шпаргалка по библиотеке **Pandas** в Python, которая поможет Вам быстро вспомнить основные компоненты и функции:

Основные компоненты библиотеки Pandas

1. **Структуры данных**:
   * **Series**: Одномерный массив, который может содержать данные любого типа (целые числа, строки, числа с плавающей запятой и т.д.). Каждому элементу присваивается индекс.
   * **DataFrame**: Двумерная таблица, состоящая из строк и столбцов. Каждый столбец может содержать данные разных типов. DataFrame можно рассматривать как словарь объектов Series.
2. **Импорт библиотеки**:

**import** pandas **as** pd

Основные операции с данными

* **Создание Series и DataFrame**:

# Создание Series

s = pd.Series([1, 2, 3, 4])

# Создание DataFrame

df = pd.DataFrame({

'A': [1, 2, 3],

'B': [4, 5, 6]

})

* **Чтение данных**:

# Чтение CSV файла

df = pd.read\_csv('file.csv')

# Чтение Excel файла

df = pd.read\_excel('file.xlsx')

# Чтение данных из JSON

df = pd.read\_json('file.json')

# Чтение данных из SQL базы данных

import sqlite3

conn = sqlite3.connect('database.db')

df = pd.read\_sql\_query('SELECT \* FROM table\_name', conn)

# Чтение данных из HTML

df\_list = pd.read\_html('http://example.com') # Возвращает список DataFrame

# Чтение данных из Parquet файла

df = pd.read\_parquet('file.parquet')

# Чтение данных из HDF5 файла

df = pd.read\_hdf('file.h5', 'key')

* **Запись данных**:

# Запись в CSV файл

df.to\_csv('file.csv', index=**False**)

# Запись в Excel файл

df.to\_excel('file.xlsx', index=**False**)

Основные методы и функции

* **Просмотр данных**:

df.head() # Первые 5 строк

df.tail() # Последние 5 строк

df.info() # Информация о DataFrame

df.describe() # Статистические характеристики

* **Индексация и выбор данных**:

df['A'] # Выбор столбца A

df.loc[0] # Выбор первой строки

df.iloc[0, 1] # Выбор элемента по индексу

* **Фильтрация данных**:

df[df['A'] > 1] # Фильтрация строк, где A > 1

* **Обработка пропусков**:

df.dropna() # Удаление строк с пропусками

df.fillna(0) # Замена пропусков на 0

* **Группировка данных**:

df.groupby('A').sum() # Группировка по столбцу A и суммирование

* **Слияние и объединение**:

pd.merge(df1, df2, on='key') # Слияние двух DataFrame по ключу

pd.concat([df1, df2]) # Объединение DataFrame

Полезные функции

* **Сортировка**:

**PythonКопировать**

df.sort\_values(by='A') # Сортировка по столбцу A

* **Применение функций**:

**PythonКопировать**

df['A'].apply(**lambda** x: x + 1) # Применение функции к столбцу A

Заключение

Pandas — мощная библиотека для анализа и манипуляции данными. Она предоставляет множество инструментов для работы с данными, что делает её незаменимой в области анализа данных и машинного обучения. Не забывайте изучать документацию для более глубокого понимания возможностей библиотеки!