도서 구매 사이트를 활용한 데이터 분석

**1. 프로젝트 개요**

1.1 주제 선정 이유

현대 사회에서 디지털 기기의 보급과 미디어 소비 증가로 인해 전통적인 독서 문화가 줄어들고 있습니다. 독서량의 감소는 개인의 성장 뿐만 아니라 사회 전체의 문화적 수준에도 부정적인 영향을 미칩니다. 이 프로젝트는 도서 구매 사이트의 데이터를 분석하여 독서 패턴을 파악하고 독서 문화를 활성화시키는 데 목적이 있습니다.

**2. 목표 및 목적**

2.1 목표

* 도서 구매 사이트를 통해 현재까지 출판된 베스트셀러 도서들을 분석
* 연령대, 성별, 분야 등을 분석하여 독서 패턴을 파악
* 독서량 감소 문제를 해결하기 위한 인사이트 제공

2.2 목적

* 출판사와 서점이 주 고객층을 설정하고 전략을 수립하는 데 도움
* 독서에 대한 사람들의 관심을 끌어올리기

**3. 데이터 수집**

* 1. 데이터 소스
* 주요 도서 구매 사이트

- ( 해당 프로젝트에서는

YES 24(<https://www.yes24.com>), 인터파크(<https://www.interpark.com>) 활용 )

3.2 크롤링 방법

* Selenium과 BeautifulSoup을 활용하여 웹 크롤링
* 크롤링 대상: 도서 제목, 저자, 출판사, 가격, 판매지수 등

3.3 데이터 수집 과정

* 웹 페이지 로딩
* 도서 목록 페이지 탐색 및 데이터 추출
* 필요한 정보가 포함된 HTML 요소 선택
* 데이터 저장 (CSV, JSON 등)

**4. 데이터 전처리**

4.1 데이터 정제

* 결측치 처리
* 중복 데이터 제거
* 데이터 형식 변환

4.2 데이터 변환

* 필요한 열 추출
* 새로운 열 생성 (예: 연령대 분류, 장르 분류)

**5. 데이터 분석**

5.1 기술 통계 분석

* 연령대별 도서 구매 패턴
* 성별에 따른 선호 도서 장르
* 분야별 베스트셀러 분석

5.2 시각화

* 파이 차트, 바 차트, 히스토그램 등 시각적 자료 활용
* 주요 인사이트 도출

**6. 결과 및 논의**

6.1 분석 결과 요약

* 주요 분석 결과 요약
* 도출된 인사이트

6.2 제안 사항

* 출판사 및 서점에 대한 전략 제안
* 독서 문화 활성화를 위한 제안

**7. 결론**

7.1 프로젝트 요약

* 프로젝트의 주요 내용 및 결과 요약

7.2 향후 연구 방향

* 추가 연구 필요성 및 제안

**8. 부록**

8.1 분석 데이터의 출처 정리

|  |  |  |  |
| --- | --- | --- | --- |
| **구분** | **데이터** | **출처** | **기준연월** |
| 행정동 | 서울 | 서울시 광장 | 21 |