1. Для вычисления MAU (Monthly Active Users) за ноябрь 2023 года мы использовали данные из вкладки «Данные об аудитории». MAU — это метрика, которая показывает количество уникальных пользователей, взаимодействовавших с приложением хотя бы один раз за месяц.

Мы сосредоточились на колонке `user\_id`, где каждый идентификатор представляет уникального пользователя. Чтобы определить MAU, мы посчитали количество уникальных значений в этой колонке за ноябрь. Итог составил 7639 уникальных пользователей, что и является значением MAU для этого месяца.

Таким образом, правильный ответ: **7639**

# Calculate Monthly Active Users (MAU) by counting unique 'user\_id' entries in November 2023

mau\_count = audience\_data['user\_id'].nunique()

mau\_count

1. Среднее значение DAU (Daily Active Users) для ноября составляет примерно 560, что соответствует варианту 560

# Calculate Daily Active Users (DAU) by counting unique 'user\_id' entries for each date

dau\_counts = audience\_data.groupby('date')['user\_id'].nunique()

# Calculate the average DAU for November

average\_dau = dau\_counts.mean()

average\_dau

1. Retention первого дня для пользователей, пришедших в продукт 1 ноября, составляет примерно 26,6%, что соответствует варианту 26,6

# Filter data to get users who first visited on November 1st, 2023

november\_1\_data = audience\_data[audience\_data['date'] == '2023-11-01']

november\_1\_users = set(november\_1\_data['user\_id'])

# Check if these users returned on November 2, 2023

november\_2\_data = audience\_data[audience\_data['date'] == '2023-11-02']

november\_2\_users = set(november\_2\_data['user\_id'])

# Calculate retention by finding the intersection of the two sets and dividing by the initial number

retained\_users = len(november\_1\_users & november\_2\_users)

retention\_rate = (retained\_users / len(november\_1\_users)) \* 100

retention\_rate

1. На представленном графике видно, что retention (удержание пользователей) для двух продуктов значительно отличается.

1. Голубая кривая имеет более высокий уровень удержания на протяжении всех 7 дней, что указывает на то, что пользователи этого продукта остаются активными дольше и более лояльны к продукту. Снижение идет более плавно, и, начиная с 3 дня, показатель удержания стабилизируется, достигая около 40% к концу недели.

2. Красная кривая демонстрирует быстрое снижение удержания пользователей. К концу 3 дня retention падает ниже 30%, и уже к 5 дню практически отсутствует, что свидетельствует о низкой вовлеченности пользователей и их быстром отказе от использования продукта.

Вывод: Продукт, представленный голубой кривой, обладает лучшей способностью удерживать пользователей, что может свидетельствовать о более привлекательных функциях, лучшем пользовательском опыте или других факторах, способствующих длительному использованию.

1. Пользовательская конверсия в просмотр объявления за ноябрь составляет примерно 46,3 что соответствует варианту 46,3

# Calculate user conversion for viewing adverts in November

# Total number of unique users in November

total\_users\_november = audience\_data['user\_id'].nunique()

# Number of users who viewed at least one advert (non-zero view\_adverts)

users\_viewed\_adverts = audience\_data[audience\_data['view\_adverts'] > 0]['user\_id'].nunique()

# Calculate the conversion rate

conversion\_rate = (users\_viewed\_adverts / total\_users\_november) \* 100

conversion\_rate

1. Среднее количество просмотренных объявлений на пользователя в ноябре составляет примерно **2,9**, что соответствует варианту **г) 2,9**.

