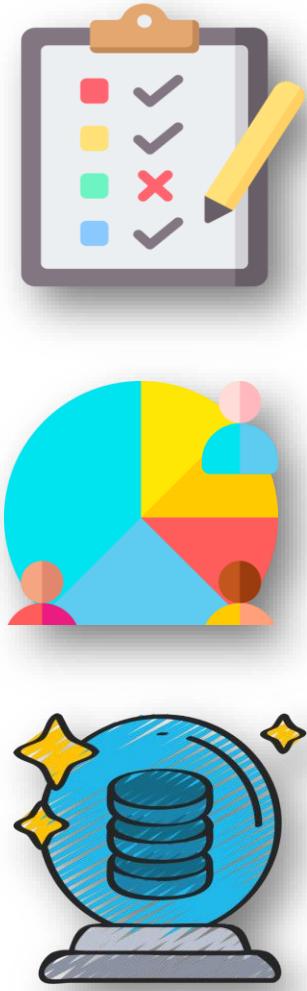




Анализ маркетинговой кампании, кластеризация клиентов и прогнозирование склонности к покупке обуви.

Афанаскина Вера, Data Analyst, Москва, 2025

Цель проекта:



Оценить эффективность маркетинговой кампании через АВ тестирование, определить влияние скидок на ключевые метрики

Сегментировать клиентскую базу для персонализации маркетинговых стратегий

Разработать модель для прогнозирования склонности клиентов к покупке обуви

Дорожная карта для роста



Вердикт по кампании: "за" или
"против"



Портреты клиентов для
точечного маркетинга



Инструмент для увеличения продаж
обуви

Предобработка и очистка данных



1. Загрузка данных из **SQLite**-базы, **csv**, **txt** файлов
2. Проверка данных на дубликаты, аномалии, удаление аномалий
3. Данные отфильтрованы по стране
4. Объединение данных по **id** клиента
5. Обработка пропусков:
 - пропуски в **gender** заполнены на основе детерминированной зависимости с **personal_coef**
 - в **colour** заполнены модой после унификации цветов
 - в **product_sex** частично заполнены на основе текста в **product** и анализа **colour**, затем построена модель классификации (**LightGBM**) для заполнения оставшихся пропусков с оценкой качества на кросс-валидации (**F1-score: ~0.9731**).
6. Преобразование текстовых признаков:
 - создана колонка **product_group**, категоризирующая товары на основе первого слова в названии и словаря категорий (Транспорт и водные виды, Обувь, Одежда, Активный отдых и спортивный инвентарь).
 - унифицированы значения в колонке **colour**.

АВ тестирование



Метрики:

1. Конверсия в покупку
2. Выручка на клиента
3. Частота покупок на клиента
4. Средний чек покупки
5. Доля клиентов с повторными покупками
6. Сегментированная выручка (по категориям товаров, полу, возрасту)
7. Частота покупок (сегментация по товарным категориям, полу, возрасту)

Статистические тесты:

1. Хи-квадрат для сравнения конверсий
2. Шапиро-Уилка, Д'Агостино-Пирсона для проверки данных на нормальность распределения
3. Манна-Уитни для сравнения средних непараметрических данных независимых групп

Результат АВ теста



✓ Успех:

1. Выручка на клиента выросла на 19.4%
2. Частота покупок на 25.6%
3. Доля клиентов с повторными покупками на 3.45%

! Вызов:

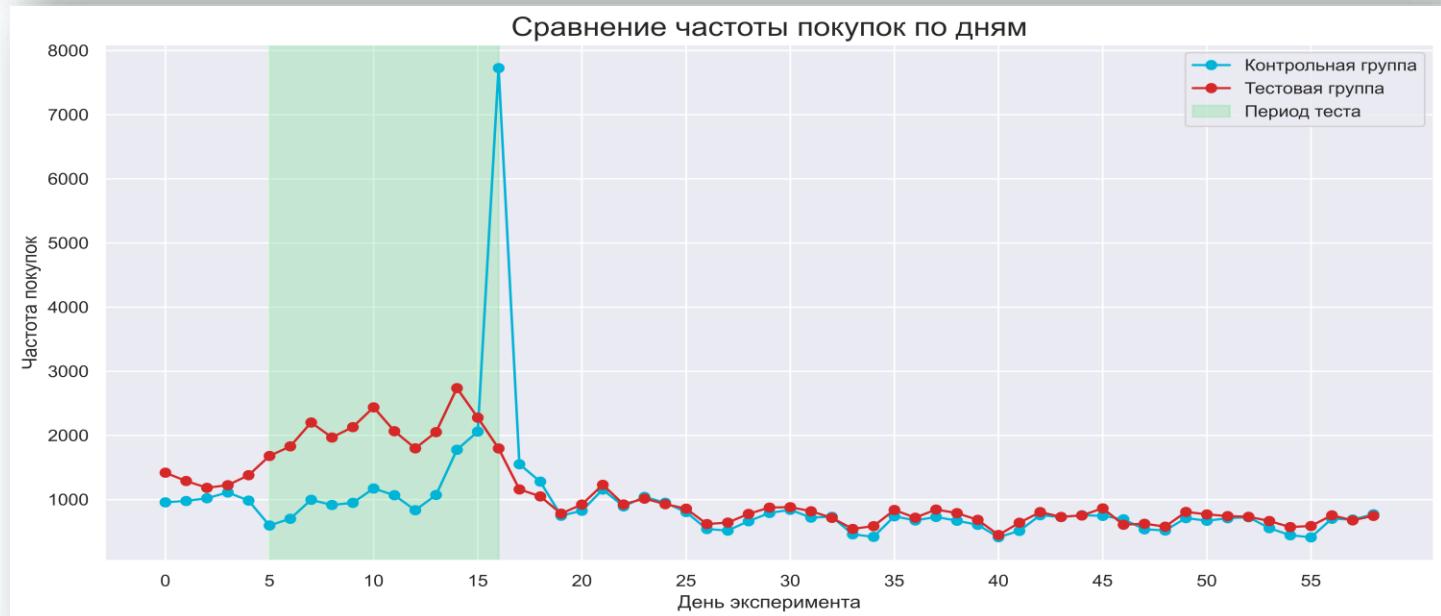
1. Конверсия снизилась на 3.7%
2. Средний чек на 5%



Ключевая аудитория:

Мужчины 19-бо лет, категории "Одежда" и "Обувь"

Выручка, частота покупок в разрезе дней



Рекомендации

Оптимизация кампании:



Продолжить персонализированные скидки для мужчин 19-60 лет на Одежду и Обувь



Увеличить средний чек



Повысить конверсию

Кластеризация клиентов

Данные:

