

Домашнее задание 1

Дедлайн сдачи работы: **19.01.2026, 23:59**

1. Создайте любой датафрейм, в котором по строкам будет 5 участников, а по столбцам следующие переменные: имя, возраст, балл за тест (любое число в диапазоне от 0 до 100, не обязательно целое). С помощью одной функции выведите типы всех переменных в датасете.
2. Выведите только возраст участников.
3. К первоначальному датасету добавьте новую факторную переменную «пол», содержащей 3 уровня: «мужчина», «женщина», «не хочу отвечать».

Для следующих заданий используйте датасет `StudentPerformance.csv`

4. Загрузите датасет и выведите информацию обо всех переменных в датасете
5. Проверьте, есть ли в нем пропущенные значения. Если есть, удалите строки с пропущенными значениями (воспользуйтесь функцией `na.omit()`, чтобы понять как она работает, посмотрите документацию: `?na.omit`)
6. Студентов какой этничности и пола в датасете больше всего?
7. Сгруппируйте данные на основе того, закончил ли студент подготовительный курс (исользуйте функцию `group_by()` из библиотеки `tidyverse`). Правда ли, что студенты, которые прошли подготовительный курс, лучше сдали экзамен по математике, чем студенты, которые его не проходили (исользуйте функцию `summarize()` из библиотеки `tidyverse`)? А по какому предмету различие между этим группами самое большое используйте функцию `summarize()` из библиотеки `tidyverse`)?

8. Отберите только студенток, которые прошли подготовительный курс и у которых оценка по математике выше среднего по всему датасету (используйте функцию `filter()` из библиотеки `tidyverse`), и сохраните отобранные данные в новый датасет. Сколько в этом датасете наблюдений?