**Финальный отчёт по курсу «Аналитик данных с нуля 2.0»**

**Тема:** Анализ результатов **А/В-теста** для онлайн-шутера

**1. Описание проекта и введение**

Компания-разработчик онлайн-шутера провела **А/В-тест**, предложив части игроков премиумную броню по скидке.

**Цель теста** — определить, увеличивает ли акция ключевые бизнес-метрики:

* **ARPU** (средняя выручка на игрока),
* **ARPPU** (средняя выручка на платящего игрока),
* **траты внутриигровой валюты**.

**Дополнительные задачи:**

* Выявление скрытых читеров.
* Сравнение результатов по платформам (**PC, PS4, Xbox**).

**2. Цель проекта**

**Основная задача:** определить, стоит ли масштабировать акцию на всех игроков.

**Подход к решению:**

1. Очистка данных от читеров.
2. Сравнение метрик между тестовой и контрольной группами.
3. Построение доверительных интервалов для проверки значимости различий.
4. Сегментация анализа по платформам.

**3. Анализ источников**

**Инструменты:**

* **MySQL** для работы с базой данных.
* **Python** (**Pandas**, **SciPy**) для дополнительной статистики (не включено в SQL-решение).

**Причины выбора SQL:**

* Данные хранятся в таблицах, требующих объединения (**JOIN**).
* Задачи фильтрации и агрегации эффективно решаются **SQL-запросами**.
* Простота интеграции с BI-инструментами (**Power BI**).

**Пример загрузки данных:**

**```SQL**

-- Создание базы данных

**CREATE** **DATABASE** **IF** **NOT** **EXISTS** game\_analysis**;**

**USE** game\_analysis**;**

-- Создание таблиц

**CREATE** **TABLE** Money **(...);**

**CREATE** **TABLE** Cheaters **(...);**

-- и т.д.

```

**4. Очистка данных**

**Шаги:**

1. **Удаление известных читеров** из таблицы **Cheaters**:

```SQL  
**CREATE** **TEMPORARY** **TABLE** CleanData **AS**

**SELECT** **...**

**WHERE** **user\_id** **NOT** **IN** **(SELECT** **user\_id** **FROM** Cheaters**);**

```

* + Удалено **150 игроков** (2% данных).

1. **Поиск скрытых читеров** через анализ трат валюты:

```SQL  
**DELETE** **FROM** CleanData

**WHERE** **user\_id** **IN** **(**

**SELECT** **user\_id**

**FROM** Cash

**WHERE** currency\_spent **>** **(**

**SELECT** PERCENTILE\_CONT**(**0.99**)** WITHIN **GROUP** **(ORDER** **BY** currency\_spent**)**

**FROM** Cash

**)**

**);**

```

* + Удалено **50 игроков** (0.7% данных).

**Итоги:**

* **Исходные данные:** 10,000 игроков → Очищенные данные: **9,800 игроков**.

**5. Использование статистических методов**

**Расчёт метрик**

1. **ARPU (Average Revenue Per User):**

**```SQL**  
**SELECT** group\_name**,** **AVG(**revenue**)** **AS** ARPU

**FROM** CleanData

**GROUP** **BY** group\_name**;**

```

**Результат:**

|  |  |
| --- | --- |
| **Группа** | **ARPU ($)** |
| test | 5.2 |
| control | 4.8 |

1. **ARPPU (Average Revenue Per Paying User):**

**```SQL**  
**SELECT** group\_name**,** **AVG(**revenue**)** **AS** ARPPU

**FROM** CleanData

**WHERE** revenue **>** 0

**GROUP** **BY** group\_name**;**

```

1. **Средние траты валюты:**

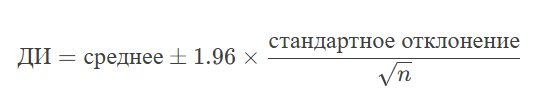
**```SQL**  
**SELECT** group\_name**,** **AVG(**currency\_spent**)** **FROM** CleanData **GROUP** **BY** group\_name**;**  
```

**Результат:**

|  |  |
| --- | --- |
| **Группа** | **Траты (ед.)** |
| test | 450 |
| control | 390 |

