|  |  |  |  |
| --- | --- | --- | --- |
| МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені Тараса Шевченка  ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ  **Кафедра програмних систем і технологій**  Дисципліна  **«Ймовірнісні основи програмної інженерії»**  **Лабораторна робота № 1**  **«Центральні тенденції та міра дисперсії»** | | | |
| **Виконав:** | Мишко Іван Леонідович | **Перевірила**: | Марцафей Анна Сергіївна |
| Група | ІПЗ-24(2) | Дата перевірки |  |
| Форма навчання | денна | Оцінка |  |
| Спеціальність | 121 |
| 2022 | | | |

**Мета**: Навчитись використовувати на практиці набуті знання про центральні тенденції та міри.

**Завдання 1**

1. Побудувати таблицю частот та сукупних частот для переглянутих фільмів.

Визначити фільм, який був переглянутий частіше за інші.

2. Знайти Моду та Медіану заданої вибірки.

3. Порахувати Дисперсію та Середнє квадратичне відхилення розподілу.

4. Побудувати гістограму частот для даного розподілу.

5. Зробити висновок з вигляду гістограми, про закон розподілу.

**Хід роботи**

**Завдання 1**

**1. Побудувати таблицю частот та сукупних частот для переглянутих фільмів. Визначити фільм, який був переглянутий частіше за інші.**

Для першої лабораторної роботи було застосовано мову програмування python та деякі бібліотеки, а саме: numpy – для зчитування вхідних даних з txt файлу, вбудовану, collections – для групування вхідних даних та matplotlib – для графічної побудови.

Вхідні дані беруться з наданих файлів до лабораторної роботи за допомогою відповідного коду (рис. 1), де вводиться назва .txt файлу та вхідні дані зберігаються у змінну.

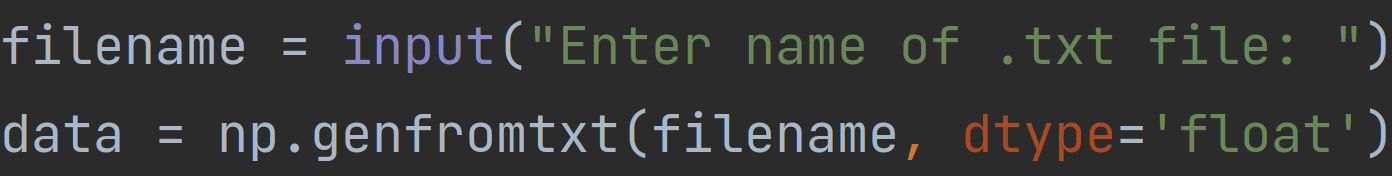


Рисунок 1 | Код

Таблиця частот (рис.3) побудована за допомогою відповідного коду (рис. 2) та складається з стовпчиків: Назва і кількість переглядів, Частота повторень відповідних даних та Сукупна частота (зображує к-сть переглядів помножену на к-сть повторень(частоту)).

