

```
In [1]: import pandas as pd  
import numpy as np
```

아래와 같은 결과를 만들어주세요

In [7]: NC\_final

Out[7]:

	안타			홈런		
연도	2013	2014	2015	2013	2014	2015
선수명						
강구성	0	-	1	0	-	0
강민국	-	0	0	-	0	0
강진성	1	-	-	0	-	-
권희동	-	63	-	-	7	-
김동건	2	-	-	1	-	-
김성욱	1	4	-	0	1	-
김종찬	1	-	-	0	-	-
김종호	129	-	125	0	-	4
김준완	-	2	10	-	0	0
김태균	-	-	107	-	-	6
김태우	-	1	-	-	0	-
김태진	-	-	0	-	-	0
나성범	98	157	184	14	30	28
노진혁	-	3	-	-	1	-
마낙길	9	0	-	0	0	-
모창민	109	110	62	12	16	6
박명환*	-	0	-	-	0	-
박민우	11	124	158	0	1	3
박정준	41	-	8	4	-	0
손시현	-	90	108	-	5	13
용덕한	-	-	12	-	-	0
이상호	31	-	-	0	-	-
이승호	0	-	-	0	-	-
이종욱	-	126	118	-	6	5
이창섭	0	-	0	0	-	0
이현곤	38	-	-	0	-	-
이호준	123	115	132	20	23	24
조영훈	107	29	35	6	6	8
조평호	21	0	8	2	0	1
지석훈	68	58	111	3	6	11
차화준	24	-	-	1	-	-
최재원	-	0	21	-	0	2
테임즈	-	152	180	-	37	47
허준	-	5	-	-	0	-

```
In [2]: # 데이터 불러오기
data = pd.read_excel('data/NC Dinos.xlsx', sheet_name=None)
NC13, NC14, NC15 = data.values()

# 필요없는 컬럼 제거
NC13.drop('Unnamed: 0', axis=1, inplace=True)
NC14.drop('Unnamed: 0', axis=1, inplace=True)
NC15.drop('Unnamed: 0', axis=1, inplace=True)

# 연도 추가
NC13['연도'] = 2013
NC14['연도'] = 2014
NC15['연도'] = 2015

NC13.set_index('선수명', inplace=True)
NC14.set_index('선수명', inplace=True)
NC15.set_index('선수명', inplace=True)
```

```
In [3]: NC = pd.concat([NC13.안타, NC14.안타, NC15.안타,
                        NC13.홈런, NC14.홈런, NC15.홈런], axis=1, sort=True).fillna('-')
NC.head()
```

```
Out[3]:
```

	안타	안타	안타	홈런	홈런	홈런
강구성	0	-	1	0	-	0
강민국	-	0	0	-	0	0
강진성	1	-	-	0	-	-
권희동	-	63	-	-	7	-
김동건	2	-	-	1	-	-

```
In [3]: import numpy as np

# np.repeat(['안타', '홈런'], 3)
[2013, 2014, 2015] * 3
```

```
Out[3]: [2013, 2014, 2015, 2013, 2014, 2015, 2013, 2014, 2015]
```

```
In [5]: NC_final = pd.DataFrame(NC.values,
                                columns=[np.repeat(['안타', '홈런'], 3), [2013, 2014, 2015]*2],
                                index=NC.index)
```

```
-----
NameError                                Traceback (most recent call last)
<ipython-input-5-4e45d570ea28> in <module>
----> 1 NC_final = pd.DataFrame(NC.values,
      2                                columns=[np.repeat(['안타', '홈런'], 3), [2013, 20
      3                                index=NC.index)

NameError: name 'pd' is not defined
```

```
In [4]: NC_final.index.set_names('선수명', inplace=True)
NC_final.columns.set_names(['', '연도'], inplace=True)
```

```
-----
NameError                                Traceback (most recent call last)
<ipython-input-4-3a5e7ac8dee0> in <module>
----> 1 NC_final.index.set_names('선수명', inplace=True)
      2 NC_final.columns.set_names('연도', level=1, inplace=True)

NameError: name 'NC_final' is not defined
```

```
In [6]: NC_final
```

```
Out[6]:
```

	안타			홈런		
연도	2013	2014	2015	2013	2014	2015
선수명						
강구성	0	-	1	0	-	0
강민국	-	0	0	-	0	0
강진성	1	-	-	0	-	-
권희동	-	63	-	-	7	-
김동건	2	-	-	1	-	-
김성욱	1	4	-	0	1	-
김종찬	1	-	-	0	-	-
김종호	129	-	125	0	-	4
김준완	-	2	10	-	0	0

