# Moduł 10
# Przypadek użycia 2
# Czujniki wykrywania zanieczyszczeń

## iBigWorld:
## Innovations for Big Data in a Real World

UBB team

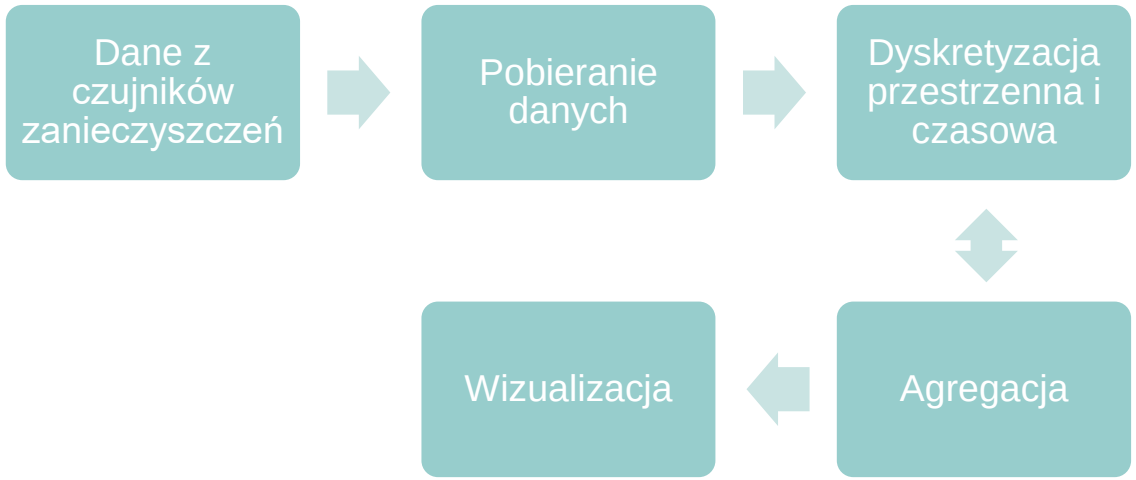


iBigWorld:2020-1-PL01-KA203-082197

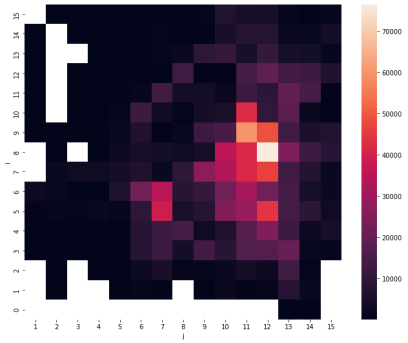
Co-funded by the Erasmus+ Programme of the European Union

# Eksploracja danych dot. zanieczyszczenia powietrza

- Analiza metodą eksploracji
  - “czyszczenie” danych geoprzestrzennych,
  - Dyskretyzacja danych z uwzględnieniem rozmiaru komórki oraz przedziału czasu na bazie statystyki
  - agregacja
  - wizualizacja




| | v | t | lat | long | j | i | \ |
|---|---|---|---|---|---|---|---|
| 207513 | 8.0 | 2021-05-29T21:33:51.010542+00:00 | 49.795803 | 18.947893 | 4 | 3 | |
| 207514 | 8.6 | 2021-05-29T21:33:49.008507+00:00 | 49.795533 | 18.947787 | 4 | 3 | |
| 207515 | 9.5 | 2021-05-29T21:33:47.028163+00:00 | 49.795290 | 18.947680 | 4 | 3 | |
| 207516 | 11.0 | 2021-05-29T21:33:45.029140+00:00 | 49.795040 | 18.947553 | 4 | 3 | |
| 207517 | 12.8 | 2021-05-29T21:33:43.049896+00:00 | 49.794758 | 18.947422 | 4 | 3 | |
| ... | ... | ... | ... | ... | ... | .. | .. |
| 2014273 | 49.5 | 2021-01-11T09:26:21.729906+00:00 | 49.792007 | 18.951485 | 4 | 3 | |
| 2014274 | 50.1 | 2021-01-11T09:26:21.370588+00:00 | 49.792007 | 18.951857 | 4 | 3 | |
| 2014275 | 51.0 | 2021-01-11T09:26:20.988369+00:00 | 49.792002 | 18.952235 | 4 | 3 | |
| 2014276 | 52.2 | 2021-01-11T09:26:20.650637+00:00 | 49.791990 | 18.952643 | 4 | 3 | |
| 2014277 | 54.6 | 2021-01-11T09:26:20.206521+00:00 | 49.791970 | 18.953118 | 4 | 3 | |



iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union

# Pobieranie danych

- Ładowanie danych:

```
import numpy as np
import pandas as pd

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

from datetime import datetime, date
df = pd.read_table('../input/measurements-pm25-17-08-2021csv/measurements_pm25_17_08_2021.csv',sep=',')
```



