ENSF 592 Spring 2023
Project Report

June 13th, 2023

Tejpreet Bal

Soumini Mohandas

This program aims to determine the countries most affected by climate change and the rate at which they are being affected. Datasets of three key indicators of climate change were used to build the data frame of this program, which include:

- Annual Surface Temperature Change [1]
- Land Cover and Land Cover Altering Indicator (Carbon Sequestration) [2]
- Climate-related Disaster Frequency [3]

To ensure data consistency, only the common years for which the data was present were merged into the data frame. The index used to identify countries is the ISO 3 code which is what the user is required to enter. The full names of the countries can be long, and tedious to type, which is why ISO 3 is used. For instance, the country Madagascar, Rep. of can be entered as MDG.

The program starts by asking for two inputs:

1. The ISO 3 code of the first country.
2. The ISO 3 code of the second country to compare with.

If an invalid input is entered, the user is prompted to re-enter a valid input. The user may also enter Exit to terminate the program, which is handled using exit() function in the Sys module. After the selection, the means for each parameter are printed for the 29-year period for both countries. Please refer to Figure 1 in Appendix A to view the screenshot when Canada and USA were selected as an example.

Next, the user is asked to select a parameter from the menu for which a pivot table is printed, and a plot generated. Temperature change is selected for USA and Canada as a continuance of the previous example, the results can be seen in Appendix B.

The final two prompts allow the user to enter Y for the aggregate statistics for the entire climate dataset, and how many countries experienced more than 20 climate-related disasters in a year. If the user pleases to skip any of these prompts, any key can be entered. The results of entering Y are shown in Appendix C. After these two prompts, an Excel file named ClimateDataExport.xlsx is exported.

To view how the specifications are met in this program, please refer to Appendix D.

References

[1] Annual Surface Temperature Change, FAO Temperature Change, February 27, 2021. [Online]. Available: https://climatedata.imf.org/datasets/4063314923d74187be9596f10d034914/explore

[2] Land Cover Accounts, FAO Land Cover; IMF staff calculations, October 26, 2021. [Online]. Available: https://climatedata.imf.org/datasets/b1e6c0ea281f47b285addae0cbb28f4b/explore

[3] Climate-related Disasters Frequency, EM-DAT, CRED / UCLouvain, Brussels, Belgium, February 27, 2021. [Online]. Available: https://climatedata.imf.org/datasets/b13b69ee0dde43a99c811f592af4e821/explore

Appendix A – Initial Prompts

```
ENSF592 Climate Data Statistics
Influence of climate change on various indicators for countries across several years

Please enter the ISO3 code of the country (example: CAN, USA) or enter 'exit' to exit the program: CAN
Now pick a second country to compare with the first country.
Please enter the ISO3 code of the country (example: CAN, USA) or enter 'exit' to exit the program: usa

You chose the countries Canada and United States

The mean aggregate statistics for the 2 countries over a period of 29 years is shown below:
For Canada:
Temperature Change              1.158759
Land Cover Index               99.571262
Extreme Temperature                  NaN
Wildfires                            NaN
Storms                          1.400000
Landslides                           NaN
Droughts                             NaN
Floods                          1.500000
Total Disasters                 2.827586
Fractional Total Disaster       0.979499
Fractional Land Cover Index     0.513028

For United States:
Temperature Change              0.863345
Land Cover Index               99.168032
Extreme Temperature                  NaN
Wildfires                            NaN
Storms                         14.827586
Landslides                      1.000000
Droughts                        1.181818
Floods                          4.571429
Total Disasters                23.172414
Fractional Total Disaster       8.066846
Fractional Land Cover Index     0.510900
```

