

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА**

Факультет комп'ютерних наук та кібернетики
Кафедра системного аналізу та теорії прийняття рішень

Кваліфікаційна робота
на здобуття ступеня бакалавра
за освітньо-професійною програмою «Системний аналіз»
за спеціальністю 124 Системний аналіз
на тему:

**Впровадження та навчання прогнозованих моделей позитивної цінності
клієнта**

Виконав студент 4-го курсу
Максим ЛОСЬ

Науковий керівник:
асистент, кандидат тех. наук
Махно Михайло Федорович

Засвідчую, що в цій роботі немає запозичень з праць
інших авторів без відповідних посилань.

Студент

Роботу розглянуто й допущено до захисту на
засіданні кафедри системного аналізу та теорії
прийняття рішень

« 07 » _____ 06 _____ 2022 р., протокол № 10

Завідувач кафедри
Олександр НАКОНЕЧНИЙ

Київ – 2022

Зміст

Зміст.....	2
Вступ.....	4
1. Предиктивна аналітика.....	7
1.1 Що таке предиктивна аналітика і чому вона важлива?	7
1.2 Етапи предиктивної аналітики.	9
2. Пожиттєва цінність клієнта і її важливість у процесі прийняття рішень.	12
2.1 Ключові терміни для розуміння проблематики.....	12
2.2 Основні теоретичні підходи до виміру CLV	13
2.3 Спільноти клієнтів, їх роль у CRM та вплив на CLV.....	15
2.4 Різновид розрахунків CLV	21
2.5 Як застосовувати показник пожиттєвої цінності клієнта	23
2.6 CLV як ключовий елемент юніт-економіки.....	25
3. Моделі, які описані в науковій роботі.	27
3.1 Pareto/Negative Binominal distribution model.	27
3.1.1 Історія.....	27
3.1.2 Припущення моделі.	28
3.2.3 Ключовий проміжний результат	30
3.2 Beta Geometric/Negative Binominal distribution model	33
3.3 Modified Beta Geometric/Negative Binominal distribution model	35
3.4 Gamma-Gamma model.....	37
3.5 RFM – аналіз.....	39
4. Розробка	43
4.1 Підготовка даних.....	43
4.2 Побудова RFM моделі	44

4.3 Використання моделі Pareto/NBD	46
4.4 Використання моделі BG/NBD	50
4.5 Використання моделі MBG/NBD	54
4.6 Порівняння результатів прогнозування.....	57
4.7 Використання результатів.....	58
Висновки	60
Список використаних джерел	61

Вступ

На сьогоднішній день, процес прийняття рішень у бізнесі можливий двома способами: методом експертних оцінок — рішення приймаються, спираючись на досвід спеціаліста, його кваліфіковану думку; data-driven підходом — рішення приймаються, базуючись на аналітиці даних. Між ними є очевидна різниця – перший метод є суб’єктивним і в цьому його основний мінус, адже зазвичай в силу особистого досвіду чи можливо світогляду, людям притаманне когнітивне спотворення дійсності; в той час як другий – об’єктивний, і дозволяє на основі даних підтверджувати думки або гіпотези і таким чином уникати неякісних рішень. Найкращою практикою є комбінація цих двох методів, коли досвідченість експерта можна підкріпити наявними даними.

Не важливо якого розміру бізнес – дані завжди мають значення, і стають основою для швидкого росту і розвитку – адже розуміння своєї бізнес-моделі через цифри надає компаніям конкурентної переваги. За даними Forbes, 59% підприємств використовують аналітику даних у різний спосіб для прийняття кращих рішень та підвищення ефективності бізнесу.

На відміну від орієнтованості бізнесу 20 століття, який більше фокусувався на продукті і тому як його покращити, сучасний бізнес має інший вектор думки – йому необхідно знайти найкраще вирішення проблеми, яка турбує користувача. І однією з основних метрик цього підходу є CLV(Customer Lifetime Value) або «пожиттєва цінність клієнта». У широкому сенсі CLV можна визначити як очікуваний валовий прибуток компанії від одного клієнта протягом усього періоду співпраці (тобто всього його життя).

Предиктивна аналітика пожиттєвої цінності клієнта допомагає бізнесу відповісти на велику кількість питань, які стосуються маркетингового таргету,

покращення утримання клієнта, стимуляція повторних покупок, підвищення рентабельності, тощо.

Задля практичної реалізації були взяті публічні данні онлайн транзакцій магазину роздрібної торгівлі в Британії за 2010-2011 та 2011-2012 роки та були проведені підготовка даних, прогноз за найбільш відомими моделями, порівняння їх результатів, використання отриманого прогнозу у практичних цілях.

Мета роботи:

- Запровадження трьох різних моделей для прогнозування пожиттєвої цінності клієнта і порівняння їх ефективності.
- На основі найбільш ефективного прогнозу зробити сегментацію клієнтів для маркетингової кампанії.

Завдання:

- Дослідження існуючих моделей прогнозування пожиттєвої цінності клієнта.
- Запровадження цих моделей на прикладі реальної бази даних.
- Порівняння результатів моделей.
- Використання прогнозованих даних задля сегментації клієнтів.

Об'єкт дослідження: дані продажів онлайн магазину роздрібної торгівлі в Британії за 2010-2012 рік.

Предмет дослідження: моделі RFM, Pareto/NBD, BG/NBD, MBG/NBD та Gamma-Gamma.

Кваліфікаційна робота складається зі вступу, чотирьох розділів, висновку, списку використаної літератури. Список використаної літератури включає в себе 24 джерела. Робота виконана на 62 сторінках друкованого тексту.

У першому розділі розглянуто поняття предиктивної аналітики, її використання задля покращення бізнес процесів і конкурентної переваги.

У другому розділі надана детальна інформація про позитивну цінність клієнта, а саме: різноманітні підходи до розрахунків, використання метрики задля покращення бізнес процесів та важливість позитивної цінності клієнта в юніт-економіці.

У третьому проаналізовано основні моделі прогнозування. Детально досліджено які принципи і гіпотези закладені в основу кожного з них.

У четвертому описаний процес побудови прогнозованих моделей і надані результати роботи кожної з них, їх порівняння та застосування.

1. Предиктивна аналітика.

1.1 Що таке предиктивна аналітика і чому вона важлива?

Предиктивна аналітика – це галузь передової аналітики, яка робить прогнози щодо майбутніх результатів, використовуючи історичні дані у поєднанні зі статистичним моделюванням, методами видобутку даних та машинним навчанням. Компанії використовують предиктивну аналітику для пошуку закономірностей у цих даних з метою виявлення ризиків та можливостей.

Предиктивна аналітика часто асоціюється з великими даними та наукою про дані. Сьогодні компанії мають великі набори даних, що зберігаються у транзакційних базах даних, лог-файлах обладнання, зображеннях, відео, датчиках та інших джерелах даних. Щоб отримати уявлення про ці дані, фахівці з аналізу даних використовують алгоритми глибокого навчання та машинного навчання для пошуку закономірностей та прогнозування майбутніх подій. До них відносяться лінійна та нелінійна регресія, нейронні мережі, машини опорних векторів та дерева рішень. Результати, отримані в ході предиктивної аналітики, можуть бути використані в рамках прескриптивної аналітики для управління процесами на основі прогнозів[1].

Однією з значних причин розвитку предиктивної аналітики є зростаюча конкуренція. У разі зростаючої конкуренції компанії прагнуть отримати перевагу під час введення товарів та послуг на переповнені ринки. Прогностичні моделі, що базуються на даних, можуть допомогти компаніям вирішити давні проблеми новими способами.

Наприклад, виробникам обладнання важко впроваджувати інновації лише у галузі апаратного забезпечення. Розробники продуктів можуть додати

можливості прогнозування існуючих рішень, щоб підвищити цінність для клієнта. Використання предиктивної аналітики для обслуговування обладнання або предиктивного обслуговування дозволяє передбачати відмови обладнання, прогнозувати потреби в енергії та знижувати експлуатаційні витрати. Наприклад, датчики, що вимірюють вібрації автомобільних деталей, можуть сигналізувати необхідність технічного обслуговування до того, як автомобіль вийде з ладу в дорозі.

Компанії також використовують предиктивну аналітику для створення більш точних прогнозів, наприклад прогнозування попиту на електроенергію в електричній мережі. Ці прогнози дозволяють ефективніше планувати ресурси.

Приклади використання предиктивної аналітики

Автомобільна промисловість - прорив у галузі автономних транспортних засобів. Компанії, що розробляють технології допомоги водієві та нові автономні автомобілі, використовують предиктивну аналітику для аналізу даних датчиків підключених автомобілів та створення алгоритмів допомоги водієві.

Аерокосмічна промисловість - моніторинг стану авіаційних двигунів. Щоб підвищити час роботи літака та знизити витрати на обслуговування, виробники двигунів створили аналітичну програму в реальному часі для прогнозування роботи підсистем для масла, палива, зльоту, механічного стану та управління.

Виробництво енергії - прогнозування цін та попиту на електроенергію. Складні програми для прогнозування використовують моделі, що відстежують доступність установок, історичні тенденції, сезонність та погоду.

Фінансові послуги – розробка моделей кредитного ризику. Фінансові установи використовують методи машинного навчання та кількісні інструменти для прогнозування кредитного ризику.

Промислова автоматизація та обладнання - прогнозування відмов обладнання. Виробник пластику та тонких плівок економить 50 тис. євро щомісяця, використовуючи додаток для моніторингу стану та прогнозування технічного обслуговування, що скорочує час простою та мінімізує відходи.

Медичні прилади - Використання алгоритмів виявлення патернів для виявлення астми та ХОЗЛ. Пристрій для лікування астми записує та аналізує звуки дихання пацієнтів та забезпечує миттєвий зворотний зв'язок через програму для смартфона, щоб допомогти пацієнтам впоратися з астмою та ХОЗЛ[1].

1.2 Етапи предиктивної аналітики.

Термін "предиктивна аналітика" визначає застосування статистичних методів або методів машинного навчання для створення кількісного прогнозу майбутнього. Часто контрольовані методи машинного навчання використовуються для прогнозування майбутнього значення або для оцінки ймовірності.

Предиктивна аналітика починається з бізнес-мети: використовувати дані для скорочення відходів, економії часу або зниження витрат. Процес використовує різноманітні, часто масивні набори даних для створення моделей, які можуть генерувати чіткі, дієві результати для досягнення цієї мети, такі як зменшення відходів матеріалів, скорочення складських запасів та виробництво продукції, що відповідає специфікаціям. Весь процес аналітики можна розділити на такі етапи[2].

1. Визначення вимог.

Предиктивна аналітика зазвичай починається з бізнес-експерта, зосередженого на вирішенні проблеми. Процес розпочинається зі складання списку питань та розставляння пріоритетів кожного питання за ступенем важливості. Фокус може бути спрямований на уточнення конкретної мети з використанням результатів. Бізнес-користувачі можуть співпрацювати з аналітиком, щоб перевірити результат та встановити метрики для виміру успіху. Також важливо помістити ціль у ширший контекст.

2. Вивчення даних.

Аналітик даних повинен визначити, які набори даних доступні та як їх можна використовувати для покращення прогнозів та вирішення інших бізнес-завдань. Необхідно враховувати актуальність, придатність, якість та чистоту даних. Знання того, як і навіщо збираються дані, допоможе виявити будь-які проблеми до того, як інформація буде введена в модель предиктивного аналізу.

3. Розробка моделі

Розглянути, як різні моделі прогнозування можуть дати найбільш цінний результат, а потім провести зворотну роботу, щоб визначити якнайкращий спосіб організації вихідних даних у відповідний набір функцій для моделі. Потрібно експериментувати з різними функціями, алгоритмами та процесами, щоб знайти правильний баланс між продуктивністю, точністю та іншими вимогами, наприклад, зрозумілістю. Фахівець з аналізу даних може використовувати інструмент для перетворення вихідних даних на нові характеристики для розрахунку прогнозу.

4. Розгортання моделі

Розроблену прогностичну модель необхідно запустити у виробництво, щоб вона давала результати. Унікальна технологія в області предиктивного аналізу цінна лише у тому випадку, коли вона вносить значні зміни у поточний процес. Після того, як спеціаліст з аналізу даних затверджує модель, інженер з обробки даних визначає, як оптимізувати процес отримання, очищення та перетворення вихідних даних, необхідних для моделі у масштабі компанії. Потім фахівець з аналізу даних та інженер повинні пов'язати результат із відповідним додатком або робочим процесом.

5. Підтвердження результатів

Якщо робота з даними була виконана добре, модель предиктивної аналітики повинна відповідати вимогам до продуктивності, точності та іншим параметрам, коли вона розгортається на реальні дані. Однак продуктивність може змінюватися з часом. У деяких випадках зміни настрою клієнтів, бізнес-клімату або інших факторів можуть вплинути на продуктивність моделі. В інших випадках зловмисники можуть спробувати навмисне підірвати точність моделі. Незалежно від причини, корисно відстежувати зміни у продуктивності моделі та встановлювати певні граничні значення продуктивності для оновлення моделей.

2. Пожиттєва цінність клієнта і її важливість у процесі прийняття рішень.

2.1 Ключові терміни для розуміння проблематики.

Customer lifetime value (CLV) - пожиттєва цінність клієнта. У випадку сучасної економічної кризи, коли організація продажів стає дедалі складнішою, неминуче зростає значимість виміру ефективності маркетингової діяльності як основи раціоналізації структури маркетингових програм. У цьому особливе значення набуває концепція CLV, оскільки вона безпосередньо пов'язує маркетингову діяльність із фінансовими показниками компанії. Акцент на концепції CLV пов'язаний з тим, що цей інструмент є основою для прийняття в поточному періоді рішень, які приносять збільшення прибутку в майбутньому. Ця концепція є особливо корисною методикою для прогнозування майбутніх показників діяльності підприємства та розгляду перспектив максимізації прибутковості компанії. У цьому випадку CLV визначають як очікуваний валовий прибуток компанії від одного клієнта, що отримується протягом усього періоду його прихильності до фірми.

Customer relationship management (CRM) - управління взаємовідносинами із клієнтами. У центрі CRM - оптимізація прибутковості клієнта і CLV[3]. У найширшому тлумаченні CRM охоплює всі рівні організації – від стратегії до окремих операцій. У вузькому визначенні CRM - це клас ІТ-рішень, що забезпечують збирання та аналіз інформації про клієнта, що підтримують оперативну взаємодію та інтеграцію клієнта в процеси компанії.