## 같은 결과를 보기 위해서 random seed를 321로 설정하기

```
In [6]: np.random.seed(321)
```

## 아래 셀은 실행시켜주세요

자세한 함수는 수업전에 설명해드리겠습니다

궁금하신분은 구글링을 통해서 미리 공부하고 와도 좋아요~

```
In [7]: # Set index multi column index
num = np.repeat(['19기', '20기', '21기'], (2, 7, 16))
dic={1:'문과', 2:'이과'}
major = [dic[x] for x in np.random.randint(1, 3, 25)]
name = ['양용준', '김현기', '김우찬', '김지윤', '박세희', '박윤정',
        '이가은', '정범영', '정예지', '강주연', '김현아', '김혜진',
        '문성희', '이세영', '이영주', '이현서', '이다빈', '천승아',
        '나윤진', '박수완', '서민혁', '이상민', '전윤아', '정현정', '조준혁']
```

## 아래 그림과 같은 결과가 나오게 만들어주세요

점수는 1~100점사이의 임의의 값을 넣으면 됩니다

row index는 만들어놨으니 column index만 설정해주세요

최종 데이터프레임 변수명은 저와 같이 python\_study로 설정해주세요

결과물은 거짓이고 컴퓨터가 임의로 정한거예요 저를 원망하지 마세요ㅠㅠ

```
In [59]: # 여기 셀 말고 아래 셀에서 작성해주세요
python_study
```

```
Out[59]:
```

			1학기			2학기		
			언어	수리	외국어	언어	수리	외국어
19기	이과	양용준	27	32	42	73	18	41
		김현기	27	89	73	84	86	2
20기	이과	김우찬	28	27	62	14	44	55
		김지윤	63	32	11	56	35	19
	문과	박세희	4	42	51	13	31	98
		박윤정	15	30	11	2	34	1
		이가은	5	76	53	4	17	64
		이과 정범영	84	40	31	92	5	44
	문과	정예지	14	38	72	1	12	11
		이과 강주연	89	8	34	65	99	86

```
In [10]: import pandas as pd
python_study = pd.DataFrame(np.random.randint(1, 100, (25, 6)),
                             columns=[np.repeat(['1학기', '2학기'], 3), ['언어', '수리', '외국어']],
                             index=[num, major, name])
```

본인의 1학기, 2학기 전체 성적을 출력해주세요

```
In [61]: python_study.loc[('20기', '문과', '김지윤')]
```

```
Out[61]: 1학기 언어      63
          수리      32
          외국어    11
2학기 언어      56
          수리      35
          외국어    19
Name: (20기, 문과, 김지윤), dtype: int32
```

20기 1학기 성적을 출력주세요

```
In [65]: python_study.loc[ '20기' , '1학기' ]
```

```
C:\Users\yyj94\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: PerformanceWarning: indexing past lexsort depth may impact performance.
    """Entry point for launching an IPython kernel.
```

Out[65]:

		언어	수리	외국어
이과	김우찬	28	27	62
	김지윤	63	32	11
	박세희	4	42	51
문과	박윤정	15	30	11
	이가은	5	76	53
이과	정범영	84	40	31
문과	정예지	14	38	72

**이과의 1학기, 2학기 성적을 출력해주세요**

```
In [11]: python_study
```

Out[11]:

			1학기				2학기		
			언어	수리	외국어	언어	수리	외국어	
19기	문과	양용준	32	11	56	35	19	4	
		김현기	42	51	13	31	98	15	
	문과	김우찬	30	11	2	34	1	5	
	이과	김지윤	76	53	4	17	64	84	
		박세희	40	31	92	5	44	14	
20기	문과	박윤정	38	72	1	12	11	89	
	이과	이가은	8	34	65	99	86	39	
	문과	정범영	50	78	3	89	77	68	
		정예지	1	20	74	61	57	70	
	문과	강주연	96	55	94	6	35	94	
		김현아	93	91	20	90	10	69	
	이과	김혜진	32	11	27	47	84	2	
	문과	문성희	30	59	56	92	63	91	
		이세영	19	58	80	79	93	4	
		이영주	28	95	64	95	99	93	
		이과	이현서	35	86	65	60	75	21
	21기		이다빈	53	34	65	10	97	4
			천승아	81	96	80	73	72	10
		문과	나윤진	81	24	43	81	77	77
			박수완	96	12	59	76	82	62
		서민혁	22	89	24	83	12	70	
		이과	이상민	77	64	2	53	14	44
			전윤아	42	8	85	59	37	70
문과			정현정	31	28	56	79	23	92
	조준혁	56	30	48	63	18	15		

```
In [68]: python_study.xls('이과', axis=0, level=1)
```

```
Out[68]:
```

		1학기			2학기		
		언어	수리	외국어	언어	수리	외국어
19기	양용준	27	32	42	73	18	41
	김현기	27	89	73	84	86	2
20기	김우찬	28	27	62	14	44	55
	정범영	84	40	31	92	5	44
	강주연	89	8	34	65	99	86
	김혜진	68	1	20	74	61	57
	이영주	69	32	11	27	47	84
21기	이현서	2	30	59	56	92	63
	천승아	4	28	95	64	95	99
	나윤진	93	35	86	65	60	75
	이상민	10	81	24	43	81	77
	전윤아	77	96	12	59	76	82
	조준혁	70	77	64	2	53	14

