Зображення, що містить текст

Автоматично згенерований опис

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 №2**

Прикладні задачі машинного навчання

**Тема:** Часові ряди і прості лінійна регресія

Виконав Перевірив:

студент групи ІП-11: Нестерук А. О

Панченко С. В.

Київ 2023

Зміст

[1 Мета лабораторної роботи 6](#__RefHeading___Toc216_587646004)

[2 Завдання 7](#__RefHeading___Toc218_587646004)

[3 Виконання 8](#__RefHeading___Toc220_587646004)

[3.1 Завантажити метеорологічні дані в 1895-2022 роках з CSV-файлу в DataFrame. Після цього дані відформатувати для використання 8](#__RefHeading___Toc222_587646004)

[3.2 Зображення лінійної регресії для 1895 по 2018. 9](#__RefHeading___Toc224_587646004)

[3.3 Спрогнозувати дані на 2019, 2020, 2021 та 2022 рік 11](#__RefHeading___Toc226_587646004)

[3.4 Оцінити за формулою, якими могли б бути показники до 1895 року 11](#__RefHeading___Toc228_587646004)

[3.5 Скористатися функцією regplot бібліотеки Seaborn для виведення всіх точок даних 12](#__RefHeading___Toc230_587646004)

[3.6 Виконати масштабування осі у 12](#__RefHeading___Toc232_587646004)

[3.7 Порівняти отриманий прогноз для 2019, 2020, 2021 та за 2022 роки з даними на NOAA «Climate at a Glance»: https://www.ncdc.noaa.gov/cag/ і зробити висновок 13](#__RefHeading___Toc234_587646004)

[4 Висновок 14](#__RefHeading___Toc236_587646004)

# Мета лабораторної роботи

Мета роботи – дослідити лінійну регресію на прикладі прогнозування січневих температур у Нью-Йорку за рокам з використанням Python.

# Завдання

1. Завантажити метеорологічні дані в 1895-2022 роках з CSV-файлу в
2. Бібліотеку Seaborn використати для графічного представлення

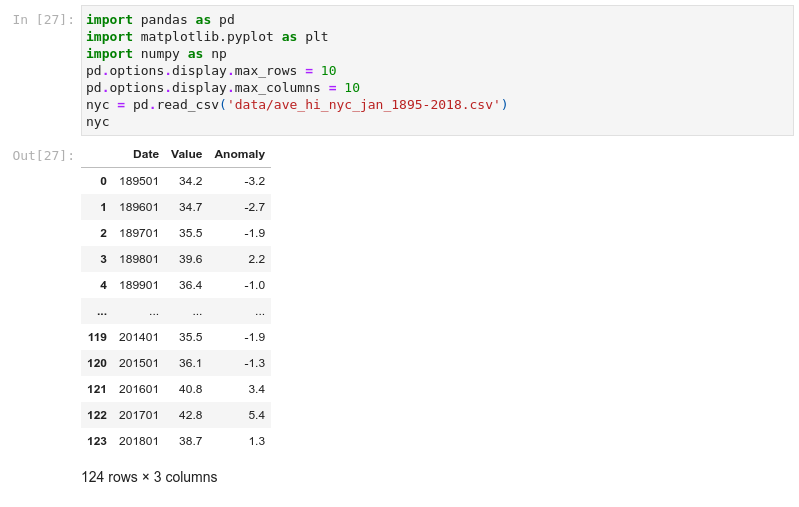
даних DataFrame у вигляді регресійної прямої, що представляє графік зміни обраних показників за період 1895-2018 років

1. Спрогнозувати дані на 2019, 2020, 2021 та 2022 рік
2. Оцінити за формулою, якими могли б бути показники до 1895 року
3. Скористатися функцією regplot бібліотеки Seaborn для виведення всіх точок даних
4. Виконати масштабування осі у
5. Порівняти отриманий прогноз для 2019, 2020, 2021 та за 2022 роки з даними на NOAA «Climate at a Glance»: https://www.ncdc.noaa.gov/cag/ і зробити висновок
6. Зробити звіт про роботу

