



Міністерство освіти і науки України

Національний технічний університет України

“Київський політехнічний інститут імені Ігоря Сікорського”

Факультет інформатики та обчислювальної техніки

Кафедра інформатики та програмної інженерії

## **Лабораторна робота №4**

Аналіз даних з використанням мови Python

**Тема:** Візуалізація даних за допомогою matplotlib та Seaborn

**Варіант:** 1

Виконав

студент групи ІП-11:

Панченко С. В.

Перевірів:

Тимофєєва Ю. С

Київ 2023

## ЗМІСТ

1 Мета лабораторної роботи.....	6
2 Завдання.....	7
3 Виконання.....	8
3.1 Побудувати графік зміни середніх денних температур: 1. загальний; 2. за 2014 рік; 3. за квітень 2013 року; 4. за листопад 2013 –травень 2015; 5. за 2015 та 2016 на одному графіку.....	8
3.2 Знайти середні значення вологості 1. за 2016 рік; 2. за кожний місяць; 3. за кожні два тижні весни та літа 2014 року. 4. Розрахувати і зобразити зміни вологості у відсотках за кожен день впродовж літа 2015 року. 5. Знайти та зобразити графічно ковзне середнє вологості за 2013 рік з вікном в місяць. .....	13
4 Висновок.....	18

## 1 МЕТА ЛАБОРАТОРНОЇ РОБОТИ

Ознайомитись з основними діаграмами та графіками, що використовуються при аналізі даних. Навчитись будувати їх за допомогою бібліотек `matplotlib` та `Seaborn`.

## 2 ЗАВДАННЯ

Варіант 1.

Файл diamonds.csv.

1. Побудувати стовпчикові діаграми, на яких відобразити а) кількість діамантів кожного з класів якості; б) максимальну ціну діамантів кожного класу якості; в) середню глибину діамантів різного класу якості з різною якістю кольору.
2. Побудувати гістограму глибини діамантів у відсотках (depth), загальну і для кожного класу якості.
3. Побудувати діаграму розмаху параметру table (загальну і в залежності від якості кольору), визначити чи присутні викиди.
4. За допомогою діаграм розсіювання зробити висновки щодо залежності між а) довжиною і шириною; б) глибиною у % і глибиною у мм. Порахувати коефіцієнт кореляції за допомогою відповідних функцій.

## 3 ВИКОНАННЯ

3.1 Побудувати графік зміни середніх денних температур: 1. загальний; 2. за 2014 рік; 3. за квітень 2013 року; 4. за листопад 2013 –травень 2015; 5. за 2015 та 2016 на одному графіку

Для початку імпортуємо модуль numpy, pandas, seaborn та matplotlib, завантажимо дані до датафрейму, використавши колонку date як індекс.

```
In [110]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
df = pd.read_csv('data/Delhi_Climate.csv',
                 index_col=['date'], parse_dates=['date'])
df
```

Out[110]:

	meantemp	humidity	wind_speed	meanpressure
date				
2013-01-01	10.000000	84.500000	0.0	1015.666667
2013-01-02	7.400000	92.000000	2.98	1017.800000
2013-01-03	7.166667	87.000000	4.633333333333334	1018.666667
2013-01-04	8.666667	71.333333	1.2333333333333334	1017.166667
2013-01-05	6.000000	86.833333	3.6999999999999997	1016.500000
...	...	...	...	...
2016-12-28	17.217391	68.043478	3.547826086956522	1015.565217
2016-12-29	15.238095	87.857143	6.0	1016.904762
2016-12-30	14.095238	89.666667	6.266666666666667	1017.904762
2016-12-31	15.052632	87.000000	7.325	1016.100000
2017-01-01	10.000000	100.000000	0.0	1016.000000

1462 rows × 4 columns

Рисунок 3.1 - Завантаження даних

Побудуємо графік денних температур застосувавши метод pandas.Series.plot.

