

# Анализ эффективности маркетинговой кампании на основе А/В-теста и сегментации клиентов

АВТОР: АНДРЕЙ ДИНКЕЛЬ.

# Описание проблемы

- ▶ Известно, что магазин проводил две маркетинговые кампании:

Первая кампания проводилась в период с 5-го по 16-й день, ID участвовавших в ней пользователей содержатся в файле `ids_first_company_positive.txt`. Эта кампания включала в себя предоставление персональной скидки 5 000 клиентам через email-рассылку.

Вторая кампания проводилась на жителях города 1 134 и представляла собой баннерную рекламу на билбордах: скидка всем каждое 15-е число месяца (15-й и 45-й день в нашем случае).

- ▶ Была запущена первая кампания для которой были выделены тестовая (`positive`) и контрольная (`negative`) группы пользователей.
- ▶ Задача: с помощью данных о покупках клиентов и их социально-демографических признаках проанализировать эффективность уже проведённых ранее маркетинговых кампаний и выявить факторы, способные повысить продажи.

# Постановка гипотезы

- ▶ **Гипотезы:**

- ▶  $H_0$ : После внедрения кампании конверсия и стоимость покупок остались неизменными.

- ▶  $H_1$ : После внедрения кампании конверсия и стоимость покупок возросли.

- ▶ **Дополнительная задача:** Провести сегментацию пользователей для выявления наиболее отзывчивых групп.

# Этапы работы

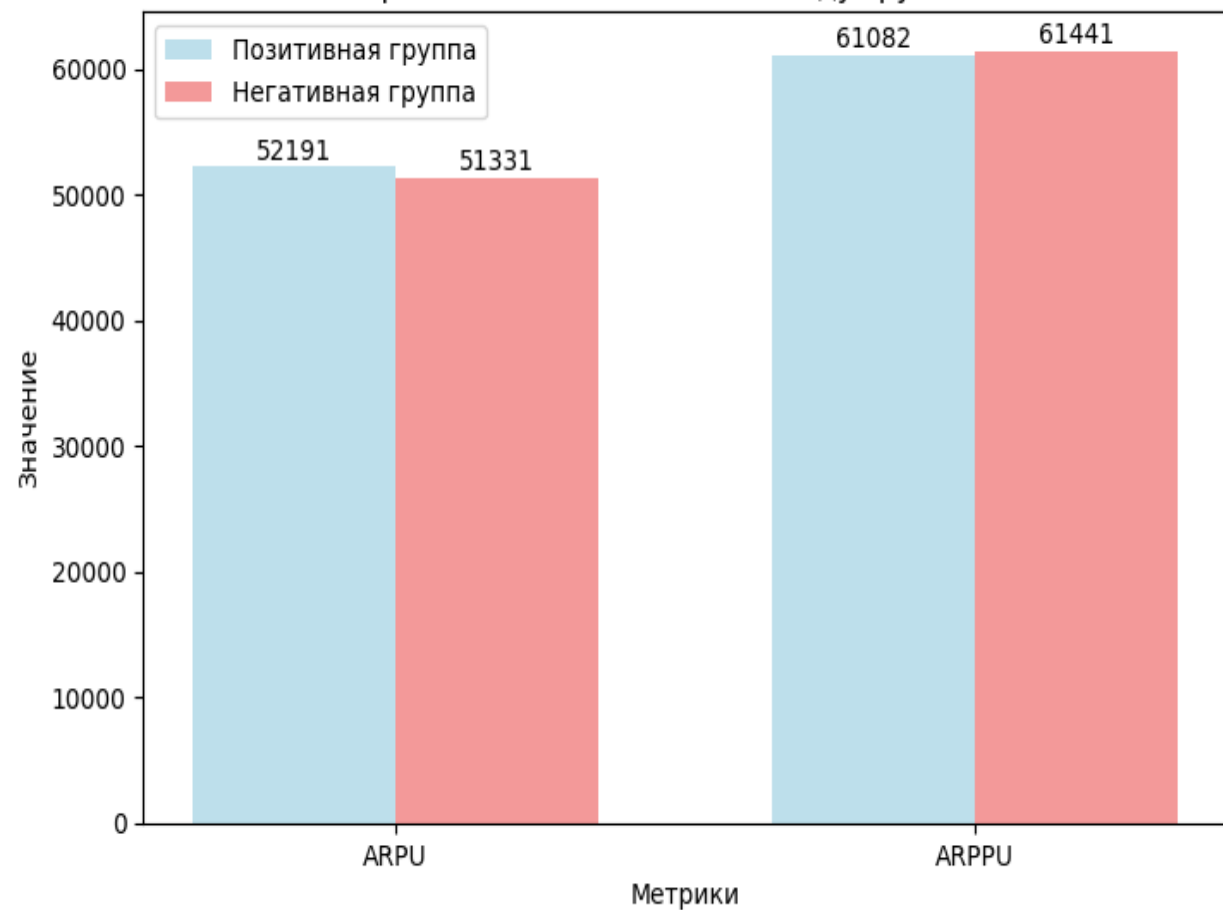
- ▶ Подготовка данных (объединение таблиц, очистка, восстановление пропусков).
- ▶ A/B-тест первой кампании (сравнение групп).
- ▶ Кластеризация клиентов (сегментация по поведению и соц.-демографии).
- ▶ Построение модели склонности клиента к покупке (ML).
- ▶ Прогноз для города 1188.