Co-funded by the Erasmus+ Programme of the European Union

# Czyszczenie danych

# Usuwanie danych spoza obszaru analizy (błędne koordynaty GPS)

- df = df[df.long >= 18.9]
- df = df[df.long <= 19.1]
- df = df[df.lat >= 49.78]
- df = df[df.lat <= 49.86]



Co-funded by the
Erasmus+ Programme
of the European Union

# Podział przestrzennych danych na obszary “kwadraty”

- cols = np.linspace(18.9,19.1, num=16)
- rows = np.linspace(49.78,49.86, num=16)

# Każda obserwacja jest przyporządkowana do danego obszaru

- df['j'] = np.searchsorted(cols, df['long'])
- df['i'] = np.searchsorted(rows, df['lat'])

iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union

# Podział danych czasowych na przedziały

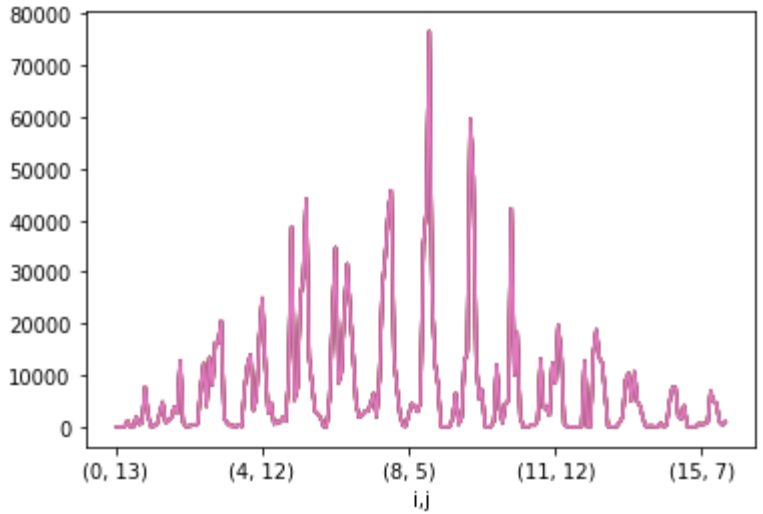
- df['t1'] = df['t']
- df['t1'] = df['t1'].apply(lambda x: datetime.fromisoformat(x).timetuple().tm_yday)
- df['t2'] = df.apply(lambda row : (row.t1 - 1) * 24 + datetime.fromisoformat(row.t).timetuple().tm_hour , axis=1)
- df['t3'] = df.apply(lambda row : row.t1 / 7, axis = 1)

iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union

# Agregacja przestrzennych danych w wykres

```
df.groupby(['i', 'j']).count().plot(legend=False)
tmp = df.groupby(['i', 'j'])['v'].count()
print(tmp)
```




iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union

# Agregacja czasowych danych dla poszczególnych obszarów

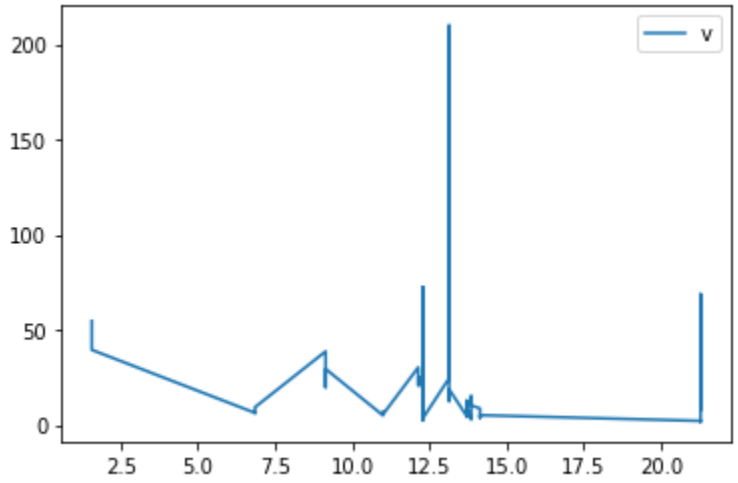
```
df_10_10 = df[(df.i == 3) & (df.j == 4)]
print(df_10_10)
df_10_10.plot(x='t3',y='v’)
df_10_10.plot.bar(x='t3', y='v', rot=0)
```

