경영전문석사학위 논문

온라인 게임 유료 이용자의 집단유형과 특성에 관한 연구

2023년 2월

서울과학종합대학원대학교

박 현 철

경영전문석사학위 논문

온라인 게임 유료 이용자의 집단유형과 특성에 관한 연구

2023년 2월

서울과학종합대학원대학교

박 현 철

온라인 게임 유료 이용자의 집단유형과 특성에 관한 연구

지도교수 김 보 영

이 논문을 경영학 석사 학위논문으로 제출함
2023년 2월
서울과학종합대학원대학교
박 현 철

박현철의 석사 학위논문을 인준함
2023년 1월

위 원 장 _____ 신호상 _____ (인)

위 원 _____ 임효숙 _____ (인)

위 원 _____ 김보영 _____ (인)

초 록

대한민국의 콘텐츠 산업에서 상당한 비중을 차지하는 게임 산업은 지난 10년 동안 꾸준히 성장하였지만 여러 문제점이 드러나기도 하였다. 특히 유료 재화의 구매를 유도하는 비즈니스 모델은 많은 비판을 받았음에도 여전히 다수의 게임에 적용되고 있다. 이용자의 요구를 충분히 반영하지 않은 일방적인 선택은 반발을 샀고 끝내 물리적인 갈등으로 이어졌다.

이에 본 연구는 게임 유저, 그중에서도 수익과 직결되는 유료 이용자에 대한 이해를 높여 유저의 니즈에 부합하는 상품과 콘텐츠의 출시에 도움을 줌으로써 비즈니스 모델의 개선에 기여하고자 진행되었다. 이를 위해 유료 이용자를 대상으로 군집분석을 시행하고 집단별 특성을 비교 분석하였다.

약 4천 개의 데이터로 구성된 기본 데이터셋을 이상치 처리, 피처 스케일링을 실시한 경우로 나누고 해당 데이터셋에 k-means, k-medoids, k-prototypes라는 3개의 알고리즘을 적용하여 군집분석을 시행한 결과, 정규화 처리가 된 데이터셋에 초기치를 4로 지정한 k-means 알고리즘을 사용하였을 때 가장 우수한 군집화 결과를 보여준다는 것을 알 수 있었다.

집단별 특성을 살펴보고자 게임의 이용 양태에 관한 변수와 온라인 게임 내 구매 총액 변수에 대하여 분산분석을, 인구통계학적 변수에 대해서 교차분석을 시행하였다. 가장 어렸을 때 게임을 접한 젊은 남성 집단인 A 그룹은 주중·주말을 가리지 않고 게임의 평균 이용시간이 가장 길었으며 과거 1년간 플레이한 온라인 게임의 개수가 제일 많았다. 또 유료 재화의 결제 금액도 평균을 웃돈다는 특성을 보였다. 반면 중·장년층 남성 위주로 구성된 B 집단은 온라인 게임을 가장 늦게 시작하였고, 주중의 평균 이용시간도 가장 짧았다.

C와 D 집단은 여성으로만 이루어졌으나 연령대에서 차이를 보였다.

중·장년층 여성이 대부분인 C 집단은 가장 늦은 나이에 온라인 게임을 접했으며 주말의 평균 이용시간도 제일 짧았다. 그리고 플레이한 온라인 게임의 개수와 유료 재화의 구매 금액도 모든 그룹 중 최하위로 나타났다. D 집단은 상대적으로 젊은 여성들로 구성되었는데 같은 여성인 C 집단과 비교하였을 때 청소년기에 온라인 게임을 접했다는 점과 더욱 긴 평균 이용시간에서 차이를 보였다. 하지만 온라인 게임 이용 개수와 온라인 게임 내 구매 총액의 평균이 전체 평균을 밑돈다는 공통점도 지니고 있었다.

본 연구는 약 4천 명의 온라인 게임 유료 이용자를 대상으로 비지도학습의 한 갈래인 군집분석을 시행하고 형성된 집단의 특성을 비교 분석하여 각 집단유형이 제각기 상이한 특성이 있다는 것을 파악하였다. 이를 바탕으로 향후 유저 친화적인 게임 운영과 마케팅 전략의 수립, 나아가 비즈니스 모델의 개선 등에 활용될 수 있을 것으로 기대한다.

목 차

제Ⅰ장 서 론 ····· 1
제 1절 연구 배경 ····· 1
제 2절 연구 목적 ····· 3
제 3절 연구 구성 ····· 4

제Ⅱ장 이론적 고찰 ····· 6
제 1절 온라인 게임의 유료화와 유료 재화의 구매 ····· 6
(1) 온라인 게임의 유료화 ····· 6
(2) 부분 유료화 모델 ····· 8
(3) 유료 재화의 구매 ····· 10
제 2절 집단유형 분류 ····· 12
(1) 게임 이용자 유형 분류 ····· 12
(2) 군집분석 ····· 14
1) 군집화 기법 ····· 15
2) 초기치 지정 및 성능평가 ····· 16

제Ⅲ장 연구 방법 ····· 19
제 1절 연구 대상 및 절차 ····· 19
제 2절 연구 방법 ····· 19
(1) 데이터 수집 ····· 19
(2) 데이터 전처리와 탐색적 데이터 분석 ····· 20
(3) 군집분석 ····· 27
(4) 군집별 특성 분석 ····· 28

제Ⅳ장 분석 결과 ····· 30
제 1절 군집분석 결과 ····· 30
(1) k-means ····· 30
(1) k-medoids ····· 36
(1) k-prototypes ····· 41
제 2절 군집별 특성 분석 결과 ····· 42
(1) 게임 이용 양태와 온라인 게임 내 구매 총액 ····· 42
(2) 인구통계학적 특성 ····· 44

제Ⅴ장 결론 ····· 47
제 1절 요약 및 시사점 ····· 47
제 2절 한계 및 향후 연구방향 ····· 49

표 목 차

<표1> 변수 요약 ····· 23
<표2> 데이터셋의 인구통계학적 특성 ····· 24
<표3> Box plot 활용 이상치 표현 ····· 25
<표4> 표준화 적용 군집분석 (k-means) ····· 31
<표5> 표준화 적용 및 이상치 처리 군집분석 (k-means) ····· 32
<표6> 정규화 적용 군집분석 (k-means) ····· 33
<표7> 정규화 적용 및 이상치 처리 군집분석 (k-means) ····· 34
<표8> 표준화 및 정규화 적용 군집분석 (k-means) ····· 35
<표9> 표준화 적용 군집분석 (k-medoids) ····· 36
<표10> 표준화 적용 및 이상치 처리 군집분석 (k-medoids) ····· 37
<표11> 정규화 적용 군집분석 (k-medoids) ····· 38
<표12> 정규화 적용 및 이상치 처리 군집분석 (k-medoids) ····· 39
<표13> 표준화 및 정규화 적용 군집분석 (k-medoids) ····· 40
<표14> 데이터셋별 군집분석 (k-prototypes) ····· 41
<표15> 군집별 분산분석표 ····· 43
<표16> 군집별 인구통계학적 특성 ····· 45

그 림 목 차

<그림1> 연구흐름도 ····· 5
<그림2> 변수 간 상관관계 ····· 27
<그림3> 최적 군집분석 결과 ····· 42

제 Ⅰ 장 서 론

제1절 연구 배경

대한민국의 콘텐츠는 국경을 뛰어넘어 해외에서도 많은 주목을 받고 있다. 먼저 음악 부문에서는 가수 싸이의 <강남스타일>이 세계적으로 유행한 이후, BTS와 블랙핑크가 강력한 팬덤을 형성하며 인기를 누리고 있다. 영화와 드라마 부문에서는 봉준호 감독의 2019년작 <기생충>이 해외 유수의 영화제와 시상식에서 작품성을 인정받았고, 드라마 <오징어 게임>은 넷플릭스의 시청 기록을 갈아치우며 흥행에 성공하였다.

콘텐츠 산업은 게임, 음악, 영화, 방송, 출판, 캐릭터 등의 10개 분야로 분류된다. 콘텐츠 시장의 크기는 2019년 기준 약 2조 4000억 달러에 육박하고 이는 세계 반도체 시장의 5배에 이르는 거대한 규모이다(김윤지, 2019). 한편 2021년과 2022년의 <대한민국 게임백서>에 따르면, 대한민국의 콘텐츠 시장은 2022년 상반기 기준, 전년 동기 7% 성장하여 약 67조 원의 가치를 지니는 것으로 평가받고 있고 방송 (16.4%), 출판 (16.4%), 게임 (15.7%) 순으로 높은 비중을 차지하고 있다. 수출의 관점에서 살펴보면 2022년 상반기에 약 55억 달러의 외화를 벌어들였다. 이 중 게임(66.5%)의 비율이 3분의 2 정도이며 그 다음으로 캐릭터 (7.2%), 음악(6.6%) 순으로 집계된다. 이처럼 대한민국의 콘텐츠 산업에서 상당한 비중을 차지하는 게임 산업은 전년 대비 21.3% 성장했고 18조 8,855억 원의 매출을 달성하였다. -0.3%의 성장률을 기록한 2013년을 제외하면 최근 10년간 지속적으로 시장 규모도 커지고 있다.

게임 산업의 양적인 성장세가 눈에 띄는 가운데 불합리한 노동자 처우, 일방적인 게임 운영, 유료 재화 판매에 치중된 비즈니스 모델 등과 같은 문제점에 대한 개선 요구도 끊이지 않고 있다. 특히 비즈니스 모델 이슈는 게임사 본사에 마차와 트럭 등을 보내 항의하는 가시적인 마찰로 이

어져 언론에 보도되기도 하였다(정호, 2022; 한경우, 2022).

따라서 본 연구에서는 2020년 기준, 게임 산업의 83.4%를 차지하는 PC와 모바일 게임을 온라인 게임이라는 하나의 범주로 묶은 데이터셋을 기반으로 온라인 게임 유료 이용자에 대한 군집분석을 시행하여 집단별 특성을 비교 분석하고자 하였다. 이를 통해 유저, 그중에서도 게임의 수익과 직결되는 유료 이용자에 대한 이해를 제고함으로써 비즈니스 모델의 개선에 기여하고 궁극적으로는 유저와 게임사의 간극을 좁히는 것에 주안점을 두었다.

우선 온라인 게임의 정의를 명확히 하였다. 온라인 게임은 게이머의 기기가 네트워크로 연결된 멀티 플레이어 분산 게임이라고 할 수 있다(Adams, 2014). 이와 유사하게 좁게는 컴퓨터를 매개로 인터넷이나 통신망을 거쳐 복수의 이용자가 서버에 접속하여 플레이하는 게임, 넓게는 네트워크를 이용해 복수의 이용자 간에 서로 영향을 주고 받을 수 있는 모든 게임이라고 정의하기도 하였다(한국게임산업개발원, 2005).

하지만 기술의 발달에 따라 전통적인 분류 체계를 적용하기 어려운 크로스 플랫폼 게임이 시장을 선도하고 있다(윤진우, 2021; 원태영, 2022). 크로스 플랫폼은 온·오프라인을 연결하여 다양한 기기를 지원함으로써 시간과 장소에 구애받지 않고 최적의 서비스를 제공하는 시스템이다(한창완, 2008). 이처럼 ICT의 발전은 게임의 분류 체계에도 상당한 영향을 미쳤고 이러한 변화를 반영하여 PC와 모바일 게임을 온라인 게임이라는 상위 범주로 포함하여 함께 다루었다.

다음으로 유료 이용자의 범위를 한정하였다. 본 연구에서 유료 이용자는 온라인 게임 내에서 유료로 판매되는 게임머니나 게임 아이템와 같은 유료 재화를 구매하는 사람을 의미한다. 재화란 표준국어대사전에 따르면 사람이 바라는 바를 충족시켜주는 모든 물건을 뜻하는데 이를 토대로 유료 재화를 스스로 바라는 것을 얻기 위하여 비용을 지불하고 구매한 무언가로 유추할 수 있다. 게임물에 대한 등급분류업무와 사후관리업무를 관장하는 문화체육관광부 산하, 게임물관리위원회의 등급분류규정 제12조 4

항에 따르면, 유료 재화를 유료 결제로 얻은 가상 재화 등 현금과 유사한 가치를 지닌 것으로 규정하고 있다. 본 연구에서는 유료로 입수한 게임머니와 아이템을 가리키는 용어로 유료 재화를 사용하며 분석을 진행하였다.

제2절 연구 목적

본 연구는 기계학습 (머신러닝, Machine Learning)의 한 갈래인 비지도학습 (Unsupervised Learning) 중에서 군집분석을 시행하여 온라인 게임 유료 이용자의 유형을 파악하고, 분산분석과 교차분석으로 형성된 집단의 특성을 비교 분석하였다.

온라인 게임 이용자의 유형을 분류한 기존 연구는 PC나 모바일 중 하나의 플랫폼 이용자를 대상으로 진행된 것이 많았다(김영주, 2022; 김종무, 2007; 김효은, 2016; 이준영, 이제성, 2017; 정다영 외, 2015). 이러한 방식의 연구는 연구 대상의 범위를 제한하여 더욱 상세하고 심층적으로 접근할 수 있다는 장점이 있겠으나 PC와 스마트폰으로 온라인 게임을 즐기는 최근의 트렌드를 반영하기 어렵고 보편적인 온라인 게임 유저의 특성을 담기에 제한이 있다고 판단하였다.

유료 재화를 구매한 유저의 특성을 살펴본 연구는 대체로 이들이 구매한 아이템의 유형에 집중하였다. 또 왜 유료 아이템에 돈을 쓰는지 구매 동기를 밝혀내고자 하는 것이 대부분이었다. 위 연구에 대해서는 다음 장에서 상술(詳述)하였다. 하지만 본 연구는 유료 아이템과 구매 동기가 아니라 어떤 유저가 온라인 게임에서 유료 재화를 구매하는지, 즉, 유료 이용자 본연의 특징에 주목하였다는 점에서 차이가 있다.

이외에도 한 가지의 군집분석 알고리즘을 사용하는 데 그치지 않고, 3개의 알고리즘으로 얻은 결과를 비교하여 더욱 정교한 군집화를 시도한 것에서도 다른 면모를 보인다. 즉, 기존 연구보다 넓은 범위의 유료 이용

자를 대상으로 복수의 알고리즘을 사용하여 가장 우수한 군집분석 결과를 확보하고자 한 것이 본 연구의 특징이다. 이렇게 도출된 온라인 게임 유료 이용자의 집단별 특성은 마케팅과 운영, 기획 등에 있어 근거자료로 활용되어 비즈니스 모델의 개선에 도움을 줄 수 있을 것으로 기대된다.

제3절 연구 구성

본 연구는 총 5장으로 구성되었고 각 장은 다음의 내용을 담고 있다.

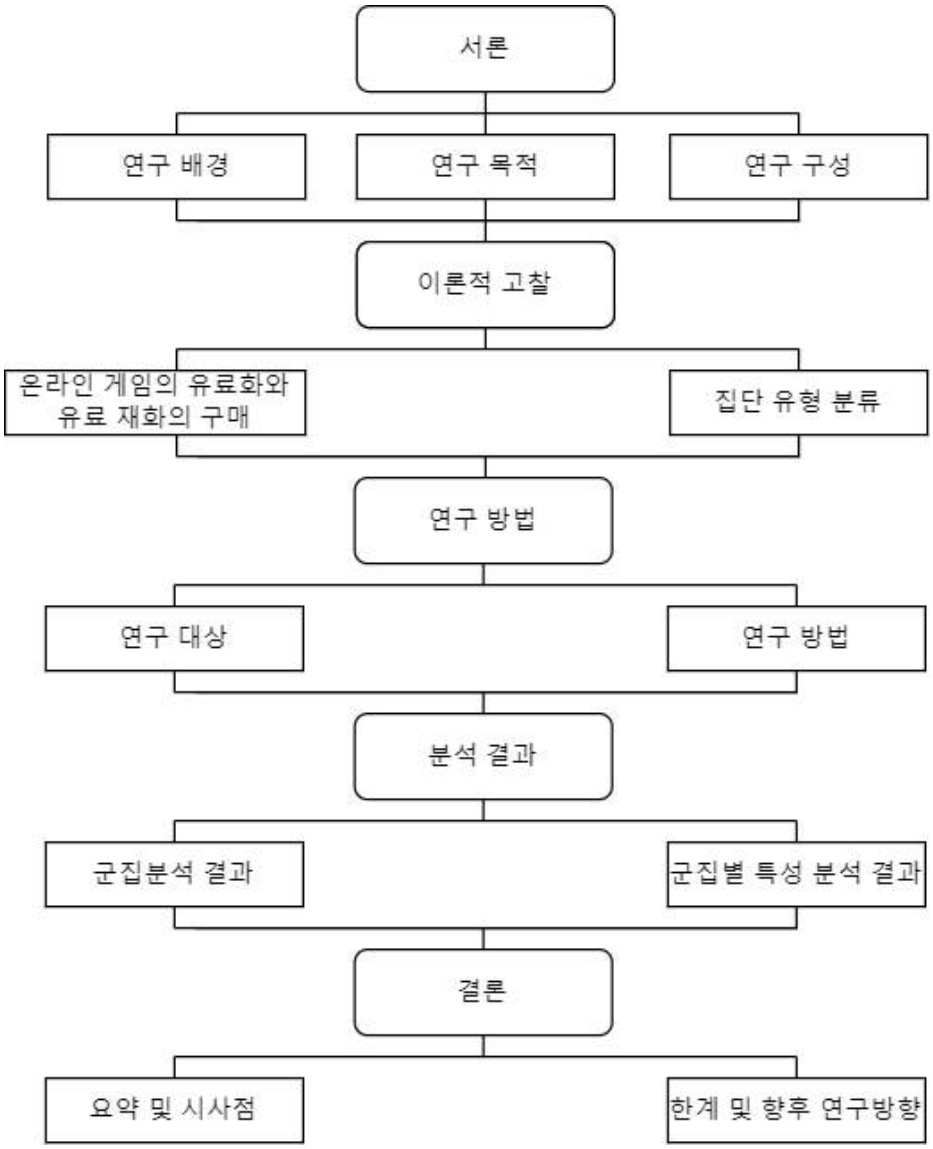
제 1장 서론에서는 연구의 배경과 목적, 그리고 전체적인 논문의 구성에 대해 정리하였다.

