**Тестовое задание.Мосеева Евгения**

**email:** [**moseeva.zh@yandex.ru**](mailto:moseeva.zh@yandex.ru) **тел.: +7902 191 9670**

[**https://t.me/EvgeniyaMoseeva**](https://t.me/EvgeniyaMoseeva)

**Продуктовая часть.**

Задание 1. Предложите 3 метрики первого уровня для мобильного и интернет-банка. Почему именно эти?

**Ответ:**

Первое, с чего нужно начать,- это ключевые метрики продукта (мобильного и интернет-банка). Было бы ошибкой сразу пытаться начать решать продуктовые или аналитические задачи, погружаться в анализ причин оттока пользователей, придумывать А/В тесты и т.д. К этому нужно вернуться, но позже. Сначала закрыть самый первый этап аналитики в компании. А именно- получить ответы на самые простые вопросы о нашем приложении.

Главный вопрос, на который должна отвечать аналитика на самом начальном этапе- это вопрос «сколько?»

Сколько у нас активных пользователей в день, неделю, месяц? Сколько просмотров сайта в определённый период?

* **Количество активных уникальных пользователей:**

Чем больше активных пользователей, тем больше вероятность того, что банк будет получать больше дохода.

Начнем с аудиторных метрик: сколько активных уникальных пользователей у нас было за конкретный период: день, неделю, месяц. То есть это DAU, WAU и MAU соответственно. Внимание заслуживает комбинация DAU к MAU или sticky factor (липкий фактор). Чем ближе значение этой метрике к единице, тем большая часть активной дневной аудитории пользуется сервисом каждый день в течение месяца.

Благодаря DAU, WAU и MAU мы знаем, сколько пользователей заходят в приложение. Хорошо бы ещё выяснить, сколько времени они там проводят.

С этим поможет пользовательская сессия — набор действий пользователя, выполняемый от открытия приложения до окончания его использования.

# **Пользовательская сессия:**

- **Найдем среднее число сессий на пользователя.**

Полезно изучить, сколько в среднем сессий приходится на одного пользователя, например, за месяц. Это ещё один хороший показатель регулярности использования приложения, помимо sticky factor.

**- Найдем среднюю продолжительность сессии**

Средняя продолжительность сессии, или ASL, показывает, сколько в среднем длится сессия пользователя.

Мы можем много смотреть на метрики активности пользователей в разные временные промежутки — на MAU, DAU, WAU. Радоваться тому, какие они большие и как они (потенциально) растут со временем, понимая, что это означает большую вовлеченность наших пользователей... Или нет? Почему в интернете можно найти много постов на тему того, что всё это — [«метрики тщеславия»](https://amplitude.com/blog/measuring-active-users), пользы от которых мало?

Проблема в том, что эта «активность пользователей» может не означать **реального интереса** пользователей к нашему продукту. Рост этих метрик может быть искусственным и не отражать **долговременную вовлеченность.** Иными словами, это легко манипулируемые метрики.

Если же мы хотим понять, сколько пользователей реально остаётся и продолжает пользоваться продуктом, мы должны посчитать **Conversion Rate**.

Чтобы оценить результаты привлечения пользователей, можно построить воронку привлечения пользователей. Для этого важно знать:

**Показы.** Сколько раз показывали баннеры;

**Клики.** Сколько пользователей по этим баннерам кликнуло;

**Переходы на сайт или мобильное приложение.** Сколько кликнувших попали на лендинг;

**Регистрации.** Сколько посетителей оставили компании свои данные, или заполнили анкету, оформили карту банка и пр.

Данные о показах и кликах получают из рекламных систем. Информацию о переходах на лендинг и регистрациях — из Яндекс.Метрики или собственной системы аналитики сайта.

Как только все данные собраны, можно считать конверсии.

