

Portfolio

데이터 분석가



성명

최호준

생년월일

1999.01.18

연락처

010-5067-8947

이메일

goya8211@naver.com

01. 이커머스 / EDA (연관 항목 분석) 3~6p

EDA 및 연관항목 분석을 통한 프로모션 성과 분석

[Git Hub](#)

02. 웹툰 / NLP (자연어 처리) 7~9p

자연어 처리를 통한 인기 웹툰 키워드 분석

및 정규 웹툰 편성 후보군 발굴

[Git Hub](#)

03. 은행 / ML (분류 머신러닝) 10~13p

은행 고객 데이터 분석 및 가입 예상 고객 분류

[Git Hub](#)

04. 게임 / AD (유저 이상탐지) 14~16p

게임 매치 데이터 분석을 통한 이상 유저 탐지

[Git Hub](#)

Project 01

이커머스 프로모션 성과 분석

2024-07-15 ~ 2024-08-21 (5주)

인원	3명 (팀 프로젝트)
분석 환경	Python
역할	프로모션별 재구매 / 재활성 분석 및 시각화, 연관 항목 및 순차 패턴 분석, 고객 RFM 분석, 기대 이익 산출
분석 목적	프로모션 별 신규 유입, 재구매 고객 분석을 통한 문제점 도출 및 해결방안 제시

Lesson & Learn	<p>1) 내가 분석한 내용으로 팀원들을 설득시키기 위해서 보다 객관적이고 정량적인 근거가 필요하다는 것을 배움</p> <p>2) 팀원들과의 분석 내용 소통, 공유를 통해 더 자연스럽게 합리적인 분석의 흐름을 찾아가는 과정을 배움</p> <p>3) EDA를 통해 데이터를 deep dive하는 과정에서 시야가 좁아지며 주제와 목적을 잃게 되는 것을 느끼고 항상 목적을 생각해야 한다는 것을 배움</p>
----------------	--

이커머스 프로모션 성과 분석

1

전처리 및 EDA

- 프로모션 정보 결합
- 판매량, GMV 현황 분석

* GMV: 매출

2

프로모션별 세부 분석

- 프로모션 기간 신규 유입 및 재구매 현황
- 프로모션 기간 비활성 고객의 재활성 분석

3

연관 항목 분석

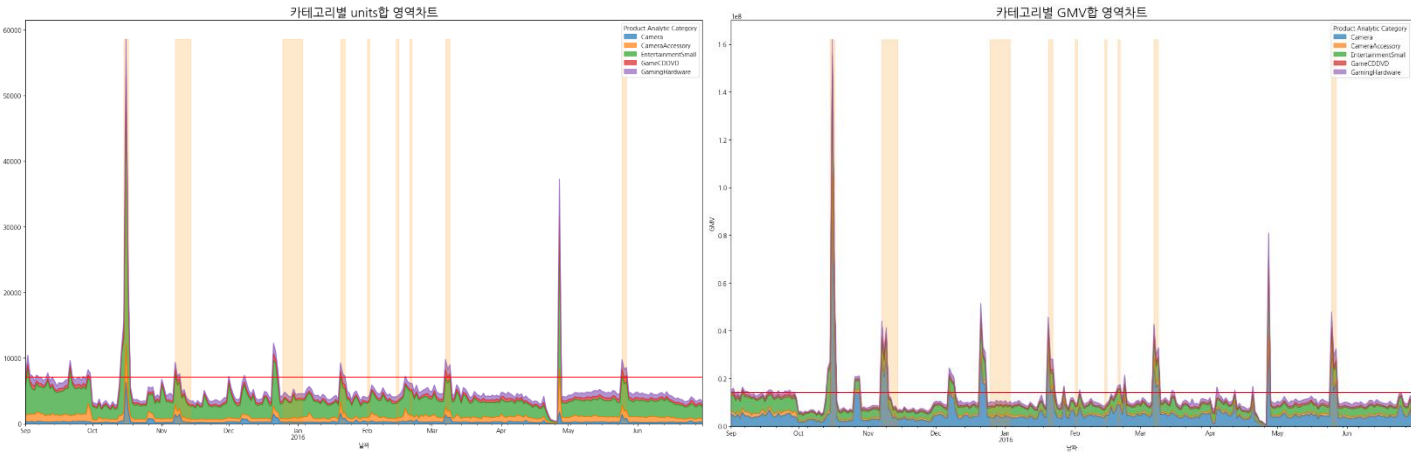
- 고객의 재구매, 재활성 유도를 목적으로 진행
- 연관 항목 조합 선정
- 항목별 재구매 증가 시 기대효과 산출

4

문제 해결 전략 제시

- 고객의 재구매, 재활성을 증가시키기 위한 전략 제시
- 충성도 높은 고객 분석

일 별 판매량 / GMV 비중 그래프



EDA를 통한 문제 정의

- 정기적인 프로모션을 수행하지만 판매량, GMV가 일시적으로만 증가
- 프로모션 기간 신규 고객 유입이 적거나 유입된 고객의 재구매 비율이 낮을 것으로 추정됨

프로모션별 고객 수 / 신규 고객 / 재구매 고객 현황

신규 유입 월별 코호트 차트

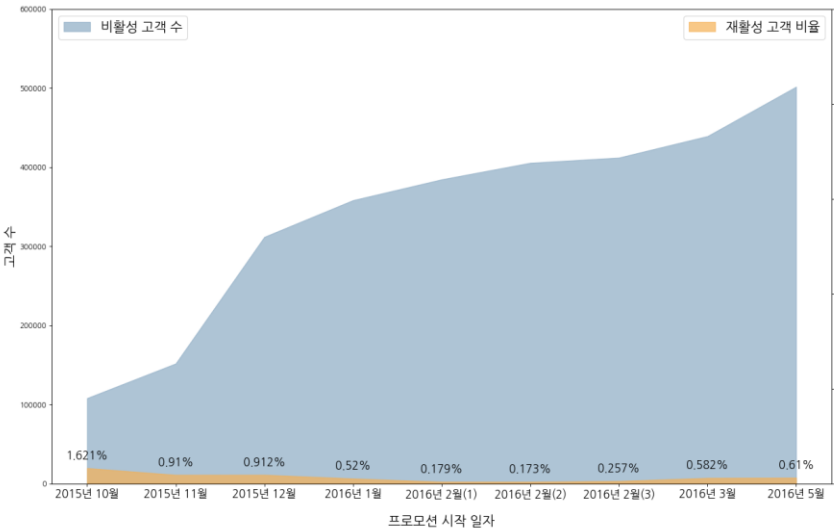
	0	1	2	3	4	5	6	7	8	9
2015-09	92532	6	3	3	2	2	2	2	2	2
2015-10	176044	4	3	2	2	2	2	2	1	
2015-11	104327	5	3	2	2	2	2	1		
2015-12	124003	5	3	2	2	2	1			
2016-01	111289	5	3	2	2	1				
2016-02	110778	5	2	2	1					
2016-03	113704	5	3	2						
2016-04	95864	5	2							
2016-05	116271	4								
2016-06	89332									

