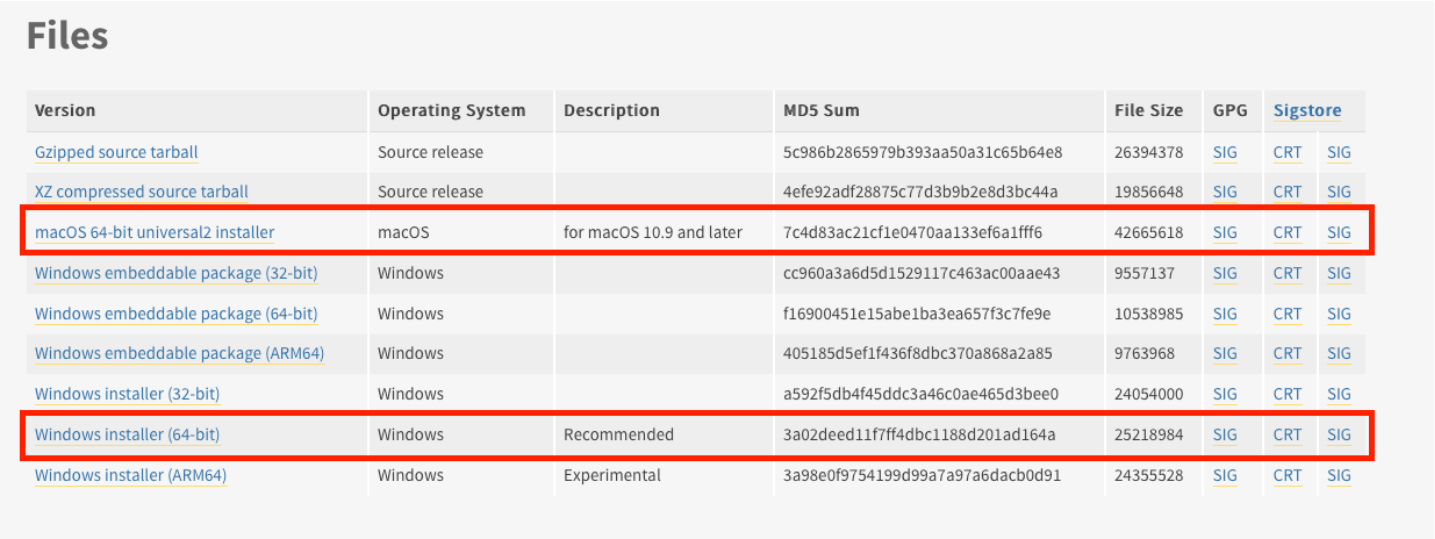
excel-condenser

An Excel condenser tool which takes a specific-format Excel file provided by Priority Group and condenses the nearly duplicate rows into a signle row with added columns.

Table of Contents

Installing Python

1. Go to https://wwww.python.org/downloads/release/python-3111/.
2. Scroll to the bottom of the page to the "Files" tables.
3. Select the correct installer for your computer architecture (most likely "Windows Installer (64-bit)" or "macOS 64-bit universal2 installer").



Files

Version	Operating System	Description	MD5 Sum	File Size	GPG	Sigstore	
Gzipped source tarball	Source release		5c986b2865979b393aa50a31c65b64e8	26394378	SIG	CRT	SIG
XZ compressed source tarball	Source release		4efe92adf28875c77d3b9b2e8d3bc44a	19856648	SIG	CRT	SIG
macOS 64-bit universal2 installer	macOS	for macOS 10.9 and later	7c4d83ac21cf1e0470aa133ef6a1fff6	42665618	SIG	CRT	SIG
Windows embeddable package (32-bit)	Windows		cc960a3a6d5d1529117c463ac00aae43	9557137	SIG	CRT	SIG
Windows embeddable package (64-bit)	Windows		f16900451e15abe1ba3ea657f3c7fe9e	10538985	SIG	CRT	SIG
Windows embeddable package (ARM64)	Windows		405185d5ef1f436f8dbc370a868a2a85	9763968	SIG	CRT	SIG
Windows installer (32-bit)	Windows		a592f5db4f45ddc3a46c0ae465d3bee0	24054000	SIG	CRT	SIG
Windows installer (64-bit)	Windows	Recommended	3a02deed11f7ff4dbc1188d201ad164a	25218984	SIG	CRT	SIG
Windows installer (ARM64)	Windows	Experimental	3a98e0f9754199d99a7a97a6dacb0d91	24355528	SIG	CRT	SIG

4. After the download is complete, run the .exe file you just downloaded (should be named

`python-3.11.1.exe` and can be double-clicked from wherever you saved it).

