**미니프로젝트**

**# 발상의 시작 :**

**당해 년도 어떤 분야의 책들이 대중들의 이목을 집중시켰을까?.**

**# 아이디어 구상**

1. **데이터 수집 단계**

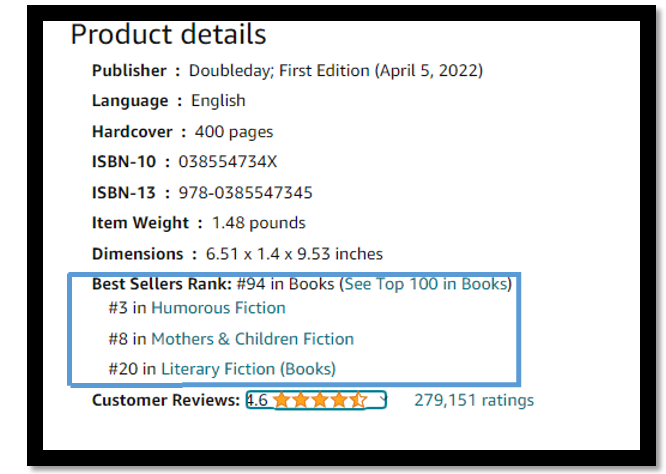
2013~2023 데이터를 web-scrapping 을 통해 분석:

1위부터 100위까지의 데이터에서 분야 별 도서가 차지 하는 비율을 파악

초기에는 한국과 미국의 연도별 베스트 셀러에 대한 분석

위 프로젝트의 핵심목표는 머신 러닝을 통해 도서 분야 별 가중치를 구하고 국가 단위에서 가중치에 대한 통계적의 유의성을 판단해보는 것으로 정하려고 하였으나

데이터 수집 단계에서 어려움 봉착

아래 그림과 같이 주석과 함께 총 3 개의 하위 카테고리로 책의 분야가 설명됨 

<https://www.amazon.com/Lessons-Chemistry-Novel-Bonnie-Garmus/dp/038554734X/ref=zg_bsar_g_books_d_sccl_6/146-5497212-4365122?psc=1>

하지만 데이터의 특성을 확인해 보면 Fiction이라는 단어가 총 3회 반복 출현되었다

이를 통해 위 책은 소설 책임을 유추할 수 있었다. 따라서



<https://product.kyobobook.co.kr/bestseller/total?period=004#?page=1&per=20&period=004&ymw=&bsslBksClstCode=A>

# 기대 결과

연도별 어떤 분야가 사랑을 받았는지에 대한 통계치 파악

1. 통계적으로 연도 별 어떤 변화가 있었고
2. 그 변화 속에서 어떤 Insight를 얻을 수 있을지.

# 활용 툴

Python

# web-scrapping : beautifulSoup/selenium

# pandas/numpy

2013~2023 연도별 책 배스트셀러들의 분야 파악하여 연도별 사람들의 최대 관심 트렌드 분석

자료 다운 :

<https://product.kyobobook.co.kr/bestseller/total?period=004#?page=1&per=20&period=004&ymw=&bsslBksClstCode=A>

교보문고 통계치 EXCEL 파일