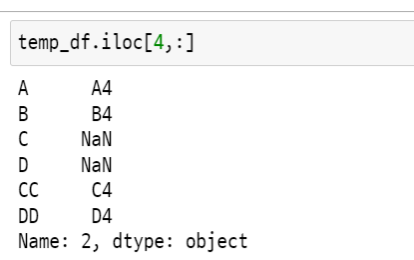
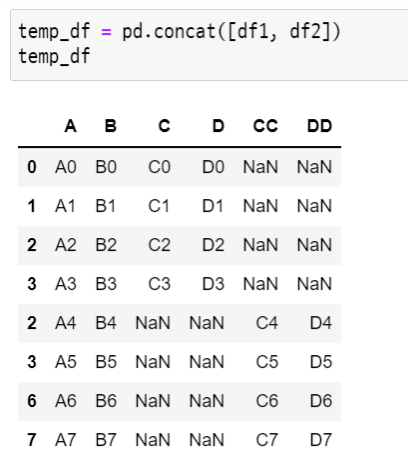
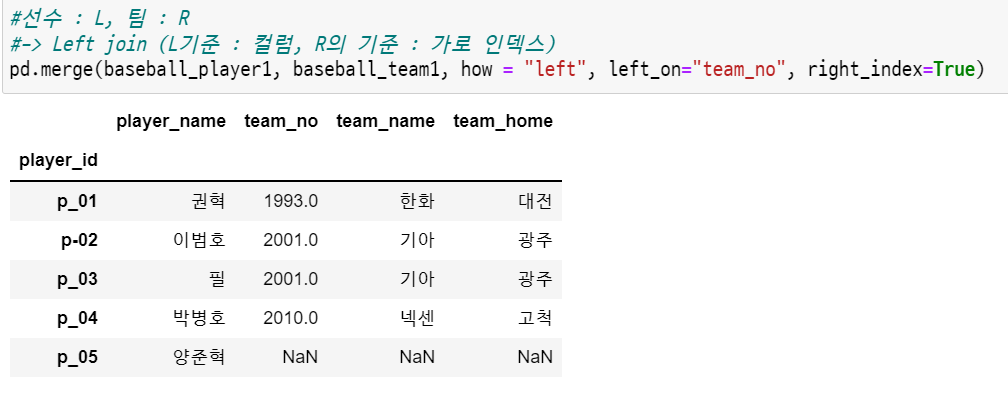
2022-02-11

Data Handling

* 변수 = pd.DataFrame( { “컬럼1“ :[“~”, “~”, ”~”], “컬럼2” : [“\*”, ”\*”, “\*”] }, index= [“1”,”2”,”3”] )
* pd.concat([df1, df2]) : concat은 리스트를 쓴 순서대로 아래로 합쳐준다
* 합칠 때 컬럼 이름이 다른 경우 -> 같은 컬럼은 아래로 붙이고 없는 컬럼은 옆으로 만들어준다( 서로컬럼 이름이 달라 없는 경우는 NaN으로 채워짐)
* 가로인덱스도겹치고 세로 인덱스도 겹치는 경우 -> 가로 인덱스에 똑같은 인덱스가 써진다 하지만 슬라이싱해서 불러올때는인덱스 순서대로 원래대로 불러오면 됨



* 중첩된 인덱스가 있어도 무시하고 원래 순서대로인 인덱스로 바꾸려면 : ignore\_index= True
* 컬럼이 달라서 옆으로 붙일 경우 : pd.concat([df1, df2], axis= 1) -> axis= 0으로 하게 될 경우 위에 처럼아래로 붙게 됨
* pd.concat([df1, df2], axis= 1).dropna() : Nan있는 값들 빼고 보여줌
* pd.concat([df1, df2], axis= 1).reindex(df2.index) :합친 후 df2인 인덱스만 보여줌
* Master의 데이터에서는 누락데이터가 없음
* Inner : 2개의 셋 모두 있는 경우에 관심 / 양쪽 정보가 다 있는 경우가 관심
* Excluding join : 데이터가 진짜 누락인지 아니면 실수가 된 것인지 관심사가 이빨 빠진쪽
* Outer join: 합집합, 전체 데이터 다 불러옴
* pd.merge(player, team, how= "left", left\_on="team\_no", right\_on= "team\_id" ) : 선수를 왼쪽으로 팀을 오른쪽으로 놓고 team\_no와 team\_id가같은 것끼리 합치려고 한다
* pd.merge(player, team, how="inner", left\_on="team\_no", right\_on= "team\_id") :선수와 팀을 team\_no와 team\_id가같은 것끼리 합치는데 빈값없는 것만 보여줌
* temp[ temp["team\_id"].isnull() ]: team\_id가없는 것만 보여줘 -> 불리언인덱싱
* pd.merge(player, team, how= "outer", left\_on="team\_no", right\_on="team\_id") :빈 값있어도 다 보여줘
* pd.merge(player1, team1, how= "left", left\_on="team\_no", right\_index=True) : 왼쪽을 선수로 기준으로 team\_no와오른쪽 team의인덱스가 같은걸기준으로 보여줌



* 위의 예시를 how= “right”로바꾸면

