



Міністерство освіти і науки України
Національний технічний університет України
“Київський політехнічний інститут імені Ігоря Сікорського”
Факультет інформатики та обчислювальної техніки
Кафедра інформаційних систем і технологій

Лабораторна робота №2
Прикладні задачі машинного навчання
“Часові ряди і прості лінійна регресія”

Виконав
студент групи ІК – 33:
Вересоцький А. Ю.

Перевірив:
асистент кафедри ІСТ
Нестерук А.О.

Київ 2025

Завдання:

1. Потрібно завантажити метеорологічні дані в 1895-2024 роках з CSV-файлу в DataFrame. Після цього дані треба буде відформатувати для використання.

Завантажую середні січніві температури в Лексінгтон штат Кентуккі з 1895 по 2024 рік через часові ряди NOAA «Climate at a Glance»:

```
1 lex = pd.read_csv('data.csv')
✓ [67] < 10 ms

1 lex.head()
✓ [68] < 10 ms
```

#	Lexington	Kentucky January Average Temperature
0	# Units: Degrees Fahrenheit	NaN
1	# Missing: -99	NaN
2	Date	Value
3	189501	28.8
4	189601	35.6

Форматування:

Видаляємо непотрібні рядки, оновлюємо індекси та змінюємо назви колонок:

```
1 lex.drop(index=[0, 1, 2], inplace=True)
✓ [69] < 10 ms

1 lex.reset_index(drop=True, inplace=True)
✓ [70] < 10 ms

1 lex.columns = ['Date', 'Temperature']
✓ [71] < 10 ms
```

```
1 lex.head()
✓ [72] < 10 ms
```

	Date	Temperature
0	189501	28.8
1	189601	35.6
2	189701	30.3
3	189801	39
4	189901	35.1

Перевіряємо типи даних та змінюємо їх за необхідності:

```
1 lex.dtypes
✓ [73] < 10 ms
```

	<unnamed>
Date	object
Temperature	object

```
1 lex.Date = lex.Date.astype(int)
2 lex.Temperature = lex.Temperature.astype(float)
✓ [74] < 10 ms

1 lex.Date = lex.Date.floordiv(100)
✓ [75] < 10 ms
```

```
1 lex.head()
✓ [76] < 10 ms
```

	Date	Temperature
0	1895	28.8
1	1896	35.6
2	1897	30.3
3	1898	39.0
4	1899	35.1

2. Використати бібліотеку Seaborn для графічного представлення даних DataFrame у вигляді регресійної прямої (лише регресійної прямої), що представляє графік зміни обраних показників за період 1895-2020 років.

Обмежуємо період до 2020 року та застосовуємо функцію `linregress` модулю `scipy.stats`:

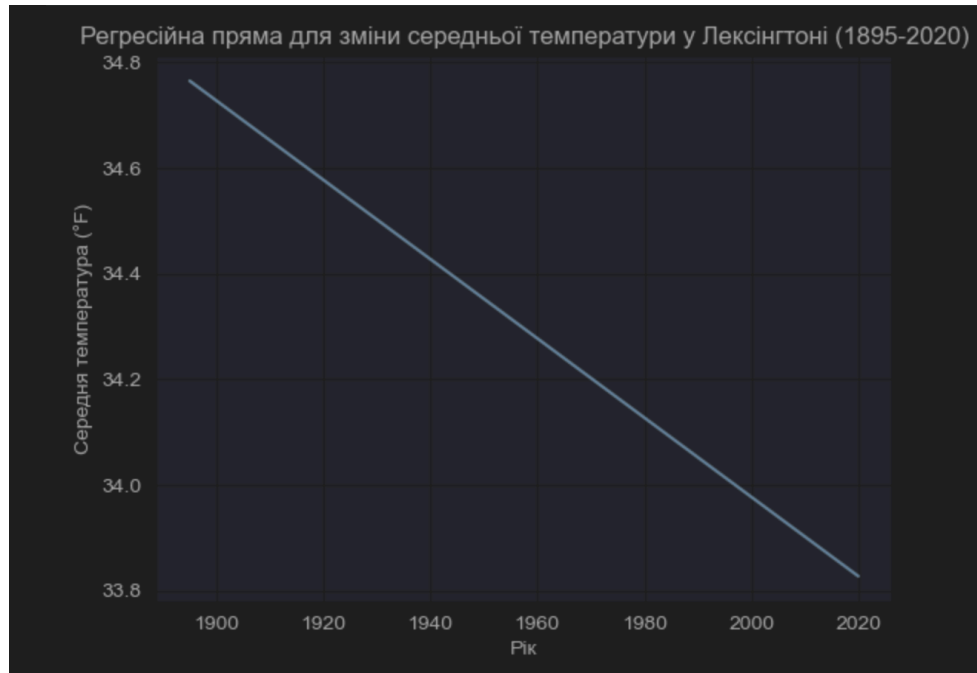
```
1 lex_filtered = lex[lex.Date <= 2020]
✓ [77] < 10 ms

1 linear_regression = stats.linregress(x=lex_filtered.Date, y=lex_filtered.Temperature)
✓ [78] < 10 ms
```