Client community - клієнтська спільнота. Щодо клієнтської спільноти, то під ним помилково можуть розуміти акаунти компанії в соціальних мережах або просту підписку на сайт. Однак клієнтська спільнота (також клієнтська, або онлайн-спільнота бренду) - це набагато ширше явище, хоча часто воно могло розвинутись з текстового форуму на сайті компанії або групи в соціальних мережах. Таким чином, клієнтське співтовариство можна визначити як спеціалізоване, негеографічно прив'язане співтовариство, засноване на структурованому наборі соціальних відносин між шанувальниками бренду.

Клієнтська спільнота виявляється набагато дешевшою і простішою для підприємства у створенні та підтримці порівняно з традиційною офлайн-спільнотою бренду. Більше того, створити таку платформу для спільної роботи може як сама фірма та її внутрішній відділ технологічних розробок на власному сайті, так і за допомогою зовнішньої (сторонньої) платформи, яку зазвичай постачають компанії, що спеціалізуються на подібному бізнесі.

2.2 Основні теоретичні підходи до виміру CLV

Перші спроби аналізу терміну CLV були пов'язані з більш широким поняттям аналізу прибутковості клієнта (CPA), який відрізняється від звіту про прибутки та збитки компанії кількістю предметів, що обговорюються - CPA розглядає прибутковість саме одного клієнта, тоді як аналіз прибутків і збитків зазвичай проводиться з урахуванням усіх клієнтів .

Одними з перших авторів, що написали про CPA як про загальний підхід, були Wang і Splegel (1994) - тобто про розрахунок валової маржі - виручка від продажів за вирахуванням витрат, що відносяться до всіх проданих продуктів -

від конкретного клієнта за певний період часу[4]. Традиційний підхід до питання CPA має широке поширення у науковому контексті і поширюється більш високі рівні розрахунків. На наступному етапі від результатів попереднього етапу віднімаються витрати на продажі, адміністративні та загальні витрати, отримані в результаті отримання операційного прибутку від одного клієнта за час його співпраці з компанією. Нарешті, ця модель береться за основу для її розвитку - розрахунку рентабельності активів клієнта або, іншими словами, рентабельність клієнта, поділена, наприклад, на суму дебіторської заборгованості та запасів.

Прибутковість клієнта набуває форми майбутніх грошових потоків і, отже, пов'язані з аналізом чистої поточної вартості. Саме в цей момент дослідники починають обговорювати позитивну цінність клієнта. У цих роботах CLV було визначено як потік очікуваних майбутніх прибутків, чистих витрат на транзакції клієнта, дисконтованих за певною відповідною ставкою до поточної чистої вартості.

Подальші дослідження зробили аналогічний висновок про аналіз CLV, пов'язавши його з іншим терміном - клієнтський капітал, який можна описати як функцію обсягу покупок клієнта, маржі на одиницю покупки, а також витрат на придбання, розвиток та утримання, пов'язаних із цим клієнтом.

Якщо розглядати конкретні моделі розрахунку CLV, то майже всі вони зводяться до двох етапів:

- 1) розрахунок чистих грошових потоків від одного клієнта з часом;
- 2) розрахунок поточної вартості цих грошових потоків.

Існує кілька формул з більш-менш схожими змінними, але різними припущеннями для оцінки CLV.

Тим не менш, традиційна думка про CLV зазнала глибокого зрушення в акцентуванні змінних моделі від фінансових до нефінансових факторів. Вважається, що такий підхід (часто званий непрямою споживчою цінністю) має найбільшу пояснювальну силу з погляду сучасних тенденцій у світовій економіці. Однак дана концепція має на увазі труднощі з точки зору переведення нефінансових показників діяльності компанії в придатні для використання індекси та статистику.

Питання визначення фінансового пояснення непрямої споживчої цінності неодноразово обговорювався та отримав широку підтримку в останніх дослідженнях. Ryals розширив традиційні знання про вимір CLV, додавши такі змінні, як вигоди від навчання, пропаганди та інновацій. Vermejo та Rodríguez-Monroy зробили акцент на споживчій цінності в секторі B2B, оцінюючи чотири типи потенціалу: зростання, база, навчання та взаємодія.

Обговорення питань стає більш конкретним у роботі Hogan(2003), які запропонували модель для виміру внеску непрямих соціальних ефектів, таких як "сарафанне радіо", у CLV.

В цілому, не існує унікального та загальноприйнятого підходу до оцінки CLV, але напрям наукових досліджень доводить тенденцію до багатомірного підходу CLV, який враховує взаємопов'язані ключові аспекти та призводить до розподілу ресурсів, а не до одновимірних розрахункам.

2.3 Спільноти клієнтів, їх роль у CRM та вплив на CLV

Говорячи про процес CRM з погляду клієнта, можна виділити чотири основні фази: поінформованість, оцінка, перша покупка та поточне утримання. Видається доцільним докладніше зупинитися кожному етапі, підкресливши їхній зв'язок з клієнтськими спільнотами.

Етап "поінформованість". Відмінною особливістю етапу "поінформованість" є те, що клієнт ще остаточно не вирішив, що він шукає. Отже, цей етап вважається найважливішим серед інших етапів CRM. Мета клієнта на етапі поінформованості - "визначити широкий список можливих варіантів, які мають повторюваний успіх у вирішенні проблеми за допомогою свого продукту або вирішення, знаходяться в рамках загальних параметрів проекту (бюджет, терміни, стандарти) і, найголовніше, заслуговують на довіру" (там ж).

Співтовариства клієнтів можуть бути використані для просування етапу «поінформованості» CRM. Насправді, 74% респондентів стратегічно використовують свої спільноти для створення контенту з метою залучення трафіку з органічного пошуку, а 50% планують використовувати свої спільноти для залучення потенційних клієнтів на більш ранніх етапах процесу продажу.

Залучення потенційного клієнта за допомогою клієнтського співтовариства на ранній стадії до фактичної покупки не є марним з кількох причин. По-перше, спільноту зазвичай легко знайти за простим запитом завдяки добре продуманій пошуковій оптимізації. По-друге, вся CRM-система підприємства дає змогу відстежувати потенційних клієнтів на всіх наступних етапах CRM. І нарешті, використання клієнтських спільнот на етапі «поінформованості» загалом забезпечує раннє знайомство з продуктом або послугою та залученням клієнтів.

Фаза оцінки. Після збору всіх можливих варіантів клієнту необхідно детальніше оцінити ці варіанти та уважно вивчити існуючі пропозиції. Його головна мета на етапі оцінки - відсіяти більшість фірм, виявлених у ході первинного дослідження на етапі обізнаності, щоб глибоко зосередитися на вибраних кількох.

Згідно з дослідженням Multichannel Customer Service Survey (2014), 65% клієнтів припиняють усі відносини з брендом після лише одного випадку поганого обслуговування. Більше того, 72% респондентів стверджують, що клієнтські спільноти забезпечують найкраще обслуговування клієнтів (Get Satisfaction, 2014).

Таким чином, клієнт з більшою ймовірністю здійснить покупку у бренду, який на етапі оцінки надає краще обслуговування. Краще обслуговування клієнтів може бути досягнуто за рахунок розгортання клієнтських спільнот такими способами:

- надання відповідей на запитання покупців на вимогу;
- встановлення та підтримання довіри покупців за рахунок вищої швидкості реагування;
- прискорення здійснення першої покупки;
- диференціація бренду за рахунок унікального сумісного досвіду.

Перша покупка. Здійснивши першу покупку, покупці можуть об'єктивно оцінити товар або послугу, пропоновані підприємством. Цей етап вважається "делікатним", оскільки очікування клієнтів до здійснення покупки дуже високі, а крива навчання/досвіду може бути крутою.

Момент відразу після першої покупки зазвичай описується як початок глибших відносин із клієнтом та можливість розширити знайомство покупця з брендом. Однак підприємства часто втрачають контроль за процесом комунікації з новим клієнтом саме на цьому етапі. Основні цілі клієнта після здійснення першої покупки включають швидке залучення в пропонований продукт або послугу і спілкування з іншими клієнтами, щоб дізнатися більше про їх власний досвід[5].

Співтовариства клієнтів можуть бути важливим інструментом залучення клієнтів на етапі першої покупки. Виділяються такі способи використання онлайн-спільнот після етапу першої покупки:

- Навчання нового клієнта (часті питання, навчальні посібники, ідеї колег, тематичні дослідження);
- Налагодження зв'язків (спілкування клієнтів із колегами, доступ до ширшої групи лідерів у компанії);
- Підтримка (швидкі відповіді на запитання від внутрішніх експертів);
- Залучення (можливість поділитися ідеями та пропозиціями);
- зміцнення довіри (оперативність реагування фірми підтверджує правильність вибору).

Поточне утримання. Так звана фаза "оновлення" - це критичний момент прийняття рішення як підприємства, так клієнта. Останній зазвичай оглядається на свій попередній досвід взаємин з фірмою, оцінюючи його за трьома основними факторами: успішність проекту або покупки, внутрішнє ставлення до бренду та підтримка клієнтів з боку компанії.

Ключові цілі задоволеного клієнта на етапі утримання полягають у підтвердженні того, що його вибір був найкращим, поглиблення відносин з фірмою шляхом укладання додаткових контрактів та надання послуг, а також у тому, щоб поділитися своєю історією успіху для отримання додаткової PR та маркетингової цінності для своєї фірми.

За даними дослідження Get Satisfaction (2014), 74% B2B фірм вважають, що клієнтські спільноти допомогли значно покращити знання компанії про своїх клієнтів. Більше того, 57% таких підприємств вважають, що спільноти

допомагають запобігти відпливу клієнтів і збільшити можливості для збільшення продажів.

На заключному етапі процесу CRM з погляду клієнта, клієнтські спільноти також відіграють важливу роль у залученні покупців у подальші відносини з організацією, що можливе завдяки розгортанню наступних дій:

- Надання клієнтам платформи, де вони можуть розповісти свою історію успіху;
- Забезпечення навчання на вимогу для покращення досвіду клієнтів у використанні продуктів або послуг компанії;
- Пропозиція додаткових товарів та послуг покупцям, що залишилися з компанією;
- Надання клієнтам можливості ініціювати різні запити щодо нових продуктів, послуг та функцій.

Споживачі очікують, що клієнтські спільноти допоможуть їм та забезпечать надійний доступ до необхідної інформації на всіх етапах придбання або вибору товарів та послуг, а також їх експлуатації у майбутньому. Інтеграція соціальних додатків дозволяє усвідомити справжню творчу природу процесу CRM, а клієнтська спільнота, таким чином, є одним з найважливіших елементів CRM усієї компанії на кожному етапі[6].

Поки не доведено, що відносини компанії з клієнтом у межах клієнтського співтовариства мають позитивний вплив на CLV. Проте підприємствам необхідно залучати клієнтів у клієнтські спільноти хоча б через такі переваги :

Зростання продажів. Інформаційні бюлетені American Express World Service доводять, що правильно організована онлайн-підтримка допомагає підвищити рівень довіри, знизити негативне ставлення та підвищити лояльність.

Було проведено опитування серед тисяч споживачів, результати якого показали, що 17% з них звертаються за допомогою до компанії через онлайн-канали в середньому раз на рік, при цьому збільшуючи свої витрати на товари та послуги компанії на 21% за умови отримання своєчасної та компетентної підтримки. Були також респонденти, які ніколи не зверталися до брендів через клієнтські спільноти та соціальні мережі, але зацікавилися самою можливістю такого виду комунікації. Ці респонденти підтвердили, що готові витратити на 11% більше за умови отримання адекватного сервісу у каналах соціальних медіа. Варто також зазначити, що 71% клієнтів із позитивним клієнтським досвідом рекомендують бренд своїм друзям[7].

Підвищення лояльності клієнтів. Публічність та людське ставлення співробітників компанії до своїх клієнтів дозволяють ефективно боротися з негативним досвідом клієнтів, підвищують лояльність клієнтів та покращують корпоративну репутацію.

Зниження витрат за підтримку. Клієнтські спільноти завжди працюють цілодобово і дозволяють клієнтам допомагати одне одному. Більш того, база знань, закладена у співтоваристві, може значно знизити навантаження на представників бренду.

Вищезгадані фактори (такі як зростання продажу та зниження витрат на підтримку), у свою чергу, позитивно впливають на збільшення CLV через саму формулу розрахунку CLV, яка включає валову маржу від кожного клієнта[8]. Підвищення лояльності клієнтів сприятливо впливає і на показники компанії: вважається, що такі покращення завжди позитивно корелюють зі зростанням прибутку фірми. Говорячи про загальний вплив клієнтських спільнот на CLV фірми, можна зробити висновок, що в цілому прямий і позитивний вплив реляційних інформаційних процесів на клієнтські та фінансові показники

демонструє стратегічний та маркетинговий зиск від інтеграції соціальних мереж до організаційних рутин[9].

2.4 Різновид розрахунків CLV

Історична позиттєва вартість клієнта

Історична CLV – це сума всього валового прибутку від минулих покупок клієнта. Для розрахунку необхідно скласти всі значення валового прибутку до останньої угоди (N), скоєної клієнтом. Необхідно виміряти CLV на основі чистого прибутку, щоб отримати справжній прибуток, який генерує даний клієнт. Сюди входять витрати на обслуговування, повернення, маркетинг, придбання тощо[10].

$$\text{Історичний CLV} = \frac{\text{Транзакція№1} + \text{Транзакція№2} + \dots + \text{Транзакція№N}}{\text{Середню валову маржу}}$$

Недоліком є те, що для отримання найбільш актуальних даних, можливо, доведеться виконувати складні математичні обчислення на індивідуальному рівні. Тим не менш, валова маржа CLV дає повне уявлення про прибутковість клієнтів на сьогоднішній день.

В принципі, цей метод зручний, якщо клієнти мають однакові переваги і однаково взаємодіють з вашим брендом протягом приблизно одного і того ж часу.

Слід пам'ятати, що розрахунок історичного CLV означає, що всі клієнти, старі та нові, мають бути поміщені в один кошик. Це може бути непросто, оскільки вони можуть відрізнитися за поведінкою та перевагами. Різниця між клієнтами може вплинути на CLV.

Прогнозована позиттєва цінність клієнта

Мета прогнозованої CLV - змоделювати транзакційну поведінку ваших клієнтів, щоб спрогнозувати, які дії вони зроблять у майбутньому. Це відмінний показник CLV, краще, ніж історичний CLV[10].

Прогнозована модель використовує алгоритми, які намагаються створити точний CLV, прогнозуючи загальну вартість клієнта. Вона працює на основі історії минулих транзакцій та дій клієнта.

Одна з простих формул має вигляд:

$$CLV = T * AOV * AGM * ALT$$

Де:

T - Середньомісячні угоди

AOV - Середня вартість замовлення

ALT - Середня тривалість життя клієнта (у місяцях)

AGM - Середня валова маржа

Аналітична платформа з CDP може швидко інтегруватися з CRM, що дозволить вам легко розрахувати AOV і ALT.

