

Тестовое задание на позицию Marketing Data Analyst



Привет!

Ниже мы собрали задание для того, чтобы проверить, насколько мы подходим друг другу. Во вложении ты найдешь 2 датасета, в которых собрана информация об установках мобильного приложения:

installs_main.csv - установки по данным внутренней атрибуции

installs_s2.csv - установки по данным второго канала привлечения (из его админки)

Описание данных:

- source - последний источник, которому атрибутирован инсталл
- contributor (1 и 2) - это два предыдущих источника, по которым был клик или показ до инсталла

Например, юзер кликнул по рекламе 3 раза, где 3ий раз - последний перед установкой:

- первый раз по рекламе источника 1 (**contributor 2** = Source 1)
- второй раз через какое-то время по рекламе источника 4 (**contributor 1** = Source 4)
- третий раз по баннеру источника 2 (**source** = Source 2).

Задача:

- 1) Предложить и проверить гипотезы, почему данные об установках со второго источника (Source 2) могут расходиться с внутренней атрибуцией
- 2) Проверить данные на наличие аномалий и найти инсайты, которые могут помочь принять решения об атрибуции установок в некоторых каналах

Инструменты и оформление результата:

Мы ожидаем, что исследование ты выполнишь в Python. Для визуализации данных кроме Python можно использовать любой BI инструмент (Tableau, Power BI или графики в Google Sheets).

Результат должен представлять из себя выводы и расчеты. Выводы нужно оформить в виде статьи в отдельном текстовом документе или в Jupyter-ноутбуке,

который будет доступен пользователю со стороны бизнеса в читаемом и понятном виде без отвлечения на кодовые блоки.

Примечание:

Это задание - аналог твоих будущих задач в нашей компании, оно поможет нам оценить твой уровень, а тебе - понять, насколько тебе будет интересен такой опыт. Оно не является оплачиваемым, его выполнение остается твоей доброй волей.

Удачи тебе!