За допомогою бібліотеки Seaborn будуємо регресійну пряму, що представляє графік зміни січневих температур у Лексінгтоні за період 1895-2020 років:

```
1 sns.set_style('darkgrid')
2 plt.plot(lex_filtered.Date, linear_regression.slope * lex_filtered.Date + linear_regression.intercept)
3 plt.title('Регресійна пряма для зміни середньої температури у Лексінгтоні (1895-2020)')
4 plt.xlabel('Рік')
5 plt.ylabel('Середня температура (°F)')
6 plt.show()
✓ [79] 56ms
```

Отримуємо наступний графік:



3. Спрогнозувати дані на 2021, 2022, 2023 та 2024 рік.

Використовуємо рівняння простої лінійної регресії ($y = mx + b$), замінюючи m на `linear_regression.slope`, x на рік для якого прогнозується температура, а b на `linear_regression.intercept`:

```
1 for year in range(2021, 2025):
2     print(f"{year} -", linear_regression.slope * year + linear_regression.intercept, "(°F)")
✓ [80] < 10 ms
```

2021 - 33.81974603174602 (°F)
2022 - 33.81224546931632 (°F)
2023 - 33.80474490688663 (°F)
2024 - 33.797244344456935 (°F)

4. Оцінити за формулою, якою могли б бути показники до 1895 року (опрацювати не менше 5 років).

Аналогічно до попереднього пункту опрацьовуємо роки з 1889 по 1894:

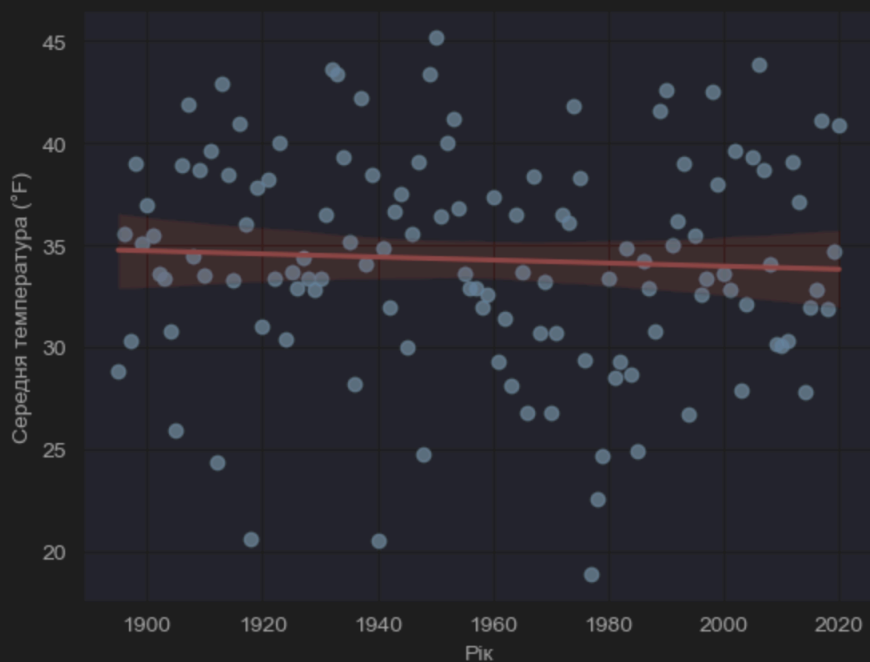
```
1 for year in range(1889, 1895):
2     print(f"{year} -", linear_regression.slope * year + linear_regression.intercept, "(°F)")
✓ [81] < 10 ms
```

1889 - 34.809820272465934 (°F)
1890 - 34.80231971003624 (°F)
1891 - 34.79481914760654 (°F)
1892 - 34.78731858517685 (°F)
1893 - 34.77981802274715 (°F)
1894 - 34.77231746031745 (°F)

5. Скористайтесь функцією `regplot` бібліотеки `Seaborn` для виведення всіх точок даних.

