Міністерство освіти і науки України

Національний технічний університет України «Київський політехнічний

інститут імені Ігоря Сікорського"

Факультет інформатики та обчислювальної техніки

Катедра інформатики та програмної інженерії

Звіт

з комп’ютерного практикуму № 7 з дисципліни

«Аналіз даних в інформаційних системах»

«Аналіз часових послідовностей»

Виконав студент ІП-11 Лесів Владислав Ігорович

Перевірила Ліхоузова Тетяна Анатоліївна

Київ 2023

**Комп’ютерний практикум 7**

**Аналіз часових послідовностей**

**Мета роботи:** ознайомитись з методами моделювання часових послідовностей.

**Основне завдання**.

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

**Додаткове завдання.**

Потрібно з'ясувати, чи є сезонна компонента в кількості опадів в Сіетлі:

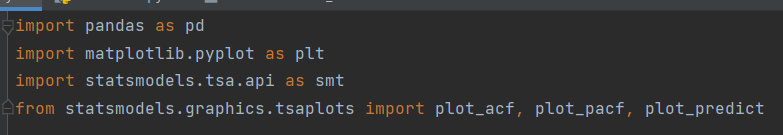
1. Градуси перевести в Цельсії.
2. Чи є кореляція між температурою та опадами?
3. Скласти прогноз опадів на 2018 рік, оцінити точність прогнозу.

**Хід роботи.**

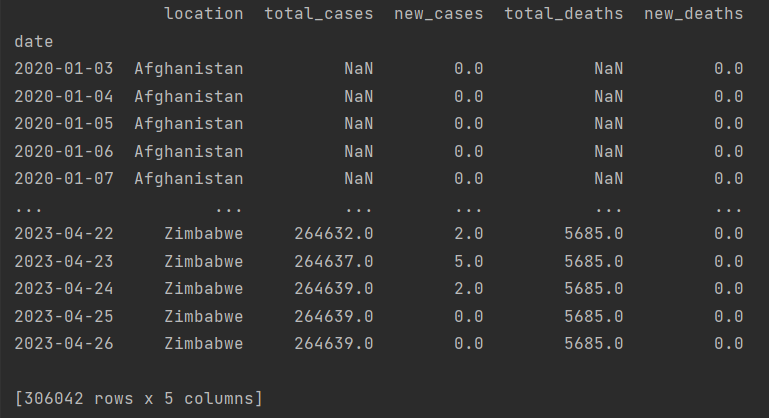
**Основне завдання.**

1. Побудувати та проаналізувати часовий ряд для статистики захворювань на Covid в двох сусідніх країнах по вашому вибору (дані взяти в інтернеті);

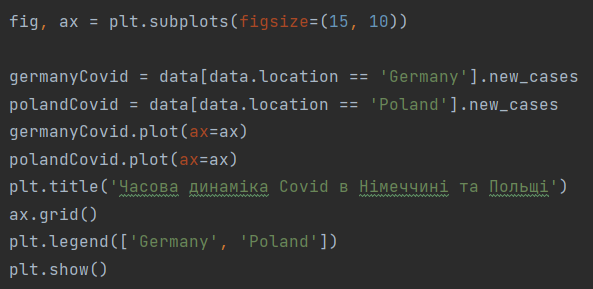
Для аналізу я обрав Німеччину та Польщу. Завантажую датасет і вводжу фактори, які можуть знадобитися для аналізу.

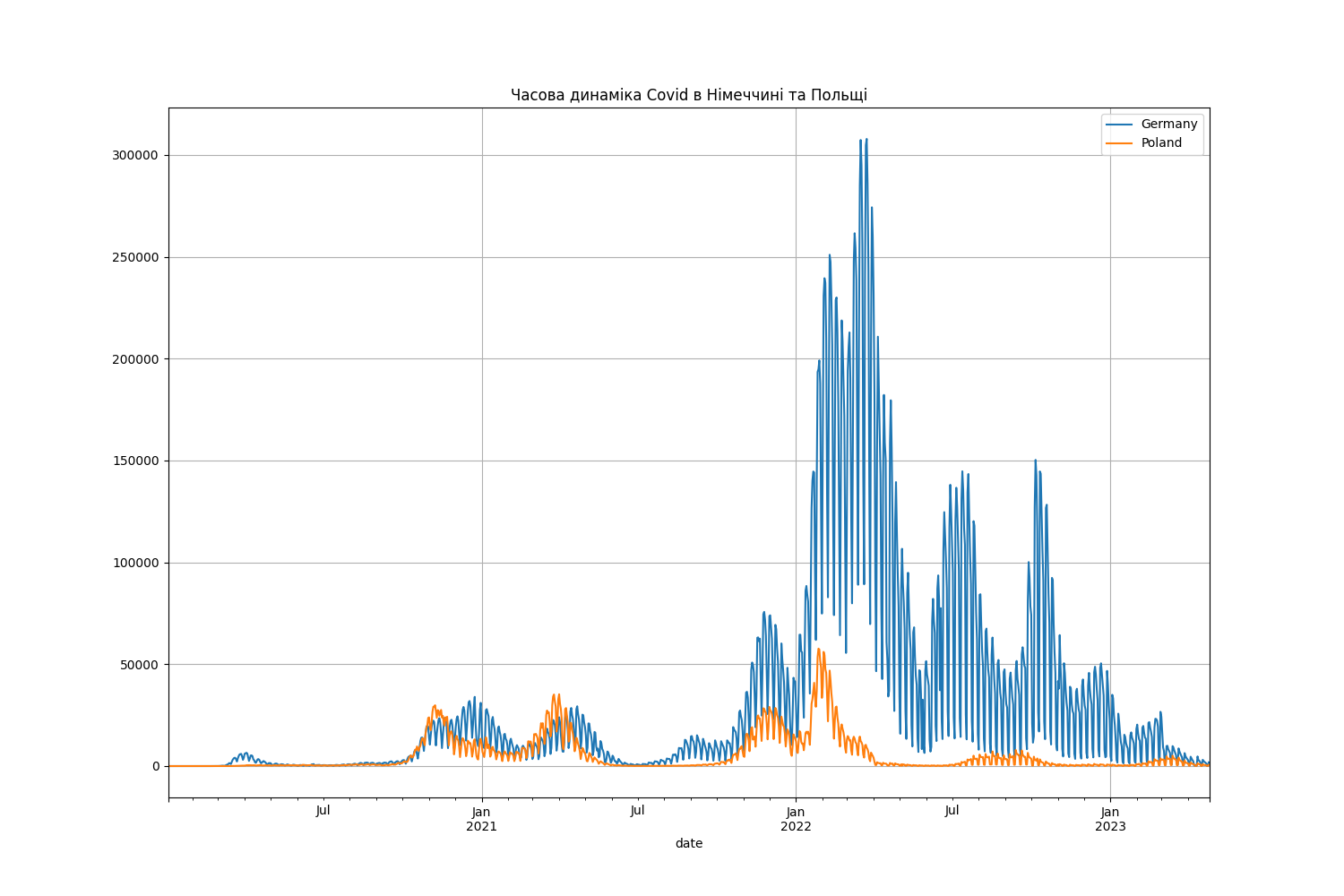






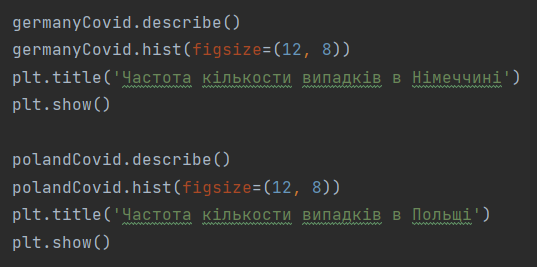
Далі візуалізую динаміку Covid в часі в Німеччині та Польщі.

