*\*Выбранные ответы мной выделены желтым хайлайтером*

**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



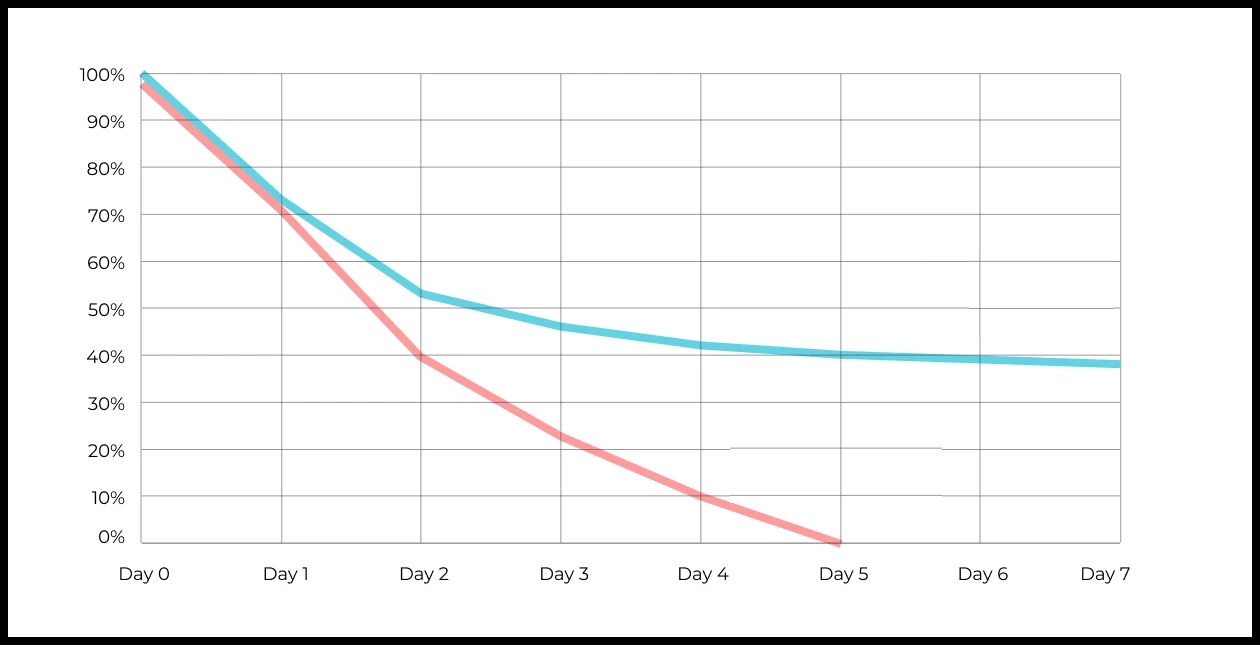
3. Используя вкладку "Данные об аудитории", посчитайте, чему будет равен retention первого дня у пользователей, пришедших в продукт 1 ноября

\*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:

Blue line

**Более высокая удержание:** Этот продукт демонстрирует значительно лучшее удержание пользователей на протяжении всего периода.

**Медленное снижение:** Кривая снижается более плавно. После резкого спада в первый день (примерно до 70%), скорость оттока замедляется. К концу первой недели (Day 7) у продукта остается около 40% пользователей, что является отличным показателем.

Red line

**Низкое удержание:** Этот продукт показывает очень слабое удержание пользователей.

**Быстрый отток:** Кривая резко падает. К концу первого дня (Day 1) продукт теряет около 30% пользователей, а к Day 5 удержание падает практически до нуля (около 1%).

Продукт, обозначенный синей кривой, гораздо более успешен в удержании своей аудитории, чем продукт с красной кривой. Продукт 2 требует срочного анализа причин массового оттока и внесения существенных изменений для повышения своей конкурентоспособности.

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:

| **Experiment** | **P-value** | **ARPU Control** | **ARPU Test** |
| --- | --- | --- | --- |
| **1** | **0,7159** | **722,46** | **665,74** |
| **2** | **0,8647** | **704,65** | **332,93** |
| **3** | **0,0617** | **663,21** | **998,67** |

Experiment 1

* Разница в ARPU между контрольной и тестовой группой небольшая (722 vs 666).
* P-value = 0,716 -> разница не статистически значима.
* Тестовая группа не показывает улучшения ARPU, можно считать, что влияние эксперимента отсутствует.

Experiment 2

* ARPU тестовой группы сильно ниже контрольной (333 vs 705).
* P-value = 0,865 → разница не статистически значима.
* На текущих данных тест не ухудшает ARPU статистически значимо. Тренд отрицательный, но нужны дополнительные данные, чтобы делать выводы.

