

## **Тема 2. Библиотека Pandas: основные объекты данных, визуализация данных**

### **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\_02.ipynb'). Безымянные работы проверяться не будут.

#### **Задача 1 (максимум 0,5 балла).**

Средствами библиотеки Pandas создать серию или фрейм данных (в зависимости от варианта). В серии должно быть не менее 20 элементов, во фрейме не менее 15 строк и не менее двух колонок. Набор данных должен иметь индексы, колонки фрейма – осмысленные названия. Полученную серию (фрейм) вывести целиком, показать длину, размерность, число элементов. Покажите тип элементов. Подсчитайте число уникальных значений каждого столбца.

#### **Задача 2 (максимум 0,5 балла).**

На просторах интернета найти файл с расширением `csv`, представляющий собой датасет реальных (а не выдуманных) данных. Для поиска можно загуглить фразы типа «Датасеты для машинного обучения и анализа данных», «20 лучших датасетов». Запрещено брать датасеты «ирисы», «преступность в Чикаго», «пингвины», «недвижимость Бостона», «пассажиры Титаника», а также те, что рассматривались на парах. Датасет должен иметь не менее 100 строчек (записей) и не менее четырёх столбцов, как минимум два из которых числовые, один строковый (категориальный) и один номинальный.

Файл загрузить во фрейм **df** (переменную типа `DataFrame`). Если в исходном фрейме много колонок, то удалить лишние, оставив не более 6, удовлетворяющих условиям из предыдущего абзаца.

Вывести на экран список колонок, число строк фрейма, размерность, общее число элементов, проверить, есть ли пустые элементы. Вывести первые и последние 3 строки фрейма.

Для каждой колонки указать тип данных и число уникальных элементов. А для каждой числовой колонки дополнительно вывести значения максимального, минимального элементов и среднее арифметическое.

### **Задача 3 (максимум 0,5 балла).**

По данным фрейма из второй задачи построить два наиболее показательных графика (диаграммы) разных типов. Обосновать, почему выбраны именно эти два типа. Графики (диаграммы) оформить максимально информативно,

### **Задача 4 (максимум 0,5 балла).**

Любым из способов сформировать осмысленный фрейм данных, содержащий несколько колонок, для визуализации разных колонок данных которых эффективно смотрятся диаграммы двух типов, указанных по номеру варианта.

Требуется вывести информацию о фрейме и изображения обоих графиков, при оформлении графиков (диаграмм) организовать наиболее информативный дизайн.