

카트라이더 유저 특성 별 맵 추천



○ 맵데이터



🍣 수집 기간 설정

평일과 주말의 게임 플레이 특징을 모두 고려하기 위한 최소 단위

일주일 치 매치 데이터 390만 개 분석에 충분하다고 판단





수집 종류 설정

일주일 간의 데이터를 조사 상위 네 개의 매치가 많이 플레이 됨

다른 매치 종류에 비해 클래식 함

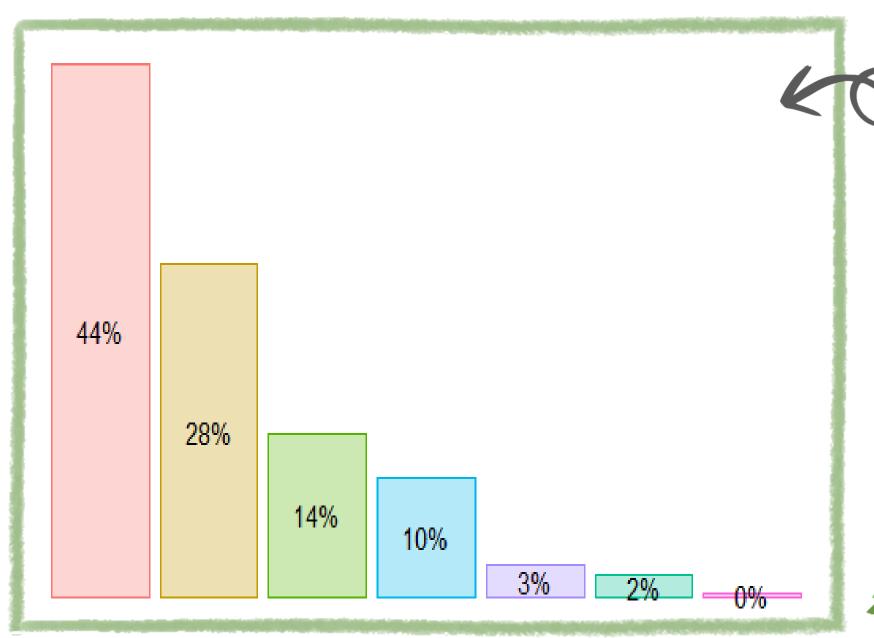
해당 매치에서 **유저들의 플레이 특성**이 잘 드러날 것이라고 생각



○ 매치 데이터

○ 맵데이터





수집 종류 설정

일주일 간의 데이터를 조사 상위 네 개의 매치가 많이 플레이 됨

다른 매치 종류에 비해 클래식 함

해당 매치에서 유저들의 플레이 특성이 잘 드러날 것이라고 생각



○ 맵데이터

○ 매치 데이터



API Key를 통해 NEXON DEVELOPER에 접근 일주일 동안의 매치 고유 코드 불러옴



분석 대상으로 설정한 네 개 <mark>매치 종류만 필터</mark>링

필터링 된 개별 매치 코드에 대해

각각 상세 매치 데이터를 받아온 후 시간 순대로 정렬

DATA

○ 매치 데이터

○ 맵데이터

channel Name	startTime	endTime	gameSpeed		matchRank	match Retired	matchWin	teamNo
speedTeamFastest	2021-10-11 0:00	2021-10-11 0:01	1		99	1	0	1
speedTeamFastest	2021-10-11 0:00	2021-10-11 0:01	1		99	1	1	2
speedIndiInfinit	2021-10-11 0:00	2021-10-11 0:01	4		1	0	1	0
speedIndiInfinit	2021-10-11 0:00	2021-10-11 0:01	4		99	1	0	0
speedIndiCombine	2021-10-11 0:00	2021-10-11 0:01	7		1	0	1	0

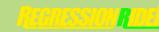


카트라이더 매치 데이터

한 플레이어가 한 게임을 할 때 생성되는 데이터로,

행 하나는 특정 유저의 경기 하나의 기록



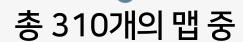


○ 매치 데이터 ○ 맵 데이터

Map Data 1

앞에서 수집한 매치 데이터로부터

맵의 특징을 나타낼 수 있는 변수를 만들어 낼 수 있지 않을까?



