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

Виконав Перевірив:

студент групи ІП-11: Нестерук А. О

Панченко С. В.

Київ 2023

Зміст

[1 Мета лабораторної роботи 3](#__RefHeading___Toc3319_204808571)

[2 Завдання 4](#__RefHeading___Toc3321_204808571)

[3 Виконання 5](#__RefHeading___Toc510_1145936286)

[3.1 Завантажити метеорологічні дані в 1895-2022 роках з CSV-файлу в DataFrame. Після цього дані відформатувати для використання 5](#__RefHeading___Toc512_1145936286)

[3.2 дані на 2019, 2020, 2021 та 2022 рік 5](#__RefHeading___Toc514_1145936286)

[3.3 за формулою, якими могли б бути показники до 1895 року 6](#__RefHeading___Toc516_1145936286)

[3.4 функцією regplot бібліотеки Seaborn для виведення всіх точок даних 7](#__RefHeading___Toc518_1145936286)

[3.5 масштабування осі у 7](#__RefHeading___Toc520_1145936286)

[3.6 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 8](#__RefHeading___Toc522_1145936286)

[4 ВИСНОВОК 10](#__RefHeading___Toc524_1145936286)

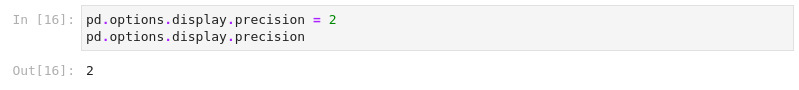
# Мета лабораторної роботи

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

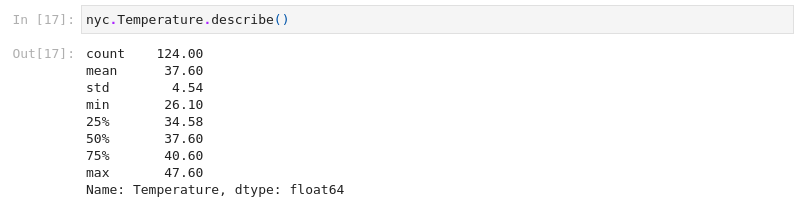
# Завдання

1. Завантажити метеорологічні дані в 1895-2022 роках з CSV-файлу в DataFrame. Після цього дані відформатувати для використання.
2. Бібліотеку Seaborn використати для графічного представлення даних DataFrame у вигляді регресійної прямої, що представляє графік зміни обраних показників за період 1895-2018 років.
3. Спрогнозувати дані на 2019, 2020, 2021 та 2022 рік.
4. Оцінити за формулою, якими могли б бути показники до 1895 року.
5. Скористатися функцією regplot бібліотеки Seaborn для виведення всіх точок даних
6. Виконати масштабування осі *у*
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.
8. Зробити звіт про роботу
9. Виконання
   1. 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 3.5 - Зчитаємо даніз CSV-файлу, уикористовуючи метод read\_csv.

Знайдемо основні статистичні показники.

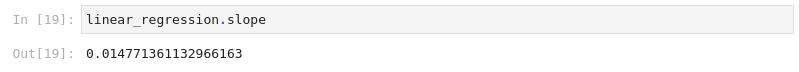
  
Рисунок 3.6 - Знайдемо основні статистичні показники.

* 1. дані на 2019, 2020, 2021 та 2022 рік

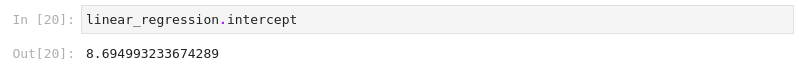
Імпортуємо модуль stats з пакету scipy та за допомогою функції linregress знайдемо лінійну регрусію, передавши в аргумети дати та температури.

  
Рисунок 3.7 - Імпортуємо модуль stats з пакету scipy та за допомогою функції linregress знайдемо лінійну регрусію, передавши в аргумети дати та температури.

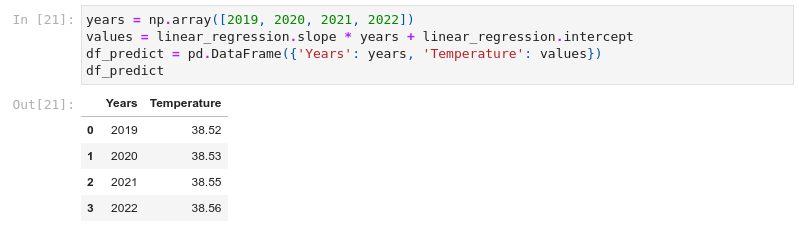
Розрахувавши регресію, дізнаємося про коефіцієнт нахилу.

  
Рисунок 3.8 - Розрахувавши регресію, дізнаємося про коефіцієнт нахилу.

