

# АНАЛИТИЧЕСКИЙ ОТЧЁТ

выполнен на основании данных базы транзакций  
интернет-магазина из Великобритании

## 1 блок: описание данных

В базе содержатся данные по транзакциям интернет-магазина (розничная торговля) за период с 1 декабря 2019г - 9 декабря 2020г (один календарный год).

База содержит данные по следующим наименованиям:

InvoiceNo - № документа/счёта, который выставляется покупателю

StockCode - код товара

Description - наименование товара

Quantity - количество

InvoiceDate - дата и время выставления (оплаты) счёта + время (часы и минуты).

Дата в формате: ММ/ДД/ГГ, время: ЧЧ:ММ

UnitPrice - стоимость за единицу товара

CustomerID - код покупателя

Country - страна осуществления сделки

Для того, чтобы сделать выводы по качеству и полноте данных, необходимо определить типы данных и соответствие, их объём, наличие / отсутствие аномалий.

InvoiceNo - категориальный тип данных (текст). Может состоять из цифр и в некоторых случаях цифр и букв. Пропусков данных нет.

StockCode - категориальный тип данных (текст). Состоит из цифр и/или букв. Пропусков данных нет.

Description - текст, категориальный тип данных, содержит описание, наименование товара.

Считаю, что здесь пустые значения нельзя признать выбросами, так как отсутствие в данных наименования товара не скажется на выводах и результатах анализа (при прочих присутствующих данных по этому товару).

В наших данных 134 отсутствующих значения в Description из общего количества 54 214 значений. Во всех этих 134 отсутствующих значениях Description, по столбцу UnitPrice соответствующее значение = 0, а по столбцу CustomerID значение также отсутствует.

Quantity - количественный (числовой) тип данных, завершённый тип.

Пустых значений нет, =0 значений нет.

Но есть отрицательные значения. Их 1 083 из общего количества 54 214. Предположим, что это возвраты товара. Есть отрицательные значения Quantity, где соответствующее значение по столбцу CustomerID отсутствует, можно предположить, что это списание товара. В любом случае данные отрицательные значения необходимо признать аномалиями, выделить их в отдельную таблицу с данными и изучать отдельно. Для начала необходимо определить природу их возникновения.

InvoiceDate - тип данных - временные ряды, содержат данные о дне и времени выставления счёта в формате ММ/ДД/ГГ ЧЧ:ММ.

Пустых значений нет, = 0 значений нет, все данные содержатся в верном формате.

При более подробном изучении видим некоторое количество пропусков по дням. Пропуски приходятся на выходные праздничные дни (например с 24.12.2019 - 3.01.2020), а также еженедельно 1 день в неделю из 7. Необходимо выяснять природу возникновения

данных пропусков, а именно - были ли продажи/возвраты/списания в эти дни, либо это действительно были выходные дни, когда интернет-магазин не осуществлял никакой деятельности.

Для изучения метрик, данные по датам можно сгруппировать по дням. Для обзора картины в целом, в разрезе года, их можно сгруппировать по месяцам.

UnitPrice - тип данных - продолжительный числовой.

Пропусков нет, но есть 225 из 54 214 значений, где UnitPrice = 0 (в них во всех данные по CustomerID отсутствуют). Можно признать эти данные аномалиями, так как для анализа объёма и сумм продаж, стоимость товара не может быть равна 0.

CustomerID - категориальный тип данных.

Значения отсутствуют в 13 571 из 54 214 случаев. Необходимо выяснять природу возникновения таких пропусков. Возможно они были не заполнены по ошибке. UnitPrice = 0 всего в 221 значении из 13 571 пропусков, во всех остальных есть и Quantity и UnitPrice и Description. Игнорировать их при анализе нельзя.

Country - текстовый тип данных, содержит наименование страны выставления счёта.

Пропусков нет, все значения со странами.

Для изучения метрик, можно сгруппировать данные по странам.

## **2 блок: описание применённого подхода к анализу данных**

В представленных данных содержатся некоторые аномалии и пропуски, ниже будет описано как выстроена с ними работа в данном аналитическом отчёте.

1) В столбце Quantity есть отрицательные значения. Их 1 083 из общего количества 54 214. Это могут быть например возвраты товара / либо списание товара. Чтобы нам сформулировать верные выводы именно по объёму проданного в 2020 году товара, данные надо разделить на положительные и отрицательные значения и изучать отдельно.

