

## import pandas, numpy module

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

아래와 같은 결과를 만들어주세요

In [2]: # 아래 셀에서 작성해주세요  
NC\_final

Out[2]:

	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
김태군	-	-	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 [33]: # 데이터 불러오기
NC = pd.read_excel("data/NC Dinos.xlsx",
                  sheet_name = None)
NC
```

```
Out[33]: OrderedDict([( '2013',
                        Unnamed: 0  선수명  팀명  경기  타석  타수  안타  홈런  득점  타
점  볼넷  삼진  도루  BABIP  \
0  모창민  NC  108  436  395  109  12  57  51  37
68  16  0.307
1  이호준  NC  126  508  442  123  20  46  87  60
109  2  0.324
2  김종호  NC  128  546  465  129  0  72  22  57
100  50  0.352
3  나성범  NC  104  458  404  98  14  55  64  33
95  12  0.279
4  조영훈  NC  120  426  380  107  6  38  39  39
56  4  0.316
5  이현곤  NC  91  161  139  38  0  10  9  16
14  2  0.304
6  이상호  NC  102  138  125  31  0  26  13  9
21  24  0.298
7  강진성  NC  3  3  2  1  0  1  0  1
0  0  0.500
0  조영훈  NC  26  26  26  70  21  2  12  7  6
```

```
In [34]: # 2013, 2014, 2015 시즌 정보를 읽어서 각각 NC13, NC14, NC15에 저장
NC13, NC14, NC15 = NC.values()
```

```
In [35]: # 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)
```

```
In [39]: NC13.set_index(['선수명'], inplace = True)
NC14.set_index(['선수명'], inplace = True)
NC15.set_index(['선수명'], inplace = True)
```

```
In [43]: # 안타(hit)
NC_hit = pd.concat([NC13['안타'], NC14['안타'], NC15['안타']], axis = 1, sort = Tr

# 행 설정
NC_hit.index.set_names(['선수명'], inplace = True)

# 열 설정
NC_hit.columns = [['안타', '안타', '안타'], [2013, 2014, 2015]]
NC_hit.columns.set_names([' ', '연도'], inplace = True)
```

In [44]: NC\_hit

Out[44]:

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

```
In [45]: # 홈런(homerun)
NC_hom = pd.concat([NC13['홈런'], NC14['홈런'], NC15['홈런']], axis = 1, sort = True)

# 행 설정
NC_hom.index.set_names(['선수명'], inplace = True)

# 열 설정
NC_hom.columns = [['홈런', '홈런', '홈런'], [2013, 2014, 2015]]
NC_hom.columns.set_names(['', '연도'], inplace = True)
```

In [46]: NC\_hom

...

```
In [48]: # 안타와 홈런 병합
NC_final = pd.concat([NC_hit, NC_hom], axis = 1).fillna('-')
NC_final
```

Out[48]:

	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

```
In [ ]: # 오답

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

# 필요한 columns만 추출
col13 = NC13[['선수명', '안타', '홈런', '연도']]
col14 = NC14[['선수명', '안타', '홈런', '연도']]
col15 = NC15[['선수명', '안타', '홈런', '연도']]

NC_middle = pd.concat([col13, col14, col15], axis = 1).fillna('-')

# pd.concat 수행
# 행으로 합치기

#NC.columns = [['안타', '홈런'], [2013, 2014, 2015]]
NC.columns = [[('안타', [2013, 2014, 2015]), ('홈런', [2013, 2014, 2015])]]

NC_middle
```

```
In [31]: # 오답
NC_middle
```

```
Out[31]:
```

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

## 같은 결과를 보기 위해서 random seed를 321로 설정하기

```
In [49]: np.random.seed(321)
```

아래 셀은 실행시켜주세요

자세한 함수는 수업전에 설명해드리겠습니다

궁금하신분은 구글링을 통해서 미리 공부하고 와도 좋아요~

```
In [50]: # 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 [3]: # 여기 셀 말고 아래 셀에서 작성해주세요
python_study
```

Out[3]:

		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	
		김지윤	76	53	4	17	64	84	
	이과	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	
21기	이과	이현서	35	86	65	60	75	21	
		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	
		이상민	77	64	2	53	14	44	
	이과	전윤아	42	8	85	59	37	70	
		문과	정현정	31	28	56	79	23	92
			조준혁	56	30	48	63	18	15

```
In [ ]: python_study = pd.DataFrame(,
                                     columns=,
                                     index=[num,major,name])
```

```
In [54]: python_study = pd.DataFrame(np.random.randint(1, 101, (25, 6)),
                                     columns = [['1학기', '1학기', '1학기', '2학기', '2학기'],
                                     index = [num, major, name])

python_study
```

```
Out[54]:
```

		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

## 본인의 1학기, 2학기 전체 성적을 출력해주세요

