

PE 프로그램 진행과정 (7.29)

01

크롤링 함수 코드 완성

- 표준 칼럼명으로 변경
- day 칼럼 타입 변환

	name	day	vs	result	bat_order	position
0	강백호	2021-04-04	한화	W 3:2	4	1B
1	강백호	2021-04-06	LG	L 2:3	4	1B
2	강백호	2021-04-07	LG	W 7:3	4	1B
3	강백호	2021-04-08	LG	L 3:7	4	1B
4	강백호	2021-04-09	@삼성	L 5:7	4	1B
...
5815	김민하	2021-06-23	@삼성	L 0:3	7	LF
5816	김민하	2021-06-24	@삼성	L 2:3	6	RF
5817	김민하	2021-06-26	KT	L 2:11	3	RF
5818	김민하	2021-07-09	@SSG	D :	5	CF
5819	김민하	2021-07-10	@SSG	D :	5	CF

5820 rows × 32 columns



	name	day	vs	result	bat_order
0	강백호	20210404	한화	W 3:2	4
1	강백호	20210406	LG	L 2:3	4
2	강백호	20210407	LG	W 7:3	4
3	강백호	20210408	LG	L 3:7	4
4	강백호	20210409	@삼성	L 5:7	4
...
5815	김민하	20210623	@삼성	L 0:3	7
5816	김민하	20210624	@삼성	L 2:3	6
5817	김민하	20210626	KT	L 2:11	3
5818	김민하	20210709	@SSG	D :	5
5819	김민하	20210710	@SSG	D :	5

5820 rows × 32 columns

PE 프로그램 진행과정 (7.29)