Данные с отрицательным значением, которые мы признали аномалиями, также необходимо отдельно проанализировать, но для этого нужно точно установить природу их появления.

В настоящем аналитическом отчёте используем только данные с положительным Quantity, их 53 131 из 54 214. Это 98% от общего объёма предоставленных данных. Получается 2% с отрицательным значением Quantity мы выделили для отдельного изучения.

2) Удаляем для анализа данные, где UnitPrice = 0 (их 225 значений). Это 0,3 % от общего объёма данных.

Есть часть данных, где Quantity с отрицательным значением и одновременно UnitPrice = 0.

После удаления из данных для анализа отрицательных значений Quantity и значений, где UnitPrice = 0, получаем 53 029 значений из общего числа 54 214. Это 97,8 % от общего количества.

Пропуски в данных по столбцу CustomerID составляют 13 571 из 54 214. Это 25% от общего количества. Это очень большой процент, данные оставим для анализа, как есть.

Пропуски в данных по столбцу InvoiceDate отсутствуют в предоставленном датасете. Этих данных либо совсем нет, если эти дни действительно были выходными в онлайн-магазине и никакая деятельность не осуществлялась. Либо вышеуказанные даты пропущены по ошибке, и продажи совершались в пропущенные даты, тогда их отсутствие может исказить полученные выводы.

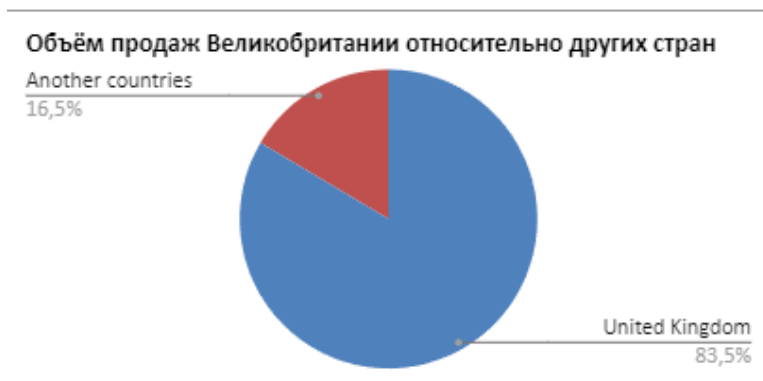
### 3 блок: описание бизнес-заказчика, его гипотетических болей и целей развития бизнеса

Бизнес-заказчик - интернет-магазин из Великобритании по розничной торговле различных предметов бытовой принадлежности.

Более 80% покупателей из Великобритании. Остальные - из Франции, Германии, Испании, Австралии, Нидерландов, Ирландия и др.

Страна	Объём выручки
United Kingdom	158 970,65
EIRE	6 284,64
Netherlands	5 968,12
Germany	4 252,50
France	2 734,99
Australia	2 575,32
Spain	1 485,53
Sweden	1 478,56
Japan	958,04
Portugal	780,51
Channel Islands	722,45
Switzerland	623,80
Cyprus	454,10
Belgium	409,03
Italy	371,05
Norway	357,43
Hong Kong	320,40
Greece	314,10
Lithuania	291,60
Poland	266,46
Singapore	186,54
Denmark	148,68
Iceland	122,75
Austria	72,20
Israel	66,70
United Arab Emirates	61,68
Finland	46,80
Lebanon	39,60
<b>Итого</b>	<b>190 364,23</b>

