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 № 1 з дисципліни

«Основи науки про дані»

«Підсумкові функції та відображення»

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

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

Київ 2024

**Завдання**

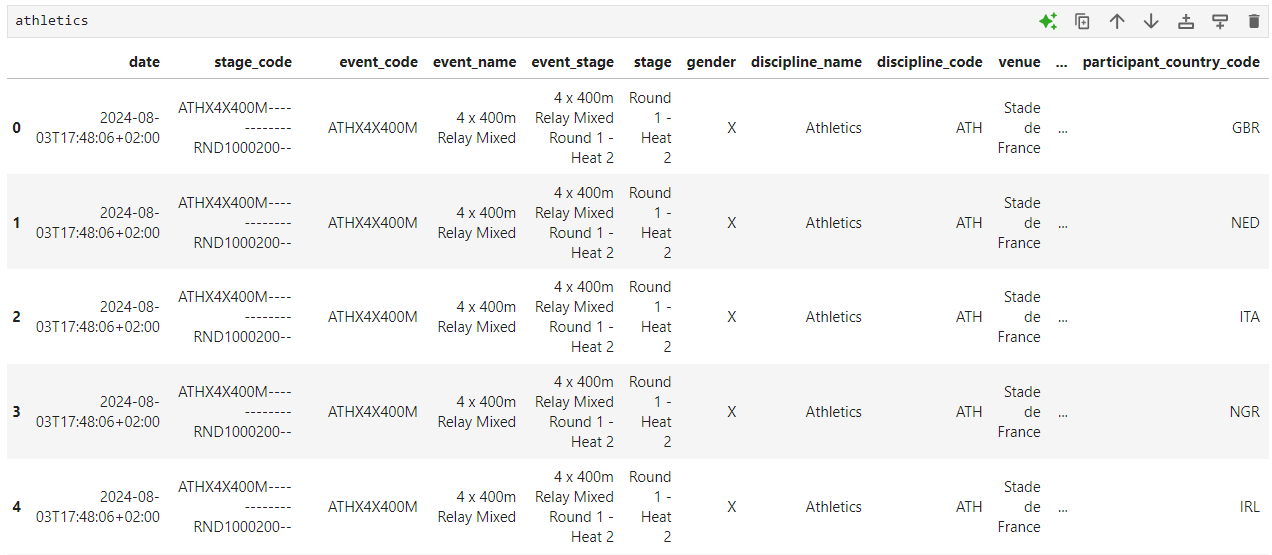
1. Детально розглянути наведені теоретичні відомості до даної лабораторної роботи.
2. Ознайомитися з кодом, що наведений у Jupyter notebook DD\_lab1.ipynb.
3. Виберіть набір даних відповідно до номеру у списку з таблиці 1.
4. Застосувати до вибраного набору всі підсумкові функції та методи відображень, які розглянуті у теоретичних відомостях до лабораторної роботи.

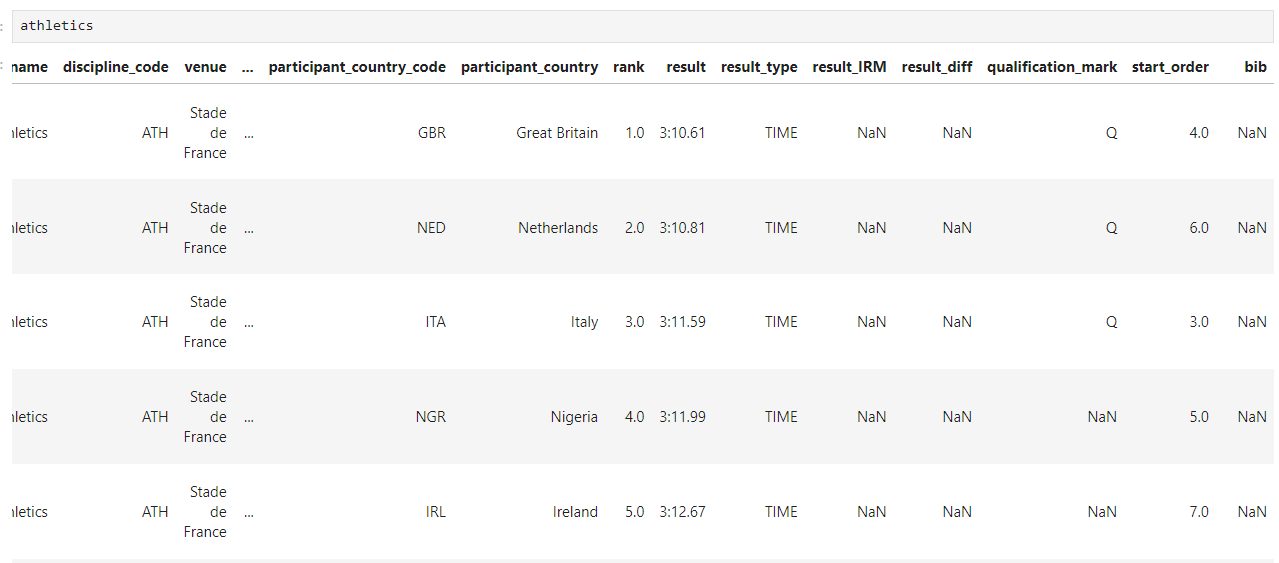
**Виконання**

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>)

