

1. 팀: (1) 시즌 참여 팀 (2) 팀의 수
2. 경기: (1) 홈 / 원정 경기 수 (2) 홈 / 원정 경기 승패 (승률)
(3) 전체 경기 승률, (4) 순위
3. 선수: (1) 팀 선수 (종합 / 투수 / 타자) (2) 팀 연령 (종합 / 투수 / 타자),
(3) 팀 연봉 (종합 / 투수 / 타자)
4. 피어슨 상관계수로 알아보는 상관관계
 - 주요 변수: 시즌 순위, 시즌 승률, 평균 연령 (종합 / 투수 / 타자), 평균 연봉 (종합 / 투수 / 타자)
5. 결과 종합

1. 팀

(1) 시즌 참여 팀

1. 팀

- 프로야구 팀 (2016 ~ 2019)

```
In [3]: 1 #2016년 시즌
2 doc2016_1 = pd.read_excel(PATH + doc2016, sheet_name = "1.팀")
3 doc2016_1
4 #2017년 시즌
5 doc2017_1 = pd.read_excel(PATH + doc2017, sheet_name = "1.팀")
6 doc2017_1
7 #2018년 시즌
8 doc2018_1 = pd.read_excel(PATH + doc2018, sheet_name = "1.팀")
9 doc2018_1
10 #2019년 시즌
11 doc2019_1 = pd.read_excel(PATH + doc2019, sheet_name = "1.팀")
12 doc2019_1
```

Out[3]:

	T_ID	T_NM
0	HH	한화
1	HT	KIA
2	KT	KT
3	LG	LG
4	LT	롯데
5	NC	NC
6	OB	두산
7	SK	SK
8	SS	삼성
9	WO	키움

(2) 시즌 별 팀의 수

- 연도별 프로야구 팀 수 (2016 ~ 2019)

```
In [4]: 1 print("2016년도 프로야구 총 팀의 수 : ", doc2016_1['T_NM'].count())
2 print("2017년도 프로야구 총 팀의 수 : ", doc2017_1['T_NM'].count())
3 print("2018년도 프로야구 총 팀의 수 : ", doc2018_1['T_NM'].count())
4 print("2019년도 프로야구 총 팀의 수 : ", doc2019_1['T_NM'].count())
```

```
2016년도 프로야구 총 팀의 수 : 10
2017년도 프로야구 총 팀의 수 : 10
2018년도 프로야구 총 팀의 수 : 10
2019년도 프로야구 총 팀의 수 : 10
```

- 2016 ~ 2019 시즌 참여 구단은 10개 구단.

- 각 시트 별 팀 아이디, 코드의 순서와 이름이 동일.

2. 경기

- 시즌 별 홈 / 원정 경기, 승률, 및 전체 경기 결과와 순위

- 2016 시즌 결과

팀	홈 경기	승률	무패	패	홈 승률	원정 경기	승률	무패	원정 승률	전체 경기	승률	무패	패	종합 승률	등수
HH	HOME: 72 [36, 2, 34]	0.514				AWAY: 72 [30, 1, 41]	0.423			TOTAL: 144	66	3	75	0.468	7
HT	HOME: 72 [38, 0, 34]	0.528				AWAY: 72 [32, 1, 39]	0.451			TOTAL: 144	70	1	73	0.49	5
KT	HOME: 72 [28, 0, 44]	0.389				AWAY: 72 [25, 2, 45]	0.357			TOTAL: 144	53	2	89	0.373	10
LG	HOME: 72 [40, 2, 30]	0.571				AWAY: 72 [31, 0, 41]	0.431			TOTAL: 144	71	2	71	0.5	4
LT	HOME: 72 [40, 0, 32]	0.556				AWAY: 72 [26, 0, 46]	0.361			TOTAL: 144	66	0	78	0.458	8
NC	HOME: 72 [39, 2, 31]	0.557				AWAY: 72 [44, 1, 27]	0.62			TOTAL: 144	83	3	58	0.589	2
OB	HOME: 72 [46, 1, 25]	0.648				AWAY: 72 [47, 0, 25]	0.653			TOTAL: 144	93	1	50	0.65	1
SK	HOME: 72 [35, 0, 37]	0.486				AWAY: 72 [34, 0, 38]	0.472			TOTAL: 144	69	0	75	0.479	6
SS	HOME: 72 [34, 0, 38]	0.472				AWAY: 72 [31, 1, 40]	0.437			TOTAL: 144	65	1	78	0.455	9
WO	HOME: 72 [44, 0, 28]	0.611				AWAY: 72 [33, 1, 38]	0.465			TOTAL: 144	77	1	66	0.538	3