제 2장 이론적 고찰에서는 온라인 게임의 유료화와 유료 재화의 구매에 관한 연구의 흐름을 살펴보고 집단 유형 분류와 관련된 연구 및 분석 기법을 소개하였다.

제 3장 연구방법에서는 분석에 사용된 데이터의 특징과 데이터 수집부터 최종 단계인 결론 도출에 이르는 연구의 프로세스에 대해 구체적으로 기술하였다.

제 4장 분석결과에서는 군집분석으로 온라인 게임 유료 이용자를 유형화하고 분산분석과 교차분석을 시행하여 집단유형별 특성을 파악하였다.

제 5장 결론에서는 본 논문을 요약하고 시사점을 도출하였다. 그리고 연구의 한계점을 언급하며 향후의 연구방향을 제시하였다.


서론
연구 배경
연구 목적
연구 구성
이론적 고찰
온라인 게임의 유료화와 유료 재화의 구매
집단 유형 분류
연구 방법
연구 대상
연구 방법
분석 결과
군집분석 결과
군집별 특성 분석 결과
결론
요약 및 시사점
한계 및 향후 연구방향

<그림 1> 연구흐름도

제 Ⅱ 장 이론적 고찰

제1절 온라인 게임의 유료화와 유료 재화 구매

(1) 온라인 게임의 유료화

초기 PC 온라인 게임 산업의 주류는 텍스트 기반의 머드 (MUD, Multi User Dungeon) 게임이었다. 머드 게임은 가상세계에서 다수의 유저와 의사소통하며 미지의 공간을 탐험하며 활동하는 게임을 의미한다. 이후, 머드 게임의 매력을 살린 다중역할수행게임 (MMORPG, Multi Player Online Role Playing Game)이 주축이 되어 온라인 게임 산업의 발전을 이끌었다. 대표적인 사례가 1998년 출시된 엔씨소프트의 <리니지>이다. 이 게임의 엄청난 성공은 아류작을 양산하였는데 이것은 게임에 적용된 비즈니스 모델 역시 답습하였다. MMORPG 장르의 주요 비즈니스 모델은 정액제였다. MMORPG의 특성상, 긴 시간 공을 들여 자신의 캐릭터를 키워야 하는데 투자한 자원에 비례하여 캐릭터는 유저의 이상(理想)과 닮아가기 시작하고 이는 게임에 대한 높은 몰입도와 충성도로 이어졌다. 아이템 교환이나 길드 활동 등과 같은 타인과의 상호작용 역시 게임에 몰두하게 만드는 요소 중 하나였다. 이와 같이 MMORPG의 특성을 고려한 정액제 모델은 온라인 게임 시장을 선점한 기업들에 막대한 부를 가져다주었다(김태경 외, 2009).

정액제는 정해진 비용을 내기만 하면 특정 기간 해당 콘텐츠를 즐길 수 있는 권리를 제공하였다. 기간마다 유저가 지불하는 개인 정액제, 시간별로 계산하는 개인 정량제, 게임사와 계약을 맺은 PC방에서 게임을 플레이하는 선불 정액제 등이 이에 속한다. <리니지>가 인기를 끌었던 90년대 후반, 2000년대 초반의 대한민국에는 블리자드의 <스타크래프트>의 흥행에 힘입어 PC방 문화가 널리 퍼져 이곳에서 게임을 접하고 즐기는

게이머가 많았다. 자연스레 이들을 목표로 PC방 업체와 제휴하며 시장이 확대되었다(김보아, 2021).

하지만 정액제를 채택한 MMORPG가 반드시 성공을 보장하는 것은 아니었다. 후발주자의 경우, 기존의 게임과 다른 콘셉트, 다채로운 볼거리 등을 제공하기 위해 고액의 개발비를 투입하였다. 여기에 신규 유저 확보를 위해 대대적인 마케팅 활동도 벌였다. 그럼에도 불구하고 기존 흥행작에 이미 몰입해버린 유저들은 쉽사리 유입되지 않았고, 무료 테스트 기간의 데이터만을 근거로 섣불리 유료화를 시도한 결과 경영위기에 빠지는 기업도 존재하였다. 이를 타개하고자 특정 콘텐츠의 이용에만 요금을 부과하는 모델을 도입하였으나 유료·무료 유저 간의 차별, 콘텐츠 최적화의 실패 등의 문제를 초래하여 오히려 유저 이탈이 가속화되었고 결국 서비스 종료를 맞이하기도 하였다(이남재 외, 2005).

모바일 온라인 게임은 스마트폰의 도입 전후로 게임의 비즈니스 모델이 크게 달라진다. 먼저 스마트폰이 주류가 되기 전인 피처폰 시절에는 통신사가 운영하는 사이트에서 유료로 콘텐츠를 내려받아 즐기는 소유형 과금 방식이 대부분이었고, 이러한 경향은 스마트폰 출시 초기까지 이어졌다(한연규, 2006). 2008년, 스마트폰이 대두되고 애플의 앱스토어로부터 애플리케이션 (앱, Application)이라는 이름으로 콘텐츠가 제공되기 시작하자 기존의 소유형 과금 방식과 더불어 게임 내 광고 방식의 수익화 모델을 채택한 게임이 등장하였다. 게임 내 광고 방식은 자본력이 부족한 소규모 개발사가 만든 게임에 주로 도입되었는데 무료로 콘텐츠를 즐길 수 있도록 하지만 그 대신 게임 내에서 광고를 시청하게 함으로써 광고주로부터 광고비를 받는 구조였다. 그러나 PC 온라인 게임 시장에서 부분 유료화 모델의 영향으로 다소 제한이 있기는 하나 무료로 콘텐츠를 즐기는 것에 익숙해진 유저들의 수요에 맞춰 모바일 온라인 게임 시장에서도 부분 유료화가 대세가 되었고 현재는 유료 다운로드 모델 (소유형 과금 방식)과 게임 내 광고 모델, 그리고 부분 유료화 모델이 병존하고 있는 상황이다(조은하, 2017).

(2) 부분 유료화 모델

부분 유료화 모델은 패키지를 구입하거나 정액 결제를 하지 않더라도 무료로 콘텐츠를 제공하지만 게임 내 재화 및 부가 콘텐츠를 유료로 판매하여 수익을 거두는 비즈니스 모델이다. 대한민국에서 서비스되는 대부분의 온라인 게임이 부분 유료화 모델을 채택하고 있고 인접한 대만과 일본, 중국뿐만 아니라 아시아를 넘어 미국, 유럽 등 타 대륙에서도 온라인 게임의 주요 비즈니스 모델로 자리잡았다(유창석 외, 2014).

부분 유료화 모델의 유형은 전통적 모델인 유료 아이템과 VIP 시스템으로 나눌 수 있다. VIP 시스템은 피로도 시스템으로 인해 일시적으로 게임 플레이가 불가능해진 게이머를 위해 고안되었다. 피로도 시스템이란 과도하게 게임을 이용하여 수치화된 피로도 심볼 (바, 게이지 등)이 고갈되었을 때 게임을 이용할 수 없게 만드는 것을 뜻한다. 유료 아이템은 게임 내의 페르소나에 영향을 주고, VIP 시스템은 게이머에게 영향을 끼치는 것으로 드러났다.

앞서 MMORPG 장르에서의 부분 유료화 모델 도입에 관하여 서술하였는데 사실 부분 유료화 모델의 시작은 MMORPG가 아닌 1999년 출시한 넥슨의 퀴즈 게임인 <퀴즈퀴즈>였다. <퀴즈퀴즈>는 처음에는 무료로 모든 콘텐츠를 제공하였으나 2000년부터 2001년까지 월 정액제를 도입하여 운영하였다. 그러나 유저들이 정액 결제를 유지할 정도의 구매 요인이 부족하였고 이에 운영진은 새로운 대책을 마련하였다. 2001년 하반기 <퀴즈퀴즈 플러스>로 개편하면서 유저들에게 인기를 끌었던 코스튬 등을 제작해 판매하였는데 이 전략이 통하여 주요 수익창출원이 되었고, 세계 최초로 부분 유료화가 도입된 사례로 남게 되었다(이동은, 이유호, 2015).

PC 온라인 게임 중 캐주얼 장르에 한하여 정액제 모델과 비교했을 때 부분 유료화 모델의 가지는 강점으로 유저들이 자신의 돈으로 원하는 콘텐츠를 구입할 수 있다는 것이 있다. 그리고 유료 아이템과 게임 이용으로 획득할 수 있는 아이템의 밸런스를 조정함으로써 유저의 편중과 이탈

을 방지하는 것이 가능하다는 점도 부분 유료화 모델의 장점으로 꼽혔다(Oh, Ryu, 2007).

그러나 정액제 모델을 보완하며 등장한 부분 유료화 모델이 완벽한 것은 아니었다. 무료로 콘텐츠를 이용할 수 있다는 부분 유료화 모델의 긍정적인 측면을 인정하면서도 게임 시장의 경쟁이 격화하며 게임사의 과도한 결제 유도가 문제가 될 수 있고 유료 재화의 구매 여부에 따른 게임 플레이의 제한, 게임 운영의 차별로 밸런스 조절의 실패가 우려된다는 주장이 제기되었다(김은비, 2019).

부분 유료화 모델의 문제점과 근본적인 한계가 지적받기도 하였다. 대체로 다른 게이머와의 경쟁에서 이기고자 하는 욕망에 기인하여 유료 재화를 구매하기 때문에 이를 구매하지 않고 플레이하는 유저 그룹과 형평성 문제를 야기할 수 있고, 또 게임 안에서 유료 결제를 하는 유저가 그렇지 않은 유저에 비해 상당히 적기 때문에 매출을 올려주는 소수의 게이머에게 최대한 많은 이익을 거두고자 게임 서비스가 편중될 수 있다는 점이 언급되었다. 즉, 부분 유료화 모델은 게임의 본질인 재미보다 매출을 위한 과금 강요나 게임 내 불평등 문제를 초래하는 근본적인 한계를 지니고 있고 이러한 한계를 가진 부분 유료화 방식을 적용한 게임이 점차 많아지며 사회적인 문제로 비화하였다는 견해를 밝혔다(전홍준, 2017).

위 주장처럼 부분 유료화 모델이 게임의 밸런스를 파괴하고 과도한 현금 결제를 유도한다는 연구도 있었다. 유저 간 아이템 거래가 가능한 게임에서 부유한 유저는 결제한 게임머니로 고성능의 유료 아이템을 다른 유저에게 구입하고 이는 게임 내의 빈부격차의 확대로 이어졌다는 실례(實例)를 들었다. 또 사행성 논란을 빚을 정도로 바람직하지 않은 유료 결제 방식을 채용하는 풍조가 만연하여 많은 유저들에게 돈을 위해서만 게임을 개발한다는 인식을 심어주고 있다는 점도 강조하였다(신종현, 신현호, 2017).

잇따른 부분 유료화 모델의 폐단을 개선하고자 다양한 연구가 진행되었다. 부분 유료화 모델이 적용된 카카오의 모바일 게임을 중심으로 게임

장르별 과금 플레이 밸런싱 전략을 제안하였다(김병규, 2014). 또, 캐릭터의 외형 (스킨)과 스킬의 이펙트 효과에만 영향을 미치는 상품만을 판매하여 큰 성공을 거둔 <리그 오브 레전드>의 흥행에 주목하였고 게임 내 밸런스와 무관한 상품을 취급하는 부분 유료화 모델이 도입된 <오버워치>, <배틀그라운드> 역시 선풍적인 인기를 끄는 데 성공하였다고 거론하며 개선책을 제시하였다(맹주홍 외, 2018).

이러한 움직임이 있었음에도 온라인 게이머들의 불만은 사그라지지 않았다. 오히려 몇몇 게임 회사는 유저의 인내심을 시험하는 듯한 운영으로 이들을 자극했고 결국 서론에서 언급한 트럭 시위 등의 형태로 쌓여왔던 분노가 표출된 것으로 보인다. 한편 2022년 하반기에는 이러한 분위기 속에 확률형 아이템 뽑기를 배제하거나 최소화한 부분 유료화 게임이 출시되어 관심을 받기도 하였다(문영수, 2022; 이정훈, 2022; 조학동, 2022).

(3)유료 재화의 구매

유료 재화를 구매하는 유저의 특성에 대하여 조사한 연구는 다음과 같았다. 게임 내 뽑기 시스템에 비용을 지출하는 유저의 심리를 분석하였는데 바라던 캐릭터가 뽑히는 순간의 쾌락을 느끼고자 하는 도박적 심리가 작용하고 특정 캐릭터를 보유하고 있지 않을 때의 패배의식과 상대적 박탈감 해소, 그리고 수집 욕구의 충족 등과 같은 개인적 만족감을 위해 유료 재화를 결제한다는 사실을 밝혔다(김소울, 2016).

Guo and Barnes(2007), Lim and Seng(2010)을 바탕으로 유료 아이템 구매 동기는 온라인 게임의 이용 동기와 비슷하고 게임 내의 성취도, 사회적 영향력, 쾌락적 가치가 유료 재화 결제에 상당한 영향을 끼친다는 것을 밝힌 연구도 있었다. 나아가 유료 아이템을 구매한 집단과 그렇지 않은 집단에 대해 분산분석을 실시하여 게임 이용행태의 차이를 살펴보았다. 유료 아이템 구매자는 무료로 게임을 즐기는 게이머에 비해 게임 몰입도와 집중도가 더 높으며 게이머 간 교류의 비중, 무료 아이템의 소비

성향이 높은 것으로 나타났다(정광재, 2015).

PC 온라인 게임 중 MMORPG 장르의 25가지 게임을 플레이하는 유저를 대상으로 자아 존중감이 유료 아이템 구매에 미치는 영향을 분석하기도 하였다. 자아 존중감은 자신의 가치를 판단하는 자신만의 감정적 평가로 정의하였다. 현실에서의 자아 존중감은 아이템 구매와 무관하였으나 게임 속 자아 존중감이 클수록 아이템 구매 비용이 높았다. 그중에서도 자신의 능력에 대한 지각을 뜻하는 유능성 자존감은 아이템 구매 비용과 정비례하였고 친구, 타인 자존감은 아이템 구매 비용과의 유의미한 관계가 발견되지 않았다(김정진, 2013).

또한 개인적인 만족감을 충족시키기 위한 욕구가 유료 아이템의 구매로 이어졌다는 시각도 존재하였다. 액세서리, 옷, 헤어처럼 아바타를 꾸밀 수 있는 아이템을 제공함으로써 자기 자신의 개성을 온라인 상에서 반영할 수 있는 기회를 부여한 것이 부분 유료화가 성공할 수 있었던 요소 중 하나라고 짚었다. 이밖에 경제적이고 심리적인 부담의 완화가 부분 유료화의 성공 요인으로 선정되었다(Wallis, 2007).

한편 모바일 소셜 네크워크 게임인 <모두의 마블>을 플레이하는 유저들의 확률형 아이템 구매 동기를 조사한 결과, 기존 연구와 유사하게 게임 속 성취도, 쾌락, 사회적 요인, 지위 및 이익 추구가 유료 재화의 구매 결정에 관여하는 요인으로 나타났다. 특히 호기심, 개인적 만족, 심미성과 같은 쾌락과 관련된 변수가 강하게 작용하였는데 이는 게임 속 성취도만을 유료 재화의 구매 요인으로 간주하는 것에 유의하여야 한다는 시사점을 준다(이일수, 2017).

모바일 캐주얼 게임 중 퍼즐 장르를 즐기는 유저의 특성을 살펴보기도 하였다. 이들은 게임을 플레이하는 이용자의 수와 이용자의 플레이 만족도. 사용 편의성을 중요하게 여기고 향후의 게임 이용 여부에 영향을 주었다. 하지만 해당 요소들이 확률형 아이템 구매에 미치는 영향은 적은 편이었다(주현택, 2019).

제2절 집단 유형 분류

(1) 게임 유저의 유형 분류

게임 유저의 유형을 분류하여 진행한 기존 연구를 정리하고자 한다. 우선 MMORPG의 시초격인 머드 게임 유저는 이들의 행동과 성향, 채팅 등을 토대로 킬러 (Killers), 성취 (Achievers), 사교 (Socialisers), 탐험 (Explorers)의 네 가지 유형으로 나누어졌고 이 내용은 사분면으로 표현되었다. X축은 Players (좌) ↔ World (우), Y축은 Interacting (하) ↔ Acting (상)으로 구성되었다.

