

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение высшего образования
«Тольяттинский государственный университет»

Институт математики, физики и информационных технологий
(наименование института полностью)

Кафедра Прикладная математика и информатика
(наименование)

09.03.03 Прикладная информатика
(код и наименование направления подготовки / специальности)

Бизнес-информатика
(направленность (профиль) / специализация)

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА (БАКАЛАВРСКАЯ РАБОТА)

на тему Разработка приложения для оценки лояльности клиентов

Обучающийся

К.А. Соколов

(Инициалы Фамилия)

(личная подпись)

Руководитель

канд. пед. наук, доцент, О.М. Гущина

(ученая степень (при наличии), ученое звание (при наличии), Инициалы Фамилия)

Тольятти 2023

Аннотация

Выпускная квалификационная работа (ВКР): 81 стр., 5 таблицы, 25 иллюстраций, 28 использованных источника, 2 приложения.

Организации торговли накапливают огромные объемы информации о потреблении и могут использовать ее для продвижения товаров. Технические достижения и доступность методов интеллектуального анализа данных позволяют делать это. Если для достижения хороших результатов маленькие компании могут обходиться без сложных инструментов, то крупным фирмам необходимо активно использовать методы data mining. Это объясняется тем, что в их случае только простые, примитивные способы учета потребностей клиентов недостаточны. Всё сказанное свидетельствует о том, что тема ВКР К. А. Соколова актуальна.

Структура работы определяется ее актуальностью, целью, задачами, объектом, предметом и методами исследования. Работа состоит из трех логически связанных глав, введения, заключения и списка использованной литературы. Главы разделены на параграфы, что позволяет сосредоточиться на отдельных проблемах в рамках конкретного вопроса.

Значимость исследования заключается в том, что результаты, полученные путем анализа научной и практической литературы и изучения современных подходов к определению и анализу лояльности, а также обобщения теоретических основ разработки систем анализа лояльности, были доведены до уровня практического применения: модель лояльности и основные идеи, лежащие в основе систем сегментации клиентов; разработана методология построения приложений для анализа лояльности клиентов. Получившееся приложение, построенное на основе RFM-анализа и машинного обучения, было протестировано. Теоретические результаты включают обобщение анализа состояния проблемы разработки таких систем и изучение методов анализа и возможности сегментации потребителей.

Оглавление

Введение.....	4
Глава 1 Теоретические основания применения методов для повышения лояльности клиентов.....	7
1.1 Понятие и сущность лояльности клиентов	7
1.2 Популярные инструменты и платформы.....	18
Глава 2 Проектирование приложения для оценки лояльности клиентов	26
2.1 Анализ лояльности клиентов на примере Direct Marketing.....	26
2.2 Методы определения лояльности клиентов	31
2.3 Использование алгоритмов машинного обучения для определения лояльности	39
Глава 3 Разработка приложения для оценки лояльности клиентов.....	46
3.1 Разработка приложения RFM-сегментирование	46
3.2 Сегментирование клиентов методом K-Means	60
Заключение	68
Список используемой литературы и используемых источников.....	70
Приложение А Макросы Excel	73
Приложение Б R-код.....	81

Введение

Различные исследования посвящены изучению лояльности клиентов. В рамках таких исследований были разработаны различные методы оценки лояльности. Как показывают статистические данные, важность лояльности в качестве фактора конкурентоспособности имеет большое значение. Глава программы поддержки бизнеса Bain Company и автор книги «Эффект лояльности» Ф.Ф. Райхельд заявил, что высокий уровень доверия в к компании увеличивает её экономическую деятельность на 25-50% или даже больше. Из-за увеличения коэффициента удерживания постоянных покупателей на 5% стоимость их покупок увеличивается на 25-100% [16]. Большинство компаний увеличивают свой доход от клиентов с расширением сотрудничества между ними и происходит это по мере расширения взаимодействия. Чтобы компенсировать потери от одного ушедшего старого клиента, обычно необходимо привлечь несколько новых клиентов.

Сегодня почти любая компания стремится к созданию собственных программ лояльности для своих клиентов. Программа лояльности - маркетинговый инструмент, направленный на улучшение отношений между клиентами и компанией. С другой стороны, программа лояльности требует обработки большого количества данных и тонкого настроя под клиента. Причем, далеко не всегда понятно, что именно анализировать и каким образом. Например, какие данные использовать для анализа имеющихся программ привлечения клиентов и какая связь между программой и изменениями товарооборотов, объемов продаж и пр. Кроме того, любая программа лояльности требуют учета разнородной информации, нередко с помощью методов data mining (интеллектуального анализа данных) [17].

Все компании признают, что постоянные клиенты являются для них приоритетом. Как представители компаний, продающих массовую продукцию, так и поставщики банковских услуг, стараются внедрить программу лояльности. Как считают специалисты в области маркетинга, это

объясняется тем фактом что «старые» клиенты имеют большую ценность для компании, по сравнению и новыми. Бизнес может сэкономить деньги как на рекламе, так и на различных акциях, для продвижения товаров и услуг, и расходах на привлечение новых клиентов [1][2].

Для продвижения товаров предприятия торговли могут использовать огромные объемы информации о клиентах и структуре их потребления, которые ни накопили. Сегодняшние технологии позволяют проводить интеллектуальный анализ данных. Для небольших компаний не требуется сложных инструментов анализа, а для крупных организаций необходимо использовать методы интеллектуального анализа. Дало в том, что для них простые способы учета потребности клиентов недостаточно эффективны. Всё сказанное свидетельствует о том, что тема выпускной квалификационной работы актуальна [25][27][8][24].

Объектом данной работы являются различные подходы к определению лояльности клиентов, которые лежат в основе маркетинговых исследований.

Предметом исследования является методология определения лояльности клиентов и ее программная реализация.

Цель исследования: разработать приложение для определения лояльности клиентов, основанное на эффективной системе сегментирования клиентов.

Для достижения поставленной цели необходимо решить задачи:

- Изучить вопросы, связанные с понятием лояльности клиентов.
- Рассмотреть различные методы определения лояльности. Описать методики сегментирования клиентов.
- Определить показатели лояльности клиентов: частота покупок, объем заказов, средний чек, продолжительность отношений с клиентом и др.
- Изучить возможности использования машинного обучения в сегментировании клиентов.
- Разработать приложение для определения лояльности клиентов.

Теоретические и методологические основы исследования. В ходе исследования изучены труды российских и зарубежных ученых в области определения лояльности клиентов и их сегментирования.

Практическая значимость:

- Проработана методика определения степени лояльности клиентов;
- Спроектированная и протестированная приложение по определению лояльности клиентов на основе RFM–метода и K–Means алгоритма.

Структура ВКР определена её целью и задачами. Работа состоит из введения, трех глав, выводов по главам, заключения, списка использованных информационных источников.

Глава 1 Теоретические основания применения методов для повышения лояльности клиентов

1.1 Понятие и сущность лояльности клиентов

С увеличением количества постоянных покупателей растет и прибыль компании, и чтобы добиться этого, организации необходимо укреплять лояльность покупателей.

Лояльность – это позитивное отношение покупателей и сотрудников к компании, магазину или бренду в результате положительного опыта, который был получен в результате проведения мероприятия повышения лояльности.

Основа лояльности – усилия компании, её осознанные и целенаправленные действия в отношении клиентов.

Основой мотивации для клиентов, которые лояльны, является чувство причастности к «избранным». Необходимо, чтобы покупатель был близок к ценностям компании. У него должно быть чувство, что его слушают и обслуживают максимально хорошо, так, как он привык [10].

Одним из самых важных источников дохода компании являются покупатели, лояльно относящиеся к бренду. Несмотря ни на что он возвращается за покупкой даже при повышении цены и более выгодной ситуации у конкурентов. В случае доверия бренду, его выгода больше потенциальной выгоды от более низких цен или выбора других продуктов. Как правило, в среднем по рынку средний чек новых клиентов ниже, чем у постоянных покупателей. Много лояльных покупателей это: стабильные доходы компании и более низкая цена продукции; быстрая ответная реакция; готовность тестирования новых продуктов и, возможно самое главное, поддержка имиджа организации в соцсетях [12]. Постоянных клиентов можно исследовать на их общие черты, а затем целенаправленно искать таких же постоянных клиентов [1].

При правильном определении уровня лояльности покупателей компания может правильно воздействовать на них и улучшать их мнение о себе. При этом фирма получает лояльного покупателя, который будет полностью удовлетворен уровнем сервиса.

Выделяют следующие виды лояльности [1][2]:

– Отсутствующая лояльность. Потребители знают о компании и ее продуктах, но не используют их для взаимодействия. В этом случае клиент не меняет свои привычки и образ жизни и относится лояльно и к другим брендам. С таким покупателем сложно договориться, так как он не всегда понимает, чего от него хотят. Он не поддается на уговоры, поэтому лучше не вступать в дискуссию с таким клиентом.

– Отрицательная лояльность. Покупатель столкнулся с негативными проявлениями в компании и остался недоволен качеством купленного товара и/или оказанной услуги. Он может начать распространять негативную информацию о компании. Выясните, что именно ему не понравилось. Такие "критики" помогут понять слабые места продукта и/или недочеты сервиса и помочь компании исправить недостатки. Если проблему сразу не решили и даже не попытались её решить, то шанс возврата такого покупателя практически отсутствует. Может оказаться, что под маской "критика" скрывается клиент с отсутствующей лояльностью. Это может быть как клиент конкурента, так и просто «капризный» человек. Важно отделять таких токсичных клиентов от настоящих критиков. Получать от них обратную связь нет смысла, но необходимо быть вежливым и честным, чтобы не оттолкнуть других потребителей.

– Ложная лояльность. Покупатель вновь и вновь возвращается для того, чтобы повторно приобрести товар или получить услугу. В этом нет ничего плохого, но это еще ничего не говорит про его приверженность бренду. Он может приобретать товары в нескольких компаниях и пользоваться услугами различных организаций. Появиться ложная лояльность может по разным причинам. Например, если продавец не имеет

конкурентов или имеет хорошие сезонные скидки. К таким клиентам относятся и те, кто не может определиться с выбором товара или услуги. Если у вас есть конкуренты с более интересными условиями, то они могут перейти к ним и даже стать их клиентами. Скорее всего, если компания поменяет место дислокации, переехав на новый, более удобный для нее адрес, такие клиенты не последуют за ней. Потребуется произвести исследование, чтобы выяснить, что вызывает теплоту и привязанность у клиента и какая ценность его привлекает. При наличии такой ценности, необходимо показать ее покупателю. Сначала необходимо объяснить ему свои отличия от конкурентов, укрепив его уверенность в качестве именно вашего товара или услуги. Внимательно слушайте, что говорят клиенты о конкурентах и делайте выводы. Это поможет узнать о конкурентах, о которых вы не слышали.

– Латентная лояльность. К компании потребитель эмоционально привязан, но редко покупает. Причин может быть множество, например: плохая доставка, высокие цены и др. или покупатель узнал о продукте нечто такое что теперь мешает ему приобретать его. Но в то же время латентные – это потенциально истинные (лояльные) покупатели. Организация может развивать их лояльность, выявляя и устраняя причины их нежелания обращаться в компанию. Можно предложить клиенту замену дорогого товара, ответить на его вопросы и снять возражение от покупки. Можно действовать по-другому. Например, узнать ценности потребителей или определить, с помощью маркетинговых стратегий, факторы привлекательности бренда для покупателя. Затем можно использовать эти факторы для роста лояльности.

– Истинная лояльность. Покупатель уже привязан к компании, он часто покупает и пользуется её услугами. Именно поэтому, настоящая (истинная) лояльность – это лучший и самый правильный вариант лояльности. Именно такие клиенты являются самым ценным активом компании. Даже если у истинного покупателя не самый высокий чек, но большая история покупок, он дороже чем латентный или ложный покупатель. Постоянная связь с клиентами должна быть сохранена в любой деятельности.

Также необходимо использовать новые маркетинговые приемы для того чтобы показать клиенту его уникальность.

Для длительного сохранения лояльности покупателей важно создать условия для того, чтобы они не хотели расставаться с компанией. Нужно уметь общаться с клиентами, которые находятся в процессе покупки или уже купили товар. Очень важно вовлечь потребителей в жизнь бренда, его развитие, создать клубы для общения и т.п.

Клиенты проявляют лояльность в несколько этапов:

Первый контакт. Как правило, потенциальный клиент пока не имеет представления о компании или товаре и услугах. И он первый раз попал в интернет-магазин или офис компании. Он к кому-то обратился за помощью, и ему порекомендовали вашу компанию, либо он нашел ее в интернете. Сначала, как правило, происходит одно из трех действий:

- Клиент ушел без покупки. Вы ему не были известны и у него еще нет доверия. Возможно, он зашел по ошибке, либо просто любопытствовал, но, возможно, его что-то насторожило. Именно такой вариант наиболее распространенный.

- Не происходит никаких покупок, клиент просто откладывает решение на неопределенный срок. В этот раз ему что-то показалось интересным и даже важным для него. Сейчас он не готов приобрести, да и в ближайшее время вряд ли будет готов к покупке. Возможно, он не может себе позволить купить, хочет больше узнать о аналогичных предложениях конкурентов или не уверен в необходимости данной покупки. В этом случае негативный фактор – риск того, что покупатель больше не вернется, но повлиять на него уже можно.

- Уже при первом контакте, клиент приобретает товар. Таким образом, такой клиент может быть переведен в разряд лояльных. Товар или услуга ему понравились, он не стал сравнивать их с товарами конкурентов, сразу приобрел и перешел на следующий уровень лояльности.

Первая покупка. Покупатель впервые совершает покупку и оценивает качество товара. Теперь у него есть свое собственное мнение о компетентности сотрудников компании., если контакт был очным. Он остался доволен, и условиями покупки в интернет-магазине, и доставкой, если покупал онлайн, но, самое главное - качеством товара. Именно этот параметр определяет, будет ли человек снова покупать что-то. При условии, что все остальные факторы так же произвели приятные впечатления компания получит лояльного покупателя. Компании следует тщательно изучить причины недовольства потребителя, если такие выясняются, постараться найти решение и/или поинтересоваться каким вариантом устроит клиента. Таким образом, клиенты указывают на недостатки и недоработки в компании, что дает возможность улучшить её сервисы и продукты. Важно сделать так, чтобы негативный опыт стал позитивным.

Повторные контакты. После покупки клиент обязательно вернется еще раз, если ему понравилось качество продукции, отношения сотрудников и услуги компании. Специалисты компании должны провести анализ удовлетворенности клиентов для выяснения факторов, которые побуждают клиентов возвращаться и что они ожидают от компании в будущем. Благодаря этому клиенту станет ясно, что его желания важны и они будут удовлетворены. Компания сможет выяснить наилучший подход к данному потребителю. Основной целью является то, чтобы потребитель убедился в том, что он сделал правильный выбор. Измерить уровень удовлетворенности можно при помощи анкет или рассылки по SMS и электронной почте, но следует предварительно получить письменное разрешение клиентов на такие действия. Также можно использовать опросы через соцсети и специализированные сервисы. К дополнительным способам сохранения клиентов можно отнести: рекламные кампании; различные конкурсы и розыгрыши призов; персональные предложения клиентам, для демонстрации особого отношения.

Как правило, методы повышения уровня лояльности делятся на эмоциональные и материальные. Если материальные методы приносят клиентам определенную выгоду, то эмоциональные методы дарят клиенту приятные эмоции. По мнению психологов, именно эмоциональная составляющая является определяющей в формировании лояльности [9].

Материальные способы, приводящие к повышению лояльности:

– Качество продукции и/или услуги. Продукт не будет пользоваться популярностью у клиентов, если он не соответствует стандартам качества. Как правило, клиенты не любят ходить в сервисы по ремонту техники и общаться со службой технической поддержки. Даже если качество продукта или услуги хорошее, все равно нужно постоянно улучшать продукт или дорабатывать его. Для клиентов важно видеть, какие усилия компании прилагаются к улучшению продуктового продукта. Это повышает его лояльность.

– Программа лояльности. Такая программа поощряет клиентов повторно покупать товары и удерживает их в течение длительного времени. Поощряемый клиент чувствует себя более ценным и важным. В целом он доволен и даже рад, что ему удалось найти такое выгодное предложение. Есть несколько вариантов программ: кешбэк; бонусы, фиксированные скидки; закрытый клуб.

– Подарки за покупку. Продукт бренда может быть подарен как в качестве подарка, так и в качестве логичного дополнения к основной покупке. Маленькие полезные подарки вызывают желание вернуться к продавцу, а ещё рассказать о подарке своим близким.

– Периодические акции и скидки. Выгодные акции и скидки на товары, которые продаются в магазинах, можно легко и быстро реализовать. По мнению авторов, это может быть очень привлекательным предложением для любого покупателя. Гарантированная возможность получения выгоды за короткое время уменьшает срок обдумывания клиентами нужности покупки.

Эмоциональные способы, повышающие лояльность:

– Ценности компании. Покупатель лоялен к компании, если она близка его взглядам на жизнь и ценности. С помощью общих ценностей, которые имеются у человека, можно сформировать эмоциональную привязанность к компании. Для привлечения внимания потребителей к ценностям фирмы можно наносить главные идеи на упаковку, использовать их как слоган. или рассказать о них в соцсетях. Также это происходит и при общении покупателей с продавцами. Важно подчеркнуть ценности компании, соответствующие главному убеждению клиента и отличающие организацию от конкурентов.

– Уровень сервиса. Если потребитель испытал негативные эмоции от некачественного сервиса, то запомнит это. В любом случае, он будет рассказывать об этом своим друзьям и знакомым. Заработает «сарафанное радио», и клиенты будут потеряны. С помощью качественного сервиса и дружелюбного отношения сотрудников компании можно добиться лояльности покупателей. Компания должна всегда соблюдать это требование. Безусловно никто не застрахован от непредвиденной ситуации, но все же нужно регулярно отслеживать ситуацию с сервисом и незамедлительно решать его проблемы. После этого необходимо провести работу по устранению недостатков, чтобы они не повторялись в дальнейшем.

– Работа с негативом. Важным является ответная реакция или действие по нейтрализации негативного опыта клиентов. В случае если компания не права, следует извиниться и подробно рассказать о том, как будет исправлена ситуация. За неудобства клиент может получить компенсацию. Если устранить проблему не удастся, например, в случае, если товар не пришел вовремя или поставщик нарушил соглашение, немедленно сообщите об этом клиенту и продолжайте информировать его о каждом шаге выполнения договора. Если клиент не прав, то нужно спокойно объяснить ему суть проблемы, указав на его ошибки, и предложить варианты решения. Покупатель почувствует себя услышанным.

