

Data Modeling Project

PUBG Finish Placement Prediction

PLAYERUNKNOWN'S
BATTLEGROUNDS
NOT JUST A GAME, THIS IS BATTLE ROYALE



30조
팀원: 최호준, 오의택

PLAY NOW



Data Description

게임 진행 방식

우승 조건

기본 사항



Feature Engineering

핵 유저 선별

경기 진행 시간 관련(맵)

Feature 결합



EDA

데이터 확인 및 전처리

데이터 별 상관 관계

킬 관련.아이템 관련.이동거리,
etc(차량파괴.로드킬 팀킬) = 0



Modeling

샘플링 후 모델 선정

XGboost,LightGBM

DicisionTree, Random Forest

목차

Data Description

기본 사항

1. 이동은 필수
2. 생존이 우승조건
3. 최대 4명이 한팀



게임 진행 방식

수송기에서 자유낙하후
낙하산 펼침

생존을 위한 무기획득 및
방어구 획득

자기장을 피해 맵의
특정구역으로 이동

마지막 최후 1인까지
살아남기

STEP 1

STEP 2

STEP 3

STEP 4

Data Description

데이터 컬럼 설명

- Datafield	
- DBNOs(Down But Not Out)	적을 기절시켰으나, 적 팀원에 의해 부활되어 실제로는 킬로 처리되지 않은 횟수
- assists	아군이 킬했을 때 같이 데미지를 넣은 수 (말 그대로 어시스트)
- boosts	부스트 아이템 사용 횟수
- damageDealt	총 넣은 데미지 (단, 팀에 준 피해나 자해는 포함되지 않음)
- headshotKills	헤드샷 킬 수
- heals	힐링 아이템 사용 횟수
- Id	플레이어 ID
- killPlace	매치에서의 킬수만 따졌을때 순위
- killPoints	킬 수로 산정한 ELO의 개념. rankpoints가 없다면 killPoints가 존재
- killStreaks	짧은 시간동안 최대한 많이 킬했을 때의 킬 수 / 죽지않고 킬을 했을때
- kills	총 킬 수
- longestKill	킬과 킬 사이의 이동한 가장 긴 거리
- matchDuration	매치의 시간 (초 단위) -> 맵 사이즈에 따른 시간이 다르다 (평균 : 30분, 23분)
- matchId	매치 ID (트레이닝 데이터와 테스트 데이터에 동일한 매치 ID는 없음)
- matchType	매치 타입 "solo", "duo", "squad", "solo-fpp", "duo-fpp", and "squad-fpp"; 나머지는 이벤트 또는 커스텀 매치

Data Description

데이터 컬럼 설명

- rankPoints	랭킹 ELO의 개념
- revives	아군 부활 횟수
- rideDistance	탈것으로 이동한 총 거리
- roadKills	탈것으로 킬한 횟수
- swimDistance	수영한 총 거리
- teamKills	팀킬 횟수
- vehicleDestroys	차량 폭파 횟수
- walkDistance	걸어간 총 거리 total : 3 map -> map 2개 : 8 X 8 km => 최대 대각선 길이 : 11.3km map 1개 : 4 X 4 km => 최대 대각선 길이 : 5.6km
- weaponsAcquired	얻은 무기의 수
- winPoints	승리 횟수로 판단한 ELO의 개념
- groupId	매치 내 그룹 아이디, 동일 그룹 구성원들이 다른 게임을 할 때마다 새로운 그룹 아이디가 부여됨
- numGroups	게임 내 총 그룹 수 (솔로 -> 듀오 -> 스쿼드로 갈 수록 줄어듦)
- maxPlace	한 경기내 최하 등수()
- winPlacePerc	예측 목표, 1에 가까워질 수록 순위가 높아진다. (0부터 1까지의 값으로 나타냄. 1은 1등을, 0은 꼴등을 의미한다.) $(\text{maxPlace} - \text{winPlace}) / (\text{maxPlace} - 1) = \text{winPlacePerc}$ 즉, 경기에 들어온 팀수에 따라 경기 등수가 결정된다.

Feature Engineering

4 X 4 (km)



사냥
평균 플레이타임 23분

8 X 8 (km)



미라마, 비켄디
평균 플레이타임 31분



전처리

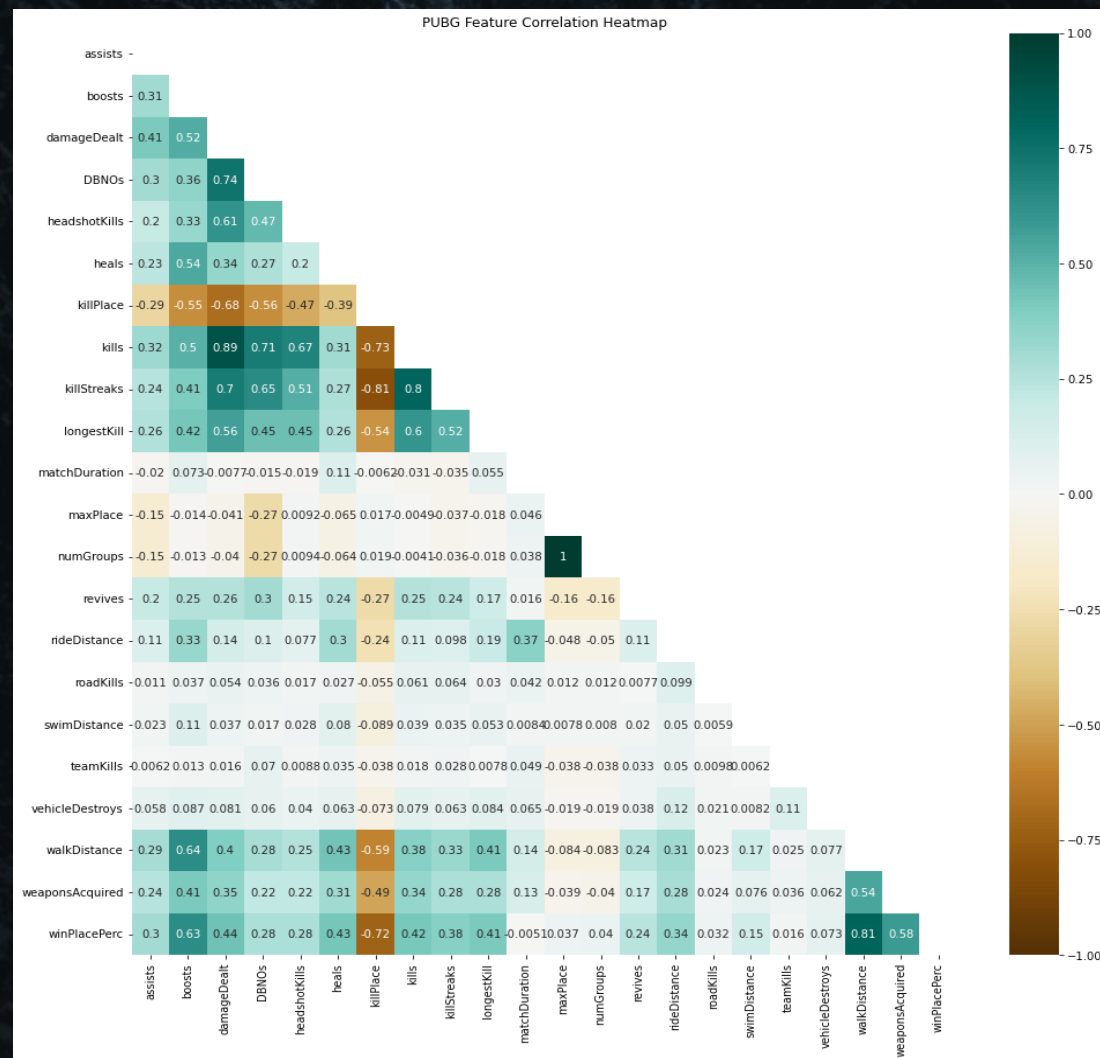
NaN 값 확인

```
df[df['winPlacePerc'].isna()]
```

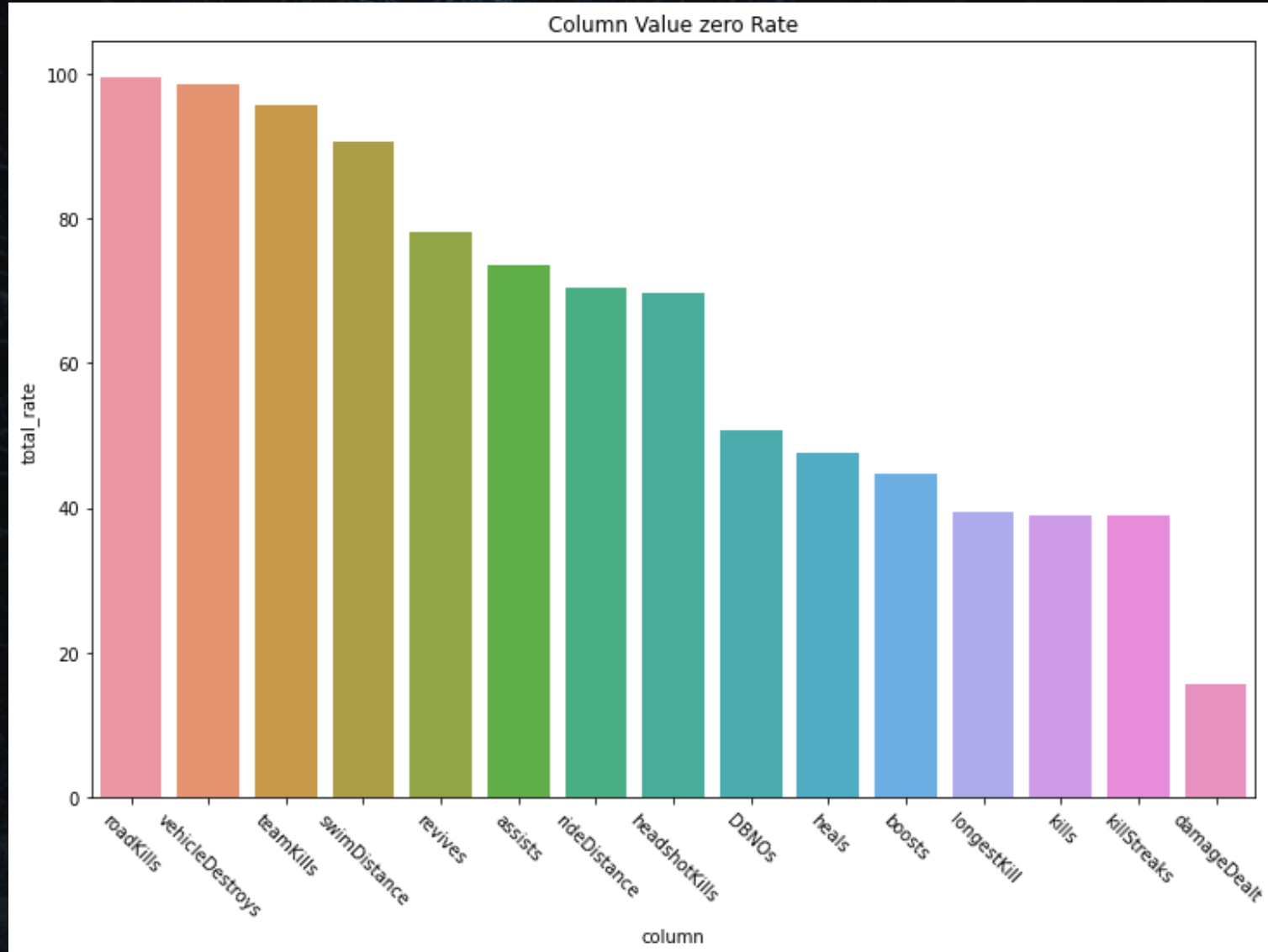
NaN값이 들어있는 경기 제외

```
df.drop(index=df[df['matchId']== '224a123c53e008'].index, inplace=True)
```

