

Поиск дубликатов

Подсчет дубликатов

```
df.duplicated().sum() # общее количество строк-дубликатов во всем датафрейме
df['name'].duplicated().sum() # общее количество значений-дубликатов в отдельном столбце
df['name'].value_counts() # количество дубликатов для уникальных значений столбца
```

Сброс индекса датафрейма

```
# drop=False (аргумент по умолчанию) создаёт в датафрейме новый столбец со старыми индексами
# drop=True удаляет старые индексы
stock = stock.reset_index(drop=True)
```

Удаление дубликатов

```
# удаляем из датафрейма все дублирующиеся строки
# по умолчанию остаются только первые вхождения дублирующихся строк
df = df.drop_duplicates()

# удаляем из датафрейма все строки, в которых встречаются дубликаты в столбце 'name'
# оставляем только последние вхождения дублирующихся строк (аргумент keep='last')
df = df.drop_duplicates(subset=['name'], keep='last')

# строки датафрейма остаются на месте, дубликаты в столбце превращаются в пропуски
# дополнительно сбрасываем индексацию изменённого датафрейма
df['name'] = df['name'].drop_duplicates().reset_index()
```

Приведение значений столбца к нижнему регистру

```
stock['item_lowercase'] = stock['item'].str.lower()
```