Однак, слід пам'ятати, що цей підхід є прогнозом, тому він має свою точність. Аби її підвищити, слід скоригувати розрахунки CLV з урахуванням специфіки галузі, в якій працює бізнес. Точне визначення CLV дає інструмент для розробки надійних маркетингових стратегій.

Традиційна позитивна цінність клієнта

Іноді традиційна, але глибока формула CLV може працювати краще. Це відбувається в тому випадку, якщо ваш річний продаж нерівномірний. Тоді важливо враховувати ставку знижки, середню валову маржу на термін життя клієнта та коефіцієнт утримання.

Ось як виглядає остаточна формула:

$$CLV = GML \left(\frac{R}{1+D-R} \right)$$

Де:

GML - валова маржа на одного клієнта

R - щомісячний коефіцієнт утримання

D - щомісячний коефіцієнт дисконтування

Ця формула розглядає можливі зміни у доходах клієнтів протягом певного періоду часу. Потім щороку коригується на ставку дисконтування обліку інфляції.

2.5 Як застосовувати показник пожиттєвої цінності клієнта

CLV можна використовувати для розробки стратегії. Ось кілька прикладів, де можна застосувати її.

- **Ефективно сегментувати своїх клієнтів:** Використовуючи моделі CLV, можна покращити як профілювання, так і сегментацію. Це дозволить адаптувати пропозиції та націлювати клієнтів на основі їхньої потенційної цінності. Це також покращує прогнозування та підвищує коефіцієнт конверсії лідів. Крім того, можна підвищити

ефективність сегментів за допомогою даних, зібраних з різних джерел таких технологій, як платформи даних про клієнтів (CDP).

- **Оптимізація придбання:** Експерти з ALTA наголошують, що прогнозований CLV дозволяє збільшити кількість придбань, оскільки він допомагає спеціалістам "переконатися у придбанні передплатників, які представлятимуть найбільшу довічну цінність у майбутньому".
- **Підвищення утримання:** CLV – це покроковий компас для навігації у збереженні лояльності клієнтів до вашого бренду. Він допоможе вам визначити пріоритети, наприклад яких клієнтів потрібно повернути, і розробити унікальну стратегію для цього.
- **Поліпшення прогнозування:** Розрахунок CLV дає інструмент для прогнозування майбутніх потреб у продукції чи послугі. Маючи під рукою цю інформацію, можна керувати своїми інвестиціями з погляду робочої сили, запасів та інших ресурсів. Детальне прогнозування є життєво важливим для зниження втрат продуктивності та більш ефективного розподілу ресурсів.
- **Розпізнавання найкращих клієнтів:** Дані про клієнтів, наприклад, про часті покупки та транзакції, дозволяють виявити тих, хто витрачає найбільше. Ця інформація підкаже, які продукти слід просувати активніше. Можна запрошувати клієнтів на спеціальні заходи та пропонувати спеціальні пропозиції для цінних клієнтів. Нарешті, можна краще подбати про своїх найцінніших клієнтів, надавши їм індивідуального помічника чи консультанта[10].

2.6 CLV як ключовий елемент юніт-економіки

Юніт-економіка - це метод економічного моделювання, який балансує між отриманим прибутком і витратами, що несе компанія в процесі реалізації свого ключового елемента торгівлі, надання послуги, залученого клієнта. Таким чином, «юніт» - це одиниця, яка генерує прибуток компанії.

Для коректного розуміння цієї моделі слід дотримуватись правила: «різний бізнес - різні нюанси». Важливо орієнтуватися тільки на актуальні саме для конкретного підприємства показники. Зазвичай, модель використовується в digital проектах.

Поява методології була спровокована «бумом доткомів»: нові підприємці - IT-фахівці - мало розуміли в монетизації проектів. Тоді бізнес-акселератори запропонували спрощену модель для оцінки бізнес-моделі на початковому етапі - розраховувати маржинальний прибуток із одиниці продажу. Незабаром популярність методології почала зростати - її почали використовувати класичні підприємці, які розвивають бізнес в інтернеті, про неї заговорили на тематичних курсах[11].

Поштовхом до розвитку юніт-економіки стало поширення інтернет-аналітики - з'явилася можливість аналізувати ефективність комунікацій та воронки продажів.

Юніт-економіка дозволяє побачити, скільки ви заробляєте з потоку клієнтів - потік складається з юнітів, кожен із яких приносить певний прибуток. Якщо вирахувати, скільки приносить кожен юніт і які витрати при цьому несе компанія, можна розрахувати, який прибуток ви отримаєте з певного потоку. За результатами розрахунку стає ясно, чи варто масштабувати бізнес, залучати інвесторів, збільшувати потік чи маржинальність угоди.

Важливо, що юнітом може називатися не тільки клієнт, який заплатив. Клієнт – прийняте визначення юніту насамперед для SaaS(Software as a Service)-проектів. Так, наприклад, у мобільних додатках та іграх це буде новий користувач. А для інтернет-видання чи сервісу – передплатник. Також як юніт можна розглядати одиницю товару[12].

Дві найважливіші метрики, які лежать в основі юніт-економіки це CLV та САС.

CLV - це чистий дохід, який клієнт приносить бізнесу за рахунок покупок протягом усього терміну свого «життя».

САС(Customer acquisition cost) – ціна залучення клієнта.

Юніт-екноміка вважається позитивною, тобто такою яка приносить прибуток коли:

$$\text{САС} < \text{CLV}$$

3. Моделі, які описані в науковій роботі.

3.1 Pareto/Negative Binominal distribution model.

3.1.1 Історія.

Вперше запропонована Шміттліном, модель Pareto/NBD була дуже успішною як інструмент аналізу клієнтської бази. Pareto/NBD орієнтована на моделювання того, чи є клієнти і чи досі вони користуються продуктами компанії та як часто вони здійснюють покупки. Клієнти здійснюють покупки відповідно до Пуассонівського розподілу доки вони «живі». Час життя клієнтів розподіляється відповідно до експоненційного розподілу. Частота покупок і схильність до виживання варіюються в популяції відповідно до окремих гамм розподілу[13].

З 1987 модель Pareto/NBD була розширена іншими дослідниками. Pareto/NBD та інші подібні моделі можуть бути використані для вирішення низки управлінських завдань, включаючи оцінку кількості "активних" клієнтів, ранжування клієнтів на основі ймовірності того, що вони "живі", та прогнозування майбутніх транзакцій. Було показано, що модель добре працює у різних умовах. Фейдер, Харді та Лі перевірили перспективні прогнози моделі для інтернет-магазину з продажу музики. Ейб застосував модель у різних умовах закупівель від електронної комерції до універмагів та великих мереж.

Крім емпіричної перевірки модель застосовувалася і в інших областях. Хопманн і Теде використовували її вивчення прогнозів відтоку за умов відсутності контрактів. Вюббен і Вангенхайм показали, що модель працює

еквівалентно або краще, ніж звичайні управлінські евристики. Гледі, Безенс та Кру розширили модель для отримання оцінок споживчої довічної цінності у кількох умовах. За цих різноманітних умов модель Парето/НБД сформувала основу широкого та успішного аналізу клієнтської бази[13].

3.1.2 Припущення моделі.

Модель Pareto/NBD заснована на таких припущеннях:

1) Клієнти проходять два етапи у взаємодії з конкретною компанією: вони "живі" – а тобто користуються послугами або продукцією компанії протягом певного період часу, а потім стають постійно неактивними.

2) Поки клієнт "живий", кількість транзакцій, вчинених ним, відповідає розподілу Пуассона зі швидкістю транзакцій λ ; тому ймовірність спостереження x транзакцій в інтервалі часу $(0, t)$ визначається наступним чином

$$P(X(t) = x | \lambda) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}, \quad x = 0, 1, 2, \dots \quad (1)$$

де t_j - час j -ї покупки.

3) Неспостережуваний "час життя" клієнта завдовжки τ (після якого він вважається неактивним) має експоненційний розподіл з коефіцієнтом відсіву μ .

$$f(t_j - t_{j-1} | \lambda) = \lambda e^{-\lambda(t_j - t_{j-1})}, \quad t_j > t_{j-1} > 0 \quad (2)$$

4) Гетерогенність швидкості транзакцій по клієнтам відповідає гамма-розподілу з формою r та параметром форми r та параметром масштабу α :

$$g(\lambda | r, \alpha) = \frac{\alpha^r \lambda^{r-1} e^{-\lambda \alpha}}{\Gamma(r)} \quad (3)$$

5) Гетерогенність у показниках відсіву за клієнтами відповідає гамма-розподілу з параметром форми з параметром форми s та параметром масштабу β .

$$g(\mu | s, \beta) = \frac{\beta^s \mu^{s-1} e^{-\mu\beta}}{\Gamma(s)} \quad (4)$$

6) Коефіцієнт транзакцій λ та коефіцієнт відсіву μ варіюються незалежно між клієнтами.

Допущення 2 і 4 дають нам модель NBD для розподілу кількості транзакцій, доки клієнт живий,

$$P(X(t) = x | r, \alpha) = \int_0^\infty P(X(t) = x | \lambda) g(\lambda | r, \alpha) d\lambda = \frac{\Gamma(r+x)}{\Gamma(r)x!} \left(\frac{\alpha}{\alpha+t}\right)^r \left(\frac{t}{\alpha+t}\right)^x \quad (5)$$

у той час як припущення 3 та 5 дають нам "розподіл Парето другого роду",

$$f(\tau | s, \beta) = \int_0^\infty f(\tau | \mu) g(\mu | s, \beta) d\mu = \frac{s}{\beta} = \left(\frac{\beta}{\beta + \tau}\right)^{s+1} \quad (6), i$$

$$F(\tau | s, \beta) = \int_0^\infty F(\tau | \mu) g(\mu | s, \beta) d\mu = 1 - \left(\frac{\beta}{\beta + \tau}\right)^s \quad (7)$$

3.2.3 Ключовий проміжний результат

У процесі роботи над похідними нам доведеться оцінити подвійний інтеграл (1) наступного виду декілька разів:

$$A = \int_0^1 \int_0^\infty \frac{\lambda^\gamma \mu^\delta}{\lambda + \mu} e^{-(\lambda + \mu)t} g(\lambda | r, \alpha) g(\mu | s, \beta) d\lambda d\mu \quad (1)$$

Розглянемо перетворення $p = \mu$ та $z = \lambda + \mu$, з відповідним зворотним перетворенням $\lambda = (1 - p)z$ та $\mu = pz$. Якобіан цього перетворення має вигляд

$$J = \begin{vmatrix} \frac{\partial \lambda}{\partial p} & \frac{\partial \lambda}{\partial z} \\ \frac{\partial \mu}{\partial p} & \frac{\partial \mu}{\partial z} \end{vmatrix} = z$$

Звідси слідує що

$$A = \int_0^1 \int_0^\infty p^\delta (1 - p)^\gamma z^{\gamma + \delta - 1} e^{-zt} g(p, z | \alpha, \beta, r, s) dz dp \quad (2)$$

де спільний розподіл p і z

$$g(p, z | \alpha, \beta, r, s) = \frac{\alpha^r \beta^s}{\Gamma(r)\Gamma(s)} p^{s-1} (1 - p)^{r-1} z^{r+s-1} e^{-z(\alpha - (\alpha - \beta)p)} \quad (3)$$

отримано за допомогою методу перетворення. Підстановка (3) до (2) дає нам

$$A = \frac{\alpha^r \beta^s}{\Gamma(r)\Gamma(s)} B \quad (4)$$

де

$$B = \int_0^1 \int_0^\infty p^{s+\delta-1} (1-p)^{r+\gamma-1} z^{r+s+\gamma+\delta-2} e^{-z(\alpha+t-(\alpha-\beta)p)} dz dp =$$

$$\frac{\Gamma(r+s+\gamma+\delta-1)}{(\alpha+t)^{r+s+\gamma+\delta-1}} \int_0^1 p^{s+\delta-1} (1-p)^{r+\gamma-1} \left[1 - \frac{(\alpha-\beta)}{(\alpha+t)} p\right]^{-(r+s+\gamma+\delta-1)} dp \quad (5)$$

що, згадуючи інтеграл Ейлера для гіпергеометричної функції Гауса,

$$= \frac{\Gamma(r+s+\gamma+\delta-1)}{(\alpha+t)^{r+s+\gamma+\delta-1}} \frac{\Gamma(r+\gamma)\Gamma(S+\delta)}{\Gamma(r+s+\gamma+\delta)} F_1\left(r+s+\gamma+\delta-1, s+\delta; r+s+\gamma+\delta; \frac{\alpha-\beta}{\alpha+t}\right) \quad (6)$$

і тому

$$A = \frac{\alpha^r \beta^s}{(\alpha+t)^{r+s+\gamma+\delta-1}} \frac{\Gamma(r+\gamma)\Gamma(S+\delta)}{\Gamma(r)\Gamma(s)} \left(\frac{1}{\Gamma(r+s+\gamma+\delta-1)}\right) F_1\left(r+s+\gamma+\delta-1, s+\delta; r+s+\gamma+\delta; \frac{\alpha-\beta}{\alpha+t}\right) \quad (7)$$

Уважно розглядаючи (4), бачимо, що аргумент z гаусівської гіпергеометричної функції, $\frac{\alpha-\beta}{\alpha+t}$, гарантовано обмежений між 0 і 1 при $\alpha \geq \beta$ (оскільки $\beta > 0$ і $t > 0$), таким чином забезпечуючи збіжність серійного уявлення функції. Однак при $\alpha < \beta$ ми можемо зіткнутися із ситуацією, коли $|z| > 1$, у такому разі ряд розходиться. Тому для випадку $\alpha \leq \beta$ розглянемо перетворення $p = \frac{\lambda}{\lambda+p}$ і $z = \lambda + \mu$, з відповідним зворотним перетворенням $\lambda = pz$ та $\mu = (1-p)z$. Якобіан цього перетворення має вигляд

$$J = \begin{vmatrix} \frac{\partial \lambda}{\partial p} & \frac{\partial \lambda}{\partial z} \\ \frac{\partial \mu}{\partial p} & \frac{\partial \mu}{\partial z} \end{vmatrix} = z$$

Звідси слідує що

$$A = \int_0^1 \int_0^\infty p^\gamma (1-p)^\delta z^{\gamma+\delta-1} e^{-zt} g(p, z | \alpha, \beta, r, s) dz dp \quad (8)$$

де спільний розподіл p і z

$$g(p, z | \alpha, \beta, r, s) = \frac{\alpha^r \beta^s}{\Gamma(r)\Gamma(s)} p^{r-1} (1-p)^{s-1} z^{r+s-1} e^{-z(\alpha - (\beta - \alpha)p)} \quad (9)$$