iBigWorld:2020-1-PL01-KA203-082197

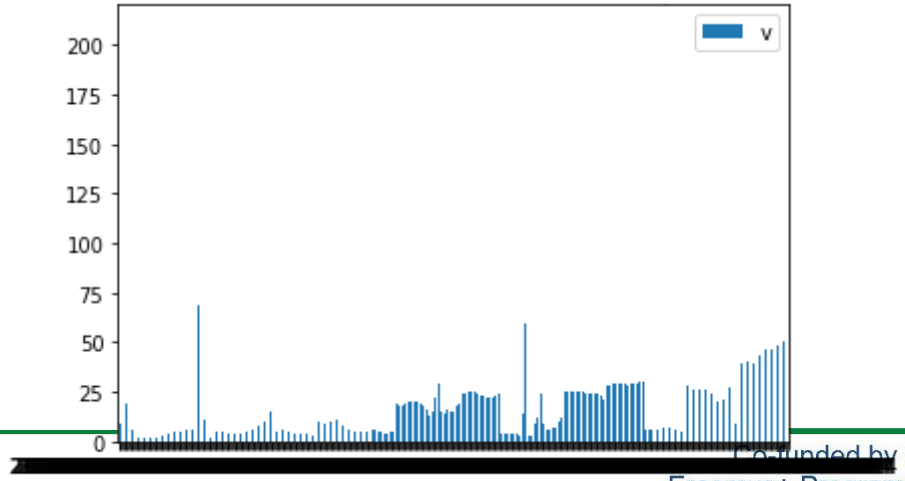
Co-funded by the
Erasmus+ Programme
of the European Union

# Agregacja czasowych danych dla poszczególnych obszarów

| | v | t | lat | long | j | i |
|---|---|---|---|---|---|---|
| 207513 | 8.0 | 2021-05-29T21:33:51.010542+00:00 | 49.795803 | 18.947893 | 4 | 3 |
| 207514 | 8.6 | 2021-05-29T21:33:49.008507+00:00 | 49.795533 | 18.947787 | 4 | 3 |
| 207515 | 9.5 | 2021-05-29T21:33:47.028163+00:00 | 49.795290 | 18.947680 | 4 | 3 |
| 207516 | 11.0 | 2021-05-29T21:33:45.029140+00:00 | 49.795040 | 18.947553 | 4 | 3 |
| 207517 | 12.8 | 2021-05-29T21:33:43.049896+00:00 | 49.794758 | 18.947422 | 4 | 3 |
| ... | ... | ... | ... | ... | .. | .. |
| 2014273 | 49.5 | 2021-01-11T09:26:21.729906+00:00 | 49.792007 | 18.951485 | 4 | 3 |
| 2014274 | 50.1 | 2021-01-11T09:26:21.370588+00:00 | 49.792007 | 18.951857 | 4 | 3 |
| 2014275 | 51.0 | 2021-01-11T09:26:20.988369+00:00 | 49.792002 | 18.952235 | 4 | 3 |
| 2014276 | 52.2 | 2021-01-11T09:26:20.650637+00:00 | 49.791990 | 18.952643 | 4 | 3 |
| 2014277 | 54.6 | 2021-01-11T09:26:20.206521+00:00 | 49.791970 | 18.953118 | 4 | 3 |




| | t1 | t2 | t3 |
|---|---|---|---|
| 207513 | 149 | 3573 | 21.285714 |
| 207514 | 149 | 3573 | 21.285714 |
| 207515 | 149 | 3573 | 21.285714 |
| 207516 | 149 | 3573 | 21.285714 |
| 207517 | 149 | 3573 | 21.285714 |
| ... | ... | ... | ... |
| 2014273 | 11 | 249 | 1.571429 |
| 2014274 | 11 | 249 | 1.571429 |
| 2014275 | 11 | 249 | 1.571429 |
| 2014276 | 11 | 249 | 1.571429 |
| 2014277 | 11 | 249 | 1.571429 |




iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union

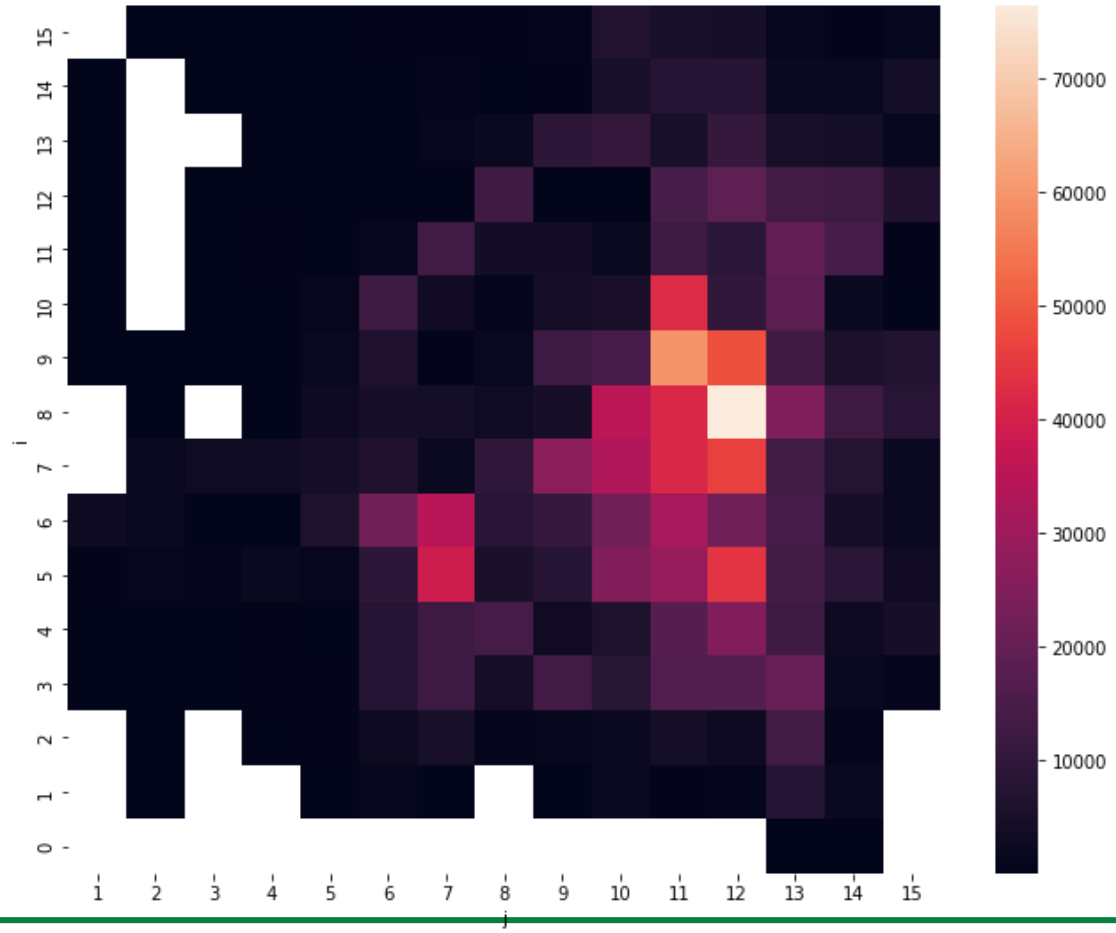
# Agregacja zarówna czasowa jak i przestrzenna

```
import numpy as npimport seaborn as sns
import matplotlib.pylab as plt
tmp = df.groupby(['i', 'j'])['v'].count().unstack()
print(tmp)
fig, ax = plt.subplots(figsize=(11, 9))
sns.heatmap(tmp)
ax.invert_yaxis()
```



Co-funded by the
Erasmus+ Programme
of the European Union

# Wykres: wynik agregacji przestrzenno-czasowej




iBigWorld:2020-1-PL01-KA203-082197

Co-funded by the Erasmus+ Programme of the European Union