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

In [2]:
```
# master table contains full info about specific player
# scoring table has seasonal information about player's records
master = pd.read_pickle(os.path.join("data","master.pickle"))
scoring = pd.read_pickle(os.path.join("data","scoring.pickle"))
```

In [3]:
```
master.head(2)
```

Out[3]:

	firstName	lastName	pos	birthYear	birthMon	birthDay	birthCountry	birthS
playerID								
aaltoan01	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN
abdelju01	Justin	Abdelkader	L	1987.0	2.0	25.0	USA	MI

In [4]:
```
scoring.head(2)
```

Out[4]:

	playerID	year	tmID	GP	G	A	Pts	SOG
0	aaltoan01	1997	ANA	3.0	0.0	0.0	0.0	1.0
1	aaltoan01	1998	ANA	73.0	3.0	5.0	8.0	61.0

In [5]:
```
# we want to join playerId (master) which is index, to the playerId(scoring) a
 normal column
pd.merge(master,scoring,left_index=True,right_on="playerID").head()
# Pandas on resulting data frame reset the whole index
```

Out[5]:

	firstName	lastName	pos	birthYear	birthMon	birthDay	birthCountry	birthState	
0	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN	Lap
1	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN	Lap
2	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN	Lap
3	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN	Lap
4	Justin	Abdelkader	L	1987.0	2.0	25.0	USA	MI	Mu

In [6]:
```
scoring.index
```

Out[6]:
```
RangeIndex(start=0, stop=28616, step=1)
```

In [7]:
```
scoring.index=scoring.index+3
```

```
In [8]: pd.merge(master,scoring,left_index=True,right_on="playerID").head()
```

Out[8]:

	firstName	lastName	pos	birthYear	birthMon	birthDay	birthCountry	birthState	
3	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN	Lap
4	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN	Lap
5	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN	Lap
6	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN	Lap
7	Justin	Abdelkader	L	1987.0	2.0	25.0	USA	MI	Mu

```
In [11]: # we want to set playerID as index
         pd.merge(master,scoring.set_index("playerID",drop=True),left_index=True,right_
         index=True).head()
```

Out[11]:

	firstName	lastName	pos	birthYear	birthMon	birthDay	birthCountry	birthS
playerID								
aaltoan01	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN
aaltoan01	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN
aaltoan01	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN
aaltoan01	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN
abdelju01	Justin	Abdelkader	L	1987.0	2.0	25.0	USA	MI

```
In [18]: # first table: left, right table: right; If i want to join the scoring and rem
         ove playerID as the index
         scoring=scoring.reset_index(drop=True)
         pd.merge(master,scoring, left_index=True,right_on="playerID").head()
```

Out[18]:

	firstName	lastName	pos	birthYear	birthMon	birthDay	birthCountry	birthState	
0	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN	Lap
1	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN	Lap
2	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN	Lap
3	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN	Lap
4	Justin	Abdelkader	L	1987.0	2.0	25.0	USA	MI	Mu

In [19]:

```
print(pd.merge(master,scoring,left_index=True,right_on="playerID").shape,
      pd.merge(master,scoring,left_index=True,right_on="playerID",how="right")
.shape)
```

```
(28616, 17) (28616, 17)
```

In [20]:

```
# drop random records
# lets drop 5 random rows from the master table using the sample method
master2=master.drop(master.sample(5).index)
master2.shape
print(pd.merge(master,scoring,left_index=True,right_on="playerID").shape,
      pd.merge(master,scoring,left_index=True,right_on="playerID",how="right")
.shape)
```

```
(28616, 17) (28616, 17)
```

In [21]:

```
#to understand this difference, we want to see additional columns on how = rig
ht
merged= pd.merge(master2,scoring,left_index=True,right_on="playerID",how="righ
t",indicator=True)
merged.head()
```

Out[21]:

	firstName	lastName	pos	birthYear	birthMon	birthDay	birthCountry	birthState	
0	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN	Lap
1	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN	Lap
2	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN	Lap
3	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN	Lap
4	Justin	Abdelkader	L	1987.0	2.0	25.0	USA	MI	Mu

In [22]:

```
merged["_merge"].value_counts()
```

Out[22]:

```
both           28587
right_only        29
left_only          0
Name: _merge, dtype: int64
```

In [24]:

```
merged[merged["_merge"].str.endswith("only")].sample(5)
```

Out[24]:

	firstName	lastName	pos	birthYear	birthMon	birthDay	birthCountry	birthState
10270	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
10276	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
20106	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25989	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25992	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [25]:

```
# outer join
merged=pd.merge(master2,scoring,left_index=True,right_on="playerID",how="oute
r",indicator=True)
merged.head()
```

Out[25]:

	firstName	lastName	pos	birthYear	birthMon	birthDay	birthCountry	birthState	
0	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN	Lap
1	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN	Lap
2	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN	Lap
3	Antti	Aalto	C	1975.0	3.0	4.0	Finland	NaN	Lap
4	Justin	Abdelkader	L	1987.0	2.0	25.0	USA	MI	Mu

In [26]:

```
merged["_merge"].value_counts()
```

Out[26]:

```
both          28587
right_only       29
left_only         0
Name: _merge, dtype: int64
```