**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: у СИНЕГО ПРОДУКТА retention намного выше, чем у КРАСНОГО ПРОДУКТА. Синий продукт теряет пользователей медленнее — даже через неделю у него удерживается около 40% пользователей. А у красного продукта почти все уходят в первые дни — через неделю остается меньше 10%.

Это значит, что у синего продукта лучше первый опыт и он интереснее людям. А красному продукту срочно нужно улучшать первые впечатления и разбираться, почему пользователи так быстро уходят.

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 контрольной группы:** 722.46
* **ARPU тестовой группы:** 665.74
* **Разница:** -56.72 (-7.8%)
* **Статистическая значимость:** p-value ≈ 0.08 (> 0.05)
* **Вывод:** Различия между группами не являются статистически значимыми. Тестовое изменение не оказало существенного влияния на выручку.
* **Рекомендация:** Изменение можно не внедрять. Рекомендуется провести дополнительное исследование для понимания причин небольшого снижения выручки.

**Эксперимент 2**

* **ARPU контрольной группы:** 704.65
* **ARPU тестовой группы:** 332.93
* **Разница:** -371.72 (-52.7%)
* **Статистическая значимость:** p-value < 0.001
* **Вывод:** Тестовое изменение привело к катастрофическому падению выручки более чем на 50%. Эффект статистически значим.
* **Рекомендация:** Немедленно откатить изменение. Провести анализ причин такого резкого падения: возможно, изменение ухудшило пользовательский опыт или нарушило процесс monetization.

**Эксперимент 3**

* **ARPU контрольной группы:** 663.21
* **ARPU тестовой группы:** 998.67
* **Разница:** +335.46 (+50.6%)
* **Статистическая значимость:** p-value < 0.001
* **Вывод:** Тестовое изменение показало исключительно положительный результат с ростом выручки на 50%. Эффект высоко статистически значим.
* **Рекомендация:** Успешное изменение. Рекомендуется к немедленному внедрению для всех пользователей. Дополнительно можно проанализировать, на каких сегментах пользователей эффект наиболее выражен.

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:

Вариант B вроде лучше: конверсия немного выросла (с 0.001002% до 0.001099%), относительный рост почти 10%. Но выборки огромные, и разница совсем маленькая. Надо бы проверить статистику — вдруг это случайность? Без теста точно сказать нельзя. Если статистика подтвердит, можно внедрять B, но экономический эффект будет очень small.