**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:

По графику retention кривых двух продуктов (красная и синяя линии) можно сделать следующие выводы:

**1. Синий продукт удерживает пользователей лучше, чем красный:**

* С первого дня (Day 1) retention синего продукта выше и убывает медленнее.
* На Day 1 retention синего продукта ≈ 70%, тогда как у красного продукта около 55-60%.
* К Day 5 retention синего продукта стабилизируется на уровне 35-40%, тогда как retention красного продукта стремится к нулю.

**2. Красный продукт теряет пользователей быстрее:**

* У красного продукта очень резкий спад retention в первые 2-3 дня.
* Уже к Day 5 почти никто из пользователей не возвращается.

**3. Стабильность синего продукта:**

* Несмотря на падение в первые дни, синяя кривая выходит на "плато" и удерживает примерно 35-40% пользователей к Day 5 и дальше.
* Это говорит о более высоком качестве продукта или лучшем пользовательском опыте.

**4. Красный продукт нуждается в доработке онбординга/первого опыта:**

* Быстрая потеря пользователей может говорить о проблемах с первичным онбордингом, полезностью продукта или плохим "первым впечатлением".

**Вывод:**

* **Синий продукт обладает сильной ретеншн-механикой и высоким качеством пользовательского опыта.**
* **Красный продукт требует оптимизации пользовательского пути, особенно в первые дни после привлечения.**

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:

**Эксперимент 1**

* **ARPU (Control):** 722.46
* **ARPU (Test):** 665.74
* **Uplift:** -7.85%
* **p-value:** 0.6890 → **нет статистически значимого эффекта**

**Вывод:**  
Тестовая группа показала падение ARPU на 7.85%, однако p-value = 0.6890 говорит о том, что разница статистически незначима. Можно сказать, что тест не оказал влияния на поведение пользователей.

**Рекомендация:**  
Не внедрять изменения, провести дополнительный анализ по сегментам (возможно, эффект скрывается в подгруппах).

**Эксперимент 2**

* **ARPU (Control):** 704.65
* **ARPU (Test):** 332.93
* **Uplift:** -52.75%
* **p-value:** 0.0011 → **статистически значимое ухудшение**

**Вывод:**  
Тестовая группа показала существенное падение ARPU на 52.75%, и это падение статистически значимо (p-value = 0.0011).

**Рекомендация:**  
Тестовое изменение негативно влияет на выручку. Не внедрять. Нужно срочно анализировать причины (возможно, технические ошибки или UX-проблемы).

**Эксперимент 3**

* **ARPU (Control):** 663.21
* **ARPU (Test):** 998.67
* **Uplift:** +50.58%
* **p-value:** 0.0603 → **эффект незначим на уровне 5%, но близко**

**Вывод:**  
Uplift в +50.58% выглядит многообещающе, однако p-value = 0.0603 чуть выше порога значимости (0.05). Это может говорить о недостаточной выборке.

**Рекомендация:**  
Провести повторный тест с увеличением объема выборки. Если тренд сохранится, тест может быть успешным. Также рассмотреть сегментацию данных.

**Общие рекомендации:**

1. **Эксперимент 2** — **отклонить гипотезу** (отрицательный эффект).
2. **Эксперимент 1** — **не внедрять** (нет эффекта).
3. **Эксперимент 3** — **доработать и протестировать повторно** (недостаточно данных, но положительная динамика).

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:

Конверсия A: 0.001003%

Конверсия B: 0.001099%

Относительный рост: 9.62%

Z-статистика: -2.1046

P-value: 0.017665

Вывод: Разница статистически значима. Вариант B лучше.

### Рекомендации

* **Внедрить вариант B**, так как он обеспечивает значимое увеличение конверсии.
* **Оценить экономический эффект:**  
  Дополнительные платежи: 1099−1003=961099−1003=96 на ~200M посетителей.  
  Если средний чек = X*X*, прибыль = 96⋅X96⋅*X*. Стоит сопоставить с затратами на изменения.
* **Проверить корректность данных:**  
  Убедиться, что в тестовой группе не было искажений (например, сегментации трафика).
* **Мониторинг долгосрочных эффектов:**  
  Иногда краткосрочный рост конверсии не сохраняется со временем.

**Итог:** Вариант B предпочтителен, но важно оценить его влияние на бизнес-метрики.