

Итоговый проект «Data Analyst Junior» Анализ эффективности проведённых маркетинговых кампаний

Денис Константинович Спиридонов

Введение:

В данном аналитическом отчете я представляю результаты исследования, проведенного в контексте маркетинговых кампаний и моделирования поведения клиентов.

Моей целью было проанализировать эффективность маркетинговых мероприятий и построить модель склонности клиентов города 1188 к покупке мужских кроссовок, так как именно в этом городе планируется следующая маркетинговая компания.

Для достижения этой цели я использовал доступные данные о покупках клиентов за два месяца, а также информацию о маркетинговых кампаниях и клиентском поведении.

Обзор данных и их качество:

Исходные данные включают информацию о покупках клиентов за два месяца и сведения о маркетинговых кампаниях.

- Id уникальный код клиента
- cost средний чек покупки
- product_sex пол потенциального покупателя продукта
- base_sale наличие скидки на момент покупки
- **dt** день покупки
- gender пол покупателя
- age возраст покупателя
- education образование покупателя
- city город
- country страна
- personal_coef персональный коэффициент покупателя
- main_color основной цвет продукта
- product_category категория продукта
- brand бренд продукта
- first_company первая рекламная компания (positive тестовая группа, negative контрольная группа)
- second_company вторая рекламная компания (positive тестовая группа, negative контрольная группа)
- target показатель того купил ли клиент мужские кроссовки (1 да, 0 нет)

Часть данных о поле клиента была утеряна.

Восстановление данной информации является очень важным шагом, так как это один из ключевых признаков для анализа. Для этой цели была обучена модель случайного леса, которая использовала имеющиеся данные для восстановления пола клиентов.

Для того чтобы выявить проблемы переобучения модели, была проведена кросс-валидация на данных.

В качестве метрики качества модели была выбрана F1-мера

Результат:

Mean F1 Score: 0.9999083659466652

Standard Deviation of F1 Score: 0.00011326997953445225

<u>Средний F1 Score</u> для модели случайного леса составляет 0.99, что очень близко к 1. Это говорит о высокой производительности модели и способности хорошо обобщать данные.

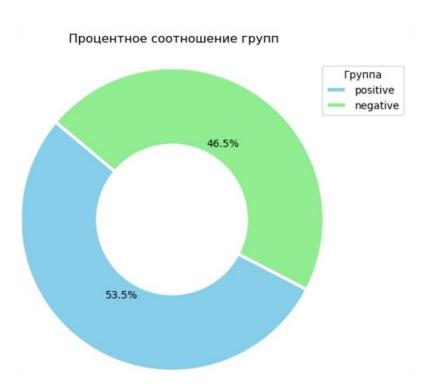
<u>Стандартное отклонение F1 Score</u> равно 0.0001, что очень низкое. Это указывает на то, что оценки F1 Score для разных фолдов близки друг к другу и имеют маленькое разброс.

На основе этих данных, можно сделать предположение, что модель не переобучилась, и она хорошо обобщает данные. Оценки F1 Score близки к идеальному значению 1, и стандартное отклонение невелико.

Маркетинговые кампании:

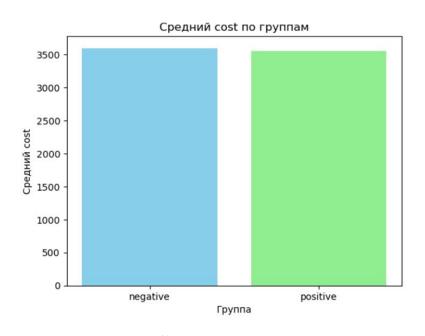
В данном разделе проведем подробный анализ двух маркетинговых кампаний.

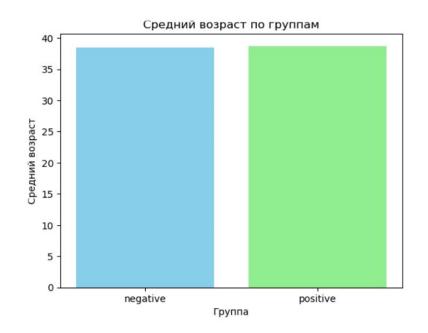
Первая кампания проводилась в период с 5-го по 16-й день. Она включала в себя предоставление персональных скидок 5 000 клиентам через email-paccылку.



Средний чек для тестовой группы составляет 3553.9, для контрольной 3595.8. Т-test показал значение р-значение: 0.077, - различие не является статистически значимым.