# A/B тест

- ▶ **Метод:** сравнение конверсий и стоимости покупок группы А (получили скидку) и группы В (контроль).
- ▶ Результаты:
  - ▶ Конверсия и стоимость покупок выше в группе А.
  - ▶ Разница статистически значима ( $p\text{-value} < 0.05$ ).
  - ▶  $Lift > 0$ . Кампания принесла результат.
- ▶ **Вывод:** Email-скидки работают, стоит внедрять кампанию. Не отвергаем  $H_0$ .

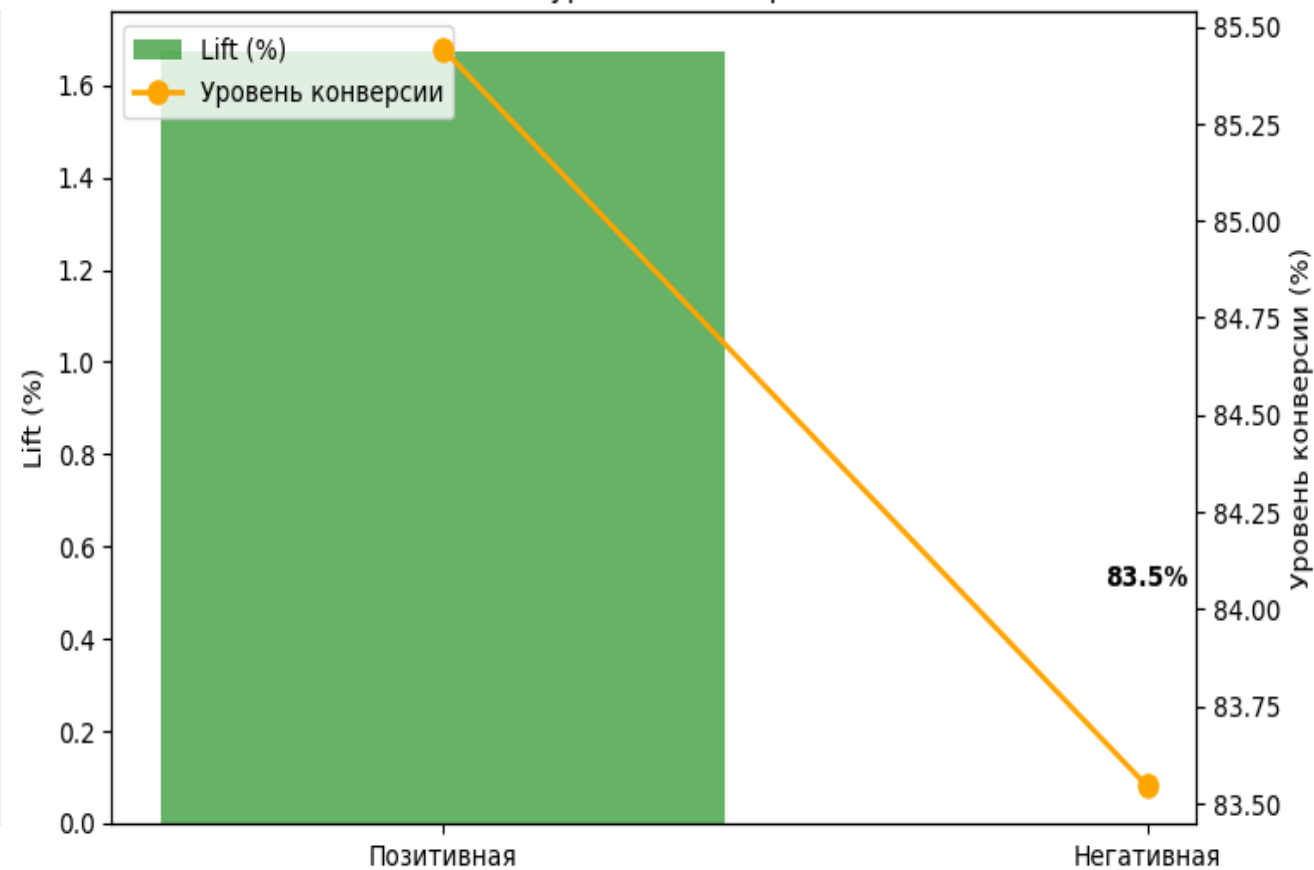
# Графики метрик.