각 유형의 특징은 다음과 같았다. 먼저 킬러 집단은 일반적으로 말이 적고 도발적인 언행을 보였다. 또 다른 유저보다 강하다는 것을 입증하면서 즐거움을 느끼기 때문에 레벨이나 업적에 신경을 쓰며 유저 간의 경쟁 (PVP, Player Versus Player) 콘텐츠를 선호하였다. 성취 집단은 킬러 그룹과 비슷하게 그들의 능력치를 올리고 목표를 완수하는 것에 주력한다. 그러나 그들은 대화나 PVP 등의 다른 유저와의 상호작용에 무관심하였다. 성취 집단은 단시간에 높은 수준의 지위에 도달하는 것을 원하기 때문에 타인과 교류는 업적과 관련이 있거나 몰랐던 부분에 대해 질문하기 위한 경우로 국한되었다. 사교 집단에 속하는 이들에게 게임 세계는 그저 설정에 불과하였으며 캐릭터의 성장과 업적의 달성보다 다른 플레이어와 교류하며 지속적인 관계를 형성하기 위해 게임을 하였다. 그들은 개인적인 고민을 거리낌 없이 드러내고 타인의 이야기에 공감하는 모습을 보였다. 마지막으로 탐험 집단은 알려지지 않은 방법을 찾고, 미지의 영역을 개척하는 등 게임 세계와의 상호작용으로 게임에서 놀라움을 느끼고자 하는 유형이다. 이들은 업적이나 레벨과 같은 수치는 제한된 지성의 표시로 여기며 기꺼이 노하우를 공유하는 특징을 가지고 있었다(Bartle, 1996).

하드코어 유저와 캐주얼 유저로 구분하는 관점도 있었다. 하드코어 유

저의 특징으로는 게임을 많이 플레이하거나 구입하며 게임에 해박하다는 것이 있었다. 또 게임이 라이프스타일로 여겨지고 도전하는 것에 흥미를 느낀다고 하였다. 반면 캐주얼 유저는 적은 게임을 플레이하고 게임에 무지하며 시간을 보내거나 휴식을 목적으로 게임을 하는 유형이었다(Bateman, Boon, 2006). 또 폭력적 게임의 선호, 최신 PC 혹은 콘솔의 보유, 유료 결제 의향, 처음 게임을 했을 때의 나이 등 15가지의 척도로 극단적 캐주얼 게임 유저, 캐주얼 게임 유저, 과도기적/보통 유저, 하드코어 게임 유저, 극단적 하드코어 게임 유저로 분류하기도 하였다(Adams, 2002).

이밖에도 유저의 집단유형을 나누는 다른 의견이 존재하였다. 게임 취향, 유저 경험, 플레이 동기라는 평가 기준에 의거하여 의지적 몰입형, 일회성 유희형, 경쟁적 유희형, 중독적 몰입형으로 구분되었다(노경희, 2008). 그리고 공간/환경, 시간, 소셜, 게임의 진행 등 총 7개 범주로 구성된 29개의 분류 기준에 따라 세 집단으로 유저의 세분화가 이루어지기도 하였다. 물리적 환경 중심 플레이어, 콘텐츠 중심 플레이어, 감성 경험 플레이어로 나누어졌고 현실의 시간이 게임의 시간에 영향을 주는지에 대한 여부는 유형의 분류에 있어 중요한 요인으로 인식되지 않는다는 것을 확인하였다(이혜정, 민애홍, 2019).

게임소비성향도 유저를 구분하는 기준이 되었다. 게임소비성향에 대한 요인분석을 시행하여 이끌어낸 가치소비, 과시적 소비, 합리적 소비 등 6가지 요인을 추려내었다. 이를 바탕으로 k-means 군집분석을 실시하여 합리적 소비형 집단, 과시적 주변 의식형 집단, 수동적 소비형 집단, 마니아 소비형 집단으로 유형화하였다. 마니아 소비형 집단을 제외한 모든 유형에서 남성의 수가 많았고, 20대가 가장 많은 비중을 차지하였다. 또 게임지출비용이 많았던 마니아 소비형과 과시적 주변 의식형 집단의 월 수입이 타 유형에 비해 적은 것으로 드러났다(서경희, 2014).

다음 연구는 온라인 게임 내에서 구입한 유료 아이템의 종류에 따라 유저의 집단유형을 나누었다. 캐릭터의 능력을 올려주는 기능형 아이템,

외형을 꾸미는 감성형 아이템, 타인과의 관계 구축을 위한 사회적 아이템으로 구분하였고 해당 아이템 구매자 집단의 특성을 분석하였다. 기능성 아이템 구매자들은 주로 RTS (Real-Time Strategy) 장르를 많이 하고, 감성형 아이템 구매자들은 롤플레잉 게임을 선호하였다(Ho&Wu, 2012). 상기 연구에서 착안하여 사회적 아이템을 제외한 두 유형의 아이템 구매자 집단을 분석하기도 하였다. 감성형 아이템 구매자는 과금 만족도, 디자인, 자기만족, 게임사 우호도의 항목이 높았고, 기능형 아이템 구매자는 타 게이머와의 상호작용 항목에서 높은 수치를 나타낸다는 것을 밝혀내었다(위정현, 김은비, 2019).

모바일 게임을 즐기는 이용자를 분류한 연구도 꽤 있었는데 그중에서도 방치형 (IDLE) 게임을 플레이하는 유저의 유형화를 시도한 사례가 있었다. IDLE 게임이란 2000년대 초에 나타난 표현으로 플레이어가 관여하지 않아도 경험치가 오르고 보상이 수집되는 등의 자동화 요소를 갖춘 콘텐츠를 가리킨다. 전문가에게 게임 요소 30개 중 7개를 선택하도록 하여 추출된 게임성, 조작감, 완성도, 오락성, 그래픽, 접근성, 난이도를 기준으로 다중 유희 군집, 단일 몰입 군집, 단기 유희 군집으로 나누었다. 다중 유희 군집은 캐릭터와 스토리에 몰입하지 못해 게임에 대한 집중도가 높지 않았으나 게임을 쉽게 포기하지 않았고 인터랙션 (상호작용)의 정도에 구애받지 않고 게임을 즐겼다. 단일 몰입 군집은 적극적으로 개입하는 것을 선호하였고 캐릭터와 게임의 시나리오에 대한 몰입도가 높았다. 단기 유희 군집은 게임 몰입도, 지식, 비용 투자에서 가장 낮은 수치를 기록했지만 모든 장르에 대한 선호도가 고르게 분포되어 넓고 얕게 게임을 즐기는 유형임을 알 수 있었다(김예은, 2021).

(2) 군집분석

군집분석 (클러스터링, Clustering)은 데이터셋을 군집으로 나누는 방법이다. 동일한 집단 안의 데이터 포인트끼리는 상당히 유사하고 다른 집

단의 데이터 포인트와는 구분되도록 데이터 군집을 형성하는 것이 목표이다. 분류 알고리즘과 비슷하게 군집 알고리즘은 각 데이터가 어떤 집단에 속하는지 할당한다(안드레아스 뮐러, 세라 가이도, 2017).

1) 군집화 기법

① k-평균 군집화

k-평균 군집화 (k-means Clustering)는 최초로 개발된 클러스터링 기법으로 알고리즘이 비교적 간단하고 데이터의 크기가 커도 사용할 수 있어 강점을 지닌다. 이 알고리즘은 데이터를 k개의 군집으로 나누는데 이때 할당된 클러스터의 평균과 속한 데이터 간의 거리 제곱합이 최소가 되도록 한다. k-means 알고리즘은 각 군집의 크기가 동일하다는 보장은 없으나 각 군집이 최대한 멀리 떨어지도록 한다(피터 브루스 외, 2018).

k-means 알고리즘은 우선 데이터 집합에서 k개의 초기 중심점을 무작위로 선택하는 초기화 단계, 중심에 인접한 모든 데이터 요소가 군집을 이루며 데이터 포인트와 모든 중심점 사이의 유클리드 거리를 기반으로 두 점 사이의 직선이 그어지고 경계선을 기준으로 군집이 나뉘는 클러스터 할당 단계를 거친다. 다음으로 센터가 필요한 클러스터가 있다면 가까운 중심점과 새로운 경계를 생성하여 군집을 만드는 중심 이동 단계에 접어드는데 클러스터 할당과 중심 이동 단계에서 중심점이 움직이지 않을 때까지 이 과정을 반복한다(이영호, 이수현, 2021).

② k-medoids 군집화

k-평균 군집화는 계산 시간의 관점에서 굉장히 효율적이지만 노이즈와 아웃라이어에 민감한 것으로 알려져 있다. 이에 k-medoids 알고리즘에서

는 중심점 대신 medoids라는 개념을 도입하여 상기 문제를 극복하고자 하였다. 클러스터에서 가장 중앙에 있는 개체를 기반으로 하기에 k-means 클러스터링에 비해 이상값에 덜 민감한 것이 특징이다. k-medoids 군집화 기법 중 PAM(Partitioning Around Medoids)이 가장 강력한 것으로 간주되는데 연산에 긴 시간이 소요되어 대용량 데이터셋에는 부적합하다는 결점이 존재한다(Park and Jun, 2009).

③ k-prototypes 군집화

보편적으로 데이터는 수치형과 범주형으로 구분되는데 k-prototypes 알고리즘은 수치형 자료에는 k-means 알고리즘을 적용하고 범주형 자료에는 k-modes 알고리즘을 적용하는 방식을 취한다. 그래서 수치형 속성과 범주형 속성이 혼합된 데이터를 군집분석하는 것에 적합하지만 두 알고리즘을 단순히 결합한 방식이기 때문에 초기치 선정 문제를 해결하지 못하였다는 문제를 안고 있다.

k-prototypes 군집화는 먼저 데이터에 대한 초깃값 k를 지정한 다음 k와 각 데이터를 속성별로 구분하고 가중치를 적용하여 유사도 거리를 측정한다. 이렇게 얻은 유사도 값에 따라 클러스터를 할당하며 군집 갱신이 이루어지는데 군집의 변화가 보이지 않을 때까지 속성별 구분부터 군집 갱신까지의 절차를 되풀이한다(오수민, 2014)

2) 초기치 지정 및 성능평가

① 엘보우 기법

위에서 언급한 군집화 기법을 사용하기 위해서는 초기치인 k를 지정해 주어야 한다. 엘보우 기법 (Elbow Method)는 클러스터의 개수를 의미하

는 k를 입력받는 군집화 알고리즘에서 적절한 군집의 수를 결정하는 방법 중 하나이다.

엘보우 기법은 초기 k값을 다소 낮게 지정하고 점차 증가시키면서 클러스터 비용을 계산한다. 클러스터 비용은 각 클러스터의 중심점과 클러스터 내부에 있는 데이터 간 거리합의 평균이다. 이 과정에서 k값에 따라 비용이 급격하게 줄어드는 구간과 완만하게 감소하는 구간이 존재하는데 해당 구간 사이의 k값을 엘보우라고 한다. 예를 들어, 직전의 k값에 대한 클러스터 비용과 현재 k값에 대한 클러스터 비용의 차가 크지 않으면 직전의 k값을 엘보우로 판단하여 최적의 군집 수로 결정할 수 있다(김동혁, 2016).

② 실루엣 분석

실루엣 (Silhouette) 분석은 실루엣 계수 (Silhouette Coefficient)를 활용하여 클러스터링의 유효성 파악하는 방법으로 군집분석의 평가와 적정 군집 수 k를 선택하는 데도 사용된다. 실루엣 계수는 데이터가 속한 클러스터에 밀집되어 있으면서도 속하지 않은 클러스터로부터는 잘 분리된 것을 객관적으로 나타내는 지표이다.

실루엣 계수를 계산하기 위해서는 응집도와 분리도를 구해야 한다. 응집도는 특정 군집에 포함된 데이터가 해당 군집의 중심에 얼마나 가까이 있는지를 측정하기 위한 지표인데 특정 클러스터에 속한 개별 데이터와 클러스터 중심 간 거리의 제곱합으로 계산된다. 클러스터 중심은 해당 클러스터에 포함된 데이터의 평균적인 위치이다. 분리도는 각각의 군집이 다른 군집으로부터 얼마나 잘 구분되는지 나타내는 지표이며 각 클러스터 중심 간 거리의 제곱합으로 계산한다. 계산식은 아래와 같다.

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)}$$

(a_i : 해당 데이터가 속한 군집의 모든 데이터 간 거리의 평균,
b_i : 해당 데이터가 속하지 않은 군집 중 가장 가까운 군집과의 평균 거리)

이렇게 도출된 실루엣 계수는 -1과 1 사이의 값으로 나타난다. 실루엣 계수가 1에 가까울수록 데이터가 속한 군집의 중심에 가깝고 타 군집으로부터 멀리 떨어져 바람직하게 군집이 형성된 것으로 해석할 수 있다. 실루엣 계수는 계산 방법 상 단위가 없고 데이터셋 전체의 이동 및 회전, 확대 및 축소에도 변하지 않으며 데이터 분포에 따른 군집 수 혹은 구조에 의한 영향만 받는 특성을 지닌다(Lee, Kim, Lee, 2012).

③ 주성분 분석

칼 피어슨에 의해 고안된 주성분 분석 (PCA, Principal Component Analysis)은 복수의 수치형 예측변수를 보다 적은 수의 변수 집합으로 나타내기 위해 사용한다. 이때 생성되는 새로운 변수들은 기존 변수들에 가중치를 적용한 선형결합을 뜻한다. 전체 변수들의 변동성을 대부분 설명할 수 있는 적은 수의 변수 집합을 주성분이라 하는데 이를 활용하여 데이터의 차원을 줄일 수 있다. 주성분을 만드는 데 사용된 가중치는 새로운 주성분을 만드는 데 있어 기존의 변수가 얼마나 기여하는지를 나타낸다. (피터 브루스 외, 2018)

본 연구에서는 군집분석 후 시각화 과정에서 여러 차원의 변수를 축소하여 그래프로 나타내기 위해 사용되었다.

제 Ⅲ 장 연구 방법

제1절 연구대상 및 절차

본 연구는 온라인 게임을 즐기는 게이머 중 유료 재화의 구입에 비용을 지불하는 유저를 복수의 군집으로 구분하고 군집별 특성을 비교 분석하여 유의미한 시사점을 도출함으로써 온라인 게임 유료 이용자에 대한 이해를 제고하는 것이 목적이다. 이를 위해 한국콘텐츠진흥원에서 2017년도부터 2021년까지 실시한 5년분의 게임이용자 실태조사 로데이터 (Raw Data)를 제공받아 활용하였다. 게임이용자 실태조사는 전국의 만 10세-65세의 국민을 대상으로 온라인과 전문 면접관에 의한 1:1 면접조사를 병행하여 이루어졌다.

전반적인 연구 프로세스는 데이터 수집, 데이터 전처리, 데이터 분석, 시각화, 결론 도출 순으로 진행하였다. 데이터 전처리부터 시각화에 이르는 과정에서 Python 3.9.12 버전을 사용하였다.

제2절 연구방법

(1) 데이터 수집

위에서 언급하였듯이 연구에 활용하고자 한국콘텐츠진흥원으로부터 게임이용자 실태조사의 5년도분 (2017년-2022년)에 해당하는 로데이터를 제공받았다. 총 15,154개의 데이터를 수집하였으며 조사항목은 게임 일반에 대한 이용 실태, 게임 분야별 이용 현황 및 특성, VR 게임 이용 경험, 게임에 대한 인식 및 태도, 응답자 특성의 다섯 가지 카테고리로 나뉜다. 이 중 연구 목적에 부합하는 게임 일반에 대한 이용 실태, 게임 분야별 이용

현황 및 특성, 응답자 특성, 3개의 구분을 토대로 데이터 전처리를 실시하였다.

(2) 데이터 전처리와 탐색적 데이터 분석

MS-Excel과 Python으로 데이터를 살펴본 결과, 조사 문항이 매년 달라지고 변수의 이름 또한 다르다는 것을 확인했다. 이에 따라 Python으로 연도별 데이터를 불러와 정제하고 이것이 마무리되면 모든 데이터를 통합하는 방향으로 진행하였다.

1) 유료 재화 구매 관련 데이터 정제

가장 먼저 PC 게임과 모바일 게임의 유료 재화 구매 관련 데이터 정제에 착수하였다. 계정이용료와 PC방 사용료 등 모든 비용이 포함된 월평균 이용 및 구매 비용, 그리고 이 중 게임머니와 게임 아이템과 같은 유료 재화에 사용한 금액을 뜻하는 월 평균 게임 내 결제비용, 약 1년 전부터 조사일까지의 확률형 아이템 총 지출액 변수의 결측치를 0으로 설정하였다. 결측치라고 하여 모두 제거하는 것보다 조금이나마 더 보존하는 것이 유의미한 결과를 도출할 수 있을 것으로 판단하였고 PC와 모바일 각 3개씩, 총 6개 항목에 대한 파생변수를 만들기 위해서 계산이 가능하도록 처리할 필요가 있었기 때문이다. PC와 모바일 게임, 각각의 항목을 합산하여 3개의 파생변수를 추가하였는데 온라인 게임 이용 및 구매 총액, 온라인 게임 내 구매 총액, 확률형 아이템 구매 총액이 그것이다.

2) 변수 선정 및 기타 데이터 정제

위 과정에서 생성된 파생변수를 포함한 컬럼 중에서 VR 게임과 같은

연구 주제와 무관한 것은 배제하였다. 또 복수선택 항목과 명목형 데이터로 표기된 항목도 제외하였고, 연령대와 가구소득처럼 비슷한 의미를 지닌 변수도 삭제하였다. 최종적으로 수치로 표현된 데이터 중 5년간의 조사에서 일관적으로 조사된 변수만 남았다고 볼 수 있다.