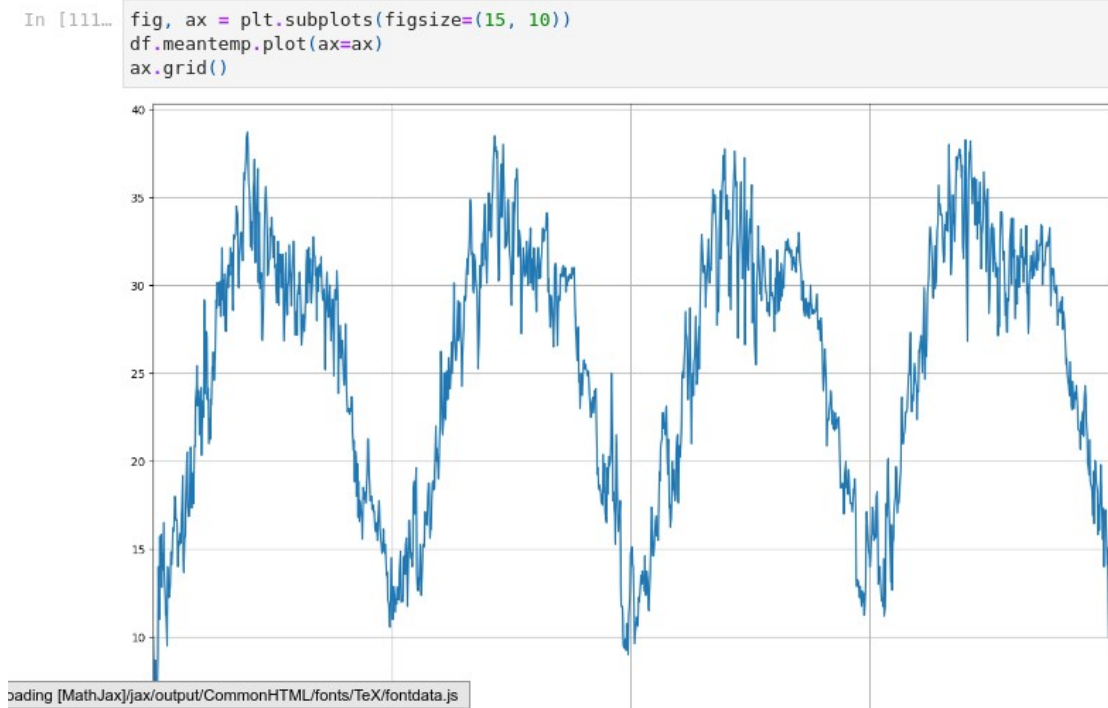
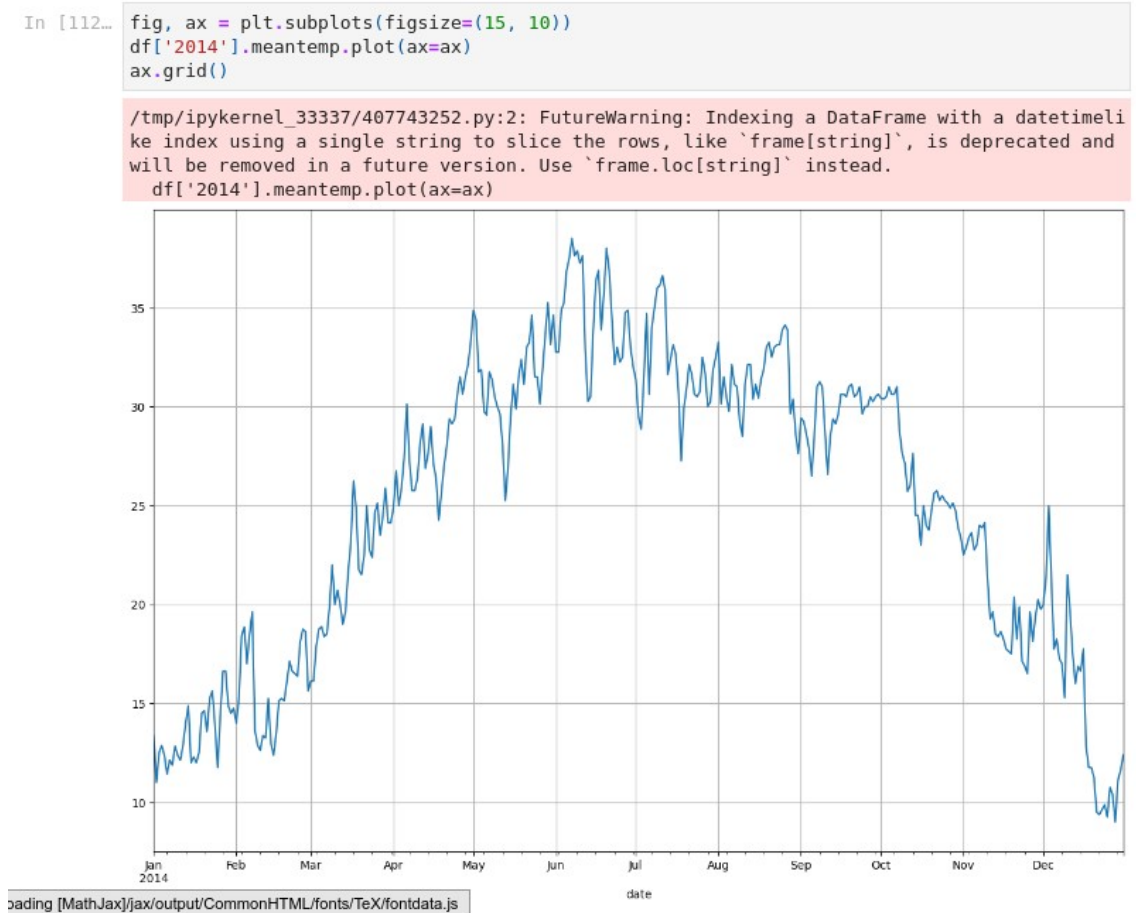