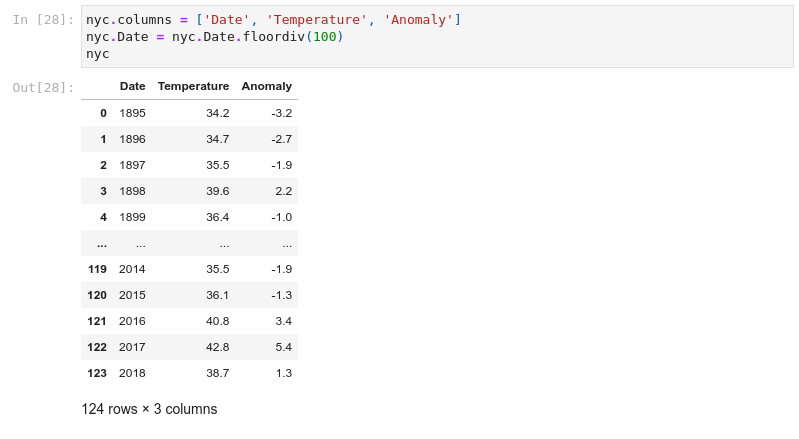
# Виконання

## Завантажити метеорологічні дані в 1895-2022 роках з CSV-файлу в DataFrame. Після цього дані відформатувати для використання

Зчитаємо даніз CSV-файлу, уикористовуючи метод read\_csv.

  
  
Рисунок 3.1 - Завантаження датасету

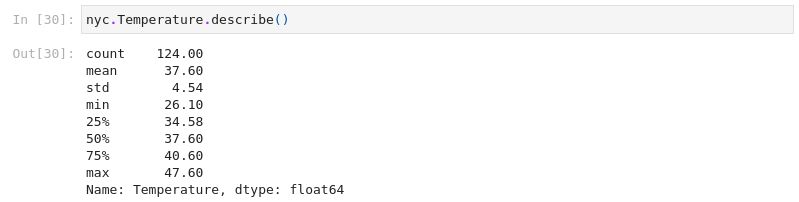
Відформатуємо датафрейм, а саме: переназвемо стовпці та застосуємо цілочисельне ділення, поділивши значенння років на 100.

  
  
Рисунок 3.2 - Форматований датафрейм

Налаштужмо точність виведення чисел.

  
  
Рисунок 3.3 - Налаштування точності виведення

Знайдемо основні статистичні показники.

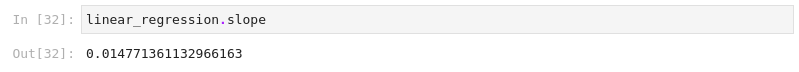
  
  
Рисунок 3.4 - Основні статистичні показники

## Зображення лінійної регресії для 1895 по 2018.

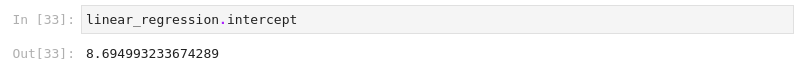
Імпортуємо модуль stats з пакету scipy та за допомогою функції linregress знайдемо лінійну регресію, передавши в аргумети дати та температури.

  
  
Рисунок 3.5 - Розрахунок лінійної регресії

Розрахувавши регресію, дізнаємося про коефіцієнт нахилу.

  
  
Рисунок 3.6 - Коефіцієнт нахилу

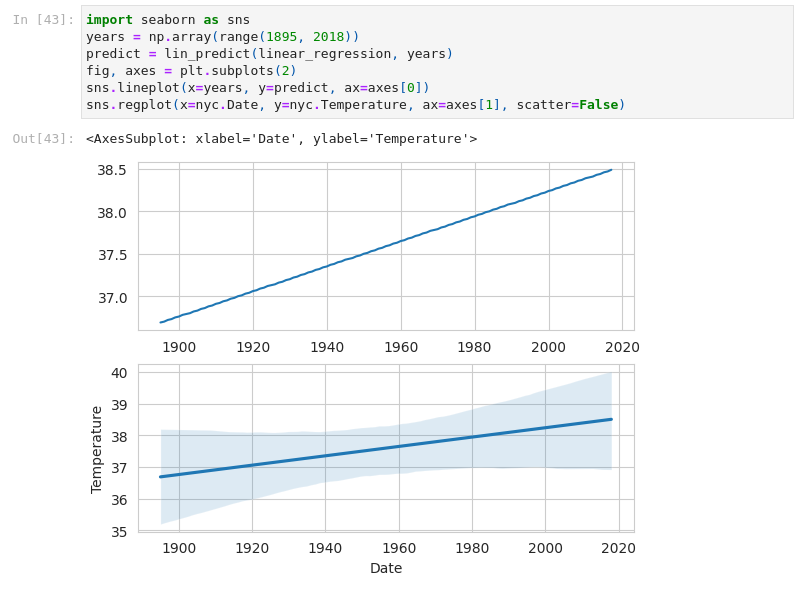
Дізнаємося про точку перетину прямої лінії