Средний возраст для тестовой группы составляет 38.7, для контрольной 38.4 года. Т-test показал значение р-значение: 0.0086, а значит различие является статистически значимым.



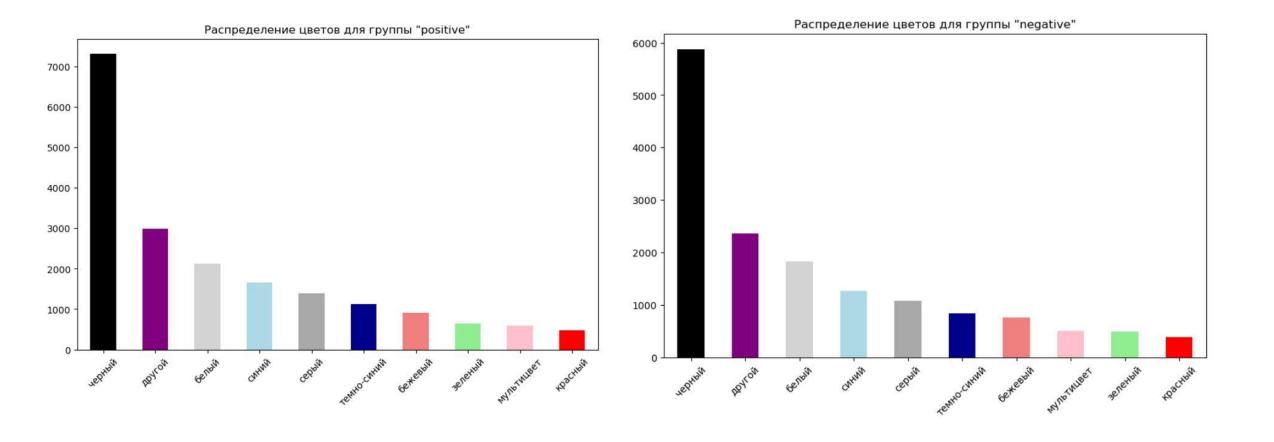


Так как средний чек между двумя группами различается незначительно, то выручка в группах будет зависеть от количества клиентов в каждой группе. (примерно 46/54)

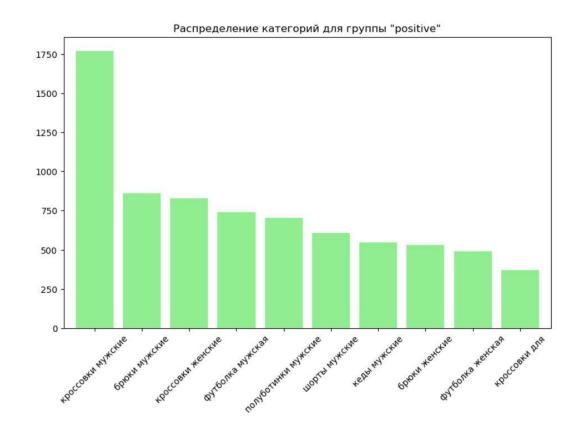
Для возраста различие хоть статистически и значимо, но в реальности возраст клиентов практически одинаков - 38 лет.

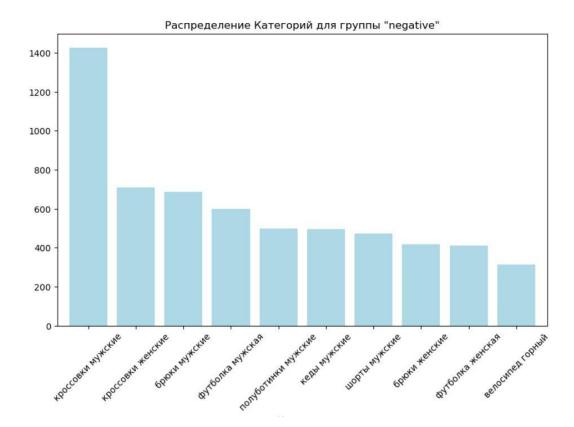
На данных графиках видно, что обе группы отдают предпочтения одним и тем же цветам продуктов: В лидерах черный, далее белый, синий и тд.

На это стоит обратить внимание при формировании складских запасов.

