paper05_data_aggregation

October 10, 2021

1 Data Aggregation Across Data Sources

We have 3 different sources of data:

1. Our sensor data: that has the Indoor Air Quality and Indoor Environmental Data.
2. SINAICA: Outdoor Air Quality Monitoring Data from the Government.
3. OpenWeatherData: Outdoor Environmental Data.

We need it to be available that data to the models we plan to train. In the following sections this process is detailed.

```
/home/jaa6766/.conda/envs/cuda/lib/python3.7/importlib/_bootstrap.py:219:
RuntimeWarning: numpy.ufunc size changed, may indicate binary incompatibility.
Expected 192 from C header, got 216 from PyObject
/home/jaa6766/.conda/envs/cuda/lib/python3.7/importlib/_bootstrap.py:219:
RuntimeWarning: numpy.ufunc size changed, may indicate binary incompatibility.
Expected 192 from C header, got 216 from PyObject
/home/jaa6766/.conda/envs/cuda/lib/python3.7/importlib/_bootstrap.py:219:
RuntimeWarning: numpy.ufunc size changed, may indicate binary incompatibility.
Expected 192 from C header, got 216 from PyObject
/home/jaa6766/.conda/envs/cuda/lib/python3.7/importlib/_bootstrap.py:219:
RuntimeWarning: numpy.ufunc size changed, may indicate binary incompatibility.
Expected 192 from C header, got 216 from PyObject
```

1.1 Indoor Data

```
                               temperature  pressure  humidity  gasResistance  \
datetime
2021-02-12 06:04:09.089621067        21.54    777.41     43.93         151328
2021-02-12 06:04:12.087778807        21.56    777.41     43.89         152702
2021-02-12 06:04:15.072475433        21.53    777.41     43.97         151328
2021-02-12 06:04:18.070170164        21.51    777.41     44.03         151464
2021-02-12 06:04:21.061994791        21.51    777.41     44.05         152425
…                                      …        …        …              …
2021-09-18 01:20:38.889113188        25.84    782.96     56.64         928867
2021-09-18 01:20:41.882042885        25.83    782.94     56.66         923130
2021-09-18 01:20:44.877856970        25.83    782.94     56.63         925034
2021-09-18 01:20:47.872255564        25.83    782.94     56.62         923130
2021-09-18 01:20:50.866486311        25.83    782.96     56.63         925034

                                 IAQ  iaqAccuracy  year  month  day  hour  \
datetime
2021-02-12 06:04:09.089621067   37.5            1  2021      2   12     6
2021-02-12 06:04:12.087778807   35.6            1  2021      2   12     6
2021-02-12 06:04:15.072475433   37.5            1  2021      2   12     6
2021-02-12 06:04:18.070170164   38.5            1  2021      2   12     6
2021-02-12 06:04:21.061994791   36.9            1  2021      2   12     6
…                               …          …    …     …   …    …
2021-09-18 01:20:38.889113188  130.8            1  2021      9   18     1
2021-09-18 01:20:41.882042885  131.5            1  2021      9   18     1
2021-09-18 01:20:44.877856970  131.3            1  2021      9   18     1
```

```
2021-09-18 01:20:47.872255564  131.9           1  2021      9   18     1
2021-09-18 01:20:50.866486311  131.6           1  2021      9   18     1

                              minute  second
datetime
2021-02-12 06:04:09.089621067       4       9
2021-02-12 06:04:12.087778807       4      12
2021-02-12 06:04:15.072475433       4      15
2021-02-12 06:04:18.070170164       4      18
2021-02-12 06:04:21.061994791       4      21
…                                 …     …
2021-09-18 01:20:38.889113188      20      38
2021-09-18 01:20:41.882042885      20      41
2021-09-18 01:20:44.877856970      20      44
2021-09-18 01:20:47.872255564      20      47
2021-09-18 01:20:50.866486311      20      50

[6285103 rows x 12 columns]
```