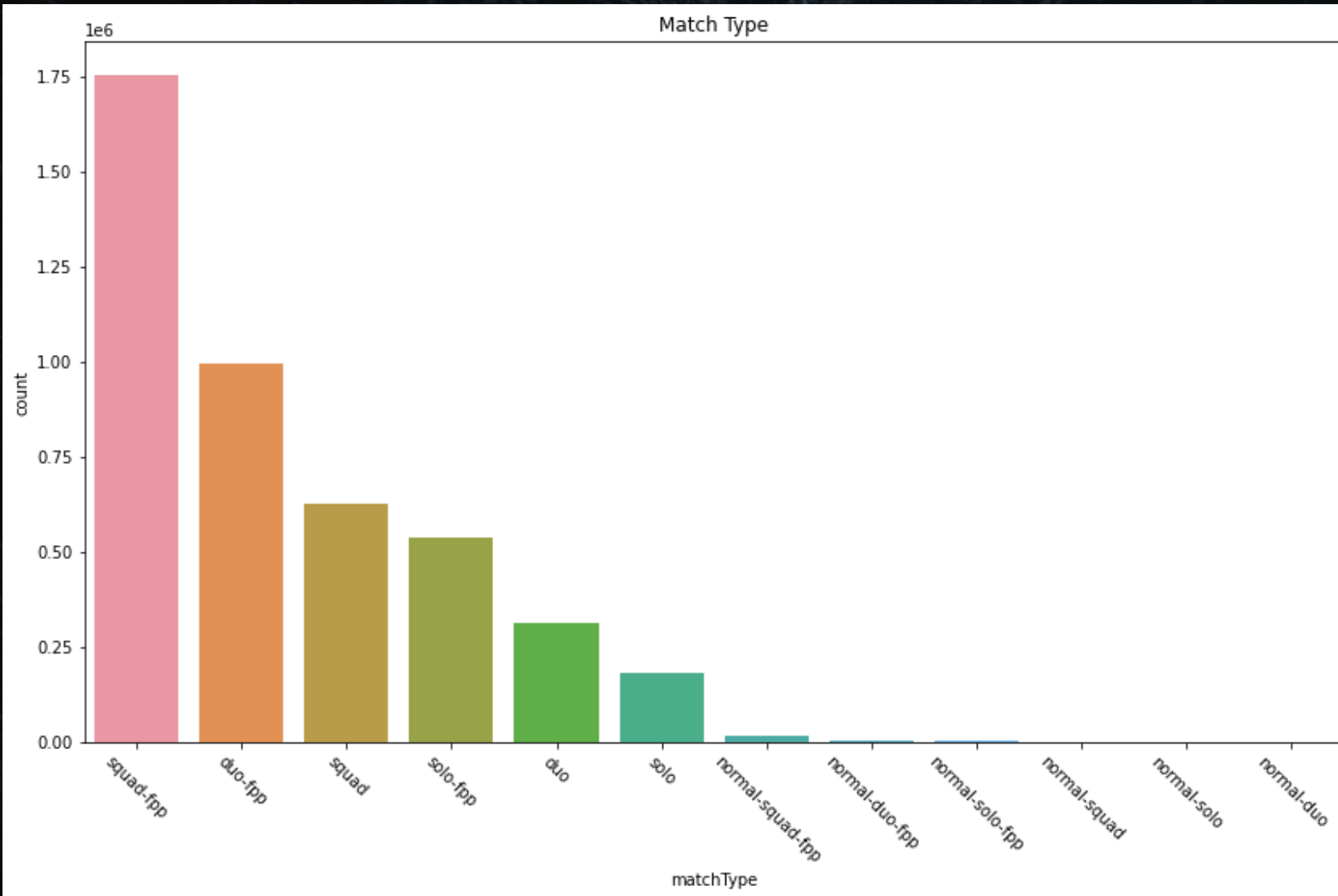
데이터 별 상관 관계



컬럼 확인(value = 0%)



Match Type 확인



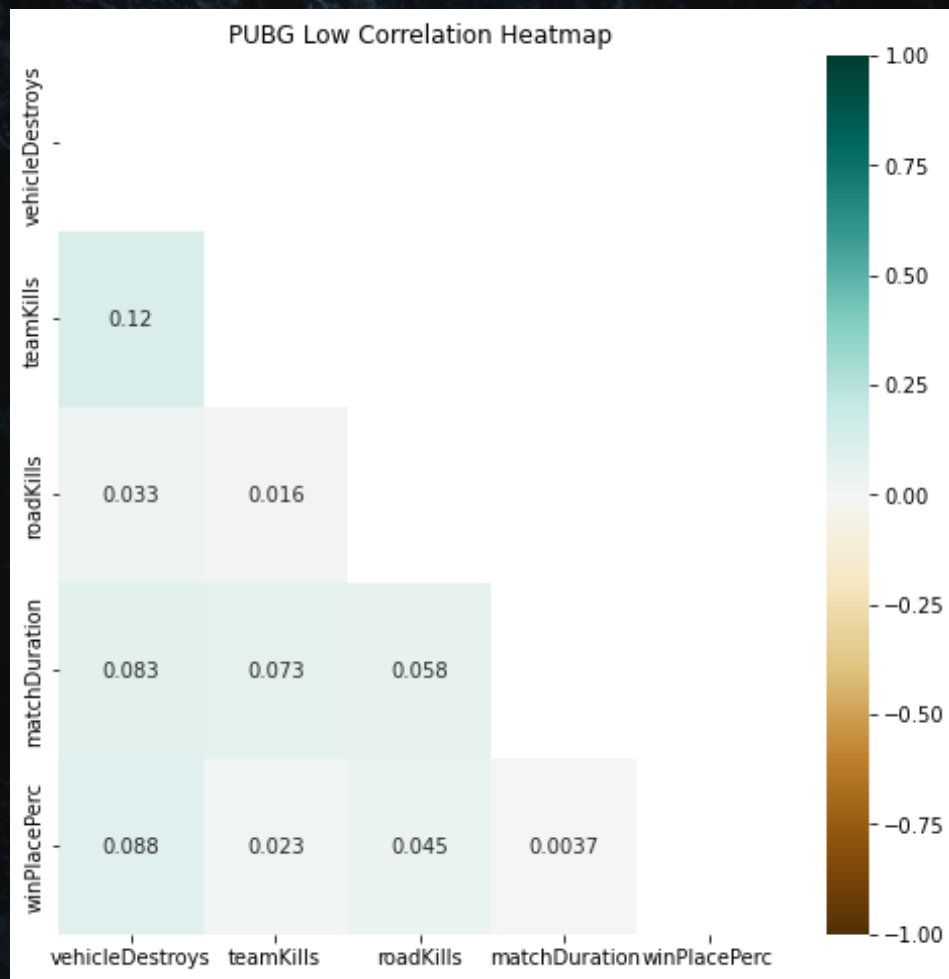
Normal type

수량 아주 적음 확인

전문가 의견

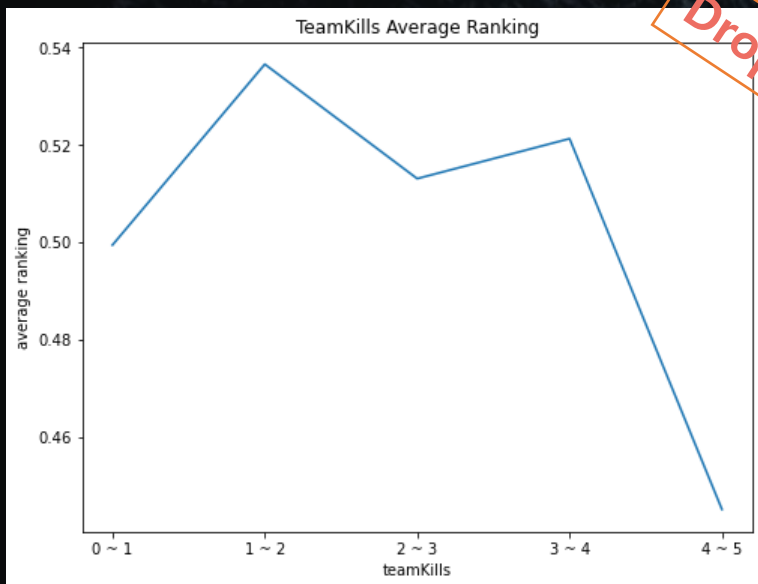
해당 게임 type은
핵유저가 다수 포함
되어있음

상관 관계가 낮은 컬럼



상관 관계가 낮은 컬럼

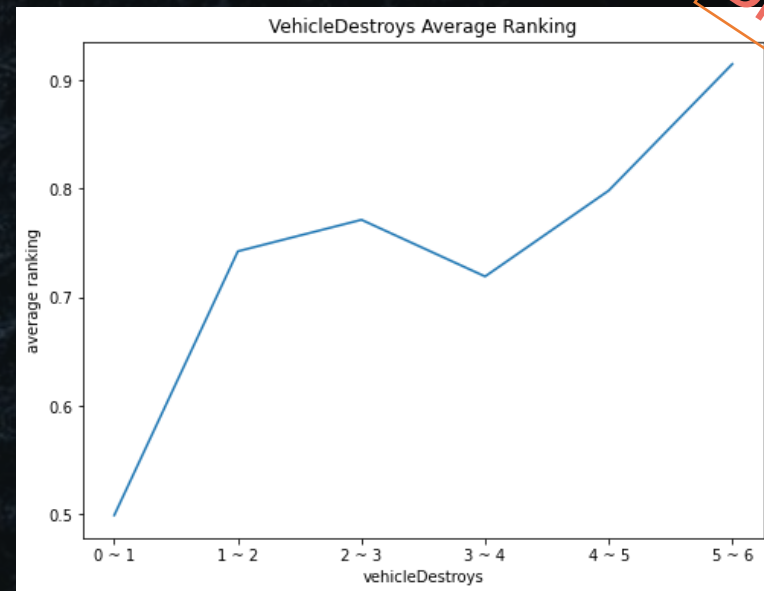
팀킬



teamKills	average ranking
0 ~ 1	0.499362
1 ~ 2	0.536485
2 ~ 3	0.513003
3 ~ 4	0.521205
4 ~ 5	0.445100

95% : 0

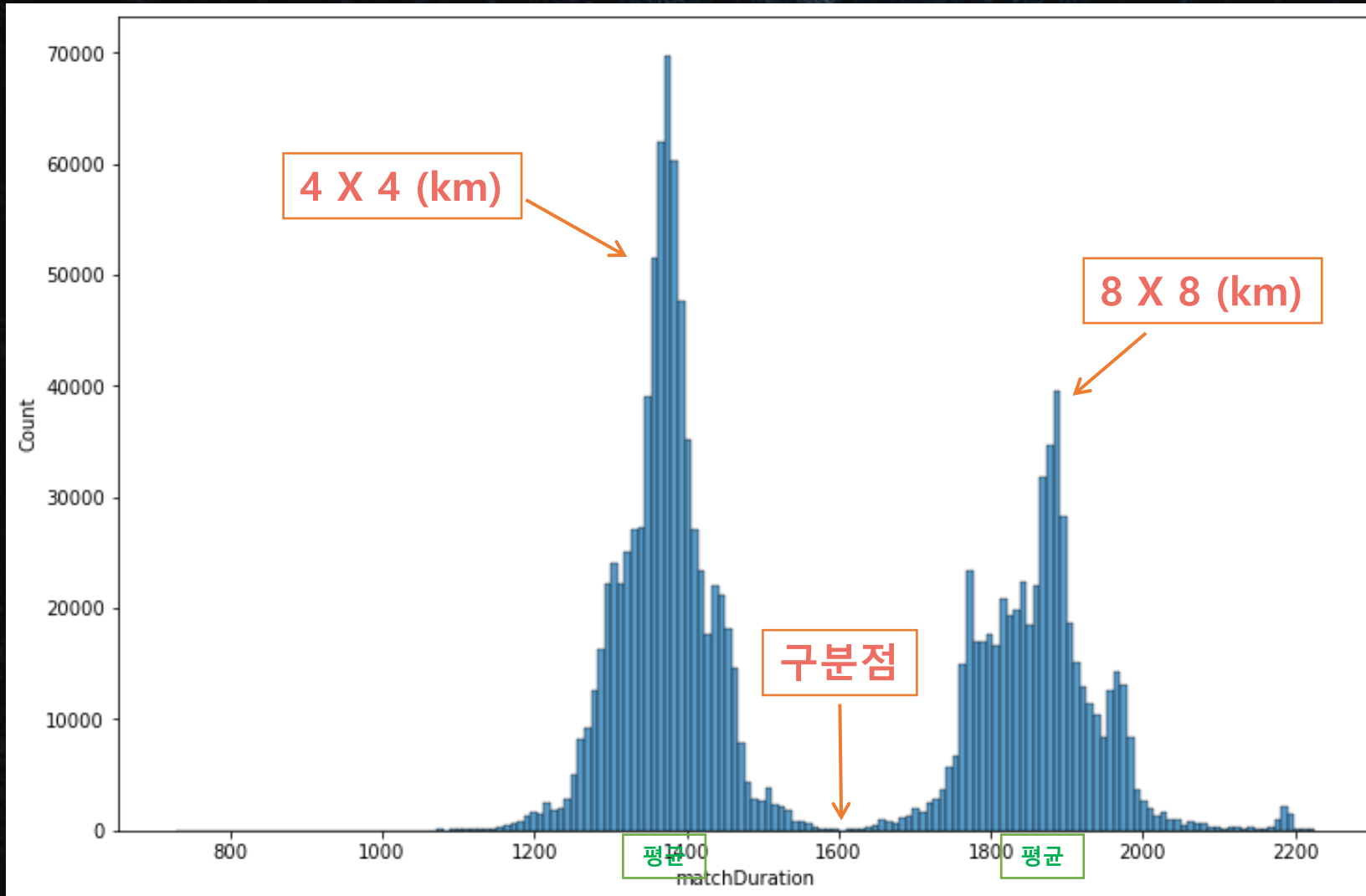
차량파괴



vehicleDestroys	average ranking
0 ~ 1	0.499124
1 ~ 2	0.742355
2 ~ 3	0.771389
3 ~ 4	0.719190
4 ~ 5	0.798300
5 ~ 6	0.91490