отримано за допомогою методу перетворення. Це дає нам

$$A = \frac{\alpha^r \beta^s}{\Gamma(r)\Gamma(s)} B \quad (10)$$

Де

$$\begin{aligned} B &= \int_0^1 \int_0^\infty p^{r+\gamma-1} (1-p)^{s+\delta-1} z^{r+s+\gamma+\delta-2} e^{-z(\beta+t - (\beta-\alpha)p)} dz dp \\ &= \frac{\Gamma(r+s+\gamma+\delta-1)}{(\beta+t)^{r+s+\gamma+\delta-1}} \int_0^1 p^{r+\gamma-1} (1-p)^{s+\delta-1} \left[1 - \frac{(\beta-\alpha)}{(\beta+t)} p \right]^{-(r+s+\gamma+\delta-1)} dp \quad (11) \end{aligned}$$

що, згадуючи інтеграл Ейлера для гіпергеометричної функції Гауса,

$$= \frac{\Gamma(r+s+\gamma+\delta-1)}{(\beta+t)^{r+s+\gamma+\delta-1}} \frac{\Gamma(r+\gamma)\Gamma(S+\delta)}{\Gamma(r+s+\gamma+\delta)} F_1(r+s+\gamma+\delta-1, r+\gamma; r+s+\gamma+\delta; \frac{\beta-a}{\beta+t}) \quad (12)$$

і тому

$$A = \frac{\alpha^r \beta^s}{(\beta+t)^{r+s+\gamma+\delta-1}} \frac{\Gamma(r+\gamma)\Gamma(S+\delta)}{\Gamma(r)\Gamma(s)} \left(\frac{1}{(r+s+\gamma+\delta-1)} \right) \times F_1(r+s+\gamma+\delta-1, r+\gamma; r+s+\gamma+\delta; \frac{\beta-a}{\beta+t}) \quad (13)$$

Ми бачимо, що аргумент z вищезгаданої гаусівської гіпергеометричної функції обмежений між 0 та 1 при $\alpha \leq \beta$. Тому ми представляємо (8) та (13) як розв'язання задачі (1): ми використовуємо (8), коли $\alpha \geq \beta$ та (13), коли $\alpha \leq \beta$.

3.2 Beta Geometric/Negative Binominal distribution model

Більшість аспектів моделі BG/NBD безпосередньо повторюють модель Pareto/NBD. Єдине відмінність у тому, як/коли клієнти стають неактивними. Тимчасова модель Парето модель передбачає, що відсіювання може статися у будь-який момент часу, незалежно від вчинення фактичної покупки. Якщо замість цього ми припустимо, що відсіювання відбувається відразу після покупки, ми можемо змоделювати цей процес за допомогою бета-геометричної (BG) моделі.

Більш формально модель BG/NBD заснована на наступних п'яти припущеннях:

1. Коли користувач активний, кількість його транзакцій за період довжиною t описується розподілом Пуассона з коефіцієнтом транзакцій λ .

Розподіл Пуассона допомагає передбачити наступ певних подій, використовуючи дані про те, як часто подія відбувалася в минулому.

2. Гетерогенність швидкості транзакцій між користувачами (тобто, як клієнти різняться по купівельній поведінці) має Гамма-розподіл з параметрами r (форма) і α (масштаб). Гамма-розподіл виникає природно у процесах з часом очікування між пуассонівськими подіями.

3. Користувач може стати неактивним після будь-якої транзакції з ймовірністю p , як точка відсіву розподіляється між покупками з допомогою геометричного розподілу. Геометричний розподіл схожий на розподіл випробувань Бернуллі та використовується для моделювання кількості випробувань до першого успіху включно.

4. Гетерогенність у ймовірності вибуття має Бета-розподіл з двома параметрами форми α та β . Бета-розподіл найкраще підходить для представлення імовірнісного розподілу ймовірностей у випадку, коли ми не знаємо заздалегідь, яка ймовірність, але у нас є деякі розумні припущення, що описуються α та β (середнє значення бета-розподілу $\alpha/(\alpha+\beta)$).

5. Швидкість транзакцій та ймовірність відсіву незалежно різняться у різних користувачів[14].

Математична нотація для представлення характеристик користувача X :

$X = x, t_x, T$, де x – кількість транзакцій у деякий період часу $(0, T]$, а $t_x(≤T)$ – час останньої покупки.

Грунтуючись тільки на цих характеристиках, модель передбачає майбутні моделі покупок клієнтів:

- $P(X(t) = x)$ - ймовірність спостереження у майбутньому угод x у період t

- $E(Y(t) | X = x, t_x, T)$ - очікувана кількість угод у період для покупця з поведінкою, що спостерігається.

Тепер ми можемо вивести ці дві основні характеристики.

Можливість бути активним

$$P(X(t) = x | \lambda, p) = (1 - p)^x \frac{(\lambda t)^x e^{-\lambda t}}{x} + \delta_{x>0} p (1 - p)^{x-1} \left[1 - e^{-\lambda t} \sum_{j=0}^{x-1} \frac{(\lambda t)^j}{j!} \right]$$

Де $\delta_{x>0} = 1$ якщо $x > 0$, 0 в інших випадках.

Очікувана кількість транзакцій

$$E(Y(t) | X = x, t_x, T, r, \alpha, a, b) = \frac{\frac{a+b+x-1}{a-1} \left[1 - \left(\frac{\alpha+T}{\alpha+T+t} \right)^{r+x} + {}_2F_1(r+x, b+x; a+b+x-1; \frac{t}{\alpha+T+t}) \right]}{1 + \delta_{x>0} \frac{a}{b+x-1} \left(\frac{\alpha+T}{\alpha+t_x} \right)^{r+x}}$$

де ${}_2F_1$ – гіпергеометрична функція Гауса.

$${}_2F_1(a; b; c; z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{j=0}^{\infty} \frac{\Gamma(a+j)\Gamma(b+j)}{\Gamma(c+j)} \frac{z^j}{j!}$$

Підсумовуючи, можна сказати, що якщо отримати оцінки параметрів моделі r , α , a , b (наприклад, за допомогою оцінки максимальної правдоподібності), можна прогнозувати очікувану кількість транзакцій для користувачів[15].

3.3 Modified Beta Geometric/Negative Binominal distribution model

У 2007 Batislam, Denizel, і Filiztekin(BDF) - виводять варіант BG/NBD, яка була розроблена Фейдером, Харді та Лі (2005). Вони вводять додаткову можливість для відхилення в нульовий момент часу, тим самим усуваючи "зсув" геометричного розподілу відхилення клієнтів. Модель одержала назву Modified BG/NBD (MBG/NBD)[16].

Цей принцип виправляє помилку в їхньому рівнянні, умовне очікуване число майбутніх угод споживача з історією покупок (x, t_x, T) -(кількість угод, час останньої угоди, час звернення). Визначаючи $Y(t)$ як випадкову змінну для транзакцій в інтервалі довжини t , що починається в момент T , та з параметрами моделі r, α, a, b , BDF виводять цей вираз як

$$E[Y(t)|x, t_x, T, r, \alpha, a, b]= \frac{\frac{b+x}{a-1} \left[1 - \left(\frac{\alpha+T}{\alpha+T+t} \right)^{r+x} F_1 \left(r+x, b+x+1; a+b+x; \frac{t}{t+T+\alpha} \right) \right]}{1 + \frac{a}{b+x} \left(\frac{\alpha+T}{\alpha+t_x} \right)^{r+x}} \quad (1)$$

тоді як правильним результатом є

$$E[Y(t)|x, t_x, T, r, \alpha, a, b]= \frac{\frac{a+b+x}{a-1} \left[1 - \left(\frac{\alpha+T}{\alpha+T+t} \right)^{r+x} F_1 \left(r+x, b+x+1; a+b+x; \frac{t}{t+T+\alpha} \right) \right]}{1 + \frac{a}{b+x} \left(\frac{\alpha+T}{\alpha+t_x} \right)^{r+x}} \quad (2)$$

Знаменники рівнянь (1) і (2) ідентичні і є оберненими до величини $P(\tau > T)$ (τ – життя клієнта). Таким чином, при виведенні скоригованого виразу основна увага приділяється чисельнику. З огляду на можливість для відхилення в нульовий момент часу, очікувана кількість угод споживача $E[X(t)]$ впливає з числа, очікуваного у разі негайного відхилення, тобто нуля, та кількості, очікуваної у разі, якщо споживач не відхиляється у нульовий момент часу. Обидва показники зважуються за ймовірності будь-якої з цих подій:

$$E(X(t)) = E[X(t)|\tau = 0] * P(\tau = 0) + E[X(t) \tau > 0] * P(\tau > 0) \quad (3)$$

Так як $p_{(\tau>0)} = \frac{b}{a+b}$ слідує, що

$$E[X(t) \tau > 0] = \frac{E[X(t)]}{P(\tau > 0)} = E[X(t)] \frac{a+b}{b}, \quad (4)$$

Де

$$E[X(t)] = \frac{b}{a-1} \left[1 - \left(\frac{\alpha}{\alpha+t} \right)^r F_1 \left(r, b+1; a+b; \frac{t}{\alpha+t} \right) \right] \quad (5)$$

$E[Y(t)]$ тепер отримано з урахуванням історії покупок клієнта. Елегантний спосіб вирішення цього завдання полягає в оновленні параметрів, запропонований Гріном. Стає очевидним, що BDF вивели рівняння (1) шляхом оновлення $E[X(t)]$. Проблема в тому, що вони мають натомість використовувати $E[X(t) | \tau > 0]$. Обидва результати можна узагальнити за допомогою наведених нижче кроків. Нехай $E[X(s, s+t)]$ означає очікувану кількість транзакцій в інтервалі від s до $s+t$. Тоді для $s=0$ це рівно $E[X(t)]$, а для $s=T - E[Y(t)]$ маємо $E[Y(t)]$:

$$E[X(s, s+t) | x, t_x, s, r', \alpha', a', b'] = \frac{\alpha'+b'}{\alpha'-1} G(r', \alpha', a', b') P(\tau > s), \quad (6)$$

Де

$$G(r', \alpha', a', b') = 1 - \left(\frac{\alpha'}{\alpha'+t} \right)^{r'} F_1 \left(r', b'+1; a'+b'; \frac{t}{\alpha'+t} \right) \quad (7)$$

І

$$P(\tau > s) = \left[1 + \frac{a'}{b'} \left(\frac{\alpha'}{\alpha+t_x} \right)^{r'} \right]^{-1} \quad (8)$$

3.4 Gamma-Gamma model

Гамма-гамма модель використовується разом з кожною наведеною вище моделлю, адже ті моделі допомагають нам знайти очікувану кількість транзакцій в той час як Гамма-Гамма дає нам можливість середню виручку за транзакцію.

Гамма-гамма модель витрат на транзакцію ґрунтується на наступних трьох загальних припущеннях:

- Грошова вартість транзакції клієнтів варіюється випадково навколо їх середньої вартості транзакції.

- Середня вартість транзакції варіюється між клієнтами, але не змінюється з часом для кожної конкретної людини.

- розподіл середньої вартості транзакцій по клієнтам не залежить від процесу транзакцій.

Для клієнта з x транзакціями, нехай z_1, z_2, \dots, z_x позначають вартість кожної транзакції. Середня вартість угоди клієнта, що спостерігається:

$$\bar{z} = \sum_{i=1}^x \frac{z_i}{x} \quad (1)$$

\bar{z} є недосконалою оцінкою їх середньої вартості угоди ζ . Наша мета - зробити висновки про ζ з урахуванням \bar{z} , які ми позначимо як $E(Z | \bar{z}, x)$.

Першим кроком нам потрібно вивести розподіл \bar{z} з урахуванням x транзакцій.

Побудова моделі.

Шміттлін та Петерсон запропонували, щоб розподіл z_i був нормальним. Дві проблеми з цим вибором розподілу полягають у тому, що а) він не обмежений нулем знизу і б) воно призводить до симетричного розподілу витрат. Оскільки дані про витрати мають тенденцію до правостороннього перекосу, ми використовуємо гамма-розподіл[17]. Більш формально,

а) ми припускаємо, що $z_i \sim \text{gamma}(p, v)$, при цьому $E(Z_i | p, v) = \zeta = p/v$.

- З огляду на властивості згортки гамми слідує, що загальна сума витрат за x транзакціями це гамма-розподіл(px, v).

- З огляду на властивість масштабування гамма-розподілу слідує, що $\bar{z} \sim \text{gamma}(px, vx)$.

б) ми припускаємо, що $v \sim \text{gamma}(q, \gamma)$.

Це призводить до того, що ми називаємо гамма-гамма моделлю грошової вартості (або витрат на транзакцію), яка також відома як бета другого роду.

3.5 RFM – аналіз

RFM(recency, frequency, monetary) - означає повторюваність, частоту та грошову вартість. RFM-аналіз - це спосіб використання даних, заснованих на поведінці існуючих клієнтів, для прогнозування того, як новий клієнт, ймовірно, поводитиметься в майбутньому. Модель RFM будується з використанням трьох ключових факторів:

- Як давно клієнти купували в останнє
- Як часто вони взаємодіяли із брендом
- Скільки грошей вони витратили на продукти та послуги

Аналіз RFM народився з маркетингу прямого поштового розсилання, зокрема, зі статті Тома Вансбіка та Яна Рупфа Бульта "Оптимальний вибір для прямого поштового розсилання", опублікованій у 1995 році в журналі Marketing Science. Їхня робота допомогла підтвердити принцип Парето - широко поширену серед маркетологів ідею про те, що 80% продажів припадає на 20% клієнтів бренду[18].

Переваги використання RFM аналізу.

RFM-аналіз дозволяє маркетологам збільшити доходи за рахунок орієнтації на певні групи існуючих клієнтів (тобто сегментації клієнтів) за допомогою повідомлень та пропозицій, які з більшою ймовірністю будуть актуальні на основі даних про певний набір моделей поведінки. Це призводить до збільшення кількості відгуків, утримання клієнтів, підвищення задоволеності клієнтів та збільшення довічної вартості клієнта (CLV).

Кожна з цих метрик RFM довела свою ефективність у прогнозуванні майбутньої поведінки клієнтів та збільшенні доходів. Клієнти, які здійснили покупку в недавньому минулому, з більшою ймовірністю зроблять це найближчим часом. Ті, хто частіше взаємодіє з брендом, з більшою ймовірністю зроблять це знову найближчим часом. А ті, хто витратив найбільше, з більшою ймовірністю витратять більше й у майбутньому.