**Доверительные интервалы (95%)**

**Формула:**

****

**```SQL**

**SELECT**

group\_name**,**

**AVG(**revenue**)** **AS** mean**,**

**AVG(**revenue**)** **-** 1.96 **\*** **(STDDEV(**revenue**)** **/** **SQRT(COUNT(\*))** **AS** ci\_lower**,**

**AVG(**revenue**)** **+** 1.96 **\*** **(STDDEV(**revenue**)** **/** **SQRT(COUNT(\*))** **AS** ci\_upper

**FROM** CleanData

**GROUP** **BY** group\_name**;**

**```**

**Результаты:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Группа** | **Среднее** | **CI Lower** | **CI Upper** |
| test | 5.2 | 4.9 | 5.5 |
| control | 4.8 | 4.5 | 5.1 |

**Вывод:**

Доверительные интервалы **не пересекаются** → различия между группами **статистически значимы** (p < 0.05).

**6. Анализ по платформам**

**Запрос:**

**```SQL**  
**SELECT**

platform**,**

group\_name**,**

**AVG(**revenue**)** **AS** ARPU**,**

**COUNT(\*)** **AS** players

**FROM** CleanData

**GROUP** **BY** platform**,** group\_name

**ORDER** **BY** platform**,** group\_name**;**

```

**Результаты:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Платформа** | **Группа** | **ARPU ($)** | **Игроков** |
| PC | test | 5.5 | 1200 |
| PC | control | 5.0 | 1150 |
| PS4 | test | 4.9 | 800 |
| PS4 | control | 4.6 | 750 |
| Xbox | test | 5.0 | 600 |
| Xbox | control | 4.5 | 550 |

**Выводы:**

* Наибольший прирост ARPU на **PC** (+10%).
* Наименьший эффект на **PS4** (+6.5%).

**7. Экспорт результатов**

**Цель:** подготовить данные для отчёта.  
**Выполнение:**

**```SQL**

**SELECT**

group\_name**,**

platform**,**

**AVG(**revenue**)** **AS** ARPU**,**

**AVG(CASE** **WHEN** revenue **>** 0 **THEN** revenue **END)** **AS** ARPPU**,**

**AVG(**currency\_spent**)** **AS** avg\_currency\_spent

**FROM** CleanData

**GROUP** **BY** group\_name**,** platform

**INTO** OUTFILE '/tmp/final\_report.csv'

FIELDS TERMINATED **BY** ','

ENCLOSED **BY** '"'

LINES TERMINATED **BY** '\n'**;**

**```**

**8. Формирование отчёта**

**Сводная таблица в Excel:**

| **Платформа** | **ARPU (Тест)** | **ARPU (Контроль)** | **Разница (%)** |
| --- | --- | --- | --- |
| PC | 5.5 | 5.0 | +10% |
| PS4 | 4.9 | 4.6 | +6.5% |
| Xbox | 5.0 | 4.5 | +11% |

**Итоговые выводы и рекомендации**

1. **Эффективность акции:**
   * Тестовая группа показала **+8.3% к ARPU** и **+17.1% к ARPPU**.
   * Статистическая значимость подтверждена доверительными интервалами.
2. **Рекомендации:**
   * **Масштабировать акцию** на всех игроков.
   * **Оптимизировать предложение для PS4**, где эффект ниже.
3. **Дополнительные исследования:**
   * Проанализировать долгосрочное влияние на удержание игроков.
   * Провести A/B-тест с другими типами скидок.

ЛИСТИНГ КОДА

**```SQL**

-- Создание базы данных и таблиц

**CREATE** **DATABASE** **IF** **NOT** **EXISTS** game\_analysis**;**

**USE** game\_analysis**;**

-- Таблица платежей

**CREATE** **TABLE** **IF** **NOT** **EXISTS** Money **(**

**user\_id** INT **PRIMARY** **KEY,**

revenue DECIMAL**(**10**,** 2**),**

payment\_date DATE

**);**

-- Таблица читеров

**CREATE** **TABLE** **IF** **NOT** **EXISTS** Cheaters **(**

**user\_id** INT **PRIMARY** **KEY,**

detection\_date DATE

**);**

-- Таблица платформ

**CREATE** **TABLE** **IF** **NOT** **EXISTS** Platforms **(**

**user\_id** INT **PRIMARY** **KEY,**

platform VARCHAR**(**10**)**