Сравнение ARPU и ARPPU между группами



1.7%  
85.4%

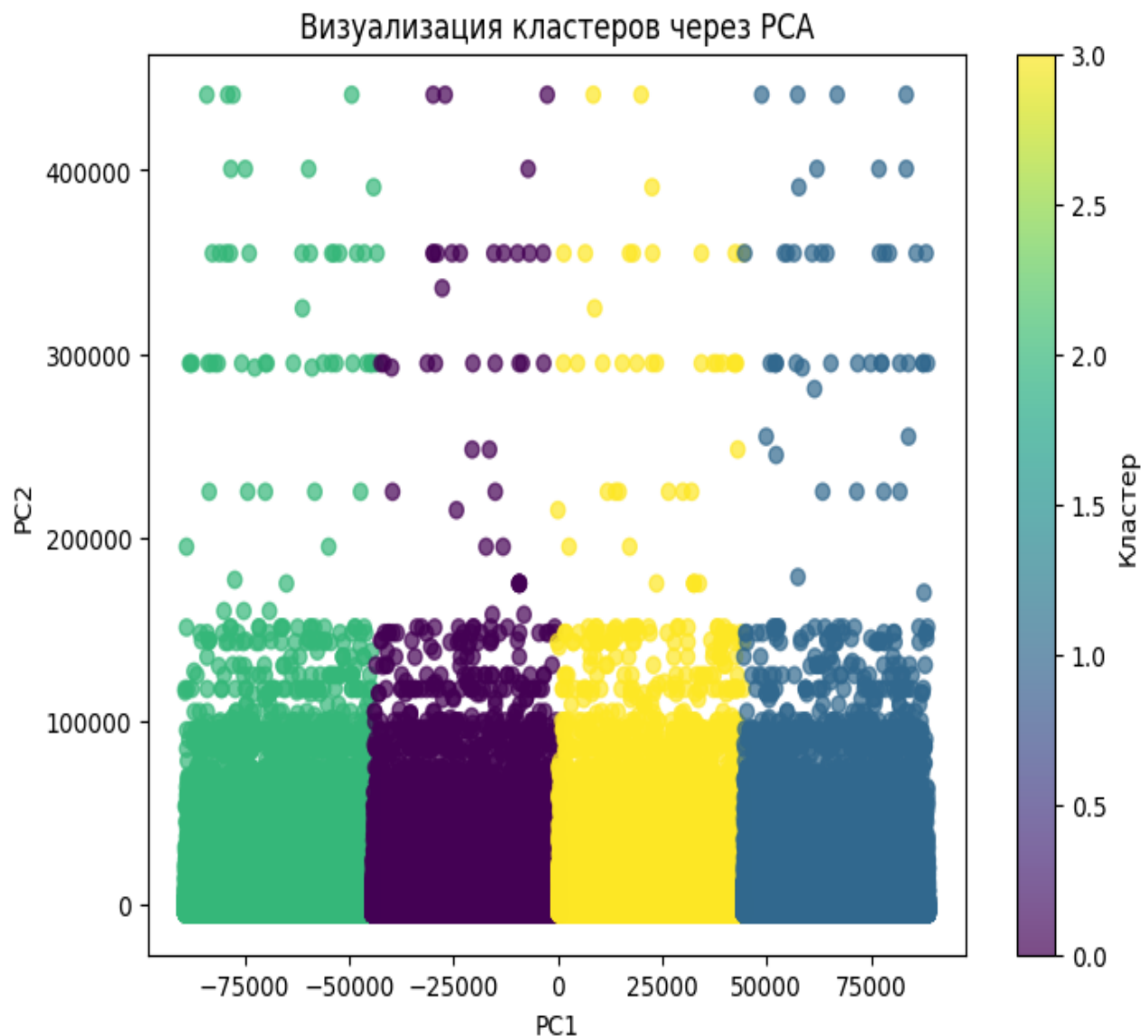