RFM-аналіз дозволяє надсилати клієнтам повідомлення, які найкраще відповідають їхнім стосункам із брендом. Наприклад, швидше за все, компанія з великим успіхом запропонує дорогі товари клієнтам, які часто й багато витрачають. З іншого боку, більше шансів підвищити цінність відносин винагороджуючи клієнтів за лояльність або пропонуючи реферальні акції[19].

Принцип роботи моделі

Дослідження ринку традиційно концентруються на демографічних та психографічних даних, які маркетологи використовують для сегментації споживачів. Ці дані потім використовуються для прогнозування поведінки споживачів у набагато більших групах населення, які мають той самий набір ознак. Однак ці методи залежать від даних, отриманих від невеликої вибірки споживачів.

З появою таких систем, як платформи даних про клієнтів (CDP), які допомагають збирати, об'єднувати та синтезувати дані про поведінку клієнтів, у

маркетологів з'явилися набагато докладніші дані про звички окремих клієнтів, які можна використовувати при сегментації. Замість того, щоб сегментувати клієнтів тільки на основі демографічних та психографічних даних, маркетологи можуть створювати сегменти на основі реальної поведінки людей, включаючи історію покупок по будь-якому каналу (онлайн або офлайн), історію переглядів, відповіді на попередні кампанії та багато іншого[20]. Не дивно, що такий тип сегментації називається поведінковою сегментацією.

І навіть базова CRM-система може виконувати елементарне відстеження трьох характеристик, що легко вимірюються, які роблять внесок в аналіз RFM:

Значення періодичності: Це показник часу, що минув з моменту останньої взаємодії клієнта з брендом, який може включати останню покупку, відвідування веб-сайту, використання мобільного додатка, "лайк" у соціальних мережах та багато іншого. Періодичність - це ключова метрика, оскільки клієнти, які взаємодіяли з брендом нещодавно, з більшою ймовірністю відреагують на нові маркетингові зусилля.

Значення частоти: Це кількість разів, коли клієнт робив покупку або іншим чином взаємодіяв з брендом протягом певного періоду часу. Частота – це ключовий показник, оскільки він показує, наскільки глибоко клієнт взаємодіє з компанією. Велика частота свідчить про більш високий рівень лояльності.

Грошова цінність: Це загальна сума, яку клієнт витратив на придбання товарів та послуг бренду за певний період часу. Грошова цінність є ключовим показником, оскільки клієнти, які найбільше витратили в минулому, з більшою ймовірністю витратять більше і в майбутньому.

Використання моделі

RFM-аналіз може бути особливо ефективним інструментом для стартапів та невеликих роздрібних компаній з обмеженими маркетинговими ресурсами завдяки:

Простоті: RFM-аналіз сам не потребує складних інструментів чи складних аналітичних можливостей. Принципи легко зрозуміти, а результати легко інтерпретувати та діяти відповідно до них.

Доступності: У багатьох випадках фахівці з маркетингу, які не мають просунутої статистичної або аналітичної підготовки, можуть проводити RFM-сегментацію клієнтів, використовуючи лише стандартну електронну таблицю.

Ефективності у прямому маркетингу: RFM-аналіз, який виріс з маркетингових баз даних та прямої поштової розсилки, показав свою ефективність у ніші недорогих цифрових стратегій прямого маркетингу, які можуть дозволити собі невеликі бренди, наприклад, маркетингова кампанія електронною поштою.

4. Розробка

4.1 Підготовка даних.

Для перевірки моделей був взятий датасет неконтрактного бізнесу з безперервними купівельними можливостями – база даних електронної комерції британського магазину за 2010-2012 роки.

Першочергово вивантажуємо дані з Excel в Jupiter і з'єднуємо їх в одну таблицю.

	Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
0	489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	2009-12-01 07:45:00	6.95	13085.0	United Kingdom
1	489434	79323P	PINK CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
2	489434	79323W	WHITE CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
3	489434	22041	RECORD FRAME 7" SINGLE SIZE	48	2009-12-01 07:45:00	2.10	13085.0	United Kingdom
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	2009-12-01 07:45:00	1.25	13085.0	United Kingdom

Таблиця 1. База даних, яка відображає діяльність магазину.

Далі необхідно видалити з таблиці записи з порожніми клітинками і дублікатами. Зробивши це отримуємо готовий для роботи набір даних.

Задамо базові операції тестування:

- За період тренування візьмемо перші півтора року. Тестування будемо проводити на останніх 6 місяцях наших даних.
- Для порівняння ефективності моделей будемо розраховувати середньоквадратичну похибку, середню похибку і корінь зі середньоквадратичної помилки прогнозу.

- Усі моделі в цій роботі задля прогнозування CLV потребують RFM модель.
- Основні бібліотеки, які були використані в роботі: pandas, lifetime, matplotlib, pyplot, seaborn, numpy.
- Маржинільність бізнесу 5%.

Для кожної моделі будуть будуватися такі графіки:

- Матриця, яка розраховує очікувану кількість транзакцій за 1 одиницю часу, яку клієнт повинен зробити в наступний період, враховуючи періодичність покупок і їхню частоту.
- Матриця вірогідності того, що клієнт ще досі «живий».
- Графік порівняння прогнозованих значень з реальними.

4.2 Побудова RFM моделі.

За допомогою методів бібліотеки lifetime створимо таблицю з RFM аналізом по кожному користувачу окремо.

	Customer ID	frequency	recency	T	monetary_value
0	12347.0	7.0	402.0	404.0	717.398571
1	12348.0	4.0	363.0	438.0	449.310000
2	12349.0	4.0	717.0	735.0	1107.172500
3	12352.0	8.0	356.0	392.0	218.182500
4	12353.0	1.0	204.0	408.0	89.000000

Таблиця 2. RFM аналіз.

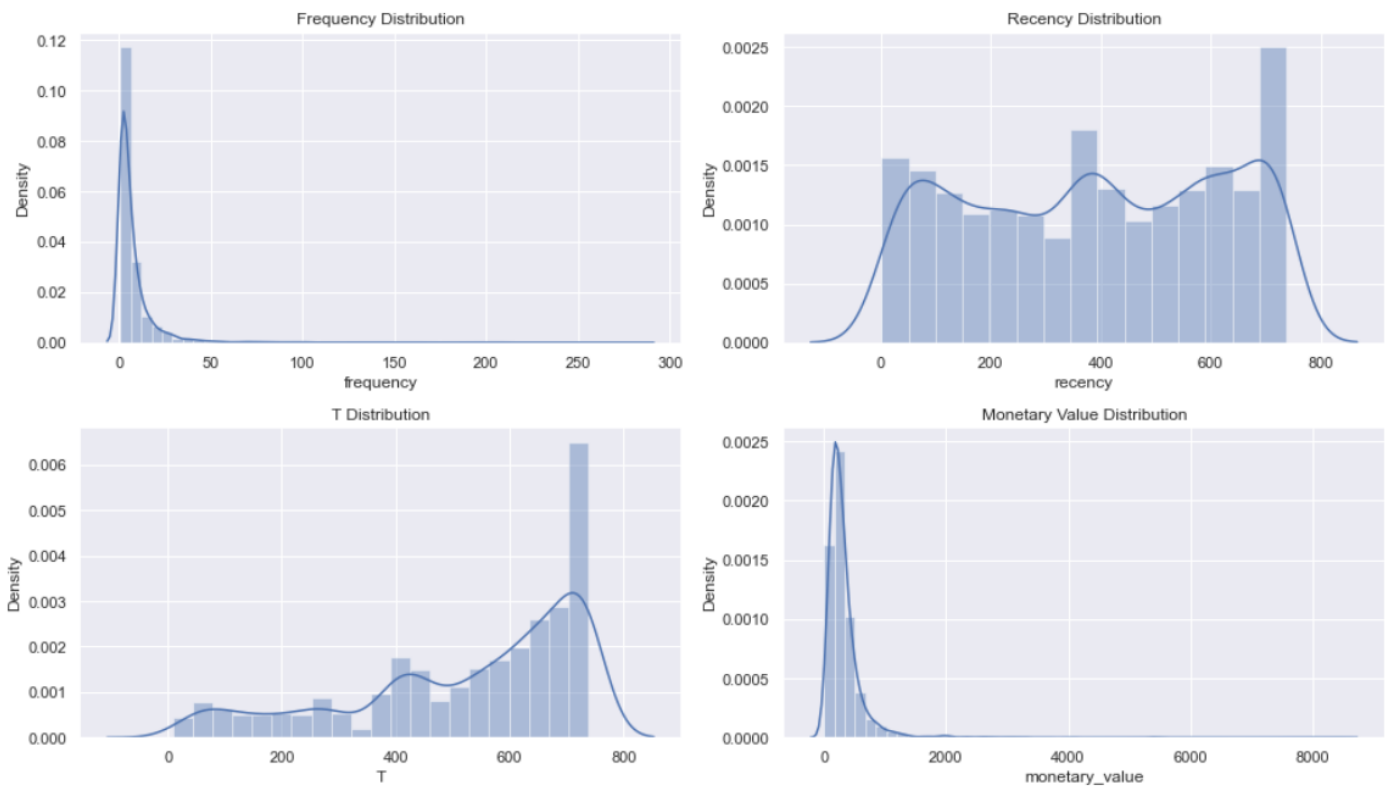
Де, T - це тривалість періоду спостереження за клієнтом, заснована на значенні, що використовується для параметра freq (за замовчуванням частота дорівнює дню). Це також можна представити як тривалість часу між першою

покупкою клієнта та кінцем періоду, що вимірюється з кроком, що визначається частотою freq;

Recency - це момент (щодо власного періоду спостереження), коли покупець зробив останню повторну покупку. Це також можна представити як проміжок часу між першою та останньою покупкою клієнта за період, що вимірюється з кроком, що визначається freq;

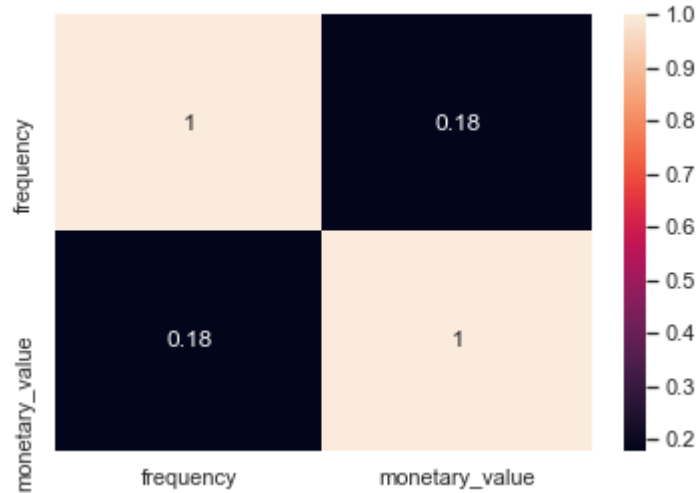
Frequency - це кількість повторних покупок, здійснених клієнтом у період спостереження;

Monetary - це середня грошова сума кожної повторної купівлі, скоєної клієнтом.



Графіки розповсюдження основних показників RFM-аналізу

Аби використати Gamma-Gamma model для знаходження середньої ціни транзакції нам необхідно впевнитись, що між значеннями частоти та грошової цінності немає кореляції.

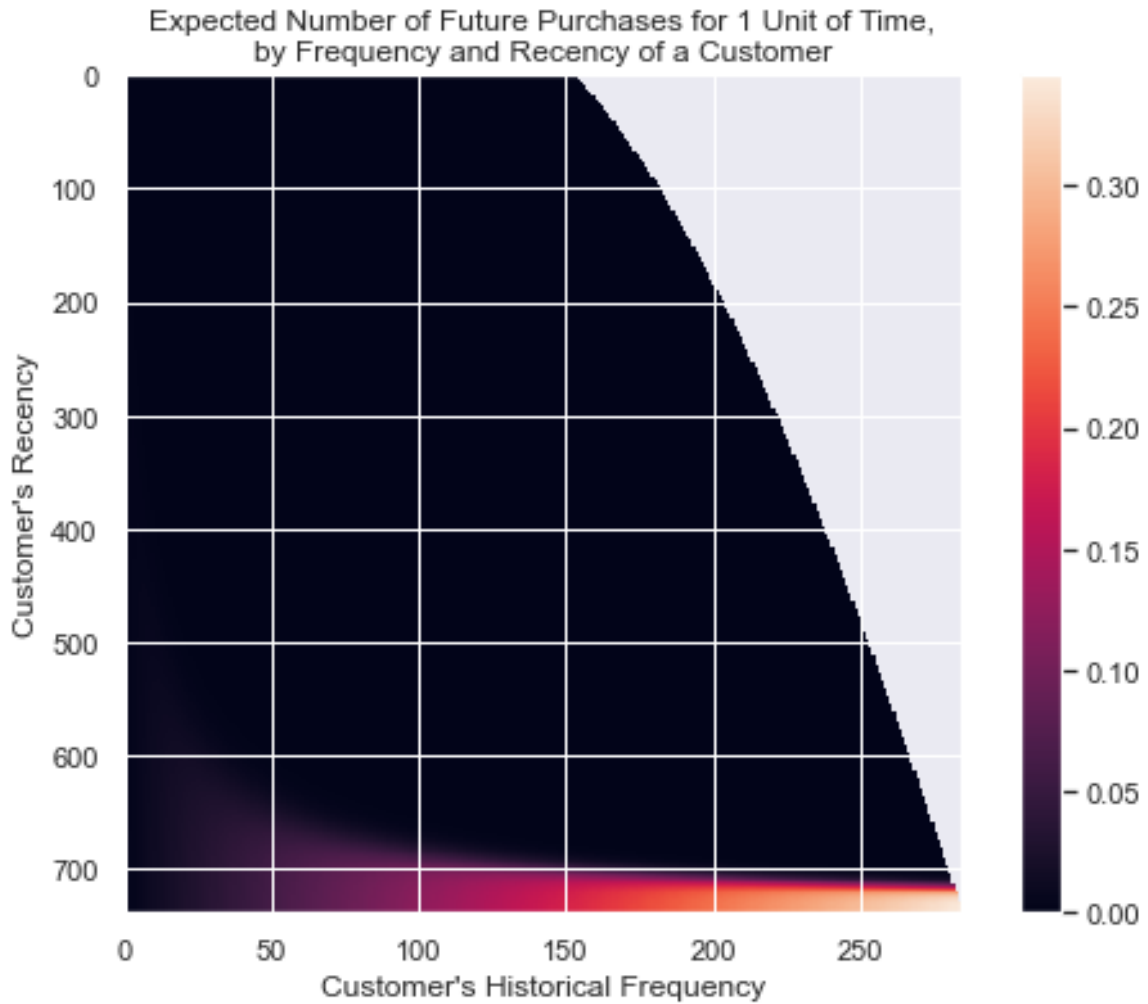


Кореляція періодичності і грошової цінності

4.3 Використання моделі Pareto/NBD

На основі даних RFM-моделі створимо Pareto/NBD модель.