02

전처리 함수 만들기

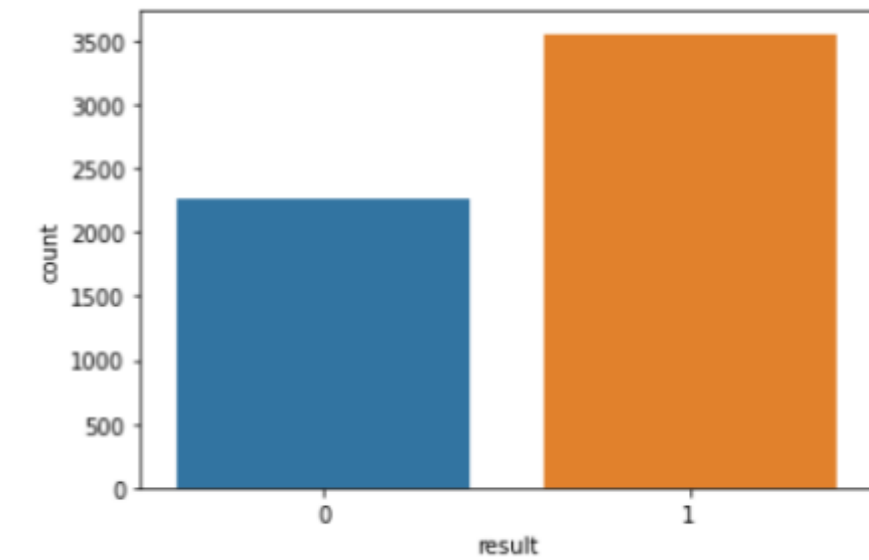
- result, double

- home/away 타율

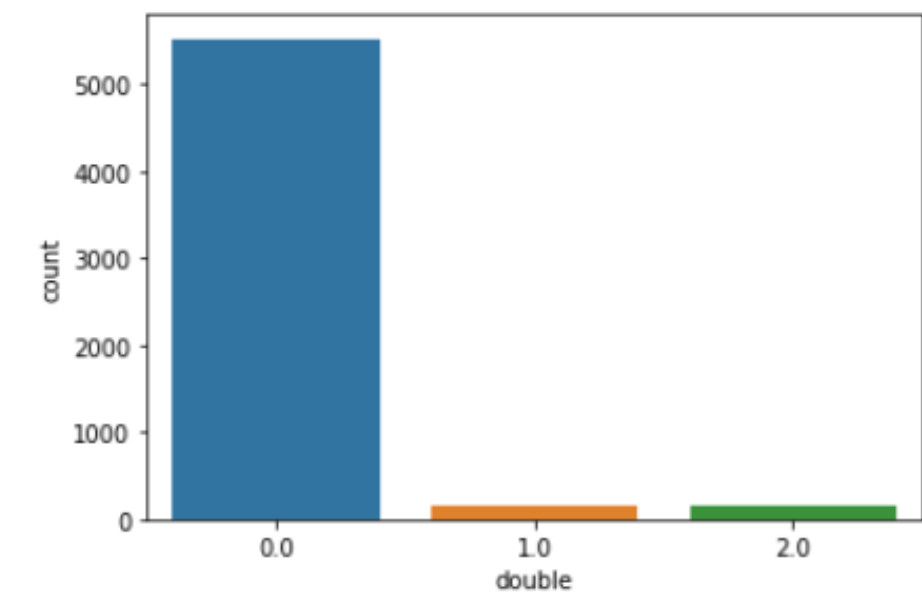
- vs 특정 팀 타율

	name	day	double	vs	vs_team	ab	hit	result
0	강백호	20210404	0.0	한화	한화	3	2	1
1	강백호	20210406	0.0	LG	LG	4	0	0
2	강백호	20210407	0.0	LG	LG	4	3	1
3	강백호	20210408	0.0	LG	LG	4	2	1
4	강백호	20210409	0.0	@삼성	삼성	5	3	1
...
5815	김민하	20210623	0.0	@삼성	삼성	3	1	1
5816	김민하	20210624	0.0	@삼성	삼성	4	0	0
5817	김민하	20210626	0.0	KT	KT	1	0	0
5818	김민하	20210709	0.0	@SSG	SSG	4	1	1
5819	김민하	20210710	0.0	@SSG	SSG	3	1	1

5820 rows × 8 columns



result (0 / 1) 분포 확인



double 분포 확인

PE 프로그램 진행과정 (7.29)

03

전처리 함수 만들기

- result, double

- home/away 타율

- vs 특정 팀 타율

	name	day	double	vs	vs_team	ab	hit	home/away	home_ab	home_hit	away_ab	away_hit	home/away_avg
0	강백호	20210404	0.0	한화	한화	3	2	home	0	0	0	0	0.000
1	강백호	20210406	0.0	LG	LG	4	0	home	3	2	0	0	0.667
2	강백호	20210407	0.0	LG	LG	4	3	home	7	2	0	0	0.286
3	강백호	20210408	0.0	LG	LG	4	2	home	11	5	0	0	0.455
4	강백호	20210409	0.0	@삼성	삼성	5	3	away	15	7	0	0	0.000
...
5815	김민하	20210623	0.0	@삼성	삼성	3	1	away	47	15	46	5	0.109
5816	김민하	20210624	0.0	@삼성	삼성	4	0	away	47	15	49	6	0.122
5817	김민하	20210626	0.0	KT	KT	1	0	home	47	15	53	6	0.319
5818	김민하	20210709	0.0	@SSG	SSG	4	1	away	48	15	53	6	0.113
5819	김민하	20210710	0.0	@SSG	SSG	3	1	away	48	15	57	7	0.123

5820 rows × 13 columns

home / away를 구분하여
전날까지의 누적 타수, 안타 값 구하기

최종적으로 홈/원정 타율 도출

PE 프로그램 진행과정 (7.29)

04

전처리 함수 만들기

- result, double
- home/away 타율
- vs 특정 팀 타율

	name	day	double	vs	vs_team	ab	hit	result	home/away	vs_ab	vs_hit	vs_avg
901	강민호	20210424	0.0	@KIA	KIA	4	2	1	away	0	0	0.000
902	강민호	20210425	0.0	@KIA	KIA	4	2	1	away	4	2	0.500
919	강민호	20210521	0.0	KIA	KIA	3	0	0	home	8	4	0.500
920	강민호	20210523	0.0	KIA	KIA	3	1	1	home	11	4	0.364
928	강민호	20210608	0.0	KIA	KIA	4	2	1	home	14	5	0.357
...
445	황재균	20210612	0.0	한화	한화	5	2	1	home	9	2	0.222
446	황재균	20210613	0.0	한화	한화	2	1	1	home	14	4	0.286
456	황재균	20210625	0.0	@한화	한화	2	0	0	away	16	5	0.312
457	황재균	20210626	0.0	@한화	한화	5	2	1	away	18	5	0.278
458	황재균	20210627	0.0	@한화	한화	4	2	1	away	23	7	0.304

5820 rows × 12 columns



vs_team 컬럼을 이용해 정렬
전날까지의 상대 전적(타수, 안타) 값 구하기

최종적으로 특정팀 상대 타율 도출

PE 프로그램 진행과정 (7.29)

05

전처리 함수 만들기

- result, double
- home/away 타율
- vs 특정 팀 타율

그 외 최근 5경기 타율

최근 안타 성공률 등..

```
recent_5days_data.loc[recent_5days_data['name']=='강민호'].loc[8791:8795,:]
```

	day	name	recent_5days_avg	double
8791	20180405	강민호	0.250000	0
8792	20180407	강민호	0.333333	0
8793	20180408	강민호	0.312500	0
8794	20180410	강민호	0.272727	0
8795	20180411	강민호	0.200000	0