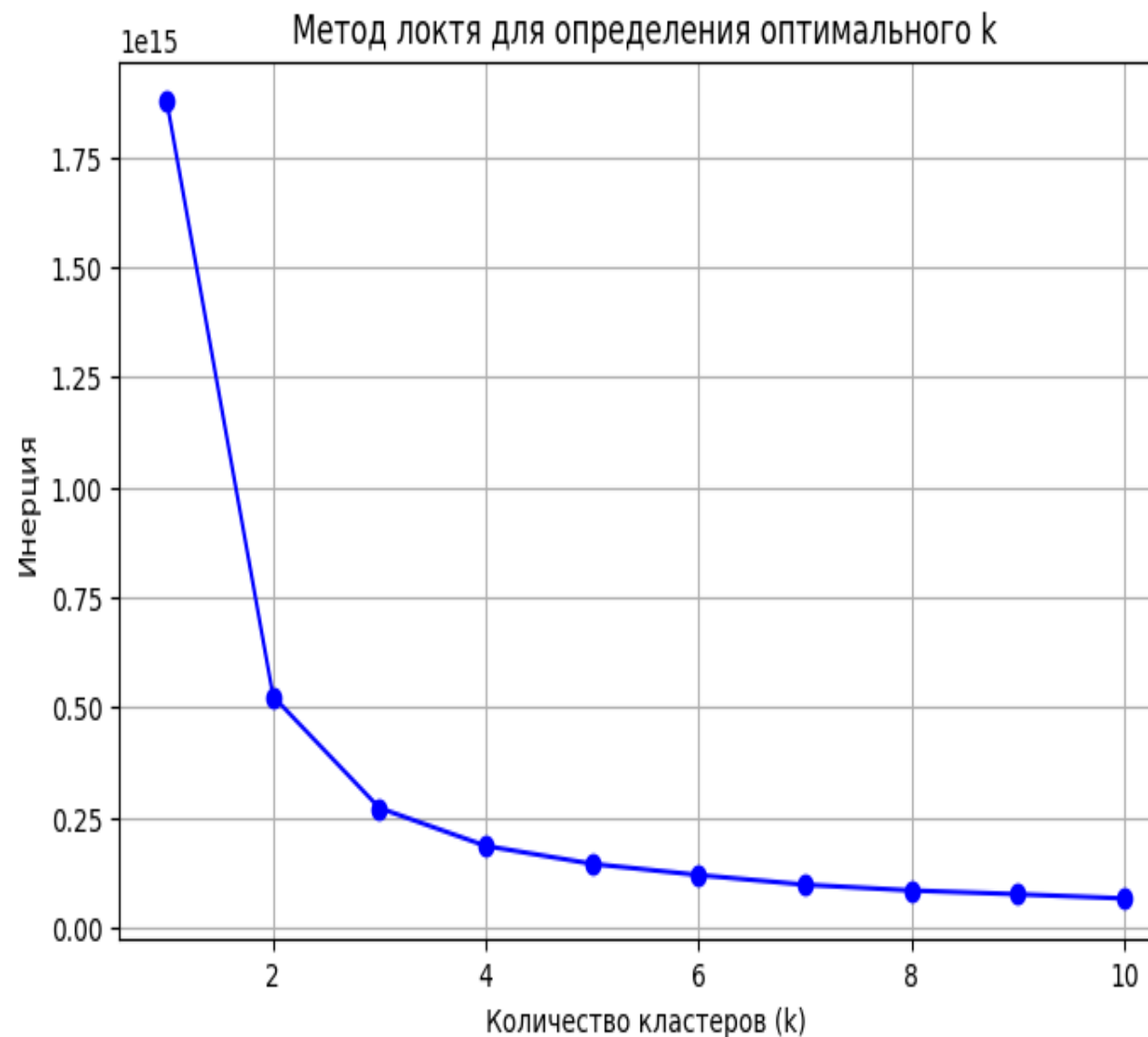
Lift и уровень конверсии



