**Старт роботи**

Першим кроком для початку аналізу є під’єднання нашої бази даних(Далі - БД) до MySQL аби в подальшому використовувати отримані запити для роботи, візуалізації, аналізу тощо та первинний аналіз БД для перевірки доречності даних.

Кожна колонка даних має певний тип, тому при додаванні БД потрібно вказати тип, що буде підходити і система розуміла дані. Вкажемо їх:

|  |  |
| --- | --- |
| RowNumber | int |
| **CustomerId** | int PK |
| Surname | text |
| CreditScore | int |
| Geography | text |
| Gender | text |
| Age | int |
| Tenure | int |
| Balance | double |
| NumOfProducts | int |
| HasCrCard | int |
| IsActiveMember | int |
| EstimatedSalary | double |
| Exited | int, |

Де int – числовий тип даних, text - текстовий тип даних, double – тип даних, що може зберігати числа з плаваючою крапкою, PK – primary key, колонка що є унікальним ідентифікатором\*

\*типи даних відрізняються в залежності від системи, за допомогою якої створюються запити – MS SQL, MySQL, PostgresSQL тощо.

Створимо перший запит, що покаже кількість клієнтів Банку:

SELECT COUNT(CustomerID) AS CountOfCustomer FROM petproject\_finance.database

Маємо 10000 записів клієнтів в нашій БД.

Перед початком роботи з базою даних потрібно проаналізувати її для розуміння набору даних. Так, дані колонок IsActiveMember та Exited суперечать одній одній, тож виключимо ці дані з нашої бази даних для отримання справедливих даних.

DELETE FROM petproject\_finance.database

WHERE IsActiveMember = '1' AND Exited = '1'

Проведемо перевірку, чи є в нашій БД код клієнта що повторюється оскільки колонка CustomerID є ключем-ідентифікатором, через який йде взаємодія з іншими даними. Створимо наступний запит:

SELECT CustomerID FROM petproject\_finance.database

GROUP BY CustomerID

HAVING COUNT(\*) > 1;

Не маємо жодного повторення, тож усі клієнти ідентифіковані різними кодами.

**KYC – Know your client або як відомості про клієнта допомагають у роботі**

Для кожного банку важливо зрозуміти, ким є його клієнти. Тому банки збирають якомога більше інформації для подальшого використання. Це може бути аналіз клієнтів, їх поведінки та інших факторів, створення персональних пропозицій для них, вирахування загальних показників роботи банку тощо.

Наша БД має широкий спектр даних про клієнтів, тож для загального розуміння клієнтів банку візуалізуємо їх.

Винесемо найбільш показові дані, такі як кількість клієнтів, вік, стать, країна, термін користування банком, баланс та вказана заробітна плата.

Для зручності візуалізації та сприйняття, згрупуємо баланс та заробітну плату на підгрупи.

Нехай підгрупи балансу будуть наступні(в у.о.):

* Пустий баланс, або «0»;
* «0 - 50тис.» (в діапазоні від 1 до 50000);
* «50тис. – 100тис.» (в діапазоні від 50000.01 до 100000);
* «100тис.+»(більше 100000).

Напишемо запит:

SELECT \*,

CASE

WHEN Balance = 0 THEN '0'

WHEN Balance BETWEEN 1 AND 50000 THEN '0-50k'

WHEN Balance BETWEEN 50000.01 AND 100000 THEN '50k-100k'

WHEN Balance BETWEEN 100000.01 AND 150000 THEN '100k-150k'

ELSE '150k+'

END AS balance\_group

FROM petproject\_finance.database

Підгрупами заробітної плати оберемо:

* Відсутні дані, або «0»;
* «0 - 50тис.» (в діапазоні від 1 до 50000);
* «50тис. – 100тис.» (в діапазоні від 50000.01 до 100000);
* «100тис. – 150тис.» (в діапазоні від 100000.01 до 150000).
* «150тис.+» (більше 150000)

Запит:

SELECT \*,

CASE

WHEN EstimatedSalary BETWEEN 1 AND 25000 THEN '0-25k'

WHEN EstimatedSalary BETWEEN 25000.01 AND 50000.01 THEN '25k-50k'

WHEN EstimatedSalary BETWEEN 50000.01 AND 100000 THEN '50k-100k'

WHEN EstimatedSalary BETWEEN 100000.01 AND 150000 THEN '100k-150k'

ELSE 'More 150k'

END AS salary\_group

FROM petproject\_finance.database

Створимо додатково запроси що покажуть кількість клієнтів за кредитним рейтингом та кількості продуктів.