На современном рынке экономической деятельности наблюдается высокая конкуренция, что приводит к росту ожиданий потребителей и возникновению их дефицита. Из-за сложившейся ситуации формируется концепция маркетинга взаимоотношений с потребителем, которая сменит классический подход. Привлечение новых клиентов было главной целью маркетинговых стратегий прошлых лет. На данный момент основное внимание уделяется сохранению существующих клиентов и формированию их лояльности и преданности компании. При этом изменилось понимание того, что долговременные отношения с клиентом являются экономически выгодными. Этот подход дает гарантию стабильного уровня покупок, требует меньших затрат на маркетинг и стимулирует рост количества постоянных клиентов. Все это гарантирует увеличение доходов, следовательно именно лояльность потребителей является важнейшим показателем успешной работы компании [9][10][12].

При оценке лояльности используются различные методы:

– Метод «разделение потребностей» (предложен Ж. Хофмайером, Б. Райсом, Д. Аакером). На основе этой методики продавцы определяют степень доверия покупателя и его лояльность к бренду на основе числовых показателей. Данный метод основан на определении лояльности к бренду, исходя из соотношения частоты и общего количества предпочтений одной торговой марки по отношению к другой. Эксперты считают, что если покупатель 7 раз из 10 выбирает конкретного производителя и в течение некоторого времени и не покупает продукцию других производителей, то его лояльность к бренду составляет 70%. Как показали исследования, лояльными покупателями являются те клиенты, которые имеют уровень повторных покупок больше 67%. Людей, которые имеют более низкий процент, эксперты называют «перебежчиками».

– Традиционный подход. Эта методика заключается в том, чтобы выявить у клиентов намерение о приобретении продукта. Если у покупателя есть сильное желание приобрести товар, то считается, что он доверяет

данному производителю и лоялен ему. Такой потребитель совершает повторную покупку, рекомендует фирму своим знакомым и поддерживает бренд в своей среде. Они не считают нужным заниматься маркетинговыми исследованиями, а также анализировать их.

Намерения покупателей отличаются, поэтому они могут оказать разный эффект на разные показатели компании. Выделяют несколько видов намерений о покупке:

- покупатель хочет совершать новые покупки, что влияет на увеличение продаж в целом;

- на рост числа клиентов влияет покупатель, рекомендуемый товар другим людям;

- если покупатель намерен приобрести больше товаров бренда, то это влияет на средний размер выручки на одного покупателя;

- Конверсионная модель, предложенная Ж. Хофмейром и Б. Райсом. Данный способ позволяет определить уровень приверженности к определенной группе товаров. Зависит от четырех показателей: удовлетворение брендом, наличие выбора из альтернативных вариантов, ценность приобретения продукта и уровень неуверенности (двойственность отношения к покупке или колебания). Стоит отметить, что продавец должен задать покупателям следующие вопросы для определения уровня приверженности потребителей:

- Что вы думаете о компании, исходя из уровня удовлетворения ваших потребностей и личных ценностей;

- Важно ли для вас решение, какую компанию выбрать;

- Какова ваша оценка других аналогичных брендов.

- Подход Д. Аакера, суть которого заключается в измерении следующих показателей:

- поведение покупателей и наблюдение за показателями повторной покупки, процента покупок и общего количества купленных торговых марок;

- учет затрат, связанных с переключением. Покупатель не приобретает продукцию другой марки, он боится «риска изменений»;
- удовлетворенность – как главный показатель оценки лояльности;
- отношение, как к фирме, так и к сотрудникам;
- приверженность – оценивается как количество взаимодействия приверженных покупателей с другими покупателями, использующими бренд.

Данный метод измерения лояльности клиентов используется многими компаниями. Для этого обычно используются количественные исследования: почтовые и телефонные опросы, вопросы в режиме on-line, интервью или карточки гостя и т. п.

- Метод NPS, разработанный Ф. Рейхельдом. По мнению автора, настоящими лояльными покупателями являются те люди, которые активно рекомендуют и рекламируют бренд своим друзьям и привлекают этим новых клиентов. По мнению авторов рекомендации является гарантией лояльности клиента к продавцу и гарантом того, что он не станет совершать покупку у другого продавца. При этом он лично гарантирует качество товаров своим знакомым. Обычно, по методу NPS, покупателю задается один вопрос: «По шкале от 0 до 10 какая вероятность того, что вы предложите нас своим друзьям или коллегам?». По ответу на этот вопрос можно сделать вывод о вероятности совершения клиентом повторной покупки или рекомендации.

- Метод фокус-группы (социологи Р. Мертон и Р. Кендалл). При этом подразумевается способ проведения глубинного фокусированного интервью с участием группы людей в форме серии групповых дискуссий, участники которых «фокусируются» на вопросах, интересующих исследователя для получения от них информации. Этот метод является высоконадежным способом оценки эффективности рекламы и отношения покупателей к бренду.

- Метод Ж. Ж. Ламбена. Для того чтобы провести исследования лояльности клиентов, необходимо пройти несколько этапов:

- исследование достоинств продукции;

- анализ работы персонала, а также качество обслуживания;
- изучение ценности продукта для покупателей и его восприятие в сознании потребителей.

Согласно правилу, при изменении степени лояльности потребителей необходимо изучить этот период более подробно (повышение цен и изменение ассортимента и т. д.) для принятия мер по укреплению её лояльности.

- Модель SERVLOYAL. Система комплексно оценивает следующие параметры: поведение покупателей; их отношение к бренду; аффективные и эмоциональные компоненты; обязательства покупателей перед брендом и их ответственность.

При выборе метода оценки лояльности покупателей компании руководствуются собственными предпочтениями. Так как они могут быть использованы в качестве основы для создания различных конструкций, то их выбор зависит от возможностей компании и её направленности. Сегодня достижение долгосрочных показателей развития бизнеса невозможно без привлечения большого количества постоянных клиентов. Причина этого проста: сильная конкуренция и большое количество предложений в сфере услуг и торговли требуют особых требований к бизнесу. Именно поэтому клиенты являются одним из самых ценных активов компании, а лояльность клиента – один из ключевых факторов, который влияет на успешность бизнеса.

Необходимо учитывать все факторы, влияющие на лояльность клиентов. Помимо этого, необходимо включать построение моделей и анализ их функционирования. Следует также решать и другие задачи: начиная от правильного учета данных (например, учёт истории отношений с клиентами), заканчивая анализом деятельности персонала для корректного составления концепции программы лояльности.

Анализ – это «мозг», который управляет всем комплексом. При отсутствии анализа вся информация о клиентах не сможет использоваться для получения новых знаний о них и увеличения прибыли компании. В этом

случае программа лояльности не будет эффективным инструментом улучшения маркетинга компании.

1.2 Популярные инструменты и платформы

Развитие области анализа данных расширяет спектр инструментов для обработки информации. В таблице 1 приводится краткое описание некоторых инструментов для анализа данных [21].

Таблица 1 – Анализ инструментов для анализа данных

Наименование	Тип инструмента	Доступность	В основном используется	Плюсы	Минусы
Excel	программное обеспечение для работы с электронными таблицами	коммерческая	обработки данных и составления отчетов	Широко используемый, с множеством полезных функций и плагинов	стоимость, ошибки в расчетах, плохая обработка больших данных
Python	Язык программирования.	с открытым исходным кодом, с тысячами бесплатных библиотек.	всего, от сбора данных до анализа и составления отчетов.	Простота в освоении, высокая универсальность, широкое применение.	объем памяти – выполняется не так быстро, как некоторые другие языки
R	Язык программирования.	с открытым исходным кодом.	статистического анализа и интеллектуального анализа данных.	Независимость от платформы, высокая совместимость, множество пакетов.	более медленный, менее безопасный и более сложный в освоении, чем Python.
Jupyter Notebook	Интерактивное программное обеспечение для разработки.	с открытым исходным кодом.	обмена кодом, создания руководств, презентации работы.	Отлично подходит для демонстрации, не зависит от языка.	не автономны и не подходят для совместной работы.

Продолжение таблицы 1

Наименование	Тип инструмента	Доступность	В основном используется	Плюсы	Минусы
Apache Spark	Платформа обработки данных.	с открытым исходным кодом.	обработки больших данных, машинного обучения.	Быстрый, динамичный, простой в использовании.	Отсутствие системы управления файлами, жесткий пользовательский интерфейс.
SAS	Статистический программный пакет.	Коммерческая.	бизнес-аналитики, многомерного и прогнозного анализа.	Легкодоступность, ориентированность на бизнес, хорошая поддержка пользователей.	высокая стоимость, плохое графическое представление.
Tableau	Инструмент визуализации данных.	коммерческая.	Создания информационных панелей и рабочих листов.	отличная визуализация, скорость, интерактивность, мобильная поддержка.	слабый контроль версий, отсутствие предварительной обработки данных.
KNIME	Платформа интеграции данных.	с открытым исходным кодом.	интеллектуального анализа данных и машинного обучения.	платформа с открытым исходным кодом, которая отлично подходит для визуального программирования.	не хватает масштабируемости, и для некоторых функций требуется техническая экспертиза.

Наиболее популярным программным обеспечением, которое используется для работы с электронными таблицами, является Microsoft Excel. Он имеет функции вычислений и построения графиков, идеально подходящие для анализа данных. При любых ваших предпочтениях и при любом другом используемом программном обеспечении, MS Excel является основным в этой

области и это основная область применения этого продукта. В его встроенные функции входят сводные таблицы (для сортировки или суммирования данных) и инструменты создания форм. Кроме того, он имеет ряд других функций, упрощающих манипулирование данными. С другой стороны, у него есть и ограничения, например, он работает довольно медленно с большим количеством информации и аппроксимирует большие числа. Это может привести к ошибкам. Как бы там ни было, MS Excel серьезный и мощный инструмент анализа данных, а благодаря множеству доступных плагинов можно легко обойти его недостатки.

Python имеет широкий спектр применения для анализа данных и является обязательным для любого аналитика. Он сфокусирован на удобстве программирования, а его популярность в области анализа означает, что многие программисты знакомы с ним. Python очень универсален; у него есть множество библиотек, которые подходят для многих задач анализа данных. С помощью библиотек NumPy и Pandas можно оптимизировать сложные задачи, а также поддерживать общие манипуляции с данными. С помощью таких библиотек, как Beautiful Soup и Scrapy можно извлекать информацию из Интернета. Matplotlib отлично подходит для визуализации данных в виде отчетов или диаграмм. Главным минусом является его скорость, он требует много памяти и медленнее других языков. Несмотря на это, при создании программного обеспечения с нуля преимущества Python значительно перевешивают его недостатки.

Язык R, как и язык Python, является очень популярным языком программирования с открытым программным кодом. Его часто используют в качестве основы для создания программного обеспечения для статистики/анализа данных. Язык R сложнее, чем язык Python, и его синтаксис более сложный. Но в то же время он был создан специально для решения сложных задач статистики и очень популярен при визуализации данных. R имеет сеть свободного доступа, которая называется CRAN, и предлагает более 10000 различных пакетов. Она отлично работает с другими языками и

системами, включая программное обеспечение для больших данных и может использовать код, созданный на языках C, C++ и FORTRAN. С другой стороны, он плохо управляет памятью и не имеет специальной команды поддержки, кроме сообщества пользователей. Следует отметить, есть отличная интегрированная среда разработки (IDE) для R под названием RStudio.

Jupyter Notebook – веб-приложение с открытым исходным кодом. С помощью него можно создавать интерактивные документы. Это сочетание живого кода, уравнения и визуализации. Представьте себе документ Microsoft Word, который гораздо более интерактивен и разработан специально для анализа данных. Его можно использовать как инструмент анализа данных: Jupyter Notebook работает в браузере и поддерживает более 40 языков, включая Python и R. Он также интегрирован с Apache Spark для создания различных выходных от HTML до изображений или видео. Вы можете получить разные выходные данные из текстового файла WebMaster. Однако, как и у любого инструмента, у него имеются свои ограничения. По мнению разработчиков дневники Jupyter Notebook плохо контролируют версии, а мониторинг изменения плохо понятен. Следовательно, это не самая подходящая среда для разработки и анализа (нужно использовать выделенную IDE, а она плохо подходит для совместной работы). Но поскольку он не является автономным, требуется предоставить любые дополнительные ресурсы (например библиотеку или систему выполнения) всем людям, с которыми делитесь документом. Однако, для целей презентации и обучения он остается отличным инструментом для анализа данных.

Apache Spark является одной из программных платформ, позволяющих аналитикам данных и специалистам по обработке информации быстро обрабатывать огромные объемы информации. Этот продукт впервые был создан в 2012 году и впоследствии передан на благотворительность некоммерческой компании Apache Software Foundation. Как инструмент анализа неструктурированных больших данных, Spark размещает сложные

задачи на множестве компьютеров. Spark работает исключительно быстро, по сравнению с другими подобными фреймворками (например, Apache Hadoop). Оперативная память в 100 раз быстрее, чем локальная, поэтому Spark примерно в сто раз быстрее Hadoop. Вот почему они широко используются при разработке моделей машинного обучения, которые содержат большой объем данных. Его библиотека алгоритмов машинного обучения MLlib включает в себя множество методов классификации, регрессии и кластеризации и это только некоторые из алгоритмов. Отметим, что использование таких больших объемов памяти означает большую вычислительную нагрузку для Spark. Кроме того, он также не имеет функции управления файлами и, следовательно, обычно требуется интеграция с другим программным обеспечением, например Hadoop.

SAS («система статистического анализа») - популярный коммерческий набор инструментов для бизнес-аналитики и аналитических исследований. Этот проект был разработан институтом SAS еще в 1960-х годах и до сих пор не теряет своей актуальности. Сейчас эти инструменты используются в основном для профилирования клиентов, составления отчетов, интеллектуального анализа данных и прогнозирования. Написанное для корпоративных рынков ПО обычно более надежное и универсальное в использовании для крупных организаций. С этим связано то обстоятельство, что они, как правило, имеют разный уровень знаний в области программирования. Цена на SAS-продукт высока. Однако, затраты приносят выгоду: в него постоянно добавляются новые модули на основе спроса клиентов. По сравнению с библиотеками Python их меньше, но они более сконцентрированы. Например, SAS предоставляет модули для конкретных целей: борьбы с отмыванием денег и аналитики для Интернета вещей.

Microsoft Power BI, с момента создания которого прошло менее десяти лет, является относительно молодым инструментом для анализа данных. Этот инструмент начал свою жизнь как подключаемый модуль Excel, но был переработан в начале 2010-х годов и стал отдельным набором инструментов

для анализа данных. С помощью PowerBI можно создать интерактивную графику и информационные панели с минимальной уровнем подготовки. Он обладает отличными возможностями - легко подключается к данным, хорошо работает с Excel (продукт Microsoft), текстовыми файлами, SQL Server и облачными источниками (Google и Facebook analytics). Этот продукт также обеспечивает надежную визуализацию данных, но у него есть необходимость в улучшении в других областях. Например, у него очень сложный пользовательский интерфейс и работа с формулами, а собственный язык анализа данных (DAX) не очень удобен для пользователя. В то же время он предлагает несколько бесплатных подписок. Это очень хорошо для тех, кто хочет изучить инструмент. Хотя у бесплатной версии есть ограничение на лимит обрабатываемых данных (до 2 ГБ).

Tableau, с помощью которого можно создавать интерактивные визуализации и информационные панели без глубоких знаний в области программирования. Это лучший доступный инструмент анализа данных. Этот пакет обрабатывает огромные массивы данных быстрее и лучше, чем многие другие инструменты BI и очень прост в использовании. У него есть визуальный интерфейс перетаскивания, что является еще одним несомненным преимуществом перед многими другими инструментами анализа данных. Но, так как в нем нет возможности создания сценариев его возможности ограничены. По этой причине Tableau не очень подходит для предварительной обработки данных и построения более сложных вычислений. Он содержит только функции манипуляции данными, но они недостаточно эффективны. Как правило, требуется использовать Python или R для скриптовых функций перед тем, как импортировать данные в Tableau. В то же время его визуализация очень хороша, что делает ее невероятно популярной даже несмотря на некоторые недостатки. Помимо этого, он может работать на любых мобильных устройствах и даже на смартфонах.

KNIME (Konstanz Information Miner), облачная платформа интеграции данных с открытым исходным кодом, является последним в списке.

Разработан немецкими программистами в 2004 году. KNIME был создан для фармацевтической промышленности, но его эффективность в сборе данных из множества различных источников привела к его использованию и в других областях. Это могут быть анализ клиентов, бизнес-аналитика, а также машинное обучение. Одним из его главных преимуществ является то, что он бесплатный и не требует от пользователя каких-либо дополнительных затрат на приобретение. Его графический пользовательский интерфейс (GUI) и возможность перетаскивания идеально подходит для визуального программирования. Поэтому пользователям не требуется специальных знаний, чтобы создавать рабочие процессы с данными. Часто говорится, что KNIME поддерживает весь спектр задач анализа данных, но на самом деле его основное преимущество - заключается в интеллектуальном анализе данных. Несмотря на то, что KNIME позволяет углубленное исследование статистики, требуются некоторое знание Python и R. Поскольку KNIME имеет открытый исходный код, он абсолютно бесплатен и легко настраивается в соответствии с потребностями организации. Поэтому он пользуется популярностью у малых компаний с ограниченным бюджетом и не вызывает нареканий со стороны клиентов, которые не могут себе позволить более дорогой продукт.

Нет единого инструмента анализа данных, который решал бы все проблемы с анализом и делал все необходимое для анализа. Хороший аналитик должен обладать значительным опытом в использовании разных языков и систем аналитики данных [21]. В приведенном списке можно выбрать инструмент, подходящий для большинства задач, но даже для небольших процессов анализа данных потребуются дополнительные инструменты. Необходимо проанализировать бизнес-потребности организации и определить, что будет использовать требуемые инструменты. Для использования аналитиками данных или нетехническими специалистами, которым требуется интерактивный и интуитивно понятный интерфейс. В вышеприведенном списке присутствуют инструменты, которые подойдут, и аналитикам, и нетехническим специалистам. Необходимо рассмотреть

возможности инструмента для моделирования данных. Обладает ли инструмент такими возможностями или, для моделирования данных, придется использовать другой инструмент. Следующий вопрос – вопрос о цене и лицензировании. Некоторые инструменты полностью бесплатные, а некоторые имеют полезные бесплатные функции, но полный продукт требует лицензии.

