

Виконала студентка ІП-13 Лисенко Анастасія

Лабораторна робота №7

- Побудувати та проаналізувати часовий ряд для статистики захворювань на Covid в двох сусідніх країнах по вашому вибору (дані взяти в інтернеті).
- Побудувати та проаналізувати часовий ряд для курсу гривня/долар або гривня/євро за останні 3 роки (дані взяти в інтернеті).

Встановлюємо потрібні бібліотеки

```
In [163]: !pip install numpy pandas matplotlib statsmodels -q

WARNING: You are using pip version 21.3.1; however, version 23.1.2 is available.
You should consider upgrading via the 'D:\DA\LAB_7\env\Scripts\python.exe -m pip install --upgrade pip' command.
```

Зчитуємо файл та працюємо з його структурою, виразимо непотрібні дані та непотрібні країни. В нашому досліджі братимуть участь Україна та Польща.

```
In [164]: import pandas as pd
data = pd.read_csv("data/covid-ua-pl.csv")

# retrieving row by loc method
df = data.loc[data['location'].isin(['Ukraine', 'Poland'])]

# Select only the desired columns
df = df.loc[:, ['location', 'date', 'total_cases', 'new_cases']]

# Pivot the data to get separate columns for each country
df_pivot = df.pivot(index='date', columns='location')

# Rename the columns to include the country name
df_pivot.columns = [f'{col[0]}_{col[1].lower()}' for col in df_pivot.columns]
# We have only 2 Nans and in the beginning of our data, so we replace it with 0
df_pivot.fillna(0, inplace=True)

df_pivot = df_pivot.reset_index()
df_pivot['date'] = pd.to_datetime(df_pivot['date'])
df_pivot.to_csv('data/covid_ua_pl.csv', index=False)
df_pivot
```

Out [164]:

	date	total_cases_poland	total_cases_ukraine	new_cases_poland	new_cases_ukraine
0	2020-03-03	0.0	1.0	0.0	1.0
1	2020-03-04	1.0	1.0	1.0	0.0
2	2020-03-05	1.0	1.0	0.0	0.0
3	2020-03-06	5.0	1.0	4.0	0.0
4	2020-03-07	5.0	1.0	0.0	0.0
...
728	2022-03-01	5680034.0	5040518.0	12980.0	0.0
729	2022-03-02	5694767.0	5040518.0	14733.0	0.0
730	2022-03-03	5708827.0	5040518.0	14060.0	0.0
731	2022-03-04	5721316.0	5040518.0	12489.0	0.0
732	2022-03-05	5734041.0	5040518.0	12725.0	0.0

