

ingest_measurement_data_02

July 2, 2025

```
[6]: # CONFIG CELL
import logging

from notebook_utils import set_root_directory

set_root_directory()
logging.getLogger().setLevel(logging.INFO)
```

```
[7]: # IMPORTS
import pandas as pd

from app.constants import TARGET_VARIABLES
from app.measurement_ingestor import MeasurementIngestor
```

```
[8]: # CONSTANTS
INPUT_DIR = "data/pomiary"
SENSOR_METADATA = "input_files/sensor_metadata.parquet"

TIME_1H_RESULT = "input_files/measurements_1h.parquet"
TIME_24H_RESULT = "input_files/measurements_24h.parquet"
```

```
[10]: # CODE
sensor_metada = pd.read_parquet(SENSOR_METADATA)

ingestor = MeasurementIngestor(
    input_dir=INPUT_DIR,
    sensor_metadata=sensor_metada,
    exclude_depoyzcja=True,
    target_variables=TARGET_VARIABLES,
)
time_1h_df, time_24h_df = ingestor.transform()
```

```
WARNING:root:Sensor name PkRzeszWIOSLang not found. Skipping.
WARNING:root:Sensor name PkRzeszWIOSLang not found. Skipping.
WARNING:root:Sensor name DsBolRynek not found. Skipping.
WARNING:root:Sensor name LbKra not found. Skipping.
WARNING:root:Sensor name PkRzeszWIOSLang not found. Skipping.
WARNING:root:Sensor name PkRzeszWIOSLang not found. Skipping.
```

```

WARNING:root:Sensor name PkRzeszWIOSLang not found. Skipping.
WARNING:root:Sensor name LbKra not found. Skipping.
WARNING:root:Sensor name LbKra not found. Skipping.
WARNING:root:Sensor name LbKra not found. Skipping.
WARNING:root:Sensor name LbKra not found. Skipping.
WARNING:root:Sensor name LbKra not found. Skipping.
WARNING:root:Sensor name LbKra not found. Skipping.
WARNING:root:Sensor name KpPASBTXBydgKujawskie not found. Skipping.
WARNING:root:Sensor name KpPASBTXBydgLPKiW not found. Skipping.
WARNING:root:Sensor name KpPASBTXBydgTorunskie not found. Skipping.
WARNING:root:Sensor name KpPASBTXBydgZachem not found. Skipping.
WARNING:root:Sensor name KpPASTorunWalyGenSik not found. Skipping.
WARNING:root:Sensor name KpBrodnicaZwirkiW not found. Skipping.
WARNING:root:Sensor name KpBrodnicaZwirkiW not found. Skipping.
WARNING:root:Sensor name KpBrodnicaZwirkiW not found. Skipping.
WARNING:root:Sensor name LbZaosChrubieszowsk not found. Skipping.
C:\Users\szymo\.conda\envs\ispies\Lib\site-
packages\openpyxl\worksheet\header_footer.py:48: UserWarning: Cannot parse
header or footer so it will be ignored
    warn("""Cannot parse header or footer so it will be ignored""")
C:\Users\szymo\.conda\envs\ispies\Lib\site-
packages\openpyxl\worksheet\header_footer.py:48: UserWarning: Cannot parse
header or footer so it will be ignored
    warn("""Cannot parse header or footer so it will be ignored""")
WARNING:root:Sensor name nan not found. Skipping.
WARNING:root:Sensor name nan not found. Skipping.
WARNING:root:Sensor name nan not found. Skipping.
C:\Users\szymo\.conda\envs\ispies\Lib\site-
packages\openpyxl\worksheet\header_footer.py:48: UserWarning: Cannot parse
header or footer so it will be ignored
    warn("""Cannot parse header or footer so it will be ignored""")
C:\Users\szymo\.conda\envs\ispies\Lib\site-
packages\openpyxl\worksheet\header_footer.py:48: UserWarning: Cannot parse
header or footer so it will be ignored
    warn("""Cannot parse header or footer so it will be ignored""")
C:\Users\szymo\.conda\envs\ispies\Lib\site-
packages\openpyxl\worksheet\header_footer.py:48: UserWarning: Cannot parse
header or footer so it will be ignored
    warn("""Cannot parse header or footer so it will be ignored""")
C:\Users\szymo\.conda\envs\ispies\Lib\site-
packages\openpyxl\worksheet\header_footer.py:48: UserWarning: Cannot parse
header or footer so it will be ignored
    warn("""Cannot parse header or footer so it will be ignored""")
C:\Users\szymo\.conda\envs\ispies\Lib\site-
packages\openpyxl\worksheet\header_footer.py:48: UserWarning: Cannot parse
header or footer so it will be ignored
    warn("""Cannot parse header or footer so it will be ignored""")
INFO:root:File 2015_depozycja.xlsx is excluded from the processing. Check

