The background is a dense, overlapping collage of various Korean movie posters. Visible titles include '삼진그룹병토의반' (Samjin Group's Secret), '슬기로운 의사생활' (Hospital Playlist), '극한직업' (Extreme Job), '변신' (Metamorphosis), '개미는 오늘도 운운' (The Ant is Still Talking), '종이집사' (Paper Man), '시지프스' (Sisyphus: The Myth), '완벽한 타인' (Perfect Stranger), '기묘한 이야기' (Strange Stories), '1호가 돌아왔다' (No. 1 has returned), and '소리호' (Soriho).

영화추천 알고리즘 고도화를 통한 고객 맞춤 서비스 제공으로 매출 증대

(주) 좋은영화 전략 3팀

A3 김경민 문찬국 이소희 이희창 전해린 최민호

- Chapter 01 추진 배경
- Chapter 02 현상 및 개선 기회
- Chapter 03 분석계획 및 결과
- Chapter 04 개선안 및 적용방안
- Chapter 05 Learned Lesson
- 첨부. 근거자료



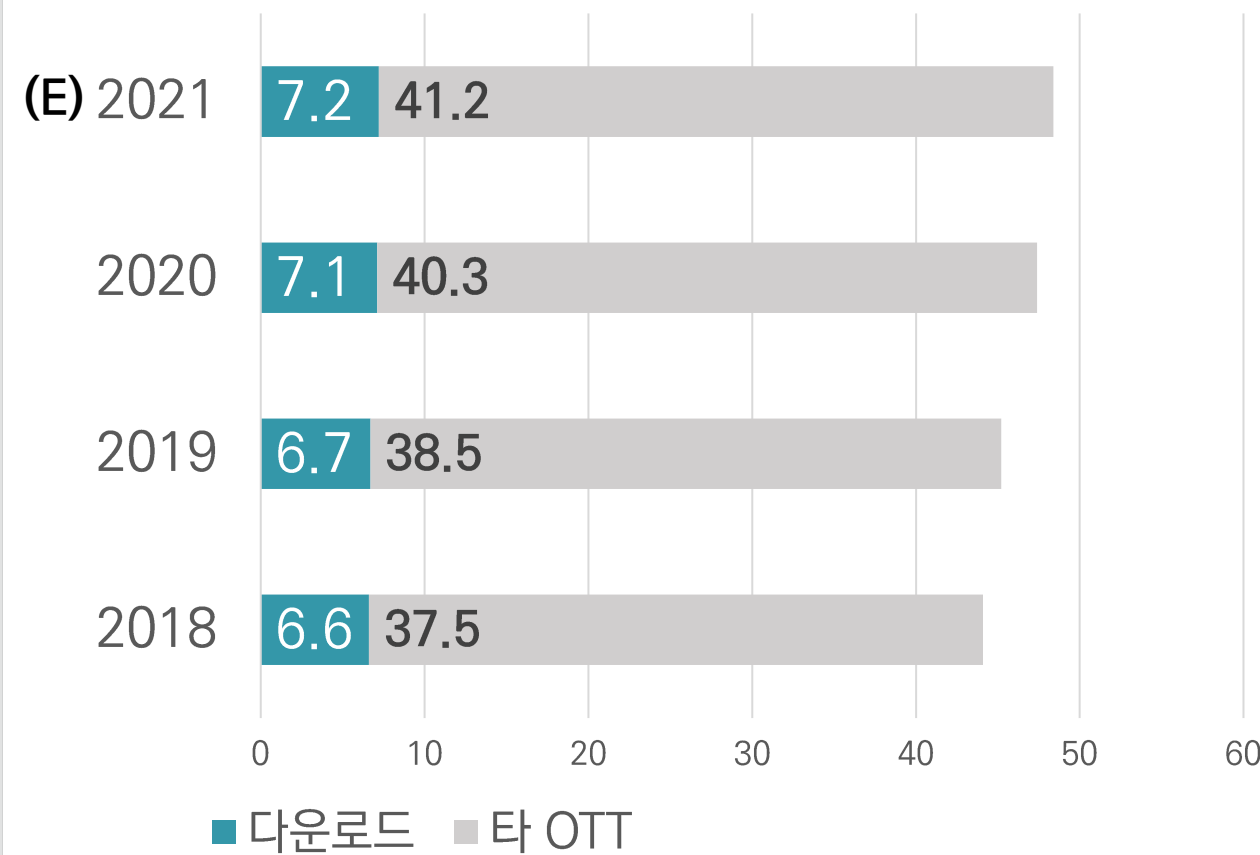
제작/배급사와 콘텐츠 계약을 통해 영화 다운로드 서비스를 제공하는 업체

OTT 시장의 경쟁 과열로 인한 좋은 영화의 매출 감소

자사 수익 구조



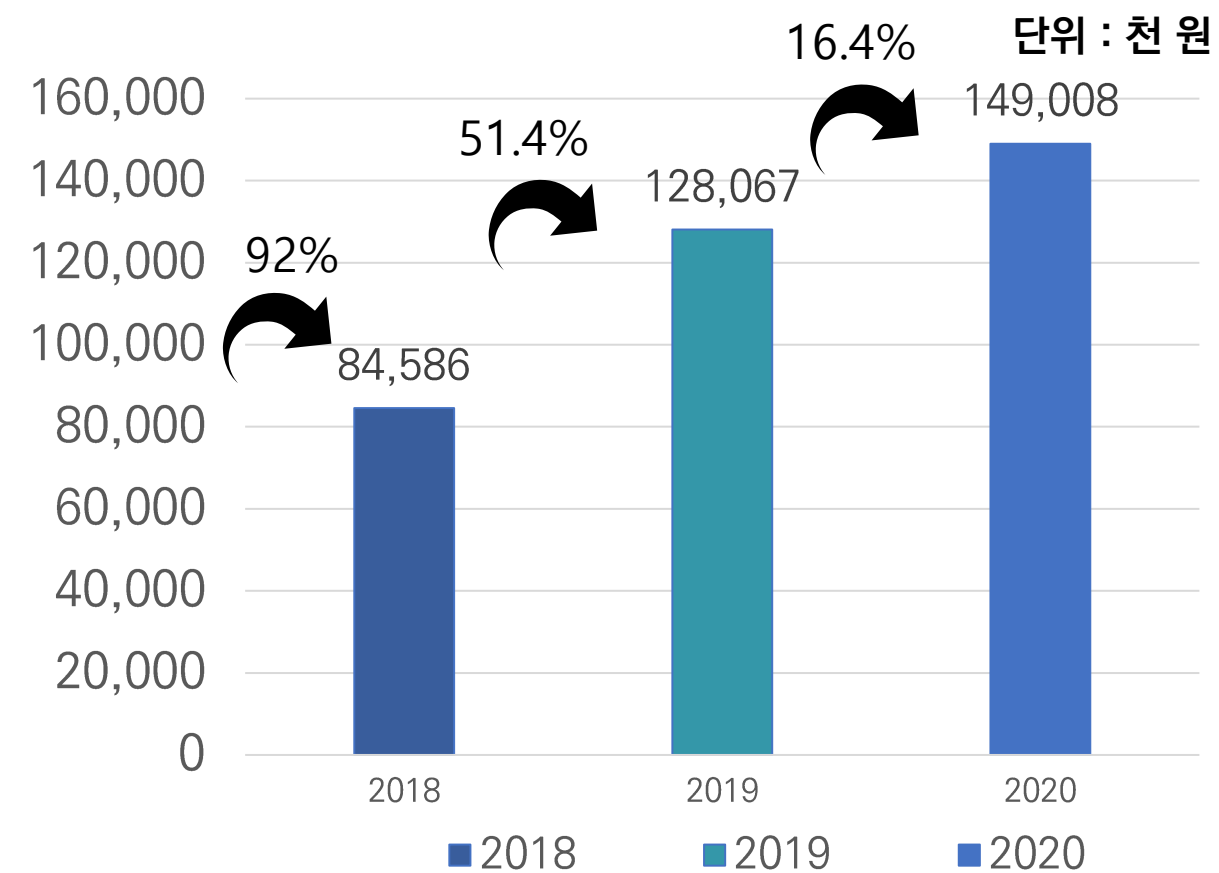