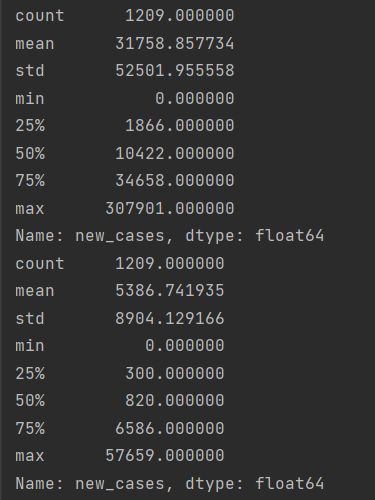




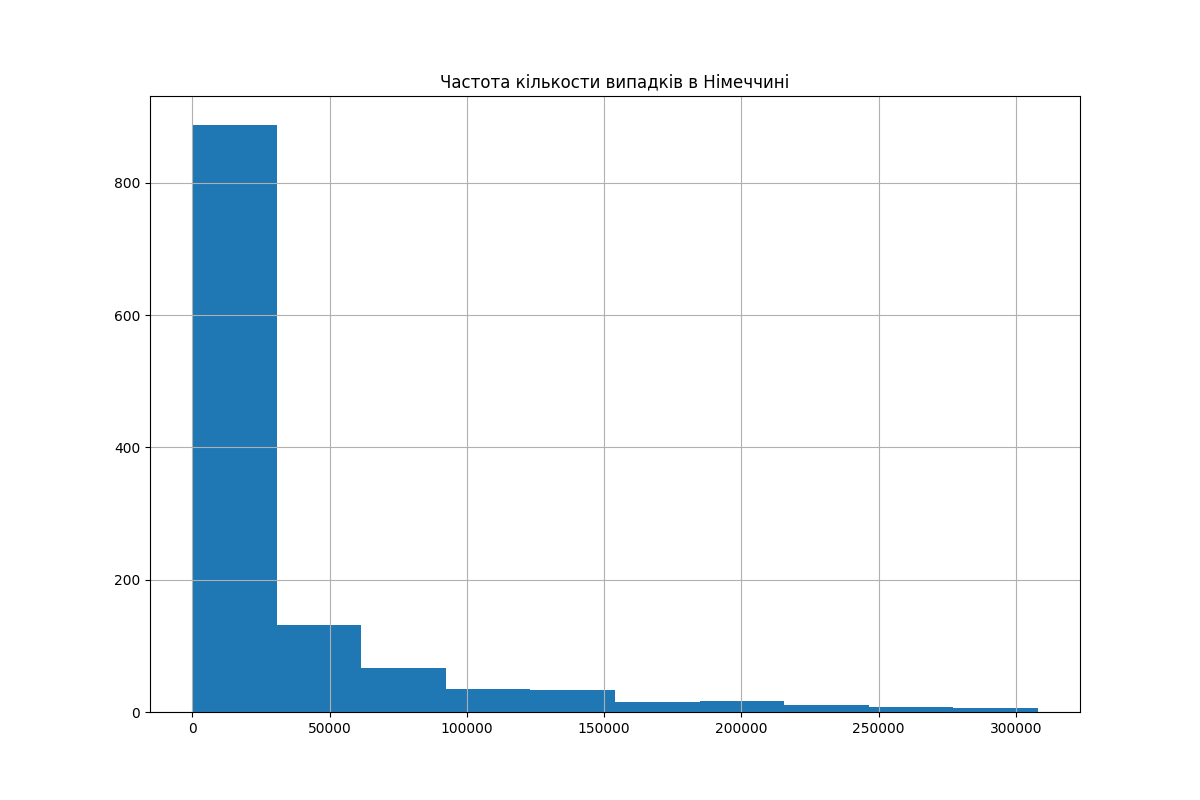
Уже можна побачити, що збільшення і зменшення захворюваності відбувається в той самий період в обох країнах, проте у Польщі кількість випадків набагато менша.

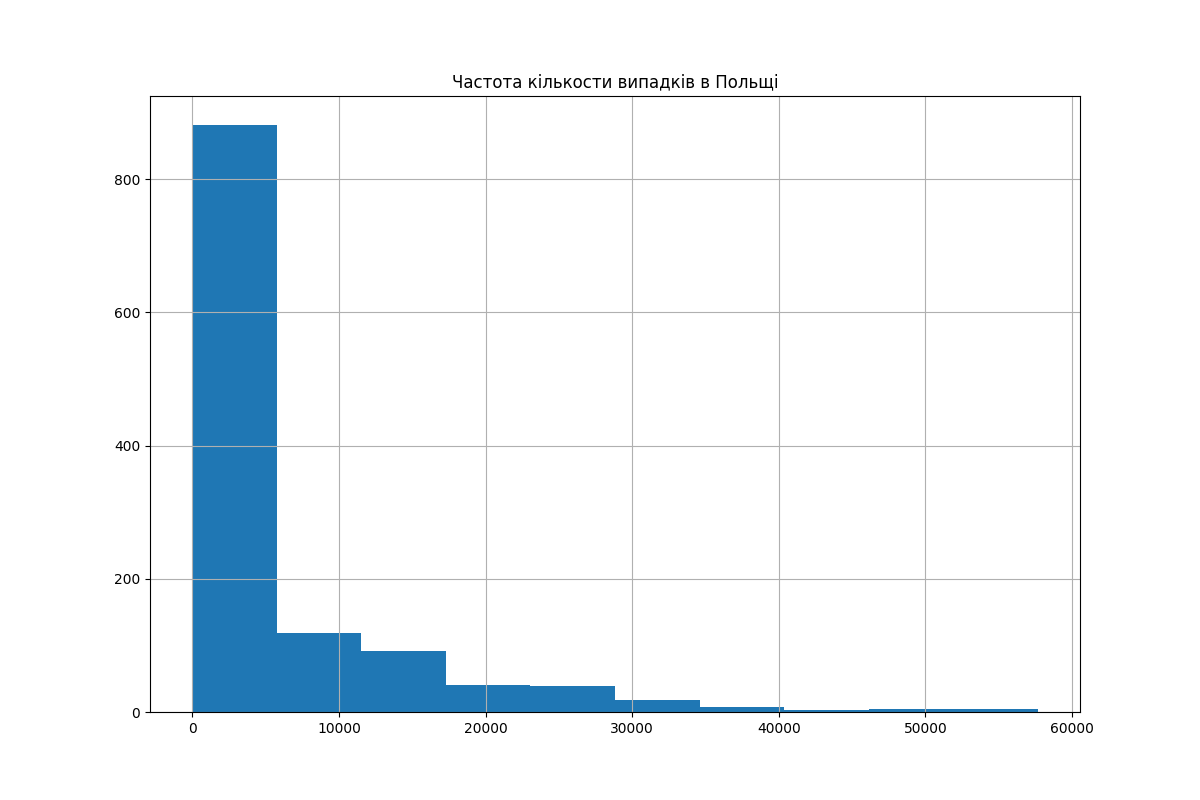
Далі досліджую часові ряди країн окремо.



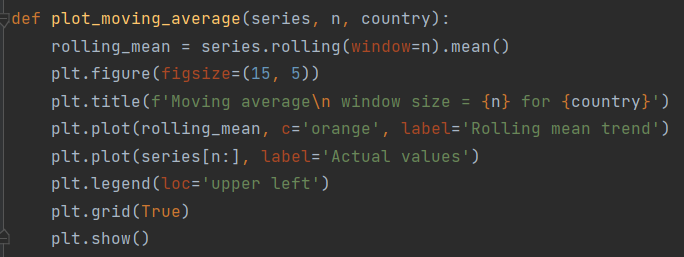


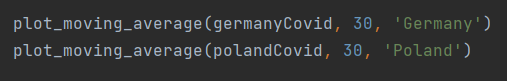
Переконуємося, що середня кількість нових випадків у Німеччині в 6 разів більша.

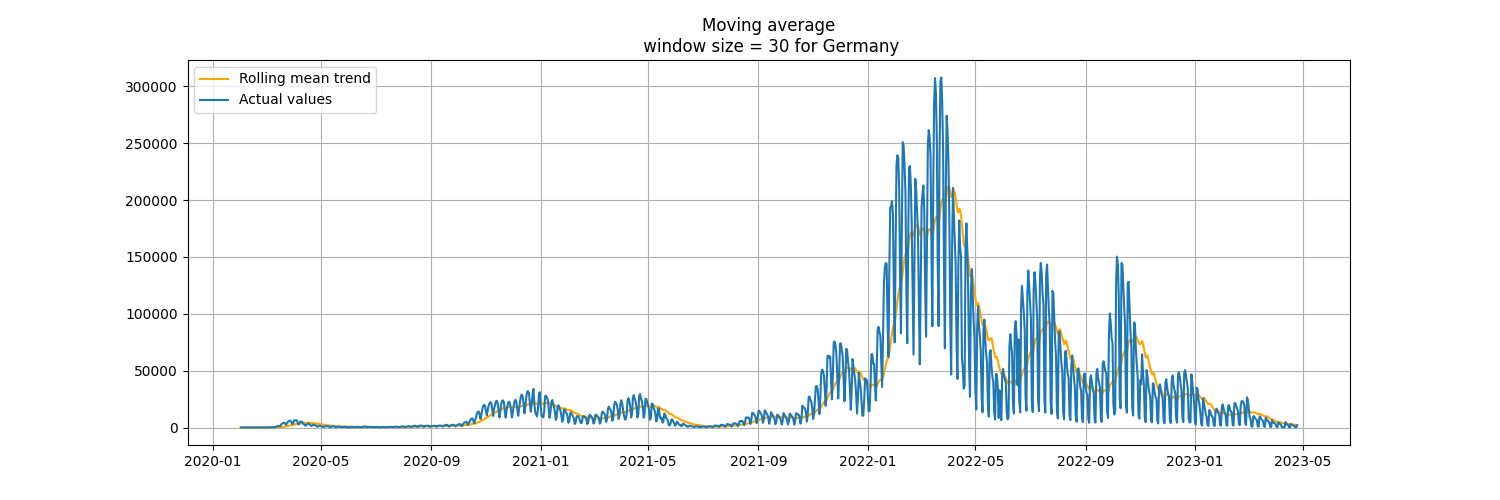


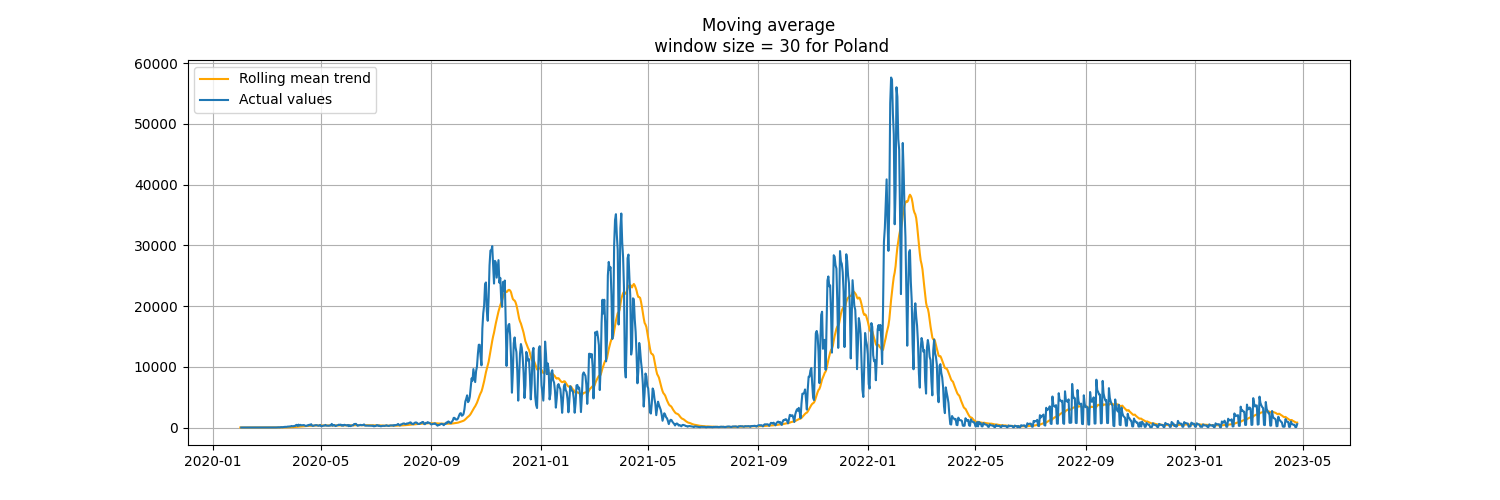


Для кращої візуалізації властивостей ряду застосовую згладжування за допомогою ковзаючого середнього.



