**User Guide - Data Management and Software for**

**Low-Cost PM2.5 Sensors Can Help Identify Driving Factors of Poor Air Quality and Benefit Communities**

Tim Keyesa,b,[[1]](#footnote-1),[[2]](#footnote-2) , Rea Domingob, Samantha Dynowskic, Royal Gravesc, Martha Kleinc, Melissa Leonardc, John Pilgrimc, Alison Sanchiricoc, Kate Trinkausc

a Evergreen Business Analytics, LLC, U.S.A.

b Sacred Heart University, U.S.A.

c Sierra Club Connecticut, U.S.A.

The following describes all data and software files and provides a guide for users.

**Contents**:

|  |  |  |  |
| --- | --- | --- | --- |
| **Site** | **File** | **Purpose** | **Use** |
| All | DEEP modeling data 2021-1H2022 formatted.csv | Waterbury FEM Monitor data | SQL input (main table) |
| All | Towantic NG Consumption Data Formatted.csv | Natural Gas consumption data for Towantic Energy Ctr | SQL input |
| All | WB\_precip formatted.csv | Waterbury precipitation data | SQL input |
| All | NG history East query.csv | Eastern U.S. Natural Gas consumption data | SQL input |
| All | EPA CAMPD Modeling Data.csv | EPA Clean Air Markets Program Data for CPV Towantic Energy Center | SQL input |
| All | OA\_Oxford.csv | Oxford “Algonquin” compressor station data | SQL input |
| All | Oxford NOAA dedupe.csv | Waterbury-Oxford Airport NOAA meteorological data | SQL input |
| Gunntown | Purple\_gunntown\_202101-202206.csv | Purple Air sensor data | SQL input |
| Gunntown | Gunntown SQL.txt | SQL code for Gunntown file processing, in text format | SQL code |
| Gunntown | WB-GT outer join 2021-1H2022 v10.csv (Excel is preferred in Python) | Result of outer join of input files | SQL output / Python input |
| Long Meadow | Purple\_longmeadow\_202101-202206.csv | Purple Air sensor data | SQL input |
| Long Meadow | Long Meadow SQL.txt | SQL code for Long Meadow file processing, in text format | SQL code |
| Long Meadow | WB-LM outer join 2021-1H2022 v10.csv (Excel is preferred in Python) | Result of outer join of input files | SQL output / Python input |
| Lake Zoar | Purple\_woti\_202101-202206.csv | Purple Air sensor data | SQL input |
| Lake Zoar | Lake Zoar SQL.txt | SQL code for Lake Zoar file processing, in text format | SQL code |
| Lake Zoar | WB-LZ outer join 2021-1H2022 v10.csv (Excel is preferred in Python) | Result of outer join of input files | SQL output / Python input |
| All | AQ Analytics 10March2023 - Heliyon.ipynb | Python Jupyter code for processing SQL output | Python Jupyter code |
| All | Waterbury PurpleAir Calibration Database.zip | Zipped file of MS Access database; can be used for all data processing | MS Access db |

**Processing Steps**:

1. Locate and download all files from <https://github.com/TKKeyes/HELIYON-D-23-18430.git>
2. Save all files to a suitable computer location in preparation for running SQL
3. For each site (Gunntown, Long Meadow, Lake Zoar), run respective SQL code on input files as noted in the Contents section above
4. Compare resulting SQL output files for each site with “SQL output / Python input” files provided
5. Convert resulting SQL output files from CSV to XLSX if needed (Python works better with \*.xlsx)
6. Launch Python/Jupyter (download instructions at <https://www.anaconda.com/download/> - both Jupyter and Spyder interfaces are provided; code provided is Jupyter notebook)
7. Python/Jupyter code is run for each site separately; “uncomment” the appropriate line of code in the section, “Data Reading and Cleaning / Load the dataset”; make sure the files created in step 4) are located in a directory path recognized by Python (i.e., change the file path in the code)
8. Run each section of code sequentially to obtain results for exploratory data analysis, local calibration regressions and local explanatory regressions

If MS Access database can be used –

1. Open database and run the following queries:
   1. WB-GT outer join 2021-1H2022 – NEW
   2. WB-LM outer join 2021-1H2022 – NEW
   3. WB-LZ outer join 2021-1H2022 – NEW
2. For each query run, save the results to CSV or XLSX with file names conforming to files in the Content section above for “SQL output / Python input”

1. 1 Corresponding author. E-mail address: [Tim.Keyes@EGBANA.com](mailto:Tim.Keyes@EGBANA.com) (T. Keyes). [↑](#footnote-ref-1)
2. 2 Tim Keyes was a paid consultant, commissioned by Sierra Club Connecticut for this work. [↑](#footnote-ref-2)