pandas 불러오기

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

1. 아래의 baseball_dic에서

밑에 있는 타율, 안타, 홈런, 타점 데이터를

dictionary 형식으로 추가해서 데이터프레임을 최종 완성해주세요

column name은 타율(avg), 안타(hit), 홈런(homerun), 타점(rbi)로 해주세요(한글 x)

In [2]:
```
baseball_dic = {'order':[x for x in range(1, 20)],
                'name':['김현수', '양의지', '이정후', '박병호', '안치홍', '전준우', '김주찬
                        '유한준', '김재환', '최주환', '이대호', '구자욱', '채은성', '러프'
                        '박건우', '허경민'],
                'team':['LG', '두산', '넥센', '넥센', 'KIA', '롯데', 'KIA', 'KIA',
                        'KT', '두산', '두산', '롯데', '삼성', 'LG', '삼성', 'SK', '롯
                        '두산', '두산']}
```

In [3]:
```
df = pd.DataFrame(baseball_dic)
df
```

Out[3]:

	order	name	team
0	1	김현수	LG
1	2	양의지	두산
2	3	이정후	넥센
3	4	박병호	넥센
4	5	안치홍	KIA
5	6	전준우	롯데
6	7	김주찬	KIA
7	8	최형우	KIA
8	9	유한준	KT
9	10	김재환	두산
10	11	최주환	두산
11	12	이대호	롯데
12	13	구자욱	삼성
13	14	채은성	LG
14	15	러프	삼성
15	16	이재원	SK
16	17	손아섭	롯데
17	18	박건우	두산
18	19	허경민	두산

In [ ]:
```
# 타율
0.362,0.358,0.355,0.345,0.342,0.342,0.34,0.339,0.339,0.334,0.333,0.333,0.333,0.

# 안타
164,157,163,138,169,190,146,179,145,176,173,181,159,175,167,134,182,159,167

# 홈런
20,23,6,43,23,33,18,25,20,44,26,37,20,25,33,17,26,12,10

# 타점
101,77,57,112,118,90,93,103,83,133,108,125,84,119,125,57,93,84,79
```

In [44]:
```
# 타율
df['avg'] = [0.362,0.358,0.355,0.345,0.342,0.342,0.34,0.339,0.339,0.334,0.333,0

# 안타
df['hit'] = [164,157,163,138,169,190,146,179,145,176,173,181,159,175,167,134,18

# 홈런
df['homerun'] = [20,23,6,43,23,33,18,25,20,44,26,37,20,25,33,17,26,12,10]

# 타점
df['rbi'] = [101,77,57,112,118,90,93,103,83,133,108,125,84,119,125,57,93,84,79]

# 최종 data frmae
df
```

Out[44]:

	order	name	team	avg	hit	homerun	rbi
0	1	김현수	LG	0.362	164	20	101
1	2	양의지	두산	0.358	157	23	77
2	3	이정후	넥센	0.355	163	6	57
3	4	박병호	넥센	0.345	138	43	112
4	5	안치홍	KIA	0.342	169	23	118
5	6	전준우	롯데	0.342	190	33	90
6	7	김주찬	KIA	0.340	146	18	93
7	8	최형우	KIA	0.339	179	25	103
8	9	유한준	KT	0.339	145	20	83
9	10	김재환	두산	0.334	176	44	133
10	11	최주환	두산	0.333	173	26	108
11	12	이대호	롯데	0.333	181	37	125
12	13	구자욱	삼성	0.333	159	20	84
13	14	채은성	LG	0.331	175	25	119
14	15	러프	삼성	0.330	167	33	125
15	16	이재원	SK	0.329	134	17	57
16	17	손아섭	롯데	0.329	182	26	93
17	18	박건우	두산	0.326	159	12	84
18	19	허경민	두산	0.324	167	10	79

(1)의 data frame을 이용해서 문제를 풀어주세요

In [13]:
```
# 1~10순위의 name과 homerun column만 불러오기
df.loc[0:9, ['name', 'homerun']]
```

Out[13]:

	name	homerun
0	김현수	20
1	양의지	23
2	이정후	6
3	박병호	43
4	안치홍	23
5	전준우	33
6	김주찬	18
7	최형우	25
8	유한준	20
9	김재환	44

In [21]:
```
# 11~20순위의 order, team, hit, rbi 불러오기
# order, team, hit, rbi는 column이 1개씩 건너띄는 column입니다 이를 생각하고 풀어주세요
# ['order', 'team', 'hit', 'rbi'] <- 이거 사용 금지!

# loc을 통한 접근 - 사용 금지 내용
df.loc[10:, ['order', 'team', 'hit', 'rbi']]

# iloc을 통한 접근
df.iloc[10:, [x for x in range(0, 7, 2)]]
```

