

2019 엠브레인 패널 빅데이터® 분석 경진대회

@Epdel.

류재원 응용통계학과
박인서 언론홍보영상학과
이승명 경영학과
임지훈 경영학과

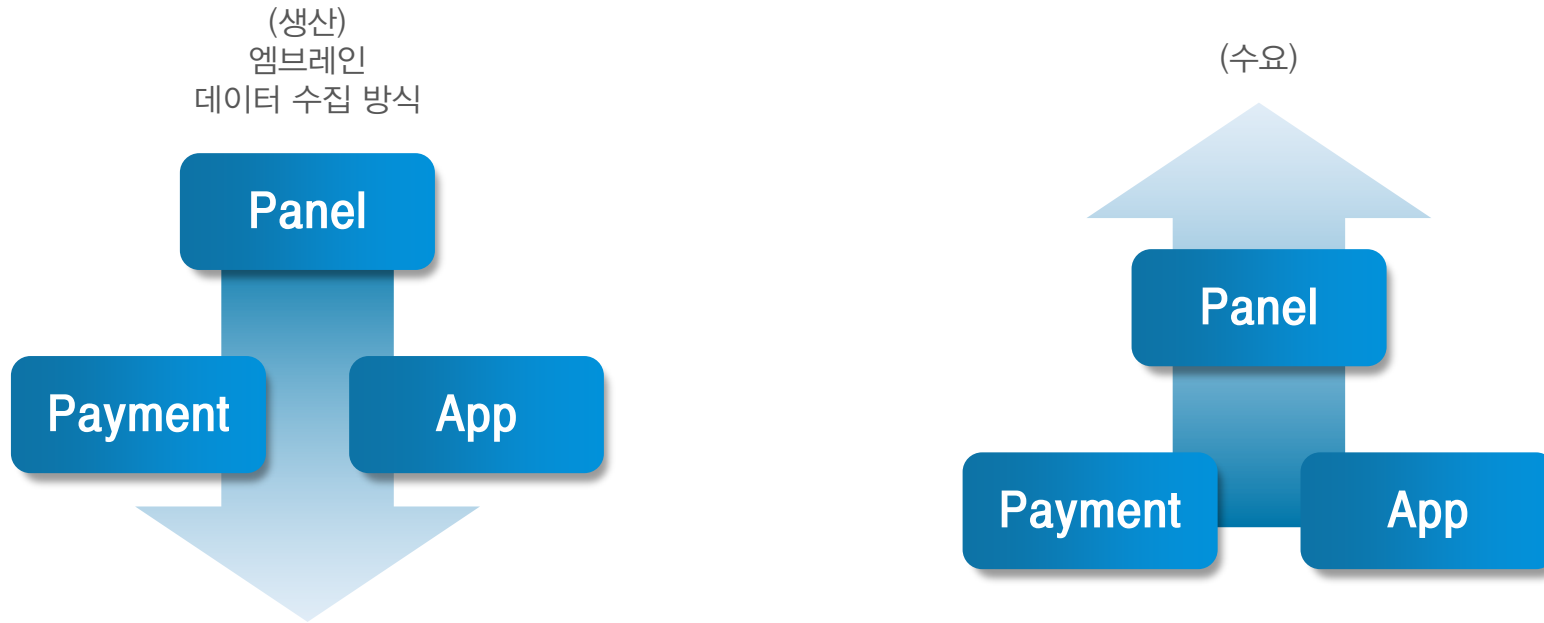
컴페토
[competō]

in classical Latin "agreement,"

경쟁이라는 단어는 라틴어에서 왔다. 말 그대로 옮기면 함께 노력한다는 뜻이다. 어
원에는 다른 사람을 패배시켜야 한다는 뜻이 전혀 없다. 그래서 협력이 곧 경쟁이다.

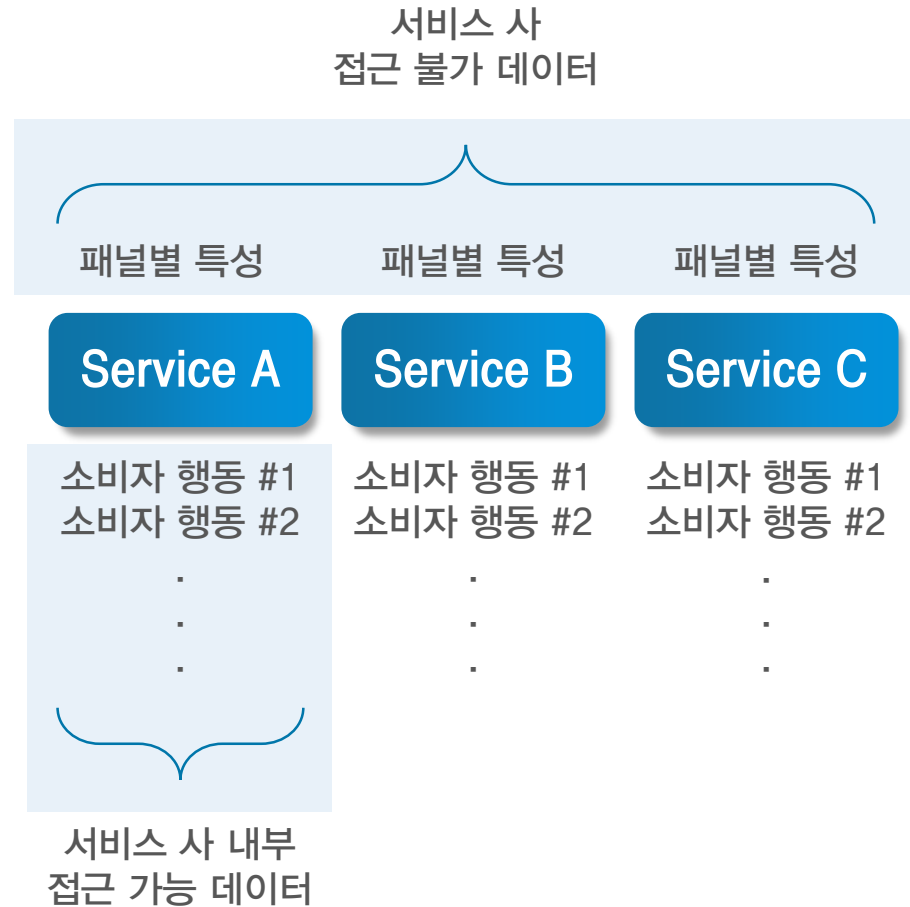
마이크 제바이스 (Mike Gervais)

엠브레인의 데이터를 실 수요자 입장에서 바라보다.



데이터의 수요자는 **자신의 서비스를 기준**으로 한 분석을 원한다.

고객 서비스 사 입장에서 엠브레인의 데이터는 접근 가능한 데이터와 불가능한 데이터로 나뉘어진다.



서비스 사 내부 접근 가능 데이터

1. 해당 서비스의 소비자의 행동 정보
Ex) [Payment] 구매상품, 구매시간, 구매금액
Ex) [App] 접속시간, 인앱결제내역, 앱내행동정보

:이미 더 높은 정확성을 가진, 더 많은 샘플을 보유할 가능성

서비스 사 내부 접근 불가 데이터

1. 해당 서비스의 소비자 정보
Ex) [패널] 나이, 성별, 지역, 소득수준, 관심사 등
2. 타 경쟁 서비스의 정보

“고객 서비스 사가 내부적으로 접근할 수 없는 데이터,
타 경쟁 서비스의 데이터를 활용한 경쟁 분석”

Agenda

A. 경쟁군 비교: 누가 내 경쟁자인가?

- (1) 카테고리간 분석 - 네트워크 분석을 통한 카테고리 간 관계
- (2) 키워드 유사성 분석 - TF-IDF를 활용한 게임 리뷰 키워드 분석 및 유사성 비교
- (3) 비교군 경쟁분석 - 비교군 내/외부의 경쟁성 비교
- (4) Switching Analysis - 전환위험지수와 진입가능지수를 통한 전환방향/크기 분석

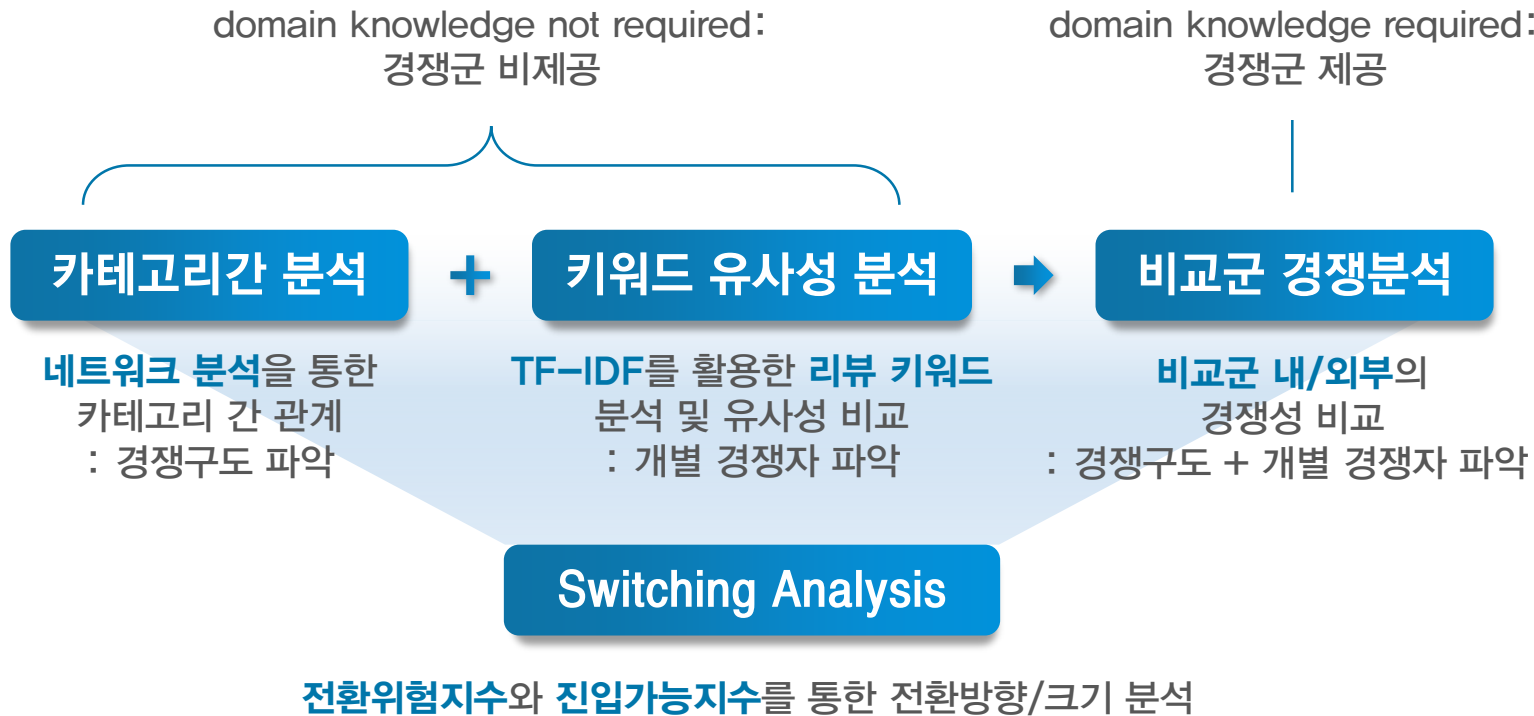
B. 협력은 또 다른 경쟁: 누구와 협력해야 하는가?

- (1) 네트워크 분석 - TF-IDF를 활용한 강남 내 결제 상점 간 유사성 및 네트워크 분석

A. 경쟁군 비교 분석 : 핵심 질문

1. 누가 내 경쟁자인가?
2. 경쟁강도는 어느 정도인가?

A. 경쟁군 비교 분석 : 핵심 질문



Who?

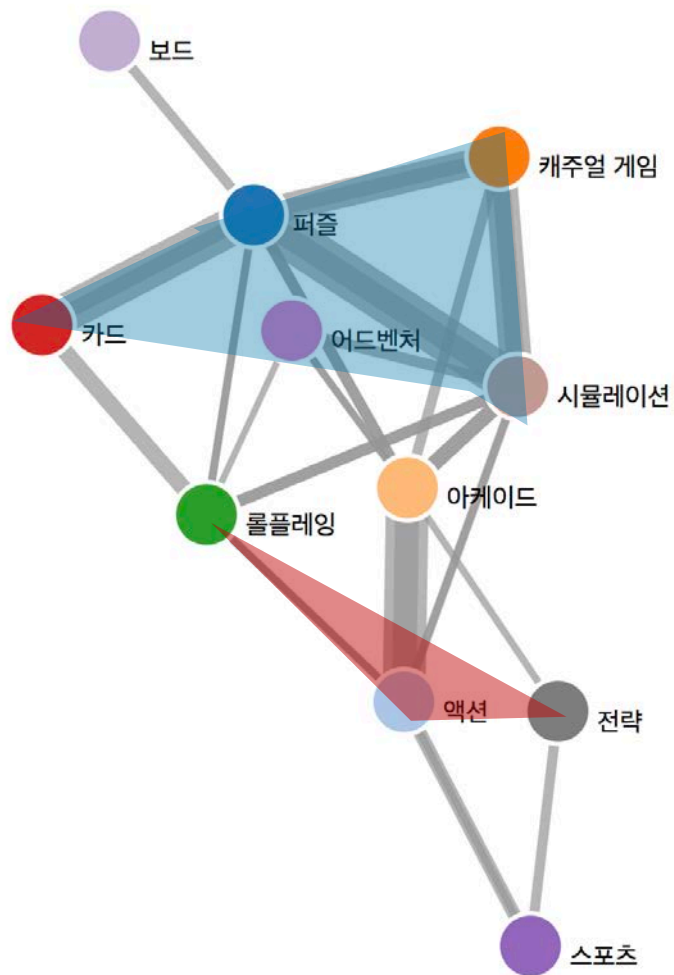
누가 경쟁자인가?

