Service Friction, Digital Disengagement and Customer Attrition in Retail Banking: Evidence from Open Banking Data and a Chatbot-Led Retention Framework for UAE Banks

[1*] Sanjib Barman, [2] Praveen Kumar Sharma, [2] Lokesh Kumar, [2] Sandeep Saxena, [2] Kuldeep Singh Gour, [3] Manish Shrivastava

[1]Research Scholar, School of Management, Suresh Gyan Vihar University, Jaipur

[2] Assistant Professor, School of Management, Suresh Gyan Vihar University, Jaipur

[3] School of Hospitality, Jaipur Engineering College & Research Centre Jaipur, India

[*] Corresponding author email: sanjib19@gmail.com

Abstract

Purpose. Banks in the United Arab Emirates (UAE) are investing heavily in AI-based chatbots, yet managers still lack evidence on which customer behaviours a chatbot can realistically change and how much those behaviours matter for retention. This study isolates the behavioural signals that precede attrition and quantifies the share of attrition risk that a conversational agent can address.

Design. A secondary, quantitative design analyses two independent open retail-banking datasets: a primary sample of 10,127 credit-card customers (1,627 attrited) and a replication sample of 10,000 bank customers (2,037 exited). Five chatbot-addressable engagement levers were operationalised: inactivity, service-contact intensity, transaction frequency, engagement momentum and relationship breadth. The analysis combines Wilson confidence intervals, chi-square tests with Cramér's V, Welch t-tests with Cohen's d, nested binary logistic regression, a composite Digital Engagement Index (DEI), bootstrap what-if scenarios and a two-axis retention priority map.

Findings. Demographics barely separated leavers from stayers (Cramér's $V \leq 0.037$), whereas the engagement levers did ($V = 0.17–0.21$; $|d|$ up to 1.09). Adding the five levers to a controls-only model raised McFadden's pseudo-R^2 from 0.070 to 0.405 and the AUC from 0.693 to 0.899 (LR $\chi^2(5) = 2{,}987$, $p < .001$). Each standard-deviation rise in service contacts raised the odds of attrition by 69% and each rise in inactivity by 61%. Attrition fell from 68.1% in the least-engaged DEI decile to 1.1% in the most-engaged decile. The pattern replicated in the second bank (active members churned at 14.3% versus 26.9%; AUC 0.752 → 0.834). A combined chatbot programme implied 636 avoided attritions per 10,000 customers (95% CI 601–672), and 29% of customers in the two engagement-risk segments accounted for 80% of all attriters.

Originality and value. The study contributes a Friction–Disengagement Matrix, a transparent DEI and a Chatbot Retention Priority Map that tell a bank where a chatbot should act alone, where it should work with a relationship manager, and where it should step aside. The evidence is translated into a staged capability roadmap for UAE retail banks.

Keywords: AI chatbots; customer retention; customer engagement; service friction; digital banking; churn analytics; UAE banking

1. Introduction

Retail banking has become one of the clearest test beds for artificial intelligence in customer service. Service theory predicts that AI first absorbs mechanical and analytical tasks and then moves, more slowly, into intuitive and empathetic work [1]. Banking fits that path closely: balance enquiries, card blocking, transfer tracking and fee explanations are rule-bound and high in volume, while complaints, hardship cases and disputes still call for judgement. Digital transformation research describes the same shift from digitising single processes to redesigning how a firm

creates value with customers [2]. The AI-based chatbot, a conversational interface that interprets a customer's free-text request and responds or acts on it [3], sits at the centre of that redesign because it is often the first and most frequent point of contact a customer has with the bank.

The United Arab Emirates (UAE) is a demanding setting for this technology. Its retail banks serve a highly connected, multilingual and largely expatriate population, compete with fast-growing digital-only entrants, and operate under a national agenda that favours digital government and digital finance. Customers in such a market switch providers easily when service disappoints, so banks increasingly justify chatbot investment by its expected effect on loyalty and retention rather than on cost alone. Empirical work on AI adoption in banking confirms that customers weigh usefulness, risk and trust when deciding whether to use AI-enabled services [4], and doctoral research presented in the UAE has started to ask whether chatbots can satisfy banking customers at all [5]. What remains unclear is whether, and through which customer behaviours, chatbots can move the retention outcome that boards actually track.

Three weaknesses in the existing evidence motivate this study. First, most chatbot studies measure attitudes, satisfaction or intention through surveys. A systematic review of chatbots and loyalty concluded that a positive link is probable but still rests largely on desk research and calls for behavioural testing [6]. Retention, however, is a behaviour, and the gap between stated intention and actual exit is well documented in services marketing. Second, studies rarely separate attrition risk that a chatbot can influence from risk that it cannot. This matters because the customers who are most likely to leave are not necessarily the customers whose behaviour an intervention can change; targeting by risk alone can waste retention budgets [7]. Third, evidence typically comes from a single bank or survey sample, which leaves its generality untested.

This study responds to these gaps with behavioural data. It asks three research questions. RQ1: Which observable engagement behaviours that a chatbot can plausibly influence are associated with customer attrition in retail banking, and how strongly? RQ2: How much explanatory power do these behaviours add beyond demographic and product attributes, and does the pattern replicate in an independent bank? RQ3: How can the evidence be converted into a decision tool that tells a UAE bank where a chatbot should act, where human staff should lead, and how large the retention gain could be? The questions operationalise the first and fifth objectives of the wider research programme (the impact of chatbots on retention, and opportunities for improving chatbot performance) and test the programme's first hypothesis, that AI-based chatbots positively influence customer retention, through five measurable sub-hypotheses.

The study analyses two open, anonymised retail-banking datasets: a primary sample of 10,127 credit-card customers and a replication sample of 10,000 bank customers. Because neither dataset comes from the UAE, the paper does not claim to describe UAE customers directly. Instead it establishes an engagement–retention evidence base and a set of transparent tools that UAE banks can calibrate on their own customer records, and it frames the planned primary survey of UAE customers and professionals as the confirmatory next step.

The paper makes four contributions. (i) It separates chatbot-addressable engagement levers from non-addressable customer attributes and quantifies the incremental explanatory power of the former in two banks. (ii) It introduces a Friction–Disengagement Matrix showing how service friction and inactivity compound. (iii) It proposes a transparent Digital Engagement Index (DEI) that ranks customers by engagement without any opaque model. (iv) It develops a Chatbot Retention Priority Map and bootstrap what-if scenarios that translate statistical findings into deployment choices. Section 2 reviews the literature and develops the hypotheses, Section 3 sets out the method, Section 4 reports and discusses the results, and Section 5 concludes with implications and limitations.

2. Literature Review and Hypotheses

2.1 AI chatbots in banking: from automation to relationship

Early chatbots followed scripted decision trees; current systems classify a customer's intent from free text, retrieve account data and execute simple transactions [3]. In service theory this places them in the analytical stage of AI, where the machine handles well-defined problems at scale while humans keep the intuitive and empathetic work [1]. For banks the appeal is not only lower cost per contact but continuous availability, consistency and the chance to reach the customer before a problem becomes a complaint. Service scholars therefore argue that automated agents change the

nature of customer engagement itself: engagement is no longer only a relationship with staff and brand but also a pattern of interaction with machines that can initiate contact, remember history and personalise responses [8]. Marketing research adds that AI makes personalised engagement possible at the level of the individual customer, adjusting content, timing and channel to observed behaviour [9]. Together these views imply that the value of a banking chatbot should be judged by whether it changes engagement behaviour, and through that, retention.

2.2 Satisfaction, continuance and loyalty: what survey evidence shows

Most banking chatbot studies test adoption or satisfaction models with questionnaires. In Turkey, a survey of 240 bank customers found that perceived performance, trust and corporate reputation drove satisfaction with a banking chatbot, while prior expectations mattered only through perceived performance [10]. In Vietnam, trust, satisfaction and perceived usefulness explained 359 users' intention to keep using bank chatbots, with trust the strongest driver [11]. Beyond banking, the quality of information and service that a chatbot provides predicts satisfaction, which in turn sustains continuance intention [12], and satisfaction with customer-service chatbots has been linked to loyalty [13]. An expectation-confirmation study of digital banking users integrated AI-specific antecedents and confirmed their role in user satisfaction [14]. In Romania, compatibility with the customer's lifestyle and perceived usefulness predicted intention to bank with a chatbot, whereas ease of use and privacy risk did not [15]. The review by Jenneboer et al. [6] synthesises this stream: chatbots with good system, service and information quality build satisfaction, trust and commitment, from which loyalty is a likely but untested consequence.