# Кластеризация клиентов

- ▶ Метод: KMeans (на выборке).
- ▶ Получено несколько кластеров.
- ▶ В кластере 0 наблюдается наибольшая средняя стоимость покупок (~6508). Группа на 56 процентов состоит из мужчин, средний возраст примерно равен 41,5. Основные продукты -- транспорт и мебель.
- ▶ Вторая по стоимости покупок группа -- кластер 1. Группа состоит на 52 процента из мужчин и средний возраст примерно равен 41,5. Основные продукты -- обувь и одежда.
- ▶ Третья по стоимости покупок группа -- кластер 2. Группа состоит на 39 процентов из мужчин и средний возраст примерно равен 41,9. Основные продукты -- аккумулятор холода, защитные спреи и мебель.
- ▶ Четвертая по стоимости покупок группа -- кластер 3. Группа состоит на 40 процентов из мужчин и средний возраст примерно равен 42,3. Основной продукт -- транспорт.

# Графики кластеризации





# Методы работы с кластерами

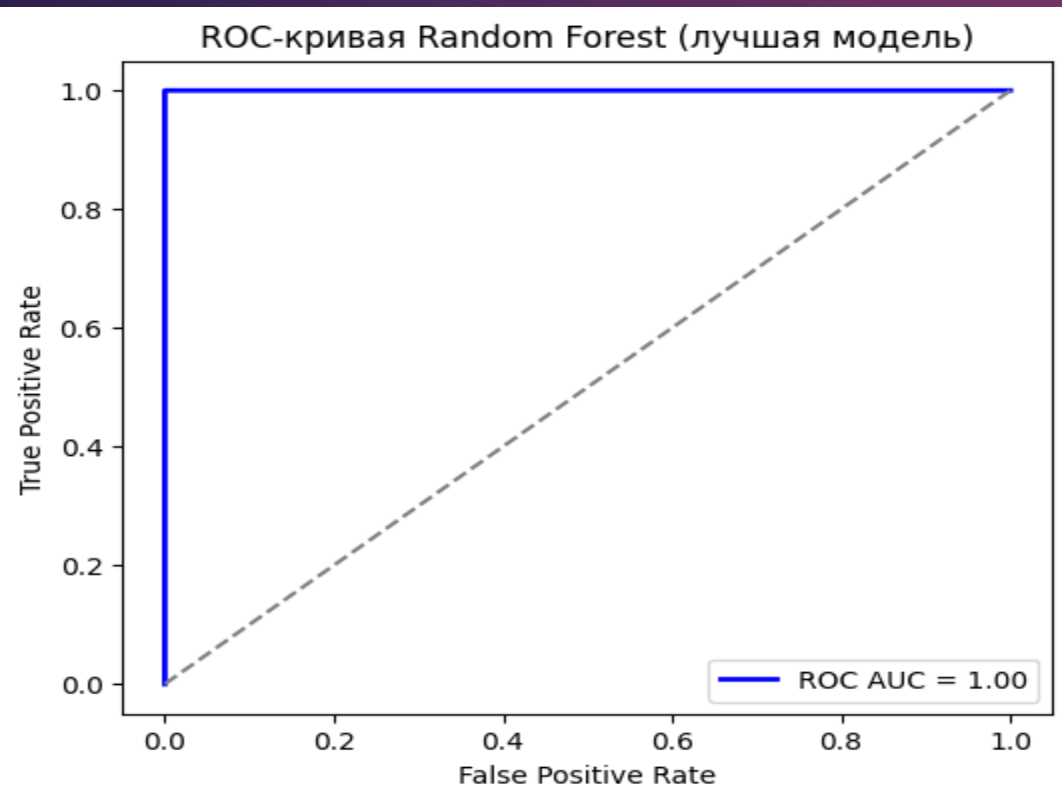
- ▶ Кластер 0: Увеличить видимость товаров через рекламу, тестировать минимальные скидки (5-10%), чтобы стимулировать интерес, нацелить на семейные наборы (например, стол + стулья), предложить бонусы (доставка) при покупке с детьми.
- ▶ Кластер 1: Предлагать премиальные товары с акцентом на качество, использовать скидки 15-20% для крупных покупок.
- ▶ Кластер 2: Нацелить на семейные наборы (например, стол + стулья), предложить бонусы (доставка) при покупке с детьми.
- ▶ Кластер 3: Акцентировать спортивные акции, предлагать скидки на аксессуары (экипировка) для повышения среднего чека.

# Влияние скидок

- ▶ Скидки значительно увеличивают вероятность покупки.