Кількість продуктів:

SELECT \*,

CASE

WHEN NumOfProducts = 1 THEN '1'

WHEN NumOfProducts = 2 THEN '2'

WHEN NumOfProducts = 3 THEN '3'

ELSE '4'

END AS numofproducts\_group

FROM petproject\_finance.database

Кредитний рейтинг:

SELECT \*,

CASE

WHEN CreditScore BETWEEN 350 AND 450 THEN '350-450'

WHEN CreditScore BETWEEN 451 AND 550 THEN '450-550'

WHEN CreditScore BETWEEN 551 AND 650 THEN '550-650'

WHEN CreditScore BETWEEN 651 AND 750 THEN '650-750'

WHEN CreditScore BETWEEN 751 AND 850 THEN '750-850'

END AS creditscore\_group

FROM petproject\_finance.database

Додамо параметр кількості людей за віком по групах для більш зручного порівняння у майбутньому:

SELECT

CASE

WHEN Age BETWEEN 18 AND 25 THEN '18-25'

WHEN Age BETWEEN 26 AND 35 THEN '26-35'

WHEN Age BETWEEN 36 AND 45 THEN '36-45'

WHEN Age BETWEEN 46 AND 55 THEN '46-60'

ELSE '61+'

END AS age\_group,

COUNT(\*) AS count\_of\_people

FROM petproject\_finance.database

GROUP BY

CASE

WHEN Age BETWEEN 18 AND 25 THEN '18-25'

WHEN Age BETWEEN 26 AND 35 THEN '26-35'

WHEN Age BETWEEN 36 AND 45 THEN '36-45'

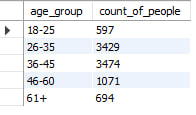
WHEN Age BETWEEN 46 AND 55 THEN '46-60'

ELSE '61+'

END

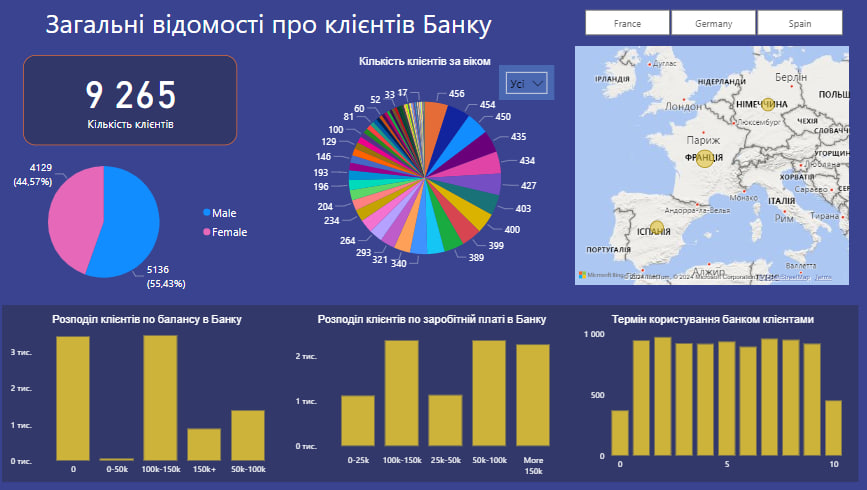
ORDER BY age\_group;

Результат:



Всі інші дані винесемо з вихідної БД – кількість людей в БД, кількість людей по віку, країни їх проживання, стать та термін перебування клієнтом банку. Зобразимо дані на візуалізації.

Фінальна візуалізація загальних даних:



Так, найбільшою групою людей за віком є від 30 до 40 років – вони складають 48% всіх записів (перевірка запитом:

SELECT COUNT(\*) FROM petproject\_finance.database WHERE Age BETWEEN 30 AND 40)

Більше половини клієнтів проживають у Франції.

Третина клієнтів не мають жодних коштів на балансі, інша третина мають на балансі суми від 100 тисяч до 150 тисяч, а найменша група налічує 67 осіб, які мають кошти на балансі, але не більше 50 тисяч.

Для банка важливо залучити клієнтів з нульовим балансом до внесення коштів, т.я. наразі вони підпадають до ризик групи і можуть покинути банк. Також важливо спонукати клієнтів з великими балансами лишатися в банку. Одним з рішень є створити сприятливі умови для депозитних вкладів.

