1. NUMPY

NUMPY에서 활용되는 함수

널값을 주는 법 🡪 np.nan, None

Ndarray 생성하기 🡪 Np.array

자료형 구하기 🡪 .dtype

자료형 변환(사이즈 크기) 🡪 .astype

배열 차원 및 행열 조정 🡪 .shape

몇 차원인지 알아보기 🡪 .ndim

배열에 수 랜덤(1~N) 배치 🡪 Np.arange

배열에 0 배치 🡪 Np.zeros

배열에 1 배치 🡪 Np.ones

배열 행열 사이즈 조정 🡪 .reshape

2. 인덱싱 INDEXING

특정 데이터 추출 🡪 [,] [], arr.ix[](대괄호 안에 들어가는 건 행,열을 나타내는 숫자이거나 문자—명칭 or 위치 기반 인덱싱, arr.iloc[](대괄호 안에 들어가는 건 오로지 숫자 기반—위치기반 인덱싱), arr.loc[](대괄호 안에 들어가는 건 오로지 문자 기반—명칭 기반 인덱싱)

슬라이싱 🡪 [n:m], [n, m:u]

불리언 인덱싱 🡪 arr3[arr3>25]

3. 정렬

오름차순, 내림차순 🡪 오름차순 같은 경우에는 sort, 내림차순 경우에는 reverse = True를 적용하면 됨. 🡪 np.sort(), array.sort()

수를 기준으로 정렬하되, 가시적으로는 인덱스가 출력. 🡪 np.argsort(), array.argsort()

4. 선형대수 연산 – 행렬 내적과 전치 행렬 구하기

행렬 내적 연산 🡪 np.dot()

전치 행렬 🡪 np.transpose()

5. 결측값

데이터 프레임에 널값이 있는지 확인하는 법 🡪 isnull, isna (여기 뒤에 sum()함수를 붙이거나, mean()함수를 붙일 수 있다. 널값이 총 몇개인지 확인 가능.)

널값 메꾸기 🡪 fillna()

6. PANDAS

pd.isnull(data).sum(0) 🡪 0이면 데이터 프레임의 열로 널값을 센다. 1이면 데이터 프레임의 행으로 널값을 센다.

데이터 프레임 생성, 외부 파일 읽어오기 🡪 pd.read.csv()

데이터 프레임의 윗부분, 아랫부분 출력 🡪 df.head(), df.tail()

데이터 프레임의 모양 표현 🡪 df.shape

데이터 프레임에 대한 대략적인 정보 출력 🡪 df.describe(), r의 str(), dplyr::glimpse()와 유사