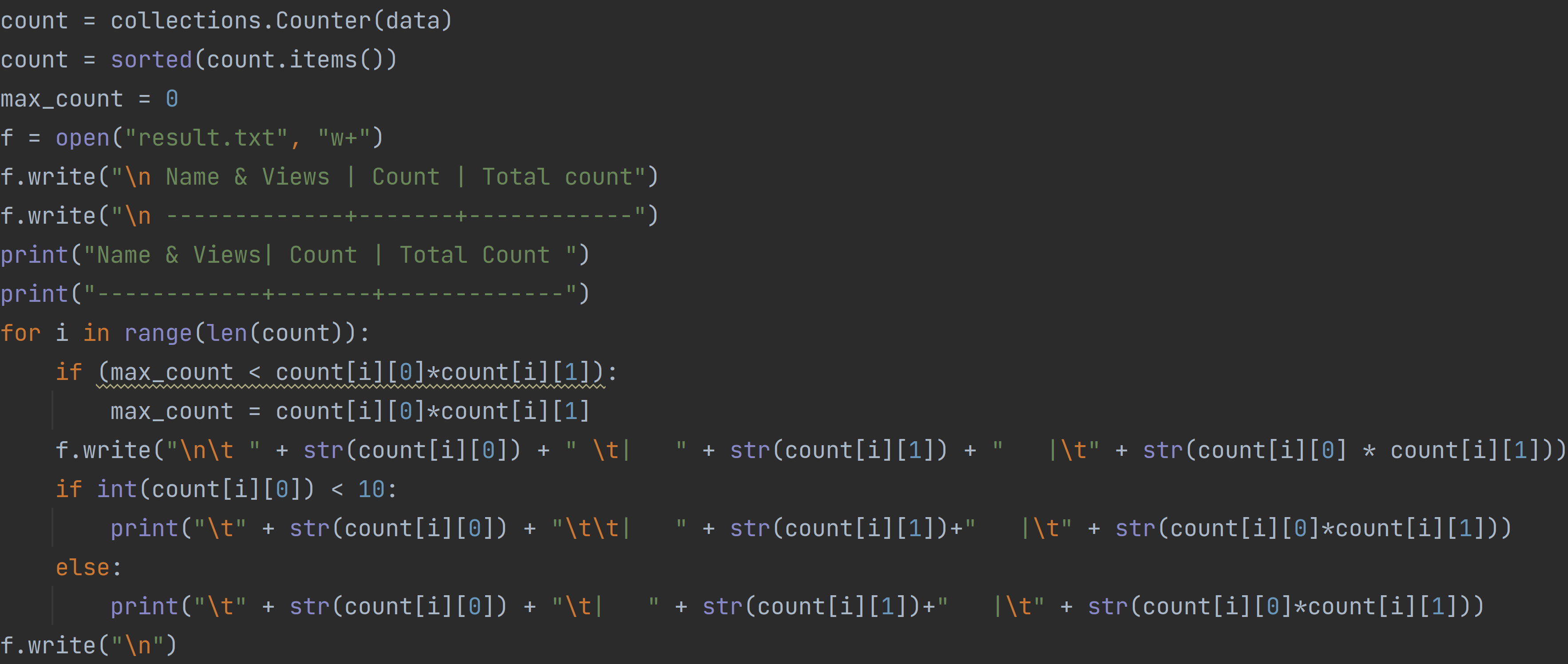


Рисунок 2 | Код

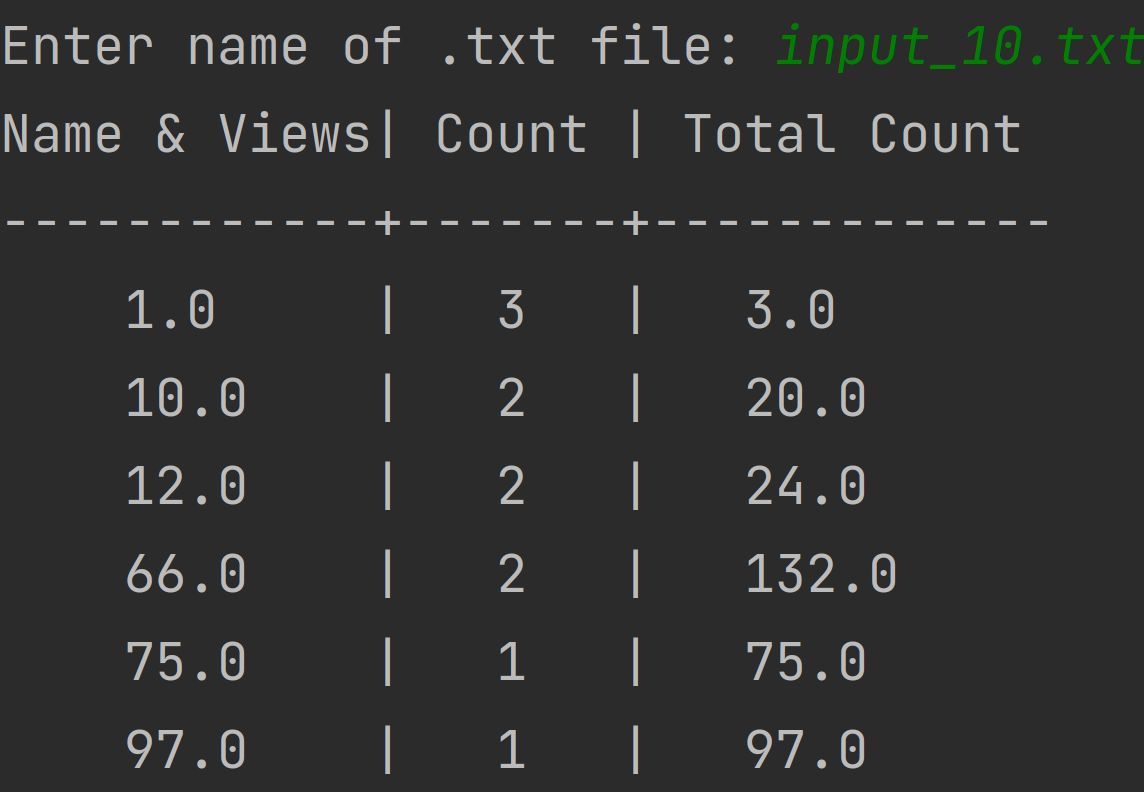


Рисунок 3 | Приклад виконання, файл input\_10.txt

Фільм з найбільшою кількістю переглядів (рис. 5) визначається відповідним кодом (рис. 4).