733 rows × 5 columns

Аналізуємо дані

```
In [165]: df_pivot.info()

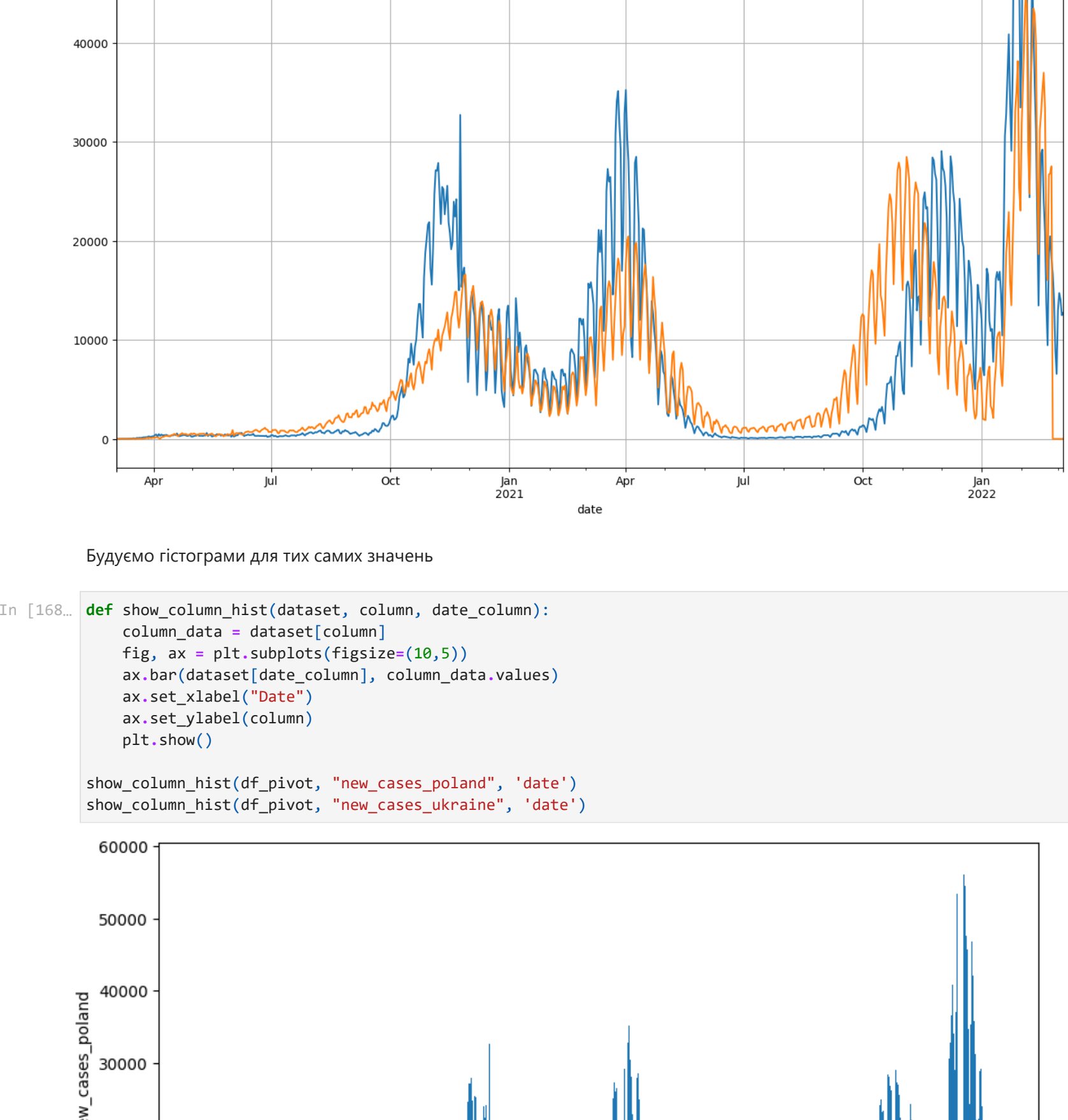
<class 'pandas.core.frame.DataFrame'>
Int64Index: 733 entries, 0 to 732
Data columns (total 5 columns):
#   Column              Non-Null Count  Dtype
---  ---
0   date                 733 non-null   datetime64[ns]
1   total_cases_poland   733 non-null   float64
2   total_cases_ukraine  733 non-null   float64
3   new_cases_poland     733 non-null   float64
4   new_cases_ukraine    733 non-null   float64
dtypes: datetime64[ns](1), float64(4)
memory usage: 28.8 KB
```

In [166]: df_pivot.describe()

