**Міністерство освіти і науки України**

**Національний технічний університет України «Київський політехнічний**

**інститут імені Ігоря Сікорського"**

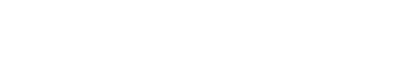
**Факультет інформатики та обчислювальної техніки**

**Кафедра інформатики та програмної інженерії**

# Звіт

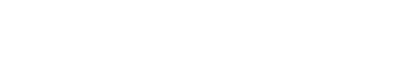
з лабораторної роботи № 2 з дисципліни

«Часові ряди і прості лінійна регресія»



**Виконав** *ІП-15, Дацьо Іван*

(шифр, прізвище, ім'я, по батькові)



**Перевірив**



прізвище, ім'я, по батькові



*Нестерук А.*

Київ 2023

# Завдання:

1. В даній лабораторній роботі Вам треба завантажити метеорологічні дані в 1895-2022 роках з CSV-файлу в DataFrame. Після цього дані треба буде 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.

1. Спрогнозуйте дані на 2019, 2020, 2021 та 2022 рік.
2. Оцініть за формулою, якою могли б бути показники до 1895 року. Наприклад, оцінка середньої температури за січень 1890 року може бути отримана наступним чином.
3. Скористайтесь функцією regplot бібліотеки Seaborn для

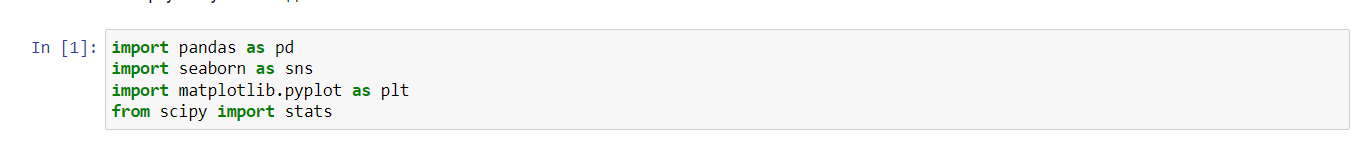
виведення всіх точок даних; дати представляються на осі x, а показники на осі y. Функція regplot будує діаграму розкиду даних, на якій точки представляють показники за заданий рік, а пряма лінія - регресійну пряму.

1. Виконайте масштабування осі *у* від (приклад від 10 до 70 градусів):
2. Порівняйте отриманий прогноз для 2019, 2020, 2021 та за 2022

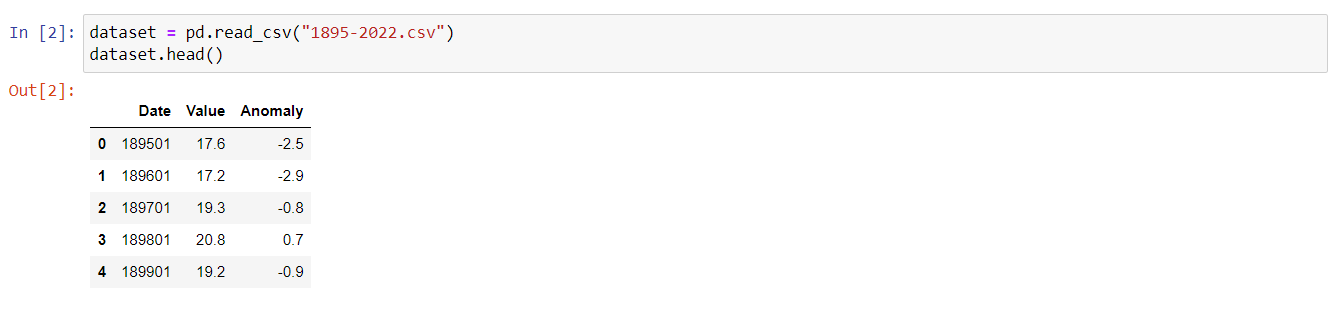
роки з даними на NOAA «Climate at a Glance»: <https://www.ncdc.noaa.gov/cag/>і зробити висновок.