* **Conversion Rate и CTR (коэффициент конверсии и CTR):**

**Conversion Rate -это центральная метрика. Показывает долю визитов на сайт, закончившихся целевым действием (например-заполнением анкеты, оформлением карты и пр.).**

Конверсию (CR) показов рекламного баннера в клики называют CTR. Этот показатель принято рассчитывать в процентах:

СTR = Клики / Показы х 100%

Конверсию из переходов по баннеру в целевое действие (регистрацию, заполнение анкеты, оформление карты и пр.) находят по стандартной формуле:

СR = Регистрация / Клики х 100%

Задание 2. Вам необходимо понять какие цели отслеживаются на странице <https://www.open.ru/cards/opencard?from=main_menu> .

* Опишите, какие цели уже отслеживаются.
* Опишите, каким образом вы поймете отслеживаются ли цели на странице.
* Какие цели на ваш взгляд необходимо отследить.

**Ответ:**

**На странице отслеживаются следующие цели:**

1. Заявка на оформление карты "Открытие". Для этого на странице есть кнопка "Оформить карту", нажав которую пользователь переходит на страницу с формой заявки.

2. Заполнение анкеты на карту "Открытие". На странице есть анкета, которую пользователь может заполнить, чтобы перейти к оформлении карты.

3. Количество переходов на страницу с формой заявки. Для этого на странице установлены специальные метрики-шаги.

**Для того, чтобы понять, отслеживаются ли цели на странице, можно использовать инструменты веб-аналитики, такие как Google Analytics.**

С помощью этого инструмента можно установить цели и отслеживать их выполнение на странице.

**На мой взгляд, необходимо отследить еще следующие цели:**

1. Количество просмотров страницы. Эта метрика позволит понять, насколько популярна страница с оформлением карты "Открытие".

2. Количество отказов от заполнения формы заявки. Эта метрика позволит понять, насколько удобна и понятна форма заявки для пользователей.

3. Количество успешно оформленных карт. Эта метрика позволит понять, насколько эффективна страница с оформлением карты "Открытие" и какие еще улучшения можно внести.

3. Вы хотите повысить конверсию на странице <https://www.open.ru/credit_cards/120days> , для этого вы хотите провести А/Б тест. Опишите процесс проведения А/Б (подготовка, запуск, итоги). Приведите в пример гипотезы, на которые вы будете опираться при тестировании страниц.

**Ответ:**

**Процесс проведения А/Б теста включает следующие этапы:**

1. Подготовка:

- Определение цели тестирования (например, увеличение конверсии на странице);

- Выбор метрик, которые будут использоваться для оценки результатов теста;

- Разработка гипотез, которые будут проверяться в рамках теста;

- Разработка вариантов страницы (A и B), которые будут тестироваться;

- Определение размера выборки и длительности теста.

2. Запуск:

- Размещение вариантов страницы на сайте;

- Разделение трафика между вариантами страницы;

- Сбор данных о поведении пользователей на страницах.

3. Итоги:

- Анализ результатов теста;

- Определение статистической значимости различий между вариантами страницы;

- Принятие решения о том, какой вариант страницы лучше с точки зрения достижения цели тестирования.

**Примеры гипотез, на которые можно опираться при тестировании страницы:**

1. Добавление видео на страницу приведет к увеличению времени, проводимого пользователями на странице, и, как следствие, к увеличению конверсии.

2. Изменение цвета кнопки "Оформить карту" на более яркий приведет к увеличению конверсии.

3. Упрощение формы заявки на карту приведет к увеличению конверсии.

**Задание 4**. Вам необходимо добавить отслеживание события перевода по номеру карты в мобильном и интернет-банке через web инструменты (выберете один из: GA, Amplitude, Я.Метрика, AppMetrica или другой на свой выбор)

* Опишите, как вы видите процесс реализации задачи от постановки задачи (вам) до внедрения.
* Опишите ТЗ для разработчика мобильного банка.
* Опишите ТЗ для разработчика интернет-банка.