How?

경쟁 방향은 어떠한가?

경쟁 강도는 어느 정도인가?

A. 경쟁군 비교 분석 : WHO? (1) 네트워크 분석을 통한 카테고리간 관계 파악



어떤 카테고리의 게임들을 동시에 플레이하는가?

패널별로 각 주차에 플레이한 게임목록 추출
카테고리 맵핑 후 카테고리쌍별 등장 횟수 계산 후 표준화

	genre1	genre2	value	normed_value
47	롤플레이밍	롤플레이밍	9409	2.593729
48	롤플레이밍	보드	1022	0.238997

[퍼즐, 캐주얼게임, 시뮬레이션, 카드]간의 공통 플레이 횟수가 높음

유사한 장르의 게임이 함께 플레이됨
: 한 판 한 판 가볍게 할 수 있는 진입장벽이 낮은 게임

[롤플레이밍, 액션, 전략]간의 공통 플레이 횟수가 낮음

유사한 장르의 게임이 함께 플레이 되지 않음
: 캐릭터/장비 등 장기적인 투자가 요구되는 진입장벽이 높은 게임

A. 경쟁군 비교 분석 : WHO? (1) 네트워크 분석을 통한 카테고리간 관계 파악

어떤 개별 게임들이 동시에 플레이되는가?

패널별로 각 주차에 플레이한 게임목록 추출
게임쌍별 등장 횟수 계산 후 각 게임별로 표준화,
n(=100)명 이상이 플레이한 게임만 추출
동시 플레이쌍의 등장 횟수가 최소 100이상인 쌍만 추출

특정 게임들간의 종속관계

- IV Go는 Pokemon Go의 부가앱
- MOBI, Zzi는 게임 사전예약, 인앱결제 가계부 등의 유틸리티앱
- 상하이 애니팡은 애니팡의 스피노프 게임 (동일 회사)
- 붐비치와 클래시오브클랜은 동일 회사의 게임

특정 게임들간의 높은 동시 플레이 횟수

- 스누피 틀린그림찾기와 상하이 애니팡
- 애니팡3와 위베어베어스 더퍼즐
: 가벼운 퍼즐 게임으로, 이를 함께 즐기는 경우가 많음

	game1	game2	normed_value1	normed_value2
1101	Pokemon GO	IV Go	0.044915	0.962199
1889	MOBI	Zzi	0.059276	0.740385
1737	상하이 애니팡	애니팡 사천성	0.520240	0.175518
1780	붐비치	클래시오브클랜	0.472483	0.080237
3515	스누피 틀린그림찾기	상하이 애니팡	0.177982	0.290855
3805	애니팡3	위베어베어스 더퍼즐	0.152647	0.103925

A. 경쟁군 비교 분석 : WHO? (2) TF-IDF를 활용한 게임 리뷰 키워드 분석 및 유사성 비교

The screenshot shows the Google Play store page for the game 'Boom Beach' (붐비치) by Supercell. The game has a 4.5-star rating from 5,781,413 reviews. The description mentions 'Boom Beach에 오신 것을 환영합니다!' and '패배의 쓰라림을 맛보지 않으려면 작전은 필수입니다!'. A user review from 'j sh' dated May 15, 2019, mentions '전함' (battleship) and '엔진' (engine) updates.

붐비치
Supercell 전략
★★★★★ 5,781,413

Boom Beach에 오신 것을 환영합니다! **패배**의 쓰라림을 맛보지 않으려면 **작전**은 필수입니다!

투노와 **용맹**을 발휘해 사악한 블랙가드를 물리치는 것이 이 멋진 전략 게임의 목표입니다. 적 기지를 공격해 섬 **원주민**들을 해방하고 열대의 낙원에 숨겨진 비밀을 밝혀보세요. 전 세계의 플레이어들과 특수 작전반을 구성해 함께 적을 몰아내세요. 정찰하고 작전을 세워 화끈한 **공격**을 퍼부어야 합니다!

4.5
★★★★★
총 5,781,413건

j sh
★★★★★ 2019년 5월 15일
오늘 **전함**에 대한 업데이트로 버그를 없앤 건 좋은데 왜 쓸데없이 별포인트에서 엔진 파괴로 바뀌었나요? 이렇게 되면 난 **엔진** 파괴를 1~2개만 하고 계속 지면서 후진하라는 얘기인가요? 시즌 2부터는 원래대로 돌리고 공평한 **매칭**을 하게 해주세요!

어떤 개별 게임들이 유사한 게임일까?

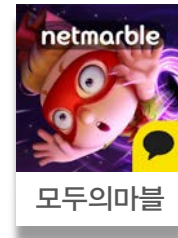
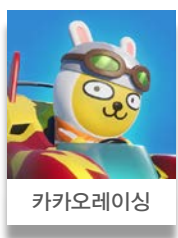
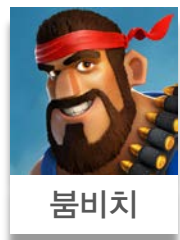
n(=100)명 이상 플레이한 254개 게임의 **플레이스토어 데이터 크롤링**
: (1) **게임 설명 텍스트** (2) **유저 리뷰 텍스트** (연관도 높은 순)

koNLPy 패키지의 Okt를 활용해 **각 텍스트의 명사를 추출**하고,
TF-IDF를 활용해 게임별 키워드 도출



공통적인 키워드를 많이 가진 게임들일수록 **유사성이 높을** 것이다.

A. 경쟁군 비교 분석 : WHO? (2) TF-IDF를 활용한 게임 리뷰 키워드 분석 및 유사성 비교



	0	1	2	game1	game2
41744	163	179	17	프렌즈레이싱	Cookie Run: OvenBreak
35085	137	150	15	Nebulous.io	서머너즈 워
12801	50	51	13	OX 퀴즈 서바이벌	월간아이돌
18996	74	126	13	애니팡 사천성	캔디크러쉬소다
45894	179	249	12	Cookie Run: OvenBreak	제5인격
41814	163	249	10	프렌즈레이싱	제5인격
25626	100	126	9	프렌즈레이싱	캔디크러쉬소다

같은 키워드가 존재하면 유사도 증가

유사성이 가장 높은 게임 쌍들의 목록

(프렌즈레이싱, 쿠키런): 아기자기한 캐릭터들의 [레이싱, 달리기] 게임
(애니팡 사천성, 캔디크러쉬소다) : 귀여운 느낌의 퍼즐게임

But 많은 결과쌍을 찾을 수는 없음

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

경쟁구도 분석: 카테고리

: 카테고리간의 관계 존재
[퍼즐, 캐주얼게임, 시뮬레이션, 카드]

: 완전하지 못한 카테고리 분류
ex) 게임정보앱-롤플레이팅
Helix Jump – 액션
배틀그라운드 – 액션

: 개별 경쟁사 탐색 불가

+

개별 경쟁사 분석: 키워드 분석

: 유사성 확인 가능

: 카테고리과 같은 경쟁 구도를 고려하지 못함

경쟁구도 하의 경쟁사 분석이 가능한 모델의 필요성

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

분석 목표

Point 1. 경쟁 구조의 파악

경쟁 구조를 파악한 후, 개별 경쟁사를 분석하는 것이 적절한 개인이 어떤 구조 하에서 게임을 선택하는가?
(ex. 카테고리, 제작사)

Point 2. 상대적 순위 도출

[절대적인] 이용 빈도 및 시간은 개인적 특성/외부요인에 종속적
이용시간 기준, 개인별로 경쟁 게임들 사이 [상대적인] 게임 순위 부여
(ex. 경쟁 게임이 7개라면 1~7위, 1위 = 기간 내 가장 많은 시간을 보낸 게임)

배속 로지스틱 회귀모형
(Nested Logistic Regression Model)

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

분석 모델 : 배속 로지스틱 회귀모형(Nested Logistic Regression Model)

다항 로지스틱 회귀모형(Multinomial Logistic Regression Model)

A model for multiple choice: 다중 클래스(multiclass)를 종속변수로 가짐

IIA (Independence of Irrelevant Alternatives) 가정:

특정 범주의 선택 여부가 다른 선택 가능한 범주들과 독립적이라고 가정
그러나 경쟁군들 사이에 **분명한 경쟁 구도가 존재**한다면,
같은 경쟁군에 속하는지 아닌지 여부가 게임 선택에 영향을 미칠 것 = IIA 가정이 깨짐

배속 로지스틱 회귀모형(Nested Logistic Regression Model)

공통된 특성에 따라(ex. 카테고리, 제작사) 경쟁군으로 묶임으로써

게임 선택이 **특정한 경쟁 구도** 하에서 이루어진다고 봄

IIA 가정을 각 **경쟁군 내부로 완화시켜 이분산성(heteroscedasticity)**을 설명,
즉 다항 로지스틱 회귀모형의 특수한 케이스

(1) **경쟁 구도의 적절성** 및 (2) **경쟁 강도**를 분석할 수 있음

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

분석 모델 : 배속 로지스틱 회귀모형(Nested Logistic Regression Model)

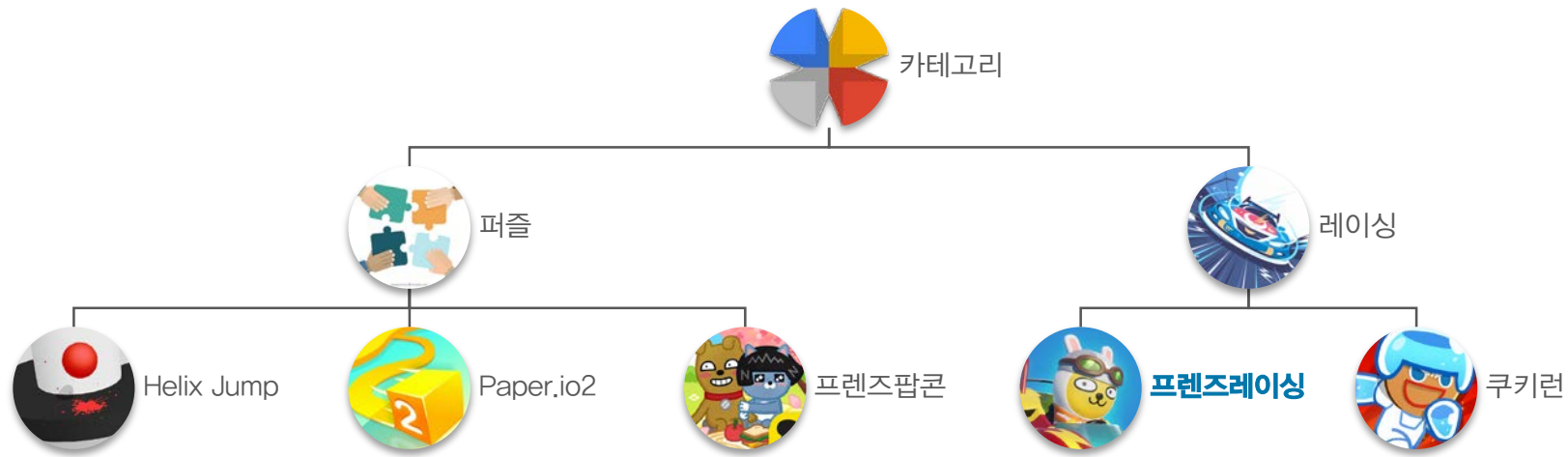
다항 로지스틱 회귀모형(Multinomial Logistic Regression Model)



$$P(\text{프렌즈레이싱}) = \sum_{i=1}^5 \frac{\text{프렌즈레이싱을 선택할 확률}}{i\text{번째 게임을 선택할 확률}}$$

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

분석 모델 : 배속 로지스틱 회귀모형(Nested Logistic Regression Model)

