

## Проект №1 “Насколько стара мидия”

Привет! Совсем недавно наша научная группа вернулась из экспедиции, в которой наконец-то собрали все необходимые данные для исследования. Спасибо еще раз большое, что согласился помочь со статической обработкой. Сбором данных руководил аспирант Ярослав (очень рассеянный парень), а помогали ему 10 студентов. Из-за этого все данные сохранены в наборе файлов, что очень неудобно для анализа. Для каждого моллюска измерены такие показатели как число колец, пол (мужской, женский и детская особь), длина моллюска, диаметр, ширина, длина раковины и другие.

В первую очередь нас будут интересовать следующие вопросы:

### 1. Нам надо как-то объединить наблюдения в единую таблицу.

Пожалуйста, напиши **пользовательскую функцию**, благодаря которой мы сможем собрать все наши наблюдения воедино. Так как экспедиции являются ежегодными, а данные всегда называются по разному, то функция должна объединять все файлы одного расширения из заданной папки, путь до которой передается как аргумент функции.

*Подсказка - суммарно у тебя должно получиться 4177 наблюдений.*

(8 баллов)

### 2. Проведите краткий EDA для полученных данных:

- визуализируй переменные на наличие выбросов (1 балл)
- оцени наличие взаимосвязей между переменными, какие гипотезы для дальнейшего исследования ты можешь сформулировать? (4 балла)
- посмотри, действительно ли все данные корректны? Если найдешь, что что-то не так, то исправь это, пожалуйста. Только объясни нам, почему ты воспользовался именно этим подходом. Может быть у него есть альтернативы? (5 баллов)

### 3. Рассчитай среднее значение и стандартное отклонение переменной Length для моллюсков разного пола.

(4 балла)

### 4. У какого процента моллюсков значение переменной Height не превышает 0.165?

(4 балла)

### 5. Чему равняется значение переменной Length, которое больше, чем у 92% от всех наблюдений?

(4 балла)

### 6. Создай новую переменную Lenght\_z\_scores и сохрани в нее значения переменной Length после её стандартизации.

(4 балла)

**7. Сравни между собой диаметр моллюсков с числом колец 5 и 15.**

К каким выводам ты пришел? (4 балла)

Пожалуйста, оформи результаты так, чтобы мы сразу могли использовать их при публикации статьи (2 балла).

**8. Нас особенно интересуют переменные Diametr и Whole\_weight. Что ты можешь про них сказать?**

Есть ли у нас основания предполагать, что они могут быть взаимосвязаны? Как ты это определил? (4 балла)

Пожалуйста, оформи результаты так, чтобы мы сразу могли использовать их при публикации статьи. (2 балла)

**9\*. Наверное, пока ты работал с данными, ты заметил в них что-нибудь интересное. Ты можешь выдвинуть парочку гипотез и проверить их.**

Мы будем рады получить свежий взгляд со стороны на наши данные. (индивидуально, по накопительной системе). (до 25 баллов)

P.S. 4 обязательных балла добавляется за соблюдение стилистики текста и графиков, разделение отчета на логические разделы и отсутствие технических сообщений.

Язык не принципиален – 2 балла за английский язык