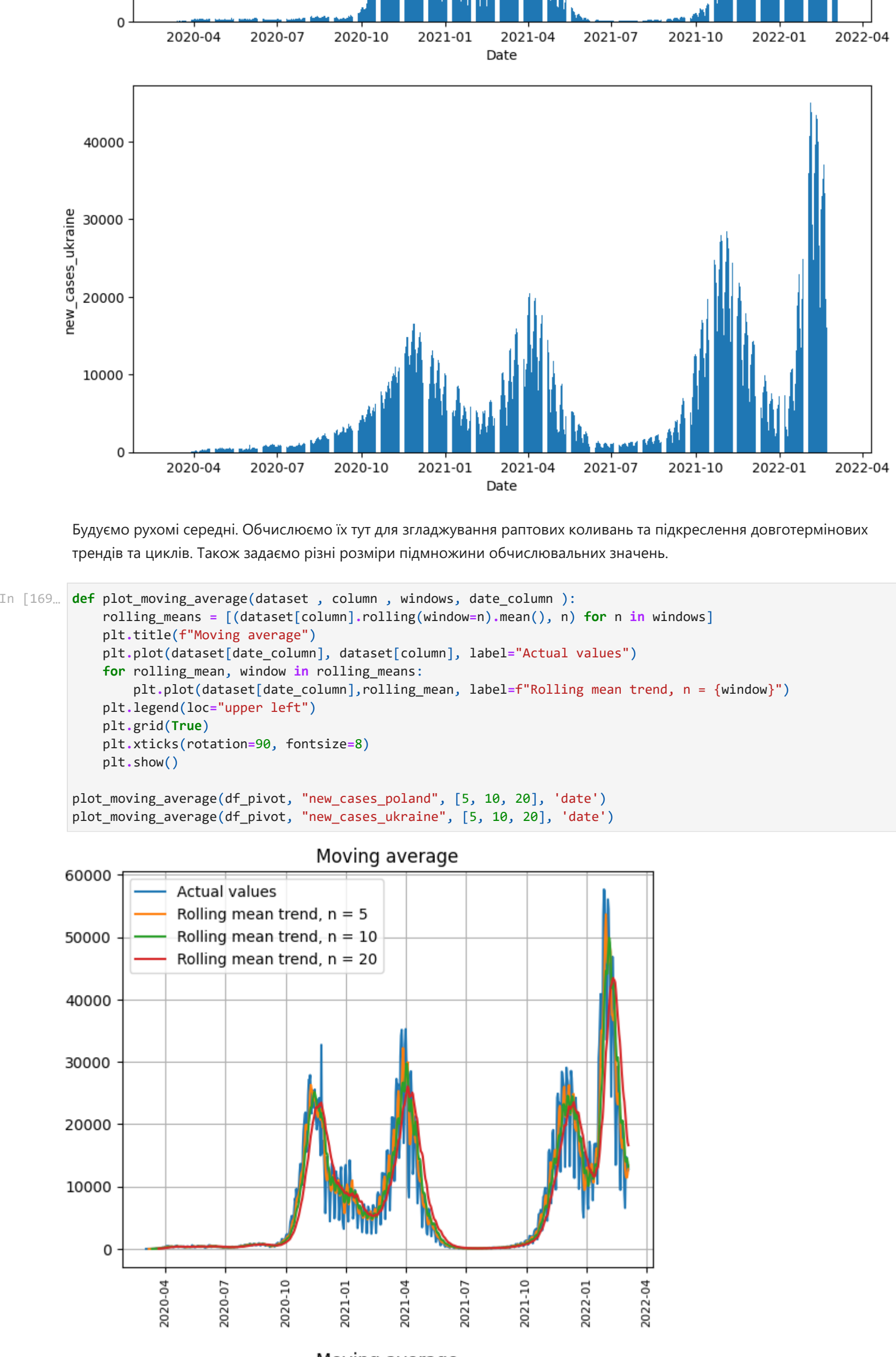
Out [166]:

	date	total_cases_poland	total_cases_ukraine	new_cases_poland	new_cases_ukraine
count	733	7.330000e+02	7.330000e+02	733.000000	733.000000
mean	2021-03-04 00:00:00.0000000256	1.859455e+06	1.619035e+06	7822.702592	6876.559345
min	2020-03-03 00:00:00	0.000000e+00	1.000000e+00	0.000000	0.000000
25%	2020-09-02 00:00:00	6.851700e+04	1.288330e+05	352.000000	1048.000000
50%	2021-03-04 00:00:00	1.750659e+06	1.422527e+06	2236.000000	3889.000000
75%	2021-09-03 00:00:00	2.889773e+06	2.393002e+06	12725.000000	10138.000000
max	2022-03-05 00:00:00	5.734041e+06	5.040518e+06	57659.000000	45022.000000
std	NaN	1.604101e+06	1.428374e+06	10537.865460	8075.914510