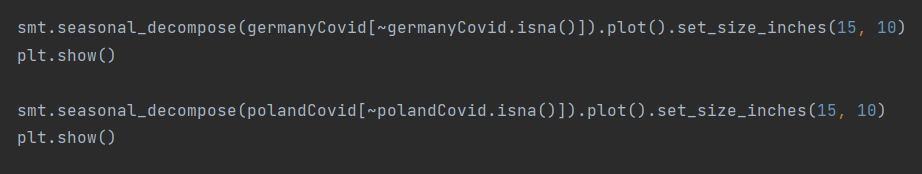


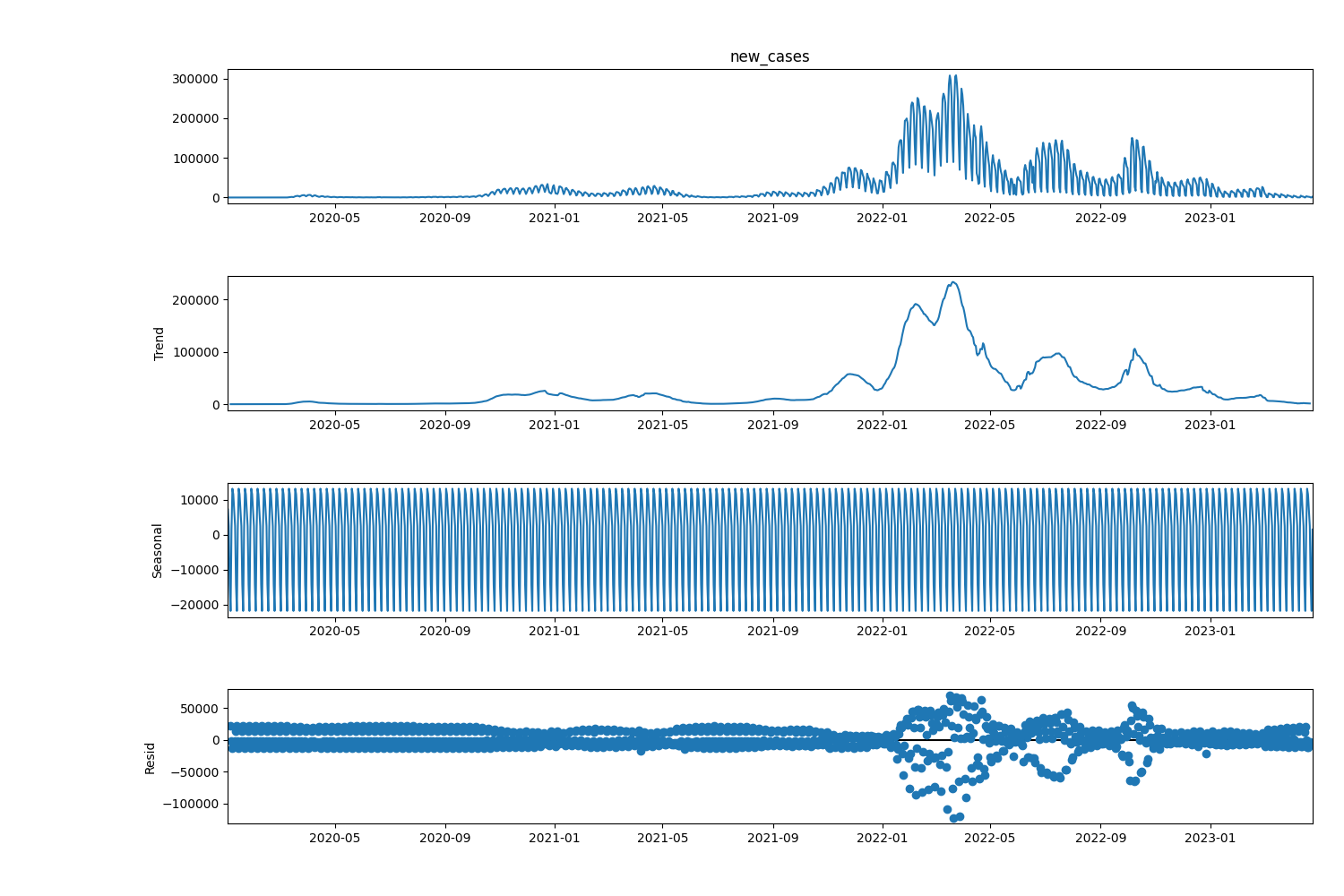


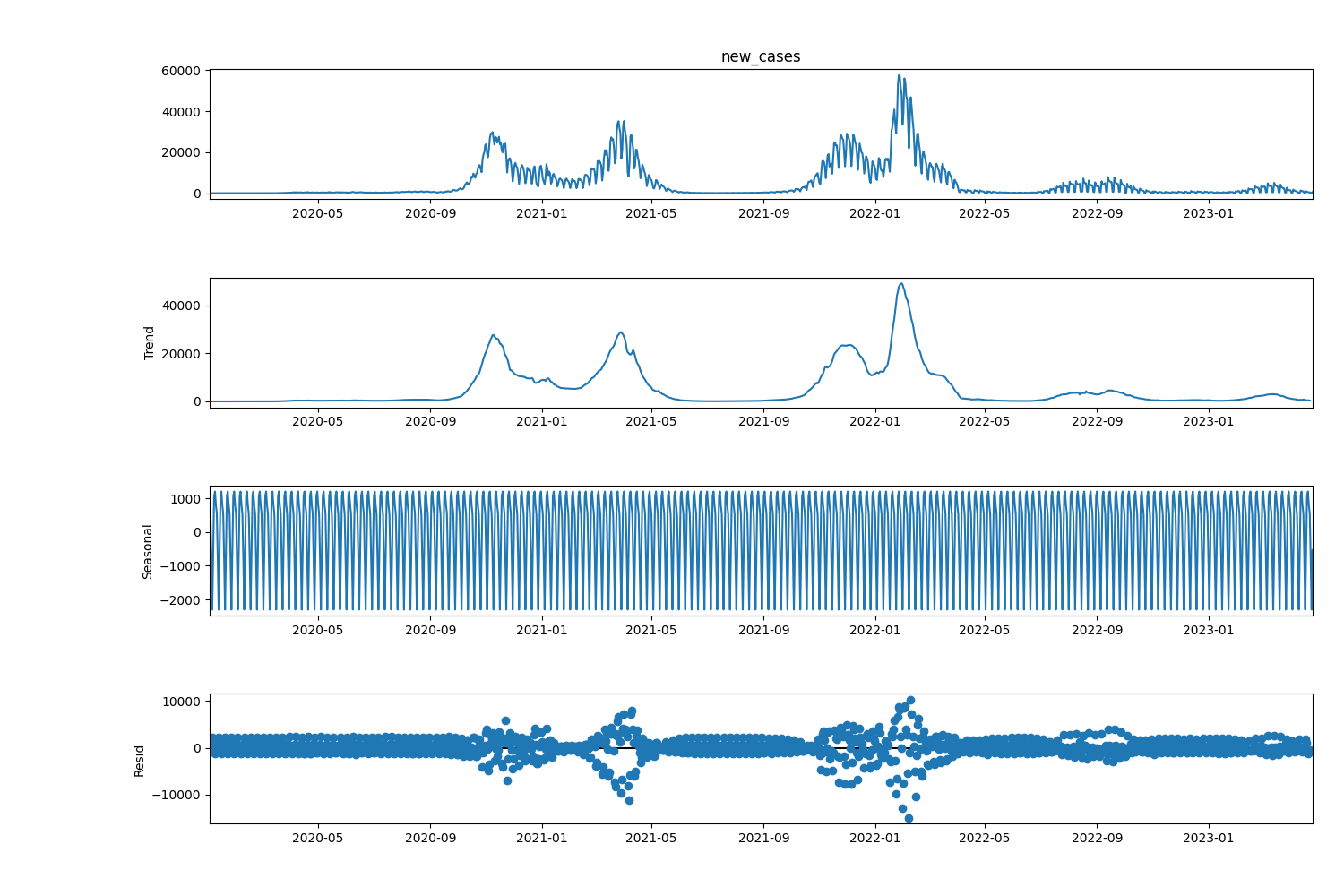


Можемо чітко побачити тренд на зростання захворюваности – кінець осени – кінець весни.

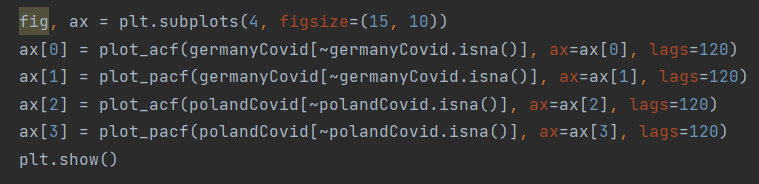
Візуалізую декомпозицію ряду на тренд, сезонність та залишки.