**Ответ:**

**Процесс реализации задачи от постановки задачи до внедрения может быть следующим:**

**Для добавления отслеживания события перевода по номеру карты в мобильном и интернет-банке через web инструмент Google Analytics 4, необходимо выполнить следующие шаги:**

1. Создать цель в Google Analytics 4. Для этого нужно зайти в административную панель Google Analytics 4, выбрать нужный аккаунт и ресурс, перейти в раздел "Цели" и нажать на кнопку "Создать цель". В открывшемся окне нужно выбрать тип цели "Событие" и указать параметры события, которые будут отслеживаться (например, "Перевод по номеру карты").

2. Настроить отслеживание события. Для этого нужно указать категорию, действие и метку события. Например, категория "Переводы", действие "Перевод по номеру карты" и метка "Успешный перевод".

3. Добавить код отслеживания на страницу перевода. Для этого нужно скопировать код отслеживания цели из Google Analytics 4 и добавить его на страницу перевода в нужном месте.

4. Проверить работу отслеживания. Для этого нужно выполнить тестовый перевод и проверить, что событие успешно отслеживается в Google Analytics 4.

**Что касается процесса реализации задачи от постановки задачи до внедрения, то он может выглядеть следующим образом:**

1. Постановка задачи. Заказчик (например, отдел маркетинга банка) формулирует задачу по добавлению отслеживания события перевода по номеру карты в мобильном и интернет-банке через web инструмент Google Analytics 4.

2. Анализ требований. Специалисты по веб-аналитике анализируют требования заказчика и определяют необходимые шаги для реализации задачи.

3. Разработка решения. Специалисты по веб-разработке разрабатывают код отслеживания цели и добавляют его на страницу перевода.

4. Тестирование. Специалисты по веб-аналитике и веб-разработке выполняют тестовый перевод и проверяют работу отслеживания.

5. Внедрение. Решение внедряется на боевом сайте и начинает отслеживать события перевода по номеру карты в мобильном и интернет-банке через web инструмент Google Analytics 4.

6. Мониторинг и анализ. Специалисты по веб-аналитике мониторят работу отслеживания и анализируют полученные данные для принятия решений по улучшению пользовательского опыта и повышению эффективности маркетинговых кампаний.

ТЗ для разработчика мобильного банка:

1. Необходимо добавить отслеживание события перевода по номеру карты в мобильном банке.

2. При переводе по номеру карты необходимо отправлять событие веб-аналитике с указанием суммы перевода и номера карты получателя.

3. Событие должно быть отправлено только в случае успешного перевода.

4. Необходимо использовать web инструмент GA для отслеживания события.

5. Код отслеживания события должен быть размещен на всех экранах, связанных с переводом по номеру карты.

ТЗ для разработчика интернет-банка:

1. Необходимо добавить отслеживание события перевода по номеру карты в интернет-банке.

2. При переводе по номеру карты необходимо отправлять событие веб-аналитике с указанием суммы перевода и номера карты получателя.

3. Событие должно быть отправлено только в случае успешного перевода.

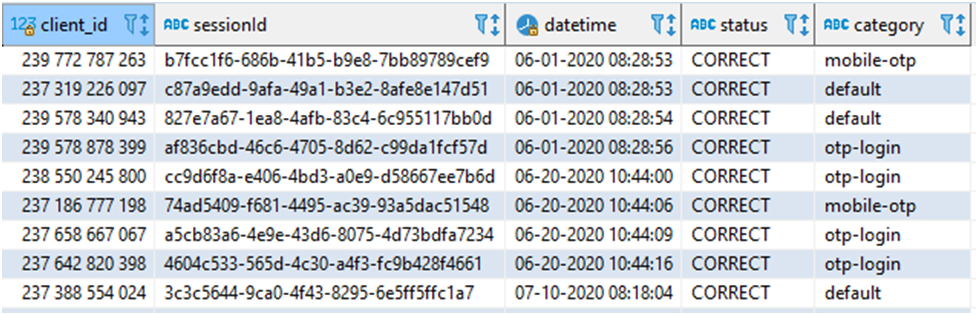
4. Необходимо использовать web инструмент GA для отслеживания события.