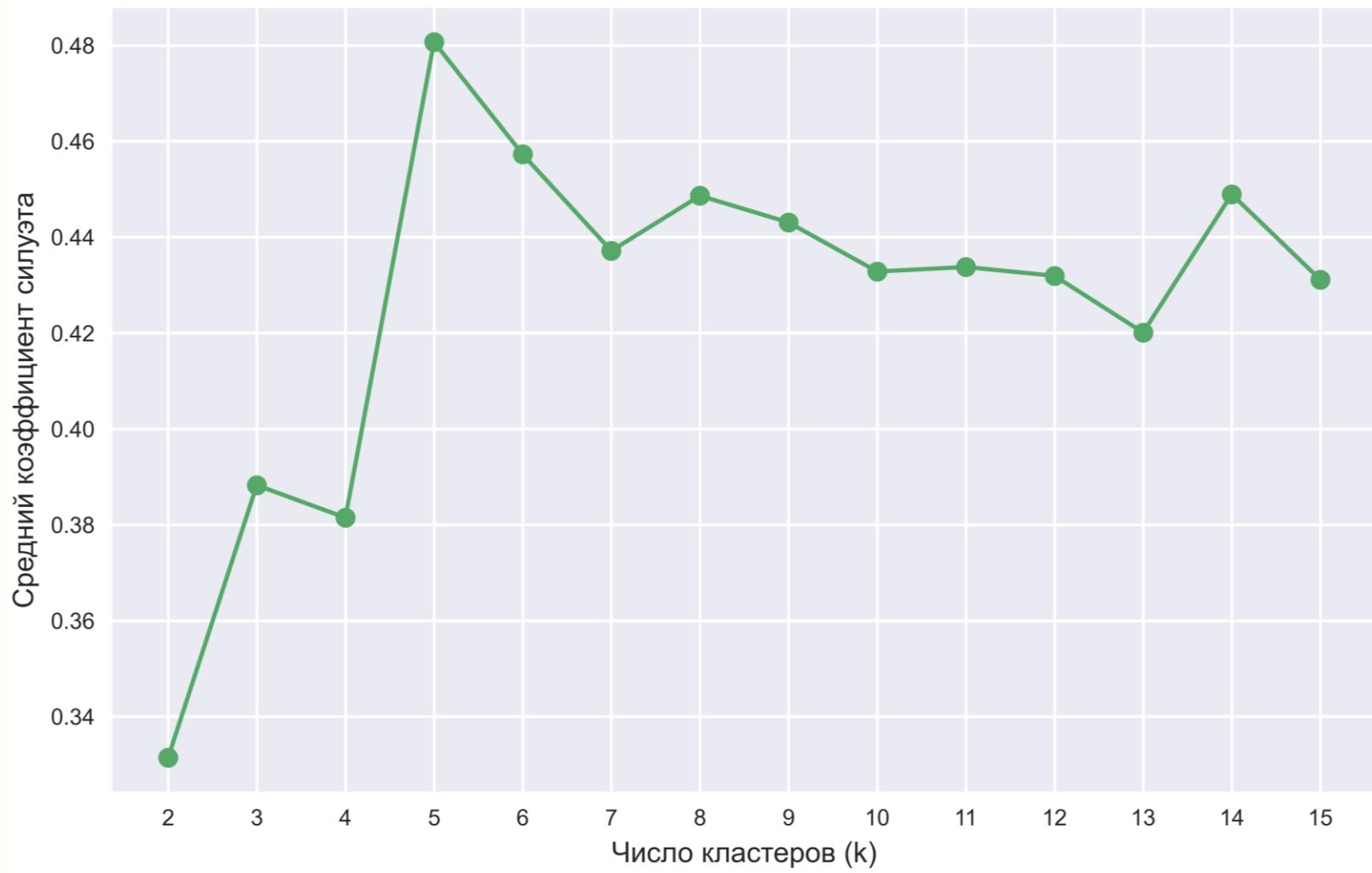
Числовые:

пол, возраст, цена товара,
скидка

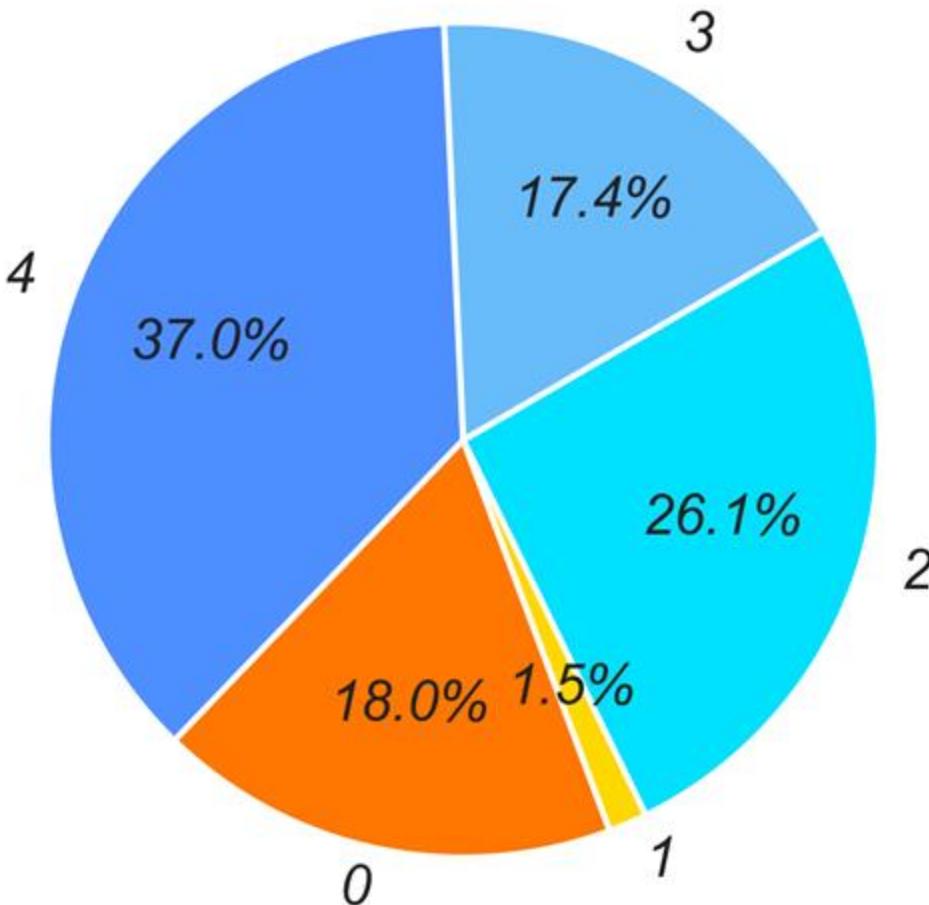
Категориальные:

категория товара, неделя
покупки

Силуэтный коэффициент для K-Prototypes



Распределение клиентов по кластерам



Портреты клиентов

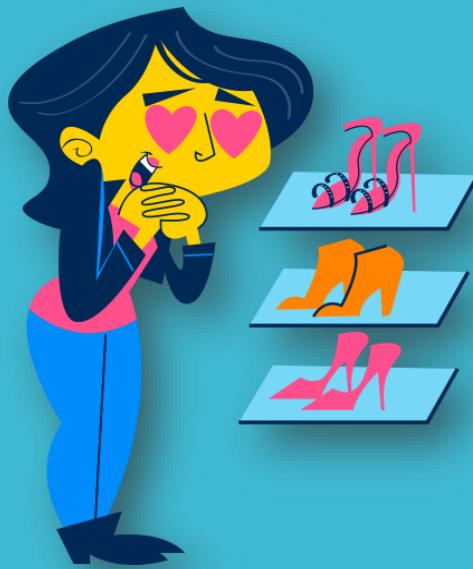
кластер	характеристика	доля	пол	Пик покупок	средний чек K₽	скидка	Категории товаров
0	 Рациональные покупательницы	18%		0-1	~3.0		Одежда Обувь
1	 Премиальные покупатели транспорта, т-ов для вод.спорта	1.5%	 82%	0-1	~66	 93%	Транспорт Водный спорт
2	 Любительницы моды	26%		0-1	~4.3		Одежда Обувь
3	 Экономные мужчины	17.5%		0-1	~4.1		Одежда Обувь
4	 Статусные мужчины	37%		0-1	~5.5		Одежда Обувь Акт.отдых

Персонализация – говорим на языке клиента



Кластер	Рекомендации
Кластер 0 <i>Рациональные покупательницы</i>	<ul style="list-style-type: none">• Уведомления о скидках, распродажах• Комплекты со скидкой, кросс-продажи аксессуаров• Накопительные баллы в обмен на покупки• Рассрочка без переплаты
Кластер 1 <i>Премиальные покупатели Транспорта и товаров для водного вида спорта</i>	<ul style="list-style-type: none">• Расширение премиального ассортимента• Допродажа комплектующих• Круглосуточная поддержка• Персональный консультант• Расширенная гарантия• Выездное обслуживание
Кластер 2 <i>Любительницы моды</i>	<ul style="list-style-type: none">• Подписка на новинки брендов• Расширение ассортимента• Кросс-продажи брендов и аксессуаров• Комплектные предложения
Кластер 3 <i>Экономные мужчины</i>	<ul style="list-style-type: none">• Бонусная программа: баллы в обмен на скидки или товар• Комплекты со скидкой• Скидки на следующие покупки
Кластер 4 <i>Статусные мужчины</i>	<ul style="list-style-type: none">• Сезонные кампании по новым коллекциям одежды, обуви• Бонусы за покупки без скидок• Кросс-продажи товаров для активного отдыха• Комплекты "одежда+обувь"• Карта постоянного покупателя

Предсказание склонности клиента к покупке



Подход (комбинированная модель):

propensity_score = вероятность покупки × ожидаемая доля обуви

Целевая переменная:



Классификатор: Бинарная переменная (1 — покупка обуви, 0 — нет).

Выявляет **потенциальных** покупателей обуви.



Регрессор: Доля покупок обуви ($shoes_mean \in [0, 1]$).

Оценивает **интенсивность** их склонности — насколько часто они выбирают обувь среди всех покупок.

Интерпретация:

Клиенты с высоким **propensity_score** указывают на приоритетных клиентов для таргетинга.

Инженерия признаков



Подготовка признаков:

Для классификатора:

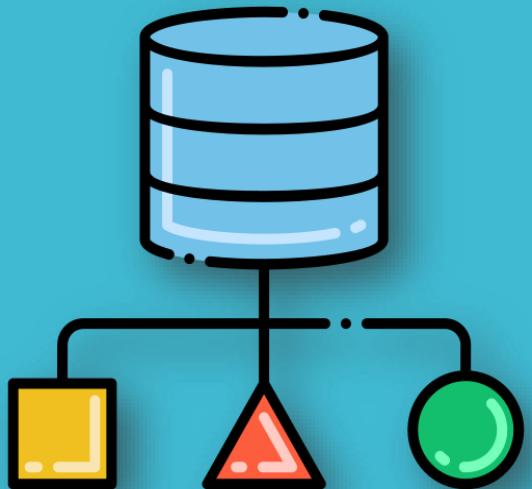
- Созданы бинарные признаки: `shoes` (1, если товар — обувь), `education_high` (1, если образование высшее).
- Возраст и цена разбиты на бины (`age_bin`, `cost_binned`).
- Агрегация по клиентам: частота покупок, медиана/сумма/размах стоимости, средний интервал между покупками, дисперсия дат, доля обуви и др.

Для регрессора:

Аналогичная инженерия признаков с добавлением логарифмического среднего цены (`cost_logmean`), отношения медианы цены к общей медиане (`cost_to_median_ratio`) и взаимодействия возраста с медианой цены (`age_cost_interaction`).

Учтены только клиенты, купившие обувь (`shoes_mean > 0`).

Моделирование



Классификатор (LGBMClassifier)

Задача: Прогнозирование факта покупки обуви.

Метод: LightGBM с использованием `class_weight`, `dart`, `adasyn` для балансировки классов, кодирование категориальных признаков (`OneHotEncoder`), масштабирование числовых (`StandardScaler`).

Метрики на тестовом наборе:

F1-macro: 0.7026

ROC-AUC: 0.7744

Регрессор (LGBMRegressor)

Задача: Прогнозирование доли покупок обуви (`shoes_mean`).

Метод: LightGBM с регрессией, кодирование, масштабирование, создание полиномиальных признаков с использованием `PolynomialFeatures`.

