

도서관 이용과 독서문화프로그램 간 상관관계에 대한 분석

CUAI 7기 DA4팀

김부영(융합공학), 송경준(수학), 임유민(예술공학)

[요약] 서울특별시 8개 도서관을 대상으로 독서문화프로그램과 도서관 이용 사이의 상관관계를 분석하였다. 행사 수와 대출건수 간 뚜렷한 상관관계는 없었으나, 대출 건수를 도서 카테고리로 세분화한 결과 일부 유의미한 상관관계가 확인되었다. 특히 접수기간과 음의 상관관계, 강의 지속시간과 양의 상관관계를 보여, 프로그램 특성이 도서관 이용에 차별적 영향을 줄 수 있음을 나타낸다.

1. 서론

도서관은 지식과 정보의 중심지로서, 지역 사회의 교육과 문화 발전에 중요한 역할을 하고 있다. 최근 들어 도서관은 전통적인 자료 제공의 역할뿐만 아니라 다양한 독서문화프로그램을 통해 사회적인 기능을 확장하고 있다. 독서문화프로그램은 지역 주민에게 다양한 기회를 제공하고 도서관 이용을 증진시키며 나아가 지역사회의 문화활동을 활발하게 전개한다는 목적을 가지고 문화(예술)강좌, 독서진흥활동, 문화행사 등의 프로그램을 제공한다 1).

본 연구는 '독서문화프로그램이 실제로 도서관 이용을 활성화 하는가'라는 의문에서 시작되었다. 선행연구에서 참여한 프로그램의 유형에 따라 도서관 이용행태에 차이가 있음을 보고한 바가 있다 2). 그러나 해당 연구는 설문지를 통해 프로그램의 유형 및 참여여부만 확인하기 때문에 세부적으로 어떤 프로그램이, 얼마나 상관관계가 있는지에 대해 정량적인 평가를 할 수 없다는 한계점이 있었다.

따라서 본 연구에서는 데이터에 기반하여 독서문화프로그램과 도서관 이용 간의 상관관계를 분석함으로써, 이러한 프로그램들이 도서관 이용 행태에 미치는 영향을 규명하고자 한다. 이를 통해 도서관 운영자들이 보다 효과적인 프로그램 기획에 기여할 수 있다. 독서문화프로그램의 효과를 구체적으로 평가하여 도서관 서비스 개선과 이용자 참여 증대에 중요한 기초

자료를 제공할 것이다.

2. 본론

1) 데이터 수집

서울특별시교육청 통합도서관 및 평생학습관에서 제공하는 독서문화프로그램 데이터를 수집하여 분석을 진행하였다. 웹 크롤링 도구인 BeautifulSoup 라이브러리를 활용하여 독서문화프로그램의 명칭, 일정, 대상 연령층, 참여 인원 등 관련 정보를 체계적으로 수집하였다. 분석 대상으로는 서울특별시교육청 소속 23개 도서관 중 강남도서관, 강동도서관, 강서도서관, 개포도서관, 고덕평생학습관, 고척도서관, 구로도서관, 남산도서관 등 8개의 도서관을 선정하였다.

도서관 대출 데이터는 도서관 정보나루에서 제공하는 8장서 및 대출 데이터를 활용하였으며, 선정한 8개 도서관의 장서 보유 현황과 대출 데이터를 다운로드 받아 분석에 사용하였다.

2) 데이터 전처리

효과적인 분석을 위해 데이터 전처리를 진행하였다. 수집된 데이터는 프로그램명, 강의분류, 접수기간, 접수방법, 강의기간, 강의시간, 강의대상, 강의요일, 강의장소, 모집인원 등 총 10개의 칼럼으로 구성되었다. 먼저, 각 칼럼 별 데이터를 주요 범주로 그룹화하였다. 강의 대상은 어린이, 청소년, 성인, 어르신, 누구나, 가족의 6가지 그룹으로 재 분류하였고, 강의 장소는 대면, 비대면, 기타 세 가지 형태로 구분하였다. 그 다음, 각 데이터를 분석에 용이하도록 인코딩 작업을 실시하였다. 접수 기간, 강의 기간, 강의 시간 등의 시간 관련 데이터는 분석의 일관성을 유지하기 위해 시간 및 분 단위로 변환되었으며, 데이터의 완전성을 확보하기 위해 각 칼럼별 결측치는 해당 칼럼의 최빈값으로 대체하였다. 이후 독서문화프로그램의 강의 기간과 요일 정보를 바탕으로, 프로그램이 개최된 날짜를 계산하여 월별로 데이터를 그룹화하였다.

