

# 2017년 삼성 라이온즈 데이터를 이용한 모의 스토브 리그

타자 기준

안용희, 안광필, 김정모

2021.05.14

# 2017 년 기준 *KBO* LEAGUE 구단 현황



넥센 히어로즈



SK와이번스



LGE트윈스



KT wiz



KIA 타이거즈



두산 베어스



한화 이글스



삼성 라이온즈



롯데 자이언츠



NC 다이노스



# 2021 **BIG Data** Project



## 삼성 라이온즈

1. 프로젝트 배경 및 목표
2. 삼성 팀 중 트레이드 선수 명단 작업
3. 영입할 선수 명단 작업
4. 연봉 예측 작업
5. 결론



# 1. 프로젝트 배경 및 목표



# ▲ 팀	경기	승	무	패	득점	승률
1. KIA	144	87	1	56	906:743	0.608
2. 두산	144	84	3	57	849:678	0.594
3. 롯데	144	80	2	62	743:701	0.563
4. NC	144	79	3	62	786:745	0.559
5. SSG	144	75	1	68	761:767	0.524
6. LG	144	69	3	72	699:677	0.490
7. 키움	144	69	2	73	789:764	0.486
8. 한화	144	61	2	81	737:820	0.431
9. 삼성	144	55	5	84	757:911	0.399
10. kt	144	50	0	94	655:876	0.347

## Project 배경

2011~2014 4년 연속우승, 2015 준우승

2016~2017 2년 연속 9위

창단 이래 2년 연속 역대 **최악의 성적**

## Project 목표

모의 스토브리그로 데이터를 비교 분석하여

2018년 성적 향상을 모색

## 2. 삼성 팀 중 트레이드 명단 작업



### 1) 삼성과 프로야구 전체 스탯 평균의 비교

	경기	안타	홈런	득점	타점	볼넷	삼진	도루	안타율	타율	출루,장타율	가중출루율	비교기여도	연봉(2017)
전체평균	77.71	62.26	6.36	33.28	30.82	19.34	42.15	3.67	0.30	0.25	0.67	0.29	0.79	19111.52
삼성평균	77.53	63.32	6.11	33.95	30.47	19.32	42.11	4.84	0.28	0.24	0.68	0.29	0.62	18494.74

Z-Score Normalization 정규화

이유: 각 스탯의 숫자 크기의 차이 큼

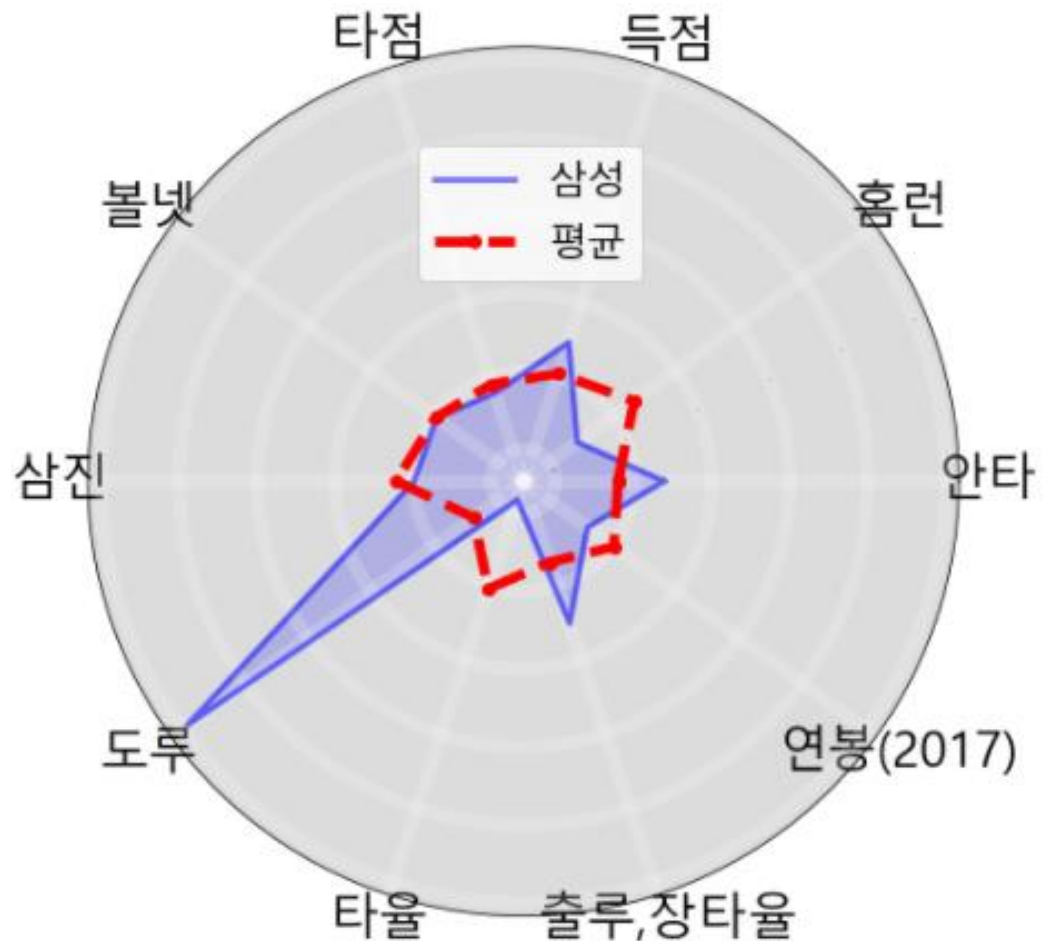


```
def standard_scaling(df, scale_columns):  
    for col in scale_columns:  
        series_mean = df[col].mean()  
        series_std = df[col].std()  
        df[col] = df[col].apply(lambda x : (x - series_mean)/(series_std))  
    return df
```