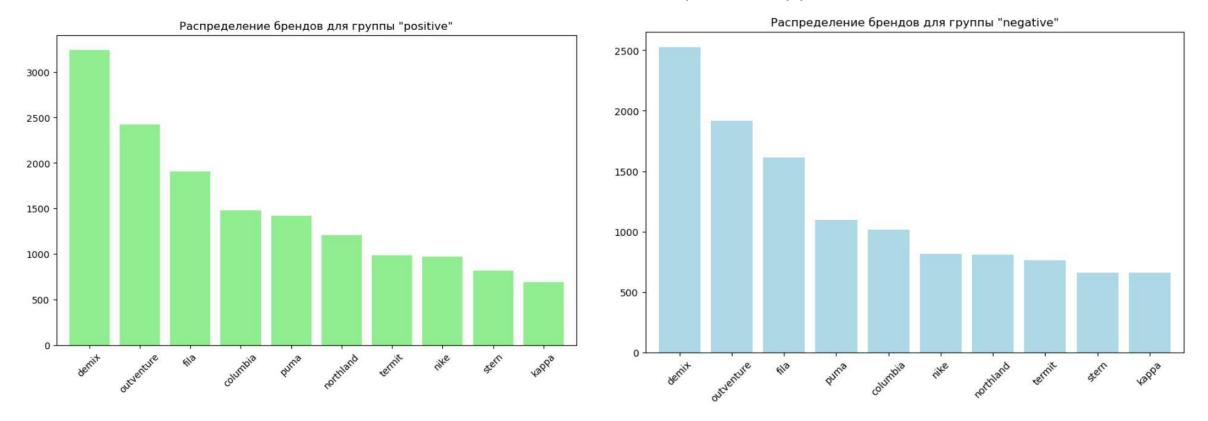


На данных графиках изображено распределение по категориям товаров тестовой и контрольной группах. Разницы как таковой нет.





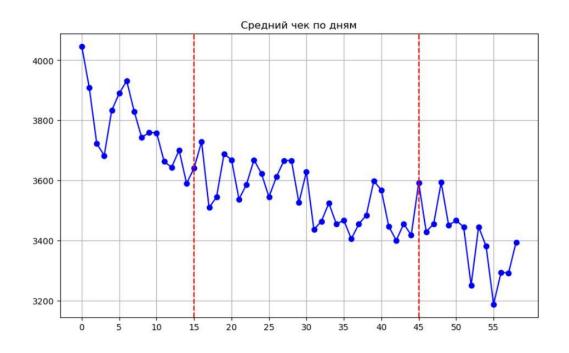
Такая же картина наблюдается и по брендам товаров. В топ-10 самых продаваемых брендов находятся одни и те же наименования и по тестовой и по контрольной группе



На основе вышеуказанных данных можно сделать вывод что разницы между тестовой и контрольной группой нет. Маркетинговая компания ощутимых результатов не принесла.

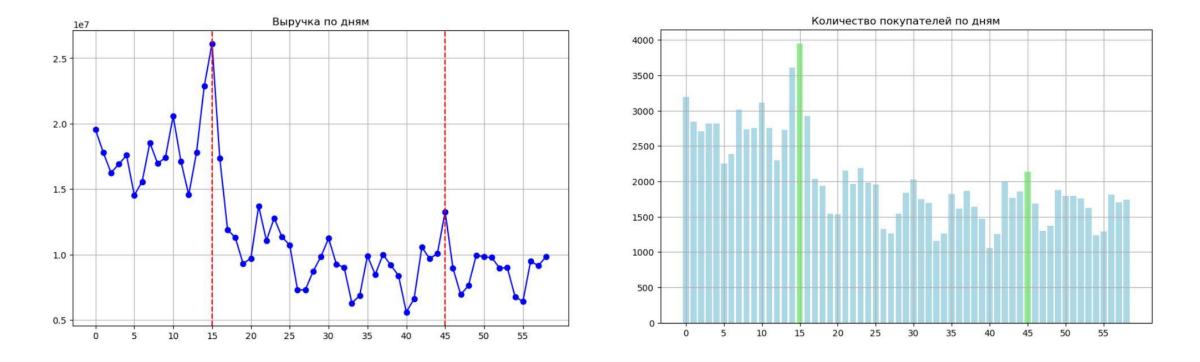
Вторая маркетинговая кампания была ориентирована на жителей города 1 134 и представляла собой баннерную рекламу на билбордах.

Скидка предоставлялась каждое 15-е число месяца (15-й и 45-й день).



На данном графике представлено распределение среднего чека по дням. Красными линиями выделены дни скидок.