# Виконання

1. Імпортуємо усі необхідні пакети:



Завантажимо дані із csv файлу та переглянемо їх:

****

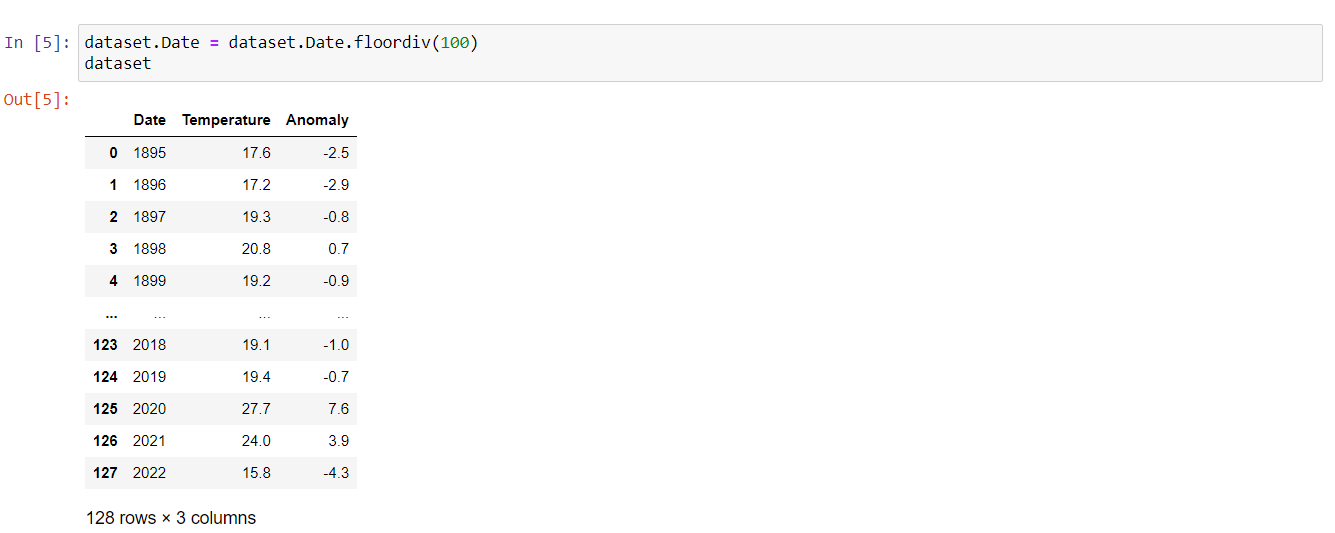
Перейменуємо назви стовпчиків для простішого використання:

****

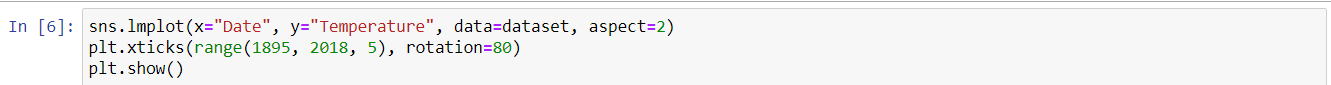
Оскільки усі місяць для усіх записів є однаковим, то для кращого читання даних видалимо місяць в колонці Date. Але спочатку перевіримо тип даних в цій колонці:

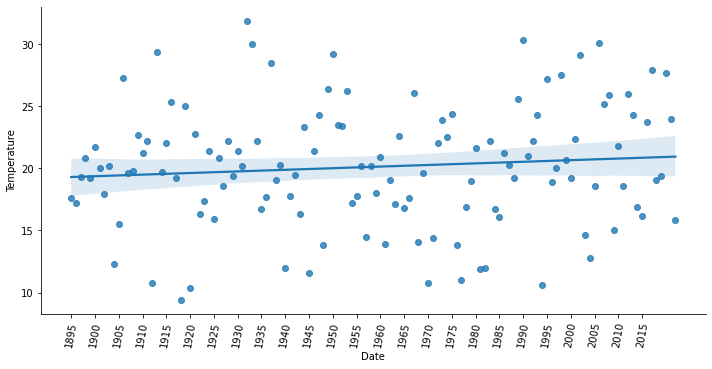
****

Оскільки дані в цій колонці є цілими числами то ми можемо розділити їх на 100 та прибрати останні два числа:

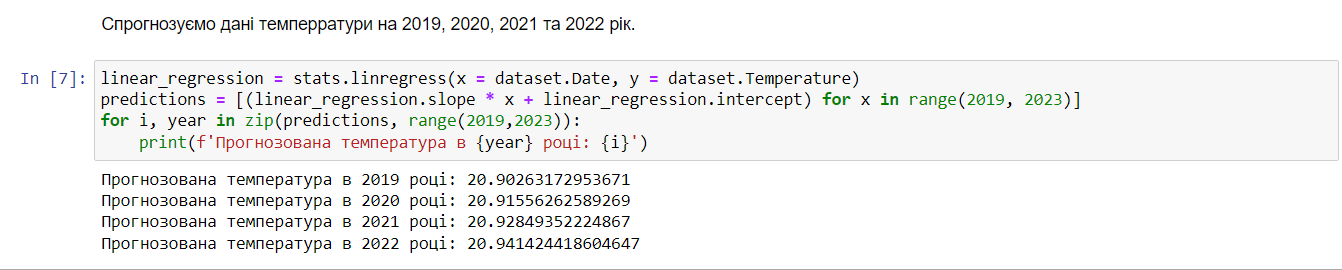


1. Бібліотеку Seaborn використати для графічного представлення даних DataFrame у вигляді регресійної прямої, за період 1895-2018 років.

