ENSF 592

Bill Yu

30045030

billyusfg@gmail.com

Final Project

**Dataset:**

The chosen dataset contains statistics and information regarding all public libraries within Ontario. The three excel files are for years: 2017, 2018 and 2019. There is a difference of a few libraries for each year, but most of the libraries contained in the data remains the same. The majority of information contained in the datasets regards the number of titles held, and number of cardholders.

**User I/O:**

The user is only required to enter in two pieces of information. The library name and the year. The user input for name must match a name contained in the ‘Library Full Name’ index. Otherwise, the user will be prompted to enter a valid name again. Same goes for year.

The various functions contained in the code are designed to output various tables and manipulate the data using functions which fulfill the requirements set out in the rubric.

**Solution for Requirements:**

Please see square points for description and line reference in code.

* Stage 1:
  + Several suggested datasets are included in the project repository. You may use the provided data or select datasets of your own choosing.
    - Used a suggested dataset
  + You must use at least three separate Excel sheets or files that can be related in some way.
    - Yes
  + Your final combined dataset (see next stage) must have at least ten columns and 200 rows.
    - Final combined dataset: 1129 Rows, 19 Columns
  + You may edit the given datasets before you begin coding, but your program should not modify the Excel files directly.
    - Yes
  + You may not hard-code/copy-paste any information into your program except for the Excel column names.
    - Yes
* Stage 2:
  + Import your chosen data into a Pandas DataFrames.
    - Yes
  + You must use at least two merge/join operations and you must delete any duplicated columns/rows that result from the merge.
    - Merged with no duplications (Line 20 and 22)
  + You must create a hierarchical index of at least two levels (row or column).
    - Yes (Line 22)
  + All data should be presented in the correctly sorted order, depending on the index.
    - Yes (Line 22)
  + You may not use global variables. You must import the data within your main function.
    - Yes (Line 154)
  + Remember to check for null values or data mismatches.
    - Yes (Function at line 104)
* Stage 3:
  + Your application must return useful information. Design an interface that allows users to search based on some sort of criteria or keywords.
    - Yes
  + The user must provide at least two pieces of information/selection (e.g. "school name" and "grade")
    - Yes: Library name and year
  + Give the user clear input instructions. If an invalid entry is given, use try/except statements to handle the error. Your program should not terminate.
    - Yes (Loops at line 161 and 167)
  + You must not hard-code any data values (the data within your spreadsheets could be changed!).
    - Yes
  + Any output information must be clearly defined using printed headers.
    - Yes
* Stage 4:
  + You may choose what data trends to presents from your data. However, you must meet the following specifications.
  + Use the describe method to print aggregate stats for the entire dataset.
    - Yes (Line 37)
  + Add at least two columns to the combined dataset.
    - Yes (Line 99 and 100)
  + Use an aggregation computation for a subset of the data.
    - Yes (Line 40, 99 and 100)
  + Use a masking operation.
    - Yes (Line 118)
  + Use the groupby operation at least once.
    - Yes (Line 40)
  + Create and print a pivot table.
    - Yes (Line 134)
  + Include at least two user-defined functions or a class that contains two methods.
    - Yes
* Stage 5:
  + Export your entire merged, hierarchical dataset to an Excel file in the working directory. Be sure to include the index and header values. The TAs will use this to verify the structure of your dataset.
    - Yes
  + Use your data to create at least one plot using Matplotlib. Save the plot as a .png file and upload to the repository.
    - Yes

**Citation:**

2017 Ontario Public Library Statistics Open Data, Government of Canada, June 2021, [Online]. Available: <http://www.icpsr.umich.edu/icpsrweb/SAMHDA/studies/30122/version/2>

2018 Ontario Public Library Statistics Open Data, Government of Canada, June 2021, [Online]. Available: <http://www.icpsr.umich.edu/icpsrweb/SAMHDA/studies/30122/version/2>

2019 Ontario Public Library Statistics Open Data, Government of Canada, June 2021, [Online]. Available: <http://www.icpsr.umich.edu/icpsrweb/SAMHDA/studies/30122/version/2>

Government of Canada website was down, could not obtain more accurate source.