

1. 공식 문서를 읽다 보니 Pandas의 다양한 기능에 놀라움을 느꼈습니다. `groupby`와 같은 여러 함수를 보면서 SQL과 비슷하다고 느꼈고, `min`, `max` 등 열 계산을 통해 Excel과 닮았다는 점이 인상 깊었습니다. 특히, 데이터를 쉽게 변환하고 간단하게 시각화할 수 있다는 점이 좋았습니다. 예를 들어, `to` 함수를 이용해 JSON 데이터를 Excel로 쉽게 변환할 수 있었고, `plot()` 함수를 사용해 간단한 시각화도 가능했습니다.

2

df.head()  
✓ 0.0s

	date	date_edit	href	title	article
0	2024-07-24 23:58:00	2024.07.25 09:08	https://www.hankyung.com/article/202407247138i	美 7월 기업활동 종합 PMI 55.0...27개월 만에 최고치	종합구매관리지수(PMI)로 본 미국의 기업 활동이 서비스 부문의 호조로 7월에 27...
1	2024-07-24 23:07:00	2024.07.25 09:08	https://www.hankyung.com/article/202407247118i	기술주 실적 실망에 나스닥, S&P500 1% 이상 하락	대형주 실적 발표 시즌이 실망스럽게 출발하면서 24일(현지 시간) 뉴욕증시는 하락으로...
2	2024-07-24 22:42:00	2024.07.25 09:08	https://www.hankyung.com/article/202407247089i	글로벌 엔케리트레이딩 청산...엔저 끝나 가나	글로벌 캐리 트레이드 수요가 줄기 시작하면서 일본 엔화가 24일(현지시간) 3일 연...
3	2024-07-24 22:00:00	2024.07.24 23:31	https://www.hankyung.com/article/202407246354g	"편의점 안 가도 되겠네"...3000원대 도시락 내놓자 '불타'	최근 대형마트를 찾는 발길이 예전만 못한 중에서도 마트의 '델리 코너' 메뉴들은 불...
4	2024-07-24 21:53:00	2024.07.25 09:08	https://www.hankyung.com/article/202407247056i	알파벳, 예상 넘는 실적에도 주가하락...이유는?	구글 모회사 알파벳이 예상을 넘는 매출과 마진을 보고하고도 24일(현지시간) 뉴욕증...

```

df.info()
✓ 0.0s

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 683 entries, 0 to 682
Data columns (total 5 columns):
#   Column      Non-Null Count  Dtype
---  -
0   date        683 non-null   datetime64[ns]
1   date_edit   683 non-null   object
2   href        683 non-null   object
3   title       683 non-null   object
4   article     683 non-null   object
dtypes: datetime64[ns](1), object(4)
memory usage: 26.8+ KB

```

3 어떤 특징 , 주로 어디에 쓰이는지 , 장단점

1 pickle

소개 : python 객체를 직렬화 하고 그 데이터를 파일로 저장하는 포맷

특징 : 기본 라이브러리로 제공, 객체를 그대로 저장하고 불러올 수 있음

사용 용도 : 객체 저장 및 불러오는데, 모델, 데이터 프레임, 사전 등 객체 저장에 사용

장단점 : 장점 : 모든 python 객체 저장 가능, 단점 : 다른 언어에서 사용하기 어려움

## 2 CSV, TSV

소개 : 텍스트 파일의 일종으로, 콤마(,) 또는 탭(tab)으로 구분하여 저장

특징 : 행과 열로 구성된 테이블 형식

사용 용도 : 데이터 교환 및 저장에 널리 사용

장단점 : 장점 : 간단하고 널리 사용됨, 단점 : 대규모 데이터에 비효율적

## 3 json

소개 : 웹 애플리케이션에서 데이터를 전송할 때 사용

특징 : 가독성이 좋고, 계층적 데이터 구조를 지원

사용 용도 : 웹 애플리케이션에서 클라이언트와 서버 간의 데이터 교환에 사용

장단점 : 장점 : 가독성이 좋고, 다양한 프로그래밍 언어에서 지원됨, 단점 : 데이터가 크면 성능저하 될 수 있음

## 4 HTML

소개 : 웹 페이지를 작성하는 언어

특징 : 웹 페이지의 구조를 정의하는 언어

사용 용도 : 웹 페이지 작성

장단점 : 장점 : 웹 브라우저에서 쉽게 데이터 확인 가능 : 단점 : 브라우저 외에는 사용할 곳이 없음

## 5 XML

소개 : 데이터를 구조화하여 저장하는 언어로 다양한 애플리케이션에서 사용됨

특징 : 태그를 사용하여 데이터의 구조를 표현

사용 용도 : 데이터 전송과 저장에 사용

장단점 : 장점 : 데이터 구조를 명확히 표현할 수 있다, 단점 : 데이터가 크면 가독성이 떨어진다

## 6 parquet

소개 : 컬럼 지향 저장 포맷으로, 대규모 데이터를 효율적으로 저장할 수 있음

특징 : 컬럼 기반 저장으로 효율적인 압축과 인코딩 제공

사용 용도 : 대규모 데이터 분석과 저장에 사용

장단점 : 장점 : I/O 성능 최적화, 단점 : 텍스트 에디터로 확인 어려움

## 7 YAML

소개 : 데이터 직렬화 포맷으로, 구성 파일 등에 널리 사용

특징 : 들여쓰기를 사용하여 데이터 계층 구조 표현

사용 용도 : 설정 파일과 데이터 직렬화에 사용

장단점 : 장점 : 가독성이 좋다, 단점 : 대규모 데이터 처리에 부적합

## 8 TOML

소개 : 간결한 데이터 직렬화 포맷

특징 : 키-값 쌍을 사용하여 데이터를 표현

사용 용도 : 설정 파일에 주로 사용됨

장단점 : 장점 : 가독성이 좋고 구성 파일로 적합, 단점 : 널리 사용되지 않고 대규모 데이터 처리에 부적합

IO 란 : 컴퓨터에서 데이터의 입력과 출력

Pandas에서의 IO : 다양한 파일을 읽고 다양한 파일로 내보내기

예) `pd.read_csv` > `df.to_json`

pickle 은 python 객체를 저장하고 불러오는 기능을 제공하는 모듈이고 네트워크를 통해 전송도 가능하다.

직렬화 : python 객체는 바이트 스트림으로 변환

역직렬화 : 바이트 스트림을 다시 원래의 python 객체로 복원하는 과정

즉 내가 만든 모듈을 다른 사람이 받을 수 있게 직렬화를 통해 바이트로 변환하여 네트워크를 (네트워크는 바이트 형태로만 보낼 수 있음) 이용해서 보내고 받을 때 역직렬화를 통해 원본의 모습으로 받는다.