	안타	홈런	득점	타점	볼넷	삼진	도루	타율	출루,장타율	연봉(2017)
전체평균	-0.010	0.017	0.002	-0.005	-0.001	0.011	-0.031	0.002	-0.016	0.001
삼성평균	0.019	-0.028	0.023	-0.011	-0.002	-0.001	0.196	-0.058	0.025	-0.020

## 2. 삼성 팀 중 트레이드 명단 작업

### 2) 삼성과 프로야구 프로야구 팀 스탯 평균의 비교



**장점 :** 출루, 장타율, 득점

**단점 :** 타율, 홈런

타율, 홈런이 좋은 선수 영입

타율과 홈런이 낮은 선수 트레이드

## 2. 삼성 팀 중 트레이드 명단 작업



### 3) 삼성 평균 타율보다 낮은 선수 명단

# 삼성 평균 타율 보다 낮은 선수 명단     bat\_ss\_df : 정규화 된 삼성 DataFrame

```
bat_ss_df[bat_ss_df['타율'] < bat_ss_df['타율'].sum().mean()]
```

	선수명	팀명	안타	홈런	득점	타점	볼넷	삼진	도루	타율	출루,장타율	연봉(2017)
132	김성윤	삼성	-1.12	-0.62	-1.02	-0.91	-0.89	-1.06	-0.62	-1.62	-0.50	-0.53
182	백상원	삼성	-1.03	-0.74	-0.98	-0.91	-0.84	-0.94	-0.62	-1.16	-1.43	-0.33

## 2. 삼성 팀 중 트레이드 명단 작업



### 4) 삼성 평균 홈런 보다 낮은 선수 명단

# 삼성 평균 홈런 보다 낮은 선수 명단

```
bat_ss_df[bat_ss_df['홈런'] < bat_ss_df['홈런'].sum().mean()]
```

	선수명	팀명	안타	홈런	득점	타점	볼넷	삼진	도루	타율	출루, 장타율	연봉(2017)
63	김성훈	삼성	-0.26	-0.74	-0.21	-0.41	-0.38	-0.79	0.06	0.67	0.27	-0.53
78	강한울	삼성	1.14	-0.74	0.83	-0.22	0.34	0.45	1.40	0.52	0.05	-0.33
105	최영진	삼성	-1.06	-0.62	-0.98	-0.85	-0.94	-1.15	-0.45	-0.48	0.46	-0.52
116	안주형	삼성	-1.08	-0.74	-1.12	-0.91	-0.99	-1.24	-0.62	0.49	0.11	-0.53
132	김성운	삼성	-1.12	-0.62	-1.02	-0.91	-0.89	-1.06	-0.62	-1.62	-0.50	-0.53
144	이현동	삼성	-1.08	-0.74	-1.08	-0.98	-0.99	-1.06	-0.45	-0.60	-0.93	-0.53
152	김민수	삼성	-1.08	-0.74	-1.05	-0.88	-0.94	-1.03	-0.62	-1.03	-1.19	-0.52
182	백상원	삼성	-1.03	-0.74	-0.98	-0.91	-0.84	-0.94	-0.62	-1.16	-1.43	-0.33
184	이지영	삼성	0.18	-0.74	-0.08	-0.15	0.03	-0.06	0.06	-0.11	-0.37	0.22