98% : 0

Match Duration



4 X 4 (km)

평균 게임시간 : 23분

8 X 8 (km)

평균 게임시간 : 31분

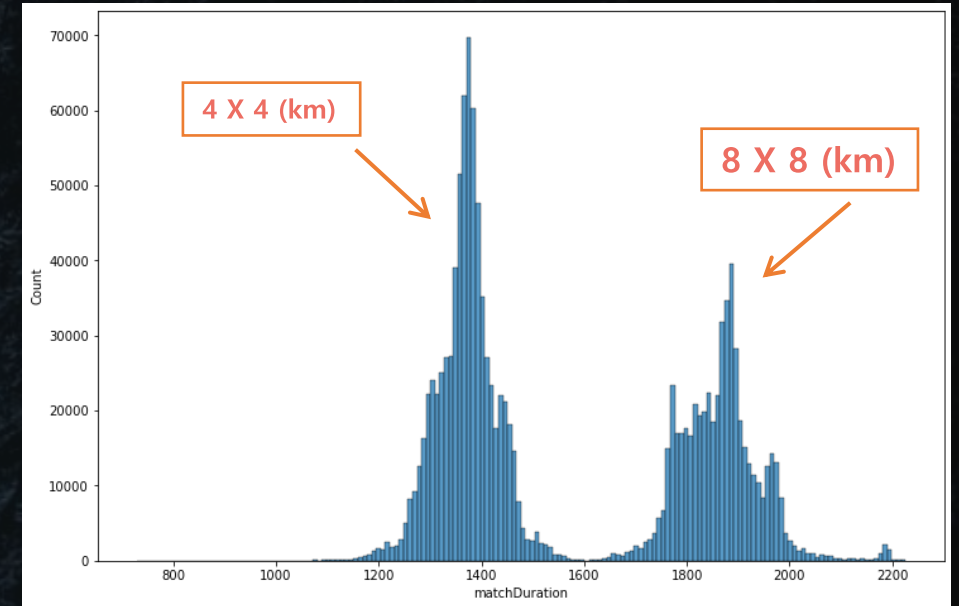
맵 확인



4 X 4 (km)



8 X 8 (km)



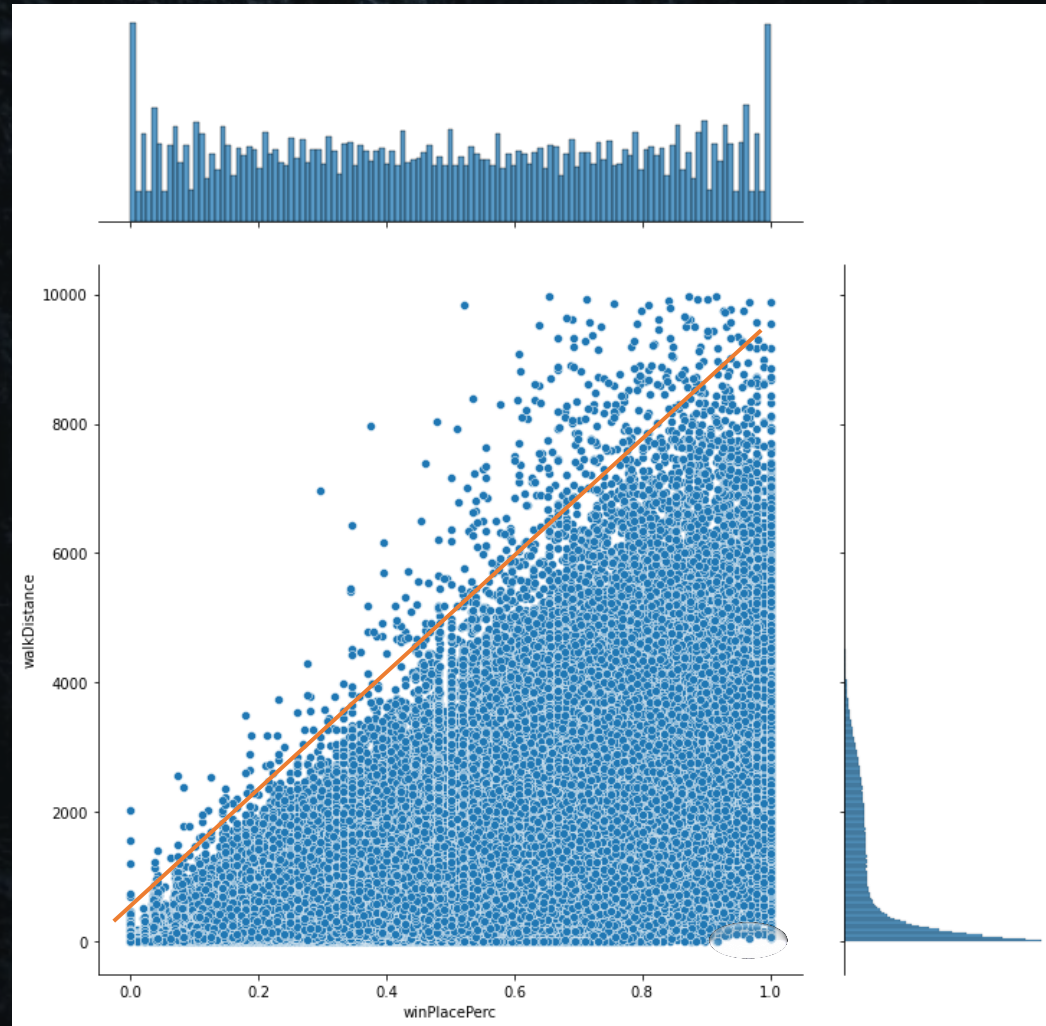
4 X 4 (km)

평균 게임시간 : 23분

8 X 8 (km)

평균 게임시간 : 31분

Walk Distance

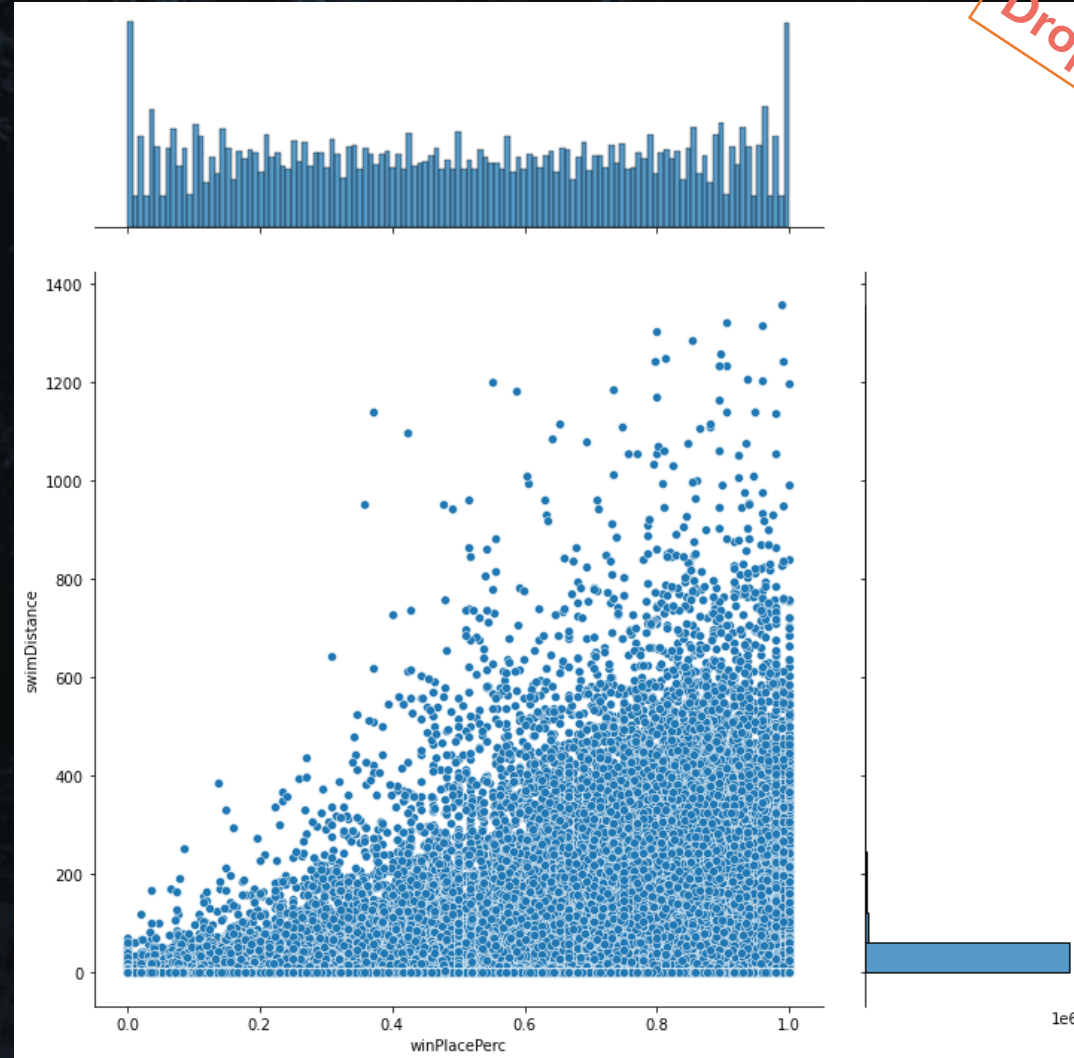


높은 상관관계 확인

EDA

Swim Distance

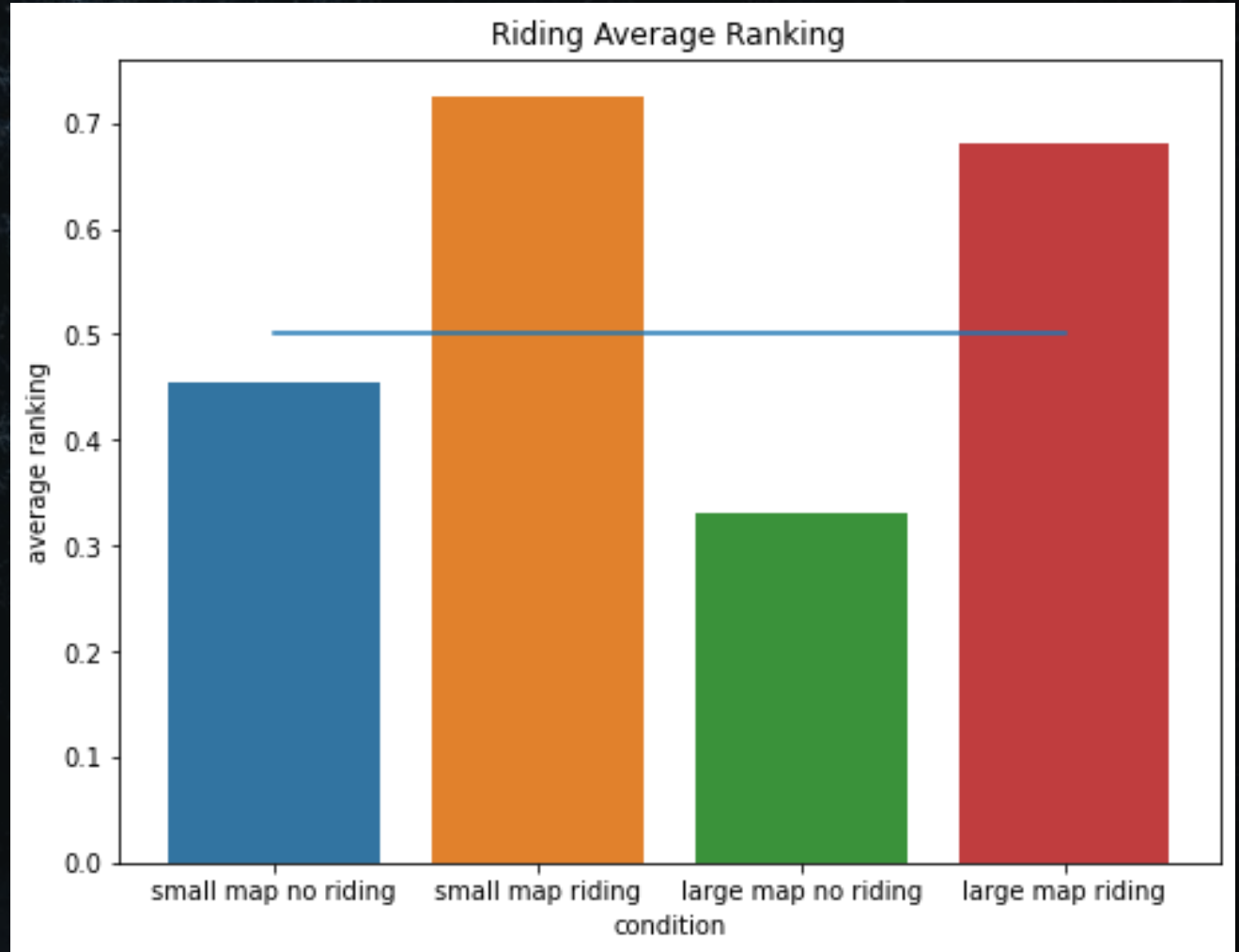
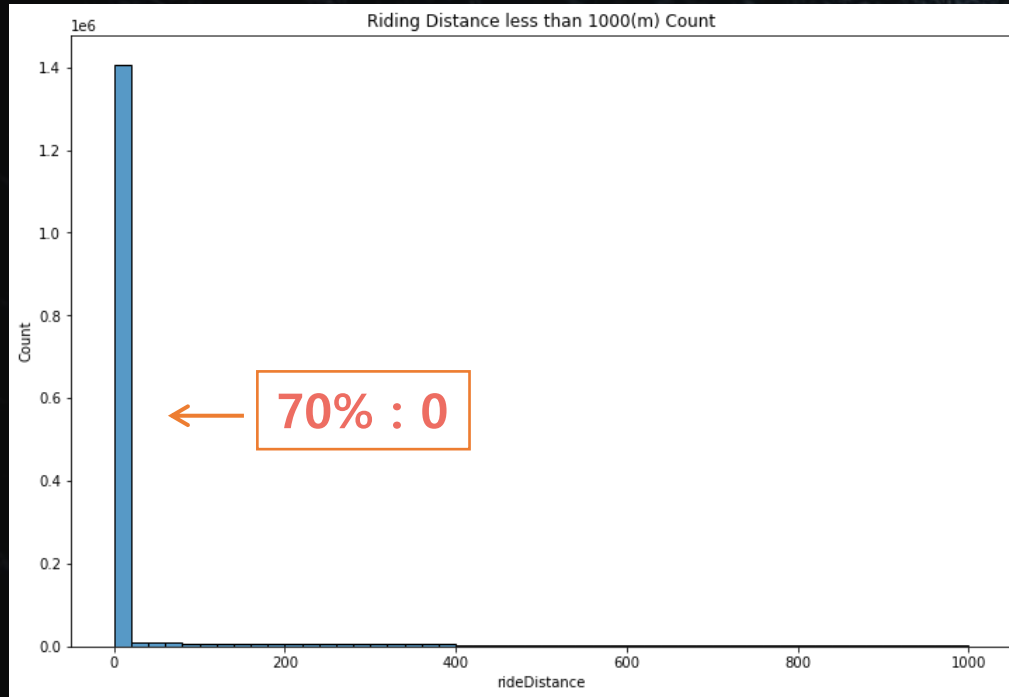
수영(X) 선수 : 승률 : 0.475
수영(O) 선수 : 승률 : 0.73



