

Тема 3. Библиотека Pandas: управление данными

Практические задания для самостоятельного выполнения

Задания выполняются по вариантам. Формулировки заданий общие для всех вариантов; конкретные условия, указанные в общей формулировке, выбираются в соответствии с номером своего варианта (для удобства представлены в отдельном файле).

Результаты выполнения заданий необходимо представить в виде двух файлов:

- 1) ноутбук в формате `ipynb`, содержащий программный код, результаты его выполнения, а также все необходимые пояснения, выводы и комментарии (в текстовых ячейках);
- 2) файл в формате `pdf`, полученный путем экспорта (или вывода на печать) ноутбука из п. 1).

Внимание: в названии файлов должна обязательно присутствовать фамилия автора (например, 'Петров_03.ipynb'). Безымянные работы проверяться не будут.

Задача 1 (максимум 0,5 балла).

Используя фрейм **df** из задачи №2 во второй лабораторной работе как донора, сформировать фрейм **df_new**, оставив строки и колонки исходного фрейма, условия для которых указаны в вариантах для каждого студента.

Полученный фрейм (без индексов строк, но с названиями колонок) сохранить в папке с файлами этого задания на компьютере под именем 'result.csv'. Фрагмент скриншота папки с файлами задания (с указанием размера файлов) вставить как рисунок в блокнот с решением данного задания этой лабораторной работы. Предварительно вывести на экран первые и последние 5 строк нового фрейма с указанием типа данных колонок.

Задача 2 (максимум 0,5 балла).

Загрузите из интернета информацию о пассажирах теплохода «Титаник» из файла “titanic.csv” (как это сделать, описано в первом блокноте 2-го занятия). Описание этого датасета есть в презентации ко второму занятию. Выполните операцию со строками этого фрейма в соответствии с номером варианта.

Пояснение:

- операция «удалить» означает, что все соответствующие строки должны быть удалены из фрейма, а не соответствующие остаться;
- операция «сохранить» означает, что во фрейме должны быть удалены строки не соответствующие условию, а остальные – остаться;

- общее (сложное) условие формируется из двух простых, при этом варианты взаимного выполнения обоих условий следующие:
 - «одновременно»: оба простых условия должны быть истинными (True);
 - «только первое»: первое простое условие должно быть True, а второе - False;
 - «только второе»: первое простое условие должно быть False, а второе - True;
 - «хотя бы одно»: или первое или второе простое условие должны быть True;
 - «ни одно»: оба простых условия должны быть False.

Следует два раза вывести (до и после операции) такой срез фрейма с названиями колонок, по которому понятно, что операция выполняется корректно.

Задача 3 (максимум 0,5 балла).

После обработки строк следует обработать столбцы фрейма из задачи 2. Вид обработки определяется вариантом студента.

Следует вывести по 4 первых и последних строки фрейма с именами колонок как до выполнения операции, так и после.

Операция «сортировать» означает переставить строки фрейма так, чтобы значения в указанной колонке возрастали или убывали (по варианту).

Объясните результат.

Задача 4 (максимум 0,5 балла).

Используя исходный фрейм с пассажирами «Титаника» из второй задачи вычислить величину, указанную в вариантах для каждого студента.

Задача 5* (максимум 0,4 балла) - творческая.

Создайте два или несколько (можно 3) фрейма данных (объекта типа Data Frame) и выполните их объединение по описанию, указанному в вариантах для каждого студента. Число фреймов также указано в вариантах. Постарайтесь организовать фреймы так, чтобы результат операции их объединения был нагляден и осмыслен. Выведите все исходные и результирующий фреймы.

Задача 6* (максимум 1 балл) - творческая.

предназначена тем студентам, кто собирается претендовать на 4 и 5.

Загрузите в виде CSV файла с сайта **rp5.ru** архив погоды в том городе, где родился кто-либо из Ваших родителей (или город неподалёку) за тот год, когда у этого родителя был юбилей. Постарайтесь корректно открыть этот файл в среде Google Colab. В блокнот с решением вставьте фрагмент скриншота со странички

сайта, где видно, как вы выбрали город и период, и каким способом вызываете загрузку. Полученный фрейм представляет данные нескольких временных рядов (температура, давление, влажность воздуха, облачности и т.д.).

Дайте описание всех колонок файла с указанием типа данных. Оставьте для работы те колонки, данные которых, как Вам кажется будут интересны для обработки. Объясните свой выбор.

Далее покажите, что можете работать с данными этих временных рядов: делать срезы, находить средние (максимальные, минимальные) значения по месяцам, изменять дискретизацию, выполнять сглаживание и т.п. Ваши операции могут выходить за рамки изученного на паре. Это приветствуется. Придумывайте, находите, реализуйте.

Обязательно надо построить как минимум три диаграммы (графика), наглядно показывающие те или иные особенности данных.