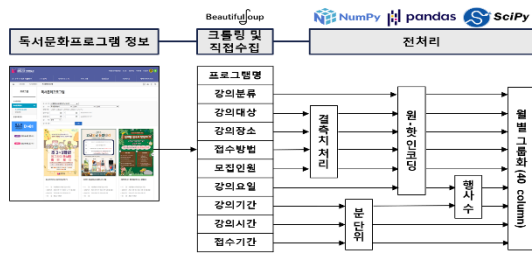


그림1. 독서문화프로그램 데이터 전처리 과정

도서관 정보나루의 장서 및 대출 데이터는 월별로 제공되며, 각 도서관의 대출 누적 건수를 포함하고 있다. 대출 데이터가 누적으로 제공되었기 때문에, 월별 대출 건수를 새로 산출하였다. 이 과정에서 주제분류번호 칼럼을 이용해 도서관을 100개의 대분류로 나누고, 대분류, 도서관수, 대출건수를 활용하여 도서관수 대비 대출건수를 계산하여 분석에 활용하였다. 대출건수 값이 음수이거나 이상치($Q3 + 3 * IQR$)로 확인된 경우, 같은 도서관 내 특정 기간에 이상치가 집중되어 있어 이를 데이터 집계 오류로 판단하고, 0보다 작은 값과 과도하게 큰 값을 중앙값으로 대체하였다.

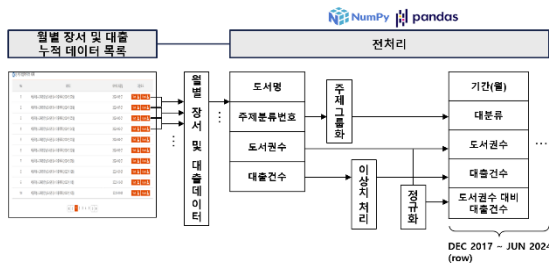


그림2. 월별 장서 및 대출데이터 전처리 과정

3) 분석방법론

분석 알고리즘으로 Pearson 상관계수, Kendall's tau를 활용하였다 3).

Pearson 상관계수는 두 변수 간의 선형성이 얼마나 강한지 측정하기 위해 사용된다. 한 변수의 변화가 다른 변수의 변화에 비례적으로 연관되어 있을 경우 선형 관계가 있다. 두 변수 간에 선형성이 강하다는 것은 변수들의 관계가 직선에 의해 잘 모델링 된다는 뜻이다.

Kendall's tau는 단조성을 평가하기 위해 사용된다. 변수는 각각 일치하는 쌍들의 수와 일치하지 않는 쌍들의 수를 나타낸다. 일치하는 쌍들의 수가 많을수록 Kendall's tau는 1에 가까워질 것이다. Kendall's tau도 절댓값이 1에 가까울수록 단조 상관성을 가지고 있고, 0에 가까울수록 단조 상관성이 없는 것이다.

4) 독서문화프로그램이 도서관 이용에 미치는 영향

우선 8개의 도서관에서 시행된 프로그램의 전반적인

요소들을 평균값과 비율을 중심으로 검토하였다. 접수 기간의 평균은 412시간이었으며, 강의 시간은 평균 143.7분으로 나타났다. 프로그램 당 모집 인원은 평균 24.4명이었다. 접수 방법은 에버러닝이 62.2%로 다른 접수 방법보다 훨씬 높은 비율을 차지하였다. 강의 대상은 어린이가 전체 대상의 29%로 가장 큰 비율을 차지했고, 그 뒤를 이어 성인과 가족이 주요 대상으로 나타났다. 또한, 토요일에 진행된 행사가 21.1%로 가장 많았다. 월별 행사 수를 살펴보면, 8월, 9월, 10월에 행사가 집중적으로 시행된 반면, 월초와 월말에는 행사 시행 횟수가 상대적으로 적었다. 마지막으로, 전체 행사 중 대면 행사가 68.8%로 가장 많이 시행된 것으로 나타났다. 이는 비대면 행사나 기타 방법을 통한 행사보다 월등히 높은 비율이다.

이후, 월 별 각 도서관에서 진행한 독서문화프로그램의 수와 도서관 이용 사이의 상관관계를 분석하였다. 도서관 이용 빈도는 대출건수와, 도서관수의 증감이 결과에 영향을 미치는 것을 방지하고자, 대출건수 비율(보유 도서관수 대비 대출건수)까지 두 가지 기준을 활용하여 판단하였다. 분석 결과 고척도서관 외 나머지 7개의 도서관에 대해서는 유의미한 상관관계가 나타나지 않았다.

