

Atlantis. Тестовое задание 1. Data Analyst

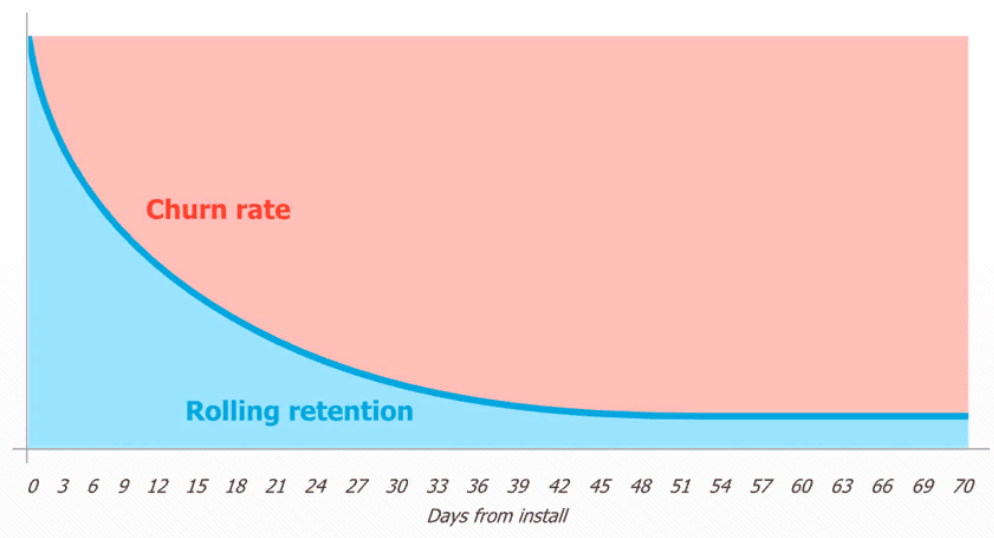
Теоретическая часть

1. Что такое LTV? Если у нас есть данные за первый месяц жизни приложения, как бы вы рассчитали LTV?

LifeTime Value - показатель, отражающей всё, что потратил пользователь в приложении за рассматриваемый период (всё время использования игры или ограниченный промежуток времени): выручка, доход от пользователя с рекламы, посты об игре в соцсетях, отзывы и оценки в AppStore/Google Play.

Стремимся к тому, чтобы LTV был больше CPI, т.е. доход с пользователя должен превышать затраты на его привлечение

Чтобы показатель LTV был наиболее репрезентативен, пользователей сегментируют на когорты (по продолжительности использования продукта / геолокации), чтобы проанализировать их поведение. Важно, чтобы качество входных данных было высоким, иначе расчеты могут ввести в заблуждение. LTV лучше использовать в совокупности с другими показателями.



1. Я бы проанализировала роллинг ретеншн 1, 7, 28 дня для похожего приложения на аналогичной платформе, в нашем частном случае это может быть Klondike Adventures;
2. Уже есть три значения, аппроксимируем, получаем нисходящую функцию, как на картинке;
3. Площадь под графиком - лайфтайм пользователя, её можно вычислить при помощи нахождения интеграла;
4. Чтобы найти LTV осталось умножить значение на выручку с DAU. Примерное значение ARPU также можем взять из похожей игры, поделив его на 30, получим ARPDau. И теперь:

$$LTV = LT * ARPPU$$

Очень приблизительное значение готово!

2. Что такое ARPPU? В результате изменений в продукте ARPPU снизился. Это хорошо или плохо? Поясните ответ.

ARPPU - Average Revenue per Paying User, показатель, отражающий средний доход с каждого платящего пользователя. Часто используется для оценки реакции пользователей на изменение цены. Рассчитывается по формуле:

$$ARPPU = ARPU / C1$$

где:

ARPU — средний доход с пользователя

C1 — показатель конверсии

Однозначно по одному показателю сказать о том, хорошо или плохо нельзя. Из формулы мы видим, что снижение ARPPU может быть вызвано возрастанием конверсии; сохранением дохода на прежнем уровне и при этом увеличением числа платящих пользователей. Важно рассмотреть причины изменения показателя.

