# 1. Описание данных

Датасет содержит описание транзакций интернет-магазина, в частности, номер транзакции, номер инвойса, артикул товара по базе, описание товара или услуги, количество единиц товара в инвойсе, дату и время получения инвойса, цену за единицу товара, идентификационный номер клиента, страну, где находится клиент.

# 2. Ограничения данных, аномалии

Имеем 54214 строк данных, при этом в некоторых местах есть пропуски – так, колонка description содержит 54080 ненулевых объектов, а колонка customerID – 40643 ненулелвых объекта. Таким образом, нам не хватает некоторого количества описаний транзакций (что мы продали или какую услугу оказали), а также довольного большого количества информации о клиентах.

Возможно, на сайте есть анализ сделать заказ без регистрации или введения логина, и часть отсутствия данных по клиентам объясняется этим. Часть транзакций с обозначением клиента как NaN, как следует из анализа данных, определяется «жизнедеятельностью» самого магазина – это печать этикеток, спецификаций и прочих бумаг на товары, а также пошлины и тарифы Амазона, поврежденные товары и так далее.

Кроме того, строка «количество» может быть отрицательной – это либо затраты магазина (см. выше), либо возвраты товаров «настоящими» клиентами. Например, описание «wrongly sold as sets» - это расходы внутри непосредственно интернет-магазина, а строка ниже:

| **202** | C536548 | 20957 | PORCELAIN HANGING BELL SMALL | -1 | 12/1/10 14:33 | 1,45 | 12472.0 | Germany |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |

это нормальный возврат обычного клиента.

Если анализировать именно финансовое состояние магазина, конечно, транзакции должны включать все, связанное с деньгами, даже если это не касается продаж напрямую. Для удобства анализа работы магазина, связанной именно с оборотом товаров, такие «внутренние» транзакции лучше исключать из рассмотрения.

# 3. Стейкхолдеры

Стейкхолдерами для данного интернет-магазина являются:

- площадка для размещения заказов (судя по всему, это Amazon). Площадка берет свою комиссию за размещение объявлений

- производители товаров, которые продаются через интернет-магазин. Возможно, у магазина есть какая-то продукция и под собственным брендом, но все, что он покупает у других продавцов, также облагается налоговой ставкой

- логистические службы, ответственные за доставку товаров от производителя на склад интернет-магазина (при наличии) и до покупателя

- возможно, существует склад или распределительный центр. В таком случае, аренда склада и завоз/развоз с него товаров тоже нужно учитывать

- другие интернет-магазины, производящие товары, имеющиеся в ассортименте нашего магазина. Они напрямую конкурируют между собой за внимание и лояльность покупателя

- покупатели

# 4. Система метрик

Для успешного функционирования магазина предлагается оценивать следующие показатели:

- CPC (Cost Per Click) — стоимость за клик. Это сумма, которую тратит рекламодатель за одного пользователя, перешедшего на сайт по рекламному объявлению.

- CTR (Click Through Rate) — соотношение количества кликов на рекламные объявления к количеству их показов. Эта метрика помогает определить эффективность той или иной площадки для размещения рекламы, эффективность рекламной кампании в целом, а также отдельных рекламных объявлений, полезно это отслеживать также и для разных устройств, с которых пользователь заходит на сайт. Так как мы имеем клиентов из 37 стран, то для разных стран могут быть более популярны разные браузеры, разные устройства, разные маркетплейсы.

- CPO (Cost Per Order) — стоимость привлечения одного заказа, т.е. отношение затрат на рекламу или маркетинговую активность к количеству полученных заказов, целевым действием считается именно покупка. Может оцениваться отдельно для разных стран, разных пользователей, разных товаров

- CRR (Cost Revenue Ratio) — это соотношение всех затрат на рекламу и выручки, которые эта реклама принесла.

- Средний чек (AOV — Average Order Value) — это отношение общей стоимости совершенных заказов (т.е. полученного дохода) к суммарному количеству заказов за определенный период. Применяется для прогнозирования дохода интернет-магазина.