최근 5일 타율 함수

```
recent_5games_data.loc[recent_5games_data['name']=='강민호'].loc[8791:8795,:]
```

	day	name	recent_5games_avg
8791	20180405	강민호	0.352941
8792	20180407	강민호	0.421053
8793	20180408	강민호	0.350000
8794	20180410	강민호	0.263158
8795	20180411	강민호	0.333333

최근 5경기 타율 함수

	day	name	success_10games
0	20180324	나성범	0.000000
1	20180325	나성범	0.000000
2	20180327	나성범	0.500000
3	20180328	나성범	0.666667
4	20180329	나성범	0.500000
...
11311	20181006	강경환	0.400000
11312	20181009	강경환	0.400000
11313	20181019	강경환	0.400000
11314	20181020	강경환	0.400000
11315	20181023	강경환	0.300000

[11316 rows x 3 columns]

최근 10경기 성공률 함수

PE 프로그램 진행과정 (7.29)

06

Join

전처리 함수를 사용해 만든 칼럼들을
Join하여 사용할 최종 데이터 생성

```
final_2021_data = pd.merge(data_2021, ex_data6, on = ['name', 'day', 'double'], how = 'inner')
final_2021_data = pd.merge(final_2021_data, hw_data, on = ['name', 'day', 'double'], how = 'inner')
final_2021_data = pd.merge(final_2021_data, final_vs_avg_data, on = ['name', 'day', 'double'], how = 'inner')
final_2021_data
```

	name	day	vs_x	result_x	bat_order	position	start_member	ab_x	score	hit_x	2_hit	3_hit	homerun
0	강백호	20210404	한화	W 3:2	4	1B	True	3	0	2	0	0	0
1	강백호	20210406	LG	L 2:3	4	1B	True	4	0	0	0	0	0
2	강백호	20210407	LG	W 7:3	4	1B	True	4	0	3	0	0	0
3	강백호	20210408	LG	L 3:7	4	1B	True	4	0	2	0	0	0
4	강백호	20210409	@삼성	L 5:7	4	1B	True	5	0	3	0	1	0
...
5815	김민하	20210623	@삼성	L 0:3	7	LF	True	3	0	1	0	1	0
5816	김민하	20210624	@삼성	L 2:3	6	RF	True	4	0	0	0	0	0
5817	김민하	20210626	KT	L 2:11	3	RF	True	1	0	0	0	0	0
5818	김민하	20210709	@SSG	D :	5	CF	True	4	0	1	0	0	0
5819	김민하	20210710	@SSG	D :	5	CF	True	3	1	1	0	0	0

5820 rows x 62 columns

vs_ab	vs_hit	vs_avg
0	0	0.000
0	0	0.000
4	0	0.000
8	3	0.375
0	0	0.000
...
10	1	0.100
13	2	0.154
7	1	0.143
14	0	0.000
18	1	0.056

...

PE 프로그램 진행과정 (7.29)

07

To do list

1. 2020, 2021년을 합친 데이터를
모델 실험에 사용
2. 지난번 MLP 모델에 적용하여
데이터 수 증가 후 정확도 비교
3. 정확도 90%
4. MLP를 보완할 모델 탐색

```
133/133 [-----] - 1s 5ms/step - loss: 0.5814 - accuracy: 0.6969
Epoch 93/100
133/133 [=====] - 1s 5ms/step - loss: 0.5814 - accuracy: 0.6969
Epoch 94/100
133/133 [=====] - 1s 5ms/step - loss: 0.5913 - accuracy: 0.6775
Epoch 95/100
133/133 [=====] - 1s 5ms/step - loss: 0.5947 - accuracy: 0.6725
Epoch 96/100
133/133 [=====] - 1s 5ms/step - loss: 0.5861 - accuracy: 0.6953
Epoch 97/100
133/133 [=====] - 1s 5ms/step - loss: 0.5812 - accuracy: 0.6939
Epoch 98/100
133/133 [=====] - 1s 5ms/step - loss: 0.5874 - accuracy: 0.6776
Epoch 99/100
133/133 [=====] - 1s 5ms/step - loss: 0.5815 - accuracy: 0.6928
Epoch 100/100
133/133 [=====] - 1s 5ms/step - loss: 0.5951 - accuracy: 0.6689
```

```
In [46]: loss_and_metrics = model.evaluate(X_test, y_test, batch_size = 32)
```

```
34/34 [=====] - 0s 2ms/step - loss: 0.5815 - accuracy: 0.6953
```