Рисунок 3.2 - Графік денних температур

Побудуємо графік температур за 2014 рік.



### Рисунок 3.3 - Графік зміни середніх денних температур за 2014 рік

Побудуємо графік зміни середніх денних температур за квітень 2013-го року.

Виконаємо знаходження потрібних даних декількома способами.

```
In [113]: april = df['2013-04']
april
```

/tmp/ipykernel\_33337/3631679302.py:1: FutureWarning: Indexing a DataFrame with a datetimelike index using a single string to slice the rows, like `frame[string]`, is deprecated and will be removed in a future version. Use `frame.loc[string]` instead.

```
april = df['2013-04']
```

Out[113]:

	meantemp	humidity	wind_speed	meanpressure
date				
2013-04-01	25.375000	45.500000	4.4	1008.500000
2013-04-02	25.166667	51.000000	8.65	1009.500000
2013-04-03	26.200000	45.600000	8.14	1009.000000
2013-04-04	24.600000	41.800000	11.120000000000001	1007.800000
2013-04-05	25.600000	31.000000	15.559999999999999	1007.000000
2013-04-06	25.857143	29.857143	11.900000000000002	1006.142857
2013-04-07	29.142857	23.285714	10.314285714285715	1005.000000
2013-04-08	28.714286	33.857143	5.3	1006.000000
2013-04-09	30.166667	30.500000	8.65	1005.333333
2013-04-10	30.000000	28.000000	6.1	1006.714286
2013-04-11	30.000000	24.200000	7.780000000000001	1006.400000
2013-04-12	28.857143	32.571429	6.342857142857143	1007.571429
2013-04-13	30.200000	29.200000	10.000000000000002	1008.400000
2013-04-14	28.250000	39.375000	6.487499999999999	1007.000000
2013-04-15	28.250000	41.375000	6.25	1003.875000
