

Часть 1

Задание 1

1. Retention rate (показатель возврата) - процент клиентов, которые возвращаются.

2. Churn rate (показатель оттока клиентов)- процент клиентов, переставших пользоваться приложением.

$\text{Churn rate} = 1 - \text{Retention rate}$

3. Daily Active Users - количество уникальных активных пользователей в конкретный момент времени (как правило, в течение дня)

4. Monthly Active Users - количество уникальных активных пользователей за месяц.

5. Daily Sessions/DAU - показатель, позволяющий понять, как часто пользователи используют приложение в течение дня.

6. CPA- показатель, показывающий насколько затратно обходится привлеченный пользователь в рамках некоторой маркетинговой кампании.

7. Stickness= DAU/MAU - чем ближе показатель к 100%, тем выше уровень вовлеченности пользователей.

8. LTV - показатель стоимости жизненного цикла клиента, дающий оценку ожидаемой прибыли от среднего клиента

$\text{LTV} = (\text{среднее значение конверсий}) * (\text{количество конверсии}) * (\text{среднее время жизненного цикла клиента})$

9. ROI - оборачиваемость инвестиций, показатель эффективности затрат к прибыли от некоторой кампании.

10. ARPU - средняя выручка от пользователя

11. ARPPU - средняя выручка от платящего пользователя

12. App Load Time - количество транзакций или вызовов за определенный момент времени.

13. App Crashes - количество сбоев в работе приложения.

14. App Latency - показатель задержки приложения.

15. 1/7/28-day retention - процент пользователей, вернувшихся в приложение через 1/7/28 дней.

16. ASL- среднее арифметическое длин всех сессий за некоторое время.

Для мониторинга "здоровья":

1. App Load Time
2. App Latency
3. App Crashes
4. Stickness
5. Churn Rate

Для "здоровья" продукта
стоит отслеживать не
только метрики качества,
но и метрики бизнеса.

а) Первые три метрики отражают качество приложения. Медленная загрузка, сбои могут сподвигнуть клиента отказаться от использования продукта, поэтому эти метрики стоит мониторить постоянно.

Stickness-показатель вовлеченности. Если он будет слишком низким, это явный сигнал о том, что стоит провести более сложный анализ и понять, почему клиенты не мотивированы использовать приложение. То же касается и Churn Rate. В случае высокого количества отказов стоит провести дополнительный анализ и выявить проблему.

б) Перечисленные метрики напрямую отражают "счастье" пользователя, поэтому при нормальных условиях их нужно отслеживать хотя бы раз в месяц. А при добавлении или изменении текущего функционала даже чаще.

с)

1. Увеличение App Load Time:

Возможно, увеличилась нагрузка приложения: им стало пользоваться больше человек или совершается больше операций. В любом случае, этот показатель должен быть в районе 2-3 секунд, что не вызовет у пользователя негативных эмоций. Однако нагрузка может быть скачкообразной для некоторых приложений, что также стоит учитывать при анализе.

2. Увеличение App Crashes сигнализирует о проблемах на стороне бэкенда и требует немедленного анализа и тестирования кода. В идеале этот показатель должен составлять 1-2%

3. На увеличение App Latency может влиять несколько факторов: среда передачи, маршрутизаторы, способ хранения и обработки данных, расстояние между взаимодействующими узлами.

4. **Stickness** подходит не всем продуктам и зависит от их цели. Например, для e-commerce приложений этот показатель может быть низким, но пользователи будут совершать дорогостоящие покупки. Иногда увеличение этого показателя говорит не о росте вовлеченности, а о росте числа пользователей. Также при запуске чего-то нового может наблюдаться сильное падение этой метрики или наоборот сильный рост благодаря маркетингу, поэтому потребуется несколько месяцев на нормализацию данных.


5. В целом, **Churn Rate** зависит от назначения приложения и может варьироваться от 10% до 50%. Увеличение CR может говорить или о личных предпочтениях клиента, или о его неудовлетворенности. Возможно, клиент не понимает, как пользоваться приложением, или его не устроила новая фича, или его жалобы игнорировались долгое время. Опять же, причин может быть много и необходимо опросить отказавшихся клиентов и провести анализ.

Задание 2

Во-первых, эти двое детей могли оказаться родственниками, в таком случае оценка не верна уже в расчетах. Во-вторых, факт отсутствия совместного проживания родителей не говорит о разводе. Один из родителей может отсутствовать по множеству причин: работа в другом городе, тюремное заключение, смерть, раздельное проживание родителей и т.д.

Задание 3

Хороший показатель MAU зависит, конечно, от целей приложения, но в среднем считается нормальным уровнем 30%-37% ежемесячно активных пользователей. Я бы запросила показатели DAU за последний месяц, чтобы определить промежутки спада и роста активности. Также, значение Churn Rate, чтобы понять, начался ли сильный отток клиентов. Если 30% - стабильная активная аудитория, то можно постараться повысить этот показатель. Ввести Push-уведомления, повысить рейтинг в Play Market и App Store, ввести ненавязчивые рассылки на почту и по sms.



