**Внимание!** Для выполнения тестовых заданий скачайте и откройте массив данных по ссылке:

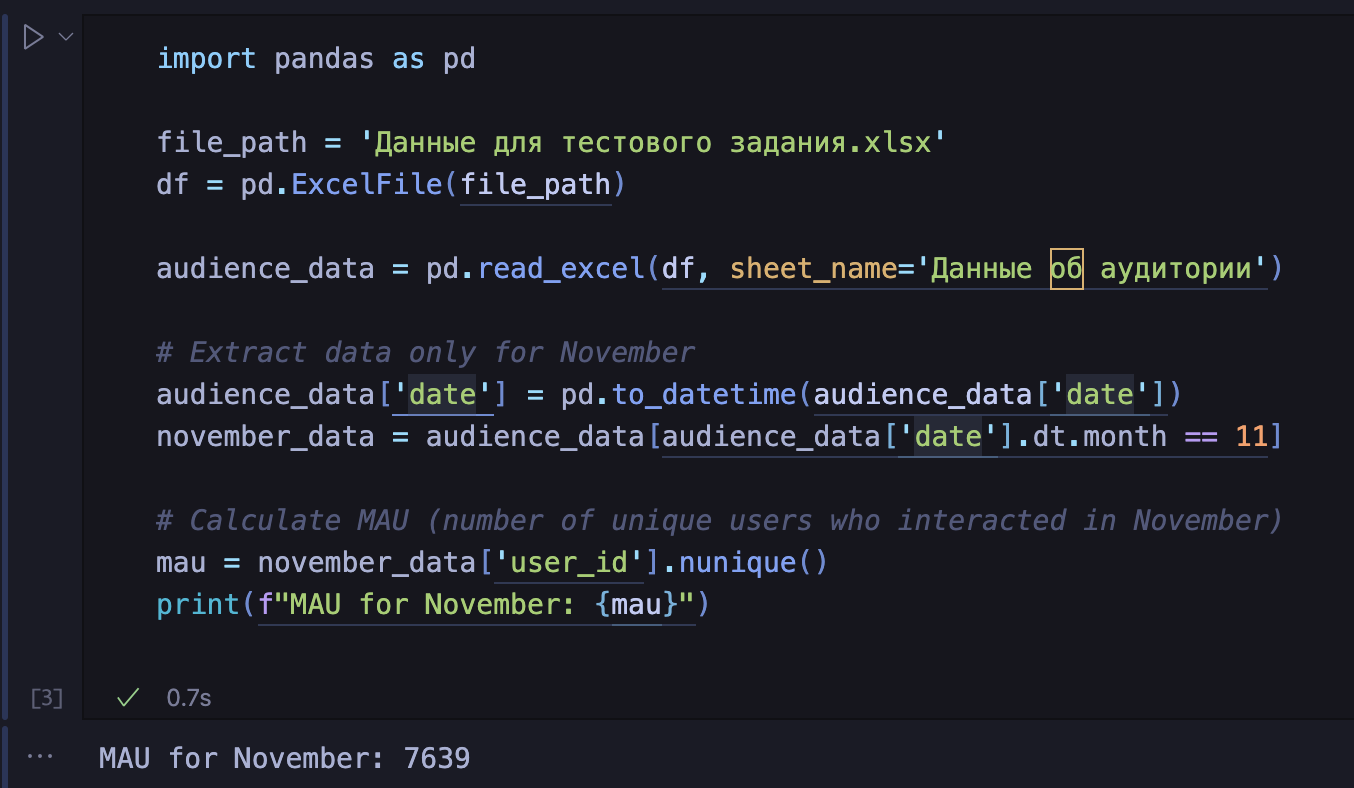
<https://docs.google.com/spreadsheets/d/1EOEmGcBpokRfYbiNBDQs5XnWG9QGmOSwYKpKiOkhQR4/edit?usp=sharing>

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



A computer screen shot of a code

Description automatically generated

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%



A screenshot of a computer program

Description automatically generated

4. На графике изображены retention кривые 2 продуктов. Какие выводы можно сделать, глядя на них?