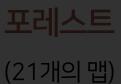
실제 플레이 된 301개 맵



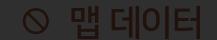
총 285개의 맵을 통해

맵 특징을 나타낼 것이라 기대하는 변수 수집 및 생성













285개 맵들에 대한 카트라이더 플레이 영상 참고하고 플레이해보며 직접 개수를 세 보자!



특징: 난이도에 대한 설명이 있는 경우만 작성



주제선정 DATA EDA 유사도 행렬



○ 맵데이터

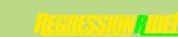


맵의 특징을 대표할 수 있는 추가 4가지 변수 설정









○ 최종 데이터셋

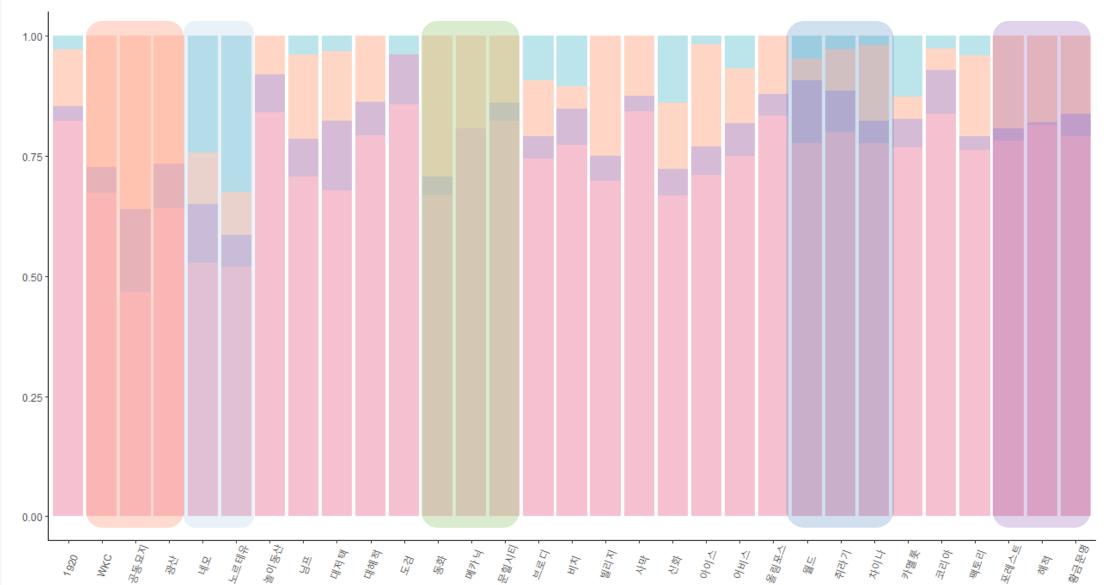
285obs. 26 variables



○ 통계적 방법

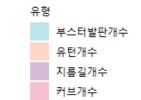
○ 시각화