Experiment 2

* ARPU тестовой группы выше контрольной (999 vs 663).
* P-value = 0,062 → близко к порогу 0,05, но не достигает статистической значимости.
* Вывод: Есть позитивный тренд, тестовая группа потенциально увеличивает ARPU, но пока нельзя утверждать это с уверенностью. Возможно, нужно увеличить выборку.

Рекомендации - у эксперимента 3 тренд положительный → можно повторить тест на более крупной выборке.

Также для всех экспериментов рекомендуется использовать медиану ARPU или непараметрические тесты (Mann–Whitney), так как выручка сильно скошена.

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

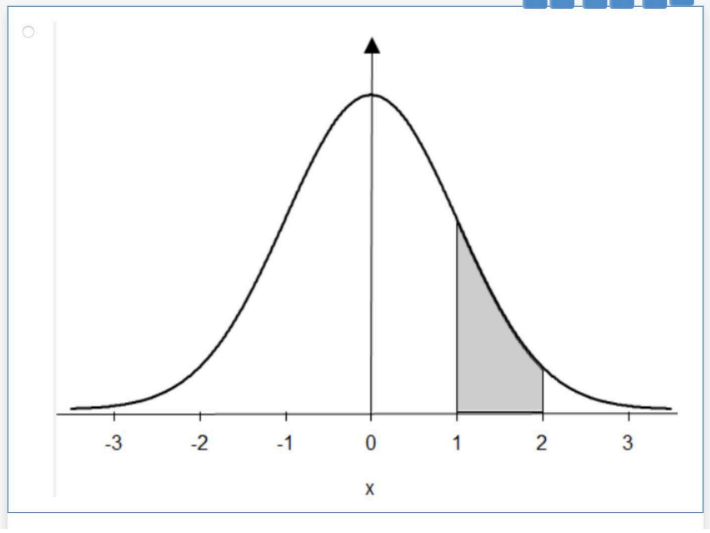
Линейный графикКруговая диаграммаЯщик с усами (box plot)



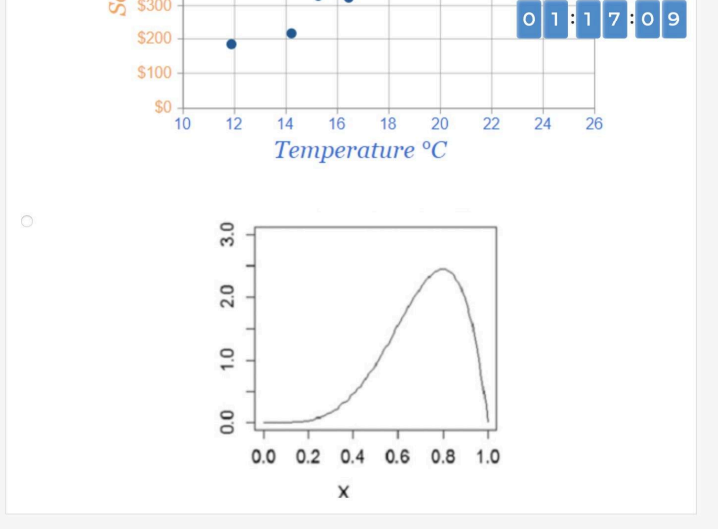
Гистограмма



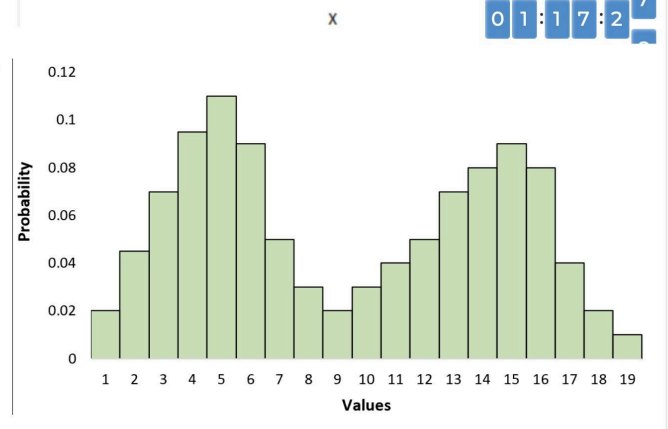
12. На каком графике бимодальное распределение?

№1

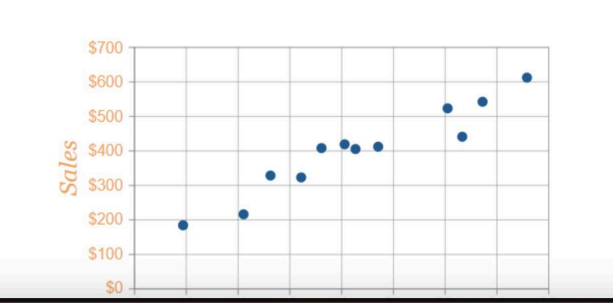


№2



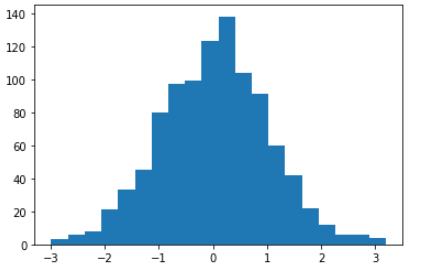
№3



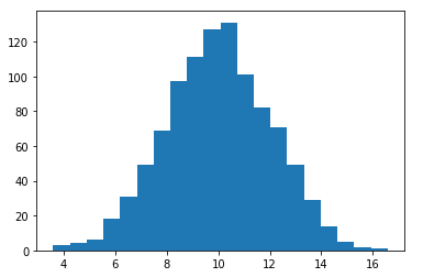
№4



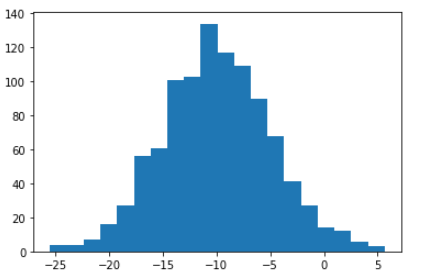
13. Какая случайная величина имеет наибольшую дисперсию данных по следующим графикам плотности распределения?

№1

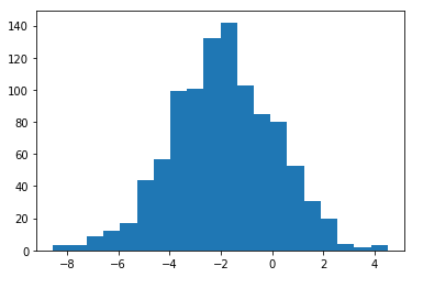


№2



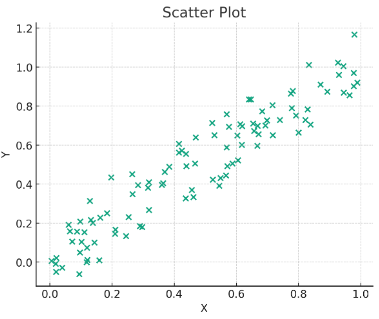
№3



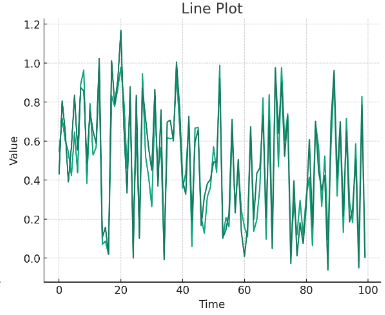
№4



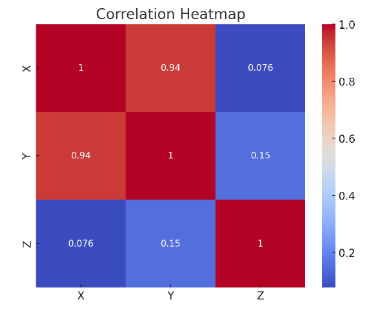
14. На каком графике можно посчитать коррелцияю?  
\*возможно несколько вариантов ответа



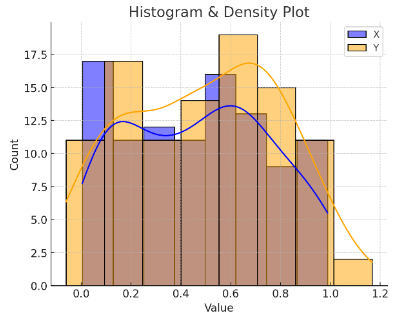














15. Что значит, если при проверке гипотез мы получили p-value = 0.05?

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:

Видно, что у тестовой группы чуть выше конверсия, но разница очень маленькая, и при таком огромном числе посетителей важно проверить статистическую значимость

Рекомендация - провести статистический тест (например, χ² или z-тест для пропорций), чтобы убедиться, что рост конверсии не является случайным шумом. Если значимости нет — оставить текущий вариант A. Если значимость подтверждается, можно рассматривать вариант B как улучшение, но эффект всё равно очень небольшой, так что стоит взвесить, оправданы ли затраты на внедрение.