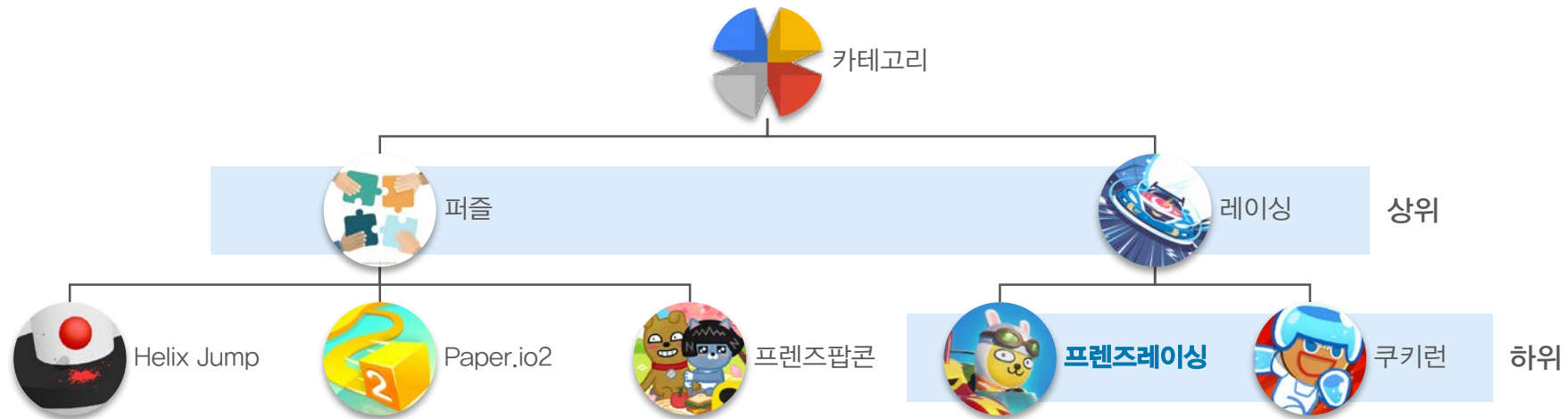


$$P(\text{프렌즈레이싱}) = P(\text{프렌즈레이싱}|\text{레이싱}) \times P(\text{레이싱})$$

$$Prob[twig_j, branch_l] = P_{jl} = P_{j|l} P_l$$

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

분석 모델 : 배속 로지스틱 회귀모형(Nested Logistic Regression Model)



$$I_l = \ln \sum_{j=1}^{J_l} e^{x'_{jl}\beta}$$

IV(Inclusive Value, 내포값)

상위 범주(Branch)별로 하위(내부, twig, 개별 범주) 경쟁이 더 치열한지,
혹은 상위(외부, branch) 경쟁이 더 치열한지 의미함
계수값 1을 기준으로 1보다 작으면 하위, 크면 상위 경쟁이 더 치열함

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

분석 모델 : 배속 로지스틱 회귀모형(Nested Logistic Regression Model)

$$P(\text{프렌즈팝콘}) = P(\text{프렌즈팝콘}|\text{퍼즐}) \times P(\text{퍼즐})$$

$$\text{Prob}[\text{twig}_j, \text{branch}_l] = P_{jl} = P_{j|l} P_l$$

$$P_{j|l} = \frac{e^{\mathbf{x}'_{j|l}\beta}}{\sum_{j=1}^{J_l} e^{\mathbf{x}'_{j|l}\beta}}$$

$$P_l = \frac{e^{\mathbf{z}'_l\gamma + \tau_l I_l}}{\sum_{l=1}^L e^{\mathbf{z}'_l\gamma + \tau_l I_l}}$$

$I_l = \ln \sum_{j=1}^{J_l} e^{\mathbf{x}'_{j|l}\beta}$ 는 각각의 상위 범주에 대해 추정

I_l 는 $P_{j|l}$ 의 분모에 자연로그를 취한 값과 동일,

곧 1번째 Nest를 선택함으로써 기대되는 이득 또는 효용(utility)으로 해석

이때 I_l 는 독립변수의 하나로 취급되며, 계수로서 τ_l 이 추정됨

$\sigma_j^2 = \frac{\pi}{\sigma \tau_j}$ 로써 경쟁군 내 이분산성(heteroscedasticity)을 설명

$\tau_l = 1$ 이면 다항 로지스틱 회귀모형으로 축소,

1보다 작으면 효용이 작으므로 개별 범주 내

경쟁관계 혹은 대체관계가 더 중요하다는 것을 알 수 있음

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

분석 모델 : 배속 로지스틱 회귀모형(Nested Logistic Regression Model)

모형 비교를 통해 어떤 경쟁 구조가 더 적절한지 분석



분석1: 어떤 경쟁 구조가 더 적절한가?

분석2: 어떤 게임이 나의 경쟁자인가?

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

분석 모델 : 배속 로지스틱 회귀모형(Nested Logistic Regression Model)

사용 변수

종속변수

게임사용순위 기반 게임 선택 여부

choice = 경쟁 게임 목록의 이용시간 기준 순위 생성 후 이분화
Ex) 게임이 5개라면 1위일 경우 TRUE, 2~5위일 경우 FALSE

독립변수

1) 범주특성 변수 (2) 2) 개체특성 변수 (5)

1) 범주특성 변수(2)

cnt = 1~17주차 **이용빈도**

cnt_playtime = 1~17주차 이용빈도 x 1/횟수별 평균 플레이 시간

평균 플레이시간이 짧은 게임에게 더 큰 가중치를 부여하여 게임별 차이 반영

2) 개체특성 변수(5)

개체특성 변수만으로 게임선택을 설명하기는 어려움

패널 변수를 통제변수로 활용: 예상하지 못한 영향력 통제

Ex. 모바일기기 이용시간이 길수록 특정 게임을 선호할 가능성?

변수 선택 => 랜덤 포레스트 모형을 이용해 가장 중요한 패널 변수 5개 선정

모델 선정 이유

모델: 이용빈도로 (이용시간 기준) 게임선택 설명,
패널 특성은 통제변수로 활용

의문: 이용빈도가 높으면 총 이용시간이 많은 건 당연한 것?
→ True

그러나 관심사는 **계수가 아닌 [경쟁구조]**

선정 이유:

논리적으로 타당한 모델을 이용해,

조건부확률에 기반을 둔 경쟁구조를 설명하려고 하는 것

cf) 패널을 통한 예측의 실패 (ROC AUC 기준)

Logistic Regression (0.398)

Random Forest (0.499)

XGBoost (0.5)




Soft Voter Ensemble (0.499)

패널 변수만으로는 게임선택 예측불가

Epdel



 Competitive Analysis <

>> Choose

>> Describe

>> Analyze

↔ Switching Analysis

Choose Your Game

===== Reference Game ===== ▾

Choose Your Competitors

Nest 1

- ☐ Helix Jump
- ☐ Paper.io 2
- ☐ 프렌즈팝콘
- ☐ 프렌즈레이싱
- ☐ Cookie Run: OvenBreak
- ☐ 배틀그라운드
- ☐ 클래시 로얄

Nest 2

- ☐ Helix Jump
- ☐ Paper.io 2
- ☐ 프렌즈팝콘
- ☐ 프렌즈레이싱
- ☐ Cookie Run: OvenBreak
- ☐ 배틀그라운드
- ☐ 클래시 로얄

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

Preprocessing : Panel 데이터 전처리 및 변수 생성

Panel 데이터

115,854명의 패널
15개 카테고리
700여 개의 binary 문항 → 패널의 특성 정량화 필요성 ↑

모바일주이용) 지난 1주일 간 게임 이용 여부	1	} mobile_digit = 2
모바일주이용) 지난 1주일 간 뉴스기사 이용 여부	0	
모바일주이용) 지난 1주일 간 동영상 이용 여부	1	



모바일게임) 지난 1년 간 RPG/MMORPG 경험 여부	1	} game_active = 3
모바일게임) 지난 1년 간 MOBA/RTS/AOS 경험 여부	1	
모바일게임) 지난 1년 간 슈팅/FPS/TPS 경험 여부	1	

전처리

1. Aggregation: 카테고리별/내 동일한 속성을 가진 문항들 통합
housework, m_cosmetic, f_cosmetic, m_fashion, f_fashion, eat_out, alcohol, health, parent_educ, stud_educ, nonstud_educ, shop, pc_shop, mobile_shop, pc_digit, mobile_digit, game_active, leisure_active, finance (19)
2. 제공군 변환: 변수 스케일 조정
통합된 19개 연속형 변수들에 대해 제공군 변환
기계적 결측치를 0으로 처리하였기에 정보 손실 방지 (ex. 남성은 여성 질문 응답 불가)
3. Age 변수 제공항 추가: 비선형적 연령 효과 고려
4. 기타 Binary, Ordinal, Categorical 변수 포함 도합 32개
Binary: female(X0001), city(X0006), married(X0007), smoke(F0057) (4)
Ordinal: age(X0002), age2, educ(X0008), family(Y0001), income(Y0008), morning(E0001), commute(H0011) (7)
Categorical: job, tv_media (2)

완성된 최종 변수 32개

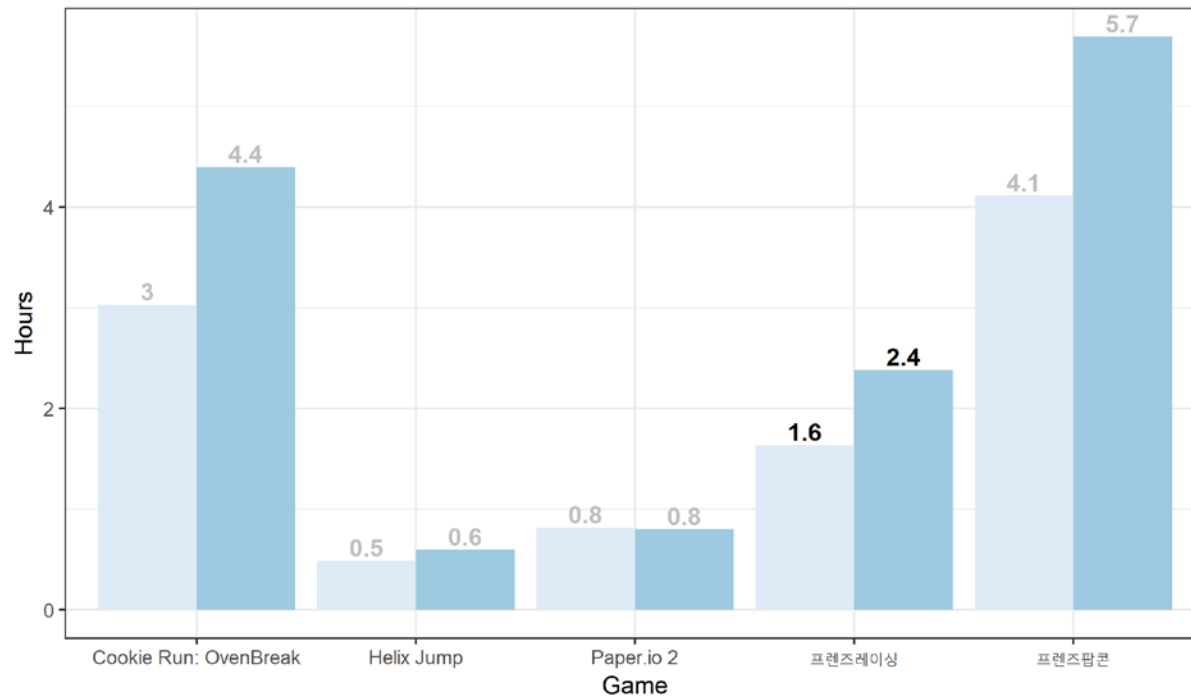
'panel_id', 'female', 'age', 'age2', 'city', 'married', 'educ',
'job', 'family', 'wage', 'housework', 'm_cosmetic', 'f_cosmetic', 'm_fashion',
'f_fashion', 'morning', 'eat_out', 'alcohol', 'health', 'smoke', 'parent_educ', 'std_educ',
'nonstud_educ', 'commute', 'shop', 'pc_shop', 'mobile_shop', 'pc_digit', 'mobile_digit',
'game_activ', 'leisure_activ', 'tv_media', 'finance'