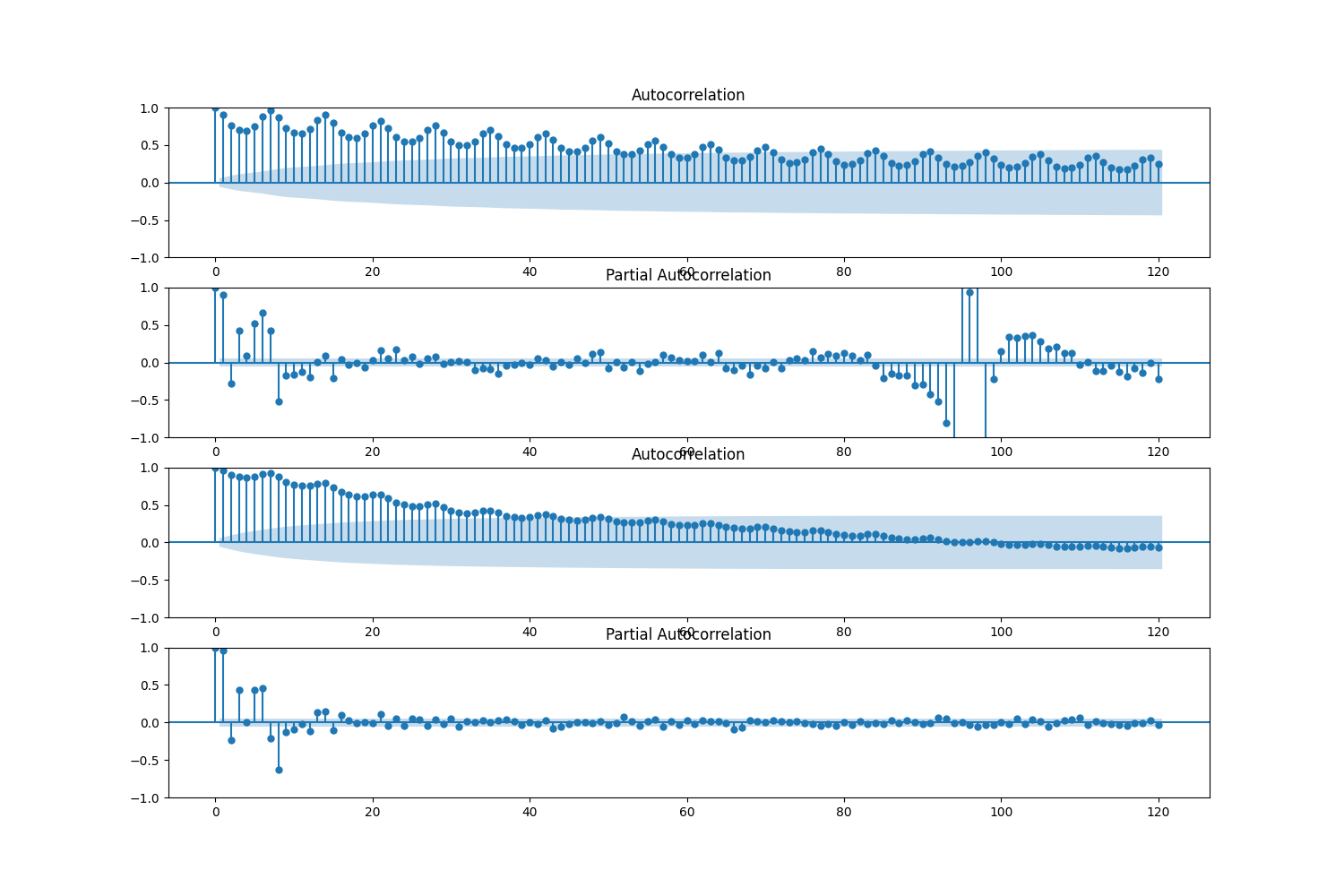




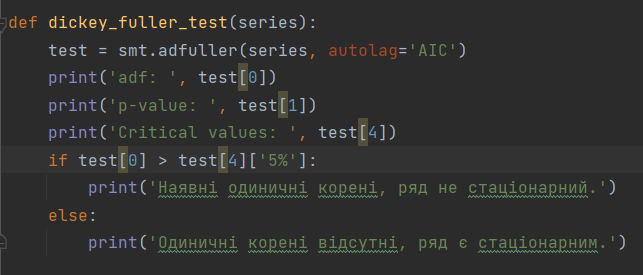


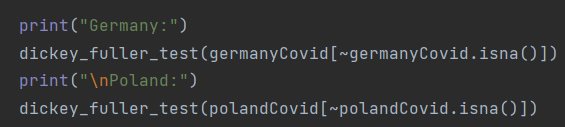
Будую графіки автокореляції та часткової автокореляції.

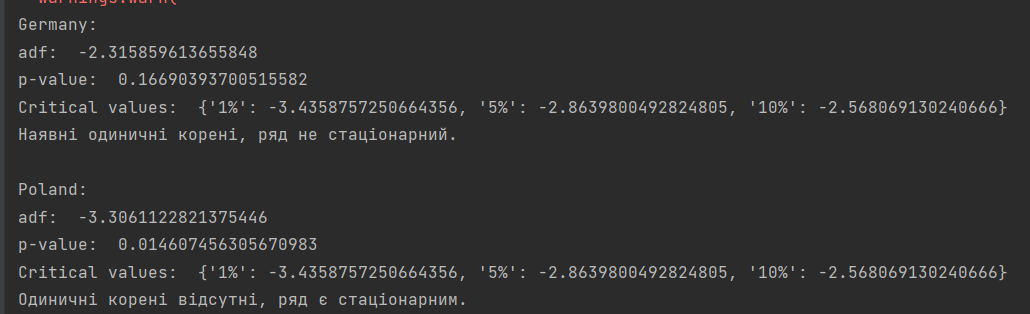




Перевіряю ряд на стаціонарність за допомогою доповненого тесту Дікі-Фуллера.







Бачимо, що ряд захворюваности в Німеччині не є стаціонарним, а в Польщі – є.

Виводжу коефіцієнт кореляції між захворюваністю в країнах.