5. The installation wizard should be running, select option "Install Now". Your computer will prompt you for permission for the installer to modify your computer. Say yes. The installation should then begin on its own.

Setting up `excel-condenser` for the First Time

1. Download this repository as a .zip file by clicking the green "Code" button and select "Download ZIP".
2. Extract the files from the .zip into the directory that you'll run the program from.
    - This directory should contain files called `condenser.py` and `requirements.txt`.

On a Windows Computer

3. Open the Windows terminal (you can run the terminal by searching "cmd" in the search box in the taskbar and selecting "Command prompt").
4. Enter the following command where `[file path]` is replaced by the location of the directory with the program files. As a check, if you enter `dir` into the terminal at this point, the Python files and input Excel sheet should be listed in the output.

```
$ cd [file path]
```

5. Install the required Python packages using the following command:

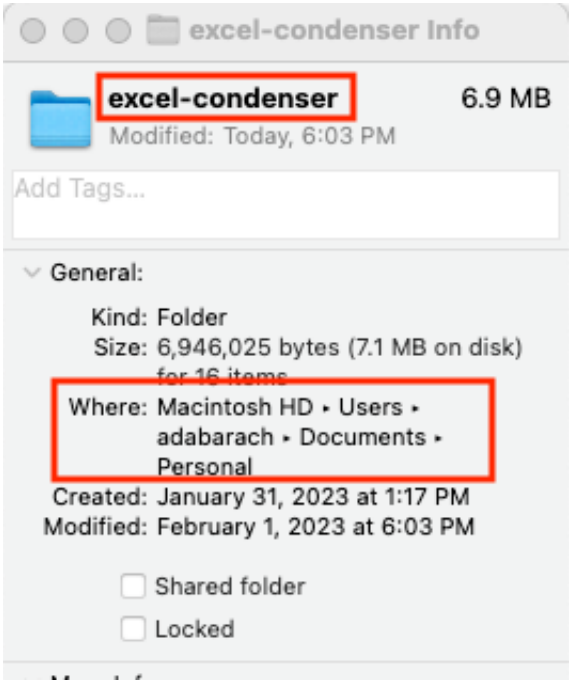
```
$ pip install -r requirements.txt
```

On a Mac Computer

3. Open the the Terminal app. This can be found in the Applications folder or by searching "terminal" in Launchpad.
4. Enter the following command where `[file path]` is replaced by the location of the directory with the program files. As a check, if you enter `ls` into the terminal at this point, the Python files and input Excel sheet should be listed in the output.
    - The file path to the directory can be found by right-clicking on the directory and

selecting "Get Info". Under "General" there will be a "Where" line containing the file path. You can ignore the "Macintosh HD" but need to add a '/' to the beginning of the path and then add '/' plus the folder name to the end. For example, the following screenshot translates to the path `/Users/adabarach/Documents/Personal/excel-condenser/`

```
$ cd [file path]
```



5. Install the required Python packages using the following command:

```
$ pip install -r requirements.txt
```

Folder Structure

- `condenser.py` - the Python program to run
- `README.md` - the User's guide, in Markdown
- `requirements.txt` - contains the required Python packages. These can be installed using the command `pip install -r requirements.txt`
- `UsersGuide.pdf` - this README as a PDF.

Running `excel-condenser`

1. Save the desired input Excel document into the same directory as above (which contains the `condenser.py` file). This file should contain only one worksheet. See the `data/data_test.xlsx` file for an example.
2. Open a terminal and enter `cd [file path]` where `[file path]` is replaced by the location of the `excel-condenser` folder.
3. Enter the following command into the terminal to run the program.

   ```
   $ python3 condenser.py --input_file [path1] --output_file [path2]
   ```

   where `[path1]` is replaced by the file path to the location of the input Excel file and `[path2]` is the file path at which you want the outputted Excel file to saved.
   - If you saved the input Excel file into the same directory as the Python file, then you can simply put the file name in place of `[path1]`.
   - If you want to save the outputted Excel file into the same directory as the Python file, then you can simply put the fiel name in place of `[path2]`.
4. The program should only take a few seconds to run.
5. The directory containing the Python file (or the directory you specified in `[path2]` above) should now contain a file with the name specified by `[path2]` above. This is the output from the Python program and can be moved or renamed to whatever you like. If you do not rename this file, it will be overwritten if the program is run again with the same `[path2]`.

