**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:

 Голубая линия показывает, что retention для первого продукта выше на всех днях, начиная с Day 1. Кривая удержания постепенно снижается, но удерживает значительную часть пользователей, что указывает на стабильное вовлечение пользователей в этот продукт.

 Красная линия демонстрирует более резкое падение retention уже на Day 1, с дальнейшим стремительным снижением в последующие дни, практически достигая нуля к Day 5. Это говорит о том, что пользователи быстрее покидают второй продукт после первоначального взаимодействия.

**Заключение**:

* **Первый продукт (голубая линия)** имеет значительно более успешные показатели удержания, что может свидетельствовать о его более долгосрочной привлекательности для пользователей.
* **Второй продукт (красная линия)** теряет пользователей значительно быстрее, что может потребовать улучшений в первых днях взаимодействия с продуктом, чтобы увеличить удержание.

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:

Control ARPU: 722.4602150537635,

Test ARPU: 665.7395833333334,

p-value: 0.688784211779017

**Интерпретация:**

1. **Среднее ARPU в контрольной группе** выше, чем в тестовой группе (722.46 против 665.74), но разница незначительна.
2. **p-value** = 0.689, что значительно больше стандартного порога 0.05, указывая на то, что различия между группами не являются статистически значимыми.

**Выводы:**

* Мы не можем отвергнуть нулевую гипотезу, что означает, что тестируемое изменение не оказало значимого влияния на ARPU.
* Различие между средними значениями ARPU в контрольной и тестовой группах могло быть вызвано случайными факторами.

**Рекомендации:**

* Поскольку различие незначительно и p-value указывает на отсутствие статистической значимости, **нет необходимости внедрять тестируемое изменение**.
* Возможно, стоит провести дополнительные эксперименты или улучшить текущий тест, чтобы получить более значимые результаты (например, увеличить выборку или пересмотреть стратегию эксперимента).

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:

1. **Рассчитаем коэффициенты конверсии**:
   * **Вариант A (контрольная группа)**: Коэффициент конверсии A=1003/100047501≈0.00001003
   * **Вариант B (тестовая группа)**: Коэффициент конверсии B=1099/100001055≈0.00001099
2. **Сравним коэффициенты конверсии**:
   * **Коэффициент конверсии A** = 0.00001003 (0.001003%)
   * **Коэффициент конверсии B** = 0.00001099 (0.001099%)

Тестовая группа (Вариант B) показывает немного более высокий коэффициент конверсии по сравнению с контрольной группой (Вариант A).

1. **Статистическая значимость**:
   * Перед тем как делать окончательные выводы, важно определить, является ли разница между этими коэффициентами статистически значимой. Для этого обычно используется **A/B тест** с такими статистическими методами, как **z-тест для пропорций** или **критерий хи-квадрат**, чтобы подтвердить, что разница не случайна.
   * Если разница окажется статистически значимой, можно рекомендовать внедрение Варианта B, так как он показывает лучший результат.
2. **Заключение**:
   * Если статистические тесты покажут значимость разницы, я бы **рекомендовал Вариант B**, так как он имеет более высокий коэффициент конверсии.
   * Если разница **не будет статистически значимой**, я бы рекомендовал продолжить использовать Вариант A или провести дополнительные тесты для получения более точных данных.

На основании предварительного анализа можно склоняться к **Варианту B**, так как он демонстрирует лучшую конверсию, но окончательные выводы необходимо подтвердить с помощью статистической проверки.