불필요한 변수를 줄인 후, 응답자 특성 카테고리에 해당하는 데이터를 정제하였다. 자녀 연령대로 구분된 컬럼은 자녀 유무라는 파생변수를 만들어 정리하였다. 자녀 연령대에 1로 표시하였다면 자녀 유무에도 1로 나타내었고 자녀 없음 항목에 체크한 경우, 0으로 부정의 의미를 담았다. 한편 1과 0 이외의 값은 전부 0, 자녀 없음으로 처리하였다. 상기 정제 과정을 거친 2017년 조사 자료 중, 자녀 연령대와 자녀 없음 항목에 모두 응답하지 않은 444명이 있었는데 평균 초혼 연령 32세 이하를 걸러낸 결과 362명, 즉 약 82%의 비율로 미혼인 상태였기 때문이다. 참고로 평균 초혼 연령은 통계청의 2017년 통계에 남성의 평균 초혼 연령은 32.9세, 여성은 30.2세로 이것의 평균을 반올림하여 32로 계산하였다.

다음으로 미혼, 기혼, 이혼/사별로 나뉜 결혼 여부 컬럼을 배우자 유무라는 파생변수로 표현하였다. 만 18세 이하는 혼인 적령 나이를 만 18세로 규정한 민법 제807조에 따라 결혼할 수 없으므로 모두 미혼으로 처리하였다. 미혼과 이혼/사별을 합하여 배우자 없음을 0, 나머지를 배우자가 존재하면 1로 나타내었다.

소득에 관한 변수는 가구소득과 개인소득 중 후자를 채택하였고 여기에 초등학생부터 고등학생를 대상으로는 한 달 용돈 데이터를 추가하여 개인소득 및 용돈이라는 파생변수를 만들었다. 원데이터에서 9개의 선택지로 구분된 용돈 컬럼의 값을 3개로 합쳤다. 3만원 미만을 1, 3만원 이상 10만원 미만을 2, 10만원 이상을 3으로 정리하였다. 그리고 개인소득의 8개 선택지에 3을 더해 용돈 데이터와 통합하였다. 4로 표기한 100만원 미만의 단계로 시작해 100만원씩 범위가 확장되고 최종 단계는 11로, 700만원 이상을 표현하도록 수정하였다.

최종학력은 원래 7개로 구분되어있던 선택지를 고졸 이하를 1, 대학교

재학 및 졸업은 2, 대학원 재학 및 졸업은 3으로 압축하였다. 고등학교 3학년에 해당하는 만 18세 이하의 결측치는 고졸 이하로 취급하였고, 이외의 결측치는 삭제하였다.

이상으로 응답자 특성에 대한 데이터 정제를 마치고 게임 분야별 이용 및 실태 카테고리의 문항 중 PC와 모바일 게임에 대한 데이터를 가다듬었다. 시간에 관한 데이터로는 주중과 주말로 나뉘어 조사된 하루 평균 이용시간과 1회 평균 이용시간이 있었다. 시간과 분 단위로 분할된 데이터를 시간 단위로 반올림해 소수점 아래 둘째 자리까지 나타내었다. 그리고 시간 단위로 표현된 각각의 데이터를 합해 온라인 게임에 투자하는 시간을 살펴보고자 하였다.

약 1년 전부터 조사일까지 이용한 모든 게임 수를 의미하는 총 이용 개수, 그중 3개월 이상 지속한 게임을 뜻하는 주 이용 개수 또한 위의 과정과 유사하게 온라인 게임 총 이용 개수, 온라인 게임 주 이용 개수라는 2개의 파생변수를 만들어 통합하였다.

이용 빈도 데이터는 빈도가 높을수록 작은 값으로 정의되어 있었다. (1 : 주 6-7일, 5 : 연 1-11회) 그래서 가독성을 높이고자 역순으로 변경하여 비례하도록 나타내었다. 즉, 적게 플레이할수록 낮은 값으로 표현되는 것이다. 결측치는 모두 0으로 대체하여 전혀 게임을 이용하지 않는 경우로 간주하였다.

온라인 게임 시작연령 데이터는 2017년을 제외하고는 PC와 모바일 게임의 시작연령 중 작은 값을 취했다. 2017년은 PC 게임의 시작연령을 조사하지 않아 모바일의 데이터만 사용하였다. 결측치는 전부 제거하였다.

마지막으로 조사 연도라는 파생변수를 추가하였으며 조사 전년 게임 이용 여부 (TYPE)와 최근 게임 이용 여부 (SQ4A) 항목이 존재하는 특정 연도에 대해서는 이 둘을 대조함으로써 데이터의 신뢰성을 향상하고자 하였다. 예를 들어 조사 전년에 이용하지 않았다고 했으나 (TYPE : 2) 조사 전년도 상반기에 게임을 한 것으로 (SQ4A : 2) 답변했다면 모순되므로 삭제하였다.

3) 변수명 수정 및 데이터 통합

결측치 대체 및 제거 등의 작업을 진행한 후 5년분의 데이터를 통합하고자 연도별로 다른 변수명을 동일하게 수정하였다. 또 앞으로 필요하지 않거나 연구 주제와 연관성이 떨어지는 변수를 제거하였다. 이 절차를 거쳐 처음 15,154건의 데이터 중 불필요 데이터와 결측치를 걸러내고 9,970건을 확보하였다. 변수는 다음 <표 1>처럼 14개로 정리하였다. 이후, 글이 불필요하게 길어지는 것을 방지하고자 가급적 <표 1>의 번호를 사용하여 변수를 표현하였다.

<표 1> 변수 요약

번호	변수	비고
1	온라인 게임 시작연령	
2	주중 하루 평균 이용시간 (온라인)	
3	주말 하루 평균 이용시간 (온라인)	
4	주중 1회 평균 이용시간 (온라인)	
5	주말 1회 평균 이용시간 (온라인)	
6	온라인 게임 총 이용 개수	조사일 기준 약 1년 전부터 플레이한 모든 게임 수
7	온라인 게임 주 이용 개수	3개월 이상 지속적으로 이용한 게임 수
8	온라인 게임 내 구매 총액 (월 평균)	게임머니/아이템 구입 비용
9	성별	1 : 남자, 2 : 여자
10	나이	
11	자녀 유무	0 : 없음, 1 : 있음
12	배우자 유무	0 : 없음, 1 : 있음
13	학력	1 : 고졸 이하 2 : 대학교 재학 및 졸업 3 : 대학원 재학 및 졸업
14	개인소득 및 용돈	한 달 기준 1 : 3만원 미만 2 : 3만원 이상 10만원 미만 3 : 10만원 이상 4 : 100만원 미만 … 11 : 700만원 이상

4) 탐색적 데이터 분석

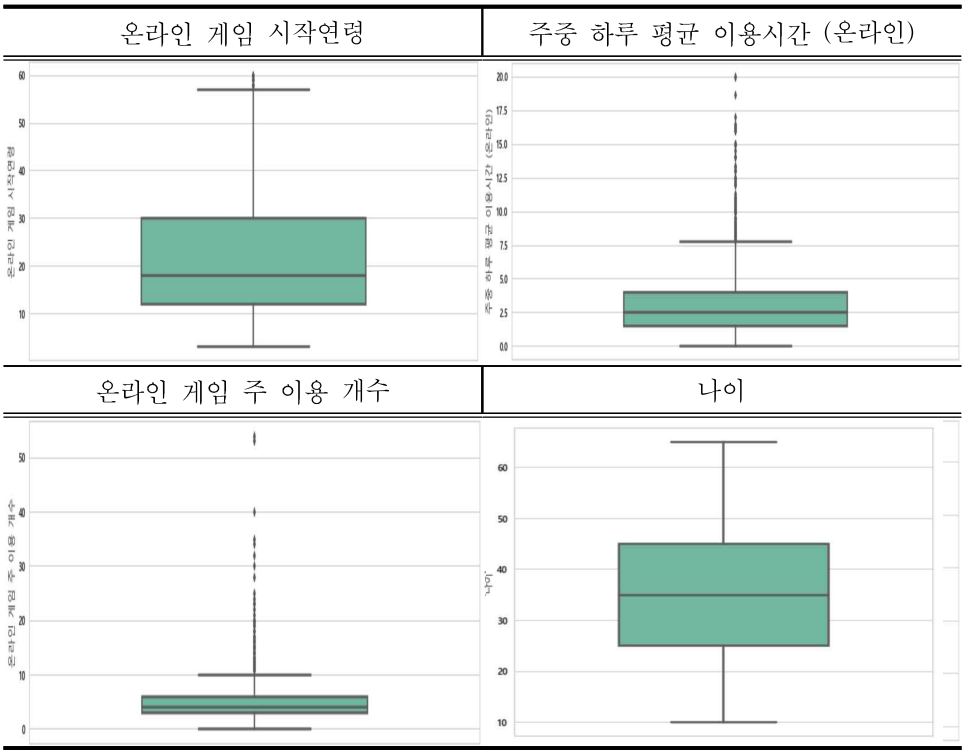
이어서 유료 이용자로 범위를 제한하기 위해 3번째 단계까지의 기초적인 데이터 전처리 과정으로 얻은 9,970건의 데이터 중 온라인 게임 내 구매 총액 변수의 값이 0보다 큰 데이터를 필터링하였다.

<표 2> 데이터셋의 인구통계학적 특성

구분		총원	비율 (%)	유료 이용자	비율 (%)
연령대	10대	1,721	17.3	503	12.6
	20대	2,112	21.2	1,012	25.4
	30대	1,992	20.0	959	24.0
	40대	1,837	18.4	888	22.3
	50대	1,700	17.1	529	13.3
	60대	608	6.1	98	2.5
성별	남성	5,298	53.1	2,433	61.0
	여성	4,672	46.9	1,556	39.0
배우자 유무	있음	4,722	47.4	1,846	46.3
	없음	5,248	52.6	2,143	53.7
자녀 유무	있음	3,694	37.1	1,330	33.3
	없음	6,276	62.9	2,659	66.7
학력	고등학교 졸업 이하	3,419	34.3	948	23.8
	대학교 재학 및 졸업	5,824	58.4	2,665	66.8
	대학원 재학 및 졸업	727	7.3	376	9.4
개인소득 및 용돈	3만원 미만	665	6.7	130	3.3
	3만-10만원 미만	647	6.5	214	5.4
	10만원 이상	197	2.0	80	2.0
	100만원 미만	2,225	22.3	710	17.8
	100만-200만원 미만	1,443	14.5	535	13.4
	200만-300만원 미만	2,030	20.4	910	22.8
	300만-400만원 미만	1,344	13.5	663	16.6
	400만-500만원 미만	676	6.8	356	8.9
	500만-600만원 미만	344	3.5	179	4.5
	600만-700만원 미만	157	1.6	90	2.3
	700만원 이상	242	2.4	122	3.1
총계		9,970	100.0	3,989	100.0

그 결과 <표 2>와 같이 5,981개의 데이터가 걸러져 3,989건의 데이터가 남게 되었다. 남성의 비율이 8%p가량 증가하였고, 10대와 50-60대의 비율은 감소하였다. 배우자와 자녀가 있는 사람의 비중은 줄어들었으며 대학교 이상의 학력을 지녔다고 응답한 인원의 비율은 상승하였다. 10대가 감소함에 따라 당연히 용돈을 받는 비율도 줄었고, 개인소득이 200만원 미만인 그룹을 제외하면 유료 이용자로 범위를 줄이면서 그 비율이 모두 상승한 것이 눈에 띄었다.

<표 3> Box plot 활용 이상치 표현



다음으로는 이상치 여부를 살펴보았다. 먼저 Tukey Fences로 알려져 있는 사분범위 (IQR)에 기초한 방법을 시도하였다. Python의 Seaborn 라이브러리를 활용하여 상자 수염 그림 (Box plot)을 그려 시각화하였고

<표 3>에 그중 일부를 정리하였다. 이상치는 1번 변수에서 12개, 2번 변수에서 208개, 7번 변수에서 187개가 발견되었고, 다른 변수에도 존재한다는 것을 확인하였다. 이 방법으로는 2,029개의 데이터가 남게 되었다.

Z-Score (표준점수)를 활용한 방법으로도 이상치 처리를 진행하였다. 식은 아래와 같다.

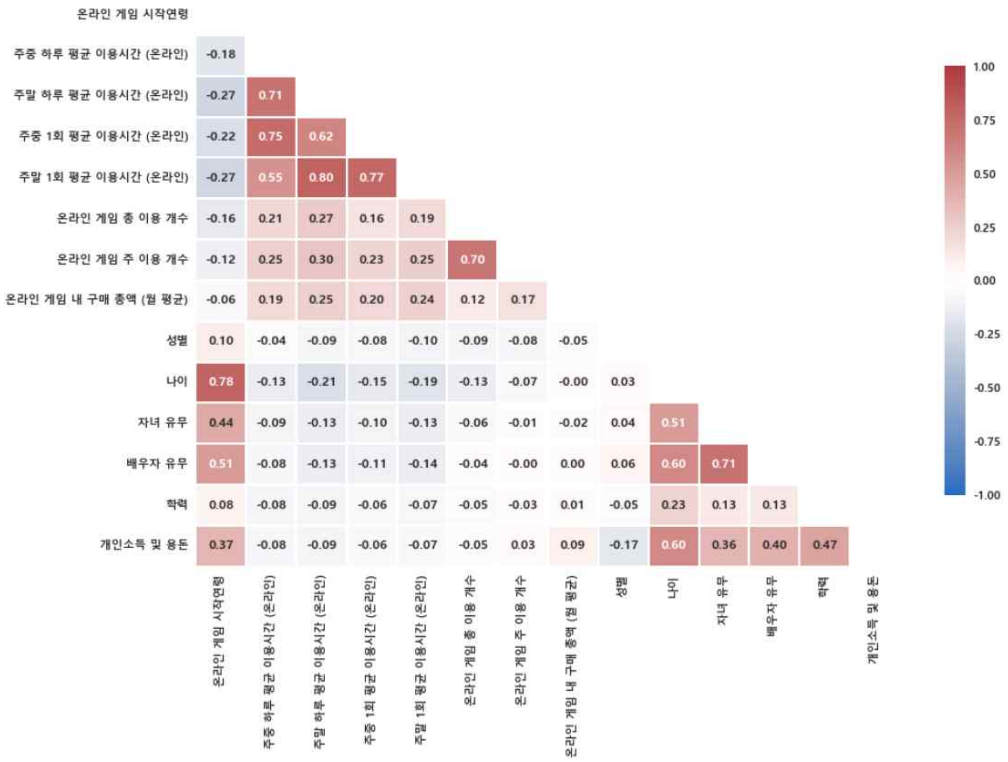
$$z = \frac{x - \mu}{\sigma}$$

(σ : 모집단의 표준편차, μ : 모집단의 평균)

최대한 많은 데이터를 보존하고자 Z-Score의 절댓값이 3을 벗어나는 것들만 이상치로 취급하였다. 그 결과, 이 방법으로는 기존의 3,989개 중 3,758개의 데이터가 잔존하는 것을 확인하였다.

변수별 이상치를 살펴본 다음, 변수 간의 상관관계를 조사하였다. 상관계수는 -1에서 1사이의 값을 가지는 피어슨 상관계수를 활용하였다. 군집분석의 성능을 높이고자 유사한 변수 중 강한 상관관계를 보이는 항목은 하나의 변수만을 취하였다.[1] 하루 평균 이용시간과 1회 평균 이용시간은 주중의 수치만을 사용하였고, 게임 이용 개수도 이에 따라 주 이용 개수만을 활용하여 군집분석을 시행하는 것으로 결정하였다.

1) 상관관계 해석 (Rea & Parker, 2005)
- 0.0-0.1 거의 관계없음
- 0.1-0.2 약한 양의 상관관계
- 0.2-0.4 보통 양의 상관관계
- 0.4-0.6 비교적 강한 양의 상관관계
- 0.6-0.8 강한 양의 상관관계
- 0.9-1.0 매우 강한 양의 상관관계



<그림 2> 변수 간 상관관계

(3) 군집분석

데이터 전처리와 탐색적 데이터 분석을 통해 군집분석을 적용할 11개의 변수 (3, 5, 7번 변수 제외)를 선정하였다. 그리고 변수 간의 범위 차이로 인한 영향을 최소화하도록 정규화와 표준화를 적용한 피처 스케일링(Feature Scailing)도 실시하였다. 정규화는 데이터를 0과 1 사이의 값이 되도록 하는 것이고, 표준화는 표준정규분포의 속성을 갖도록 조정하는 것이다. 본 연구에서는 Python의 머신러닝 라이브러리인 scikit-learn에서 MinMaxScaler와 StandardScaler를 불러와 진행하였다.

11개의 변수가 있는 데이터셋에 표준화만 적용한 경우, 정규화만 적용

한 경우, 두 가지를 모두 적용한 경우, 이상치를 제거한 경우 등으로 나누어 k-means, k-medoids, k-prototypes 알고리즘을 적용하여 각각의 성능을 비교하였다. k-means, k-medoids 알고리즘은 scikit-learn 라이브러리의 KMeans, KMedoids 모듈을 활용하였다. 언제나 동일한 결과를 얻기 위해 KMeans의 초매개변수 (하이퍼파라미터, Hyperparameter) 중 하나인 random_state값을 5로 지정하였다. KMedoids는 random_state값은 동일하였고 추가로 method를 'pam', init을 'k-medoids++'로 설정하였다. k-prototypes 알고리즘을 적용하고자 kmodes라는 라이브러리에서 KPrototypes 모듈을 불러왔고 하이퍼파라미터는 random_state값을 5, init을 'Huang'으로 하였다.

