

Анализ данных и оформление результатов

Группировка

```
# группировка по column_1
grouped = df.groupby('column_1')

# группировка по column_1 и выбор столбца column_2
col = df.groupby('column_1')['column_2']

# количество значений в каждой группе (без учёта пропусков)
df.groupby('column_1').count()

# сумма значений в каждой группе
df.groupby('column_1').sum()
```

Сортировка

```
# сортировка датафрейма по возрастанию
sorted_df = df.sort_values(by='column')

# сортировка датафрейма по убыванию
sorted_df = df.sort_values(by='column', ascending=False)

# сортировка Series по возрастанию
sorted_series = df['column'].sort_values()

# сортировка Series по убыванию
sorted_series = df['column'].sort_values(ascending=False)
```

Анализ данных и оформление результатов

Подсчёт метрик

```
# сумма значений в столбце
df['column'].sum()

# количество значений в столбце (без учёта пропусков)
df['column'].count()

# минимальное значение в столбце
df['column'].min()

# максимальное значение в столбце
df['column'].max()

# среднее значение в столбце
df['column'].mean()

# медианное значение в столбце
df['column'].median()
```

Глоссарий

Split-apply-combine — формула, которой выражают стадии группировки:

- **разделить (split)** — сначала данные разбивают на группы по определённому критерию;
- **применить (apply)** — затем к каждой группе применяют методы вычисления, например считают элементы группы методом `count()` или их суммы методом `sum()`;
- **объединить (combine)** — наконец, результаты сводят в новую структуру данных, таблицу или `Series`.