Выводы главы 1

Достичь стабильного роста в бизнесе невозможно без создания сообщества лояльных клиентов. Причины этого довольно просты: растущая конкуренция и огромное разнообразие предлагаемых товаров и услуг предъявляет особые требования к организации бизнеса в целом. В связи с этим важным активом компании является постоянный клиент, а лояльность клиентов становится одним из ключевых показателей успеха компании.

Проанализировав предметную область, можно сделать выводы:

- пока что нет единой методики, которая бы оценивала покупательскую лояльность;
- для создания комплекса программ с целью эффективного управления лояльностью покупателей, необходимо использовать ключевые методы диагностики и анализ поведения покупателей.

Когда фирма проводит анализ своего потенциального клиента на предмет его требований к компании и ее продукции, она анализирует его с помощью данных, которые были получены от них. Анализ клиентов, на основе опросов покупателей — это целенаправленные исследования поведения потребителей. Исследователи вступают во взаимодействие с клиентами и затрагивают "болевые точки" целевой аудитории с целью увеличения общего объема продаж. Перед тем, как клиент начнет покупать компания изучает определенные факторы для определения критериев покупки своей целевой аудитории.

Глава 2 Проектирование приложения для оценки лояльности клиентов

2.1 Анализ лояльности клиентов на примере Direct Marketing

Исследование клиентов включает в себя качественные и количественные методики исследования, используемые для определения того как покупатель выбирал компанию и его покупательское поведение. Анализ позволяет оптимизировать процесс принятия решений и составления успешного маркетингового плана. Его цель сообщить о потребностях активных покупателей, текущих или новых фокус-групп для предложения продуктов и услуг по индивидуальным заказам для увеличения продаж компании и повышения лояльности клиентов [1][8][14][28].

При анализе клиентов выявляют:

- своих самых прибыльных клиентов;
- определение потребностей клиентов и проблем, с которыми они сталкиваются («болевые точки»). Больные точки клиента — это состояния дискомфорта и беспокойства, которые возникают у клиента в связи с его неудовлетворенной потребностью;
- выявление новых групп покупателей, которые имеют схожие черты и поведение;
- разработка и создание профилей (персон) идеального покупателя;
- понимание моделей принятия решений о покупке новыми клиентами и их улучшение;
- создание индивидуального контента и совершенствование стратегии продаж, с использованием наиболее подходящие каналов маркетинга;
- улучшенное принятие решения, путем предложения существенных изменений для продукта или услуги, с целью повышения соответствия потребностям клиентов.

Рассмотрим подробнее, как можно сегментировать клиентов. Один из самых популярных методов исследования лояльности Direct Marketing - разработка предложений клиентам с учетом их предпочтений. Концепция заключается в том, чтобы нужный товар был предложен именно тем людям, которые его ищут.

При проведении маркетинговой кампании можно выделить следующие этапы [14][1][24].

Сегментация клиентской базы. Главный этап при разработке программы стимулирования - сегментация клиентов и отбор наиболее подходящего клиента. Необходимо выяснить, кто из потребителей является потенциальным клиентом. С помощью опроса выяснить у клиентов их предпочтения в отношении того или иного товара. Подборка клиентуры по простым правилам типа «Пол и возраст» не могут отразить реальную структуру предпочтений клиентов. Возможно сегментировать покупателей по разным категориям:

- Сведения о возрастной группе населения, доходах, полу и других демографических данных.
- Географическая сегментация, которая зависит от места расположения клиента.
- Основой для формирования психографических сегментов является ценности, интересы и убеждения.
- Технические сегменты, которые зависят от устройства или платформы клиента.
- Поведенческая сегментация, связанная с различными способами совершения повседневных действий.
- Сегментации по потребности в продукте и его свойствах.
- Сегментация на основе формулы оценки ценности, обычно измеряемая CLV («Customer lifetime value» - прогноз прибыли каждого привлеченного покупателя или клиента в будущем).
- Сфера деятельности клиента влияет на отраслевую сегментацию.

– Сегменты различают в зависимости от количества сотрудников и объема выручки.

Очевидно, что на предпочтения клиентов влияют многие факторы в очень сложной комбинации, начиная от географии и заканчивая средним доходом. После сегментации необходимо использовать всю информацию из описанных выше шагов для создания пользовательских образов (персонажей) целевого клиента. С помощью этих образов покупателей можно сконцентрировать внимание на демографии клиентов, моделях поведения и целях клиента. Они помогают сосредоточить внимание в первую очередь именно на потенциальных покупателях. Теперь вы можете построить отношения с определенными категориями клиентов и создать для каждой группы специализированную программу маркетинга.

Выбор целевой аудитории. Согласно принципу Парето, для компании около 80% бизнеса приходится на 20% клиентов, поэтому очень важно найти своих ценных покупателей. На первом месте должен стоять клиент, который наиболее вероятно будет привлечен продуктами и услугами организации. При помощи исследования модели поведенческих реакций клиентов компания может определить своих потенциальных покупателей. Как правило, верный выбор приносит хороший результат и эффективность маркетингового мероприятия вырастает в несколько раз. Поэтому после обнаружения сегмента необходимо оценить его и оценить показатель эффективности мероприятия:

– Измерив сегменты и вычислив статистические характеристики для каждого, можно оценить размеры сегмента и вычислить статистическую характеристику для него.

– Вычислить финансовый потенциал и привлекательность сегмента.

– Опираясь на какие характеристики определить, что это за сегмент и в чем его особенность.

– Определить, какая оптимальную стратегию для каждого конкретного сегмента.

Формирование адресных предложений. Клиенты остаются лояльными к брендам, которые помогли им получить качественные продукты и их требования были удовлетворены компанией. Нужно четко знать, что хочет клиент и идти ему навстречу, но это невозможно без участия самого клиента. Покупатели доверяют бренду, который постоянно улучшается согласно их пожеланиям. Очевидно, что, если четко определить продукты и услуги для каждого сегмента рынка, то можно рассчитывать на больший эффект от адресных обращений. Продуманные и качественные продукты увеличивают лояльность клиентов, а некачественные могут вызвать недовольство. Вместо отправки одинаковых электронных писем всем потенциальным клиентам из базы данных, рекомендуется адаптировать сообщения для конкретного персонажа и к тому, что известно о каждом конкретном покупателе.

Анализ отклика клиентов. Отслеживание отзывов помогает выявить лучшие способы продвижения товара и их эффективность. Даже если предложения потребителям сформированы на основе лучших алгоритмов анализа, нет гарантии получения хорошего отклика. Это связано с тем, что в любом экономическом процессе есть множество факторов, влияющих на его развитие. Анализируйте отклики на маркетинговое действие, чтобы выявить факторы, влияющие на процесс. Необходимо учитывать отрицательные ответы и причины отказа, чтобы избежать ошибок в дальнейшем.

Мало придумать и систему повышения лояльности, также нужно контролировать её и следить за её эффективностью. Это очень сложная задача и она не может быть сведена к расчету среднего чека. В большинстве случаев программа лояльности не контролируется руководством компании или ее акционерами. В эффективности и смысле программы никто не уверен, потому что нет необходимого контроля за ее выполнением. Для того, чтобы руководство могло проанализировать эффективность системы и в денежном эквиваленте, и в оценке количества клиентов, необходимо хранить статистический материал в системе управления [1][8][10].

Имеется различные критерии, которые следует использовать при оценке эффективности программ лояльности:

– Пожизненная стоимость клиента. LTV (life time value) – доход, получаемый с одного клиента за все время (1).

$$LTV = \frac{D-Z}{K} \quad (1)$$

где:

D – доходы за период;

Z – стоимость рекламы;

K – число клиентов.

– Коэффициент повторных покупок. RPR (repeat purchase rate) – показывает, сколько человек вернулись после первой покупки в процентном соотношении ко всей клиентской базе (2).

$$RPR = \frac{K}{All} \cdot 100 \quad (2)$$

где:

K – количество людей, совершивших повторные покупки;

All – всего покупателей за отчетный период.

– Коэффициент удержания клиентов. CRR (customer retention rate) – помогает проанализировать ситуацию с оттоком покупателей за определенный период времени (3).

$$CRR = \frac{K-P}{N} \cdot 100 \quad (3)$$

где:

K – число клиентов к концу срока измерения;

P – клиенты, пришедшие за этот срок;

N – число клиентов на начало срока.

– Коэффициент оттока клиентов. CR (churn rate) – коэффициент оттока клиентов (4).

$$CR = \frac{N-K}{K} \cdot 100 \quad (4)$$

где:

N – число клиентов на начало срока измерения;

K – число клиентов на конец срока измерения.

– Индекс потребительской лояльности. NPS (net promoter score) – все больше компаний в качестве обратной связи в своей триггерной электронной почте просят клиентов ответить на вопросы анкеты и указать по шкале от 1 до 10, насколько им нравится сервис, обратятся ли они снова к компании и будут ли рекомендовать её своим знакомым [25][7][12].

Направление и сфера деятельности организации, сезонность спроса и социально-демографические характеристики потребителей влияют на общий уровень лояльности и должны учитываться при оценке общего уровня поддержки и доверия.

С помощью базы данных клиентов так же можно получить ценную информацию. Например, о личных данных целевой аудитории (узнать их возраст и биографию), чтобы использовать правильные способы общения с ними.

2.2 Методы определения лояльности клиентов

Исследование потребителей является способом анализа и использования данных о клиентах для разработки бизнес-плана организации. При анализе клиента выявляется важная информация о них (например, их потребностях и предпочтениях), что позволяет сегментировать покупателей во время проведения маркетинговых кампаний продукта или услуги с целью оптимизации роста продаж и удержания клиентов. Главная цель этого исследования - узнать предпочтения большого количества целевых клиентов

и почему они выбирают тот или иной продукт. В процессе анализа клиентской базы компании могут лучше понять, как удовлетворять конкретные потребности аудитории [1][2].

По сути, сегодня сегментирование потребительского рынка больше не является дополнительным инструментом маркетинга. Это ключевой фактор успеха компании. Разбиение клиентов на отдельные группы выгодно, поскольку это поможет разработать стратегию маркетинга сразу для целой категории покупателей. При сегментировании клиентов используются несколько факторов: возраст, местоположение, язык, пол и т.д., а затем определяются требования клиентов и разных групп [14].

С помощью анализа потребительского поведения компания анализирует поведение клиентов и определяет критерии сегментации для запуска персонализированных рекламных кампаний для целевой аудитории определенных сегментов, таких как географическое положение, возрастной диапазон или склонности.

Эффективная маркетинговая кампания поможет повысить лояльность клиентов, что поспособствует процветанию бизнеса. Потребители обычно выбирают товары или услуги бренда, имеющего хороший имидж, и положительные отзывы других клиентов.

Метод NPS. Net Promoter Score (NPS) или Индекс потребительской лояльности – это показатель, позволяющий узнать о том сколько клиентов готовы рекомендовать услуги и товары компании. Стоит заметить, что оценка актуальна только сейчас, поэтому опросы следует проводить регулярно, отслеживая динамику [25][7][10][12].

Показатель NPS – это отношение клиентов и сотрудников к компании. Его можно использовать для выявления слабых и сильных сторон компании, определения стратегии развития и отслеживания эффективности внедренных изменений. Можно оценивать не весь бизнес, а различные его сегменты: доставку и продажи; обслуживание клиентов и пр. В итоге все пользователи будут разделены на тех людей, которые довольны работой и людей, имеющих

замечания. На основе этого можно выработать стратегию поведения с различными клиентами: одних поощрять, других слушать и предлагать решение проблем и в результате делать их сторонниками. В идеале лучше всего, чтобы пользователь не только оценивает, но и оставлял комментарий по поводу того почему поставил именно этот балл. Эта информация может помочь в выявлении конкретных проблем и предложить пути их решения. В формулу оценки лояльности входят такие показатели, как уровень доверия к компании, уровень удовлетворенности клиентов и уровень удовлетворенности сотрудников. Часто оценка лояльности проводится по формуле (5).

$$\frac{П - К}{n} \cdot 100\% \quad (5)$$

где:

- П – количество промоутеров,
- К – количество критиков,
- n – количество участников опроса.

Полученный результат может находиться в пределах от -100 до 100%:

- -100 % – 0 % – среди опрошенных преобладают критики;
- 1 % – 30 % – доля критиков высока;
- 31 % – 50 % – критики и промоутеры находятся примерно в равном соотношении;
- 51 % – 70 % – число промоутеров превышает критиков;
- 71 % – 100 % – большинство опрошенных промоутеры, компания высоко ценится на рынке.

Например, удалось опросить 1500 клиентов из имеющейся базы. Оценку от 9 до 10 баллов поставили 35% опрошенных, 15% поставили 7–8 баллов, 50% – от 0 до 6 баллов. Тогда: $NPS = \frac{(525 - 750)}{1500} \cdot 100 = -15\%$

Показатель NPS выше 30% является положительным. Но если посмотреть на эту цифру в отрыве от ситуации, то это будет неправильно.

Несмотря на то, что NPS вашего предприятия составил 23%, а у конкурентов оценка не поднимается выше 15–20%, то результаты могут быть оценены как положительные, но еще есть куда расти.

По мнению Ф. Райхельда, важны не цифры и показатели, а стратегии по превращению нейтральных и критиков в промоутеров. Необходимо смотреть не только на итоговые цифры, но и на динамику. Рост оценки по отношению к предыдущему периоду – это нормально. Тип клиентской лояльности по методу исследования NPS:

- Промоутеры (оценка 9 или 10). Люди, которые довольны продукцией или услугой компании. Рекомендуют компанию другим и сами готовы обратиться повторно. В большинстве случаев промоутеры являются адвокатами бренда: они готовы поддержать компанию в случае, если кто-то критикует её.

- Нейтралы (оценка 7 или 8). Эти люди могут не оставить негативный отзыв, но и рекомендовать компанию не будут. Если им удастся получить скидку у конкурента или же ему сделают выгодное предложение по снижению цены, они уйдут к нему.

- Критики или детракторы (оценки от 0 до 6). Не любят продукты, которые производит компания, или услуги, которые она оказывает. Скорее всего, они не будут повторно обращаться в компанию, и даже могут написать негативный отзыв. Если клиент уйдет молча, то вам никогда не узнать причины его поведения. Его замечания говорят о том, что он ожидает изменений в работе и готов к диалогу и при грамотном вступлении в такой диалог из него можно сделать промоутера.

Критики. Посетители, которые оставляют отзывы (даже отрицательные) – готовы к контакту и дают обратную связь, которая поможет улучшению бизнеса. Следовательно, единственным правильным решением является установление личного контакта со всеми критиками. Этот вариант можно осуществить в e-mail-рассылке, задавая уточняющие вопросы и собирая предложения по изменениям. Например, может быть стоит сменить

транспортную компанию или курьерскую службу, если продукт устраивает, а доставка нет.

Нейтралы. Этот сегмент является одним из самых сложных, поскольку неопределившиеся клиенты обычно не дают развернутой обратной связи. Они ищут более выгодные предложения и будут с вами только до тех пор, пока не найдут что-то лучшее. Нужно выяснить, что их больше всего в данный момент интересует, а что отталкивает. Можно сделать им уникальное предложение для того, чтобы подчеркнуть их значимость. Не забывайте напоминать о себе, например, рассылкой со специальными предложениями или новыми товарами и услугами.

Промоутеры. Нужно почаще поощрять эту часть аудитории. Желательно поблагодарить человека и сказать ему пару слов благодарности или сделать подарок за отзыв в соцсетях. Также, в качестве подарка для промоутеров можно использовать брендированную продукцию или предоставить скидку на покупку товара клиентам, которые будут покупать у вас по рекомендациям ваших промоутеров.

RFM метод. RFM-анализ определяет клиентов, которые имеют наибольшее и меньшее значение для организации. Анализ делается на основе давности покупок, их частотой и денежной стоимостью и позволяет предсказать, кто из клиентов в будущем вероятнее всего совершат покупки снова. [27][15][26].

Идея заключается в том, чтобы разделить клиентов на группы, а затем работать с ними по-разному. Тогда можно предлагать различным группам различные товары и услуги, в наибольшей степени отвечающие требованиям покупателей. Подчеркнем также то, что даже для небольшого проекта повышения лояльности необходимо учитывать интересы десятков разных групп. Например, с помощью сегментирования по критерию «Пол» (мужчина или женщина), получаем две группы. Далее введите в группу «Возраст», разделенную на пять сегментов, и получим уже десять сегментов ($2*5=10$). Введем критерий «Доход» (добавив ещё пять категорий), тогда проявляется

уже 50 сегментов. Легко представить на сколько увеличится количество сегментов при добавлении нового критерия, и как это увеличит трудоемкость процесса сегментирования.

Модель RFM имеет три основных фактора:

- Recency (Новизна, недавность) – когда последний раз клиент совершал сделку;
- Frequency (Частота) – как много покупок совершается;
- Monetary Value (Денежная стоимость) – сумма расходов на все совершенные покупки.

Для клиентов может быть создана таблица со столбцами из ФИО клиента, даты последней покупки и его расходах.

Имеются различные способы количественного определения значения RFM. Наилучшие подходы будут зависеть от "пути" клиента в организацию и бизнес-модели. В качестве одного из подходов в RFM используется присвоение оценки каждому из трех факторов по шкале от 1..10, где наилучшее значение - 10. Тогда бизнес по предоставлению услуги может быть основан на расчетах:

- Давность = 10 – сколько времени прошло с момента последней покупки клиента. В частности, «давность» может быть разделена на три категории.: – те кто совершил покупки в течение последних 90 дней; – от 91 до 365 дней; – более 365 дней.;

- Частота = максимальное количество покупок, совершенных клиентом за последние 12 месяцев с ограничением 1..10;

- Денежная = стоимость заказов клиента, выраженная по отношению к определенной контрольной сумме.

Если Денежный фактор дал 10 баллов ежегодному расходованию более 500 долларов клиенту, который сделал три покупки в прошлом году и последнюю покупку 3 месяца назад, потратив 600 долларов за год, то его оценки будут: R = 7; F = 3; M - 10. Подобные категории могли бы быть

получены из правил или используя методы интеллектуального анализа данных для поиска значимых разрывов.

После определения всех атрибутов согласно их категориям, на пересечении их возможных значений возникает сегмент.д Если для каждой категории имеется три уровня, то результирующая матрица будет иметь двадцать семь комбинаций. В одном известном коммерческом подходе используется пять значений, тогда получим 125 сегментов. Для того, чтобы создать предложение интересное для компании и привлекательное для клиентов, необходимо учесть особенности каждого из возможных сегмента.

