전체 열,행 기준 합 a열 평균값 확인

: df.sum(axis=0(or1)) : df['a'].mean()

# tip 데이터프레임 확인

tip\_df=pd.DataFrame(stock, index=dates, columns=names)

df 기본 통계 메서드

집계하기

결측값 조회

a열 별 b합을 구하고 싶을 때

df.groupby(by='a',as\_index=True)['b'].sum()

인덱스 값을 day로 지정하고 싶을 때

df.groupby(by='a',as\_index=False)['b'].sum()

df.info()로 결측치 여부 확인

isnull()=isna() : 결측치 True

notnull()=notna() : 결측치 False

df.isna().sum():열의 결측치 개수 조회

df의 date날짜 값을 날짜형식으로 변환

: pd.to\_datetime( df['date'])

날짜형식에서 년도 뽑아내기

df[‘date’].dt.year

날짜 형식에서 달 뽑아내기

df[‘date’].dt.month

날짜 입력

결측값 제거

dropna() : 결측치가 있는 열이나 행 제거 메서드

1) 행 제거

: df.dropna(axis=0,inplace=True)

2) 열 제거

: df.dropna(axis=1,inplace=True)

3) subset 옵션을 이용해 a열의 결측치가 있는 행 제거

: df.dropna(subset=['a'], axis=0 ,inplace=True)

a열을 가변수화 할 때

dumm\_cols=['a']

df=pd.get\_dummies(df, columns=dumm\_cols,drop\_first=True ,dtype=int)

여러 열을 가변수화 할 때

dumm\_cols=['a','b','c']

df=pd.get\_dummies(df,columns=dumm\_cols,drop\_first=True ,dtype=int)

가변수 만들기

reset\_index()메서드이용

drop=True(인덱스값제거)

inplace=True 사용가능.

특정열 특정 행

df.[‘a’] df[1]

여러 열 여러 행

df.[[‘a’,’b’]] df[1,3,5]

특정 조건 여러 조건 between

df.loc[df[’a’]>=80] df.loc[df[‘a].between(n\_1,n\_2)]

열 조회

인덱스 초기화

1) fillna()로 mean\_a값을 이용해 결측치 채우기

df['a'].fillna(mean\_a,inplace=True)

2) 특정값을 이용해 결측치 채우기

: df['a'].fillna(0,inplace=True)

3) a열의 결측값을 앞의 값으로 채우기

df['a'].fillna(method='ffill',inplace=True)

4) a열의 결측값을 뒤의 값으로 채우기

df['a'].fillna(method='bfill',inplace=True)

5) 선형보간법으로 a열 채우기

df ['a'].interpolate(method='linear',inplace=True)

결측값 채우기

concat(): 인덱스 값을 기준으로 프레임 합치기(기본axis1)

가로로 합치기: pd.concat([(값1),(값2)],axis=1,join='outer')

세로로 합치기: pd.concat([(값1),(값2)],axis=0,join='outer')

세로로 합칠 시 중복되는 인덱스 초기화

: reset\_index(drop=True, inplace=True) 사용

merge : 같은 이름의 열이 있을 시 on 없이도 그 열을 기준으로 조인 가능

pd.merge( 값1 , 값2 ,on='year',how='inner')

파일 합치기

파일 불러오기. 저장

파일을 불러올 때 : pd.read\_csv(path)값을 입력

파일을 저장할 때 : (파일 명).to\_csv(저장할 이름)름),index=True)