This body of work is valuable but shares a structural limit. Its dependent variables are attitudes or intentions measured at one point in time, usually among people who already use the chatbot. It therefore cannot show which customers actually leave, nor whether the behaviours a chatbot can influence are the ones that precede exit.

2.3 Barriers: trust, privacy and disclosure

Research on barriers explains why adoption alone does not guarantee benefit. Interviews with Nigerian banking customers showed that the classic adoption factors still apply and that perceived expertise, responsiveness and security weigh especially heavily in how customers judge a banking chatbot [16]. Willingness to share personal information with banking chatbots depends on trust, which makes privacy a precondition for any personalised retention dialogue [17]. A field experiment found that chatbots that were not disclosed as machines sold as effectively as proficient human agents, but that disclosing the bot before the conversation sharply reduced purchases [18]. Adoption research in emerging-market banking similarly identifies perceived risk and trust as gatekeepers of AI service use [4]. These findings caution that a chatbot can damage the relationship it is meant to protect when it fails on security, competence or transparency, which is directly relevant to the security and usability concerns raised in the wider research programme.

2.4 Retention analytics and the behavioural turn

Churn research in operations and analytics takes the opposite route to the survey literature: it predicts exit from behavioural records. Hybrid models that combine logistic regression with decision trees show that interpretable churn models can be nearly as accurate as complex ones [19], and training on several time slices of behavioural data improves prediction because engagement changes over time [20]. Yet this stream rarely asks which predictors a specific intervention can move. Ascarza [7] demonstrated that targeting the customers with the highest predicted churn can be ineffective when those customers do not respond to the intervention; what matters is the part of the risk the intervention can change. For a chatbot, that part consists of behaviours that conversation can plausibly influence: activity, the experience of contacting the bank, usage frequency and the breadth of the relationship.

2.5 Conceptual framework and hypotheses

Building on these streams, the study proposes that a chatbot affects retention through five engagement levers. Each lever is observable in bank records, each can be influenced by conversational interventions, and each has a theoretical reason to relate to exit.

Disengagement (inactivity). Months without activity weaken habit and reduce the practical cost of leaving; the customer drifts before deciding to close the account. Proactive chatbot messages, reminders and contextual offers are designed to interrupt this drift [9].

> ***H1a.*** *Longer inactivity is positively associated with customer attrition.*

Service friction (contact intensity). Repeated contacts with the bank usually signal unresolved problems: a delayed card, a disputed fee, a failed transfer. Each extra contact is both a cost to the bank and a moment of irritation for the customer. A chatbot that resolves the issue at first contact, shows case status or hands over smoothly to a human can reduce this friction [12, 16].

> ***H1b.*** *Higher service-contact intensity is positively associated with customer attrition.*

Transactional engagement and momentum. Frequent transactions indicate that the bank is part of the customer's routine; a falling transaction count between quarters signals that routine moving elsewhere. In-chat insights, bill reminders and usage prompts target both the level and the trend [8].

> ***H1c.*** *Higher transaction frequency and positive engagement momentum are negatively associated with customer attrition.*

Relationship breadth. Each additional product raises switching costs and deepens the relationship. Chatbots can guide needs-based product discovery, provided the advice respects the customer's interests and consent [17].

> ***H1d.*** *A broader product relationship is negatively associated with customer attrition.*

Generality. If these mechanisms are behavioural rather than bank-specific, the association between active engagement and retention should appear in an independent bank with different customers and products.

> ***H1e.*** *The engagement–retention association replicates in an independent retail-bank sample.*

2.6 Research gap

The literature therefore offers strong evidence that chatbot quality shapes satisfaction and intention [6, 10–15], and strong evidence that behaviour predicts churn [19, 20], but little that connects the two. No study located for this review separates chatbot-addressable from non-addressable attrition risk, tests the separation in two banks, and turns the result into a deployment tool. The present study is designed to fill that gap and to give UAE banks a framework they can apply to their own data.

3. Research Methodology

3.1 Research design

The study follows a positivist, explanatory and cross-sectional design based on secondary quantitative data. It mirrors the analytical plan of the wider research programme, which combines descriptive statistics (frequencies, percentages and means) with inferential tests (chi-square, t-tests and regression), and it adds effect sizes, replication and simulation to strengthen inference. Open behavioural data were chosen deliberately for this first phase. They record what customers actually did, they are large enough for stable estimates, and anyone can re-run the analysis. The planned stratified survey of UAE bank customers and professionals is positioned as the confirmatory phase that will test the same levers with perception and satisfaction measures.

3.2 Data sources

Dataset A (primary sample). The Credit Card Customers dataset contains 10,127 customers of a single retail bank observed over twelve months, of whom 1,627 (16.07%) closed their accounts. It records demographics, card tier, credit profile and behavioural fields including months of inactivity, number of contacts with the bank, transaction counts and amounts, quarter-on-quarter change in transactions, and number of products held. A public copy is archived on Zenodo, and the analysis notebook downloads it automatically.

Dataset B (replication sample). The Bank Customer Churn dataset contains 10,000 customers of a retail bank operating in France, Germany and Spain, of whom 2,037 (20.37%) exited. It records age, gender, country, credit score, balance, estimated salary, tenure, number of products, card ownership and an active-member flag. It shares two levers with Dataset A (activity and relationship breadth) and therefore offers an independent test of the engagement–retention association.

Both datasets were selected on four criteria: an observed exit outcome, behavioural fields that map onto chatbot-addressable levers, a sample above 5,000, and open public availability. Integrity checks found no missing values and no duplicated customer identifiers in either dataset. Identifier and surname fields were removed before analysis.

3.3 Variables and operationalisation

The dependent variable is binary attrition (1 = closed the relationship, 0 = retained). Five engagement levers are the focal predictors, and six customer attributes serve as controls, together with income band and card tier as dummy variables. Table 1 defines each variable and compares retained and attrited customers.

Table 1. Variables, operationalisation and descriptive statistics (Dataset A, n = 10,127)

Variable	Construct / role	Operationalisation	All, mean (SD)	Retained (n = 8,500), mean (SD)	Attrited (n = 1,627), mean (SD)
Months inactive	Disengagement (lever)	Months with no activity in the last 12 months (0–6)	2.34 (1.01)	2.27 (1.02)	2.69 (0.90)
Service contacts	Service friction (lever)	Customer–bank contacts in the last 12 months (0–6)	2.46 (1.11)	2.36 (1.08)	2.97 (1.09)
Transaction count	Transactional engagement (lever)	Number of transactions in the last 12 months	64.86 (23.47)	68.67 (22.92)	44.93 (14.57)
Transaction momentum	Engagement momentum (lever)	Transaction count in Q4 ÷ count in Q1	0.71 (0.24)	0.74 (0.23)	0.55 (0.23)
Products held	Relationship breadth (lever)	Number of bank products held (1–6)	3.81 (1.55)	3.91 (1.53)	3.28 (1.58)
Age	Control	Age in years	46.33 (8.02)	46.26 (8.08)	46.66 (7.67)
Male	Control	1 = male, 0 = female	0.47 (0.50)	0.48 (0.50)	0.43 (0.49)
Dependants	Control	Number of dependants	2.35 (1.30)	2.34 (1.30)	2.40 (1.28)
Tenure	Control	Relationship length in months	35.93 (7.99)	35.88 (8.02)	36.18 (7.80)
Credit limit (log)	Control	Natural logarithm of credit limit	8.60 (0.93)	8.62 (0.93)	8.51 (0.96)
Utilisation ratio	Control	Average revolving utilisation (0–1)	0.27 (0.28)	0.30 (0.27)	0.16 (0.26)

Note. Income band (six categories, reference: below $40K) and card tier (four categories, reference: Blue) enter the models as dummy variables. In Dataset B, the active-member flag operationalises activity, the number of products (1, 2, 3–4) operationalises breadth, and tenure is measured in years.

