**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 изображены кривые удержания (retention) двух продуктов за период 7 дней.

1. **Начальный уровень удержания**:
   * Оба продукта начинают с 100% удержания в первый день, что означает, что в начале оба продукта имеют полный набор пользователей, которые впервые воспользовались продуктом.
2. **Скорость падения удержания**:
   * Красная кривая (второй продукт) показывает более быстрое падение удержания по сравнению с синей кривой (первый продукт). Это говорит о том, что пользователи второго продукта быстрее теряют интерес или не находят ценности в продукте, поэтому они покидают его быстрее.
   * Синяя кривая более пологая, что свидетельствует о лучшем удержании пользователей. Она демонстрирует, что продукт лучше удерживает пользователей, и они продолжают использовать его на протяжении более длительного времени.
3. **Уровень удержания на конец периода**:
   * К концу 7-дневного периода первый продукт (синяя кривая) удерживает около 30-40% пользователей, что является хорошим показателем.
   * Второй продукт (красная кривая) к концу периода удерживает менее 10% пользователей, что указывает на проблемы с удержанием и, возможно, требует изменений в продукте или улучшений для повышения интереса пользователей.

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:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

1. **Эксперимент 1**:
   * **ARPU (Control):** 722.46
   * **ARPU (Test):** 665.74
   * **p-value:** 0.68896
   * **Вывод:** P-value значительно выше уровня значимости 0.05, что указывает на отсутствие статистически значимой разницы между контрольной и тестовой группами. Это означает, что изменения, внесенные в тестовой группе, не оказали существенного влияния на ARPU. **Рекомендация**: нет необходимости внедрять изменения на основе этого теста.
2. **Эксперимент 2**:
   * **ARPU (Control):** 704.65
   * **ARPU (Test):** 332.92
   * **p-value:** 0.01128
   * **Вывод:** P-value меньше уровня значимости 0.05, что указывает на статистически значимую разницу между контрольной и тестовой группами. ARPU тестовой группы существенно ниже, чем у контрольной. Это говорит о том, что изменения в тестовой группе негативно повлияли на выручку. **Рекомендация**: не внедрять изменения, так как они снизили ARPU.
3. **Эксперимент 3**:
   * **ARPU (Control):** 663.21
   * **ARPU (Test):** 998.67
   * **p-value:** 0.00212
   * **Вывод:** P-value также меньше 0.05, что указывает на статистически значимую разницу. В этом случае ARPU тестовой группы выше, чем у контрольной. Это свидетельствует о положительном влиянии изменений в тестовой группе на выручку. **Рекомендация**: рассмотреть возможность внедрения изменений, так как они способствовали увеличению ARPU.

**Общие рекомендации:**

* **Эксперимент 1**: Нет статистически значимого эффекта — изменения не нужны.
* **Эксперимент 2**: Изменения в тестовой группе привели к значительному снижению ARPU, поэтому лучше оставить текущую версию без изменений.
* **Эксперимент 3**: Изменения положительно повлияли на ARPU, поэтому их можно рекомендовать к внедрению.

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:

 Рассчитаем конверсию для каждой группы:

* **Конверсия A** = (1003 / 100 047 501) × 100 ≈ 0.001003%
* **Конверсия B** = (1099 / 100 001 055) × 100 ≈ 0.001099%

 **Сравнение конверсий**:

* Конверсия в группе B немного выше, чем в группе A. Разница составляет около 0.000096%.

 **Рекомендации**:

* Несмотря на небольшую разницу в конверсии, для окончательного вывода необходима статистическая проверка (например, z-тест или t-тест) на значимость этой разницы.
* Если разница статистически значима (p-value ≤ 0.05), можно рекомендовать использование варианта B, так как он показывает немного лучшую конверсию.
* Если разница не является значимой, можно считать, что оба варианта имеют схожую эффективность, и нет необходимости в изменениях.