Совокупность сегментов может быть упорядочена от наиболее ценных, где самые высокие новизна, частота и стоимость, до самых низких значений. В некоторых случаях компании могут решить свернуть некоторые сегменты, если количество клиентов в них кажется слишком маленьким для того, чтобы они были полезными.

На сегодняшний день существует множество вариантов реализации метода RFM.

RFD (Новизна, частота, продолжительность) – модификация RFM, которая используется для анализа поведения потребителей при просмотре контента или другой читательской активности. Например, можно узнать о том, сколько онлайн времени провел человек в Википедии.

RFE (Новизна, частота, вовлеченность) – более широкая версия RFD, в которой вовлеченность определена как длительность посещения, количество изученных страниц за одно посещение или другие аналогичные показатели.

RFM-I (Новизна, частота, денежная стоимость взаимодействия) – данный формат применяется для учета новизны и частоты маркетинговых контактов с клиентами. Как правило это происходит в тех случаях, когда необходимо контролировать возможные сдерживающие эффекты от рекламных контактов.

RFMTC (Новизна, частота, денежная стоимость, время, коэффициент оттока) – новая модификация, созданная И. Ченгом и др., представленная в

2009 году. Данная модель, используя теорию вероятностей, предлагает формулы, определяющие вероятность покупки клиентами при следующей рекламной кампании. Модель была использована А. Иоаннидисом, при обработке данных о переливании крови и CDNOW [17].

Составление RFM-сегментов достаточно сложная работа. При правильной коммуникации с сегментами повышается эффективность сохранения клиентов и увеличение повторных покупок. На протяжении всего времени существования принципа Вильфредо Парето 80/20 он очень хорошо работал с продажами и маркетингом. Он означает, что, изучив 20% своих лучших клиентов и узнав их лучшие качества легче понять как привлечь других клиентов и увеличить свой доход.

RFM анализ предоставляет возможность посмотреть на клиентов под новым углом: увидеть группы самых лояльных и привлекающих больше всего денег пользователей, а также группы наиболее пассивных клиентов. Можно использовать RFM-анализ для организации коммуникации с клиентами и способствовать их перемещению в другие более лояльные группы. RFM анализ поможет организации выставить акценты и ответить на четыре основных вопроса:

- С какими потребителями необходимо расстаться в первую очередь, чтобы не тратить на них впустую время и деньги?
- Какие группы потребителей наиболее вероятно откликнутся на рекламу?
- Какая группа покупателей является лучшей и принесёт максимальную прибыль компании?
- Какова эффективность рекламных кампаний, которые проводятся для различных групп потребителей?

По мнению экспертов RFM можно использовать в любом бизнесе независимо от направления деятельности фирмы. Для того чтобы получить максимальную наглядность, необходимо иметь базу от 10 000 адресов.

Главный плюс этого подхода – его простота, а также прозрачность сегментации.

2.3 Использование алгоритмов машинного обучения для определения лояльности

На данный момент фирмы решают проблему повышения лояльности клиентов только с помощью максимального её упрощения. В этом случае, как правило, просто предоставляется скидка всем клиентам. Если руководители организации решают, что нужно изменить свою работу с клиентами и разработать программу повышения лояльности, то для ее запуска необходимо решить ряд вопросов [8][10][14][1][24]:

- определить наиболее важных клиентов, сегменты на которые они делятся и отличия одной группы от другой;
- выявить причины поведения покупателей и определить структуру их потребления;
- определить способы воздействия на поведение клиентов;
- определить уровень лояльности клиентов и факторы, которые могут ее изменить;
- другое.

Для этой задачи часто используют различные алгоритмы машинного обучения. Например: регрессионный анализ, случайные леса и нейронные сети и др. [11][18][19][20][23]. Необходимые данные могут быть получены из демографических данных клиентов, а также информации об их истории покупок клиентов или пользованием услугами организации. В конце концов, цель определения лояльности клиентов — это предсказание будут ли клиенты продолжать отношения с организацией или же прервут их и уйдут к конкурентам. Для бизнеса этот вопрос очень важен, т. к. уровень лояльности клиентов увеличивает прибыль компании.

Как следует из названия, машинное обучение предоставляет машине возможность обучаться самостоятельно на основе опыта и наблюдениях на заданном наборе данных, и, главное, без явного программирования. Обратим внимание на то, что в машинном обучении используется набор данных. На нем компьютер будет учиться распознавать шаблоны из заданного набора и самостоятельно принимать решения на основе информации, полученной из набора данных [3][13][28].

Компьютеры учатся также, как это делают обычные люди. Люди получают знания об определенных вещах, затем, имея эти знания, могут определять эти же вещи снова. Имея прошлый опыт, когда уже принимались правильные решения, есть большая вероятность опять принять их. Именно поэтому при передаче информации/данных на ЭВМ, они делятся по две неравные части: информация для обучения (training data) и данных для тестирования (testing data). На данных для обучения компьютер учится принимать правильные решения по шаблонам и особенностям, которые были получены в ходе обучения. Чтобы выяснить, насколько точно машины могут принимать такие решения и как они работают, их прогнозы тестируются на оставшихся данных, не участвовавших в обучении [20].

Все алгоритмы машинного обучения можно разделить на:

- Контролируемые (управляемые) алгоритмы (Supervised Algorithms), обучение с учителем: регрессия, машина опорных векторов, деревья решений, случайный лес и др.;
- Неуправляемые алгоритмы (Unsupervised Algorithms), обучение без учителя: K-means кластеризацию, нейронные сети и др.

Набор данных, который используется машиной для обучения в алгоритмах обучения с учителем, состоит из помеченных данных. Это значит, что в состав входят входные параметры и правильный результат. Если управляемые (контролируемые) алгоритмы обучения имеют дело с помеченными данными для изучения, то неконтролируемые алгоритмы

машинного обучения (обучение без учителя) не используют помеченные данные [19][20].

Данные определенной группы будут объединяться в кластеры на основе сходства между переменными (факторами). Используются при этом, как правило, такие алгоритмы машинного обучения, как кластеризация K-средних и нейронные сети.

Кластерирование – поиск закономерности для разделения совокупности объектов на однородные группы (кластеры), т.е. поиск существующих структур данных в данных. Построение кластеров (кластеризация) — это объединение точек данных, которые сгруппированы вместе из-за их сходства.

В методе K-means точка относится к одному из центроидов (центральная точка каждого кластера). Их число определяется человеком для каждой конкретной задачи. Алгоритм кластеризации K-means работает следующим образом (рисунок 1):

- Определить значение K. Если K равно 2, то будет два центроида.
- Случайно выбирается K точек, значения которых выбираются в качестве начальных центроидов.
- Вычисляется расстояние от каждой точки данных до каждого центроида, а затем точку данных назначается центроиду, с которым она имеет минимальное расстояние. Таким образом формируется группа похожих точек данных.
- Пересчитывается центроид для каждого вновь сформированного кластера, и точка данных переназначается кластеру, центроид которого находится на минимальном расстоянии от точки данных.

Чтобы оптимизировать алгоритм можно ограничить количество итераций на последнем шаге.

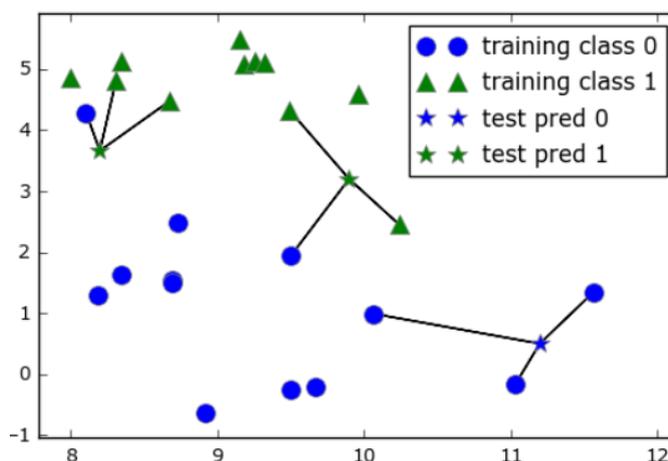


Рисунок 1 – K-means кластеризация [20]

Этапы Data Mining [11][19][20][28]:

- Очистка данных (Clean Data). Необходимо очистить информацию перед тем, как формировать матрицу. Удаляем выбросы, пустые значения и др., все, что является информационным шумом и не влияет на поведение покупателей. Когда данные подготовлены приступаем к созданию таблицы (матрицы) клиентов.

- Формируем матрицу с входными данными. Записываем в таблицу входные данные и формируем матрицу. Результаты кластеризации сильно зависят от того какой период времени выбран. Если выбрать короткий период, то в результате получим только текущие тренды. Перед Новым годом появятся кластеры, невидимые на протяжении длительного времени. Например, кластер «Любителей салата Оливье» и «Сельди под шубой». Кластеризация длительного периода позволит видеть общую картину и выделить клиентов со стабильными поведением. Можно кластеризировать клиентов на разных уровнях категоризации, например: «Студенты», «Домохозяйки», «Пенсионеры» и т. д. Отметим, что количество кластеров будет зависеть от размеров созданной матрицы и числа факторов.

- Выбираем оптимальное количество кластеров. Число кластеров рассчитывается экспериментально на основе опыта. Малое количество

кластеров не эффективно. Это может привести к тому, что будет получено, например, всего два «мегакластера», в которые будут включены 98 % клиентов и несколько маленьких, но бесполезных кластера для оставшихся 2 % покупателей. В случае большого количества кластеров возможно будет слишком много маленьких групп. Никто не пожелает заниматься анализом 500 отдельных мелких кластеров, ведь в каждом конкретном случае нужно учитывать особенности и нюансы данной группы клиентов.

- Выполняем кластеризацию. Выбрав алгоритм K-Means, запускаем кластеризацию по полученной матрице.

- Анализируем результаты кластеризации. В результате выполнения алгоритма все клиенты программы лояльности будут маркированы в зависимости от их поведения или покупок. Примем, те клиенты, которые имеют схожие поведенческие характеристики, попадающие в один кластер.

Когда кластеризация проводится на всем периоде, все клиенты компании станут её участниками. Выполнение кластеризации за определенный период (год, месяц и т.п.) будет проводиться только над клиентами, совершившими покупки в заданном промежутке времени.

С помощью анализа характеристик всех составляющих кластеров, можно понять, какие из них наиболее подходят для данной цели [24]:

- Название. В принципе, не обязательно давать кластерам названия, достаточно просто пронумеровать их. Но лучше все-таки дать им названия в зависимости от поведения клиентов внутри каждого сектора. В качестве примера можно привести такие названия как: «Домохозяйки», «Молодожены», «Холостяки».

- Оборот. Позволяет оценить количество и качество кластеров, приносящих наибольший доход.

- Доля кластера в обороте. Отношение к общему обороту в выбранном периоде.

- Общее число клиентов в кластере и число новых клиентов в нем.

- Процентное соотношение мужчин и женщин в кластере. Благодаря этому можно учесть типично мужские и специфические женские покупки.
- Общее число чеков в кластере и число чеков на одного клиента. Позволяет отследить, сколько раз клиент делал повторные покупки за выбранный период.
- Среднее число товаров и средняя сумма в чеке. Определяет кластер, в котором продаются самые дорогие товары.

После анализа каждого кластера и формирования «образа» покупателя можете перейти к персонализированной рассылке. При использовании кластеризации вы получите четкую систему рекомендаций для персонала, что именно и когда продавать.

Выводы главы 2

По мнению экспертов, исследование лояльности должно учитывать направление и отрасль деятельности компании. Сезонность спроса на товары и социально-демографические характеристики потребителей, так же влияет на общий уровень лояльности.

Компания может создавать профили клиентов, учитывая различные факторы: возраст, пол, местожительства, уровень доходов, язык и образование, хобби и др. Клиенты распределяются в отдельные сегменты. С помощью профиля клиента бренд может сосредоточиться на определенной группе клиентов, определить их потребности и решить проблемы.

В таблице 2 представлено сравнение методов, которые используются на практике.

Наиболее широко используется модель RFM, особенно в области розничной торговли и корпоративных услуг. Это классический метод классификации потребительской ценности.

Таблица 2 – Сравнение методов

Метод	Трудоемкость	Уровень подготовленности	Объем данных
NPS	Низкая	Не требуется	Небольшой
RFM	Средняя	Опытный пользователь MS Excel	Средний
K-means	Высокая	Аналитик данных	Большой

В исследовании RFM наиболее ценные сегменты будут выявлены с помощью скрытых взаимосвязей в данных о продажах. По мнению сторонников этого метода, его преимущество заключается в простоте: не нужно специального статистического программного обеспечения, а результаты легко понятны руководству. Даже если нет других методов таргетирования, это поможет повысить процент откликов на рекламу.

Обсуждая эффективность программы лояльности, используются различные методы оценки, учитывающие разные аспекты поведения клиентов: частота закупок; время последней покупки; отклик на предложения товаров и услуг.

Методы Data Mining и статистическая обработка данных для оценки эффективности позволяют использовать различные оценки лояльности. Например: переход клиентов из сегмента «лояльные клиенты» в категорию «неблагонадежных» и обратно.

Метод K-means используется при длительных периодах и большого количества учитываемых факторов. Известно, если клиентам дать оценку по трем показателям в RFM-методе, то будет сформировано 27 сегментов. Использование алгоритма K-Means позволяет уменьшить количество сегментов до ограниченного количества сегментов (обычно 5 или более). Сегмент покупателя определяется стоимостью покупок, частота покупок, давность покупок. Также могут учитываться другие факторы.

Глава 3 Разработка приложения для оценки лояльности клиентов

3.1 Разработка приложения RFM-сегментирование

Для проведения RFM-анализа нужны данные о покупках, совершенных клиентами. Как правило, они загружаются из CRM или платформ для аналитики. Зачастую данные необходимо предварительно обрабатывать.

Все строки, которые содержат данные, являются отдельными транзакциями. Число столбцов в таблице может быть разным, но обязательно должны быть следующие:

- идентификатор клиента (уникальное значение, например, e-mail, номер телефона, id-счетчик);
- дата совершения покупки;
- стоимость покупки.

Эти данные позволят сформировать таблицу для анализа:

- идентификатор клиента (уникальное значение, например, e-mail, номер телефона, id-счетчик);
- дата последней покупки;
- количество покупок;
- стоимость покупок.

Действовать можно по следующему алгоритму.

Шаг 1. Сбор информации по покупателям.

На каждого покупателя собирается информация для анализа: промежуток времени со дня последней покупки (Recency), число покупок (Frequency) и стоимость всех покупок (Monetary). Чтобы получить требуемые данные, нужно сформировать запрос к исходной таблице и выполнить его.

Шаг 2. Определяем диапазоны сегментирования.

Необходимо выбрать диапазон сегментирования по каждому показателю «Recency», «Frequency» и «Monetary». Имеется как минимум три подхода к этому:

– равномерное сегментирование по диапазонам значений. По этому принципу деление на сегменты происходит исходя из заданных значений показателей. Посмотрите на диапазон значений по каждому показателю и определите максимальное значение в каждом показателе. Разделим его на количество сегментов, получаем сегменты с равным диапазоном по каждому из показателей. Удобство подхода в том, что можно выделить сегменты, но при этом распределение клиентов по разным сегментам будет, скорее всего, неравномерным.

– равномерное сегментирование по количеству клиентов. Для этого механизма деление производится таким образом, чтобы в каждый сегмент попало одинаковое число клиентов. Данная сегментация дает возможность быстро выделить отдельные сегменты, но так чтобы не было сильного дисбаланса между группами клиентов.

– сегментирование с фиксированным диапазоном. При сегментации с фиксированным диапазоном необходимо самостоятельно определить границы значений по каждому показателю. Преимущество такой сегментации состоит в том, что деление действительно качественное, и клиенты равномерно попадают в свои сегменты. Процесс подбора сегментов трудоемкий, требует много времени и привлечения специалиста. Необходимо учитывать: – сезонность; – время между совершенными покупками; – среднюю продолжительность активности клиента; – акции; – праздники, каникулы и выходные; – др. Для их учета специальной формулы нет, это чисто логические размышления.

Определяя границы интервалов, можно опираться: в «Recency» – Средний период между заказами и, отдельно, время между первым, и вторым заказом.; в «Monetary» – на значение среднего чека; во «Frequency» – учитывать среднее число заказов отдельного клиента. Например.

– «Recency»: последняя покупка была: более 6 месяцев назад; 3–6 месяцев назад; 0–3 месяца назад.

– «Frequency»: зависит от средней частоты покупок и может иметь значения: 1–2 заказа; 3–5 заказов; более 6 заказов.

– «Monetary»: в зависимости от средней суммы покупок, может быть: 52 000 руб.; 36 000 руб.

При определении «Frequency», необходимо помнить о том, что среднее время между первым и вторым заказом и всеми остальными заказами будет разным. Клиент ещё не привык к бренду и будет дольше решаться на вторую покупку, если вообще решится. В случае использования просто среднего времени между заказами «Новички» попадут в группу «развивающихся» клиентов, и работа с ними будет построена неправильно. Кроме этого, для «Frequency» нужно учитывать сезонность и тип продукта.

Чтобы определить параметры диапазонов для «Monetary», требуется учитывать такие факторы как: тип бизнеса, цена, средний чек и сезонность. Совокупность покупок – 50 000 тысяч в год для магазина косметики, со средней суммой около 10 000 тысяч рублей - очень хорошо, а для мебельного салон не подходит.

Шаг 3. Формирование сегментов по показателям: R, F, M.

Когда определены диапазоны для показателей, нужно присвоить им оценку (коэффициент качества). Можно использовать любой вариант, например от 1 до 3. Количество может быть больше и зависит от того насколько детально будет прорабатываться каждый сегмент. Три - стандартное значение.

Примем, что значение 3 – самое плохое, а значение 1 – лучшее. В итоге получается три оценки по каждому из трех показателей. На основе значений трех показателей складывается сегменты с одинаковыми оценками:

- постоянные: R1 – F1 – M1;
- развивающиеся: R1 – F2 – M(1..3);
- новички: R1 – F3 – M(1..3);
- лояльные спящие клиенты: R2 – F1 – M(1..3);
- спящие: R2 – F(2, 3) – M(1..3);

- лояльные клиенты, потерявшие активность: R3 – F(1, 2) – M(1..3);
- потерянные: R3 – F3 – M(1..3).

