

프로야구 배럴(Barrel)을 통한 타자 성적 예측

「2021 빅콘테스트」 데이터 분석 분야 (챔피언리그 : 스포츠테크)

CONTENTS

01. 회사소개

02. 문제 출제 목적

03. 대회 문제 소개

04. 제공 데이터

05. 참고정보

회사소개

스포츠투아이 주식회사는 **IT기술력(빅데이터 / 솔루션 / 인공지능)**을 기반으로

스포츠의 즐거움을 제공하는 동시에 새로운 영역을 개척해 나가고자 하는 “**스포츠 IT 기업**”입니다.



스포츠데이터 전문기업

국내 최고의 스포츠데이터 전문가
국내 최대 규모 스포츠데이터 확보/가공 능력
스포츠 IT 분야 23년 경험



고도화 기술 연구 그룹

부속 전문 연구소 운영
국내 최고의 산학연 파트너십 구축
문화체육관광부 혁신기반 조성(R&D) 기업



스포츠산업 선도기업

국민체육진흥공단 스포츠산업 선도기업 선정
100억 매출 규모 스포츠 사업 역량
글로벌 시장 진출 추진

회사소개

대한민국 [스포츠&IT 산업 국가대표] 기업 (주요성과 Reference)

문화체육관광부 (국민체육진흥공단)

- ◎ 대한민국, 스포츠산업 선도기업 선정 (2019~2021)
- ◎ 대한민국, 스포츠산업종합지원 기업 선정 (2021~2022)
- ◎ 대한민국, 스포츠산업 혁신기반 조성(R&D) 주관기업 지정 (2021~2024)

금융위원회

- ◎ 대한민국, 혁신기업 국가대표 1000 기업 선정 (2021) * 문체부/국민체육진흥공단 추천 기업

과학기술정보통신부 (한국지능정보사회진흥원, 한국정보통신진흥협회)

- ◎ 대한민국, 디지털경제 선도 파트너 기업 (2020~2021)

중소벤처기업부 (창업진흥원)

- ◎ 대한민국, 대-스타 해결사 플랫폼 사업 협력 추진 기업 (2020~2021)

회사소개

《《《 스포츠투아이 주요 상품 및 서비스 《《《

상품/서비스 01

스포츠 데이터 기반
기술 시스템

자동 추적 시스템

로봇심판 시스템

AI 로봇기사 시스템

...

상품/서비스 02

스포츠 데이터 기반
기술 솔루션

종목별 기록 수집 솔루션

전력분석 솔루션

기록실 솔루션

연봉고과 솔루션

...

상품/서비스 03

스포츠 데이터 기반
콘텐츠

3D 그래픽 콘텐츠

Live Strikezone 콘텐츠

영상+데이터 융합 콘텐츠

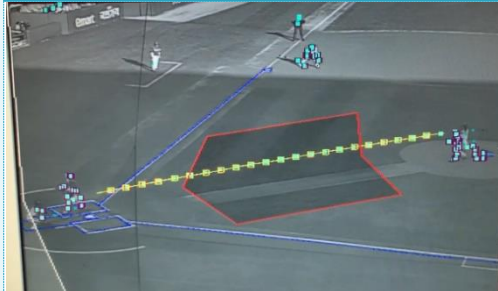
AI 투구예측 콘텐츠

...

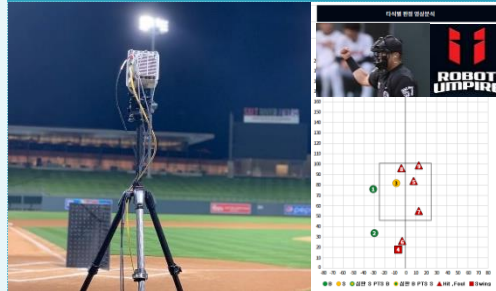
회사소개

스포츠 데이터 기반
기술 시스템

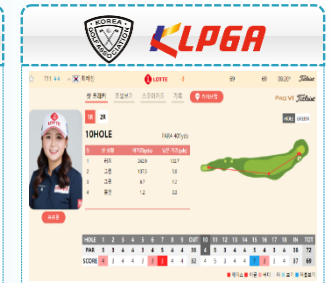
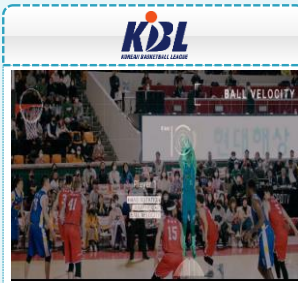
자동 추적 시스템



