**Внимание!** Для выполнения тестовых заданий скачайте и откройте массив данных по ссылке:

<https://docs.google.com/spreadsheets/d/1EOEmGcBpokRfYbiNBDQs5XnWG9QGmOSwYKpKiOkhQR4/edit?usp=sharing>

1. Во вкладке "Данные об аудитории" информация о пользователях, посетивших наше приложение в ноябре. Чему равен MAU продукта?

\***MAU (Monthly Active Users)** — это метрика, используемая для измерения активности пользователей в течение одного месяца. Она показывает количество уникальных пользователей, которые взаимодействовали с продуктом, сервисом или приложением хотя бы один раз за последний месяц.

7639168141048216529

2. Используя вкладку "Данные об аудитории", посчитайте, чему будет равен DAU

\***DAU (Daily Active Users)** — это метрика, которая показывает количество уникальных пользователей, которые взаимодействовали с продуктом, приложением или сервисом хотя бы один раз в течение дня. DAU помогает понять, сколько пользователей активно пользуются продуктом каждый день.

255490560483

3. Используя вкладку "Данные об аудитории", посчитайте, чему будет равен retention первого дня у пользователей, пришедших в продукт 1 ноября

\*Retention (удержание пользователей) — это метрика, которая показывает, сколько пользователей продолжает пользоваться продуктом через определенный промежуток времени после первоначального взаимодействия. Retention можно рассчитать как процент пользователей, вернувшихся в продукт через определенное время (например, через 1 день, 1 неделю, 1 месяц) от количества всех новых пользователей.

28,3%26,6%38,5%32,7%

4. На графике изображены retention кривые 2 продуктов. Какие выводы можно сделать, глядя на них?



Ваш ответ:

На графике показаны retention кривые для двух продуктов. Из графика видно, что продукт представленный голубым цветом, имеет более высокую retention, чем продукт представленный розовым цветом, т.к. на протяжении всего срока наблюдения, его кривая сохраняется выше, чем кривая второго продукта.

**Сравнение уровня удержания**: Голубая кривая значительно выше розовой, это может свидетельствовать о том, что один продукт лучше удерживает пользователей. Это может быть связано с качеством продукта, пользовательским опытом или маркетингом.

**Скорость снижения удержания**: Розовая кривая показывает резкое падение на ранних этапах, это может означать, что пользователи не находят ценности в продукте и быстро отказываются от него. Более плавное снижение может свидетельствовать о лучшем удержании.

**Долгосрочные тренды**: Если обе кривые начинают схоже, но затем расходятся, это может указывать на различия в том, как пользователи взаимодействуют с продуктами в долгосрочной перспективе.

**Сравнение между сегментами**: Если данные представлены по сегментам пользователей (например, по типу подписки или демографическим характеристикам), можно увидеть, какие группы удерживаются лучше и какие требуют доработки.

**Выявление проблем**: Если наблюдаются аномалии, такие как резкие скачки в retention, это может указывать на проблемы, связанные с обновлениями продукта или изменениями в маркетинговой стратегии.

Важно понимать, что это только визуальная оценка, для точного сравнения двух продуктов нужны дополнительные данные, например, о размере выборки, о причинах оттока пользователей, о факторах влияющих на retention.

5. Во вкладке "Данные об аудитории" есть информация о том, сколько объявлений посмотрел каждый пользователь (view\_adverts). Посчитайте пользовательскую конверсию в просмотр объявления за ноябрь? (в пользователях)

\* Пользовательская конверсия — это метрика, которая показывает, какой процент пользователей выполнил целевое действие по отношению к общему количеству пользователей. В контексте веб-сайтов это может быть действие, такое как просмотр объявления или клик по рекламному баннеру.

41,8%54,7%46,3%39%

6. Используя информацию из вкладки "Данные об аудитории", посчитайте среднее количество просмотренных объявлений на пользователя в ноябре

4,96,25,32,9

7. Мы провели опрос среди 2000 пользователей. Из них 500 «критики», 1200 «сторонники» и 300 «нейтралы». Посчитайте, чему будет равен NPS

\*NPS (Net Promoter Score) — это метрика, которая измеряет лояльность пользователей к компании или продукту и делит их на три группы: Сторонники (Promoters) , Нейтралы (Passives), Критики (Detractors). NPS высчитывается как (% сторонников - % критиков).

30%43%40%35%

8. Во вкладке "Данные АБ-тестов" результаты трех несвязанных АБ тестов для ARPU (общая выручка/общее количество пользователей).  
Посмотрите на результаты тестов и интерпретируйте их. Напишите значения p-value, которые вы получили.  
Подготовьте выводы и рекомендации.   
  
experiment\_num - номер эксперимента  
experiment\_group - группа, в которую попал пользователь  
user\_id - id пользователя  
revenue - выручка, которую сгенерировал пользователь, купив платную услугу продвижения

Ваш ответ:

p-values: {1: nan, 2: nan, 3: nan}

experiment\_num experiment\_group total\_revenue user\_count arpu

1 control 335944 465 722.460215

1 test 319555 480 665.739583

2 control 327664 465 704.653763

2 test 159806 480 332.929167

3 control 308391 465 663.206452

3 test 479361 480 998.668750

**P-значения**: Для всех экспериментов (1, 2, 3) значения p не определены (nan). Это может указывать на то, что для анализа статистической значимости недостаточно данных или возникла ошибка в расчетах.

**Основные наблюдения:**

* В первом эксперименте контрольная группа имеет более высокий ARPU, хотя общий доход в тестовой группе немного ниже.
* Во втором эксперименте наблюдается резкое снижение как общего дохода, так и ARPU в тестовой группе.
* В третьем эксперименте тестовая группа значительно превзошла контрольную по общему доходу и ARPU.

### Выводы:

Данные показывают вариативность результатов между группами, но отсутствие p-значений делает невозможным статистический анализ и выводы о значимости различий между группами.

**Рекомендации:**

1. Провести повторный анализ данных
2. Увеличить выборку
3. Изучить причины различий
4. Проверить методологию тестирования
5. Определить дополнительные метрики

9. По датасету с листерами посчитайте средний доход на пользователя

121.2156.470.930.7средняя здесь не применима

10. По датасету с листерами посчитайте медиану возраста пользователя

27,422827,9327медиана здесь не применима

11. Какой график лучше всего подходит для отображения разброса цен на товары в разных магазинах?  
\*возможно несколько вариантов ответа

Линейный графикКруговая диаграммаЯщик с усами (box plot)

Гистограмма

12. На каком графике бимодальное распределение?

№1

№2

№3

№4

13. Какая случайная величина имеет наибольшую дисперсию данных по следующим графикам плотности распределения?

№1

№2

№3

№4

14. На каком графике можно посчитать коррелцияю?  
\*возможно несколько вариантов ответа









15. Что значит, если при проверке гипотез мы получили p-value = 0.05?

Это означает, что нет никакой статистически значимой разницы между группами

Есть 5% вероятность случайно получить такой или еще более экстремальный результат, если нулевая гипотеза верна

Это означает, что результаты эксперимента на 95% точны

Это говорит о том, что альтернативная гипотеза верна с вероятностью 95%

16. Какой метод наиболее подходит для проверки гипотезы о равенстве средних двух выборок из нормального распределения?

t-тестХи-квадрат тестАнализ дисперсии (ANOVA)Корреляция Пирсона

17. Как интерпретировать квартили в распределении доходов пользователей?

Показывают максимальный и минимальный доход

Делят данные на четыре равные части

Указывают на наиболее часто встречающийся доход

График плотности распределения вещества во вселенной

18. Были получены следующие результаты. Коллеги просят вас подтвердить их и сделать окончательный вывод по эксперименту.

* + - * Вариант A (контрольная группа) — 100 047 501 посетитель, 1003 платежа.
      * Вариант B (тестовая группа) — 100 001 055 посетителей, 1099 платежей.

Какие рекомендации вы бы дали, основываясь на этих данных?

Ваш ответ:

Чтобы сделать окончательный вывод по эксперименту, нужно рассчитать коэффициент конверсии для каждой группы и оценить, есть ли статистически значимая разница между ними.

1. **Рассчитаем коэффициенты конверсии**:

Для варианта A (контрольная группа):

Коэффициент конверсии A=1003/100047501≈0.00001003 (или 0.001003%)

Для варианта B (тестовая группа):

Коэффициент конверсии B=1099/100001055≈0.00001099 (или 0.001099%)

**2. Сравнение коэффициентов**:

Коэффициент конверсии для варианта B выше, чем для A, что может свидетельствовать о положительном влиянии изменений, внедренных в тестовую группу.

**3. Статистическая значимость**:

Чтобы подтвердить, что наблюдаемое различие является статистически значимым, необходимо провести тест (t- тест).

**4. Рекомендации**:

Если результаты статистического теста подтвердят значимость (p-value < 0.05), рекомендуется внедрить изменения из варианта B в продуктивную среду, поскольку они могут улучшить коэффициент конверсии.

Если результаты окажутся незначительными, стоит рассмотреть возможность дальнейших тестов с изменениями или увеличением выборки для получения более точных данных.