2013-04-16	32.125000	24.625000	10.424999999999999	1000.000000
2013-04-17	29.200000	24.200000	6.659999999999999	1002.200000
2013-04-18	30.285714	30.285714	4.757142857142858	1002.857143
2013-04-19	28.285714	31.285714	3.971428571428571	1002.571429
2013-04-20	30.625000	29.000000	7.6375	1003.375000
2013-04-21	27.666667	38.666667	13.883333333333333	1006.833333
2013-04-22	27.375000	45.375000	7.650000000000001	1010.000000
2013-04-23	28.625000	44.125000	4.625000000000001	1009.875000
2013-04-24	30.285714	41.714286	2.1142857142857143	1008.571429
2013-04-25	31.142857	38.285714	3.7	1007.714286
2013-04-26	29.875000	45.875000	6.7125	1008.375000
2013-04-27	31.142857	31.428571	13.485714285714286	1007.000000
2013-04-28	30.571429	28.000000	12.971428571428572	1005.428571
2013-04-29	32.125000	26.375000	7.875	1004.875000
2013-04-30	31.142857	32.000000	7.928571428571429	1004.857143

### Рисунок 3.4 - Фільтрування за допомогою масок

Побудуємо графік середніх температур за квітень 2013 року.

```
In [114... fig, ax = plt.subplots(figsize=(15, 10))
april.meantemp.plot(ax=ax)
ax.grid()
```

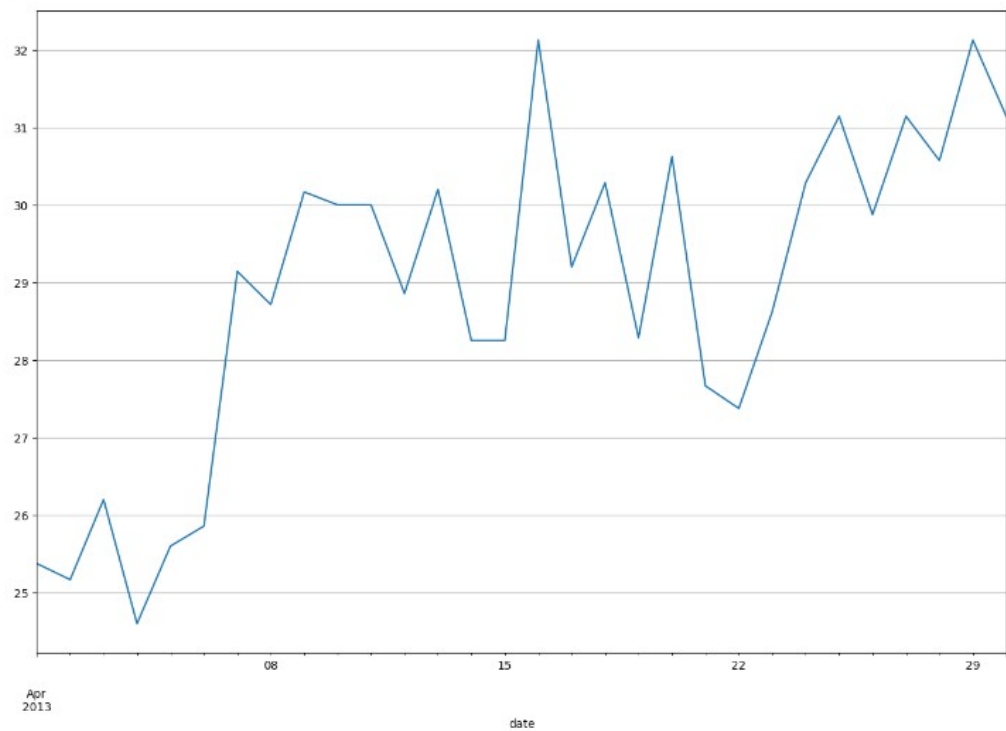


Рисунок 3.5 - Графік середніх температур за квітень 2013 року

Виконаємо знаходження даних за допомогою атрибута `loc` для побудови графіку середніх денних температур з листопада 2013-го по травень 2015-го.



```
In [115... nov_may = df.loc['2013-11':'2015-05']
fig, ax = plt.subplots(figsize=(15, 10))
nov_may.meantemp.plot(ax=ax)
ax.grid()
```