Common Error Messages and Solutions

1. The below error indicates that the program cannot find the input Excel file. This likely occurred due to an incorrect file path (`[path1]` above).

```
adabarach@Adas-Air excel-condenser % python3 condenser.py --input_file data.xlsx --output_file data1.xlsx
Traceback (most recent call last):
  File "/Users/adabarach/Documents/Personal/excel-condenser/condenser.py", line 174, in <module>
    original_data = read_data_and_group(args.input_file)
  File "/Users/adabarach/Documents/Personal/excel-condenser/condenser.py", line 25, in read_data_and_group
    df = pd.read_excel(filepath)
  File "/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/pandas/util/_decorators.py"
, line 211, in wrapper
    return func(*args, **kwargs)
  File "/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/pandas/util/_decorators.py"
, line 317, in wrapper
    return func(*args, **kwargs)
  File "/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/pandas/io/excel/_base.py",
line 483, in read_excel
    io = ExcelFile(io, storage_options=storage_options, engine=engine)
  File "/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/pandas/io/excel/_base.py",
line 1629, in __init__
    ext = inspect_excel_format(
  File "/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/pandas/io/excel/_base.py",
line 1502, in inspect_excel_format
    with get_handle(
  File "/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/pandas/io/common.py", line
866, in get_handle
    handle = open(handle, ioargs.mode)
FileNotFoundError: [Errno 2] No such file or directory: 'data.xlsx'
adabarach@Adas-Air excel-condenser %
```

2. This error occurs if the requirements from the `requirements.txt` file were not installed properly. Repeat stepf f) from Setting up `excel-condenser` for the First Time.

```
adabarach@Adas-Air excel-condenser % python3 condenser.py --input_file data_test.xlsx --output_file data_modified.x
lsx
Traceback (most recent call last):
  File "/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/pandas/compat/_optional.py"
, line 142, in import_optional_dependency
    module = importlib.import_module(name)
  File "/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/importlib/__init__.py", line 126, in impo
rt_module
    return _bootstrap._gcd_import(name[level:], package, level)
  File "<frozen importlib._bootstrap>", line 1050, in _gcd_import
  File "<frozen importlib._bootstrap>", line 1027, in _find_and_load
  File "<frozen importlib._bootstrap>", line 1004, in _find_and_load_unlocked
ModuleNotFoundError: No module named 'openpyxl'

During handling of the above exception, another exception occurred:

Traceback (most recent call last):
  File "/Users/adabarach/Documents/Personal/excel-condenser/condenser.py", line 174, in <module>
    original_data = read_data_and_group(args.input_file)
  File "/Users/adabarach/Documents/Personal/excel-condenser/condenser.py", line 25, in read_data_and_group
    df = pd.read_excel(filepath)
  File "/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/pandas/util/_decorators.py"
, line 211, in wrapper
    return func(*args, **kwargs)
  File "/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/pandas/util/_decorators.py"
, line 317, in wrapper
    return func(*args, **kwargs)
  File "/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/pandas/io/excel/_base.py",
line 483, in read_excel
    io = ExcelFile(io, storage_options=storage_options, engine=engine)
  File "/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/pandas/io/excel/_base.py",
line 1672, in __init__
    self._reader = self._engines[engine](self._io, storage_options=storage_options)
  File "/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/pandas/io/excel/_openpyxl.p
y", line 548, in __init__
    import_optional_dependency("openpyxl")
  File "/Library/Frameworks/Python.framework/Versions/3.10/lib/python3.10/site-packages/pandas/compat/_optional.py"
, line 145, in import_optional_dependency
    raise ImportError(msg)
ImportError: Missing optional dependency 'openpyxl'  Use pip or conda to install openpyxl.
adabarach@Adas-Air excel-condenser %
```

   - Alternatively, you can manually install the requirement. The red box highlights the

package that needs to be installed. In this example, the package is `openpyxl`. To install the requirement, enter the following command into the terminal. The installation may prompt you to confirm the download by entering `y` for yes.

```
$ pip install openpyxl
```

Author

- Ada Barach