## 1학기 때 언어, 수리 평균은 문과/이과 중 누가 더 높을까?

**HINT:** 처음부터 이과이면서 1학기 성적을 추출하려 하지말고 이과를 출력한 다음에 과목 성적 추출

```
In [16]: python_study.xls('문과', axis=0, level=1)[('1학기', '언어')].mean()
```

```
Out[16]: 52.0
```

```
In [84]: def check(subject):
    society = python_study.xls('문과', axis=0, level=1)[('1학기', subject)].mean()
    science = python_study.xls('이과', axis=0, level=1)[('1학기', subject)].mean()

    if society > science:
        print('%s 과목은 문과의 성적이 이과보다 높습니다' % subject)
    elif society < science:
        print('%s 과목은 이과의 성적이 문과보다 높습니다' % subject)
    else:
        print('%s 과목은 문과의 성적과 이과의 성적은 같습니다' % subject)
```

```
In [86]: check('언어')
```

언어 과목은 이과의 성적이 문과보다 높습니다

```
In [83]: check('수리')
```

수리 과목은 이과의 성적이 문과보다 높습니다



```
In [19]: python_study.xls('1학기', axis=1, level=0)['언어']
python_study.xls('2학기', axis=1, level=0)['언어']

python_study.xls('언어', axis=1, level=1)
```

Out[19]:

		언어	수리	외국어		
19기	문과	양용준	35	19	4	
		김현기	31	98	15	
	문과	김우찬	34	1	5	
		이과	김지윤	17	64	84
	박세희		5	44	14	
20기	문과	박윤정	12	11	89	
	이과	이가은	99	86	39	
	문과	정범영	89	77	68	
		정예지	61	57	70	
	문과	강주연	6	35	94	
		김현아	90	10	69	
	이과	김혜진	47	84	2	
	문과	문성희	92	63	91	
		이세영	79	93	4	
	21기	문과	이영주	95	99	93
			이현서	60	75	21
		이과	이다빈	10	97	4
천승아			73	72	10	
문과		나윤진	81	77	77	
		박수완	76	82	62	
이과		서민혁	83	12	70	
		이상민	53	14	44	
문과		전윤아	59	37	70	
		정현정	79	23	92	
		조준혁	63	18	15	

```
In [96]: a = python_study.xls('언어', axis=1, level=1)
b = python_study.xls('수리', axis=1, level=1)
c = python_study.xls('외국어', axis=1, level=1)
```

```
In [108]: d = pd.concat([a.mean(axis=1), b.mean(axis=1), c.mean(axis=1)], axis=1)
d.columns = [['평균']*3, ['언어', '수리', '외국어']]
pd.concat([python_study, d], axis=1)
# python_study(['평균', '언어'])
```

Out[108]:

			1학기				2학기				평균
			언어	수리	외국어	언어	수리	외국어	언어	수리	외국어
19기	이과	양용준	27	32	42	73	18	41	50.0	25.0	41.5
		김현기	27	89	73	84	86	2	55.5	87.5	37.5
	이과	김우찬	28	27	62	14	44	55	21.0	35.5	58.5
		김지윤	63	32	11	56	35	19	59.5	33.5	15.0
20기	문과	박세희	4	42	51	13	31	98	8.5	36.5	74.5
		박윤정	15	30	11	2	34	1	8.5	32.0	6.0
		이가은	5	76	53	4	17	64	4.5	46.5	58.5
	이과	정범영	84	40	31	92	5	44	88.0	22.5	37.5
	문과	정예지	14	38	72	1	12	11	7.5	25.0	41.5
	이과	강주연	89	8	34	65	99	86	77.0	53.5	60.0
	문과	김현아	39	50	78	3	89	77	21.0	69.5	77.5
	이과	김혜진	68	1	20	74	61	57	71.0	31.0	38.5
	문과	문성희	70	96	55	94	6	35	82.0	51.0	45.0
		이세영	94	93	91	20	90	10	57.0	91.5	50.5
	이과	이영주	69	32	11	27	47	84	48.0	39.5	47.5
		이현서	2	30	59	56	92	63	29.0	61.0	61.0
21기	문과	이다빈	91	19	58	80	79	93	85.5	49.0	75.5
	이과	천승아	4	28	95	64	95	99	34.0	61.5	97.0
		나윤진	93	35	86	65	60	75	79.0	47.5	80.5
	문과	박수완	21	53	34	65	10	97	43.0	31.5	65.5
		서민혁	4	81	96	80	73	72	42.0	77.0	84.0
	이과	이상민	10	81	24	43	81	77	26.5	81.0	50.5
		전윤아	77	96	12	59	76	82	68.0	86.0	47.0
	문과	정현정	62	22	89	24	83	12	43.0	52.5	50.5
	이과	조준혁	70	77	64	2	53	14	36.0	65.0	39.0