인터넷 동영상 사용자 규모 및 전망



다운로드 시장은 규모와 성장률이 작음

출처: DMC미디어 2020 인터넷 동영상 시장 현황 및 전망

자사 매출 현황



매출 증가율이 매년 하락, 현재 16.4%에 정체

포화된 OTT 시장 속 회사의 변화 필요

포화된 OTT 시장 속 회사의 변화 필요

기존 추천 알고리즘 개선



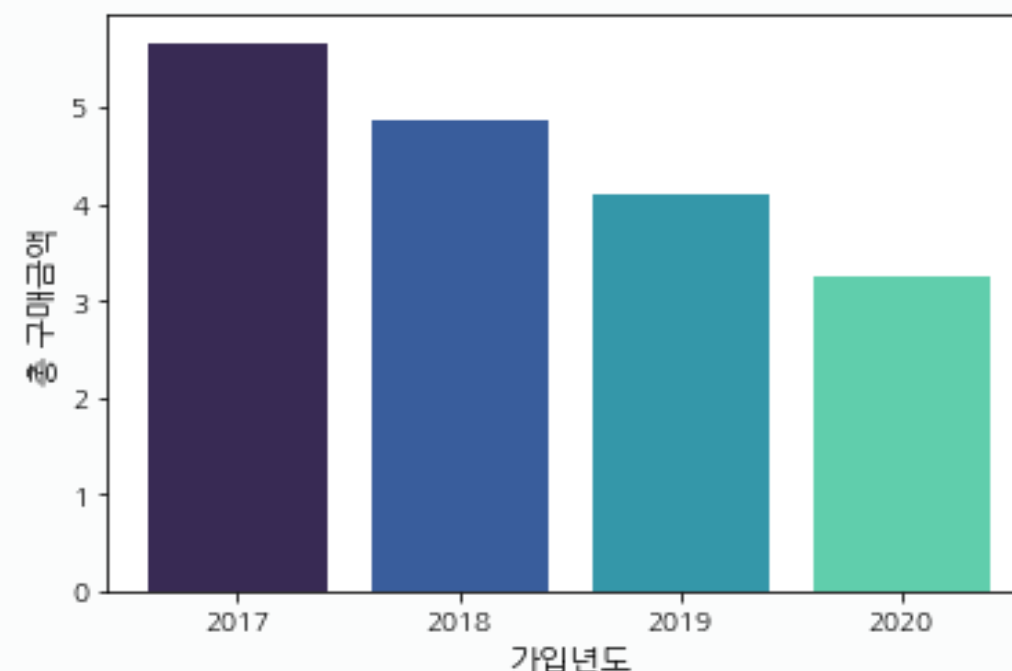
국내 OTT플랫폼 상위 회사는 이미
고객 맞춤 추천 알고리즘 제공

넷플릭스의 경우 고객 콘텐츠 이용 비율 중
추천 콘텐츠 이용 비율 75%

고객 개인 특성을 고려한 추천 알고리즘 개발

충성도 높은 고객 관리

고객 가입년도별 2020년 총 구매 금액

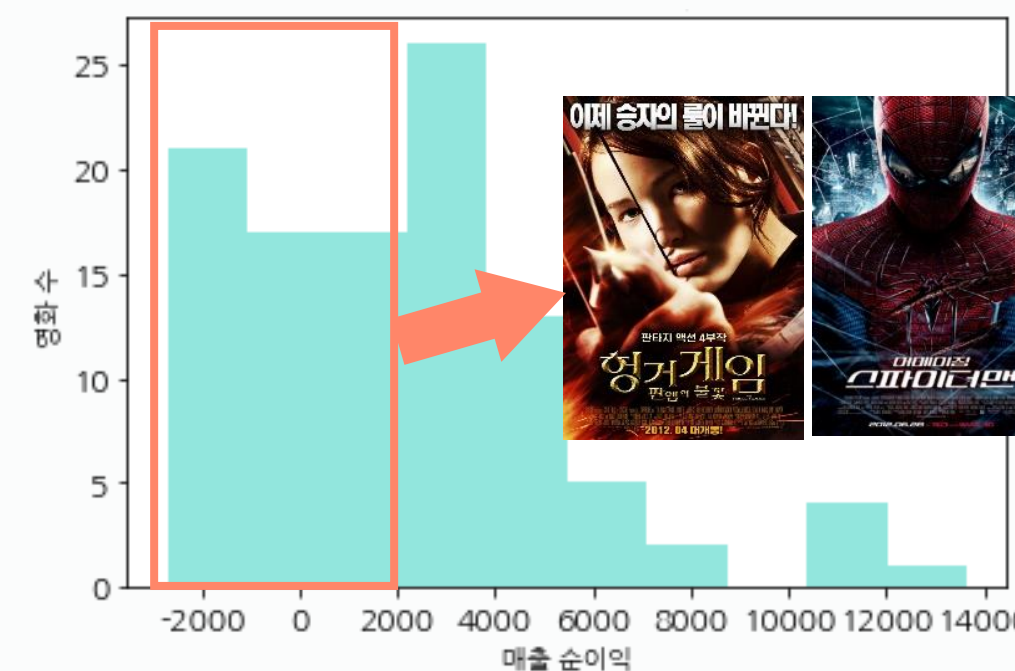


가입 기간이 길수록 매출에 끼치는 영향이 큼
자사의 경우 기존 고객 관리를 하지 않음

기존 고객의 매출 유지 및 향상을 위한 관리

판권계약 프로세스 개편

영화 별 매출 순이익



매출 이익을 내지 못하는 영화가 많음
흥행했던 영화도 자사에서는 매출이 없음

고객 수요가 있는 판권 계약



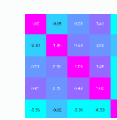
알고리즘 개발, 고객 관리와 판권 계약 프로세스 개선 → 1년 이내 매출 2억 달성

Data Set



고객 정보(3277, 11), 판권 보유 영화 정보(106, 32), IMDB 영화 정보(1668, 30), 고객 다운로드 기록(116464, 32) (매출에 관련된 칼럼 별도 추가)
- 18년 ~ 20년 약 12만 개의 데이터로 분석 진행