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

EDA: 게임별 패널별/횟수별 평균 이용시간

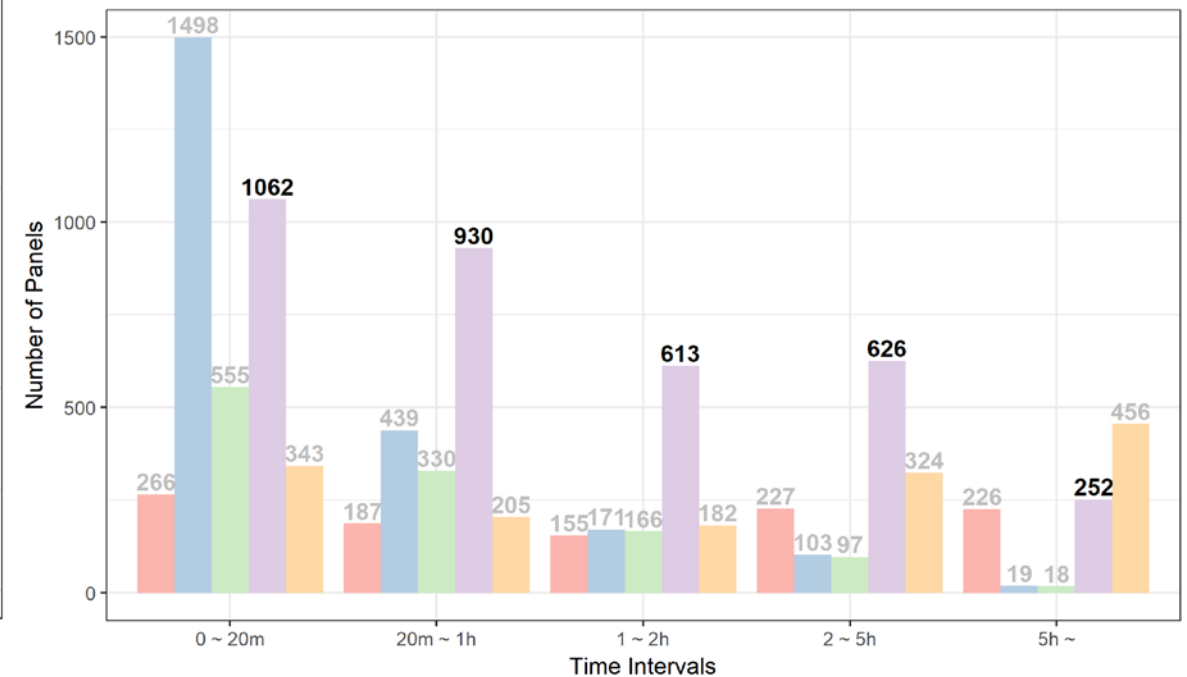
Average Weekly Play Hours

Per Panel Play



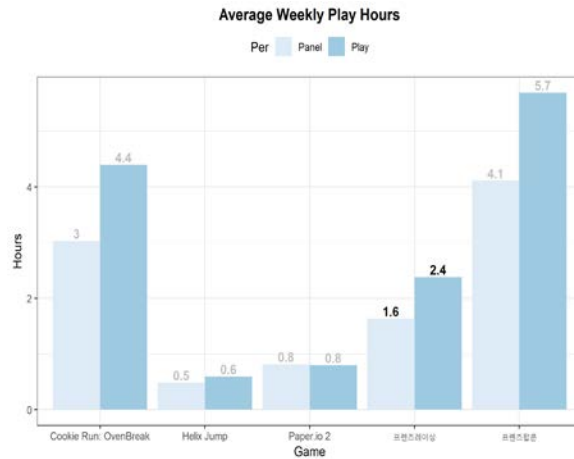
Number of Panels in Each Time Interval

Game Cookie Run: OvenBreak Helix Jump Paper.io 2 프렌즈레이싱 프렌즈팝콘



A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

EDA: 게임별 패널별/횟수별 평균 이용시간



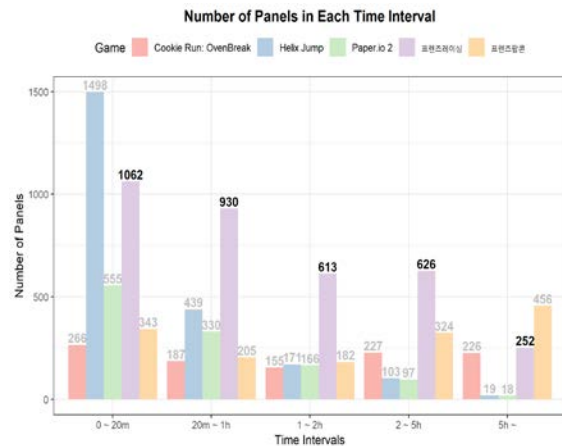
패널별 평균 이용시간 < 횟수별 평균 이용시간

Figure 1.

Helix Jump, Paper.io2는 [횟수별 평균 이용시간/패널별 평균 이용시간]의 크기가 큰 편
Figure 2.

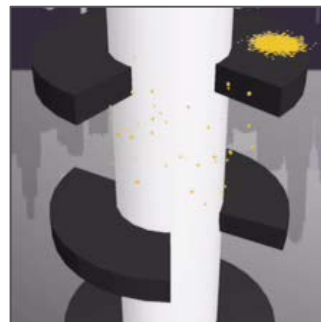
Helix Jump와 Paper.io2는 오른쪽 꼬리가 긴 분포를 보임

특히 Helix Jump는 20분 이내로 이용시간이 매우 짧은 경우가 대다수
[모델의 가중치: 횟수별 평균 이용시간의 역수]

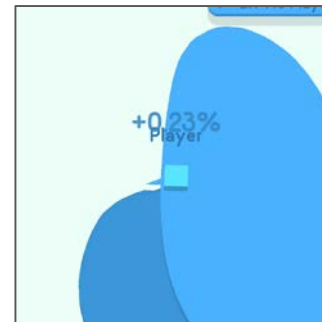


게임 이용시간이 극단적으로 긴 충성고객의 수가 매우 적음

즉, 게임에 집중 혹은 중독될 만한 요소(ex. 캐릭터수집, 유저들간 커뮤니케이션 등)가 부족



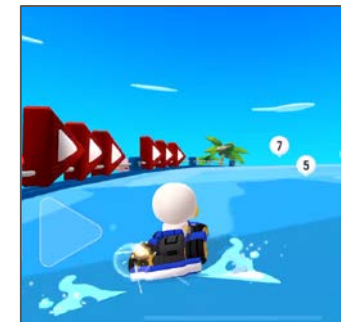
Helix Jump



Paper.io2



Cookie Run



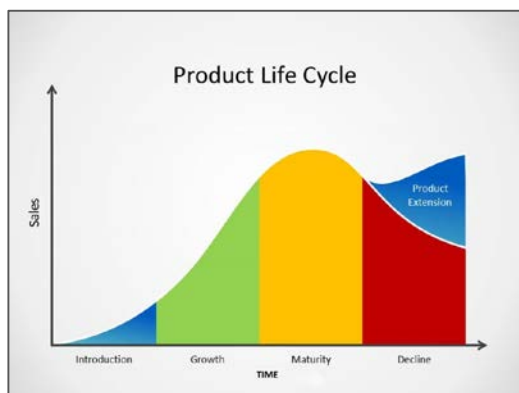
프렌즈 레이싱

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 - 비교군 내/외부의 경쟁성 비교

EDA: 게임별 패널별/횃수별 평균 이용시간

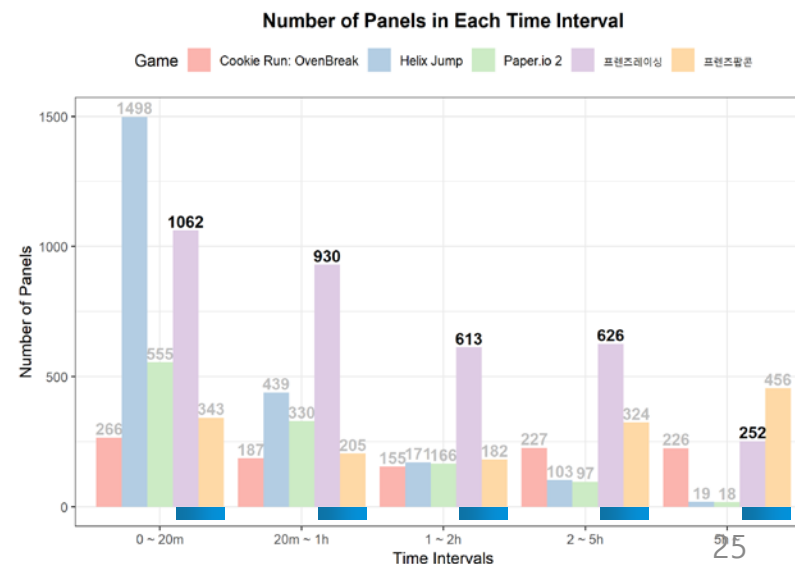
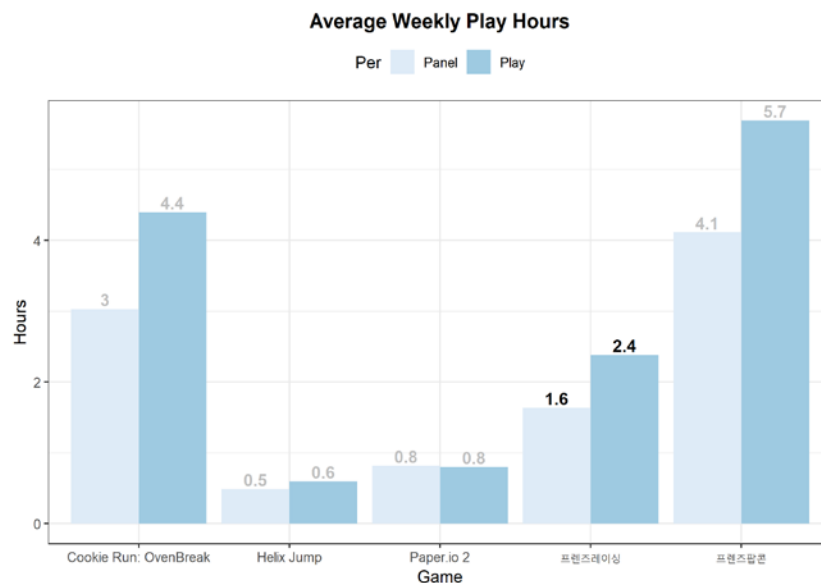
[Insight: 프렌즈레이싱]

패널 및 횃수별 평균 이용시간은 프렌즈팝콘과 쿠키런보다 낮으나, 이용빈도는 더 높음
즉 프렌즈팝콘, 쿠키런은 소수의 충성고객으로 생존 중 = 성숙기 진입



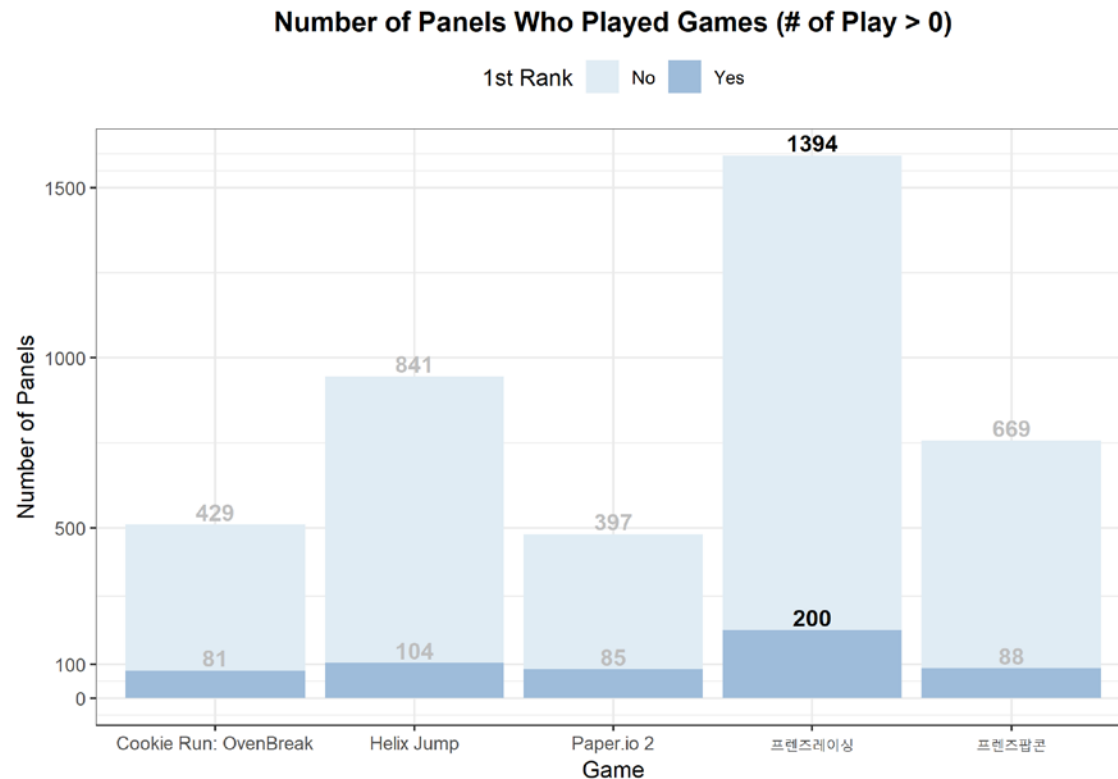
제품수명주기이론

[프렌즈레이싱: 2018년 10월 출시]
[프렌즈팝콘: 2016년 10월 출시]
[쿠키런(OvenBreak): 2016년 10월 출시]



A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

EDA: 게임별 플레이한 패널 수 / 1순위 선정 패널 수



[Insight]

해당 게임에 가장 많은 시간을 보낸 사람의 수는 한 번이라도 게임을 해본 사람들의 수와 관계없이 비슷함

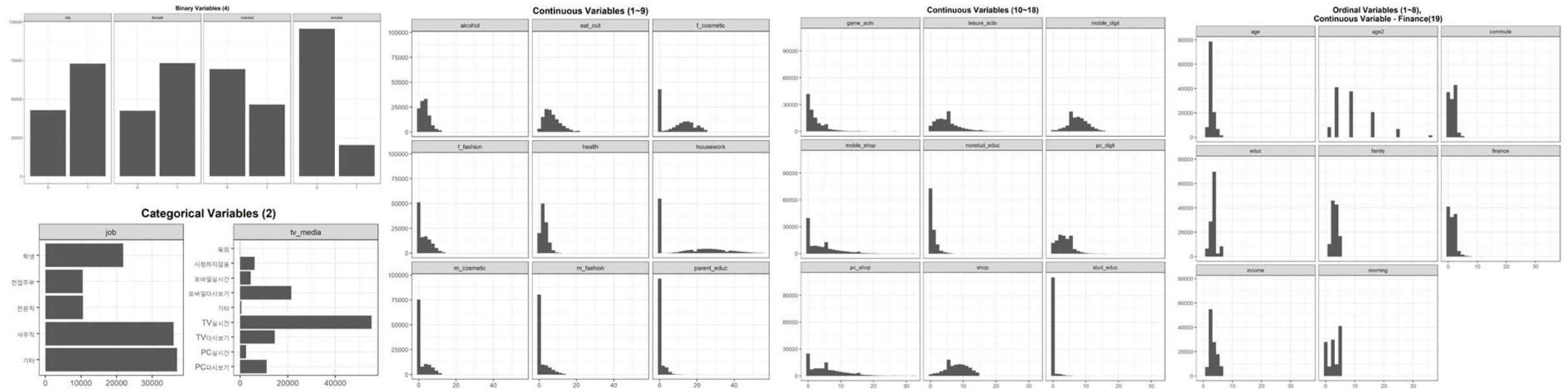
[프렌즈레이싱]