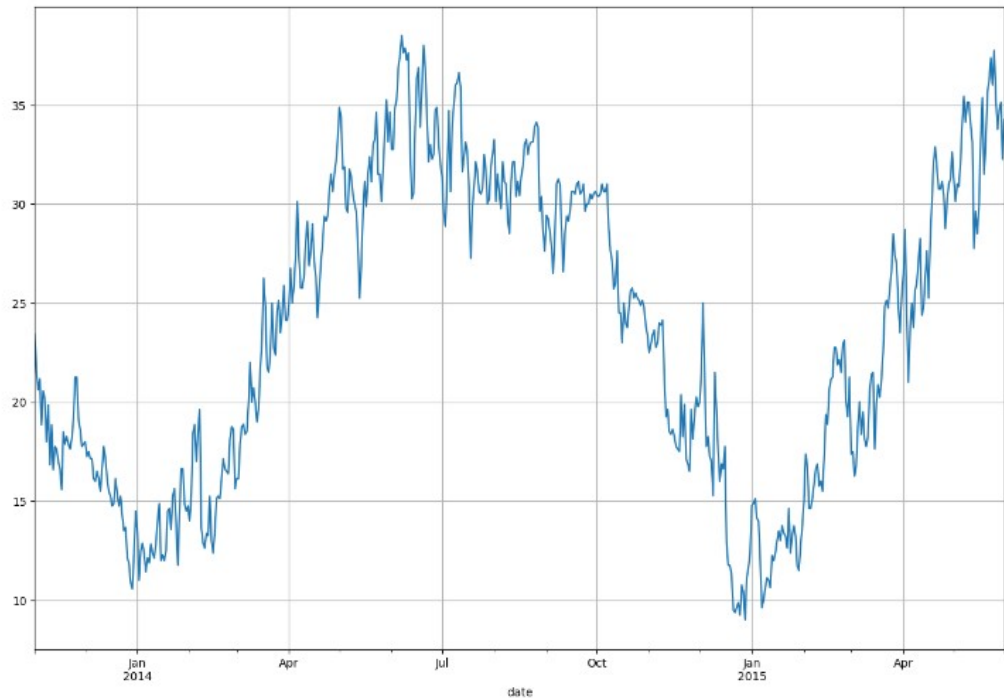


Рисунок 3.6 - Графік середніх денних температур з листопада 2013-го по травень 2015-го

Побудуємо на одному графіку дані середніх денних температур за 2015 та 2016 рік.

