# Демонстрационная задача по анализу данных

Каждый поступающий в Школу Чёрной Магии должен сдать либо латынь, либо каллиграфию (только один из предметов). Куратор Школы Михаил долгое время сам определял, кому что сдавать, но потом ему это надоело и, собрав достаточно данных из анкет поступающих, он обучил классификатор на основе логистической регрессии, который теперь делает это за него. Представьте, что вы секретарь Школы и ваша задача — следить, чтобы распределение работало нормально. Сегодня Михаил взял отгул, и именно в этот день вы решили подредактировать форму регистрации, но что-то пошло не так и данные пришли с большим количеством пропущенных значений. Теперь вам надо восстановить пропуски так, чтобы Михаил — который, конечно же, проверит, как вы работаете без него — по возможности ничего не заметил.

**Входные данные**

В файле `spoiled_data.csv` находятся данные студентов за 9 последних дней. В столбцах 0–9 находятся анонимизированные признаки (защита персональных данных!); в столбце 10 стоит номер дня. Пропуски возникали в течение дня номер 8. Пропущенные значения кодируются символом '-'.

**Что требуется сделать**

Вы должны заменить все пропущенные значения числами. На письменном экзамене полученную таблицу нужно было бы загрузить в Яндекс.Контест. Проверка выглядела бы так:

- Для строк, соответствующих дню номер 8, вычислялись бы предсказания (`predict`) классификатора, описанного в файле `model.py` (на экзамене у вас бы не было доступа к классификатору) и вычислялась бы точность (`accuracy`) прогноза;
- Точность модели на исходных данных без пропусков порядка 0.92. Ваша задача — получить как можно более близкие значения точности. Ваша оценка за это задание была бы равна

$$\frac{accuracy\_score - 0.6}{0.25}$$

**Поскольку это всё равно пока не настоящий экзамен...**

То мы вам даём возможность проверить себя и публикуем исходные данные вместе с истинными значениями таргета `clean_data_with_target.csv`.

Удачи!