

Тестовое задание состоит из нескольких задач, которые необходимо реализовать на языке Python.

Допускается применение дополнительных пакетов, будет плюсом распараллелить задачи для ускорения выполнения там где это требуется.

Все тестовые датасеты загружаем и процессим (делаем выборки, джойны, фильтруем и т.д.) в коде.

Если у вас не получилось сделать какие-то шаги, но вы понимаете все остальное - пропускайте их и делайте то в чем разбираетесь.

Результат выполнения - ссылка на github с jupyter notebook. Высылаем все ответы / ссылки на fediaeva@skytecgames.com. Так же напишите сколько часов у вас ушло на выполнение.

При возникновении вопросов по тестовому заданию - пишем туда же, на fediaeva@skytecgames.com

Необходимо сделать минимум 3 задания на ваш выбор. Чем больше заданий сделали, тем лучше.

Тестовый датасет прилагается к письму sqlite файлом testcase.db.zip

1 Успешность прототипа

Известно что компания конкурент выпустила 1000 прототипов игр, из которых было 5 успешных. Наша компания выпустила 200 прототипов из которых ни один не был успешен.

Какова вероятность что следующий (201й) наш прототип будет успешен? Решите задачу классическим и баесовским подходами.

2 Сравнение групп платящих игроков

В мобильной игре около 10% игроков совершает платежи в первый месяц с момента установки игры (база расчета - 500 игроков).

Геймдизайнеры разработали обновление которое по их экспертной оценке должно увеличить процент плательщиков с 10% до 11%. Цель эксперимента - подтвердить или опровергнуть их гипотезу.

Геймдизайнеры разработали обновление которое по их экспертной оценке должно увеличить процент плательщиков с 10% до 11%. Цель эксперимента - подтвердить или опровергнуть их гипотезу.

1. Опишите оптимальный по вашему мнению дизайн эксперимента
2. Рассчитайте длительность эксперимента при условии что каждый день в игру приходит около 100 новых игроков.
3. Самостоятельно сгенерируйте датасет с около 10% плательщиков (контроль) и рассчитайте 95% HDI / CI.
4. Сгенерируйте несколько вариантов экспериментальной группы (хуже, лучше, без эффекта), рассчитайте силу эффекта, HDI / CI и ваши выводы для каждого варианта.
5. Решите задачу 2 разными подходами: frequentist / bayesian

Необходимо все ответы сопровождать комментариями почему вы выбрали ту или иную методику расчета, алгоритм и т.д.

3 Предсказание оттока

В мобильной игре, после установки и запуска игры начинается tutorial - игрок проходит обучение игре, которое состоит из нескольких последовательных шагов.

Данные представляют собой набор строк - уникальный id игрока, datetime и номер шага tutorialа который пройден игроком.

Геймдизайнеры выдвинули гипотезу - если мы с высокой вероятностью сможем предсказать шаг на котором игрок уйдет из игры заранее, то запустив альтернативное продолжение tutorialа с этого шага мы избежим оттока и увеличим прохождение tutorialа.

1. Как бы вы решали задачу предсказания оттока игрока?
2. Как оценить качество полученной модели? Опишите методы которые знаете и какой считаете оптимальным и почему.
3. Предположим что вы разработали модель которая с достаточной вероятностью предсказывает отток. Опишите дизайн эксперимента и методологию по проверке гипотезы геймдизайнеров

4 Эффективность рекламных кампаний

Отдел маркетинга оперирует некоторым множеством рекламных кампаний. Первая часть оперирования состоит из запуска тестовых рекламных кампаний, у которых есть требования по KPI при достижении которых тест считается успешным.

Успешные тесты переходят в разряд постоянных кампаний и вторая часть оперирования состоит из мониторинга постоянных кампаний - управление бюджетом кампании (COST) и мониторинг ее доходов (REVENUE), а так же соотношением доходов к расходам.

Главный KPI для постоянных рекламных кампаний - ROAS на 60й день, т.е. какой процент от расходов кампания возвращает на 60й день.

Для выполнения задачи необходимо использовать тестовый датасет (2 таблицы по расходам и доходам).

1. Отдел маркетинга руководствуется гипотезой - чем больший COST расходуется на кампанию тем ниже ROAS 60го дня (рост CPI при увеличении объема закупки при сохранении того же LTV 60го дня). На основании данных подтвердите или опровергните эту гипотезу. Исходите из того что мы точно знаем что CPI зависит от объема нелинейно, вопрос в том как эту зависимость описать и учесть во 2м вопросе.
2. По каждой рекламной кампании рассчитайте суточный рекламный бюджет который максимизирует абсолютную маркетинговую прибыль рекламной кампании (REVENUE 60 дня минус COST).
3. Исходя из пункта 2 по каждой рекламной кампании дайте ваше заключение о том насколько нужно увеличить / уменьшить ее суточный бюджет либо вообще остановить.
4. Решите проблему рекламных кампаний по которым еще нет полных 60 дней. Объясните почему ваше решение оптимально

5 Связь рекламного траффика и органического

Приток новых игроков делится на 2 основных источника - рекламный и органический. Отдел маркетинг выдвигает гипотезу о связи рекламного и органического траффика: когда маркетологи привлекают больше рекламного траффика, растет и органический. Для выполнения задачи необходимо использовать тестовый датасет (таблица source_comparison).

1. Используя Causal Impact Analysis подтвердите или опровергните гипотезу маркетологов. Если вы ее опровергаете, то сформулируйте гипотезу которая кажется вам наиболее вероятной и проверьте ее
2. Количественно рассчитайте вероятность что гипотеза маркетологов (или ваша) верна