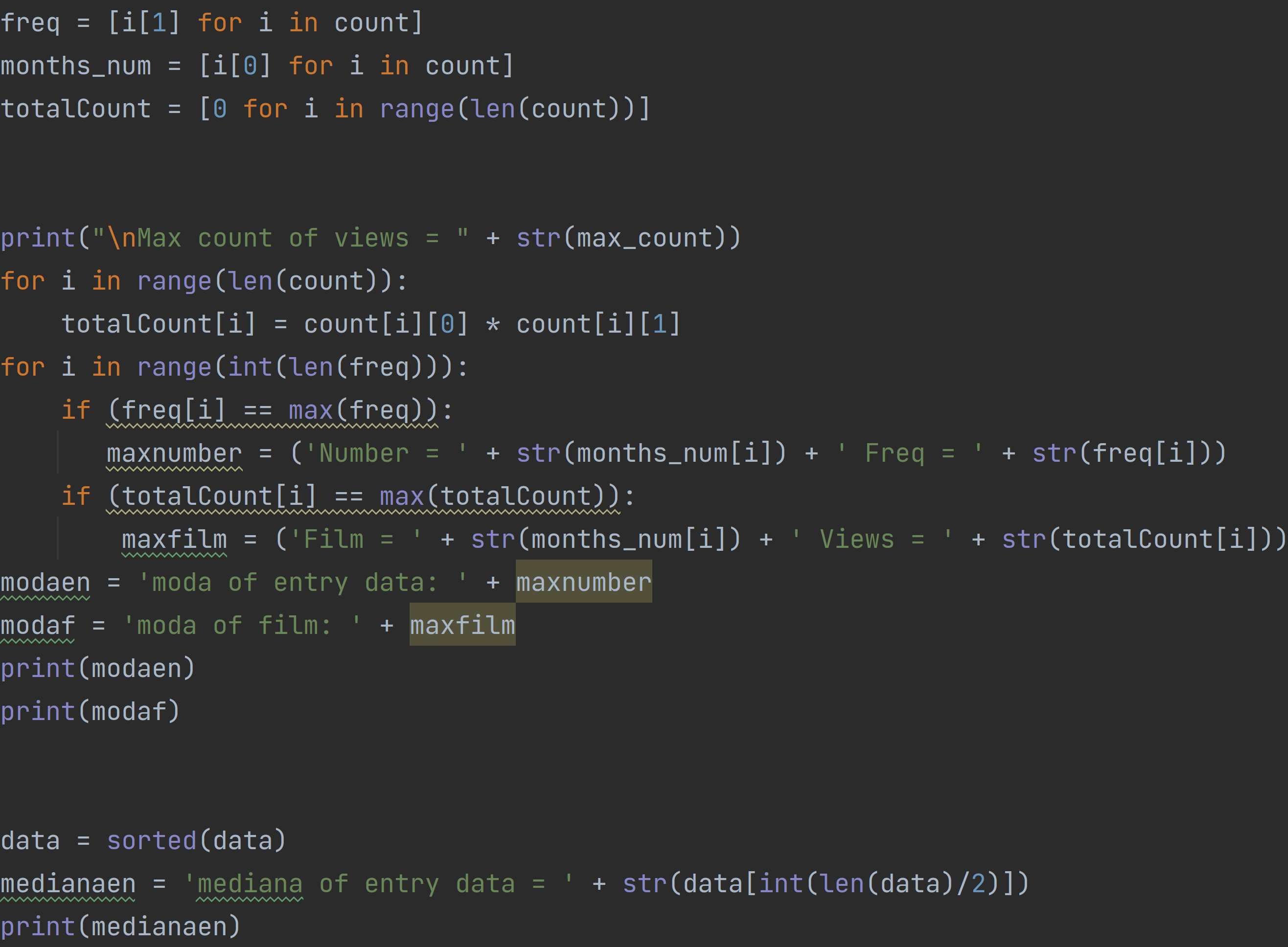


Рисунок 4 | Код



Рисунок 5 | Фільм з найбільшою к-стю переглядів, файл input\_10.txt

На рис. 5 можна побачити, що результат в цілому дорівнює моді вибірки фільмів та сукупної частоти.

