

Сегментация клиентской базы: Online Shoppers Purchasing Intention

Руководитель: Глазков А.А.

Выполнила: Ишутина Полина

БКНАД211

Описание проекта

Dataset: Online Shoppers Purchasing Intention.

География: Турция.

Основные задачи:

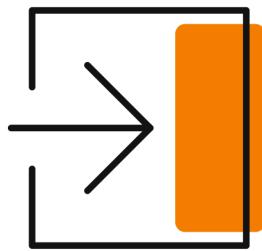
- разработать модель данных, исследовать качество;
- сегментировать клиентскую базу аналитически;
- подготовить витрины данных для моделей сегментации;
- построить модели сегментации, определить минимальное количество сегментов и исследовать модели на качество;
- выделить профили клиентов и сформировать Next Best Offer;
- оценить финансовую эффективность кампании.



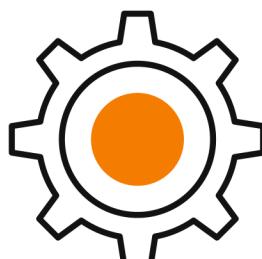
В датасете хранятся данные о посещениях сайта



Численные признаки: количество страниц разных типов (Administrative, Informational, Product Relates) и время, которое пользователь на них провел, доля отказов и выходов, средняя ценность страницы перед транзакцией, близость к праздникам.



Качественные признаки: месяц, операционная система, браузер, регион, источник трафика, тип пользователя.



Индикаторы: выходные дни, успешное завершение транзакции.

Очищаем данные и готовим витрину

01

Обрабатаем численные и категориальные переменные.

Проверим описательные статистики, уберем пропуски, нормируем данные, удалим и перегруппируем лишние категории.

02

Определим корреляцию величин и введем новые признаки

Построим тепловые карты, чтобы оценить корреляцию. Добавим признаки для подробного анализа ProductRelated-страниц.

03

Декодируем значения категориальных признаков

Заменим числа на осмысленные названия городов, операционных систем, браузеров и источников трафика.



соберем витрину



Зафиксируем выручку в стандартной ситуации

| | |
|---------------------|-----------------------|
| Пользователей | 12330 |
| Доля покупок | 0,14 |
| Стоимость товара | 1 000,00 ₽ |
| Выручка | 1 726 200,00 ₽ |
| Расходы на рассылку | 57 600,00 ₽ |
| Прибыль | 1 668 600,00 ₽ |



Удалось выделить 5 основных сегментов

- 1. Стамбул + OS Windows.**
- 2. Анталия + OS X.**
- 3. Остальные города + Windows.**
- 4. Остальные города + OS X.**
- 5. Все города + iOS.**

У каждого из сегментов клиентской базы есть выраженный предпочтаемый источник трафика и период в году.



Сегмент «Стамбул & Win»

Next Best Offer: скидка 15% на любой товар при покупке с устройства с ОС Windows в Стамбуле в летний период.

2200 + посещений
сайта

Приобретают товары
круглый год, но чаще
всего – в ноябре и в мае

Любимый трафик: поисковые
системы и реферальные программы.

Сегмент «Анталия & OS X»

Next Best Offer: скидка 5% на любой товар при покупке с устройства с ОС X в Анталье в зимний период.

Немного посещений сайта, но доля покупки 60%

Чаще всего приобретают товары в зимний период

Любимый трафик: email-рассылки.

Сегмент «not Стамбул & Win»

Next Best Offer: скидка 10% на любой товар при покупке с устройства с ОС Windows в любом городе, кроме Стамбула, в летний период.

3600+ посещений
сайта

Чаще приобретают
товары летом, любимый
месяц – май.

Любимый трафик: поисковые
системы и реферальные программы.

Сегмент «not Анталья& OS X»

Next Best Offer: скидка 15% на любой товар при покупке с устройства с ОС X в любом городе, кроме Антальи, в летний период.

5706+ посещений сайта

Приобретают товары круглый год, но чаще всего – в летнее время.

Любимый трафик: реферальные программы.

Сегмент «all regions & iOS»

Next Best Offer: скидка 5% на любой товар при покупке с устройства с ОС Windows в Стамбуле в зимний период.



2100 + посещений
сайта

Приобретают товары
круглый год, но чаще
всего — зимой.

Любимый трафик: социальные
сети и поиск.

Для сегментации использовала три основных метода

Каждый из методов дал свою пользу,
но с ними также возникали некоторые
сложности.



01

К-means, или метод локтя.

02

Иерархическая кластеризация.

03

Аналитическая сегментация базы
по тепловым картам и известному
количеству сегментов.

Какие сложности возникли?

≡

- K-means сегменты вышли однородными.
- Иерархическая кластеризация тоже – силуэт обеих моделей оказался близок к нулю.
- РСА показал, что данные хорошо перемешаны.

Решение: принимаем минимальное количество сегментов равным 5 и ищем сегменты аналитически по тепловым картам.

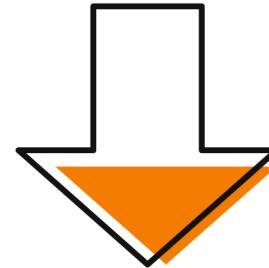
Оценим новую стоимость товара

| | Зима | Лето |
|---|------|------|
| 1 | 0,49 | 0,51 |
| 2 | 0,8 | 0,2 |
| 3 | 0,44 | 0,56 |
| 4 | 0,46 | 0,54 |
| 5 | 0,52 | 0,48 |

| Стоимость зимой | Стоимость летом | Матожидание цены товара |
|-----------------|-----------------|-------------------------|
| 1000 | 1150 | 1076 |
| 1050 | 1000 | 1040 |
| 1000 | 1100 | 1056 |
| 1000 | 1150 | 1081 |
| 1050 | 1000 | 1026 |

Оцениваем прибыль

1 668 600 – исходная прибыль



1 968 948 – итоговая прибыль

Эффективность сегментации составила 18%

Спасибо!