

WHAT STRATEGIES DO WE NEED TO SURVIVE FOR A LONG TIME?

# PUBG Survival Time Prediction

## Data

### PlayerUnknown's BattleGround 상위 랭커 데이터

85,000명의 게임 플레이와 관련된 변수들로 구성  
승률, 데미지, Top 10 달성률, 차량 이동 시간 등



## Environment

Python 3.6+ / pandas, numpy, matplotlib, seaborn, scipy, scikit-learn

## Issues

1. 특정 타입의 게임만 고집하여 하는 **편향적인 플레이어** 존재  
(솔로, 듀오, 스쿼드 중 하나만 하는 플레이어들)
2. 입력변수들과 출력변수 사이의 **데이터 누출** 의심  
(오래 생존하였기 때문에 승률이 높다고 설명할 수 있지만,  
승률이 높다고 하여 오래 생존했다고 추론할 수 없음)
3. 기존 데이터에 이미 **파생 변수**들이 많이 존재함
4. 각 변수들의 분포에서 **왜도(skewness)**가 크게 나타남



## Purpose

게임 플레이 변수들을 이용하여 특정 플레이어의 **스쿼드 평균 생존시간을 예측**  
**변수 설명**이 가능한 알고리즘을 사용하여 **생존에 중요한 변수를 탐색**

## Methodology

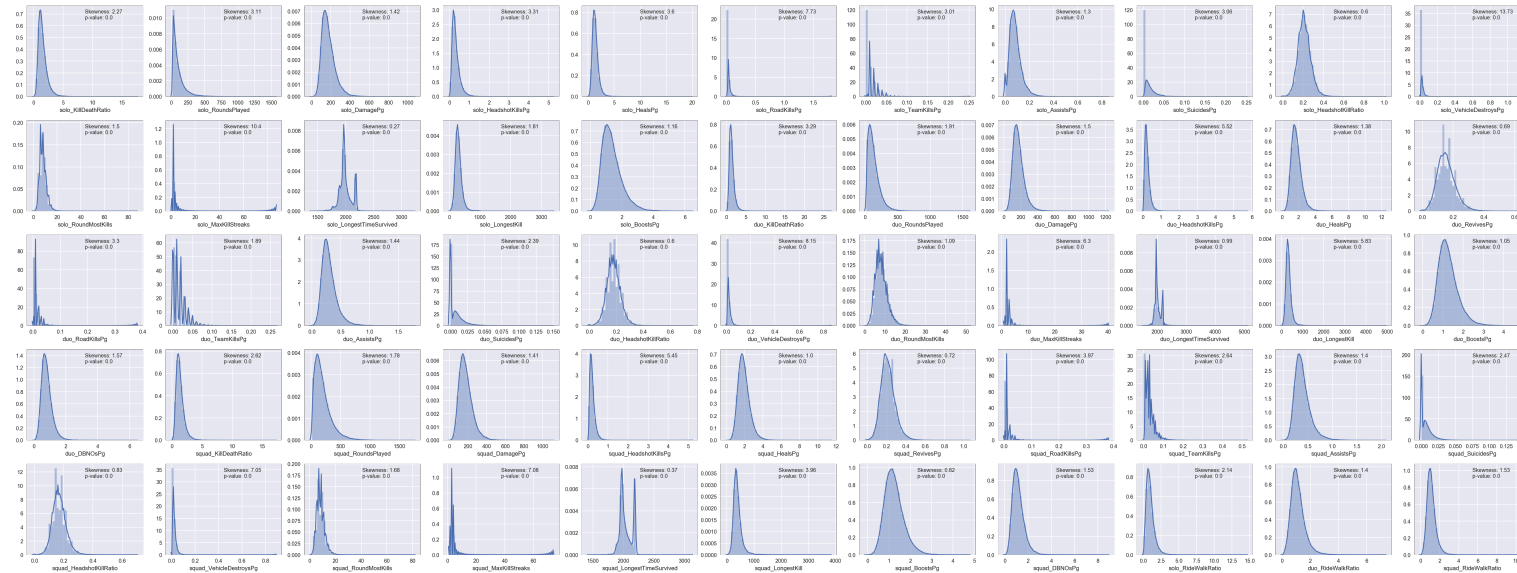
알고리즘 성능의 비교를 위하여 다양한 알고리즘을 사용 / RMSE,  $R^2$ 로 모델 평가  
최소제곱법(OLS), 능형 회귀(Ridge regression), Lasso,  
랜덤 포레스트(Random Forest), 그라디언트 부스팅(Gradient boosting)