**2. Знайти Моду та Медіану заданої вибірки.**

За допомогою коду, зображеному на рис. 4, можна спостерігати відповідні результати (рис. 6).

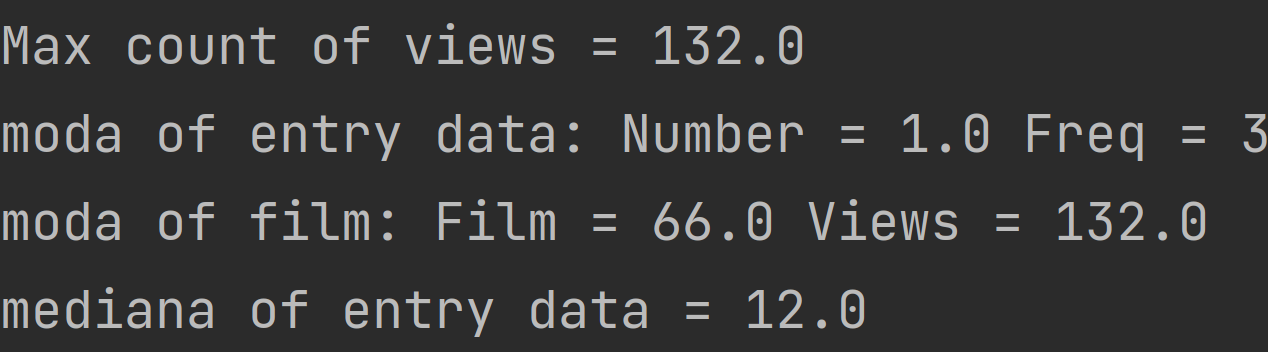


Рисунок 6 | Приклад виконання, файл input\_10.txt

Потрібно зазначити, що на рис. 6 наведено дві моди, а саме: мода з вхідних даних та мода з вибірки фільмів та сукупної частоти (к-сті переглядів), також медіана розрахована з вхідних даних.

Розрахунки були проведені відповідними формулами (рис. 7) (рис. 8).

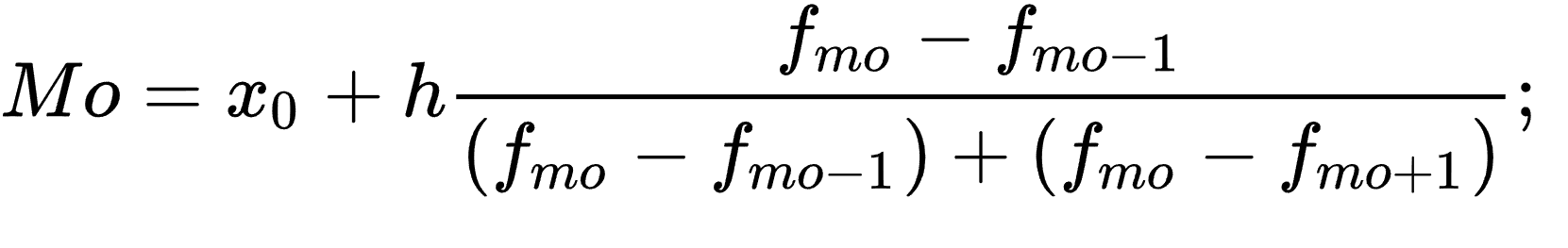


Рисунок 7 | Мода

х0 та h — нижня межа та ширина модального інтервалу,

fmo, fmo-1, fmo+1 — частоти (частки) відповідно модального, передмодального та післямодального інтервалів.

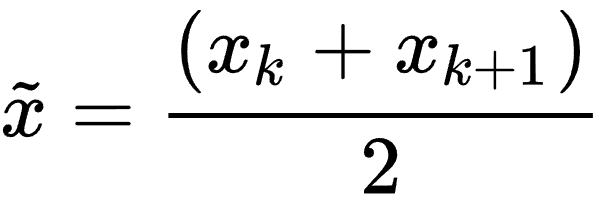


Рисунок 8 | Пошук номеру елементу медіани

Наприклад, для ряду 2 3 5 6 7 медіана дорівнює 5; для ряду 2 3 5 6 7 9 медіана дорівнює (5 + 6)/2 = 5.5.