Из графика видно, что наличие скидки не оказало никакого эффекта на средний чек покупателей. Зато заметен явный тренд на снижение среднего чека клиентов на протяжении двух, исследуемых, месяцев.



Если посмотреть на выручку в динамике, то можно увидеть всплески на 15 и 45 день. Данные всплески прямо пропорциональны количеству покупателей в указанные дни.

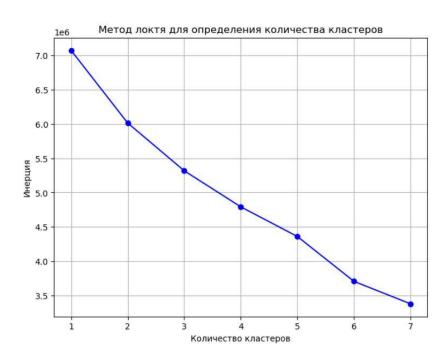
Вывод:

<u>Вторая маркетинговая компания удалась.</u> Выручка, как и количество покупателей в день акции возросли. !Но, если смотреть на картину в целом, то виден явный тренд на снижение покупательского интереса. Такие показатели как средний чек, выручка, количество покупателей снижаются на всем периоде исследования.

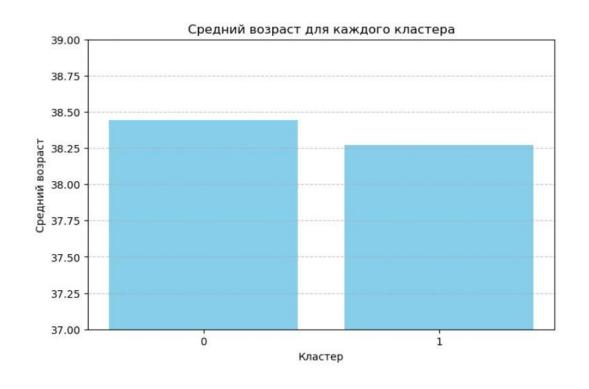
Кластеризация клиентов:

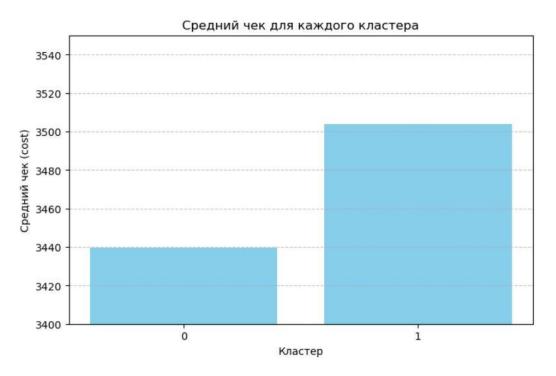
Кластеризация клиентов позволяет компании лучше понимать разнообразие своей клиентской базы и оптимизировать стратегии привлечения и удержания клиентов для каждой группы.

Методом локтя было определено, что оптимальное количество кластеров составляет 2. Затем, применяя метод К-средних с указанным количеством кластеров, данные были разделены на два кластера.

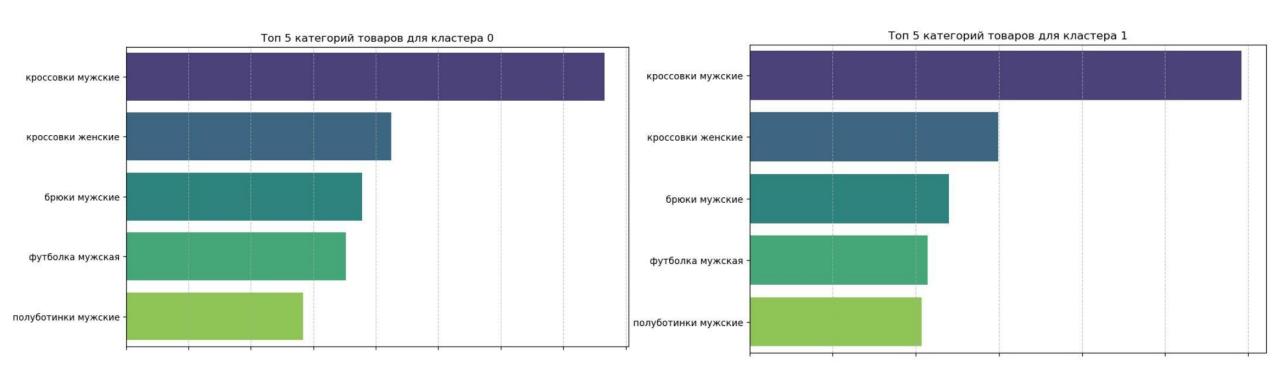


На данных графиках показано различие между кластерами по возрасту и среднему чеку