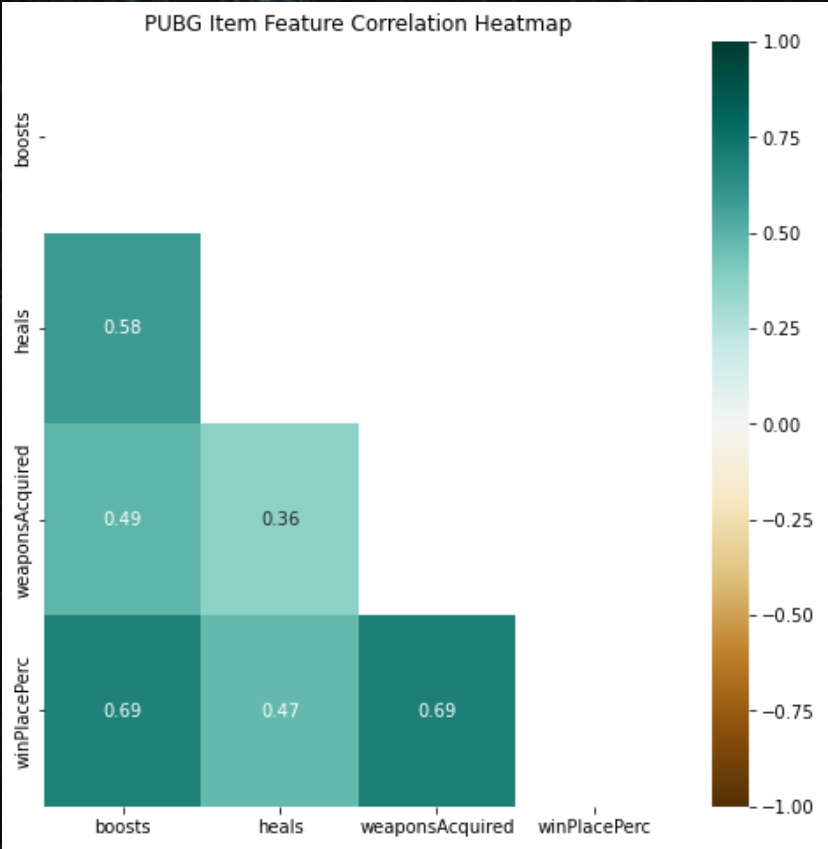
← 90% : 0

EDA

Ride Distance



Heal and Boost



	boosts	heals	weaponsAcquired	winPlacePerc
boosts	1.000000	0.579211	0.491401	0.686819
heals	0.579211	1.000000	0.359407	0.471152
weaponsAcquired	0.491401	0.359407	1.000000	0.688473
winPlacePerc	0.686819	0.471152	0.688473	1.000000

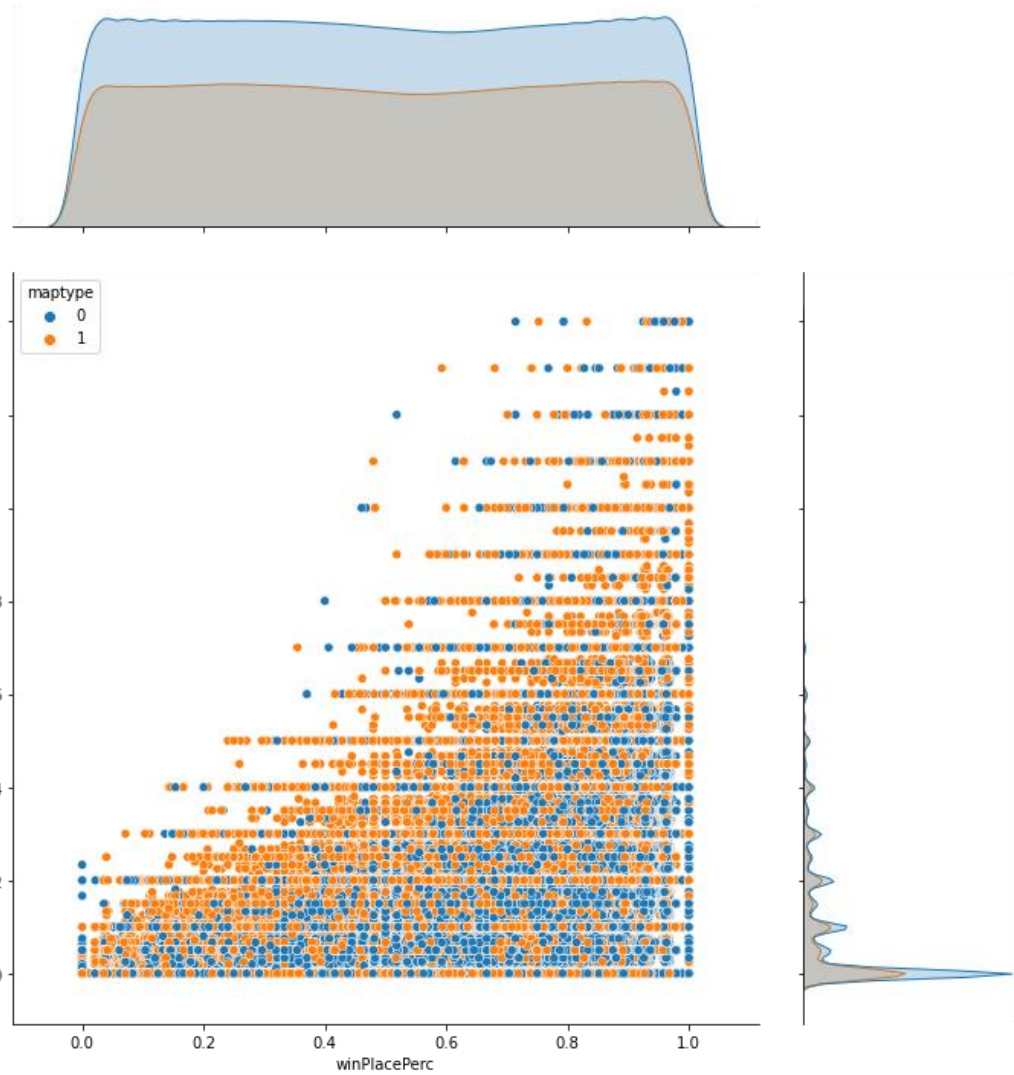
Boost item

작은 맵에서
더 많은 아이템
획득



전문가 의견

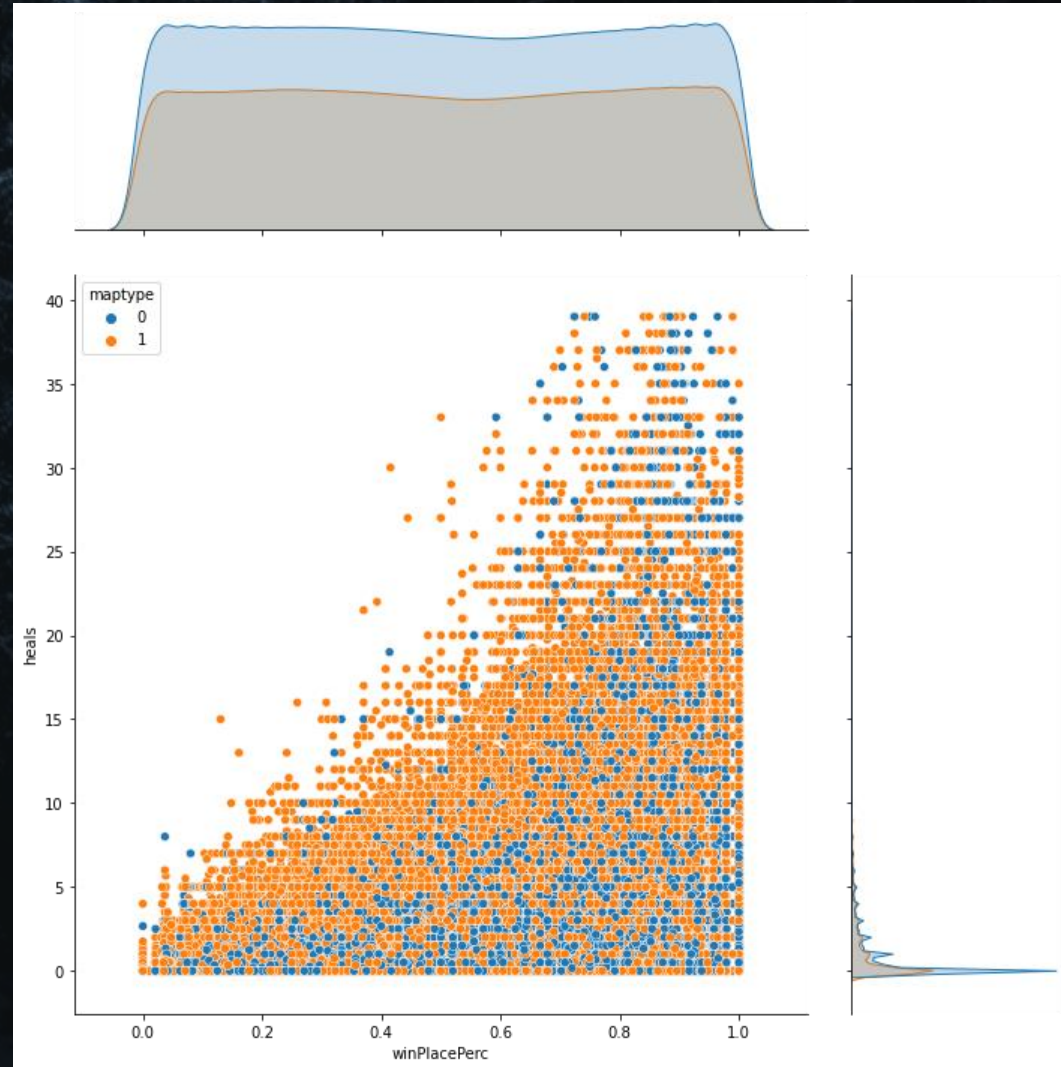
작은 맵에서
더 많은
아이템이
드랍되었음
-> 유저들이
큰 맵보다는
작은 맵을
더 선호



