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

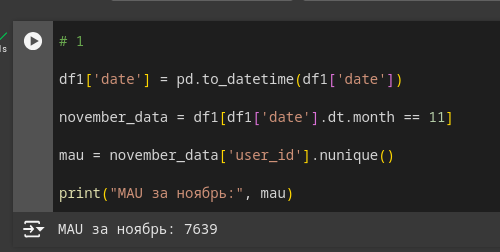
<https://docs.google.com/spreadsheets/d/1EOEmGcBpokRfYbiNBDQs5XnWG9QGmOSwYKpKiOkhQR4/edit?usp=sharing>

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.

**7639** 168141048216529

**7639**

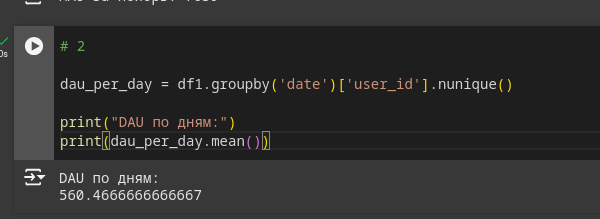


2. Используя вкладку "Данные об аудитории", посчитайте, чему будет равен DAU

\*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.

255490 **560** 483

**560**

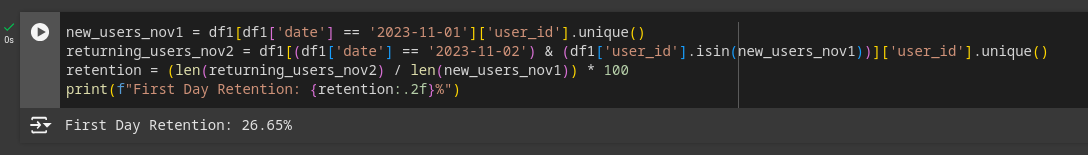


3. Используя вкладку "Данные об аудитории", посчитайте, чему будет равен retention первого дня у пользователей, пришедших в продукт 1 ноября

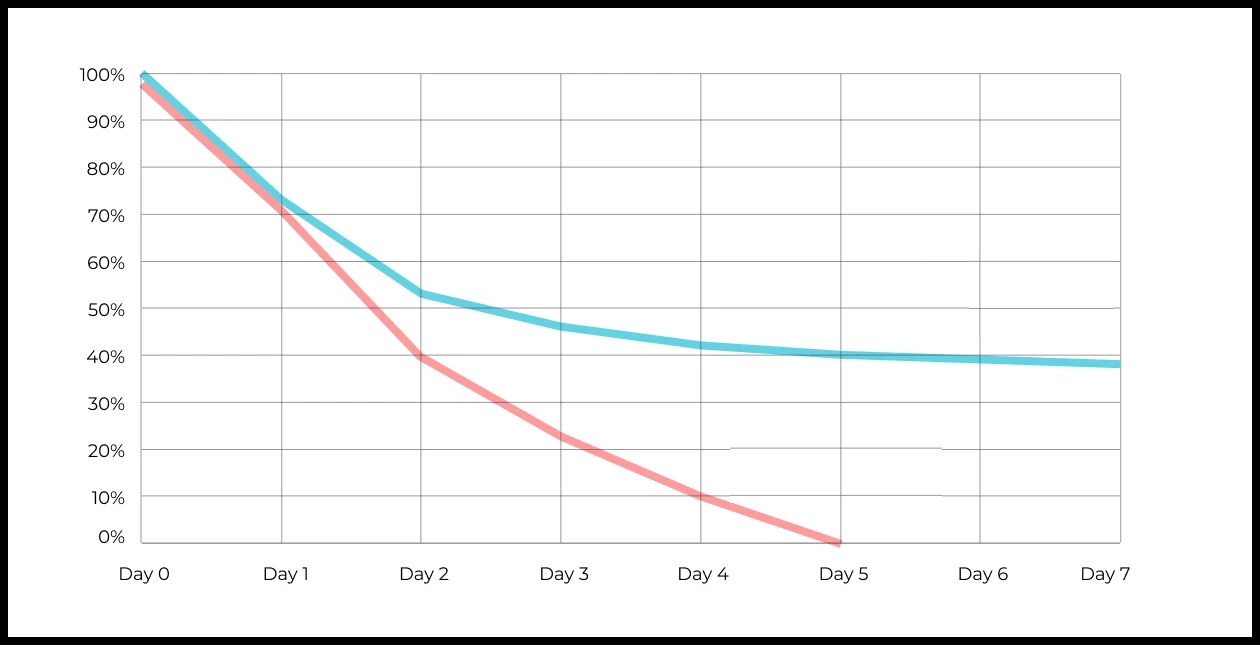
\*Retention (удержание пользователей) — это метрика, которая показывает, сколько пользователей продолжает пользоваться продуктом через определенный промежуток времени после первоначального взаимодействия. Retention можно рассчитать как процент пользователей, вернувшихся в продукт через определенное время (например, через 1 день, 1 неделю, 1 месяц) от количества всех новых пользователей.

