**Тестовое задание на позицию продуктового аналитика**

Выполнила Перетятько София

1. **Проводится А/Б тестирование двух форм регистрации: на одной пикантное фото девушки на фоне, а на другой - улыбающаяся пара. Для этого аудитория делится случайным образом на две равные группы и каждой из групп показывается страница с одной из форм регистрации. Тестирование проводилось в течение двух дней и дало следующие результаты – см. Excel файл “1. Результаты теста”.**

**a. Как Вы думаете, какая форма регистрации соответствует какому номеру группы разбиения?**

Пикантное фото девушки на фоне привлекает внимание, соответственно пользователи дольше находились на странице и рассматривали фото. Значит, группе «ID 1» показывали фото девушки, потому что среднее время на странице намного больше, чем у группы «ID 2».

У группы «ID 2» выше конверсия в регистрацию на сайте, но меньше среднее время на странице. Предположительно фото пары вызывает больше доверия, и пользователи чаще регистрируются.

**b. Какие выводы можно сделать, на основе имеющейся статистики? Статистически значимы ли результаты?**

На статистическую значимость с помощью z-теста для сравнения двух пропорций можно проверить конверсию в регистрацию. Остальные данные (выручка, среднее время на странице) на статистическую значимость не проверить, потому что не хватает значений в выборках для какого-либо статистического теста и многие тесты требуют нормальное распределение, известную дисперсию.

Проверим количество успешных регистраций на статистическую значимость. Уровень значимости 5%, достоверность результата 95%. Гипотеза - статистически значимый результат отсутствует,- есть статистически значимая разница.

Формула:

где – конверсии в регистрацию для каждой группы на протяжении всего AB теста,

– общая конверсия, – количество уникальных просмотров в каждой группе.

По z-таблице, найдем что значит конверсия в группе ID2 значимо выше, чем в ID1.

**c. Сколько дней нужно проводить данное тестирование, чтобы получить статистически значимые результаты?**

За два дня тест показал статистически значимый результат, но важно обратить внимание, что проведение теста совпало с выходными днями, что могло искусственно завысить конверсии: у людей больше свободного времени, выше вовлеченность, выше конверсии. Для того чтобы результаты можно было обобщать на все дни недели необходимо проводить тест минимум 7 дней, чтобы попали и будни и выходные.

**d. Какие изменения Вы бы внесли в тестирование для получения лучших результатов?**

Для получения лучших результатов можно:

1. Увеличить время проведения теста: проводить тест минимум 7 дней, чтобы охватить все дни недели (лучше 14 дней, если конверсия нестабильна или трафик непредсказуем)
2. Обеспечить правильную сегментацию и рандомизацию: важно, чтобы группы были одинаковы по всем признакам, кроме формы регистрации. Нужно убедиться, что разбиение по группам происходит случайно и равномерно. Также можно добавить контроль по признакам: регион, возрастная когорта.
3. Собрать больше метрик, например, выручка на пользователя (ARPU). Для этого нужно хранить данные на уровне пользователя.
4. Провести А/А тест, чтобы убедиться в корректности измерений.

**e. Какие еще показатели можно добавить в исследование, помимо указанных?**

Помимо указанных метрик можно добавить в исследование:

1. CTR по кнопке регистрации
2. Разница между CTR и конверсией в регистрацию. Позволяет построить воронку регистрации и понять сколько пользователей не завершили регистрацию.
3. ARPU

**2. Обработка логов в аналитической платформе Loginom**

**a. Установите программу (https://loginom.ru/download), либо откройте демо-версию программы онлайн (https://demo.loginom.ru/) и импортируйте приложенный файл “2.**