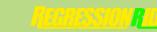




테마별로 각 유형별 비율에 차이 존재

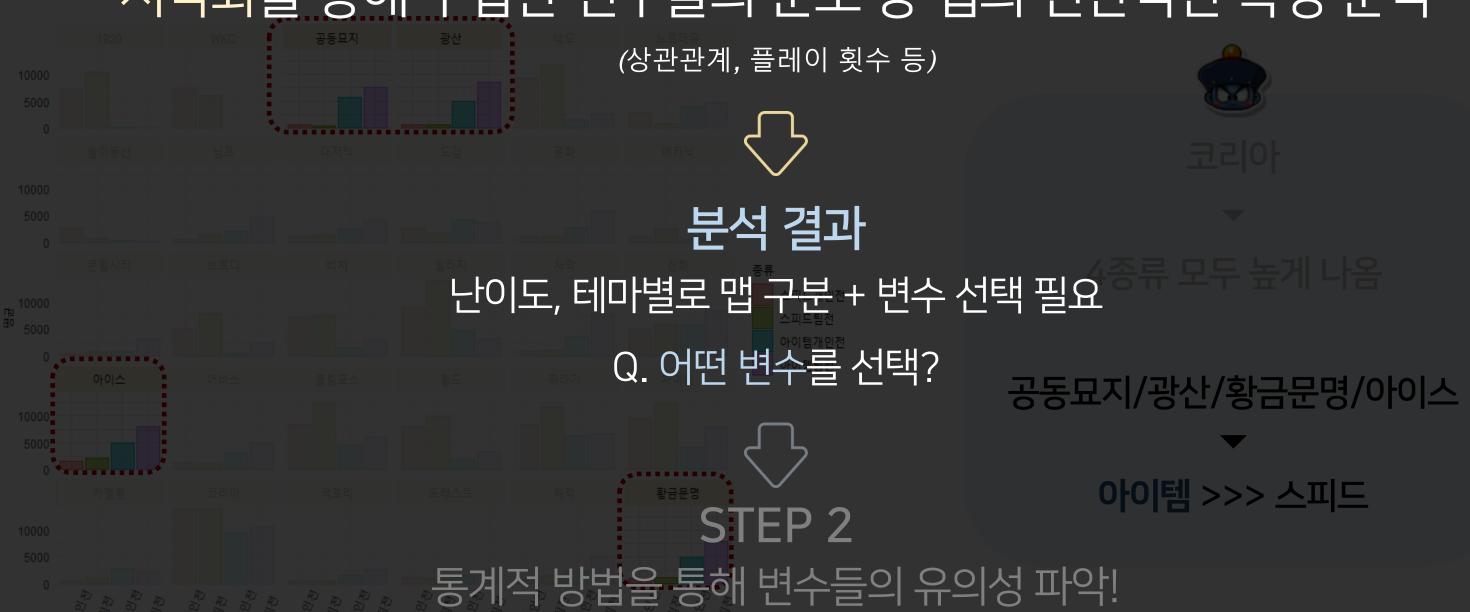
해당 변수가 맵의 특징을 구분 짓는데 의미가 있을 것이라 추측





○ 시각호STEP 됩





○ 시각호STEP 됩

₩ 테마별플레이횟수 평균 시각화를 통해 수집한 변수들의 분포 등 맵의 전반적인 특징 분석

(상관관계, 플레이 횟수 등) 분석 결과

난이도, 테마별로 맵 구분 + 변수 선택 필요

Q. 어떤 변수를 선택?

공동묘지/광산/황금문명/아이스



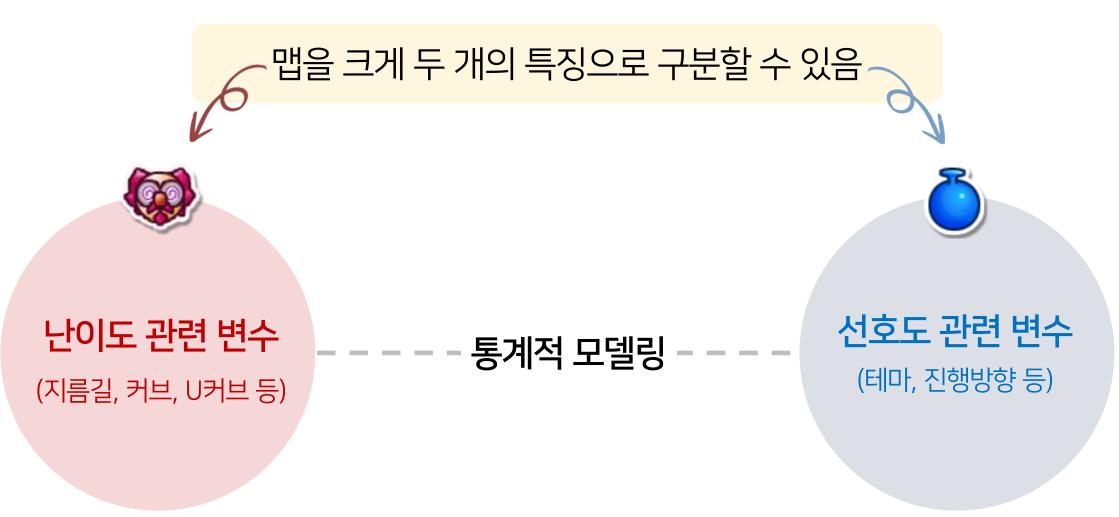
아이템 >>> 스피드

통계적 방법을 통해 변수들의 유의성 파악!



○ 통계적 방법





해당 요소가 맵 난이도를 잘 반영하였는가?

Cumulative logit model 이용

요소가 맵 내 유저 선호도를 잘 반영하였는가?

