Анализ данных и оформление результатов

Синтаксис

**Метод groupby ('название столбца')
для группировки данных**

```
In  df.groupby('название столбца')

    # группировка по столбец_1
    # и вывод столбец_2
    df.groupby('столбец_1')['столбец_2']

    # подсчёт количества в группе
    df.groupby('название столбца').count()

    # подсчёт суммы в группе
    df.groupby('название столбца').sum()
```

**Метод sort_values (by = 'название столбца')
для сортировки таблицы по указанному столбцу**

```
In  # сортировка по возрастанию
    # (значение по умолчанию)
    df.sort_values(by = 'название столбца')

    # сортировка по убыванию
    df.sort_values(by = 'название столбца',
        ascending = False)
```

**Метод max ( ) для определения
максимального значения**

```
In  df['название столбца'].max()
    # максимальное значение в столбце
```

**Метод min ( ) для определения
минимального значения**

```
In  df['название столбца'].min()
    # минимальное значение в столбце
```

**Метод mean ( ) для расчёта
среднего арифметического**

```
In  df['название столбца'].mean()
    # среднее значение по столбцу
```

Метод median ( ) для расчёта медианы

```
In  df['название столбца'].median()
    # медиана по столбцу
```

Словарь

Группировка
Разбиение данных на группы по какому-либо признаку

Стадии группировки можно описать формулой:
split-apply-combine

- *split* (разделить) — разбиение на группы по определённому критерию
- *apply* (применить) — применение какого-либо метода к каждой группе в отдельности, например, подсчёт численности группы методом count() или суммирование вызовом sum()
- *combine* (объединить) — сведение результатов в новую структуру данных. В зависимости от условий разделения и выполнения метода это бывает DataFrame и Series

Логический оператор **&** (аналог **AND**)
служит для соединения нескольких условий в одно при логической индексации

Минимум и **максимум**
Это наименьшее и набольшее числа в наборе.
Показатель максимума или минимума обычно вычисляют по отдельному признаку

Среднее и **Медиана**
Используются для оценки значения в центре выборки. Если выборка равномерна и в ней нет значений, слишком отличающихся от остальных, — среднее подойдет. Но когда есть оторванные от основной массы значения, они сильно смещают среднее вверх. В таком случае используется медиана

Правильно оформленные результаты помогут донести главную мысль и ценность исследования.
Несколько советов по оформлению:

- Показывая, как меняется какой-нибудь параметр во времени, поместите его значения в строке, а столбцами задавайте временные промежутки
- Если нужно показать разнородные признаки для конкурирующих категорий (например, для жанров), то каждой категории отведите отдельную строку, а значения признаков размещайте по столбцам
- Не старайтесь обязательно собрать все данные в одну таблицу: лучше несколько таблиц, чтобы каждая отражала одну важную идею
- Отлично работает детализация от большего к меньшему. К общей сводной таблице прилагайте более подробные. Например, сначала обзорная таблица по всему сервису, затем более детальные: сводки по группам пользователей, по городам и т.п.