- 총 고객 수 중 94 ~ 78% 가 신규 고객
- 신규 고객 중 이후 재구매를 하는 고객은 약 10~20%에 불과함

- 첫 달은 유입 수이며 나머지는 재구매율%
- 유입 후 바로 다음 달 재구매율은 약 5%, 이후 1~3%로 감소

Project 01 – 이커머스 프로모션 성과 분석

프로모션 기간 재활성 고객 수 / 비활성 고객 수



- 비활성 / 재활성의 기준: 30일로 설정
* 30일: 프로모션 사이 평균 간격, 평균 재구매 기간
- 첫번째 프로모션을 제외하면
재활성 고객의 비율은 모두 1% 미만
- 비활성 고객 수는 계속해서 증가하지만
재활성 고객 수는 낮은 수준을 유지 중

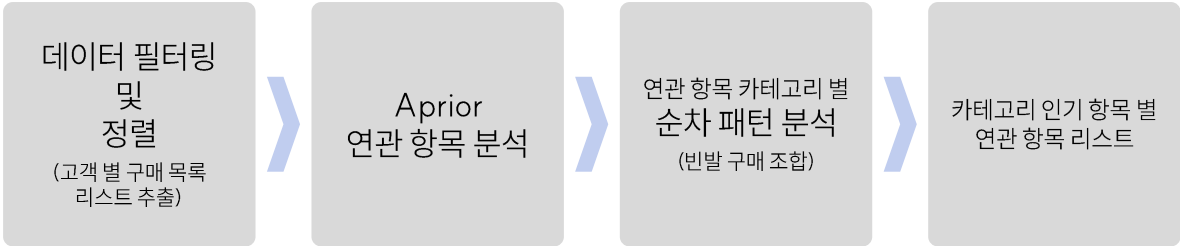
문제 상황 세분화

1) 프로모션 기간 높은 신규 고객 비율 대비 낮은 재구매율

2) 늘어나는 비활성 고객, 낮은 수준의 재활성

연관 항목 분석 및 순차 패턴 분석을 통한 문제 해결 방안 제시

- 특정 상품을 구매한 고객에게 연관 항목을 이용한 마케팅으로 재구매 / 추가 구매를 유도
- 연관 항목의 순서 및 2회 이상의 재구매 목록을 분석하기 위해 순차 패턴 분석 진행



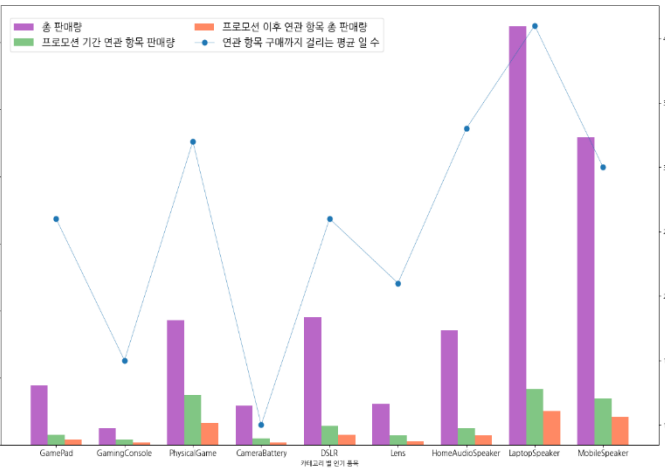
- 순차 패턴 분석: 고객 별 구매 항목 리스트에서 [A, B] 또는 [A, C] 등 각 조합의 횟수를 카운트
- 횟수를 기준으로 진행하여 판매량이 높은 항목만 등장하는 왜곡 발생
- 연관 항목 분석 결과를 반영하여 연관성이 있는 카테고리끼리 분류

GamingHardware, GameCDDVD > Game / Camera, CameraAcc > Camera / EntertainmentSmall

주요 항목별 추가구매 빈발 항목 리스트

First	Second
PhysicalGame	PhysicalGame, GamingConsole, GamingAccessoryKit, GamePad
GamePad	GamePad, GamingConsole, PhysicalGame
MobileSpeaker	MobileSpeaker, LaptopSpeaker, HomeAudioSpeaker
DSLR	Filter, CameraTripod, Lens, DSLR
LaptopSpeaker	AudioMP3Player, HomeAudioSpeaker, MobileSpeaker, LaptopSpeaker
Lens	CameraTripod, Lens, Filter, DSLR, CameraBag
HomeAudioSpeaker	LaptopSpeaker, HomeAudioSpeaker, MobileSpeaker
GamingConsole	GamePad, GamingConsole, PhysicalGame
CameraBattery	CameraBattery, CameraBatteryCharger