1.2 Outdoor Air Quality Data

```
                            NO        CO       NO2       NOx        O3  \
Fecha
2021-01-01 00:00:00  0.006000  1.000000  0.032000  0.036000  0.006000
2021-01-01 01:00:00  0.021000       NaN       NaN       NaN       NaN
2021-01-01 02:00:00  0.013000  1.100000  0.032000  0.039000  0.004000
2021-01-01 03:00:00  0.031000  1.200000  0.033000  0.043000  0.001000
2021-01-01 04:00:00  0.005000  1.200000  0.031000  0.039000  0.002000
…                       …       …       …       …       …
2021-10-04 00:00:00  0.008292  0.545833  0.019083  0.026875  0.015833
2021-10-05 00:00:00  0.010000  0.563158  0.019722  0.030500  0.012278
2021-10-06 00:00:00  0.007571  0.672222  0.026111  0.035611  0.011000
2021-10-07 00:00:00  0.011565  0.713636  0.028636  0.040318  0.017909
2021-10-08 00:00:00  0.023778  0.758824  0.029412  0.050588  0.017941

                           PM10      PM2.5       SO2  year  month  day  hour  \
Fecha
2021-01-01 00:00:00  31.000000  19.000000  0.003000  2021      1    1     0
2021-01-01 01:00:00        NaN        NaN       NaN  2021      1    1     1
2021-01-01 02:00:00  37.000000  24.000000  0.003000  2021      1    1     2
2021-01-01 03:00:00  49.000000  39.000000  0.003000  2021      1    1     3
2021-01-01 04:00:00  80.000000  65.000000  0.003000  2021      1    1     4
…                       …       …       …    …     …   …   …
2021-10-04 00:00:00  11.826087   7.913043  0.000750  2021     10    4     0
2021-10-05 00:00:00  11.090909   6.772727  0.000556  2021     10    5     0
2021-10-06 00:00:00  18.722222  11.833333  0.000111  2021     10    6     0
2021-10-07 00:00:00  26.772727  17.000000  0.001045  2021     10    7     0
2021-10-08 00:00:00  29.000000  17.705882  0.008176  2021     10    8     0

                     minute
Fecha
2021-01-01 00:00:00       0
2021-01-01 01:00:00       0
2021-01-01 02:00:00       0
2021-01-01 03:00:00       0
2021-01-01 04:00:00       0
…                       …
2021-10-04 00:00:00       0
2021-10-05 00:00:00       0
```

```
2021-10-06 00:00:00       0
2021-10-07 00:00:00       0
2021-10-08 00:00:00       0

[2352 rows x 13 columns]
```

1.3 Outdoor Weather Data

```
                     temperature  feels_like  pressure  humidity  wind_speed  \
dt
2021-02-12 07:00:00        13.87       12.46      1020        44        0.00
2021-02-12 08:00:00        12.81       11.37      1020        47        0.00
2021-02-12 09:00:00        10.83        9.35      1019        53        1.54
2021-02-12 10:00:00         6.40        3.51      1019        61        4.12
2021-02-12 11:00:00         6.23        6.23      1019        57        0.00
…                           …           …         …           …          …
2021-09-27 19:00:00        21.51       20.89      1006        45        0.89
2021-09-27 20:00:00        23.18       22.81      1005        48        0.45
2021-09-27 21:00:00        22.21       21.69      1025        46        6.17
2021-09-27 22:00:00        21.03       20.68      1004        57        0.45
2021-09-27 23:00:00        20.17       19.81      1004        60        5.66

                     wind_deg weather_main
dt
2021-02-12 07:00:00         0        Clear
2021-02-12 08:00:00         0        Clear
2021-02-12 09:00:00        60        Clear
2021-02-12 10:00:00        40        Clear
2021-02-12 11:00:00         0        Clear
…                         …           …
2021-09-27 19:00:00       139        Clear
2021-09-27 20:00:00       224         Rain
2021-09-27 21:00:00       220         Rain
2021-09-27 22:00:00       242         Rain
2021-09-27 23:00:00       140       Clouds

[5601 rows x 7 columns]
```