Здесь может быть много более тонких исследований – например, вычислять наиболее выгодных клиентов и делать им какие-то мелкие подарки. Также можно оценивать количество товаров в одном инвойсе и/или суммарную стоимость одного инвойса

- Конверсия в покупки (Conversion Rate) — показывает, какая доля посетителей сайта совершила покупку.

- соотношение возвратов к покупкам (в денежном или в штучном эквиваленте)

Позволит оценить, какие товары возвращают.

**-** ARPPU доход на платящего пользователя

- САС Rate (Customer Acquisition Cost) — совокупные маркетинговые расходы на привлечение одного нового пользователя.

- Коэффициент «Брошенных корзин» (Abandoned Basket, Shopping Cart Abandonment) отражает количество посетителей, которые добавили товар в корзину, но не завершили оформление заказа.

- LTV (Lifetime Value) — это показатель совокупной прибыли, которую может принести нам клиент за время взаимодействия с нашим сайтом.

- Retention Rate — коэффициент удержания клиентов.

- Сhurn Rate - показатель «оттока» количества клиентов.

# 5. Рекомендации по модели данных

Для оценки рекомендуемых метрик в датасете не представлена маркетинговая информация – с какого сайта пришел клиент, когда он бы зарегистрирован (чтобы оценить LTV).

Имеем только информацию по совершенным покупкам, но не знаем, сколько всего клиентов проходило мимо магазина, но почему-то не купили. Нужно иметь информацию также и по клиентам, не совершившим покупку (для оценки CR).

Для оценки удержания и оттока также нужна более расширенна информация по клиентам, так как из датасета не представляется возможным оценить retention и отток.

# 6. Выводы и рекомендации для бизнес-заказчика

- разработка рекомендательной системы

Может быть разница в предпочтениях между разными странами относительно товаров, которые там покупают (например, по тепловой карте в прилагаемом ноутбуке видно, что большинство клиентов покупают товары низкой ценовой категории – значит, именно их и стоит показывать)

- работа с возвратами товаров – определить, какие товары возвращают и почему. Возможно, была бракованная партия какого-то определенного товара. Возможно, есть «вредные» клиенты, которые ответственны за основную долю возвратов

- работа с удержанием и оттоком клиентом

Нужно понимать, как долго у нас задерживаются клиенты и что можно предпринять, чтобы они оставались у нас

- бонусы для «хороших» клиентов

Часто 80% заказов делают 20% клиентов (и прибыль мы получаем в основном с них). Для таких клиентов можно предусмотреть скидки, бонусы, приятные мелкие подарки с каждым 10м заказом.

- работа с логистикой

Нужно учитывать, что цена заказа складывается не только из стоимости товара, но и его доставки. Возможно, не во все страны и не для всех клиентов имеется оптимальное по соотношению цена/качества решение по доставке, что может приводить к потере части клиентов

- работа с ценой и средним чеком

Из данных видно, что есть клиенты, которые заказывают много позиций, но каждая из них очень дешевая, а есть клиенты, которые заказывают по одной-две позиции, но при этом дорогих. При этом средний чек может быть одинаковым. Однако, сборка заказа из многих недорогих позиций может занимать больше времени – с этой точки зрения, возможно, стоит отдавать предпочтение (например, делать небольшую скидку), клиентам, которые заказывают немного позиций, но дорогих

- работа с ассортиментом

Видно из данных, что, например, подвеску-держалку в форме сердца купили у нас 243 раза, в то время как зеленые серьги – всего 1 раз.

WHITE HANGING HEART T-LIGHT HOLDER 243

DROP DIAMANTE EARRINGS GREEN 1

Нужно делать анализ, какие товары расходятся лучше всего и корректировать ассортимент в их сторону. Например, если покупают подвеску-держалку в форме сердца, то, возможно, в форме звездочки она тоже будет хорошо продаваться. А серьги, возможно, не наша история и стоит от них отказаться как от категории.

- мониторинг конкурентов

Оценивать доступную информацию по конкурентам.

Понятно, что навряд ли мы сможем оценить CTR/CPO/ARPPU чужого магазина, так как просто не будем иметь всех данных. Но сравнивать цены, ассортимент и удовлетворенность клиентов по отзывам в открытым источникам мы в состоянии.