**KYC 2.0**

Окрім загальних відомостей про клієнтів потрібно розуміти, ким є середньостатистичний клієнт банку. Це дозволить керувати різноманітними процесами, такими як стратегічне планування, керування ризиками, маркетинг, якість обслуговування, аналіз конкурентоспроможності, розробка продуктів тощо.

Для кращого розуміння варто робити аналіз середньостатистичного клієнта в розрізі країни, т.я. вони мають різний економічний та фінансовий стан.

Визначимо середнього клієнта з Іспанії. До уваги візьмемо усі показники та їх середнє значення.

Створимо запит:

SELECT

NULL AS RowNumber,

NULL AS CustomerId,

NULL AS Surname,

AVG(CreditScore) AS CreditScore,

'Spain' AS Geography,

(SELECT Gender FROM petproject\_finance.database WHERE Geography = 'Spain' GROUP BY Gender ORDER BY COUNT(\*) DESC LIMIT 1) AS Gender,

AVG(Age) AS Age,

AVG(Tenure) AS Tenure,

AVG(Balance) AS Balance,

AVG(NumOfProducts) AS NumOfProducts,

(SELECT HasCrCard FROM petproject\_finance.database WHERE Geography = 'Spain' GROUP BY HasCrCard ORDER BY COUNT(\*) DESC LIMIT 1) AS HasCrCard,

(SELECT IsActiveMember FROM petproject\_finance.database WHERE Geography = 'Spain' GROUP BY IsActiveMember ORDER BY COUNT(\*) DESC LIMIT 1) AS IsActiveMember,

AVG(EstimatedSalary) AS EstimatedSalary,

(SELECT Exited FROM petproject\_finance.database WHERE Geography = 'Spain' GROUP BY Exited ORDER BY COUNT(\*) DESC LIMIT 1) AS Exited

FROM petproject\_finance.database

WHERE Geography = 'Spain';

В свою чергу, ідеальний клієнт має ті показники клієнта, до яких банк має прагнути(в ідеалі).

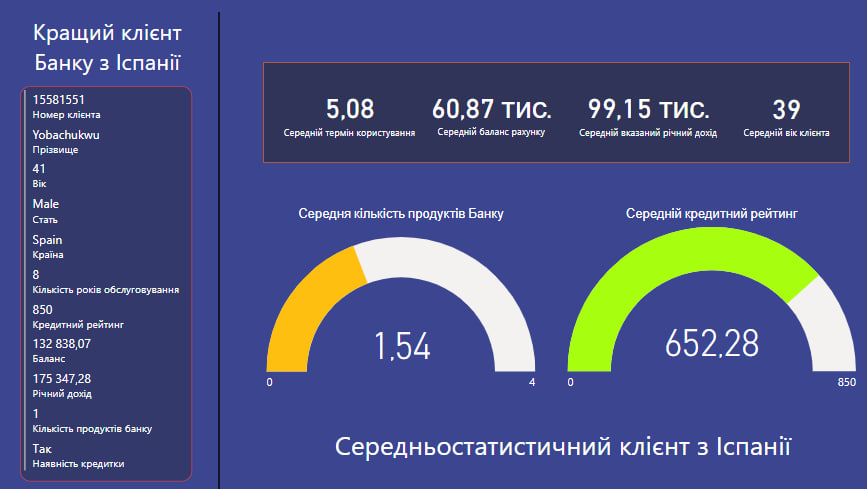
Для визначення такого клієнта було взято до уваги активного клієнта з Іспанії, що обслуговується в Банку більше 5 років, має баланс більше 100000 у.о., користується кредитною картою та має найвищий кредитний рейтинг. Запит для ідеального клієнта:

SELECT \* FROM petproject\_finance.database

WHERE Tenure > '5' AND IsActiveMember = '1' AND Balance > '100000.%' AND HasCrCard = '1' AND Geography ='Spain'

ORDER BY CreditScore DESC, EstimatedSalary DESC

LIMIT 1;



**Біль кожної компанії – відтік клієнтів**

Одним з викликів, з яким зіштовхується будь-який бізнес є відтік клієнтів. Причинами можуть бути прямі та непрямі фактори впливу на клієнтів.

До прямих можна віднести ті, що впливають безпосередньо на клієнтів – якість обслуговування, процентні ставки по депозитах та кредитах, високі комісії та збори, відсутність інновацій та погані цифрові сервіси(Інтернет-банкінг тощо) та інші.

До непрямих можна віднести загальний стан фінансового ринку, наявність особистих обставин, тобто ті, на які банк самостійно не може повпливати.