## 2. 삼성 팀 중 트레이드 명단 작업



### 5) 타율과 홈런이 프로야구 평균 보다 낮은 선수 명단

# 타율이 전체 평균 보다 낮은 선수 들중에서 타율과 홈런 내림차순

```
a = bat_ss_df.query("타율 < 0.002 and 홈런 < 0.017").sort_values(['타율', '홈런'])  
a[['선수명', '타율', '홈런', '연봉(2017)']].sort_values(by = ['연봉(2017)', '타율', '홈런'], ascending = False)
```

	선수명	타율	홈런	연봉(2017)
184	이지영	-0.11	-0.74	0.22
182	백상원	-1.16	-0.74	-0.33
159	정병곤	-0.32	-0.51	-0.51
87	권정웅	-0.36	-0.04	-0.52
105	최영진	-0.48	-0.62	-0.52
152	김민수	-1.03	-0.74	-0.52
144	이현동	-0.60	-0.74	-0.53
132	김성윤	-1.62	-0.62	-0.53

Query 함수는 아래 6가지 기능을 포함하고 있습니다.

- 1) 비교 연산자(==, >, >=, <, <=, !=)
- 2) in 연산자(in, ==, not in, !=)
- 3) 논리 연산자(and, or, not)
- 4) 외부 변수(또는 함수) 참조 연산
- 5) 인덱스 검색
- 6) 문자열 부분검색( str.contains, str.startswith, str.endswith )

## 2. 삼성 팀 중 트레이드 명단 작업



### 6) 삼성 라이온즈에서 트레이드 할 선수 명단

**트레이드 명단**

**이지영 선수, 김성윤 선수**



### 3. 영입할 선수 명단 작업



#### 1) 이지영 선수 대체 할 선수 영입

##### 영입 조건

- 트레이드 선수끼리의 몸값은 비슷해야 한다.
- 프로야구선수 전체 평균 보다 높은 선수여야 한다. (타율, 홈런 기준)

```
# 연봉 0.22 , -0.53
```

bat\_df : 정규화 된 모든 프로야구 팀 DataFrame

```
bat_df[bat_df['연봉(2017)'] == 0.02]
```

```
#bat_df[['선수명', '타율', '홈런', '연봉(2017)']].head()
```

	선수명	팀명	안타	홈런	득점	타점	볼넷	삼진	도루	타율	출루, 장타율	연봉(2017)
20	오재일	두산	1.16	2.29	0.96	1.84	1.31	1.14	-0.45	0.55	1.05	0.02

### 3. 영입할 선수 명단 작업

#### 2) 김성윤 선수 대체 할 선수 영입

	선수명	팀명	타율	홈런
31	정현	KT	0.49	-0.04
18	이정후	넥센	0.72	-0.51
79	강진성	NC	0.45	-0.62
68	황진수	롯데	0.40	-0.62
160	도태훈	NC	-0.57	-0.62
132	김성윤	삼성	-1.62	-0.62
167	김태연	한화	-1.96	-0.62

```
b = bat_df[bat_df['연봉(2017)'] == -0.53]
c = b.query('타율 > 0.002 or 홈런 > 0.017').sort_values(['타율', '홈런'])
#c[['선수명', '타율', '홈런']].sort_values(by = ['타율', '홈런'], ascending = False)
b[['선수명', '팀명', '타율', '홈런']].sort_values(by = ['홈런', '타율'], ascending = False)
```