1.4 Merging the 3 Datasets: Indoor Data, Outdoor Air Quality Data, Outdoor Weather Data.

1.4.1 Merging Air Quality and Weather Data

```
                           NO        CO       NO2       NOx        O3  \
2021-02-12 07:00:00  0.244000  2.500000  0.035000  0.205000  0.002000
2021-02-12 08:00:00  0.146000  1.600000  0.030000  0.089000  0.004000
2021-02-12 09:00:00  0.099000  1.500000  0.039000  0.072000  0.012000
2021-02-12 10:00:00  0.024000  1.200000  0.030000  0.047000  0.025000
2021-02-12 11:00:00  0.009000  0.900000  0.016000  0.026000  0.033000
…                       …         …         …         …         …
2021-09-14 00:00:00  0.017000  0.716667  0.015125  0.024625  0.013333
2021-09-15 00:00:00  0.027458  0.954167  0.028167  0.048625  0.019375
2021-09-16 00:00:00  0.006875  0.883333  0.028000  0.034458  0.022792
2021-09-17 00:00:00  0.010250  0.947619  0.032700  0.042750  0.026650
2021-09-18 00:00:00  0.009174  0.765217  0.027826  0.039043  0.015304

                          PM10      PM2.5       SO2  temperature  feels_like  \
2021-02-12 07:00:00  57.000000  25.000000  0.005000        13.87       12.46
2021-02-12 08:00:00  67.000000  33.000000  0.003000        12.81       11.37
```

```
2021-02-12 09:00:00  50.000000  28.000000  0.002000        10.83        9.35
2021-02-12 10:00:00  40.000000  21.000000  0.002000         6.40        3.51
2021-02-12 11:00:00  33.000000  19.000000  0.001000         6.23        6.23
…                          …          …         …            …           …
2021-09-14 00:00:00   7.047619   4.500000  0.000125        15.85       14.95
2021-09-15 00:00:00  24.416667  17.333333  0.000875        17.95       17.08
2021-09-16 00:00:00  46.833333  40.041667  0.001542        18.45       17.63
2021-09-17 00:00:00  29.666667  24.875000  0.003350        18.34       17.69
2021-09-18 00:00:00  20.304348  16.391304  0.001304        17.61       17.85

                     pressure  humidity  wind_speed  wind_deg weather_main
2021-02-12 07:00:00    1020.0      44.0        0.00       0.0        Clear
2021-02-12 08:00:00    1020.0      47.0        0.00       0.0        Clear
2021-02-12 09:00:00    1019.0      53.0        1.54      60.0        Clear
2021-02-12 10:00:00    1019.0      61.0        4.12      40.0        Clear
2021-02-12 11:00:00    1019.0      57.0        0.00       0.0        Clear
…                         …         …           …          …           …
2021-09-14 00:00:00    1025.0      56.0        4.12     170.0         Rain
2021-09-15 00:00:00    1023.0      49.0        7.72     130.0       Clouds
2021-09-16 00:00:00    1022.0      49.0        6.17     130.0        Smoke
2021-09-17 00:00:00    1024.0      56.0        4.63     300.0         Rain
2021-09-18 00:00:00    1015.0      93.0        1.37     199.0         Rain

[1341 rows x 15 columns]

                 NO        CO       NO2       NOx        O3       PM10  \
2021-09-16  0.006875  0.883333  0.028000  0.034458  0.022792  46.833333
2021-09-17  0.010250  0.947619  0.032700  0.042750  0.026650  29.666667
2021-09-18  0.009174  0.765217  0.027826  0.039043  0.015304  20.304348

               PM2.5       SO2  temperature  feels_like  pressure  humidity  \
2021-09-16  40.041667  0.001542        18.45       17.63    1022.0      49.0
2021-09-17  24.875000  0.003350        18.34       17.69    1024.0      56.0
2021-09-18  16.391304  0.001304        17.61       17.85    1015.0      93.0

            wind_speed  wind_deg weather_main
2021-09-16        6.17     130.0        Smoke
2021-09-17        4.63     300.0         Rain
2021-09-18        1.37     199.0         Rain
```