**Авторизации”, где**

**crc – уникальный идентификатор пользователя**

**date – дата авторизации**

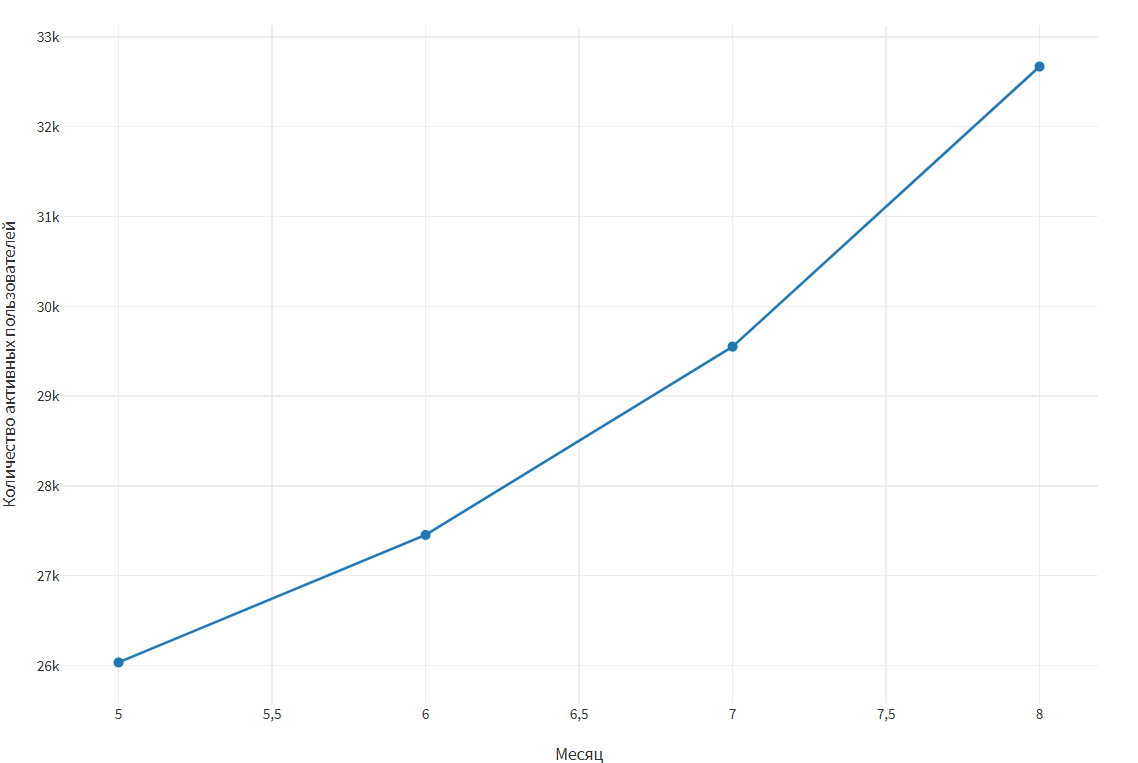
**session\_count – кол-во сессий в конкретный день конкретного пользователя**