◆ 분석 내용

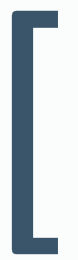
영화 상영 당시 흥행한 영화와 매출액 관계 확인

-  막대 그래프 자사 매출 상,하위 10%, 개봉 영화관 수, 관람객 리뷰 수, 미국에서의 매출 비교
-  통계적 검정 자사 매출과 평점, 개봉 영화관 수, 관람객 리뷰 수, 미국에서의 매출 사이 관계 분석
(카이 제곱, 상관 분석)
-  히트맵 영화 흥행과 매출 간 상관 관계


고객 특성과 고객별 다운로드 수의 관계 파악

-  막대 그래프 연령대와 다운로드 수의 관계
-  파이 차트 성별, 12세 이하 자녀 유무,거주 지역, 결혼 여부와 고객별 다운로드 수 간의 관계

매출에 영향을 미치는 핵심인자 도출

-  다중선형회귀
Decision Tree
Random forest
Gradient Boosting 가장 정확도가 높은 모형 선정
새로운 판권을 계약할 때 핵심인자 반영

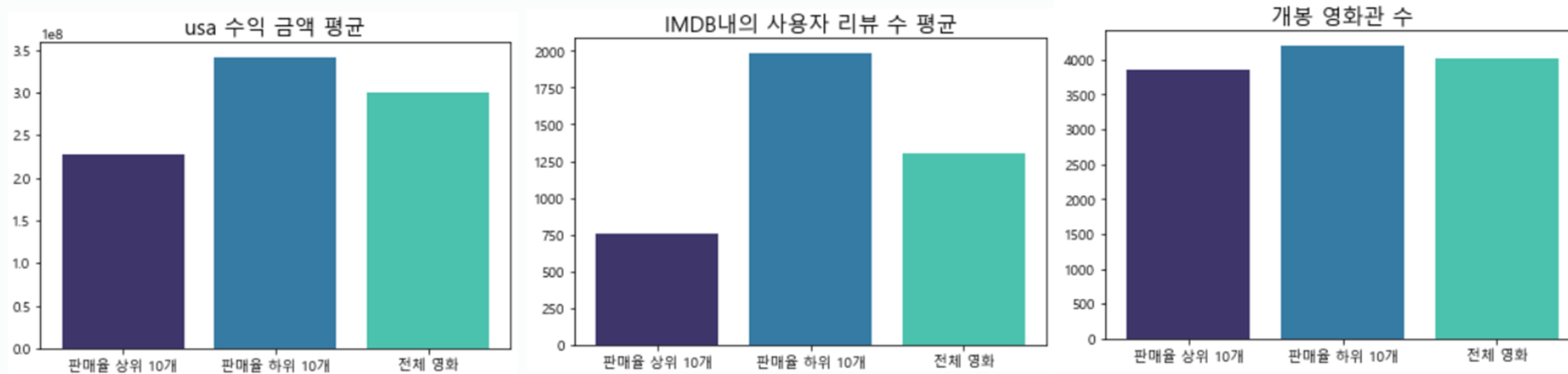
고객과 영화 특성을 고려해 영화 추천

-  하이브리드 추천
(협업 기반 필터링,
콘텐츠 기반 필터링) 하이브리드 추천 알고리즘을 통해
고객의 특성을 고려한 영화 추천

영화 상영 당시 흥행한 영화와 매출액 관계 확인

탐색적 분석

자사 매출 상위 · 하위 10% 영화 특성 비교



자사 매출 상위 10%는 오히려 상영 당시 상대적으로 **흥행하지 않은** 영화들

통계 검정

개봉당시 매출액, 개봉 영화관 수, 관람객 리뷰 수 와 매출 사이 상관 검정

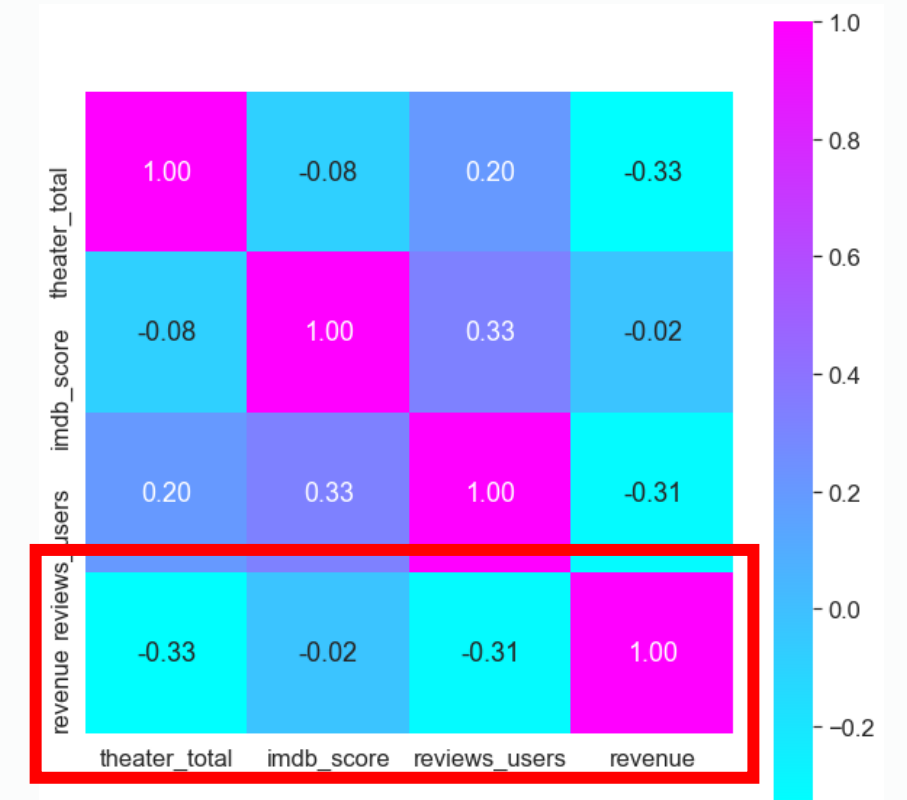
가설	변수별 매출과 연관성 유무	변수	상관계수	P-value
귀무가설	없다	개봉당시 매출	-0.301	0.002
대립가설	있다	개봉 영화관수	-0.332	0.001
		관람객 리뷰 개수	-0.307	0.001

통계적으로 유의미하므로 해당 변수들은 매출과 연관성 有

- 박스오피스 기준 영화 흥행 척도 사용 : 매출액, 관객수
- 매출액 : 개봉 당시 영화 매출액(미국)
- 관객수 : 개봉 영화관 수, 유저 리뷰 개수

히트맵

상관관계 분석



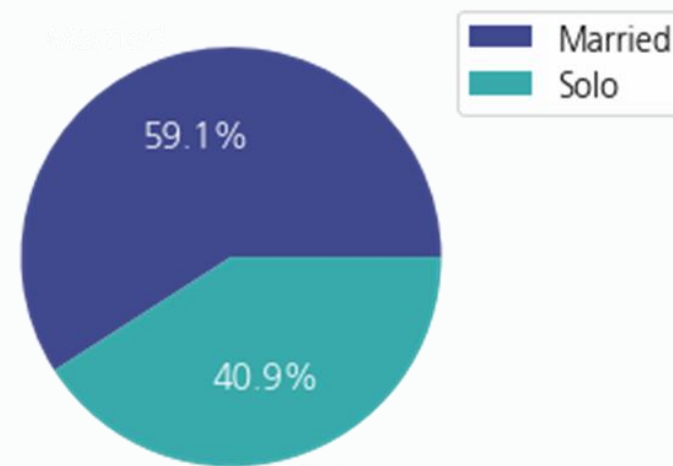
수입은 미국 내 수익, 개봉 영화관, 리뷰한 고객 수와 **음의 상관관계**