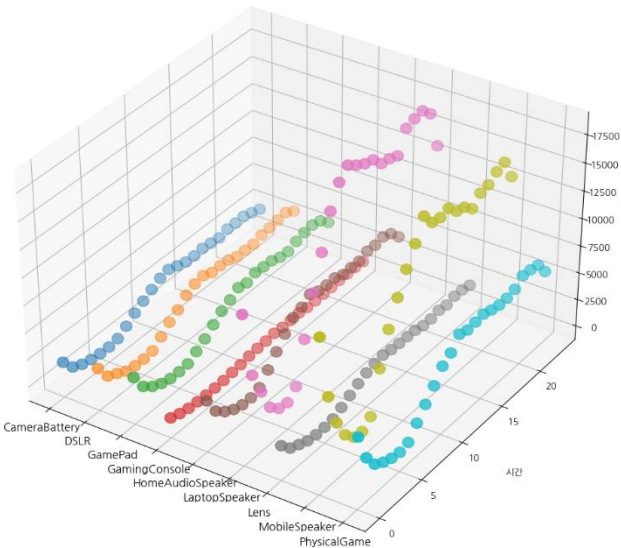
인기 항목 기간별 연관 항목 판매량



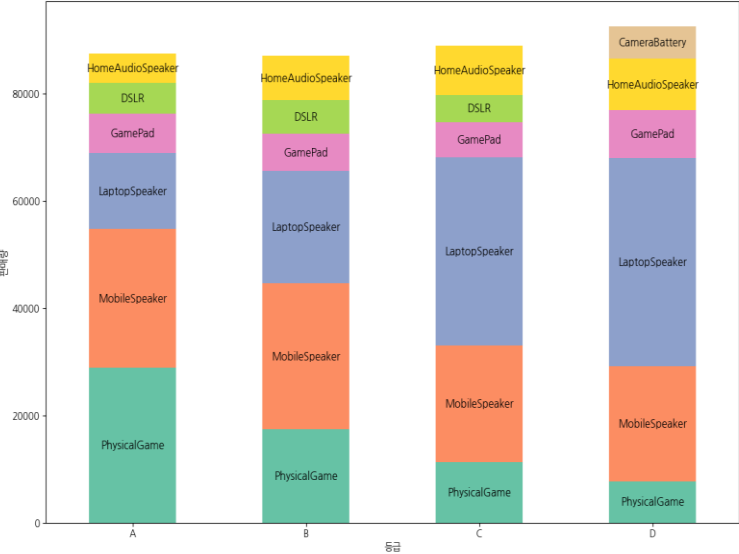
• 연관항목 구매율 평균 20%, 피지컬 게임 57%

Project 01 – 이커머스 프로모션 성과 분석

항목별 판매 시간대 현황



고객 등급별 구매 항목



주요 항목 고객 활동 시간 파악

- 1) 항목별 판매량 차이는 있으나 주요 활동 시간대는 유사함
- 2) 12시 ~ 22시 높은 거래량 유지



재구매 / 재할성 증가 전략 1

- 빈발 항목 리스트를 활용하여 주요 상품 구매 시 연관 항목 추가 구매 유도
- 주요 활동 시간대를 공략하여 연관 항목 마케팅
피지컬 게임 등 게임 관련 항목: 20~22시
스피커, 카메라 관련 항목: 11~13시, 20~22시

고객 등급별 구매 항목 차이 파악

- 1) RFM 점수 사분위수를 기준으로 등급 부여
- 2) 등급이 높아질수록 피지컬 게임 비중 증가



재구매 / 재할성 증가 전략 2

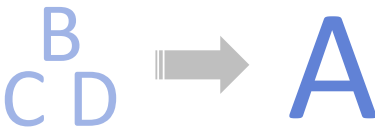
- 충성도 높은 고객 증가로 간접적인 해결
 - 게임 이외 카테고리 구매 고객에게 재구매율이 높은 피지컬 게임 할인/지급
 - 재구매율 효율적으로 증가

기대 효과

항목별 판매 증가 시 기대 이익

	First	3%	5%	10%
PhysicalGame	22136238.0	36893730.0	73787460.0	
GamePad	14919794.0	24866324.0	49732647.0	
MobileSpeaker	10357133.0	17261889.0	34523777.0	
DSLR	10292464.0	17154107.0	34308214.0	
LaptopSpeaker	7539036.0	12565060.0	25130121.0	
Lens	6409680.0	10682800.0	21365600.0	
HomeAudioSpeaker	3351028.0	5585047.0	11170094.0	
GamingConsole	1936961.0	3228269.0	6456538.0	
CameraBattery	658881.0	1098136.0	2196271.0	

등급별 피지컬 게임 판매 규모 성장 시 기대 이익



피지컬 게임 고객 유입 전략을 통하여
B, C, D 등급 피지컬 게임 고객의 재구매율이
A등급만큼 성장하였을 때 판매 증가에 따른 기대이익

Grade	3%	5%	10%
B	17737613	18082033	18943082
C	14387801	14667176	15365613
D	13649681	13914724	14577330

고객 활동 시간에 맞춘 연관 항목의 판매 촉진 전략을
통하여 판매가 증가됨에 따라 얻게 되는 연간 기대 이익

Project 02

웹툰 데이터 EDA 및 키워드 분석

2024-06-10 ~ 2024-06-17 (1주)

인원	1명 (개인 프로젝트)
분석 환경	Python, Power BI
주요 분석	데이터 EDA, 한국어 자연어 처리, 키워드 빈도와 중요도 측정 및 유사도 측정
분석 목적	웹툰 장르별 트렌드 분석 및 트렌드에 맞는 베스트 도전 작품 선정

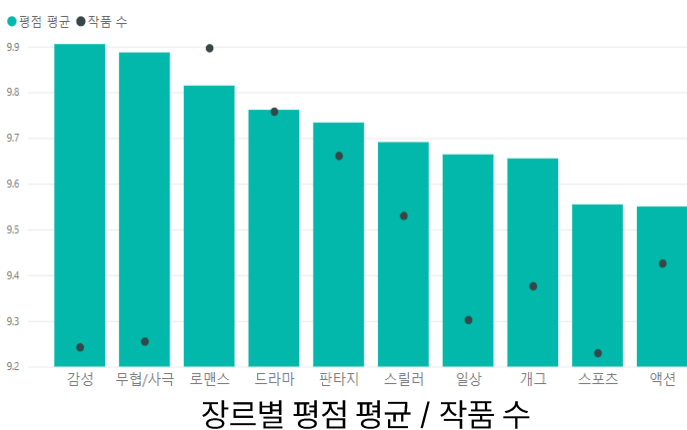
Lesson & Learn	<p>1) 보다 정교하게 키워드를 처리하기 위해서는 의미가 있는 부분만, 똑같은 형태로 만들어야 하며 한국어에서는 띄어쓰기를 옳게 교정해야 정확한 어근 추출이 가능하다는 것을 배움.</p> <p>2) 분석 목적과 텍스트의 성질에 따라 명사만 사용할지, 동사도 사용할지 등 분석에 사용할 품사의 선택이 분석 결과에 큰 영향을 준다는 것을 배움.</p> <p>3) 크롤링을 통해 평점 외의 다른 지표를 수집하려 했으나 실패함. 평점이라는 하나의 지표만 활용한 것에 큰 아쉬움이 남음.</p>
----------------	--

웹툰 데이터 EDA 및 키워드 분석

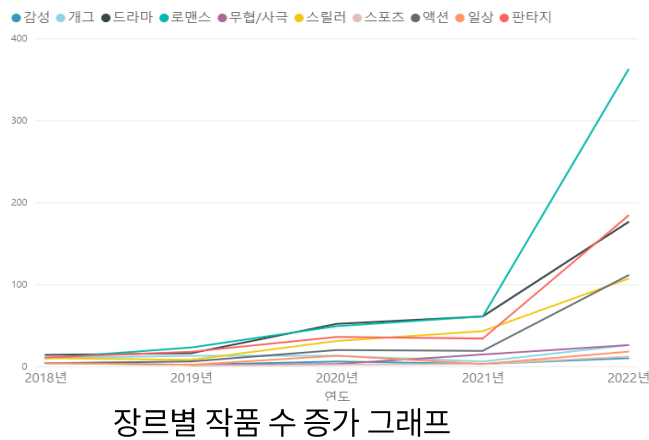
분석 과정

- 1) 작품 평점을 기준으로 웹툰 장르 특성 분석 및 인사이트 추출
- 2) 인기 웹툰 키워드 분석을 통한 트렌드 파악
- 3) 베스트 도전 웹툰 키워드 분석을 통해 잠재된 인기 작품 선정