EDA

Heal item

Heal 사용(X) : 승률 : 0.33
Heal 사용(O) : 승률 : 0.654



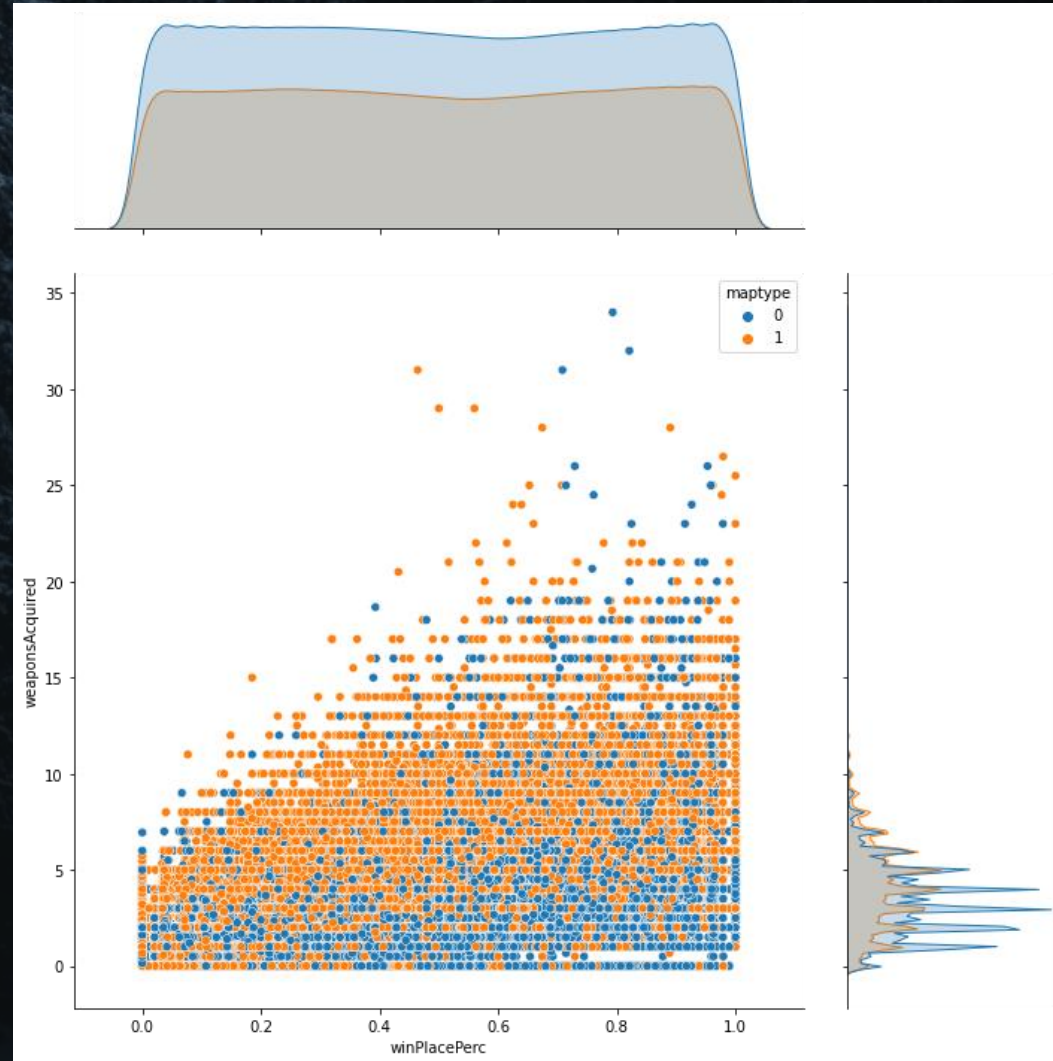
EDA

Weapon Acquired

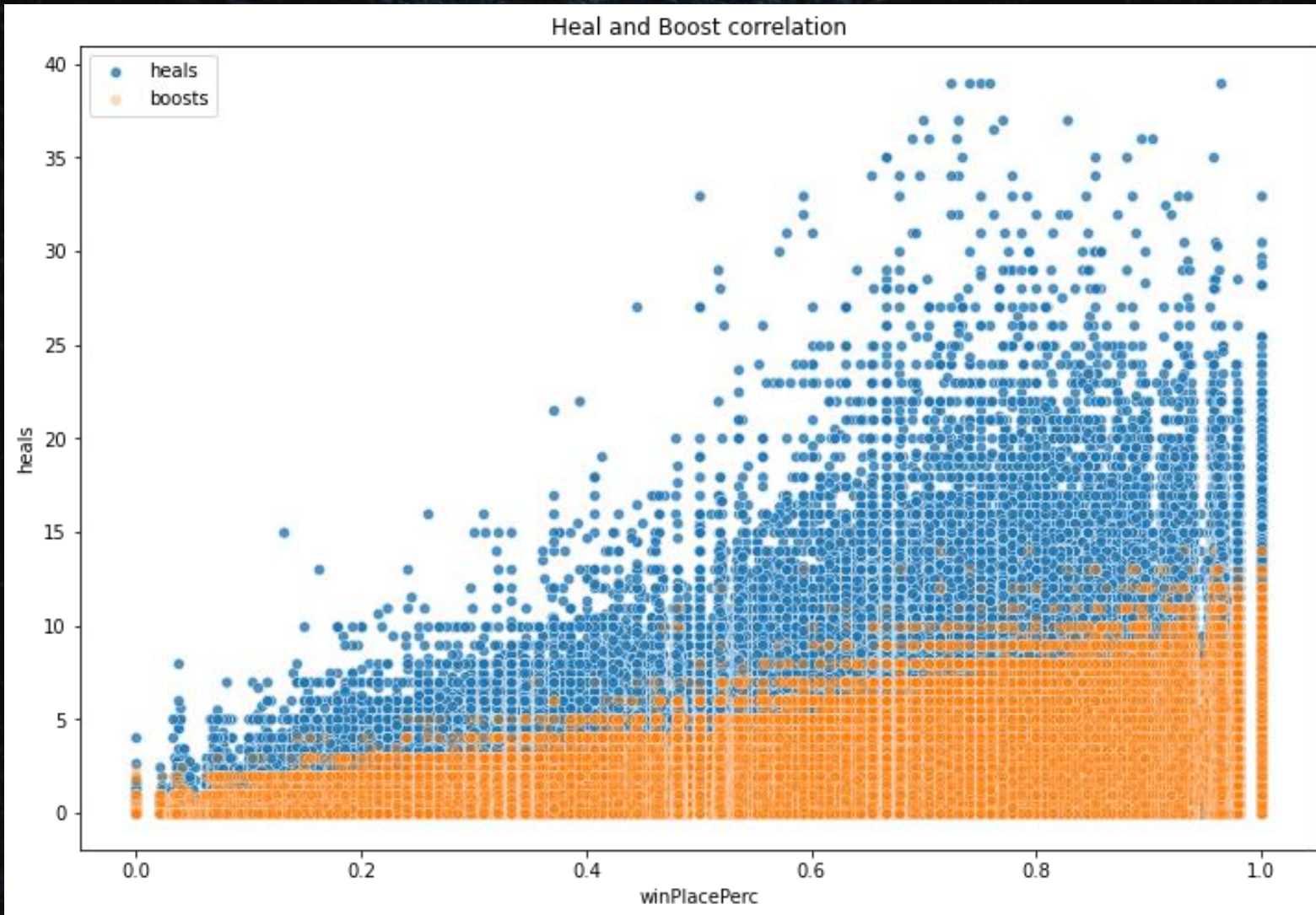
Weapon 사용(X) : 승률 : 0.1
Weapon 사용(O) : 승률 : 0.508