개봉 당시 흥행했던 영화는 매출 부진 → 판권 계약 시 해당 변수 고려

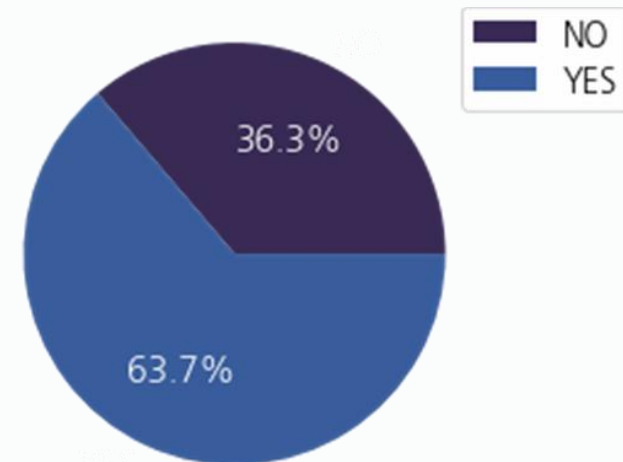
고객 특성과 고객별 다운로드 수의 관계 파악

자사 고객 특성에 따른 다운로드 현황

결혼 여부에 따른 1인 평균 다운로드 비율



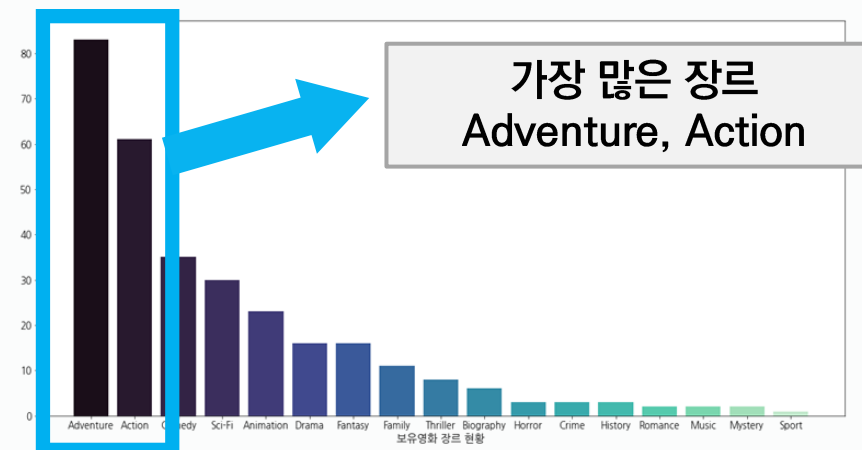
아동 유무에 따른 1인 평균 다운로드 비율



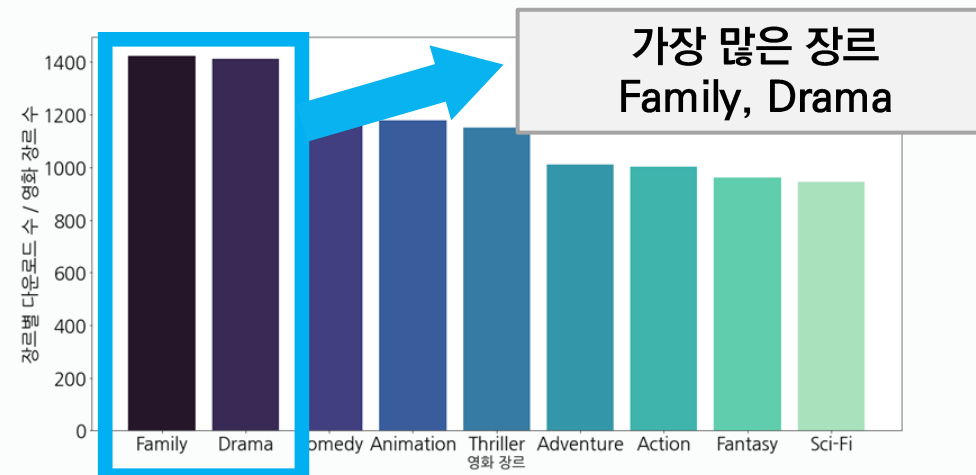
기혼, 아동 O 비율 높음
→ 가족 고객이 많음

자사 고객 장르 선호도 파악

좋은영화사 보유 영화 장르 수



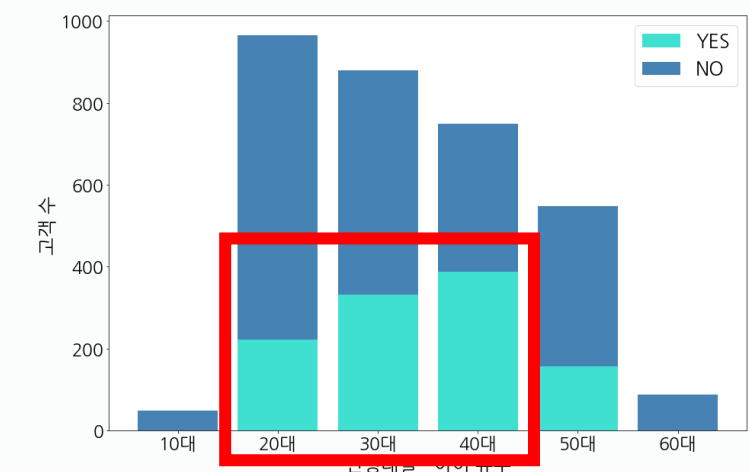
(고객이 다운로드한 영화 장르 수) / (전체 영화 장르 수)



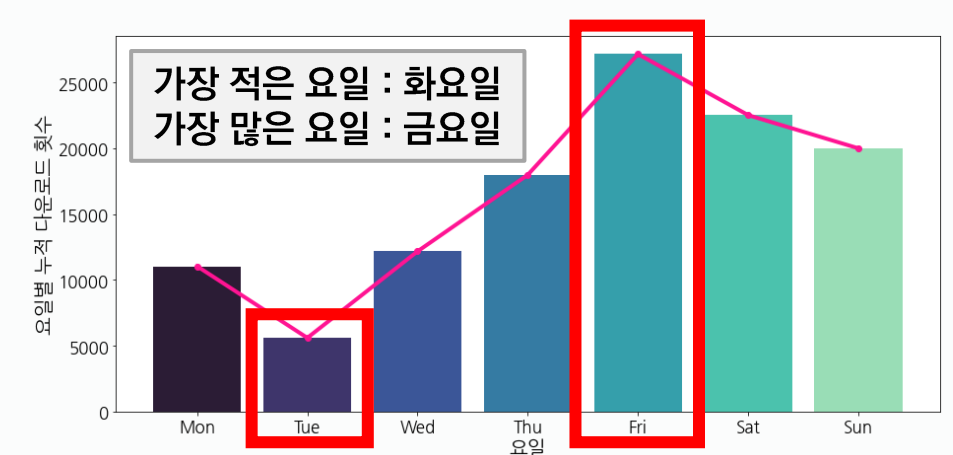
보유 영화 장르 수 대비 고객 선호 영화 장르 파악
→ Action, Adventure 이외 Family, Drama 가 상위 차지