Допустим, ARPPU снизился, так как возросло количество платящих пользователей, но платят они не так много как раньше, это может значить, что они будут готовы платить и дальше, если цена возрастет. В таком случае разница между ARPPU и ARPU сократится, что является хорошим показателем для бизнеса
Хорошо!

Допустим, мы изменили цены в приложении, ARPPU снизился, ввиду того, что снизился совокупный доход, который приносила наша игра, и количество платящих пользователей стало меньше.

Это плохо 😞

3. При тесте новой маркетинговой кампании значительно снизилась стоимость установки, но упал и ретеншн 1 дня. Какие метрики стоит смотреть и какую картину по этим метрикам вы ожидаете увидеть чтобы понять, это хорошо или плохо?

При падении Retention 1 есть вероятность того, что рекламная кампания некорректно настроена и привлекает нерелевантных пользователей. Они скачивают, но приложение им неинтересно. Возможно, в игре есть какой-то баг, и она не пускает пользователей

Я бы предложила посмотреть Retention 7, Retention 28, они дают более широкое представление об основном цикле игры. Если по результатам Retention 28 игроки не возвращаются, нужно пересматривать маркетинговую стратегию. И ещё можно сравнить метрики retention для разных когорт игроков, провести А/Б-тестирование. Например, контрольная группа - те, кто пришел не по новой кампании, экспериментальная - новые игроки. Сравнить длительность и частоту сессий, ARPPU, ARPPU, Sticky Factor.

Практическая часть

Дана SQL-база, содержащая данные за первые 10 дней жизни в игре по когорте пользователей. Посмотрите и проанализируйте:

- Где и когда игроки отваливаются
- Где и когда игроки платят

Анализ датасета и графики приведены в Jupyter notebook 😊

Качественное исследование

С целью выявления гипотез для дальнейшего анализа я использовала индуктивный метод и устроила небольшое качественное исследование: провела интервью с несколькими игроками и собрала фидбэк. Главной целью опроса было выяснить, что может спровоцировать уход пользователя, а также что мотивирует его остаться. Были опрошены игроки, не совершавшие покупки в приложении. Полезные тезисы, которые были выявлены из интервью (прим. орфография и пунктуация автора сохранены):

- энергия имеет свойство заканчиваться, это напрягает;
- хочется больше возможностей рефилить энергию;
- нравится задумка и сюжет, это заставляет возвращаться и играть дальше;
- классный дизайн и fabrics;
- иногда напрягают уведомления, слишком часто приходят, приходили бы реже - ок;
- хотелось бы иметь возможность пропустить диалог полностью, а не тапать на каждую реплику;
- очень сильно нравится мини-игра с ракушками.

На вопрос “что могло бы стать триггером для вас как для игрока, чтобы закончить играть” были получены следующие ответы:

- мало энергии, за полную шкалу продвигаешься очень мало;
- не совсем понятно, откуда брать ресурсы (об энергии) помимо просмотра рекламы.

Feature 1.

Я бы предложила генерировать дополнительную фичу, которая будет отражать поведение пользователя после того, как у него закончилась энергия. Пусть, это будет **quit_flag**. Если бы в логе отражалось время окончания энергии (время получения игроком уведомления о том, что у него недостаточно энергии для выполнения к-л действия), оно сравнивалось бы с **session_close.close_time**. Был бы принят допустимый временной промежуток, допустим 30 сек.:

- H0: run out of energy влияет на завершение сессии;
- H1: не влияет.

Так, если игрок проведет в приложении меньше установленного времени, принимаем нулевую гипотезу, `quit_flag = 1`; иначе - альтернативная гипотеза верна, игрок не покидает приложение после иссякания энергии, здесь можно ещё посмотреть, возможно, он совершит покупку в приложении, т.е. докупит энергию.

