



Добрый день!

Еще раз благодарим тебя за интерес к стажировке в М.Тех. И поздравляем с прохождением на третий этап нашего отбора.

Перед тобой задание для вакансии «[Младший исследователь данных \(DS\)](#)». **Подробнее с описанием вакансии ты можешь ознакомиться по ссылке из ее названия (зажав Ctrl и кликнув).**

Вернемся к заданию. На его выполнение у тебя есть 72 часа с момента отправки письма.

Задания сформулированы исчерпывающе и не требуют дополнительных уточнений. Пожалуйста, отвечай на них так, как понимаешь.

Если у тебя возникнут организационные вопросы по стажировке, можешь задать их, написав на почту Alexey.stolyarov@mvideo.ru (Алексей)

Желаем удачи!

Тестовое задание

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 проверить следующие гипотезы:

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

Все необходимые данные содержатся в файле «М.Тех_Данные_к_T3_DS».

Решение необходимо предоставить:

- 1) В виде jupyter notebook (с аккуратно оформленным кодом, графиками и описанной логикой решения).
- 2) В виде дашборда на Streamlit с простым функционалом:
 - a. должна быть возможность загрузить .csv, в формате аналогично файлу «Статистика»
 - b. должна быть возможность задать параметры age и work_days
 - c. указан результат проверки гипотез

- d. указана логика получения результата (должны быть отрисованы графики распределений, указаны критерии проверки (стат. тесты, статистики, уровень значимости т.п.)

Код проекта должен быть обернут в docker выложен в Gitlab/Github репозиторий предоставлена ссылка на него, сохраненная в файл формата .doc, .docx, .pdf, tss. Файл назови по следующему принципу: Фамилия_Имя_Название_вакансии. Будет обращать внимание на качество кода, код стайл, оформление ноутбука, кода и репозитория. Будет плюсом, если дашборд будет поднят на сервере и на него будет предоставлена ссылка.