```

parameter exclude_depozycja=True.
INFO:root:No sensor names found in file 2016_C6H6_1g.xlsx. Trying to extract in another way.
INFO:root:Found row with station codes
INFO:root:No sensor names found in file 2016_C6H6_24g.xlsx. Trying to extract in another way.
INFO:root:Found row with station codes
INFO:root:No sensor names found in file 2016_CO_1g.xlsx. Trying to extract in another way.
INFO:root:Found row with station codes
INFO:root:File 2016_Depozycja.xlsx is excluded from the processing. Check parameter exclude_depozycja=True.
INFO:root:No sensor names found in file 2016_NO2_1g.xlsx. Trying to extract in another way.
INFO:root:Found row with station codes
INFO:root:No sensor names found in file 2016_NO2_24g.xlsx. Trying to extract in another way.
INFO:root:Found row with station codes
INFO:root:No sensor names found in file 2016_NOx_1g.xlsx. Trying to extract in another way.
INFO:root:Found row with station codes
INFO:root:No sensor names found in file 2016_O3_1g.xlsx. Trying to extract in another way.
INFO:root:Found row with station codes
INFO:root:No sensor names found in file 2016_PM10_1g.xlsx. Trying to extract in another way.
INFO:root:Found row with station codes
INFO:root:No sensor names found in file 2016_PM10_24g.xlsx. Trying to extract in another way.
INFO:root:Found row with station codes
INFO:root:No sensor names found in file 2016_SO2_1g.xlsx. Trying to extract in another way.
INFO:root:Found row with station codes
INFO:root:No sensor names found in file 2016_SO2_24g.xlsx. Trying to extract in another way.
INFO:root:Found row with station codes
INFO:root:File 2017_Depozycja.xlsx is excluded from the processing. Check parameter exclude_depozycja=True.
INFO:root:No sensor names found in file 2018_C6H6_1g.xlsx. Trying to extract in another way.
INFO:root:Found row with station codes
INFO:root:No sensor names found in file 2018_C6H6_24g.xlsx. Trying to extract in another way.
INFO:root:Found row with station codes
INFO:root:No sensor names found in file 2018_CO_1g.xlsx. Trying to extract in another way.
INFO:root:Found row with station codes
INFO:root:File 2018_Depozycja.xlsx is excluded from the processing. Check

```

parameter exclude_depozycja=True.
INFO:root:No sensor names found in file 2018_NO2_1g.xlsx. Trying to extract in
another way.
INFO:root:Found row with station codes
INFO:root:No sensor names found in file 2018_NO2_24g.xlsx. Trying to extract in
another way.
INFO:root:Found row with station codes
INFO:root:No sensor names found in file 2018_NOx_1g.xlsx. Trying to extract in
another way.
INFO:root:Found row with station codes
INFO:root:No sensor names found in file 2018_O3_1g.xlsx. Trying to extract in
another way.
INFO:root:Found row with station codes
INFO:root:No sensor names found in file 2018_PM10_1g.xlsx. Trying to extract in
another way.
INFO:root:Found row with station codes
INFO:root:No sensor names found in file 2018_PM10_24g.xlsx. Trying to extract in
another way.
INFO:root:Found row with station codes
INFO:root:No sensor names found in file 2018_PM25_1g.xlsx. Trying to extract in
another way.
INFO:root:Found row with station codes
INFO:root:No sensor names found in file 2018_PM25_24g.xlsx. Trying to extract in
another way.
INFO:root:Found row with station codes
INFO:root:No sensor names found in file 2018_SO2_1g.xlsx. Trying to extract in
another way.
INFO:root:Found row with station codes
INFO:root:No sensor names found in file 2018_SO2_24g.xlsx. Trying to extract in
another way.
INFO:root:Found row with station codes
INFO:root:File 2019_Depozycja.xlsx is excluded from the processing. Check
parameter exclude_depozycja=True.
INFO:root:File 2020_Depozycja.xlsx is excluded from the processing. Check
parameter exclude_depozycja=True.
INFO:root:File 2021_Depozycja_1m.xlsx is excluded from the processing. Check
parameter exclude_depozycja=True.
INFO:root:File 2022_Depozycja_1m.xlsx is excluded from the processing. Check
parameter exclude_depozycja=True.
INFO:root:File 2023_Depozycja_1m.xlsx is excluded from the processing. Check
parameter exclude_depozycja=True.