표1. 독서문화프로그램 수와 도서관 이용 사이의 상관관계

| 도서관 명 | 대출건수 | 대출건수 비율 |
|---------|--------|---------|
| 고척도서관 | 0.828 | 0.831 |
| 강동도서관 | -0.092 | -0.149 |
| 고덕평생학습관 | -0.297 | -0.274 |
| 강남도서관 | -0.207 | -0.178 |
| 개포도서관 | 0.236 | 0.193 |
| 구로도서관 | 0.127 | 0.097 |
| 강서도서관 | 0.247 | 0.237 |
| 남산도서관 | 0.210 | 0.121 |

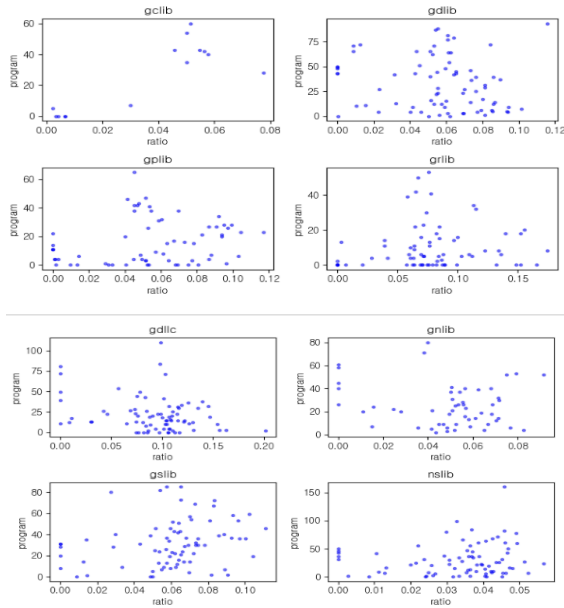


그림3. 도서관 별 도서관 대출건수 비율(ratio)와 프로그램 시행 횟수(program) 사이의 산점도

상관분석 결과 고척도서관이 타 도서관에 비해 독서문화프로그램 수와 도서관 대출건수와의 상관계수는 0.828, 대출건수 비율과의 상관계수는 0.831로 매우 높은 수치로 나타났다. 고척도서관은 타 도서관과 비교했을 때 독서문화프로그램에서의 특이한 특징을 보였는데, 고척도서관의 평균 강의시간은 264.5분으로 전체 8개 도서관의 평균인 143.7분보다 훨씬 높은 수치를 기록했다. 반면, 독서문화프로그램의 모집인원은 평균 15.6명으로, 8개 도서관 중 가장 낮은 수치를 나타냈다.

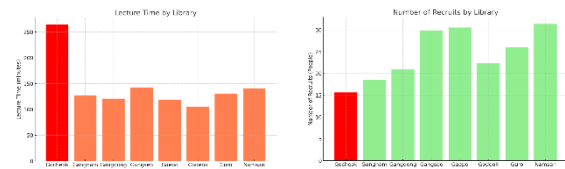


그림4. 도서관 별 시행된 독서문화프로그램의 강의시간과 평균 모집인원 (빨간색-고척도서관)

또한, 독서문화프로그램이 도서 대출건수에 영향을 미친다면, 전반적인 도서 대출건수보다 프로그램과 관련된 특정 분류의 도서 이용에 영향을 줄 가능성이 높다고 판단하였다. 서울특별시교육청 통합도서관 자료 분류를 참고하여 전체 도서를 100개의 카테고리 나눈 뒤 도서관의 월 별 독서문화프로그램 시행 횟수가 도서 분류별 대출건수, 대출건수 비율에 미치는 영향을 분석하고자 하였고, 0.3 이상의 유의미한 상관관계를 가지는 카테고리 수를 도서관 별로 정리하였다. 고척도서관의 경우 88개, 개포도서관은 25개, 강서도서관은 8개의 카테고리가 독서문화프로그램 시행

횟수와 유의미한 상관관계를 보임을 확인했다.

표2. 독서문화프로그램 수와 유의미한 상관관계를 보이는 도서 대출 카테고리 개수

| 도서관 명 | 대출건수 | 대출건수 비율 |
|---------|------|---------|
| 고척도서관 | 88 | 88 |
| 강동도서관 | 0 | 0 |
| 고덕평생학습관 | 0 | 0 |
| 강남도서관 | 1 | 1 |
| 개포도서관 | 25 | 13 |
| 구로도서관 | 3 | 1 |
| 강서도서관 | 8 | 7 |
| 남산도서관 | 2 | 2 |

여기서 도서관 별 유의미한 상관관계를 가지는 카테고리 수의 차이가 어떤 요인으로부터 발생하는지 알아보고자 하였다. 각 도서관 별 독서문화프로그램의 접수기간, 강의시간, 모집인원의 중앙값(median)과 유의미한 상관관계를 가지는 도서 카테고리 수 사이의 상관관계 분석을 진행하였다. 표본이 작아 Pearson 상관계수 뿐만 아니라 Kendall's tau 또한 활용하였다. 분석 결과, 두 경우 모두 접수기간과는 음의 상관관계를 보였고, 모집인원과는 유의미한 상관관계를 보이지 않았다. 또한, 강의시간과는 양의 상관관계를 보였다.