В статье [11] клиенты распределены по сегментам (таблица 3) и для различных сегментов «XXX» предлагается рекламная стратегия. Средние значения («двойки») отдельно не выделяются, так как к ним подходят стандартные маркетинговые стратегии.

Таблица 3 – Распределение клиентов по сегментам

Сегмент	Характеристики сегмента	Маркетинговая стратегия
111 – «Ядро»	Часто совершали покупки на крупные суммы.	Можно предложить что-то ценное, пригласить на мероприятие, выразить благодарность. Только не надо предлагать скидки.
x1x – «Лояльные»	Тратят понемногу, зато регулярно.	Прорекламирровать сопутствующие товары, предложить подарок или бонусы за покупку, чтобы стимулировать показатель «Monetary».
xx1 – «Киты»	Тратят большие суммы, но редко – ценный «зверь».	Получить информацию о том какие товары были куплены клиентами и предложить товары из аналогичной категории. Узнать, чего бы они хотели от фирмы в будущем. Заинтриговать чем-то необычным, ценным, дорогим, но скидки предлагать не стоит.
13x при низкой оценке «3» – «Новички»	Нет гарантии того, что все они станут лояльными. Однако именно на этот сегмент стоит обратить особое внимание.	Подсказать, как правильно выбрать товар, дать полезную информацию. В качестве варианта – поздравить с покупкой, пригласить в социальную сеть или же предложить выгодную акцию. Укрепление отношений с клиентами на долгие годы – это то, что нужно сделать для того, чтобы наладить отношения с ними всерьёз и надолго.
33x при низкой оценке «3» – «Сони»	Были хорошими покупателями.	Разобраться в причинах отсутствия активности и предложить новый продукт или бесплатный пробный период. Как можно больше рассказать о проводимых в данный момент акциях или распродажах. Возможно, это поможет наладить отношения с ними.
333 – «Потерянные»	Перестали совершать покупки.	Попробовать спровоцировать, чтобы вернуть в число активных клиентов. Главное, не настаивать, если нет отклика, то есть не тратить время зря.

В результате получается RFM-матрица (таблица 4) и каждый клиент оказывается присвоен конкретному RFM-сегменту «XXX».

Таблица 4 – RFM-матрица

111 недавние, частые, высокий чек	121 недавние, редкие, высокий чек	131 недавние, разовые, высокий чек
112 недавние, частые, средний чек	122 недавние, редкие, средний чек	132 недавние, разовые, средний чек
113 недавние, частые, низкий чек	123 недавние, редкие, низкий чек	133 недавние, разовые, низкий чек
211 спящие, частые, высокий чек	221 спящие, редкие, средний чек	231 спящие, разовые, высокий чек
212 спящие, частые, средний чек	222 спящие, редкие, средний чек	232 спящие, разовые, средний чек
213 спящие, частые, низкий чек	223 спящие, редкие, низкий чек	233 спящие, разовые, низкий чек
311 уходящие, частые, высокий чек	321 уходящие, редкие, высокий чек	331 уходящие, разовые, высокий чек
312 уходящие, частые, средний чек	322 уходящие, редкие, средний чек	332 уходящие, разовые, средний чек
313 уходящие, частые, низкий чек	323 уходящие, редкие, низкий чек	333 уходящие, разовые, низкий чек

Довольно часто для анализа сегментов используют по два показателя из RFM-анализа:

– RF-анализ. Анализ на основе давности покупок и частоты их совершения (рисунок 2). Поскольку стоимость покупки (Monetary) зависит от частоты покупок (Frequency), иногда его можно не учитывать. На основе анализа RF сегментирование покажет, насколько часто клиент совершает покупку и позволяет выявить тех клиентов компании, которые недавно обращались за покупками или постоянно сотрудничали с компанией.

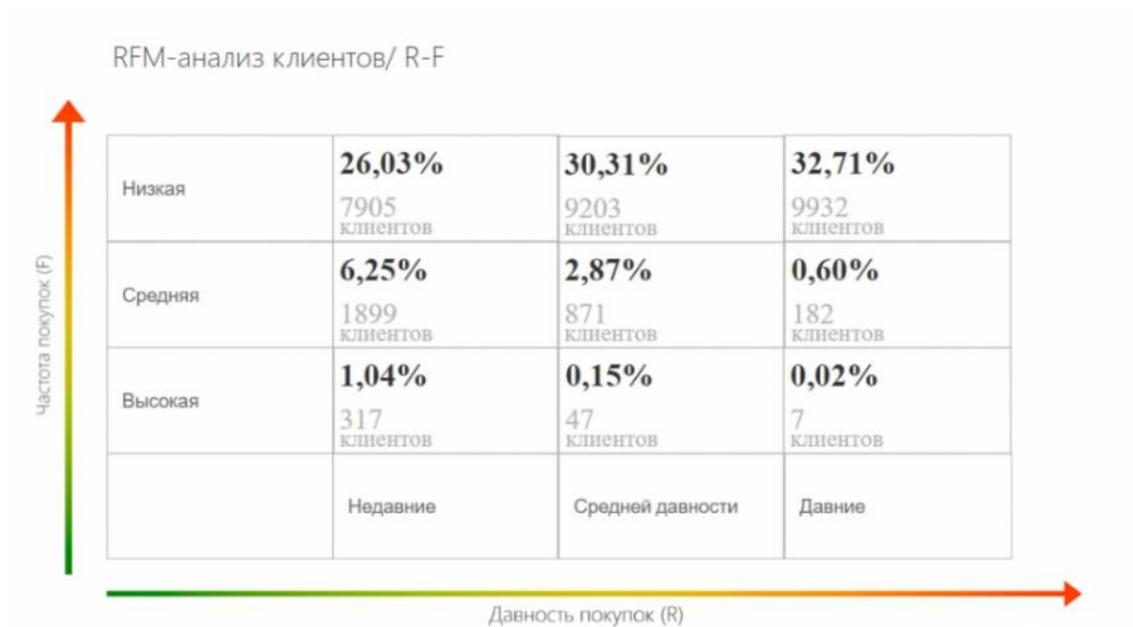


Рисунок 2 – Анализ по давности покупок и частоте покупок

– RM-анализ. Анализ демонстрирует распределение покупателей по параметрам «Recency» и «Monetary» (рисунок 3). В результате анализа можно выявить клиентов, которые приносят наибольшую прибыли и тех, вклад которых невелик.

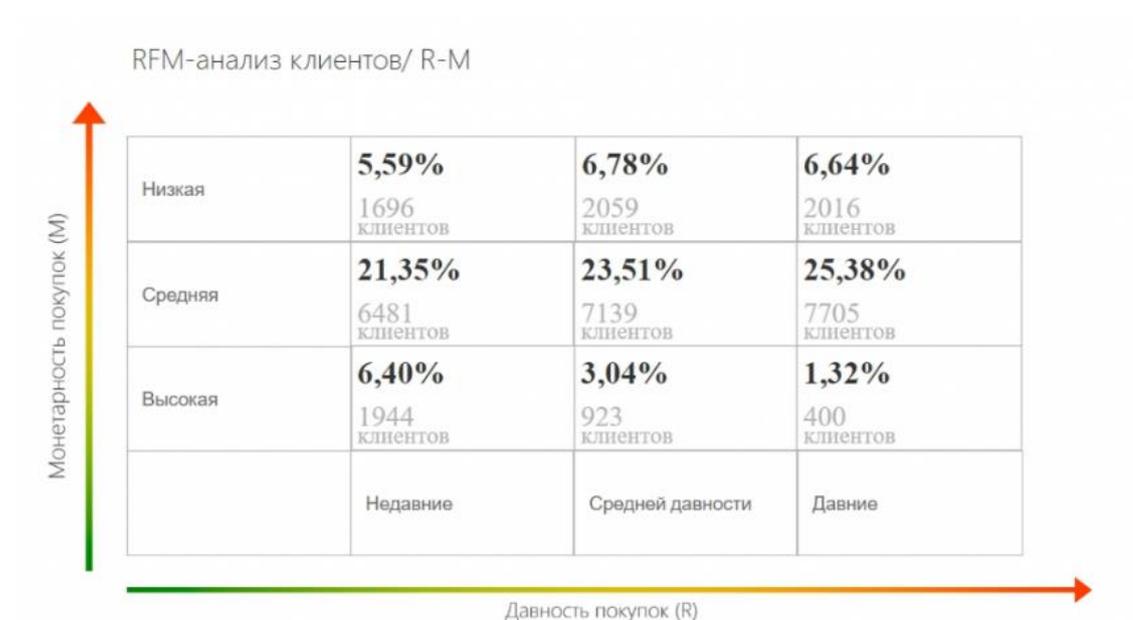


Рисунок 3 – Анализ по давности покупок и сумме покупок

– FM-анализ. На рисунке 4 показано распределение клиентов в зависимости от частоты (Frequency) и суммы покупок (Monetary), Такой анализ дает возможность определить покупателей с небольшими чеками, но частыми покупками, и клиентами с редкими покупками, но на большие суммы.

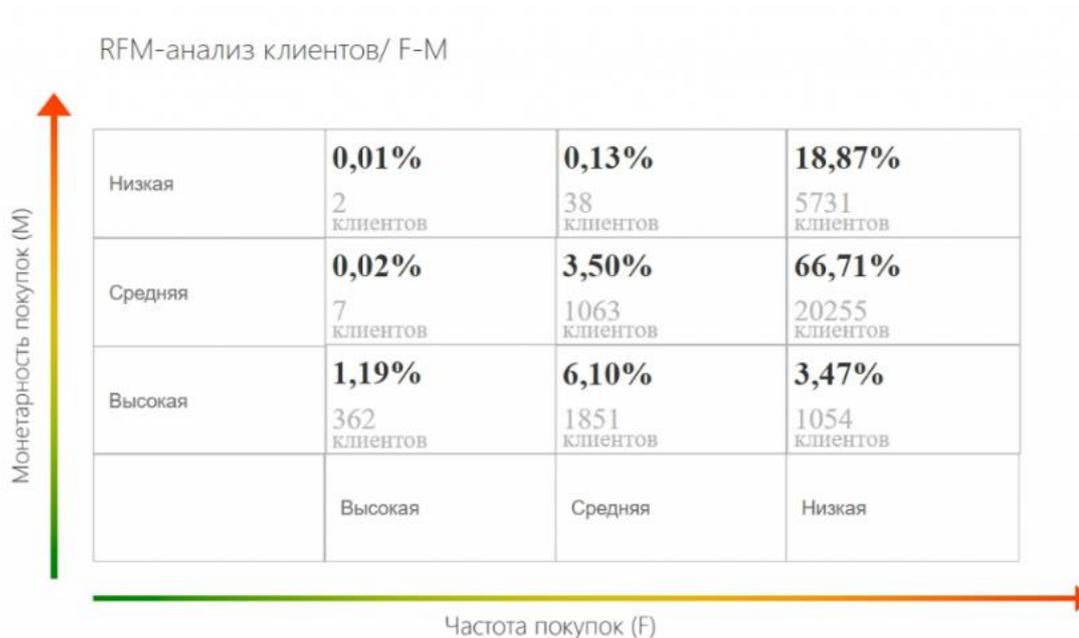


Рисунок 4 – Анализ по частоты покупок и сумме покупок

UML (Unified Modeling Language – унифицированный язык моделирования). UML содержит соглашения, которые облегчают процесс проектирования и обмена информацией в проектной группе. С помощью стандартизированной нотации можно сократить время на усвоение информации, облегчить общение и взаимодействие, упростить процесс документирования.

UML является графической нотацией, используемой при моделировании и описании процессов, возникающих при разработке. В основу UML входят различные типы диаграмм, предназначенных для моделирования различных аспектов разработки.

Диаграммы вариантов использования. Проектируемая система - это множество сущностей или актеров, которые взаимодействуют с системой при помощи прецедентов (рисунок 5).

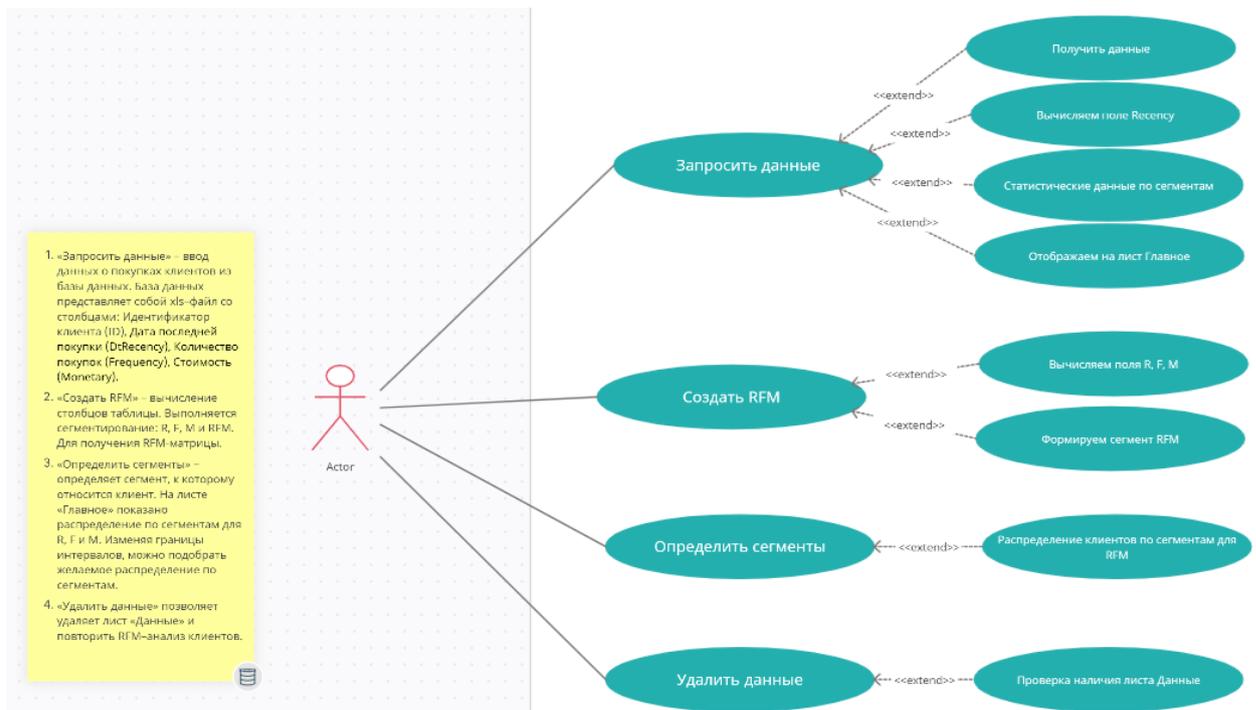


Рисунок 5 – Диаграммы вариантов использования

Актером (actor) или действующим лицом называют любую сущность, взаимодействующую с системой извне. В каждом варианте использования система определяет набор действий, которые должна совершить система при диалоге с актером. Причем, не говорится о том, как будет реализовано взаимодействие акторов и системы. [4].

Другой важной диаграммой UML для описания динамических элементов системы является диаграмма деятельности. Диаграмма действий – это блок-схема, показывающая, как одно действие ведет к другому. Действие можно назвать системной операцией. Одна операция приводит к следующей в потоке управления. Этот поток может быть параллельным, современным или разветвленным. Диаграммы действий используют множество функций, таких как разветвление, объединение и т. д., чтобы справиться со всеми типами управления потоком. Диаграммы деятельности служат схожим основным целям. Он фиксирует динамическое поведение системы (рисунок 6).

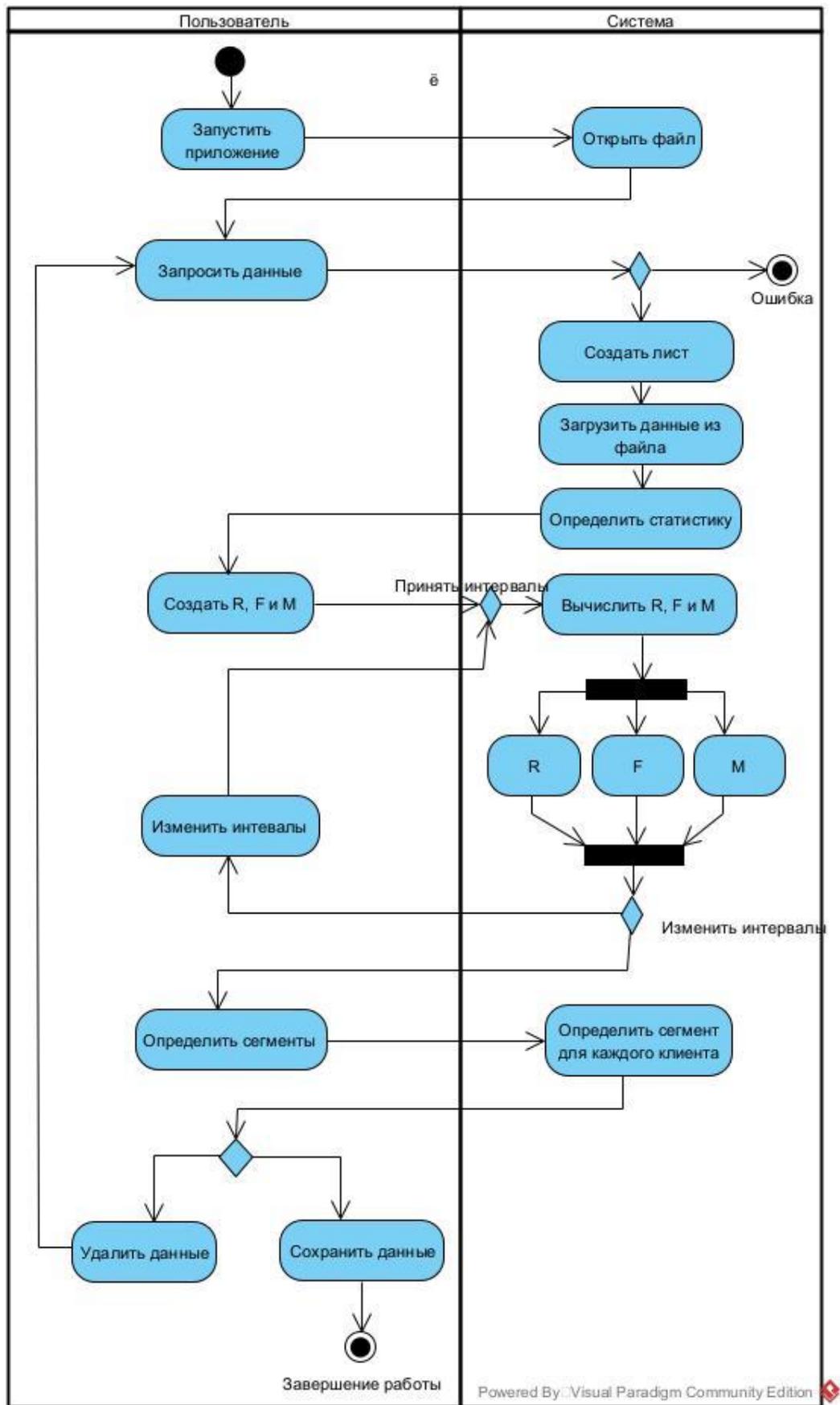


Рисунок 6 – Диаграмма действий

Работа начинается с первого листа. Всего листов три:

- Главная – отображает статистические данные и определяет интервалы сегментирования.
- Данные – на листе выводятся данные о покупках клиентов и вычисляемые значения давности, частоты и стоимости покупок.
- Сегменты клиентов (лист скрыт) – хранит таблицу наименований сегментов и их специфику (рисунок 7).

