 있는 분석 Point 에 대한 문제점을 파악하고 새로운 분석 Tool을 연구하고 개발하여 더욱 다양하고 많은 고객 군을 대상으로 분석하여 다가올 무한 경쟁 시대에 새로운 제주 카지노의 생존 경쟁 도구로 사용 될 수 있었으면 하는 바램이다. 다만 서론에서도 언급했듯 고객 분석을 위해서는 보다 더 많은, 디테일한 데

이터가 필요한 반면 개인정보보호법 등의 사생활 관련 법이 강화되면 될수록 이런 데이터에 대한 수집이 점점 어려워지고 있는 것이 현실이다. 한시라도 빨리 이러한 모순적인 문제점들이 해결되어 데이터 분석에 대한 더 많은 방법과 결과들이 연구될 수 있는 환경이 마련되었으면 하는 바람도 있다.

## 참 고 문 헌

- [1] 김형수,김영걸, CRM고객관계관리 전략 원리와 응용, 2009.
- [2] 이규웅,윤지환,최규완, 「DEA 기법을 활용한 카지노 기업간의 경영 효율성 분석」, 2010.
- [3] 박창이, 김진석, 「R을 이용한 데이터마이닝」, 2008.
- [4] 박찬주,박윤선, 「고객가치(Customer Value) 산정 방법론에 관한 연구」, 2002.
- [5] 전희주, 「고객 세분화에 기반한 생존분석을 활용한 고객수명 예측 모델」, 2011.
- [6] 이희찬, 「내국인 카지노갬블 수요분석」, 2009.
- [7] 이도현, 「데이터 마이닝을 이용한 CRM」, 2000.
- [8] 최종후,서두성, 「데이터마이닝 의사결정나무의 응용」, 1999.
- [9] 박명하, 「데이터마이닝을 활용한 CRM사례」, 2005.
- [10] 하성호,이재신, 「데이터마이닝을 활용한 동적인 고객분석에 따른 고객관계관리 기법」, 2003.
- [11] 김주성,박유라, 「매출액과 고객생애가치(CLV)기반 고객가치평가 기법의 결과 비교」, 2011.
- [12] 정은영, 「분위수 회귀분석을 이용한 카지노 고객의 국적 및 방문형태가 카지노 매출액에 미치는 영향 분석」, 2011.
- [13] 문화체육관광부, 「2015년기준 관광동향에 관한 연차 보고서」, 2016.
- [14] 김병수,배화수, 석경하, 조대현, 최국렬, 『SAS Enterprise Miner 6.2를 이용한 데이터마이닝 입문』,(교우사,2013).
- [15] 김연형, 김재훈, 이석원, 『고객관계관리와 데이터마이닝』(교우사,2010).
- [16] [네이버 지식백과] 「마카오의 경제」, (홍콩·마카오 개황, 외교부, 2009.8).
- [17] [네이버 블로그] 「카지노사업」, <http://blog.daum.net/costmgr/9203>.
- [18] [네이버 블로그] 「고객가치분석 및 활용」, <http://haam1004.tistory.com/40>
- [19] [네이버 블로그] 「데이터마이닝 정리」, <http://blog.naver.com/khaisilver/88803394>.
- [20] [네이버 지식백과] 「데이터 마이닝」, (컴퓨터인터넷IT용어대사전,2011.1.20).
- [21] 권세혁(한남대학교 비즈니스통계학과교수) 홈페이지 <http://wolfpack.hnu.ac.kr/>