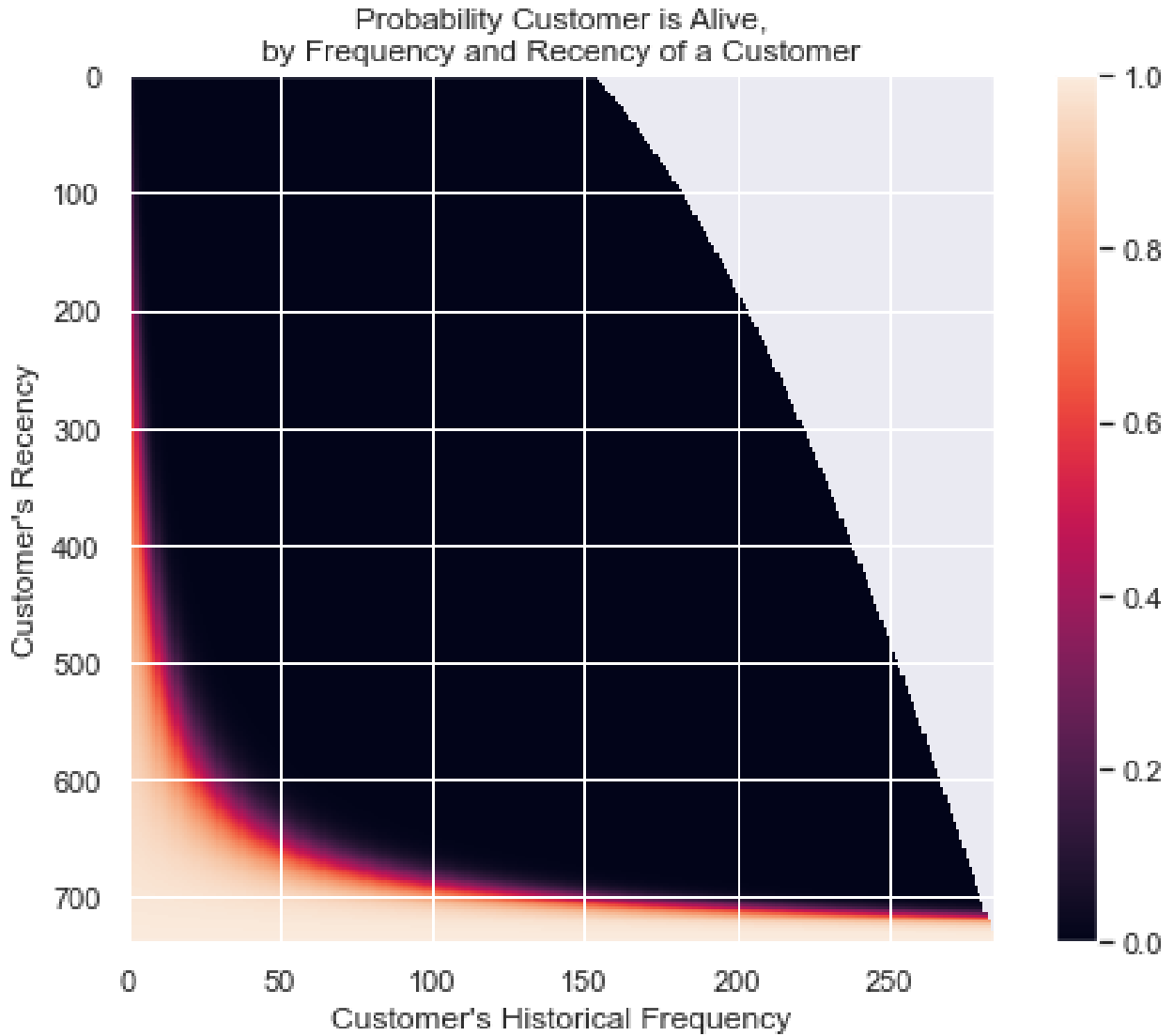
Маємо такі значення параметрів: $\alpha - 87.62$, $\beta - 9263.73$, $r - 1.43$, $s - 3.06$.



Очікувана кількість транзакцій за 1 одиницю часу(Pareto/NBD model)

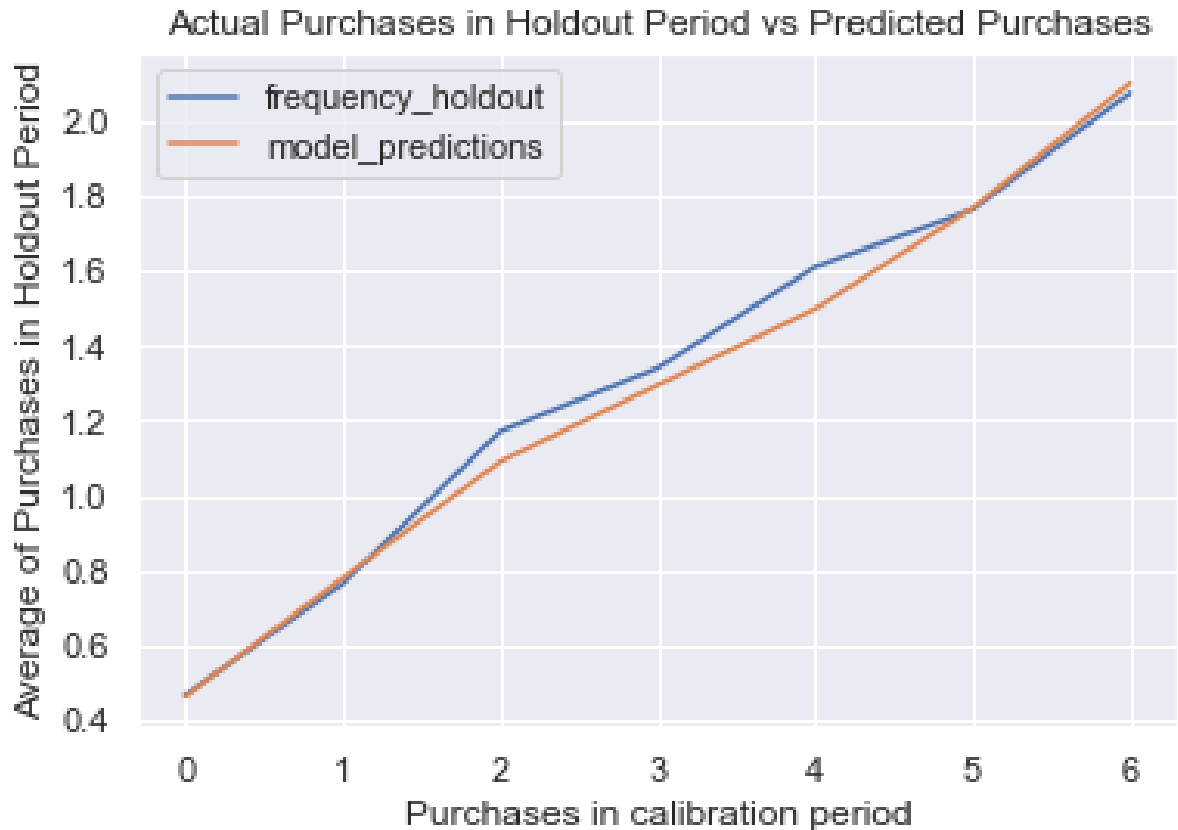
З графіку можемо винести, що найкращими клієнтами для нас є ті які мають частоту 250+ і періодичність 700+. Найхолодніші клієнти – це ті, які знаходяться у правому верхньому кутку: вони швидко зробили багато покупок, та вже давно не купували нічого.

Також маємо «хвіст» між значеннями періодичності 50 та 250. Це клієнти, які купують нечасто, але ми бачили їх недавно, так що вони здійснити транзакцію знову. Ми не впевнені, чи «померли» вони, чи просто знаходяться в періоді між купівлею.



Графік вірогідності того що клієнт досі користується послугами компанії(Pareto/NBD model)

Як ми бачимо на цьому графіку близько 75% клієнтів уже не купують у цьому магазині.



Графік порівняння прогнозованої кількості покупок з реальними (Pareto/NBD model)

Побудуємо Gamma-Gamma модель.

Customer ID	frequency	recency	T	monetary_value	p_not_alive	p_alive	predicted_purchases	actual_30_p	p_error	expected_avg_sales_p	
0	12347.0	7.0	402.0	404.0	717.398571	0.000644	0.999356	0.511782	0.522388	0.010606	621.092737
1	12348.0	4.0	363.0	438.0	449.310000	0.035940	0.964060	0.297484	0.330579	0.033095	406.903061
2	12349.0	4.0	717.0	735.0	1107.172500	0.005842	0.994158	0.196041	0.167364	-0.028677	820.949271
3	12352.0	8.0	356.0	392.0	218.182500	0.016647	0.983353	0.577401	0.674157	0.096756	244.728338
4	12353.0	1.0	204.0	408.0	89.000000	0.116063	0.883937	0.129509	0.147059	0.017550	261.607852

Таблиця з прогнозованими середніми грошовими витратами за покупку (Pareto/NBD model).

Маємо такі дані похибки:

Середня прогнозовані продажі: 330.6

Реальні продажі: 340.87

Різниця: -10.24

Середньоквадратична помилка: 23078

Корінь з середньоквадратичної помилки: 151.9

Тепер на основі отриманих розрахунків ми можемо спрогнозувати позиттєву цінність клієнта.

	Customer ID	frequency	recency	T	monetary_value	p_not_alive	p_alive	predicted_purchases	actual_30_p	p_error	expected_avg_sales_p	predicted_clv_p	CLV_p
0	12347.0	7.0	402.0	404.0	717.398571	0.000644	0.999356	0.511782	0.522388	0.010606	621.092737	6451.358984	322.567949
1	12348.0	4.0	363.0	438.0	449.310000	0.035940	0.964060	0.297484	0.330579	0.033095	406.903061	2274.185355	113.709268
2	12349.0	4.0	717.0	735.0	1107.172500	0.005842	0.994158	0.196041	0.167364	-0.028677	820.949271	3172.596116	158.629806
3	12352.0	8.0	356.0	392.0	218.182500	0.016647	0.983353	0.577401	0.674157	0.096756	244.728338	2841.238747	142.061937
4	12353.0	1.0	204.0	408.0	89.000000	0.116063	0.883937	0.129509	0.147059	0.017550	261.607852	498.869763	24.943488

Таблиця. Розрахунок CLV(Pareto/NBD model)

Похибка прогнозування CLV:

Середня помилка значень: 105.6

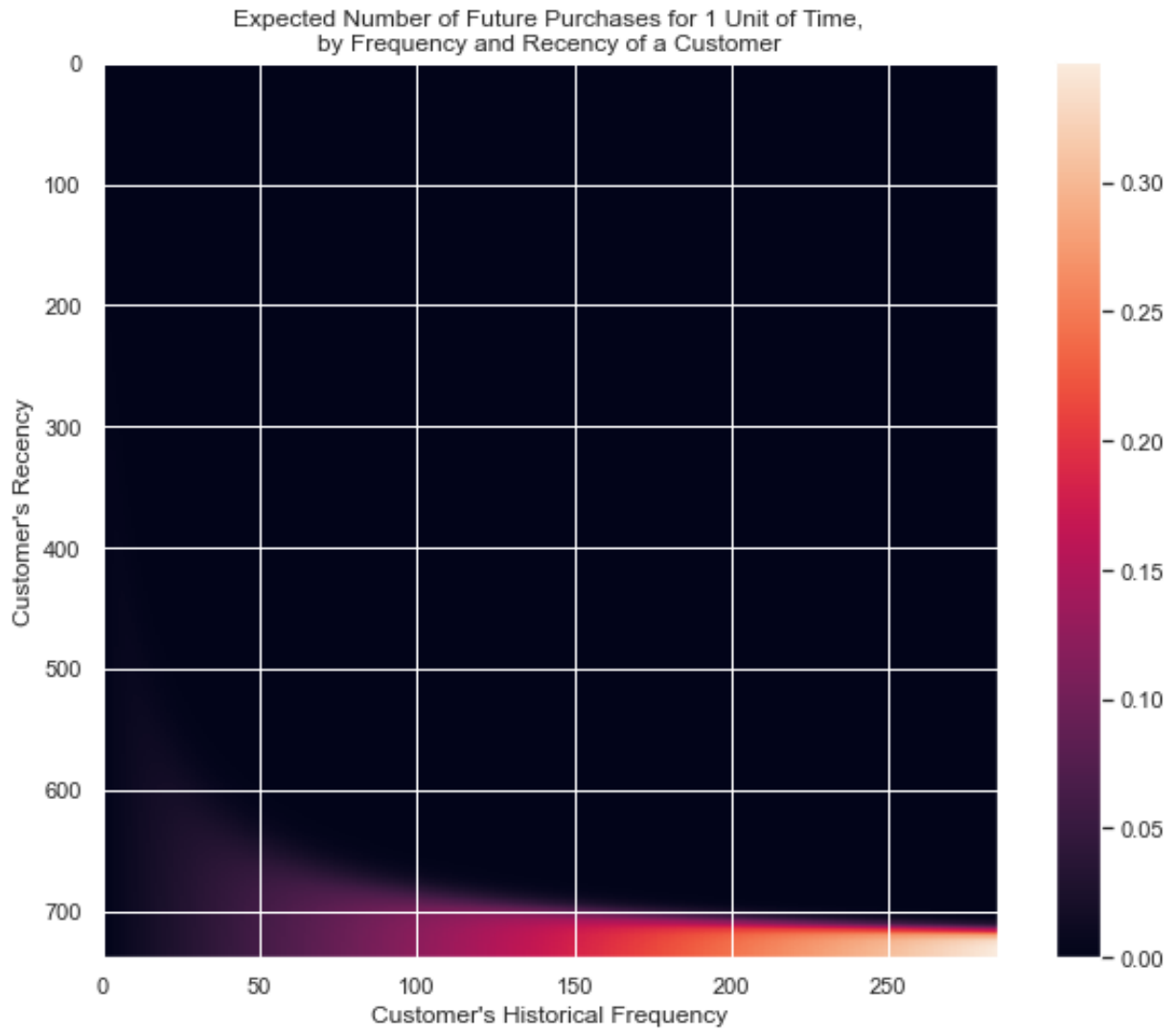
Середньоквадратична помилка: 916742.

Корінь з середньоквадратичної помилки: 957.