- 2017 시즌 결과

팀	홈 경기	승률	무패	패	홈 승률	원정 경기	승률	무패	원정 승률	전체 경기	승률	무패	패	종합 승률	등수
HH	HOME: 72 [30, 1, 41]	0.423				AWAY: 72 [31, 1, 40]	0.437			TOTAL: 144	61	2	81	0.43	8
HT	HOME: 72 [45, 0, 27]	0.625				AWAY: 72 [42, 1, 29]	0.592			TOTAL: 144	87	1	56	0.608	1
KT	HOME: 72 [27, 0, 45]	0.375				AWAY: 72 [23, 0, 49]	0.319			TOTAL: 144	50	0	94	0.347	10
LG	HOME: 72 [38, 0, 34]	0.528				AWAY: 72 [31, 3, 38]	0.449			TOTAL: 144	69	3	72	0.489	6
LT	HOME: 72 [45, 2, 25]	0.643				AWAY: 72 [35, 0, 37]	0.486			TOTAL: 144	80	2	62	0.563	3
NC	HOME: 72 [44, 1, 27]	0.62				AWAY: 72 [35, 2, 35]	0.5			TOTAL: 144	79	3	62	0.56	4
OB	HOME: 72 [39, 3, 30]	0.565				AWAY: 72 [45, 0, 27]	0.625			TOTAL: 144	84	3	57	0.596	2
SK	HOME: 72 [42, 0, 30]	0.583				AWAY: 72 [33, 1, 38]	0.465			TOTAL: 144	75	1	68	0.524	5
SS	HOME: 72 [28, 2, 42]	0.4				AWAY: 72 [27, 3, 42]	0.391			TOTAL: 144	55	5	84	0.396	9
WO	HOME: 72 [39, 2, 31]	0.557				AWAY: 72 [30, 0, 42]	0.417			TOTAL: 144	69	2	73	0.486	7

- 2018 시즌 결과

팀	홈 경기	승률	무패	패	홈 승률	원정 경기	승률	무패	원정 승률	전체 경기	승률	무패	패	종합 승률	등수
HH	HOME: 72 [43, 0, 29]	0.597				AWAY: 72 [34, 0, 38]	0.472			TOTAL: 144	77	0	67	0.535	3
HT	HOME: 72 [43, 0, 29]	0.597				AWAY: 72 [27, 0, 45]	0.375			TOTAL: 144	70	0	74	0.486	6
KT	HOME: 72 [31, 2, 39]	0.443				AWAY: 72 [28, 1, 43]	0.394			TOTAL: 144	59	3	82	0.418	9
LG	HOME: 72 [39, 1, 32]	0.549				AWAY: 72 [29, 0, 43]	0.403			TOTAL: 144	68	1	75	0.476	8
LT	HOME: 72 [34, 0, 38]	0.472				AWAY: 72 [34, 2, 36]	0.486			TOTAL: 144	68	2	74	0.479	7
NC	HOME: 72 [33, 0, 39]	0.458				AWAY: 72 [25, 1, 46]	0.352			TOTAL: 144	58	1	85	0.406	10
OB	HOME: 72 [51, 0, 21]	0.708				AWAY: 72 [42, 0, 30]	0.583			TOTAL: 144	93	0	51	0.646	1
SK	HOME: 72 [42, 0, 30]	0.583				AWAY: 72 [36, 1, 35]	0.507			TOTAL: 144	78	1	65	0.545	2
SS	HOME: 72 [33, 3, 36]	0.478				AWAY: 72 [35, 1, 36]	0.493			TOTAL: 144	68	4	72	0.486	6
WO	HOME: 72 [37, 0, 35]	0.514				AWAY: 72 [38, 0, 34]	0.528			TOTAL: 144	75	0	69	0.521	4