Банк N зіштовхнувся з великою кількістю клієнтів що перестали бути клієнтами банку. Створимо запит, що покаже кількість клієнтів, що покинули банк і їх кількість від загального числа:

SELECT

COUNT(DISTINCT CASE WHEN Exited ='1' THEN CustomerId END) AS churned\_clients,

COUNT(DISTINCT CustomerId) AS total\_clients,

(COUNT(DISTINCT CASE WHEN Exited ='1' THEN CustomerId END) / COUNT(DISTINCT CustomerId)) \* 100 AS churn\_rate

FROM petproject\_finance.database

Як результат, ми маємо 1302 клієнти що перестали бути клієнтами банку, а Churn Rate – показник відтоку клієнтів більше 14%. Тобто більше ніж 14% клієнтів покинули банк.

Виведемо статистику по групах для клієнтів що покинули за допомогою запитів. Використаємо групи, що застосовували для визначення клієнтів на початку дослідження.

Кількість покинувших клієнтів у розрізі заробітньої плати:

SELECT \*,

CASE

WHEN EstimatedSalary BETWEEN 1 AND 25000 THEN '0-25k'

WHEN EstimatedSalary BETWEEN 25000.01 AND 50000.01 THEN '25k-50k'

WHEN EstimatedSalary BETWEEN 50000.01 AND 100000 THEN '50k-100k'

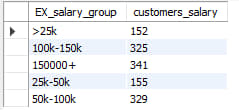
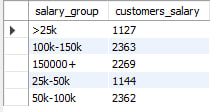
WHEN EstimatedSalary BETWEEN 100000.01 AND 150000 THEN '100k-150k'

ELSE 'More 150k'

END AS salary\_group

FROM petproject\_finance.database

WHERE Exited = '1'



Кількість покинувших клієнтів у розрізі балансу:

SELECT \*,

CASE

WHEN Balance = 0 THEN '0'

WHEN Balance BETWEEN 1 AND 50000 THEN '0-50k'

WHEN Balance BETWEEN 50000.01 AND 100000 THEN '50k-100k'

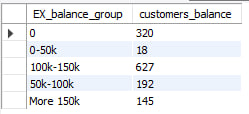
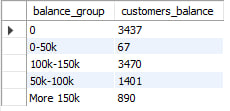
WHEN Balance BETWEEN 100000.01 AND 150000 THEN '100k-150k'

ELSE '150k+'

END AS balance\_group

FROM petproject\_finance.database

WHERE Exited = '1'



Кількість покинувших клієнтів у розрізі віку:

SELECT

CASE

WHEN Age BETWEEN 18 AND 25 THEN '18-25'

WHEN Age BETWEEN 26 AND 35 THEN '26-35'

WHEN Age BETWEEN 36 AND 45 THEN '36-45'

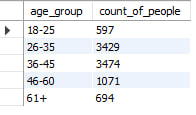
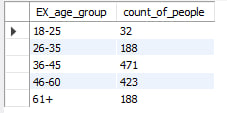
WHEN Age BETWEEN 46 AND 55 THEN '46-60'

ELSE '61+'

END AS age\_group

FROM petproject\_finance.database

WHERE Exited = '1'

Так наприклад, третину покинувших клієнтів і 40% від загальної кількості вікової категорії складають люди у віці від 46 до 60 років. Майже половина з 423 колишніх клієнтів зберігало на рахунку від 100 тис. до 150 тис. у.о. Людям старшого віку стає дедалі важче з урахуванням темпів цифровізації та прискоренням процесів, що відбуваються навколо. Вони також є більш чутливими на процеси, що впливають на банк ззовні, тож можуть приймати нелогічні рішення щодо взаємодії з банком.

Наступною аномалією що наглядається є велика кількість людей з балансами 100тис. – 150тис., що покинула банк на 9 рік обслуговування. Їхня доля складає лише 7% від загальної кількості, але порівняно з людьми що обслуговуються довше 5 років з тої самої балансової категорії, їх на 70 відсотків більше.



Для оптимізації роботи з поточними клієнтами та мінімізацією їхньої співпраці з банком складемо список, що заснований на кредитному рейтингу, а також додаткових параметрах як віку, активності, кількості продуктів у використанні:

SELECT \* FROM petproject\_finance.database

WHERE Exited = 0 AND (CreditScore < 600 OR Age > 60 OR IsActiveMember = 0 OR NumOfProducts = 1)

ORDER BY CreditScore ASC, Age DESC;

**Німецький феномен – можливість покращення**

Під час проведення дослідження було виявлено феномен. Кожен з клієнтів банку у Німеччині має кошти на балансі. Перевіримо це:

SELECT Geography, COUNT(Balance) FROM petproject\_finance.churn

WHERE Balance = '0'

GROUP BY Geography;

Як результат, жоден з клієнтів у Німеччині не має пустого балансу, навіть ті, що активно не користуються банком. Тож є пропозиція розширити кредитну лінію банку продуктом для неактивних жителів країни з балансом, але не перестали бути клієнтами банку і мають фінансове забезпечення.

