Matthew Hileman & Sydney Fowler

CS 3030.001

Final Project Report

15 December 2019

**PROJECT TITLE**

The purpose of this project was to create a command-line tool including various functions that could be used to work with Excel files. The idea behind this tool is to automate simple utilities commonly used with Excel files (sharing, compressing, importing, etc.), automate common cleanup procedures (standardizing dates, changing state names to state codes, truncating to a specified character limit, etc.), and eliminate human error in validation (email addresses, web addresses, etc.). The tool’s functions include:

* + - * Import: Converts a .csv to .xlsx.
      * Export: Converts a .xlsx to .csv.
      * Cleanup: Allows the user to select a set of cleanup/validation rules for each column in their file and applies said cleanup to a new version of the file. The cleanup rules include phone numbers, email addresses, state codes, zip codes, dates, web addresses, creating a list of unique entries, checking entries of a column against a list, character limits, and data types.
      * Compress: Allows the user to compress a file or folder.
      * Email: Emails a copy of an Excel file to a list of email addresses.
      * Duplicate Removal: Removes duplicate data from rows, columns, or entire sheets.
      * Analysis: Allows the user to select from the following analysis options for each column in their file: Sum, Count, Max, Min, Unique Values, and Average. For each sheet analyzed, a new analysis sheet is created within the workbook containing the results for each analyzed column within that sheet.

To run the tool:

* + - * Open your command line
      * Navigate to project folder
      * Run python3 main.py

This will bring up the main menu, allowing you to choose from the functionalities listed above. Once you have selected a function, the tool will provide you with instructions on how to use that function and the function’s requirements. You will then have the option to proceed with that function, return to the main menu, or quit the program. If you choose to proceed, you will be asked to provide the path of your Excel file, and potentially some additional information required for that function. For example, the Email function requires a text file of recipient email addresses, and an email address and password for the account you wish to use. On the other hand, For example, the data validation and cleanup tool would take a file as input, produce a list of headers, and allow the user to apply cleanup/validation rules to each column of data in the file.

the Cleanup function would take a file as input, produce a list of headers, and allow the user to apply cleanup/validation rules to each column of data in the file. Once the tool has received valid input for all required fields, the tool will perform the function and redirect you to the function’s menu, allowing you to repeat the function, return to the main menu, or quit the program.

To develop this tool, the project was divided by feature, resulting in 8 total tasks: main menu plus 7 features. The work was divided as follows:

| Matthew Hileman | Sydney Fowler |
| --- | --- |
| duplicate\_removal.py | analysis.py |
| file\_in.py | cleanup.py |
| file\_out.py | compress.py |
| main.py | share.py |

The remaining files were created as supporting files for the main features of the project and were developed by the team member in charge of the corresponding main feature:

* + - custom\_dictionaries.py
    - custom\_regular\_expressions.py
    - main\_info.py
    - menus.py
    - quit.py
    - exel\_funcs.py
    - quit.py

The code for this project can be accessed with the following GitHub link: <https://github.com/sydneyfowler/CS3030>