

▼ 20191121

▼ team 승수. 팀명. salary 연봉. 팀명 2016년 groupby(팀명) 30개 팀 평균 데이터를 구하시

```
1 import pandas as pd
2 import numpy as np
3 from google.colab import drive
4 drive.mount('/content/drive')
```

↳ Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive")

```
1 team=pd.read_csv('/content/drive/My Drive/통계적방법19_박초연/Teams.csv')
2 salary=pd.read_csv('/content/drive/My Drive/통계적방법19_박초연/Salaries.csv')
```

```
1 team = team[team.yearID == 2016]
2 team.groupby(by='teamID')
3 team = team[['teamID', 'W']]
4 team = team.replace('.', np.nan)
```

```
1 salary = salary[salary.yearID == 2016]
2 salary = salary[['salary', 'teamID']]
3 salary = salary.groupby(by='teamID').mean()
4 salary = salary.replace('.', np.nan)
```

```
1 t_s = pd.merge(team, salary, on='teamID', how='inner')
2 t_s = t_s.replace('.', np.nan)
```

```
1 t_s
```

↳

	teamID	W	salary
0	ARI	69	3.363041e+06
1	ATL	68	2.362010e+06
2	BAL	89	5.581498e+06
3	BOS	93	6.501578e+06
4	CHA	78	4.519947e+06
5	CHN	103	5.312678e+06
6	CIN	68	3.066899e+06
7	CLE	94	2.752293e+06
8	COL	75	3.413487e+06
9	DET	86	6.286338e+06
10	HOU	84	3.389061e+06
11	KCA	81	4.534039e+06
12	LAA	74	5.278897e+06
13	LAN	91	6.322525e+06
14	MIA	79	2.761222e+06
15	MIL	73	2.292508e+06
16	MIN	59	4.274300e+06
17	NYA	84	7.689579e+06
18	NYN	87	4.958857e+06
19	OAK	69	2.893541e+06
20	PHI	71	2.033793e+06
21	PIT	78	3.706387e+06
22	SDN	68	3.756475e+06
23	SEA	86	4.845834e+06
24	SFN	87	6.890151e+06
25	SLN	86	4.614629e+06
26	TBA	68	2.039190e+06
27	TEX	95	6.070301e+06
28	TOR	89	4.782817e+06
29	WAS	95	5.448179e+06

▼ 그리고 1승을 하기 위하여 연봉을 얼마나 더 지급해야 하나? optional.

```

1 import statsmodels.api as sm
2 y=t_s['salary']
3 X=sm.add_constant(t_s['W'])
4 model=sm.OLS(y, X).fit()
5 model.summary()

```

↳ /usr/local/lib/python3.6/dist-packages/numpy/core/fromnumeric.py:2495: FutureWarning: Method .return_ptp(axis=axis, out=out, **kwargs)

OLS Regression Results

Dep. Variable:	salary	R-squared:	0.374
Model:	OLS	Adj. R-squared:	0.352
Method:	Least Squares	F-statistic:	16.73
Date:	Thu, 21 Nov 2019	Prob (F-statistic):	0.000329
Time:	12:32:55	Log-Likelihood:	-462.61
No. Observations:	30	AIC:	929.2
Df Residuals:	28	BIC:	932.0
Df Model:	1		

Covariance Type: nonrobust

	coef	std err	t	P> t	[0.025	0.975]
const	-2.789e+06	1.77e+06	-1.576	0.126	-6.41e+06	8.36e+05
W	8.876e+04	2.17e+04	4.091	0.000	4.43e+04	1.33e+05

Omnibus: 1.039 **Durbin-Watson:** 1.771
Prob(Omnibus): 0.595 **Jarque-Bera (JB):** 0.370
Skew: 0.251 **Prob(JB):** 0.831
Kurtosis: 3.212 **Cond. No.** 635.

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

1승을 하기 위해서는 887.6천불 더 지급해야하며,

이 결과는 Prob (F-statistic)가 0.001보다 작으므로 유의하다.

R-squared가 0.374인 것을 보아 연봉은 승수에 37.4% 의 영향을 미치고,

나머지(62.6%)는 승수 외의 다른 요인으로 인해 변동된다.

```
1 (8.876e+04)/100
```

↳ 887.6