Figure 1: Selecting Canada and USA using ISO 3

Appendix B: Pivot Table and Figure

```
Please select the parameter or climate change indicator you want to compare from the list.
Please enter the code for the parameter you want to choose from the list below or enter 'exit' to exit the program
0: Temperature Change
1: Land Cover Index
2: Total Disasters
0
You picked the Temperature Change parameter to compare the two countries.

A pivot table showing the Temperature Change statistics for the 2 countries for the period from 1992 to 2020 is displayed below and a corresponding bar gr
aph is generated as well showing the trend:
Country  Canada  United States
Year
F1992     0.100          0.395
F1993     0.342          0.004
F1994     0.467          0.466
F1995     0.938          0.630
F1996    -0.125         -0.011
F1997     0.426          0.384
F1998     2.470          1.272
F1999     1.694          0.758
F2000     1.291          1.000
F2001     1.421          0.786
F2002     0.540          0.951
F2003     1.240          0.990
F2004     0.485          0.863
F2005     1.282          1.153
F2006     2.343          1.026
F2007     1.333          1.143
F2008     0.851          0.212
F2009     0.539          0.398
F2010     2.915          0.695
F2011     1.438          0.537
F2012     2.144          1.448
F2013     1.182          0.614
F2014     0.288          0.501
F2015     1.231          1.531
F2016     2.373          2.224
F2017     1.480          1.433
F2018     0.477          1.276
F2019     1.311          1.034
F2020     1.128          1.324
Plotting the data.
```

Figure 2: Pivot table showing yearly temperature change for Canada and USA

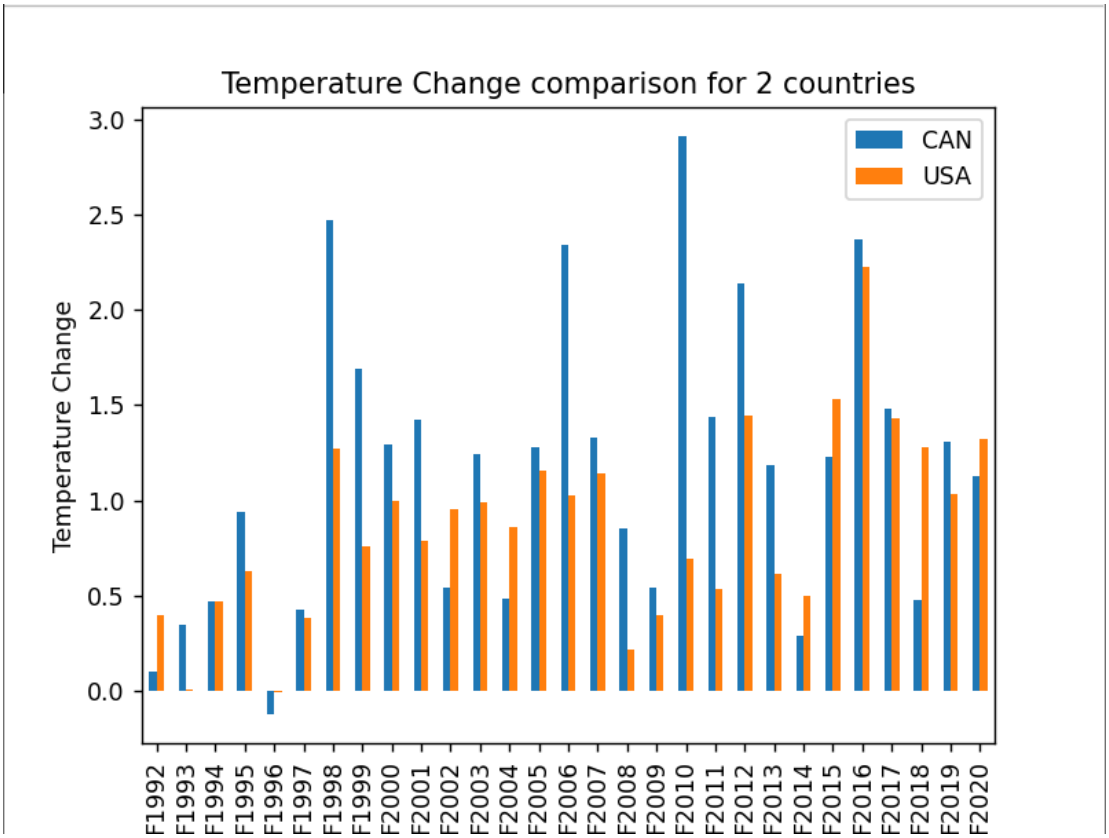


Figure 3: Plot generated using Matplotlib

Appendix C

```
Would you be interested in knowing the aggregate statistics for the entire climate data dataset?
Enter 'Y' for yes or enter any other key to skip this step and continue: Y

The aggregate stats for the entire Climate Data dataset is:
       Temperature Change  Land Cover Index  Extreme Temperature  ...  Total Disasters  Fractional Total Disaster  Fractional Land Cover Index
count                 0.0               0.0                  0.0  ...              0.0                        0.0                          0.0
mean                  NaN               NaN                  NaN  ...              NaN                        NaN                          NaN
std                   NaN               NaN                  NaN  ...              NaN                        NaN                          NaN
min                   NaN               NaN                  NaN  ...              NaN                        NaN                          NaN
25%                   NaN               NaN                  NaN  ...              NaN                        NaN                          NaN
50%                   NaN               NaN                  NaN  ...              NaN                        NaN                          NaN
75%                   NaN               NaN                  NaN  ...              NaN                        NaN                          NaN
max                   NaN               NaN                  NaN  ...              NaN                        NaN                          NaN

[8 rows x 11 columns]
```