5. Код отслеживания события должен быть размещен на всех страницах, связанных с переводом по номеру карты.

**SQL.**

Дана таблица по вводу ОТП кодов клиентами (файл: OTP\_session\_id.csv)

**Задача:** решить максимальное количество заданий используя SQL.



Поля таблицы:

* ***client\_id*** - номер клиента
* ***sessionId*** – идентификатор сессии
* ***datetime*** - дата время
* ***status*** – статус ввода кода
* ***category*** – текстовый код ОТП

**Задание 1.** Показать категорию, по которой было введено наибольшее число кодов

**Ответ:**

SELECT category

FROM OTP\_session\_id

GROUP BY category

ORDER BY COUNT(\*) DESC

LIMIT 1

**Задание 2.** Добавить индикатор, который будет выделять следующие значения:

* Если otp для категории платежей (payment), то - 1
* Если otp для категории переводов (transfer), но не для переводов с использованием счетов (acc), то - 2

Все остальные заявки не должны попасть в результат выполнения запроса.

**Ответ:**

Для добавления индикатора, который будет выделять значения в соответствии с условиями, можно использовать функцию CASE WHEN в запросе:

SELECT

client\_id,

sessionId,

datetime,

status,

category,

CASE

WHEN category = 'payment' THEN 1

WHEN category = 'transfer' AND category != 'acc' THEN 2

ELSE NULL

END AS indicator

FROM

OTP\_session\_id

WHERE

category IN ('payment', 'transfer')

AND status = 'CORRECT'

В данном запросе мы выбираем поля `client\_id`, `sessionId`, `datetime`, `status`, `category` из таблицы `OTP\_session\_id `. Далее, используя функцию CASE WHEN, мы добавляем новое поле `indicator`, которое будет принимать значение 1, если категория равна 'payment', и значение 2, если категория равна 'transfer', но не равна 'acc'. Все остальные значения будут равны NULL.

Также мы добавляем условия WHERE, чтобы выбрать только заявки с категориями 'payment' и 'transfer' и со статусом ' CORRECT '.

**Задание 3.** Посчитать метрику Month-of-Month (прирост текущего месяца к предыдущему) по уникальным клиентам с кодами otp-login.

**Ответ:**

Для решения данной задачи можно использовать SQL-запрос, который будет группировать данные по месяцам и считать количество уникальных клиентов с кодами otp-login для каждого месяца. Затем можно вычислить прирост текущего месяца к предыдущему, используя функцию LAG().

SELECT

DATE\_TRUNC('month', datetime) AS month,

COUNT(DISTINCT CASE WHEN category = 'otp-login' THEN client\_id END) AS unique\_clients,

LAG(COUNT(DISTINCT CASE WHEN category = 'otp-login' THEN client\_id END)) OVER (ORDER BY DATE\_TRUNC('month', datetime)) AS prev\_month\_unique\_clients,

ROUND((COUNT(DISTINCT CASE WHEN category = 'otp-login' THEN client\_id END) - LAG(COUNT(DISTINCT CASE WHEN category = 'otp-login' THEN client\_id END)) OVER (ORDER BY DATE\_TRUNC('month', datetime))) / LAG(COUNT(DISTINCT CASE WHEN category = 'otp-login' THEN client\_id END)) OVER (ORDER BY DATE\_TRUNC('month', datetime)) \* 100, 2) AS mom\_growth

FROM

OTP\_session\_id

GROUP BY

month

ORDER BY

month

В данном запросе используется функция DATE\_TRUNC() для группировки данных по месяцам, функция COUNT(DISTINCT) для подсчета количества уникальных клиентов с кодами otp-login, функция LAG() для вычисления значения предыдущего месяца и функция ROUND() для округления значения метрики до двух знаков после запятой.