```

```

[11]: time_1h_df.to_parquet(TIME_1H_RESULT, index=False)
      time_24h_df.to_parquet(TIME_24H_RESULT, index=False)

```

```

[12]: time_1h_df.shape

```

```
[12]: (97197408, 11)
```

```
[13]: time_1h_df.head()
```

```
[13]:
```

	ds	unique_id	NO2	NOx	O3	S02	C6H6	CO	PM10	PM2.5	\
0	2000-01-01 01:00:00	10	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1	2000-01-01 02:00:00	10	6.0	7.0	35.0	NaN	NaN	NaN	NaN	NaN	
2	2000-01-01 03:00:00	10	3.0	3.0	42.0	NaN	NaN	NaN	NaN	NaN	
3	2000-01-01 04:00:00	10	4.0	4.0	39.0	NaN	NaN	NaN	NaN	NaN	
4	2000-01-01 05:00:00	10	3.0	3.0	44.0	NaN	NaN	NaN	NaN	NaN	


```
NO
```

0	NaN
1	NaN
2	NaN
3	NaN
4	NaN

```
[14]: time_24h_df.shape
```

```
[14]: (7775442, 7)
```

```
[15]: time_24h_df.head(5)
```

```
[15]:
```

	ds	unique_id	NO2	PM10	S02	C6H6	PM2.5
0	2000-01-01	102	3.6	NaN	NaN	NaN	NaN
1	2000-01-02	102	4.9	NaN	NaN	NaN	NaN
2	2000-01-03	102	5.6	NaN	NaN	NaN	NaN
3	2000-01-04	102	5.3	NaN	NaN	NaN	NaN
4	2000-01-05	102	6.9	NaN	NaN	NaN	NaN

```
[16]: time_24h_df.query("unique_id == 1")
```

```
[16]:
```

	ds	unique_id	NO2	PM10	S02	C6H6	PM2.5
3594060	2000-01-01	1	NaN	NaN	NaN	NaN	NaN
3594061	2000-01-02	1	NaN	NaN	NaN	NaN	NaN
3594062	2000-01-03	1	NaN	NaN	NaN	NaN	NaN
3594063	2000-01-04	1	NaN	NaN	NaN	NaN	NaN
3594064	2000-01-05	1	NaN	NaN	NaN	NaN	NaN
...
3602821	2023-12-27	1	NaN	NaN	NaN	NaN	NaN
3602822	2023-12-28	1	NaN	NaN	NaN	NaN	NaN
3602823	2023-12-29	1	NaN	NaN	NaN	NaN	NaN
3602824	2023-12-30	1	NaN	NaN	NaN	NaN	NaN
3602825	2023-12-31	1	NaN	NaN	NaN	NaN	NaN

```
[8766 rows x 7 columns]
```

[]: