**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.

**7639**168141048216529



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.

255490**560**483



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:

Глядя на кривые retention двух продуктов, можно сделать вывод что с течением времени retention rate у продукта 1 ухудшается намного стремительнее достигая 0% всего за 5 дней, в то время как у продукта 2 retention rate стабилизируется примерно на 30-35% от первоначальных пользователей. Продукт 1, в отличие от продукта 2, вообще не цепляет долгосрочных пользователей, что может быть обусловлено различными причинами (не нужный функционал, неудобный интерфейс, баги, некрасивое оформление и тд.). Пользователи продукта 2 напротив отсеиваются медленнее и в итоге примерно 30-35% от первоначальных пользователей в итоге могут перейти в категории постоянных пользователей, которых можно будет монетизировать и проводить различные кампании с ними. Такой процент удержания явно показывает более высокое качество продукта по сравнению с продуктом 1, но в то же время указывает на необходимость развития, чтобы увеличить процент удержания первоначальных пользователей скажем до 50% и выше.

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%



Ответ на этот вопрос не выходил в google sheets. Попробовал в MySQL, получился ответ.

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

4,96,25,3**2,9**



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:

A screenshot of a computer

AI-generated content may be incorrect.

Как показал анализ, все пользователи участвовали на всех трех экспериментах. Первый эксперимент не показал никакой статистически значимой разницы в ARPU между группами. Второй эксперимент показал, что ARPU экспериментальной группы значительно меньше ARPU контрольной группы. Третий же эксперимент, наоборот, показал, что ARPU экспериментальной группы значительно выше ARPU контрольной группы. Предполагая, что эти эксперименты проводились с целью испробовать новые функционалы продукта или их комбинации и увидеть влияние на ARPU, и что каждый эксперимент использовался для тестирования определенного функционала или же комбинаций функционалов, можно сделать вывод что функционал или комбинация функционалов тестируемые в эксперименте 3 значительно повлияли на увеличение ARPU. Обратная картина наблюдается в эксперименте 2, в то время как в эксперименте 1 не наблюдается существенной разницы. Отсюда следуют рекомендации, что в целях увеличения выручки с продукта, стоит запустить функционалы продукта, которые тестировались в эксперименте 3; убрать и не внедрять функционал из эксперимента 2 из-за его плохого влияния на выручку; по похожей причине, не внедрять функционал из эксперимента 1, несмотря на то что он не влияет на выручку, разработка и внедрение этого функционала будет стоить ресурсов, которые можно будет потратить на дальнейшее развитие функционала продукта из эксперимента 3.

9. По датасету с листерами посчитайте средний доход на пользователя

121.2**156.4**70.930.7средняя здесь не применима



10. По датасету с листерами посчитайте медиану возраста пользователя

27,42**28**27,9327медиана здесь не применима



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

Линейный графикКруговая диаграмма**Ящик 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





(можно посчитать, но коэффициент корреляции будет очень маленьким)



(heatmap здесь уже показывает корреляции между переменными)







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-тест**Хи-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:

Мы сравниваем конверсию между двумя выборками. Используя гугл, ютуб, проводим Two Proportions Z-test чтобы узнать есть ли разница между их конверсиями.

Двусторонний тест:



Односторонний тест (альтернативная гипотеза: конверсия контрольной группы меньше):



Как видно по результатам теста, конверсии у них разные и конверсия контрольной группы значительно меньше конверсии экспериментальной группы. Отсюда есть вероятность сделать вывод что функционал продукта, который мы тестировали, увеличивает конверсию. Однако, если мы обратим внимание на выборки, у нас нереально огромные выборки с очень маленьким количеством платежей в обоих выборках. Из-за чувствительности теста к огромным выборкам, даже при ничтожно маленькой разнице конверсии мы можем получить статистически значимый результат. Чтобы посмотреть на более практическую значимость нашего теста, посмотрим на размер эффекта *h* для Two Proportions Z-test:



Отсюда видим, что мы имеем ничтожно маленькое улучшение в конверсии, что не стоит внедрения. Отсюда финальный вывод: не внедрять новый функционал в продукт, так как ничтожный прирос в конверсиях не оправдывает усилий. Да и в целом, если посмотреть на конверсию в контрольной группе, она так же очень мала. Будет целесообразно в целом заменить систему или концепцию продукта, чтобы улучшить конверсию.