**sex – пол (1=м, 2=ж)**

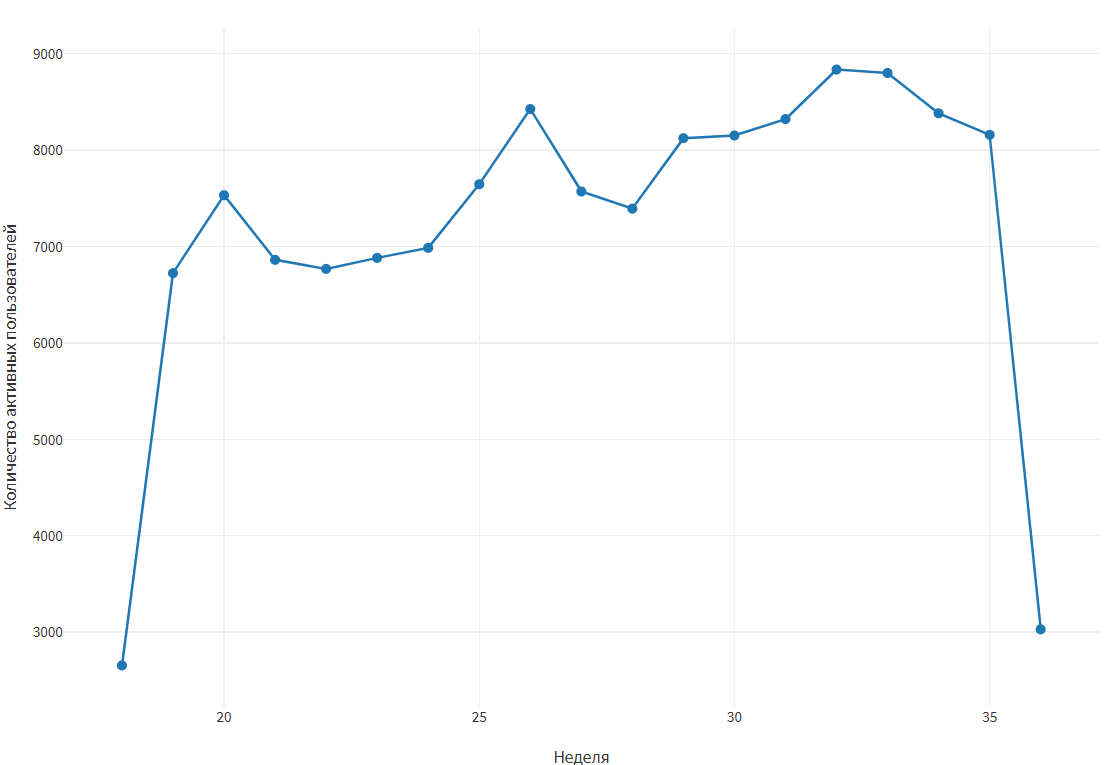
**reg\_date – дата регистрации**

**b. Посчитайте MAU/WAU/DAU.**

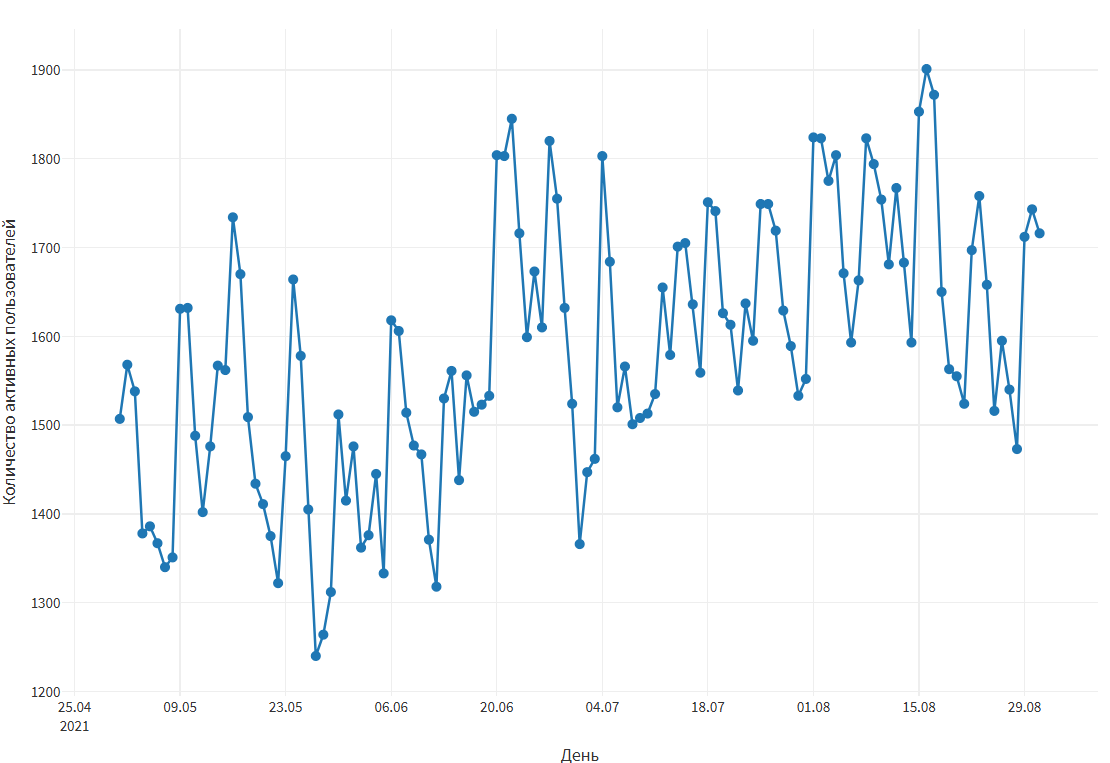
**MAU:**

****

**WAU:**

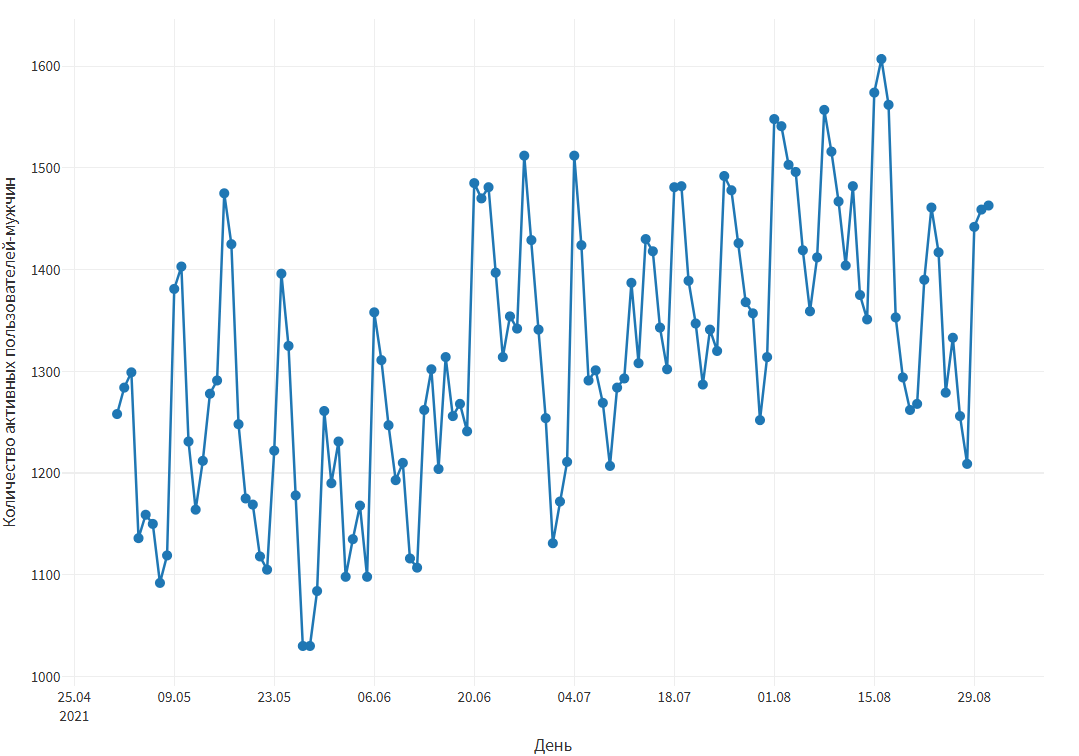
****

**DAU:**

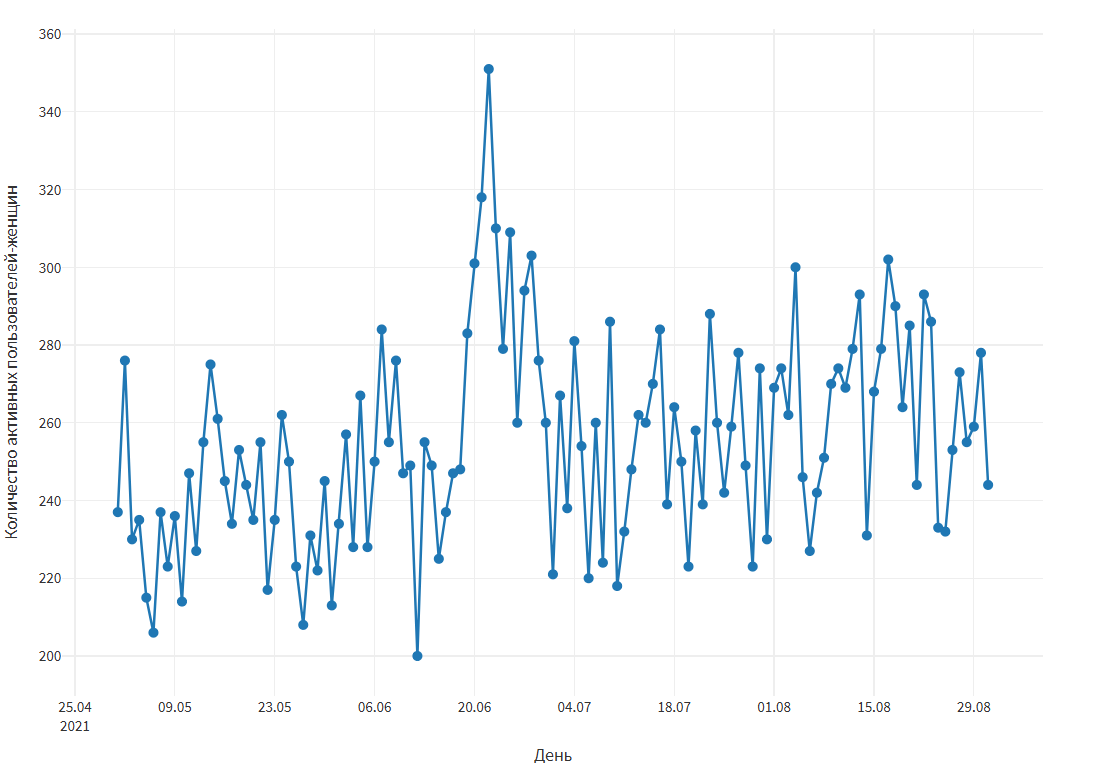
****

**c. Посчитайте DAU отдельно для мужчин и женщин в динамике по дням.**

**DAU для мужчин:**

****

**DAU для женщин:**

****

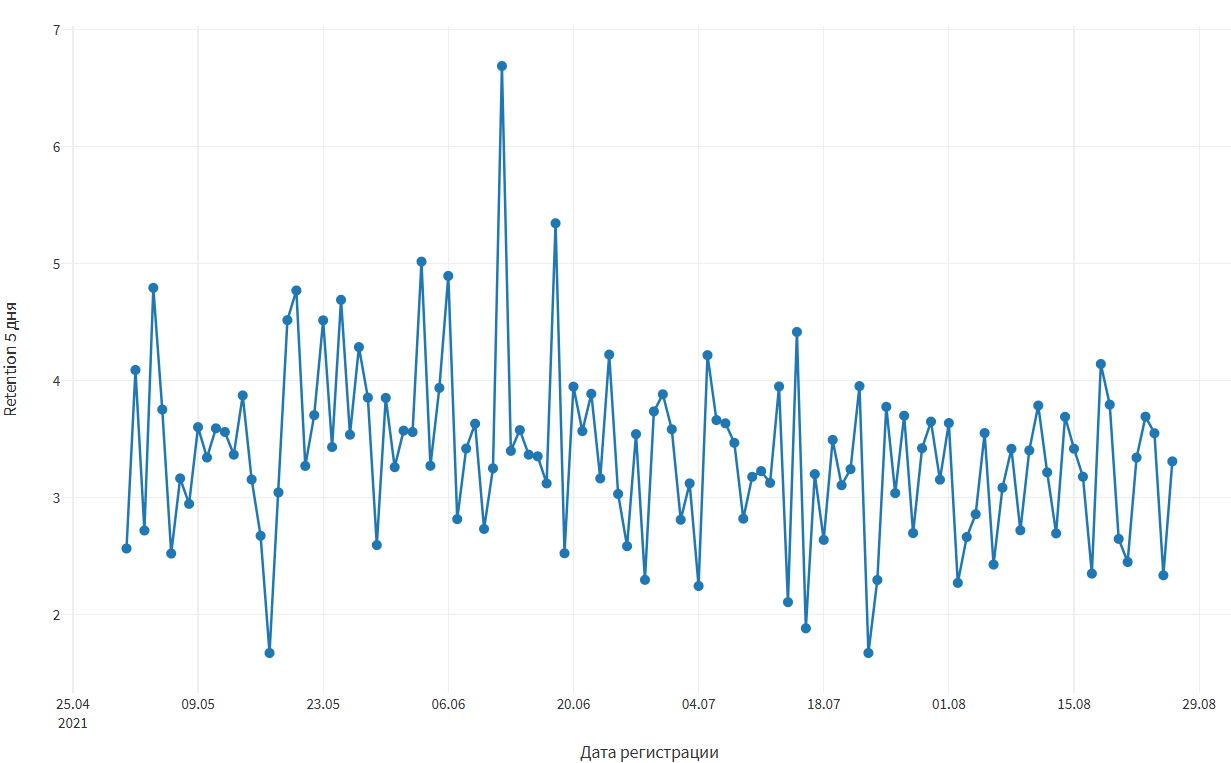
**d. Какие можно сделать выводы?**

Мужчины ведут себя более предсказуемо, чем женщины. DAU по мужчинам показало, что их пик активности происходит каждое воскресенье и активность увеличивается начиная с пятницы.

**e. Какое можно ожидать DAU 5 сентября? 30 сентября?**

С помощью модели ARIMAX был получен прогноз. На 5 сентября – 1646, на 30 сентября – 1572.