프로모션 타겟 선정 관련 분포

연령대별 아이 유무



요일 별 다운로드 분포



아이 있는 고객 중 3-40대 고객 비율이 높음
다운로드 횟수는 주중이 낮고 주말이 높음

기혼 고객이나 아이가 있는 고객이 선호하는 장르의 영화 계약 & 해당 고객 대상으로 프로모션 진행

모델링을 통해 매출에 영향을 주는 고객의 특성 파악

◆ 핵심영향인자 확인

Target: 매출액

변수 중요도에 따른 영향 인자 도출

순위	Decision Tree	Random Forest	Gradient Boosting	Multiple Regression
1	가입 년도	가입 년도	가입 년도	자녀 여부
2	자녀 여부	자녀 여부	자녀 여부	가입 년도
3	연령대	연령대	연령대	결혼 여부

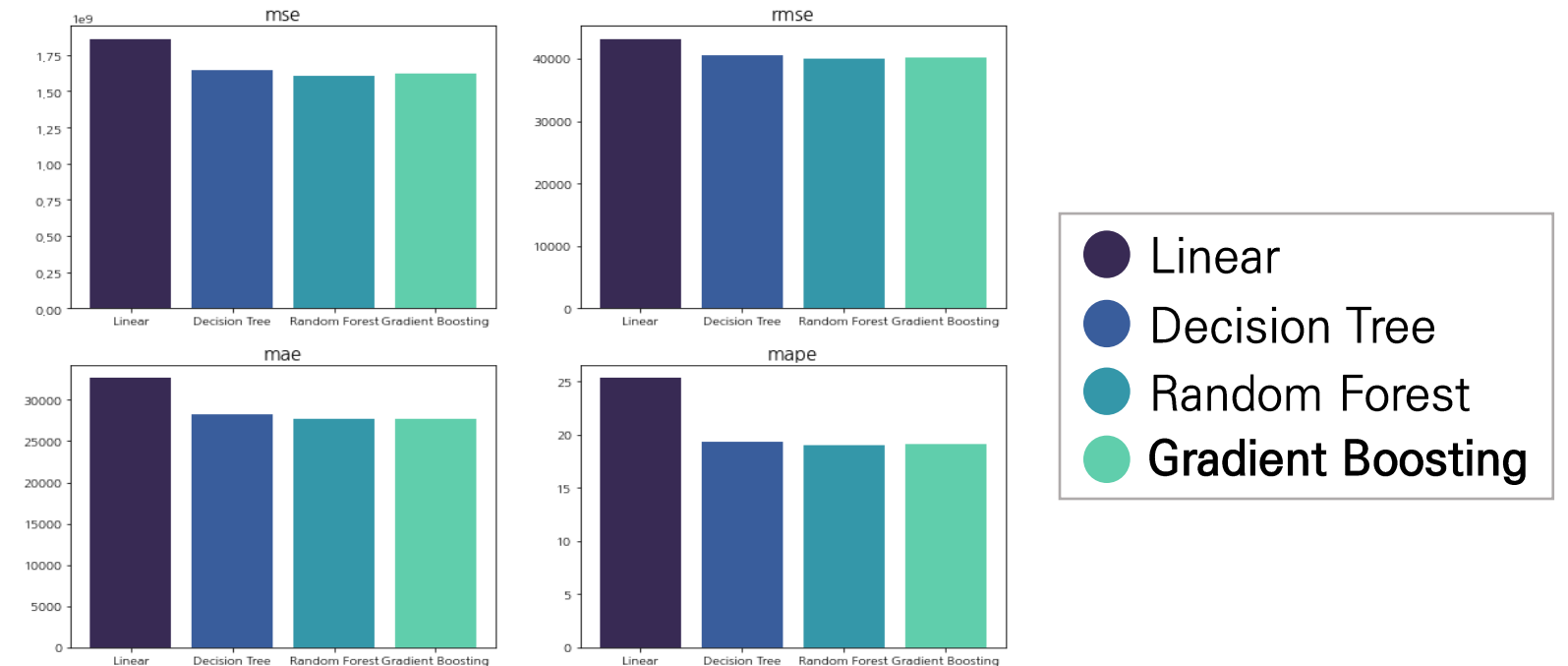
• 핵심 영향인자

1
가입 년도

2
12세 이하 자녀 여부

3
연령대

◆ 모델 성능 비교



분석 모델	Train data 설명력	Test data 설명력
Decision Tree	0.843	0.809
Random Forest	0.849	0.812
Gradient Boosting	0.858	0.811