- 2019 시즌 결과

팀	홈 경기	승률	무패	패	홈 승률	원정 경기	승률	무패	원정 승률	전체 경기	승률	무패	패	종합 승률	등수
HH	HOME: 72 [31, 0, 41]	0.431				AWAY: 72 [27, 0, 45]	0.375			TOTAL: 144	58	0	86	0.403	9
HT	HOME: 72 [39, 0, 33]	0.542				AWAY: 72 [23, 2, 47]	0.329			TOTAL: 144	62	2	80	0.437	7
KT	HOME: 72 [41, 1, 30]	0.577				AWAY: 72 [30, 1, 41]	0.423			TOTAL: 144	71	2	71	0.5	6
LG	HOME: 72 [41, 1, 30]	0.577				AWAY: 72 [38, 0, 34]	0.528			TOTAL: 144	79	1	64	0.552	4
LT	HOME: 72 [29, 1, 42]	0.408				AWAY: 72 [19, 2, 51]	0.271			TOTAL: 144	48	3	93	0.34	10
NC	HOME: 72 [40, 2, 30]	0.571				AWAY: 72 [33, 0, 39]	0.458			TOTAL: 144	73	2	69	0.514	5
OB	HOME: 72 [47, 0, 25]	0.653				AWAY: 72 [41, 1, 30]	0.577			TOTAL: 144	88	1	55	0.615	2
SK	HOME: 72 [44, 1, 27]	0.62				AWAY: 72 [44, 0, 28]	0.611			TOTAL: 144	88	1	55	0.615	2
SS	HOME: 72 [36, 0, 36]	0.5				AWAY: 72 [24, 1, 47]	0.338			TOTAL: 144	60	1	83	0.42	8
WO	HOME: 72 [45, 1, 26]	0.634				AWAY: 72 [41, 0, 31]	0.569			TOTAL: 144	86	1	57	0.601	3

3. 선수

- 2016 시즌 결과

팀	총선수	투수	타자	평균 연령	투수 평균	타자 평균	평균 연봉	투수 평균	타자 평균	시즌 순위
HH	74	32	42	29.23	29.44	29.07	19172.74	22915.09	16321.43	7
HT	67	32	35	27.37	27.25	27.49	15308.96	19509.38	11468.57	5
KT	65	30	35	27.06	25.97	28.0	10793.85	10309.33	11214.29	10
LG	66	31	35	27.94	28.1	27.8	13760.61	13645.16	13862.86	4
LT	70	35	35	27.4	27.2	27.6	15135.71	15434.29	14837.14	8
NC	67	32	35	25.94	24.66	27.11	12667.16	8737.5	16260.0	2
OB	67	36	31	26.9	27.08	26.68	14655.22	15255.56	13958.06	1
SK	68	38	30	27.13	26.89	27.43	13760.29	11284.21	16896.67	6
SS	67	34	33	27.94	28.0	27.88	16267.16	16941.18	15572.73	9
WO	64	34	30	25.58	25.41	25.77	8843.75	7455.88	10416.67	3

- 2017 시즌 결과

팀	총선수	투수	타자	평균 연령	투수 평균	타자 평균	평균 연봉	투수 평균	타자 평균	시즌 순위
HH	70	36	34	27.89	28.25	27.5	12497.14	15255.56	9576.47	8
HT	68	33	35	27.25	27.73	26.8	16480.88	12648.48	20094.29	1
KT	67	29	38	27.0	25.07	28.47	13347.76	9348.28	16400.0	10
LG	65	30	35	26.58	27.2	26.06	9810.77	9893.33	9740.0	6
LT	67	33	34	26.91	26.58	27.24	13961.19	15509.09	12458.82	3
NC	66	29	37	27.71	27.17	28.14	13406.06	8917.24	16924.32	4
OB	64	35	29	26.3	25.97	26.69	18592.19	16362.86	21282.76	2
SK	62	31	31	28.56	27.48	29.65	13541.94	10164.52	16919.35	5
SS	66	32	34	26.65	27.28	26.06	15040.65	17546.34	12682.35	9
WO	66	36	30	25.53	25.89	25.1	11018.18	13433.33	8120.0	7