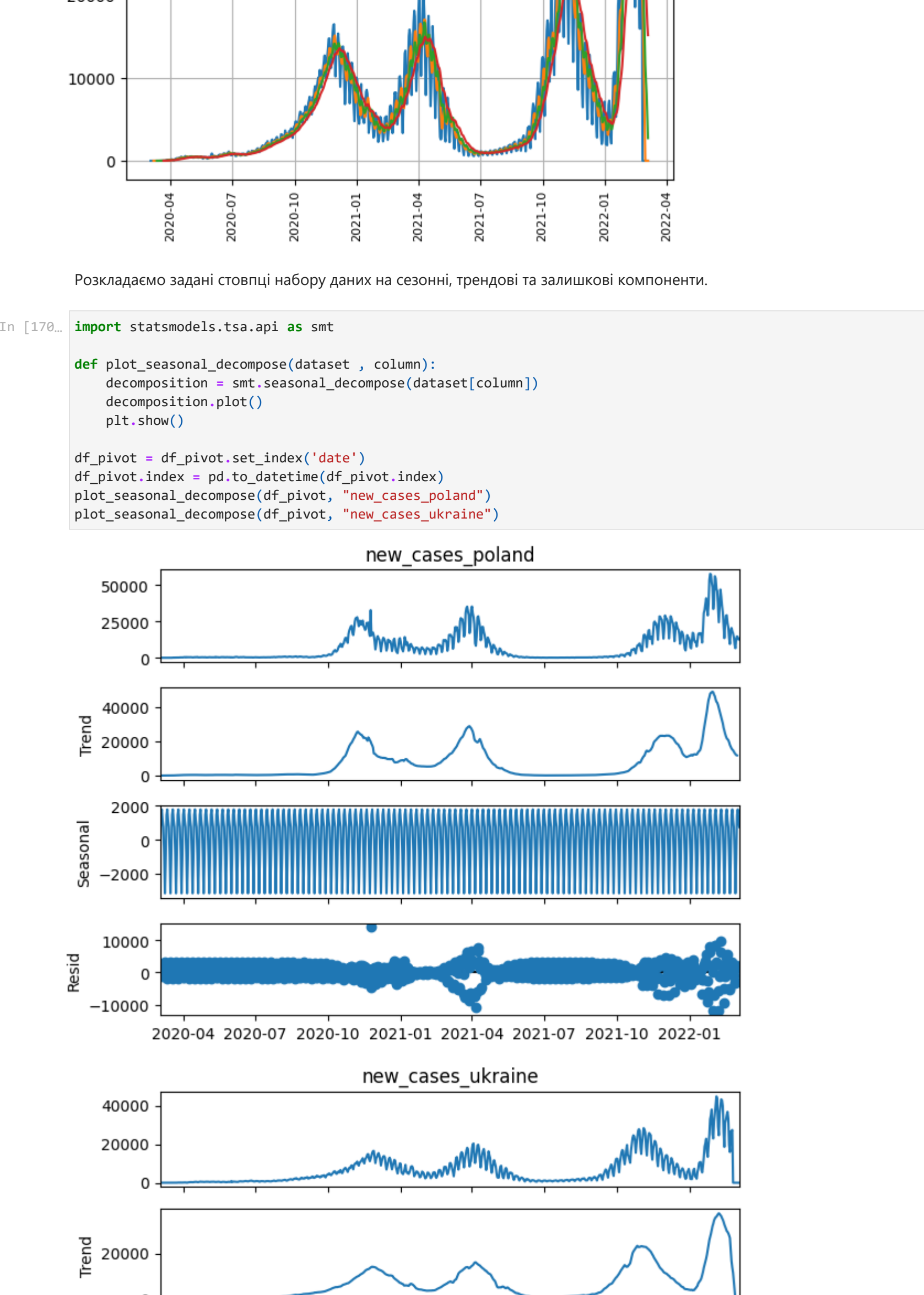
Будемо лінійний графік із двома лініями, однією для стовпця «new_cases_poland», а іншою — для стовпця «new_cases_ukraine», проти значень осі X зі стовпця «date».