3.4 Analytical procedure

Figure 1 summarises the five-stage procedure, from data acquisition to the translation of results into chatbot strategy.

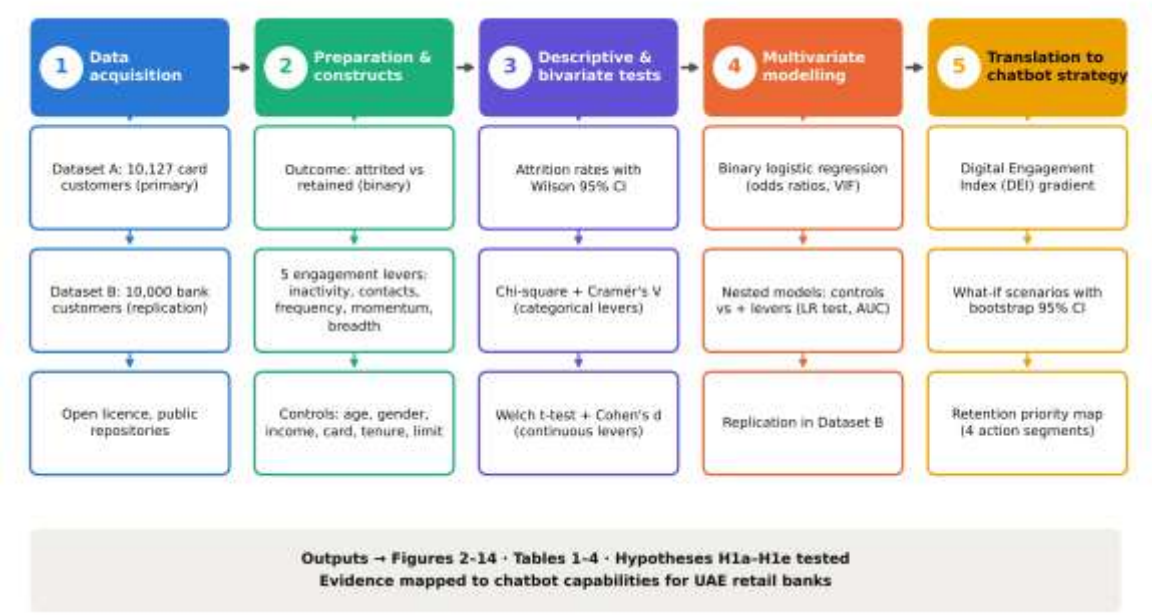


Figure 1. Methodological flowchart: five analytical stages, the techniques used at each stage and the outputs they produce.

The stages were executed in the following order.

1. **Acquisition and integrity checks.** Both datasets were downloaded from public repositories, checked for missing values and duplicates, and stripped of identifiers.
2. **Coding.** The outcome was coded 0/1; categorical controls were converted to dummy variables; credit limit was log-transformed; continuous predictors were standardised (z-scores) so that odds ratios compare a one-standard-deviation change across levers.
3. **Descriptive analysis.** Attrition rates were computed for every level of each lever and for demographic segments, with Wilson 95% confidence intervals, which remain accurate for small cells and extreme proportions.
4. **Bivariate hypothesis tests.** Chi-square tests of independence with Cramér's V were applied to the categorical form of each lever (upper categories pooled at 5+ to avoid sparse cells). Welch t-tests, which do not assume equal variances, compared means of continuous levers, with Cohen's d as the effect size. Mann–Whitney U tests served as a distribution-free robustness check and returned $p < .001$ for all five levers.
5. **Multivariate modelling.** Binary logistic regression estimated Model 1 (controls only) and Model 2 (controls plus the five levers). A likelihood-ratio test, McFadden's pseudo-R^2 and the area under the ROC curve (AUC) measured the incremental contribution of the levers. Variance inflation factors (VIF) checked multicollinearity; transaction amount was excluded because it largely duplicates transaction count.
6. **Replication.** The same nested comparison was repeated in Dataset B with its own controls and its available levers.
7. **Digital Engagement Index.** The five standardised levers were combined into an equally weighted index (signs aligned so that higher values mean more engagement) and attrition was examined across index deciles.
8. **What-if scenarios.** Model 2 was used to estimate how attrition would change under four chatbot interventions and their combination. Uncertainty was estimated with 300 bootstrap resamples, refitting the model each time.
9. **Retention priority map.** Each customer received a profile-risk score from Model 1 and an engagement-risk score from a levers-only model; the portfolio base rate split each axis into four action segments.

3.5 Statistical specification

The Wilson interval for an attrition rate p observed in n customers, with z = 1.96, is:

$$CI = [p + z^2/(2n) \pm z\cdot\sqrt{(p(1-p)/n + z^2/(4n^2))}] \div [1 + z^2/n] \quad (1)$$

Cramér's V for a contingency table with n observations and k = min(rows, columns), and Cohen's d for attrited (1) versus retained (0) customers with pooled standard deviation s_p, are:

$$V = \sqrt{(\chi^2 / [n(k-1)])} \quad (2)$$

$$d = (\bar{x}_1 - \bar{x}_0) / s_p \quad (3)$$

Model 2 expresses the log-odds of attrition for customer i as a function of the five standardised levers L and the controls C; Model 1 omits L:

$$ln[\, p_i / (1-p_i)\,] = \beta_0 + \Sigma_j \beta_j \cdot L_{ij} + \Sigma_k \gamma_k \cdot C_{ik}\,,\ j = 1, \ldots, 5 \quad (4)$$

$$R^2(McFadden) = 1 - ln\,\hat{L}(model) / ln\,\hat{L}(null) \quad (5)$$

The Digital Engagement Index aligns the signs of the standardised levers so that higher values always mean stronger engagement:

$$DEI_i = (1/5) \times (-z(inactive)_i - z(contacts)_i + z(transactions)_i + z(momentum)_i + z(products)_i) \quad (6)$$

The effect of a scenario s is expressed as attritions avoided per 10,000 customers, where x_i is the observed profile and $x_i^{(s)}$ the profile after the intervention:

$$\Delta_s = 10{,}000 \times (1/n) \times \Sigma_i [\, \hat{p}(x_i) - \hat{p}(x_i^{(s)}) \,] \quad (7)$$

3.6 Validity, reliability and ethics

Construct validity is supported by the direct, behavioural nature of each lever: inactivity, contacts and transactions are recorded events rather than self-reports, which removes common-method bias. The DEI is a formative index of distinct behaviours, so internal-consistency statistics are not appropriate; its validity is assessed by its association with the outcome. Multicollinearity was acceptable (maximum VIF = 3.48). External validity is addressed through replication in a second bank, and statistical uncertainty through confidence intervals and bootstrapping with a fixed random seed (42). Both datasets are anonymised and publicly released, so no human-subject approval was required.

All analyses were run in Python 3 (pandas, SciPy, statsmodels, scikit-learn and matplotlib) in Google Colab. The tests are the same procedures available in SPSS, and every table is exported to Microsoft Excel. A single notebook downloads the data and regenerates all fourteen figures and four tables.

4. Results and Discussion

4.1 Sample profile: who the customer is explains little

The primary sample has an attrition rate of 16.07% (1,627 of 10,127). Figure 2 shows that demographic and product segments sit close to this base rate. Women left slightly more often than men (17.4% versus 14.6%; $\chi^2(1) = 13.87$, $p < .001$), but the effect is negligible (Cramér's V = 0.037). Income band was marginally significant ($\chi^2(5) = 12.83$, $p = .025$, V = 0.036), while age band ($p = .305$), card tier ($p = .525$), education ($p = .051$) and marital status ($p = .109$) were not. The wide interval for Platinum cardholders reflects only 20 customers.