28,3% **26,6%** 38,5%32,7%

**26.66%**



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



Ваш ответ:

The retention curves displayed in the chart represent user retention rates over a 7-day period for two distinct products, visualized through two color-coded lines. The initial point (Day 0) shows both products at 100% retention, indicating a full user base at the outset.

The key observation from the curves is that retention drops substantially over the first few days for both products. However, the product represented by the blue line demonstrates higher retention over time compared to the product represented by the red line. By Day 3, the red line's retention rate is below 20%, showing a steep decline that continues until near-zero retention by Day 5. In contrast, the blue line experiences a slower, more gradual decline, stabilizing around 20-30% by Day 7.

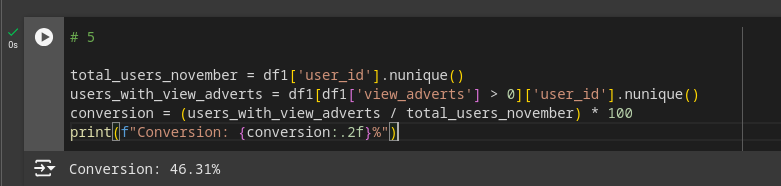
This suggests that users engage with the blue-line product for a longer period than with the red-line product, indicating a potentially higher user satisfaction or a more engaging product experience. The red-line product, with its rapid drop-off, may face issues with initial user retention, possibly due to unmet user expectations, lack of engagement, or competition with other offerings.

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%

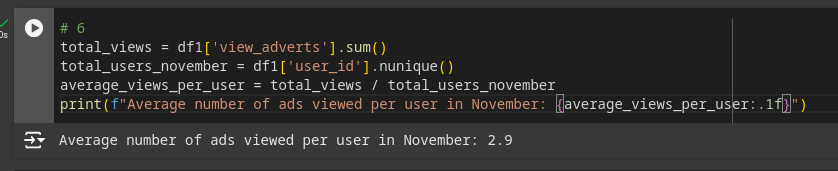
**46.3%**



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

4,96,25,3 **2,9**

**2.9**

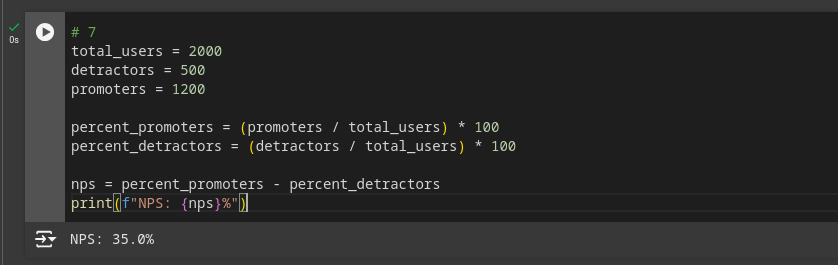


7. Мы провели опрос среди 2000 пользователей. Из них 500 «критики», 1200 «сторонники» и 300 «нейтралы». Посчитайте, чему будет равен NPS

\*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).

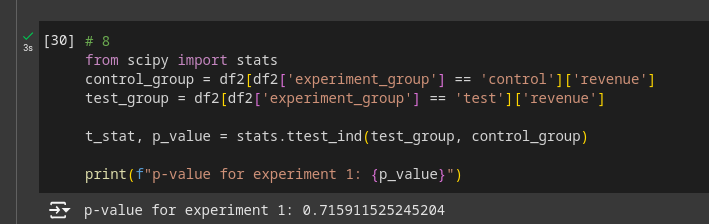
30%43%40% **35%**

**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:**



The **p-value for Experiment 1 is 0.7159**, which is **greater than 0.05**, meaning the difference between the test and control groups is **not statistically significant**.

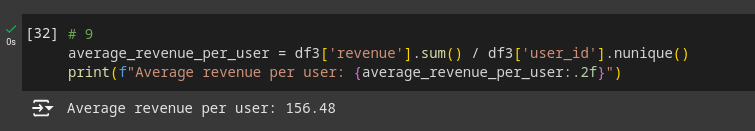
Recommendations:

* **Review changes** made in the test group; they might not have had enough impact.
* **Consider refining or testing other variations** to improve ARPU.
* **Increase the sample size** for more reliable results.
* **Run additional tests** to find a strategy that works better.

9. По датасету с листерами посчитайте средний доход на пользователя

121.2 **156.4** 70.930.7средняя здесь не применима

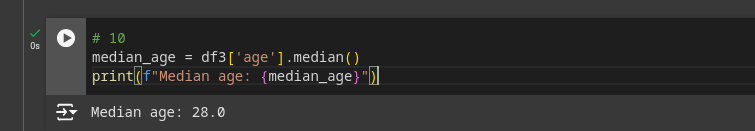
**156.4**



10. По датасету с листерами посчитайте медиану возраста пользователя

27,42 **28** 27,9327медиана здесь не применима

**28**



11. Какой график лучше всего подходит для отображения разброса цен на товары в разных магазинах?  
\*возможно несколько вариантов ответа

Линейный графикКруговая диаграмма

**Ящик с усами (box plot)** **Гистограмма**

For displaying the **price distribution of products across different stores**, the best options are:

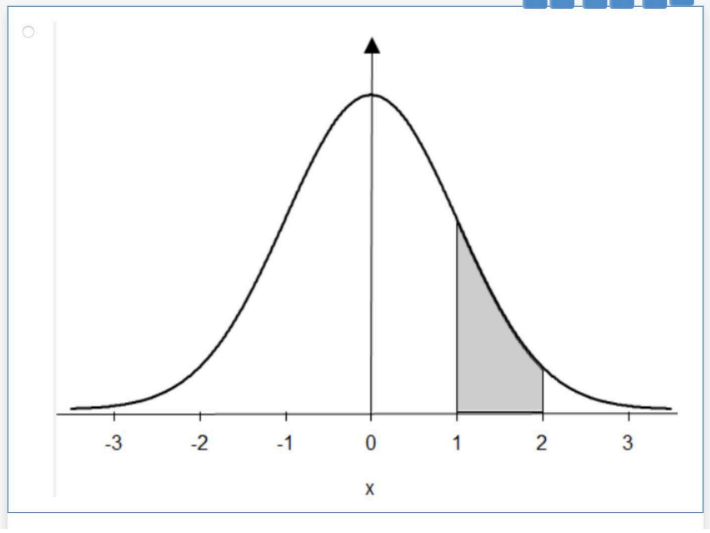
1. **Box plot** – This is ideal for visualizing the spread of data, including the median, quartiles, and any potential outliers. It effectively shows the distribution of prices and their variation across stores.
2. **Histogram** – Useful for displaying the frequency of prices within different ranges. It helps in understanding how common certain price ranges are for products.

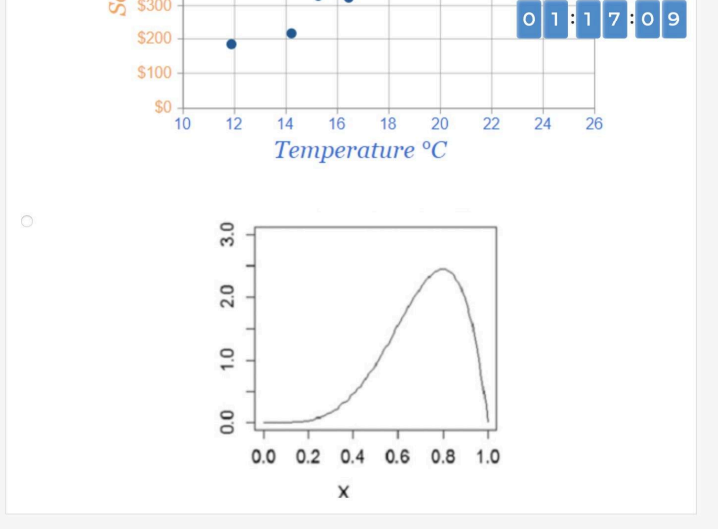
*Less suitable options:*

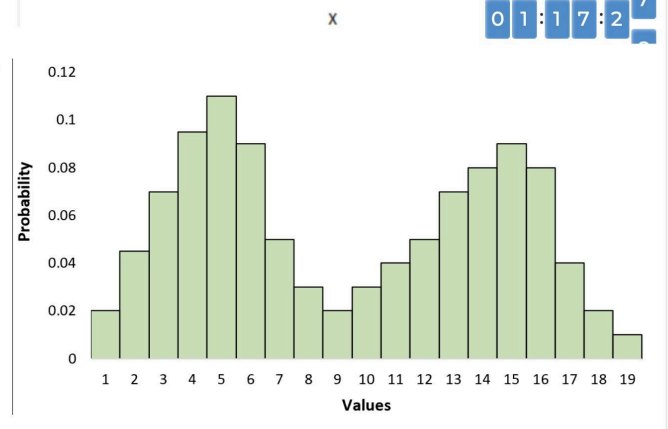
* **Line graph** – More suitable for showing changes over time rather than comparing prices across stores.
* **Pie chart** – Best for showing parts of a whole but not ideal for visualizing price distribution.

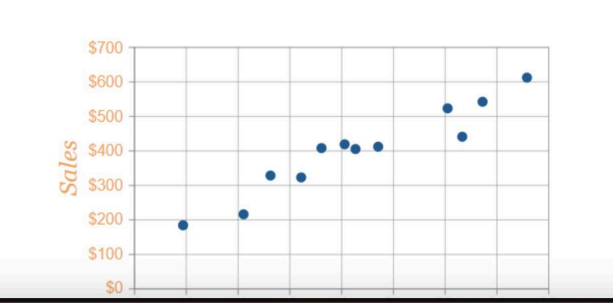
So, the **best options** are **box plot** and **histogram**.

12. На каком графике бимодальное распределение?

№1

№2

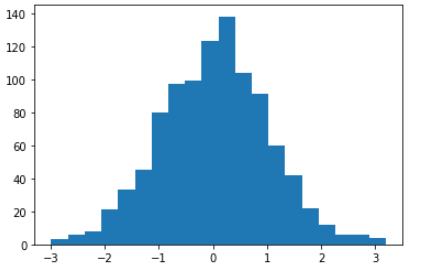
**№3**

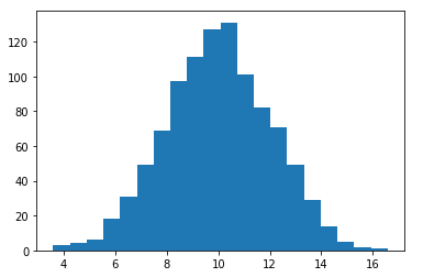
№4

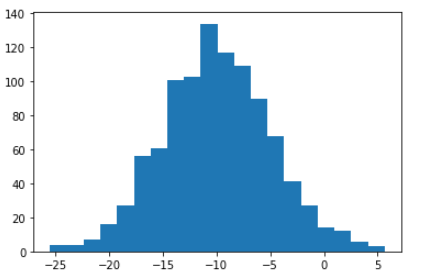
**On 3rd**

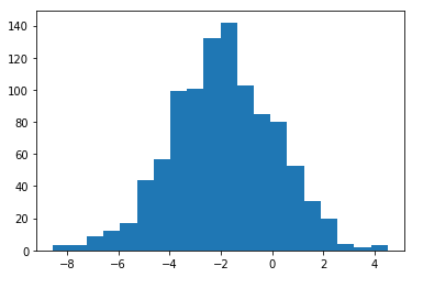
**The bimodal distribution is characterized by the presence of two distinct peaks in the data.**

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

№1

№2

№3

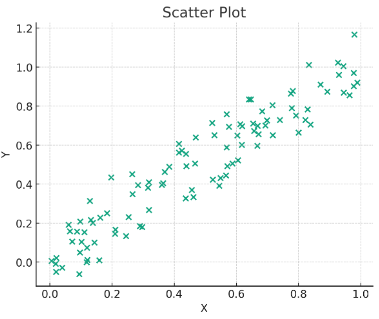
№4

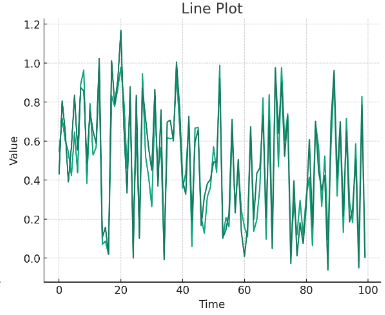
A graph with wider tails or a wider range of values has a greater variance.