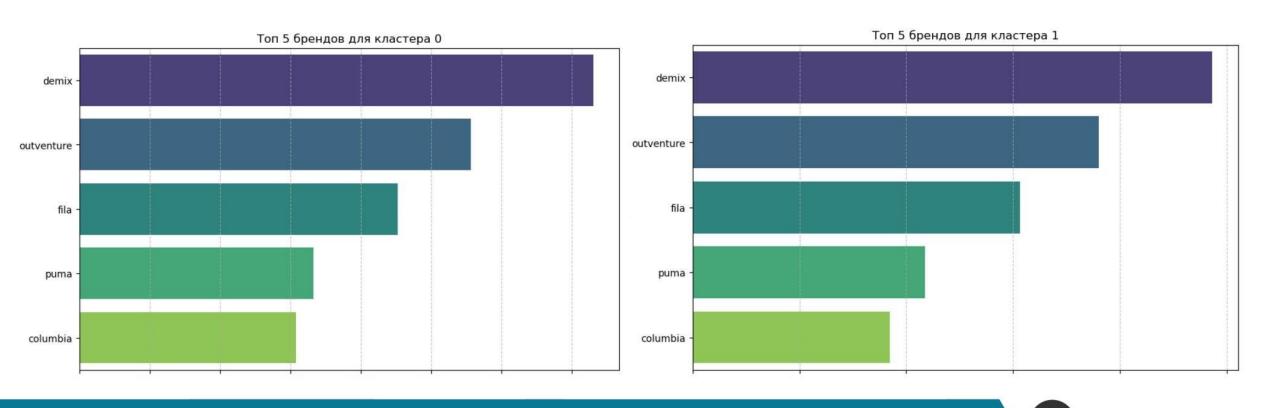




Из данных графиков видно, что оба кластера клиентов предпочитают идентичные категории товаров



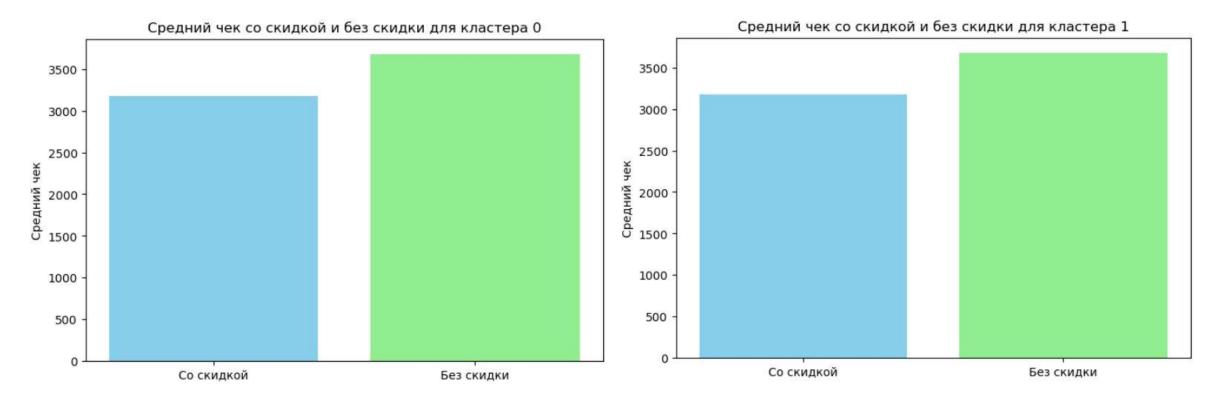
Такая же картина наблюдается и по брендам