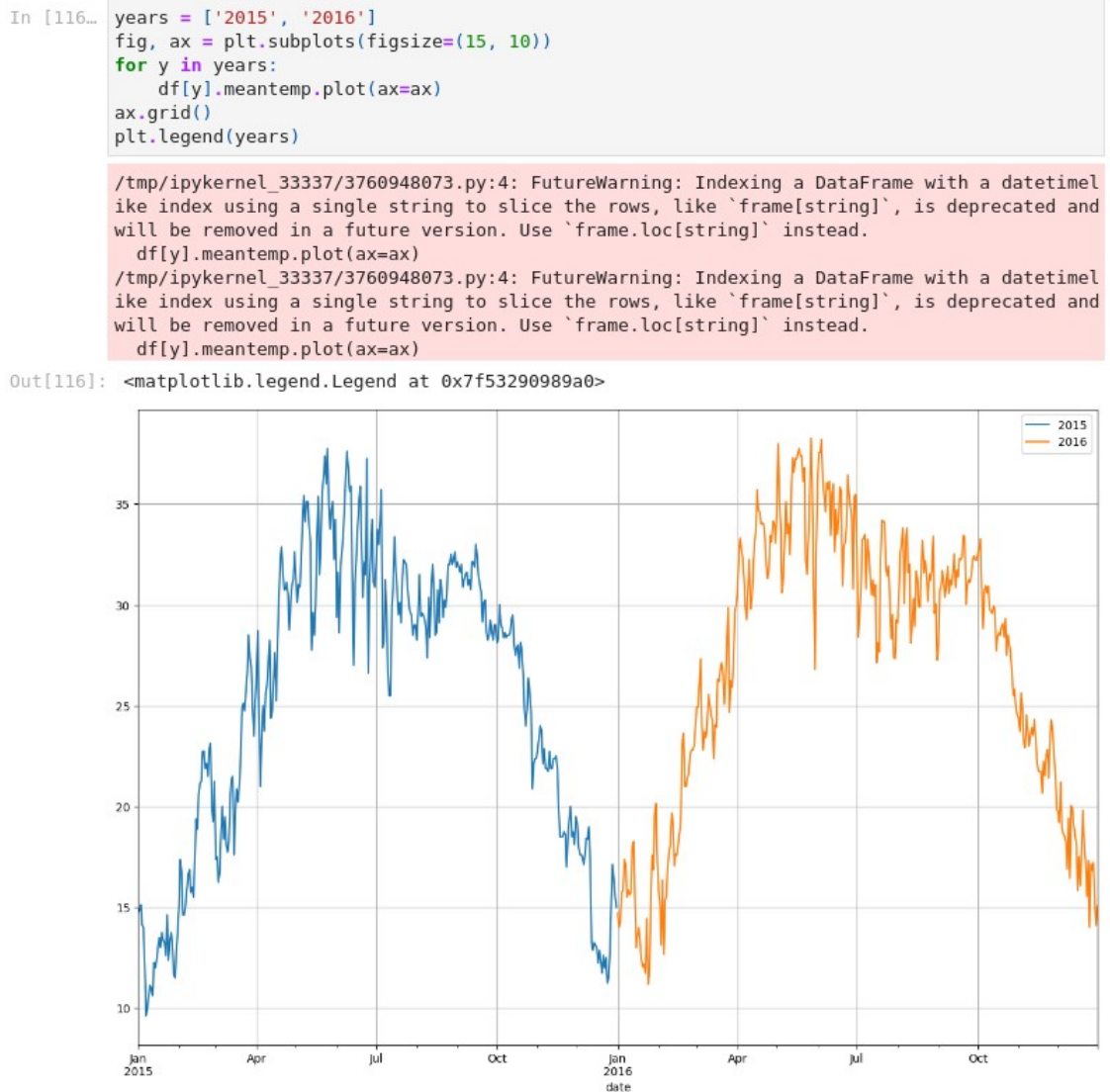


Рисунок 3.7 - Графік середніх денних температур за 2015-ий та 2016-ий роки

3.2 Знайти середні значення вологості 1. за 2016 рік; 2. за кожний місяць; 3. за кожні два тижні весни та літа 2014 року. 4. Розрахувати і зобразити зміни вологості у відсотках за кожен день впродовж літа 2015 року. 5. Знайти та зобразити графічно ковзне середнє вологості за 2013 рік з вікном в місяць.

