



유저 샷 분석

Introduction

QED 콘솔에 내재화 시킬때 필요한 Q-Fit 사용자 Input 데이터의 평균값 신뢰성을 파악하기 위해서 샷 분석을 하였습니다. 주어진 데이터는 2022년 6, 8월 샷 데이터를 통일된 컬럼들에 맞추어 병합하였고 대략 200만개의 데이터 행(전체 유저의 전체 샷 개수)을 활용하였습니다. 또한, 전체 샷 데이터중에서 30~100개의 유저수는 6485명, 100~200개의 유저수는 3572명, 200~300개의 유저수는 1279명, 300개 이상의 유저수는 1279명입니다. 드라이버만을 분류하기 위한 데이터 전처리에는 total distance와 back spin의 기준으로 데이터 유형을 나누어 분석을 하였습니다. 원본 데이터, 전처리한 데이터를 활용하여 각각의 데이터 유형마다 랜덤으로 1000명을 선택하여 back spin, club speed, launch angle, face angle, club path, side distance 각각의 특성들에 대해서 1000명 중에 정규성을 가진 유저의 수를 파악하였습니다. 아래의 분석 결과를 통해서 어떤 경우에 정규성 비율이 높은지 낮은지 확인 할 수 있습니다.

Analysis

데이터 전처리

번호	데이터 유형	total distance	back spin(spin rate라고 가정)
1	원본 데이터	전체	전체
2	드라이버만	상위 90%	하위 90%
3	드라이버만	상위 80%	하위 80%
4	드라이버만	상위 70%	하위 70%

→ 4가지 유형으로 분류하여 원본 데이터, 각각의 total distance와 back spin를 기준으로 전처리한 데이터입니다. 예를들어, 번호 2번 같은 경우에 한 유저의 샷데이터에서 total distance 하위 10%와 back spin 상위 10%를 제거하여 둘의 조건에 만족하는 샷데이터를 의미합니다.

정규성(%)

데이터 유형마다의 랜덤으로 선택한 1000명 중 정규성 가진 유저수의 비율(%)

원본 데이터 - 번호 1번

전체 유저수	전체 샷데이터 수	데이터 샷 갯수	back spin	club speed	launch angle	face angle	club pat
6485명	386148명	30~100개	36.3%	18.5%	36.1%	60.8%	60.1%
3572명	502145명	100~200개	12.9%	5.6%	12.0%	35.2%	32.5%
1325명	321903명	200~300개	5.1%	2.6%	5.1%	19.8%	17.2%
1279명	586027명	300개 이상	1.2%	0.9%	1.8%	10.7%	9.1%

→ 전반적으로 정규성 비율이 낮고 한 유저당 샷의 갯수가 많으면 많을수록 점점 정규성이 떨어지는 것을 볼 수 있습니다.

드라이버만(비거리 상위 90%, Back Spin 하위 90%) - 번호 2번

전체 유저수	전체 샷데이터 수	데이터 샷 갯수	back spin	club speed	launch angle	face angle	club pat
6485명	306699명	30~100개	62.1%	32.7%	66.2%	78.5%	72.6%
3572명	403200명	100~200개	25.8%	15.0%	38.1%	61.1%	52.3%
1325명	259283명	200~300개	8.2%	7.6%	20.6%	47.2%	35.3%

1279명	473384명	300개 이상	2.4%	3.0%	13.0%	33.6%	22.8%
-------	---------	---------	------	------	-------	-------	-------

→ 원본 데이터에 비해서 상대적/전반적으로 각 특성마다 정규성을 가진 유저의 비율이 2배 이상 가까이 올라간 것을 볼 수 있습니다.

드라이버만(비거리 상위 80%, Back Spin 하위 80%) - 번호 3번

전체 유저수	전체 데이터 수	데이터 샷 갯수	back spin	club speed	launch angle	face angle	club pat
6485명	242170명	30~100개	58.8%	43.7%	71.5%	84.0%	77.9%
3572명	318515명	100~200개	18.9%	22.1%	47.1%	72.6%	59.5%
1325명	204833명	200~300개	4.6%	9.9%	29.8%	58.1%	47.2%
1279명	374058명	300개 이상	0.4%	3.9%	20.0%	46.0%	30.8%

→ back spin을 제외한 나머지 특성들의 정규성을 가진 유저의 비율이 위의 표들 보다 상대적/전반적으로 높아진 것을 볼 수 있습니다.

드라이버만(비거리 상위 70%, Back Spin 하위 70%) - 번호 4번

번호	전체 유저수	전체 데이터 수	데이터 샷 갯수	back spin	club speed	launch angle	face ang
4	6485명	186182명	30~100개	63.5%	54.9%	80.8%	88.0%
4	3572명	244355명	100~200개	20.4%	30.4%	61.1%	78.6%
4	1325명	157009명	200~300개	4.7%	16.7%	41.5%	67.4%
4	1279명	286522명	300개 이상	0.6%	7.1%	24.3%	55.6%

→ 위와 같이 back spin을 제외한 나머지 특성들의 정규성 가진 유저의 비율이 위의 표들 보다 상대적/전반적으로 높아진 것을 볼 수 있습니다.

Result

데이터 유형 4가지 중에서 total distance와 back spin의 이상치를 제거할수록 1000명 중에서 각각의 특성들의 정규성 가진 유저수의 비율이 높아지는 것을 확인할 수 있었습니다(반면에, back spin은 정규성 비율이 떨어졌습니다). 이번 분석은 30~100개, 100~200개, 200~300개, 300개 이상을 가진 각각의 유저들을 비교하였지만 다른 경우의 수를 가진 유저수에게는, 예를들어 500개 이상, 1000개 이상, 혹은 10000만개 이상, 이 결과를 통해 추론은 가능하지만 다른 결과를 가질 가능성도 있습니다. 결과적으로, 데이터 total distance와 back spin의 이상치 처리를 크게 하면 할수록 정규성을 가진 유저수의 비율은 높아지는 것을 알수있지만 데이터 손실량의 크기도 커지는 것을 우려하여 적정선에서 데이터 전처리를 한다면 QFit Input 특성들에 대해 신뢰성을 판단하는데 있어서 유의미한 결과가 나올 것으로 예상됩니다.

다음 단계는 정규성 비율뿐만 아니라 각 상황에 따른 평균, 표준편차, 최대값, 최소값등을 구하여 드라이버/웨지/아이언 샷중에 드라이버샷 분류를 할 예정입니다.

Data Source

- 2208_150m_driver.xlsx(2022년 8월 샷데이터, sheet1/sheet3)
- 드라이버추출_220610.xlsx(2022년 6월 샷데이터, 스윙옵션/스윙옵틱스)