1.4.2 Merging Indoor and Outdoor (Air Quality and Weather) Data

```
                            temperature  pressure  humidity  gasResistance  \
datetime
2021-09-16 23:59:42.445474625        25.86    780.44     57.00         877365
2021-09-16 23:59:45.439809322        25.85    780.40     57.04         874512
2021-09-16 23:59:48.434415102        25.85    780.42     57.07         877365
2021-09-16 23:59:51.428925753        25.84    780.42     57.07         872810
2021-09-16 23:59:54.423571348        25.84    780.42     57.07         872244
2021-09-16 23:59:57.418201685        25.84    780.44     57.06         868302
2021-09-17 00:00:00.412962675        25.84    780.42     57.02         869987
2021-09-17 00:00:03.407538652        25.84    780.44     57.02         877365
2021-09-17 00:00:06.402314186        25.83    780.42     57.05         868302
2021-09-17 00:00:09.396840096        25.84    780.42     57.00         875651

                              IAQ  iaqAccuracy  year  month  day  hour  \
datetime
2021-09-16 23:59:42.445474625  246.4           1  2021      9   16    23
2021-09-16 23:59:45.439809322  246.7           1  2021      9   16    23
```

```
2021-09-16 23:59:48.434415102  245.9            1  2021      9   16    23
2021-09-16 23:59:51.428925753  246.9            1  2021      9   16    23
2021-09-16 23:59:54.423571348  247.7            1  2021      9   16    23
2021-09-16 23:59:57.418201685  249.6            1  2021      9   16    23
2021-09-17 00:00:00.412962675  250.0            1  2021      9   17     0
2021-09-17 00:00:03.407538652  248.1            1  2021      9   17     0
2021-09-17 00:00:06.402314186  249.8            1  2021      9   17     0
2021-09-17 00:00:09.396840096  248.6            1  2021      9   17     0

                               …       PM10   PM2.5      SO2  \
datetime                       …
2021-09-16 23:59:42.445474625  …        NaN     NaN      NaN
2021-09-16 23:59:45.439809322  …        NaN     NaN      NaN
2021-09-16 23:59:48.434415102  …        NaN     NaN      NaN
2021-09-16 23:59:51.428925753  …        NaN     NaN      NaN
2021-09-16 23:59:54.423571348  …        NaN     NaN      NaN
2021-09-16 23:59:57.418201685  …  29.666667  24.875  0.00335
2021-09-17 00:00:00.412962675  …        NaN     NaN      NaN
2021-09-17 00:00:03.407538652  …        NaN     NaN      NaN
2021-09-17 00:00:06.402314186  …        NaN     NaN      NaN
2021-09-17 00:00:09.396840096  …        NaN     NaN      NaN

                               temperature_outdoor  feels_like  \
datetime
2021-09-16 23:59:42.445474625                  NaN         NaN
2021-09-16 23:59:45.439809322                  NaN         NaN
2021-09-16 23:59:48.434415102                  NaN         NaN
2021-09-16 23:59:51.428925753                  NaN         NaN
2021-09-16 23:59:54.423571348                  NaN         NaN
2021-09-16 23:59:57.418201685                18.34       17.69
2021-09-17 00:00:00.412962675                  NaN         NaN
2021-09-17 00:00:03.407538652                  NaN         NaN
2021-09-17 00:00:06.402314186                  NaN         NaN
2021-09-17 00:00:09.396840096                  NaN         NaN

                               pressure_outdoor  humidity_outdoor  wind_speed  \
datetime
2021-09-16 23:59:42.445474625               NaN               NaN         NaN
2021-09-16 23:59:45.439809322               NaN               NaN         NaN
2021-09-16 23:59:48.434415102               NaN               NaN         NaN
2021-09-16 23:59:51.428925753               NaN               NaN         NaN
2021-09-16 23:59:54.423571348               NaN               NaN         NaN
2021-09-16 23:59:57.418201685            1024.0              56.0        4.63
2021-09-17 00:00:00.412962675               NaN               NaN         NaN
2021-09-17 00:00:03.407538652               NaN               NaN         NaN
2021-09-17 00:00:06.402314186               NaN               NaN         NaN
2021-09-17 00:00:09.396840096               NaN               NaN         NaN

                               wind_deg  weather_main
datetime
2021-09-16 23:59:42.445474625      NaN           NaN
2021-09-16 23:59:45.439809322      NaN           NaN
2021-09-16 23:59:48.434415102      NaN           NaN
2021-09-16 23:59:51.428925753      NaN           NaN
2021-09-16 23:59:54.423571348      NaN           NaN
2021-09-16 23:59:57.418201685    300.0          Rain
2021-09-17 00:00:00.412962675      NaN           NaN
2021-09-17 00:00:03.407538652      NaN           NaN
2021-09-17 00:00:06.402314186      NaN           NaN
```

```
2021-09-17 00:00:09.396840096       NaN           NaN

[10 rows x 27 columns]
```