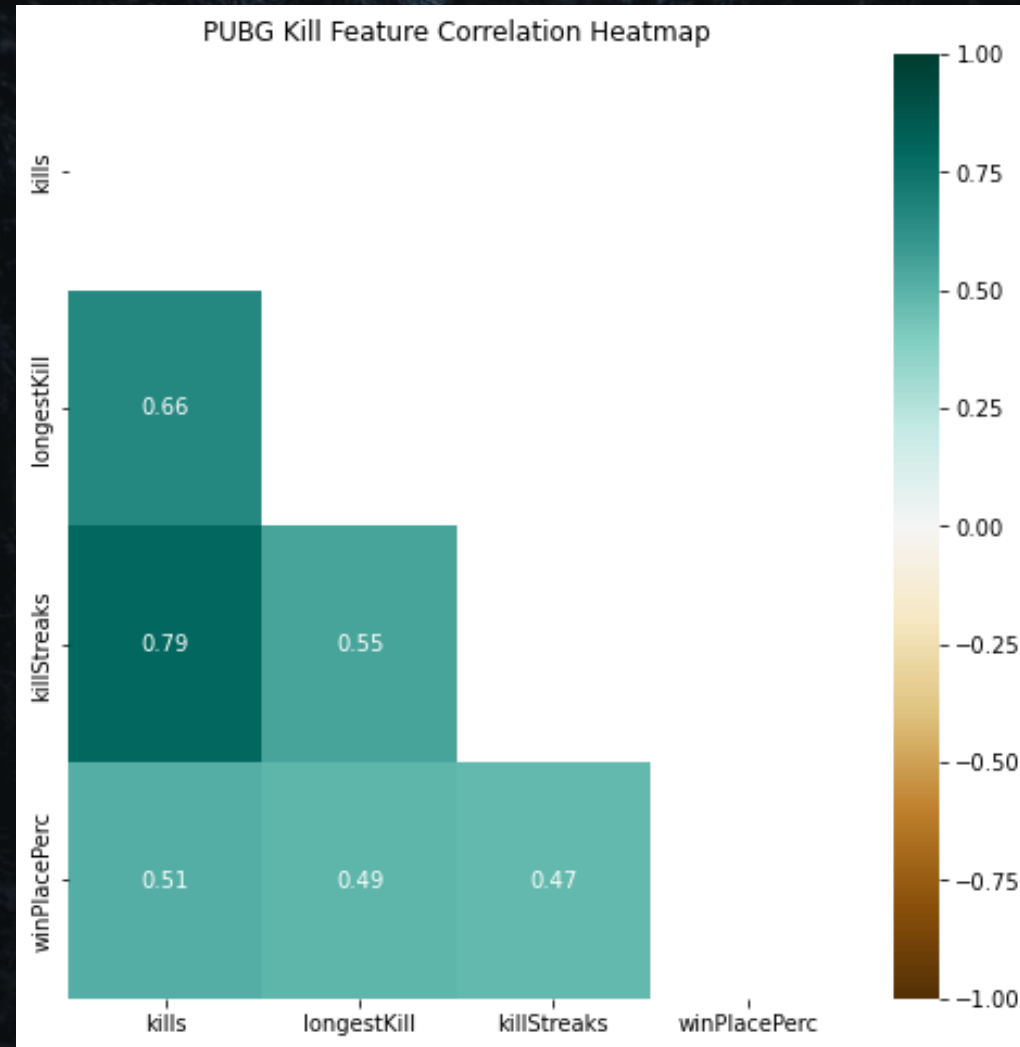
무기 사용시 평균적 성적
하지만
무기 획득에 실패하면
최하등급 예상



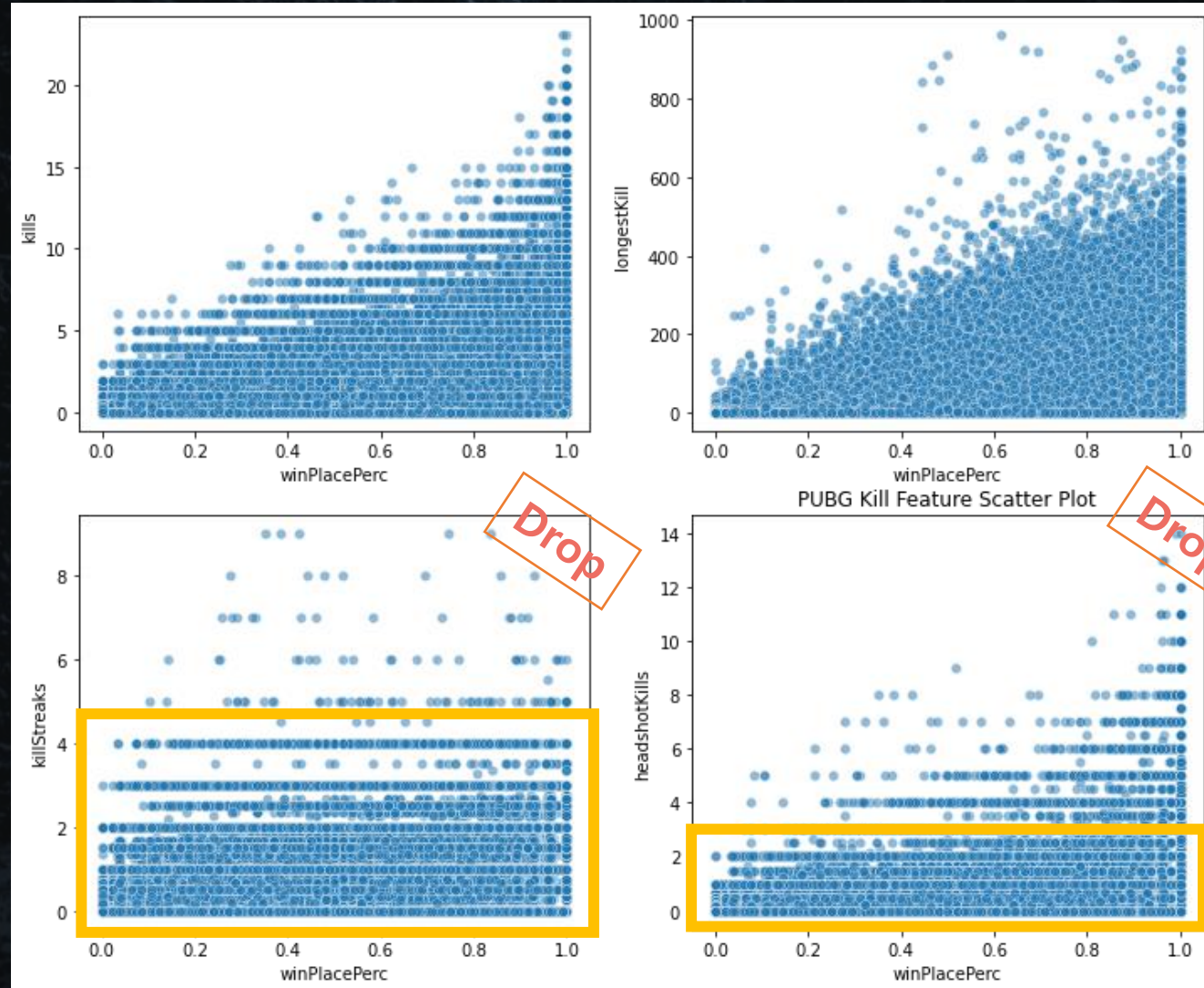
Heal and Boost



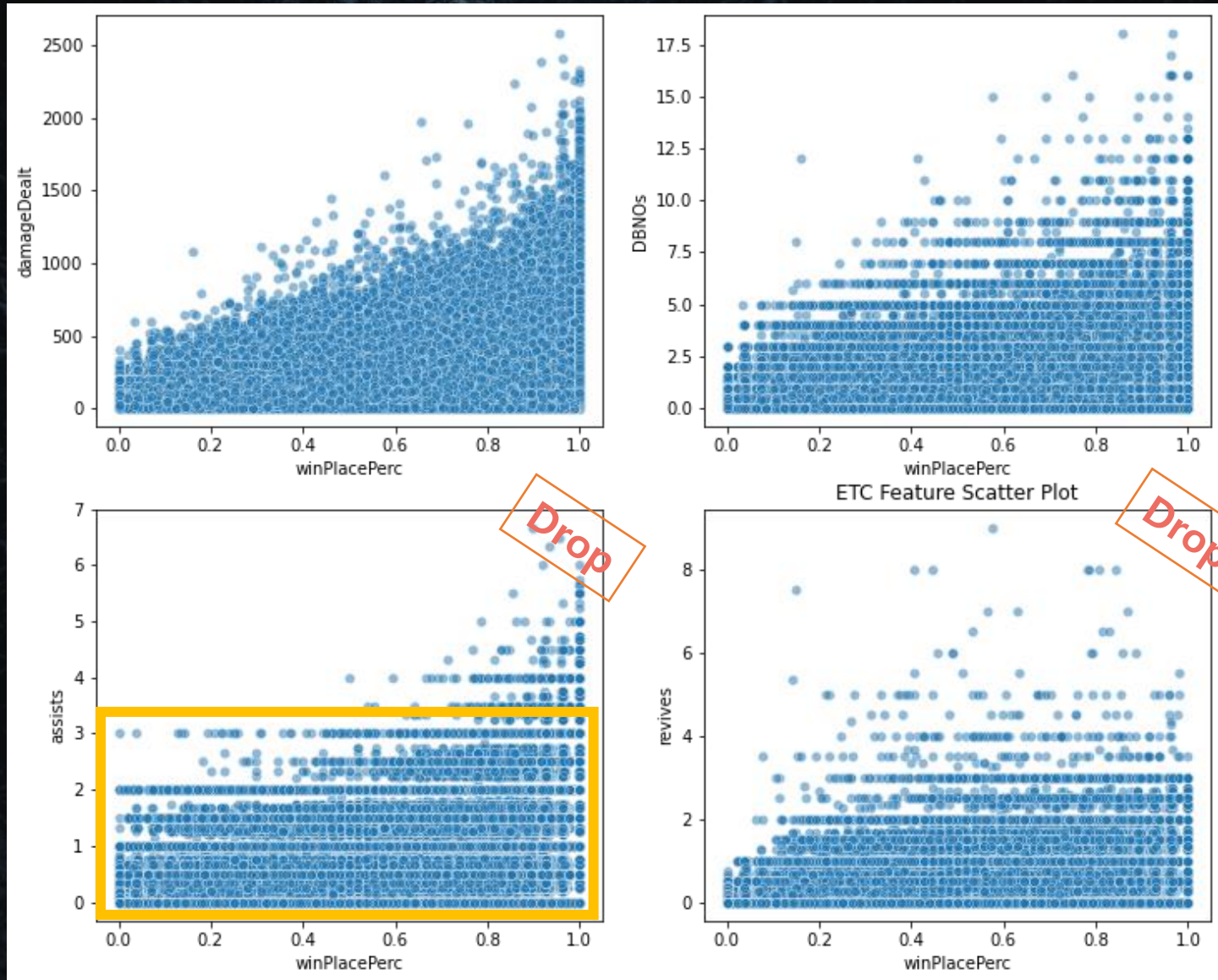
Kill Feature



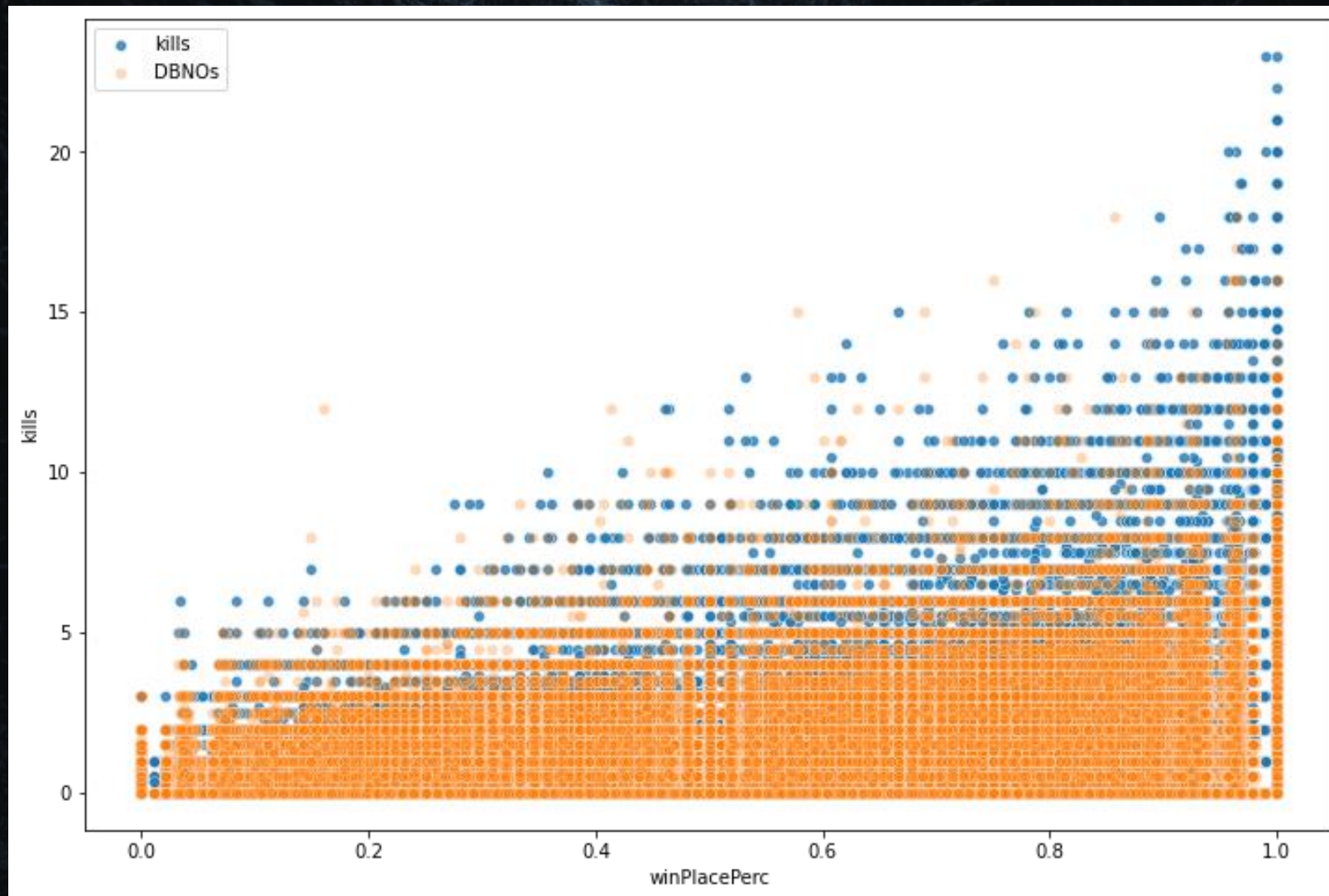
Kill Feature



ETC Feature

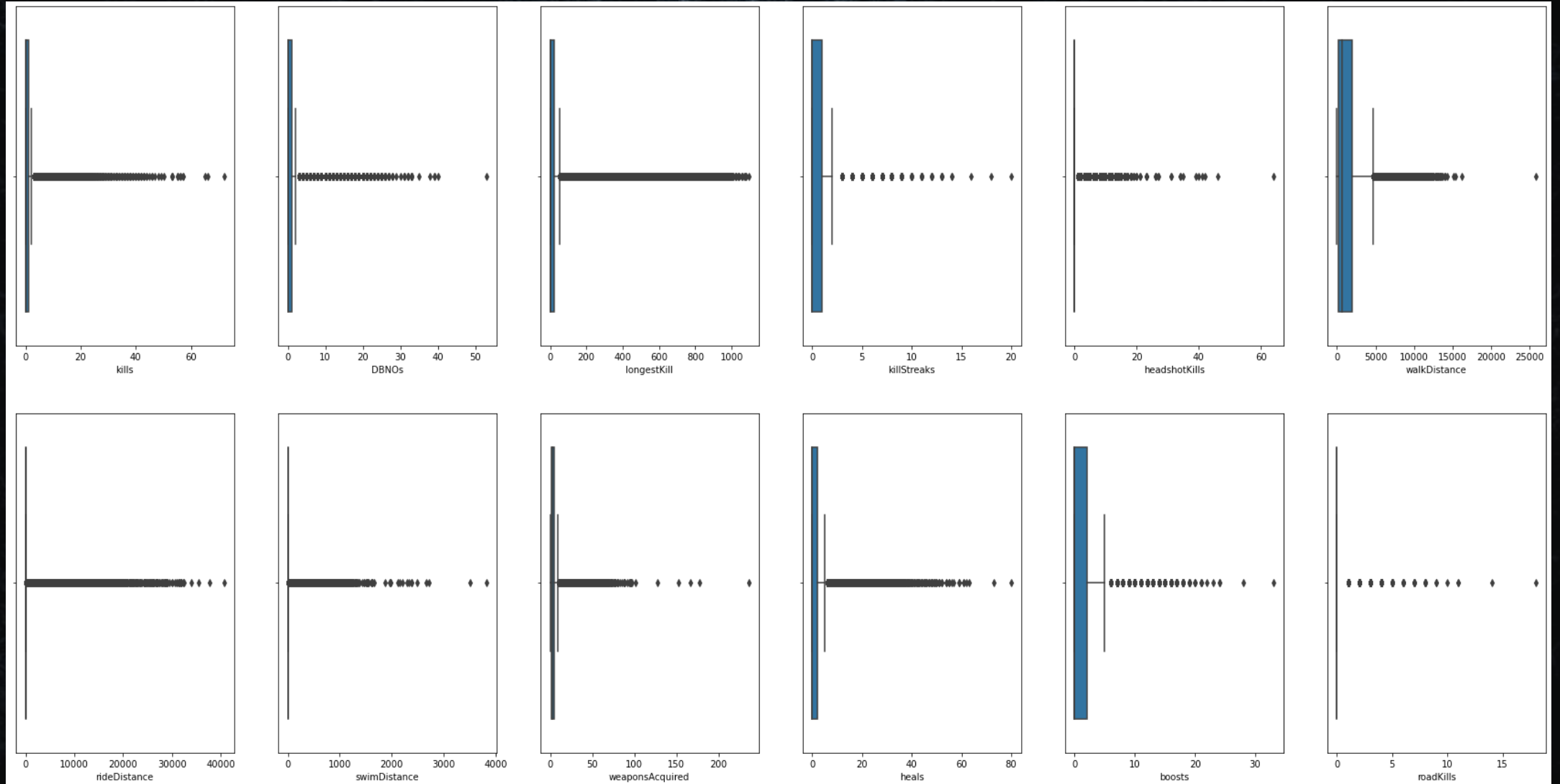


Kill and DBNOs

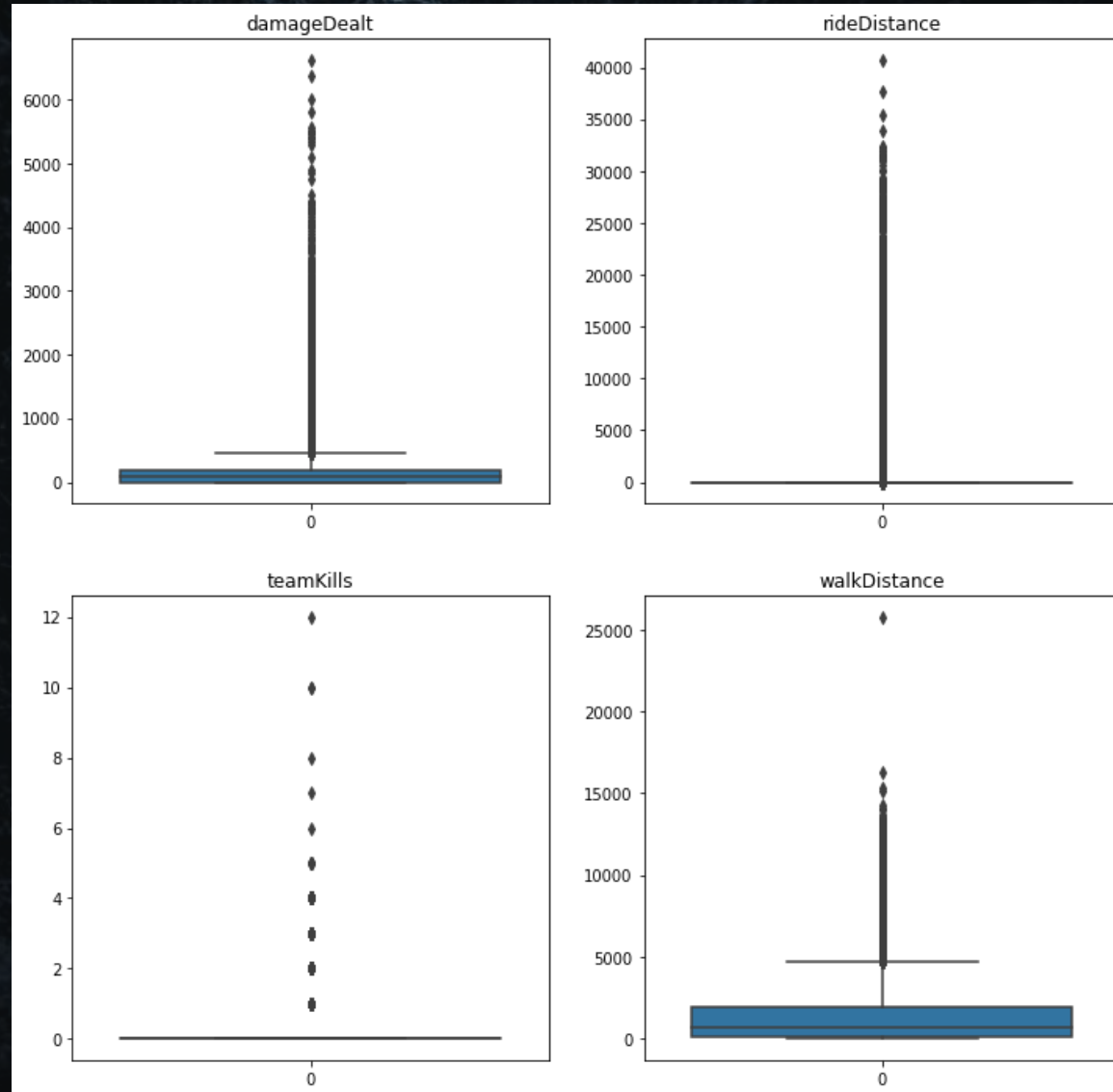


EDA

Feature Engineering



Box Plot Top 4



Feature Engineering

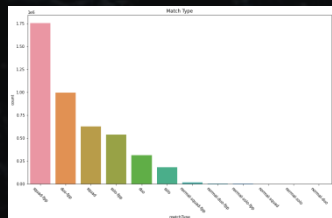
Hack Users

Roadkill	5회 이상
Kill	40회 이상
Headshotkill	15회 이상
LongestKill	1 km 이상
DBNOs	20회 이상
WalkDistance	10 km 이상
RideDistance	25 km 이상
SwimDistace	1.5 km 이상
Heal	40회 이상 사용
Boosts	15회 이상 사용
Revives	10회 이상
Weapon	50회 이상 획득
Teamkill	5회 이상

WalkDistance	10 m 이하일때
Kill	5회 이상
Heal	5회 이상
Headshotkill	5회 이상
Damage	10000이상
Superman(?)	
Damage	15000 이상
Kill	15회 이상
Heal	0

Feature Engineering

Zombi Mode
(normal game)



	damageDealt	kills	weaponsAcquired	walkDistance	winPlacePercent	maxPlace
					(==0)	(<=5)
20384	63.03	0	3	3033.0	0.0	4
22420	0.00	0	6	841.4	0.0	4
33256	579.60	3	12	0.0	0.0	2
41242	746.00	10	17	2400.0	0.0	3
51361	100.00	1	2	129.3	0.0	2
...
4435972	1248.00	10	17	614.2	0.0	5
4438715	700.90	8	22	784.4	0.0	3
4440898	90.83	1	8	0.0	0.0	2
4441098	661.50	7	13	186.6	0.0	2
4444523	0.00	0	4	4299.0	0.0	2

전처리

Normal 경기 drop 전 핵유저 현황