  
  
Рисунок 3.7 - Точка перетину прямої лінії

Створимо функцію lin\_predict, яка буде видавати спрогнозовані значення для лінійної регресії.

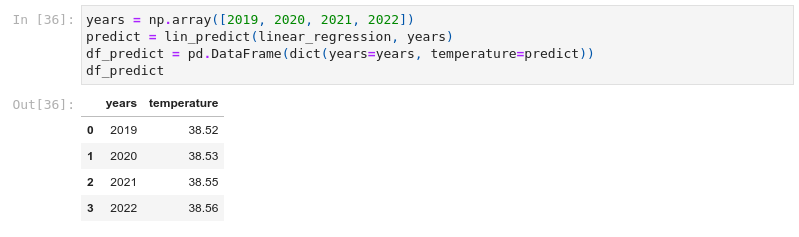
  
  
Рисунок 3.8 - Функція lin\_predict

Імпортуємо бібліотеку seaborn та застосуємо функцію regplot для відображення лінійної регресії. Передамо в неї в якості аргумента дату, значення температуру та параметр за замовчуванням scatter=False, щоб відобразити лише пряму.

  
  
Рисунок 3.9 - Лінійна регресія для 1895 по 2018

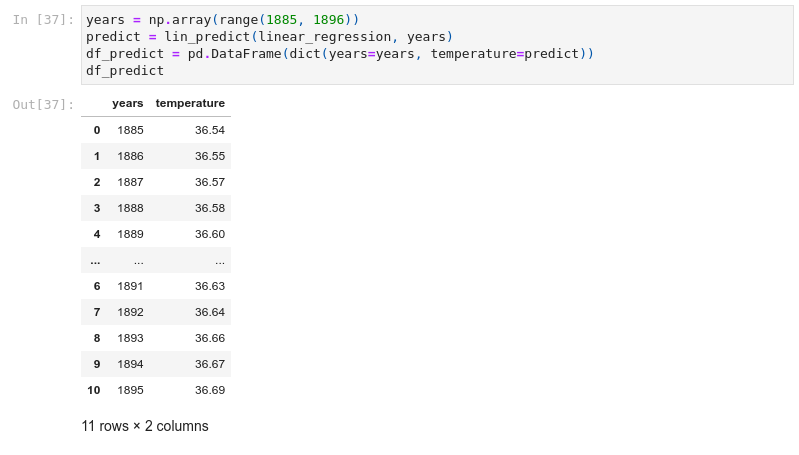
## Спрогнозувати дані на 2019, 2020, 2021 та 2022 рік

Спрогнозуємо дані для наступних років. Тобто підставимо роки у формулу лінійної регресії.

  
  
Рисунок 3.10 - Прогнозовані температури за роками

## Оцінити за формулою, якими могли б бути показники до 1895 року

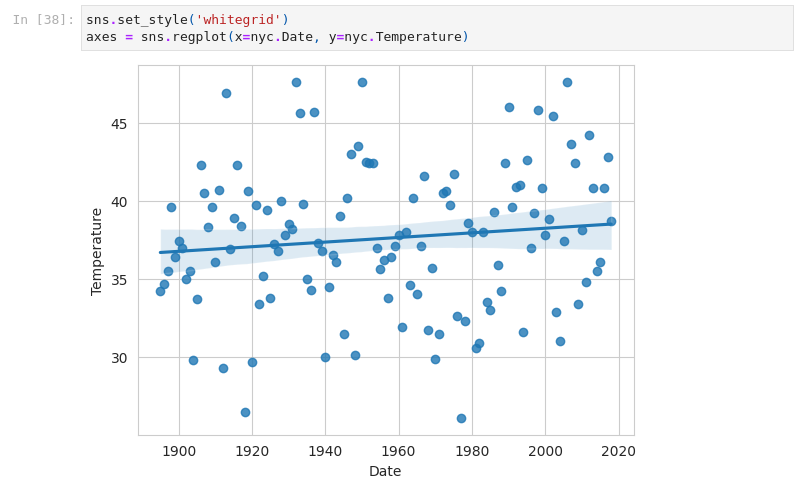
Обчислимо показники до 1895 року.