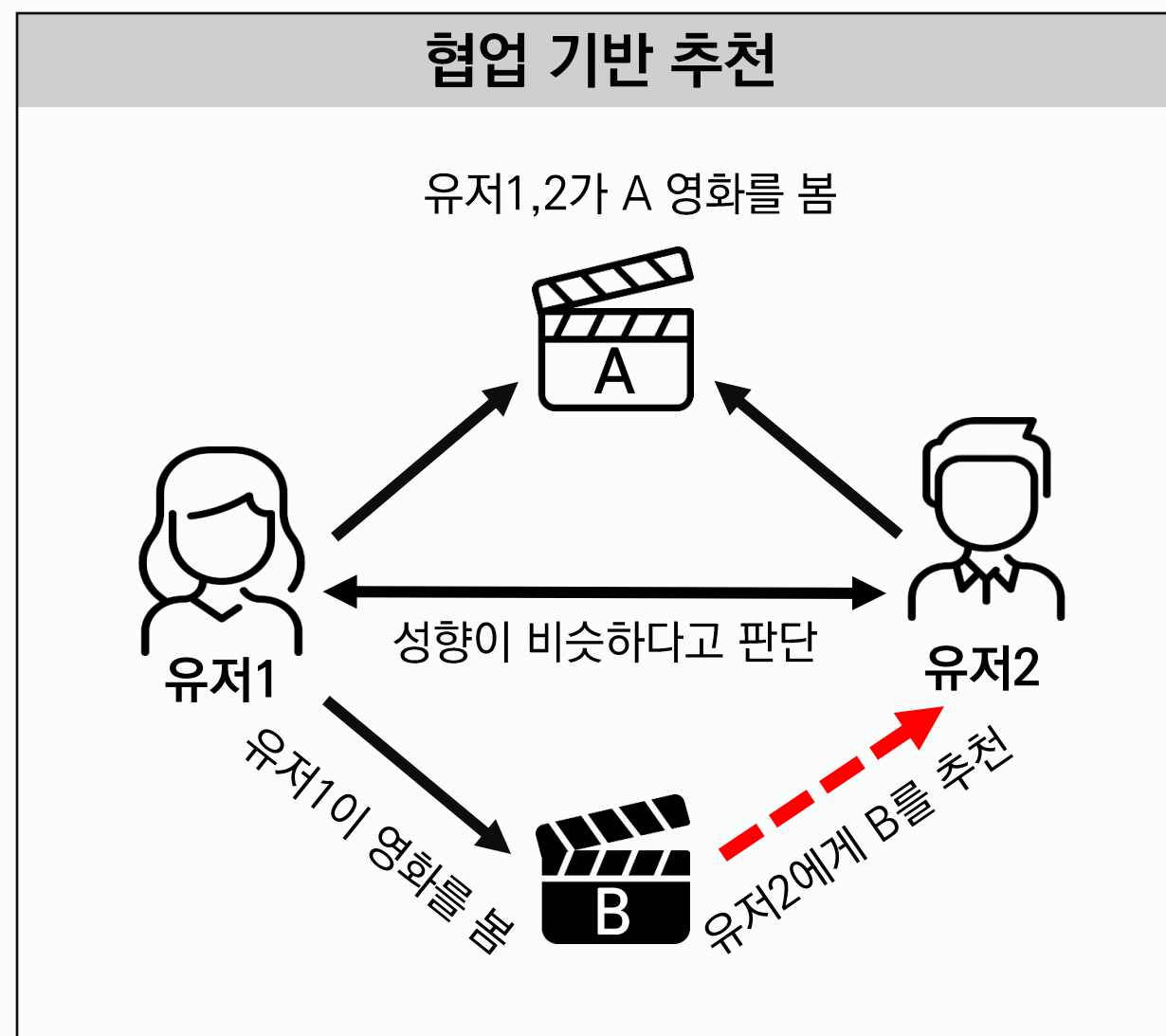
Gradient Boosting 모델링 사용

가입기간이 길고, 12세 이하 자녀가 있는 고객을 대상으로 하는 프로모션 진행

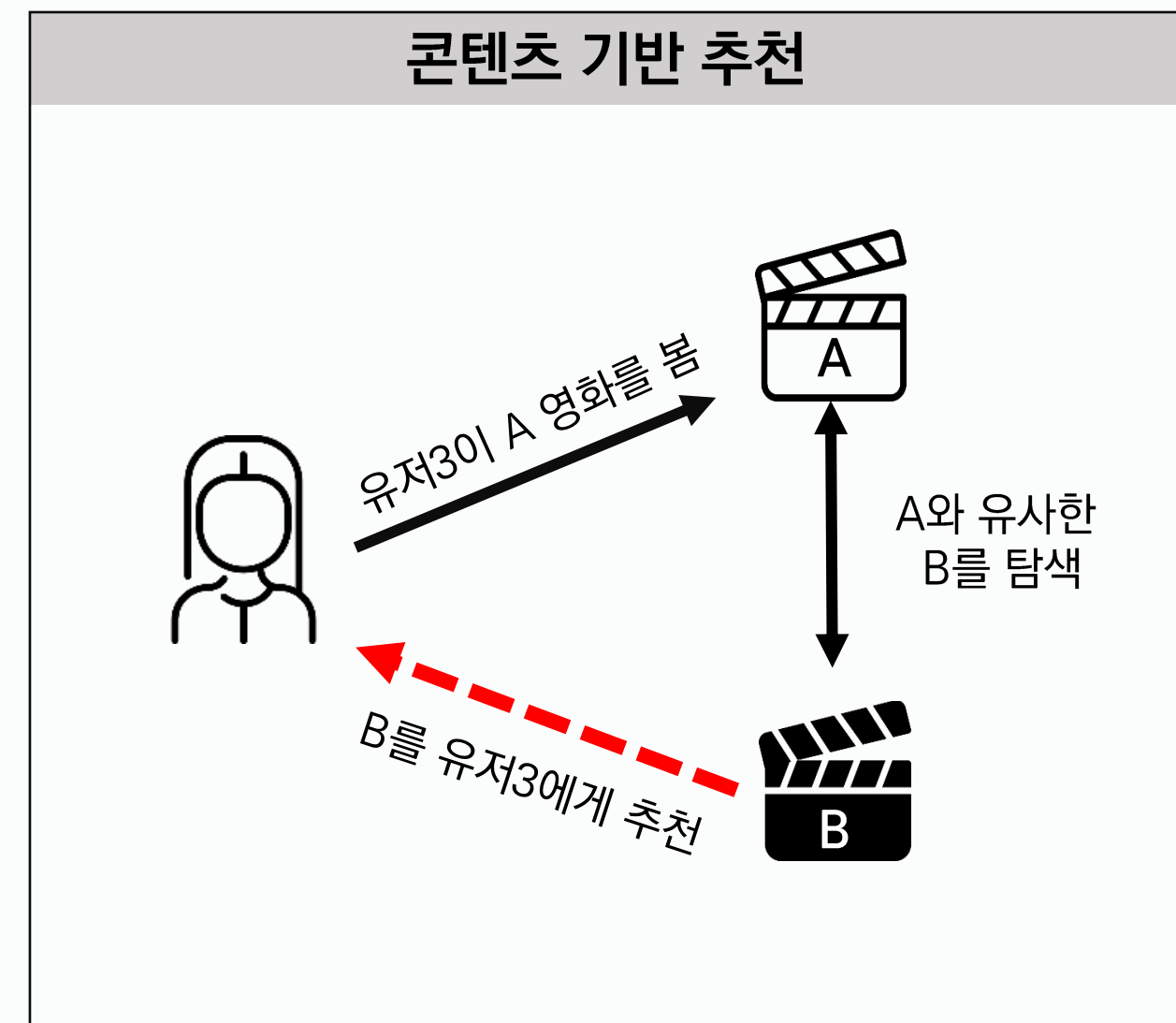
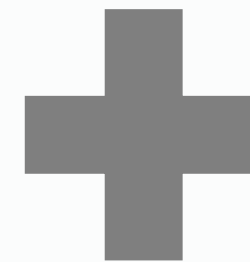
하이브리드 추천 알고리즘을 통해 고객의 특성에 맞는 영화 추천

* 하이브리드 추천 알고리즘

협업 기반 추천으로 고객 간 특성을 비교하여 추천하고,
콘텐츠 기반 추천으로 고객이 많이 다운로드했던 영화와 비슷한 영화를 추천하는 방식



고객의 선호도를 고려한 영화 추천



이전에 구매했던 영화와 비슷한 영화 추천

하이브리드 추천 알고리즘을 통해 고객의 특성에 맞는 영화 추천

1 고객의 id를 입력 & 입력 받은 고객의 정보 출력

회원번호 : C7122
성별 : M
연령대 : 20대
결혼 여부 : S
12세 이하 자녀 유무 : N
선호 장르 : Adventure Action Sci-Fi



고객의 성별
연령대
혼인여부
12세 이하 자녀 유무

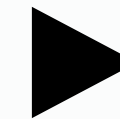
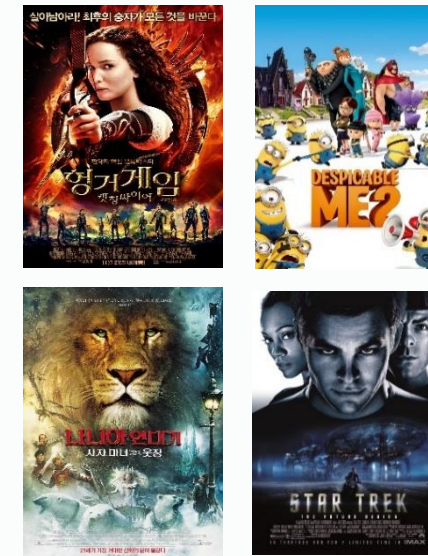
2 고객의 특성군에 속하는 고객들이 시청한 영화 목록 출력

title	Genre_1	Genre_2	Genre_3	actor_1	director
The Hunger Games: Catching Fire	Action	Adventure	Mystery	Jennifer Lawrence	Francis Lawrence
The Chronicles of Narnia: The Lion, the Witc	Adventure	Family	Fantasy	Georgie Henley	Andrew Adamson
Despicable Me 2	Animation	Adventure	Comedy	Steve Carell	Pierre Coffin
Monsters University	Animation	Adventure	Comedy	Billy Crystal	Dan Scanlon
Star Trek	Action	Adventure	Sci-Fi	Chris Pine	J.J. Abrams