프렌즈레이싱은 한 번 이상 게임을 해본 사람들의 수가 1394명으로 압도적으로 많으므로, 1위 선택 역시 200명으로 많음

Yes로 표시된 게임이 본 모델의 패널별 종속변수의 값으로 사용될 것

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

EDA: 패널 변수 소개



A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

변수 선택: 랜덤 포레스트 모델을 이용한 중요 패널 변수 5개 선정

사용 변수

종속변수

게임사용순위 기반 게임 선택 여부

choice = 경쟁 게임 목록의 이용시간 기준 순위 생성 후 이분화
Ex) 게임이 5개라면 1위일 경우 TRUE, 2~5위일 경우 FALSE

독립변수

1) 범주특성 변수(2) 2) 개체특성 변수 (32)

게임 선택에 직접적인 영향을 미치는 cnt, cnt_playtime (2) 제외
미디어 이용 변수와 상관성이 높은 범주형 변수 tv_media 제외

지니 불순도(Gini Impurity)

CART(classification and regression tree) 알고리즘에서 최적의 분류 트리를 찾기 위해 사용되는 측정기준(metric) 중 하나

$$I_G(t) = \sum_{i=1}^c p(i|t)(1 - p(i|t))$$

모형 선정 이유

Tree based 모형은 기준(지니불순도, 엔트로피, 분류오차 등)에 따른 변수의 중요성(Feature Importance) 제공

모형 추정이 안정적이고, 빠르고, 일정 수준 이상의 정확도를 보장하며, 과적합 문제가 적음

I_G 는 클래스가 균일하게 섞여있을 때 최대가 되므로, 이를 최소화함으로써 서로 다른 클래스를 구분하고 순도(homogeneity)를 높이고자 함

지니 불순도를 크게 감소시킨 순으로 변수의 중요도 추출

패널 특성만으로는 게임 선택을 예측하는 데 한계가 있음을 인지
그러나 주어진 정보를 활용하고, 예상하지 못한 영향력 통제한다는 점에서 패널 특성 변수를 고정된 숫자인 5개로 선택한 것

뿐만 아니라 변수의 효과를 해석하는 측면에서 이점을 줄 것

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

최종 선택 변수 : 랜덤 포레스트 모델을 이용한 중요 패널 변수 5개 선정

PANEL_ID	female	age	age2	city	...	game_activ	leisure_activ	finance	choice
H0038172	0	5	25	1	...	1.73	2.83	0	g1
H0064982	0	1	1	0	...	3.16	2	1	g1
H0090553	0	4	16	0	...	0	1.41	1	g1
H0100978	0	2	4	1	...	2	2.45	1.41	g1
H0101130	1	1	1	0	...	3.46	3.74	1	g1
H0105249	0	4	16	0	...	2	1	0	g1
H0111769	1	1	1	0	...	1.41	1.73	0	g1
H0147298	1	1	1	1	...	0	2.45	0	g1
H0162412	1	3	9	1	...	0	2.65	1	g1

$choice \sim cnt + cnt_playtime |$
 $eat_out + mobile_digit + shop + game_active + leisure_active$

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

최종 선택 변수 : 랜덤 포레스트 모델을 이용한 중요 패널 변수 5개 선정

Predicted True	g1	g2	g3	g4	g5
g1	18	9	1	8	1
g2	3	17	0	3	2
g3	0	2	27	9	1
g4	1	0	1	57	2
g5	1	0	0	0	23

Train(66.6%) Test로 데이터셋을 나눈 후
랜덤 포레스트 학습 (좌)
OOB(Out-of-Bag) error = $1 - 142/186 = 23.7\%$

Predicted True	g1	g2	g3	g4	g5
g1	43	23	3	30	5
g2	13	56	1	13	2
g3	1	3	67	14	3
g4	7	3	12	170	8
g5	4	1	2	9	65

모든 데이터를 학습하여 모델을 추정한 결과 (우)
OOB error = $1 - 403/558 = 28.1\%$

우측 모형에 기반을 두어 변수 중요도를 추출

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

최종 선택 변수 : 랜덤 포레스트 모델을 이용한 중요 패널 변수 5개 선정

PANEL_ID	eat_out	mobile_digit	shop	game_activ	leisure_active	cnt	cnt_playtime	choice	alt	chid
H0038172	1.73	1.41	2.45	1.73	2.83	1	0.30	TRUE	g1	1
H0038172	1.73	1.41	2.45	1.73	2.83	0	0	FALSE	g2	1
H0038172	1.73	1.41	2.45	1.73	2.83	1	0.23	FALSE	g3	1
H0038172	1.73	1.41	2.45	1.73	2.83	0	0	FALSE	g4	1
H0038172	1.73	1.41	2.45	1.73	2.83	0	0	FALSE	g5	1
H0064982	1.73	3	1.73	3.16	2	2	0.60	TRUE	g1	2
H0064982	1.73	3	1.73	3.16	2	0	0	FALSE	g2	2
H0064982	1.73	3	1.73	3.16	2	0	0	FALSE	g3	2
H0064982	1.73	3	1.73	3.16	2	0	0	FALSE	g4	2

$choice \sim cnt + cnt_playtime |$
 $eat_out + mobile_digit + shop + game_active + leisure_active$

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

분석 모델 : 배속 로지스틱 회귀모형(Nested Logistic Regression Model)

분석 결과 1 모형 비교 결과, 카테고리보다 유통사 기준 분류가 더 적합



```
> lmttest::lrtest(mn1, mn2)
```

Likelihood ratio test

Model 1: choice ~ cnt + cnt_playtime | eat_out + mobile_digit + shop + game_activ + leisure_activ

Model 2: choice ~ cnt + cnt_playtime | eat_out + mobile_digit + game_activ + shop + leisure_activ

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	28	-327.83			
2	28	-319.73	0	16.193	< 2.2e-16 ***

Likelihood ratio test 결과,
주어진 데이터에는 유통사 기준 분류가 카테고리보다 적합
즉 카테고리보다는 유통사:카카오 여부에 따른
경쟁 구조 하에서 게임 선택이 이루어짐

카카오가 플레이스토어 외에도 배타적인 유통 채널을 가진 점,
프로모션이나 이모티콘, 소셜 미디어 연결 등
타 게임들과 질적으로 다른 마케팅 서비스를 제공한다는 점에서
이 분류는 상식과 부합

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

분석 모델 : 배속 로지스틱 회귀모형(Nested Logistic Regression Model)

분석 결과 2 프렌즈 레이싱 vs 프렌즈 팝콘



Coefficients :

	Estimate	Std. Error	z-value	Pr(> z)	
iv:nest1	2.227378	0.317512	7.0151	2.298e-12	***
iv:nest2	0.720333	0.096716	7.4479	9.481e-14	***

```
> nest_ivtest(mn2, 2)
[1] "iv:nest1 = 2.2274 (***)"
[1] "IV >= 1, so branches are more competitive."
[1] "iv:nest2 = 0.7203 (**)"
[1] "IV < 1, so twigs are more competitive."
```

H0: IV = 1 vs. H1: IV > 1 (or IV < 1)

R에서 제공하는 함수는 IV = 0 여부를 테스트하므로 새로운 함수 생성
 $\alpha=0.05$ 기준으로 양측검정 후 통계적 유의성 확인

경쟁군 1(카카오): IV가 1보다 유의미하게 크므로 하위 경쟁 < 상위 경쟁
경쟁군 2(비카카오): IV가 1보다 유의미하게 작으므로 하위 경쟁 > 상위 경쟁

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

분석 모델 : 배속 로지스틱 회귀모형(Nested Logistic Regression Model)

분석 결과 2

프렌즈 레이싱 vs 프렌즈 팝콘 : 패널 특성 비교

Coefficients :				
	Estimate	Std. Error	z-value	Pr(> z)
g1:(intercept)	1.013126	1.134690	0.8929	0.371929
g2:(intercept)	1.109724	1.208086	0.9186	0.358315
g3:(intercept)	-2.111297	3.770895	-0.5599	0.575553
g5:(intercept)	0.090309	1.163797	0.0776	0.938147
cnt	5.438750	0.538001	10.1092	< 2.2e-16 ***
cnt_playtime	-16.007285	1.735299	-9.2245	< 2.2e-16 ***
g1:eat_out	-0.379558	0.261191	-1.4532	0.146174
g2:eat_out	-0.137871	0.286866	-0.4806	0.630794
g3:eat_out	-1.803472	0.757608	-2.3805	0.017290 *
g5:eat_out	0.364187	0.346535	1.0509	0.293287
g1:mobile_digit	-0.173425	0.365970	-0.4739	0.635588
g2:mobile_digit	-0.223002	0.390845	-0.5706	0.568296
g3:mobile_digit	-0.341171	1.055993	-0.3231	0.746634
g5:mobile_digit	-0.320209	0.338712	-0.9454	0.344469
g1:game_activ	0.217514	0.190510	1.1417	0.253559
g2:game_activ	0.228920	0.205895	1.1118	0.266210
g3:game_activ	-0.060010	0.507296	-0.1183	0.905836
g5:game_activ	0.157468	0.250105	0.6296	0.528952
g1:shop	0.262502	0.319839	0.8207	0.411799
g2:shop	0.322081	0.341601	0.9429	0.345754
g3:shop	2.416377	0.922628	2.6190	0.008818 **
g5:shop	-0.298577	0.444963	-0.6710	0.502210
g1:leisure_activ	-0.153860	0.282511	-0.5446	0.586018
g2:leisure_activ	-0.397533	0.310593	-1.2799	0.200574
g3:leisure_activ	-1.067455	0.650889	-1.6400	0.101006
g5:leisure_activ	0.020925	0.356372	0.0587	0.953178
iv:nest1	2.227378	0.317512	7.0151	2.298e-12 ***
iv:nest2	0.720333	0.096716	7.4479	9.481e-14 ***

g1: Helix Jump
g2: Paper io.2
g3: 프렌즈팝콘
g4: 프렌즈레이싱
g5: 쿠키런

패널 특성(외식)에 따라 기준 범주인 프렌즈레이싱(g4)에 비해 프렌즈팝콘(g3)를 선택할 확률이 유의미하게 낮아짐

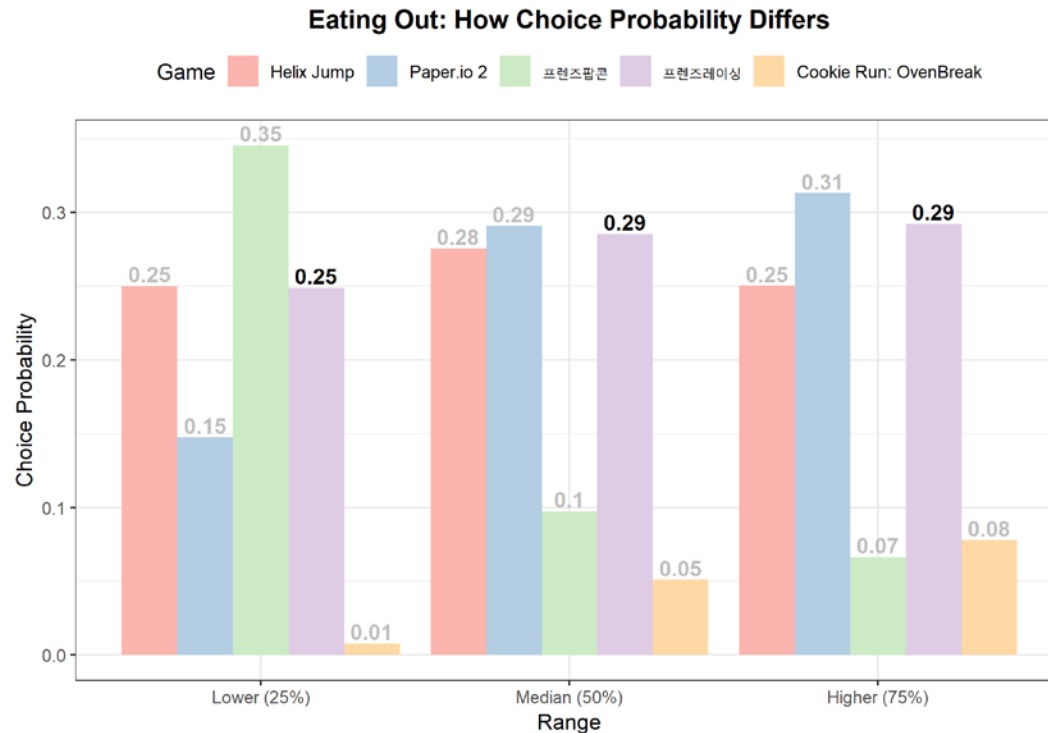
패널 특성(쇼핑)에 따라 기준 범주인 프렌즈레이싱(g4)에 비해 프렌즈팝콘(g3)를 선택할 확률이 유의미하게 높아짐

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

분석 모델 : 배속 로지스틱 회귀모형(Nested Logistic Regression Model)

분석 결과 2

프렌즈 레이싱 vs 프렌즈 팝콘 : 패널 특성 비교 – 외식



패널 특성(외식) vs. 게임 선택

외식을 자주 하는 사람일수록
프렌즈레이싱 선택확률(보라)은 **증가**하지만
프렌즈팝콘(초록)은 **감소**

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

분석 모델 : 배속 로지스틱 회귀모형(Nested Logistic Regression Model)

