9. Pandas для анализа данных

Синтаксис

Импорт библиотеки

```
import pandas as pd
```

Конструктор `DataFrame()` для создания таблицы

```
pd.DataFrame(data=data, columns=columns)
# аргумент data — список с данными,
# аргумент columns — список с названиями столбцов
```

Метод `head()` для вывода первых строк таблицы

```
df.head() # первые 5 строк
df.head(10) # первые 10 строк
```

Метод `tail()` для вывода последних строк таблицы

```
df.tail() # последние 5 строк
df.tail(15) # последние 15 строк
```

Чтение файлов формата `.csv`

```
df = pd.read_csv('путь к файлу')
```

Атрибуты и методы датафрейма

```
df.columns # выводит названия столбцов
df.shape # выводит размер таблицы
df.dtypes # выводит информацию о типах данных в таблице
df.info() # метод выводит сводную информацию о таблице
```

Полная и сокращённая записи при индексации датафрейма

	Полная запись	Сокращённая запись
Один столбец	.loc[:, 'genre']	df['genre']
Несколько столбцов	.loc[:, ['genre', 'Artist']]	df[['genre', 'Artist']]
Все строки, начиная с заданной	.loc[0:]	df[0:]
Все строки до заданной	.loc[:3] (включая 3)	df[:3] (не включая 3)
Несколько строк подряд (срез)	.loc[2:5] (включая 5)	df[2:5] (не включая 5)
Одна ячейка	.loc[7, 'genre']	-
Одна строка	.loc[1]	-
Несколько столбцов подряд (срез)	.loc[:, 'total_play': 'genre']	-

Индексация в Series

	Полная запись	Сокращённая запись
Один элемент	total_play.loc[7]	total_play[7]
Несколько элементов	total_play.loc[[5, 7, 10]]	total_play[[5, 7, 10]]
Несколько элементов подряд (срез)	total_play.loc[5:10] (включая 10)	total_play[5:10] (не включая 10)
Все элементы, начиная с заданного	total_play.loc[1:]	total_play[1:]
Все элементы до заданного	total_play.loc[:3] (включая 3)	total_play[:3] (не включая 3)

Глоссарий

Библиотека — набор готовых методов для решения распространённых задач.

`.csv` — формат файла (от англ. Comma-Separated Values, «значения, разделённые запятой»). Каждая строка представляет собой одну строку таблицы, где данные разделены запятыми.

Датафрейм — двумерная структура данных pandas, где у каждого элемента есть два индекса: по строке и по столбцу.

Объект `Series` — одномерная структура данных pandas, её элементы можно получить по индексу.

Конструктор — это метод, который создаёт новые объекты. Например, конструктор `DataFrame()` в pandas создаёт датафреймы.