  
  
Рисунок 3.11 - Показники температур з 1885 по 1895 роки включно

Як можна побачити, температура поступово зростає, і з періоду 1885 по 2023 роки спостерігається збільшення на 2 градуси.

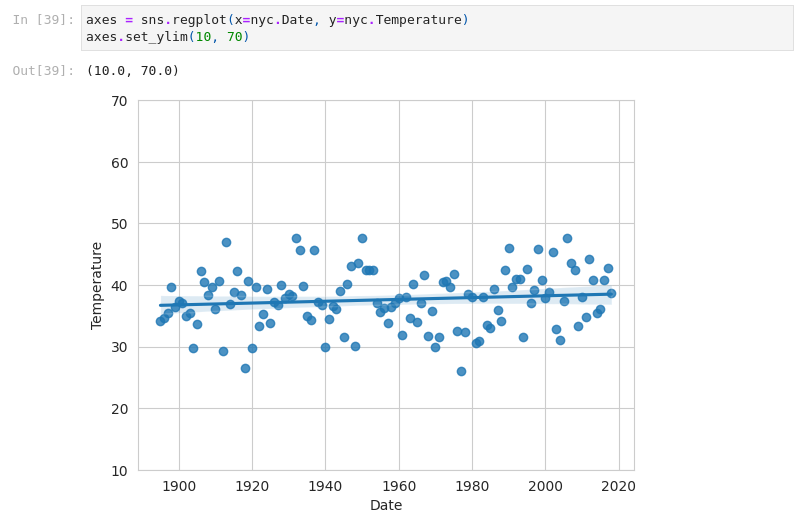
## Скористатися функцією regplot бібліотеки Seaborn для виведення всіх точок даних

Встановлюємо стиль відображення, побудуємо графік роки-температури. Побачимо, що дані доволі розкидані.

  
  
Рисунок 3.12 - Графік лінійної регресії роки-температури

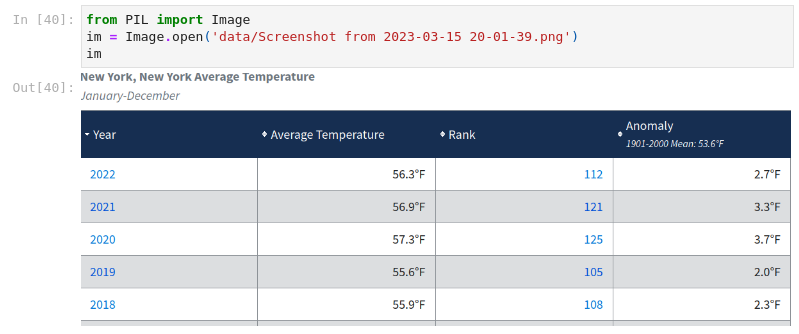
## Виконати масштабування осі у

За допомогою методу set\_ylim вкажемо межі від 10 до 70 градусів.

  
  
Рисунок 3.13 - Масштабований графік від 10 до 70 градусів

## Порівняти отриманий прогноз для 2019, 2020, 2021 та за 2022 роки з даними на NOAA «Climate at a Glance»: https://www.ncdc.noaa.gov/cag/ і зробити висновок

Подивимося на сайті дані за період 2019-2022 років. Побачимо, що фактичні дані сильно відрізняються від того, що спрогнозувала лінійна регресія. Можна зробити висновок, що треба давати їй іще якісь дані для кращого прогнозування, наприклад: кількість опадів, кількість CO2 тощо.

  
  
Рисунок 3.14 - Справжні зафіксовані дані середніх температур

# Висновок

Під час виконання цієї лабораторної роботи здобув базові навички

використання пакету scipy мови Python, досліджуючи середні температури в

січня у Нью-Йорку з 1895 до 2022 років, обчисливши лінійну регресію та

зробивши прогноз. У результаті спрогнозовані дані не збігалися з фактичними

даними. Отже, точність низька та потрібно врахувати додаткові параметри.