**f. Посчитайте Retention 5 дня в динамике (по дням) по регистрациям с мая 2021 по август 2021, приложить скриншот графика. Чему равен Retention 5 дня за август 2021?**



Retention 5 дня за август 2021 равен 3,14%

**К результатам необходимо прикрепить файл\ссылку Loginom с решением, скриншоты графиков и ответ из пункта f.**

**3. В базе есть две таблицы: regs и payments (описание и сэмплы в Excel-файле “3. Для расчета LTV”). Ваша задача:**

**a. Написать SQL запрос, который посчитает среднюю выручку, которую приносит 1 зарегистрированный пользователь за первые 30 дней после регистрации для мужчин и женщин.**

**СУБД PostgreSQL**

with revenue\_per\_user as (

select

crc,

sex,

coalesce(sum(money\_netbeznds), 0) as revenue

from usr\_wrk.regs as regs

left join usr\_wrk.payments as payments

on regs.crc = payments.uid

and payments.date between regs.date and regs.date + interval '30 days'

group by crc, sex

)

select

sex,

avg(revenue) as avg\_revenue\_per\_user

from revenue\_per\_user

group by sex

order by sex asc

;

**b. Написать SQL запрос, который посчитает среднюю выручку, которую приносит 1 зарегистрированный пользователь, который совершил хотя бы 1 платеж, за первые 30 дней после регистрации для мужчин и женщин.**

with revenue\_per\_user as (

select

crc,

sex,

sum(money\_netbeznds) as revenue

from usr\_wrk.regs as regs

inner join usr\_wrk.payments as payments

on regs.crc = payments.uid

and payments.date between regs.date and regs.date + interval '30 days'

group by crc, sex

)

select

sex,

avg(revenue) as avg\_revenue\_per\_user

from revenue\_per\_user

group by sex

order by sex asc

;

**c. Какие настройки должны быть у таблиц, чтобы запросы работали быстрее при больших объемах?**

Для обеспечения эффективности работы SQL-запросов при больших объёмах данных, можно использовать:

1. Индексацию

Для таблицы regs индексация по столбцам:

crc – для соединения с payments.

sex – для группировки.

date – если потребуется фильтрация по дате регистрации.

Для таблицы payments индексация по столбцам:

(uid, date) – составной индекс для фильтрации платежей в 30-дневном окне после регистрации.

1. Партиционирование:

Таблицу payments можно партиционировать по дате платежа — например, по месяцам или кварталам. Это сократит объём данных при выборке по интервалам.