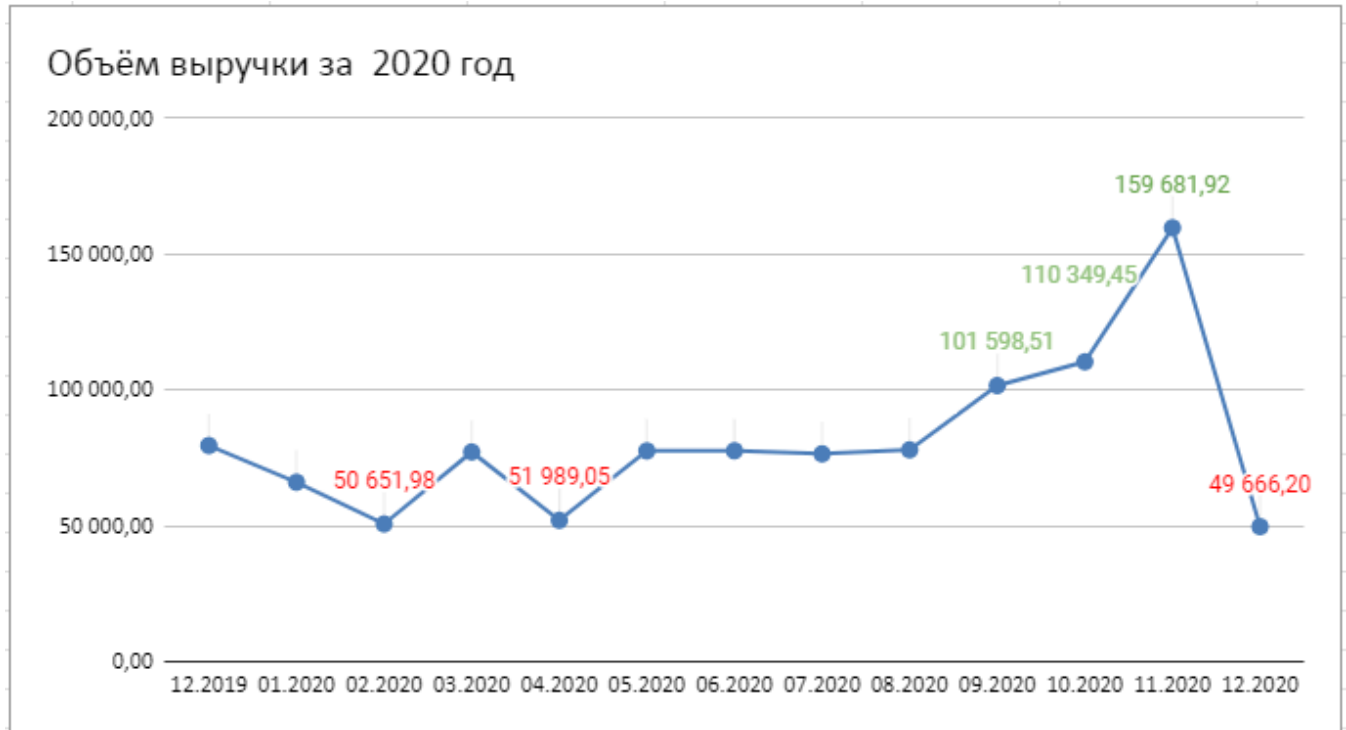
United Kingdom	158 970,65
Another countries	31 393,58



1) Есть данные о продажах интернет-магазином в другие страны. Значит, у бизнес-заказчика есть задача расширять территориальные границы. При этом мы видим, что доля в продажах в других, кроме Великобритании стран - всего 16,5%. Следовательно, необходимо эту долю увеличивать, искать пути выхода на рынки других стран, вкладываться в рекламу. И главное, в настоящий момент, необходимо разработать и использовать систему метрик для контроля всего пути клиента, от просмотра рекламы до перехода на сайт и совершения покупки.

2) Все компании стремятся к постоянному ежемесячному росту выручки. У нашего бизнес-заказчика мы видим постоянную ежемесячную выручку в районе 75 000 - 80 000, три точки спада - в феврале 2020г, в апреле 2020г и в декабре 2020г., и резкий рост с сентября по ноябрь 2020г (более чем в два раза), а затем снова спад в декабре 2020г.

В связи с этим можно сделать вывод о возможной сезонности продаваемых товаров. Необходим более тщательный детальный анализ клиентского поведения и продаваемых товаров: на пике роста и в точках спада.



#### 4 блок: стейкхолдеры отчёта и выбранные метрики (система метрик)

В качестве стейкхолдеров данного отчёта можно рассматривать лиц, занимающихся продажами, маркетингом в компании, и кто имеет полномочия на принятие решений в соответствующих областях.

Возможно стейкхолдером будет сам собственник(и) бизнеса, если размер компании небольшой и собственник сам выполняет данный функционал.

На основе данных, имеющихся в предоставленном датасете, были разработаны и использованы в отчёте следующие метрики.

1) Выручка, полученная по каждой позиции товара по каждому счёту, выставленному покупателю.