GAM model 이용



○ 시각화

○ 통계적 방법



₩ 맵 유사도 행렬을 위한 최종 변수

[맵이름	난이도	지름길 개수	커브 개수	유턴 개수	플레이 횟수	평균 리타율	트랙길이	테마
į	유령의 계곡	3	3	6	0	22559	0.27555	7.5	공동묘지
	꼬불꼬불 다운힐	5	2	15	13	4639	0.22804	5.2	광산
į	익스프레스	5	5	34	9	12810	0.21208	8.8	노르테유
GREAT!									
7	ᄼᄼ	3	2	2	7	16458	0.21934	6.0	빌리자



Similarity

벡터로 표현된 두 데이터 간 닮은 정도를 나타내는 지표 상대적인 크기 비교로 유사도 비교가 가능함



① 코사인 유사도

$$sim(A,B)_{cos} = \frac{\langle A,B \rangle}{||A|| \cdot ||B||}$$

② 피어슨 유사도

$$sim(A, B)_{pearson} = \frac{Cov(A, B)}{\sqrt{V(A) \cdot \sqrt{V(B)}}}$$

③ 유클리드 유사도

$$sim(A,B)_{euclidian} = \frac{1}{||A - B|| + \epsilon}$$



Similarity

벡터로 표현된 두 데이터 간 닮은 정도를 나타내는 지표 상대적인 크기 비교로 유사도 비교가 가능함



유사도를 구할 맵 데이터의 경우

각각의 특징 변수들의

| 지름길 개수, 커브 개수, 난이도, 트랙 길이 등 |

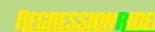
크기가 반드시 고려되어야 하므로

유클리드 유사도를 사용하여 맵 유사도 행렬 계산!

③ 유클리드 유사도

$$sim(A,B)_{euclidian} = \frac{1}{||A - B|| + \epsilon}$$

벡터 원소 크기 고려 O



유사도 행렬



※ 최종 맵 유사도 행렬 | 테마 고려 X

범주형 변수는 OneHot encoding을 거친 후 변수들끼리 Min-Max scaling적용 후 유사도 계산

	공동묘지유령의견	呼R S ^{と 上} は が と と と と と と と と と と と と と と と と と と	R 박 버려진 오아	从 <u> </u>	이나서안병마용	
공동묘지 유령의 계곡 R	0	2.405	1.216	1.191	0.761	
광산 보석채굴장 R	2.405	0	1.603	1.503	0.851	
사막 버려진 오아시스	1.216	1.603	0	5.957	0.742	
포레스트 골짜기	1.191	1.503	5.957	0	0.733	
:						

● 관심 유저 정의 ▶



관심 유저 기준 선정

관심 유저로 선별될 유저들은

개인의 특징을 나타내는 최소한의 플레이 기록을 남겨야 함

① 게임 플레이 시간

일정 시간 이상 플레이 한 유저는 카트라이더에 애정을 갖고 활동하는 유저

② 게임 플레이 횟수

단순히 접속만 하는 것이 아닌 직접 게임 플레이도 하고 플레이 횟수가 충분한 유저

● 유저 특징 변수 ▶ ● 유저의 선호도 ▶



도전 지수란?

