4/27 강의노트

테이블이(가) 표시된 사진

자동 생성된 설명

시리즈 연산

시리즈끼리 더하기 = 인덱스를 정렬한 후 같은 인덱스끼리 더한다 (단순히 보이는 순서대로 하는것이 아님)

테이블이(가) 표시된 사진

자동 생성된 설명

시리즈끼리 곱하기 = 덧셈과 같은 메커니즘

테이블이(가) 표시된 사진

자동 생성된 설명

시리즈에 숫자를 더할 경우 모든 값에 대상 값을 더함 (리스트에 숫자 더하는건 지원 안하는데 판다스 시리즈에는 연산 가능)

테이블이(가) 표시된 사진

자동 생성된 설명

시리즈에 숫자 곱하기

텍스트이(가) 표시된 사진

자동 생성된 설명

인덱스 수가 다른 시리즈끼리 연산 = 인덱스가 동일한 데이터끼리만 연산되고 인덱스가 한쪽만 있는 쪽은 NaN을 반환함

위 연산은 원본이 바뀌지 않는 비파괴 명령어

텍스트이(가) 표시된 사진

자동 생성된 설명sort\_index로 인덱스 기준 정렬이 가능

텍스트이(가) 표시된 사진

자동 생성된 설명

순서를 뒤집어도 인덱스가 같은 값끼리 더하므로 최종 결과는 같다

테이블이(가) 표시된 사진

자동 생성된 설명

데이터프레임 다루기

조건을 달아서 특정 데이터만 가져오기

테이블이(가) 표시된 사진

자동 생성된 설명

조건식은 시리즈의 형태로 나타나서 이와 데이터프레임과 결합하여 특정 값만을 뽑아냄

텍스트이(가) 표시된 사진

자동 생성된 설명

데이터프레임 연산 : 문자열은 곱셈시 반복, int64는 곱셈시 해당 연산 수행

텍스트이(가) 표시된 사진

자동 생성된 설명

판다스의 to\_datetime 명령어 : 문자를 format으로 형식을 지정하여 날짜 타입으로 바꿔줌

%Y : 2019 등 4자리수로 표현된 연도, %y : 03 등 2자리수로 표현된 연도

파이썬 공식 홈페이지 등에서 확인 가능

텍스트이(가) 표시된 사진

자동 생성된 설명

날짜 형식을 추가하여 dtypes 조회

테이블이(가) 표시된 사진

자동 생성된 설명

날짜끼리 연산이 가능하여 뺄셈을 하여 두 날짜 사이의 일 수를 계산 가능

테이블이(가) 표시된 사진

자동 생성된 설명

시간끼리 연산하는 timedelta64 타입은 초, 분, 일수 등을 구분하기 위하여 days 등의 단위가 붙어서 나옴 (타입의 고유 특성)

텍스트이(가) 표시된 사진

자동 생성된 설명

열의 삭제 : drop 명령어로 삭제 가능, (행 또는 열의 이름, axis = 0 (열 삭제), 1 (행 삭제))

해당 명령어는 비파괴 명령어라서 원본은 바뀌지 않음

테이블이(가) 표시된 사진

자동 생성된 설명

Inplace = True를 추가하면 파괴적 명령어로 간주되어 원본데이터가 바뀔 수 있다 (메모리 절약 차원에서 사용 가능)

텍스트이(가) 표시된 사진

자동 생성된 설명

데이터의 저장

피클 : 파이썬에서만 사용하는 파일 형식으로 변수 등까지 모두 저장됨 (개발자간에 파일이라 파면 보통 피클을 의미함)

테이블이(가) 표시된 사진

자동 생성된 설명

Read\_pickle 명령어로 피클 불러오기 가능

텍스트이(가) 표시된 사진

자동 생성된 설명

Csv 저장 – 불러오기 : 저장할때 인덱스가 같이 저장이 되었는데 읽을 때 인덱스 넘버를 컬럼으로 불러들였음

테이블이(가) 표시된 사진

자동 생성된 설명

해당 이슈를 방지하기 위하여 저장시 index = False로 인덱스를 빼고 저장함 (인덱스가 단순 넘버링일 경우 사용), 매우 중요함

Or 불러올 때 index\_col로 지정

테이블이(가) 표시된 사진

자동 생성된 설명

Xlsx로 저장 및 불러오기

테이블이(가) 표시된 사진

자동 생성된 설명

데이터 연결 : concat 명령어로 데이터프레임을 연결할 수 있음, 연결 시 같은 컬럼끼리 연결됨

테이블이(가) 표시된 사진

자동 생성된 설명

컬럼명이 일치하지 않으면 (시리즈거나) 데이터가 제대로 이어지지 않고 없는 부분은 모두 NaN이 됨

텍스트이(가) 표시된 사진

자동 생성된 설명

컬럼명을 일치시켜 데이터 붙임 : concat은 비파괴 명령어임

텍스트이(가) 표시된 사진

자동 생성된 설명

딕셔너리는 append 명령어를 이용하여 데이터프레임에 데이터 추가 가능 (ignore\_index로 인덱스 넘버를 무시하고 넣어야 함, 일단 데이터를 쑤셔넣고 인덱스 넘버를 다시 부여하는 방식)

테이블이(가) 표시된 사진

자동 생성된 설명

데이터프레임 연결시에도 인덱스 넘버가 중복이 일어나지 않게 할 때 ignore\_index 사용 가능

달력이(가) 표시된 사진

자동 생성된 설명

Concat에 axis 매개변수를 1로 지정하면 열 방향으로 데이터가 연결됨 (인덱스가 같은 데이터끼리 합쳐짐)

테이블이(가) 표시된 사진

자동 생성된 설명

같은 컬럼명을 쓰는 여러 컬럼이 있을 경우 열 이름으로 추출시 모두 추출됨

텍스트이(가) 표시된 사진

자동 생성된 설명

직접 컬럼명을 지정하여 데이터 추가

달력이(가) 표시된 사진

자동 생성된 설명

Ignore\_index로 열 이름 다시 지정 (연결부터 하고 기존 열 이름 무시하고 0부터 자동으로 기입)

달력이(가) 표시된 사진

자동 생성된 설명

컬럼 명이 다른 데이터를 그냥 연결하면 데이터가 엉망이 됨

테이블이(가) 표시된 사진

자동 생성된 설명

Concat에 join 매개변수를 inner로 지정하면 공통된 열만 합칠 수 있음 (SQL의 내부 조인과 비슷)

테이블이(가) 표시된 사진

자동 생성된 설명

공통된 컬럼명이 없으면 컬럼이 없는 데이터프레임이 생성됨

테이블이(가) 표시된 사진

자동 생성된 설명

Merge (데이터 병합) : SQL의 inner join과 유사함

테이블이(가) 표시된 사진

자동 생성된 설명

특정 열을 기준으로 (on으로 지정) 두 데이터프레임을 병합, df2에 없는 키 (c)는 출력되지 않음

테이블이(가) 표시된 사진

자동 생성된 설명

누락값 (결측치, NaN, NULL)은 값이 알수 없는 상태이므로 isnull 명령어를 통해서만 True를 판별 가능하며 그 이외의 모든 값과도 동일하지 않다 (NaN 자신과도)