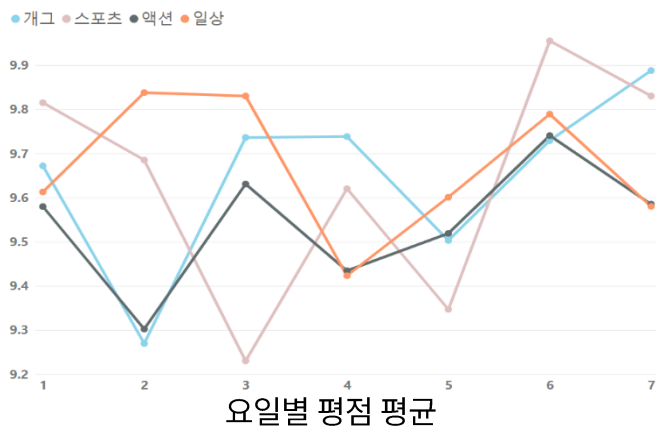
* 베스트 도전 웹툰: 정규 웹툰이 되기 위한 신인, 무명 작가들의 작품



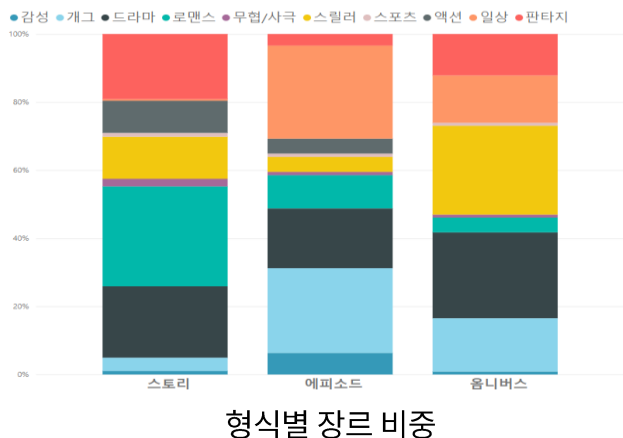
- 감성, 무협/사극 아주 적은 작품 수에도 높은 평점
- 로맨스, 드라마, 판타지 많은 작품 수와 높은 평점
- 나머지 장르는 상대적으로 낮은 작품 수와 평점
평점 7점 이하의 작품들이 존재



- 조금씩 증가하는 추세를 보이다 2022년 급증
- 로맨스, 판타지, 드라마, 스릴러, 액션 장르
특히 급증



- 요일별 평점 평균 차이가 있는 장르만 표시
- 스포츠, 액션 토요일에 가장 높은 평점
- 개그, 액션은 화, 스포츠는 수요일 가장 낮은 평점
- 해당 장르들은 비교적 불안정한 장르로
요일별, 작품 별로 평점의 편차가 큼

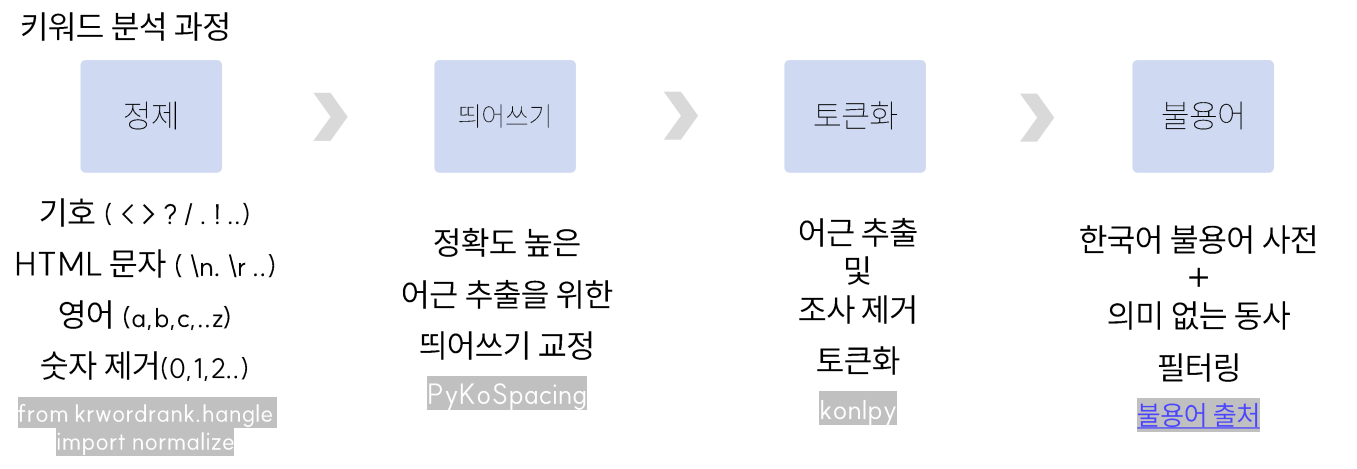


- 형식별 주요 장르에 큰 차이가 있음
- 일상, 개그 장르 에피소드 형식에서 큰 비중
- 스릴러 장르 유니버스에서 큰 비중
- 각 장르별 주요 형식이 존재

인기 웹툰 장르별 분석

- 1) 감성, 무협/사극 장르는 100개 미만의 적은 작품 수와 높은 평가로 성장 가능성이 가장 큰 장르
- 2) 로맨스, 드라마, 판타지 장르는 400개 이상의 많은 작품, 높고 안정적인 평점으로 가장 안정적인 시장을 가진 장르
- 3) 나머지 장르는 특정 형식, 작품, 작가 등의 영향을 크게 받는 불안정한 장르