	B	C	D	E
1	111	111 недавние, частые, высокий чек	«Ядро»	Совершали покупки часто, недавно и на большие суммы. Войдут в
2	112	112 недавние, частые, средний чек	«Лояльные»	Тратят немного, но регулярно.
3	113	113 недавние, частые, низкий чек	«Лояльные»	Тратят немного, но регулярно.
4	121	121 недавние, редкие, высокий чек	«Киты»	Тратят большие суммы, но от случая к случаю. Это ценный «зверь».
5	122	122 недавние, редкие, средний чек	Развивающиеся	
6	123	123 недавние, редкие, низкий чек	Развивающиеся	
7	131	131 недавние, разовые, высокий чек	«Новички»	Тратят большие суммы, но от случая к случаю. Это ценный «зверь».
8	132	132 недавние, разовые, средний чек	«Новички»	
9	133	133 недавние, разовые, низкий чек	«Новички»	Не гарантируется, что все из них станут лояльными. Но на этот
10	211	211 спящие, частые, высокий чек	«Киты»	Тратят большие суммы, но от случая к случаю. Это ценный «зверь».
11	212	212 спящие, частые, средний чек	«Лояльные»	Тратят немного, но регулярно.
12	213	213 спящие, частые, низкий чек	«Лояльные»	Тратят немного, но регулярно.
13	221	221 спящие, редкие, средний чек	«Киты»	Тратят большие суммы, но от случая к случаю. Это ценный «зверь».
14	222	222 спящие, редкие, средний чек	Спящие	
15	223	223 спящие, редкие, низкий чек	Спящие	
16	231	231 спящие, разовые, высокий чек	«Киты»	Тратят большие суммы, но от случая к случаю. Это ценный «зверь».
17	232	232 спящие, разовые, средний чек	Спящие	
18	233	233 спящие, разовые, низкий чек	Спящие	
19	311	311 уходящие, частые, высокий чек	«Киты»	Тратят большие суммы, но от случая к случаю. Это ценный «зверь».
20	312	312 уходящие, частые, средний чек	«Лояльные»	Тратят немного, но регулярно.
21	313	313 уходящие, частые, низкий чек	«Лояльные»	Тратят немного, но регулярно.
22	321	321 уходящие, редкие, высокий чек	«Киты»	Тратят большие суммы, но от случая к случаю. Это ценный «зверь».
23	322	322 уходящие, редкие, средний чек	Потерявшие активность	
24	323	323 уходящие, редкие, низкий чек	Потерявшие активность	
25	331	331 уходящие, разовые, высокий чек	«Сони»	В прошлом были хорошими покупателями.
26	332	332 уходящие, разовые, средний чек	«Сони»	В прошлом были хорошими покупателями.
27	333	333 уходящие, разовые, низкий чек	«Потеряны»	Перестали совершать покупки.
28				

Рисунок 7 – Сегменты клиентов (лист хранит таблицу наименований сегментов и их специфику)

База данных представляет собой xls–файл со столбцами:

- Идентификатор клиента (ID),
- Дата последней покупки (DtRecency),
- Количество покупок (Frequency),
- Стоимость (Monetary).

Управление системой выполняется на странице «Главная» нажатием кнопок (рисунок 8).

	A	B	C	D	E	F	G	H
1	Дата проверки	10.10.2019	Начать с	1	11.10.2019			
2		Давность	Частота	Сумма				
3	Верхняя граница					Запросить данные		
4	Нижняя граница							
5	Среднее значение							
6								
7	Интервалы	Давность	Частота	Сумма		Создать RFM		
8	Граница верхняя	90	4	54 000				
9	Граница нижняя	240	2	32 000				
10								
11		Recency	Frequency	Monetary		Определить сегменты		
12								
13								
14						Удалить данные		
15								
16								
17								
18								
19								

Рисунок 8 – Страница «Главная»

При нажатии кнопки «Запросить данные» – ввод данных о покупках клиентов из базы данных.

При выполнении макроса создается лист «Данные» (рисунок 9) и вычисляется число дней с последней покупки (Recency). Значение вычисляется как разность даты формирования отчета ($E1=B1+D1$) и даты последней покупки ($DtRecency$ – столбец скрыт).

	A	B	C	D	E	F	G	H	I	J	K
1	Верхняя граница		9	8	760000					Дата	11.10.2019
2	Нижняя граница		639	1	10000						
3	ID	DtRecency	Recency	Frequen	Moneta						
4	Клиент 1	10.01.2018	639	1	140000						
5	Клиент 10	11.05.2018	518	1	10000						
6	Клиент 11	10.01.2018	639	3	25000						
7	Клиент 12	16.08.2019	56	3	76000						
8	Клиент 13	02.01.2019	282	5	129000						
9	Клиент 14	08.03.2019	217	7	56000						
10	Клиент 15	13.12.2018	302	4	85000						
11	Клиент 16	10.07.2019	93	1	760000						
12	Клиент 17	16.01.2019	268	2	34000						
13	Клиент 18	02.02.2019	251	4	21000						
14	Клиент 19	27.01.2019	257	6	96000						
15	Клиент 2	16.09.2019	25	3	30000						
16	Клиент 20	30.05.2019	134	4	11000						
17	Клиент 3	02.10.2019	9	7	43000						
18	Клиент 4	15.01.2019	269	3	56000						
19	Клиент 5	13.05.2018	516	1	87000						
20	Клиент 6	10.01.2018	639	8	170000						
21	Клиент 7	13.08.2019	59	5	30000						
22	Клиент 8	02.04.2019	192	3	36000						
23	Клиент 9	15.07.2018	453	6	76000						
24											
25											

Рисунок 9 – Страница «Данные»

При этом на листе «Главное» появляются статистические данные («Верхняя граница», «Нижняя граница» и «Среднее значение»), которые могут быть использованы для определения интервалов сегментирования (рисунок 10).

	A	B	C	D	E	F	G	H
1	Дата проверки	10.10.2019	Начать с	1	11.10.2019			
2		Давность	Частота	Сумма				
3	Верхняя граница	9	8	760000				
4	Нижняя граница	639	1	10000				
5	Среднее значение	291	4	98550				
6								
7	Интервалы	Давность	Частота	Сумма				
8	Граница верхняя	90	4	54 000				
9	Граница нижняя	240	2	32 000				
10								
11		Recency	Frequency	Monetary				
12								
13								
14								
15								
16								
17								
18								
19								

Рисунок 10 – Страница «Данные»

При нажатии кнопки «Создать RFM» – вычисление столбцов таблицы. Выполняется сегментирование: R, F, M и RFM (рисунок 11). Для получения RFM-матрицы.

	A	C	D	E	F	G	H	I	J	K	L
1	Верхняя граница	9	8	760000	90	4	54000		Дата	11.10.2019	
2	Нижняя граница	639	1	10000	240	2	32000				
3	ID	Recency	Frequen	Monetal	R	F	M	RFM			
4	Клиент 1	639	1	140000	3	3	1	331			
5	Клиент 10	518	1	10000	3	3	3	333			
6	Клиент 11	639	3	25000	3	2	3	323			
7	Клиент 12	56	3	76000	1	2	1	121			
8	Клиент 13	282	5	129000	3	1	1	311			
9	Клиент 14	217	7	56000	2	1	1	1211			
10	Клиент 15	302	4	85000	3	2	1	321			
11	Клиент 16	93	1	760000	2	3	1	231			
12	Клиент 17	268	2	34000	3	2	2	322			
13	Клиент 18	251	4	21000	3	2	3	323			
14	Клиент 19	257	6	96000	3	1	1	311			
15	Клиент 2	25	3	30000	1	2	3	123			
16	Клиент 20	134	4	11000	2	2	3	223			
17	Клиент 3	9	7	43000	1	1	2	112			
18	Клиент 4	269	3	56000	3	2	1	321			
19	Клиент 5	516	1	87000	3	3	1	331			
20	Клиент 6	639	8	170000	3	1	1	311			
21	Клиент 7	59	5	30000	1	1	3	113			
22	Клиент 8	192	3	36000	2	2	2	222			
23	Клиент 9	453	6	76000	3	1	1	311			
24											
25											

Рисунок 11 – Страница «Данные»

Кнопка «Определить сегменты» – определяет сегмент, к которому относится клиент (рисунок 12).

	A	C	D	E	F	G	H	I	J	K	L
1	Верхняя граница	9	8	760000	90	4	54000		Дата	11.10.2019	
2	Нижняя граница	639	1	10000	240	2	32000				
3	ID	Recency	Frequen	Monetal	R	F	M	RFM	Сегмен		
4	Клиент 1	639	1	140000	3	3	1	331	«Сони»		
5	Клиент 10	518	1	10000	3	3	3	333	«Потерянные»		
6	Клиент 11	639	3	25000	3	2	3	323	Потерявшие активность		
7	Клиент 12	56	3	76000	1	2	1	121	«Киты»		
8	Клиент 13	282	5	129000	3	1	1	311	«Киты»		
9	Клиент 14	217	7	56000	2	1	1	1211	«Киты»		
10	Клиент 15	302	4	85000	3	2	1	321	«Киты»		
11	Клиент 16	93	1	760000	2	3	1	231	«Киты»		
12	Клиент 17	268	2	34000	3	2	2	322	Потерявшие активность		
13	Клиент 18	251	4	21000	3	2	3	323	Потерявшие активность		
14	Клиент 19	257	6	96000	3	1	1	311	«Киты»		
15	Клиент 2	25	3	30000	1	2	3	123	Развивающиеся		
16	Клиент 20	134	4	11000	2	2	3	223	Спящие		
17	Клиент 3	9	7	43000	1	1	2	112	«Лояльные»		
18	Клиент 4	269	3	56000	3	2	1	321	«Киты»		
19	Клиент 5	516	1	87000	3	3	1	331	«Сони»		
20	Клиент 6	639	8	170000	3	1	1	311	«Киты»		
21	Клиент 7	59	5	30000	1	1	3	113	«Лояльные»		
22	Клиент 8	192	3	36000	2	2	2	222	Спящие		
23	Клиент 9	453	6	76000	3	1	1	311	«Киты»		
24											
25											

Рисунок 12 – Страница «Данные»

На листе «Главное» показывается распределение клиентов по сегментам для значений R, F и M (рисунок 13).

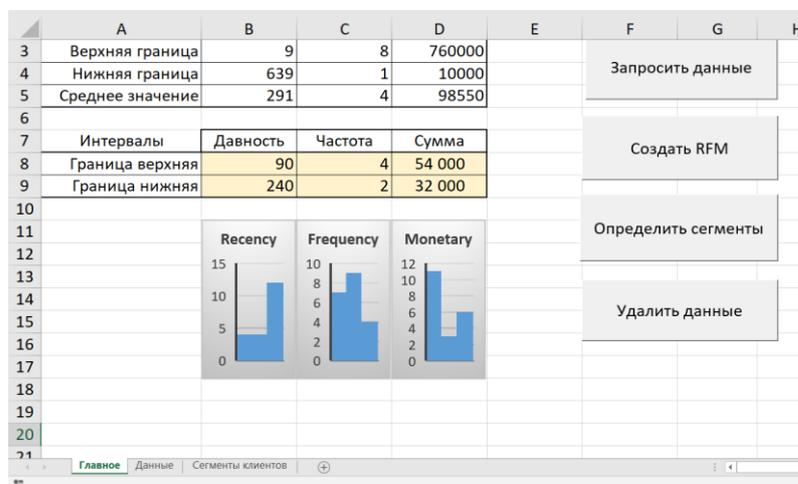


Рисунок 13 – Страница «Данные»

Изменяя границы интервалов, можно подобрать желаемое распределение клиентов по сегментам.

Кнопка «Удалить данные» позволяет удаляет лист «Данные» и повторить RFM–анализ клиентов (рисунок 14).

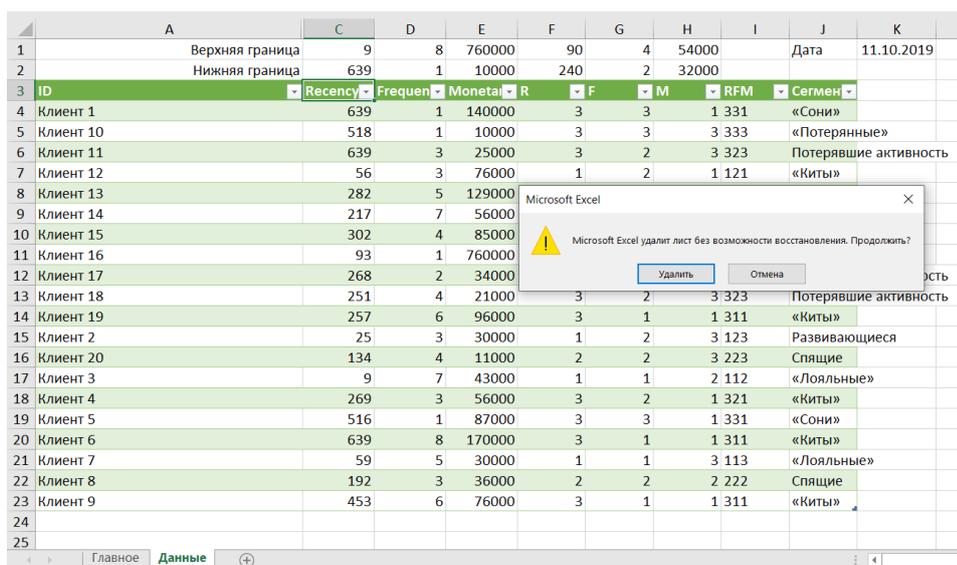


Рисунок 14 – Удаление страницы «Данные»

RFM - анализ предназначен для разработки маркетинговой стратегии компании с индивидуальным подходом ко всем сегментам. В этом случае можно сэкономить на «мертвых» и больше заработать на «живых».

3.2 Сегментирование клиентов методом K-Means

Например, тренировочный набор данных можно найти на сайте <https://www.kaggle.com/> [6]. Данные взяты из UCI: Online Retail, который содержит транзакции, произошедшие в онлайн-магазине розничной торговли, зарегистрированном в Великобритании в период с 1 декабря 2010 года по 9 декабря 2011 года. Компания в основном продает уникальные подарки на полный рабочий день. Многие клиенты компании являются оптовиками. Атрибуты набора данных приведены в таблице 5.

Таблица 5 – Атрибуты набора данных Online Retail

Имя поля	Краткая справка	Пояснение
InvoiceNo	номер счета-фактуры	6-значное целое число, уникально выделяемое для каждой транзакции. Если этот код начинается с буквы «с», это означает отмену.
StockCode	код продукта	5 - значное целое число, однозначно присваиваемое каждому отдельному продукту
Description	название продукта	строка текста
Quantity	количество каждого продукта в транзакции	число
InvoiceDate	дата и время выставления счета	генерирует дату и время каждой транзакции
UnitPrice	цена единицы измерения	цена товара в фунтах стерлингов
CustomerID	номер клиента	5 - значное целое число, однозначно присваиваемое каждому клиенту
Country	название страны	название страны или региона клиента

Как было отмечено ранее, при большом числе клиентов рекомендуется использовать методы Data Mining, и, конкретно, методы кластеризации (например, K – means). Для демонстрации работы метода будем использовать

свободно распространяемый пакет didroRFM. В нем реализован RFM–метод, разбивающий клиентов на пять групп. Имеется возможность гибкой настройки параметров сегментации. Пакет реализован на языке R.

Как уже оговорилось, требуется предварительная обработка данных. Для обработки данных будем использовать редактор Power Query (рисунок15). Создадим запрос для формирования требуемого набора данных.

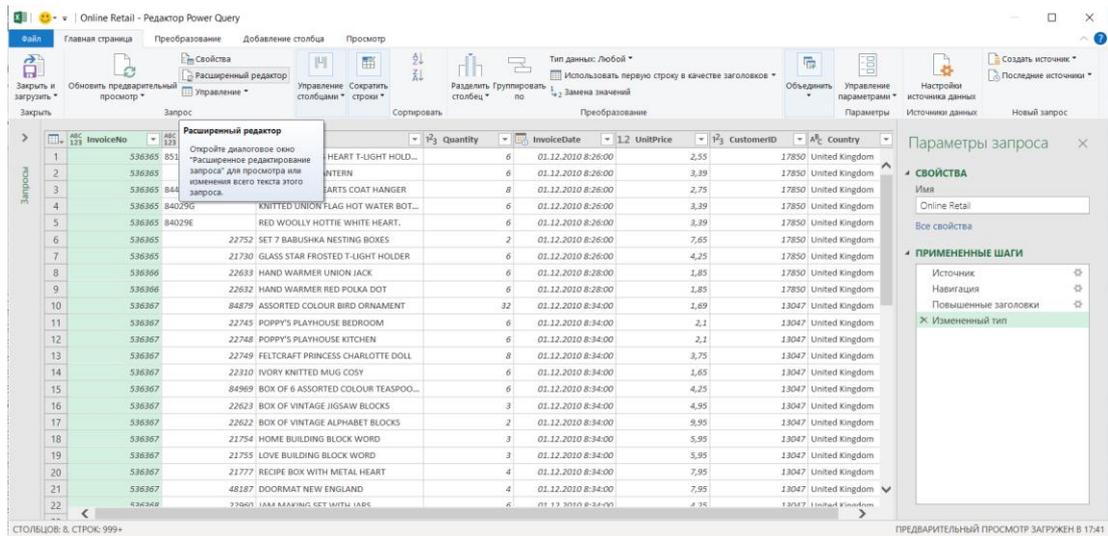


Рисунок 15 – Загрузка данных в редактор запросов Power Query

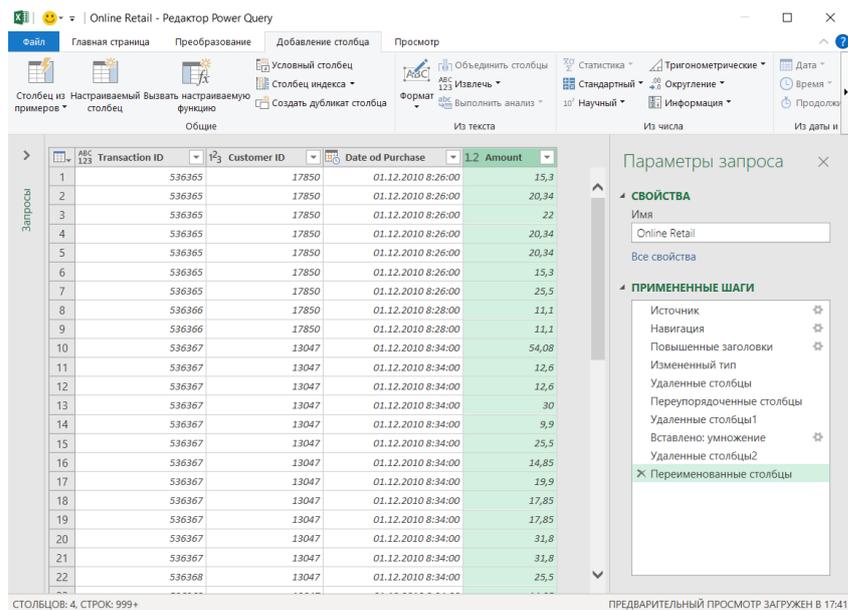


Рисунок 16 – Сформированный набор данных

Чтобы сформировать нужный для обработки набор данных преобразуем набор Online Retail с помощью запроса (рисунок 17).