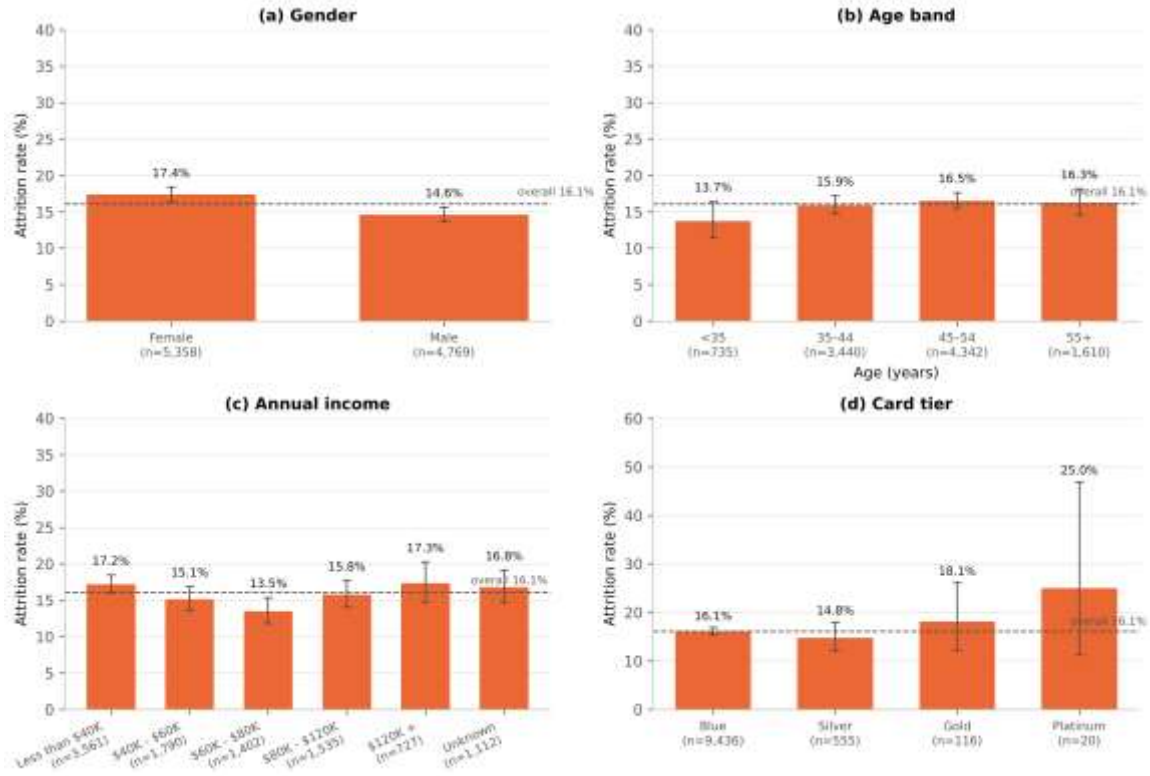


Figure 2. Attrition rate (%) with Wilson 95% confidence intervals by (a) gender, (b) age band, (c) annual income and (d) card tier, Dataset A. The dashed line marks the overall rate of 16.1%.

The practical message is that segmentation by demographics, the default in many retention programmes, would barely separate leavers from stayers in this bank. This sets a low benchmark against which the behavioural levers can be judged.

4.2 Disengagement and service friction raise attrition, and they compound

Figure 3 shows the two levers that a chatbot touches most directly. Attrition climbs steadily with inactivity, from 4.5% after one idle month to 15.4%, 21.5% and 29.9% after two, three and four months. Beyond four months the rate eases to 18.0% (n = 178) and 15.3% (n = 124), a pattern consistent with long-dormant accounts that customers keep open at no cost rather than close. The 29 customers with no inactive month show 51.7% attrition, but the interval is wide (34%–69%) and the group too small for interpretation.

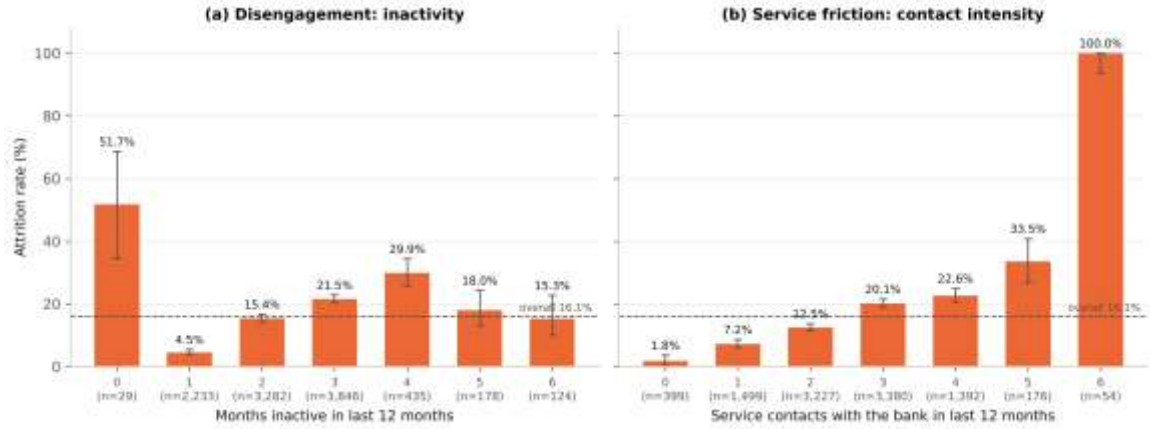


Figure 3. Attrition rate (%) by (a) months inactive and (b) service contacts in the last 12 months, Dataset A, with Wilson 95% confidence intervals and group sizes.

The contact gradient is strictly monotonic. Customers who did not contact the bank left at 1.8%; the rate rises to 7.2%, 12.5%, 20.1%, 22.6% and 33.5% for one to five contacts, and every one of the 54 customers with six contacts left. The rank correlation with attrition is slightly stronger for contacts (Spearman $\rho = 0.189$) than for inactivity ($\rho = 0.172$),

both p < .001. The six-contact group should be read with care: part of that contact volume is probably the exit process itself. The broader gradient, however, starts long before that extreme and supports reading contacts as a marker of unresolved service friction.

Figure 4 crosses the two levers in a Friction–Disengagement Matrix. The levers compound rather than substitute. Customers with at least three inactive months and at least three contacts left at 29.6% (n = 2,317), more than four times the 6.9% observed among customers with at most two of each (n = 2,859). Individual cells reach 42% (three idle months with five contacts, n = 73; four idle months with three contacts, n = 161), whereas recently active customers with no contact show no attrition at all (n = 100).