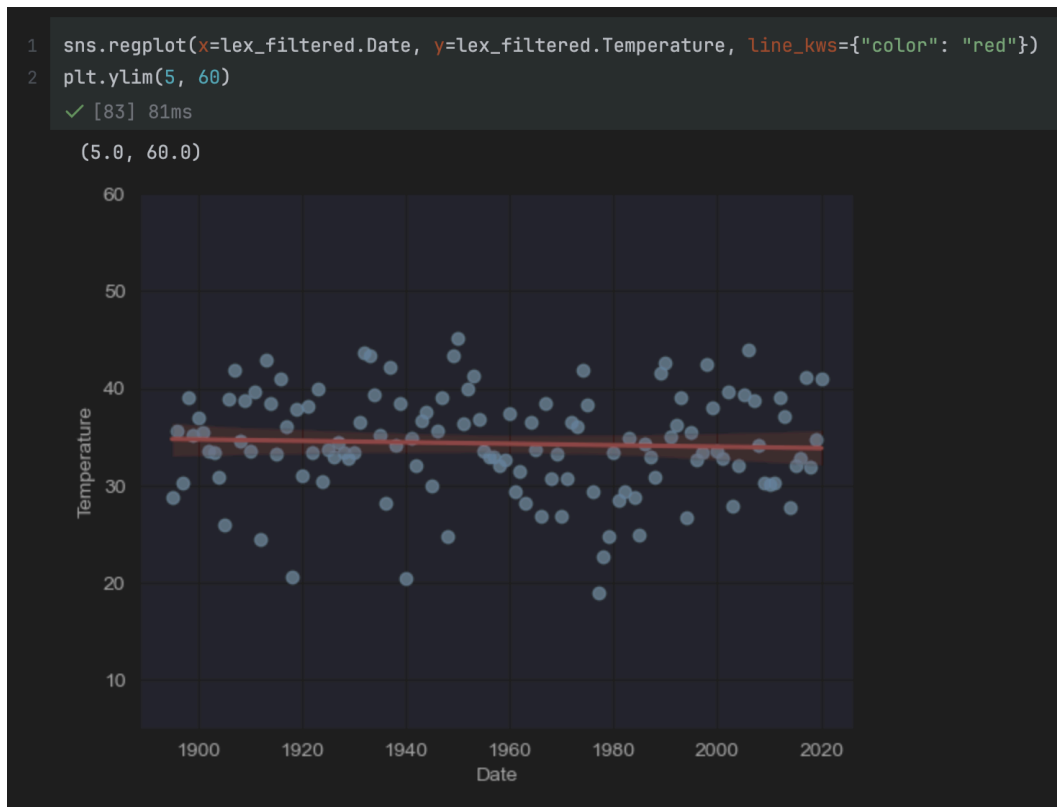
За допомогою функції `regplot` будемо діаграму розкиду даних, на якій точки представляють показники за заданий рік, а пряма лінія - регресійну пряму:

```
1 sns.regplot(x=lex_filtered.Date, y=lex_filtered.Temperature, line_kws={"color": "red"})
2 plt.xlabel('Рік')
3 plt.ylabel('Середня температура (°F)')
4 plt.show()
✓ [82] 76ms
```



6. Виконати масштабування осі у:

Функцією `ylim` обмежемо вісь температур від 5 до 60 градусів:



7. Порівняти отриманий прогноз для 2021, 2022, 2023 та за 2024 роки з даними на NOAA «Climate at a Glance» і зробити висновок про те, чому результати співпадають чи не співпадають.

Використовуємо отримані у другому пункті результати функції `linregress` для побудови регресійної прямої для періоду 2021-2024 і накладасмо цю пряму на графік січневих температур у Лексінгтоні у той же період (дані взяті з NOAA):

```
1 lex_predict = lex[lex.Date > 2020]
2 sns.set_style('darkgrid')
3 real = lex_predict.plot(x='Date', y='Temperature', style='.-', label="Дані NOAA")
4 estimated = plt.plot(lex_predict.Date, linear_regression.slope * lex_predict.Date + linear_regression.intercept,
5                       label="Прогноз (лінійна регресія)")
6 plt.title('Зміна середньої температури у Лексінгтоні (2020-2024)')
7 plt.xlabel('Рік')
8 plt.ylabel('Середня температура (°F)')
9 plt.legend()
10 plt.show()
✓ [84] 58ms
```

Отримуємо наступний графік:



Реальні дані NOAA значно варіюються: спостерігається різке зниження у 2022 році, стрибок у 2023 році, а потім зменшення у 2024 році.

Прогноз лінійної регресії є майже постійним: значення змінюються незначно ($\approx 33.8^{\circ}\text{F}$), що свідчить про дуже слабкий тренд.

Проте останній результат за 2024 рік знаходиться доволі близько до реально визначених даних.

Причини розбіжностей:

- Лінійна регресія припускає сталий тренд і не враховує коливання даних.
- Реальні температури NOAA мають значну річну варіативність, яку проста лінійна модель не може передбачити.

Висновок: У цій лабораторній роботі було проведено аналіз середніх січневих температур у Лексінгтоні (Кентуккі) за період 1895–2024 років. Використовуючи метод лінійної регресії, було здійснено прогноз температур на 2021–2024 роки та порівняння отриманих результатів із реальними даними NOAA. Як ми визначили, метод лінійної регресії є корисним для загального аналізу трендів, проте його недостатньо для точного прогнозування кліматичних змін у короткостроковій перспективі.