

**Пояснительная записка к финальной работе по курсу АИС**  
**Денис Иванов**  
**13 сентября 2020**

**Первичное исследование полей датасета:**

- InvoiceDate**
- дата покупки, формат даты,
  - данные представлены в полном объеме . пропусков нет
  - период с декабря 2010 года по декабрь 2011
  - данные за декабрь 2011 приведены только за 2 недели . несопоставим с другими месяцами
  - поле потребуется для группировки продаж по неделям, месяцам, кварталам и т д/
  - возможно углубить анализ в разрезе дней недели
- CustomerID**
- Идентификатор уникального покупателя
  - данные представлены только по 3777 уникальным Покупателям, по остальным записям данные пропущены
- Допущение:
- Покупатели, по которым указан ID это покупатели зарегистрированные на сайте, по которым в распоряжении Интернет магазина есть дополнительная информация/ Анкета
  - Покупатели, по которым НЕ указан ID это покупатели с улицы - осуществившие " покупку без регистрации" / без Анкеты или не оставившие (не разрешившие Интернет магазину пользоваться) свои персональные данные.
- Установить, что такой покупатель пришел повторно - невозможно.
- Покупки такого покупателя ограничены одним уникальным InvoiceNo
- Логика преобразования признака CustomerID:
- всем полям CustomerID с отсутствующими значениями, соответствующим уникальному InvoiceNo, присваивается свой новый код в интервале значений ~(0:1500)
  - вводим дополнительный (бинарный) показатель Customer\_reg, отражающий различие: была ли покупка осуществлена зарегистрированным покупателем или покупателем с улицы
- Country**
- Идентификатор страны происхождения Покупателя
  - все поля заполнены пропусков нет
  - ~90% продаж это Великобритания, остальные продажи распределены между большим количеством стран, каждая из которых не существенна в объеме продаж и статистически не значима
- Логика преобразования признака Country:
- вводим дополнительный признак, используемый в анализе . отчетности, отражающий различие: продажи внутри страны (Великобритании) и продажи за рубежом (в целом)
- InvoiceNo**
- StockCode**
- Идентификаторы номера счета и кода товара
  - все поля заполнены пропусков нет
  - одному счету Invoice могут соответствовать один или несколько товаров Stock
  - каждая строка соответствует позиции товара, если в счете несколько товаров то номера Счета повторяются в нескольких последовательных записях товаров, соответствующих одному счету
  - Дополнительных преобразований нет

Quantity	<p>Количество единиц товара в счете  ~ по 1000 записей значение по полю отрицательно, при том, что отрицательные значения не правдоподобны (отрицательное значение по модулю может быть больше чем количество товаров того же кода проданных за период наблюдений)  Возможно это возвраты, с неполной / искаженной информацией. Как к ним относиться не знаем  Записи с отрицательными значениями по количеству товара исключены из исследования</p>
UnitPrice	<p>- цена за единицу товара  - все поля заполнены пропусков нет  - части записей соответствует значения = 0,  записи UnitPrice = 0 соответствуют (как правило) НЕзарегистрированным покупателям , это точно не факт состоявшейся покупки  Записи с нулевыми значениями по цене товара исключены из исследования</p>
Description	<p>- описание товара  количество товаров с разным описанием &gt; 3000, наблюдение по каждому уникальному описанию статистически не значимо  - анализ (наиболее характерных) отдельных слов в составе описания товара не позволяет выделить товарную группу  Поле из анализа исключено. совсем.</p>

### **Ограничения, выявленные в ходе работы с датасетом;**

- для маркетингового анализа - нет анкеты клиента
- период наблюдений - 1 год, не видно сезонности

Описание бизнес-заказчика и его гипотетических болей и целей развития текущего бизнеса