분석 결과 2

프렌즈 레이싱 vs 프렌즈 팝콘 : 패널 특성 비교 – 쇼핑



패널 특성(쇼핑) vs. 게임 선택

쇼핑을 자주 하는 사람일수록
프렌즈레이싱 선택확률(보라)은 감소하지만,
프렌즈팝콘(초록)은 증가

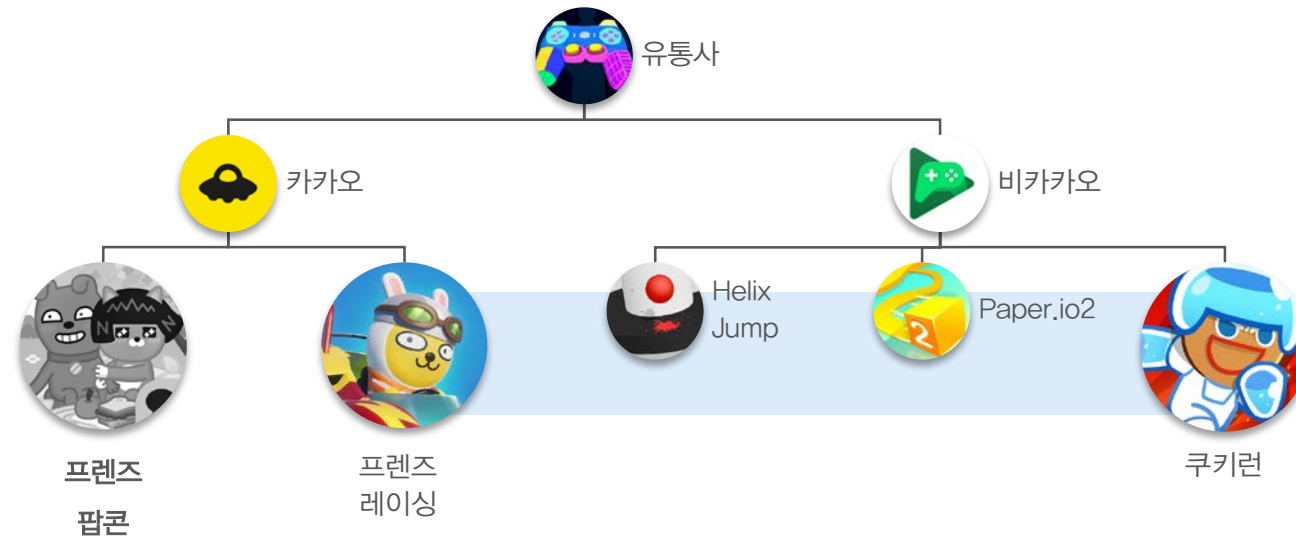
앞에 비해 증감폭이 크지 않음

A. 경쟁군 비교 분석 : WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

분석 모델 : 배속 로지스틱 회귀모형(Nested Logistic Regression Model)

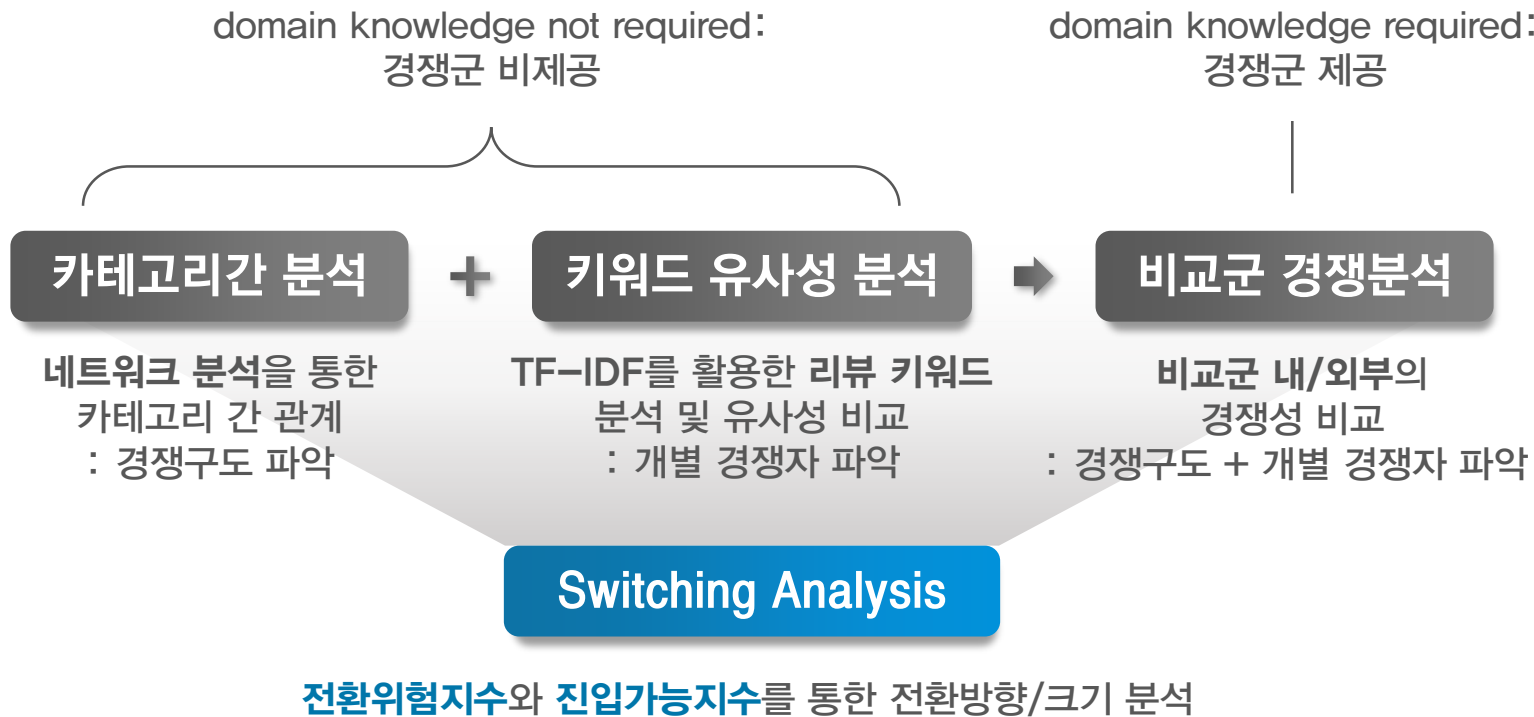
분석 결과 2

프렌즈 레이싱 vs 프렌즈 팝콘 : 경쟁구도 분석



프렌즈레이싱의 관점:
같은 카카오 게임인 **프렌즈팝콘에 비해**
쿠키런 (비카카오 게임)와의 경쟁이 더 치열

A. 경쟁군 비교 분석 : HOW? 전환 분석- 전환위험지수와 진입가능지수를 통한 전환방향/크기 분석



Who?
누가 경쟁자인가?

How?
경쟁 방향은 어떠한가?
경쟁 강도는 어느 정도인가?

A. 경쟁군 비교 분석 : HOW? 전환 분석- 전환위험지수와 진입가능지수를 통한 전환방향/크기 분석

전환 분석 목표

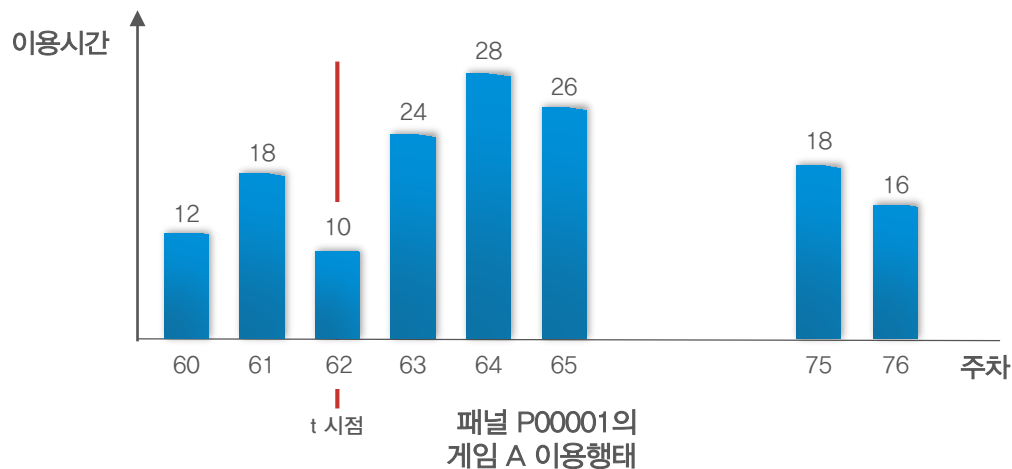
HOW? 경쟁의 방향은 어떠한가? 경쟁의 강도는 어느 정도인가?

경쟁 게임 간 유저 전환 수를 계산하여 경쟁사 간 비교우위/비교열위 파악

경쟁구도별 유저 전환 비율을 계산하여 더 치열한 경쟁구도 확인

전환 분석 방법

key idea: 유저별 A게임의 이용행태와 B게임의 이용행태를 복합적으로 고려하여 A↔B 전환 여부를 파악



: 해당 주차 직전까지 자신의 평균 이용시간과 해당 주차의 이용시간의 비율을 통해 게임전환위험도 및 게임진입가능도 계산

$$\text{전환위험지수}_t = \frac{\frac{\sum_{1}^{t-1} \text{이용시간}}{t-1}}{\text{이용시간}_t} = \frac{\left(\frac{12+18}{2}\right)}{10} = \frac{3}{2}$$

$$\text{진입가능지수}_t = \frac{1}{\text{전환위험지수}_t} = \frac{\text{이용시간}_t}{\frac{\sum_{1}^{t-1} \text{이용시간}}{t-1}} = \frac{10}{\left(\frac{12+18}{2}\right)} = \frac{2}{3}$$

1. 유저 개인별 이용행태를 반영하여 전환/진입을 수치화할 수 있다
2. 비슷한 시기에 A게임 전환지수 & B게임 진입지수가 동시에 가파르게 상승한다면 A→B로 전환된 것이라고 유추할 수 있다.
3. 실제 이용시간 추이를 통해 검증할 수 있다.
4. Longitudinal 데이터 분석: 시간의 흐름에 따른 이용률 변화 관찰

A. 경쟁군 비교 분석 : HOW? 전환 분석- 전환위험지수와 진입가능지수를 통한 전환방향/크기 분석

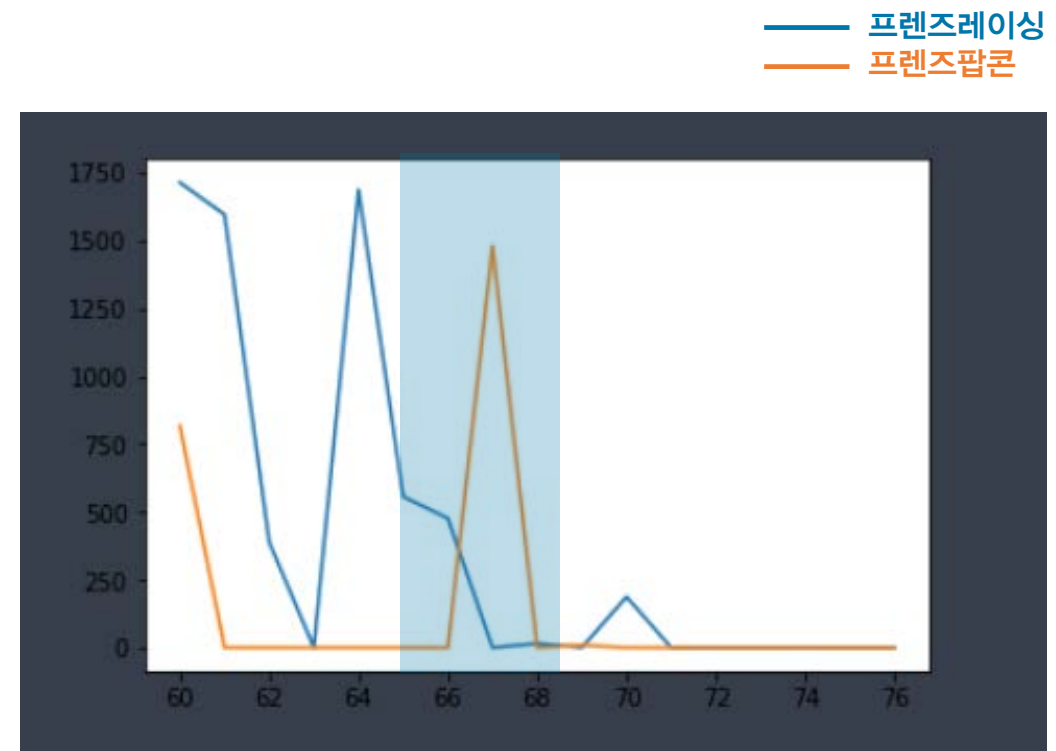
전환 사례



유저의 주차별 게임 전환위험지수와 진입가능지수

프렌즈레이싱에서 **전환할 가능성이 높은 동시에**
프렌즈팝콘으로 **진입할 가능성이 높은** 시점

프렌즈레이싱 → 프렌즈팝콘으로 **유저가 전환한 것**으로 판단



유저의 주차별 게임 이용시간

실제 이 유저의 게임별 이용시간을 살펴보면
67주차에 **프렌즈레이싱은 이용시간이 줄어든 반면**
프렌즈팝콘의 이용시간은 급격히 상승했다

A. 경쟁군 비교 분석 : HOW? 전환 분석- 전환위험지수와 진입가능지수를 통한 전환방향/크기 분석

전환 분석 방법

1. 경쟁구도를 확인하고자 하는 **두 게임을 선택** (프렌즈레이싱 vs 프렌즈팝콘)
2. **게임을 아예 안한 주차**에 대해선 임의로 1초를 부여 (전환위험지수와 진입가능지수를 계산하기 위함)
3. 게임별 주차별 **전환위험지수 및 진입가능지수** 계산
4. A 게임의 주차별 전환위험지수와 B 게임의 주차별 진입가능지수를 **추이 그래프로 표현**
5. A 전환위험지수 & B 진입가능지수가 **동시에 상승하는 구간**을 찾음
 1. **두 지수의 기울기가 모두 1 이상인 모든 구간**을 찾는다
 2. 선택된 구간들 중 **기울기의 차이가 가장 작은 구간**을 택한다
 3. 선택된 구간의 두 기울기 비가 **1:3 이하**면 해당 시점에 **전환이 일어났다고 판단**한다.
6. A게임과 B게임 간 전환 유저의 수를 확인

