

Аналитический отчет

Магазин спортивной атрибутики

Теребиленко Матвей

Структура отчета:

- 1.Первичный анализ данных и представление о деятельности компании
- 2.Проблема ,ее сущность, влияние на бизнес, формирование вопроса
- 3.Решение проблемы и ответ на поставленный вопрос
- 4.Аргументация
- 5.Заключение и рекомендации

1. Первичный анализ данных и представление о деятельности компании

Для анализа были предоставлены следующие данные: `ids_first_company_negative` и `ids_first_company_positive` (уникальные идентификаторы клиентов к которым были применена скидка и к которым нет), `shop_database.db` (база данных содержащая информацию о ID клиентов, их пол, возраст, образование, страна и город проживания, данные с персональными коэффициентами клиентов, которые рассчитываются по некоторой закрытой схеме, данные о покупках: ID покупателя, название товара, цвет, стоимость, гендерная принадлежность потенциальных покупателей товара, наличие скидки), `personal_data.csv.gz` (файл с утерянными данными о покупателе). Данные были объединены в единую таблицу, основные пропуски были в `product_sex` (для какого пола предназначен предмет) и потеряны данные в `sex` (пол покупателя), стоит поменять систему сбора этих данных, данные были восстановлены с помощью классификационной модели. Компания продает весь спектр спортивной атрибутики, во многих городах, страна анализировалась с названием 32. Размер итогового набора данных после обработки составил 790000 строк, был создан дополнительный столбец о наличии покупке, он будет использоваться как целевой, в процессе анализа выборка будет уменьшаться вследствие нехватки мощностей (до 100000 строк).

2. Проблема ,ее сущность, влияние на бизнес, формирование вопроса

Следующие этапы работы были выполнены :

- предобработки данных,
- бинарной классификации,
- проведения А/В-тестирования,
- кластеризации,
- построения модели склонности клиента к покупке

Все эти этапы связаны аналитикой данных в целях того, чтобы увеличить выручку компании. Были проведены а/в тесты, чтобы увидеть принесет ли больше прибыли определенные изменения в продукте, их результаты были так же проанализированы и проверены статистическими критериями. Так же был проведен кластерный и ассоциативный анализ для того, чтобы понимать какие продукты выгоднее продавать. На основе всех данных была построена предсказательная модель, чтобы больше клиентов доводить до конверсии в покупку. Все это нужно для увеличения выручки фирмы. Из этого следуют вопросы. Как интегрировать и интерпретировать результаты экспериментов? Какие продукты больше продавать? Как применить все результаты исследования для увеличения выручки?

3. Решение проблемы и ответ на поставленный вопрос

Решение проблемы полностью основывается на результатах анализа данных. Увеличению выручки способствует применение результатов эксперимента,нахождение взаимосвязей между товарами и предсказание покупательской способности потребителя. Начну с результатов эксперимента, первое тестирование гипотезы о влияние скидки на выручку показало более высокую выручку у контрольной группы(у которой не было скидки),поэтому изменения разворачивать не стоит(скидку применять ко всем клиентам не стоит), дальнейшие рекомендации и гипотезы сделанные из тестирований будут выдвинуты в 5 разделе отчета. Ассоциативный анализ связал конверсию в покупку с такими вещами,как:Брюки женские Outventure,Брюки мужские Demix,Брюки мужские FILA,Брюки мужские Outventure,Купальник женский Joss ,Лиф женский Joss и футболки этих фирм. Чаще не совершают покупку люди со средним образованием. Так же, известно, что следующая маркетинговая компания будет запускаться в городе 1188,ассоциативный анализ связал с этим городом и конверсией в покупку следующие вещи: Брюки женские Outventure,Брюки мужские Demix,Брюки мужские Outventure,Купальник женский Joss,Лиф женский Joss,Плавки женские Joss,Футболка женская/мужская Demix,шорты Demi,на эти товары стоит сделать акцент так как они пользуются большим спросом.Так же была создана модель, которая предсказывает будет ли совершенна покупка, она поможет увеличить выручку, так как будет предсказывать клиентов готовых совершить покупку и им можно предложить больше товаров и новые товары, остальной

группе можно будет предложить скидку или предложить пройти опрос и тем самым можно будет улучшить воронку и конверсию в покупку.

4. Аргументация

Аргументация по результатам проведенных экспериментов:

1-е тестирование: Были рассмотрены две группы, для которых была предложена скидка и для которых не было предложено скидки, проверка результатов эксперимента статистическим критерием и анализом выручки показала, что контрольная группа (без скидки) превосходит тестовую по p -value (статистический критерий не прошел порог значимости, значит принимаем нулевую гипотезу), так же средняя выручка контрольной группы выше ($5570 > 5312$) и тестовая группа больше контрольной на 5000 покупателей, то есть общий объем выручки больше у контрольной группы, но клиентов больше пришло в тестовую группу (25000, в контрольной 20000), поэтому метрика выручки ниже в тестовой группе, но метрика количества покупок выше, моей рекомендацией будет провести эксперимент с меньшей скидкой и этот эксперимент не разворачивать (не применять ко всем пользователям), так как метрика выручки снизилась и в общем объеме и в среднем, p -value не прошел порог значимости.

2-е тестирование: Выборка состоит из 10864 покупателей, средняя выручка по всей выборке 5800, нет данных для сравнения, поэтому нужно сравнить это с выручкой за предыдущие схожие сроки и сделать выводы, в сравнении с выручкой по первому эксперименту, это хороший результат, он выше, чем в первом эксперименте, но по этому сравнению сделать выводы нельзя.

Аргументация по товарам на которые стоит обратить большее внимание: Эти товары были выведены путем ассоциативного

анализа, в котором устанавливается связь между всеми категориальными фичами, то есть чаще всего с одним объектом идет другой, так можно понять какие товары чаще покупают и другие взаимосвязи (анализ проводился с помощью Apriori с уровнем в 0.4, это помогло установить достаточное количество взаимосвязей)

Аргументируя работоспособность метода использования предсказательной модели покупки для увеличения прибыли, хочу указать какие фичи она использует: id, gender, age, city, education, country, personal_coef, product, colour, cost, product_sex, base_sale, dt, целевой фичей является buy. то есть основываясь на большем разнообразии информации можно предсказать покупку, это повысит лояльность клиента и успешность маркетинговых компаний. Обучалась модель логистической регрессии, так как она лучше показывает себя с бинарной целевой переменной и лучше показала себя на кросс-валидации, модель предсказывает с вероятностью 99,7 на кросс-валидации.

5. Заключение и рекомендации

Проведен большой анализ и создано много инструментов на основе данных. Вот рекомендации бизнес-стратегии: вследствие проведения кластерного анализа, можно сказать, что идеальное количество разбиений на кластеры это 3, то есть при следующих экспериментах делите выборку на 3 кластера. Не применяйте скидку ко всей выборке клиентов, это не дало прироста в выручке, но в группе, где была скидка клиентов больше на 5000, поэтому рекомендую провести A/B тест для проверки следующей гипотезы — Если предложить клиентам скидку в два раза меньше, чем в предыдущем эксперименте, выручка вырастет на 5%.

Сконцентрировать маркетинговую компанию на товарах из ассоциативного анализа. Применять модель для определения склонности к покупке и при прогнозировании успешности дальнейших экспериментов.