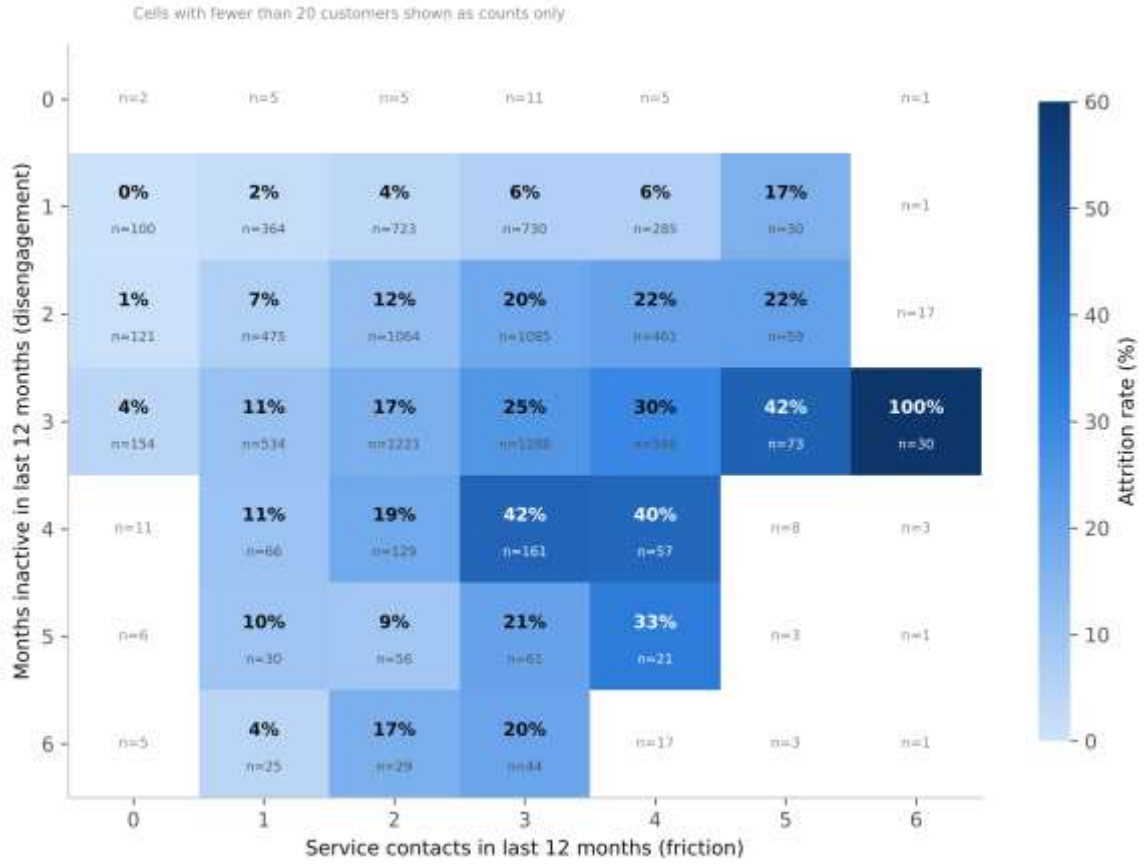


Figure 4. Friction–Disengagement Matrix: attrition rate (%) for each combination of months inactive and service contacts, Dataset A. Cells with fewer than 20 customers show counts only.

For a bank, the matrix defines a trigger rule a chatbot can execute: a customer who crosses into the lower-right region, with repeated contacts and growing inactivity, should receive a proactive, resolution-focused conversation rather than a generic marketing message.

4.3 Usage frequency, momentum and breadth protect the relationship

Figure 5 compares the distributions of transaction frequency and momentum. Attriters made a median of 43 transactions in twelve months against 71 for retained customers, and their distribution is concentrated in the 30–60 range. Momentum separates the groups as clearly: attriters' fourth-quarter transaction count was a median 0.53 of their first-quarter count, versus 0.72 for retained customers. In other words, leavers had largely stopped using the card before they closed the account, which is precisely the window in which an engagement intervention can act.

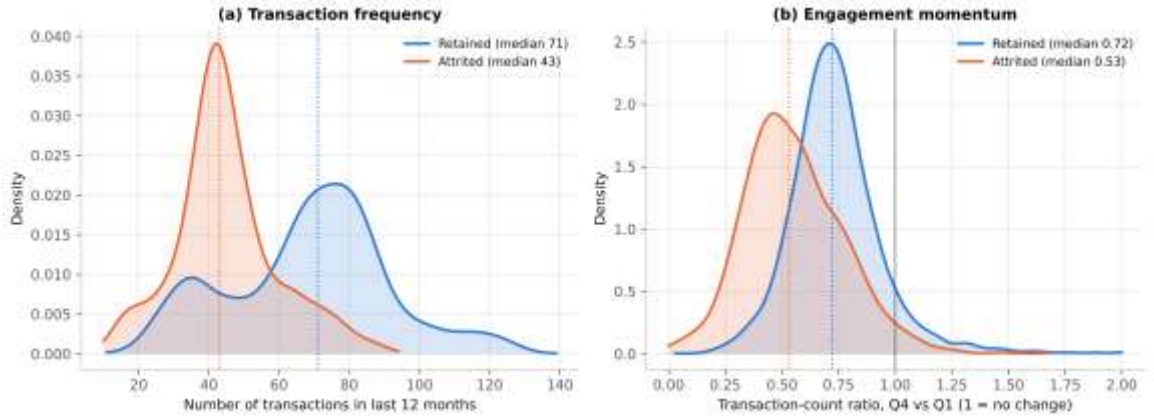


Figure 5. Kernel density of (a) annual transaction count and (b) the Q4/Q1 transaction-count ratio for retained and attrited customers, Dataset A. Dotted lines mark group medians.

Relationship breadth shows a threshold rather than a straight line (Figure 6). Customers with one or two products left at 25.6% and 27.8%, those with three at 17.4%, and those with four to six at 10.5%–12.0%. The protective effect of breadth therefore comes mainly from moving customers from two to four products, after which it levels off.

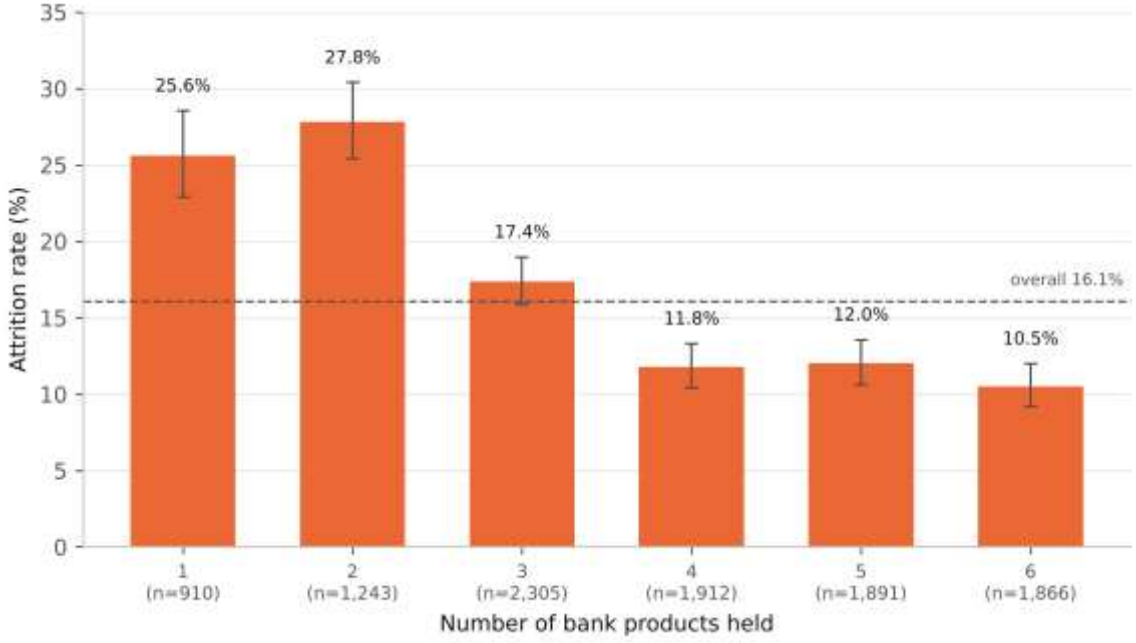


Figure 6. Attrition rate (%) by number of products held, Dataset A, with Wilson 95% confidence intervals.

4.4 Bivariate hypothesis tests

Table 2 reports the formal tests. Every lever in Dataset A is significant at $p < .001$. The categorical associations for contacts ($V = 0.211$), inactivity ($V = 0.198$) and breadth ($V = 0.167$) are five to ten times larger than those of the demographic variables in Section 4.1.