Рассчитана по формуле:  $PRICE = UnitPrice * Quantity$

2) Объём выручки, полученный по каждой из стран, в которые осуществлялись продажи.

Получен путём суммирования продаж (PRICE) по каждой стране за 2020 год.

3) Количество товара, проданного в каждой из стран.

Получено путём суммирования количества товара (Quantity) по каждой стране за 2020 год.

4) Объём выручки, полученный в целом за каждый месяц 2020 года.

Получен путём суммирования продаж (PRICE) по месяцам.

5) Количество товара, проданного за каждый месяц 2020 года.

Получено путём суммирования количества товара (Quantity) по месяцам.

6) Объем продаж с разбивкой по каждому в отдельности товару.

Данная метрика позволит нам выявить самые популярные, продаваемые товары, на которые стоит делать акцент. А также самые непродаемые товары, от которых возможно отказаться.

Получен при помощи суммирования количества товара (Quantity) по каждому коду товара (StockCode) за 2020г.

Самые продаваемые/популярные товары:

Код товара	общее количество
21212	5241
85099B	5129
84879	4447
84077	4276
21915	3834
15036	3777
85123A	3432
16014	3075
22386	3052
22492	2960
22197	2936
84568	2768
17003	2631
18007	2439
22595	2243
22355	2189
23199	2166
23084	2165
23203	2127
20724	2116
85099C	2111
21977	2044
22961	2032
84946	2027
22086	2007
20725	2001

7) Можно объем продаж по каждому из товаров посмотреть с дополнительной разбивкой по странам.

Это необходимо для того, чтобы понять в каких странах какие товары наиболее / наименее популярны. В разных странах могут различаться реализуемые товары.

8) Сумма выручки, с разбивкой по покупателям.

Позволит выяснить кто из покупателей совершил у нас покупки в 2020 году на наибольшую сумму.

Считаем как сумму продаж (PRICE) по каждому из покупателей (CustomerID).

Самые крупные покупатели (по показателю суммы продаж):

CustomerID	Сумма его покупок		
18102	26 533,44	12931	6 404,12
14646	25 519,16	17857	6 259,50
17450	23 000,10	13089	6 190,59
12415	16 102,89	14088	5 660,96
14911	15 238,45	15061	5 630,52
14156	12 092,71	14298	5 343,62
17949	12 048,08	16422	4 891,05
17511	9 534,21	17404	4 398,70
14096	8 487,80	17841	3 913,63
15769	8 129,04	13798	3 892,80
13694	7 722,13	15195	3 861,00
16684	7 512,92	16333	3 545,20
15311	7 242,23	14680	3 447,65
16029	6 739,52	17389	3 424,86
		12748	3 220,03

9) Количество товаров, купленных каждым покупателем.

Позволит выяснить кто из покупателей совершил у нас покупки в 2020 году в наибольшем количестве.

Считаем как количество товаров (Quantity) по каждому из покупателей (CustomerID).

Самые крупные покупатели (по показателю количества приобретенных товаров):

CustomerID	Количество товаров
14646	20333
12415	9565
14911	8448
17450	8036
17511	6998
13694	6893
18102	6587
17949	5528
14298	5404
16422	4919
14156	4919
15311	4749
16684	4560
15769	4536
17857	4251
17404	4195
17381	3976
12931	3886
16029	3412
13089	3229
16333	3112

10) Из предыдущей метрики можно выделить всех покупателей, которые в 2020 году совершали покупки в интернет-магазине более 1 раза.

Это те, кому понравилось. Их можно считать постоянными покупателями и для повышения их лояльности к компании необходимо разрабатывать для них систему скидок, различных бонусов и акций.

11) Если мы предполагаем, что есть определённая сезонность в продажах интернет-магазина, то можно отслеживать метрику: в какие месяцы какие категории товаров пользуются наибольшим спросом.

Для этого необходимо вывести количество (Quantity) конкретных проданных товаров (StockCode) с разбивкой по месяцам (InvoiceDate).

Можно повысить выручку от продаж, увеличив в нужные месяцы объём товаров, которые пользуются спросом (сезонные товары).

Если интернет-магазин не отслеживает никаких других (кроме имеющихся в предоставленном датасете) метрик, показателей, например, характеристики клиентов, их путь до покупки, статистики по разным каналам сбыта (если кроме сайта интернет-магазина есть другие), затраты на товары, размещение, доставку и проч., зарплата сотрудников, хозяйственные затраты и др., предлагаю внедрить и использовать очень удобную систему метрик AARRR.