핵유저수 : 3169	전체 참가인원 : 4437085	비율 : 0.071
핵유저 포함 팀수 : 2262	전체 참가팀수 : 2022540	비율 : 0.112
핵유저팀 포함된 인원수 : 7591	전체 참가인원 : 4437085	비율 : 0.171
핵유저팀 포함된 경기수 : 1565	전체 경기수 : 47849	비율 : 3.271

Drop Normal 경기 수 대비 핵유저수

핵유저 포함 normal 경기수 : 206 삭제한 경기수 : 663 비율 : 39.2156

Normal 경기 drop 후 핵유저 현황

핵유저수 : 1810	전체 참가인원 : 4411698	비율 : 0.041
핵유저 포함 팀수 : 1343	전체 참가팀수 : 2014090	비율 : 0.067
핵유저 포함된 인원수 : 3419	전체 참가인원 : 4411698	비율 : 0.077
핵유저 포함된 경기수 : 1305	전체 경기수 : 47185	비율 : 2.766

핵유저 우승 횟수: 1005

핵유저팀 평균 등수 : 0.771

핵유저 평균 등수 : 0.891

핵유저 참여 경기수: 1305

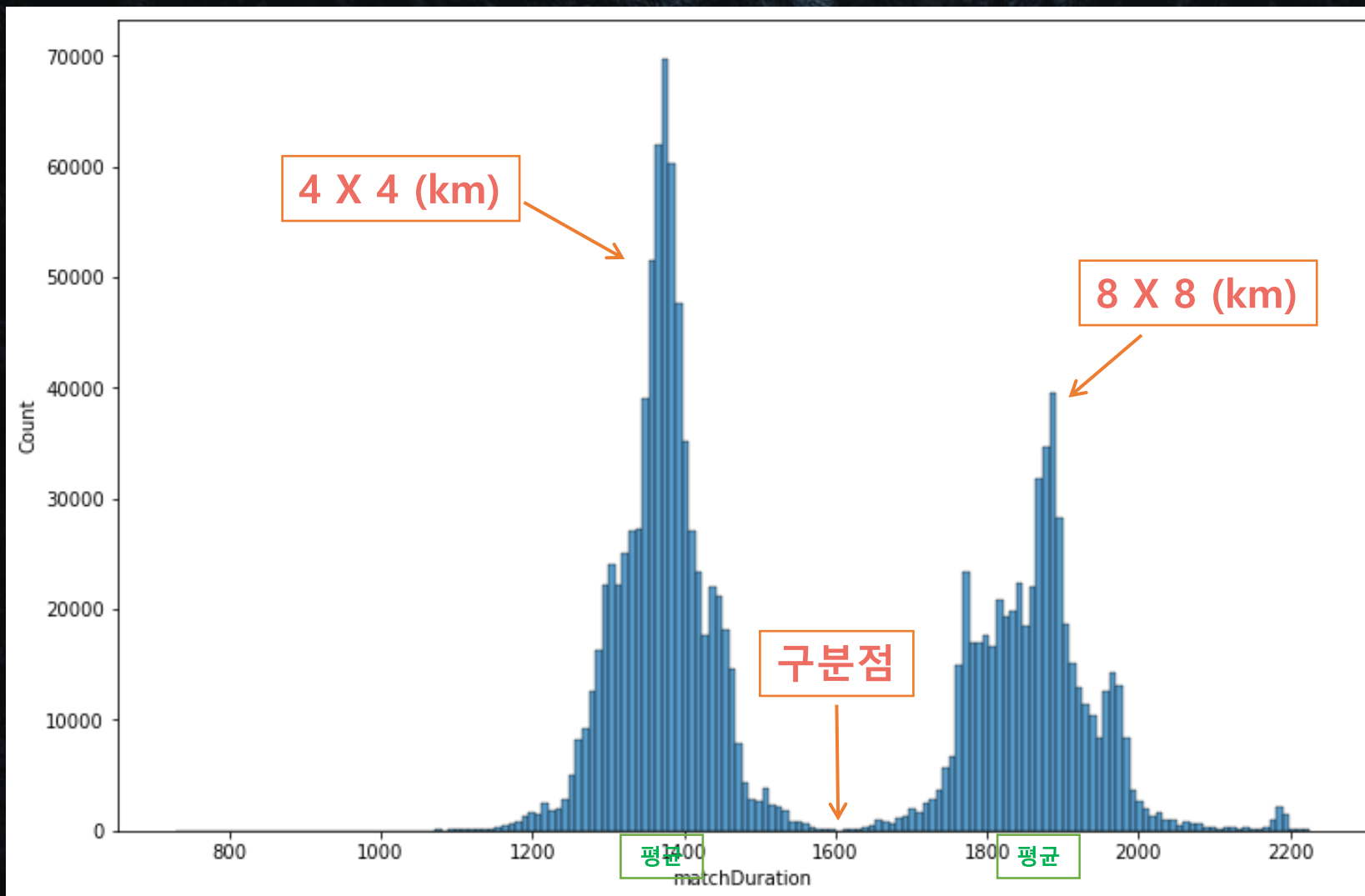
일반 유저팀 평균 등수 0.472

일반 평균 등수 : 0.473

우승 확률 : 77.011

- 핵유저의 우승 확률이 77% 가 넘는 수치가 나옴
- 핵유저를 포함한 팀은 성적 상승
- 핵유저는 평균 등수 상승
- 핵유저가 참석한 경기 -> 공정한 경기로 보기 어려움
- 핵 사용 여부 컬럼 생성 후 1인 컬럼 drop

결과 : 핵유저가 참석한 경기와 같은 팀들의 상황을 분석하고 해당 경기 제외



4 X 4 (km)

평균 게임시간 : 23분

8 X 8 (km)

평균 게임시간 : 31분

경기 시간 1600초
기준으로 맵 분류
맵 컬럼 생성

Feature Engineering

Kill Feature

Kill + DBNOs

Kill / WalkDistance

상관관계 영향 X

Distance Feature

WalkDistance + RideDistance

Item Feature

Heal + Boosts

Weapon / WalkDistance

WalkDistance + Heal

WalkDistance + Boosts

Sample Data Modeling

(No Drop Killplace)

Mean Absolute Error

1. Linear Regression	0.0894
2. Lasso	0.2584
3. Ridge	0.0894
4. XGBoost	0.0604
5. LightGBM	0.0534
6. DecisionTreeRegressor	0.1019
7. RandomForestRegressor	0.0549