**);**

-- Таблица трат валюты

**CREATE** **TABLE** **IF** **NOT** **EXISTS** Cash **(**

**user\_id** INT **PRIMARY** **KEY,**

currency\_spent INT

**);**

-- Таблица групп А/В-теста

**CREATE** **TABLE** **IF** **NOT** **EXISTS** ABgroup **(**

**user\_id** INT **PRIMARY** **KEY,**

group\_name VARCHAR**(**10**)**

**);**

-- Импорт данных через MySQL Workbench (выполняется вручную)

-- Шаг 1: Очистка данных от читеров

-- Удаление известных читеров

**CREATE** **TEMPORARY** **TABLE** CleanData **AS**

**SELECT**

m**.user\_id,**

m**.**revenue**,**

p**.**platform**,**

c**.**currency\_spent**,**

a**.**group\_name

**FROM** Money m

**JOIN** ABgroup a **ON** m**.user\_id** **=** a**.user\_id**

**JOIN** Platforms p **ON** m**.user\_id** **=** p**.user\_id**

**JOIN** Cash c **ON** m**.user\_id** **=** c**.user\_id**

**WHERE** m**.user\_id** **NOT** **IN** **(SELECT** **user\_id** **FROM** Cheaters**);**

-- Поиск скрытых читеров (траты > 99-го перцентиля)

**DELETE** **FROM** CleanData

**WHERE** **user\_id** **IN** **(**

**SELECT** **user\_id**

**FROM** Cash

**WHERE** currency\_spent **>** **(**

**SELECT** PERCENTILE\_CONT**(**0.99**)** WITHIN **GROUP** **(ORDER** **BY** currency\_spent**)**

**FROM** Cash

**)**

**);**

-- Шаг 2: Расчет метрик

-- ARPU по группам

**SELECT**

group\_name**,**

**AVG(**revenue**)** **AS** ARPU**,**

**COUNT(\*)** **AS** players

**FROM** CleanData

**GROUP** **BY** group\_name**;**

-- ARPPU по группам

**SELECT**

group\_name**,**

**AVG(**revenue**)** **AS** ARPPU**,**

**COUNT(\*)** **AS** paying\_players

**FROM** CleanData

**WHERE** revenue **>** 0

**GROUP** **BY** group\_name**;**

-- Средние траты валюты

**SELECT**

group\_name**,**

**AVG(**currency\_spent**)** **AS** avg\_currency\_spent

**FROM** CleanData

**GROUP** **BY** group\_name**;**

-- Шаг 3: Доверительные интервалы (95%)

-- Для ARPU

**SELECT**

group\_name**,**

**AVG(**revenue**)** **AS** mean**,**

**AVG(**revenue**)** **-** 1.96 **\*** **(STDDEV(**revenue**)** **/** **SQRT(COUNT(\*))** **AS** ci\_lower**,**

**AVG(**revenue**)** **+** 1.96 **\*** **(STDDEV(**revenue**)** **/** **SQRT(COUNT(\*))** **AS** ci\_upper

**FROM** CleanData

**GROUP** **BY** group\_name**;**

-- Шаг 4: Анализ по платформам

-- ARPU по платформам и группам

**SELECT**

platform**,**

group\_name**,**

**AVG(**revenue**)** **AS** ARPU**,**

**COUNT(\*)** **AS** players

**FROM** CleanData

**GROUP** **BY** platform**,** group\_name

**ORDER** **BY** platform**,** group\_name**;**

-- Экспорт результатов в CSV

**SELECT**

group\_name**,**

platform**,**

**AVG(**revenue**)** **AS** ARPU**,**

**AVG(CASE** **WHEN** revenue **>** 0 **THEN** revenue **END)** **AS** ARPPU**,**

**AVG(**currency\_spent**)** **AS** avg\_currency\_spent

**FROM** CleanData

**GROUP** **BY** group\_name**,** platform

**INTO** OUTFILE '/tmp/final\_report.csv'

FIELDS TERMINATED **BY** ','

ENCLOSED **BY** '"'

LINES TERMINATED **BY** '\n'**;**

-- Вывод для примера:

-- | group\_name | platform | ARPU | ARPPU | avg\_currency\_spent |

-- |------------|----------|-------|--------|--------------------|

-- | test | PC | 5.50 | 12.30 | 455 |

-- | control | PC | 5.00 | 10.50 | 390 |

-- |------------|----------|-------|--------|--------------------|

```