### 3. 영입할 선수 명단 작업

#### 3) 오재일 선수와 정현 선수 영입

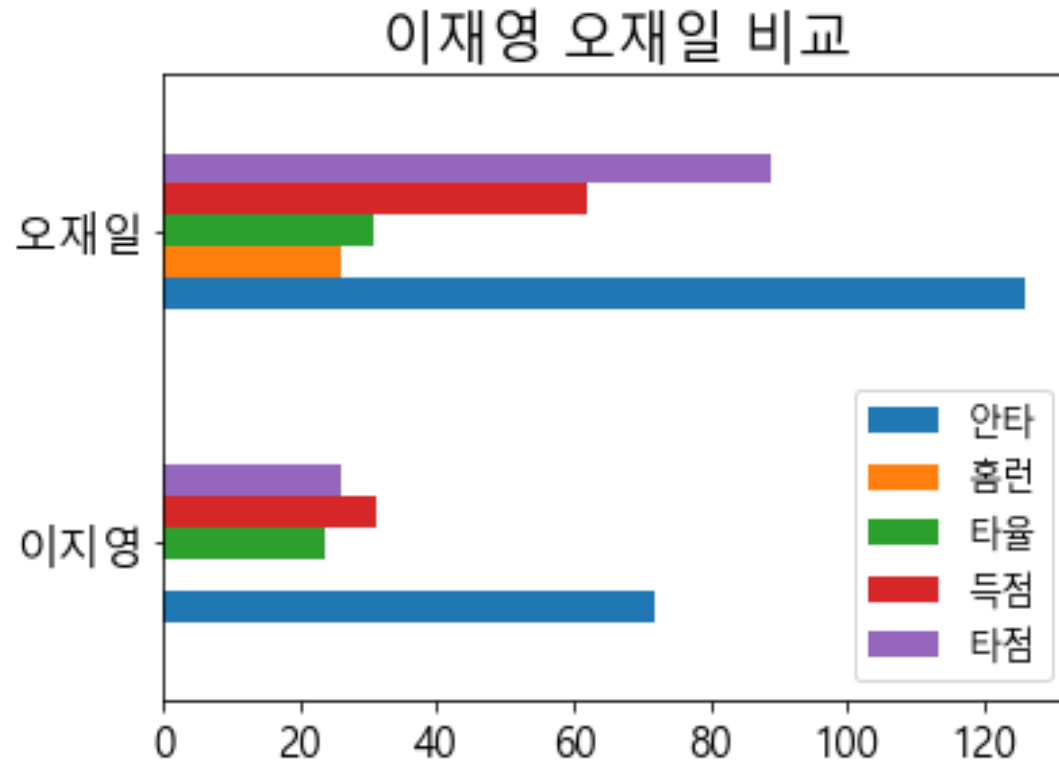
	선수명	팀명	안타	홈런	득점	타점	볼넷	삼진	도루	타율	출루, 장타율	연봉(2017)
184	이지영	삼성	0.18	-0.74	-0.08	-0.15	0.03	-0.06	0.06	-0.11	-0.37	0.22
132	김성운	삼성	-1.12	-0.62	-1.02	-0.91	-0.89	-1.06	-0.62	-1.62	-0.5	-0.53



	선수명	팀명	안타	홈런	득점	타점	볼넷	삼진	도루	타율	출루, 장타율	연봉(2017)
20	오재일	두산	1.16	2.29	0.96	1.84	1.31	1.14	-0.45	0.55	1.05	0.02
31	정현	KT	0.78	-0.04	0.39	0.35	-0.07	0.81	0.06	0.49	0.48	-0.53

