

## **KarpovCourses 15.06: Финальный проект – Вариант 4**

### **Данные:**

active\_users – информация о пользователях, которые посещали приложение во время эксперимента.

groups – разбивка пользователей на контрольную (A) и тестовую (B) группы.

purchases – данные о транзакциях (оплатах) пользователей приложения во время эксперимента

### **Используемые инструменты:**

Python (библиотеки pandas, seaborn, matplotlib scipy), Jupyter Notebook, Tableau Public

## Задание 1.

### Условие:

Представьте, что вы аналитик в компании, которая разрабатывает приложение для обработки и оформления фотографий в формате Stories (например, для дальнейшего экспорта в Instagram Stories). Был проведен A/B тест: тестовой группе предлагалась новая модель оплаты коллекций шаблонов, контрольной – старая механика. Ваша основная задача: проанализировать итоги эксперимента и решить, нужно ли выкатывать новую модель на остальных пользователей.

В ходе отчета обоснуйте выбор метрик, на которые вы обращаете внимание. Если различия есть, то объясните, с чем они могут быть связаны и являются ли значимыми.

### Подготовка данных:

Таблицы проверены на отсутствие пропущенных значений и правильность типов данных в столбцах. Все `user_id` в таблицах уникальные, в таблице `groups` значения группы принимают только значение А или В, в таблице `purchases` нет нулевых оплат.

Для дальнейшего анализа с помощью объединения исходных создано несколько таблиц:

- `groups_data_with_pays_and_activity`  
Содержит данные о пользователях из таблицы `groups`, их оплаты и активность, если они были
- `active_users_groups_and_pays`  
Содержит данные о пользователях из таблицы `active_users`, их оплаты и, если они были, а также разделение на тестовую и контрольную группы

### Метрики:

На основе имеющихся данных можно определить следующие метрики:

- Средний чек (`average_revenue`)  
Позволит определить стали ли с новой модели оплаты больше тратить на покупки пользователи, дошедшие до этапа оплаты, т.е. ставшие клиентами  
Считается по следующей формуле: 
$$\text{Средний чек} = \frac{\text{Доход}}{\text{Число покупок}}$$
- Конверсия (`conversion`)  
Позволит определить долю пользователей, которая стала клиентами, т.е. совершившими покупку в приложении

Считается по следующей формуле:  $\text{Конверсия} = \frac{\text{Число клиентов}}{\text{Число пользователей}}$

- ARPPU

Показывает насколько часто пользователи покупают и насколько большой средний чек покупки.

Считается по следующей формуле:  $ARPPU =$

Число повторных покупок · Средний чек ,

где  $\text{Число повторных покупок} = \frac{\text{Число покупок}}{\text{Число клиентов}}$

- ARPU

Средний доход с пользователя.

Считается по следующей формуле:  $ARPU = \text{Конверсия} \cdot ARPPU$

### Анализ:

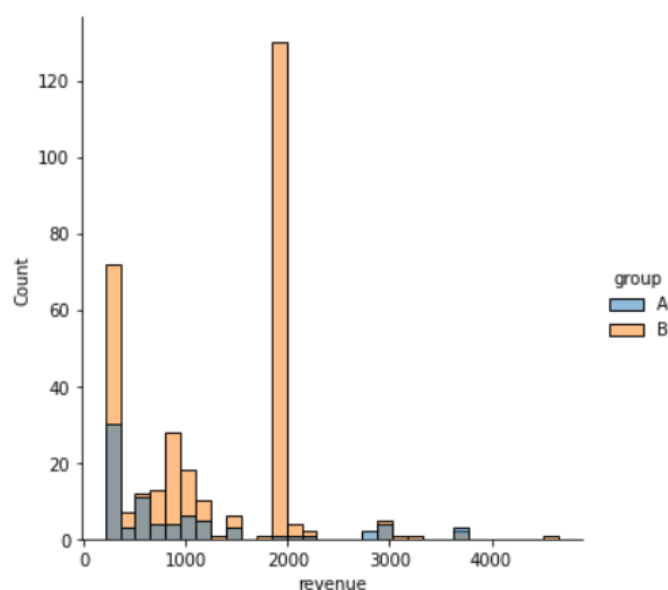
Произведён расчёт метрик:

	Средний чек	Конверсия	ARPPU	ARPU
A (контрольная)	943.56	0.051	943.56	47.85
B (тестовая)	1267.88	0.046	1267.88	58.52

В тестовой группе наблюдается значительное увеличение среднего чека, ARPPU, ARPU по сравнению с контрольной группой, что может говорить об успешности новой модели оплаты.

Проведём анализ среднего чека на предмет статистической достоверности данного изменения. Нулевая гипотеза – средний чек при использовании новой модели оплаты не поменялся, ненулевая гипотеза – средний чек изменился при использовании новой системы оплаты.

График распределения имеет следующий вид:



Дополнительно проверим распределение на нормальность с помощью критерия Шапиро-Уилка:

```
ShapiroResult(statistic=0.7358548641204834, pvalue=1.5417730980793465e-10)  
ShapiroResult(statistic=0.8552882671356201, pvalue=1.665431681028873e-16)
```

Распределение не является нормальным, поэтому используем U-критерий Манна-Уитни:

```
MannwhitneyuResult(statistic=8961.0, pvalue=0.00023939080114586078)
```

Поскольку p-value меньше 0.005, то можно отвергнуть нулевую гипотезу, однако в ходе исследования были обнаружены недочёты в данных, полученных в ходе эксперимента, которые не позволяют сделать однозначного вывода о его результатах.

