Project Report

By: Timothy Mok and Aron Saengchan

Summary

This project focuses on a data set of statistics for the Ontario Public Library System from 2017 to 2019, made available by the Government of Canada [1]. Approximately 380 libraries self-report the data, providing general information such as the library names, codes, and addresses, as well as specific statistics on the number of cardholders and the number of English and French resources available at each branch. The proposed project incorporates these data sets into an interactive portal for users to use as part of the Ontario Public Library System.

The program opens to a **Main Menu** where the user has four options to choose from:

1. **Branch Information Search –** The user is prompted to enter a library name or code and the program will provide them specific information, including the address, website, and number of print/electronic resources available of that branch.
2. **Library Locator** – The user is prompted to enter a postal code and searches for any nearby libraries, while also considering the user's needs. If nearby branches are found, the program prints a list of them and allows the user to obtain information about a specific branch.
3. **Access Yearly Archives** – The user is prompted to enter a year between 2017-2019 and the program displays various statistical information from that year, from statistics of all the libraries to specific records of each column in the data set.
4. **Quit –** Exits the program.

After each action, the user can perform the previous action again or return to the main menu.

Program Specifications

To address the given specifications, the table below describes how each requirement was achieved in this project.

Specifications		Line	Description
Dataset Selection	Importing Data	359	Data sets were imported from the three spreadsheets located in the *Ontario Public Library Datasets* directory
DataFrame Creation	Merge/Join Operation	364	*merge()* was used twice to combined the three data sets together
	Hierarchical Index	371	Indices were sorted by *'Library Full Name' → 'Library Code' → 'Year'*
User Entry	User Selection	53	Three options are listed in the *Main Menu* (see *Summary* section)
Analysis and Calculations	Describe Method	324	*describe()* was used to print some general statistics in the *Yearly Archives*
	Adding Two Columns	387	'Total Resources' and 'Resources per Cardholder' columns were added to the data set
	Aggregation	318 342	*sum()* was used to aggregate a row in the described data and sliced data for plotting; *max()* was used to find library records in the *Yearly Archives*
	Masking Operation	202	Masking was used to filter postal codes that matched the user's input
	Groupby Operation	406	*groupby()* was used to calculate the sum by year for the described data and also used to help construct the plot from Matplotlib
	Pivot Table	329	Pivot table was used to compare the *'Average Resources per Cardholder'* by *'Service Region'* and *'Service Type'* in the *Yearly Archives*
	Functions/Methods	34	Three user-defined functions and seven class methods were created
Export and Matplotlib	Exported Data	440	DataFrame was exported to *'exported_statistics.xlsx'* in the directory
	Create Plot	414	Plot was generated on Matplotlib to illustrate the yearly comparison of print and e-resources by language

References

[1] *Ontario public library statistics,* Government of Canada, June 2021. [Online]. Available: https://open.canada.ca/data/en/dataset/363fff31-6a07-41eb-9922-e9b64192b08b