Dataset with Indoor and Outdoor Data: * 6285103 Rows * 27 Columns.

```
                            temperature  pressure  humidity  gasResistance  \
datetime
2021-02-12 06:59:59.987502337        21.51    777.30     43.78         143943
2021-02-12 07:59:58.990879536        21.01    776.94     42.43         152841
2021-02-12 08:59:57.738294601        20.41    776.35     42.60         153259
2021-02-12 09:59:59.458741903        20.28    776.20     42.18         145689
2021-02-12 10:59:58.053189993        19.92    776.23     42.24         141519
…                                    …        …       …             …
2021-09-13 23:59:59.861400843        24.32    780.40     55.44         872810
2021-09-14 23:59:58.549029827        25.52    779.40     54.90         937936
2021-09-15 23:59:58.141078472        27.09    778.30     48.42        1221617
2021-09-16 23:59:57.418201685        25.84    780.44     57.06         868302
2021-09-17 23:59:59.829545736        26.38    782.02     54.55         947851

                              IAQ  iaqAccuracy  year  month  day  hour  \
datetime
2021-02-12 06:59:59.987502337   96.2            1  2021      2   12     6
2021-02-12 07:59:58.990879536   80.4            1  2021      2   12     7
2021-02-12 08:59:57.738294601   99.9            1  2021      2   12     8
2021-02-12 09:59:59.458741903  177.1            1  2021      2   12     9
2021-02-12 10:59:58.053189993  214.2            1  2021      2   12    10
…                              …          …   …    …  …   …
2021-09-13 23:59:59.861400843  230.0            3  2021      9   13    23
2021-09-14 23:59:58.549029827  125.9            1  2021      9   14    23
2021-09-15 23:59:58.141078472  148.6            3  2021      9   15    23
2021-09-16 23:59:57.418201685  249.6            1  2021      9   16    23
2021-09-17 23:59:59.829545736  112.9            1  2021      9   17    23

                            …       PM10      PM2.5       SO2  \
datetime                    …
2021-02-12 06:59:59.987502337  …  57.000000  25.000000  0.005000
2021-02-12 07:59:58.990879536  …  67.000000  33.000000  0.003000
2021-02-12 08:59:57.738294601  …  50.000000  28.000000  0.002000
2021-02-12 09:59:59.458741903  …  40.000000  21.000000  0.002000
2021-02-12 10:59:58.053189993  …  33.000000  19.000000  0.001000
…                           …        …         …        …
2021-09-13 23:59:59.861400843  …   7.047619   4.500000  0.000125
2021-09-14 23:59:58.549029827  …  24.416667  17.333333  0.000875
2021-09-15 23:59:58.141078472  …  46.833333  40.041667  0.001542
2021-09-16 23:59:57.418201685  …  29.666667  24.875000  0.003350
2021-09-17 23:59:59.829545736  …  20.304348  16.391304  0.001304

                            temperature_outdoor  feels_like  \
datetime
2021-02-12 06:59:59.987502337                13.87       12.46
2021-02-12 07:59:58.990879536                12.81       11.37
2021-02-12 08:59:57.738294601                10.83        9.35
2021-02-12 09:59:59.458741903                 6.40        3.51
2021-02-12 10:59:58.053189993                 6.23        6.23
…                                             …        …
2021-09-13 23:59:59.861400843                15.85       14.95
2021-09-14 23:59:58.549029827                17.95       17.08
2021-09-15 23:59:58.141078472                18.45       17.63
2021-09-16 23:59:57.418201685                18.34       17.69
```

```
2021-09-17 23:59:59.829545736               17.61       17.85

                            pressure_outdoor  humidity_outdoor  wind_speed  \
datetime
2021-02-12 06:59:59.987502337           1020.0              44.0        0.00
2021-02-12 07:59:58.990879536           1020.0              47.0        0.00
2021-02-12 08:59:57.738294601           1019.0              53.0        1.54
2021-02-12 09:59:59.458741903           1019.0              61.0        4.12
2021-02-12 10:59:58.053189993           1019.0              57.0        0.00
…                                          …                 …           …
2021-09-13 23:59:59.861400843           1025.0              56.0        4.12
2021-09-14 23:59:58.549029827           1023.0              49.0        7.72
2021-09-15 23:59:58.141078472           1022.0              49.0        6.17
2021-09-16 23:59:57.418201685           1024.0              56.0        4.63
2021-09-17 23:59:59.829545736           1015.0              93.0        1.37

                            wind_deg  weather_main
datetime
2021-02-12 06:59:59.987502337      0.0         Clear
2021-02-12 07:59:58.990879536      0.0         Clear
2021-02-12 08:59:57.738294601     60.0         Clear
2021-02-12 09:59:59.458741903     40.0         Clear
2021-02-12 10:59:58.053189993      0.0         Clear
…                                  …           …
2021-09-13 23:59:59.861400843    170.0          Rain
2021-09-14 23:59:58.549029827    130.0        Clouds
2021-09-15 23:59:58.141078472    130.0         Smoke
2021-09-16 23:59:57.418201685    300.0          Rain
2021-09-17 23:59:59.829545736    199.0          Rain

[1189 rows x 27 columns]
```

1.5 Imputations

We found that the resulting dataframe after merging 2 datasets (Outdoor Data that is sampled every 1 hour and Indoor Data that is sampled every 3 seconds) contains repeated records on the columns of hourly data: SINAICA Gov't Air Quality Monitoring and OpenWeatherData.

We think that the repeated data can be an issue, as the data moves very abruptly from a record call it at 10:57 and 11:00. This is relevant as the real world is not represented by the data correctly. Temperature, pressure and general natural features move slowly from one value to other. But we don't have that data, and it's not easily obtainable.

