

Background

This is a plan for a data mining project aiming to gain insights on how Ontario libraries can be successful in the 21st century.

In this project I will be looking into a dataset from the Ontario [Data Website](#).

The data is self-reported data from approximately 380 public libraries, First Nation public libraries and contracting organizations.

Looking through the dataset, I found out it has extensive data from 2011 to 2022, the data includes about 317 columns that describes key information about the Ontario Libraries, mostly I will be looking only at the columns describes the following:

- Library General and demographic Information
- Library Holding Information (English And French, and Other)

Problem statement

To enhance the success of Ontario libraries in the 21st century, this data mining project aims to extract insights about the representation of English, French, and Other materials in libraries.

By identifying the patterns, the analysis will inform strategic decisions, and help with ensuring inclusive and culturally relevant library services for Ontario's diverse communities.

Business Objective

The objective of this project is to dive deeply into the dataset, specifically focusing on the representation of English, French, and other materials in Ontario libraries.

By examining the patterns and trends related to language-specific holdings, my aim is to provide actionable insights that empower Ontario libraries to make informed decisions and be successful in the 21st century.

Through this analysis, I strive to enhance the libraries' ability to cater to the diverse linguistic needs of their communities, fostering a more inclusive and accessible library experience for all patrons.

Success Criteria

- To me, a successful library in the 21st century is a library that beholds a rich amount of varieties of materials in both English, French, and other languages, taking in consideration the demographic diversity of their communities.

- By promoting linguistic diversity and cultural inclusivity, libraries will strengthen community bonds, encourage cross-cultural understanding, and ultimately enrich the social fabric of the region.
- To achieve this, the analysis will dive deep in the data set to find the language representation of the hold materials in Ontario Libraries and give insights on the current representation to help for good decision making in which libraries need more diverse or centric language representation.

Research question / Hypothesis

- English materials are expected to be predominant, given the widespread use of English in Ontario.
- French materials might be more prominent in certain regions with a significant Francophone population.
- Other languages could vary based on the diversity of local communities, such as Mississauga, Brampton, and Markham.
- The analysis will explore the balance between digital and physical materials, aiming to optimize this mix to accommodate diverse reading preferences within the community.

Data selection:

For a nuanced analysis of language representation in Ontario libraries, strategic selection of pertinent columns is paramount. The chosen columns from year 2019 to year 2022, including but not limited to:

Demographic data: (N for numerical, C for Categorical)

- Survey Year From [C]
- Library Full Name [C]
- Library Number [C]
- A1.3 Ontario Library Service (OLS) Region [C]
- A1.4 Type of Library Service [C]
- A1.5 Mailing Address [C]
- A1.9 Street Address [C]
- A1.10 City/Town [C]
- A1.11 Province [C]
- A1.12 Postal Code [C]
- A1.13 Web Site Address [C]
- A1.14 No. of Active Library Cardholders [N]

Materials in hold related data: (N for numerical, C for Categorical)

- Print Titles [N]
- Print Volumes [N]
- CDs and DVDs Titles [N]
- CDs and DVDs Copies [N]

- E-books Titles [N]
- E-book Copies [N]
- Audio Book Titles [N]
- Audio Book Copies [N]
- Special Collections - Original Formats [N]
- Print Periodical Titles [N]
- Databases and Database Subscriptions [N]

Assumptions:

- **Data Accuracy:** This analysis assumes that the data provided is accurate and has undergone necessary validation processes, ensuring reliability in the representation of library holdings.
- **Data Consistency:** This analysis assumes consistent data formatting and categorization across all libraries, allowing for uniform analysis and comparison.
- **Data Completeness:** This analysis assumes that the dataset includes comprehensive information about English, French, and other language materials, enabling a holistic overview of library holdings.
- **Data Currency:** This analysis assumes that the data is up-to-date and reflective of the current state of library collections, considering recent acquisitions and removals.
- **Data Integrity:** This analysis assumes that there is no absence of significant data entry errors or biases that could skew the analysis results.

Analysis plan:

Following Crisp-DM, the analysis plan for the language representation in ontario libraries goes through the following steps:

1. Data Understanding:

- **Data Collection:** Using R, I will take a subset of the dataset, containing only the columns related to library holdings, focusing on English, French, and other language materials.
- **Data Exploration:** I will explore the subset dataset to understand the structure, identify columns, data types, and assess data quality.

2. Data Preparation:

- **Data Cleaning:** Using R, I will apply data cleaning techniques like removing rows with missing values or imputing missing values using appropriate techniques will be used to handle missing values
- Using R, I will detect the outliers if they exist and handle them.

- **Data Transformation:** Using R, I will apply some techniques to transform raw data into usable formats, standardize columns, and create derived variables if necessary.

3. **Exploratory Data Analysis:**

Using R, I will apply some techniques to understand more about the data, This might include the following:

- **Descriptive Analysis:** Calculate basic statistics for English, French, and other languages, such as mean, median, and mode, to understand the distribution of materials.
- **Comparative Analysis:** Compare the representation of English and French materials across various categories, including print, digital (CD and DVD, Audio, E-books), and special collections.
- **Correlation Analysis:** Investigate correlations between language representation and the demographic area.

4. **Evaluation:**

- I will be evaluating the analysis results against the predefined objective.

5. **Deployment and Presentation:**

- I will create visualizations (charts, graphs) to present the findings clearly.
- I will prepare a detailed report summarizing the analysis, insights gained, and recommendations for libraries to improve language inclusivity.

6. **Iteration and Feedback:**

- If possible, I will gather feedback from stakeholders and domain experts.
- If needed, I will iterate on the analysis based on feedback, refining the approach and conducting additional analyses if required.

7. **Conclusion and Recommendations:**

- In this step, I will summarize the key findings, including insights into language representation of the behold materials.
- Depending on what I understand from the findings, I will provide actionable recommendations for libraries to enhance language inclusivity, and to optimize their collections based on community demographics.

8. **Documentation:**

- During the analysis, I will document the analysis process, the methodologies employed, and the tools used for future reference and reproducibility.

Benefits of the analysis:

- The analysis of language representation in Ontario libraries holds significant benefits for both the libraries and the communities they serve.
- By understanding how English, French, and other languages are represented in libraries' collections, libraries can tailor their acquisitions to better align with the linguistic diversity of their communities.

- This targeted approach ensures that library resources are inclusive and accessible to a wider audience, fostering a sense of belonging and cultural representation.
- Additionally, insights derived from this analysis can inform strategic decisions, enabling libraries to allocate resources more effectively, enhance community engagement, and promote cultural exchange.
- Ultimately, this analysis empowers libraries to fulfill their role as cultural hubs, promoting literacy, understanding, and unity among diverse linguistic groups.