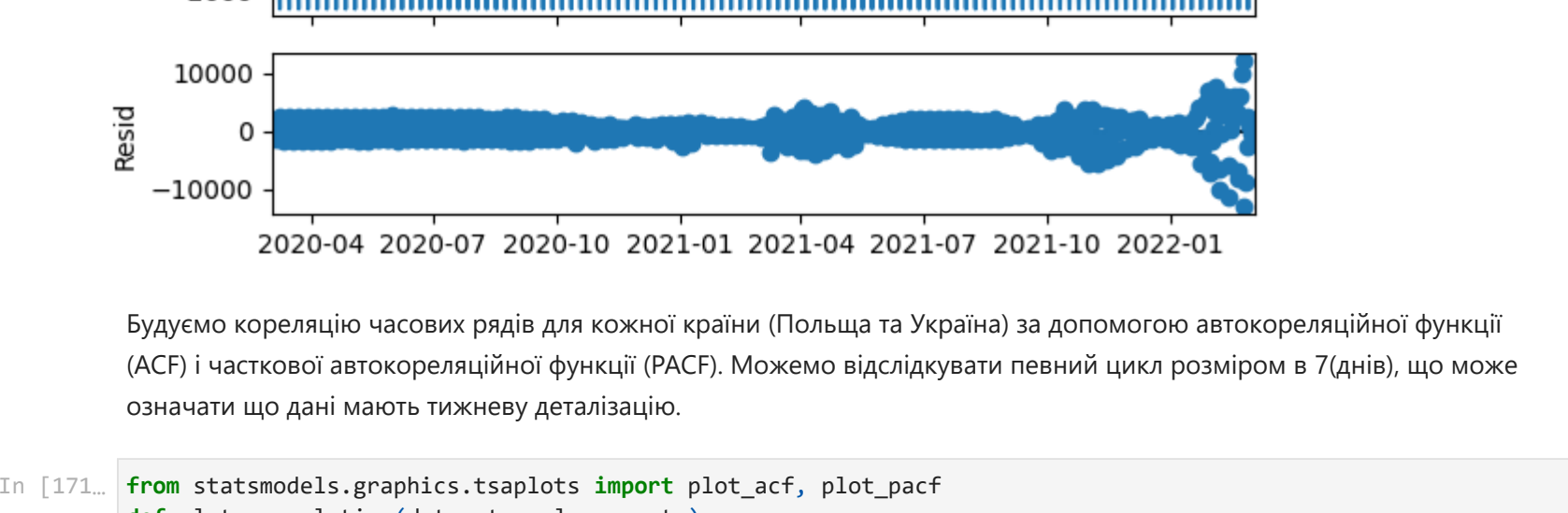
Будемо гістограми для тих самих значень



Будемо рухомі середні. Обчислюємо їх тут для згладжування раптових коливань та підкреслення довготермінових трендів та циклів. Також задаємо різні розміри підмножини обчислювальних значень.



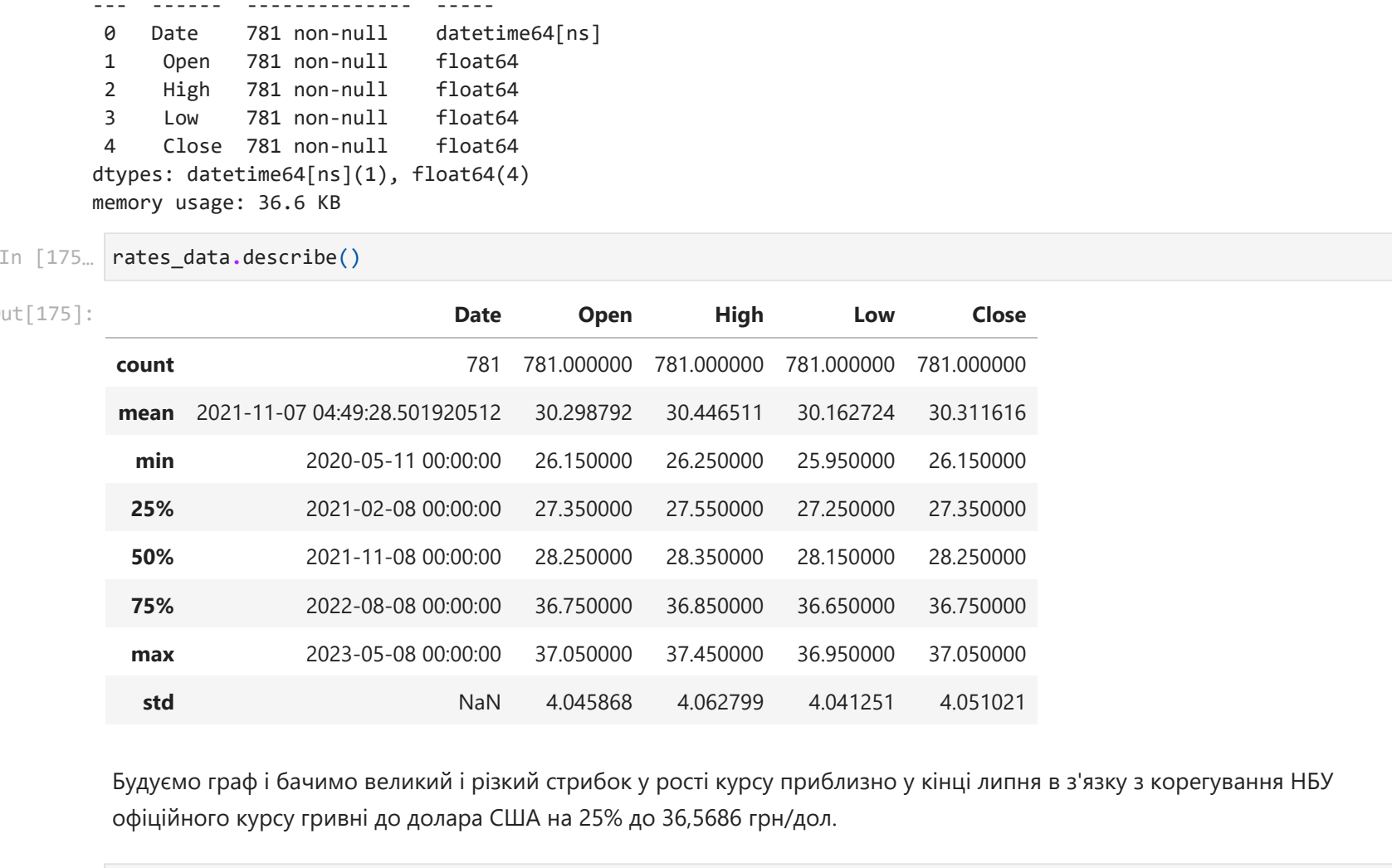
Розкладемо задані стовпці набору даних на сезонні, трендові та залишкові компоненти.