**3. Порахувати Дисперсію та Середнє квадратичне відхилення розподілу.**

За допомогою наведеного коду (рис. 9) було обраховано дисперсію та середнє квадратичне відхилення розподілу (рис. 10).



Рисунок 9 | Код

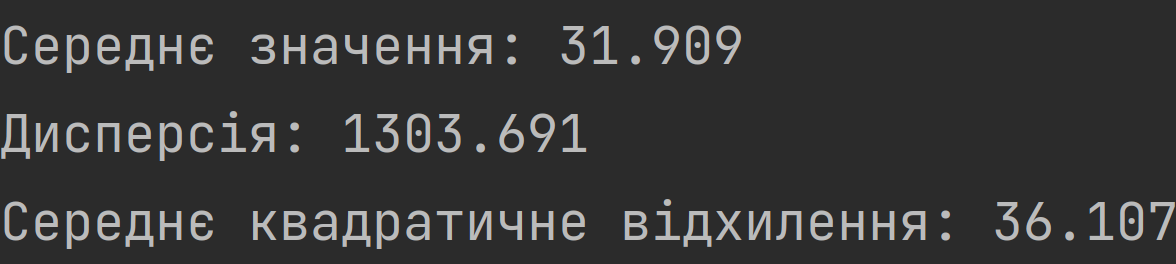


Рисунок 10 | Приклад виконання, файл input\_10.txt

Розрахунки були виконані за відповідними формулами (рис. 11) (рис. 12).

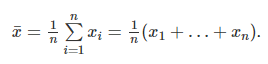


Рисунок 11 | Середнє арифметичне вибірки

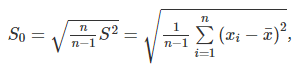


Рисунок 12 | Середньоквадратичне відхилення

Дисперсія буде дорівнювати підкореневому виразу зображеному на рис. 12.

**4. Побудувати гістограму частот для даного розподілу.**

За допомогою відповідного коду (рис. 13) було побудовану потрібну гістограму частот (рис. 14), де за віссю x знаходяться назви фільмів, у вигляді номерів, а за віссю y к-сть переглядів.