- 2018 시즌 결과

팀	총선수	투수	타자	평균 연령	투수 평균	타자 평균	평균 연봉	투수 평균	타자 평균	시즌 순위
HH	67	35	32	27.9	28.4	27.34	18698.51	26062.86	10643.75	3
HT	67	33	34	26.84	26.52	27.15	11955.22	11963.64	11947.06	6
KT	65	32	33	26.38	25.31	27.42	14927.43	16468.75	13432.82	9
LG	63	30	33	25.92	25.27	26.52	10015.87	4593.33	14945.45	8
LT	64	35	29	27.77	28.74	26.59	16650.0	17377.14	15772.41	7
NC	68	28	40	27.25	27.46	27.1	13458.82	12564.29	14085.0	10
OB	68	37	31	25.56	25.65	25.45	11952.94	14291.89	9161.29	1
SK	65	35	30	27.77	27.37	28.23	11016.92	10368.57	11773.33	2
SS	66	32	34	27.74	28.72	26.82	14446.97	17856.25	11238.24	6
WO	59	31	28	26.58	25.16	28.14	13191.53	8135.48	18789.29	4

- 2019 시즌 결과

팀	총선수	투수	타자	평균 연령	투수 평균	타자 평균	평균 연봉	투수 평균	타자 평균	시즌 순위
HH	66	30	36	27.17	27.27	27.08	13465.15	12060.0	14636.11	9
HT	70	31	39	26.7	27.06	26.41	14435.71	13287.1	15348.72	7
KT	64	36	28	26.78	27.28	26.14	12978.12	10344.44	16364.29	6
LG	69	35	34	27.36	27.26	27.47	15356.52	15542.86	15164.71	4
LT	67	34	33	26.82	27.12	26.52	12716.42	15702.94	9639.39	10
NC	62	30	32	27.68	27.33	28.0	12243.55	14190.0	10418.75	5
OB	66	33	33	26.59	26.48	26.7	11837.88	11424.24	12251.52	2
SK	66	31	35	28.08	27.77	28.34	16427.27	12845.16	19600.0	2
SS	67	31	36	25.78	25.84	25.72	10608.96	11025.81	10250.0	8
WO	59	30	29	26.92	26.83	27.0	15791.24	13170.0	18502.86	3

- 평균 연봉의 경우 선수 별로 엔화/달러/원 단위로 구성이 되어있다. 그래서 환율로 원으로 변환하여 계산하였다. 단위는 (만)원이다.

- 시즌 하나를 놓고 봤을 때 2016년 기준 선수 투자 연봉에 가장 많은 돈을 쓴 한화 (19,172.74)는 시즌 7위를, 가장 적은 돈을 쓴 키움 (8843.75)를 시즌 3위를 기록하는 결과를 나타냈다.

- 이후 시즌 들에도 다음과 같은 연관성이 나타났는지는 상관계수를 통해 살펴 볼 수 있겠지만, 높은 팀 연봉에도 2017년 삼성(9위), 2018년 롯데 (7위)와 같이 극단적인 결과가 나타나서 유의미한 결과를 보지 못할 것으로 예측한다.

- 이는, 팀은 가치 있는 선수, 잘하는 선수를 가지고 있을 때 좋은 성적을 가질 수 있다는 생각에 기반하여, 선수의 가치를 연봉으로 생각한 기존의 내 추측이 잘못되었음을 보여준다. 타 기준으로 팀 능력비교 필요.

4. 피어슨 상관계수로 알아보는 상관관계

(1) 정규화 (normalization): 데이터의 범위를 0과 1로 변환하여 데이터 분포 조정

- 전체 평균 연봉 / 투수 평균 연봉 / 타자 평균 연봉 / 전체 평균 연령 / 투수 평균 연령 / 타자 평균 연령

[1.0, 0.49, 0.405, 0.647, 0.499, 0.099, 0.362, 0.425, 0.647, 0.0]
 [1.0, 0.542, 0.274, 0.72, 0.531, 0.0, 0.506, 0.467, 0.699, 0.157]
 [1.0, 0.521, 0.676, 0.615, 0.555, 0.406, 0.276, 0.503, 0.639, 0.0]
 [1.0, 0.626, 0.189, 0.476, 0.609, 0.37, 0.563, 0.476, 0.719, 0.0]
 [1.0, 0.78, 0.184, 0.4, 0.516, 0.083, 0.505, 0.248, 0.614, 0.0]
 [0.911, 0.162, 0.123, 0.532, 0.682, 0.902, 0.547, 1.0, 0.796, 0.0]

