

2024.10.12. 1-3장 공부

- **공개 데이터 세트**는 기업이나 정부 등이 무료로 공개하는 데이터 세트입니다. 누구나 저작권에 상관없이 다운로드하여 데이터 분석이나 제품 개발에 활용할 수 있습니다.
- **CSV 파일**은 콤마(,)로 구분된 텍스트 파일입니다. 한 줄이 하나의 레코드이며, 레코드는 콤마로 구분된 여러 필드 혹은 열로 구성됩니다. 데이터가 엑셀처럼 표 형태를 가져야 하기 때문에 레코드에 있는 필드 개수는 모두 동일해야 합니다.
- **판다스**는 표 형식 데이터를 위한 편리한 도구를 다양하게 제공하는 강력한 데이터 분석 패키지입니다. CSV나 엑셀 파일을 읽거나 쓸 수 있으며 데이터 분석을 위한 많은 기능을 제공합니다.
- **데이터프레임**은 판다스의 핵심 데이터 구조로 행과 열로 구성됩니다. CSV 파일이나 엑셀 파일로부터 데이터프레임을 만들 수 있습니다.
- **시리즈**는 1차원 배열과 흡사한 판다스의 객체입니다. 시리즈는 한 종류의 데이터만 포함할 수 있습니다.

<code>pandas.read_csv</code>	CSV 파일을 읽어 데이터프레임을 만듭니다.
<code>DataFrame.head</code>	데이터프레임에서 처음 다섯 개의 행을 반환합니다.
<code>DataFrame.to_csv()</code>	데이터프레임을 CSV 파일로 저장합니다