**Практическое задание**

**Предварительный анализ данных**

Материалы

[Уроки по Pandas](https://bitbucket.org/hrojas/learn-pandas)

Постановка задачи

Имеются данные регионального этапа олимпиады школьников по информатике. Для сохранения персональных данных реальные имена участников заменены именами «Ученик 1», «Ученик 2» и т.д. Олимпиада проходила в 2 тура. В каждом туре предлагалось решить 4 задачи. Максимальный балл за задачу – 100. В первый тур относятся задачи A,B,C и D, а во второй – E, F, G, H. Данные о баллах, набранных участниками, представлены в файле «Результаты олимпиады.csv».

Требуется проанализировать представленные данные и попробовать выявить закономерности.

План работы

1) Загрузите набор данных. Учитывайте, что данные сохранены в кодировке Windows (cp1251).

2) Заполните пропущенные значения в результатах решения задач. Каким значением лучше заполнять пропущенные значения (исходите из специфики набора данных)?

3) Проанализируйте суммарные баллы по каждой задаче. Отсортируйте задачи по убыванию суммарных баллов и постройте столбиковую диаграмму, на которой задачи будут расположены в порядке убывания суммарных баллов.

4) Попробуем понять какой из туров олимпиады оказался сложнее.

а) посчитаем общую сумму баллов, набранных участниками за каждый из туров олимпиады

б) отсортируем участников по убыванию итогового балла и на столбиковой диаграмме отразим результаты их суммарные результаты за первый и второй тур олимпиады. Опишите, какую закономерность Вы увидели на этой диаграмме.

5) Отфильтруйте набор данных, оставив только победителей и призеров. Определите сколько победителей и призеров в той или иной школе и расположите школы в порядке убывания этого количества. Представьте эти данные графически.

6) Творческое задание. Попробуйте самостоятельно определить: чем отличаются 9-ти, 10-ти и 11-классники (восьмиклассников можно рассмотреть отдельной категорией или считать их частью девятиклассников). Может быть одиннадцатиклассники лучше справлялись с какой-то конкретной задачей, или 9-классники в среднем пробовали решать большее число задач, или что-то еще. Возможно, для решения этой задачи стоит обратиться к исходному набору данных, в котором еще были пропущенные значения. Отобразите найденные закономерности графически.