

## 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* (или *html*), 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.

Обратите внимание, что все необходимые для выполнения задания программные конструкции рассмотрены в учебных ноутбуках, размещенных в системе LMS. После изучения этих материалов выполнение задания не потребует больших усилий.

Максимальная оценка за выполнение задания вне аудитории – 1 балл. Дополнительные баллы (от 0 до 4) можно будет получить на следующем практическом занятии по результатам тестирования – общий тест по лаб. работам 2 и 3.

Внимание: самостоятельное и вдумчивое выполнение задания серьезно повышает вероятность успешного прохождения теста (будет проверяться понимание работы принципов работы с инструментарием и, в частности, умение понимать программный код).

#### Задача 1.

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.**

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 должно 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.**

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.**

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