Figure 4: Statistics of the entire climate data

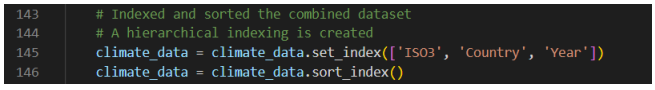
```
Would you be interested in knowing how many countries experienced more than 20 climate-related disasters in a year?
Enter 'Y' for yes or enter any other key to skip this step and continue: Y

The countries (along with their ISO3 codes and year) that experienced more than 20 climate-related disasters a year are:
('CHN', 'China, P.R.: Mainland', 'F2000')
('CHN', 'China, P.R.: Mainland', 'F2001')
('CHN', 'China, P.R.: Mainland', 'F2002')
('CHN', 'China, P.R.: Mainland', 'F2005')
('CHN', 'China, P.R.: Mainland', 'F2006')
('CHN', 'China, P.R.: Mainland', 'F2008')
('CHN', 'China, P.R.: Mainland', 'F2009')
('CHN', 'China, P.R.: Mainland', 'F2010')
('CHN', 'China, P.R.: Mainland', 'F2012')
('CHN', 'China, P.R.: Mainland', 'F2013')
('CHN', 'China, P.R.: Mainland', 'F2014')
('CHN', 'China, P.R.: Mainland', 'F2015')
('CHN', 'China, P.R.: Mainland', 'F2016')
('CHN', 'China, P.R.: Mainland', 'F2017')
('IDN', 'Indonesia', 'F2020')
('IND', 'India', 'F2005')
('IND', 'India', 'F2018')
('PHL', 'Philippines', 'F2009')
('PHL', 'Philippines', 'F2011')
('USA', 'United States', 'F1992')
('USA', 'United States', 'F1993')
('USA', 'United States', 'F1997')
('USA', 'United States', 'F1998')
('USA', 'United States', 'F1999')
('USA', 'United States', 'F2000')
('USA', 'United States', 'F2001')
('USA', 'United States', 'F2002')
('USA', 'United States', 'F2003')
('USA', 'United States', 'F2006')
('USA', 'United States', 'F2007')
('USA', 'United States', 'F2008')
('USA', 'United States', 'F2011')
('USA', 'United States', 'F2012')
('USA', 'United States', 'F2013')
('USA', 'United States', 'F2015')
('USA', 'United States', 'F2016')
('USA', 'United States', 'F2017')
('USA', 'United States', 'F2020')
Please wait. The program is exporting the dataframe to an excel file.
Exporting completed. Thank you for your patience.
```

Figure 5: More than 20 climate-related disasters in a year

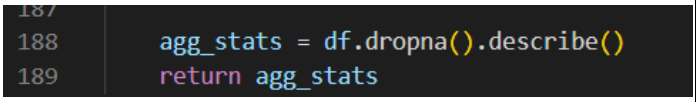
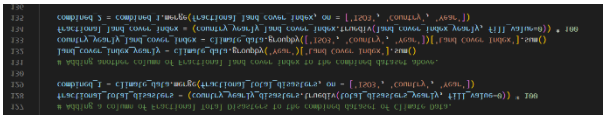
Appendix D – Specification Justification

Table 1: Table summarizing the compliance of the program with the given specifications