Project 02 – 웹툰 데이터 EDA 및 인기 웹툰 키워드 분석



키워드 추출	description	keywords
소꿉친구와 하루아침에 커플이 되었다?!통제가 강한 부모님들을 안심 시키고 자유롭게 생활하기 위해 동맹을 맺은 두 모범생 '윤세이'와 '김재하'동맹을 들릴 위기를 모면하기 위한 두 사람의 가짜 커플 행세가 시작된다!		[소꿉친구, 하루아침, 커플, 통제, 강하다, 부모님, 안심, 자유롭다, 생활, 동맹, 맺다, 모범생, 세이, 김재, 동맹, 들, 키다, 위기, 모면, 사람, 가짜, 커플, 행세, 시작]
고려, 조선시대를 거쳐 바리스타가 된 계룡산 선녀님(女, 699세)환생한 서방님과 날개웃을 찾을 수 있을까		[고려, 조선시대, 거처다, 바리스타, 계룡산, 선녀, 환생하다, 서방, 날개, 찾다]
정체불명의 3인조, 천진난만 켄, 카리스마 선녀, 쉬크노노wn루키루가 만들어내는 황당무계하지만 따뜻한 이야기들. 지금부터 루키루와 함께 고고		[정체, 불명, 난만, 카리스마, 선녀, 크다, 루키루, 만들다, 당무, 따뜻하다, 지금, 루키루, 고고]

- 정규 웹툰 소개 (description)의 키워드 추출, 분석
- 정규 웹툰 중 평점이 장르별 평균 이상인 작품 인기 작품으로 선정



베스트 도전 웹툰 선정

- 베스트 도전 웹툰 description과 정규 웹툰 간의 키워드 유사도 분석
- 각 작품의 TF-IDF 벡터 생성 후 인기 작품 키워드와 코사인 유사도 측정
- 장르별 예상 인기 작품 top10, 예상 인기 작품 top50 최종 선정
- 정규 웹툰 승급 시 인기 작품 될 것으로 예상되는 베스트 도전 작품을 선정

	title	author	summary	genre	format
	짐승의 이빨처럼	shad****	작수를 죽인다.	action, drama	스토리
	Baby Boy Brother	roze****	연하남 9인 공작기	fantasy, pure	스토리
	밤을 거둔 신기루는	kims****	부적 남주 정취기	fantasy, pure	스토리
	1년 동안 잘 부탁해!	mono****	우당탕탕 기속학교!	drama, pure	스토리
	취사병 전설이 되다	leej****	군대요리 판타지만화	fantasy, drama	스토리
	스크류 드라이버	cenc****	아이를 잃은 아버지	drama, thrill	스토리
	원포인트	joko****	배구를통한성장드라마	drama, sports	스토리
	대공녀는 거기에 없었다	mihy****	가상중세시대를	drama, historical	스토리
	고민상담특별부	vhxm****	랭킹 시스템 학교	comic, drama	스토리
	바람을 실은 메아리 (바람과 소원)	vari****	정식연재공지!	drama	스토리

Project 03

은행 고객 분석 및 가입 예상 고객 분류

2024-06-19 ~ 2024-06-28 (1주)

인원	1명 (개인 프로젝트)
분석 환경	Python, Power BI
주요 분석	고객 특성 데이터 & 경제 지표 EDA, 분류 머신러닝
분석 목적	정기 예금 가입 증가, 마케팅 비용 최소화를 위한 고객 특성 분석 및 가입자 예측 모델 구현

Lesson & Learn	<p>1) 타겟 변수 클래스 불균형이 심할 경우 다수의 클래스로만 예측이 이루어짐. 따라서 정확도는 높지만 실제로는 올바르게 예측을 하지 못하는 상황을 알게됨.</p> <p>2) 오버 샘플링, 언더 샘플링, 하이퍼 파라미터 튜닝을 통해 클래스 불균형 문제를 해소하기 위해 다양한 시도 경험. 그 과정에서 Confusion Matrix를 활용하여 모델을 해석하고 더 알맞은 전처리를 탐색함.</p> <p>3) 모델링의 목적과 데이터 형태에 따라 정확도, 정밀도, 재현율 중 어떤 수치가 중요한 지 어떻게 높은 수치를 얻을 수 있을지 고민하여 결과를 도출함.</p>
-------------------	--

은행 고객 분석 및 가입 예상 고객 분류

- 분석 과정
- 1) 경제 지표와 정기 예금 가입률 관계 파악

2) 고객 정보 EDA를 통해 정기 예금 가입 고객 특성 파악

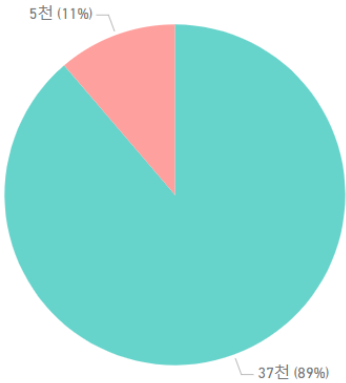
3) 가입 예측 모델 구축 및 예상 가입자 선정

데이터

고객 특성	경제 지표	목표 변수
나이, 직업, 결혼 여부, 교육 수준, 신용 카드 유무, 주택 소유 등... 총 15개 칼럼	고용 변동률, 소비자 물가 지수, 소비자 신뢰 지수, 유리보 3개월 비율 지역 고용 수 총 5개 칼럼	정기 예금 가입 여부

- 총 21개의 칼럼
- 고객 특성, 경제적 지표 등의 정보

가입률

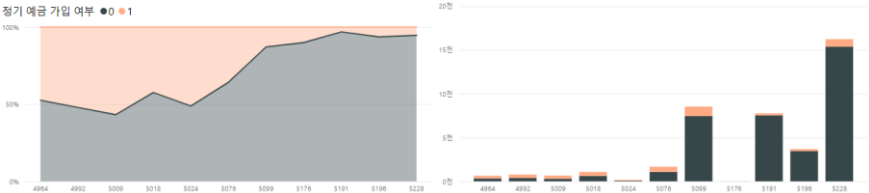


- 정기 예금 가입률은 약 11%

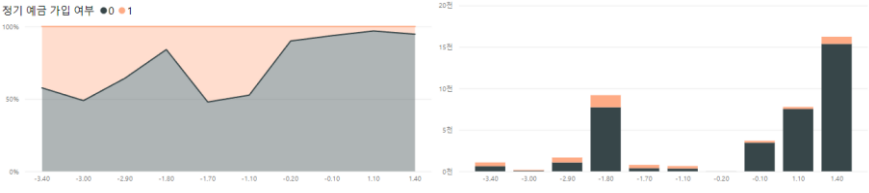
경제 지표

정기 예금 가입 여부 ●No ●Yes

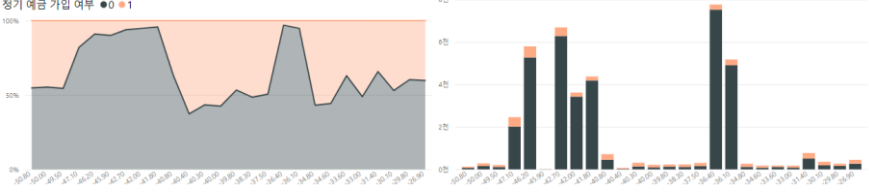
지역 고용 수별 가입률 / 고객 수



고용 변동률별 가입률 / 고객 수



소비자 신뢰 지수별 가입률 / 고객 수



- 고객 수가 많아질수록 가입률이 낮음
- 지역 고용 수, 고용 변동률 낮아질수록 가입률 증가 높아질수록 고객 수 증가
- 소비자 신뢰 지수 높아질수록 가입률 증가 낮아질수록 고객 수 증가