Будемо кореляцію часових рядів для кожної країни (Польща та Україна) за допомогою автокореляційної функції (ACF) і часткової автокореляційної функції (PACF). Можемо відслідкувати певний цикл розміром в 7(днів), що може означати що дані мають тижневу деталізацію.



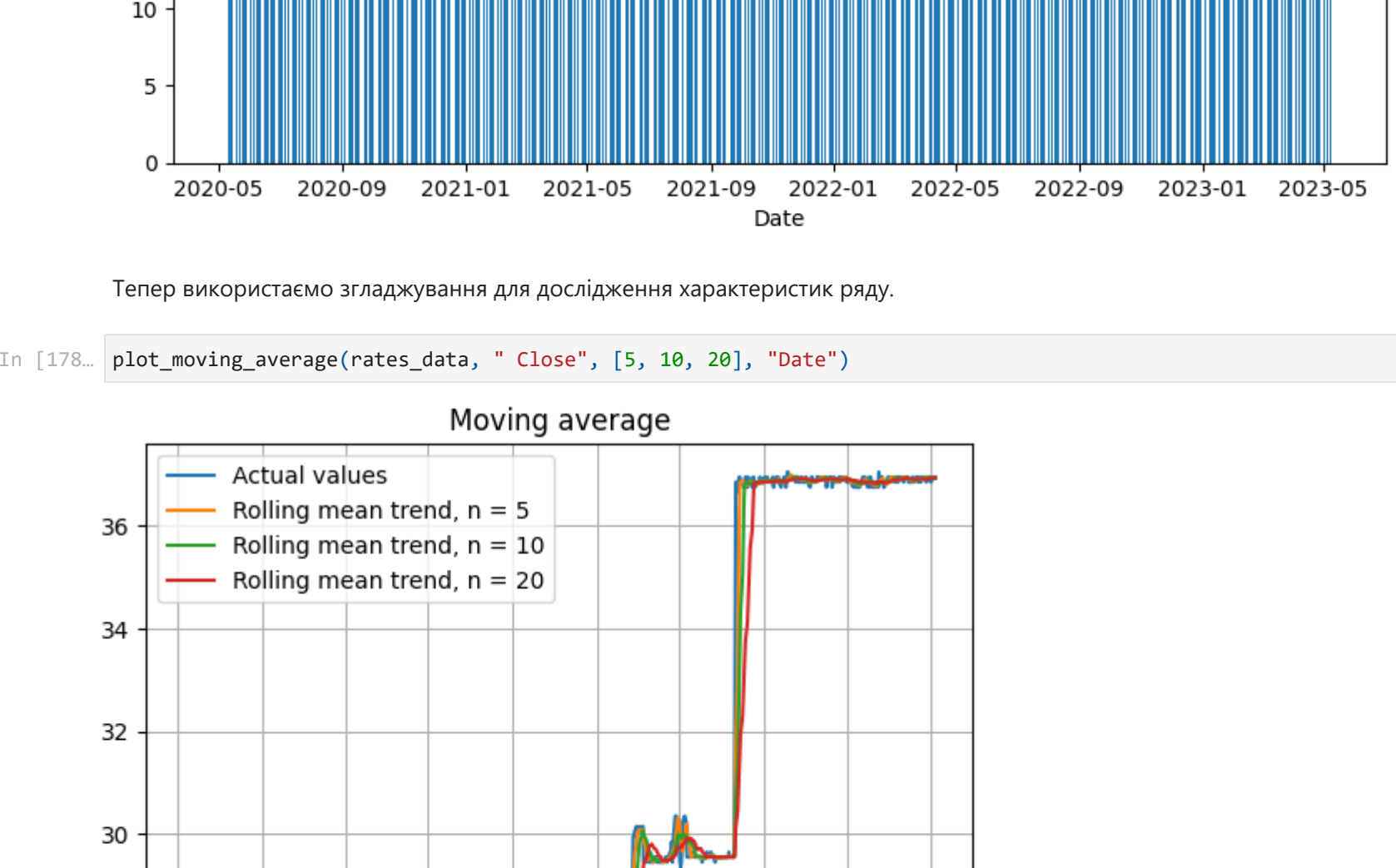
Виконуємо тест Дікі-Фуллера на стаціонарність та бачимо, що обидва набори даних стаціонарні.