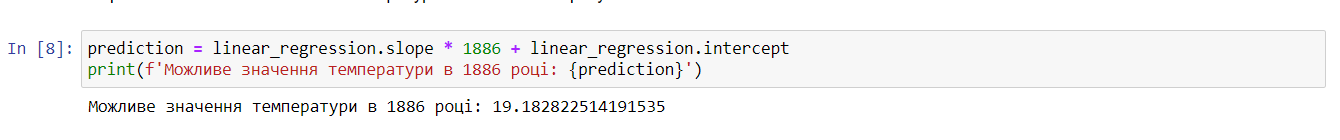


****

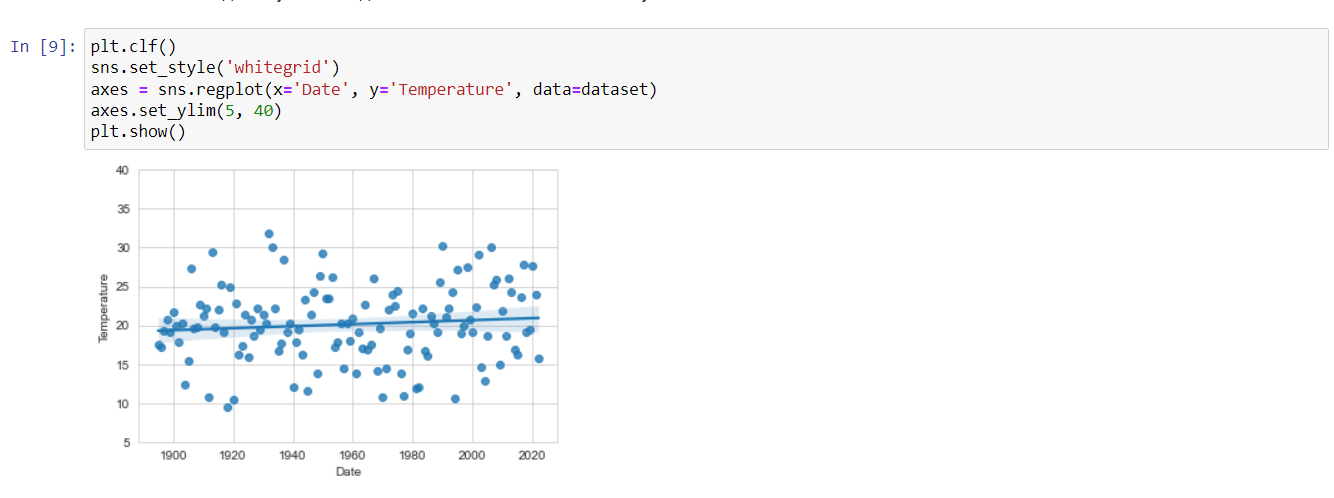
1. Спрогнозуємо дані температури на 2019-2022 рік.



1. Оцінимо за формулою, яка середня температура могла бути отримана в 1886 році :



**5 та 6.** Скористаємось функцією regplot бібліотеки Seaborn для виведення всіх точок даних та змінимо масштабування для осі Y від 5 до 40 для кращого відображення:



**7.** Порівняємо отримані результати із реальними даними:

|  |  |  |  |
| --- | --- | --- | --- |
| Рік | NOAA «Climate at a  Glance» | Отримані результати | Різниця |
| 2019 | 19.4 | 20.90 | +1.5 |
| 2020 | 27.7 | 20.91 | -6.79 |
| 2021 | 24 | 20.92 | -3.08 |
| 2022 | 15.8 | 20.94 | +5.14 |

Можна помітити що прогнозовані дані відрізняються суттєво. Це зумовлено тим що даний підхід не враховує додаткові чинники, які могли вплинути на температуру. Даний спосіб є корисним, коли потрібно знайти приблизне значення величини.

**Висновок**

Виконавши дану лабораторну я ознайомився з бібліотекою seaborn, яка була використана для побудови лінійної регресії. Було спрогнозовано середні температури у січні, використовуючи для цього дані із попередніх років. Був зроблений висновок, що даний метод передбачає лише приблизні значення які в реальності значно відрізняються.