Table 2. Bivariate hypothesis tests

Dataset	Hypothesis	Indicator	Test	Statistic	p-value	Effect size	Result
A	H1a	Months inactive (0–5+)	Chi-square	$\chi^2(5) = 396.07$	< .001	$V = 0.198$	Significant
A	H1b	Service contacts (0–5+)	Chi-square	$\chi^2(5) = 451.22$	< .001	$V = 0.211$	Significant

Dataset	Hypothesis	Indicator	Test	Statistic	p-value	Effect size	Result
A	H1d	Products held (1–6)	Chi-square	$\chi^2(5) = 284.07$	< .001	V = 0.167	Significant
A	H1a	Months inactive	Welch t-test	t = 16.86	< .001	d = 0.420	Significant
A	H1b	Service contacts	Welch t-test	t = 20.90	< .001	d = 0.569	Significant
A	H1c	Transaction count	Welch t-test	t = −54.14	< .001	d = −1.089	Significant
A	H1c	Transaction momentum (Q4/Q1)	Welch t-test	t = −30.61	< .001	d = −0.825	Significant
A	H1d	Products held	Welch t-test	t = −14.94	< .001	d = −0.413	Significant
B	H1e	Active membership (yes/no)	Chi-square	$\chi^2(1) = 242.99$	< .001	V = 0.156	Significant
B	H1e	Products held (1–4)	Chi-square	$\chi^2(3) = 1,503.63$	< .001	V = 0.388	Significant
B	H1e	Tenure (years)	Welch t-test	t = −1.38	.166	d = −0.035	Not significant

Figure 7 ranks all continuous attributes by Cohen's d. Transaction frequency (d = −1.09) and momentum (d = −0.83) show large effects, service contacts a medium effect (d = 0.57), and inactivity and breadth small-to-medium effects (0.42 and −0.41). Age, tenure and dependants are close to zero ($|d| \leq 0.05$). Revolving balance and utilisation also separate the groups, but they describe credit behaviour that a service chatbot cannot directly change, so they are kept as controls.

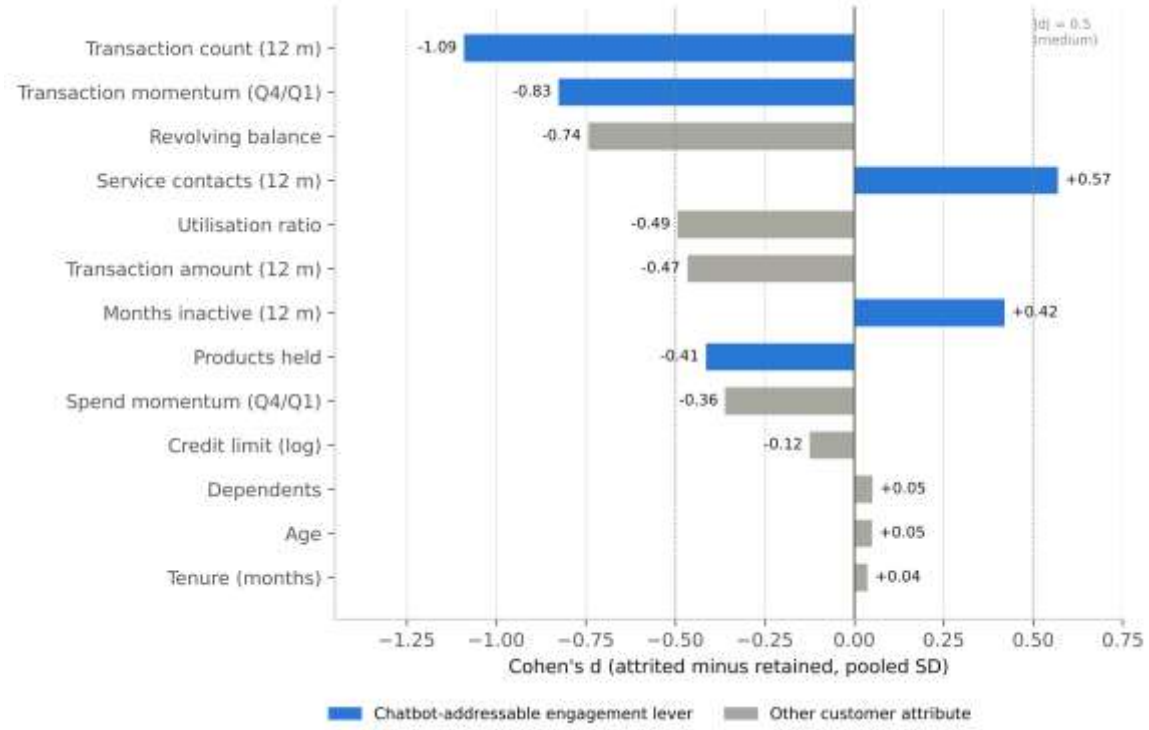


Figure 7. Standardised mean differences (Cohen's d, attrited minus retained) for continuous attributes, Dataset A. Blue bars are chatbot-addressable levers; dotted lines mark |d| = 0.5.

4.5 Multivariate evidence: the levers carry most of the explainable risk

Table 3 reports the logistic regression results. Holding all controls constant, a one-standard-deviation increase in service contacts raised the odds of attrition by 69% (OR = 1.685, 95% CI 1.566–1.813) and in inactivity by 61% (OR = 1.613). The protective levers were stronger still: one standard deviation more transactions, about 23 a year, cut the odds by 77% (OR = 0.228), higher momentum by 48% (OR = 0.518), and broader relationships by 56% (OR = 0.440). All variance inflation factors were below 3.5.

Table 3. Binary logistic regression of attrition (Model 2, full specification)

Dataset	Predictor	Role	Odds ratio [95% CI]	p-value	VIF
A	Months inactive (per SD)	Lever	1.613 [1.504, 1.731]	< .001	1.01
A	Service contacts (per SD)	Lever	1.685 [1.566, 1.813]	< .001	1.04
A	Transaction count (per SD)	Lever	0.228 [0.207, 0.251]	< .001	1.14
A	Transaction momentum (per SD)	Lever	0.518 [0.479, 0.561]	< .001	1.03
A	Products held (per SD)	Lever	0.440 [0.407, 0.477]	< .001	1.08
A	Age (per SD)	Control	0.990 [0.884, 1.109]	.861	2.68
A	Male	Control	0.491 [0.374, 0.645]	< .001	3.48
A	Dependants (per SD)	Control	1.151 [1.072, 1.237]	< .001	1.04
A	Tenure (per SD)	Control	0.913 [0.816, 1.023]	.116	2.66
A	Credit limit, log (per SD)	Control	0.618 [0.559, 0.684]	< .001	2.36

Dataset	Predictor	Role	Odds ratio [95% CI]	p-value	VIF
A	Utilisation ratio (per SD)	Control	0.463 [0.426, 0.503]	< .001	1.43
A	Income $40K–$60K	Control	0.860 [0.685, 1.080]	.195	1.50
A	Income $60K–$80K	Control	1.049 [0.741, 1.485]	.786	2.72
A	Income $80K–$120K	Control	1.398 [0.988, 1.979]	.059	3.04
A	Income $120K+	Control	1.756 [1.181, 2.610]	.005	2.15
A	Income unknown	Control	0.881 [0.686, 1.132]	.323	1.29
A	Card: Silver	Control	2.006 [1.428, 2.819]	< .001	1.24
A	Card: Gold	Control	3.526 [1.863, 6.673]	< .001	1.06
A	Card: Platinum	Control	4.803 [1.292, 17.851]	.019	1.01
B	Active member	Lever	0.332 [0.294, 0.375]	< .001	–
B	Holds 2 products (vs 1)	Lever	0.207 [0.180, 0.239]	< .001	–
B	Holds 3–4 products (vs 1)	Lever	16.145 [11.464, 22.737]	< .001	–
B	Tenure (per SD)	Lever	0.947 [0.894, 1.002]	.060	–
B	Age (per SD)	Control	2.113 [1.997, 2.237]	< .001	–
B	Male	Control	0.592 [0.528, 0.665]	< .001	–
B	Germany (vs France)	Control	2.678 [2.313, 3.100]	< .001	–
B	Spain (vs France)	Control	1.054 [0.908, 1.223]	.491	–
B	Credit score (per SD)	Control	0.935 [0.883, 0.991]	.023	–
B	Balance, log (per SD)	Control	0.920 [0.855, 0.990]	.027	–
B	Has credit card	Control	0.935 [0.825, 1.061]	.298	–
B	Estimated salary (per SD)	Control	1.026 [0.968, 1.087]	.382	–
A	Model fit		McFadden R^2 0.070 → 0.405; AUC 0.693 → 0.899; LR $\chi^2(5)$ = 2,987.0, p < .001		
B	Model fit		McFadden R^2 0.115 → 0.264; AUC 0.752 → 0.834; LR $\chi^2(4)$ = 1,501.3, p < .001		

Note. Reference categories: income below $40K, Blue card, France, one product. Model fit rows compare Model 1 (controls only) with Model 2 (controls plus levers).