Ваш ответ:

Продукт с синей кривой демонстрирует более высокое удержание пользователей на протяжении всей недели, с уровнем удержания около 60% на 2-й день и примерно 40% на 7-й день, в то время как продукт с красной кривой показывает быстрое снижение, почти полностью теряя пользователей к 4-му дню. Это может указывать на более высокую вовлеченность и ценность синего продукта для пользователей, в то время как красному продукту стоит улучшить функциональность или опыт использования, чтобы повысить удержание.

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%



A screen shot of a computer program

Description automatically generated

6. Используя информацию из вкладки "Данные об аудитории", посчитайте среднее количество просмотренных объявлений на пользователя в ноябре

4,96,25,32,9



A screen shot of a computer

Description automatically generated

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%



A computer screen with text and numbers

Description automatically generated

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 program

Description automatically generated

**Эксперимент 1**: Значение p-value для первого эксперимента составляет 0.68897, что значительно выше 0.05. Это означает, что различия между контрольной и тестовой группой статистически незначимы. Таким образом, изменения в тестовой группе, вероятно, не повлияли на ARPU, и мы не отвергаем нулевую гипотезу.

**Эксперимент 2**: p-value в этом эксперименте равно 0.00113, что значительно ниже порогового значения 0.05. Это указывает на статистически значимые различия между контрольной и тестовой группой. Мы отвергаем нулевую гипотезу и можем с уверенностью утверждать, что изменения в тестовой группе повлияли на ARPU. Следовательно, рекомендуется рассмотреть внедрение изменений, тестировавшихся в этом эксперименте, на более широкую аудиторию.

**Эксперимент 3**: В этом эксперименте p-value составляет 0.06032, что близко к 0.05, но все еще выше этого порога. Это означает, что различия между контрольной и тестовой группой не являются статистически значимыми на уровне значимости 5%. Следовательно, на данный момент мы не отвергаем нулевую гипотезу, но для получения более уверенных результатов необходимо провести дополнительные исследования.

**Выводы и рекомендации**:

Результаты эксперимента 1 показывают, что изменения не повлияли на ARPU, и дальнейшее внимание к этому эксперименту не требуется. Эксперимент 2, напротив, показал значимый эффект, и на основе этих результатов стоит рассмотреть возможность внедрения изменений, так как они оказали положительное влияние на ARPU. В случае с экспериментом 3 результат оказался на грани статистической значимости, поэтому рекомендуется провести повторный тест с более широким набором данных для уточнения результатов.

В целом, рекомендуется сосредоточиться на эксперименте 2 и использовать его результаты для принятия решений, а также продолжить анализ по эксперименту 3, чтобы подтвердить или опровергнуть текущие выводы.

9. По датасету с листерами посчитайте средний доход на пользователя

121.2156.470.930.7средняя здесь не применима  
A screen shot of a computer program

Description automatically generated



10. По датасету с листерами посчитайте медиану возраста пользователя

27,422827,9327медиана здесь не применима  
A screen shot of a computer code

Description automatically generated



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



Box plot, гистограмма

12. На каком графике бимодальное распределение?

№1



№2



№3



№4



#3

13. Какая случайная величина имеет наибольшую дисперсию данных по следующим графикам плотности распределения?

№1



№2



№3



№4



#3

14. На каком графике можно посчитать коррелцияю?  
\*возможно несколько вариантов ответа

















#1

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%



Ответ: Есть 5% вероятность случайно получить такой или еще более экстремальный результат, если нулевая гипотеза верна

16. Какой метод наиболее подходит для проверки гипотезы о равенстве средних двух выборок из нормального распределения?

t-тестХи-квадрат тестАнализ дисперсии (ANOVA)Корреляция Пирсона



t-тест

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



Делят данные на четыре равные части

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:

В представленных данных у тестовой группы конверсия немного выше, чем у контрольной группы. Однако чтобы сделать окончательный вывод, нужно провести тест, который покажет, является ли эта разница значимой или могла возникнуть случайно. Если тест покажет, что вероятность случайности мала (менее 5%), можно рекомендовать вариант B, так как он действительно работает лучше. Если разница окажется случайной, то внедрять вариант B смысла нет, и можно оставить вариант A. Важно также проверить корректность данных и другие показатели, чтобы иметь полное представление о результате эксперимента.