

Анализ маркетинговых кампаний для спортивного ритейлера

Описание проекта

Этот проект посвящён анализу поведения клиентов и эффективности маркетинговых кампаний на основе данных крупного спортивного ритейлера. С помощью инструментов анализа данных и машинного обучения были исследованы ключевые метрики, протестированы гипотезы и предложены улучшения для будущих маркетинговых активностей.

Этапы выполнения проекта

1. Загрузка и первичная фильтрация данных

- Распакованы и загружены данные.
 - Обнаружены пропуски в поле **пол**.
 - Оставлены только данные по **стране 32**.
-

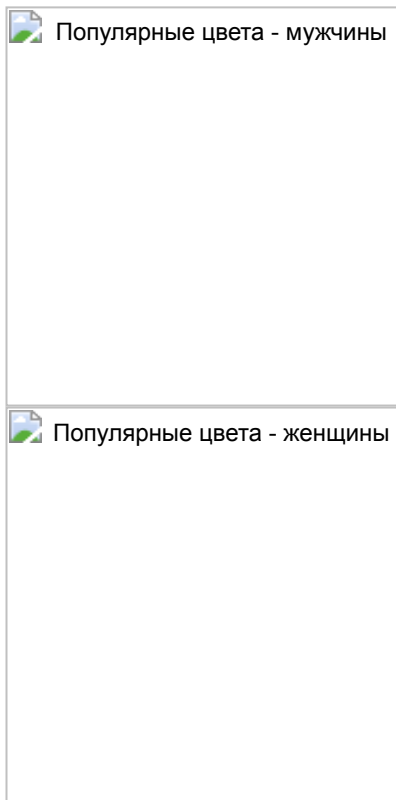
2. Подготовка данных

Анализ таблицы `Purchase`:

- Найдены несоответствия в названиях цветов.
- Пропущенные значения цвета.

Решения:

- Замена «Ё» на «Е»
- Приведение к **нижнему регистру**
- Очистка от лишних символов ('-', '/', ',')
- Построены графики распределения популярных цветов товаров отдельно для мужчин и женщин.
- Пропущенные значения заменены на самый часто встречающийся цвет для каждого пола на основе визуализации:



3. Создание новых признаков

Добавлены признаки:

- **Company** — название компании продукта. Извлечено как первое английское слово из строки описания товара.
- **Product_type** — тип товара. Извлечён как первое слово описания, после чего убраны последние три символа для устранения вариативности по родам/падежам.
- **Price_category** — категория цены товара:
 - Высокий уровень
 - Средний уровень
 - Низкий уровень

4. Модель для восстановления пола

Цель: восстановить пропущенные значения пола на основе данных о покупках.

Подход:

- Построена бинарная модель классификации (мужчина / женщина).
- Использованные признаки:
 - Цвет товара
 - Компания
 - Тип продукта

- Категория цены
- Возраст клиента
- Проведено кодирование категориальных признаков и обучение модели Random Forest Classifier.

Результаты:

- Модель достигла точности около 80%.
- Предсказанные значения пола были присвоены клиентам с отсутствующими данными.
- Данные объединены в одну итоговую таблицу с восстановленными значениями.

5. Подготовка данных к анализу кампаний

- Загрузка таблиц с клиентами, участвовавшими и не участвовавшими в маркетинговых кампаниях.
- Очистка и объединение данных с основной таблицей покупок.
- Распределение пользователей на контрольную и тестовую группы по кампаниям: *first_campaign* / *second_campaign*

6. Ключевые метрики для A/B-тестов

Проведён анализ основных метрик для каждой кампании с целью подготовки к статистическим проверкам гипотез:

Первая кампания:

- Средний чек на клиента (ARPU)
- Средний чек на одну покупку (AOV)
- Общий доход каждой группы
- Разница в сумме покупок между группами

Вторая кампания:

- Количество уникальных пользователей
- Средний чек
- Конверсия (доля клиентов, совершивших покупку)
- Доля пользователей, воспользовавшихся скидкой
- Доля товаров, купленных со скидкой по сумме

7. ☐ Проведение A/B-тестов

Проведён анализ эффективности маркетинговых кампаний с использованием A/B-тестов по ключевым метрикам.

Использовались **t-тесты**, направленные на выявление статистически значимых различий между контрольной и тестовой группами.

☐ Первая кампания

□ Метрика: ARPU (Average Revenue Per User)

- Нулевая гипотеза (H_0): Средний доход на пользователя одинаков в контрольной и тестовой группе.
 - Альтернативная гипотеза (H_1): Средний доход на пользователя отличается между группами.
 - Результат t-теста:
($p\text{-value} < 0.05$)
 - Вывод:
ARPU значительно выше у пользователей, участвовавших в скидочной кампании
-

□ Метрика: AOV (Average Order Value)

- Нулевая гипотеза (H_0): Средний чек на покупку одинаков в обеих группах.
 - Альтернативная гипотеза (H_1): Средний чек на покупку отличается между группами.
 - Результат t-теста:
($p\text{-value} < 0.05$)
 - Вывод:
Средний чек на покупку (AOV) немного ниже в группе со скидками, что ожидаемо, так как клиенты платили меньше за товары.
-

□ Дополнительно:

Построен график общего дохода по двум выборкам, демонстрирующий итоговый финансовый эффект от кампании.
(место для визуализации столбчатой диаграммы или pie chart)

Вторая кампания

Метрика: AOV (Average Order Value)

- Нулевая гипотеза (H_0): Средний чек на покупку одинаков в контрольной и тестовой группе.
 - Альтернативная гипотеза (H_1): Средний чек одной покупки различается между группами.
 - Результат t-теста:
($p\text{-value} = 0.268$)
 - **Вывод: скидка НЕ повлияла на средний чек
- **
-

□ Дополнительно:

Создана таблица с разбивкой клиентов на тех, кто **воспользовался скидкой**, и кто **не воспользовался**, с анализом их поведения и среднего чека.

8. Общие выводы по маркетинговым кампаниям

- **Первая кампания:**

*Средний чек на покупку (AOV) немного ниже в группе со скидками, что ожидаемо, так как клиенты платили меньше за товары. Средний чек (ARPU) в группе со скидками значительно выше, чем в группе без скидок, что говорит о том, что клиенты у которых есть скидка тратились гораздо больше. Это говорит нам об успехе данной компании. В качестве подтверждения выступает график общего дохода по двкм выборкам, где клиеты со скидкой принесли больший доход компании *

- **Вторая кампания:**

Тест показал, что средний чек из 2 выборок не имеет статистически значимое значение, что говорит о неэффективности данной кампании. Все подтверждается графиками

- **Вывод:**

Первая кампания сработала успешно, что говорит о том, что можно масштабировать ее. Вторая кампания не принесла нужных результатов. Стоит ее пересмотреть.

9. Построение модели склонности клиента из города 1188 к покупке определённого товара

Подготовка данных:

- Отфильтрованы клиенты только из города 1188.
- Проведена очистка: удалены непопулярные категории товаров, которые редко покупаются — это уменьшает шум в модели.

Обучение моделей:

- Построены две модели машинного обучения:
 - Decision Tree Classifier
 - Random Forest Classifier
- Сравнены результаты на валидационной выборке. Оценивалась точность предсказания.
- Выбрана модель дерева решений по соображениям интерпретируемости и достаточной точности.

Экспорт модели:

- Модель дерева решений была **сохранена/скачана** для дальнейшего использования в продакшене или демонстрации.