AARRR проста в использовании и отлично подходит для диджитал продуктового мира (в нашем случае интернет-магазин розничной торговли).

AARRR:

Acquisition	Привлечение
Activation	Активация
Retention	Удержание
Revenue	Доход
Refferal	Рекомендация

Данная система позволяет отследить все необходимые метрики на каждом этапе жизненного цикла клиента, а также посчитать конверсию клиентов по каждому этапу воронки.

## **5 блок: рекомендации по модели данных бизнес-заказчика**

В связи с вышеуказанными рекомендациями по системе метрик AARRR, считаю важным отслеживать следующие метрики.

### **1 этап : привлечение аудитории на сайт интернет-магазина: метрики**

- 1) Характеристика приходящей аудитории, её интересы.
- 2) Количество переходов с канала на сайт за определенный промежуток времени (день, неделя, месяц).
- 3) Количество конверсий за определенный промежуток времени (день, неделя, месяц).
- 4) Стоимость одного перешедшего на сайт пользователя с этого канала.

Берём число переходов на сайт с определённого канала и делим на затраченные деньги на этот канал.

### **2 этап : активация пользователей на сайте: метрики**

- 1) Число тех, кто прошел регистрацию на сайте.
- 2) Число тех, кто воспользовался поиском товаров по каталогу.
- 3) Число переходов по вкладкам сайта/по категориям товаров.
- 4) Сколько человек попыталось оплатить корзину в личном кабинете.
- 5) Число просмотров определенных страниц магазина и конкретного товара и т.п.

6) Отношение числа пользователей, совершивших активацию к общему числу пользователей, перешедших на сайт (конверсия).

### **3 этап : удержание посетителей сайта: метрики**

1) Какими путями и как часто пользователи возвращаются.

2) Сколько времени они проводят на сайте и как распределяется их активность в течение суток, недели и месяца.

3) Отношение числа пользователей, кто вернулся на сайт (побывал на нем более 1 раза) к числу пользователей, перешедших на сайт единожды (конверсия).

### **4 этап : доход от сайта: метрики**

1) Количество пользователей сайта, которые приобрели продукт.

2) Сумма выручки от продаж с сайта.

3) Средний чек на одного покупателя.

4) Прибыль от работы сайта: сумма выручки от продаж за месяц минус сумма затрат на привлечение потока клиентов за месяц.

5) Отношение числа пользователей, кто приобрёл товар на сайте, к общему числу пользователей, активировавшихся на сайте (конверсия).

### **5 этап : рекомендация нашего сайта друзьям / знакомым: метрики**

1) Отношение числа пользователей, кто рекомендовал наш сайт друзьям / знакомым, к общему числу пользователей, активировавшихся на сайте (конверсия).

## **6 блок: выводы и рекомендации для бизнес-заказчика**

1) Предоставленный бизнес-заказчиком датасет содержит некоторые аномалии и пропуски, подробно описанные в настоящем отчёте. Для более глубокого погружения в анализ данных необходимо изучить природу возникновения пропусков и аномалий, отрицательных значений в столбце Quantity, и проанализировать их отдельно.

2) Если у бизнес-заказчика есть цель: увеличение объема продаж товаров в другие, кроме Великобритании страны, то необходимо искать пути выхода на рынки других стран, вкладываясь в рекламу. И главное, в настоящий момент, необходимо разработать и использовать систему метрик для контроля всего пути клиента, от просмотра рекламы до перехода на сайт и совершения покупки (в разрезе каждой отдельно взятой страны).

3) Бизнес-заказчику необходим детальный продуктовый анализ, если действительно есть проблема с сезонностью продаваемых товаров. Необходим более тщательный анализ клиентского поведения и продаваемых товаров: на пике роста и в точках спада.

4) В сложившейся ситуации стейкхолдерам в настоящем отчёте была предложена для внедрения и развития система метрик AARRR. Она позволит интернет-магазину отследить все необходимые метрики на каждом этапе жизненного цикла клиента, а также посчитать конверсию клиентов по каждому этапу воронки. Что в свою очередь сделает работу компании более клиентоориентированной, и следовательно поспособствует увеличению выручки.