로봇심판 시스템



AI 로봇기사 시스템

스포츠 데이터 기반
기술 솔루션스포츠 데이터 기반
콘텐츠

3D 그래픽 콘텐츠



Live Strikezone 콘텐츠



영상+데이터 융합 콘텐츠



OUR CLIENTS

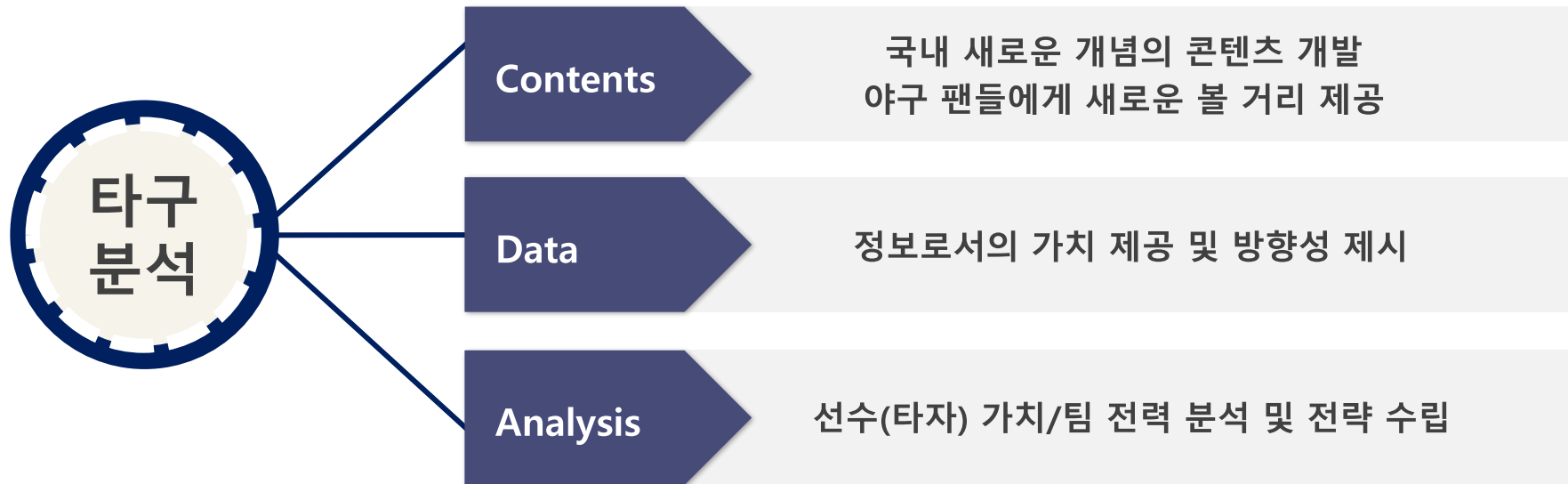
스포츠투아이는 국내외 50여개의 CLIENTS와 함께하고 있습니다.



문제 출제 목적

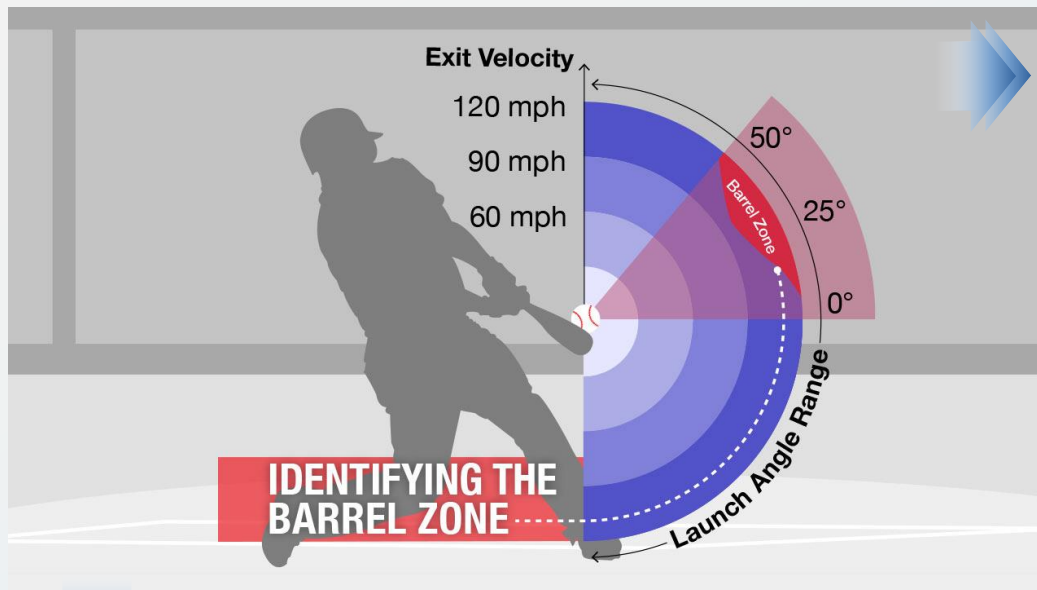
Q : “과연 어떤 타구가 좋은 타구일까?”

