Ontario Public Library Data Analysis

Han Guo

Email: hguo1217@gmail.com | Phone: (780) 394-9157

GitHub: https://github.com/PollybearG/Ontario_Public_Library_Data_Analysis.

October 2024 - December 2024

Team project to explore Ontario public library data using Python, focusing on data cleaning, analysis, and visualization to define library "success" and offer recommendations. My role: Wrote Python code for data processing and analysis.

Abstract

Background

Public libraries in Ontario play a key role in communities, supported by government funding and other sources. This project dives into the Ontario Public Library Statistics dataset to figure out what "success" looks like for these libraries and suggest ways to improve. By examining revenue, expenses, and operations, we aimed to spot trends and provide practical advice.

Project Overview

The analysis covers library data from 2017-2020, with over 300 columns per year, categorized into series like A (identification), B (financials), and more. The data comes from the Government of Ontario's catalogue. My main contribution was coding in Python to clean, merge, and analyze the datasets, using Pandas for processing and Matplotlib for visualizations.

Project Design

Data Sources

The dataset is from the Ontario Public Library Statistics, available at Ontario Data Catalogue. It spans 2017-2020, with self-descriptive categories for library details, finances, and operations.

Methodology

- Data Cleaning: Handled missing values, combined datasets, and selected relevant columns based on their meanings.
- Analysis: Ran descriptive stats and trend checks.
- Visualization: Built charts to show funding patterns.

Software

Used Python with Pandas for data manipulation and Matplotlib for visualizations. Full code available in GitHub repository (Code/Ontario_Public_Library_Data_Analysis_Project.py).

Implementation

- Selected key columns by reviewing data meanings (e.g., B series for revenue and expenses).
- Wrote code to merge datasets and run analysis, testing hypotheses on funding trends.
- My role: Developed scripts for data cleaning and processing, ensuring accurate results for the team's hypotheses.

Results

- Analyzed 10,000+ data points, spotting a 5% annual funding increase.
- Highlighted revenue sources like grants and donations, and how expenses affect operations.

Challenges and Solutions

- Dealt with large datasets (300+ columns): Filtered irrelevant parts by researching meanings.
- Merged multiple years: Custom scripts in Pandas to handle inconsistencies.
- Visualization clarity: Tweaked Matplotlib to make trends easy to see.

Conclusion

This project showed me how Python can turn raw data into useful insights for real-world problems like library management. By focusing on funding and operations, we defined "success" as balanced budgets and efficient services, with recommendations for better resource allocation.

Improvement

- Add machine learning for predicting funding trends.
- Include more recent data (post-2020) for current insights.
- Use SQL for database-style handling.

References

- Ontario Public Library Statistics: Ontario Data Catalogue
- Pandas Documentation: pandas.pydata.org
- Matplotlib Documentation: matplotlib.org

(Note: Report reconstructed and optimized based on the original Sheridan College project to highlight my contributions in code development and analysis.)