## Solution

1. 각 게임 타입을 **20회 이상** 한 플레이어에 한해서 분석  
(87,898명에서 57,593명으로 대상을 약 34% 줄임)
2. 승률, 레이팅 등의 **직접적인 변수들을 제거**  
평균도보시간과 평균차량탑승시간은  
**도보와 차량탑승의 비율**로 새로운 파생변수 생성
3. 누적값인 변수는 **총 게임수로 나누어** 파생변수 생성
4. 왜도가 2보다 큰 변수들은 **로그를 취하여** 왜도를 제거  
(너무 큰 왜도를 갖는 경우, 이를 이용하여도 왜도가 제거되지 않는 경우 존재)

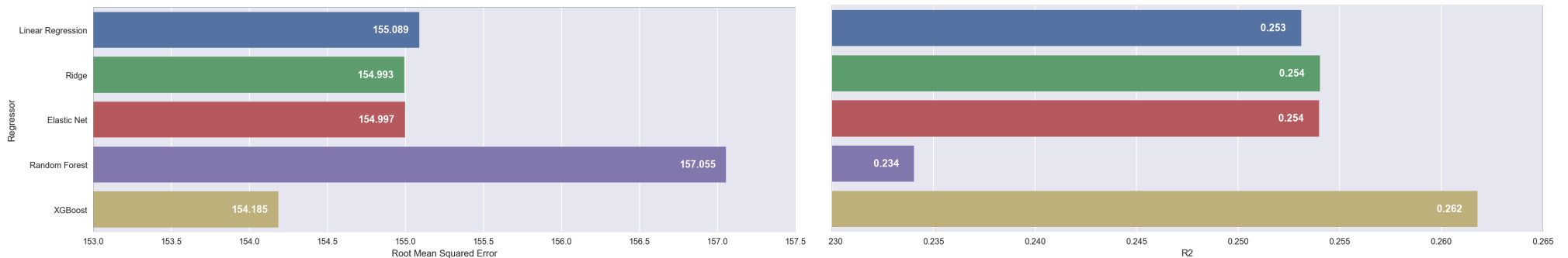
WHAT STRATEGIES DO WE NEED TO SURVIVE FOR A LONG TIME?