### 3. 영입할 선수 명단 작업

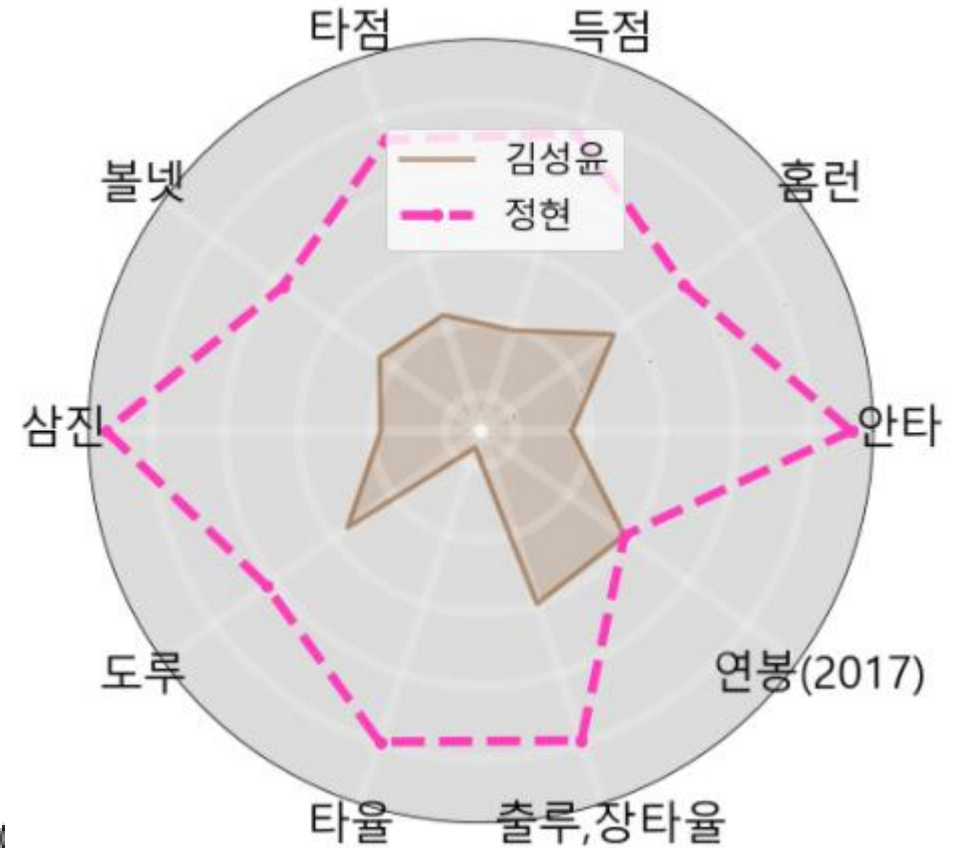
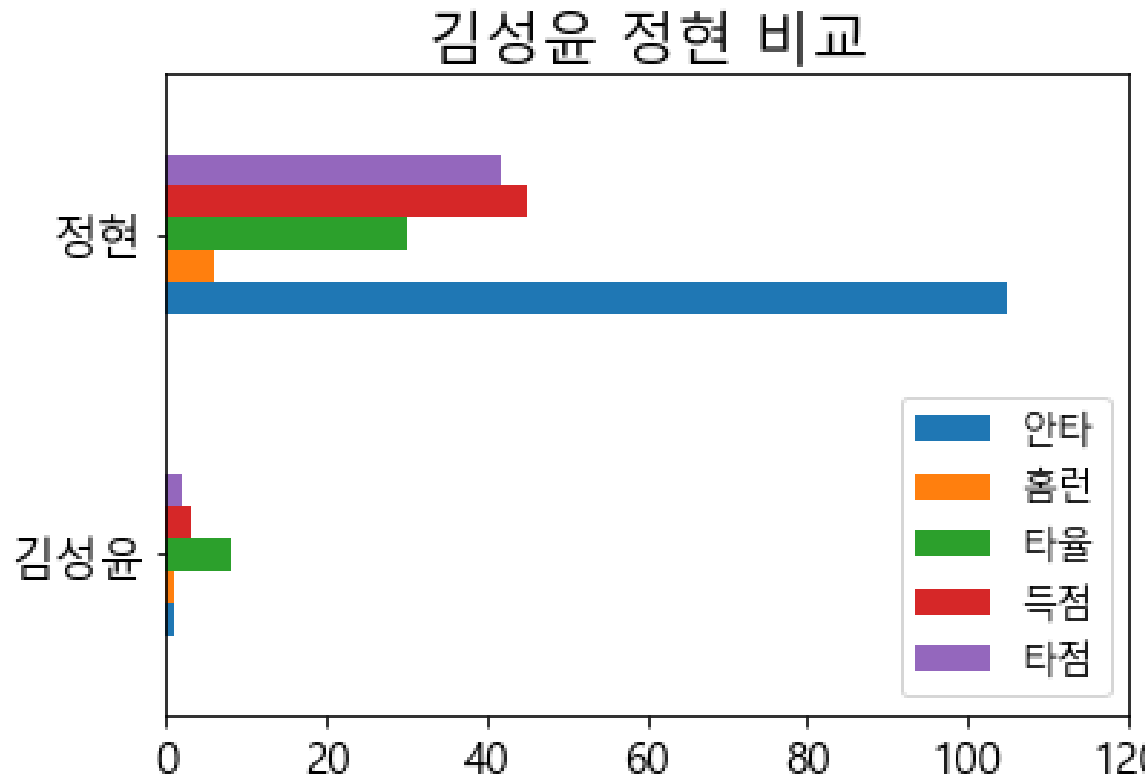
#### 4) 오재일 선수와 이지영 선수 스탯 비교