Знайдемо середні значення вологості за 2016-ий рік.

```

In [117]: df[df.index.year == 2016].humidity.mean()

Out[117]: 58.74017392187424

```

Рисунок 3.8 - Середнє значення вологості за 2016-ий рік

Знайдемо середнє значення вологості за кожен місяць кожного року. Використаємо метод `resample` для агрегування значень у періоди.

```
In [118]: df.resample('1M').humidity.mean()

Out[118]: date
2013-01-31    73.028802
2013-02-28    71.938563
2013-03-31    57.417243
2013-04-30    34.612103
2013-05-31    28.938249
2013-06-30    58.758135
2013-07-31    74.873387
2013-08-31    76.757104
2013-09-30    64.552738
2013-10-31    70.255607
2013-11-30    66.245952
2013-12-31    79.134409
2014-01-31    84.572581
2014-02-28    73.336097
2014-03-31    62.323733
2014-04-30    40.657738
2014-05-31    41.742512
2014-06-30    42.675000
2014-07-31    62.846368
2014-08-31    61.418262
2014-09-30    63.385714
2014-10-31    59.094662
2014-11-30    55.333333
2014-12-31    69.948157
2015-01-31    78.189516
2015-02-28    64.376913
2015-03-31    68.015553
2015-04-30    51.632143
2015-05-31    33.320276
2015-06-30    49.322024
2015-07-31    69.376152
2015-08-31    70.710685
2015-09-30    53.948690
2015-10-31    59.313748
2015-11-30    66.825000
2015-12-31    71.646313
2016-01-31    77.932892
2016-02-29    62.385095
2016-03-31    55.037324
2016-04-30    29.244164
2016-05-31    38.958726
2016-06-30    51.759739
2016-07-31    75.429839
2016-08-31    72.993345
2016-09-30    61.011111
2016-10-31    56.218619
2016-11-30    54.051350
2016-12-31    68.840384
2017-01-31    100.000000
Freq: M, Name: humidity, dtype: float64
```

Рисунок 3.9 - Дані про середню вологість за кожний місяць з включно 2013 по 2018  
ВИКЛЮЧНО

Знайдемо середні значення вологості за кожні два тижні весни та літа 2014 року.

```
In [119]: df['2014-03':'2014-08'].resample('2W').mean()
```

/tmp/ipykernel\_33337/3871361821.py:1: FutureWarning: The default value of numeric\_only in DataFrameGroupBy.mean is deprecated. In a future version, numeric\_only will default to False. Either specify numeric\_only or select only columns which should be valid for the function.  
df['2014-03':'2014-08'].resample('2W').mean()

```
Out[119]:
```

	meantemp	humidity	meanpressure
date			
2014-03-02	16.133929	83.142857	1016.205357
2014-03-16	19.881378	63.723214	1015.801020
2014-03-30	23.904337	59.250000	1010.687500
2014-04-13	26.669643	41.543367	1008.798469
2014-04-27	28.258929	42.857143	1008.910714
2014-05-11	31.594388	36.492347	1003.422194
2014-05-25	30.678571	42.383929	1004.741071
2014-06-08	34.625000	36.991071	999.603316
2014-06-22	35.178571	42.982143	995.107143
2014-07-06	32.212127	59.030987	999.063884
2014-07-20	33.008929	58.241071	996.562500
2014-08-03	31.381378	67.325255	998.734694
2014-08-17	30.779562	66.380233	999.182379
2014-08-31	31.998724	55.081633	1002.774235

Рисунок 3.10 - Середні значення вологості за кожні два тижні весни та літа 2014 року

Розрахуємо і зобразимо зміни вологості у відсотках за кожен день впродовж літа 2015 року. Використаємо метод `pct_change` для знаходження зміни у відсотках.

```
In [120]: dd = df['2015-06':'2015-08'].humidity.pct_change()
```

Рисунок 3.11 - Знаходження зміни вологості у відсотках

Зобразимо вище знайдені дані.

```
In [121]: dd.plot()
```

```
Out[121]: <AxesSubplot: xlabel='date'>
```

