### 1) Объяснение кода

1. **Загрузка данных**: Данные считываются из файла ab\_stats.csv.
2. **Расчет ARPPU**: Вычисляется ARPPU для каждой группы.
3. **Проверка нормальности**: Используется тест Шапиро-Уилка для проверки нормальности.
4. **Тестирование**: В зависимости от нормальности данных выполняется t-тест или U-тест.
5. **Вывод результатов**: Печатаются p-value, информация о значимости и рекомендация по версии.

**Рекомендации менеджеру**

* Если результат **статистически значим** и группа A имеет более высокий ARPPU, стоит рассмотреть внедрение изменений из группы A.
* Если различия **не значимы**, возможно, стоит проанализировать дополнительные параметры или протестировать другие изменения.

2) Для проведения А/Б-теста по трем источникам трафика потребуется примерно 258 086 пользователей. С учетом вашего месячного трафика в 40 000 пользователей, на тестирование гипотезы потребуется около 581 дня.

**Рекомендации:**

1. **Тест займет очень много времени** — 581 день — это больше полутора лет. Чтобы ускорить процесс, можно:
   * Увеличить трафик, направляемый на тест.
   * Рассмотреть снижение уровня доверия или мощности, если это допустимо для вашей задачи.
   * Рассмотреть возможность проведения параллельных тестов для других источников трафика или гипотез.
2. **Улучшение точности** — можете пересмотреть целевой прирост конверсии. Если ожидаемый прирост будет больше, время на тестирование сократится.

### 3). Интерпретация результатов:

* **Z-статистика** и **p-значение** показывают, насколько сильно конверсии отличаются от среднего значения.
* Если **p-значение** для одного из источников меньше 0.05 (или другого уровня значимости), то различия считаются статистически значимыми.

**Рекомендации для менеджера:**

1. Если различия значимы, то имеет смысл изучить, какой источник показал наилучший результат и выяснить причины его эффективности.
2. Если различия незначимы, это может означать, что источники работают примерно одинаково и стоит рассмотреть другие параметры для оптимизации.

### 4) Выводы по результатам теста

1. **Нормальность распределения**:
   * **Результат**: Отвергаем нулевую гипотезу о нормальности распределения.
   * Это означает, что данные CPA в обеих группах не имеют нормального распределения.
2. **Рекомендация по выбору статистического критерия**:
   * **Рекомендованный тест**: Лучше использовать тест Манна-Уитни (U-тест), так как данные не нормально распределены.

**Обоснование выбора теста**

Тест Манна-Уитни является непараметрическим и позволяет сравнивать два независимых выборки, что делает его подходящим в условиях ненормального распределения.

### 5) Результаты А/Б-теста

1. **Z-статистика**: z≈11.24z \approx 11.24z≈11.24
2. **p-value**: p≈2.69×10−29p \approx 2.69 \times 10^{-29}p≈2.69×10−29
3. **Статистическая значимость**: Результат является статистически значимым на уровне доверия 80% (alpha = 0.20).

**Рекомендация**

* **Версия для продакшн**: Рекомендуется выкатить версию **A**, так как её среднее время, проведенное пользователями, (360) значительно выше, чем у версии B (352).

Таким образом, результаты теста показывают, что версия A имеет значительно лучшее среднее время, что делает её предпочтительной для внедрения

### 6) ехническая архитектура проекта

#### 1. Управленческий процесс по договоренностям с внешними партнерами

* **Шаг 1**: Определение ключевых метрик успеха (например, конверсия в подписку, прирост пользователей).
* **Шаг 2**: Переговоры с банком о 99% кэшбэке на первый месяц.
* **Шаг 3**: Заключение соглашения.
* **Шаг 4**: Реализация и тестирование кэшбэка.
* **Шаг 5**: Оценка влияния на конверсию.

#### 2. Архитектура данных

* **Системы для скачивания данных**:
  + **Система управления пользователями**: данные о новых подписках, отписках.
  + **Система аналитики**: данные о поведении пользователей на платформе.
  + **CRM-система**: информация о взаимодействии с пользователями.
  + **Система платежей**: данные о транзакциях и кэшбэке.
* **Хранилище данных**: Data Warehouse для агрегирования данных из разных источников.

#### 3. Внутрикомандное взаимодействие

* **Команда разработки**: отвечает за интеграцию и реализацию кэшбэка.
* **Команда аналитиков**: проводит A/B-тестирование и анализирует результаты.
* **Маркетинг**: информирует пользователей о предложении кэшбэка.
* **Команда поддержки**: отвечает на вопросы пользователей.