### 3. 영입할 선수 명단 작업

#### 5) 정현 선수와 김성윤 선수 스탯 비교



### 3. 영입할 선수 명단 작업

#### 6) 오재일 선수와 정현 선수 영입 후 삼성과 프로야구 팀 비교



#### 분석

스탯 중 제거한 항목이 많아 부정확  
하지만 목표인 홈런과 타율은 **상승**

2017년 영입 한 선수들의 기량이 상승하여  
다음 시즌에는 연봉이 상승 할 것으로 보아  
팀 연봉은 **상승** 할 것으로 예상



# 4. 연봉 예측 작업

## 1) 원-핫 인코딩을 통한 팀 연봉 예측

```
# 범주형 데이터 '팀명'을 정규화 : 원핫인코딩 - get_dummies()  
team_scaling = pd.get_dummies(batman_df['팀명'])  
batman_df = batman_df.drop('팀명', axis = 1)  
batman_df = batman_df.join(team_scaling)  
team_scaling.head()
```

	KIA	KT	LG	NC	SK	넥센	두산	롯데	삼성	한화
0	0	0	0	0	1	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	1	0	0	0
3	0	0	0	1	0	0	0	0	0	0
4	0	0	0	0	0	0	0	1	0	0

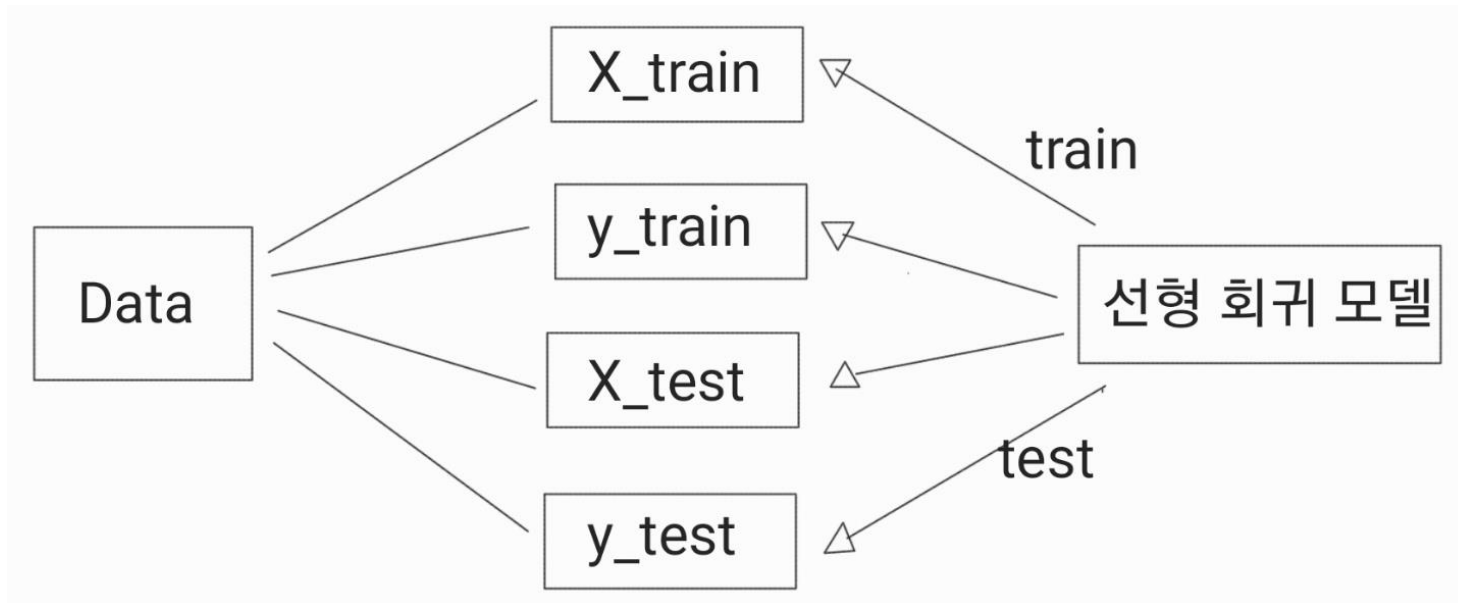
```
def sum_salary(x):  
    a = []  
    for i in x:  
        b= batman_df[batman_df['선수명']==i]['예측연봉(2018)'].values  
        a.append(float(b))  
    c = sum(a)  
    return c
```

