**Фінальний проект**

**Частина 1. Building an End-to-End Streaming Pipeline**

**Изображение выглядит как текст, меню, снимок экрана

Автоматически созданное описание**

Зчитуємо дані атлетів з бази.

**Изображение выглядит как текст, меню, снимок экрана

Автоматически созданное описание**

Видаляємо пусті значення

**Изображение выглядит как снимок экрана, текст

Автоматически созданное описание**

Записуємо в кафка-топік та зчитуємо дані по результатам змагань з перетворенням даних у датафрейм.

**Изображение выглядит как текст, снимок экрана, меню

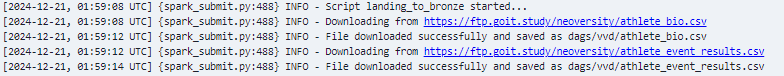
Автоматически созданное описание**

Результат потокового обчислення середніх ваги та росту атлетів для кожного виду спорту, типу медалі або її відсутності, статі, країни.

**А ось п.6 з першої частини завдання я не зміг побороти.**

**Частина 2. Building an End-to-End Batch Data Lake**

**landing\_to\_bronze.py:**



Зчитуємо таблиці

Изображение выглядит как текст, Шрифт, линия, График

Автоматически созданное описание

Скрін відпрацьованого DAG-файлу

Изображение выглядит как текст, документ, меню, снимок экрана

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, меню, документ

Автоматически созданное описание

Після запису у форматі parquet робимо виводи в консоль по 30 рядків таблиць та їхні структури.

**silver\_to\_gold.py:**

Изображение выглядит как текст, документ, меню, бумага

Автоматически созданное описаниеИзображение выглядит как текст, меню, документ, число

Автоматически созданное описание

Очищений текст з дедуплікацією.

**silver\_to\_gold.py:**

Изображение выглядит как текст, меню, документ

Автоматически созданное описание

Цільовий файл. У структурі видно типи стовпчиків (переведено у double)