

Тест на стажировку в анализе данных в Авито

Требования к решению: просим оформлять решение в едином документе любого формата (doc/docx/pdf): в шапке текста нужно указать ФИО. Далее пронумеровать решения и расположить по порядку просмотра.

Тест для стажеров-аналитиков

1. Аналитик в компании "Рога и Ко" за день написал некоторое количество SQL-запросов. Все запросы можно охарактеризовать количественной характеристикой сложности выполнения - условная "память". Три "тяжелейших" (7/20 общей затраченной памяти) запроса превысили максимальное время выполнения и были отключены системой. Три "легчайших" (5/13 затраченной памяти всех оставшихся запросов) выполнялись менее 10 минут каждый. Остальные запросы выполнялись менее получаса каждый.

1.1 Найдите общее количество запросов, которое написал аналитик

1.2 Дайте оценку на среднее время успешно выполненного запроса

2. В команде есть два стажера-аналитика. Правильный ответ каждый из стажеров получает в 14 из 17 случаев. Чтобы быть более уверенным в важном решении, менеджер решил дать одну и ту же задачу сразу обоим аналитикам: если оба получают одинаковый ответ, то менеджер его использует, а если ответы разные - выберет один из них наугад. Насколько такой способ повышает шансы менеджера принять верное решение?

3. В мешке лежат три кубика: 6-гранный, 12-гранный, 20-гранный. Мы достали один кубик на удачу, подкинули его, и на нем выпало число N . Какова вероятность того, что если мы достанем еще один кубик из оставшихся в мешке и подкинем его, на нем выпадет меньше N ?

1) $N = 12$

2) $N = 4$

4. Чтобы между пользователями Авито было больше доверия, а жизнь мошенников стала сложнее, мы решили попробовать ввести систему отзывов: покупатель может оставить отзыв на продавца. Отзыв может быть просто рейтингом (1-5 звездочек), а может содержать дополнительно какой-то произвольный текст.

а. Предложите метрики и поясните, каким образом их можно использовать для отслеживания прогресса и успешности проекта?

б. Поскольку Авито не магазин, а площадка для связи покупателя и продавца, мы в большинстве случаев не знаем, произошла ли в действительности сделка и на каких условиях. Для большинства сделок последнее, что нам известно - покупатель нажал кнопку просмотра телефона продавца или связался с ним в чате. Также мы всегда знаем логин (привязанный к email и телефону) продавца, но покупатель может быть незалогиненным. В связи с этим кажется, что есть большой риск накрутки отзывов и недобросовестного использования системы: например, профессиональные участники будут пытаться оставлять негативные отзывы на своих конкурентов и хвалебные на себя. Предположим, что система некоторое время уже работает и у нас есть данные по отзывам и всей активности клиентов: что продавал, что искал, на каких объявлениях смотрел телефоны и т.п. Как оценить масштабы накрутки, т.е. долю фальшивых отзывов?