최적의 클러스터 수를 나타내는 k값을 찾기 위해 엘보우 기법과 실루엣 분석을 활용하였고, 실루엣 계수는 군집분석의 성능 척도로써 사용되기도 하였다. 또 주성분 분석을 적용하여 2차원 도표로 시각화하는 것으로 군집이 제대로 형성되었는지 확인하였다. 주성분 분석을 위해 scikit-learn 라이브러리의 PCA 모듈, 시각화를 위해 Seaborn과 Matplotlib 라이브러리를 불러와서 사용하였다.

(4) 군집별 특성 분석

형성된 군집 간의 차이점을 크게 두 가지 측면에서 조사하였다. 먼저 게임을 어떻게 플레이하는지에 대하여 이용 양태의 관점에서 접근하였다. 게임 시작연령, 주중·주말의 하루 및 1회 이용시간, 플레이하는 게임의 개수 등이 이와 관련된 변수이고, 해당 변수를 대상으로 분산분석(ANOVA)을 시행하여 각 군집 간에 유의미한 차이가 있는지 살펴보았다. 분산분석을 실시하기 위하여 Python에서 statsmodel 라이브러리의 anova_lm 모듈을 활용하였다.

다음으로 인구통계학적 관점에서도 네 클러스터의 특성을 파악하고자 하였다. 성별, 연령대, 배우자 유무, 자녀 유무, 학력, 개인소득 및 용돈 등

의 변수에 대하여 교차분석을 진행하였다. pandas로 각 변수와 군집이 들어간 crosstab을 만들고 scipy 라이브러리에 있는 chi2_contingency 모듈을 사용하여 카이제곱 통계량과 p-value, 자유도 등을 도출하였다.

제 Ⅳ 장 분석 결과

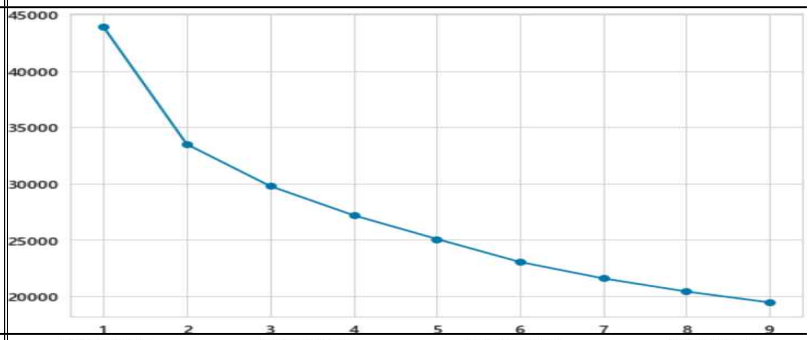
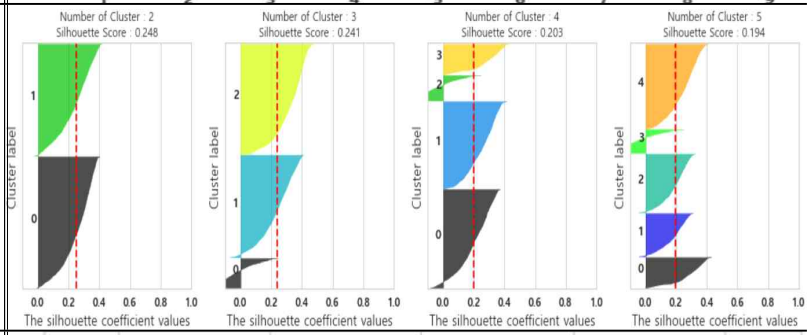
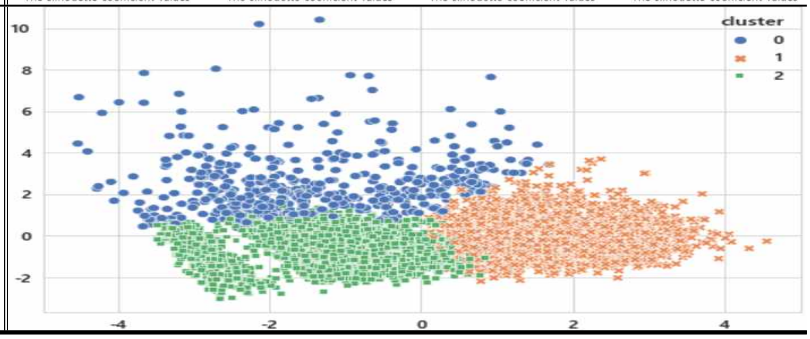
제1절 군집분석 결과

군집분석의 결과는 3장에서 언급하였듯이 세 가지의 군집화 알고리즘과 다양한 데이터 전처리 기법을 활용하여 성능을 비교한 내용을 도식화하여 정리하였다. 군집화 알고리즘의 종류와 더불어 표준화, 정규화, 이상치 처리 여부에 따라서도 군집분석 결과와 성능이 다른 것을 확인하였다.

(1) k-means

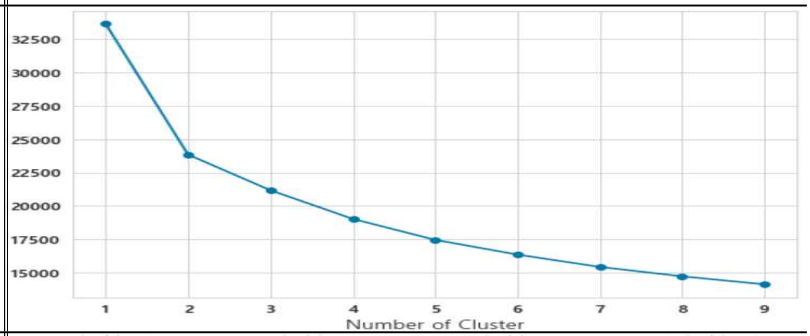
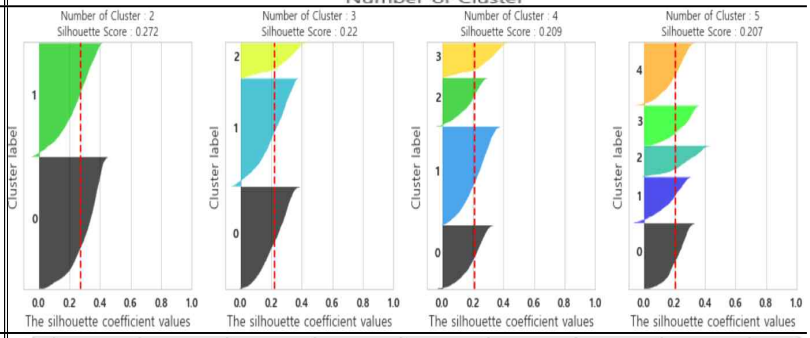
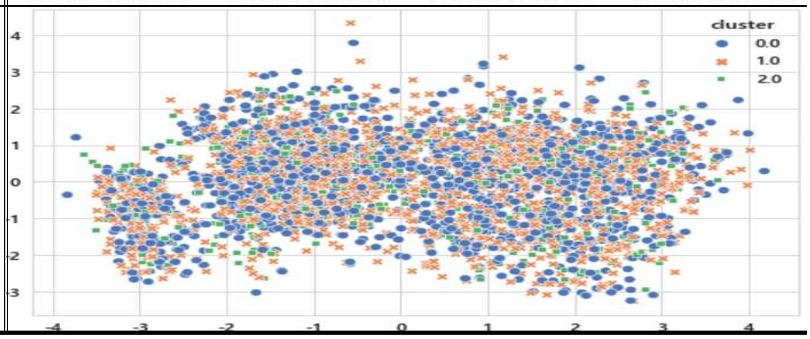
먼저 준비된 데이터셋에 표준화만 적용하여 군집분석을 시도하였다. 엘보우 기법과 실루엣 분석에 근거한 적절한 군집 수 k값은 2로 보이지만 단 2개의 군집으로 나누면 유저의 특성을 반영하는 것이 어렵다고 판단하여 3개로 설정하였다. 그 결과, <표 4>와 같은 내용을 확인하였다. 전반적으로 실루엣 계수가 낮았고 파란색 동그라미로 표시된 0번 군집의 데이터가 광범위하게 분포하고 있으나 주황색 가위표로 표현된 1번과 초록색 네모로 그려진 2번 군집에 속하는 데이터는 비교적 밀집하여 있는 것을 알 수 있었다.

<표 4> 표준화 적용 군집분석 (k-means)

데이터셋	A. 표준화
엘보우 기법	
실루엣 분석	
군집 시각화 (k=3)	

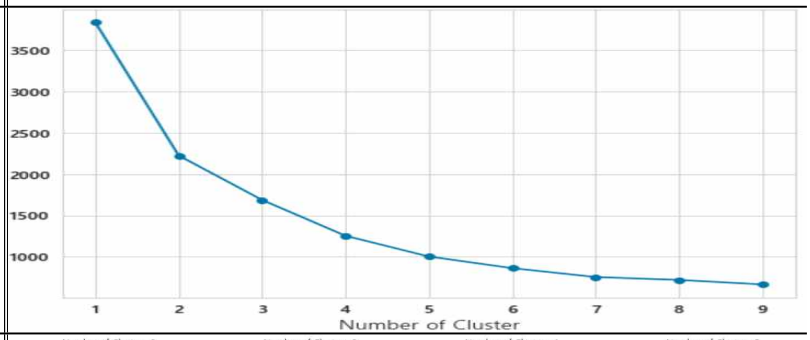
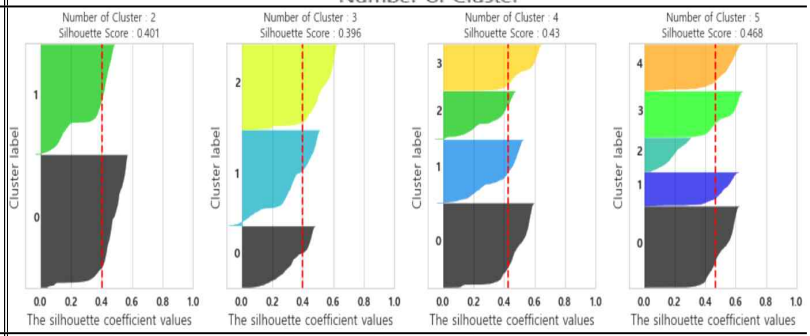
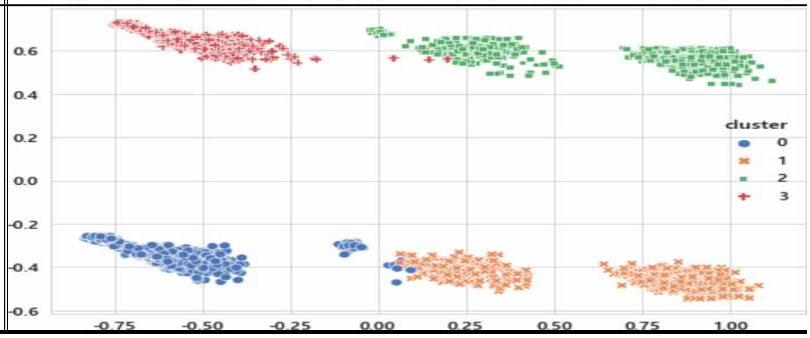
두 번째로는 표준화 적용 후, Z-Score가 절댓값 3을 초과하는 데이터를 아웃라이어로 간주한 데이터셋으로 군집분석을 실시하였다. 상기 케이스와 마찬가지로 적정 군집 개수를 3으로 설정하였다. 다만 첫 번째 경우와 다르게 데이터끼리의 군집이 전혀 보이지 않고 모두 섞여버리는 양상을 보였다.

<표 5> 표준화 적용 및 이상치 처리 군집분석 (k-means)

데이터셋	B. 표준화 적용 및 이상치 처리 (Z-Score)
엘보우 기법	
실루엣 분석	
군집 시각화 (k=3)	

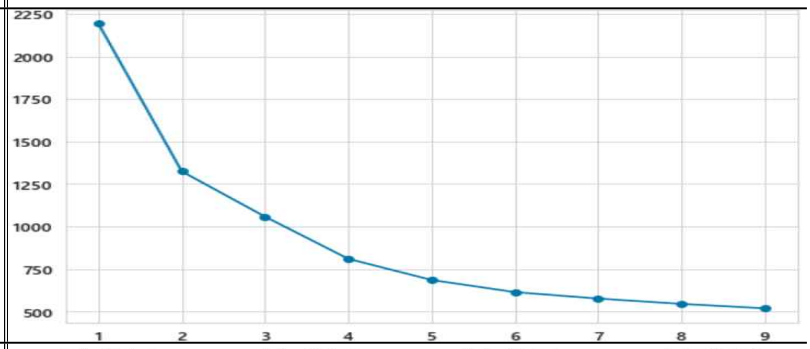
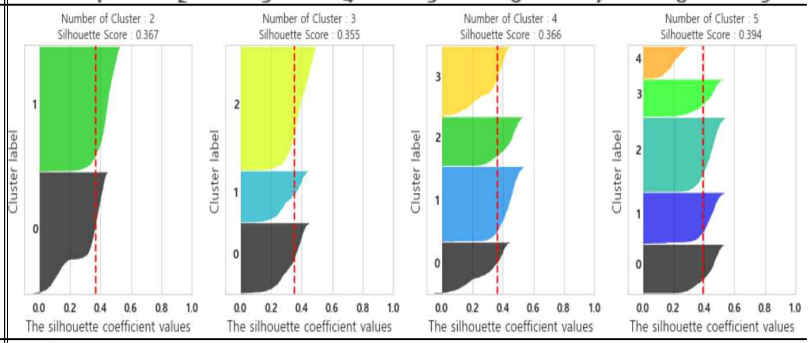
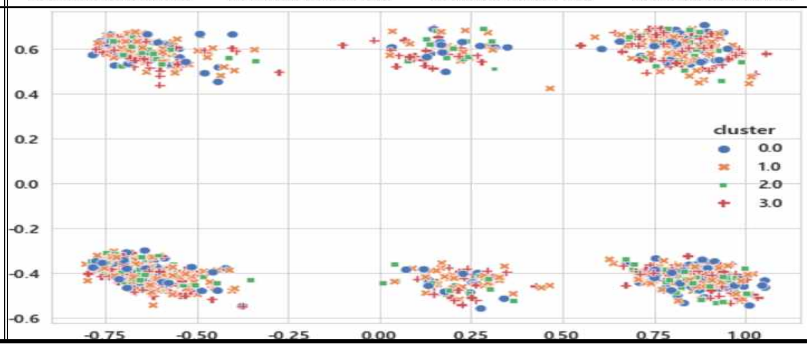
세 번째로는 데이터셋에 정규화만을 적용하여 군집분석 결과를 조사하였다. 이전과 달리 k가 3인 경우를 제외하면 실루엣 스코어가 0.4를 넘어 성능이 향상된 것을 알 수 있었다. k값은 4로 하였는데 실루엣 계수가 높으면서도 군집별 실루엣 계수의 평균 차이가 비교적 균일하다고 생각했기 때문이다. 시각화 결과로도 군집이 형성된 것을 확인하였다.

<표 6> 정규화 적용 군집분석 (k-means)

데이터셋	C. 정규화 적용
엘보우 기법	
실루엣 분석	
군집 시각화 (k=4)	

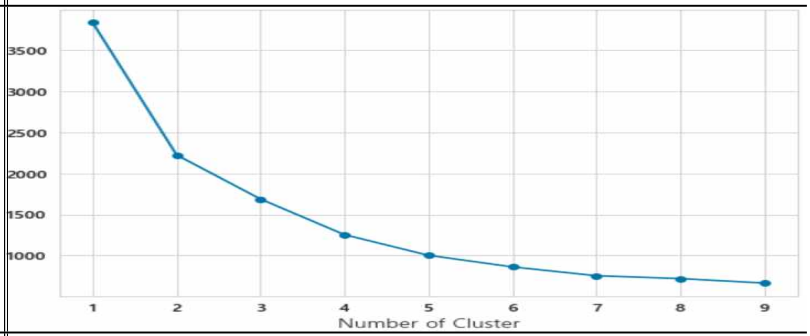
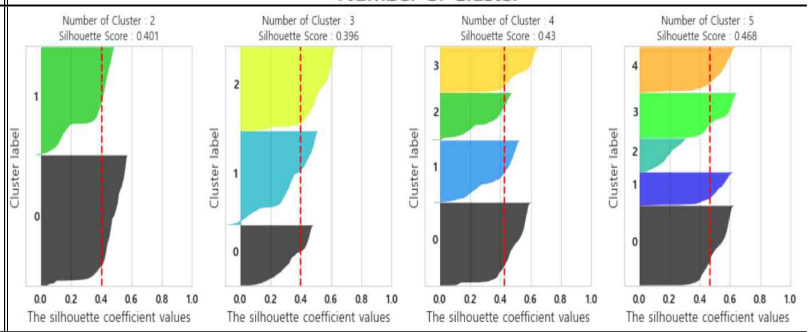
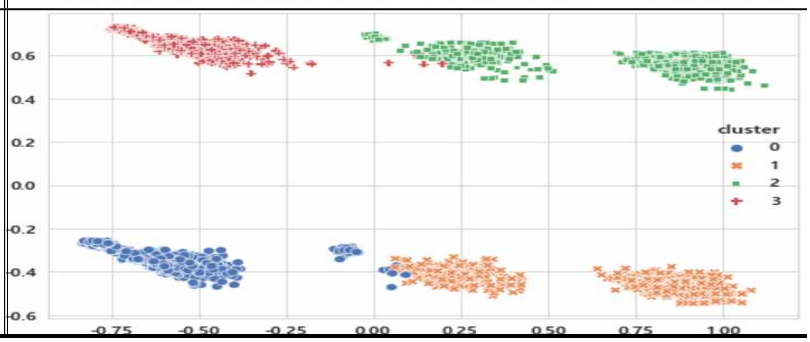
다음으로는 IQR에 기반한 이상치 처리 후 정규화를 실시한 데이터셋에 군집분석을 시도하였다. 이상치 제거를 하지 않고 정규화를 한 세 번째 데이터셋에 비해 전반적으로 실루엣 계수가 감소하였고, 군집은 만들어지지만 데이터가 섞이는 것을 알 수 있었다.