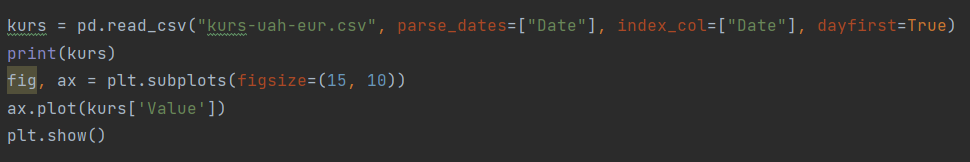


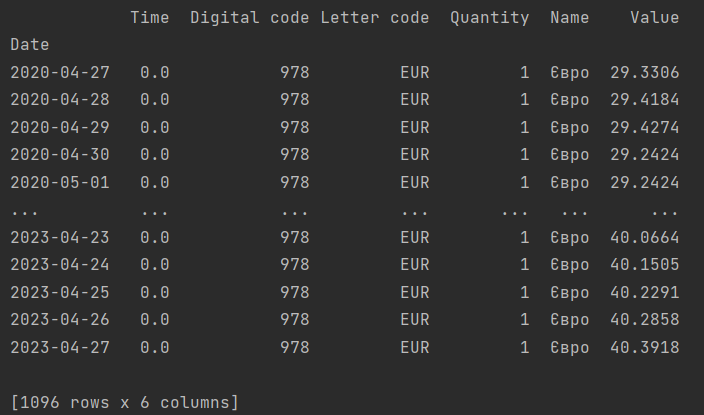


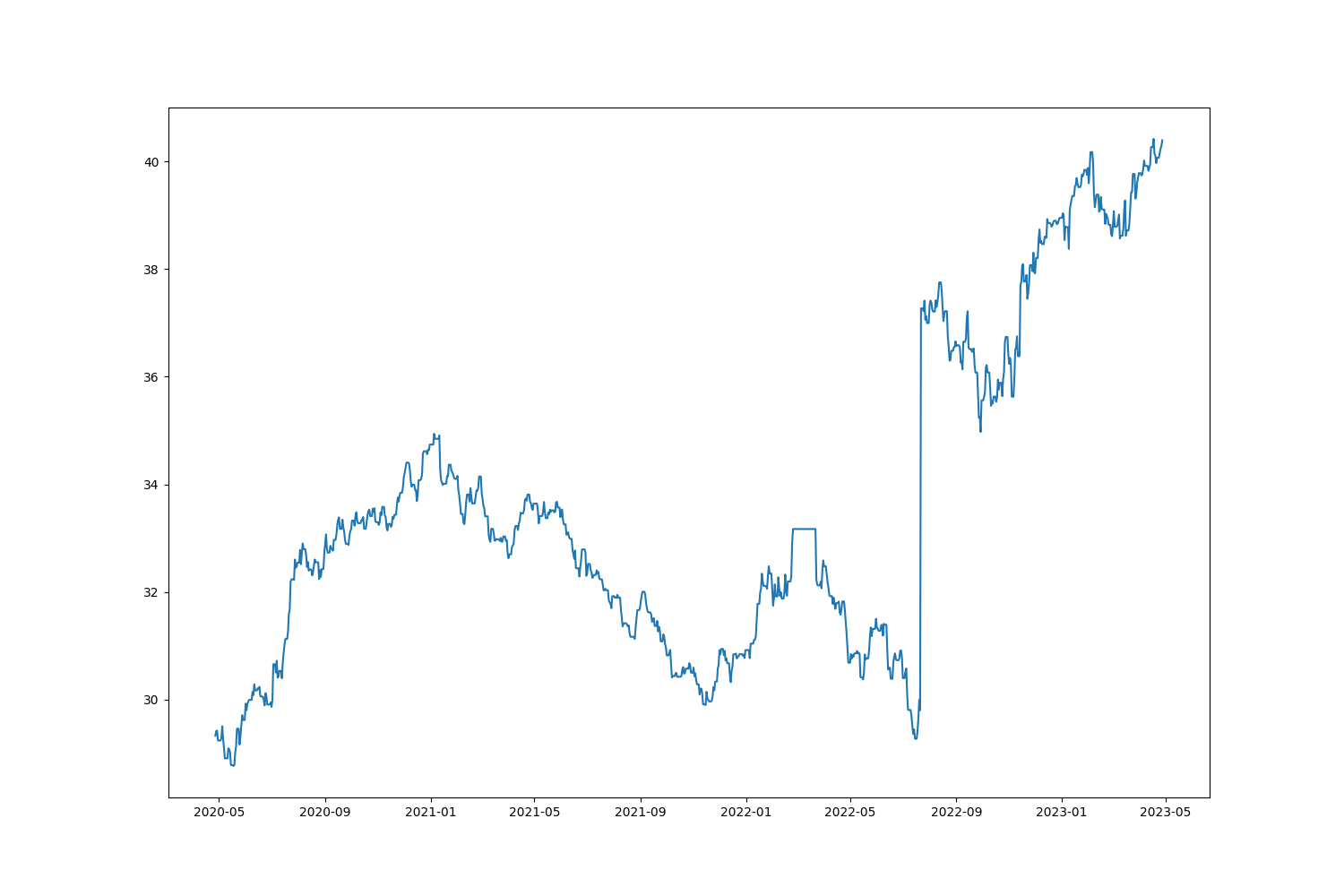
Коефіцієнт кореляції не дуже високий, тож захворюваності помірно корелюють, і це не дивно, адже це дві сусідні країни і ситуації в них мають в чомусь бути схожими.

1. Побудувати та проаналізувати часовий ряд для курсу гривня/долар або гривня/євро за останні 3 роки (дані взяти в інтернеті).

Завантажую датасет курсу гривня/євро і візуалізую його.

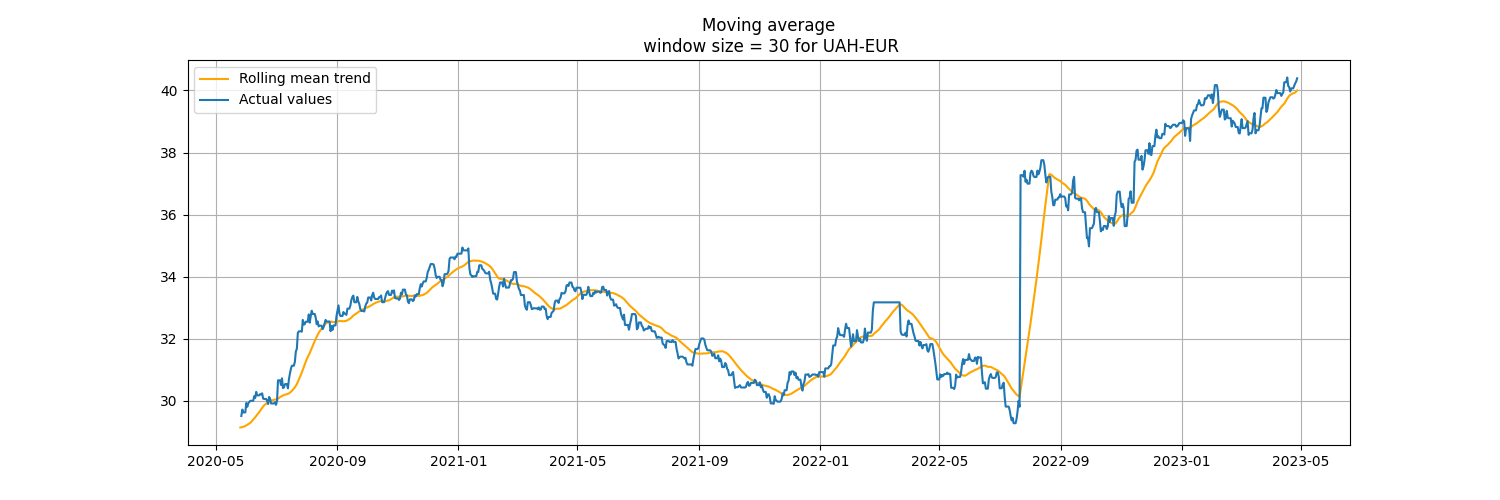






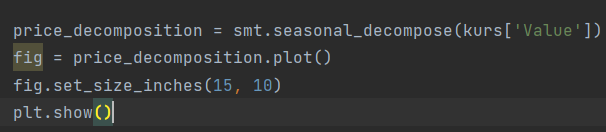
Застосовую згладжування за допомогою ковзаючого середнього.

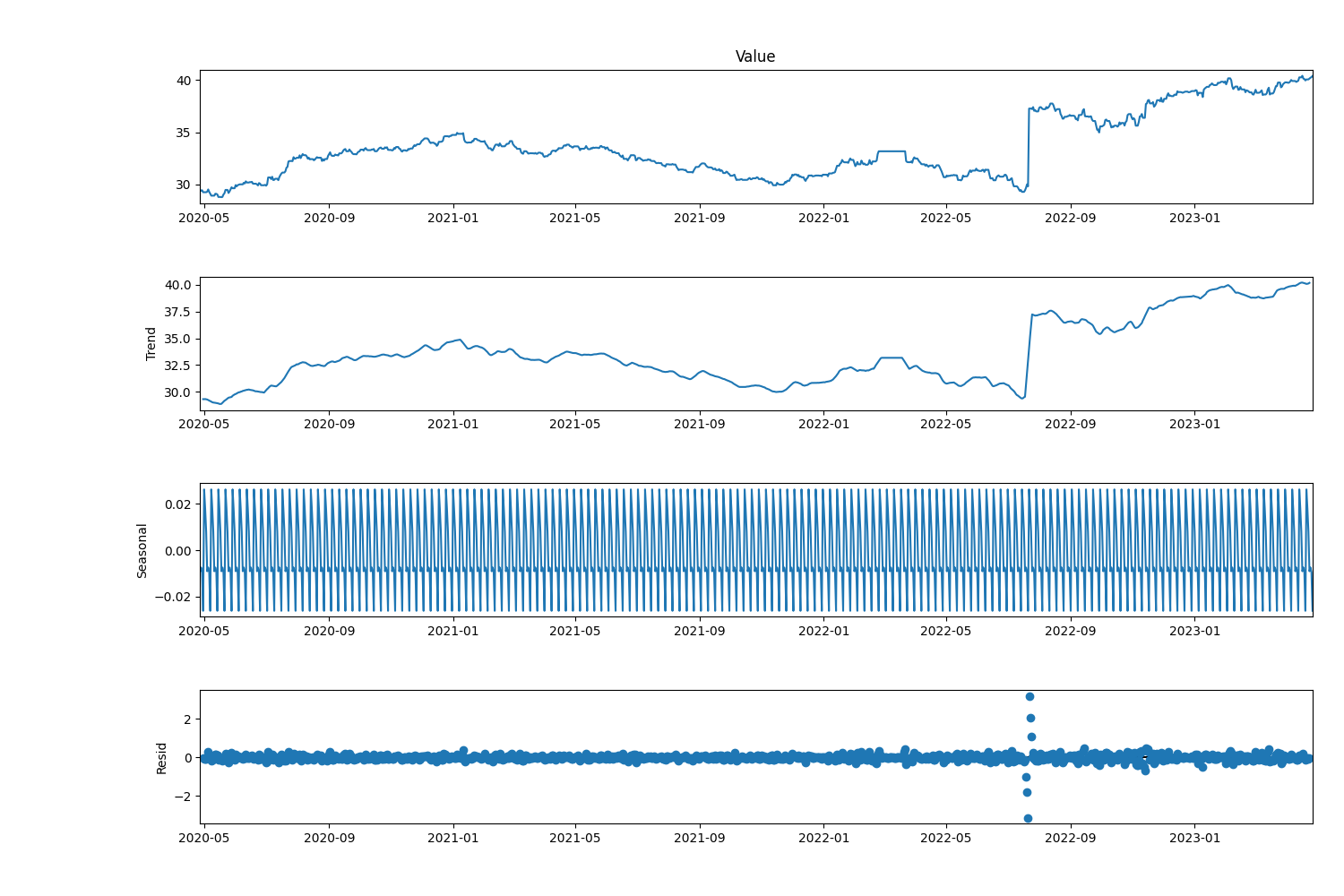




Бачимо, як на початку повномасштабного вторгнення курс був зафіксований, потім на літо він трохи спадав і знову почав сильно рости.

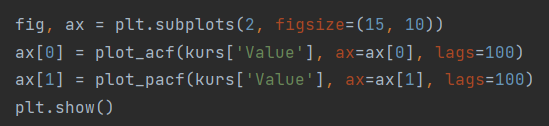
Візуалізую декомпозицію ряду на тренд, сезонність та залишки.