4.4 Використання моделі BG/NBD

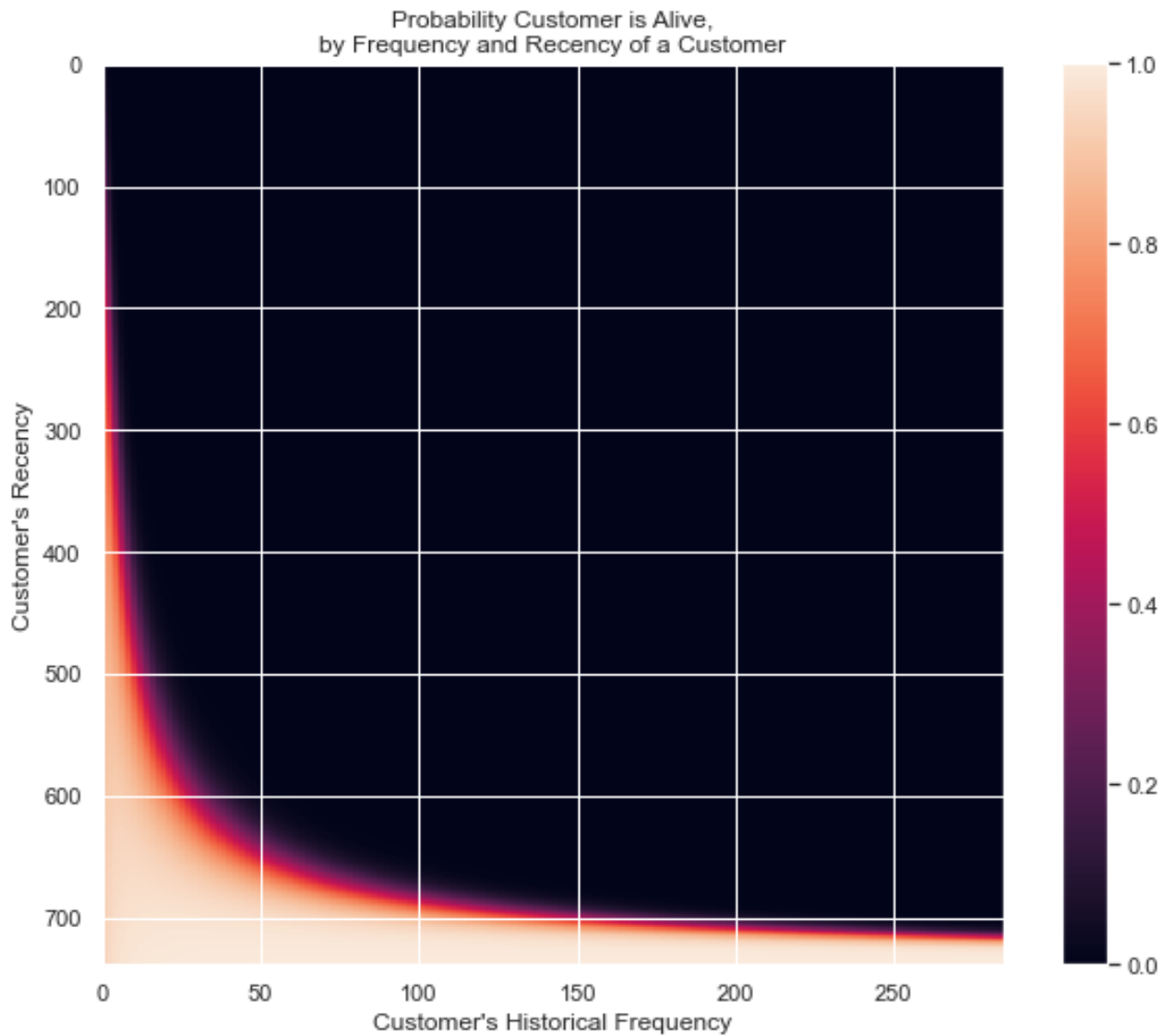
На основі наявних даних створимо Beta Geometry модель.

Маємо такі значення параметрів: $a = 0.16$, $\alpha = 85.43$, $b = 2.45$, $r = 1.5$.



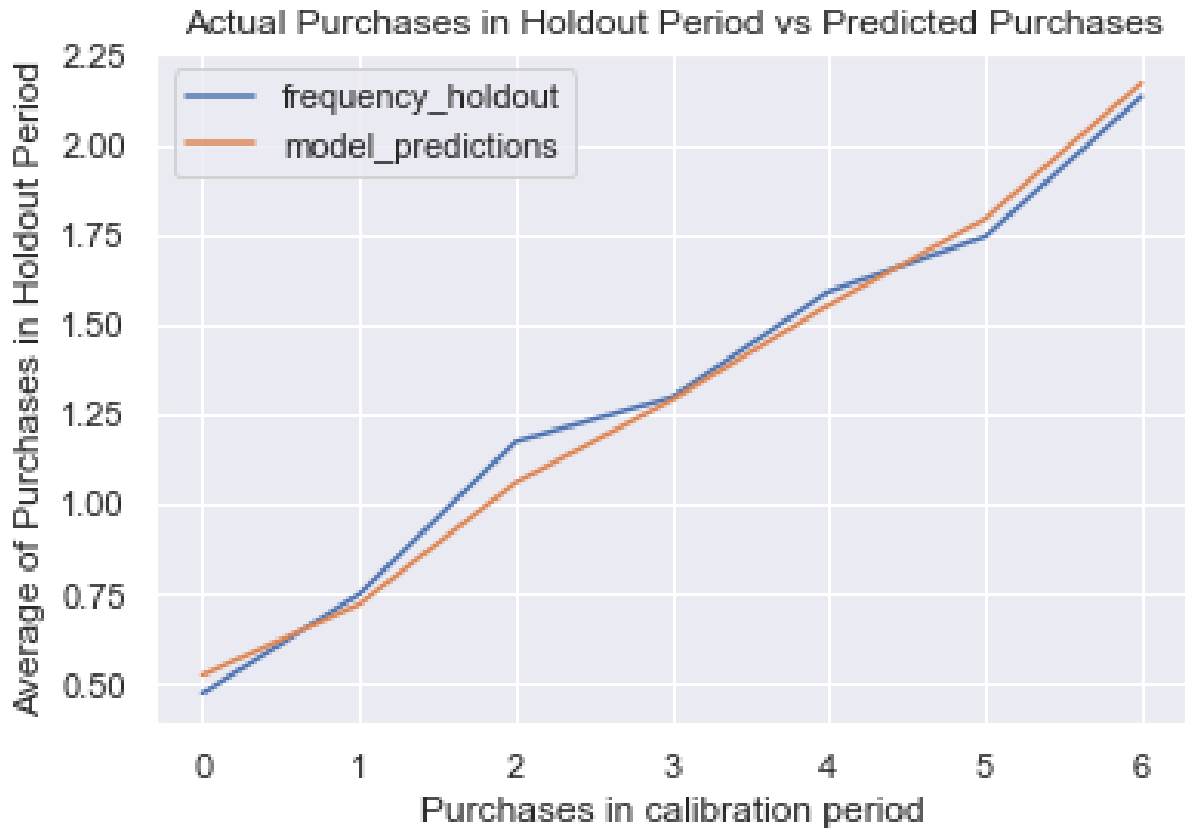
Очікувана кількість транзакцій за 1 одиницю часу(BG/NBD model)

Матриця дуже схожа на матрицю моделі Pareto/NBD. Отримали майже ті самі значення.



Графік вірогідності того що клієнт досі користується послугами компанії(BG/NBD model)

Як ми бачимо на цьому графіку близько 80% наших клієнтів уже не купують у цьому магазині.



Графік порівняння прогнозованої кількості покупок з реальними(BG/NBD model)

Тепер застосуємо Гамма-гамма модель..

	Customer ID	frequency	recency	T	monetary_value	predicted_purchases	actual_30	Error	Expected_Avg_Sales
0	12347.0	7.0	402.0	404.0	717	0.508087	0.522388	0.014301	627.806581
1	12348.0	4.0	363.0	438.0	449	0.293047	0.330579	0.037531	409.124871
2	12349.0	4.0	717.0	735.0	1107	0.193825	0.167364	-0.026461	838.789505
3	12352.0	8.0	356.0	392.0	218	0.572903	0.674157	0.101254	242.371147
4	12353.0	1.0	204.0	408.0	89	0.120620	0.147059	0.026439	255.679502

.Таблиця з прогнозованими середніми грошовими витратами за покупку(BG/NBD model).

Похибка прогнозу:

Середня прогнозовані продажі: 339.2

Реальні продажі: 330.2

Різниця: -9.6

Середньоквадратична помилка: 21244

Корінь з середньоквадратичної помилки: 144

	Customer ID	frequency	recency	T	monetary_value	predicted_purchases	actual_30	Error	Expected_Avg_Sales	predicted_clv	CLV
0	12347.0	7.0	402.0	404.0	717	0.508087	0.522388	0.014301	627.806581	7021.114308	355.020998
1	12348.0	4.0	363.0	438.0	449	0.293047	0.330579	0.037531	409.124871	2495.788566	125.531007
2	12349.0	4.0	717.0	735.0	1107	0.193825	0.167364	-0.026461	838.789505	3363.804924	171.867956
3	12352.0	8.0	356.0	392.0	218	0.572903	0.674157	0.101254	242.371147	3135.694192	155.364086
4	12353.0	1.0	204.0	408.0	89	0.120620	0.147059	0.026439	255.679502	550.760862	26.913998

Таблиця. Розрахунок CLV(BG/NBD model)

Порівняємо значення прогнозованих значень з реальними і запишемо похибку:

Середня помилка значень: 107.

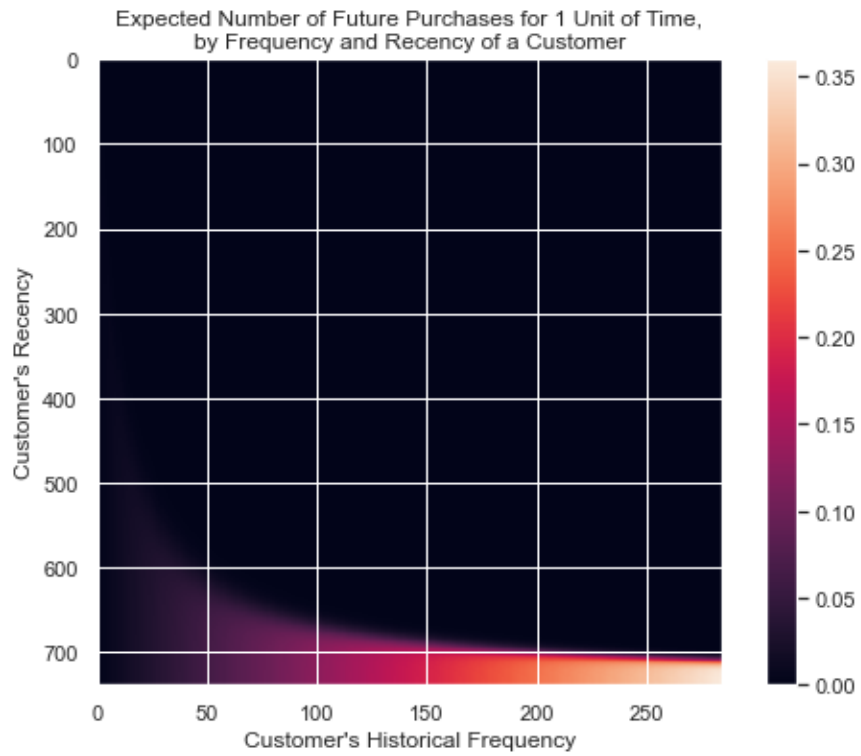
Середньоквадратична помилка: 919946.

Корінь з середньоквадратичної помилки: 957.

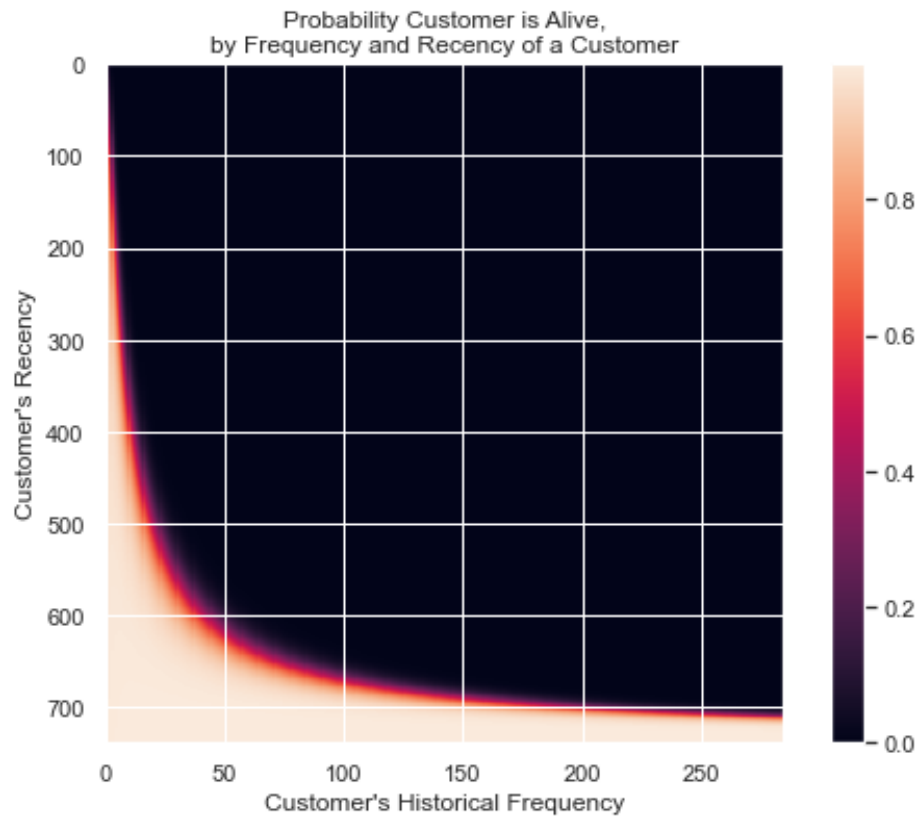
4.5 Використання моделі MBG/NBD

На основі наявних даних створимо MBG/NBD модель.

Маємо такі значення параметрів: $a = 0.02$, $\alpha = 52.7$, $b = 0.4$, $r = 0.87$.