★ 경제 지표는 고객의 가입 시점을 기준으로 함
★ 지역 고용 수: 고객이 속한 지역의 고용 수

경제 지표와 정기 예금 가입률

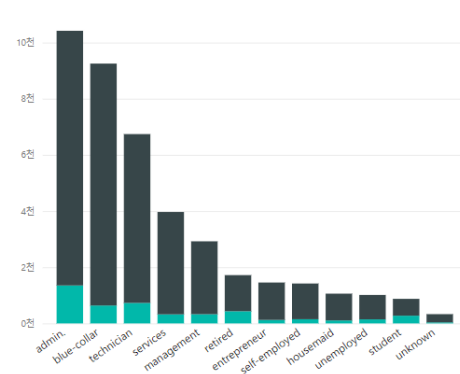
- 1) 경제 지표별 가입률과 대략적인 추세 존재
- 2) 고객 수가 많은 시기 가입률이 가장 낮음 (3~5%), 고객 수가 낮은 시기 가입률이 높음 (40~50%)
- 3) 각 지표별 가입률의 차이는 크지만 가입 고객 수의 차이는 비교적 낮은 수준 (400명 이하)

Project 03 – 은행 고객 분석 및 가입 예상 고객 분류

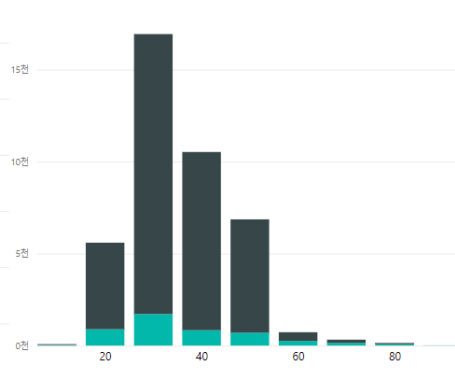
고객 특성

정기 예금 가입 여부 ●No ●Yes

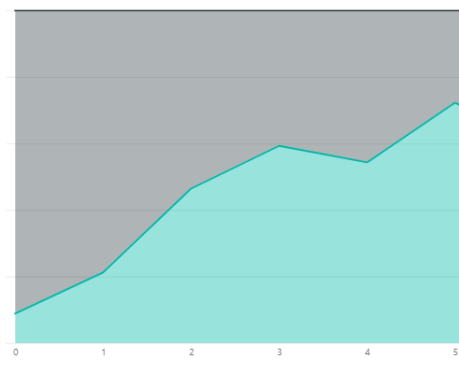
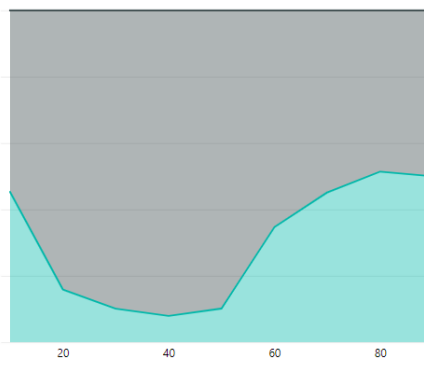
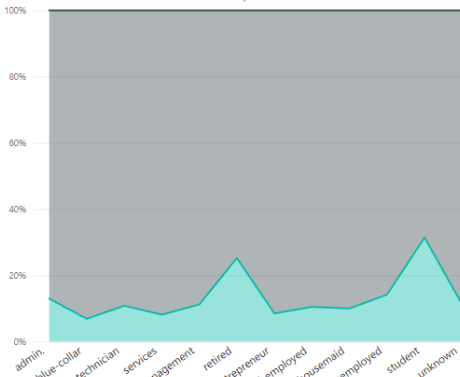
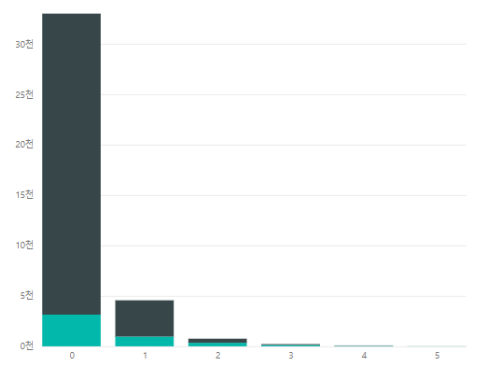
직업별 고객 수 / 가입률



나이대별 고객 수 / 가입률



연락 횟수별 고객 수 / 가입률

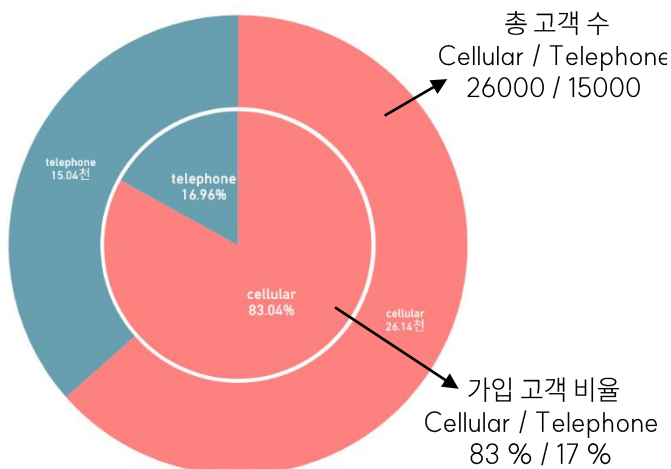


- 직업별 가입률은 10% 미만
- student, retired 가입률 각각 31, 25%

- 30, 40, 50대 고객 수 Top 3
- 고객 수가 적은 그룹 가입률 높음
- 20, 30대 고객 수와 가입률 높은 편

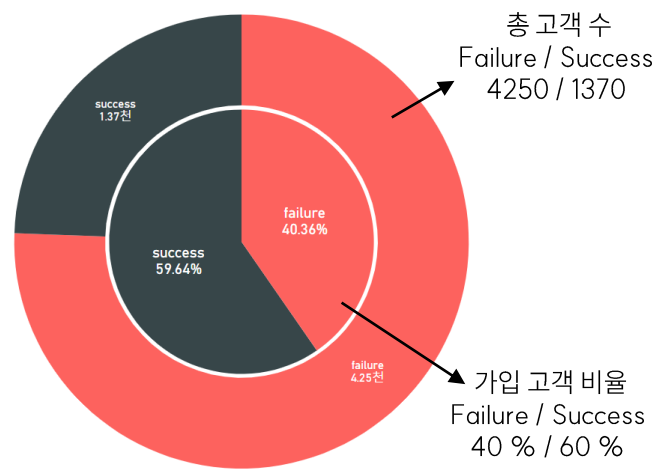
- 대부분 연락 횟수 0회 (35000명)
- 연락 횟수가 증가할수록 가입률 크게 증가 (8 -> 70%)

