

Домашнее задание 5

Дедлайн сдачи работы: **16.02.2026, 23:59**

Для выполнения заданий используйте датасет HR_Employees.xlsx

В датасете приведены данные о работниках некоторой компании, в частности социально-демографические данные, опыт работы, должность и т.д. Расшифровка кодирования категориальных предикторов приведена на второй странице файла. Задача: научиться предсказывать зарплату (переменная MonthlyIncome) сотрудника.

1. Загрузите датасет, проведите описательную статистику переменных и сделайте выводы о данных
2. Выберите 2 категориальные переменные и не менее 2 количественных переменных, чтобы научиться предсказывать доход работников (переменная MonthlyIncome).
3. Изучите связь между выбранными переменными и зависимой переменной.
4. Изучите связь между выбранными предикторами. (Есть ли она? Если есть, на что это может повлиять?).
5. Постройте модель множественной линейной регрессии со всеми выбранными предикторами без взаимодействий.
6. Проверьте допущения и проинтерпретируйте результаты.
7. Проинтерпретируйте значимость модели в целом и каждого предиктора по отдельности. Какую долю дисперсии объясняет эта модель?
8. Содержательно проинтерпретируйте каждый предиктор.
9. Какой предиктор сильнее всего связан с зависимой переменной?
(Подсказка: постройте такую же модель, но на стандартизованный

переменных)

10. Постройте график, демонстрирующий полученные результаты.

BONUS: Попробуйте добавить в модель взаимодействие между двумя категориальными предикторами. Улучшило ли это модель?