Результатом выполнения запроса будет таблица со следующими полями:

- month - месяц

- unique\_clients - количество уникальных клиентов с кодами otp-login за текущий месяц

- prev\_month\_unique\_clients - количество уникальных клиентов с кодами otp-login за предыдущий месяц

- mom\_growth - прирост текущего месяца к предыдущему в процентах

**Задание 4.** Одним запросом сформируйте:

* Количество успешно введённых ОТП кодов в разрезе категории кода ОТП
* Долю каждой категории по убыванию
* Количество с накопительным итогом
* Общее количество введённых ОТП кодов
* Последний отчётный месяц (полный месяц, от текущего)

**Ответ:**

SELECT

category,

COUNT(\*) AS count\_successful,

ROUND(COUNT(\*) / (SELECT COUNT(\*) FROM OTP\_session\_id), 2) AS category\_share,

COUNT(\*) OVER (ORDER BY category\_share DESC ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW) AS cumulative\_count,

COUNT(\*) OVER () AS total\_count,

DATE\_TRUNC('month', NOW() - INTERVAL '1 month') AS last\_reporting\_month

FROM OTP\_session\_id

WHERE status = 'CORRECT'

GROUP BY category

ORDER BY category\_share DESC;

Комментарии:

1. Выбираем категорию и количество успешно введенных ОТП кодов в разрезе категории.

2. Добавляем долю каждой категории по убыванию, округленную до двух знаков после запятой.

3. Добавляем количество с накопительным итогом с помощью функции Т-SQL.

4. Добавляем общее количество введенных ОТП кодов.

5. Добавляем последний отчетный месяц (полный месяц, от текущего).

Фильтруем только успешные статусы. Группируем по категории. Сортируем по доле категории по убыванию.

**Python - дополнительно.**

**Задание 5\*.**

Представьте решение заданий 1-3 с помощью методов библиотеки pandas (исходный файл загрузите как датафрейм). Пришлите решение в виде файла Jupyter Notebook или ссылкой на Datalore.

**Решения к 1 и 2 заданиям:** <https://disk.yandex.ru/d/4k5e5l5ClJ9gIw>

**Решения к 3 заданию**:

**Планирование и запуск А/В тестирования**

**Описание задачи №1:**

У нас есть данные АА-теста с '2022-11-25' по '2022-12-01'. Нам нужно сделать симуляцию, как будто мы провели 10000 АА-тестов. На каждой итерации нам нужно сформировать подвыборки без повторения в 500 юзеров из 2 и 3 экспериментальной группы. Провести сравнение этих подвыборок t-testом.

Построить гистограмму распределения получившихся 10000 p-values

Посчитать, какой процент p values оказался меньше либо равен 0.05

Написать вывод по проведенному АА-тесту, корректно ли работает наша система сплитования

**Вот что у нас получилось:**

 Проведение АА-тестов на пользователях ленты новостей [**https://disk.yandex.ru/d/TX3hT8y4iSs2BA**](https://disk.yandex.ru/d/TX3hT8y4iSs2BA)

 P-values оказалось менее 5%, что свидетельствует о том, что наша система сплитования работает корректно и АА-тест пройден успешно

 Несмотря на то, что процент "больших" p-value больше 5, это еще не будет значить, что сплитование некорректно. Тут как раз правильно смотреть на совокупность: равномерность распределение - небольшая доля маленьких pvalue

**Описание задачи №2:**

Пришло время проанализировать результаты эксперимента, который мы провели вместе с командой дата сайентистов. Эксперимент проходил с 2022-12-02 по 2022-12-08 включительно. Для эксперимента были задействованы 2 и 1 группы.

В группе 2 был использован один из новых алгоритмов рекомендации постов, группа 1 использовалась в качестве контроля.

Основная гипотеза заключается в том, что новый алгоритм во 2-й группе приведет к увеличению CTR.