Часть 2

① C_{12}^5 - кол-во сочетаний

$$C_{12}^5 = \frac{12!}{5! \cdot 7!} = \frac{12 \cdot 11 \cdot 10 \cdot 9 \cdot 8}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 792$$

Ответ: 792

② $CTR_1 = \frac{26}{38} \cdot 100 = \frac{1300}{19}$

$$CTR_2 = \frac{42}{98} \cdot 100 = \frac{300}{7}$$

$H_0: p_1 = p_2$ - пром. одинаковые эффективность

$H_1: p_1 \neq p_2$

$$\Phi(u_{кр.}) = \frac{1 - 0,05}{2} = 0,475$$

По таблице Лапласа

$u_{кр.} \approx 1,96$. Если $|u| < u_{кр.} \Rightarrow$ принимаем H_0 .
иначе отвергаем.

$$u_{набл.} = \frac{\frac{13}{19} - \frac{3}{7}}{\sqrt{\frac{26+42}{38+98} \left(1 - \frac{26+42}{38+98}\right) \left(\frac{1}{38} + \frac{1}{98}\right)}} \approx 2,67 > u_{кр.}$$

$\Rightarrow H_0$ отвергаем

Ответ: На уровне значимости 0,05
разница CTR наблюдается

3



$$E = \sum x_i \cdot p_i$$

x_i - попытка, при которой
выбывается шип

$$\begin{array}{l} x_i \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \\ p_i \quad \frac{1}{5} \quad \frac{4}{5} \cdot \frac{1}{4} \quad \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} \quad \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} \end{array}$$

$$E = \frac{1}{5} + 2 \cdot \frac{1}{5} + 3 \cdot \frac{1}{5} + 4 \cdot \frac{1}{5} + 5 \cdot \frac{1}{5} = (1+2+3+4+5) \cdot \frac{1}{5} = 15 \cdot \frac{1}{5} = 3$$

Ответ: 3

4

	USA	India	Brazil
Тренд	+	+	-
Возможность	+	-	+
Цикличность	+	-	-

5

Сформулирует гипотезы:

Но: "Разницы между группами нет"

H_1 : "Разница между группами есть".

Ошибка I-го рода (ложно положительная):

мы считаем, что разница в результатах есть, но на деле она случайна.

Чтобы избежать ошибки первого рода члены групп должны быть

независимы, мы должны установить уровень достоверности

(например, 5%). Ошибка может еще возникнуть при одновременном тесте трех групп, поэтому надежнее проводить его попарно.

Часть 5

1) Удержание клиентов имеет значение не только на стадии роста продукта, но и формирует стабильную базу пользователей, которые будут использовать услуги сервиса регулярно. А значит сервис будет конкурентноспособным на рынке.

2) Здесь мы решаем задачу классификации, потому что хотим предсказать, вернется ли пользователи на 7-й день (True/ False), на основе множества признаков.

3) Набор признаков, при которых клиент обычно возвращается на 7-й день.

- 4) 1. Сколько клиент платит за услуги?
2. С какими фичами взаимодействует?

	Фича 1	Фича 2	Фича 3	...
Взаимо-ет?	Да(1)	Нет(0)	Нет(0)	...

3. Из какой страны?(категориальная переменная)

	РФ	ИША	Италия
User1	1	0	0
User2	0	0	1

Таблица №2

4. Вернулся ли на 7-й день?

Предварительный анализ: проверю, нет ли пропусков, выбросов в данных. Приведу категориальные признаки к численному виду

Учёт категориальных переменных: См. таблицу № 2

5) Алгоритм логистической регрессии.

6) i-индекс

{User_i} - пользователи со всевозможными признаками

$$\text{User } i: \bar{x}_i(x_1^i, x_2^i, \dots, x_n^i)$$

↓

$$a(\bar{x}_i) = a_1 x_1^i + a_2 x_2^i + \dots + x_n^i$$

↓

Пронесем $a(\bar{x}_i)$ через
функцию сигмоиды

↓

Результат (1 - вернулся,
0 - нет)

Количество объектов: должно быть не менее 10 случаев с наименее частым исходом с вероятностью p для каждой из n объясняющих переменных, тогда размер выборки должен быть не менее

$$N = \frac{10 \cdot n}{p}$$

7) Применение методов регуляризации: Ridge/ LASSO/ Elastic-Net

Необходимо отслеживать меткость, точность, полноту, специфичность, AUC-ROC, AUC-PR, Log Loss, коэффициент Мэтьюса

8) Этот датасет можно использовать для решения задачи регрессии.

Можно использовать линейную регрессию или метод К-ближайших соседей. Может дать информацию о связи ожидаемой прибыли от клиента с используемым функционалом и 7d-retention.