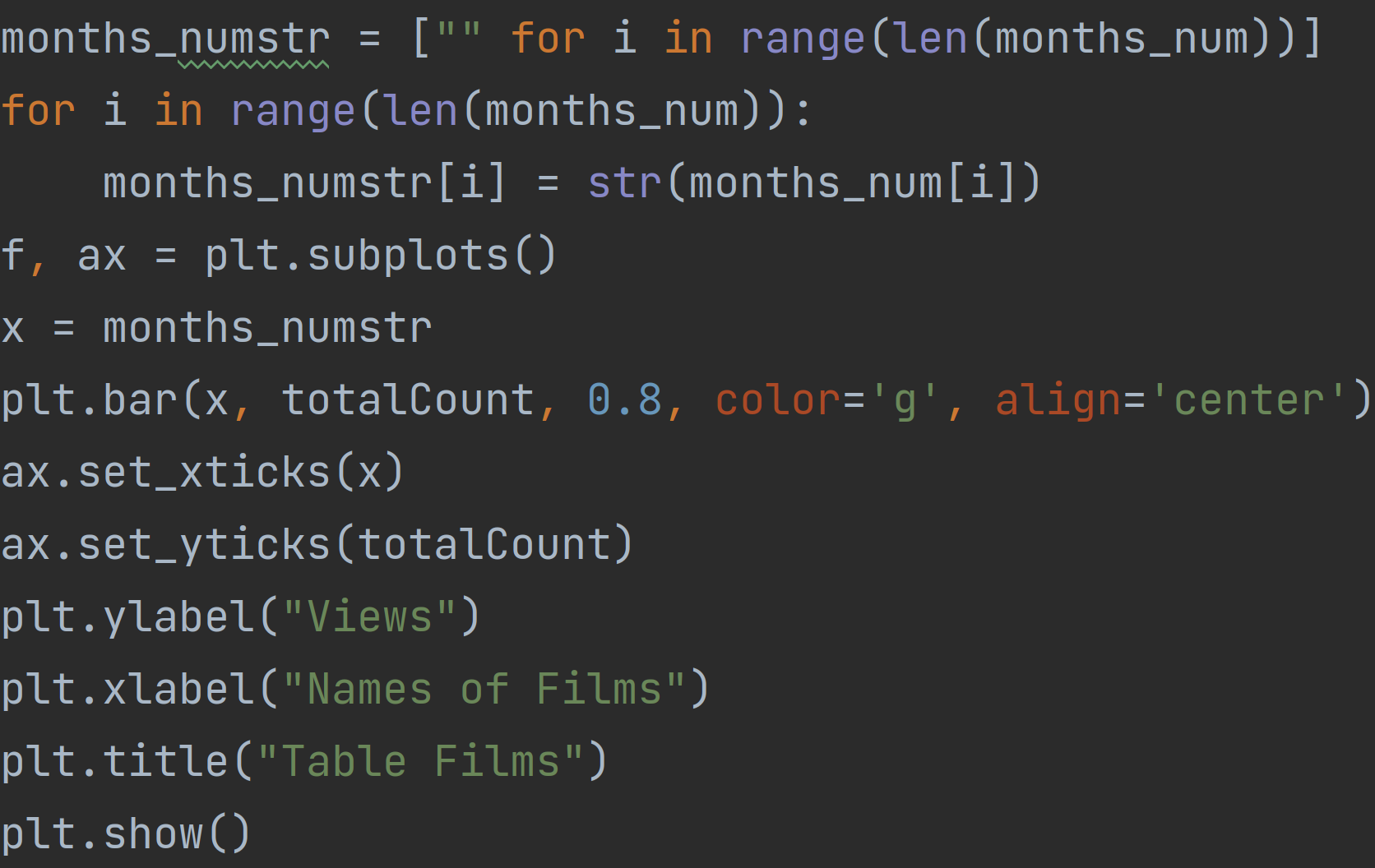


Рисунок 13 | Код

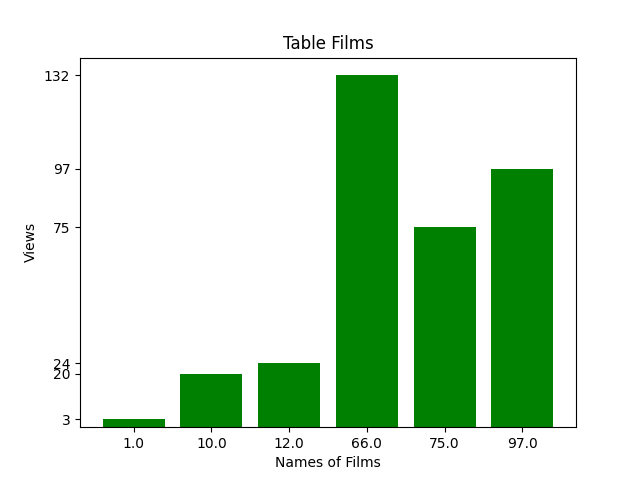


Рисунок 14 | Гістограма частот, файл input\_10.txt

Також, потрібно зазначити, що додатково була пророблена робота з додавання певних даних у відповідний .txt документ (рис. 15).

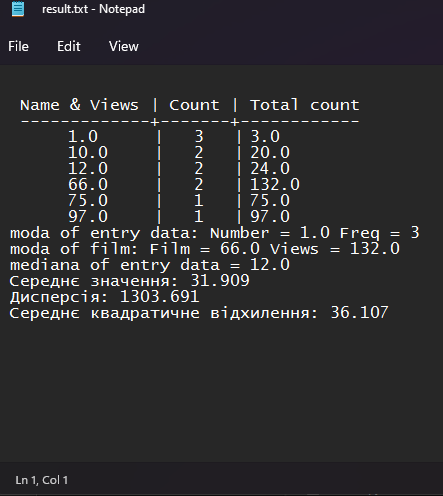


Рисунок 15 | Приклад виконання, файл input\_10.txt

**5. Зробити висновок з вигляду гістограми, про закон розподілу.**

Проаналізувавши отримані гістограми під час виконання лабораторної роботи, можна прийти до висновку, що закон розподілу характеризує випадкову величину з точки зору теорії ймовірностей. Розподіл імовірностей тісно зв'язаний з рядами розподілу частот. Якщо розглядати ряди розподілу (користаючись термінологією теорії ймовірностей) як перелік можливих результатів або груп вимірів і відповідних їм частот кожного результату, то аналогічне визначення можна дати і розподілу ймовірностей. Це перелік можливих результатів або груп вимірів, але, замість спостережуваної частоти, тут вказані ймовірності появи кожного результату.

**Висновок:** Під час виконання першої лабораторної роботи було опановано методи обчислення деяких статистичних даних, використано на практиці набуті знання про центральні тенденції та міри, отримано нові навички з побудови діаграм, проаналізовано виконану роботу та повторено використання бібліотек у мові програмування Python.