Створимо запит потенційних учасників, беручи до уваги кредитний рейтинг клієнта та його заробітну плату:

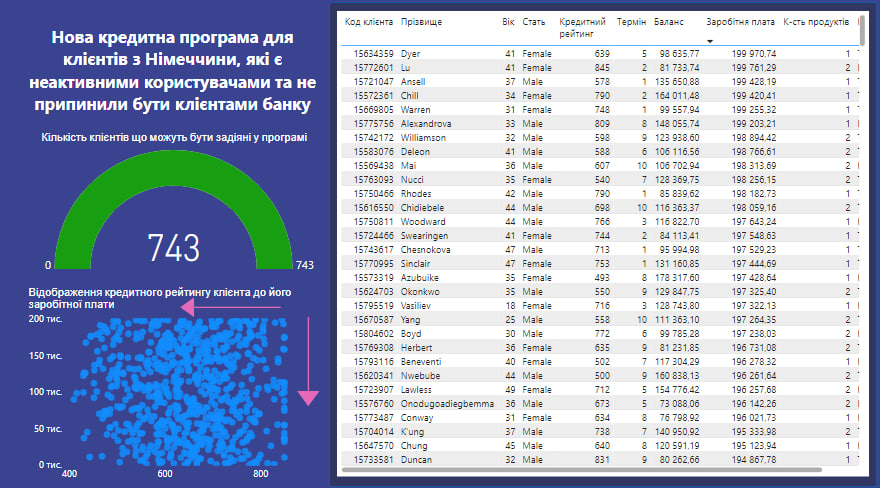
SELECT \*, IF(HasCrCard = 1, 'Так', 'Ні') AS HasCrCard\_v2 FROM petproject\_finance.database

WHERE IsActiveMember = '0' and Geography = 'Germany' and Exited = '0'

ORDER BY CreditScore DESC, EstimatedSalary DESC;

Як результат, ми можемо потенційно повернути 743 клієнта, запропонувавши їм продукт .

Стрілочками показано, в яких напрямках буде рухатися список кандидатів. Тобто кандидати у верхньому правому кутку будуть проінформовані першими, кандидати у лівому нижньому – останніми.



**Додаткові кошти для банку**

Для збільшення оборотних активів банку потрібно активніше залучати гроші клієнтів банку. Майже половина клієнтів БД є з Франції, 50% яких мають баланс від 0 до 50000у.е., і заробляють хоча б 10000у.е.

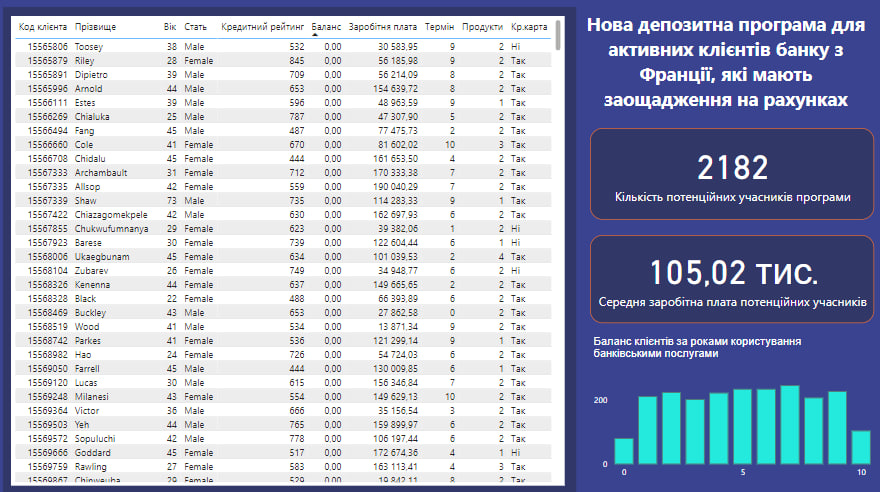
Робимо запит:

SELECT \*, IF(HasCrCard = 1, 'Так', 'Ні') AS HasCrCard\_v2 FROM petproject\_finance.database

WHERE Balance BETWEEN 0 AND 50000 AND Geography = 'France' AND EstimatedSalary > 10000

ORDER BY CreditScore DESC, NumOfProducts DESC, Tenure DESC, Balance DESC;

Отримуємо дані 2182 клієнтів з Франції, яким ми можемо запропонувати прийняти участь у новій депозитній програмі банку.