[프로야구 배럴(Barrel) 기준 정의 및 타자의 OPS(장타율+출루율) 성적 예측]



대회 문제 소개

(사례) MLB Barrel 개념 및 기준



<<https://www.mlb.com/glossary/statcast/barrel>>

MLB Barrel 기준 정의

The Barrel classification is assigned to batted-ball events whose comparable hit types (in terms of exit velocity and launch angle) have led to a **minimum .500 batting average and 1.500 slugging percentage** since Statcast was implemented Major League wide in 2015.

----> 타율 0.500 ↑ & 장타율 1.500 ↑ 를 생산 가능한 타구



규정
범위

“타율 0.500, 장타율 1.500 이상을 기록할 수 있는 타구의 속도와 발사 각도의 집합체”

- 98마일(mph) 이상의 타구 속도가 26°~30° 범위의 발사각을 충족하는 경우
[98마일(mph) 초과인 타구 속도에서는 발사각 범위도 더 확대됨(위 사진 참조)]
- 116 마일(mph) 이상의 타구 속도가 8°~50° 범위의 발사각을 충족하는 경우

대회 문제 소개

(사례) MLB 2021 전반기 배럴 Top 5

Rk.	Player	Year	타구속도(평) 발사각(평)		배럴%
			Avg EV (MPH)	Avg LA (°)	Barrel%
1	 Ohtani, Shohei	2021	93.7	17.8	26
2	 Tatis Jr., Fernando	2021	93.5	13	21.7
3	 Acuna Jr., Ronald	2021	93.8	18.2	20.3
4	 Judge, Aaron	2021	95.8	10.2	19.7
5	 Gallo, Joey	2021	92.4	21.3	19.6

올해 MLB 에서 배럴 타구를 가장 많이 생산한 타자는 누구일까?

놀랍게도 MLB 전체 1위는 '투타 겸업' 선수

오타니 쇼헤이(LA 에인절스)이다.

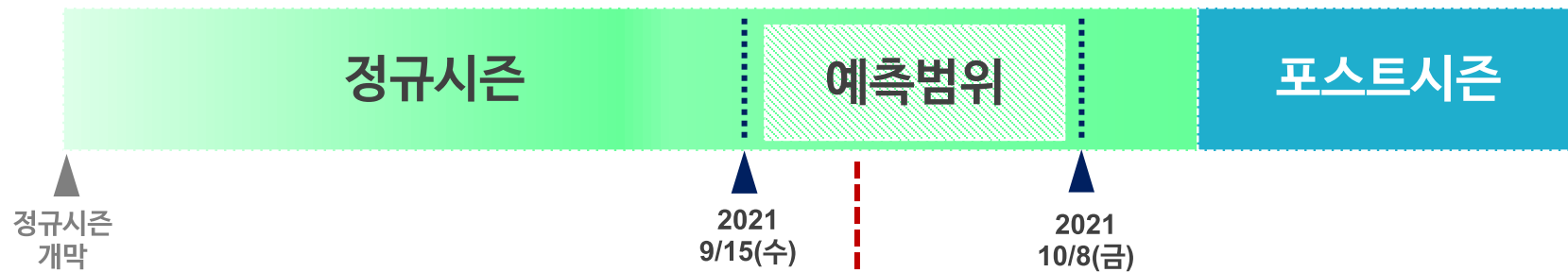
2021시즌 전반기, 26%의 배럴 타구 생산 비율로 MLB에서 가장 많은 배럴 타구를 만들어 내고 있다.

문제 01

프로야구 배럴(Barrel) 정의

"프로야구에서는 어떤 타구가 배럴기준을 충족할 수 있는가?"

2021 프로야구 정규시즌



제공 데이터 범위(p.13 참조)

- 데이터 항목
 - 타자 기본(클래식) 데이터
 - 타구 트래킹(HTS) 데이터
 - 경기 스케줄 (예측범위 : 9/15~10/8)
- 데이터 기간
 - 2018~2020시즌
 - 2021시즌(전반기)
- 포맷 및 방식 : csv 파일 / 연도별로 파일 생성



문제 02

성적 예측 모형 개발(모델링) 통한
타자 OPS(장타율+출루율) 예측

OPS

=

장타율

+

출루율

대회 문제 소개

출제 문제 요약

문제 01

프로야구 배럴(Barrel) 정의

"프로야구에서는 어떤 타구가 배럴기준을 충족할 수 있는가?"