연락 수단별 총 고객 수 / 가입 고객 비율



- 가입 고객 Cellular 비중이 압도적
- Cellular 가입률: 15%
- Telephone 가입률: 5%

이전 마케팅 결과별 총 고객 수 / 가입 고객 비율



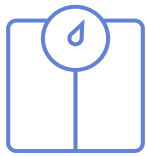
- 이전 마케팅 성공 여부가 정기 예금 가입에 큰 영향
- 이전 마케팅 성공 고객의 65%가 이후 정기예금에 가입

고객 특성과 정기 예금 가입률

- 1) 고객 수가 많은 admin, blue-collar, technician / 가입률이 높은 student, retired 타겟 전략 필요
- 2) 가입률이 높은 student, retired 고객과 나이대 (10대, 60대 이상) 고객이 중복 / 해당 특성의 고객 유입 증가
- 3) 직업, 나이대 고객 수 높은 집단 대상 연락, 마케팅 횟수 증가

Project 03 – 은행 고객 분석 및 가입 예상 고객 분류

머신 러닝 알고리즘



XGBoost Classifier

- max_depth, n_estimators 등 과적합 방지 파라미터 튜닝 가능
- scale_pos_weight로 클래스 가중치 조정 가능

- 클래스 불균형
- 36548 : 4640 (9:1)

- 데이터 수: 41188
- 범주형 변수 10개
- 연속형 변수 10개

기본 모델 정확도 보고서

* 클래스 0 : No / 1 : Yes

	Accuracy	Precision / 0	Recall / 0	Precision / 1	Recall / 1
Train set	0.933	0.94	0.99	0.90	0.46
Test set	0.896	0.91	0.97	0.58	0.28

- 0.037 정도의 과적합
- 정확도에 비하여 클래스 1의 정밀도, 재현율이 현저히 낮음

개선 목표

- 실제 가입 확률이 높은 고객을 놓치지 않는 것이 중요
- 따라서 클래스 1의 실제 양성을 놓치지 않아야 함

파라미터 튜닝

- GridSearchCV 사용
- Recall 스코어 기준으로 서치

개선 모델 정확도 보고서

	Accuracy	Precision / 0	Recall / 0	Precision / 1	Recall / 1
Train set	0.856	0.99	0.85	0.44	0.92
Test set	0.854	0.99	0.85	0.43	0.92
Threshold	0.906	0.97	0.92	0.56	0.77

Confusion Matrix

10136	832	9280	1688
322	1067	105	1284
Threshold		Test set	

- 임계값 조정 시 정확도 0.85 -> 0.9 상승
- 하지만 Recall 0.15 하락으로 인하여 기존 모델 채택

잠재 고객 선정

- 정기 예금 가입하지 않은 고객 데이터 이용
- 모델 활용하여 가입 확률 산출
- 가입 확률이 85% 이상인 고객 선정

잠재 고객 수: 3725

ID	%
40713	0.978329
39783	0.976785
40836	0.974522
40257	0.974485

결론 및 기대효과

- 경제 지표별 고객 수를 이용하여 마케팅 기간 설정 [고용 변동률 -0.1 이상 / 소비자 신뢰 지수 <= -41.8 이하]
- 고객 수가 많은 직업 및 나이 고객 대상 정기 예금 가입 마케팅 [admin, blue-collar, technician / 20~50대] **가입률 증가**
- 가입률이 높은 직업 및 나이 고객 신규 유입 증가 전략 [student, retired / 10, 60대 이상] **고객 수 증가**

	1%	3%	5%	10%
customer	359	1078	1796	3592
• 가입률 증가 시 정기 예금 가입 기대 고객 수				
	100	500	1000	1500
customer	880	1036	1231	1426
• 고객 수 증가 시 정기 예금 가입 기대 고객 수				

Project 04

게임 데이터 분석을 통한 이상 유저 탐지

2024-07-01 ~ 2024-07-08 (1주)

인원	1명 (개인 프로젝트)
분석 환경	Python, Power BI
주요 분석	게임 데이터 EDA, 이상 행동 정의, 유저 행동 이상탐지
분석 목적	각종 어뷰즈 유저의 특성 파악 및 이상 유저 탐지

Lesson & Learn	<p>1) 이상치의 기준점을 정하기 위해서 데이터를 여러 시각에서 살펴보고 고민함. 적용 분야와 상황에 따라 통계적 방법만으로는 한계가 있음을 배움,</p> <p>2) 게임 데이터 와 같이 특수한 데이터를 분석하기 위해서는 단순히 해당 분야의 지식 뿐만 아니라 보다 깊게 해당 게임을 직접 플레이 해본 경험 또한 필수적이라는 것을 느낌.</p>
----------------	--

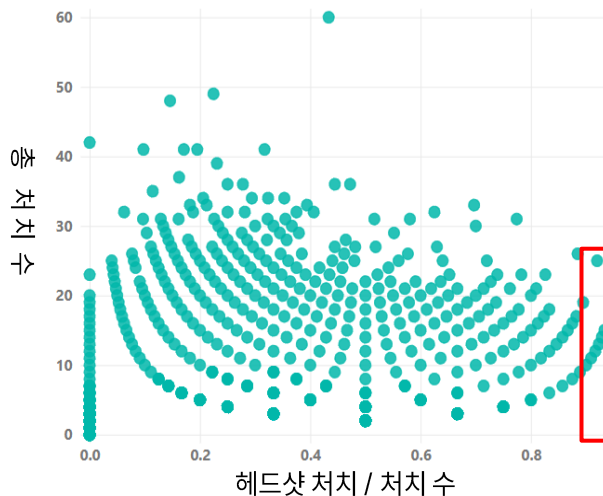
게임 데이터 분석을 통한 이상 유저 탐지

- 분석 과정
- 1) 이상 행동 정의 및 이상치 탐색
 - 2) 이상 행동 의심 유저 선정

1

높은 수준의 헤드샷 처치

총 처치 수와 헤드샷 처치율 분포



총 처치 수별 유저 수 변화

* 헤드샷 처치율 0.9 이상



- (헤드샷 처치 수 / 총 처치 수) → 헤드샷 처치율 도출
 - 헤드샷 처치율이 0.9 이상을 이상치로 정의
 - 하지만 총 처치 수가 적다면 이상치가 아닐 수 있음
ex) 총 처치 수: 3, 헤드샷 처치 수: 3
 - 총 처치 수 상위 0.1%(12킬)의 제한 추가