3 감독, 배우 및 장르를 고려하여 유사도가 높은 영화 중 가장 인기 있는 작품 추천

	title	Genre_1	Genre_2	Genre_3	actor_1	director
0	Mad Max: Fury Road	Action	Adventure	Sci-Fi	Tom Hardy	George Miller
1	Star Trek	Action	Adventure	Sci-Fi	Chris Pine	J.J. Abrams
2	I Am Legend	Adventure	Drama	Sci-Fi	Will Smith	Francis Lawrence

고객과 성향이 비슷한
고객들이 시청한 영화 목록



유사도가 높은 영화 중 가장 인기 있는 작품 추천



가중치

- 영화 순위 계산 시 사용
- IMDB에서 사용하는 가중치 공식 사용

$$\left(\frac{v}{v+m} \cdot R\right) + \left(\frac{m}{v+m} \cdot C\right)$$

v : 영화의 득표 수
m : 차트에 나열되기 위한 최소한의 득표 수
→ 관람객 리뷰 수로 대체
R : 영화의 평균 평점
C : 전체 차트의 평균 평점 → 매출(미국)으로 대체

코사인 유사도 (Cosine Similarity)

- 두 영화 사이의 유사성을 계산
- 고객 특성들 사이에서 유사도가 높을수록 더 큰 값을 부여

$$\text{cosine}(x, y) = \frac{x \cdot y^T}{\|x\| \cdot \|y\|}$$

고객 취향 분석과 유사도 높은 영화 추천 알고리즘으로 다운로드 수 증가를 통한 매출 증대

고객의 특성을 고려한 효율적인 프로모션으로 매출 증대, 회원 등급제 도입으로 고객 관리

회원 등급제 도입



구매 금액별 고객 혜택 세분화
- 다운로드 시 할인 쿠폰을 제공

베를린, 칸 3%, 베니스 5%, 아카데미 7%



상위 등급 고객을 위한 이벤트 진행

기존 고객에게 혜택 제공

충성도 높은 고객 유치

고객 맞춤형 프로모션 진행



기혼, 자녀여부, 연령대 등을 고려해 고객
맞춤 프로모션 진행



아이와 같이 영화관에 가지 못하는 고객을
대상으로 신작 콘텐츠 제공

고객 맞춤형 프로모션으로 주 고객층 이탈 방지

요일별 판매 프로모션 실시



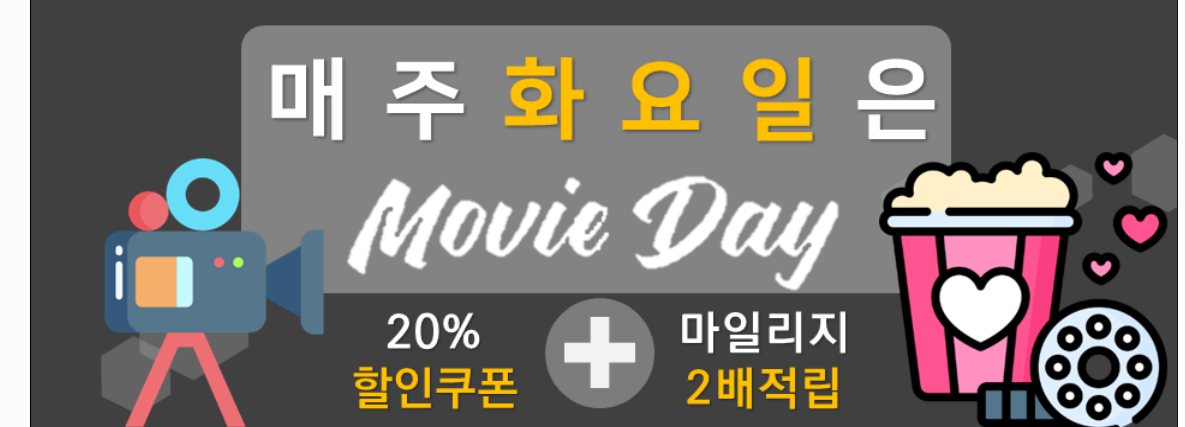
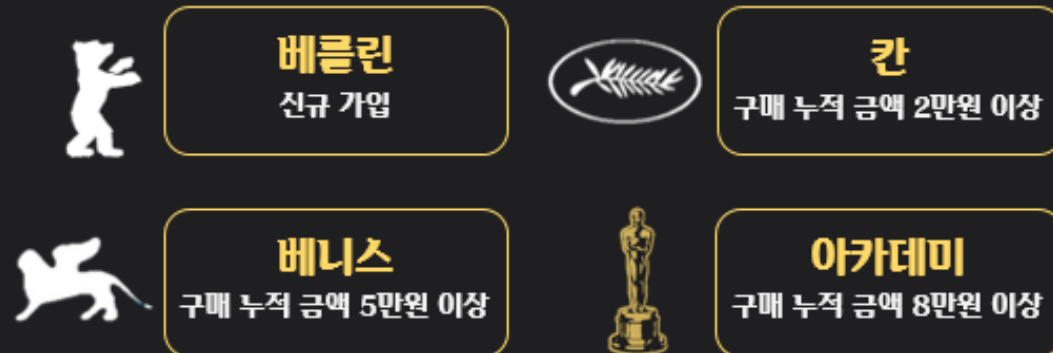
주중에는 알림을 통해 다양한 영화 홍보 및
기간 한정 쿠폰, 시리즈 등 할인 판매



주말에는 매출이 부진한 영화를 할인쿠폰
제공

매출이 부진한 요일에 다운로드 유도로 매출 증대

다운로드 수가 낮은 영화 접근성 향상



영화 판권 계약 시 고객 선호도 반영

고객 선호도를 반영한 판권 계약



매출비중이 높은 12세 이하 자녀가 있는 40대 고객이 선호하는
콘텐츠 계약 (family, drama 장르의 영화들의 판권 계약)

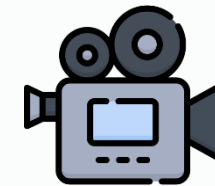


주고객층 중 아이들 때문에 영화관을 가지 못하는 고객층
대상으로 신작 콘텐츠 제공

판권 계약 프로세스 개선



자사 고객은 오히려 크게 흥행이 된 영화를 비선호



영화 흥행 척도 기준을 나눠 상대적으로 덜 흥행한 영화
판권계약을 진행

◆ 판권 계약 프로세스 개선 적용 시뮬레이션



제목: Mr. Bean's Holiday
장르: Family, Comedy
계약금 : 2,195달러
매출 : 6,190달러
다운로드 수 : 884회



제목: Evan almighty
장르: Family, Comedy, Fantasy
계약금 : 2,295달러
매출 : 5,750달러
다운로드 수 : 821회



제목: Glass
장르: Drama, Sci Fi, Thriller
계약금 : 2,195달러
매출 : 5,310달러
다운로드 수 : 758회

새로운 판권 계약 프로세스 적용 → 합리적인 가격의 판권 계약을 통해 매출 증대

추천 알고리즘에 대한 개선 방안

기술 경쟁력 향상

기존 고객 반응이 높은 선별 능력 알고리즘에서
고객 개인의 취향을 반영한 알고리즘으로 개선

➔ OTT 시장 내 기술 경쟁력 확보

고객의 만족도 증가

고객의 취향을 잘 반영하여 선호도에 맞는 영화 추천

- ➔ 고객의 서비스 이용 만족도 증가
- ➔ 기존 고객 이탈 감소

구매율 증가

고객의 취향 범위 내에서 기존에 접하던 영화 외
새롭게 접할 수 있는 영화 추천

➔ 사용자의 구매 결정시간 단축

하이브리드 추천 알고리즘



웹 개발 시연 (신규회원)



Good Movies

로그인 회원가입

좋은 영화 회원 등급제 도입!

