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

«Основи науки про дані»

«Індексування, вибір, редагування набору даних»

Виконав студент \_\_\_\_ІП-11 Трикош І. В\_\_\_\_

Перевірив \_\_\_Новотарський М. А.\_\_\_

Київ 2024

**Завдання**

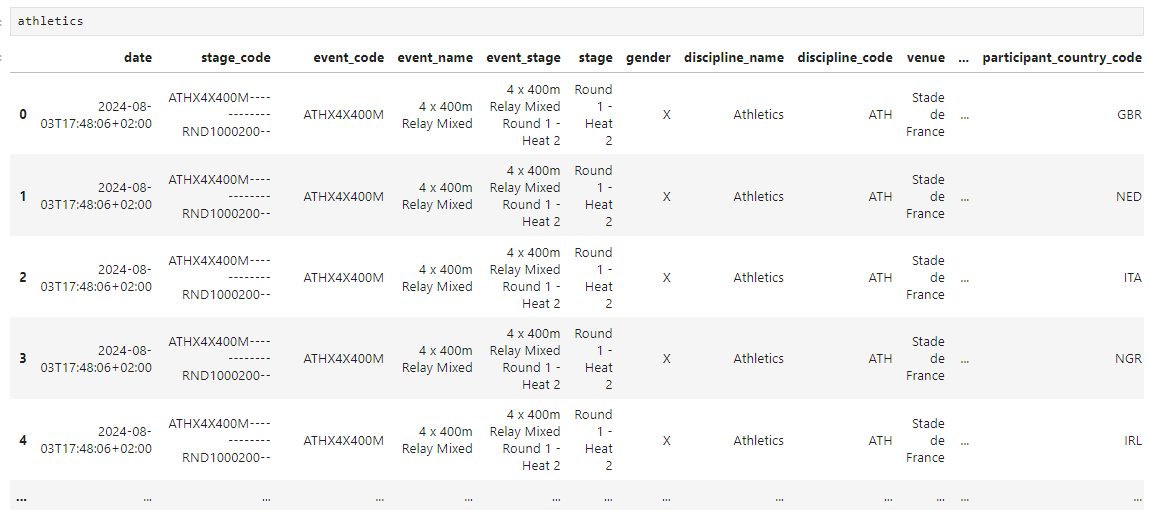
1. Детально розглянути наведені теоретичні відомості до даної лабораторної роботи.
2. Ознайомитися з кодом, що наведений у Jupyter notebook DD\_lab2.ipynb.
3. Застосувати до вибраного набору всі методи індексування, вибору та редагування набору даних.
4. Виберіть набір даних відповідно до номеру у списку з таблиці 2.

**Виконання**

Для виконання лабораторної роботи візьмемо датасет про Олімпійські ігри 2024: <https://www.kaggle.com/datasets/muhammadehsan02/olympic-summer-games-paris-2024/data> (<https://drive.google.com/file/d/1rtZkHbKiEhNY6wtvAbPYGKzfu28n0EVm/view?usp=sharing>)

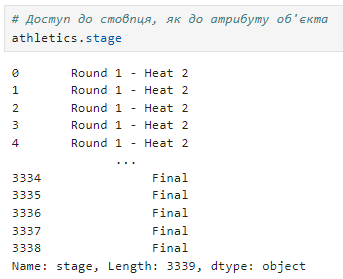
Працюватимемо з файлом Athletics.csv, що містить дані про змагання з атлетики.

Датасет має такий вигляд:

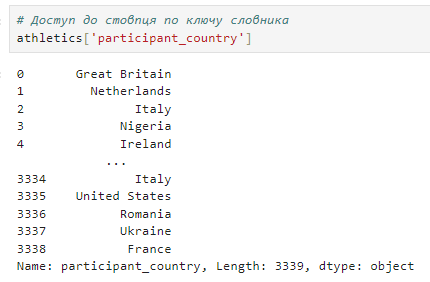




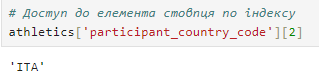
Отримаємо доступ до стовпця stage як до атрибуту об’єкта з точковою нотацією:



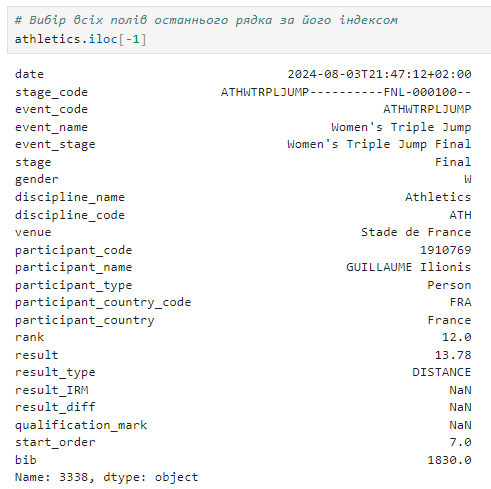
Отримаємо доступ до стовпця як до елемента словника, де ключем є назва стовпця participant\_country:



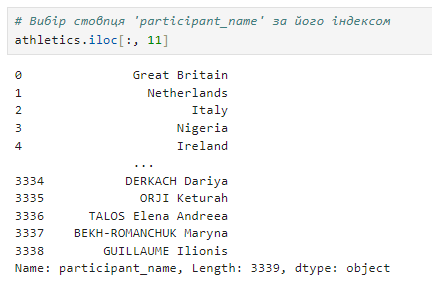
Отримаємо доступ до елемента з індексом 2 стовпця participant\_country\_code:



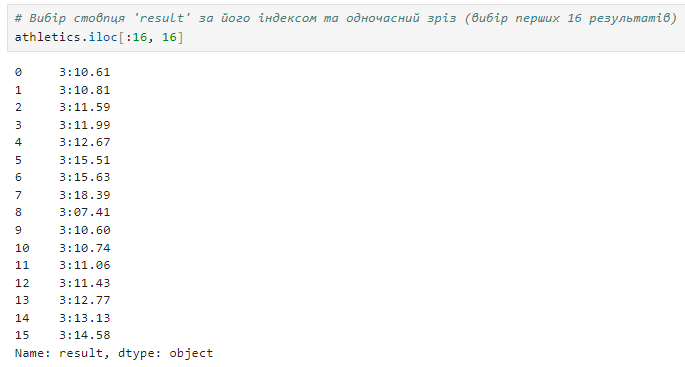
Виберемо всі поля останнього рядка з допомогою iloc. Цей метод – це індексно орієнтований вибір даних. Перший параметр – індекс рядка:



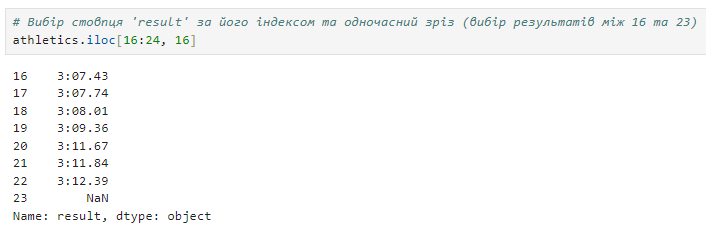
Виберемо стовпець participant\_name на основі його індексу допомогою iloc. Перший параметр методу – індекси рядків (тут обрану усі рядки), а другий – індекс стовпця:



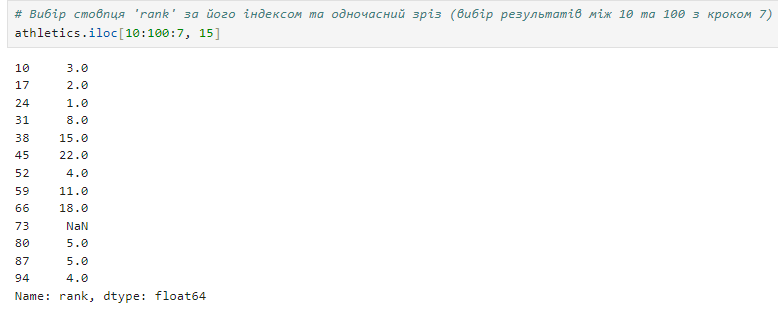
Виберемо стовпець result за його індексом та його перші 16 результатів з допомогою iloc:



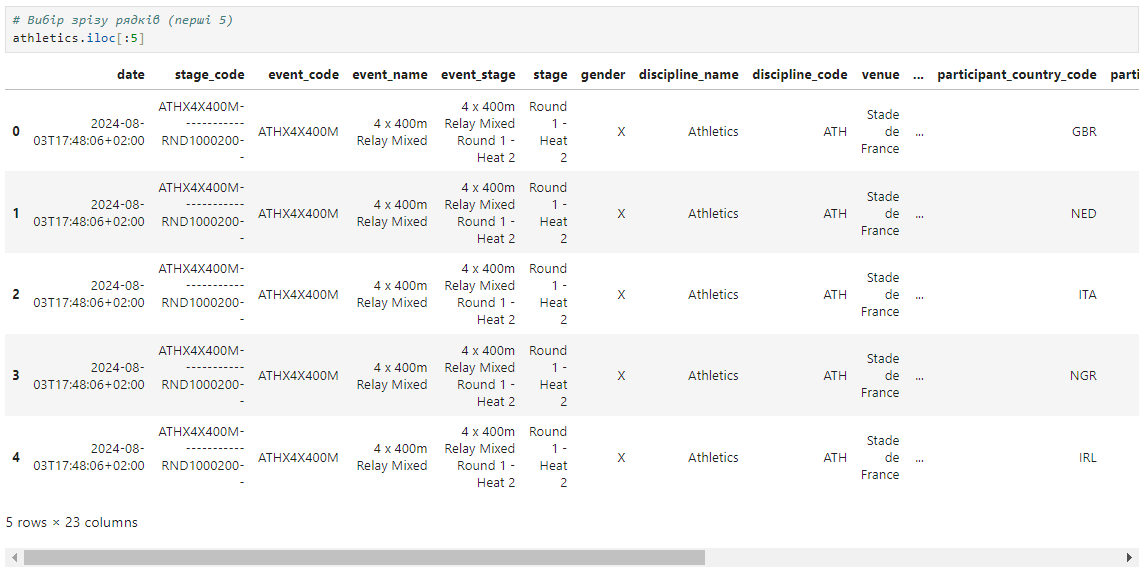
Виберемо стовпець result за його індексом та результати між 16 та 23 з допомогою iloc:



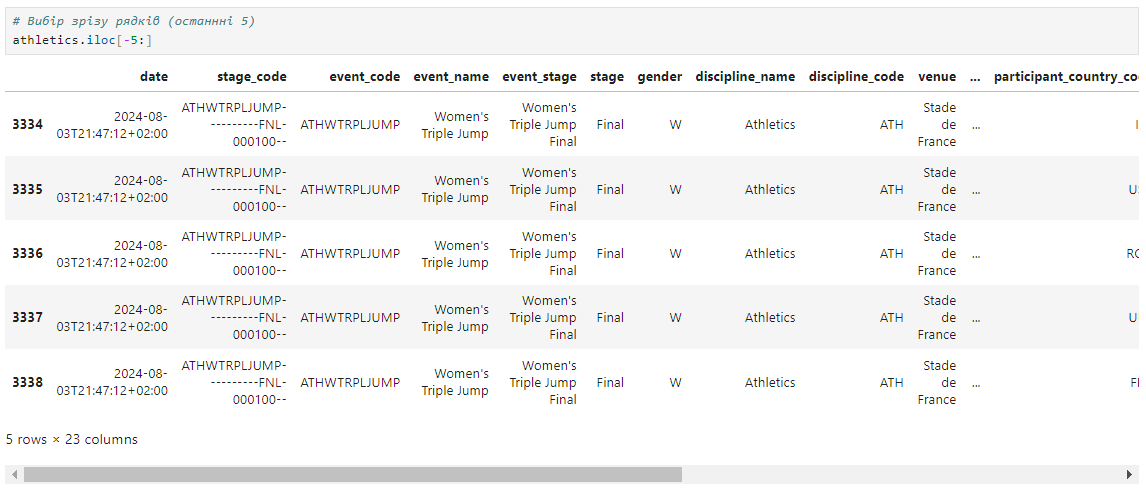
Виберемо стовпець rank та його результати між 10 та 100 з кроком 7 з допомогою iloc:



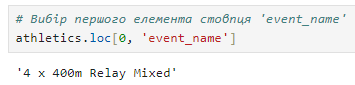
Виберемо перші 5 рядків з допомогою зрізу, використовуючи iloc:



Виберемо останні 5 рядків з допомогою зрізу, використовуючи iloc:



Оберемо перший елемент стовпця event\_name, використовуючи loc. Це мітко-орієнований вибір даних. Перший параметр методу – індекс рядка, другий – назва стовпця:



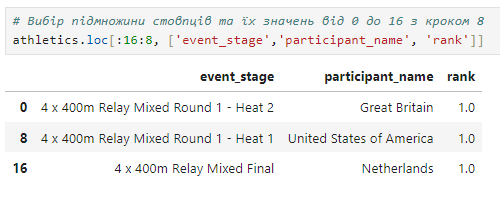
Оберемо підмножину стовпців та усіх їх значень з допомогою loc. Перший параметр методу – індекси рядків (тут обрано усі рядки), другий – назви стовпців, що задані списком:



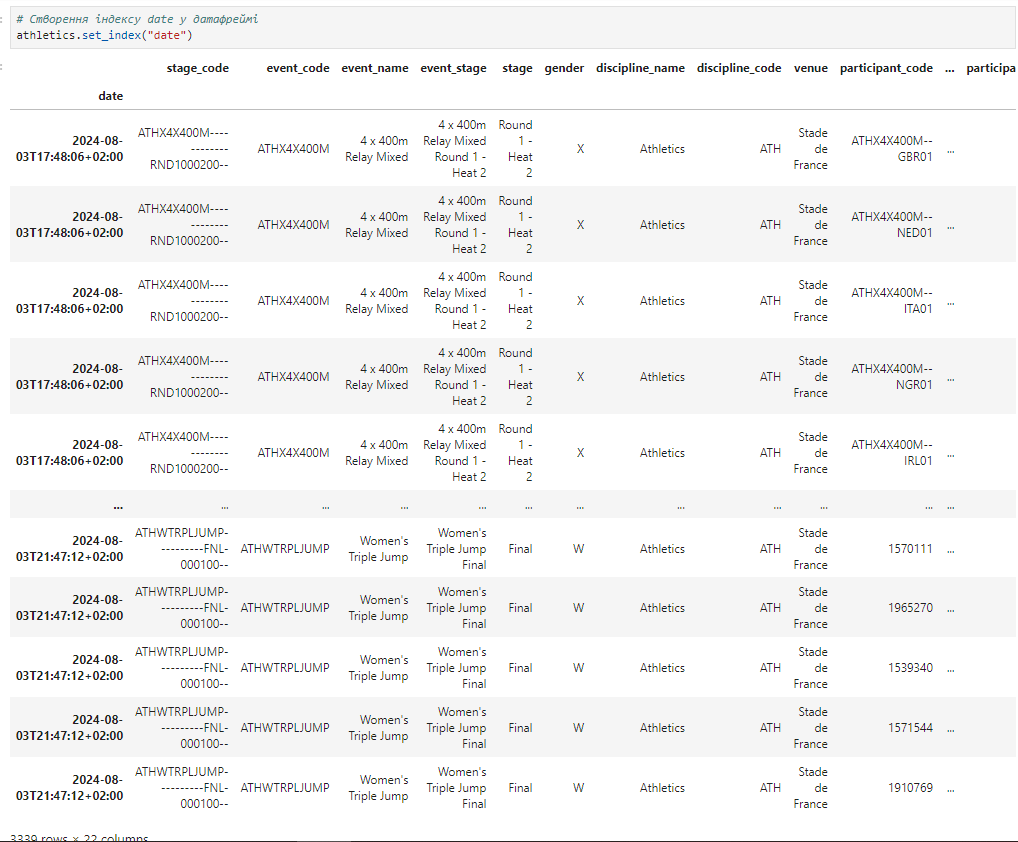
Оберемо підмножину стовпців та їх перших 8 значень з допомогою loc:



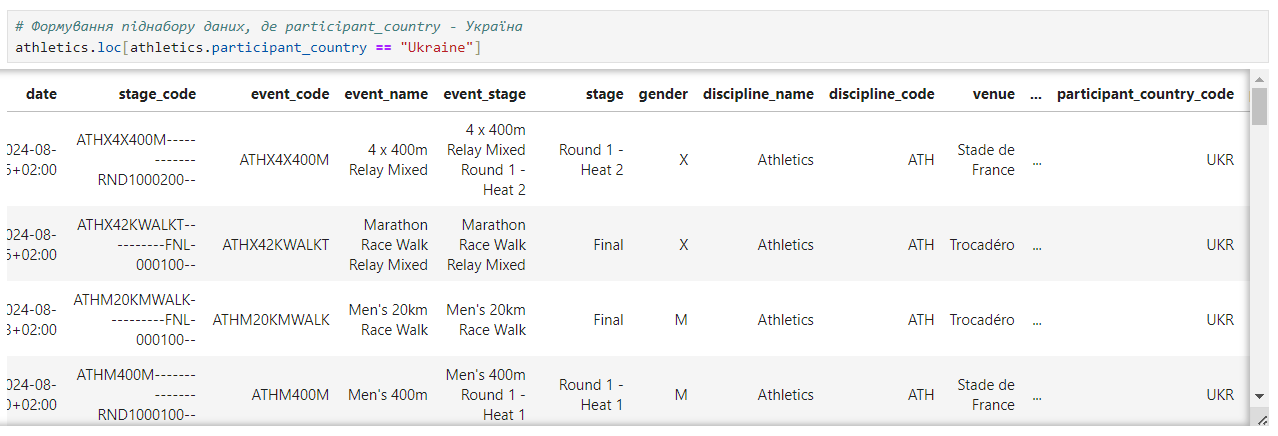
Оберемо підмножину стовпців та їх значень з індексом від 0 до 16 з кроком 8 з допомогою loc:



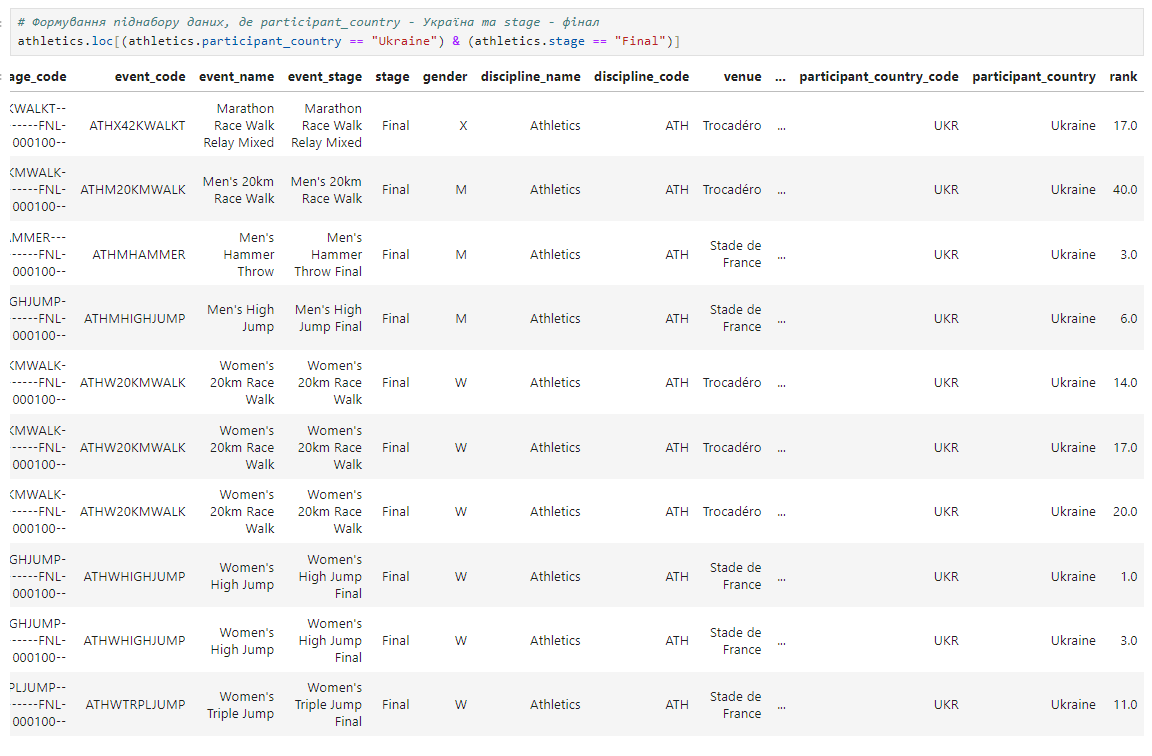
Створимо індекс за датою для наших даних з допомогою методу set\_index. Він встановлює значення індексів для рядків:



Отримаємо усі події зі змагань, де брала участь Україна з допомогою loc та простої умови:



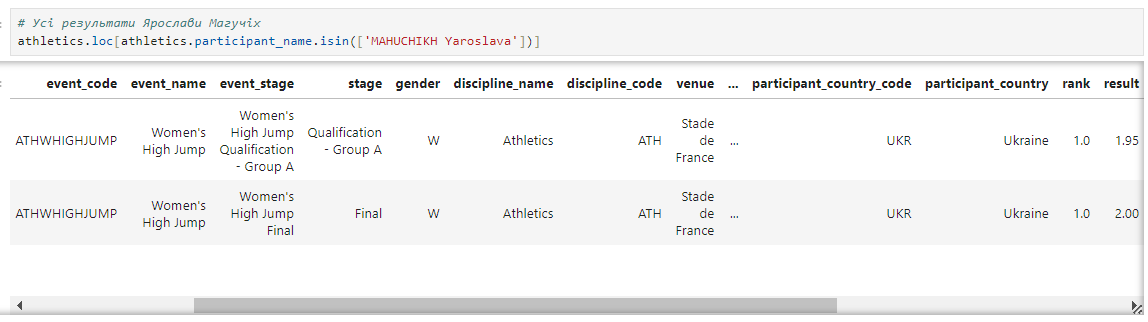
Отримаємо усі виступи України у фіналах змагань з атлетики з допомогою loc та складної умови (оператора &):



Отримаємо усі виступи у жіночих стрибках у висоту або ті, які мають змішаний тип з допомогою loc та складної умови (оператора |):



Отримаємо усі виступи Ярослави Магучіх на змаганнях з допомогою loc та умовного селектора isin, що дозволяє вибрати дані, значення яких є у списку значень:



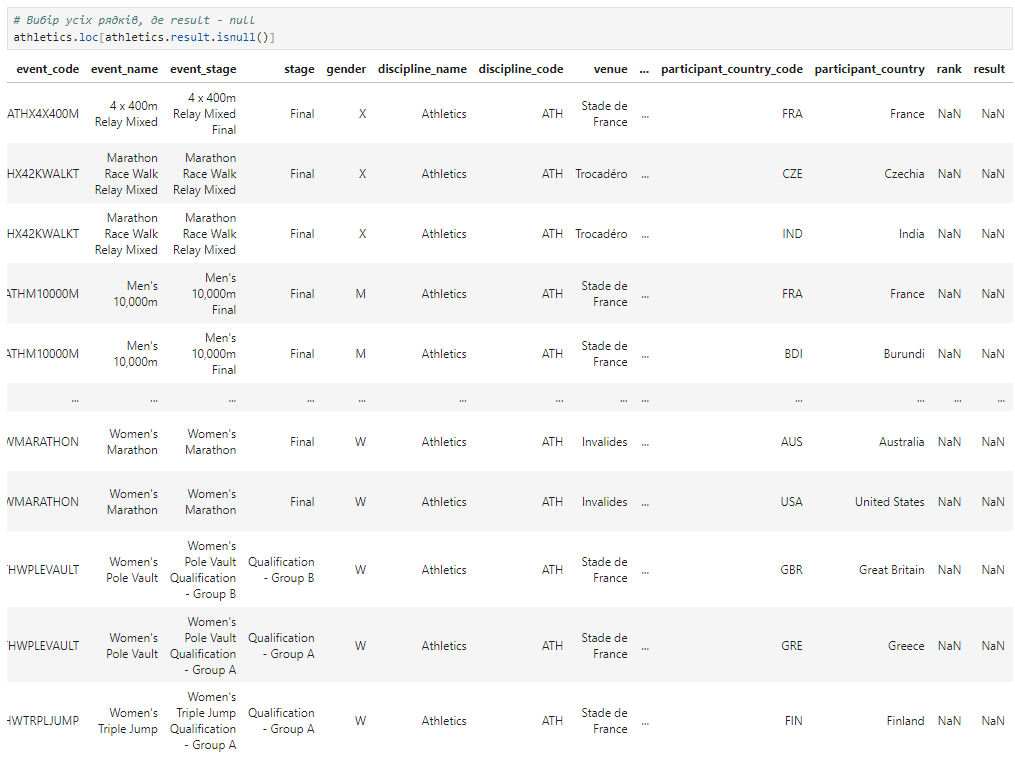
Усі виступи у чоловічих та жіночих стрибках у висоту з допомогою loc та умовного селектора isin:



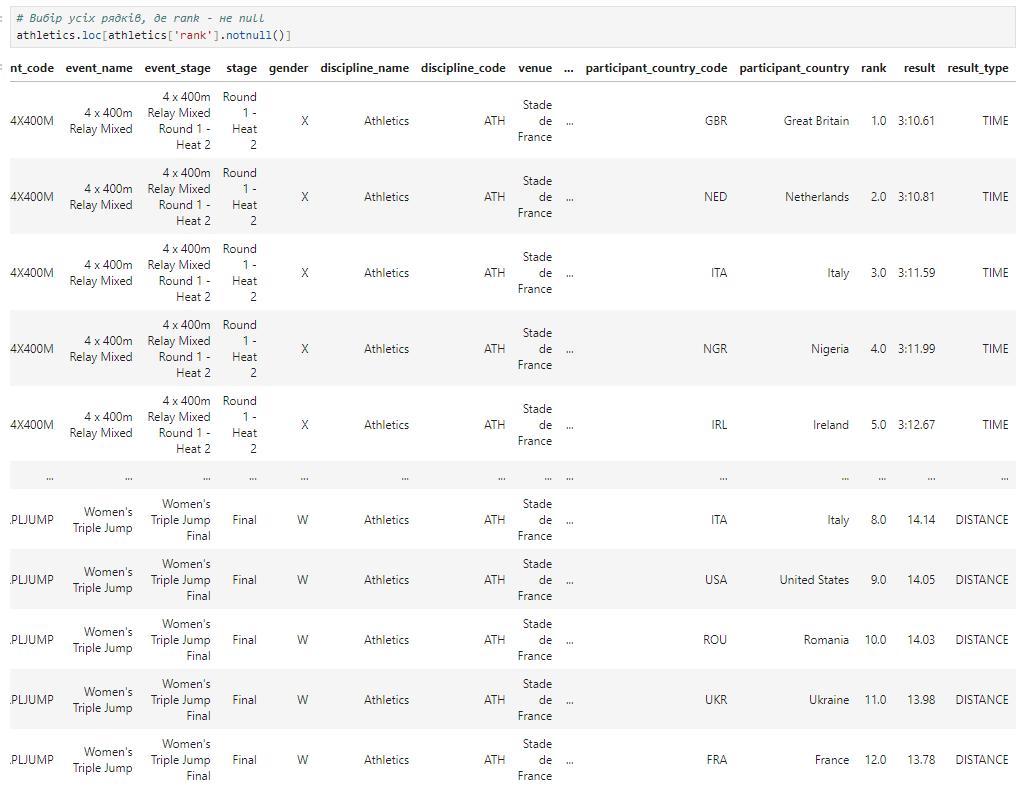
Виведемо усі виступи без змішаних дисциплін з допомогою loc та умовного селектора isin та заперечення ~:



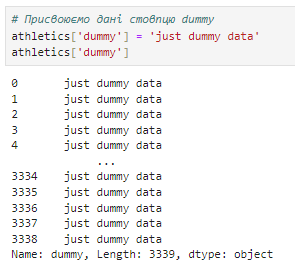
Виберемо усі виступи, які мають результат null з допомогою loc та isnull, який виділяє значення, які є порожніми:



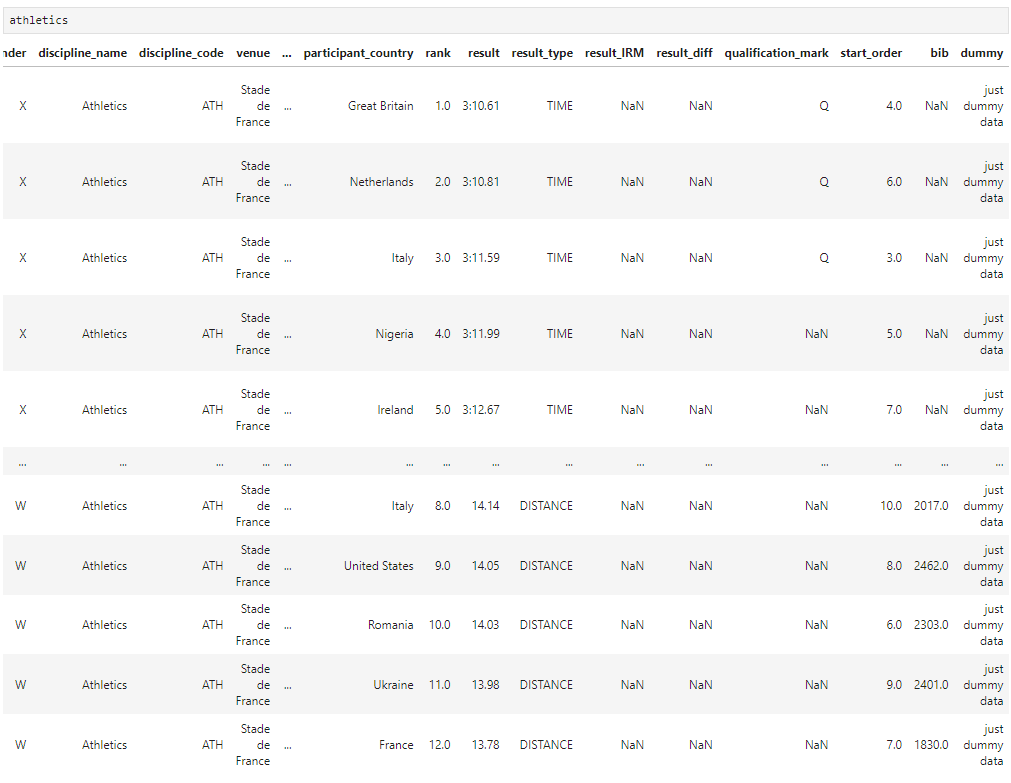
Виберемо усі виступи, де rank не є null з допомогою loc та notnull, який виділяє значення, які не є порожніми:



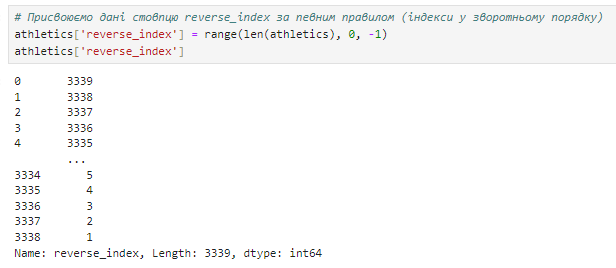
Присвоїмо рядкові дані стовпцю dummy:



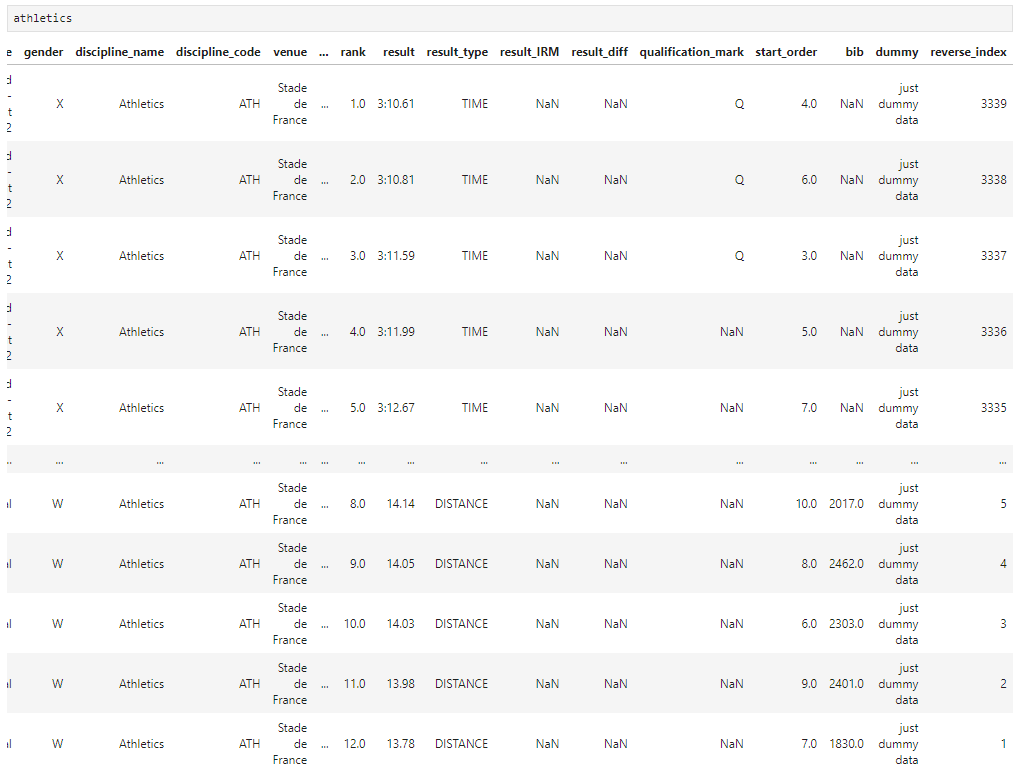
Бачимо, що дані присвоєно успішно:



Присвоїмо стовпцю reverse\_index індекси рядків у зворотному порядку:



Бачимо, що дані присвоєно успішно:



**Висновки**

Під час виконання лабораторної роботи я освоїв основні навички із застосування методів індексування, вибору та редагування набору даних.