* Наша задача — проанализировать данные АB-теста
* Выбрать метод анализа и сравнить CTR в двух группах (мы разбирали t-тест, Пуассоновский бутстреп, тест Манна-Уитни, t-тест на сглаженном ctr (α=5) а также t-тест и тест Манна-Уитни поверх бакетного преобразования)
* Сравнить данные этими тестами. А еще посмотрите на распределения глазами. Почему тесты сработали так как сработали?
* Описать потенциальную ситуацию, когда такое изменение могло произойти
* Написать рекомендацию, будем ли мы раскатывать новый алгоритм на всех новых пользователей или все-таки не стоит

**Вот что у нас получилось:**

* Оценка результатов проведения А/В тестирования данных пользователей ленты новостей <https://disk.yandex.ru/d/oR6eRCUqcblCzw>
* Оценка тестами результатов А/В тестирования показал различные результаты и однозначных выводов сделать нельзя. В связи с чем, эксперимент будем считать неудачным, и в продакшн такой алгоритм запускать нельзя.

**Описание задачи №3:**

Относительно недавно (в 2018-м году) исследователи из Яндекса разработали классный метод анализа тестов над метриками-отношениями (прямо как у нас) вида 𝑥/𝑦 (У нас likes/clicks).

**Идея метода заключается в следующем:**

* Вместо того, чтобы заталкивать в тест «поюзерные» CTR, можно сконструировать другую метрику и анализировать ее, но при этом гарантируется (в отличие от сглаженного CTR), что если тест на этой другой метрике «прокрасится» и увидит изменения, значит изменения есть и в метрике исходной (то есть в лайках на пользователя и в пользовательских CTR
* При этом метод сам по себе очень прост. Что это за метрика такая?
  + Считаем общий CTR в контрольной группе 𝐶𝑇𝑅𝑐𝑜𝑛𝑡𝑟𝑜𝑙=𝑠𝑢𝑚(𝑙𝑖𝑘𝑒𝑠)/𝑠𝑢𝑚(𝑣𝑖𝑒𝑤𝑠)
  + Посчитаем в обеих группах поюзерную метрику 𝑙𝑖𝑛𝑒𝑎𝑟𝑖𝑧𝑒𝑑\_𝑙𝑖𝑘𝑒𝑠=𝑙𝑖𝑘𝑒𝑠−𝐶𝑇𝑅𝑐𝑜𝑛𝑡𝑟𝑜𝑙∗𝑣𝑖𝑒𝑤𝑠
  + После чего сравним t-тестом отличия в группах по метрике 𝑙𝑖𝑛𝑒𝑎𝑟𝑖𝑧𝑒𝑑\_𝑙𝑖𝑘𝑒𝑠
* Метод простой, гарантируется, что при приличном размере выборки (как у нас — подойдет) можно бесплатно увеличить чувствительность вашей метрики (или, по крайней мере, не сделать хуже). Это ОЧЕНЬ круто.

###### ***Наша задача***:

* Проанализировать тест между группами 0 и 3 по метрике линеаризованных лайков. Видно ли отличие? Стало ли 𝑝−𝑣𝑎𝑙𝑢𝑒 меньше?
* Проанализировать тест между группами 1 и 2 по метрике линеаризованных лайков. Видно ли отличие? Стало ли 𝑝−𝑣𝑎𝑙𝑢𝑒 меньше?

###### ***Вот что у нас получилось***:

* Анализ АВ теста по метрике линеаризованных лайков <https://disk.yandex.ru/d/P_gfVsAH9fr_mA>
* Для экспериментальных групп 0 и 3 t-тест показал статистически значимые различия с очень малыми p-value как на линеаризованных лайках, так и на CTR'ах. Но критерий стат. значимости после t-теста на линеаризованных лайках стал на 9 порядков меньше, чем после теста на CTR'ах, что свидетельствует о бОльшей чувствительности метрики линеаризованных лайков по сравнению с чистыми CTR'ами
* Для экспериментальных групп 1 и 2 t-тест НЕ показал статистически значимых различий на метрике CTR, но тест на линеаризованных лайках дал p-value << 0.05, следовательно показал статистически значимые различия между группами. Такие различия в результатах между тестами еще больше убеждают меня в том, что метрика линеаризованных лайков гораздо чувствительнее чистых CTR'ов.