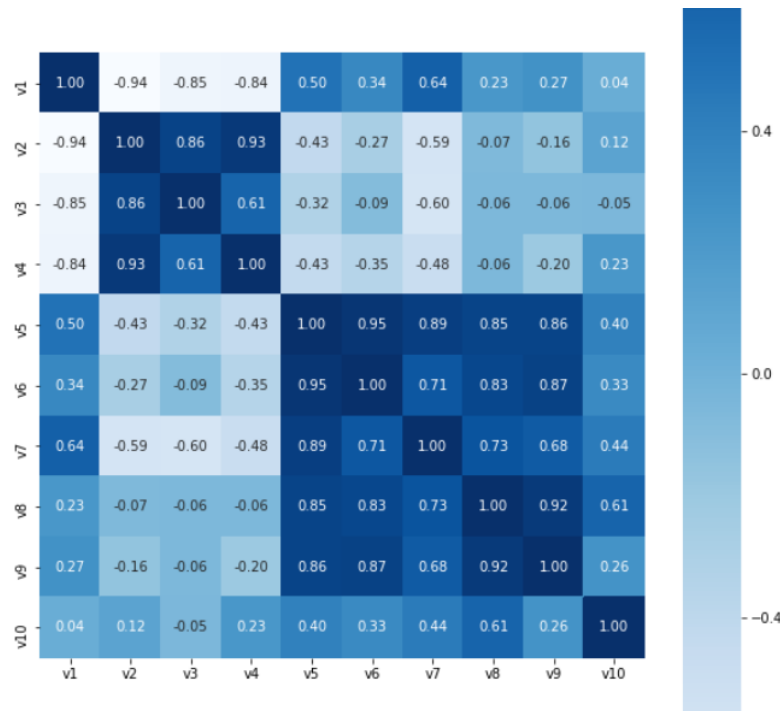
[0.779, 0.568, 0.485, 0.347, 0.455, 0.719, 0.254, 1.0, 0.37, 0.0]
 [1.0, 0.836, 0.0, 0.67, 0.475, 0.66, 0.283, 0.758, 0.695, 0.258]
 [0.527, 0.374, 0.741, 0.211, 0.47, 0.668, 0.349, 1.0, 0.211, 0.0]
 [0.306, 0.76, 0.403, 0.0, 0.473, 0.409, 1.0, 0.425, 0.596, 0.137]
 [0.735, 0.432, 0.05, 0.113, 0.764, 0.0, 0.863, 0.145, 1.0, 0.523]
 [0.166, 1.364, 0.943, 0.185, 0.494, 1.003, 1.5, 1.003, 0.52, 0.0]

[1.0, 0.547, 0.35, 0.154, 0.944, 0.722, 0.0, 0.944, 0.932, 0.436]
 [0.905, 0.38, 0.042, 0.031, 1.0, 0.642, 0.137, 0.617, 0.994, 0.0]
 [0.68, 0.612, 0.709, 0.385, 0.41, 0.594, 0.0, 1.0, 0.493, 0.968]
 [1.0, 0.223, 0.566, 0.0, 0.764, 0.397, 0.223, 0.115, 0.51, 0.366]
 [1.0, 0.343, 0.553, 0.0, 0.595, 0.371, 0.452, 0.269, 0.618, 0.165]
 [0.154, 0.289, 0.444, 0.601, 0.687, 0.511, 0.0, 0.271, 0.216, 1.0]

[0.604, 0.4, 0.435, 0.687, 0.452, 0.826, 0.352, 1.0, 0.0, 0.496]
 [0.741, 0.632, 0.746, 0.736, 0.663, 0.772, 0.332, 1.0, 0.0, 0.513]
 [0.519, 0.263, 0.16, 0.668, 0.305, 0.87, 0.374, 1.0, 0.0, 0.489]
 [0.491, 0.658, 0.407, 0.816, 0.362, 0.281, 0.211, 1.0, 0.0, 0.891]
 [0.32, 0.549, 0.0, 0.97, 1.0, 0.718, 0.202, 0.467, 0.127, 0.527]
 [0.502, 0.573, 0.675, 0.555, 0.0, 0.078, 0.262, 1.0, 0.061, 0.89]