이상 유저 선정

헤드샷 처치율이 0.9 이상이며
총 처치 수가 상위 0.1% 이상인 유저
총 22명의 이상 유저 데이터 출력

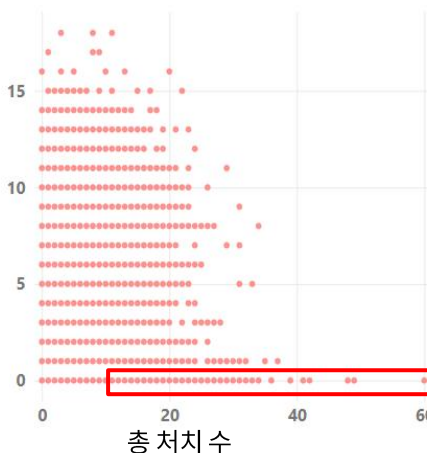
2

무기 / 아이템 사용 없이 높은 처치 수

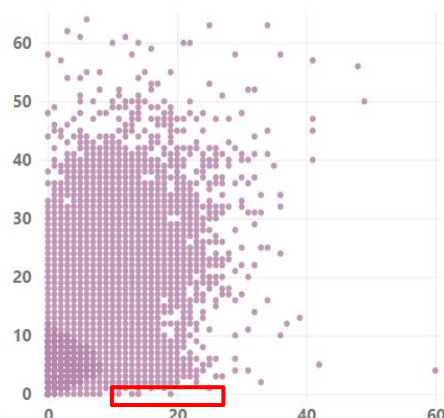
Heal아이템



Boost 아이템



무기



- Heal, Boost 아이템 사용 없이
총 처치 수 상위 0.1%(12킬) 이상
- 무기 획득 없이 총 처치 수 상위 0.1%(12킬) 이상
- 위의 두 경우를 이상 행동으로 정의



이상 유저 선정

아이템 사용 0회 / 무기 획득 0회
&
총 처치 수가 상위 0.1% 이상인 유저
총 352명의 이상 유저 데이터 출력

Project 04 – 게임 데이터 분석을 통한 이상 유저 탐지

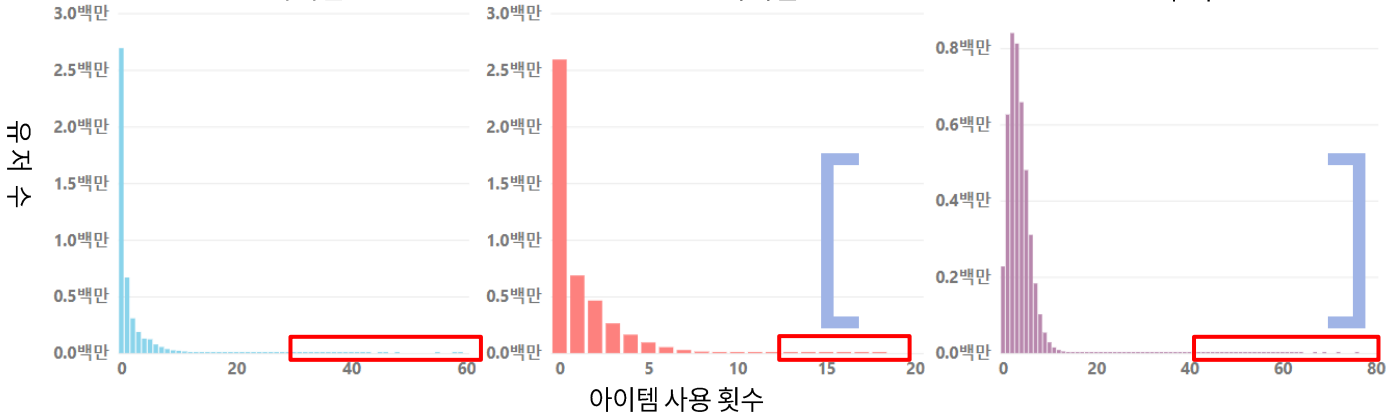
3

무기 / 아이템 사용 횟수 이상

Heal 아이템

Boost 아이템

무기



- 아이템 / 무기 획득 및 사용량이 상위 0.01% 이상인 경우 이상치로 정의
- Heal: 30개 / Boost: 13개 / 무기: 42개
- [총 처치 수 1 이상] 조건을 추가하여
아이템 및 무기 획득의 기회가 존재했던 유저 선별

이상 유저 선정
아이템 / 무기 획득 및 사용량
상위 0.01% 이상이며
총 처치 수가 1 이상인 유저
총 600명의 이상 유저 데이터 출력

4

이동 없이 무기 / 아이템 획득 및 사용

Heal 아이템

Boost 아이템

무기



- (걸음 이동거리 + 수영 이동거리 + 탑승 이동거리)
→ 총 이동거리 도출
- 이동거리가 0 이지만 아이템 / 무기를
획득 및 사용한 것은 분명히 이상 행동임
- 그 중 아이템 / 무기 사용량 상위 0.01% 이상 선별

이상 유저 선정
총 이동거리가 0이며
아이템 / 무기 획득 및 사용량
상위 0.01% 이상인 유저
총 27명의 이상 유저 데이터 출력

추가 정보 수집 계획

1

헤드샷 이상 유저

처치에 관한 정보를 넣어
명중률 등의 정보가 있다면
보다 정밀한 탐지가
가능할 것으로 보임

2 아이템/무기 미사용

플레이어 킬 로그와 같은
데이터를 활용하여
처치 시 행동, 도구 등을
분석한다면 이상 행동을
명확히 파악할 것으로 추정

3 아이템/무기 과사용

매치별 아이템/무기 수량,
Drop 확률, 분포 지역 등을
활용하여 이상치의 정의와
이상 행동 패턴을 더 깊게
파악할 수 있을 것임

4

이동 없이 장비 획득

아이템/무기 drop 지역,
밀도 등을 확인하여
이상 행동임을 명확히 확인
해당 현상 모니터링 필요