

Домашнее задание 4

Дедлайн сдачи работы: **09.02.2026, 23:59**

Для выполнения заданий используйте датасет

Updated Quality of Life Data.csv

Описание датасета можно найти по ссылке: Кликабельная ссылка на описание датасета

1. Загрузите датасет и посмотрите на данные. Проверьте, есть ли в нем пропущенные значения. Если есть, удалите строки с пропущенными значениями.
2. Проведите описательную статистику переменных и сделайте выводы о данных. Есть ли в данных выбросы или какие-то странные значения? Если есть, исключите их из анализа.
3. Есть ли в данных какая-то связь между полом респондента и выбранной профессией? Выберите и обоснуйте подходящий анализ. Проверьте гипотезу и проинтерпретируйте результат.
4. Есть ли связь между тем, сколько времени респондент тратит на занятия спортом и продолжительностью жизни? Выберите и обоснуйте подходящий анализ. Проверьте гипотезу и проинтерпретируйте результат.
5. Различается ли продолжительность жизни у респондентов разных профессий? Выберите и обоснуйте подходящий анализ. Проверьте гипотезу и проинтерпретируйте результат.
6. Визуализируйте результаты предыдущего пункта.
7. Постройте модель простой линейной регрессии, чтобы предсказать продолжительность жизни по количеству рабочих часов в день. Проинтерпретируйте значимость модели в целом, коэффициент

детерминации и каждый предиктор в отдельности. Опишите, как связан каждый предиктор с зависимой переменной.

8. Проверьте допущения линейной регрессии и проинтерпретируйте результаты.
9. Постройте такую же модель на нормализованных переменных. Чем она отличается от первой модели?
10. Постройте график, демонстрирующий полученные результаты.