Therefore, we propose an approach similar to the imputations using the interpolation incorporating noise, that could avert the overfitting issue on our machine learning and deep learning training.

```
CPU times: user 476 ms, sys: 166 ms, total: 642 ms
Wall time: 639 ms

                            temperature_outdoor  feels_like  \
datetime
2021-02-12 06:59:59.987502337                13.87       12.46
2021-02-12 07:59:58.990879536                12.81       11.37
2021-02-12 08:59:57.738294601                10.83        9.35
2021-02-12 09:59:59.458741903                 6.40        3.51
2021-02-12 10:59:58.053189993                 6.23        6.23
…                                             …           …
2021-09-13 23:59:59.861400843                15.85       14.95
2021-09-14 23:59:58.549029827                17.95       17.08
2021-09-15 23:59:58.141078472                18.45       17.63
2021-09-16 23:59:57.418201685                18.34       17.69
2021-09-17 23:59:59.829545736                17.61       17.85
```

```
                            pressure_outdoor  humidity_outdoor  wind_speed  \
datetime
2021-02-12 06:59:59.987502337            1020.0              44.0        0.00
2021-02-12 07:59:58.990879536            1020.0              47.0        0.00
2021-02-12 08:59:57.738294601            1019.0              53.0        1.54
2021-02-12 09:59:59.458741903            1019.0              61.0        4.12
2021-02-12 10:59:58.053189993            1019.0              57.0        0.00
…                                           …                 …           …
2021-09-13 23:59:59.861400843            1025.0              56.0        4.12
2021-09-14 23:59:58.549029827            1023.0              49.0        7.72
2021-09-15 23:59:58.141078472            1022.0              49.0        6.17
2021-09-16 23:59:57.418201685            1024.0              56.0        4.63
2021-09-17 23:59:59.829545736            1015.0              93.0        1.37

                            wind_deg weather_main
datetime
2021-02-12 06:59:59.987502337      0.0        Clear
2021-02-12 07:59:58.990879536      0.0        Clear
2021-02-12 08:59:57.738294601     60.0        Clear
2021-02-12 09:59:59.458741903     40.0        Clear
2021-02-12 10:59:58.053189993      0.0        Clear
…                                  …          …
2021-09-13 23:59:59.861400843    170.0         Rain
2021-09-14 23:59:58.549029827    130.0       Clouds
2021-09-15 23:59:58.141078472    130.0        Smoke
2021-09-16 23:59:57.418201685    300.0         Rain
2021-09-17 23:59:59.829545736    199.0         Rain

[1323 rows x 7 columns]
```

1.6 Resampling

To reduce training time we propose to have a resampling of the data.

In the following subsections we create those resampled-data dataframes.

1.6.1 1 Minute Resampling

```
CPU times: user 4.91 s, sys: 166 ms, total: 5.07 s
Wall time: 5.07 s

                     temperature    pressure   humidity  gasResistance  \
datetime
2021-02-12 06:04:00    21.530000  777.410000  43.974000  151849.400000
2021-02-12 06:05:00    21.526250  777.408750  43.840000  152790.000000
2021-02-12 06:06:00    21.693000  777.409000  43.426000  152220.550000
2021-02-12 06:07:00    21.759000  777.410500  43.245500  151978.450000
2021-02-12 06:08:00    21.750500  777.390500  43.056000  150300.400000
…                            …          …          …             …
2021-09-18 01:16:00    25.871000  782.805000  56.607000  921467.050000
2021-09-18 01:17:00    25.861000  782.832000  56.587500  921211.600000
2021-09-18 01:18:00    25.850000  782.866000  56.597500  922348.700000
2021-09-18 01:19:00    25.836190  782.900000  56.683810  921997.095238
2021-09-18 01:20:00    25.836471  782.932941  56.662941  922768.882353

                           IAQ  iaqAccuracy    year  month   day  hour  \
datetime
2021-02-12 06:04:00   37.200000          1.0  2021.0    2.0  12.0   6.0
2021-02-12 06:05:00   32.162500          1.0  2021.0    2.0  12.0   6.0
```

```
2021-02-12 06:06:00   34.325000          1.0  2021.0    2.0  12.0   6.0
2021-02-12 06:07:00   36.190000          1.0  2021.0    2.0  12.0   6.0
2021-02-12 06:08:00   46.600000          1.0  2021.0    2.0  12.0   6.0
…                          …          …     …    …   …    …
2021-09-18 01:16:00  133.940000          1.0  2021.0    9.0  18.0   1.0
2021-09-18 01:17:00  134.815000          1.0  2021.0    9.0  18.0   1.0
2021-09-18 01:18:00  133.850000          1.0  2021.0    9.0  18.0   1.0
2021-09-18 01:19:00  134.190476          1.0  2021.0    9.0  18.0   1.0
2021-09-18 01:20:00  133.876471          1.0  2021.0    9.0  18.0   1.0

                    minute     second  temperature_outdoor  feels_like  \
datetime
2021-02-12 06:04:00     4.0  15.000000                  NaN         NaN
2021-02-12 06:05:00     5.0  45.875000                  NaN         NaN
2021-02-12 06:06:00     6.0  30.500000                  NaN         NaN
2021-02-12 06:07:00     7.0  30.500000                  NaN         NaN
2021-02-12 06:08:00     8.0  30.500000                  NaN         NaN
…                        …        …                    …         …
2021-09-18 01:16:00    16.0  28.500000                  NaN         NaN
2021-09-18 01:17:00    17.0  28.500000                  NaN         NaN
2021-09-18 01:18:00    18.0  28.500000                  NaN         NaN
2021-09-18 01:19:00    19.0  29.619048                  NaN         NaN
2021-09-18 01:20:00    20.0  26.000000                  NaN         NaN

                    pressure_outdoor  humidity_outdoor  wind_speed  wind_deg
datetime
2021-02-12 06:04:00              NaN               NaN         NaN       NaN
2021-02-12 06:05:00              NaN               NaN         NaN       NaN
2021-02-12 06:06:00              NaN               NaN         NaN       NaN
2021-02-12 06:07:00              NaN               NaN         NaN       NaN
2021-02-12 06:08:00              NaN               NaN         NaN       NaN
…                                …                 …         …         …
2021-09-18 01:16:00              NaN               NaN         NaN       NaN
2021-09-18 01:17:00              NaN               NaN         NaN       NaN
2021-09-18 01:18:00              NaN               NaN         NaN       NaN
2021-09-18 01:19:00              NaN               NaN         NaN       NaN
2021-09-18 01:20:00              NaN               NaN         NaN       NaN

[313637 rows x 18 columns]
```

