**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 7639

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 два разных продукта

* В первый день (Day 0) retention у обоих продуктов составляет 100%, что логично — все пользователи только пришли.
* У розовой линии падение происходит очень быстро: уже к пятому дню почти никто из пользователей не остается.
* У голубой линии снижение более плавное: даже через неделю около 38–40% пользователей продолжают пользоваться продуктом.
* Продукт с голубой линией показывает лучшее удержание пользователей — он, вероятно, более полезный или интересный для аудитории.
* Продукт с розовой линией быстро теряет пользователей, что может указывать на проблемы с качеством продукта, неудобством использования или недостаточной ценностью для пользователей.
* Если выбирать между этими двумя продуктами, продукт с голубой кривой выглядит предпочтительнее.

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:

**Выводы по A/B тестам**

**Эксперимент 1:**

* ARPU в контрольной группе: **722.46**
* ARPU в тестовой группе: **665.74**
* p-value = **0.688966**

Статистически значимых различий между контрольной и тестовой группой нет (p-value > 0.05). Нововведения не оказали влияния на выручку.

**Эксперимент 2:**

* ARPU в контрольной группе: **704.65**
* ARPU в тестовой группе: **332.93**
* p-value = **0.001128**

В этом тесте разница между группами получилась значимой (p-value < 0.05). При этом в тестовой группе ARPU заметно снизился, значит изменение повлияло на метрику негативно.

**Эксперимент 3:**

* ARPU в контрольной группе: **663.21**
* ARPU в тестовой группе: **998.67**
* p-value = **0.060315**

Хотя тестовая группа показывает более высокий ARPU, уровень значимости чуть выше порога (p-value около 0.06). Можно сказать, что результат пограничный, и чтобы сделать окончательные выводы, стоит провести тест повторно на большем числе пользователей.

* По **первому тесту**: изменений внедрять не стоит, так как значимых отличий нет.
* По **второму тесту**: нововведения лучше отменить, так как они ухудшают выручку.
* По **третьему тесту**: рекомендую провести повторный тест на расширенной выборке для более точной оценки эффекта.

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 №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 = 1003/100047501~~ 0,001003%

Конверсия  B = 1099/100001055 ~~ 0,001099 %

Конверсия в группе B немного выше, однако в целом оба значения крайне малы (~~0,001%). Разница в количестве платежей (1003 против 1099) на фоне более чем 100 миллионов пользователей является незначительной.

z-статистика = 2,10 , а p-value ~~ 0,035

Так как p-value меньше 0,05, можно сказать, что разница между A и B есть и она значимая.   
То есть вариант B лучше, и его можно рекомендовать.  
Но улучшение совсем небольшое, так что в реальности оно может быть почти незаметным.