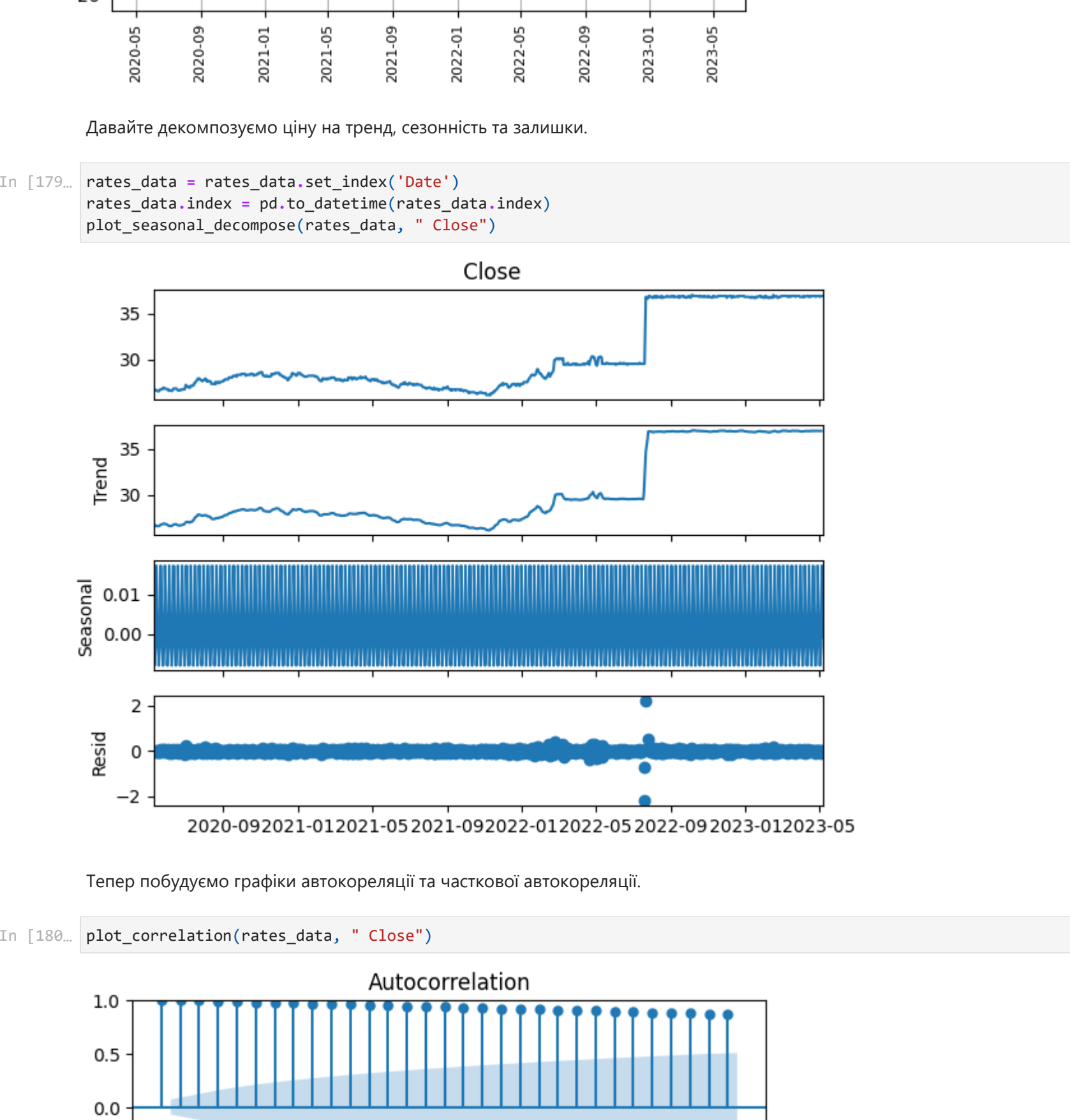
Аналізуємо дані



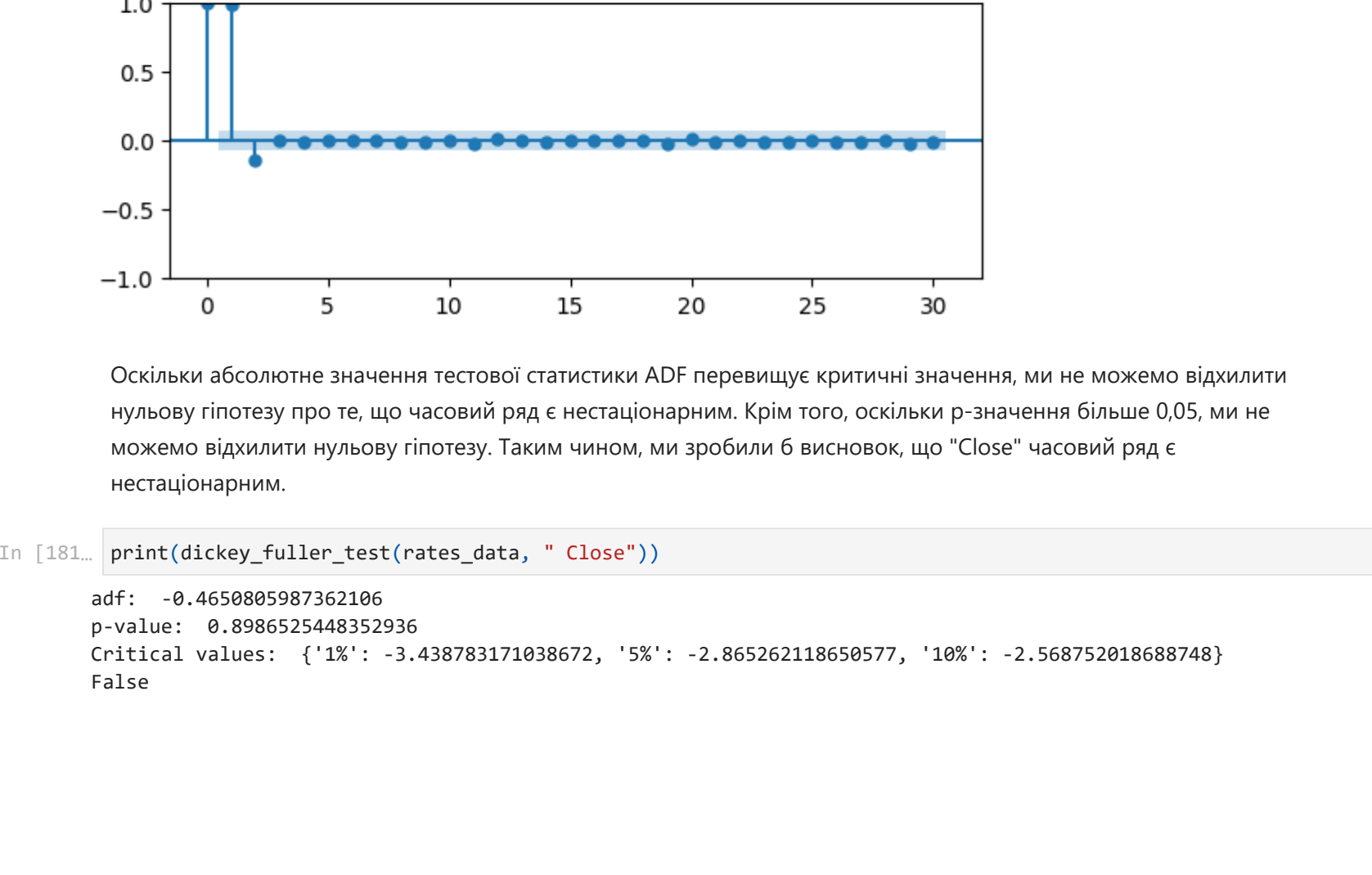
Будемо граф і бачимо великий і різкий стрибок у рості курсу приблизно у кінці липня в з'язку з корегування НБУ офіційного курсу гривні до долара США на 25% до 36.5686 грн/дол.



При побудові гістограм бачимо таку ж динаміку



Тепер використаємо згладжування для дослідження характеристик ряду.



Давайте декомпозируємо ціну на тренд, сезонність та залишки.

Тепер побудуємо графіки автокореляції та часткової автокореляції.

Оскільки абсолютне значення тестової статистики ADF перевищує критичні значення, ми не можемо відхилити нульову гіпотезу про те, що часовий ряд є нестационарним. Крім того, оскільки p-значення більше 0.05, ми не можемо відхилити нульову гіпотезу. Таким чином, ми зробили б висновок, що "Close" часовий ряд є нестационарним.