유저의 플레이 성향이

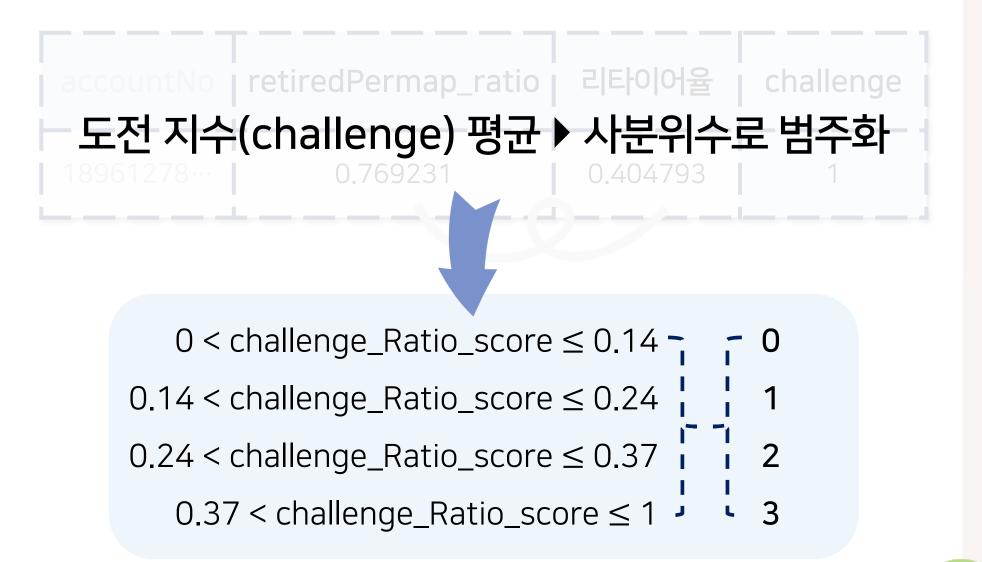
도전적인지를 판단하는 지수



유저의 맵별 리타이어율 & 맵 평균 리타이어율 이용

유저의 리타이어율 > 맵 리타이어율

도전적이라고 판단



● 유저 특징 변수 ▶ ● 유저의 선호도 ▶



accountNo	선호시간대_1	선호시간대_2	라이센스	선호 인원 수
1006691919	주중_야간	없음	3	L
1376244258	주중_오후	주말_야간	5	М

스피드	도전정도	아이템스피드 선호	팀개인 선호	레벨	
일반	3	아이템	팀전	gloveG	REAT!!
무한부스터	1	스피드	개인전	meda	

31890명의 관심 유저의 특징을 나타내는 총 10개의 변수

31890 X 10

●유저 특징 변수 ▶ ●유저의 선호도 ▶



카트라이더의 랜덤 맵도 **유저의 선호도**를 반영할 수 있나요? 협업 필터링 알고리즘 활용을 위해

맵 평점 데이터에 기반한 유저의 맵 선호도 측정 필요

랜덤 맵이라도 몇몇 개의 특정한 맵에서 랜덤으로 나온다는 점에서

유저가 해당 맵의 플레이 여부를 결정하는 과정 자체가

카트라이<mark>선호도를빞영</mark>한타고(볼-수)있음이지 않음

유저 별 맵 플레이 판수로 선호도에 해당하는 지표를 대신!

추천 알고리즘

추천 기준 설정

매치 데이터로부터 추천기준 설정

추천기준에 따른 유저 프로필 구축 유저 필터링

추천 받고 싶은 유저와 유사한 특성을 가진 유저 필터링

> 가장 유사한 유저 1명 추출

선호도 기반 맵 필터링

유저-맵 플레이횟수 파악

앞 단계에서 추출된 유저 1명이 가장 선호하는 맵 파악 최종 추천

맵-맵 유사도를 통해 해당 맵과 유사한 맵 N개 추천



추천 시스템

● 추천 시스템 방법 ▶ ● 최종 추천 알고리즘 ▶

추천 알고리즘

추천 기준 설정

매치 데이터로부터 추천기준 설정

추천기준에 따른 유저 프로필 구축 유저 필터링

유사도 기반 필터링

선호도 기반 맵 필터링

협업 필터링 최종 추천

콘텐츠 기반 필터링



● 추천 시스템 방법 ▶ ● 최종 추천 알고리즘 ▶



유저-아이템 행렬에 잠재요인이 있다고 가정

* 잠재요인 : 유저가 해당 아이템을 선택하게 한 요인

유저가 해당 맵을 플레이하는데 영향을 미친 잠재요인이 존재함

∴ 해당 잠재요인을 추천에 활용





● 추천 시스템 방법 ▶ ● 최종 추천 알고리즘 ▶

우리의 유저-맵 데이터는 NA값이 다수 존재 (미플레이)

	포레스트	공 동 묘지	아이스	빌리지	노르테유	광산	브로디
유저 1	1 3	NA NA	2	4	NA I	1 1	NA
유저 2	2	1	NA	1	4	NA	18
유저 3	5	13	NA	NA	NA	3	NA
유저 4	7	21	5	1	2	2	NA

SVD를 이용하려면 NA값이 존재하면 안됨



Gradient Descent와 Alternating Least Squares 방법을

모두 적용해 예측 성능 비교!