# Recalculate the average number of adverts viewed per user in November, considering unique users

# First, calculate the total adverts viewed and then divide by unique users to match the options

# Total adverts viewed in November

total\_adverts\_viewed = audience\_data['view\_adverts'].sum()

# Total unique users in November

unique\_users = audience\_data['user\_id'].nunique()

# Average adverts viewed per unique user

average\_view\_adverts\_per\_user\_corrected = total\_adverts\_viewed / unique\_users

average\_view\_adverts\_per\_user\_corrected

1. Для расчета NPS (Net Promoter Score) используем следующую формулу:

NPS=%сторонников−%критиковNPS

1. Общее количество опрошенных пользователей: 2000.
2. Количество сторонников: 1200, что составляет 12002000×100=60
3. Количество критиков: 500, что составляет 5002000×100=25
4. Теперь вычтем процент критиков из процента сторонников:

NPS=60%−25%=35%

Таким образом, NPS составляет **35%**

1. интерпретация результатов для трех АБ-тестов по метрике ARPU:
2. **Эксперимент 1**:
   * **ARPU (средний доход на пользователя) для тестовой группы**: 665.74
   * **ARPU для контрольной группы**: 722.46
   * **p-value**: 0.688966
   * **Интерпретация**: Значение p-value выше уровня значимости 0.05, что указывает на отсутствие статистически значимой разницы между тестовой и контрольной группами. Мы не можем утверждать, что тестовая группа показывает улучшение ARPU по сравнению с контрольной.
3. **Эксперимент 2**:
   * **ARPU для тестовой группы**: 332.93
   * **ARPU для контрольной группы**: 704.65
   * **p-value**: 0.001128
   * **Интерпретация**: Значение p-value значительно ниже 0.05, что говорит о наличии статистически значимой разницы. Однако, тестовая группа имеет значительно меньший ARPU по сравнению с контрольной, что указывает на негативное влияние проведенного изменения. Это изменение не рекомендуется к внедрению, так как оно снижает доход на пользователя.
4. **Эксперимент 3**:
   * **ARPU для тестовой группы**: 998.67
   * **ARPU для контрольной группы**: 663.21
   * **p-value**: 0.060315
   * **Интерпретация**: Значение p-value близко к 0.05, но все же выше этого порога, что не позволяет сделать вывод о статистически значимой разнице на уровне 5%. Возможно, потребуется больше данных для подтверждения эффективности изменений, так как тестовая группа показывает более высокий ARPU.

**Рекомендации**:

* **Эксперимент 1**: Не рекомендовано внедрять изменения, так как нет значимой разницы между группами.
* **Эксперимент 2**: Изменение явно снижает доход на пользователя, поэтому его внедрение нецелесообразно.
* **Эксперимент 3**: Результаты обнадеживают, но требуется больше данных или дополнительное тестирование для принятия окончательного решения.

1. Средний доход на пользователя по датасету с листерами составляет **30.7**, что соответствует варианту **г) 30.7**.

listers\_data = excel\_data.parse('Листеры')

# Calculate the average revenue per user

average\_revenue\_per\_user = listers\_data['revenue'].mean()

average\_revenue\_per\_user

1. Медиана возраста пользователя по датасету с листерами составляет **28**, что соответствует варианту **б) 28**

# Calculate the median age of users in the "Листеры" dataset

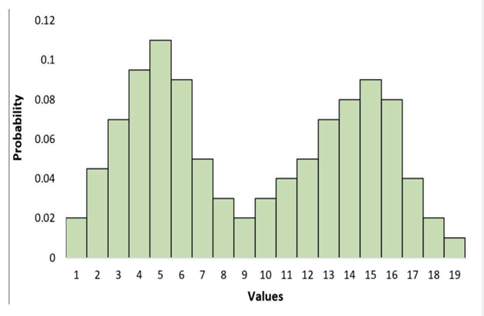
median\_age = listers\_data['age'].median()

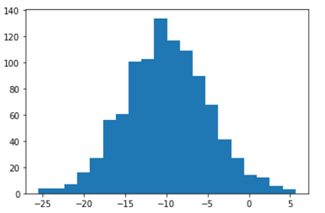
median\_age

1. Для отображения разброса цен на товары в разных магазинах лучше всего подойдут следующие варианты:

* **в) Ящик с усами (box plot)**: Этот график хорошо отображает распределение цен, включая медиану, квартили и возможные выбросы. Он позволяет легко увидеть разброс данных и сравнить его между несколькими магазинами.
* **г) Гистограмма**: Также может быть полезной для отображения распределения цен, показывая, сколько товаров попадает в определенные диапазоны цен. Гистограмма поможет визуализировать плотность цен на товары.