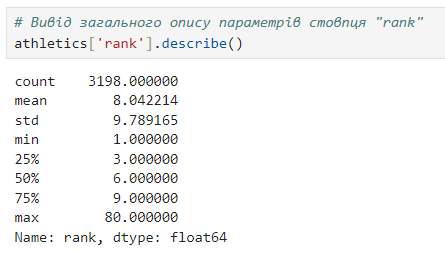
Працюватимемо з файлом Athletics.csv, що містить дані про змагання з атлетики.

Датасет має такий вигляд:





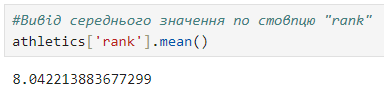
Застосуємо до стовпця “rank” метод describe. Цей метод виводить загальний опис параметрів стовпця залежно від типу даних. Стовпець “rank” містить чисельні дані, тому серед описаних параметрів бачимо ті, які характерні для числових величин, наприклад середнє значення та середнє квадратичне відхилення:



Застосуємо до стовпця “participant\_country” метод unique. Цей метод виводить тільки унікальні елементи у стовпці. При застосуванні цього методу ми дізнаємося список усіх країн, що брали участь у змаганнях з атлетики.



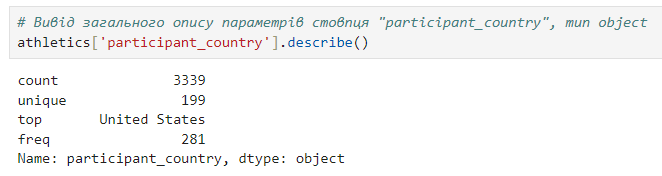
Застосуємо до стовпця “rank” метод mean. Цей метод визначає середнє значення по числовому стовпцю. Так ми дізнаємося середнє місце по таблиці.



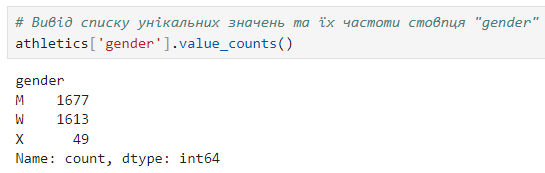
Застосуємо до стовпця “event\_name” метод unique. При застосуванні цього методу ми дізнаємося список усіх видів змагань з атлетики на Олімпійських іграх.



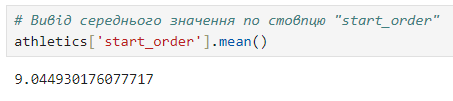
Застосуємо до стовпця “participant\_country” метод describe. Стовпець “rank” містить рядкові дані, тому серед описаних параметрів кількості значень, кількості унікальних значень, значення, яке найбільше зустрічається та його частота:



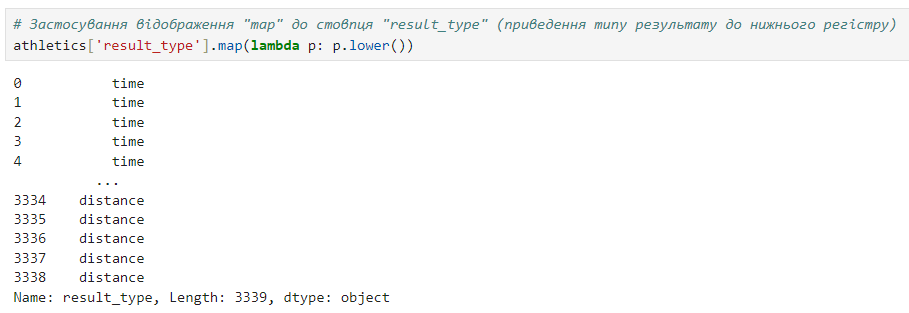
Застосуємо до стовпця “gender” метод value\_counts. Цей метод обраховує частоту унікальних даних для стовпця. Так ми дізнаємося, скільки змагань було проведено для чоловіків, жінок, чи змішаного типу.



