

Pandas

- Python 대표 데이터 분석/처리 오픈소스 라이브러리
- 2008년 wes Mckinney의 개인 프로젝트로 시작
- 다양한 데이터 입/출력, 변환, 선택, 결합, 클리닝, 그룹핑 지원

Pandas 객체

- DataFrame: 테이블 데이터 객체
- Series : 벡터(리스트형태 - 1차원) 데이터 객체(인덱스까지 같이 저장됨)

Pandas function

- head(), tail() -> 첫/마지막 데이터 출력
- [], iloc[] (위치를 가지고), loc[] (컬럼, 인덱스 이름을 가지고) -> 행/열 선택
- apply() -> 행/열에 특정 함수 적용
- groupby() -> 특정 열로 그룹핑
- sort_values() -> 값에 따라 정렬
- sort_index() -> 인덱스에 따라 정렬
- drop(), drop_duplicates() -> 행/열 삭제
- Shape() -> 열/열 사이즈 리턴
- Count(), sum(), cumsum()
- mean(), median(), min(), max(), std(), var()
- sample() -> 개수 or 퍼센트를 지정해서 샘플을 만듦
- unique() -> 해당 열에 unique한 값이 몇종류 있는지 리턴
- nunique() -> 해당 열에 unique한 값이 몇개 있는지 리턴
- value_counts() -> 각각의 값이 몇번 보이는지 리턴
- set_index(), reset_index() -> 인덱스값 변경
- read_csv(), to_csv -> csv파일 입출력
- merge(), concat() -> 서로 다른 dataframe 결합
- hist(), plot() -> 데이터 히스토그램/시각화
- dt.day/weekday/month/year/weekofyear -> datetime[ns] 타입 데이터의 일/요일/월/년/주
- Resample() -> datetime[ns] 타입 데이터 리샘플링
- Pivot_table() -> 피벗테이블 생성
- corr() -> 상관관계 계산
- describe() -> 데이터 개요 출력
- dtype() -> 데이터 타입을 보여줌
- Quantile() -> 각각 값에대한 퍼센트지를 보여줌