Требования к решению заданий:

1. Язык программирования: R или Python.
2. Воспроизводимость: пользователь Вашего решения должен иметь возможность запустить Ваш код и получить результаты схожие с Вашими.
3. По первому заданию необходимо подготовить аналитический отчет с программным кодом, по второму заданию – аналитический отчет с программным кодом и csv-файлом с предсказаниями на тестовой выборке (предсказания требуется дать в виде «вероятностей», чисел от 0 до 1).

Задание 1.

Руководство компании обратило внимание на то, что сотрудники старше 35 лет болеют чаще, чем более молодые сотрудники. Кроме этого, среди мужчин количество пропусков рабочих дней в связи с больничным выше, чем среди женщин. В связи с этой ситуацией, руководство организации планирует ввести дополнительные медицинские осмотры среди групп риска.

Вам необходимо обосновать, что:

1. Мужчины пропускают в течение года более 2 рабочих дней по болезни значимо чаще женщин.
2. Работники старше 35 лет пропускают в течение года более 2 рабочих дней по болезни значимо чаще своих более молодых коллег.

Все необходимые данные содержатся в файле «Статистика за 2016 год».

Задание 2.

Для автоматической сортировки одного из ключевых продуктов предприятия по двум категориям Вам необходимо разработать модель, которая позволит предсказывать категории этого продукта по показаниям датчиков. Вы договорились, что показателем, характеризующим качество Вашей модели, будет AUC. Все необходимые данные лежат по ссылке: <https://www.dropbox.com/s/u9mby8wlb9f8xyr/case2.7z?dl=0>