<표 7> 정규화 적용 및 이상치 처리 군집분석 (k-means)

데이터셋	D. 정규화 적용 및 이상치 처리 (IQR)
엘보우 기법	
실루엣 분석	
군집 시각화 (k=4)	

마지막으로 표준화와 정규화를 모두 적용한 데이터셋의 군집분석 결과를 살펴보았다. 정규화만을 적용했을 때와 동일한 결과가 도출되었다. 엘보우 기법과 실루엣 분석, 군집 시각화 자료에서도 차이점을 발견하지 못하였다.

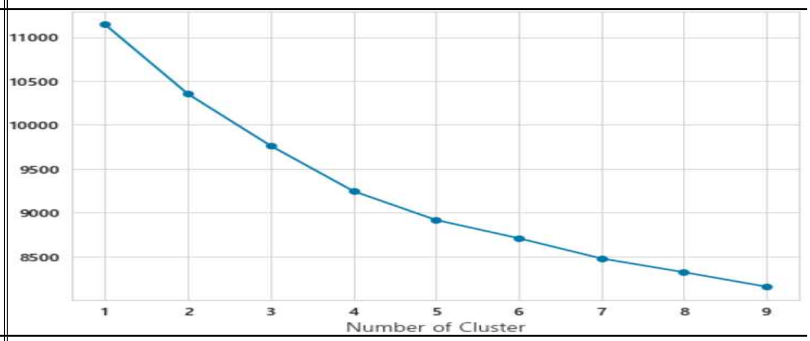
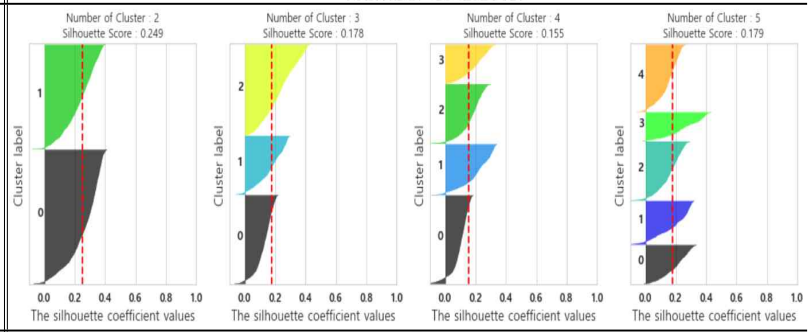
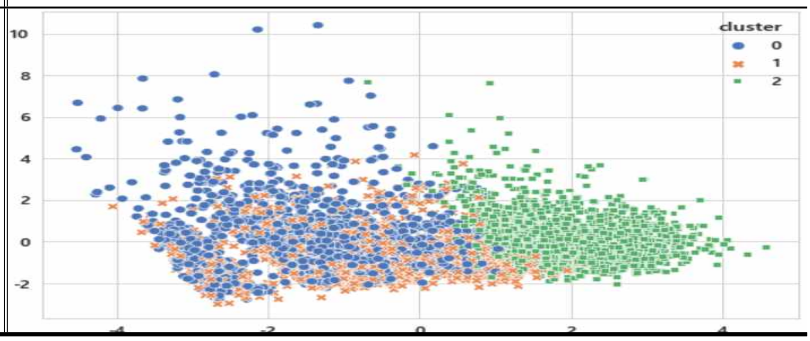
<표 8> 표준화 및 정규화 적용 군집분석 (k-means)

데이터셋	E. 표준화 및 정규화 적용
엘보우 기법	
실루엣 분석	
군집 시각화 (k=4)	

(2) k-medoids

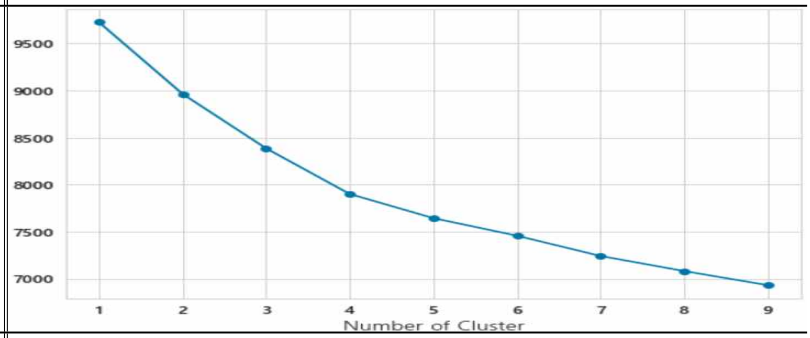
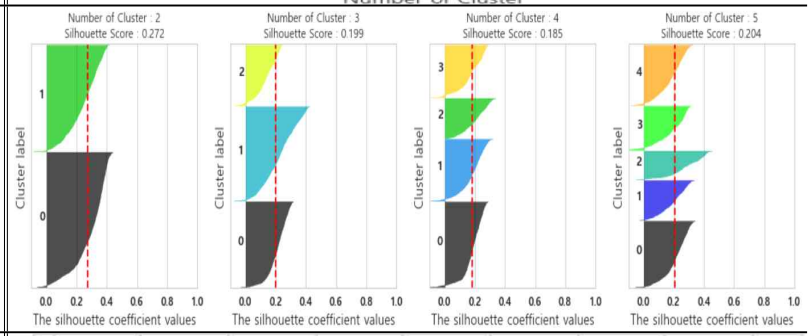
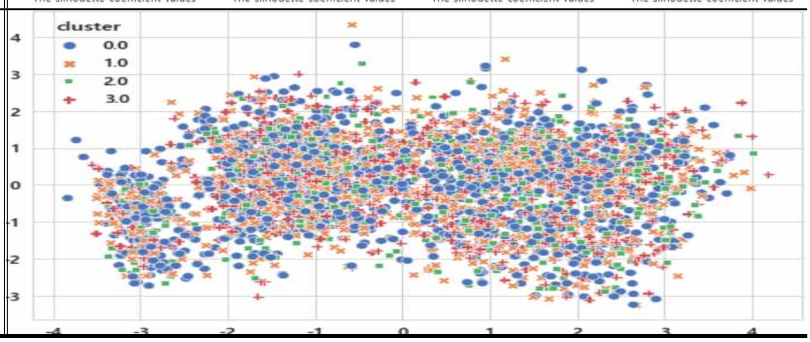
k-medoids 알고리즘을 사용한 분석도 k-means 클러스터링과 동일한 절차를 거쳐 진행하였다. 표준화만 적용된 데이터셋에 적정 군집수를 3으로 설정하여 군집분석을 실시하였다. 그 결과, k-means 알고리즘을 실시했을 때와 다르게 데이터 간의 경계가 불분명하게 나타났다.

<표 9> 표준화 적용 군집분석 (k-medoids)

데이터셋	A. 표준화 적용
엘보우 기법	
실루엣 분석	
군집 시각화 (k=3)	

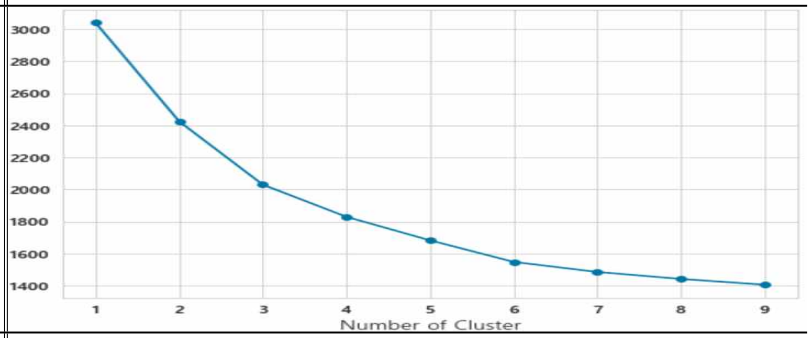
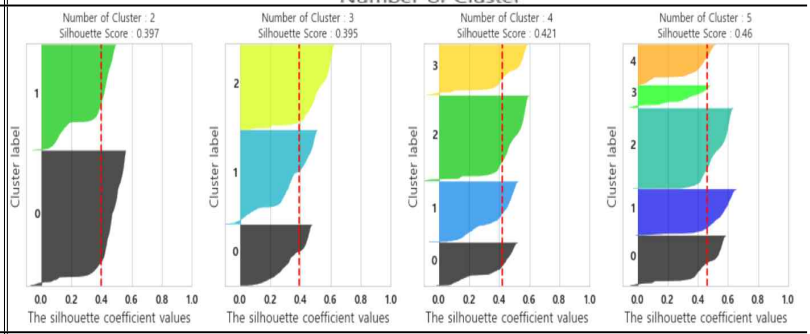
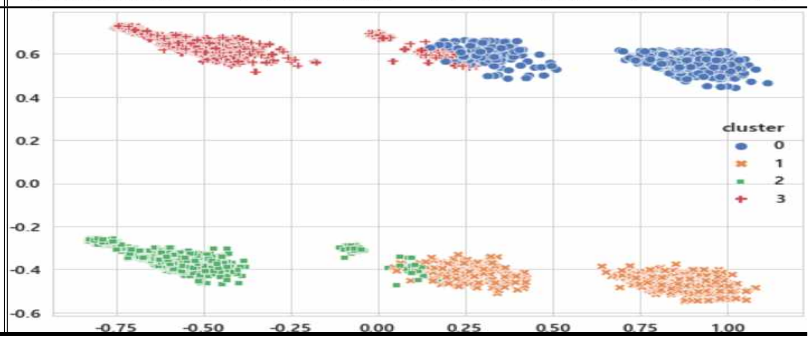
두 번째로는 표준화 적용 후, Z-Score가 절댓값 3을 넘는 데이터를 이상치 처리한 데이터셋으로 군집분석을 실시하였다. 빨간 선으로 표시되는 전체 실루엣 계수의 평균은 위의 A데이터셋보다 소폭 상승하였으나 시각화 결과, 군집을 전혀 이루지 못하고 있다는 것을 확인하였다.

<표 10> 표준화 적용 및 이상치 처리 군집분석 (k-medoids)

데이터셋	B. 표준화 적용 및 이상치 처리 (Z-Score)
엘보우 기법	
실루엣 분석	
군집 시각화 (k=4)	

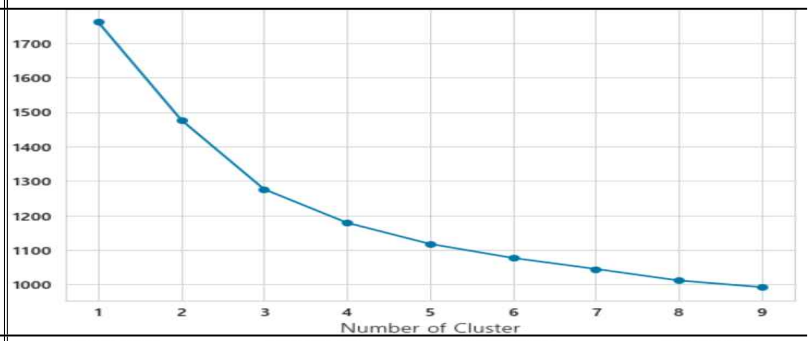
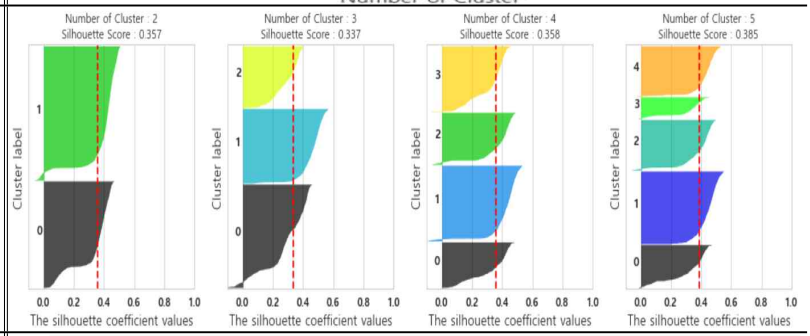
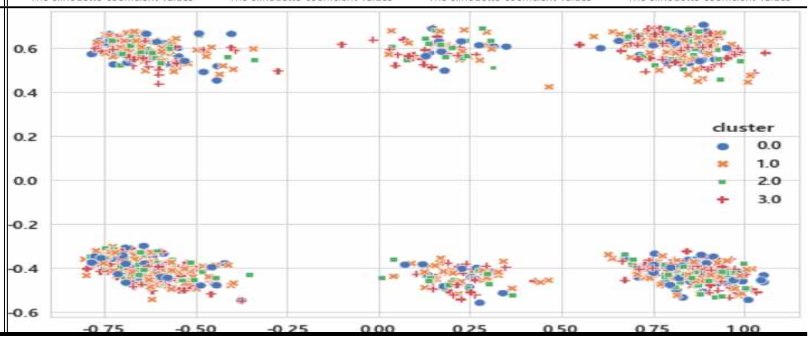
정규화만 적용된 C데이터셋을 실루엣 분석 결과에 따라 k값을 4로 정하여 군집분석을 진행하였더니 다음 <표 11>과 같은 결과를 볼 수 있었다. k-means와 k-medoids 모두 데이터가 군집을 이루고 있었으나 후자는 그래프 중앙 부분에서 데이터가 더욱 뒤섞인 모습을 보였다.

<표 11> 정규화 적용 군집분석 (k-medoids)

데이터셋	C. 정규화 적용
엘보우 기법	
실루엣 분석	
군집 시각화 (k=4)	

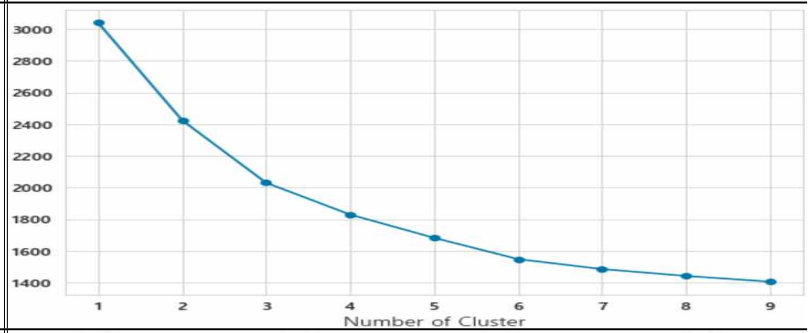
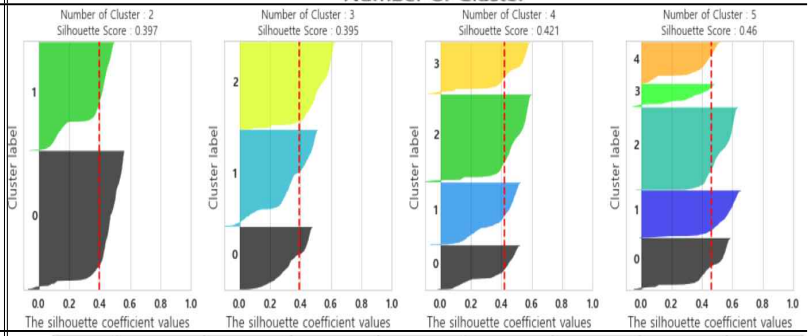
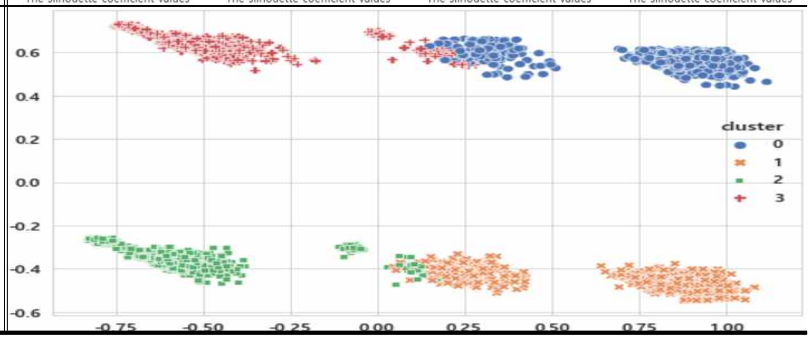
이상치 제거 후 정규화를 적용한 D데이터셋은 실루엣 계수와 시각화 내용에 있어 k-means와 k-medoids, 두 알고리즘이 비슷한 결과를 나타내었다. 실루엣 계수의 차이가 거의 없고 군집은 이루지만 각 군집에서 데이터가 섞인 유사한 양상을 찾을 수 있었다.

<표 12> 정규화 적용 및 이상치 처리 군집분석 (k-medoids)

데이터셋	D. 정규화 적용 및 이상치 처리 (IQR)
엘보우 기법	
실루엣 분석	
군집 시각화 (k=4)	

마지막으로 표준화와 정규화를 모두 적용한 데이터셋의 군집분석 결과를 조사하였다. k-means 군집화와 마찬가지로 정규화만 적용한 C데이터셋과 동일한 결과를 얻을 수 있었다. 그러나 k-means 군집화 알고리즘의 평균 실루엣 계수가 k-medoids 알고리즘의 그것을 웃도는 수치를 기록하였다.