1.6.2 2 Minute Resampling

```
CPU times: user 2.32 s, sys: 112 ms, total: 2.43 s
Wall time: 2.43 s

                    temperature    pressure   humidity  gasResistance  \
datetime
2021-02-12 06:04:00    21.527692  777.409231  43.891538  152428.230769
2021-02-12 06:06:00    21.726000  777.409750  43.335750  152099.500000
2021-02-12 06:08:00    21.686250  777.365250  43.291500  147429.200000
2021-02-12 06:10:00    21.499500  777.302000  43.106250  149288.475000
2021-02-12 06:12:00    21.628250  777.279500  42.830750  149325.975000
…                            …          …          …            …
2021-09-18 01:12:00    25.913500  782.776500  56.513250  920534.625000
2021-09-18 01:14:00    25.892750  782.793500  56.555500  921024.825000
2021-09-18 01:16:00    25.866000  782.818500  56.597250  921339.325000
2021-09-18 01:18:00    25.842927  782.883415  56.641707  922168.609756
2021-09-18 01:20:00    25.836471  782.932941  56.662941  922768.882353

                         IAQ  iaqAccuracy    year  month   day  hour  \
```

```
datetime
2021-02-12 06:04:00   34.100000          1.0  2021.0    2.0  12.0   6.0
2021-02-12 06:06:00   35.257500          1.0  2021.0    2.0  12.0   6.0
2021-02-12 06:08:00   72.652500          1.0  2021.0    2.0  12.0   6.0
2021-02-12 06:10:00   69.505000          1.0  2021.0    2.0  12.0   6.0
2021-02-12 06:12:00   71.237500          1.0  2021.0    2.0  12.0   6.0
…                           …          …     …    …   …    …
2021-09-18 01:12:00  134.397500          1.0  2021.0    9.0  18.0   1.0
2021-09-18 01:14:00  134.380000          1.0  2021.0    9.0  18.0   1.0
2021-09-18 01:16:00  134.377500          1.0  2021.0    9.0  18.0   1.0
2021-09-18 01:18:00  134.024390          1.0  2021.0    9.0  18.0   1.0
2021-09-18 01:20:00  133.876471          1.0  2021.0    9.0  18.0   1.0

                       minute     second  temperature_outdoor  feels_like  \
datetime
2021-02-12 06:04:00   4.615385  34.000000                  NaN         NaN
2021-02-12 06:06:00   6.500000  30.500000                  NaN         NaN
2021-02-12 06:08:00   8.500000  30.500000                  NaN         NaN
2021-02-12 06:10:00  10.500000  29.900000                  NaN         NaN
2021-02-12 06:12:00  12.500000  29.500000                  NaN         NaN
…                          …         …                    …          …
2021-09-18 01:12:00  12.500000  28.500000                  NaN         NaN
2021-09-18 01:14:00  14.500000  28.500000                  NaN         NaN
2021-09-18 01:16:00  16.500000  28.500000                  NaN         NaN
2021-09-18 01:18:00  18.512195  29.073171                  NaN         NaN
2021-09-18 01:20:00  20.000000  26.000000                  NaN         NaN

                    pressure_outdoor  humidity_outdoor  wind_speed  wind_deg
datetime
2021-02-12 06:04:00              NaN               NaN         NaN       NaN
2021-02-12 06:06:00              NaN               NaN         NaN       NaN
2021-02-12 06:08:00              NaN               NaN         NaN       NaN
2021-02-12 06:10:00              NaN               NaN         NaN       NaN
2021-02-12 06:12:00              NaN               NaN         NaN       NaN
…                                …                 …           …         …
2021-09-18 01:12:00              NaN               NaN         NaN       NaN
2021-09-18 01:14:00              NaN               NaN         NaN       NaN
2021-09-18 01:16:00              NaN               NaN         NaN       NaN
2021-09-18 01:18:00              NaN               NaN         NaN       NaN
2021-09-18 01:20:00              NaN               NaN         NaN       NaN

[156819 rows x 18 columns]
```