매년 다운로드 누적 금액별 회원 등급제 도입



베를린
신규 가입



칸
구매 누적 금액 2만원 이상



베니스
구매 누적 금액 5만원 이상



아카데미
구매 누적 금액 8만원 이상



매 주 화 요 일 은 Movie Day



20%
할인쿠폰



마일리지
2배적립



1일 동안 보지 않음

닫기



웹 개발 시연 (기존 회원)



Good Movies

로그인 회원가입

좋은 영화
회원 등급제 도입 !
매년 다운로드 누적 금액별 회원 등급제 도입



베를린
신규 가입



칸
구매 누적 금액 2만원 이상



베니스
구매 누적 금액 5만원 이상



아카데미
구매 누적 금액 8만원 이상



모든 것을 압도할
진짜 이야기의 시작



(주) 좋은영화 전략 3팀

대표이사 이소희

빅데이터 분석 프로젝트를 통해 빅데이터를 실제로 직접 만져보고 모델링을 해보면서 중요인자를 찾는 과정이 재밌었습니다. 또한 프로젝트를 통해 한층 더 성장해서 기분이 좋습니다.

사장 전혜린

빅데이터 교육 한 주 한 주가 배울 점이 많고 유익한 시간이었습니다. 교육기간 동안 좋은 가르침을 주신 교수님들께 감사의 말씀 전합니다!

부장 문찬국

밤 늦게까지 조원들이랑 같이 프로젝트를 진행하며 데이터에 관해 의견을 주고받고, 도메인 지식을 쌓기 위해 공부하며, 의견을 조율하는 과정에서 많은 것을 배울 수 있었습니다.

팀장 최민호

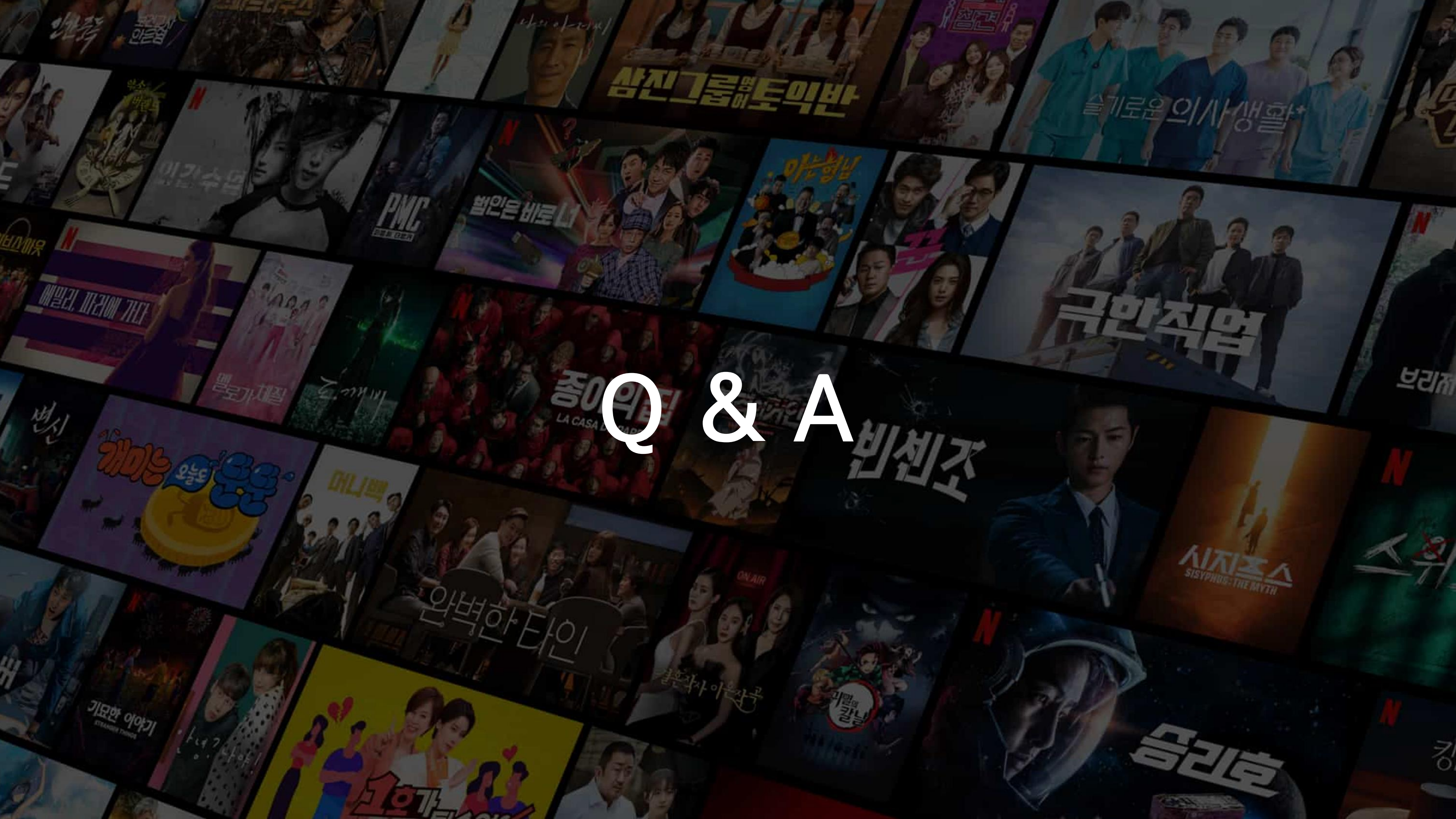
데이터 분석의 어려움을 느낌과 동시에 인사이트를 발견하는 재미를 느낄 수 있었고 추천 알고리즘의 원리를 이해하고 개발했을 때 보람 있었습니다.

대리 김경민

빅데이터 프로젝트를 처음 경험해 봤는데 가설을 세우고 분석 하면서 예상과는 다른 결과를 만나게 재미있었습니다.^.^ 다음에도 데이터에서 좋은 인사이트를 발견해보고 싶습니다.

인턴 이희창

데이터를 분석하면서 회사의 주 고객층이 누구인지 알게 되었을 때 예상과 다른 결과가 나와 놀라웠습니다. 가설을 너무 믿지 말아야겠다는 생각이 들었습니다.



Q & A