Feature 2.

Ввиду того, что рассылаемые уведомления зачастую могут напрягать пользователей, я предложила бы внедрить аплифт-модель для рассылки push-уведомлений. Персонализация это классно.

Пусть, наше целевое действие(оно же Y) - клик на push-уведомление и дальнейший переход в приложение. А флаг наличия коммуникации - факт отправления push-уведомления(он же W).

Наша задача оптимизировать рассылку уведомлений, отправляя их только тому сегменту игроков, которому коммуникация необходима для совершения целевого действия.



$P(Y=0 | W=1)$ - игроки, которые могут покинуть приложение из-за приходящих уведомлений;

$P(Y=0 | W=0)$ - те, кто не играет, вне зависимости того, придет пуш или нет;

$P(Y=1 | W=0)$ - те, кто в любом случае играет, они же лояльные пользователи;

$P(Y=1 | W=1)$ - те, кого коммуникация сподвигнет совершить целевое действие, этот сегмент представляет для нас наибольший интерес

Для построения модели я бы включила в датасет поля **target_flag**(Y), **treatment_flg**(W). В качестве фич для построения модели я бы дополнительно сгенерировала поведенческие характеристики на основании **Recency**, **Frequency**, **Monetary**. Для recency: найти последнюю дату для всего датасета, а затем рассчитать разницу между ней и последней датой взаимодействия с приложением для каждого игрока; frequency: время взаимодействия игрока с игрой разделить на количество сеансов; monetary - потрачено в приложении на уровне игрока(такое поле уже рассчитано). И ещё можно рассчитать **first_session_delay** между установкой и первым взаимодействием с игрой

или Rolling Retention первого дня **retention_1D**, чтобы оценить, как быстро игроки приступают к игре. В целом, важно посмотреть поведение игрока во время первой сессии максимально подробно(длительность прохождения квестов, взаимодействие или невзаимодействие с рекламой, где получают игровую валюту), согласно проведенному анализу, на первых этапах воронки отсеивается большое кол-во игроков. В их игровом поведении может кроиться ответ, почему так происходит.

Начать сбор данных из логов по имеющимся переменным. Построить модель(в похожей задаче я использовала дерево решений с градиентным бустингом). На основании полученных данных оценить лояльность клиента, вероятность воздействия уведомления на него. После провести А/Б-тест с целью проверить, работает ли модель так, как нам нужно: не теряем ли мы пользователей(это возможно, если ошиблись с сегментированием), динамику длительности и частоты сессий.

Feature 3.

Согласно обратной связи игроков, можно попробовать добавить кнопку, которая пропускает диалог. Далее провести А/Б-тестирование, где проанализировать показатели пользователей контрольной и тестовой группы(например, длительность сессий, ARPU7) и исходя из нужд бизнеса согласовать, нужно ли вводить данную опцию.

Feature 4.

Касаемо события покупки, совершаемой пользователем. Я бы собирала параметры пользователя: какой девайс использует, геоположение, - могут пригодиться для сегментации; параметры события: что купил, сам или мы его мотивировали, была ли скидка/акция - будет полезно для экономики игры, анализа трендов ресурсов.

Говоря об экономике игры, можно еще собирать инфу об источниках получении игровой валюты(квесты, повышение уровня, банк, акции). Чем больше доля банка в этом соотношении, тем интереснее нам с точки зрения монетизации, но тем менее интересно игрокам, которые не готовы платить. Согласно мнению опрошенных мной игроков, им как раз-таки хотелось бы больше внутриигровых источников получения валюты. Можно было бы провести количественное исследование, оттолкнувшись от данной гипотезы.

Feature 5.

Соотношение величины выручки, полученной из разных каналов: in-app purchases/реклама в приложении(eCPM от rewarded video и баннерной рекламы). Метрика была бы полезна при разработке стратегий монетизации. Если мы имеем характеристики устройства, с которого совершаются сессии + гео игрока, можно строить ещё более продвинутую аналитику