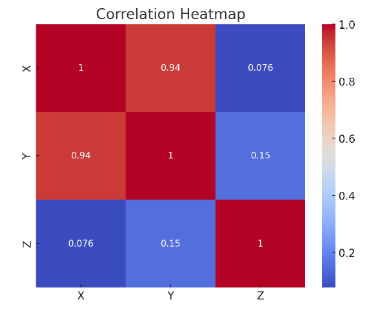
Distributions with smaller peaks and more "flat" densities indicate higher variance because the data is more distributed.

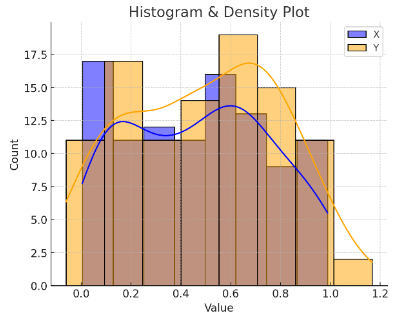
**3rd graph**

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









1. **Scatter plot** – This is the best graph for visualizing the relationship between two variables. It shows individual data points, allowing you to visually assess the strength, direction, and linearity of the correlation.
2. **Correlation heatmap** – A heatmap is great for showing the correlation between multiple variables at once. It provides a color-coded matrix where the strength of the correlation is represented by color intensity. This is ideal if you have several variables and want to see pairwise correlations.

Less suitable for calculating correlation:

* **Line plot** – This is useful for showing trends over time, but it doesn't specifically help in calculating or visualizing correlation between two variables.
* **Histogram** – This graph is for displaying the distribution of a single variable, so it doesn't help with correlation.
* **Density plot** – Similar to a histogram, a density plot shows the distribution of a single variable, not the relationship between two variables.

Best options for calculating correlation: **Scatter plot and Correlation heatmap.**

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%

**2nd**

P-value represents the probability of observing the data (or something more extreme) given that the null hypothesis is true.

If the p-value is less than the pre-set significance level (e.g., 0.05), it suggests that the results are statistically significant, and we can reject the null hypothesis.

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

**t-тест** Хи-квадрат тестАнализ дисперсии (ANOVA)Корреляция Пирсона

**1st**

t-тест используется для сравнения средних двух групп, если данные в этих группах подчиняются нормальному распределению. Этот тест позволяет проверить, есть ли статистически значимая разница между средними двух выборок.

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.

Квартили — это статистические величины, которые делят данные на четыре равные части:

Первый квартиль (Q1): 25% данных ниже этого значения.

Второй квартиль (Q2) или медиана: 50% данных ниже этого значения.

Третий квартиль (Q3): 75% данных ниже этого значения.

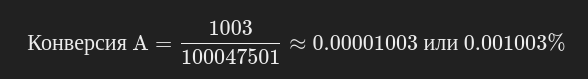
Четвертый квартиль — это все данные, которые больше третьего квартил

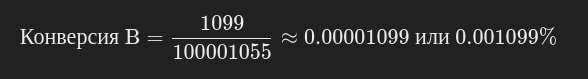
18. Были получены следующие результаты. Коллеги просят вас подтвердить их и сделать окончательный вывод по эксперименту.

* + - * 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 группе B больше людей, которые совершают конверсию, чем в группе A. Но разница небольшая. Это может быть важно для бизнеса, но разница такая маленькая, что нужно проверить, насколько она важна.

Если вероятность того, что разница есть, больше 0,05, то она не очень важна. Тогда не нужно ничего менять.

Если вероятность меньше 0,05, то разница важна. Тогда можно использовать вариант B.

Нужно проверить разницу с помощью статистики. Если она важна, то группа B лучше.