PART 5. 데이터 사전 처리

1. 누락 데이터 처리

누락 데이터 확인

-데이터셋 로드한 이후, console 에서 df.head(): NaN 포함한 열 있는지 확인

df.info() : 각 열 별 유효 데이터 개수 df로 반환

-df[‘ 열이름 ’].value\_counts(dropna=False) : 옵션 default True시 NaN 제외한 유효 개수 반환

- 누락 데이터 직접 찾기

Df.isnull() : 누락 데이터면 True 반환, 유효 데이터 존재 시 False 반환

Df.notnull: 유효 데이터 존재 시 True 반환, 누락 데이터면 False 반환

해당 데이터프레임의 각 원소 자리에 T/F 입력된 df 반환

-누락 데이터 개수 구하기

Df. head().isnull().sum(axis=0)

Isnull은 True시 1, False 시 0으로 판별

Sum(axis=0): 참(1)의 합, 즉 참의 개수, 즉 NaN의 개수 확인 가능

(<http://taewan.kim/post/numpy_sum_axis/> )

이후 내용부터 예제 코드 주석에 정리.