Метрики на тестовом наборе:

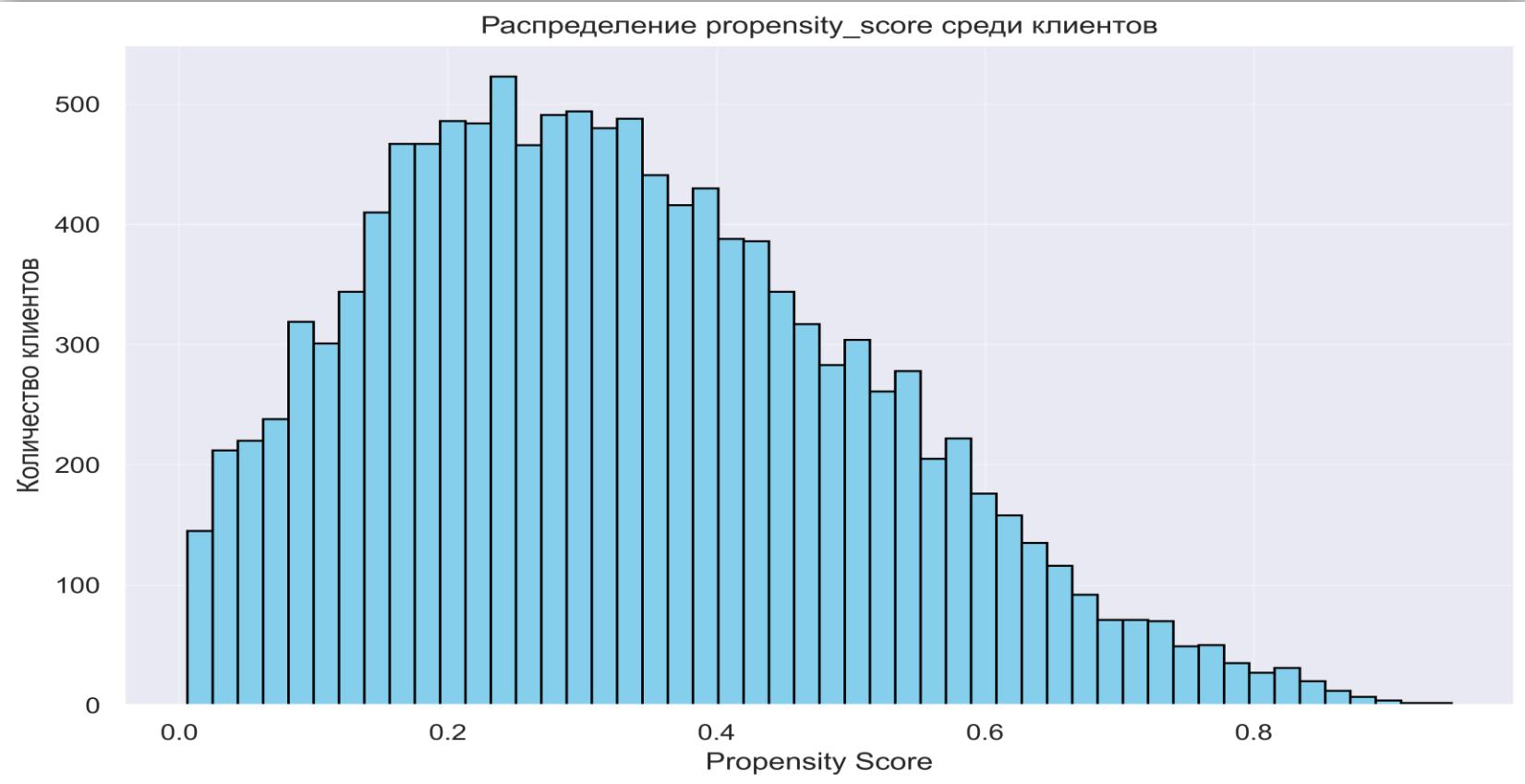
MAE: 0.1690

MSE: 0.0446

RMSE: 0.2112

R²: 0.4419.

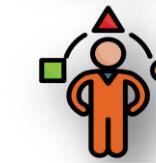
Таргетинг по «горячим» клиентам



Рекомендации.



1. Использовать *propensity_score* для таргетинга клиентов с высокой склонностью к покупке обуви.



2. Тестировать персонализированные предложения (скидки, комплекты) для клиентов с высоким *propensity_score*.

Спасибо за внимание!