<표 13> 표준화 및 정규화 적용 군집분석 (k-medoids)

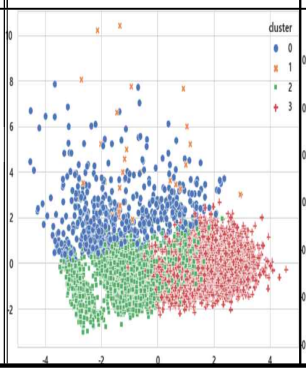
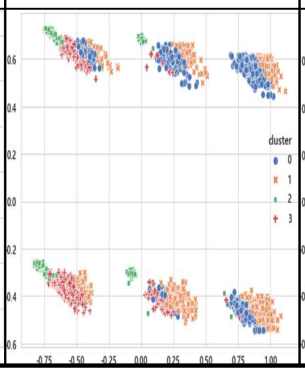
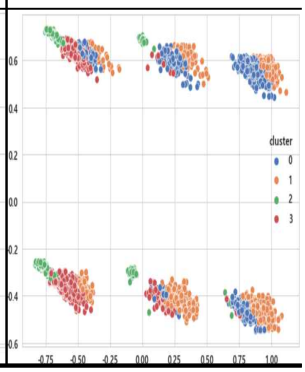
데이터셋	E. 표준화 및 정규화 적용
엘보우 기법	
실루엣 분석	
군집 시각화 (k=4)	

(3) k-prototypes

(2)번 목차까지의 군집분석으로 정규화를 적용하는 경우의 성능이 뛰어난 편이라는 것을 확인하였고 비교를 하면서 계속 같은 절차를 반복하고 있기에 k-prototypes 알고리즘은 이를 간소화하여 결과를 정리하려고 한다.

표준화와 정규화, 그리고 두 가지 피처 스케일링 기법이 모두 적용된 데이터셋을 선정해 k-prototypes 군집화를 실시한 결과는 아래와 같았다. 적정 군집 수를 4로 하였을 때의 실루엣 계수는 모두 0.2를 넘지 않았고, 시각화 결과로도 제대로 군집이 형성되지 않았다는 것을 엿볼 수 있었다. 또한 C와 E데이터셋으로 군집분석을 시행한 결과가 동일한 것도 공통점이었다.

<표 14> 데이터셋별 군집분석 (k-prototypess)

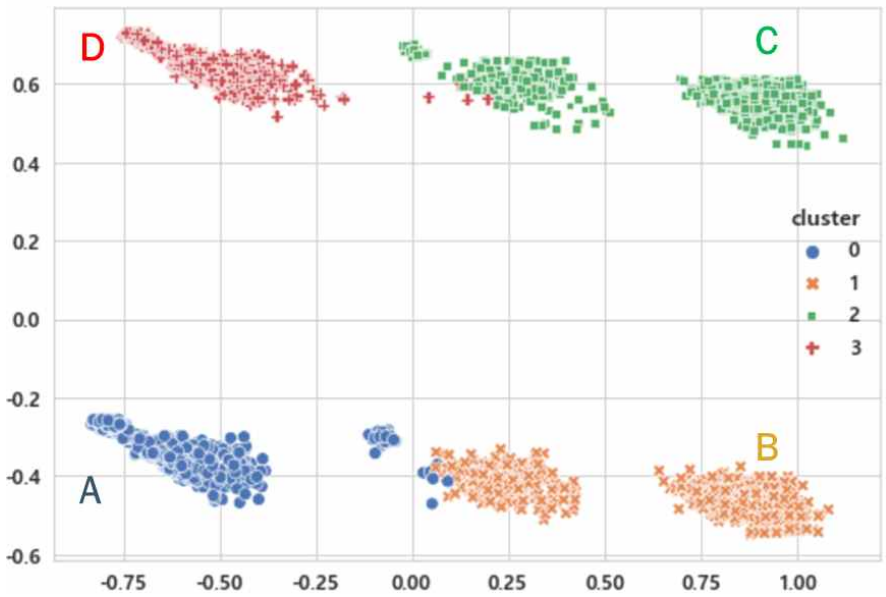
데이터셋	A. 표준화 적용	C. 정규화 적용	E. 표준화 및 정규화 적용
실루엣 계수*	0.193	0.150	0.150
군집 시각화 (k=4)			

* 소수점 넷째 자리에서 반올림

(4)군집별 특성 분석 결과

지금까지의 결과를 토대로 가장 군집화가 잘 진행된 케이스는 이상치를 제거하지 않은 채 정규화를 실시한 C 또는 E데이터셋에 k-means 군집화 알고리즘을 적용한 경우인 것을 확인하였다. 해당 방법을 사용한 군집화 결과를 게임의 이용 양태와 유료 재화의 구매, 인구통계학적 특성이라는 2개의 측면에서 살펴보았다.

1) 게임 이용 양태와 온라인 게임 내 구매 총액



<그림 3> 최적 군집분석 결과

<그림 3>의 좌측 하단에 있는 0번 군집 (파랑, A)에 1,392명, 그 우측의 1번 군집 (주황, B)에 1,041명, 우측 상단의 2번 군집 (초록, C)에 794명, 3번 군집 (빨강, D)에 762명 순으로 속했다. 군집별로 게임 이용 양태를 나타내는 변수와 유료 재화의 구입에 투자한 비용을 뜻하는 온라인 게

임 내 구매 총액을 <표 15>로 정리하였다.

<표 15> 군집별 분산분석표

변수	군집	A	B	C	D	전체	F값
온라인 게임 시작연령	평균	14.403017	27.847262	28.875315	16.251969	21.145399	595.51102
	표준편차	7.372833	11.970104	12.010853	8.340930	11.931092	***
주중 하루 평균 이용시간	평균	3.458441	2.764025	2.951159	2.969003	3.082753	19.918938
	표준편차	2.482569	1.969661	2.455866	2.387602	2.351642	***
주말 하루 평균 이용시간	평균	5.427155	4.301460	4.122557	4.677244	4.730456	40.305328
	표준편차	3.266187	2.743897	3.165531	3.182398	3.146601	***
주중 1회 평균 이용시간	평균	2.557593	1.912315	1.962506	2.036430	2.171191	35.572817
	표준편차	1.980876	1.449118	1.563928	1.845626	1.770414	***
주말 1회 평균 이용시간	평균	4.064167	3.142065	2.896134	3.391404	3.462519	45.69682
	표준편차	2.685921	2.311138	2.374895	2.615975	2.560191	***
온라인 게임 총 이용 개수	평균	8.647989	7.810759	6.527708	7.262467	7.742793	14.429412
	표준편차	8.312616	8.431153	6.144323	6.232414	7.627816	***
온라인 게임 주 이용 개수	평균	4.639368	4.647454	4.037783	4.116798	4.421910	9.656323
	표준편차	3.139343	4.061027	2.851266	2.539208	3.265305	***
온라인 게임 내 구매 총액	평균	36985.889	41525.841	28110.977	31972.697	35446.495	4.726757
	표준편차	80678.448	99079.241	67148.369	67729.753	81398.882	**

* $p < .05$. ** $p < .01$. *** $p < .001$.

온라인 게임 시작연령, 주중과 주말의 하루 평균 이용시간과 1회 평균 이용시간 등과 온라인 게임 내 구매 총액을 대상으로 분산분석을 실시한 결과, <표 15>와 같이 제시된 8개의 변수에 대해 집단별로 유의미한 차이가 있는 것으로 나타났다.

A 집단은 가장 이른 나이에 온라인 게임을 접하였고 주중과 주말을 불문하고 하루 평균 이용시간과 1회 평균 이용시간이 가장 길다는 특징을 지니고 있었다. 특히 평균 게임 이용시간이 전체를 웃도는 것은 이 군집이 유일하였다. 또 설문 실시일을 기준으로 지난 1년 동안 제일 많은 게임을 플레이한 것으로 드러났고 온라인 게임 내 구매 총액의 평균은 4개

의 집단 중 2위였다.

B 집단은 20대 후반에 온라인 게임을 시작하였고 주중의 평균 이용시간이 가장 짧은 그룹이었다. 3개월 이상 플레이를 지속해온 게임의 수가 약 4.6개로 가장 많았으며 강한 몰입도는 월 평균 약 4만원의 지출로 이어지지 않았을까 짐작하게 하였다.

C 집단은 가장 늦은 나이에 온라인 게임을 접하였으며 주말의 평균 이용시간이 제일 짧다는 특성을 보였다. 이용하는 온라인 게임의 개수도 가장 적었고 게임에 투자하는 비용 역시 월 평균 2만 8천원으로 전체 4개의 집단 중 가장 작은 것으로 나타났다.

청소년기에 게임을 시작한 D 집단은 C 집단과 동일하게 온라인 게임 내 구매 총액과 온라인 게임 이용 개수의 평균이 전체의 수치를 밑도는 모습을 보였다. 그러나 C 집단과 달리 게임을 플레이하는 시간은 긴 편에 속하는 것을 확인할 수 있었다.

2) 인구통계학적 특성

다음으로 각 군집과 인구통계학적 특성을 나타내는 변수 간에 교차분석을 시행하여 <표 16>으로 요약하였다. 괄호 안에는 각 집단의 전체 인원 대비 해당 범주의 수치를 반올림하여 소수 첫째 자리까지 표기하였다. 분석 결과, 6개의 인구통계학적 변수 중 모든 변수에서 통계적으로 유의미한 차이가 존재하였다.

A, B 집단은 남성으로만 이루어졌고 C, D 집단은 여성으로만 구성되었다. 그중 A와 D 집단은 비교적으로 젊은 연령층이 많았으며 B와 C 집단은 중·장년층이 다수를 차지하였다. 배우자와 자녀 유무 측면에서 살펴보면 A 집단과 D 집단은 대부분 배우자와 자녀가 없었지만 B와 C 집단은 배우자와 자녀를 가진 구성원의 비율이 높았다.

<표 16> 군집별 인구통계학적 특성

변수	A	B	C	D	전체
성별	x^2 = 3,989.0, df = 3, p < .001				
남성	1,392 (100%)	1,041 (100%)	0	0	2,433
여성	0	0	794 (100%)	762 (100%)	1,556
연령대	x^2 = 1844.920, df = 15, p < .001				
10대	320 (23.0%)	0	36 (4.5%)	147 (19.3%)	503
20대	594 (42.7%)	28 (2.7%)	38 (4.8%)	352 (46.2%)	1,012
30대	319 (22.9%)	270 (25.9%)	199 (25.1%)	171 (22.4%)	959
40대	125 (9.0%)	414 (39.8%)	283 (35.6%)	66 (8.7%)	888
50대	32 (2.3%)	269 (25.8%)	205 (25.8%)	23 (3.0%)	529
60대	2 (0.1%)	60 (5.8%)	33 (4.2%)	3 (0.4%)	98
배우자 유무	x^2 = 3672.147, df = 3, p < .001				
있음	46 (3.3%)	1,017 (97.7%)	783 (98.6%)	0	1,846
없음	1,346 (96.7%)	24 (2.3%)	11 (1.4%)	762 (100%)	2,143
자녀 유무	x^2 = 2267.152, df = 3, p < .001				
있음	7 (0.5%)	764 (73.4%)	554 (69.8%)	5 (0.7%)	1,330
없음	1,385 (99.5%)	277 (26.6%)	240 (30.2%)	757 (99.3%)	2,659
학력	x^2 = 201.278, df = 6, p < .001				
고졸 이하	447 (32.1%)	100 (9.6%)	210 (26.4%)	191 (25.1%)	948
대학교 재학 및 졸업	848 (60.9%)	782 (75.1%)	522 (65.7%)	513 (67.3%)	2,665
대학원 재학 및 졸업	97 (7.0%)	159 (15.3%)	62 (7.8%)	58 (7.6%)	376
개인소득 및 용돈	x^2 = 1559.322, df = 30, p < .001				
3만원 미만	87 (6.3%)	0	3 (0.4%)	40 (5.2%)	130
3만-10만원 미만	153 (11.0%)	0	13 (1.6%)	48 (6.3%)	214
10만원 이상	44 (3.2%)	0	16 (2.0%)	20 (2.6%)	80
100만원 미만	321 (23.1%)	11 (1.1%)	188 (23.7%)	190 (24.9%)	710
100만-200만원 미만	178 (12.8%)	38 (3.7%)	157 (19.8%)	162 (21.3%)	535
200만-300만원 미만	341 (24.5%)	169 (16.2%)	214 (27.0%)	186 (24.4%)	910
300만-400만원 미만	181 (13.0%)	309 (29.7%)	102 (12.8%)	71 (9.3%)	663
400만-500만원 미만	49 (3.5%)	232 (22.3%)	48 (6.0%)	27 (3.5%)	356
500만-600만원 미만	44 (3.2%)	139 (13.4%)	22 (2.8%)	6 (0.8%)	179
600만-700만원 미만	11 (0.8%)	65 (6.2%)	11 (1.4%)	3 (0.4%)	90
700만원 이상	15 (1.1%)	78 (7.5%)	20 (2.5%)	9 (1.2%)	122

한편, 모든 집단에서 대학교 재학 및 졸업이라고 응답한 사람들이 압도적으로 많았고 대학원까지 진학한 고학력자는 B 집단에서 가장 많이 발견되었으며 그 비율도 가장 높았다. 개인소득 및 용돈 측면에서도 B 집

단은 월 300만원 이상 400만원 미만의 범위에 속하는 구성원이 가장 많았고 700만원 이상의 월 소득을 수령하는 고소득자도 78명, 7.5%의 비율로 제일 높은 수치를 기록하였다. 이와 달리 나머지 집단은 100만원 이상 300만원 미만의 범위에서 주로 분포하였다.

제 Ⅴ장 결론

제1절 요약 및 시사점

본 연구는 온라인 게임 유료 이용자를 대상으로 군집분석을 시행하여 이들을 집단별로 나누고 각 집단을 비교 분석함으로써 특성을 알아보았다. 데이터 전처리 과정을 거쳐 얻은 3,989개의 데이터를 군집분석한 결과, 이상치 처리를 진행하지 않은 채 정규화를 실시한 데이터셋에 k-평균 알고리즘을 적용하면 가장 우수한 군집화 성능을 보인다는 것을 확인할 수 있었다. 표준화 여부는 영향을 주지 않았고, 이때 초기치 k값은 4, 실루엣 계수는 0.43이었다.

4개로 구분된 집단은 아래의 특성을 보였다. 먼저 A 집단은 1,392명으로 그 수가 가장 많았다. 그들은 20대가 약 43%를 차지하는 젊은 남성 집단이었고 대부분 배우자와 자녀는 없는 것으로 드러났다. 온라인 게임을 처음 접하는 나이가 약 21세였는데 A 집단은 약 14세의 나이에 온라인 게임을 처음 시작하였다. 또한 주중과 주말에 구애받지 않고 가장 긴 하루 평균 이용시간, 1회 평균 이용시간을 기록하였다. 설문 실시일 기준으로 지난 1년 동안 플레이해본 온라인 게임 수가 약 9개로 가장 많았고 온라인 게임의 유료 재화에 쓰는 비용은 3만 7천원가량으로 전체 평균보다 1,500원 정도 많이 투자하였다. 4개 집단 중 고졸 이하의 학력을 가진 인원의 비율이 제일 높았고 월 200만원 이상 300만원 미만의 개인소득을 보이는 유형이 가장 많았다.

B 집단은 1,041명의 남성으로 이루어졌으며 그중 30대부터 50대까지의 유료 이용자가 대부분이었는데 40대가 가장 많은 비율을 차지하였다. B 집단에서 약 97%의 구성원이 배우자가 있었고 73% 정도가 자녀를 가지고 있었다. 청소년기에 온라인 게임을 처음 접한 A 집단과 달리 이들은 약 27세의 나이에 시작하였으며 주중의 하루 및 1회 플레이타임이 각각

2.8시간, 1.9시간으로 가장 짧았다. 3개월 동안 지속해서 이용한 게임의 수는 약 4.6개로 가장 많았고 온라인 게임 내 구매총액은 약 4만 2천원으로 전체 평균보다 6천원 높은 것으로 나타났다. 대학원 재학 이상의 고학력자 비율이 약 15%, 월 소득 700만원 이상의 고소득자 비율이 7.5% 가량으로 모든 집단 중 제일 높았으며 월 300만원 이상 400만원 이하의 개인소득을 벌어들이는 구성원이 많았던 것도 타 집단과의 차이점이었다.

C 집단은 794명의 여성으로 구성되었고 연령대는 B 집단과 유사하였다. 다만 B 집단보다 배우자가 있다고 답한 이들의 비율은 다소 높았고, 자녀가 있는 것으로 응답한 비율은 떨어졌다. 이들은 약 29세라는 가장 늦은 나이에 온라인 게임을 접하였으며 주말의 하루 및 1회 평균 이용시간이 제일 짧다는 특성이 있었다. 조사일 기준으로 지난 1년간 이용했던 게임 수와 유료 재화 구매 금액 역시 가장 적은 것으로 드러났다. 대학교 재학 이상의 학력을 가진 사람이 약 74%로 네 개 집단 중 세 번째였고 월 소득 700만원 이상의 고소득자 비율은 두 번째로 높았다.

D 집단은 가장 적은 762명이었고 모두 여성이었다. 연령대와 가족 구성, 개인소득 및 용돈은 A 집단과 비슷하였다. 그렇지만 학력은 D 집단이 A 집단보다는 높은 것으로 나타났다. D 집단은 약 16세에 온라인 게임을 시작하였고, 주중과 주말의 평균 이용시간이 두 번째로 높은 집단이었지만 전체 평균보다는 낮았다. 온라인 게임의 이용 개수, 온라인 게임 내 구매 총액 역시 최하위는 아니었으나 전체 평균을 밑돌았다.