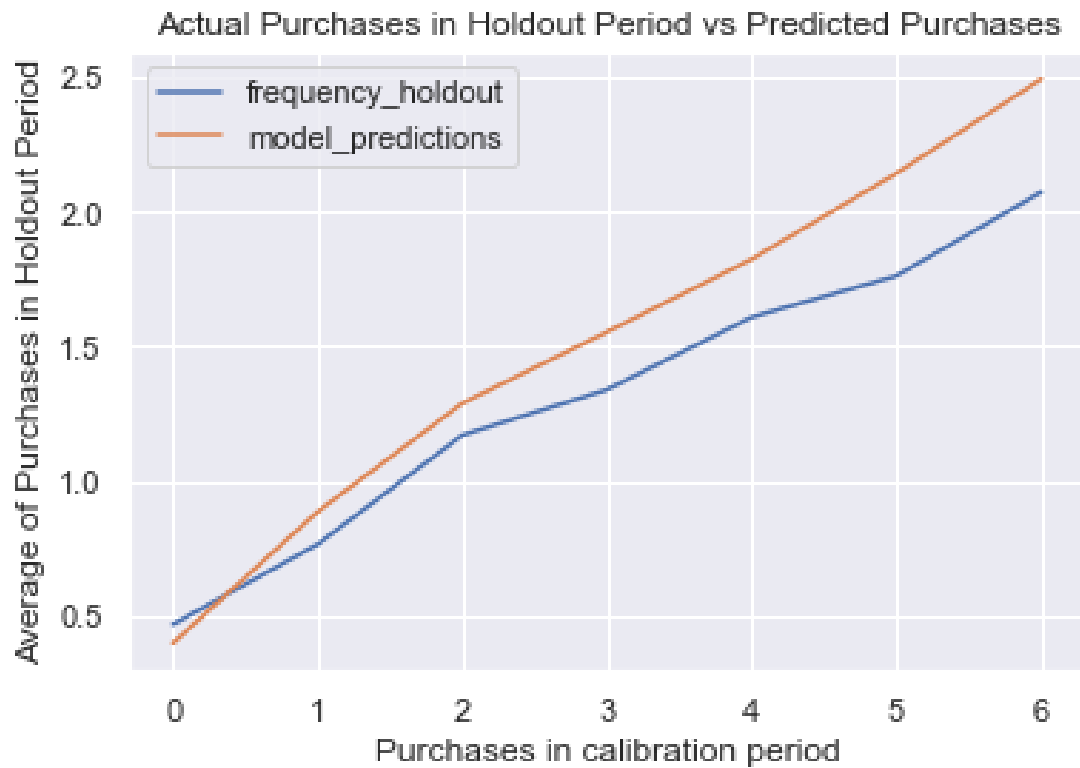


Очікувана кількість транзакцій за 1 одиницю часу(MBG/NBD model)



.Графік вірогідності того що клієнт досі користується послугами компанії(BG/NBD model)

Як ми бачимо на цьому графіку близько 85% наших клієнтів уже не купують у цьому магазині.



Графік порівняння прогнозованої кількості покупок з реальними(MBG/NBD model)

Тепер застосуємо Гамма-гамма модель.

	Customer ID	frequency	recency	T	monetary_value	predicted_purchases	actual_30	Error	Expected_Avg_Sales
0	12347.0	7.0	402.0	404.0	717.398571	0.491460	0.522388	0.030928	621.092737
1	12348.0	4.0	363.0	438.0	449.310000	0.263757	0.330579	0.066821	406.903061
2	12349.0	4.0	717.0	735.0	1107.172500	0.170717	0.167364	-0.003353	820.949271
3	12352.0	8.0	356.0	392.0	218.182500	0.558933	0.674157	0.115225	244.728338
4	12353.0	1.0	204.0	408.0	89.000000	0.089574	0.147059	0.057485	261.607852

Таблиця з прогнозованими середніми грошовими витратами за покупку(MBG/NBD model).

Маємо таку похибку:

Середня прогнозовані продажі: 340.9

Реальні продажі: 330.6

Різниця: -10.29

Середньоквадратична помилка: 23078

Корінь з середньоквадратичної помилки: 151.9

	Customer ID	frequency	recency	T	monetary_value	predicted_purchases	actual_30	Error	Expected_Avg_Sales	predicted_clv	CLV
0	12347.0	7.0	402.0	404.0	717.398571	0.491460	0.522388	0.030928	621.092737	7897.078134	394.853907
1	12348.0	4.0	363.0	438.0	449.310000	0.263757	0.330579	0.066821	406.903061	2872.043291	143.602165
2	12349.0	4.0	717.0	735.0	1107.172500	0.170717	0.167364	-0.003353	820.949271	3652.250189	182.612509
3	12352.0	8.0	356.0	392.0	218.182500	0.558933	0.674157	0.115225	244.728338	3598.495800	179.924790
4	12353.0	1.0	204.0	408.0	89.000000	0.089574	0.147059	0.057485	261.607852	646.607503	32.330375

Таблиця. Розрахунок CLV(MBG/NBD model)

Похибка розрахунку CLV:

Середня помилка значень: 107.54

Середньоквадратична помилка: 916946.

Корінь з середньоквадратичної помилки: 957,6.

4.6 Порівняння результатів прогнозування.

Тепер порівняємо помилки

	BG-NBD	Pareto-NBD	MBG-NBD
MSE Purchase Error	2.475712	2.455403	2.474593
RMSE Purchase Error	1.573440	1.566973	1.573084
Avg Purchase Error	0.360071	0.352917	0.372685

Таблиця. Порівняння помилок моделей прогнозування

Як ми бачимо, немає великої різниці у продуктивності цих моделей, але модель Pareto/NBD показує трішки кращі результати.

4.7 Використання результатів.

Візьмемо прогнозовані дані моделі Pareto/NBD і сегментуємо користувачів залежно від прогнозованих продаж, аби зрозуміти на кому сфокусувати маркетинговий таргет.

	Customer ID	frequency	recency	T	monetary_value	p_not_alive	p_alive	predicted_purchases	actual_30_p	p_error	expected_avg_sales_p	predicted_clv_p	CLV_p
0	12347.0	7.0	402.0	404.0	717.398571	0.000644	0.999356	0.511782	0.522388	0.010606	621.092737	6451.358984	322.567949
1	12348.0	4.0	363.0	438.0	449.310000	0.035940	0.964060	0.297484	0.330579	0.033095	406.903061	2274.185355	113.709268
2	12349.0	4.0	717.0	735.0	1107.172500	0.005842	0.994158	0.196041	0.167364	-0.028677	820.949271	3172.596116	158.629806
3	12352.0	8.0	356.0	392.0	218.182500	0.016647	0.983353	0.577401	0.674157	0.096756	244.728338	2841.238747	142.061937
4	12353.0	1.0	204.0	408.0	89.000000	0.116063	0.883937	0.129509	0.147059	0.017550	261.607852	498.869763	24.943488

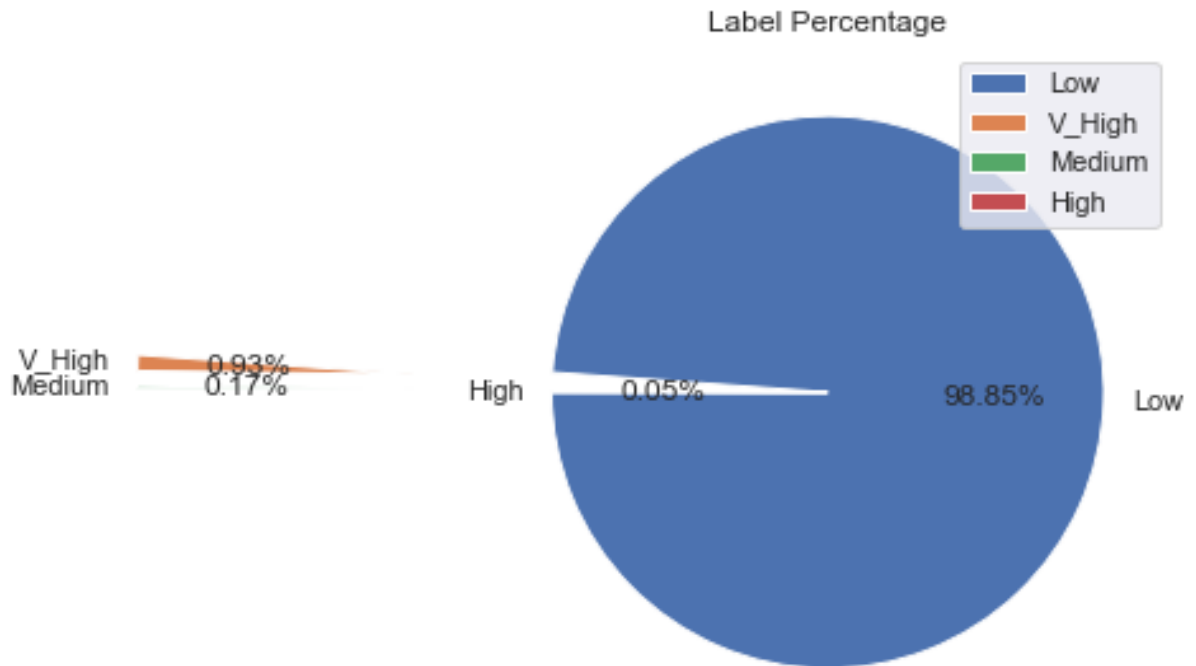
Таблиця. Дані прогнозу Pareto/NBD моделі

Розділимо усі наші значення на градації.

Labels	High	Low	Medium	V_High
Customer ID	16374.000000	15302.435234	15166.461538	14890.428571
frequency	83.000000	6.956742	68.128205	94.142857
recency	737.000000	385.801112	671.256410	566.142857
T	737.500000	523.651764	676.717949	573.428571
monetary_value	6952.630519	315.169142	1103.292669	3244.581676
p_not_alive	0.000162	0.155350	0.005650	0.003835
p_alive	0.999838	0.844650	0.994350	0.996165
predicted_purchases	3.055491	0.382696	2.631190	3.908173
actual_30_p	3.379286	0.735102	2.965654	4.674437
p_error	0.323795	0.352406	0.334464	0.766264
expected_avg_sales_p	6758.956897	326.493802	1052.377347	3045.494321
predicted_clv_p	441102.895667	2472.128671	47493.394713	184770.184271
CLV_p	22055.144783	123.606434	2374.669736	9238.509214

Таблиця. Поділ клієнтів на градації базуючись на отриманих прогнозованих значеннях

Маємо такий розподіл:



Графік відсотку розподілу клієнтів за майбутньою цінністю

Висновки

В кваліфікаційній роботі був розглянутий метод сегментації клієнтської бази та на його основі впроваджені найпопулярніші моделі для прогнозування кількості транзакцій, а саме Pareto/NBD, BG/NBD, MBG/NBD і Gamma-Gamma модель для прогнозування середньої ціни транзакції, які в сукупності дали можливість спрогнозувати пожиттєву цінність клієнта.

Для реалізації і порівняння результатів роботи різних моделей був використаний найскладніший випадок прогнозування пожиттєвої цінності клієнта – випадок неконтрактного бізнесу з безперервними купівельними можливостями. Датасет відображає онлайн транзакції магазину роздрібною торгівлі в Британії в період 2010-2012 років. Задля якісного аналізу над базою даних були виконані такі кроки: підготовка даних, сегментація клієнтської бази, прогноз пожиттєвої цінності клієнта за трьома моделями та використання отриманого прогнозу.

Результатом впровадження і тренування моделей стали дуже схожі прогнози. Та все ж найбільшу ефективність показала модель Pareto/NBD. Надані результати показують високу точність прогнозу, а тому можуть бути застосовані як фундамент для будування майбутньої стратегії, прийнятті бізнес-рішень, оптимізації процесів, тощо.

На основі прогнозованої моделі була створена база клієнтів, які є найбільш цінними для компанії – ті які мають найбільші показники прогнозованих значень на найближчий період. Ці дані може використовувати відділ маркетингу для ефективно реалізації маркетингового бюджету.

Список використаних джерел

1. <https://www.mathworks.com/discovery/predictive-analytics.html>
2. <https://www.techtarget.com/searchbusinessanalytics/tip/5-step-predictive-analytics-process-cycle>
3. Kumar, V. & Reinartz, W. (2006). Customer Relationship Management: A Databased Approach. – Wiley. – 352 p.
4. Wang, P. & Splegel, T. (1994). Database marketing and its measurements of success // Journal of Direct Marketing. – 8. – pp. 73-81
5. Vasuk, E.N. (2013). Klientskie soobshchestva kak instrument privlechenija i uderzhanija klientov // Klienting i upravlenie klientskim portfelem. – 3. – pp. 208-
6. Akulich, M. (2016) Lojalnost' klienta: Ponjatija, programma, analiz, primery, sposoby povyshenija. – Izdatelskie reshenija. – 202 p./
7. McAlexander, J.H., Schouten, J.W. & Koenig, H.F. (2002). Building brand community // Journal of Marketing. – 66(1). – pp. 38-54
8. Ryals, L. (2008). Determining the indirect value of a customer // Journal of Marketing Management Special Issue Marketing / Finance Interface. – 24(7-8). –
9. Bermejo, G. & Rodríguez Monroy, C. (2010). How to measure customer value and its relationship with shareholder value in a business-to-business market // Intangible Capital. – 6(2). – pp. 142-161
10. <https://piwik.pro/blog/customer-lifetime-value-important-for-your-business/#how-to-apply-customer-lifetime-value-clv>
11. <https://www.uplab.ru/blog/unit-economics-how-to-evaluate-the-success-of-the-business/>
12. <https://www.profitwell.com/recur/all/unit-economics>

13. http://www.brucehardie.com/notes/009/pareto_nbd_derivations_2005-11-05.pdf
14. http://www.brucehardie.com/notes/039/bgnbd_derivation_2019-11-06.pdf
15. <https://medium.com/bolt-labs/understanding-the-customer-lifetime-value-with-data-science-c14dcafa0364>
16. <https://dokumen.tips/documents/erratum-on-the-mbgnbd-model.html?page=2>
17. http://www.brucehardie.com/notes/025/gamma_gamma.pdf
18. <https://www.actioniq.com/blog/what-is-rfm-analysis/>
19. <https://www.techtarget.com/searchdatamanagement/definition/RFM-analysis>
20. <https://www.investopedia.com/terms/r/rfm-recency-frequency-monetary-value.asp>
21. <https://towardsdatascience.com/buy-til-you-die-predict-customer-lifetime-value-in-python-9701bfd4ddc0>
22. <https://towardsdatascience.com/predicting-customer-lifetime-value-with-buy-til-you-die-probabilistic-models-in-python-f5cac78758d9>
23. <https://towardsdatascience.com/data-driven-growth-with-python-part-3-customer-lifetime-value-prediction-6017802f2e0f>
24. <https://colab.research.google.com/drive/1yxUkWZSBJWIoXBai7MIxvKwKNgr0NXDH#scrollTo=3LqDPrTdBtvX>