



Домашнее задание: Фильтрация, сортировка и преобразование табличных данных (pandas)

Домашнее задание

Цель

Сделать операции фильтрации, сортировки и преобразования табличных данных с использованием логических условий и методов библиотеки `pandas`.

Задание

Шаг 1. Подготовка данных

1. Сделать загрузку набора данных с клиническими случаями:
 - использовать `DataFrame` из предыдущего домашнего задания
2. Сделать проверку структуры таблицы.
3. Приложить вывод первых строк таблицы.

Шаг 2. Условия и логические операции

1. Сделать фильтрацию клинических случаев:
 - пациенты старше 60 лет;
 - пациенты младше 40 лет **и** с определённым диагнозом;
 - пациенты с возрастом от 30 до 70 лет включительно.
2. Использовать логические операторы `>`, `<`, `>=`, `<=`, `==`, `&`, `|`.
3. Приложить результаты каждой фильтрации.

Шаг 3. Фильтрация строк

1. Сделать выборку:

- всех клинических случаев с выбранным диагнозом;
 - всех случаев с неблагоприятным исходом лечения;
 - всех случаев, где возраст пациента отсутствует (если есть пропуски).
2. Использовать разные способы фильтрации (`[]` , `.loc`).
 3. Приложить результаты.

Шаг 4. Сортировка данных

1. Сделать сортировку:
 - по возрасту пациента по возрастанию;
 - по возрасту пациента по убыванию;
 - по диагнозу и возрасту одновременно.
2. Сделать сортировку с указанием порядка для каждого столбца.
3. Приложить результат сортировки.

Шаг 5. Группировка данных

1. Сделать группировку данных:
 - по диагнозу;
 - по лечащему врачу.
2. Для каждой группировки сделать:
 - подсчёт количества клинических случаев;
 - средний возраст пациентов.
3. Приложить результаты группировки.

Шаг 6. Добавление и удаление столбцов

1. Сделать добавление нового столбца, например:
 - возрастная группа пациента;
 - признак пациента старше 60 лет.
2. Сделать удаление одного из добавленных столбцов.
3. Приложить результат.

Шаг 7. Изменение типов данных

1. Сделать проверку типов данных всех столбцов.

2. Сделать изменение типа данных хотя бы одного столбца (например, `age` → `int`, `episode_id` → `str`).
3. Приложить результат до и после изменения типов.

Шаг 8. Переименование столбцов

1. Сделать переименование:
 - одного столбца;
 - нескольких столбцов одновременно.
2. Использовать метод `rename`.
3. Приложить результат.

Подсказки по ключевым частям

- Для фильтрации используйте логические выражения и методы `.loc`.
- Для сортировки:
 - `.sort_values()`
- Для группировки:
 - `.groupby()`
 - `.agg()`
- Для работы со столбцами:
 - добавление: `df["new_column"] = ...`
 - удаление: `.drop()`
- Для изменения типов:
 - `.astype()`
- Для переименования:
 - `.rename()`

Что проверить перед отправкой (чек-лист)

- Сделана фильтрация с использованием логических условий.
- Сделана сортировка данных разными способами.
- Сделана группировка с агрегатными функциями.
- Сделано добавление и удаление столбцов.
- Сделано изменение типов данных.

- Сделано переименование столбцов.
- Код выполняется без ошибок.
- Приложены результаты выполнения всех шагов.

Советы по улучшению работы

- Делайте отдельные ячейки для каждого шага.
- Добавляйте краткие комментарии к сложным операциям.
- Проверяйте результаты после каждого преобразования данных.

Ответ

Ссылка на решение:

https://colab.research.google.com/drive/1tkgYeG9eiNFsXtrna1IFe8Ge9vf7ss0_?usp=sharing