Modeling

Sample Data Modeling

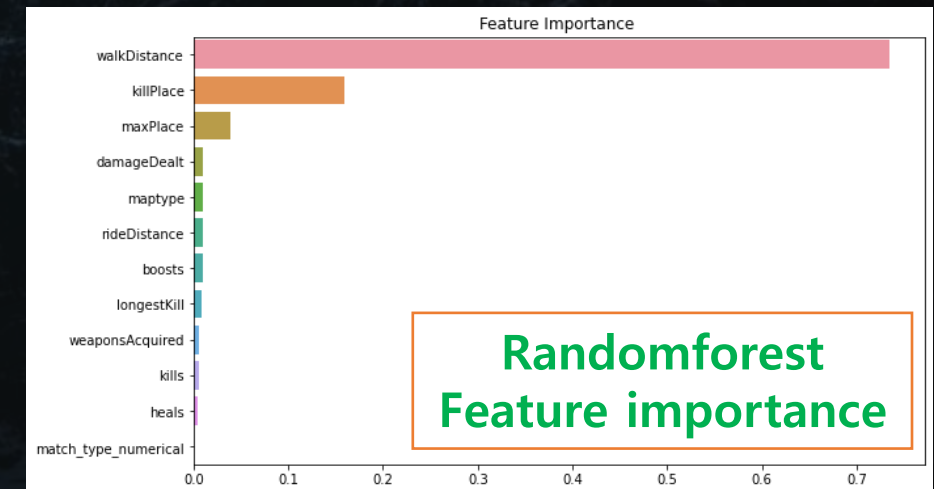
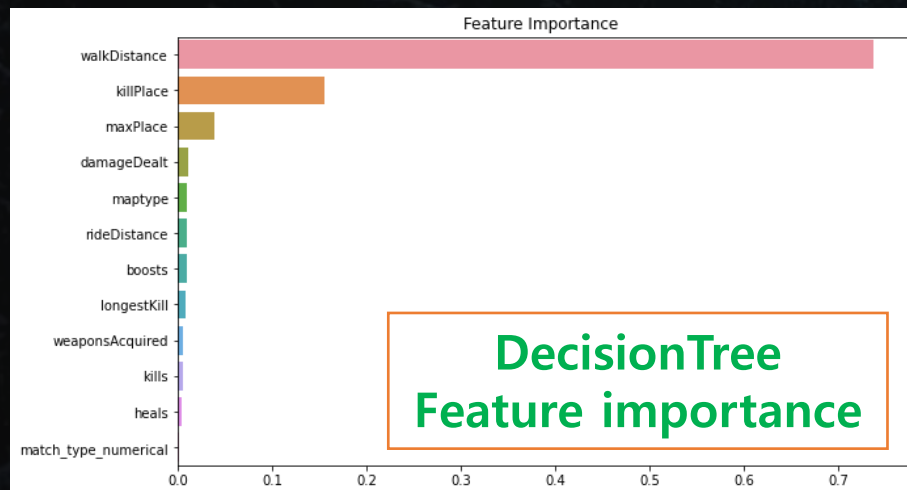
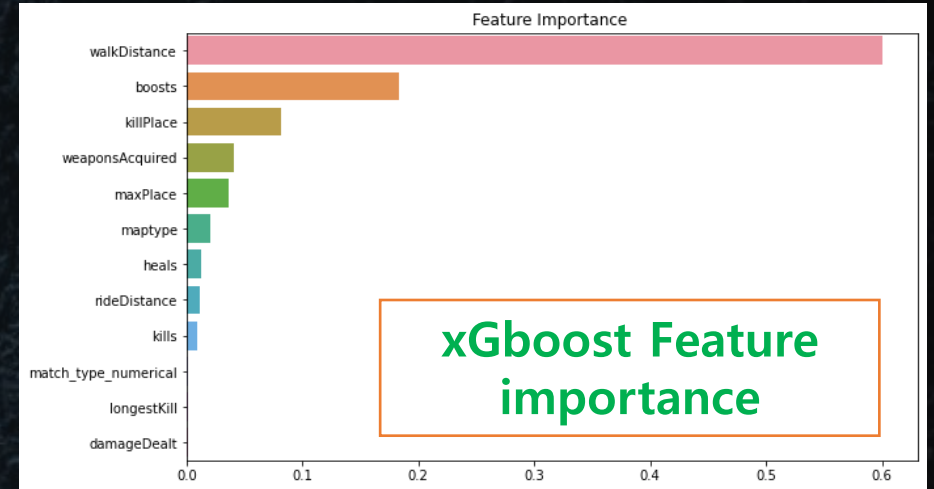
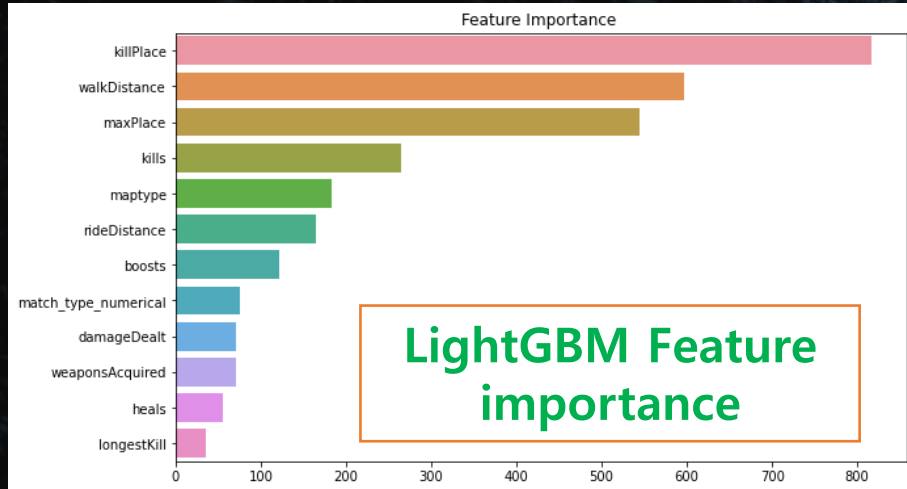
(Drop Killplace)

Mean Absolute Error

1. Linear Regression	0.1024
2. Lasso	0.2584
3. Ridge	0.1024
4. XGBoost	0.0740
5. LightGBM	0.0716
6. DecisionTreeRegressor	0.0758
7. RandomForestRegressor	

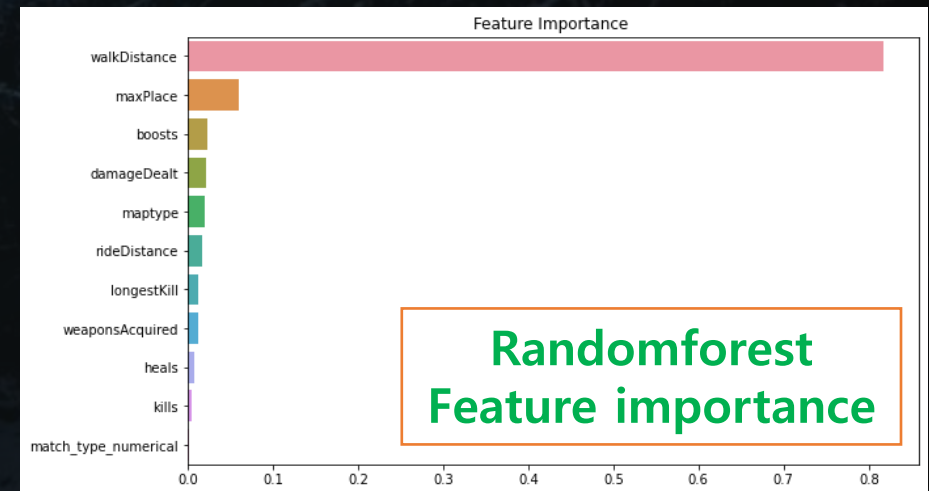
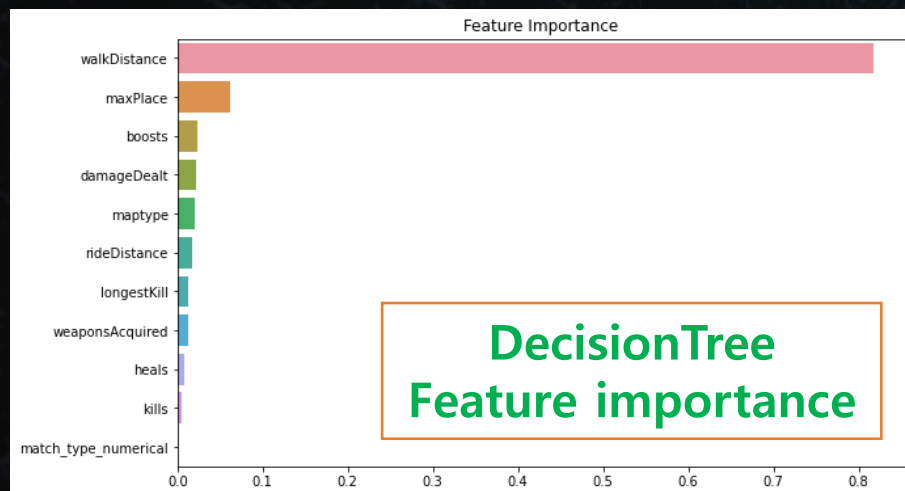
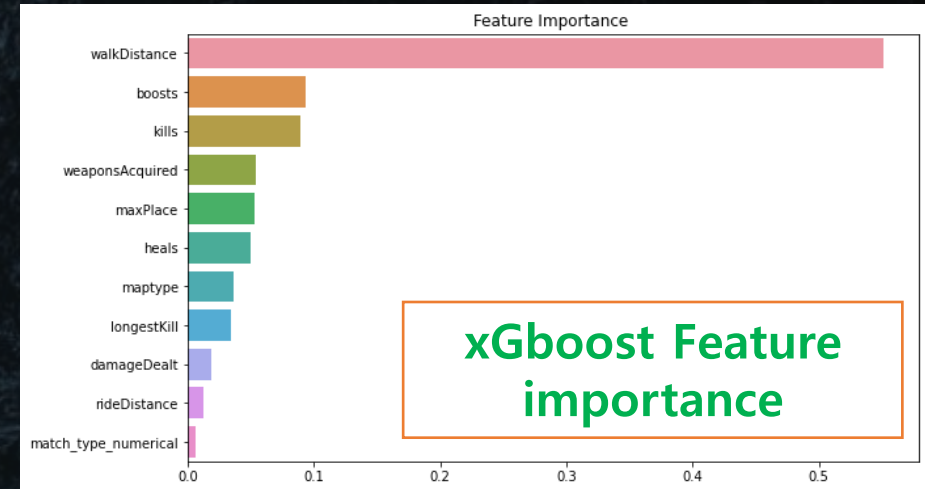
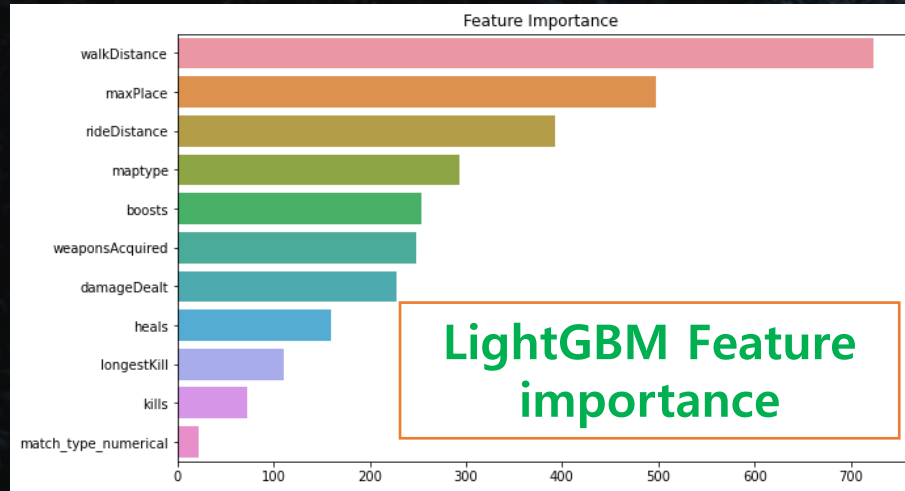
Modeling

Sample Data Feature Importance (No Drop Killplace)



Modeling

Sample Data Feature Importance (Drop Killplace)



원본 Data Modeling

Mean Absolute
Error
(Mean Error Player)

1. Linear Regression	0.0841	
2. Lasso	0.1145	
3. Ridge	0.0841	
4. XGBoost	0.0607	-> 0.2333
5. LightGBM	0.0535	-> 0.0712
6. DecisionTreeRegressor	0.0001	
7. RandomForestRegressor	0.0287	-> 0.0962

The background image is a firefighter's helmet and gear resting on a bed of straw. The image is split vertically down the middle. The left half is a solid yellow color, and the right half is a solid dark blue color. The helmet is positioned in the lower-left quadrant, with its visor facing upwards. The gear, including a strap and a small label, is visible on the right side of the helmet. The straw is a light tan color and is scattered around the gear.

THANK

YOU