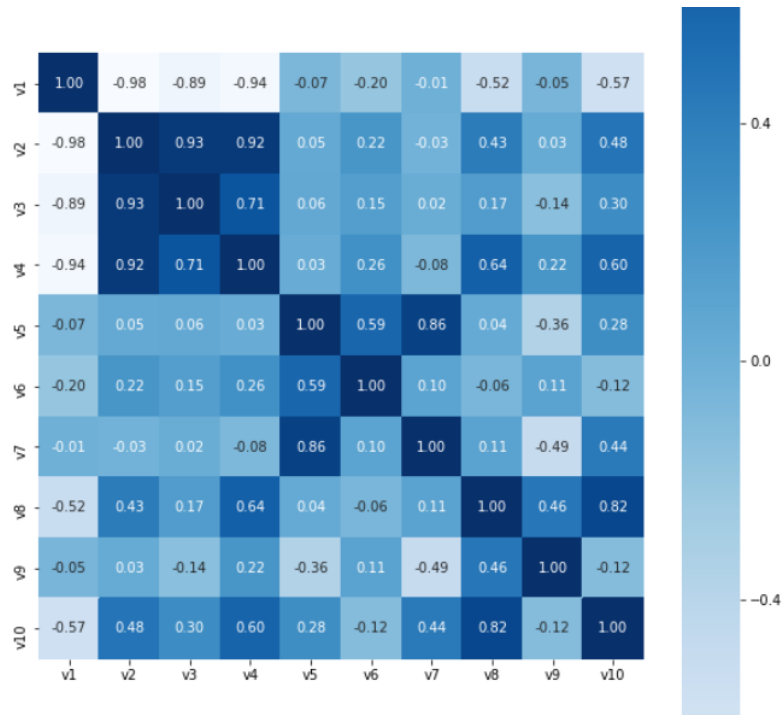
(2) 2016년도 시즌

- 순위 / 전체 승률 [홈 승률, 원정 승률] / 연령 평균 [투수, 타자] / 연봉 평균 [투수, 타자] 상관관계



(3) 2017 시즌

- 순위 / 전체 승률 [홈 승률, 원정 승률] / 연령 평균 [투수, 타자] / 연봉 평균 [투수, 타자] 상관관계



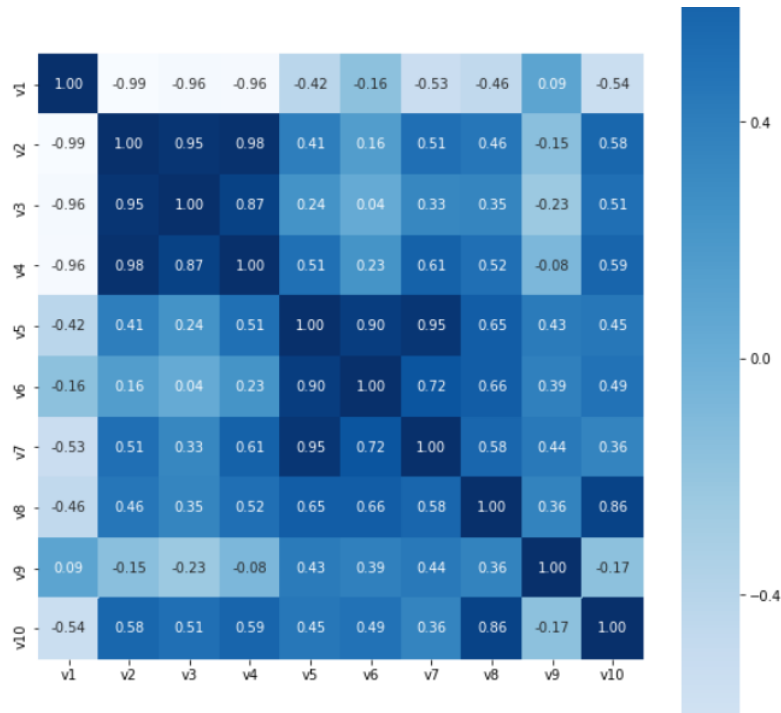
(4) 2018 시즌

- 순위 / 전체 승률 [홈 승률, 원정 승률] / 연령 평균 [투수, 타자] / 연봉 평균 [투수, 타자] 상관관계



(5) 2019 시즌

- 순위 / 전체 승률 [홈 승률, 원정 승률] / 연령 평균 [투수, 타자] / 연봉 평균 [투수, 타자] 상관관계



- 피어슨 상관 계수는 상관관계 분석에서 기본적으로 두 변수 간의 관련성을 구하기 위해 사용된다.
- 양/음(+/-)는 관계의 방향을 나타내며, 절대값 구간마다 관계의 강도를 나타낸다.
- 보통 강한 관계는 0.7 ~ 1.0, 뚜렷한 관계는 0.3 ~ 0.7, 약한 관계는 0.1 ~ 0.3 정도의 수치를 나타낸다.

5. 결과 종합

[1] 승률 - 연령

- 승률과 연령[전체/투수/타자]의 시즌 별 피어슨 상관계수는 다음과 같다.

2016 [-0.43, -0.27, -0.59]		2017 [0.05, 0.22, -0.03]
2018 [-0.24, -0.10, -0.32]		2019 [0.41, 0.16, 0.51]

- 나이가 많을수록 승률에 좋다(+) / 나이가 많을수록 승률에 좋지 않다(-)
- 시즌 전체로 살펴보면, 승률과 일정한 양이나 음의 관계를 갖지 않는다. 어느 해(2017, 2019)는 나이가 많을수록 승률에 좋은 영향을 미친 반면, 어느 해(2016, 2018)는 반대의 결과를 나타낸다.
- 영향력의 규모를 살펴봤을 때, 2017년을 제외하고, 투수의 영향력보다 타자의 영향력이 더 높게 나왔다. 특히 2016년, 2018년, 2019년의 기록을 살펴보면, 절대치 0.59, 0.32, 0.51로 어떠한 뚜렷한 관계를 나타낸다. 팀 승률에 좋은 영향을 미치든, 나쁜 영향을 미치든 타자의 영향력이 더 높다.
