

## ▼ HW05 (제출기한-5월3일자정)


Datascience 모듈의 Table 연습을 위한 문제입니다.

- 반드시 수업시간 배운 내용만을 사용하여 코드를 작성합니다.
- 한 셀에 여러줄의 코드를 사용해도 무방합니다.

제출시 다음 사항에 유의하기 바랍니다.

- 텍스트 셀에 설명되어 있는 각 문항을 잘 읽은 뒤, '답안을 작성하시오'라고 적힌 코드 셀에 적절한 코드를 작성합니다.
- '지우지 마시오'라고 적힌 코드 셀은 절대로 지우면 안 됩니다.
- 작성 후 파일명에 학번 을 자신의 학번으로 고친 후 ULMS 해당과제 제출란에 업로드합니다.

```
1 from datascience import *
2 import matplotlib
3 %matplotlib inline
4 import matplotlib.pyplot as plots
5 plots.style.use('fivethirtyeight')
6 import numpy as np
```

 /usr/local/lib/python3.7/dist-packages/datascience/tables.py:17: MatplotlibDeprecationWarning  
matplotlib.use('agg', warn=False)  
/usr/local/lib/python3.7/dist-packages/datascience/util.py:10: MatplotlibDeprecationWarning:  
matplotlib.use('agg', warn=False)

## ▼ 수업시간에 사용하였던 다음 NBA 연봉 데이터를 이용합니다.

```
1 nba = Table.read_table('https://raw.githubusercontent.com/data-8/materials-sp18/master/lec/nba_s
2 nba
```

PLAYER	POSITION	TEAM	SALARY
Aaron Brooks	PG	Chicago Bulls	2.25

## ▼ 문제1

POSITION이 'PG'이면서 연봉이 가장 높은 TOP-5 선수 정보를 *PG\_TOP5* 이름의 Table 변수에 저장 하시오.

```
1 ### 답안을 작성하시오.
2 PG_TOP5 = nba.where('POSITION', 'PG').sort('SALARY', descending=True).take(np.arange(5))
3
4
```

Alec Burks	SG	Utah Jazz	9.46348
------------	----	-----------	---------

```
1 ### 지우지 마시오.
2 PG_TOP5
```

## ▼ 문제2

SALARY의 최대값과 최소값을 구하여 각각 *max\_salary*, *min\_salary* 변수에 저장하시오.

```
1 ### 답안을 작성하시오.
2 max_salary = nba.select('SALARY').max()
3 min_salary = nba.select('SALARY').min()
```

```
/usr/local/lib/python3.7/dist-packages/datascience/tables.py:192: FutureWarning: Implicit col
warnings.warn("Implicit column method lookup is deprecated.", FutureWarning)
```



```
1 ### 지우지 마시오.
2 max_salary, min_salary
```

```
(SALARY
 25, SALARY
 0.030888)
```

## ▼ 문제3

각 TEAM별 선수 수를 감소순으로 정렬하여 *team\_ct* 이름의 Table 변수에 저장하시오.

```
1 ### 답안을 작성하시오.  
2  
3 team_ct = nba.group('TEAM').sort('count',descending=True)
```

```
1 ### 지우지 마시오.  
2 team_ct
```

#### ▼ 문제4

연봉이 가장 높은 200명 중 POSITION별 평균 연봉을 Table변수 *top200* 에 저장하시오.

```
1 ### 답안을 작성하시오.  
2  
3 top200 = nba.select('POSITION','SALARY').sort('SALARY',descending=True).take(np.arange(200)).gro  
4  
5
```

```
1 ### 지우지 마시오.  
2 top200
```

## ▼ 문제5

위와 같은 식으로 연봉이 가장 높은 100명 중 POSITION별 평균 연봉을 Table변수 top100에 저장 하시오.

```
1 ### 답안을 작성하시오.
2
3 top100 = nba.select('POSITION', 'SALARY').sort('SALARY', descending=True).take(np.arange(100)).gro
4
```

```
1 ### 지우지 마시오.
2 top100
```

## ▼ 문제6

위에서 구한 두 Table을 합쳐 Table 변수 top을 생성하시오. 이 때, column의 이름을 top100, top200 으로 고치되 출력 순서도 top100, top200 이 되도록 하시오.

```
1 ### 답안을 작성하시오.
2
3 top = top100.join('POSITION', top200).relabelled("SALARY mean", 'top100').relabelled("SALARY mean_2"
4
```

```
1 ### 지우지 마시오.
2 top
```

## ▼ 문제7

top100과 top200의 POSITION별 상대적 SALARY비중을 비교하기 위하여 top100과 top200 column 각각에 대하여 백분율(소수점 2번째자리에서 round-off)한 결과를 top에 추가하시오.

```
1 ### 답안을 작성하시오.
2
3 top=top.with_columns("top100 Percent",np.round(top.column('top100')/np.sum(top.column('top100'))
4
5
```

```
1 ### 지우지 마시오.
2 top
```

## ▼ 문제8

top100 Percent 와 top200 Percent 에 대해서 수평 Bar 그래프를 그리시오.

```
1 ### 답안을 작성하시오.
2
3 top.select('POSITION',"top100 Percent","top200 Percent").barh('POSITION')
4
```

## ▼ 문제9

Table 변수 `nba` 의 `SALARY` column에 대해서 histogram을 그리시오. (단, bins의 개수를 50으로 설정

```
1 ### 답안을 작성하시오.  
2  
3 nba.hist('SALARY',bins=50)  
4
```

## ▼ 문제10

Position과 Salary를 입력받으면 `nba` Table에서 해당 Position의 Salary 중 몇 등에 해당하는지를 계산하는 함수 `position_ranking` 을 작성하시오. (함수 정의는 그대로 사용할 것)

```
1 ### 답안을 작성하시오.  
2 def position_ranking(nba, p, s):  
3  
4  
5  
6
```

```
1 ### 지우지 마시오.  
2 position_ranking(nba, 'C', 15.0)
```