**ЗАГАЛЬНІ ВИСНОВКИ**

На основі проведеного аналізу клієнтів банку можна зробити кілька важливих висновків, які допоможуть покращити стратегію обслуговування клієнтів, оптимізувати продукти і послуги, а також зменшити відтік клієнтів.

1. Демографічна картина:

Основна частина клієнтів банку знаходиться у віковому діапазоні від 26 до 45 років, а саме 6903 клієнта або 75%, яку можна вважати найбільш платоспроможною. За статевою ознакою, перевага чоловіків є більшою на 10% - 55,4% проти 44,5% у жінок.

Однак варто ураховувати, що з них 1302 покинули банк, а у категорії 46-60 років відбувся відтік 40% від їх загальної кількості, що висуває проблеми роботи з більш старшим поколінням.

1. Кредитний рейтинг:

Значна частина клієнтів має середній кредитний рейтинг, від 550 до 750 пунктів, однак є певний відсоток клієнтів з низьким кредитним рейтингом, що потребує підвищеної уваги з боку Банку для зменшення кредитних ризиків. Таких клієнтів наразі 170, які мають кредитний рейтинг менше 450, при чому 80% з них заробляє більше ніж 50 тисяч у.о. на рік.

Високий кредитний рейтинг спостерігається у клієнтів з довготривалим стажем роботи з банком, що свідчить про лояльність і надійність таких клієнтів.

1. Фінансова активність:

Клієнти з високим балансом на рахунках частіше користуються додатковими продуктами банку, якими можуть бути інвестиційні програми або преміальні послуги.

Майже всі клієнти банку мають лише 1 або 2 продукти, що свідчить про можливість розширення використання банківських послуг через крос-продажі.

Клієнти з нульовим балансом складають майже третину, і деякі з них вже покинули банк, та є ризик втрати інших клієнтів з такою ситуацією.

1. Активність клієнтів:

Високий відсоток активних членів(більше 50%) свідчить про зацікавленість клієнтів у співпраці з банком, однак є значна кількість клієнтів що вже покинули банк і неактивні клієнти, що можуть бути під ризиком відтоку.

Потрібно звернути увагу на останніх, для того щоб не втратити лояльність клієнтів Для збільшення їх активності пропоную застосувати програми лояльності і регулярні акції що вплине і на задоволення клієнтів.

Можливі причини відтоку клієнтів:

* Несприятлива економічна ситуація на ринку;
* Відсутність інновацій;
* Відсутність адаптації до сучасних трендів та потреб клієнтів;
* Високі комісії за перекази, комунальні та інші операції;
* Проблеми з безпекою Банку;
* Відсутність якісного обслуговування клієнтів;
* Складність проведення операцій та внутрішніх процесів, тощо.

Рекомендації:

Покращення обслуговування. Потрібна перевірка поточної якості обслуговування клієнтів, впровадження програм навчання персоналу для покращення якості обслуговування. Оптимізація часу обробки запитів і підвищення доступності допоможе краще надавати підтримку клієнтам.

Адаптація продуктів. Розробка нових продуктів, що відповідають потребам різних сегментів клієнтів. Запропоновані депозитна та кредитна програми взаємодіятимуть з 30% клієнтів з бази та можуть підвищити лояльність клієнтів. Потрібна додаткова розробка інших програм лояльності для підвищення задоволеності та утримання клієнтів.

Цифрові інновації. Поліпшення мобільних додатків та онлайн-банкінгу підвищить взаємодію з клієнтами та зручність користування Банку в цілому. Впровадження нових цифрових сервісів, таких як чат-боти і персоналізовані рекомендації.

Сегментація ринку. Проведення детальнішого аналізу кожного сегменту клієнтів дасть можливість розробити спеціальні програми та пропозицій для кожного сегменту, що позитивно вплине на взаємодію клієнтів з банку та відобразиться на його загальних результатах.

Загалом, розуміння профілю середньостатистичного клієнта банку, а також причин відтоку дозволяє розробляти ефективні стратегії для утримання клієнтів, покращення продуктів і послуг, а також оптимізації внутрішніх процесів банку.