



Домашнее задание: Работа с табличными данными (pandas) на примере клинических случаев

Домашнее задание

Цель

Сделать базовый анализ табличных данных с использованием библиотеки `pandas`, закрепив навыки:

- создания `DataFrame` из стандартных структур Python и файлов;
- первичного осмотра данных;
- работы с индексами;
- доступа и изменения данных в таблице.

Задание

Шаг 1. Подготовка среды

1. Сделать настройку среды выполнения:
 - Google Colab
 - или локальное окружение (Python + VS Code).
2. Сделать импорт библиотеки `pandas`.
3. Приложить результат в виде работающего ноутбука или скрипта.

Шаг 2. Создание DataFrame из стандартных структур Python

1. Сделать набор данных по клиническим случаям, используя стандартные структуры Python:
 - список словарей **или**

- словарь списков.
2. Таблица должна содержать не менее **8–10 записей** и следующие столбцы:
 - `episode_id` — номер клинического случая
 - `patient_name` — имя пациента
 - `age` — возраст
 - `diagnosis` — диагноз
 - `doctor` — лечащий врач
 - `outcome` — исход лечения
 3. Сделать `DataFrame` из подготовленных данных.
 4. Приложить результат (вывод таблицы).

Шаг 3. Первичный осмотр данных

1. Сделать вывод:
 - первых строк таблицы;
 - информации о типах данных;
 - размерности таблицы.
2. Сделать проверку:
 - наличия пропущенных значений;
 - корректности типов данных.
3. Приложить результаты выполнения команд.

Шаг 4. Работа с индексами

1. Сделать установку столбца `episode_id` в качестве индекса.
2. Сделать:
 - сброс индекса;
 - повторную установку индекса.
3. Сделать краткий комментарий в коде, зачем используется индекс.
4. Приложить результат.

Шаг 5. Доступ к данным

1. Сделать выборку:
 - данных одного пациента по имени;
 - всех клинических случаев с выбранным диагнозом;

- пациентов старше 60 лет.
2. Сделать примеры доступа:
 - через имя столбца;
 - с использованием `.loc`;
 - с использованием `.iloc`.
 3. Приложить результаты выборок.

Шаг 6. Создание DataFrame из файла

1. Сделать CSV-файл с аналогичной структурой данных **или** использовать файл, предоставленный преподавателем.
2. Сделать загрузку файла в `pandas.DataFrame`.
3. Сделать первичный осмотр данных и установку индекса.
4. Приложить результат.

Подсказки по ключевым частям

- Для создания таблицы используйте `pd.DataFrame()`.
- Для первичного осмотра:
 - `.head()`
 - `.info()`
 - `.shape`
- Для проверки пропусков:
 - `.isna().sum()`
- Для работы с индексами:
 - `.set_index()`
 - `.reset_index()`
- Для фильтрации используйте логические условия.

Что проверить перед отправкой (чек-лист)

- Сделано создание `DataFrame` из стандартных структур Python.
- Сделано создание `DataFrame` из CSV-файла.
- Приложены результаты первичного осмотра данных.
- Сделана работа с индексами.

- Сделаны выборки данных разными способами.
- Код выполняется без ошибок.
- Результат оформлен аккуратно и понятно.

Советы по улучшению работы

- Делайте комментарии к каждому логическому блоку кода.
- Используйте понятные названия переменных и столбцов.
- Дополнительно можно сделать:
 - сортировку данных;
 - подсчёт количества клинических случаев по диагнозам;
 - анализ исходов лечения.

Ответ

Ссылка на решение:

<https://colab.research.google.com/drive/1dynBhgyn0Q2LgekidPBDGHB32UohjfgJ?usp=sharing>