도서 구매 사이트를 활용한 데이터 분석

1. 프로젝트 개요

1.1 주제 선정 이유

1.1 주제 선정 이유

현대 사회에서 디지털 기기의 보급과 미디어 소비 증가로 인해 전통적인 독서 문화가 줄어들고 있습니다. 독서량의 감소는 개인의 성장뿐만 아니라 사회 전체의 문화적 수준에도 부정적인 영향을 미칩니다. 이 프로젝트는 도서 구매 사이트의 데이터를 분석하여 고객들의 분류에 맞는 독서 패턴을 파악하여 Needs를 분석하여 고객 뿐만 아니라 출판사를 대상으로 서비스를 제공하여 독서 문화를 활성화시키는 데 목적이 있습니다.

1.2 목표 및 목적

2.1 목표

도서 구매 사이트를 통해 현재까지 출판된 베스트셀러 도서들을 분석

연령대, 도서 분야별 도서 구매량을 분석하여 도서 소비 트렌드를 파악

독서량 감소 문제 해결 및 도서 소비 증가를 위한 인사이트 제공

2.2 목적

출판사와 서점이 주 고객층을 설정하고 전략을 수립하는 데 도움

독서에 대한 사람들의 관심을 끌어올리기

2. 데이터 수집

2.1 데이터 소스 추출 사이트 선정

주요 도서 구매 사이트- ( 해당 프로젝트에서는 예스24(https://www.yes24.com), 인터파크 도서(https://book.interpark.com/) 활용 )

온라인 시장 점유율 : 예스24 (약 40%, 업계 1위), 인터파크 도서 (약 10%, 업계 4위)

2.2 크롤링 방법

Selenium과 BeautifulSoup을 활용하여 웹 크롤링

크롤링 대상: 도서정보(제목, 저자, 출판사, 가격, 판매지수, 평점, 판매량 등)

고객정보(연령대, 성별)

2.3 데이터 수집 과정

3. 데이터 전처리

3.1 데이터 정제

결측치 확인 & 처리

목적에 불필요한 데이터 제거

데이터 형식 변환

비슷한 도서 카테고리 통합

3.2 데이터 변환

필요한 열 추출

새로운 열 생성 (예: 날짜) 주제

4. 데이터 분석

4.1 기술 통계 분석

연령대별 도서 구매 패턴

연령대별 독서량 변화 확인

연령대별 선호 도서 종류

4.2 시각화

파이 차트, 바 차트, 히스토그램 등 시각적 자료 활용

주요 인사이트 도출

5. 결과 및 논의

5.1 분석 결과 요약

주요 분석 결과와 인사이트를 요약합니다.

5.2 제안 사항

출판사 및 서점에 대한 전략과 독서 문화 활성화 방안을 제안합니다.

독서 문화 활성화를 위한 제안

향후 연구 방향 필요성 및 제안

6. 참고

6.1 참고한 뉴스, 서적, 웹사이트 등

코로나 시대의 '명암'…성장하는 예스24 vs 쇠락하는 교보문고 (https://www.ibtomato.com/Mobile/mView.aspx?no=5460)

책 덮은 대한민국, 당신은?…1년간 단 한권도 안읽은 성인 53% (https://www.mk.co.kr/news/society/10482768)

여유가 부족한 현대 사회...독서가 더 필요한 이유

(https://www.stcarollo.or.kr/0401/2046)