Out[21]:

	order	team	hit	rbi
10	11	두산	173	108
11	12	롯데	181	125
12	13	삼성	159	84
13	14	LG	175	119
14	15	삼성	167	125
15	16	SK	134	57
16	17	롯데	182	93
17	18	두산	159	84
18	19	두산	167	79

In [54]:
```
# for문을 사용해서 타점(rbi)이 100이상인 사람이 몇명인지 출력해보기
rbi_n = 0

for i in range(100, df.rbi.max()+1):
    for j in range(0, len(df.rbi)):
        if df.rbi[j] == i:
            rbi_n += 1

print(rbi_n)
```

```
9
```

In [67]:
```
# 타율이 0.3이상이고 홈런이 30개이상인 선수들의 평균 타율, 평균 안타, 평균 홈런, 평균 타점을 출력해주
# 타율(avg), 안타(hit), 홈런(homerun), 타점(rbi)
mean_avg = 0
mean_hit = 0
mean_homerun = 0
mean_rbi = 0

for i in range(0, len(df)):
    if (df.avg[i] >= 0.3) & (df.homerun[i] >= 30):
        mean_avg += df.avg[i]
        mean_hit += df.hit[i]
        mean_homerun += df.homerun[i]
        mean_rbi += df.rbi[i]

print('Players with a batting average of 0.3 or more and 30 or more home runs.\
      'Average batting average: ', mean_avg, '\n',
      'Average hits: ', mean_hit, '\n',
      'Average home runs: ', mean_homerun, '\n',
      'Average RBIs: ', mean_rbi)
```

```
Players with a batting average of 0.3 or more and 30 or more home runs.
 Average batting average:  1.6840000000000002
 Average hits:  852
 Average home runs:  190
 Average RBIs:  585
```

In [68]:
```
# hh라는 이름으로 hit * 1 + homerun * 4 column 추가하기
df['hh'] = df['hit'] + 4 * df['homerun']
df
```

Out[68]:

	order	name	team	avg	hit	homerun	rbi	hh
0	1	김현수	LG	0.362	164	20	101	244
1	2	양의지	두산	0.358	157	23	77	249
2	3	이정후	넥센	0.355	163	6	57	187
3	4	박병호	넥센	0.345	138	43	112	310
4	5	안치홍	KIA	0.342	169	23	118	261
5	6	전준우	롯데	0.342	190	33	90	322
6	7	김주찬	KIA	0.340	146	18	93	218
7	8	최형우	KIA	0.339	179	25	103	279
8	9	유한준	KT	0.339	145	20	83	225
9	10	김재환	두산	0.334	176	44	133	352
10	11	최주환	두산	0.333	173	26	108	277
11	12	이대호	롯데	0.333	181	37	125	329
12	13	구자욱	삼성	0.333	159	20	84	239
13	14	채은성	LG	0.331	175	25	119	275
14	15	러프	삼성	0.330	167	33	125	299
15	16	이재원	SK	0.329	134	17	57	202
16	17	손아섭	롯데	0.329	182	26	93	286
17	18	박건우	두산	0.326	159	12	84	207
18	19	허경민	두산	0.324	167	10	79	207

In [73]:
```
# 각 팀에 해당되는 인원이 몇명이 있는지 확인해보기
# data frame에 있는 두산, 롯데, KIA, 삼성, 넥센, LG, KT, SK에 각각 몇명이 있는지 출력
team_ds = 0   # Doosan
team_lt = 0   # Lotte
team_ka = 0   # KIA
team_ss = 0   # Samsung
team_nx = 0   # Nexen
team_lg = 0   # LG
team_kt = 0   # KT
team_sk = 0   # SK

for i in range(0, len(df)):
    if df.team[i] == '두산':
        team_ds += 1

    if df.team[i] == '롯데':
        team_lt += 1

    if df.team[i] == 'KIA':
        team_ka += 1

    if df.team[i] == '삼성':
        team_ss += 1

    if df.team[i] == '넥센':
        team_nx += 1

    if df.team[i] == 'LG':
        team_lg += 1

    if df.team[i] == 'KT':
        team_kt += 1

    if df.team[i] == 'SK':
        team_sk += 1

print('In the data frame, there are\n',
     team_ds, ' Doosan players.\n',
     team_lt, ' Lotte players.\n',
     team_ka, ' KIA players.\n',
     team_ss, ' Samsung players.\n',
     team_nx, ' Nexen players.\n',
     team_lg, ' LG players.\n',
     team_kt, ' KT player.\n',
     team_sk, ' SK player.')
```

```
In the data frame, there are
 5  Doosan players.
 3  Lotte players.
 3  KIA players.
 2  Samsung players.
 2  Nexen players.
 2  LG players.
 1  KT player.
 1  SK player.
```