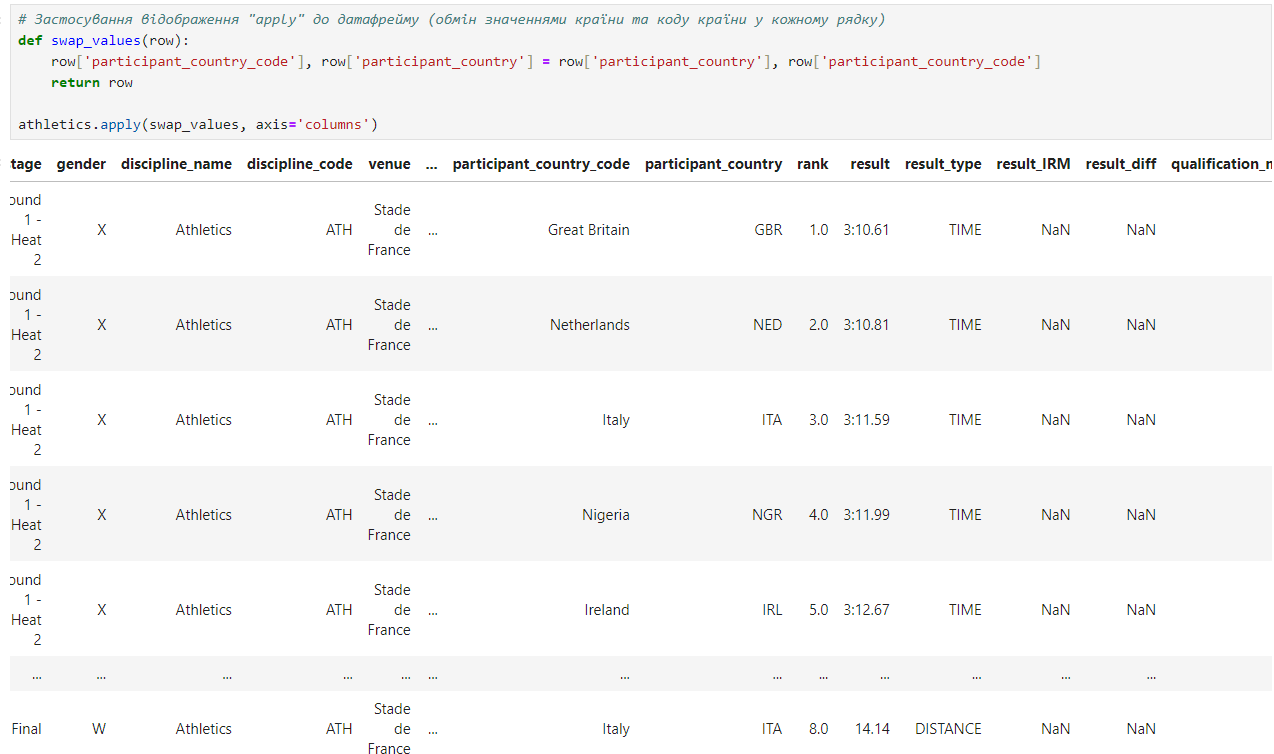
Застосуємо до стовпця “start\_order” метод mean. Так ми дізнаємося середню стартову позицію у таблиці.



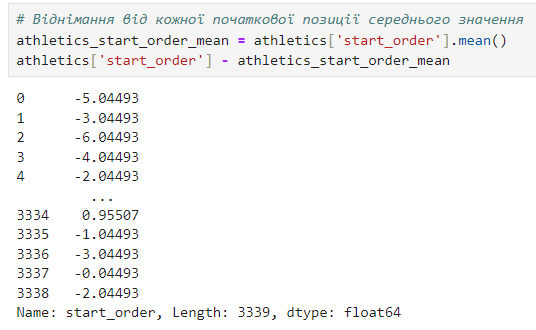
Застосуємо до стовпця “result\_type” відображення map. Це відображення (тобто певна задана функція перетворення) застосовується до кожного елемента стовпця та повертає нову серію з перетвореними значеннями. Для демонстрації приведемо усі значення стовпця до нижнього регістру.



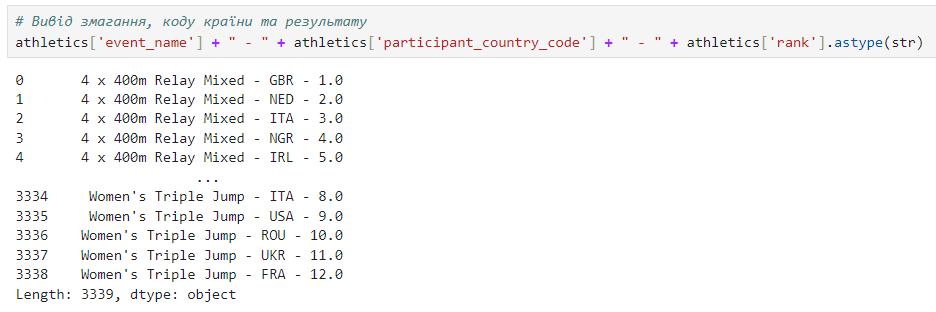
Застосуємо відображення apply. Це відображення (тобто певна задана функція перетворення) застосовується до кожного рядка (якщо задано параметр axis=“columns”) або стовпця (якщо задано параметр axis=“index”) та повертає новий датафрейм з перетвореними даними. Для демонстрації обміняємо значення країни та її коду у кожному рядку.



Застосуємо також вбудовані операції. Спочатку знайдемо середнє значення для start\_order, а потім віднімемо це значення від реальних позицій.



Виведемо назву змагання, код країни та результат з допомогою вбудованих операцій.



**Висновки**

Під час виконання лабораторної роботи я освоїв основні навички із застосування підсумкових функцій, таких як mean, describe, unique, value\_counts, відображень map та apply і вбудованих операцій над стовпцями.