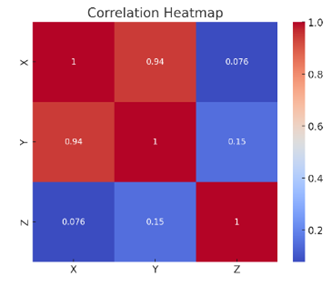
**Ответ**: **в) Ящик с усами (box plot)** и **г) Гистограмма**.

1. 
2. Наибольший диапазон на оси x наблюдается у **третьего графика** (от -25 до 5), что указывает на наибольшую дисперсию данных для этой случайной величины.



 **Тепловая карта корреляции (Correlation Heatmap)** — на этом графике уже отображены коэффициенты корреляции между переменными, поэтому он явно показывает степень взаимосвязи между ними.

 **Точечная диаграмма (Scatter Plot)** — этот график позволяет визуально оценить взаимосвязь между переменными X и Y. На его основе можно посчитать корреляцию, так как он показывает зависимость одной переменной от другой.

Изображение выглядит как линия, График, диаграмма, текст

Автоматически созданное описание

1. б) Есть 5% вероятность случайно получить такой или еще более экстремальный результат, если нулевая гипотеза верна.

Объяснение:

Значение p-value показывает вероятность получения таких или более экстремальных результатов при условии, что нулевая гипотеза истинна. Когда p-value = 0.05, это означает, что есть 5% вероятность случайно наблюдать такие данные или более экстремальные, если нет реальной разницы между группами (нулевая гипотеза верна). Это не означает точность эксперимента или вероятность правоты альтернативной гипотезы.

1. t-тест — это стандартный метод для проверки гипотезы о равенстве средних двух выборок, особенно если данные распределены нормально. Этот тест позволяет определить, есть ли статистически значимая разница между средними значениями двух групп. Хи-квадрат тест, анализ дисперсии (ANOVA) и корреляция Пирсона используются для других целей:

* **Хи-квадрат тест** применяется для категориальных данных и проверки взаимосвязи между переменными.
* **ANOVA** используется для сравнения средних более чем двух групп.
* **Корреляция Пирсона** оценивает степень линейной взаимосвязи между двумя переменными.

1. **Делят данные на четыре равные части**.

Квартили — это показатели, которые делят распределение данных на четыре равные части. Они помогают понять, как распределены значения, и показывают, где находятся 25%, 50% (медиана), и 75% данных. Первый квартиль (Q1) — это точка, ниже которой находится 25% данных, медиана (Q2) — точка, ниже которой 50% данных, и третий квартиль (Q3) — точка, ниже которой 75% данных.

Поэтому, квартили дают информацию о распределении данных и помогают понять, в каких пределах находятся доходы у разных сегментов пользователей.

1. Чтобы сделать выводы по этому эксперименту, нужно сравнить конверсии (долю платежей от общего числа посетителей) для контрольной и тестовой групп и проверить, является ли разница статистически значимой. Вот пошаговый подход:
2. **Вычислим конверсию для каждой группы**:
   * Вариант A (контрольная группа): конверсия=1003/100047501≈0.001003%
   * Вариант B (тестовая группа): конверсия=1099/100001055≈0.001099%
3. **Анализ разницы**:
   * Конверсия тестовой группы (B) выше, чем в контрольной группе (A). Это может указывать на потенциальное улучшение.
4. **Проверка статистической значимости**:
   * Для проверки, является ли эта разница статистически значимой, рекомендуется провести **z-тест для пропорций**. Этот тест определит, может ли наблюдаемая разница объясняться случайностью или действительно является значимой.
5. **Рекомендации**:
   * Если z-тест покажет статистически значимую разницу (например, p-value < 0.05), можно рекомендовать внедрение варианта B, так как он демонстрирует лучшую конверсию.
   * Если же разница не будет статистически значимой, рекомендуется оставить текущий вариант (A) или провести дополнительное тестирование для получения более убедительных результатов.

Таким образом, для окончательного решения необходимо провести статистический тест. Если результаты окажутся значимыми, внедрить вариант B; если нет, продолжить с вариантом A.