1.6.3 3 Minute Resampling

```
CPU times: user 918 ms, sys: 119 ms, total: 1.04 s
Wall time: 1.04 s

                    temperature    pressure   humidity  gasResistance  \
datetime
2021-02-12 06:00:00    21.530000  777.410000  43.974000  151849.400000
2021-02-12 06:05:00    21.689773  777.389432  43.361477  150039.409091
2021-02-12 06:10:00    21.538300  777.285200  42.909800  149975.940000
2021-02-12 06:15:00    21.563900  777.269000  42.704100  150897.020000
2021-02-12 06:20:00    21.616931  777.223960  42.695545  149963.910891
…                            …         …         …            …
2021-09-18 01:00:00    25.987500  782.832000  56.333000  918660.290000
2021-09-18 01:05:00    25.966100  782.800200  56.379800  920302.740000
2021-09-18 01:10:00    25.915100  782.782200  56.511300  920836.000000
2021-09-18 01:15:00    25.860099  782.840594  56.613069  921519.306931
```

```
2021-09-18 01:20:00    25.836471  782.932941  56.662941  922768.882353

                           IAQ  iaqAccuracy    year  month   day  hour  \
datetime
2021-02-12 06:00:00   37.200000          1.0  2021.0    2.0  12.0   6.0
2021-02-12 06:05:00   51.973864          1.0  2021.0    2.0  12.0   6.0
2021-02-12 06:10:00   67.172000          1.0  2021.0    2.0  12.0   6.0
2021-02-12 06:15:00   65.798000          1.0  2021.0    2.0  12.0   6.0
2021-02-12 06:20:00   71.275248          1.0  2021.0    2.0  12.0   6.0
…                           …          …      …    …   …    …
2021-09-18 01:00:00  135.811000          1.0  2021.0    9.0  18.0   1.0
2021-09-18 01:05:00  134.243000          1.0  2021.0    9.0  18.0   1.0
2021-09-18 01:10:00  134.184000          1.0  2021.0    9.0  18.0   1.0
2021-09-18 01:15:00  134.372277          1.0  2021.0    9.0  18.0   1.0
2021-09-18 01:20:00  133.876471          1.0  2021.0    9.0  18.0   1.0

                        minute     second  temperature_outdoor  feels_like  \
datetime
2021-02-12 06:00:00   4.000000  15.000000                  NaN         NaN
2021-02-12 06:05:00   7.272727  31.897727                  NaN         NaN
2021-02-12 06:10:00  12.000000  29.660000                  NaN         NaN
2021-02-12 06:15:00  17.000000  28.890000                  NaN         NaN
2021-02-12 06:20:00  22.009901  29.029703                  NaN         NaN
…                          …          …                    …          …
2021-09-18 01:00:00   2.000000  29.710000                  NaN         NaN
2021-09-18 01:05:00   7.000000  29.500000                  NaN         NaN
2021-09-18 01:10:00  12.000000  28.600000                  NaN         NaN
2021-09-18 01:15:00  17.019802  28.732673                  NaN         NaN
2021-09-18 01:20:00  20.000000  26.000000                  NaN         NaN

                     pressure_outdoor  humidity_outdoor  wind_speed  wind_deg
datetime
2021-02-12 06:00:00               NaN               NaN         NaN       NaN
2021-02-12 06:05:00               NaN               NaN         NaN       NaN
2021-02-12 06:10:00               NaN               NaN         NaN       NaN
2021-02-12 06:15:00               NaN               NaN         NaN       NaN
2021-02-12 06:20:00               NaN               NaN         NaN       NaN
…                                 …                 …           …         …
2021-09-18 01:00:00               NaN               NaN         NaN       NaN
2021-09-18 01:05:00               NaN               NaN         NaN       NaN
2021-09-18 01:10:00               NaN               NaN         NaN       NaN
2021-09-18 01:15:00               NaN               NaN         NaN       NaN
2021-09-18 01:20:00               NaN               NaN         NaN       NaN

[62729 rows x 18 columns]
```

1.7 References

- https://scikit-learn.org/stable/modules/linear_model.html#generalized-linear-regression
- https://pythonhealthcare.org/2018/05/03/81-distribution-fitting-to-data/
- https://medium.com/@amirarsalan.rajabi/distribution-fitting-with-python-scipy-bb70a42c0aed
- https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KernelDensity.html?highlight=kernel%20density#sklearn.neighbors.KernelDensity