Online Retail

```
let
    Источник = Excel.Workbook(File.Contents("C:\Users\Azerty\YandexDisk\Диплом\Кирил Соколов\Online Retail.xlsx"), null, true),
    #"Online Retail Sheet" = Источник{[Item="Online Retail",Kind="Sheet"]}[Data],
    #"Повышенные заголовки" = Table.PromoteHeaders(#"Online Retail Sheet", [PromoteAllScalars=true]),
    #"Измененный тип" = Table.TransformColumnTypes(#"Повышенные заголовки",{{"InvoiceNo", type any}, {"StockCode", type any}, {"Description", type any}}, {"InvoiceNo", "StockCode", "Description"}),
    #"Удаленные столбцы" = Table.RemoveColumns(#"Измененный тип",{"StockCode", "Description"}),
    #"Переупорядоченные столбцы" = Table.ReorderColumns(#"Удаленные столбцы",{"InvoiceNo", "CustomerID", "InvoiceDate", "Quantity", "UnitPrice"}),
    #"Удаленные столбцы1" = Table.RemoveColumns(#"Переупорядоченные столбцы",{"Country"}),
    #"Вставлено: умножение" = Table.AddColumn(#"Удаленные столбцы1", "Умножение", each [UnitPrice] * [Quantity], type number),
    #"Удаленные столбцы2" = Table.RemoveColumns(#"Вставлено: умножение",{"Quantity", "UnitPrice"}),
    #"Переименованные столбцы" = Table.RenameColumns(#"Удаленные столбцы2",{{"InvoiceNo", "Transaction ID"}, {"CustomerID", "Customer ID"}}, {"Transaction ID", "Customer ID"}),
in
    #"Переименованные столбцы"
```

Рисунок 17 – Текст запроса для формирования набора данных

Полученный набор позволит сформировать данные необходимые для анализа в пакете «didrooRFM. Окончательно получим таблицу с данными, приведенную на рисунке 18.

	A	B	C	D	E
1	Transaction ID	Customer ID	Date of Purchase	Amount	
2	536365	17850	01.12.2010 8:26	15,3	
3	536365	17850	01.12.2010 8:26	20,34	
4	536365	17850	01.12.2010 8:26	22	
5	536365	17850	01.12.2010 8:26	20,34	
6	536365	17850	01.12.2010 8:26	20,34	
7	536365	17850	01.12.2010 8:26	15,3	
8	536365	17850	01.12.2010 8:26	25,5	
9	536366	17850	01.12.2010 8:28	11,1	
10	536366	17850	01.12.2010 8:28	11,1	
11	536367	13047	01.12.2010 8:34	54,08	
12	536367	13047	01.12.2010 8:34	12,6	
13	536367	13047	01.12.2010 8:34	12,6	
14	536367	13047	01.12.2010 8:34	30	
15	536367	13047	01.12.2010 8:34	9,9	
16	536367	13047	01.12.2010 8:34	25,5	
17	536367	13047	01.12.2010 8:34	14,85	
18	536367	13047	01.12.2010 8:34	19,9	
19	536367	13047	01.12.2010 8:34	17,85	
20	536367	13047	01.12.2010 8:34	17,85	
21	536367	13047	01.12.2010 8:34	31,8	
22	536367	13047	01.12.2010 8:34	31,8	
23	536368	13047	01.12.2010 8:34	25,5	
24	536368	13047	01.12.2010 8:34	14,85	
25	536368	13047	01.12.2010 8:34	14,85	
26	536368	13047	01.12.2010 8:34	14,85	

Рисунок 18 – Таблица данных для проведения сегментации клиентов

Обязательно должны быть следующие поля:

- Уникальный номер транзакции (Transaction ID = InvoiceNo);
- уникальный идентификатор клиента (Customer ID = CustomerID);
- дата последней покупки (Date of Purchase = InvoiceDate);
- сумма покупок (Ammount = UnitPrice·Quantity).

Теперь можно выполнить сегментацию клиентов в среде RStudio.

Загрузим данные из csv-файла командой (рисунок 19, Приложение Б).

```
> data<-read.csv2('Customer RFM data.csv', header = TRUE, sep = ";", dec = ",")
```

Рисунок 19 – Команда для загрузки данных из csv-файла

На рисунке 20 показано окно RStudio после загрузки данных из файла Customer RFM data.csv. Было загружено 541 909 записей.

The screenshot shows the RStudio environment. The main window displays a data table with the following columns: Transaction.ID, Customer.ID, Date.of.Purchase, and Amount. The table contains 14 rows of data, with the first row being (1, 536365, 17850, 01.12.2010 8:26, 15.30). The console shows the following R commands and their output:

```
> data<-read.csv2('Customer RFM data.csv', header = TRUE, sep = ";", dec = ",")
Error in read.table(file = file, header = header, sep = sep, quote = quot
e, :
  'file' должен быть текстовой строкой строка или соединением
> data<-read.csv2('Customer RFM data.csv', header = TRUE, sep = ";", dec
= ",")
> View(data)
> View(data)
> View(Customer_RFM_data)
> data<-read.csv2('Customer RFM data.csv', header = TRUE, sep = ";", dec
= ",")
> View(Customer_RFM_data)
Error in View: object 'Customer_RFM_data' not found
> View(data)
```

The Environment pane on the right shows a data object named 'data' with 541909 observations and 4 variables. The Files pane shows the file 'read.csv' is open. The Data Input pane shows the 'read.table' function documentation.

Рисунок 20 – Таблица данных для проведения сегментации клиентов

Следующим шагом надо очистить данные. Необходимо преобразовать строки из столбца с Date.of.Purchase к типу Date. Выполним команды (рисунок 21).

```
> library(dplyr)
> str(data)
'data.frame':  541909 obs. of  4 variables:
 $ Transaction.ID  : chr  "536365" "536365" "536365" "536365" ...
 $ Customer.ID     : int   17850 17850 17850 17850 17850 17850 17850 17850
17850 13047 ...
 $ Date.of.Purchase: chr   "01.12.2010 8:26" "01.12.2010 8:26" "01.12.2010
8:26" "01.12.2010 8:26" ...
 $ Amount         : num  15.3 20.3 22 20.3 20.3 ...
```

Рисунок 21 – Команды для преобразования строки из столбца

Из ответа видно, что Date.of.Purchase: chr, т.е. поле имеет символьный тип. Исправим это командами (рисунок 22).

```
> data<-mutate(data, Date.of.Purchase = as.Date(Date.of.Purchase,
"%m.%d.%Y"))
> str(data)
'data.frame':  541909 obs. of  4 variables:
 $ Transaction.ID  : chr  "536365" "536365" "536365" "536365" ...
 $ Customer.ID     : int   17850 17850 17850 17850 17850 17850 17850 17850
17850 13047 ...
 $ Date.of.Purchase: Date, format: "2010-01-12" "2010-01-12" "2010-01-12" ...
 $ Amount         : num  15.3 20.3 22 20.3 20.3 ...
```

Рисунок 22 – Код для исправления типа поля

Теперь Date.of.Purchase: Date типа Date. Кроме того, в наборе встречаются строки с пустыми значениями. Имеются различные подходы к исправлению этой ошибки. Воспользуемся самым простым, просто удалим из набора все такие строки. Недостатком такого способа является то, что могут

быть удалены много строк. Осталось 172782 записи:

```
> data<-na.omit(data)
```

Проинсталлируем пакет didrooRFM и активируем его

```
> install.packages("didrooRFM").
```

Пакет 'didrooRFM' успешно распакован, MD5-суммы проверены. Скачанные бинарные пакеты находятся в C:\Users\Azerty\AppData\Local\Temp\Rtmpe8o4Ed\downloaded_packages

```
> library(didrooRFM)
```

Выполним сегментацию клиентов командой (рисунок 23).

```
> rfm<-findRFM(data)
> str(rfm)
rowws_df [3,125 × 16] (S3: rowwise_df/tbl_df/tbl/data.frame)
 $ CustomerID      : chr [1:3125] "12347" "12348" "12350" "12352" ...
 $ MeanValue       : num [1:3125] 24 73.4 19.7 17.5 35.3 ...
 $ LastTransaction : Date[1:3125], format: "2011-09-06" "2011-05-04"
 "2011-02-02" ...
 $ NoTransaction   : int [1:3125] 106 5 17 26 13 21 131 19 147 110 ...
 $ MonetaryPercentile : num [1:3125] 0.749 0.933 0.642 0.533 0.857 ...
 $ FrequencyPercentile : num [1:3125] 0.873 0.104 0.332 0.477 0.257 ...
 $ RecencyPercentile : num [1:3125] 0.546 0.22 0.101 0.16 0.539 ...
 $ MonetaryScore    : num [1:3125] 4 5 4 3 5 4 5 5 4 4 ...
 $ FrequencyScore   : num [1:3125] 5 1 2 3 2 3 5 2 5 5 ...
 $ RecencyScore     : num [1:3125] 3 2 1 1 3 3 2 5 5 5 ...
 $ MonetaryWeightedScore : num [1:3125] 16 20 16 12 20 16 20 20 16 16 ...
 $ FrequencyWeightedScore: num [1:3125] 20 4 8 12 8 12 20 8 20 20 ...
 $ RecencyWeightedScore : num [1:3125] 12 8 4 4 12 12 8 20 20 20 ...
 $ FinalScore       : num [1:3125] 48 32 28 28 40 40 48 48 56 56 ...
 $ FinalWeightedScore : num [1:3125] 4 2.67 2.33 2.33 3.33 ...
 $ FinalCustomerClass : chr [1:3125] "Class-4" "Class-2" "Class-2" "Class-2" ...
```

Рисунок 23 – Код для сегментации клиентов

На рисунке 24 и 25 приведены результаты работы. Выводится гистограмма распределения клиентов по сегментам. В наборе данных rfm видно, что получено 3125 клиентов.

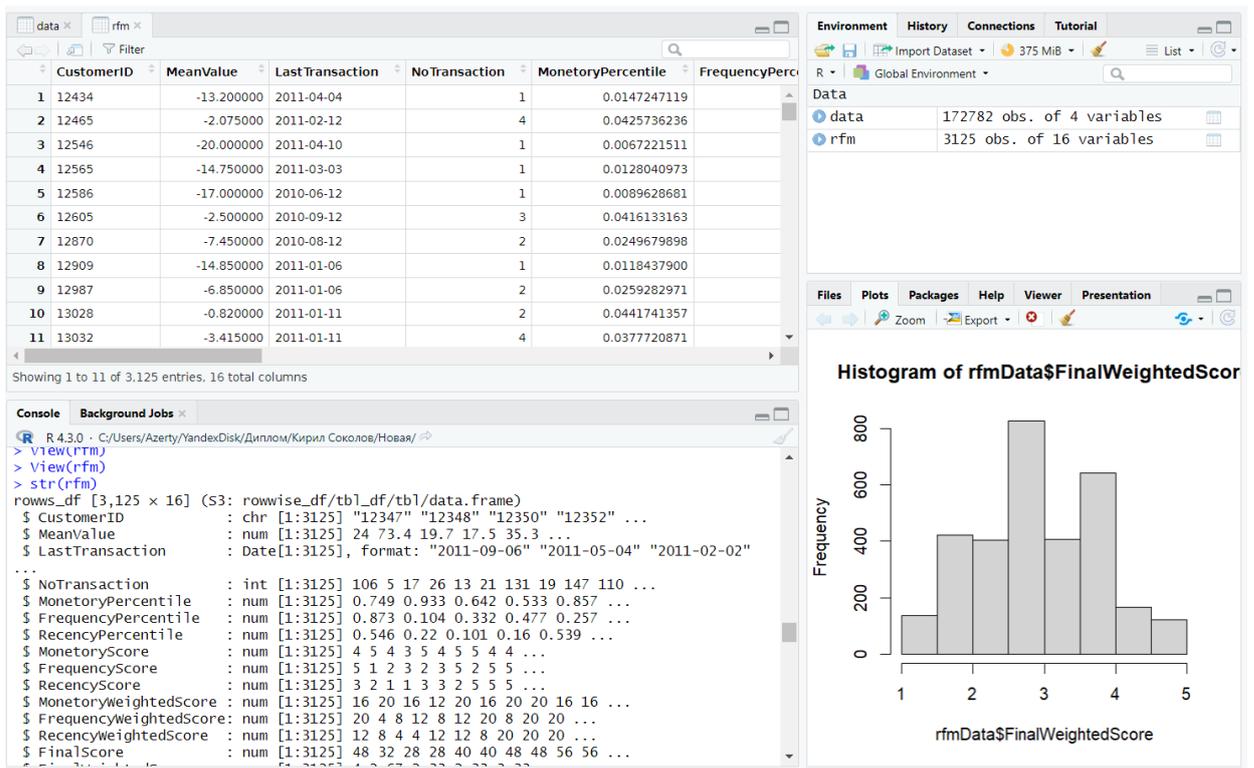


Рисунок 24 – Результат работы пакета didrooFRM/

	FrequencyWeightedScore	RecencyWeightedScore	FinalScore	FinalWeightedScore	FinalCustomerClass
6	20	12	48	4.000000	Class-4
0	4	8	32	2.666667	Class-2
6	8	4	28	2.333333	Class-2
2	12	4	28	2.333333	Class-2
0	8	12	40	3.333333	Class-3
6	12	12	40	3.333333	Class-3
0	20	8	48	4.000000	Class-4
0	8	20	48	4.000000	Class-4
6	20	20	56	4.666667	Class-4
6	20	20	56	4.666667	Class-4
2	8	16	36	3.000000	Class-3
8	12	4	24	2.000000	Class-2
2	8	8	28	2.333333	Class-2
6	16	16	48	4.000000	Class-4
6	16	20	52	4.333333	Class-4
6	8	16	40	3.333333	Class-3
6	8	4	28	2.333333	Class-2
6	8	12	36	3.000000	Class-3
2	20	4	36	3.000000	Class-3
4	4	8	16	1.333333	Class-1

Рисунок 25 – Информация о классах клиентов

Выводы глава 3

Благодаря подробному анализу клиента можно понять поведение клиентов и их модели покупок, поскольку это помогает узнать о вашем целевом клиенте.

Разделение клиентов на отдельные группы выгодно, поскольку это поможет в разработке стратегии маркетингового плана для конкретной группы сразу. Данные о клиентах разделяются на основе нескольких факторов, таких как возраст, местоположение, пол, язык и т.д., а затем определяются их требования. Персонализированные маркетинговые кампании приведут к повышению лояльности клиентов, что поможет процветанию бизнеса.

Разработчики маркетинговой стратегии должны использовать RFM-анализ для разработки индивидуальной стратегии с индивидуальным подходом к каждому сегменту. Это позволит сэкономить на «мертвых» и больше заработать на «живых». Например, даем клиентам оценку по трем показателям (R, F, M), тогда в общей сложности будет 27 комбинаций.

Для длительных периодов и большого числа кластеров лучше использовать метод K-means, так как он позволяет уменьшить количество сегментов до определенного ограниченного числа сегментов (обычно 5).

Применение кластеризации позволяет бизнесу выстраивать отношения с клиентами и вести работу с ними по-новому. Персонализированный подход к каждому клиенту будет способствовать повышению лояльности потребителей, и несомненно поможет увеличить прибыль.

Заключение

На данный момент добиться стабильного роста рентабельности компании и запланированного уровня продаж невозможно без привлечения значительного числа постоянных клиентов. Для компании клиентская база является одним из самых важных активов, а доверие клиентов один из основных показателей успешности бизнеса. Не существует единого инструмента анализа данных, который бы решал все возможные проблемы с анализом данных.

Рассмотренные методики позволяют сделать вывод, что единого метода оценки потребительской лояльности не существует, а осуществление комплексного программно-ориентированного управления лояльностью клиентов возможно только при сочетании ключевых методик диагностики и контроля лояльности.

Разделение клиентов на отдельные группы выгодно, поскольку это поможет в разработке стратегии маркетингового плана для конкретной группы сразу. Данные о клиентах разделяются на основе нескольких факторов, таких как возраст, местоположение, пол, язык и т.д., а затем определяются их требования. Персонализированные маркетинговые кампании приведут к повышению лояльности клиентов, что поможет процветанию бизнеса.

Благодаря подробному анализу клиента можно понять поведение клиентов и их модели покупок, поскольку это помогает узнать о вашем целевом клиенте.

При анализе лояльности необходимо учитывать направление и сферу деятельности компании, сезонность спроса на товары и социальные характеристики потребителей, так как данные показатели влияют на общий уровень лояльности.

Разработчики маркетинговой стратегии должны использовать RFM-анализ для разработки индивидуальной стратегии с индивидуальным

подходом к каждому сегменту. Это позволит сэкономить на «мертвых» и больше заработать на «живых».