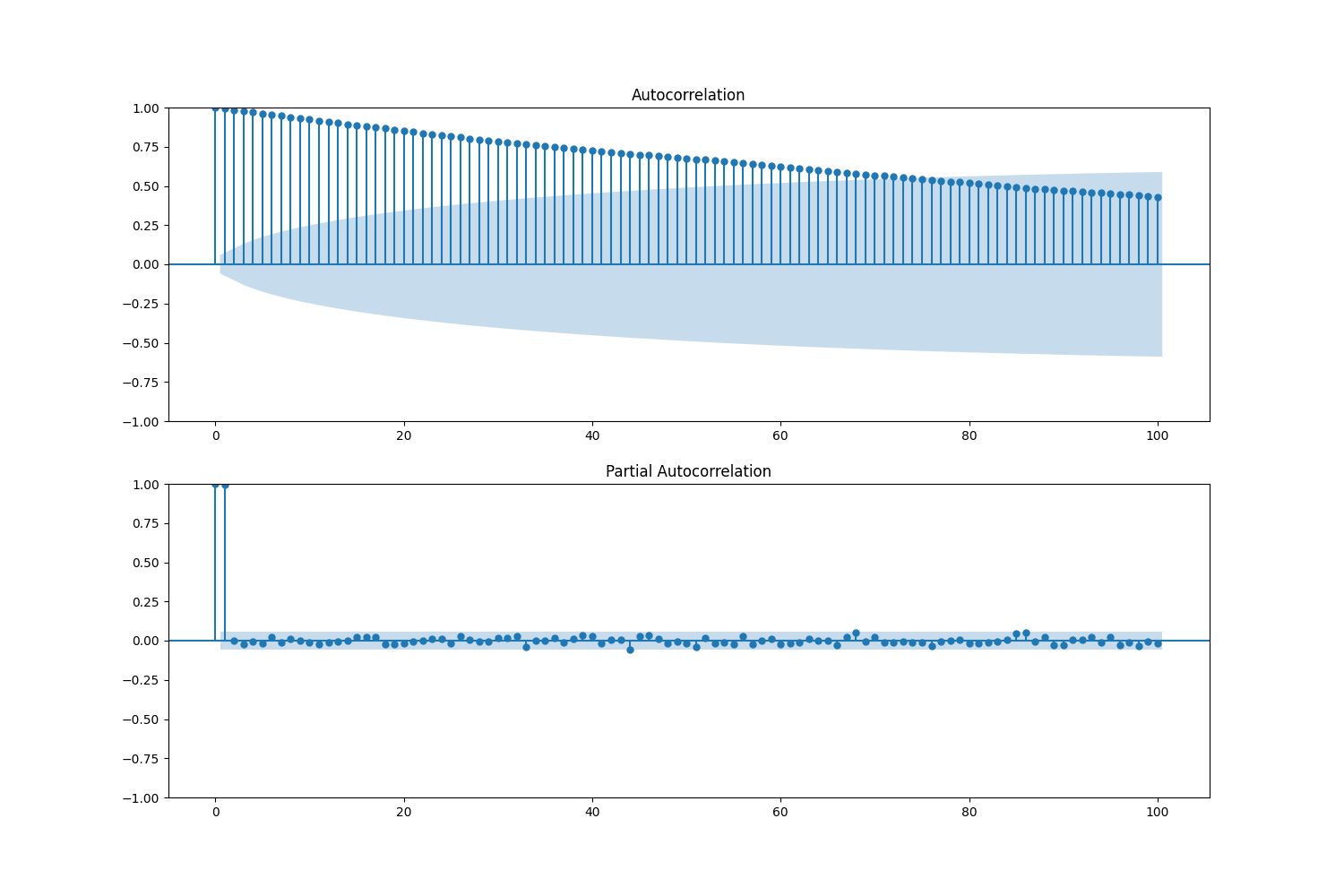




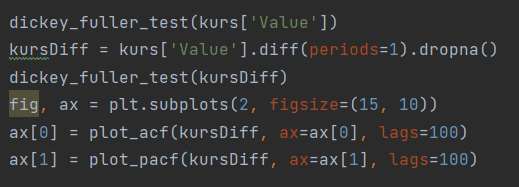
Бачимо, що ряд не має чіткої сезонности.

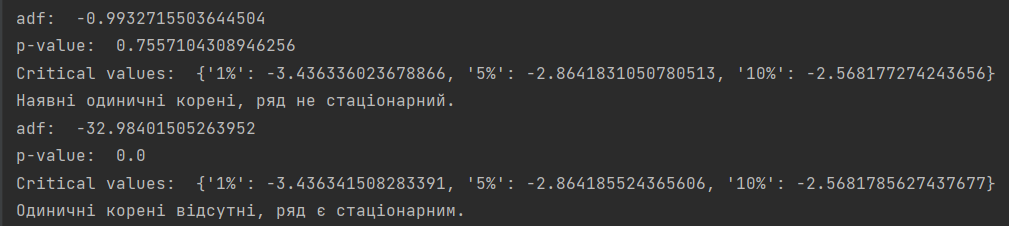
Будую графіки автокореляції та часткової автокореляції.

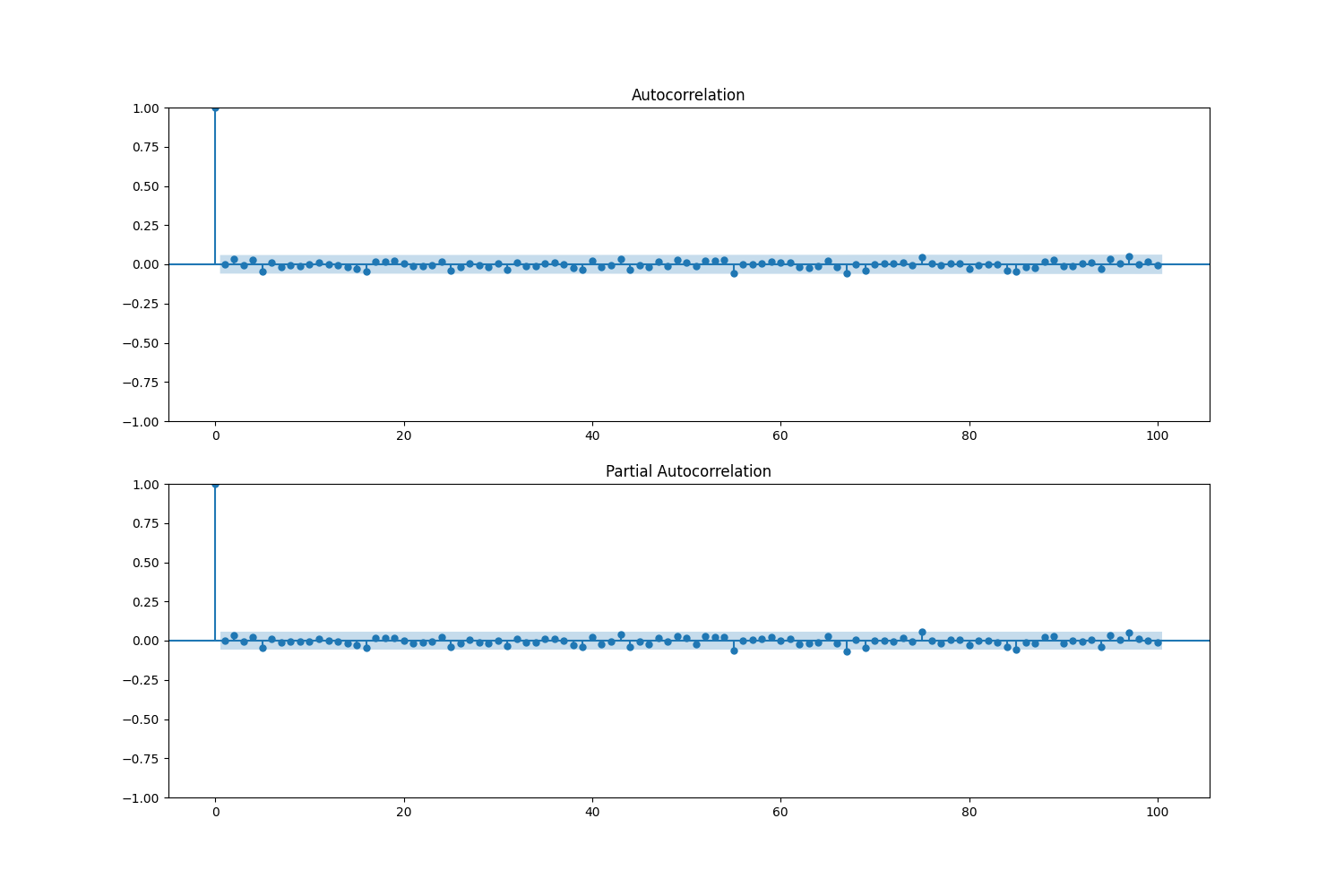




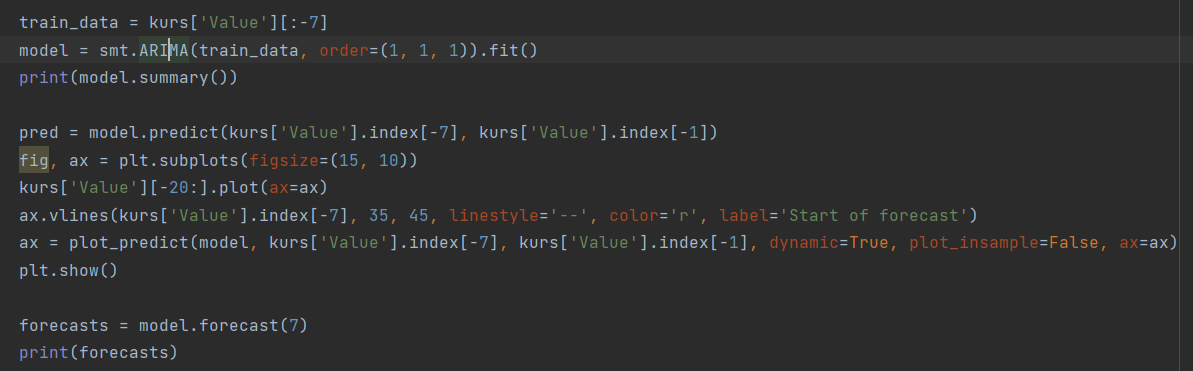
Проводжу тест на стаціонарність, бачимо, що ряд не є стаціонарним, тому використовую різницю.

