How to Run EMMA Data Wrangling Python Code

# Setup

1. Install Python
2. Install Pip
3. Open a console window in the src directory
4. Run, pip install -r requirements.txt, to get all python packages to run code
5. Run python data\_wrangling.py **with command line arguments**

# Command Line Arguments

EMMA data wrangling requires command line arguments to run completely.

|  |  |
| --- | --- |
| Command Line Argument | Definition |
| --interactions filename | The name of the interactions excel dataset. **This file must be in the data/input directory.** |
| --events filename | The name of the events excel dataset. **This file must be in the data/input directory.** |
| --debug [0 or 1] | If 1, then debugging and processing information will be printed to the console throughout the program runtime. |

Example:

python data\_wrangling.py --interactions "Cohort1\_6mo\_interactions\_0\_2022-04-03\_2022-05-

01.xlsx" --events "Cohort 1 Events\_6mo\_noPHI.xlsx" --debug 1

# Directories

|  |  |
| --- | --- |
| ./data | Contains input and output files |
| ./tests | Contains test scripts and test data. Do not use the data directory “./tests/test data” as input datasets. |
| ./src | Contains all python scripts to run data wrangling. Run python calls from this directory. |

# Demo

You can execute the file run\_demo.bat in the src file to see a test run of the command above.