- ▶ Сильнее всего реагируют клиенты из нулевого кластера (Группа на 56 процентов состоит из мужчин, средний возраст примерно равен 41,5. Основные продукты -- транспорт и мебель. ).
- ▶ В сегменте дорогих товаров скидки менее важны (главное — бренд).

# Модель склонности

- ▶ Использовали **Random Forest**, т.к. модель показала наилучший результат на кросс-валидации.
- ▶ Метрики: Accuracy, confusion matrix, F1, ROC-AUC.
- ▶ Возможность таргетинга → выделяем клиентов с  $p > 0.5$ .
- ▶ **Вывод:** модель успешно предсказывает покупателей → можно использовать для кампаний.



# Значения метрик

- ▶ Accuracy = 0,9999
- ▶ Confusion matrix =  $\begin{bmatrix} 0 & 1 \\ 0 & 204124 \end{bmatrix}$
- ▶ F1 score = 0,9999

# Прогноз для города 1188

- ▶ Большинство клиентов имеют высокую вероятность покупки.
- ▶ Основные интересы: обувь, одежда, фитнес.
- ▶ Вывод статистики из Python:



# ВЫВОДЫ.

- ▶ Данные показывают, что наша аудитория и так уже имеет максимальную склонность к покупке. Вместо того чтобы тратить ресурсы на ее предсказание, мы рекомендуем сосредоточиться на персонализации предложений на основе кластерного анализа и оптимизации маркетинговых механик для дальнейшего роста среднего чека и лояльности.
- ▶ По основной гипотезе: Маркетинговая кампания показала **положительный бизнес-эффект** в виде роста конверсии на 1.9 п.п. и лифта ARPU на 1.68%. По сегментации: С помощью кластеризации были выявлены три четких сегмента пользователей. **Кластер 0** (мужчины ~41 года с высокой средней стоимостью покупок) является наиболее перспективным для таргетирования в будущих кампаниях. T-test не выявил статистические различия между группами.
- ▶ **Рекомендация:** Следует постепенно внедрять кампанию.