A. 경쟁군 비교 분석 : HOW? 전환 분석- 전환위험지수와 진입가능지수를 통한 전환방향/크기 분석

전환 분석 방법

1. 경쟁구도를 확인하고자 하는 두 게임을 선택
(프렌즈레이싱 vs 프렌즈팝콘)

	panel_id	reg_cnt	app_name_x	total_time_stamp
28733137	P0031278	60	프렌즈레이싱	2592
28734052	P0031278	61	프렌즈레이싱	4359
28735833	P0031278	62	프렌즈레이싱	1452
28736238	P0031278	63	프렌즈레이싱	1338
28740081	P0031278	66	프렌즈레이싱	116
25573484	P0031278	61	프렌즈팝콘	16145
25574503	P0031278	62	프렌즈팝콘	22747
25574752	P0031278	63	프렌즈팝콘	9415
25575744	P0031278	64	프렌즈팝콘	1141
25577122	P0031278	66	프렌즈팝콘	4169
25585051	P0031278	70	프렌즈팝콘	2892
25587111	P0031278	71	프렌즈팝콘	2892

2. 게임을 아예 안한 주차에 대해선 임의로 1초를 부여
(이후 전환위험지수와 진입가능지수를 계산하기 위함)

	panel_id	reg_cnt	app_name_x	total_time_stamp
0	P0031278	60	프렌즈레이싱	2592
1	P0031278	61	프렌즈레이싱	4359
2	P0031278	62	프렌즈레이싱	1452
3	P0031278	63	프렌즈레이싱	1338
5	P0031278	64	프렌즈레이싱	1
6	P0031278	65	프렌즈레이싱	1
4	P0031278	66	프렌즈레이싱	116
7	P0031278	67	프렌즈레이싱	1
8	P0031278	68	프렌즈레이싱	1

A. 경쟁군 비교 분석 : HOW? 전환 분석- 전환위험지수와 진입가능지수를 통한 전환방향/크기 분석

전환 분석 방법

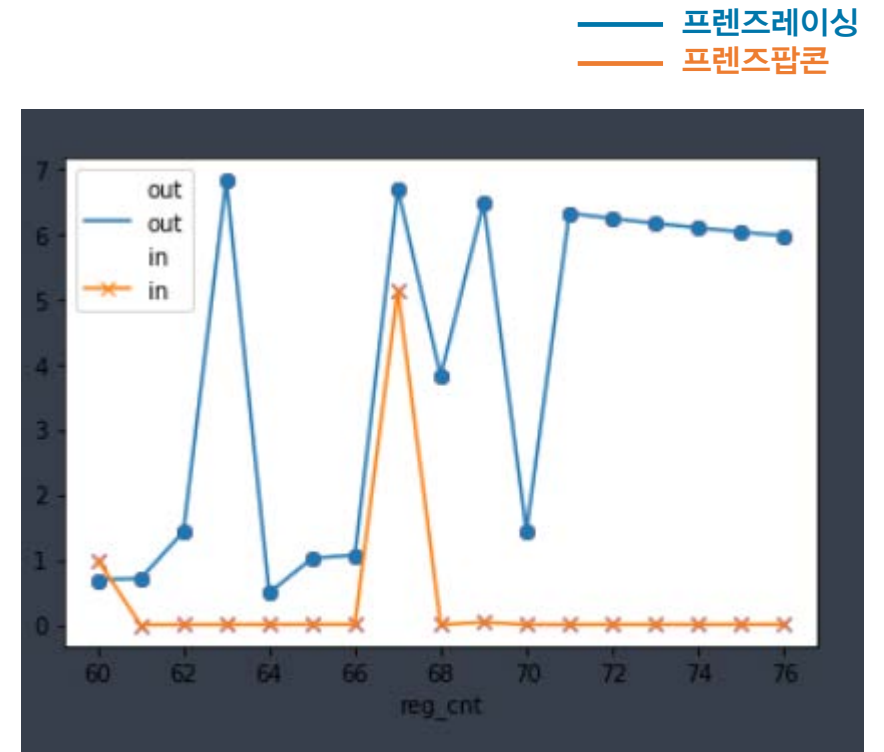
3. 전환위험지수 및 진입가능지수 계산

	panel_id	reg_cnt	app_name_x	total_time_stamp	out	in
0	P0031278	60	프렌즈레이싱	2592	0.693147	1.000000
1	P0031278	61	프렌즈레이싱	4359	0.586294	1.254208
2	P0031278	62	프렌즈레이싱	1452	1.074683	0.518386
3	P0031278	63	프렌즈레이싱	1338	1.036761	0.549430
5	P0031278	64	프렌즈레이싱	1	7.575277	0.000513
6	P0031278	65	프렌즈레이싱	1	7.393161	0.000616
4	P0031278	66	프렌즈레이싱	116	2.575785	0.082361
7	P0031278	67	프렌즈레이싱	1	7.117611	0.000811
8	P0031278	68	프렌즈레이싱	1	7.000031	0.000913

$$\text{전환위험지수}_t = \frac{\frac{\sum_1^{t-1} \text{이용시간}}{t-1}}{\text{이용시간}_t} = \frac{\left(\frac{12+18}{2}\right)}{10} = \frac{3}{2}$$

$$\text{진입가능지수}_t = \frac{1}{\text{전환위험지수}_t} = \frac{\text{이용시간}_t}{\frac{\sum_1^{t-1} \text{이용시간}}{t-1}} = \frac{10}{\left(\frac{12+18}{2}\right)} = \frac{2}{3}$$

4. A 게임의 주차별 전환위험지수와 B 게임의 주차별 진입가능지수를 추이 그래프로 표현



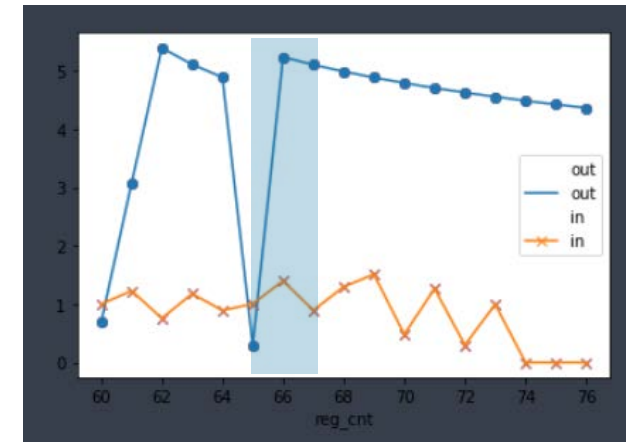
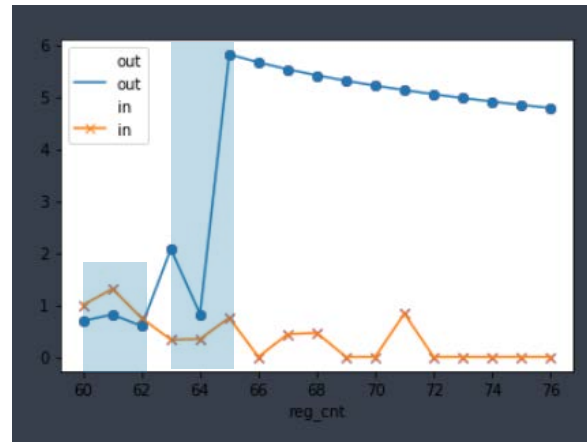
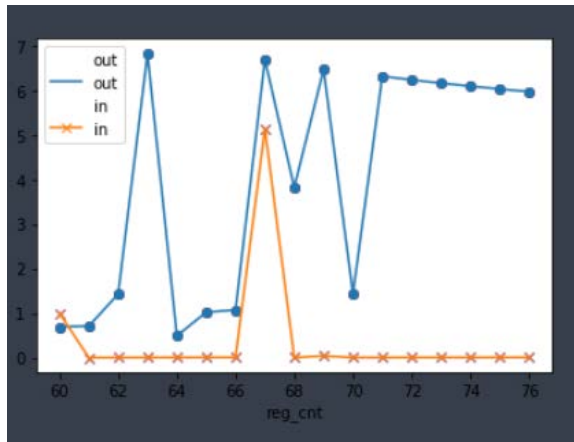
유저의 주차별 게임 전환위험지수와 진입가능지수

A. 경쟁군 비교 분석 : HOW? 전환 분석- 전환위험지수와 진입가능지수를 통한 전환방향/크기 분석

전환 분석 방법

5. A의 전환위험지수 & B의 진입가능지수가 동시에 상승하는 구간을 찾음

1. 두 기울기가 모두 1 이상인 구간들을 찾는다
2. 선택된 구간들 중 기울기의 차이가 가장 작은 구간을 택한다
3. 선택된 구간의 두 기울기 비가 1:3 이하면 해당 시점에 전환이 일어났다고 판단한다.

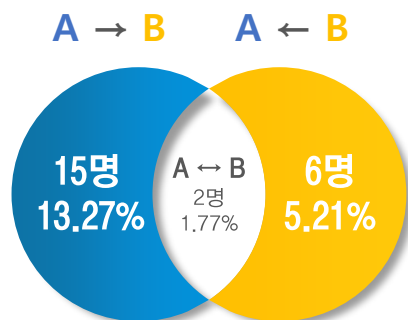
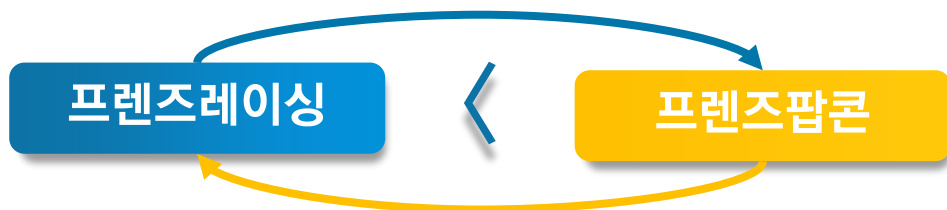


A. 경쟁군 비교 분석 : HOW? 전환 분석- 전환위험지수와 진입가능지수를 통한 전환방향/크기 분석

전환 분석 방법

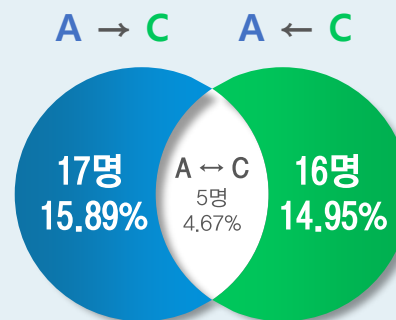
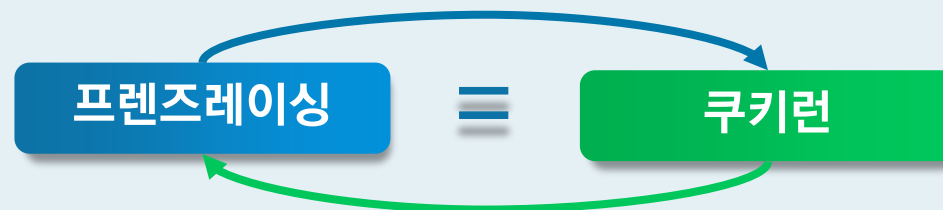
6. 게임 A와 게임 B간 전환 유저의 수를 확인