Дізнаємося про точку перетину прямої лінії

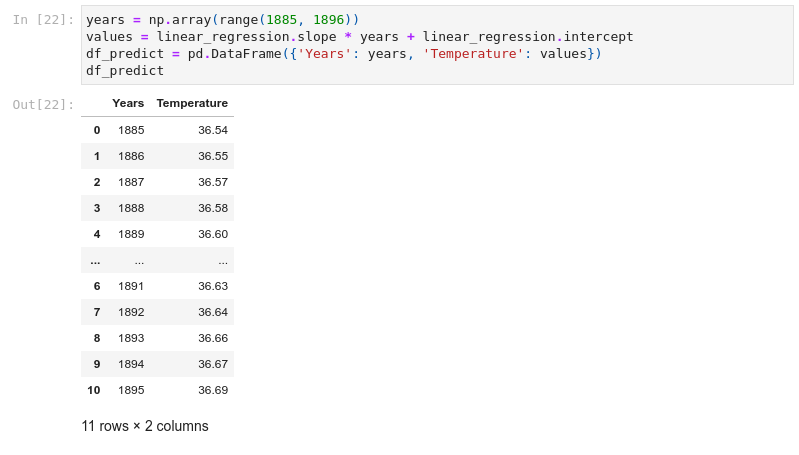
  
Рисунок 3.9 - Дізнаємося про точку перетину прямої лінії

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

  
Рисунок 3.10 - Спрогнозуємо дані для наступних років. Тобто підставимо роки у формулу лінійної регресії.

* 1. за формулою, якими могли б бути показники до 1895 року

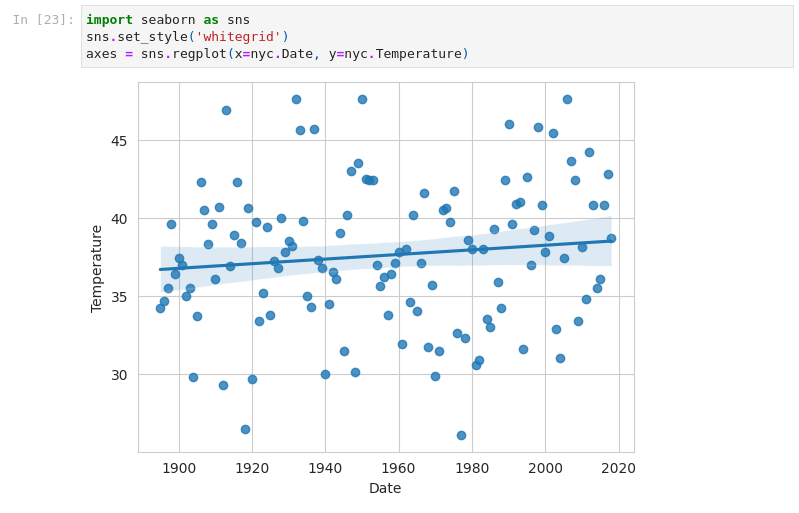
Обчислимо показники до 1895 року

  
Рисунок 3.11 - Обчислимо показники до 1895 року

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

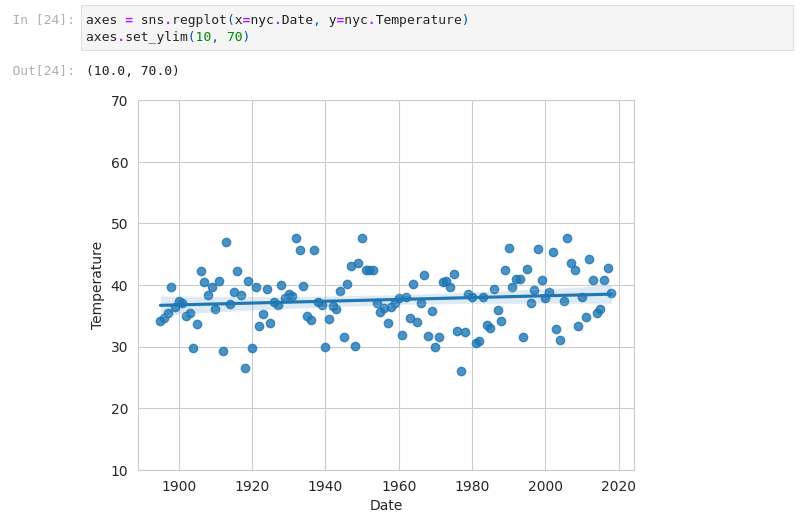
* 1. функцією regplot бібліотеки Seaborn для виведення всіх точок даних

Імпортуємо Seaborn, встановлюємо стиль відображення, побудуємо графік роки-температури. Побачимо, що дані доволі розкидані.

  
Рисунок 3.12 - Імпортуємо Seaborn, встановлюємо стиль відображення, побудуємо графік роки-температури. Побачимо, що дані доволі розкидані.

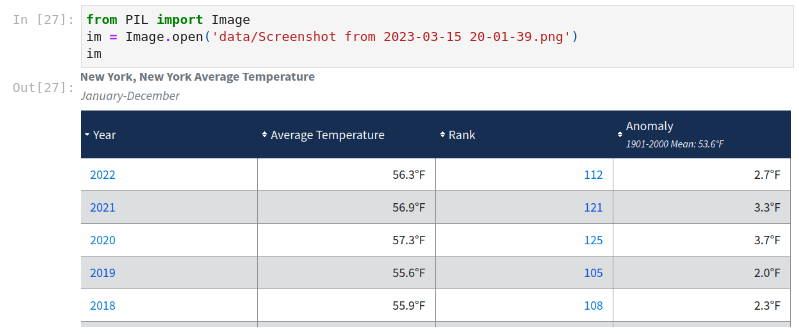
* 1. Масштабування осі у

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

  
Рисунок 3.13 - За допомогою методу set\_ylim вкажемо межі від 10 до 70 градусів.

* 1. 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 2019-2023 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 — Фактичні дані середніх температур

# 4 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.