```
In [56]: python_study.loc['20기', '이과', '박세희']
```

```
Out[56]: 1학기 언어      40
          수리      31
          외국어    92
2학기 언어      5
          수리     44
          외국어    14
Name: (20기, 이과, 박세희), dtype: int64
```

## 20기 1학기 성적을 출력주세요

```
In [57]: python_study.xs('1학기', axis = 1, level = 0)
```

```
Out[57]:
```

		언어	수리	외국어	
19기	문과	양용준	32	11	56
		김현기	42	51	13
	문과	김우찬	30	11	2
	이과	김지윤	76	53	4
		박세희	40	31	92
20기	문과	박윤정	38	72	1
	이과	이가은	8	34	65
	문과	정범영	50	78	3
		정예지	1	20	74
21기	문과	강주연	96	55	94
		김현아	93	91	20



## 이과의 1학기, 2학기 성적을 출력해주세요

```
In [58]: python_study.xls('이과', axis = 0, level = 1)
```

```
Out[58]:
```

		1학기			2학기		
		언어	수리	외국어	언어	수리	외국어
	김지윤	76	53	4	17	64	84
20기	박세희	40	31	92	5	44	14
	이가은	8	34	65	99	86	39
	김혜진	32	11	27	47	84	2
	이세영	19	58	80	79	93	4
	이영주	28	95	64	95	99	93
	이현서	35	86	65	60	75	21
21기	이다빈	53	34	65	10	97	4
	천승아	81	96	80	73	72	10
	서만혁	22	89	24	83	12	70

## 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 [64]: # 이과(natural sciences)

nat = python_study.xls('이과', axis = 0, level = 1)
nat_first = nat.xls('1학기', axis = 1, level = 0)
nat_lang = nat_first['언어'].mean()
nat_math = nat_first['수리'].mean()
```

```
In [65]: # 문과(liberal arts)

lib = python_study.xls('문과', axis = 0, level = 1)
lib_first = lib.xls('1학기', axis = 1, level = 0)
lib_lang = lib_first['언어'].mean()
lib_math = lib_first['수리'].mean()
```

```
In [66]: print("The average language of natural sciences is ", nat_lang, ".")
print("The average math of natural sciences is ", nat_math, ".")
print("The average language of liberal arts is ", lib_lang, ".")
print("The average math of liberal arts is ", lib_math, ".")
```

```
The average language of natural sciences is 42.75 .
The average math of natural sciences is 54.916666666666664 .
The average language of liberal arts is 52.0 .
The average math of liberal arts is 41.69230769230769 .
```

## 아래와 같이 각 사람마다의 1학기 2학기 각 과목 평균을 구해주세요

```
In [4]: # 밑에 셀에서 작성해주세요
mean
```

Out [ 4 ] :

			평균		
			언어	수리	외국어
19기	문과	양용준	33.5	15.0	30.0
		김현기	36.5	74.5	14.0
	문과	김우찬	32.0	6.0	3.5
		김지윤	46.5	58.5	44.0
	이과	박세희	22.5	37.5	53.0
20기	문과	박윤정	25.0	41.5	45.0
	이과	이가은	53.5	60.0	52.0
	문과	정범영	69.5	77.5	35.5
		정예지	31.0	38.5	72.0
	문과	강주연	51.0	45.0	94.0
		김현아	91.5	50.5	44.5
	이과	김혜진	39.5	47.5	14.5
	문과	문성희	61.0	61.0	73.5
		이세영	49.0	75.5	42.0
		이영주	61.5	97.0	78.5
	이과	이현서	47.5	80.5	43.0
21기		이다빈	31.5	65.5	34.5
		천승아	77.0	84.0	45.0
		나윤진	81.0	50.5	60.0
	문과	박수완	86.0	47.0	60.5
		서민혁	52.5	50.5	47.0
	이과	이상민	65.0	39.0	23.0
		전윤아	50.5	22.5	77.5
	문과	정현정	55.0	25.5	74.0
		조준혁	59.5	24.0	31.5

```
In [68]: # first / second semester
python_1st = python_study.xls('1학기', axis = 1, level = 0)
python_2nd = python_study.xls('2학기', axis = 1, level = 0)

python_mean = (python_1st + python_2nd) / 2
python_mean
```

Out[68]:

			언어	수리	외국어	
19기	문과	양용준	33.5	15.0	30.0	
		김현기	36.5	74.5	14.0	
	문과	김우찬	32.0	6.0	3.5	
		이과	김지윤	46.5	58.5	44.0
			박세희	22.5	37.5	53.0
20기	문과	박윤정	25.0	41.5	45.0	
	이과	이가은	53.5	60.0	52.0	
		문과	정범영	69.5	77.5	35.5
			정예지	31.0	38.5	72.0
	문과	강주연	51.0	45.0	94.0	
			김현아	91.5	50.5	44.5