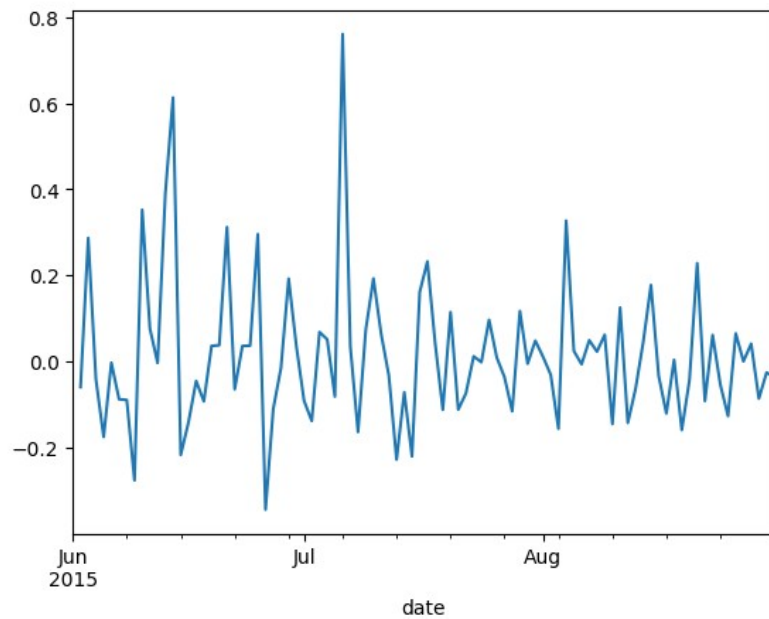


Рисунок 3.12 - Uhfascr зміни вологості у відсотках

Знайдемо та зобразимо графічно ковзне середнє вологості за 2013 рік з вікном в місяць.

```
In [123]: df['2013'].rolling(30).humidity.mean().plot()
```

```
/tmp/ipykernel_33337/309220255.py:1: FutureWarning: Indexing a DataFrame with a datetimeli
ke index using a single string to slice the rows, like `frame[string]`, is deprecated and
will be removed in a future version. Use `frame.loc[string]` instead.
  df['2013'].rolling(30).humidity.mean().plot()
```

```
Out[123]: <AxesSubplot: xlabel='date'>
```

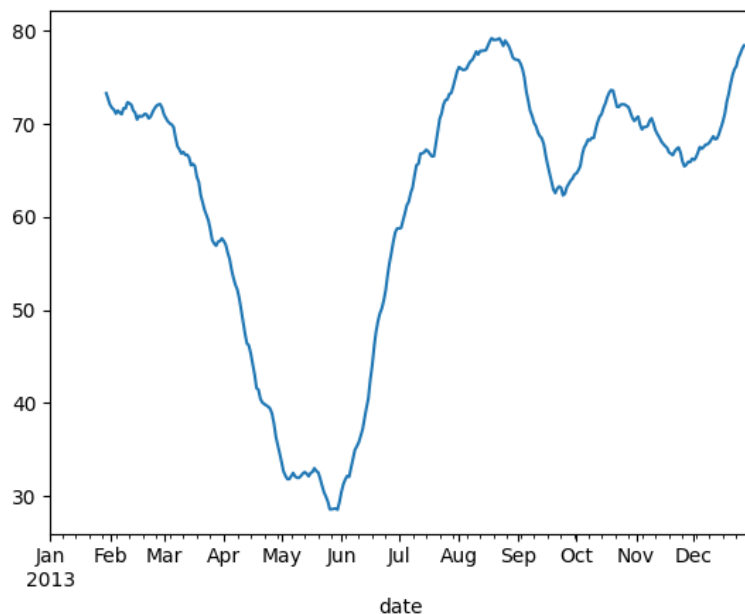


Рисунок 3.13 - Графік ковзного середнього вологості за 2013 рік з вікном в місяць

## 4 ВИСНОВОК

Під час виконання даної лабораторної роботи я ознайомився з основними представленнями часових рядів в Pandas та операціями з ними.

У завданнях було застосовано знаходження проміжків за прямою індексацією за датою, метод `resample` агрегування даних за періодами, `pct_change` для знаходження зміни вологості у відсотках, `rolling` для знаходження ковзного середнього вологості. Графіки наведені.