**Финальный отчёт по A/B-тесту**

**Раздел 1. Цель проекта**

**Задача**

Основная задача - определить, принесла ли тестовая группа, участвовавшая в акции, больше прибыли по сравнению с контрольной. Для этого мы рассчитаем и сравним:

* **ARPU (Average Revenue Per User)**: Среднюю прибыль на одного игрока.
* **ARPPU (Average Revenue Per Paying User)**: Среднюю прибыль на одного платящего игрока.
* **Среднюю трату внутриигровой валюты (cash)**: Оценим влияние акции на внутреннюю экономику игры.

**Подход к решению**

Мы будем использовать язык **Python** и его библиотеки для проведения всего анализа. Наш подход включает:

1. **Загрузку и очистку данных**: Объединим все файлы в один датафрейм и удалим дубликаты.
2. **Анализ данных**: Выявим известных и потенциальных читеров.
3. **Расчёт метрик**: Рассчитаем ключевые метрики и их доверительные интервалы.
4. **Визуализацию**: Построим графики для наглядного представления результатов.
5. **Формирование выводов**: Сделаем заключение на основе полученных данных, чтобы дать рекомендацию по будущему проведению акции.

**Раздел 2. Анализ источников**

**Выбор инструмента: Python**

Для решения задачи был выбран **Python**, так как он предоставляет полный набор инструментов для всех этапов анализа.

**Загрузка данных**

Начнём с загрузки всех файлов и проверки их на наличие дубликатов, что является критически важным шагом для обеспечения достоверности данных.

import pandas as pd

import numpy as np

**# Загрузка данных**

df\_abgroup = pd.read\_csv('ABgroup.csv')

df\_cash = pd.read\_csv('Cash.csv')

df\_cheaters = pd.read\_csv('Cheaters.csv')

df\_money = pd.read\_csv('Money.csv')

df\_platforms=pd.read\_csv('Platforms.csv')

**# Удаление дубликатов**

df\_abgroup.drop\_duplicates(inplace=True)

df\_cash.drop\_duplicates(inplace=True)

df\_cheaters.drop\_duplicates(inplace=True)

df\_money.drop\_duplicates(inplace=True)

df\_platforms.drop\_duplicates(inplace=True)

|  |  |  |
| --- | --- | --- |
| **Файл** | **Количество строк до удаления дубликатов** | **Количество строк после удаления дубликатов** |
| ABgroup | 8640000 | 1080000 |
| Cash | 8640000 | 8640000 |
| Cheaters | 8640000 | 1080000 |
| Money | 8640000 | 8640000 |
| Platforms | 8640000 | 1080000 |

**Раздел 3. Очистка данных**

**Описание процесса очистки**

Перед началом анализа необходимо вычистить данные от читеров, которые могут исказить результаты. Процесс очистки состоит из двух этапов:

1. **Удаление известных читеров**: Мы используем файл Cheaters.csv, чтобы найти и удалить всех, кто был помечен как читер.
2. **Поиск скрытых читеров**: На основе **анализа известных читеров**, мы знаем, что они играют на **PC**, не совершают платежей реальными деньгами (money == 0), но при этом тратят аномально большое количество внутриигровой валюты (cash). Для выявления таких аномалий мы используем **Z-score > 3**, что является стандартным методом для обнаружения выбросов.  
   **Вывод:**   
   Найдено известных читеров: 353

Найдено потенциальных читеров: 343

Количество пользователей до очистки: 1080000

Количество пользователей после очистки: 1079304

**Раздел 4. Использование статистических методов**

**Доверительный интервал** - это диапазон значений, в котором с 95% вероятностью находится истинное среднее значение. Если доверительные интервалы для двух групп **не пересекаются**, это означает, что разница между их средними значениями **статистически значима**.  
  
**# Расчет ARPU (Average Revenue Per User) по money**

arpu = df\_clean.groupby('group')['money'].mean()

**# Расчет ARPPU (Average Revenue Per Paying User) по money**

paying\_users = df\_clean[df\_clean['money'] > 0]

arppu = paying\_users.groupby('group')['money'].mean()

**# Расчет средней траты внутриигровой валюты (cash)**

avg\_cash\_spent = df\_clean.groupby('group')['cash'].mean()

**# Расчет ARPU по группам и платформам**

arpu\_by\_platform = df\_clean.groupby(['group', 'platform'])['money'].mean()

**# Функция для расчета доверительного интервала**

def get\_ci(data, confidence=0.95):

mean = data.mean()

se = st.sem(data)

h = se \* st.t.ppf((1 + confidence) / 2, len(data) - 1)

return mean, mean - h, mean + h

ci\_control = get\_ci(df\_clean[df\_clean['group'] == 'control']['money'])

ci\_test = get\_ci(df\_clean[df\_clean['group'] == 'test']['money'])

**Раздел 5. Формирование отчёта**

**Сравнение метрик**

Ниже представлены ключевые метрики, рассчитанные для обеих групп.

* **ARPU (money)**:

Контрольная группа: 5.829467

Тестовая группа: 6.162276

* **ARPPU (money)**:

Контрольная группа: 5.831109

Тестовая группа: 6.163144

* **Средняя трата cash**:

Контрольная группа: 5800.73

Тестовая группа 6229.60

**95% доверительных интервалов ARPU:**

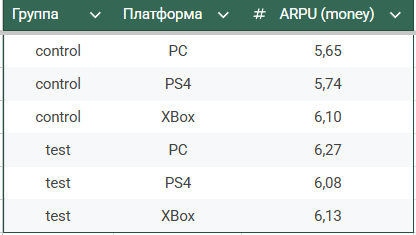
Контрольная группа: [5.82, 5.83]

Тестовая группа: [6.16, 6.17]

**Вывод:**

Тестовая группа показала лучшие результаты по всем метрикам.

**ARPU по платформам**



**Вывод**: Мы видим, что на платформе **PC** тестовая группа значительно превзошла контрольную. На PS4 и Xbox разница менее выражена.

Изображение выглядит как текст, рукописный текст, линия, Шрифт

Контент, сгенерированный ИИ, может содержать ошибки.

**Вывод:**

**ARPU (money) по дням**: График показывает, что ARPU тестовой группы стабильно выше, чем у контрольной, на протяжении всего периода.

**Средняя трата внутриигровой валюты (cash)**: Этот график также показывает, что в тестовой группе игроки тратили больше внутриигровой валюты, что может говорить о возросшей активности.

**Финальное заключение**

Проведенный анализ показывает, что акция оказала положительное влияние на финансовые показатели.

* **ARPU и ARPPU**: В тестовой группе средняя прибыль на пользователя (и на платящего пользователя) выше, чем в контрольной.
* **Доверительные интервалы**: 95% доверительные интервалы для ARPU не пересекаются ([5.82, 5.83] для контрольной группы и [6.16, 6.17] для тестовой), что говорит о **статистически значимом** результате.
* **Траты внутриигровой валюты**: Игроки из тестовой группы также тратили больше внутриигровой валюты, что может свидетельствовать о возросшем интересе и вовлеченности в игру.
* **Платформы**: Наибольший прирост прибыли наблюдается на платформе **PC**.

**Рекомендация**: На основе полученных данных, акция со скидкой на премиумную броню оказалась **успешной** и принесла **статистически значимый** прирост прибыли. Рекомендуется повторить акцию в дальнейшем, с особым акцентом на платформу **PC**.