# PUBG Survival Time Prediction



각 변수들의 분포를 나타내는 그래프

Prediction performance for each model

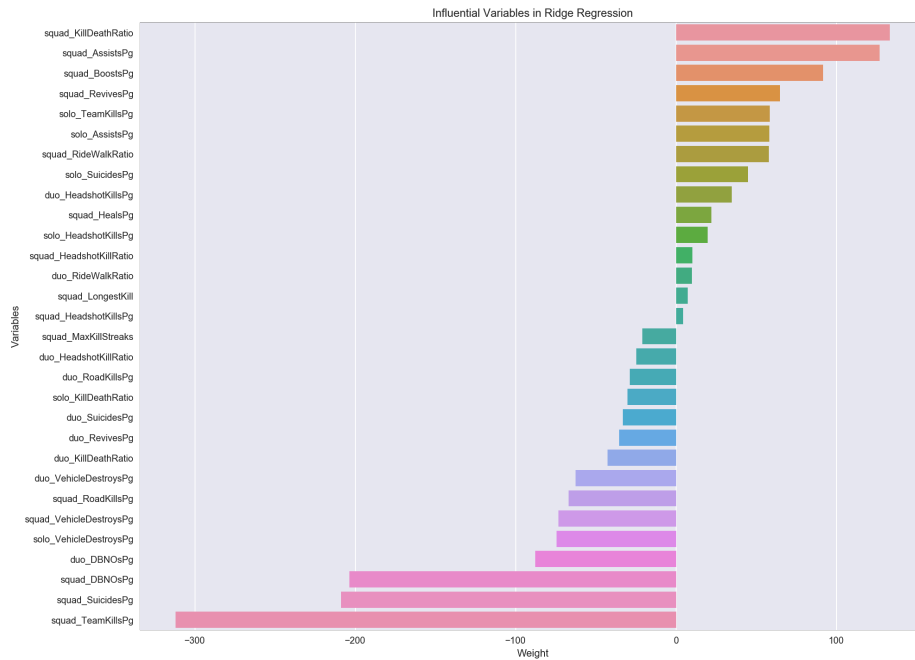


각 모델들의 성능 비교

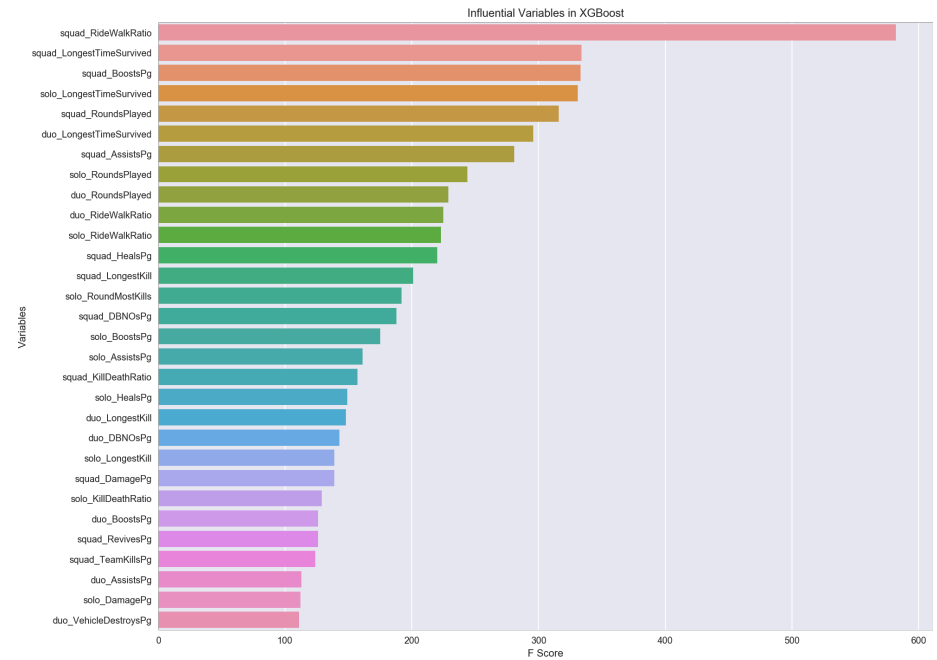
그라디언트 부스팅의 성능이 가장 좋으며, 랜덤 포레스트의 성능이 가장 나쁨을 알 수 있다.

WHAT STRATEGIES DO WE NEED TO SURVIVE FOR A LONG TIME?

# PUBG Survival Time Prediction



능형 회귀로 얻은 변수들의 가중치



그라디언트 부스팅으로 얻은 변수들의 중요도

에너지드링크 등의 **부스팅**과 **어시스트**가 생존 시간을 크게 늘려주는 변수이며  
 걷는 시간보다 **차량을 이용한 시간이 긴 경우** 역시 생존 시간을 크게 늘려줄 수 있다.

추가적으로 **헤드샷 비율**이 높은 경우, 빠른 적 사살로 인해 생존 시간에 도움이 됐다.

반대로 **팀킬**이나 **DBNO(Down But Not Out)**가 많은 경우는 생존 시간을 줄이는 것으로 확인됐다.