원-핫 인코딩을 통하여  
삼성 팀의 총 연봉을 쉽게 보기 위하여 사용

팀 선수 이름 입력만 하면  
총 연봉을 계산

# 4. 연봉 예측 작업

## 2) 삼성 라이온즈 연봉 예측 작업



train 평가 정확도 : 93%

test 평가 정확도 : 96%

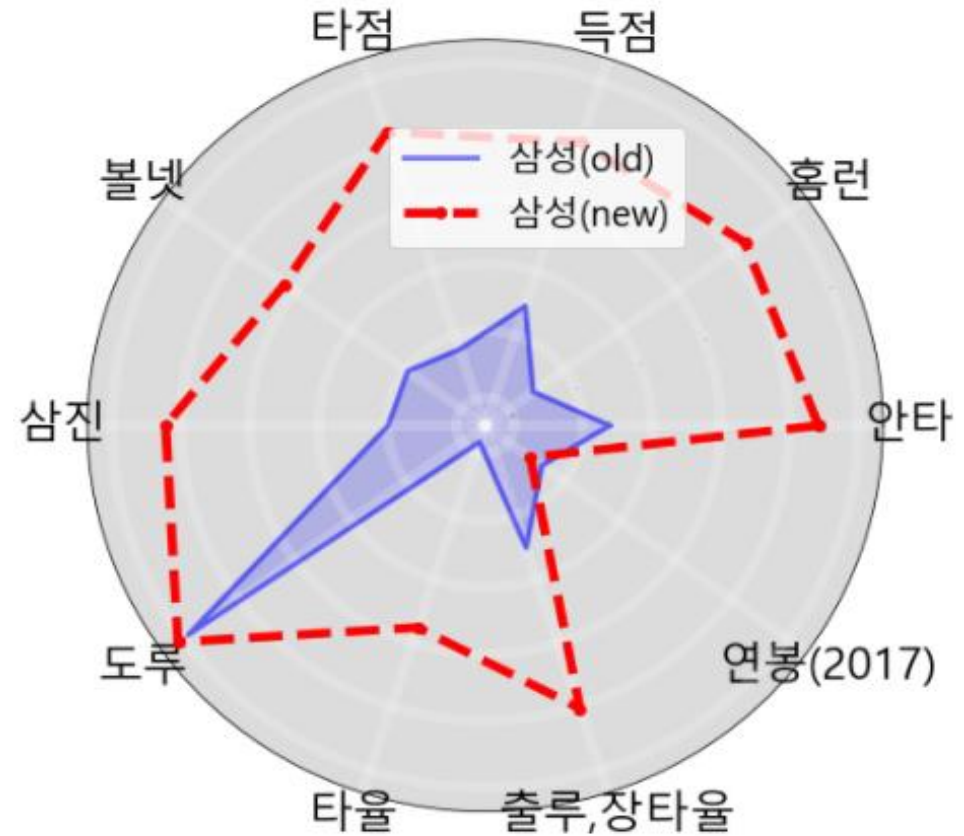
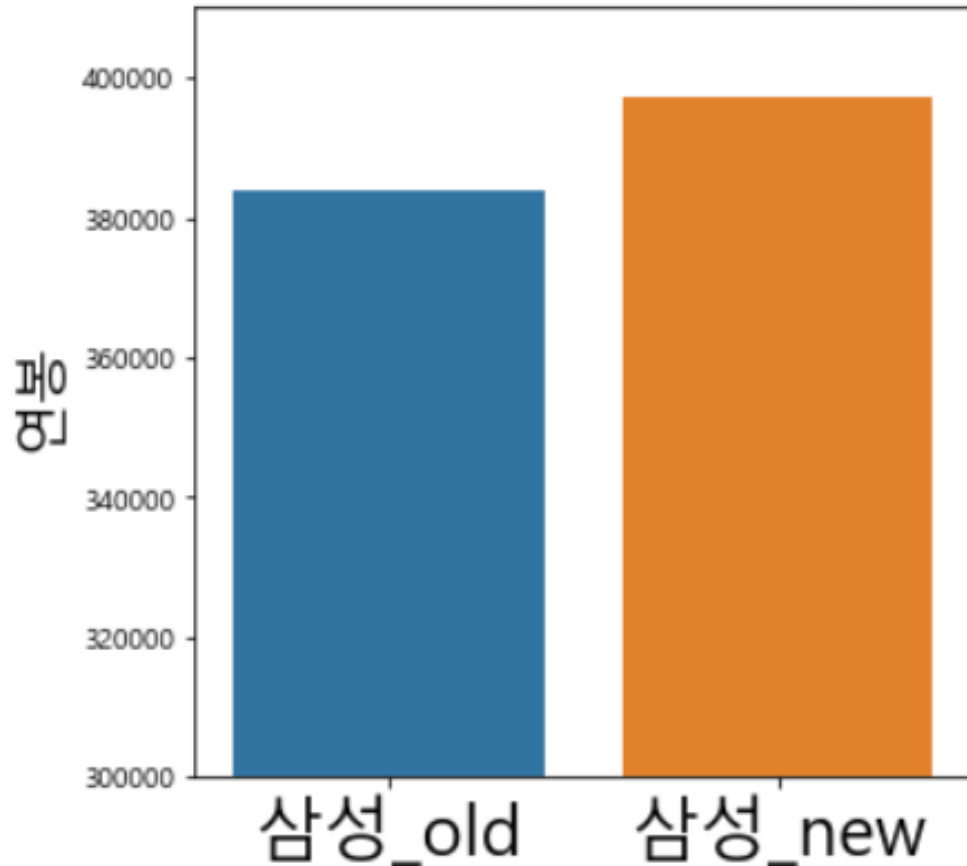
2017년 데이터를 위 그림 처럼 train 데이터와 test 데이터 생성  
선형회귀 모델로 데이터를 학습하고 평가로 다음 시즌 평균 연봉을 예측 계산