기존 연구에 따르면 전국의 중·고등학생을 대상으로 일일 게임 이용시간을 살펴본 결과, 남학생이 여학생보다 긴 것으로 밝혀졌는데 A 집단의 하루 및 1회 평균 이용시간이 D 집단보다 긴 것으로 나타난 본 연구의 결과와 동일하였다(김경미, 염유식, 2016). 그러나 게임 지출비용이 높은 그룹에 비해 낮은 그룹에 상대적으로 고학력자가 많다는 연구와 다르게 온라인 게임 내 구매 총액이 가장 컸던 B 집단에서 대학교 재학 이상의 학력을 가진 구성원의 비율이 가장 높은 것으로 드러나 차이를 보였다(서경희, 2014).

소득과 게임 이용시간의 관계에 관한 연구는 이견이 존재하였다. 먼저, 경제적인 여유가 있어야 게임을 즐길 수 있다는 시각이 있었다(최성락, 박준휘, 2018). 하지만 경제력이 풍부한 유저는 게임 내 단순한 콘텐츠를 반복하지 않고 현실의 자본으로 해결하는 것이 가능하므로 불필요한 시간 낭비를 줄일 수 있다는 주장이 제기되기도 하였다(김효은, 2016). 본 연구에서는 개인소득이 가장 많았던 B 집단은 게임 이용시간이 짧은 편에 속하였고, 가장 경제적 여유가 없었던 A 집단은 게임 이용시간이 제일 길어 후자에 부합하는 결과가 도출되었다.

이처럼 온라인 게임 유료 이용자를 집단별로 구분하고 특성을 파악하였을 뿐만 아니라 그 과정에서 머신러닝을 활용하여 최적의 군집분석 알고리즘을 선정한 것에서도 학문적 의의를 가진다. 본 연구 역시 대부분의 게임 이용자 분류 연구와 같이 k-means 알고리즘을 사용하였으나 여러 군집분석 알고리즘의 성능을 비교하여 가장 우수한 결과를 도출하는 절차를 거쳤다는 점이 종래의 연구와 상이하다고 할 수 있다. 또한 본 연구는 유저 친화적인 게임 운영과 마케팅 전략의 수립, 나아가 비즈니스 모델 개선의 자료로 활용될 수 있다는 점에서 실무적 의의를 찾을 수 있다.

예컨대 주말에 게임을 많이 플레이하는 집단을 목표로 한 이벤트를 개최할 수도 있고 그 반대로 주말 이용시간이 적은 집단의 접속률을 끌어올리기 위한 콘텐츠를 기획할 수도 있을 것이다. 또 유료 재화 구매 금액이 적은 여성 집단을 대상으로 취향에 맞는 아이템을 출시할 수도 있다. 이와 같이 본 연구는 온라인 게임 유료 이용자에 대한 이해를 높임으로써 유저의 니즈에 부합하는 상품 및 콘텐츠를 제공하는 데 도움을 줄 것으로 기대한다.

제2절 한계 및 향후 연구방향

본 연구의 한계는 먼저 한국콘텐츠진흥원에서 제공한 로데이터의 모든

변수를 반영하여 군집분석을 시행하지 못했다는 것이다. 원래의 데이터에는 선호 장르, 기기, 직업 등과 같이 게임 이용자의 특성을 나타낼 수 있는 변수가 있었으나 연구 주제와의 연관성, 데이터 전처리의 어려움 등을 사유로 선택하지 못하였다. 그중에서도 특히 장르와 확률형 아이템 구매 여부는 유저의 특성과 유료 재화의 구매에도 많은 영향을 줄 것으로 추정되는데 향후 연구는 해당 변수를 고려하여 이루어져야 한다.

또 실제 게임의 로그를 토대로 진행한 것이 아니므로 게임사에서 본 연구의 결과를 바로 적용하는 것에 신중하여야 한다. 전국의 모든 국민을 대상으로 도출한 데이터이기에 보편적인 특성으로써 참고할 수 있겠으나 게임에 적용하기 위해서는 기업이 가지고 있는 데이터를 기반으로 이용자의 특성을 고찰하는 것이 요구된다.

연구에서 활용된 변수 간의 관계를 심도 있게 규명하지 못하고 집단별 특성을 나열하는 데 그치고 있다는 점도 본 연구의 한계이다. 예를 들어 온라인 게임 내 구매 총액에 영향을 주는 변수와 그 비중을 규명하거나 성별은 다르지만 유사한 연령대의 구성원으로 이루어진 B와 C 집단은 주중과 주말의 게임 평균 이용시간에 있어 다른 양상을 보이는데 그 원인을 파악하지 못하였다. 후속 연구에서는 회귀분석 등의 다른 통계기법을 적용하거나 본 연구에서 규명하지 못한 부분에 대한 심층적인 조사가 진행될 필요가 있다.

참고문헌

<논문>

김경미, 염유식(2016), “청소년의 게임이용과 학교생활”, *한국콘텐츠학회 논문지*, 16(1), 753-765.

김동혁(2016), k-medians 클러스터링과 Elbow method를 이용하여 정확도가 향상된 Snort rule classifier, 한양대학교 대학원 석사학위논문.

김병규(2014), 모바일 부분유료화 게임의 과금 플레이 밸런싱 전략 연구, 한양대학교 대학원 석사학위논문.

김보아(2021), 게임 구독 서비스의 비즈니스 모델 연구, 한양대학교 대학원 석사학위논문.

김소울(2016), “게임 내 캐릭터 뽑기 사용자의 과금 심리 분석”, *한국게임학회 논문지*, 16(3), 77-86.

김영주(2022), 모바일 게임 이용자 유형이 유료 아이템 구매에 미치는 영향, 동국대학교 대학원 석사학위논문.

김예은(2021), 모바일 게임 인터랙션 정도에 따른 유저분류와 게임 정의의 차이, 경희대학교 대학원 석사학위논문.

김은비(2019), 온라인 게임 내의 유료 아이템 특성과 유저의 과금 성향에 따른 관계 연구, 중앙대학교 대학원 석사학위논문.

김정진(2013), 게임 이용자의 자아 존중감이 유료 아이템 구매동기에 미치는 영향, 홍익대학교 영상대학원 석사학위논문.

김종무(2013), “모바일 게임 이용 형태에 관한 연구”, *한국디자인문화학회지*, 19(3), 109-117.

김태경, 장정주, 장현영(2009), “온라인 게임 비즈니스의 재편과 도전”, *Korea Business Review*, 12(3), 111-133.

김효은(2016), 온라인 게임 이용자의 자본에 따른 게임 이용 특성 연구, 한양대학교 대학원 박사학위논문.

노경희, 이태일, 조성현(2007), “게임 이용자의 특성 분류를 통한 게임 캐릭터

선호도에 관한 조형 연구", *한국게임학회 논문지*, 7(4), 23-31.

맹주홍, 정민수, 서재원, 박윤조, 김효남(2018), "온라인 게임에서의 새로운 과금 모델에 대한 연구", *한국컴퓨터정보학회 하계학술대회 논문집*, 26(2), 102-103.

서경희(2014), 게임소비성향에 따른 게임활용 유형화 분석, 홍익대학교 영상대학원 석사학위논문.

신종현, 신현호, 임세희, 이동헌, 부정철(2017), "온라인 게임에서의 캐시아이템 사용에 관한 연구", *한국컴퓨터정보학회 하계학술대회 논문집*, 25(2), 69-72.

오수민(2014), 데이터 마이닝에서의 범주형데이터 군집분석을 위한 초기치 선정방법, 제주대학교 대학원 박사학위논문.

위정현, 김은비(2019), "온라인 게임 아이템 특성과 이용자 특성의 관계 분석", *한국게임학회 논문지*, 19(3), 113-121.

유창석, 정재기, 이좌지자(2014), "경제학 관점에서 부분유료화 게임 비즈니스 모델 분석 및 사례 연구", *한국콘텐츠학회 논문지*, 14(11), 883-892.

이남재, 설남오, 이광재(2005), "MMORPG 개발 후발업체의 과금방식에 관한 연구", *한국게임학회 논문지*, 5(6), 19-27.

이동은, 이유호(2015), "디지털 게임의 부분유료화 모델에 관한 연구", *한국게임학회 논문지*, 15(6), 107-120.

이준영, 이제성(2017), "게임 플레이어 유형에 따른 몰입 및 충성도 연구", *한국게임학회 논문지*, 17(5), 71-79.

이일수(2018), 모바일 소셜 네트워크 게임에서 확률형 아이템의 구매 동기 요인 분석, 인천대학교 경영대학원 석사학위논문.

이혜정, 민애홍(2019), "디지털 게임 플레이어의 주관성 연구", *한국콘텐츠학회 논문지*, 19(3), 275-287.

전홍준(2017), 게임의 부분 유료화 개선 방안에 대한 연구, 광운대학교 대학원 석사학위논문.

정광재(2015), "온라인 게임에서의 무료 이용자와 유료 이용자의 행태 비교", *경영관리연구*, 8(1), 57-71.

정다영, 윤선희, 이용우, 김승인(2015), “모바일 게임 사용자 유형별 분석에 따른 유료결제 서비스 활성화 방안”, *디지털디자인학연구*, 15(1), 569-577.

조은하(2017), “게임 산업 수익화(monetization) 방식과 한중일 스마트폰 게임의 국가별 수용 양상”, *한국게임학회 논문지*, 17(2), 75-83.

주현택(2019), 구조 방정식 모형을 활용한 모바일 캐주얼 게임에서 아이템 구매 동기에 관한 연구, 가천대학교 게임대학원 석사학위논문.

최성락, 박준휘(2018), “개인의 사회적 요소들이 게임 수행에 미치는 영향 분석”, *한국콘텐츠학회 논문지*, 18(4), 613-621.

한창완(2008), “플랫폼 통합에 따른 디지털콘텐츠 제작기술 경향연구”, *한국애니메이션 연구*, 14, 151-164.

Bartle, R. (1996). Hearts, clubs, diamonds, spades: Players who suit MUDs. *Journal of MUD Research*, 1(1), 19.

Cheng-Hsun Ho and Ting-Yun Wu. (2012). Factors Affecting Intent to Purchase Virtual Goods in Online Games. *International Journal of Electronic Business Management*, 10(3), 204-212.

Guo, Y. and Barnes. S. J. (2007). Consumer Purchasing Behavior in Virtual Worlds. *Advances in Information Systems*.

Gyhwan Oh., TaiYoung Ryu. (2007). Game Design on Item-selling Based Payment Model in korean Online Games. *DiGRA*.

Lim, R. and Seng, E. (2011). Virtual Goods in Social Games: An Exploratory Study of Factors that Drive Purchase of In-Game Items. *Proceeding of the Ninth International Conference on e-Business 2010,* Faculty of Science Kasetsart University.

Lee, C. H., Kim, M. S., Lee, J. S. (2012). System and method for analyzing cluster results of large amounts of data, http://www.google.com/patents/WO2013151221A1?cl=ko.

Park, H. S. and Jun, C. H. (2009). A simple and fast algorithm for k-medoids clustering. *Expert systems with applications*, 36(2), 3336-3341.

<단행본 및 보고서>

김윤지(2019), "'문화'를 넘어 '산업'으로 자리매김하다", 연중기획 12월호, https://eiec.kdi.re.kr/publish/naraView.do?fcode=000020000400001000 08&cidx=12324&sel_year=2020&sel_month=03&pp=20&pg=1

이영호, 이수현(2021), 파이썬으로 경험하는 빅데이터 분석과 머신러닝, 생능출판.

안드레아스 뮐러, 세라 가이도(2017), Introduction to Machine Learning, 한빛미디어 (박해선 역).

피터 브루스, 앤드루 브루스, 피터 게데크(2018), Practical Statistics for Data Scientists, 한빛미디어 (이준용 역).

한국게임산업개발원(2005), "2005 대한민국 게임백서".

한국콘텐츠진흥원(2017), "2017 게임이용자 실태조사 보고서".

한국콘텐츠진흥원(2018), "2018 게임이용자 실태조사 보고서".

한국콘텐츠진흥원(2019), "2019 게임이용자 실태조사 보고서".

한국콘텐츠진흥원(2020), "2020 게임이용자 실태조사 보고서".

한국콘텐츠진흥원(2021), "2021 게임이용자 실태조사 보고서".

한국콘텐츠진흥원(2021), "2021 대한민국 게임백서".

한국콘텐츠진흥원(2022), "2022 대한민국 게임백서".

한국콘텐츠진흥원(2022), "2022년 상반기 콘텐츠 산업 동향분석 보고서".

Adams, E. (2002). From Casual to Core: A Statistical Mechanism for Studying Gamer Dedication. Gamasutra.

Adams, E. (2014). Fundamental of Game Design. New Riders.

Bateman, C. R. and Boon. R. (2006), 21st Game Design. Charles River Media.

Rea, L. M. and Parker, R. A. (2005). Designing & conducting survey research: A comprehensive guide (3rd Edition). Jossey-Bass.

Wallis, A. (2007). Q&A: Nexon America talks Maple Story. Gamasutra.

<기사>

문영수(2022), "확률형 과금 없어도 되네…'대항해시대 오리진' 톱10 목전", 아이뉴

스24, https://www.inews24.com/view/1515158
원태영(2022), “‘PC·모바일 경계 사라진다’ 대세로 떠오른 크로스 플랫폼”, 이코노미스트, https://economist.co.kr/2022/02/03/it/general/20220203200101527.html
윤진우(2021), “모바일게임을 PC로 즐긴다…게임시장 대세로 떠오른 크로스 플랫폼”, 조선비즈, https://biz.chosun.com/it-science/ict/2021/05/22/JODSORG4AZCPBFNA7PUWX5P3O4/?utm_source=naver&utm_medium=original&utm_campaign=biz
이정훈(2022), “확률형 아이템을 제거했는데 매출 상위권? 착한 과금 게임이 보여준 기적”, 게임와이, http://www.gamey.kr/news/articleView.html?idxno=3002834
정호(2022), “고객 서비스 불만 터지는 모바일게임 속, ‘부분유료화’”, 뉴스워치, http://www.newswatch.kr/news/articleView.html?idxno=60165
조학동(2022), “확률형 아이템 없앤 '카트라이더', 넥슨의 새로운 시도에 주목해야”, 동아일보, https://www.donga.com/news/It/article/all/20230117/117484767/1
한경우(2022), “과금에 '역풍' 맞았던 엔씨소프트, 신작으로 불명예 씻을까”, 한경코리아마켓, https://www.hankyung.com/finance/article/202211189506i
한연규(2006), “게임빌, '2006 프로야구'도 밀리언셀러 등극”, 게임동아, https://game.donga.com/30019/

Abstract

A Study on the Group Types and Characteristics of Online Games Paid Users

Park, Hyunchul
Seoul School of Integrated Sciences and Technologies
Advisor: Kim, Bo Young

The game industry accounts for a significant portion of the content industry in South Korea and has grown steadily over the past 10 years. However, a few problems have been reported. In particular, the business model that induces the purchase of paid goods has received much criticism but is still applied to several games. The one-sided choice that did not fully reflect users' needs drew backlash and eventually led to physical conflict.

Therefore, the purpose of this study was to improve the business model by increasing the understanding of game users, especially paid users who are directly related to profits, which can help to release products and content that meet the needs of users. For this purpose, a cluster analysis was conducted for paid users, and the characteristics of each group were compared and analyzed.

Data preprocessing, and feature scaling such as normalization and standardization were performed on the basic dataset that consisted of about 4,000 data. Performance was compared by applying three algorithms, k-means,

k-medoids, and k-prototypes, for that dataset. Consequently, it was found that applying the k-means algorithm with an initial value of 4 to the normalized dataset showed the best clustering performance.

For the statistical analysis, an ANOVA(Analysis of Variance) was performed on variables related to the use of games and the total online game purchases in order to understand each cluster's characteristics. Also, a cross-tabulation analysis was performed on the participants' demographic variables. Results showed that Group A, which was a group of young men who encountered games when they were the youngest, had the longest average playtime of games on both weekdays and weekends, and the largest number of online games played in the past year. The group's payment amount also exceeded the average. However, Group B, composed of middle-aged men, started online games the latest, and the average playtime during the week was the shortest.

Both Groups C and D were comprised of only women, but they showed differences in age. Group C, consisting of middle-aged women, encountered online games at the latest age and had the shortest average weekend playtime. In addition, the number of online games played, and the purchase amount of paid goods was also the lowest among all groups. Group D was composed of relatively young women. Compared to Group C, Group D showed a longer average playtime and earlier game experience during adolescence. However, Group C and D showed similar characteristics, such that both groups' average number of online games played and the total amount of online game purchases were below the overall average.

In sum, this study found that each group had different characteristics, by conducting a cluster analysis, a type of unsupervised learning, on about 4,000 paid users of online games, and by conducting a group comparison analysis. The findings of this study can be used to operate user-friendly games, estab-

lish marketing strategies, and improve business models in the future.

key words: Online Games, Cluster Analysis, Characteristics of Paid Users
Student Number: 2125418002

감사의 글

우선 데이터를 제공해주신 한국콘텐츠진흥원에 감사드린다. 그리고 미흡한 점이 많음에도 친절하고 상세하게 지도해주신 김보영 교수님께 깊은 감사를 전하고 싶다. 조언을 아끼지 않으신 aSSIST의 교수님들께도 감사드린다. 또 바쁜데도 짬을 내서 영문초록을 검수하여 논문의 완성도를 높여준 유비에게도 고마울 뿐이다. 끝으로 돌아가신 며칠 전에 큰아버지의 명복을 빈다.