- 제시된 사례(MLB Barrel 개념 및 기준) 참고,
프로야구 Barrel 개념(기준) 정의
- 스포츠투아이(주)에서 제공하는 데이터 활용
(타자 기본정보, 타자 트래킹 데이터 등)

문제 02

성적 예측 모형 개발(모델링)을 통한
타자 OPS(장타율+출루율) 예측

- 타자 성적(OPS) 예측 모형 개발
(예측 모형은 정의한 Barrel 기준을 기반으로 개발해야함)
- 예측 모형을 통한 2021시즌 타자 성적(OPS) 예측

- 대상 : 우측 리스트 선수(10명)

- ▶ 세부정보는 제공 데이터 파일 참조
- ▶ 평가, 예측값 제출 : 아래 조건 참조

No	PCODE	No	PCODE
1	76232	6	78224
2	68050	7	78513
3	75847	8	76290
4	67341	9	79215
5	79192	10	67872

* (평가) 특정 기간(9/15~10/8) 기록이 없는 선수 제외 후 평가(공통)

* (경가) 2021 정규시즌 규정타석 기준(팀경기수*3.1) 미충족 선수
제외 후 평가(공통)

* (동점자 처리 기준) OPS>장타율>출루율 순서로 순위 구분
: 예측한 타자 성적 제출 시, OPS/장타율/출루율 구분 해야함

(작성 예시)

[소수점 셋째 자리까지 표기]

구분	OPS	장타율	출루율
홍길동	1.097	0.680	0.417

제공 데이터

2018년~2020년, 2021년(전반기) 기간, 타자 데이터

[타자 기본(클래식) 데이터] 시트

컬럼	GYEAR	PCODE	GAMENUM	PA	AB	BA	HIT	HR	TOTB	SLG	SF	BB	KK	IB	HP	GD
컬럼 설명	연도	선수코드	경기	타석	타수	타율	안타	홈런	루타	장타율	희생플라이	볼넷	삼진	고의4구	사구	병살타
샘플 데이터	2018	60100	70	169	152	0.243	37	3	63	0.414	1	12	36	0	4	5
	2018	60184	6	10	10	0.200	2	0	3	0.300	0	0	3	0	0	0
	2018	60288	1	1	1	0.000	0	0	0	0.000	0	0	0	0	0	0
	2018	60343	83	174	162	0.216	35	8	63	0.389	0	8	49	0	3	3
	2018	60456	4	8	6	0.500	3	0	3	0.500	0	2	0	0	0	0

고의4구(IB) 포함한 수치

[선수] 시트

컬럼	GYEAR	PCODE	NAME	T_ID	POSITION	AGE_VA	MONEY
컬럼 설명	연도	선수코드	선수명	팀코드	포지션	나이	연봉
샘플 데이터	2018	61102	유강남	LG	포	26	18500만원
	2018	61186	이천웅	LG	외	30	8000만원
	2018	62404	구자욱	SS	외	25	25000만원
	2018	62415	박해민	SS	외	28	29000만원
	2018	62869	문승원	SK	투	29	9000만원

제공 데이터

2018년~2020년, 2021년(전반기) 기간, 타자 데이터

[타자의 타구 트래킹 데이터(HTS)] 시트

컬럼	GYEAR	G_ID	PIT_ID	PCODE	T_ID	INN	HIT_VEL	HIT_ANG_VER	HIT_RESULT	PIT_VEL	STADIUM
컬럼 설명	연도	경기코드	타구코드	선수코드	타자 팀 코드	이닝	타구속도 (km/h)	발사각도	타격결과	상대투수 투구구속 (km/h)	해당 구장
샘플 데이터	2018	20180324HHWO0	180324_140436	62797	HH	1	131.50	42.7	플라이	149.59	고척
	2018	20180324HHWO0	180324_140514	76753	HH	1	135.18	9.9	1루타	148.78	고척
	2018	20180324HHWO0	180324_140647	71752	HH	1	152.41	2.1	1루타	148.59	고척
	2018	20180324HHWO0	180324_140911	62700	HH	1	113.72	13.5	1루타	139.13	고척
	2018	20180324HHWO0	180324_142050	68730	HH	2	54.11	16.4	번트안타	134.07	고척



타구코드
각 타구가 가지는 고유의 코드(구분 값)



타격 결과 항목(총 17가지)	
1루타	야수선택
2루타	인필드플라이
3루타	직선타
내야안타(1루타)	파울플라이
땅볼아웃	플라이
번트아웃	홈런
번트안타	희생번트
병살타	희생플라이
삼중살타	

▲ 국내 프로야구 공식페이지
(www.koreabaseball.com – 기록 – 기록실)

▲ 해외 프로야구 공식페이지
(<https://www.mlb.com>)

▲ App (KBO STATS)

감사합니다

[Transformation 2021 : Reboot 2i]