train와 test 각각 93%, 96% => **우수**하다고 판단

## 4. 연봉 예측 작업

### 3) 현재 연봉과 예측 연봉 및 스탯 비교 시각화

2018 삼성 라이온즈 연봉 예측 비교





# 5. 결론



## 결론

- 트레이드 전후 예측 연봉이 13,415.27 ( 만원 ) 차이가 났다.
- 팀의 능력치가 프로야구팀 전체 평균보다 크게 **향상**되었다.
- 2018 프로야구는 좋은 성적을 기대해 볼 수 있을 것 같다.

## 실제

- 이지영 선수가 넥센 히어로즈 이적
- 오재일 선수가 삼성 라이온즈 이적

# 5. 결론



## 이지영 넥센·고종욱 SK·김동엽 삼성행...삼각 트레이드 성사

송고시간 | 2018-12-07 14:20



넥센 히어로즈 유니폼을 입은 이지영.  
[연합뉴스 자료사진]

## [SC핫이슈]오재일 삼성행으로 '두산발 FA 빅 3' 모두 계약, 정수빈 차례?

스포츠조선=김진희 기자  
입력 2020.12.14 16:27



두산 오재일. 고척=박재만 기자 pjn@sportschosun.com



## 참고 및 출처

야구 순위 이미지: <https://www.flashscore.co.kr/baseball/south-korea/kbo-2017/standings/>

야구 역대 순위 : [https://www.samsunglions.com/history/stat\\_5.asp/](https://www.samsunglions.com/history/stat_5.asp/)

정규화 사이트: <http://hleecaster.com/ml-normalization-concept/>

Query 함수 : <https://m.blog.naver.com/wideeyed/221867273249/>

Matplotlib :  
[https://matplotlib.org/stable/api/\\_as\\_gen/matplotlib.pyplot.plot.html#matplotlib.pyplot.plot/](https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.plot.html#matplotlib.pyplot.plot/)

Polar plot : <https://velog.io/@khnn/TIL-Matplotlib으로-차트-꾸미기/>

이지영 기사: <https://www.yna.co.kr/view/AKR20181207096200007/>

오재일 기사:  
[https://www.chosun.com/sports/sports\\_photo/2020/12/14/A70PQQW7740P3FPJAXVLZ3243A/](https://www.chosun.com/sports/sports_photo/2020/12/14/A70PQQW7740P3FPJAXVLZ3243A/)

# 감사합니다