

Домашнее задание №3

Pandas

Для выполнения задания используйте jupyter notebook. Результат загрузите в github.

Шаг 1. Импортируйте необходимые библиотеки

Шаг 2. Импортируйте набор данных с

https://drive.google.com/file/d/10eu07g_vT3BDHxq0V2ATzhgiV0DGBAKH/view?usp=sharing

Шаг 3. Назначьте его переменной с именем df.

Шаг 4. Для этого упражнения разделите фрейм данных от «**school**» до столбца «**guardian**».

Шаг 5. Создайте лямбда-функцию, которая будет капитализировать строчные буквы.

Шаг 6. Используйте функцию для капитализации значений в Mjob и Fjob.

Шаг 7. Распечатайте последние элементы набора данных (tail).

Шаг 8. Вы заметили, что оригинальный набор данных все еще строчный? Почему? Исправьте это (замените значения Mjob и Fjob на прописные буквы).

Шаг 9. Выясните, в каком процентном отношении находятся студенты (по возрасту).

Шаг 10. Каков средний уровень потребления алкоголя в каждой из возрастных групп (15, 16, 17 лет и т.д.) отдельно по выходным и будним дням (поля Walc и Dalc)?

Шаг 11. Каков средний уровень потребления алкоголя в каждого из полов отдельно по выходным и будним дням (поля Walc и Dalc)?

Опционально. Используя набор данных, попробуйте ответить на вопрос «Есть ли связь между потреблением алкоголя и академической успеваемостью студентов?» Аргументируйте ваш ответ.

См. также <https://www.kaggle.com/uciml/student-alcohol-consumption>

P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance. In A. Brito and J. Teixeira Eds., Proceedings of 5th FUTURE BUSINESS TECHNOLOGY CONFERENCE (FUBUTEC 2008) pp. 5-12, Porto, Portugal, April, 2008, EUROSIS, ISBN 978-9077381-39-7.