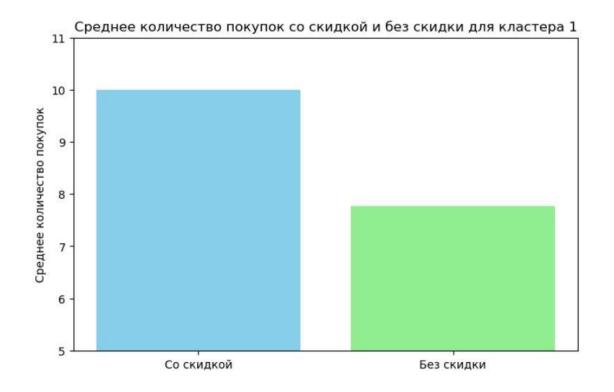


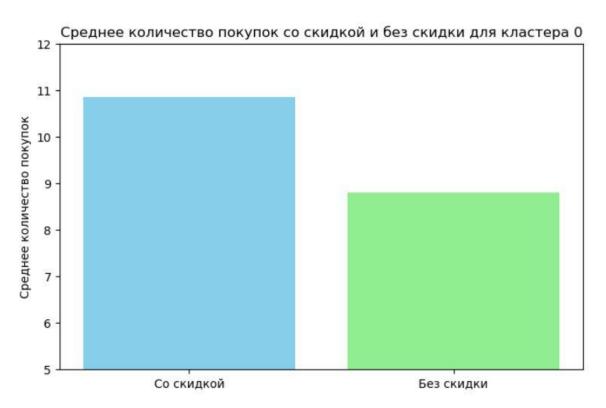
Из данного графика видно, что оба кластера клиентов тратят чуть меньше с наличием скидки, нежели без нее.

Средний чек со скидкой для кластера 0: 3088.9 Средний чек без скидки для кластера 0: 3659.7

Средний чек со скидкой для кластера 1: 3183.6 Средний чек без скидки для кластера 1: 3680.7







Среднее количество покупок со скидкой в кластере 0: 10.9 Среднее количество покупок без скидки в кластере 0: 8.8

Среднее количество покупок со скидкой в кластере 1: 10.0 Среднее количество покупок без скидки в кластере 1: 7.8

Наличие скидки положительно влияет на количество покупок клиентов. <u>Скидка дает прирост в 20-25% к количеству покупок.</u> Что не может не оказать положительного влияния на бизнес.

Модель склонности клиента к покупке мужских кроссовок:

Данная модель позволяет лучше понимать, какие клиенты более вероятно будут приобретать данное продуктовое предложение, и позволяет настраивать маркетинговые усилия для увеличения конверсии. Это поможет компании сэкономить ресурсы и повысить уровень обслуживания клиентов, что в свою очередь может привести к увеличению выручки и укреплению клиентской базы.

В качестве модели были выбраны два метода:

- метод логистической регрессии
- метод случайного леса

Метрики качества логистической регрессии

Accuracy: 0.90

F1-Score: 0.91

ROC-AUC: 0.90

Accuracy (Точность): 0.90. Это означает, что около 90% всех предсказаний, сделанных этой моделью, верны. Очень высокий показатель точности.

F1-Score (F1-мера): 0.91. Это высокий показатель F1-меры, который объединяет точность и полноту. Он близок к 1, что указывает на хорошее сбалансированное соотношение между точностью и полнотой.

ROC-AUC: 0.90. ROC-AUC близок к 1, что означает, что модель имеет высокую способность разделять классы.

Метрики качества случайного леса

Accuracy: 0.82

F1-Score: 0.84

ROC-AUC: 0.82

Ассигасу (Точность): 0.82. Это означает, что около 82% всех предсказаний, сделанных этой моделью, верны.

F1-Score (F1-мера): 0.84. Это хороший показатель F1-меры, который также близок к 1.

ROC-AUC: 0.82. ROC-AUC также является хорошим, но немного ниже, чем у модели логистической регрессии.

Модель логистической регрессии имеет более высокие метрики точности и F1-меры, что может быть связано с тем, что данные могут быть хорошо разделимы линейной гиперплоскостью. Она также имеет более высокий ROC-AUC, что свидетельствует о хорошей способности модели разделять классы.

Выводы:

В ходе данного исследования:

- Был проведен анализ данных о покупках клиентов за два месяца, включая информацию о маркетинговых кампаниях, сегментации клиентов и модели склонности к покупке мужских кроссовок.
- В результате анализа было установлено, что первая маркетинговая кампания с персональными скидками не принесла ощутимого эффекта, а вторую кампанию с баннерной рекламой следует оптимизировать.
- Сегментация клиентов позволяет более эффективно настраивать маркетинговые стратегии для разных групп клиентов.
- Модель склонности к покупке мужских кроссовок предоставляет инструмент для персонализации маркетинговых усилий и увеличения вероятности успешных продаж.