- 이에, 만일 팀의 승률을 위한 전력의 수치를 계산하기 위해서는 타자의 영향력을 더 높게 생각해야한다고 생각이 된다.
- 만약 2020년도 예측을 위해 계산을 한다면, 해당 시즌은 어떤 방향으로 투수가 팀 승률에 영향을 미치는지 그 규모와 방향을 살펴보고 팀 전력 계산에 사용할 수 있을지 한편으로 생각해본다.

[2] 승률 - 연봉

- 승률과 연봉[전체/투수/타자]의 시즌 별 피어슨 상관계수는 다음과 같다.

2016 [-0.07, -0.16, 0.12]		2017 [0.43, 0.03, 0.48]
2018 [-0.15, 0.07, -0.45]		2019 [0.46, -0.15, 0.58]

- 2016년, 2018년 그리고 2017년, 2019년 이렇게 관계의 방향성이 다르다. 일관성이 없어 보인다.
- 여기서도 비싼 타자, 가치 있는 타자의 영향력이 높아 보인다. 연봉이 실력을 대변하지는 않지만, 타자에 투자를 많이 할수록 전체 승률에 큰 영향 2017[0.48], 2018[-0.45], 2019[0.58]을 미친다. 비록 2018년에는 타자가 값을 못했지만, 어쨌든 영향력!
- 결론적인 생각...? 당해 승률 예측이나 어떤 예측에도 연봉수치는 고려하지 말자

[3] 승률 - [홈 승률 / 어웨이 승률]

- 전체 승률과 홈 승률 / 어웨이 승률의 시즌 별 피어슨 상관계수는 다음과 같다.

2016 [0.86, 0.93]		2017 [0.93, 0.93]
2018 [0.88, 0.85]		2019 [0.95, 0.98]

- 홈 경기와 원정 경기 중 어떤 경기가 전체 승률에 더 큰 영향력을 미치는지 비교했을 때, 비교적 원정 경기에서 잘 할수록 수치상으로는 더 큰 영향을 미치지만, 그 유효성이 그렇게 커 보이지 않는다.
- 남은 경기의 승률을 계산하는 것이므로, 내 생각에 팀 별 승률 계산을 기준으로 기대 승수를 계산하는 접근이 더 맞는 접근이지 않을까 생각한다. -끝-