상호 전환 비율이 더 높은 경쟁구도가 더 치열한 경쟁구도이다



113명

$$(15 + 6 - 2) / 113 = 16.81\%$$



107명

$$(17 + 16 - 5) / 107 = 26.17\%$$

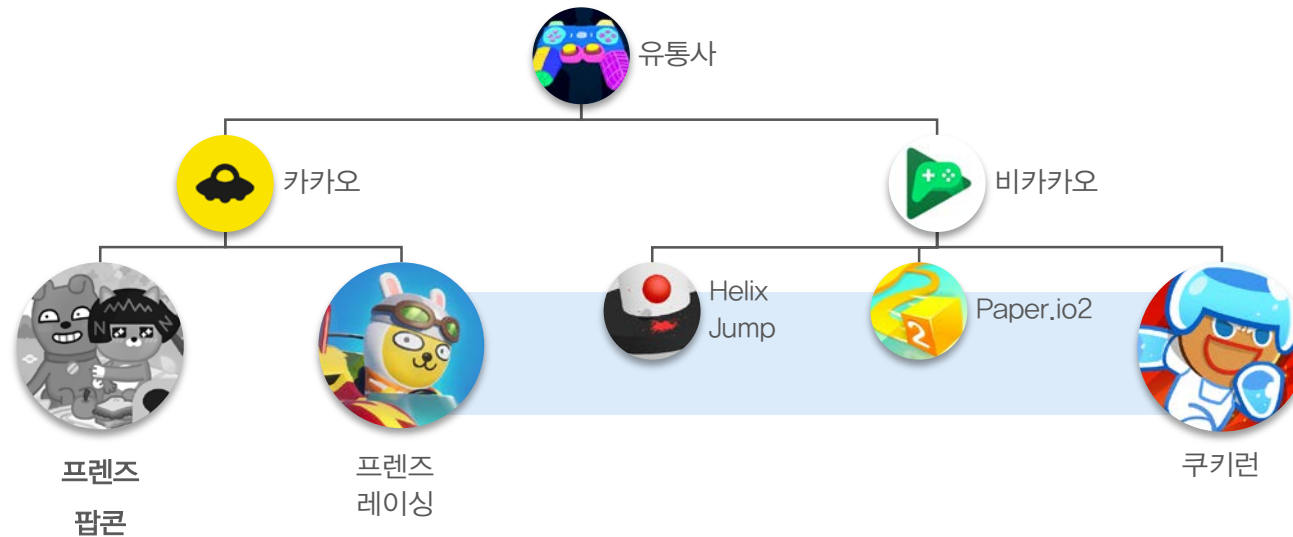
전환이 더 빈번해
더 적절한 경쟁구도

A. 경쟁군 비교 분석 : [REMIND] WHO? (3) 비교군 경쟁분석 – 비교군 내/외부의 경쟁성 비교

분석 모델 : 배속 로지스틱 회귀모형(Nested Logistic Regression Model)

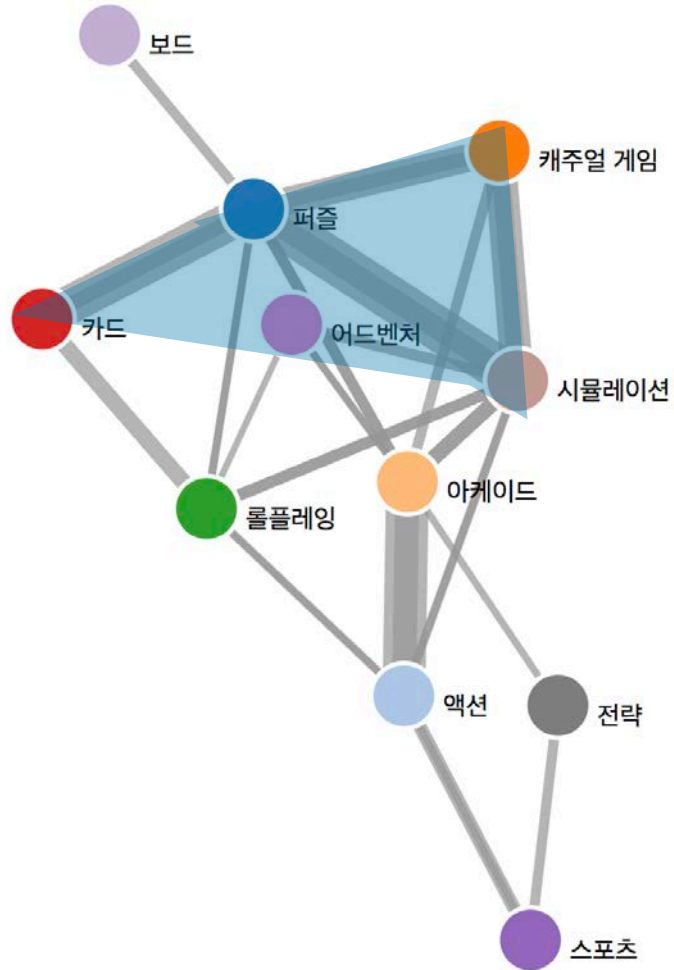
분석 결과 2

프렌즈 레이싱 vs 프렌즈 팝콘 : 경쟁구도 분석



프렌즈레이싱의 관점:
같은 카카오 게임인 **프렌즈팝콘에 비해**
쿠키런 (비카카오 게임)와의 경쟁이 더 치열

B. 협력사 분석 : 협력은 또 다른 경쟁



동시에 플레이된 게임은 경쟁관계인가? 협력관계인가?

[게임]이라는 같은 분야 내에서 함께 사용되었기 때문에, 서로의 대체재로써 작용 가능

일반적으로 게임간의 협력 관계 구축은 어려움

네트워크 분석으로 협력 관계를 찾을 수 있는가?

서로 다른 분야의 서비스 동시 이용은 보완적인 관계 형성 가능

B. 협력사 분석 : 핵심 질문

1. 누구와 협력해야 하는가?
2. 어떻게 협력해야 하는가?

B. 협력사 분석 : 협력은 또 다른 경쟁



강남 데이터 선정 이유

실질적인 거래를 다루는 오프라인 결제 데이터 사용

지리적 요인으로 인해 같은 상권 내에서 협력/경쟁이 용이
가장 대표적인 상권인 강남에서 성공적인 분석이 가능하다면,
국내 다른 상권에도 적용 가능하다고 판단

활용 데이터 추출

1. Payment 내의 GPS 정보가 강남 상권에 속한 23,903건의 데이터 중
2. Unique한 방문 패널의 수가 $n(>10)$ 이상인 상점 데이터 추출
3. 강남 내 결제 횟수가 $n(>10)$ 건 이상인 패널의 결제 데이터 추출

주요 전처리 내용

APPROVAL_TYPE = LA(국내 승인), LT01(간편결제 승인)

CATEGORY_CODE = 오프라인 결제라 볼 수 없는 것들 삭제
(예 801: 통신, 1201~1204: 공과금 등)

CARD_PAYMENT_TYPE = PCT(신용카드), PCK(체크), PA(간편결제)

B. 협력사 분석 : 협력은 또 다른 경쟁

TF-IDF를 활용한 유사도

TF-IDF를 활용하여 강남 내 가게 간 유사도를 계산
같은 주기에 함께 방문(등장)한 패널이 많을 수록 두 가게간의 유사도가 높다.

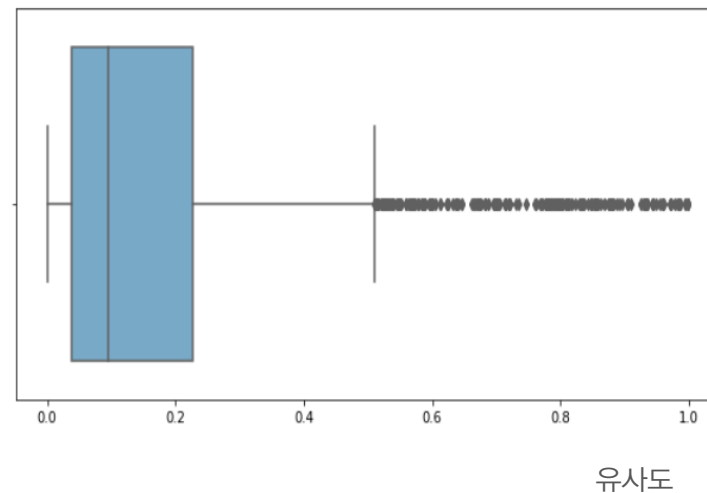
EDA : 유사도 분포

공유 고객이 전혀 없는 경우와, 자기 자신과의 유사도를 제외하면
2971건의 링크가 발견됨

유사도는 0과 1사이의 값으로 표현

0.5 이상의 유사도는 표본 수가 너무 적어 발생하거나,
단골 고객 한 명의 영향을 크게 받는 아웃라이어

약 0.2 이상의 유사도를 갖는 관계가 강한 관계,
0.1 이상인 경우 유의미한 고객군을 공유한다고 판단



B. 협력사 분석 : 협력은 또 다른 경쟁

유사도 matrix

	버거킹	한솔	뮤즈의원	에프알엘코리아(주)강남점	씨제이올리브네트웍스	스타벅스	씨유	김밥천국	GS25	우리집김밥
버거킹	1.000	0.000	0.000	0.000	0.032	0.043	0.057	0.000	0.027	0.052
한솔	0.000	1.000	0.000	0.000	0.000	0.036	0.000	0.000	0.005	0.000
뮤즈의원	0.000	0.000	1.000	0.197	0.093	0.000	0.000	0.000	0.022	0.000
에프알엘코리아(주)강남점	0.000	0.000	0.197	1.000	0.057	0.066	0.014	0.000	0.001	0.000
씨제이올리브네트웍스	0.032	0.000	0.093	0.057	1.000	0.113	0.087	0.000	0.185	0.000
스타벅스	0.043	0.036	0.000	0.066	0.113	1.000	0.100	0.058	0.140	0.016
씨유	0.057	0.000	0.000	0.014	0.087	0.100	1.000	0.016	0.071	0.009
김밥천국	0.000	0.000	0.000	0.000	0.000	0.058	0.016	1.000	0.024	0.000
GS25	0.027	0.005	0.022	0.001	0.185	0.140	0.071	0.024	1.000	0.007
우리집김밥	0.052	0.000	0.000	0.000	0.000	0.016	0.009	0.000	0.007	1.000

358x358 matrix

강남 상권 관계도

2018.11 ~ 2019.02

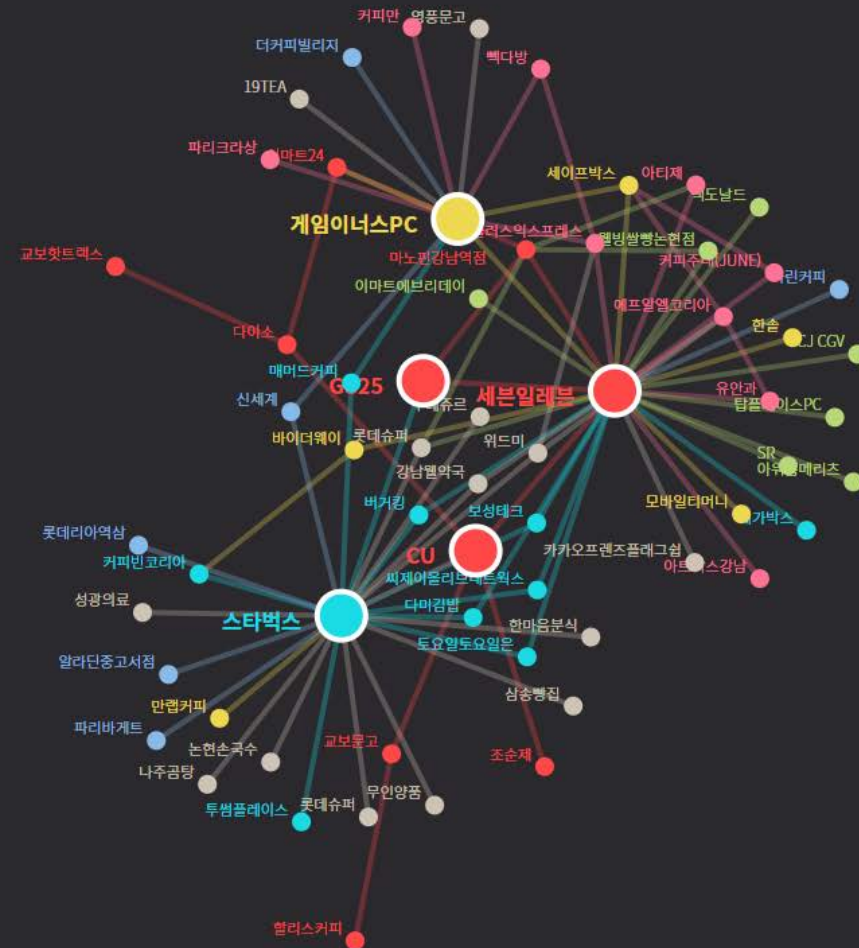
[보는 방법]

1. 전체보기 버튼을 누르면 네트워크가 나타납니다.
2. 각 상호와 관계선을 클릭하면 세부정보가 나타납니다.
3. 각 상호를 클릭하면 관계선이 강조됩니다. 강조선을 해제하려면 한번 더 누르세요.
4. 네트워크가 겹치면 드래그해서 움직여보세요.

상호 혹은 연결선을 클릭하세요

- 편의점
- 카페/프랜차이즈
- 식당/프랜차이즈
- 카페/비프랜차이즈
- 식당/비프랜차이즈
- 쇼핑
- 기타

기준 : NOV. 2018 ~ FEB. 2019



Z. 활용 방안

A. 경쟁 분석

네트워크 분석, 키워드 분석을 통한 경쟁구도 및 경쟁자 파악 가능
배속 로지스틱 회귀모형을 통한 경쟁구도 하 경쟁 분석 가능
전환위험지수를 통한 비교우위/열위의 파악과 경쟁강도 측정 가능

엠브레인의 설문조사 플랫폼을 활용하여 게임과 관련된 추가적인 패널 설문을 진행한다면,
더 설명력 있는 모델의 설계가 가능

B. 협력 분석

TF-IDF를 활용한 사업자 간 유사성 계산
유사도를 통한 추가적인 사업자간 협력관계 규명 및 패널별 특성을 활용한 마케팅 제언
유사성 기반 네트워크 웹 시각화

추가적인 데이터 수집을 통한 표본 대표성의 확보