Применение кластера позволит бизнесу выстраивать новые отношения с клиентами и вести работу в новом ключе. Наличие индивидуального подхода повысит лояльность потребителей, а также увеличит прибыль.

Используем K-means для длительных периодов и большого количества кластеров. Если клиентам дается оценку по трем показателям с тремя уровнями, то в общей сложности будет 27 комбинаций. Посредством алгоритма K-Means, число сегментов можно уменьшить до определенного количества (обычно пять категорий).

Разработаны два приложения, реализующие сегментацию клиентов. Первое (Приложение А) позволяет работать с относительно небольшой базой данных о клиентах средствами MS Excel. Достоинством является то, что от сотрудника не требуется специальных навыков, кроме навыков работы с MS Excel. Второе приложение (Приложение Б) требует специальных навыков по работе с данными.

Получены результаты кластеризации клиентов.

Таким образом, можно говорить, что все поставленные задачи решены, цель работы достигнута.

Список используемой литературы и используемых источников

1. Анализ клиента. URL: <https://www.marketing91.com/customer-analysis/> (дата обращения 17.04.2023)/
2. Барнс Д. Путь к сердцу клиента. Стратегия отношений, когда лояльности мало. М.: Изд-во Юрайт, 2009.
3. Бринк Хенрик, Ричардс Джозеф, Феверолф Марк. Машинное обучение. – СПб.: Питер, 2017. – 336 с.: ил. – (Серия «Библиотека программиста»).
4. Буч Гради, Рамбо Джеймс, Якобсон Ивар Язык UML. Руководство пользователя. – М.: ДМК Пресс, 2007. – 388 с.
5. Венэблз У.В. Н., Смит Д.М. Рабочая группа разработки R Введение в R Заметки по R: среда программирования для анализа данных и графики Версия 3.0.2 Patched (2014-01-27) Перевод и редакция А.А.Фоменко. – Москва, 2014. 109 с. – (серия технической документации).
6. Данные о покупках клиентов с платформы Kaggle: <https://www.kaggle.com/jihyeseo/online-retail-data-set-from-uci-ml-repo> (дата обращения 17.04.2023)/
7. Индекс лояльности NPS: подробный гайд и советы экспертов. URL: <https://skillbox.ru/media/marketing/my-izmerili-indeks-loyalnosti-klientov-nps/> (дата обращения 17.04.2023).
8. Кластеризация: расскажи мне, что ты покупаешь, и я скажу кто ты? URL: <https://habr.com/ru/companies/datawiz/articles/248863/> (дата обращения 17.04.2023)/
9. Ламбен Ж.Ж. Менеджмент, ориентированный на рынок // Маркетинг в России и за рубежом, 2008. 146 с.
10. Лояльность клиентов: что это и как её увеличить? URL: https://kontur.ru/compass/spravka-compass/25088-loyalnost_klientov_chno_eto_i_kak_eyo_uvelichit (дата обращения 17.04.2023)

11. Мастицкий С.Э., Шитиков В.К. (2014) Статистический анализ и визуализация данных с помощью R. URL: <http://r-analytics.blogspot.com> (дата обращения 17.04.2023).
12. Методы оценки лояльности покупателей торговых компаний. URL: <https://web.snauka.ru/issues/2017/01/77372> (дата обращения 17.04.2023).
13. Мюллер А., Гвидо С. - Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными – 2017.pdf. 393.с. URL: https://github.com/amueller/introduction_to_ml_with_python (дата обращения 17.04.2023).
14. Применение Data Mining для повышения лояльности клиентов. URL: <https://basegroup.ru/community/articles/data-mining-loyalty> (дата обращения 17.04.2023).
15. Разделяем клиентов по лояльности с помощью RFM-анализа. URL: <https://habr.com/ru/articles/497356/> (дата обращения 17.04.2023).
16. Райхельд Ф., Томас Т. Движущие силы экономического роста, прибыли и непреходящей ценности. М.: Издательский дом Вильямс, 2008. 345 с.
17. Реализация с использованием набора данных CDNOW. URL: <https://github.com/it21208/RFMTC-Implementation-Using-the-CDNOW-dataset> (дата обращения 17.04.2023).
18. Роберт И. Кабаков R в действии. Анализ и визуализация данных в программе R / пер. с англ. Полины А. Волковой. – М.: ДМК Пресс, 2014. – 588 с.
19. Шипунов А. и др. Наглядная статистика. Используем R! – М.: ДМК Пресс, 2014. – 298 с.
20. Шитиков В.К., Мастицкий С.Э. (2017) Классификация, регрессия и другие алгоритмы Data Mining с использованием R. 351 с. URL: <https://github.com/ranalytics/data-mining> (дата обращения 17.04.2023)
21. 9 лучших инструментов анализа данных для аналитиков данных в 2023 году <https://careerfoundry.com/en/blog/data-analytics/data-analytics-tools/> с

22. A Beginner's Guide to Customer Behavior Analysis. URL: <https://blog.hubspot.com/service/customer-behavior-analysis> (дата обращения 17.04.2023).

23. An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics) 1st ed. 2013, Corr. 7th printing 2017 Edition, Kindle Edition. 426 с.

24. How to Create Detailed Buyer Personas for Your Business [Free Persona Template]. URL: https://blog.hubspot.com/marketing/buyer-persona-research?hubs_content=blog.hubspot.com%2Fservice%2Fcustomer-behavior-analysis&hubs_content-cta=buyer%20personas (дата обращения 17.04.2023).

25. NPS: зачем, как и когда измерять лояльность к бренду. URL: <https://blog.click.ru/growthhacking/nps-zachem-kak-i-kogda-izmeryat-loyalnost-k-brendu/> (дата обращения 17.04.2023).

26. RFM (Recency, Frequency, Monetary Value), Explained in 800 Words or Less. URL: https://blog.hubspot.com/service/rfm-analysis?hubs_content=blog.hubspot.com%2Fservice%2Fcustomer-behavior-analysis&hubs_content-cta=RFM%20analysis (дата обращения 17.04.2023).

27. RFM-анализ: что такое в маркетинге + шаблон в Excel. URL: <https://in-scale.ru/blog/rfm-analiz/> (дата обращения 17.04.2023).

28. Stergiou C., Siganos D. Neural Networks [Electronic resource] // Neural Networks. URL: http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html (дата обращения 17.05.2023).

Приложение А
Макросы Excel

Function SheetExist(WbName As String, ShName As String) As Boolean

'Проверка наличия листа

Dim mySheet As Worksheet

For Each mySheet In Workbooks(WbName).Sheets

If mySheet.Name = ShName Then

SheetExist = True

Exit Function

End If

Next

End Function

Sub УдалитьЛистДанные()

'

If SheetExist(ThisWorkbook.Name, "Данные") Then

Sheets("Данные").Select

ActiveWindow.SelectedSheets.Delete

End If

Sheets("Главное").Select

Range("B3:D5").Select

Selection.ClearContents

Range("B1").Select

End Sub

Sub ЗагрузитьДанные()

Продолжение Приложения А

```
If SheetExist(ThisWorkbook.Name, "ДАННЫЕ") Then
```

```
    Sheets("ДАННЫЕ").Select
```

```
    ActiveWindow.SelectedSheets.Delete
```

```
End If
```

```
Sheets("Главное").Select
```

```
Range("B3:D5").Select
```

```
Selection.ClearContents
```

```
Application.CutCopyMode = False
```

```
ActiveWorkbook.Worksheets.Add
```

```
With ActiveSheet.ListObjects.Add(SourceType:=0, Source:= _
```

```
    "OLEDB;Provider=Microsoft.Mashup.OleDb.1;Data
```

```
Source=$Workbook$;Location=ДАННЫЕ;Extended Properties=" _
```

```
    , Destination:=Range("$A$1")).QueryTable
```

```
    .CommandType = xlCmdSql
```

```
    .CommandText = Array("SELECT * FROM [ДАННЫЕ]")
```

```
    .RowNumbers = False
```

```
    .FillAdjacentFormulas = False
```

```
    .PreserveFormatting = True
```

```
    .RefreshOnFileOpen = False
```

```
    .BackgroundQuery = True
```

```
    .RefreshStyle = xlInsertDeleteCells
```

```
    .SavePassword = False
```

```
    .SaveData = True
```

```
    .AdjustColumnWidth = True
```

```
    .RefreshPeriod = 0
```

```
    .PreserveColumnInfo = True
```

```
    .ListObject.DisplayName = "ДАННЫЕ"
```

Продолжение Приложения А

```
.Refresh BackgroundQuery:=False
End With
Selection.Worksheet.Name = "ДАННЫЕ"
Worksheets("Главное").Move Before:=Worksheets("ДАННЫЕ")

Sheets("ДАННЫЕ").Select
' Range("ДАННЫЕ[[#Headers],[ID]]").Select

Rows("1:1").Select
Selection.Insert Shift:=xlDown, CopyOrigin:=xlFormatFromLeftOrAbove
Selection.Insert Shift:=xlDown, CopyOrigin:=xlFormatFromLeftOrAbove

Range("I1").Select
ActiveCell.FormulaR1C1 = "Дата"
Range("J1").Select
ActiveCell.FormulaR1C1 = "=Главное!R1C5"
Selection.NumberFormat = "m/d/yyyy"
Columns("J:J").ColumnWidth = 9.56

Columns("C:C").Select
Selection.Insert                               Shift:=xlToRight,
CopyOrigin:=xlFormatFromLeftOrAbove
Range("ДАННЫЕ[[#Headers],[Столбец1]]").Select
ActiveCell.FormulaR1C1 = "Recency"
Range("C4").Select
Application.CutCopyMode = False
ActiveCell.FormulaR1C1 = "=R1C11-[@DtRecency]"
```

Продолжение Приложения А

```
Range("Данные[Recency]").Select
```

```
Selection.NumberFormat = "0"
```

```
Columns("B:B").Select
```

```
Selection.EntireColumn.Hidden = True
```

```
Range("C1").Select
```

```
ActiveCell.FormulaR1C1 = "=MIN(Данные[Recency])"
```

```
Range("C2").Select
```

```
ActiveCell.FormulaR1C1 = "=MAX(Данные[Recency])"
```

```
Range("D1").Select
```

```
ActiveCell.FormulaR1C1 = "=MAX(Данные[Frequency])"
```

```
Range("D2").Select
```

```
ActiveCell.FormulaR1C1 = "=MIN(Данные[Frequency])"
```

```
Range("E1").Select
```

```
ActiveCell.FormulaR1C1 = "=MAX(Данные[Monetary])"
```

```
Range("E2").Select
```

```
ActiveCell.FormulaR1C1 = "=MIN(Данные[Monetary])"
```

```
Range("A1").Select
```

```
ActiveCell.FormulaR1C1 = "Верхняя граница"
```

```
Range("A2").Select
```

```
ActiveCell.FormulaR1C1 = "Нижняя граница"
```

```
Range("F1").Select
```

```
With Selection
```

```
    .HorizontalAlignment = xlRight
```

```
    .VerticalAlignment = xlBottom
```

```
    .WrapText = False
```

```
    .Orientation = 0
```

Продолжение Приложения А

```
.AddIndent = False  
.IndentLevel = 0  
.ShrinkToFit = False  
.ReadingOrder = xlContext  
.MergeCells = False
```

End With

```
Range("A1:A2").Select
```

With Selection

```
.HorizontalAlignment = xlRight  
.VerticalAlignment = xlBottom  
.WrapText = False  
.Orientation = 0  
.AddIndent = False  
.IndentLevel = 0  
.ShrinkToFit = False  
.ReadingOrder = xlContext  
.MergeCells = False
```

End With

```
Sheets("ДАННЫЕ").Select
```

```
Columns("A:A").ColumnWidth = 20.33
```

```
Columns("A:A").ColumnWidth = 34.22
```

```
Columns("C:E").Select
```

```
Selection.ColumnWidth = 8.56
```

```
Range("K1").Select
```

```
Sheets("Главное").Select
```

```
Range("B3").Select
```

Продолжение Приложения А

```
ActiveCell.FormulaR1C1 = "=MIN(Данные[Recency])"  
Range("B4").Select  
ActiveCell.FormulaR1C1 = "=MAX(Данные[Recency])"  
Range("B5").Select  
ActiveCell.FormulaR1C1 = "=Average(Данные[Recency])"  
Range("C3").Select  
ActiveCell.FormulaR1C1 = "=MAX(Данные[Frequency])"  
Range("C4").Select  
ActiveCell.FormulaR1C1 = "=MIN(Данные[Frequency])"  
Range("C5").Select  
ActiveCell.FormulaR1C1 = "=Average(Данные[Frequency])"  
Range("D3").Select  
ActiveCell.FormulaR1C1 = "=MAX(Данные[Monetary])"  
Range("D4").Select  
ActiveCell.FormulaR1C1 = "=MIN(Данные[Monetary])"  
Range("D5").Select  
ActiveCell.FormulaR1C1 = "=Average(Данные[Monetary])"  
  
Range("D5").Select  
  
Range("E1").Select  
End Sub  
  
Sub СоздатьRFM()  
,  
  
Sheets("Данные").Select  
Range("F3").Select  
ActiveCell.FormulaR1C1 = "R"
```

Продолжение Приложения А

Range("F4").Select

'ЕСЛИ([@Recency]>=F\$2;3;ЕСЛИ([@Recency]<=F\$1;1;2))

ActiveCell.FormulaR1C1 = _

"=IF([@Recency]>=R2C6,3,IF([@Recency]<=R1C6,1,2))"

Range("G3").Select

ActiveCell.FormulaR1C1 = "F"

Range("G4").Select

ActiveCell.FormulaR1C1 = _

"=IF([@Frequency]<R2C7,3,IF([@Frequency]>R1C7,1,2))"

Range("H3").Select

ActiveCell.FormulaR1C1 = "M"

Range("H4").Select

ActiveCell.FormulaR1C1 = _

"=IF([@Monetary]<R2C8,3,IF([@Monetary]>R1C8,1,2))"

Range("I3").Select

ActiveCell.FormulaR1C1 = "RFM"

Range("I4").Select

ActiveCell.FormulaR1C1 = _

"=TEXT([@R],""#"") & TEXT([@F],""#"") & TEXT([@M],""#"")"

Range("F1").Select

ActiveCell.FormulaR1C1 = "=Главное!R[7]C[-4]"

Range("F2").Select

ActiveCell.FormulaR1C1 = "=Главное!R[7]C[-4]"

Range("F1:F2").Select

Продолжение Приложения А

```
Selection.Copy
Range("G1").Select
ActiveSheet.Paste
Range("H1").Select
ActiveSheet.Paste
' Range("ДАННЫЕ[#Headers],[M]").Select

Range("K1").Select
End Sub

Sub НазначитьСегменты()
'
Sheets("ДАННЫЕ").Select
Range("J3").Select
ActiveCell.FormulaR1C1 = "Сегмент"
Range("J4").Select
ActiveCell.FormulaR1C1 = "=VLOOKUP([@RFM],'Сегменты
клиентов'!R1C2:R27C5,3,0)"
End Sub
```

Приложение Б

R-код

```
> data<-read.csv2('Customer RFM data.csv', header = TRUE, sep = ";", dec = ",")
> View(data)
> library(dplyr)
> str(data)
'data.frame':  541909 obs. of  4 variables:
 $ Transaction.ID : chr  "536365" "536365" "536365" "536365" ...
 $ Customer.ID   : int   17850 17850 17850 17850 17850 17850 17850 17850 178
50 13047 ...
 $ Date.of.Purchase: chr  "01.12.2010 8:26" "01.12.2010 8:26" "01.12.2010 8:2
6" "01.12.2010 8:26" ...
 $ Amount        : num   15.3 20.3 22 20.3 20.3 ...
> data<-mutate(data, Date.of.Purchase = as.Date(Date.of.Purchase, "%m.%d.%Y")
)
> str(data)
'data.frame':  541909 obs. of  4 variables:
 $ Transaction.ID : chr  "536365" "536365" "536365" "536365" ...
 $ Customer.ID   : int   17850 17850 17850 17850 17850 17850 17850 17850 178
50 13047 ...
 $ Date.of.Purchase: Date, format: "2010-01-12" "2010-01-12" "2010-01-12" ...
 $ Amount        : num   15.3 20.3 22 20.3 20.3 ...
> data<-na.omit(data)
> install.packages("didrooRFM")
пакет 'didrooRFM' успешно распакован, MD5-суммы проверены

Скачанные бинарные пакеты находятся в
C:\Users\Azerty\AppData\Local\Temp\Rtmp8o4Ed\downloaded_packages

> library(didrooRFM)
> rfm<-findRFM(data)
> str(rfm)
rows_df [3,125 × 16] (S3: rowwise_df/tbl_df/tbl/data.frame)
 $ CustomerID      : chr [1:3125] "12347" "12348" "12350" "12352" ...
 $ MeanValue       : num [1:3125] 24 73.4 19.7 17.5 35.3 ...
 $ LastTransaction : Date[1:3125], format: "2011-09-06" "2011-05-04" "2
011-02-02" ...
 $ NoTransaction   : int [1:3125] 106 5 17 26 13 21 131 19 147 110 ...
 $ MonetaryPercentile : num [1:3125] 0.749 0.933 0.642 0.533 0.857 ...
 $ FrequencyPercentile : num [1:3125] 0.873 0.104 0.332 0.477 0.257 ...
 $ RecencyPercentile : num [1:3125] 0.546 0.22 0.101 0.16 0.539 ...
 $ MonetaryScore    : num [1:3125] 4 5 4 3 5 4 5 5 4 4 ...
 $ FrequencyScore    : num [1:3125] 5 1 2 3 2 3 5 2 5 5 ...
 $ RecencyScore      : num [1:3125] 3 2 1 1 3 3 2 5 5 5 ...
 $ MonetaryWeightedScore : num [1:3125] 16 20 16 12 20 16 20 20 16 16 ...
 $ FrequencyWeightedScore: num [1:3125] 20 4 8 12 8 12 20 8 20 20 ...
 $ RecencyWeightedScore : num [1:3125] 12 8 4 4 12 12 8 20 20 20 ...
 $ FinalScore       : num [1:3125] 48 32 28 28 40 40 48 48 56 56 ...
 $ FinalWeightedScore : num [1:3125] 4 2.67 2.33 2.33 3.33 ...
 $ FinalCustomerClass : chr [1:3125] "Class-4" "Class-2" "Class-2" "Class-
2" ...
```