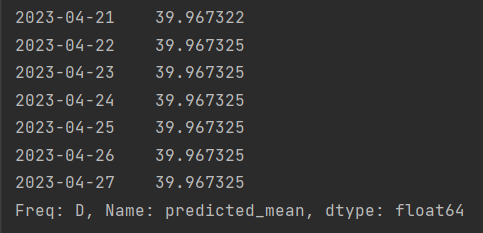


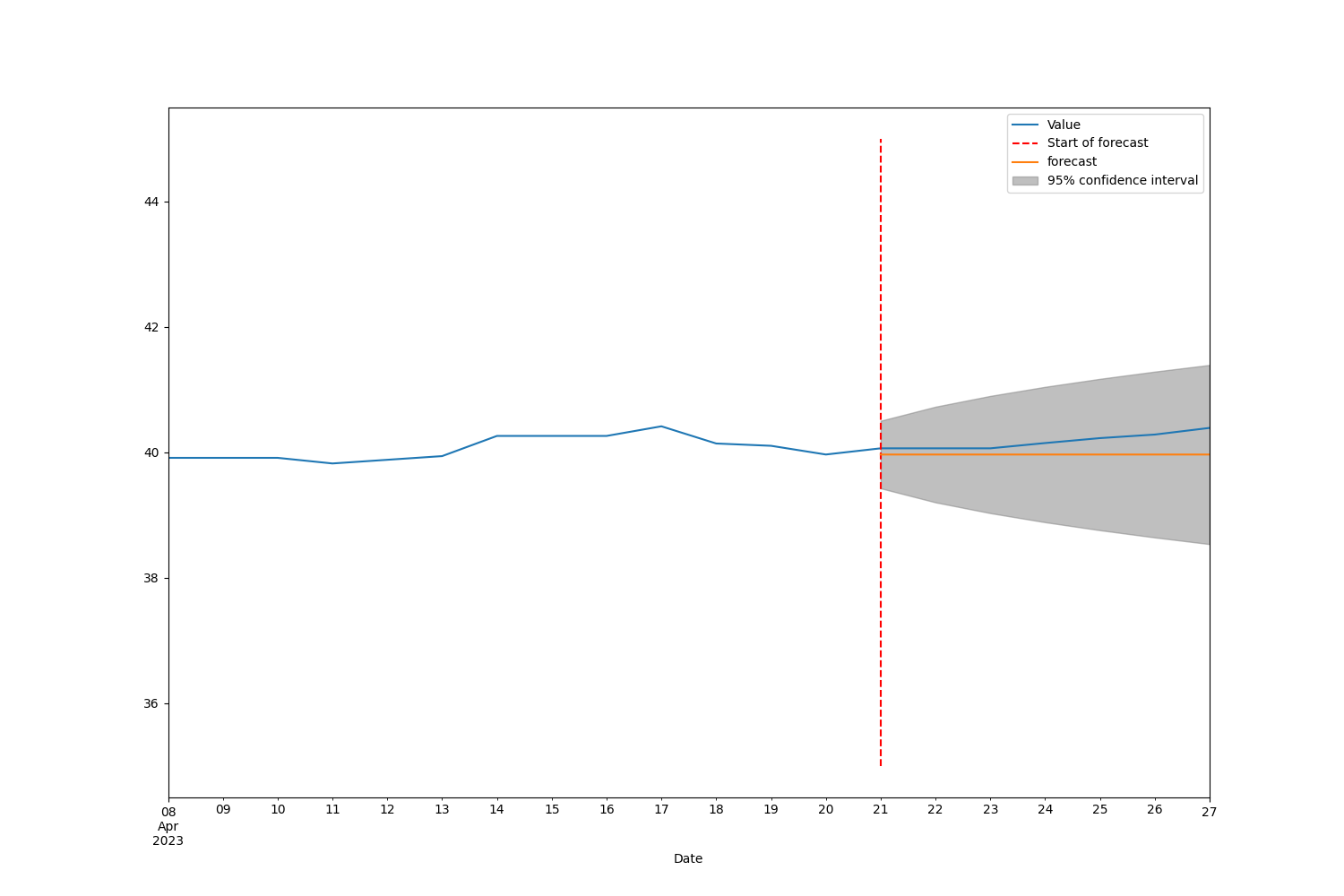




Будую модель ARIMA для прогнозу значення ціни на тиждень вперед.







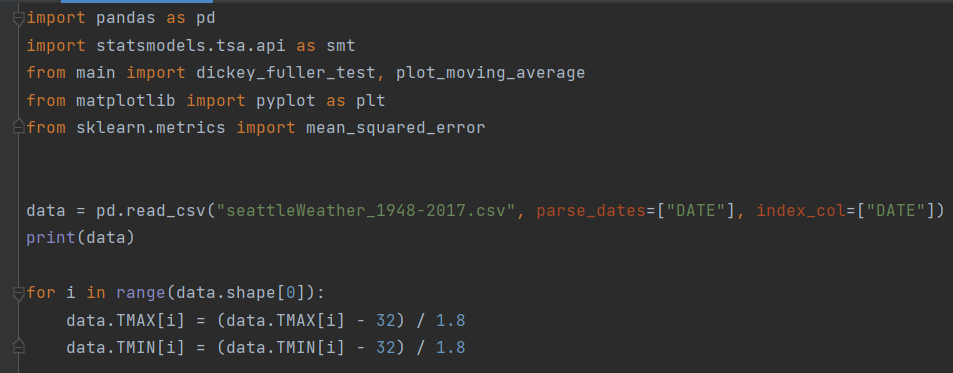
Бачимо, що прогноз є доволі точний, від початку прогнозу він взагалі майже збігався з реальним значенням, далі трошки почало розходитися, але на зовсім невеликі значення. Тож модель працює добре.

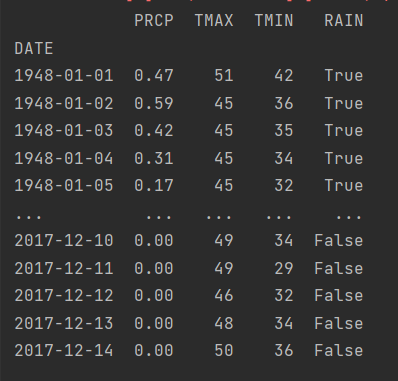
**Додаткове завдання.**

Потрібно з'ясувати, чи є сезонна компонента в кількості опадів в Сіетлі:

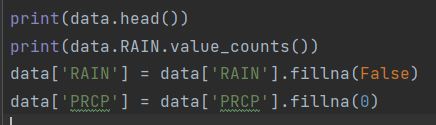
1. Градуси перевести в Цельсії.
2. Чи є кореляція між температурою та опадами?
3. Скласти прогноз опадів на 2018 рік, оцінити точність прогнозу.

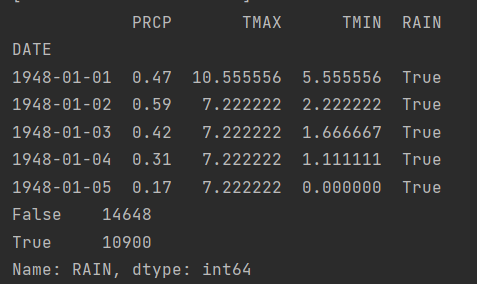
Завантажую датасет і за допомогою формули переведення Фаренгейтів у Цельсії виконую перетворення температури.



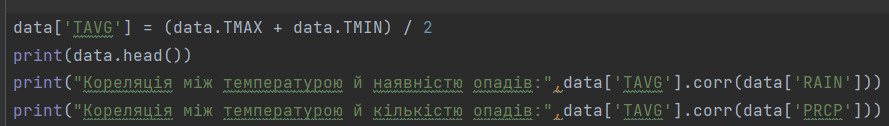


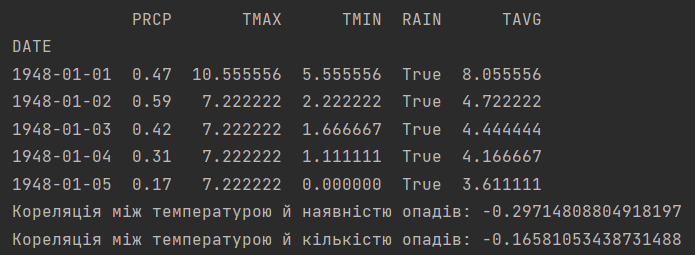
Далі дивлюся, частіше йде дощ чи ні і відповідно замінюю пропущені значення відсутністю дощу і 0 опадів.





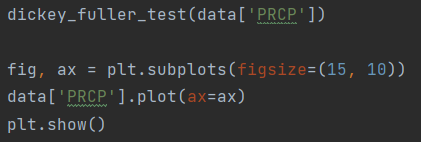
Для дослідження кореляції між температурою й опадами я додаю колонку середньої температури за день. Знаходжу кореляцію між цією температурою і наявністю і кількістю опадів.