исследование производится для функционала:	Маркетинговый аналитик
в рамках достижения цели:	- увеличение количества клиентов удержание клиентской базы
для выявления точек роста в рамках	структуры клиентской базы, в разрезе потребительских предпочтений и опыта управления программами лояльности
а еще ставил цель:	
разработка масштабируемой (на вырост) системы метрик, характеризующих и влияющих на период \продолжительность жизненного цикла Покупателя (определение момента прекращения жизненного цикла - ухода покупателя (в пределах заданного набора входящих данных)	

В рамках представленной задачи представляется следующая система метрик

Метрики текущего поведения Покупателей:

- Количество активных покупателей за период
- Средняя сумма, потраченная активным покупателем за период
- Среднее количество товарных позиций, приобретенных активным покупателем за период

Средняя цена товарной позиции (сегментация отчета по группам товарных позиций от стоимости)

Метрики опыта Покупателей / накопленной лояльности

показатели в отношении всех Покупок, совершенных Покупателем до периода наблюдения, необходимые для понимания на каком этапе жизненного цикла Покупателя совершаются покупки, Все потребители за отчетный месяц сегментируются по группам в разрезах:

Количество покупок (покупали часто / средне / редко)

Сумма покупок (тратили на покупки много / средне / мало)

Средняя сумма чека (покупали за раз помногу / средне / помалу)

Далее сумма покупок за месяц (динамика помесечно) может быть разбита, в плоскостях:

- любой показатель (социального или поведенческого) профиля Покупателя
- показатель накопленной лояльности

и

точки усиления мероприятий по усилению лояльности тут и проявятся (возможно)

Метрика стабильности / регулярности взаимодействия с Покупателем

Количество месяцев (недель) в которые Покупатель был активен за период наблюдений (в условиях малых данных датасета производился расчет количества активных месяцев за последние 3 месяца до даты отчета)

Метрики оценки ушедших Покупателей - Метрика стабильности взаимодействия с Покупателем (или иная качественная метрика) со знаком минус зафиксированная в момент ухода (месяц отсутствия) Покупателя

## Дашборд

<https://datastudio.google.com/reporting/1dd6056f-01ed-4066-8762-122dcf2baa6d>

Дашборд предназначен для квалифицированного пользователя - Аналитика, задача которого разобраться, какова структура покупателей, прекративших покупать в Магазине ( в прошлом месяце покупка была, в исследуемом месяце - нет), Насколько выбывшие Покупатели ценны для Магазина, Как влияют структурные факторы оттока Покупателей на выручку Магазина. Расширение дашборда - добавить данные Анкеты Клиента - Модель оценки влияния отдельных факторов на решение о прекращении покупок

Структура дашборда:

Фильтры представления данных:

Customer\_reg :

1 - информация в Дашборде представлена только по зарегистрированным Покупателям (CustomerID проставлен в первичном датасете), предполагается, что это более лояльные Клиенты, уже зарегистрированные на сайте и предоставившие информацию о себе, по которым можно фиксировать повторную покупку

0 - Покупатели с улицы

Sales type:

Domestic sales - продажи в Великобритании

Foreign sales - продажи за рубежом - в любой другой стране

Заголовок :

Ушло покупателей за месяц - количество покупателей, которые делали покупки в предпоследнем месяце и не делали покупок в месяце наблюдений  
Потери выручки: выручка за последний месяц наблюдений, поступившая от ушедших покупателей

Раздел: Базовые цифры

расчет доли покинувших покупателей и утраченных продаж в динамике ежемесячно.  
представлен расчет и общий тренд по показателям:

- Customers\_away\_% - ушло клиентов %
- Sales\_away\_% - потери продаж (резерв за счет укрепления лояльности) %

Раздел: Распределения покинувших клиентов (графики слева)

В основе раздела представлена динамика распределения ключевых показателей, поставленных в заглавие дашборда:

- сумма текущих потерь / утерянной выручки (ежемесячно) - графики слева
- количество ушедших клиентов - графики справа

Распределения представлены в отдельных графиках в разрезе показателей накопленной лояльности, разбитых по группам:

- LTVyruchka - выручка за период сотрудничества с клиентом (~ LTV)
- LTUnitPrice - ценовая группа товара, наиболее характерная для ушедшего клиента
- LTPeriod - период сотрудничества с клиентом в месяцах
- перечень разрезов распределений расширяем в зависимости от задач исследования и наличия данных

Раздел: Живые примеры

Представлены реальные кейсы отдельных ушедших клиентов, расположенных в порядке убывания потерь за месяц, в разрезе сочетания (групп) отдельных показателей накопленной лояльности, в числе которых дополнительно:

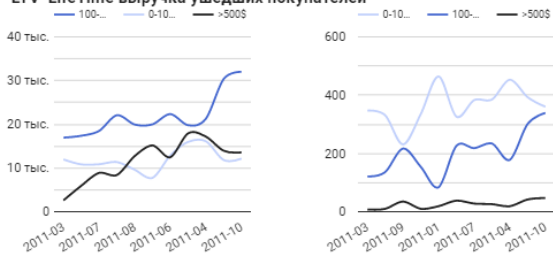
- среднее количество покупок за lifetime,
- средняя сумма чека

Раздел предназначен для более глубокого эмпирического анализа сочетаний параметров, на которых можно построить детализацию анализа и выводов.

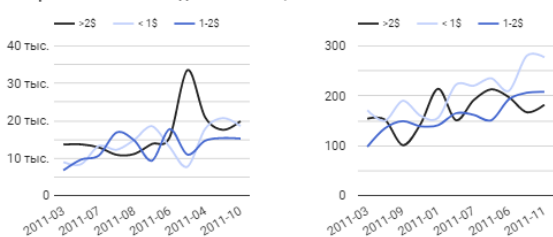
Дашборд качественно расширяем, если к метрикам накопленной лояльности добавить социальные и поведенческие метрики

Распределение покинувших клиентов по сумме текущих потерь / по количеству клиентов

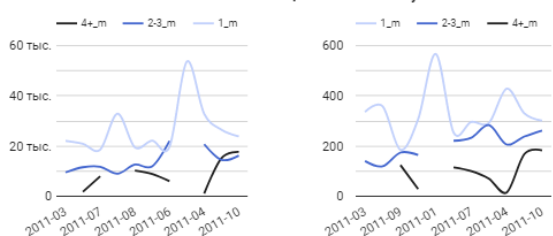
LTV LifeTime выручка ушедших покупателей



LTUp Life UnitPrice - Медиана UnitPrice,



LTP Life Time Period - количество месяцев активных покупок



Ушло покупателей за месяц 736

Потери выручки: 56 131

Sales\_type

Customer\_reg

Базовые цифры

	Period	Customers	Customers_away	Customers_away_%	Sales_gross	Sales_customers_away	Sales_away_%
1.	2011-06	861	645	74.9	77615.4	47665.6	61.4
2.	2011-07	839	629	75	76444.3	38219.5	50
3.	2011-08	770	590	76.6	77937.6	42277.8	54.2
4.	2011-09	1094	483	44.1	101598.5	42899	42.2
5.	2011-10	1178	746	63.3	110349.4	57700.9	52.3
6.	2011-11	1436	736	51.3	159681.9	56130.5	35.2

1 - 6 / 6

Живые примеры

	CustomerID	LTAOVg	LTAUPg	LTPg	LTVg	LTAOV	LMTV
1.	17450	1	1	2	1	4 282,67	9 807
2.	15769	1	1	2	1	3 577,51	5 869,8
3.	378	1	1	1	1	4 350,37	4 350,4
4.	17857	2	2	2	2	2 722,08	4 073,5
5.	12931	1	1	1	1	1 806,28	3 612,6
6.	12415	1	1	1	1	1 710,48	2 931,2
7.	18102	1	1	2	1	3 326,8	2 753,1
8.	53	1	1	1	1	2 645,7	2 645,7

1 - 100 / 4188

Распределение потерь за последний месяц по среднему чеку

Потери выручки / потери покупателей



50+ \$

20-50\$

0-20\$



20-50\$

50+ \$

0-20\$