표3. 각 도서관 별 유의미한 상관관계를 보이는 카테고리 수와 독서문화프로그램 요인 사이의 상관관계

| 요인(median) | Pearson | Kendall |
|------------|---------|---------|
| 접수시간 | -0.32 | -0.44 |
| 강의시간 | 0.56 | 0.37 |
| 모집인원 | -0.01 | 0.22 |

3. 결론

본 연구를 통해 독서문화프로그램의 시행이 도서관 이용 빈도와 유의미한 상관관계가 있는지를 분석하고자 하였다. 초기 분석에서는 프로그램 운영 수와 대출건수 간에 유의미한 상관관계가 나타나지 않았으나, 대분류를 기준으로 상관관계를 도출해본 결과, 상당 수의 도서관

대분류 도서 대출건수와 프로그램 운영 수 간에 유의미한 관계를 확인할 수 있었다. 상관관계를 보이는 카테고리의 수가 많은 도서관에서 진행된 프로그램에 어떠한 특징이 있는지 추가 분석한 결과는 다음과 같다. 첫째, 접수기간은 프로그램과 유의미한 상관관계를 가지는 도서 분류 수와 음의 상관관계를 가졌다. 이는 상시형 프로그램보다 한정된 기간 내에 모집하는 프로그램이 도서관 이용 촉진에 더 도움이 될 수 있음을 시사한다. 둘째, 강의 지속시간은 프로그램과 유의미한 상관관계를 가지는 카테고리 수와 양의 상관관계를 가졌다. 이는 장시간의 강의가 도서관 이용에 긍정적인 영향을 줄 수 있음을 의미한다. 마지막으로, 모집인원과는 유의미한 상관관계가 발견되지 않았다.

연구팀은 이 결과를 독서문화프로그램이 학부모들의 교육열과 연관될 수 있다고 해석하였다. 8개 도서관은 어린이를 대상으로 한 독서문화프로그램이 다수였으며, 이는 학부모들의 관심과 참여를 유도하는 데 중요한 역할을 했을 가능성이 크다고 볼 수 있다. 접수 기간이 짧고 한정된 기간 동안 운영된 프로그램은 특히 저학년 자녀를 둔 학부모들에게 효과적으로 작용했을 것으로 보이며, 자녀의 교육에 대한 즉각적인 반응을 이끌어내는 동시에 학부모들이 자녀 교육에 적극적으로 참여하도록 자극했을 가능성이 있다. 또한, 프로그램 시간이 길어질수록 충분한 학습과 토론 활동으로 자녀의 독서 습관 형성에 긍정적인 영향을 미칠 수 있다. 고척도서관의 사례는 이러한 해석을 뒷받침하는데, 프로그램의 규모가 작지만 집중적으로 운영된 결과 학부모들의 기대에 더 부합했을 가능성을 보여준다.

그러나 본 연구는 몇 가지 한계점을 가진다. 첫째, 독서문화프로그램 데이터를 수집할 때, 크롤링을 시도하였으나 포스터에만 명시된 몇몇 요소들은 수동으로 데이터를 수집해야 했다. 이러한 데이터 수집의 어려움으로 인해 분석에 사용된 도서관의 수가 제한적이었고, 결과의 유의성을 저하시켰을 가능성이 있다. 둘째, 도서관 이용에 관한 별도의 수치화된 지표가 존재하지 않아, 대출건수만으로 판단하였다. 향후 연구에서는 분석 범위를 확대하여 서울특별시교육청 도서관뿐만 아니라 타 지역의 도서관을 포함해 결과의 유의성을 높일 수 있을 것이다. 또한, 도서관 이용 지표로 대출건수가 아닌 순수 이용 건수를 활용할 수 있다면 더욱 정확한 상관관계가 도출될 수 있을 것으로 기대된다.

참고 문헌

- 1) 황금숙, 김수경, 박미영, “공공도서관 문화프로그램 현황 분석과 활성화 방안”, 한국도서관·정보학회지, 2008
- 2) 이해운, 이지연, “공공도서관 교육문화프로그램 참여와 도서관 이용의 관계 연구”, 정보관리학회지, 2014
- 3) 김재호, 김장영, “코로나19와 서울 소상공인 상권의 상관관계 분석”, 2021