**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: Давайте для начала обозначим эти продукты как продукт “Синий” и продукт “Красный”. Мы видим что градация обоих продуктов к окончанию первого дня приблизительно равна, но в промежутке 1-2 дней имеется расхождение в retention кривых обоих продуктов. Мы видим, что активность пользователей, который по сути и показывает retention value, красного продукта стремительно уменьшается к нулевому проценту, что означает полную неактивность пользователей, что нельзя сказать про синий продукт. Его активность конечно упала, но в период с 5-7 дни мы видим, что изменения в активности некритичны и, очень грубо говоря, практические неизменны, что может свидетельствовать о стабильности.

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

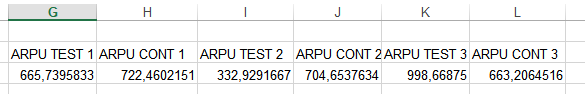
7. Мы провели опрос среди 2000 пользователей. Из них 500 «критики», 1200 «сторонники» и 300 «нейтралы». Посчитайте, чему будет равен NPS

\*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: После расчёта ARPU для каждой группы в каждом эксперименте и проведения т-теста на 3 экспериментах, мы получили следующие данные:



Мы видим, что средняя выручка за клиента в первых двух экспериментах контрольной группы выше(особенно во втором эксперименте), что может говорить о том, что, если считать контрольную группу как ту, в которой после тестовой внесли некоторые изменения, эти самые изменения могут пойти на пользу, так как в итоге ср. выручка на клиента увеличилась. В 3 же эксперименте тестовая(пробная) группа показала более лучшие результаты, что может говорить наоборот о том, что внесенные изменения ведут лишь к ухудшению.

Данные т-теста  


Здесь мы провели т-тест по 2 группам эксперимента 1-2-3 и получили следующие p-значения. Сразу говорим, что использовали готовую функци “ТТЕСТ/T.TEST” для этой операции. Наша нулевая гипотеза состояла в том, что между двумя группами нет фактической разницы. Из 3 экспериментов мы получили приблизительно одинаковые значения в 71,8%. Это определённо больше a=0,05, что является стандартным уровнем значимости, а это значит, что вероятность получить такие или более экстремальные различия в ARPU составляет как раз 71.8%

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: Даже невооруженным глазом мы можем видеть, что Варинт В показывает себя чуть лучше, чем Вариант А. Можно так же сравнить их конверсии или коэффициенты конверсии. То есть, в нашем случае, найти отношение количества действий, совершенных клиентами, на само общее количество клиентов. (1003/100047501)≈0,001%, (1099/100001055)≈0,0011%. Мы может наблюдать аналогичный исход того, что Вариант В немного лучше. Это означает, что изменения контрольной группы А были успешны.