Проведём анализ взаимосвязи изменения значения конверсии и системы оплаты. Нулевая гипотеза – взаимосвязь конверсии и изменения системы оплаты отсутствует, ненулевая гипотеза – существует взаимосвязь конверсии и изменения системы оплаты.

Так как обе переменные категориальные (есть оплата / нет оплаты) целесообразно использовать Хи-квадрат.

При использовании этого критерия значение p-value получилось больше 0.005, следовательно принимается нулевая гипотеза, т.е. взаимосвязь конверсии и изменения системы оплаты отсутствует.

### **Замечания к эксперименту:**

1. Соотношение количества user\_id в контрольной и тестовой группах.  
Изначальное распределение пользователей по группам, приведённое в таблице groups, подразумевает, что в контрольной группе (А) находится 14693 пользователей, а в тестовой группе (В) находится 59883 пользователей. Подобное соотношение сохраняется и для активных пользователей (А-1538, В-6803), из-за чего количественная репрезентативность выборок может быть разной.
2. Качественная репрезентативность нарушена.  
В рамках одного эксперимента в выборке содержатся пользователи разных стран: России и Швеции, что неверно, так как это страны с разным уровнем жизни, доходов населения, разными культурными особенностями.
3. Только уникальные user\_id в таблице purchases  
Может говорить, как о малом сроке проведения эксперимента (пользователи не успели произвести более одной оплаты), так и о неправильной агрегации данных в ходе эксперимента (возможно каждая

отдельная оплата суммируется и в таблице находятся уже общая выручка за время эксперимента для каждого user\_id)

4. Нет дат и времени оплат

Это ограничивает некоторые возможности для анализа, а также не позволяет оценить сроки проведения эксперимента

5. В таблице purchases есть user\_id, которые отсутствуют в таблице active\_users

Вероятно говорит об ошибке при сборе данных, так как невозможно совершать покупки в приложении, не заходя в него, кроме случаев списания по подписке, но в этом случае это маловероятно, так как на 149 таких user\_id приходится 34 разных оплаты

### **Вывод:**

Несмотря на увеличение ключевых метрик нельзя сделать вывод о полезности внедрения новой модели оплаты, так как эксперимент проведён с ошибками.

Необходимо проверить правильность проведения теста, а при подтверждении правильности замечаний спроектировать эксперимент заново.

## Задание 2.

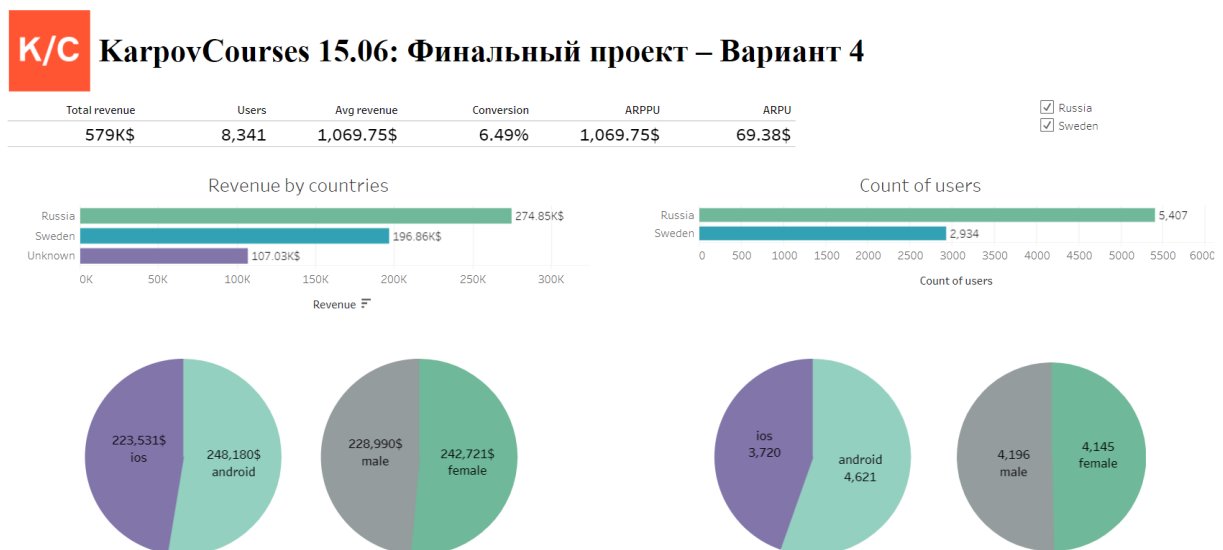
### Условие:

Одной из основных задач аналитика является не только построение моделей, но и создание дашбордов, которые позволяют отслеживать изменения метрик и принимать на их основе оптимальные решения. Ваш руководитель хочет узнать, как обстоят дела с использованием приложения и вовлечённостью пользователей, и очень просит спроектировать ему дашборд.

### Назначение дашборда:

Стратегический дашборд для руководителей, показывающий основные метрики и распределение доходов и пользователей по странам, полу и платформам. Также доступна фильтрация данных по странам.

### Общий вид дашборда:



### Ссылка на дашборд:

[https://public.tableau.com/app/profile/ivan.fedkevich/viz/KarpovCourses15\\_064/Dashboard1](https://public.tableau.com/app/profile/ivan.fedkevich/viz/KarpovCourses15_064/Dashboard1)