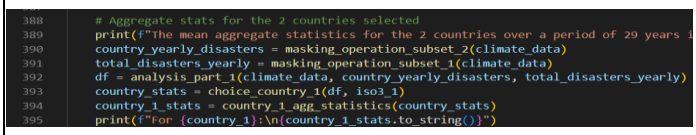
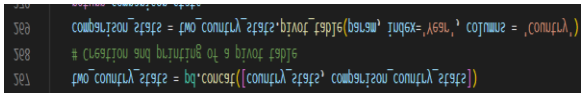
Stage 1: Dataset Selection ○ You must use at least three separate Excel sheets or files that can be related in some way. ○ Your final combined dataset (see next stage) must have at least ten columns and 200 rows. ○ You may edit the given datasets before you begin coding, but your program should not modify the Excel files directly. ○ You may not hard-code/copy-paste any information into your program except for the Excel column names.	• 3 Excel file names: Annual_Surface_Temperature_Change.csv, Climate-related_Disasters_Frequency.csv, Land_Cover_Accounts.csv • The data frame has: `[5771 rows x 12 columns]` • The Excel files were not modified and nothing was hard coded.
Stage 2: DataFrame Creation ○ Import your chosen data into a Pandas DataFrame. ○ You must use at least two merge/join operations and you must delete any duplicated columns/rows that result from the merge. ○ You must create a hierarchical index of at least two levels (row or column). ○ All data should be presented in the correctly sorted order, depending on the index. ○ You may not use global variables. You must import the data within your main function.	• After importing and formatting the data using Pandas, outer merge operations were done at line 73 in the climate_data.py file:  • Hierarchical index of two levels was created at line 144:  • All data was sorted according to the country name and the years ranging from 1992 to 2020. • Data was imported in the main function by calling the create_dataframe() function. • All null values were replaced with NaNs.

```
71    # All the above sub datasets of disaster are joined to create a dataset of disasters.
72    dis_1 = pd.merge(d_temp, d_wildfire, on = ['ISO3', 'Country', 'Year'], how = 'outer')
73    dis_2 = pd.merge(d_storm, d_landslide, on = ['ISO3', 'Country', 'Year'], how = 'outer')
74    dis_3 = pd.merge(d_drought, d_flood, on = ['ISO3', 'Country', 'Year'], how = 'outer')
75    dis_4 = pd.merge(dis_1, dis_2, on = ['ISO3', 'Country', 'Year'], how = 'outer')
76    dis_5 = pd.merge(dis_4, dis_3, on = ['ISO3', 'Country', 'Year'], how = 'outer')
77    disaster_data = pd.merge(dis_5, d_total, on = ['ISO3', 'Country', 'Year'], how = 'outer')
```

```
143    # Indexed and sorted the combined dataset
144    # A hierarchical indexing is created
145    climate_data = climate_data.set_index(['ISO3', 'Country', 'Year'])
146    climate_data = climate_data.sort_index()
```

○ Remember to check for null values or data mismatches.	
Stage 3: User Entry ○ Your application must return useful information. Design an interface that allows users to search based on some sort of criteria or keywords. ○ The user must provide at least two pieces of information/selection (e.g. “school name” and “grade”). ○ Give the user clear input instructions. If an invalid entry is given, use try/except statements to handle the error and continue to prompt for user input. ○ You must not hard-code any data values (the data within your spreadsheets could be changed!). ○ Any output information must be clearly defined using printed headers (DataFrame tables) or sentences (scalar values).	• The program returns data based on the user input for two different countries. It also returns the entire dataset is required. Please see Appendices A, B and C for the user entry stage. • For each input, try/except is used to print the appropriate message and allow for re-entry. • Nothing was hard coded, and the output is clearly defined.
Stage 4: Analysis and Calculations ○ You may choose what data trends to present from your data. However, you must meet the following specifications. ○ Use the describe method to print aggregate stats for the entire dataset. ○ Add at least two columns to the combined dataset.	• The describe method was used at line 188:  • Two columns were added at line 128 and below: 

```
188     agg_stats = df.dropna().describe()
189     return agg_stats
```

○ Use an aggregation computation for a subset of the data. ○ Use a masking operation. ○ Use the groupby operation at least once. ○ Create and print a pivot table. ○ Include at least two user-defined functions or a class that contains two methods.	• At line 218, aggregation computation for a subset of the data was done using .mean(). • Masking operation was done at line 390:  • Pivot table at line 269:  • The two user defined function are called when the user enters two ISO 3 codes for comparison
• Stage 5: Export and Matplotlib ○ At the conclusion of your program, 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 and your added columns. ○ Use your data to create at least one plot using Matplotlib. Save the plot as a .png file and upload to the repository.	• The Excel file is exported at the conclusion of the program and a plot can also be generated as shown Appendix B.