● 추천 시스템 방법 ▶ ● 최종 추천 알고리즘 ▶



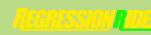
```
regularization = [0.1, 0.3, 0.5, 1, 3, 5, 10, 25, 40, 50, 75]
factor = list(range(10,151,10)) #튜닝할 factor 값 넣기 알아서 여기만 바꿔도 됨.
#그냥 그리드 서치할 데이터 프레임 만들어 주는 함수임.
import itertools
def expandgrid(*itrs):
    product = list(itertools.product(*itrs))
    return {'Yar{}'.format(i+1):[x[i] for x in product] for i in range(len(itrs))}
expandgrid(regularization, factor)
expand_df = pd.DataFrame.from_dict(expandgrid(regularization, factor))
expand_df.columns =['regularization', 'factor']
expand_df['rmse'] = np.nan
expand_df
```

Grid Search

그리드 서치를 이용하여 하이퍼 파라미터 튜닝

파라미터 목록은 여러 예시와 논문을 토대로 대체로 결과가 좋았던 파라미터들로 리스트를 구성

Performance Evaluation of Latent Factor Models for Rating Prediction (Lan Zheng, 2015) Collaborative Filtering for Implicit Feedback Datasets (Yifan Hu, 2008)



추천 시스템

협업 필터링과 콘텐츠 기반 필터링을 결합한

하이브리드 추천 시스템

① 취향이 편향된 아이템만을 추천해 준다는

최종 맵 추천 (10개)

콘텐츠 기반 필터링의 단점 극복

~ 저희가 추천할 만한 맵은 다음과 같습니다 ~

포레스트 지그재그 | 노르테유 !

노르테유 부스② 협업 필터링에서 발생하는

아이스 하프파이프

빌리이d start문제를 해결할 수 있음 전의 선택한 후 비에 출력

노르테유 익스프레스

Cold start 한? 유서에 대한 충분한 정보가 수집된 상태가 아니어서

WKC 코리아 서킷

해당 유서들에게 적절한 제품을 추천해주지 못하는 문제 수 지 오늘로 내

문힐시티 터널 속 무법자

차이나 상해 동방명주

어디까지나 예시

◀ 시뮬레이션 ▶

카트라이더 유저 맵 추천

카트라이더를 해보셨나요?

 \circ \sqcup

이 아니오

도전적인 정도를 표시해주세요 (3이 가장 도전적)

 \circ

O 1 O 2

○ 3

팀전과 개인전 중 뭘 더 선호하시나요?

○ 팀전

○ 개인전

아이템전과 스피드전 중 뭘 더 선호하시나요?

○ 아이템전

○ 스피드전

속도감은 어느 정도를 원하시나요?

○ 보통

다음

카트라이더 유저 맵 추천

카트라이더 플레이 시 사람이 많은 것과 적은 것 중 어느 것을 선호하시나요?

○ 적음 (1명~4명)

○ 보통(5명~6명)

○ 많음 (7명~8명)

주로 플레이 하는 시간대는 언제인가요?

○ 주말 오전(7시-11시) ○ 주중 오전(7시-11시)

○ 주말 점심(12시-13시 ○ 주중 점심(12시-13시)

○ 주말 오후(14시-22시) ○ 주중 오후(14시-22시)

○ 주말 야간(22시-06시) ○ 주중 야간(22시-06시)

그 다음으로 플레이 하는 시간대는 언제인가요?

○ 주말 오전(7시-11시) ○ 주중 오전(7시-11시)

○ 주말 점심(12시-13시 ○ 주중 점심(12시-13시)

○ 주말 오후(14시-22시) ○ 주중 오후(14시-22시)

○ 주말 야간(22시-06시) ○ 주중 야간(22시-06시)

이전

다음

카트라이더 유저 맵 추천

플레이 경험이 있다면 라이센스가 무엇인가요?

○ 초보 ○ L3 ○ 루키○ L2

O L1

○ 없음/모름

O PRO

플레이 경험이 있다면 레벨이 무엇인가요?

○ 장갑

○ 별장갑 ○ 메달

○ 별

○ 만렙

○ 없음/모름

카트라이더 맵에는 맵의 분위기를 정하는 '테마'가 존재합니다 맵을 고를 때 테마를 중요시 하나요?

O 네

○ 아니오

이전

추천 시작