Figure 8 displays the same odds ratios. The levers cluster far from the line of no effect with narrow intervals, while most controls sit near it or carry wide intervals. Among the controls, men left less often than women (OR = 0.491), a higher credit limit and higher utilisation were protective, and premium card tiers carried higher odds once usage was held constant, suggesting that premium customers who stop using their card are especially likely to leave. Age and tenure had no independent effect.

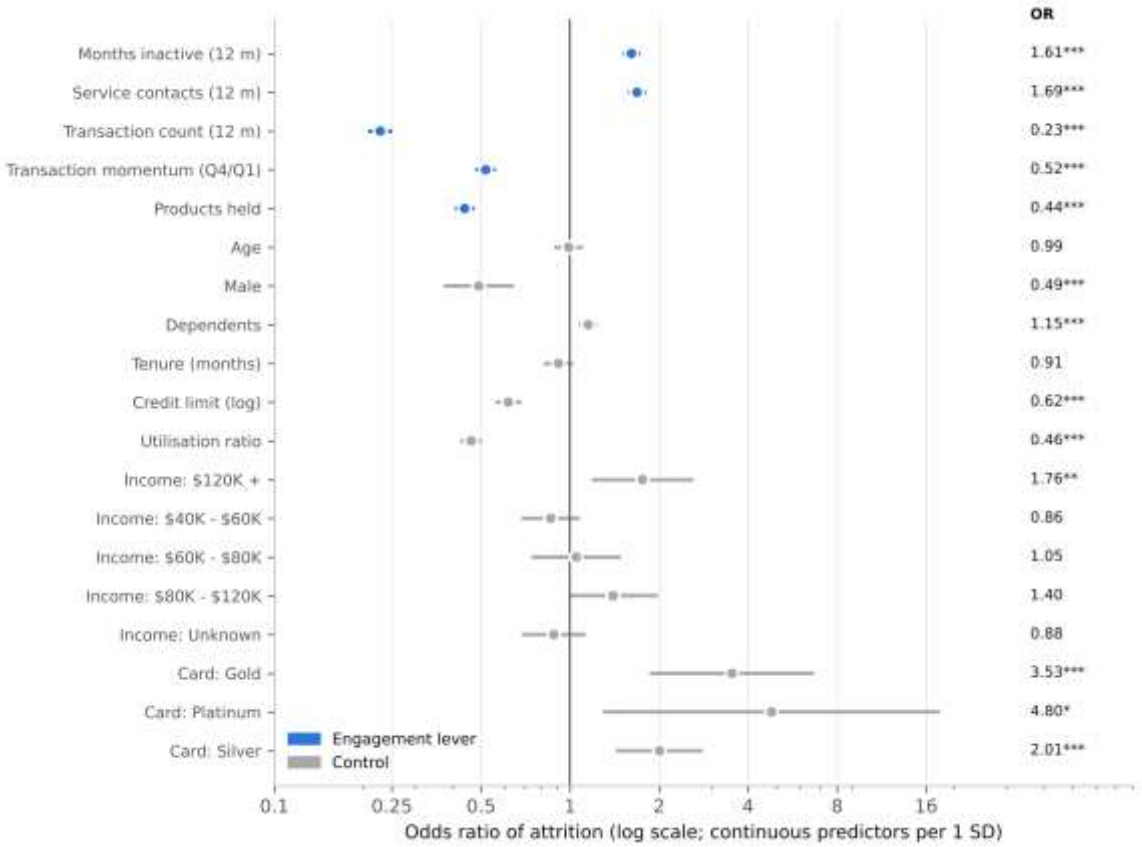


Figure 8. Odds ratios with 95% confidence intervals from Model 2, Dataset A (log scale; continuous predictors per one standard deviation). *** p < .001, ** p < .01, * p < .05.

Figure 9 converts the model into predicted probabilities for a reference customer. With one idle month, risk rises from 1.3% with no contacts to 18.6% with six; with four idle months, from 5.3% to 48.6%. At three contacts, moving from one to four idle months raises risk from 5.3% to 18.7%. Because the logistic link is multiplicative on the odds scale, the two forms of friction amplify each other in probability terms, confirming the pattern of Figure 4 after adjustment for all controls.

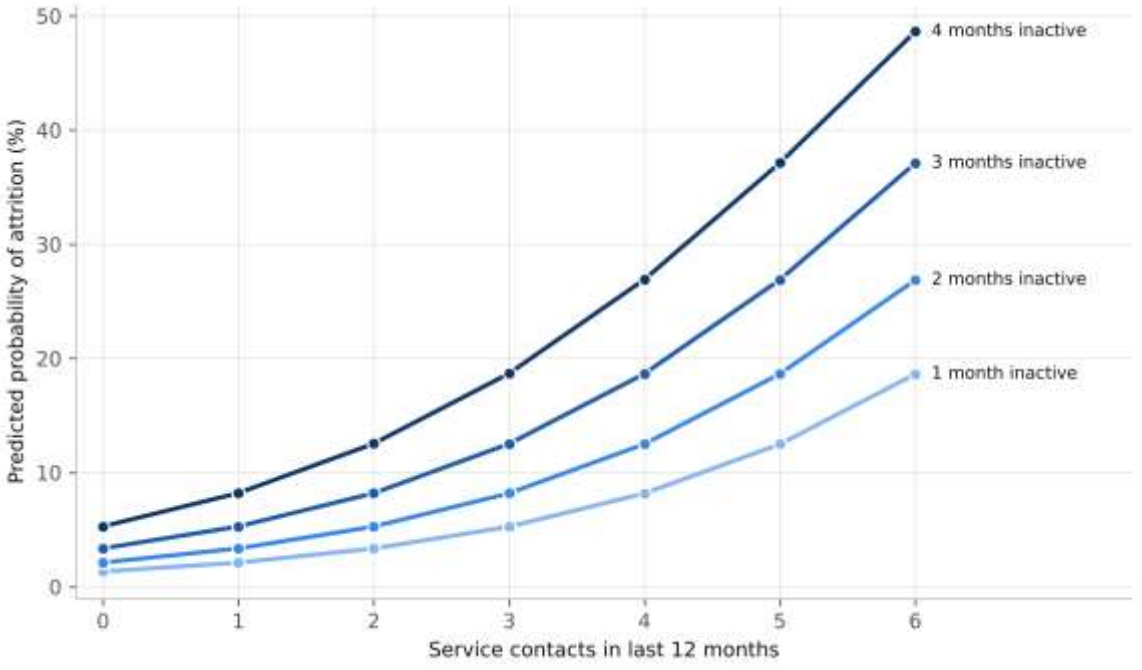


Figure 9. Predicted probability of attrition by number of service contacts for four levels of inactivity (Model 2; other predictors at sample means and reference categories).

The nested comparison in Figure 10 is the study's central test of incremental value. Adding the five levers raised McFadden's pseudo-R^2 from 0.070 to 0.405, a near six-fold gain, and the AUC from 0.693 to 0.899 (LR $\chi^2(5)$ = 2,987.0, $p < .001$). In the replication bank, adding activity, breadth and tenure raised pseudo-R^2 from 0.115 to 0.264 and the AUC from 0.752 to 0.834 (LR $\chi^2(4)$ = 1,501.3, $p < .001$). In both banks, most of the explainable attrition risk is behavioural rather than demographic.