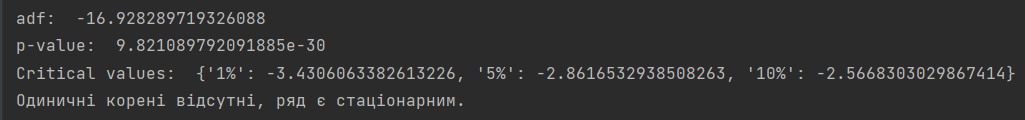


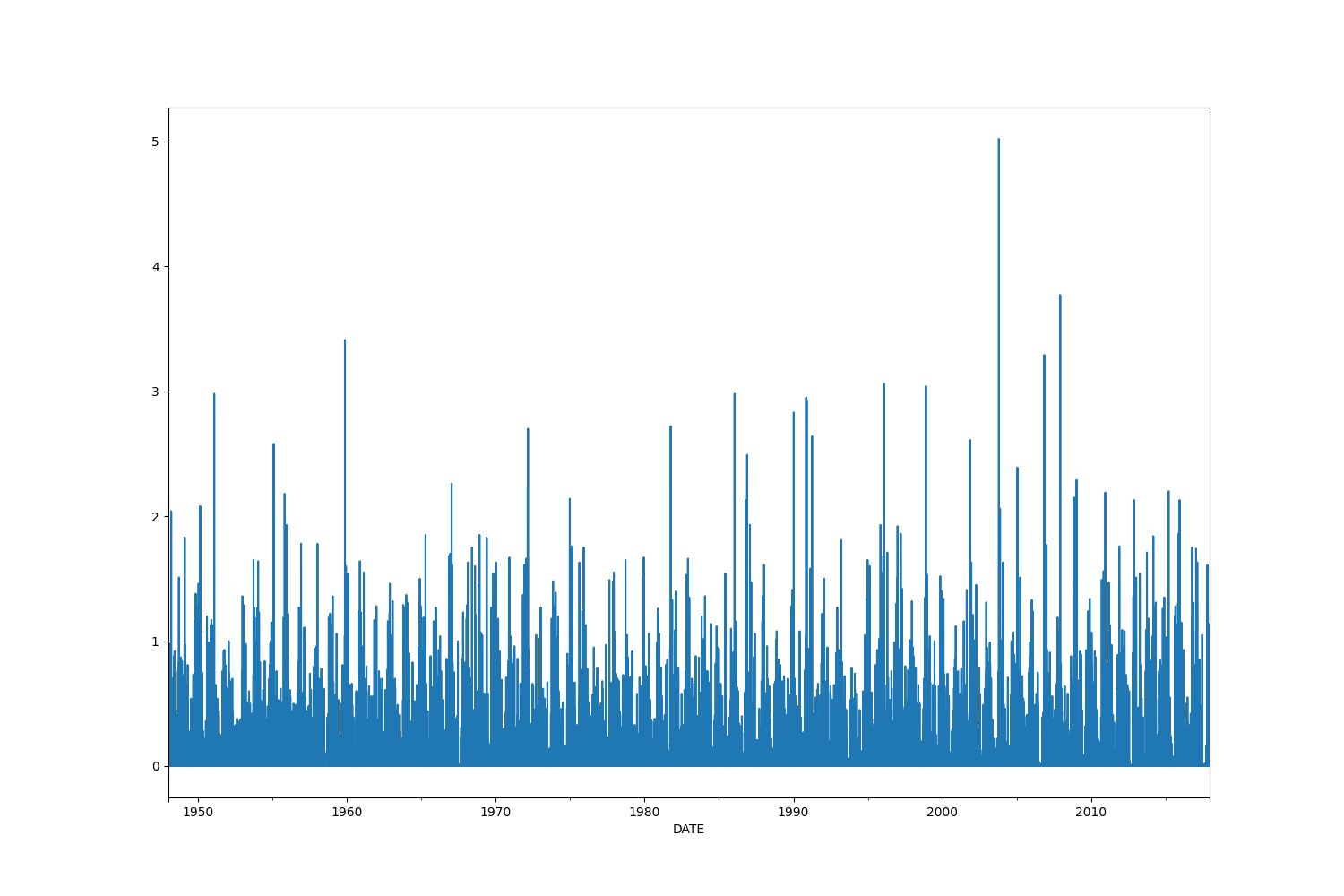


Маємо слабку негативну кореляцію, тобто при збільшенні температури дощ іде рідше й менше, але ця залежність дуже слабка.

Перевіряю ряд опадів на стаціонарність і будую його.

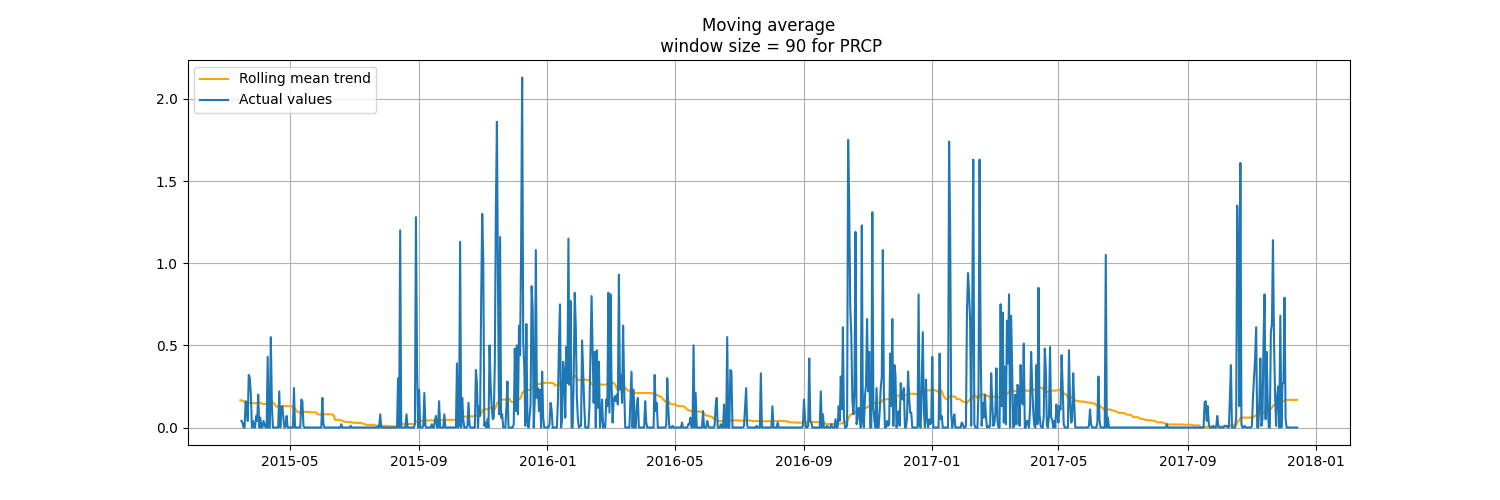






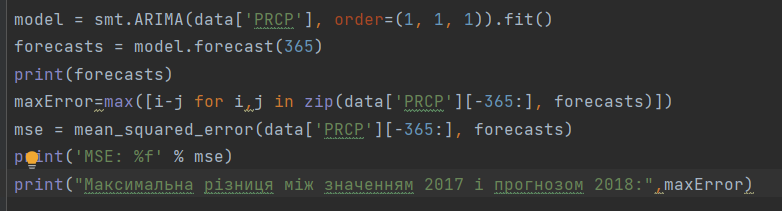
Далі застосовую згладжування для останніх трьох років опадів.

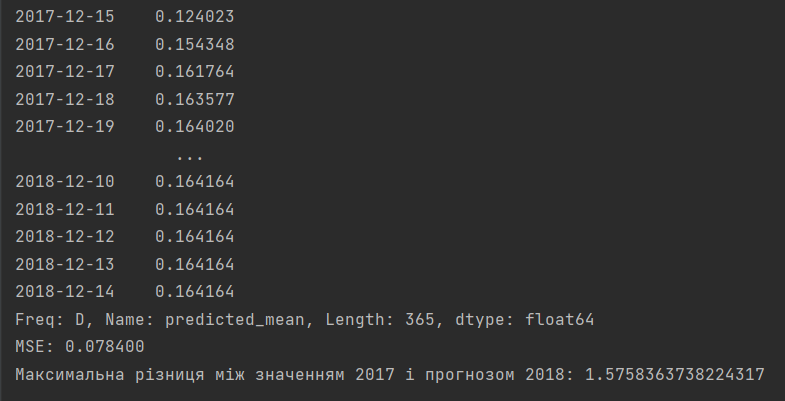




Можна побачити, що є тренд на зменшення кількости опадів приблизно з січня і до вересня, а потім восени і на початку зими кількість опадів зростає.

Застосую модель ARIMA для прогнозу опадів на наступний рік.





Маємо такі значення опадів. Для оцінювання точності я обрав порівняння з попереднім роком оцінювання за допомогою різниці між значенням того дня попереднього року і прогнозованого та середньої квадратичної похибки. Отримуємо хороші результати, прогноз не сильно відрізняється від значень попереднього року, тож можна сказати, що модель працює добре.

**Висновок.**

Отже, у цій роботі я ознайомився з методами моделювання часових послідовностей.

У результаті лабораторної роботи було використано різні DataFrames, проведено їх дослідження, побудовано й проаналізовано часові ряди захворювання на Covid у Німеччині та Польщі, курсу гривня-євро та опадів у Сієтлі. Було передбачено опади на наступний рік. Отримали, що сусідні країни у плані захворюваности помірно кореляюють, але в Німеччині кількість випадків була набагато більшою. Але періоди зростання і спаду у цих країнах схожі. Також проаналізувавши ряд курсу валют побачили, що ряд не має сезонності і за допомогою моделі ARIMA зробили передбачення за останній тиждень даних. Передбачення вийшло доволі ефективним, модель працює добре.

У додатковому завданні градуси Фаренгейта перевів у Цельсія, дослідив, що кореляція між температурою та опадами дуже слабка, а також зробив прогноз на наступний рік. Порівнюючи зі значеннями попереднього року можна прийти до висновку, що модель працює правильно, адже значення кардинально не відрізнялися. Використовуючи PyCharm, отримуємо коректний результат.