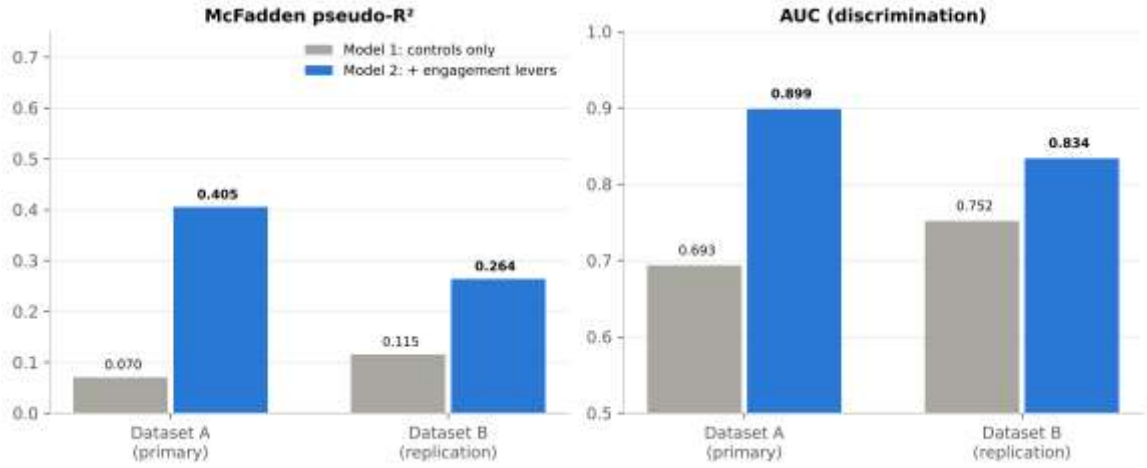


Figure 10. McFadden pseudo-R^2 and AUC for Model 1 (controls only) and Model 2 (controls plus engagement levers) in both datasets.

4.6 A transparent Digital Engagement Index

Banks often hesitate to act on scores they cannot explain to staff or regulators. The DEI addresses this by averaging the five standardised levers with equal weights, so every customer's score can be traced to five observable behaviours. Figure 11 shows a steep gradient: attrition falls from 68.1% in the least-engaged decile to 33.1% and 22.0% in the next two, and to 1.1% in the most-engaged decile, a 63-fold difference (Spearman $\rho = -0.459$, $p < .001$). Used alone, the index reaches an AUC of 0.861, close to the full regression model without any estimated weights. The three lowest deciles hold 30% of customers but 76.7% of all attriters, which gives a bank a simple, explainable shortlist for chatbot outreach.

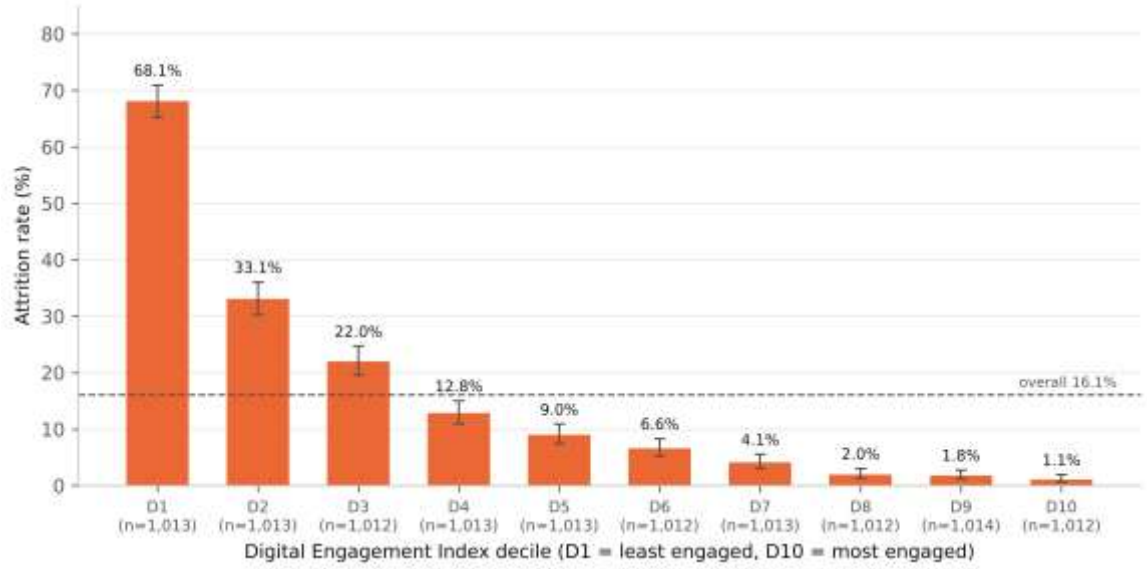


Figure 11. Attrition rate (%) by Digital Engagement Index decile, Dataset A (D1 = least engaged), with Wilson 95% confidence intervals.

4.7 Replication in an independent bank

Figure 12 tests whether the engagement–retention link holds in Dataset B. Inactive members churned at 26.9% against 14.3% for active members, and activity remained strongly protective after controls (OR = 0.332, 95% CI 0.294–

0.375). The activity gap appears at every product level and in every age band and widens with age: among customers aged 55 or over, 88% of inactive members left against 20% of active members, and among those aged 45–54, 60% against 35%.

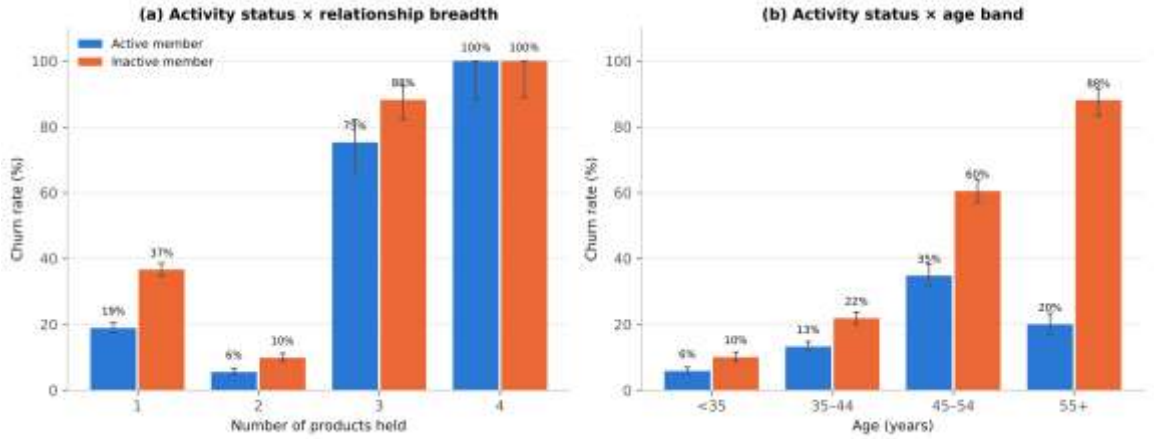


Figure 12. Churn rate (%) by activity status and (a) number of products held or (b) age band, Dataset B (n = 10,000), with Wilson 95% confidence intervals.

Dataset B also qualifies the breadth hypothesis. Moving from one to two products cut churn from 27.7% to 7.6% (OR = 0.207), but customers with three products (n = 266) churned at 82.7% and all 60 customers with four products left (OR = 16.1 for three or four products). Holding many products is therefore not protective in itself; a likely explanation is that some of these products were sold rather than needed, so they added cost or complexity instead of value. Tenure was not significant in either bank (p = .116 in A, p = .060 in B), which suggests that how long a customer has banked matters less than how actively they bank now.

4.8 What-if scenarios for a chatbot programme

Figure 13 translates Model 2 into four interventions that map onto chatbot capabilities, with 95% bootstrap intervals from 300 resamples. Reactivation nudges that shorten inactivity by one month for customers idle three months or more imply 189 avoided attritions per 10,000 customers (CI 164–213). First-contact resolution that removes one contact for customers with three or more implies 213 (187–243). Usage prompts that lift transactions by 10% imply 230 (218–242), and needs-based guidance that adds one product for customers with two or fewer implies 81 (69–91). The combined programme implies 636 avoided attritions per 10,000 customers (601–672), about 40% of the 1,607 attritions expected per 10,000 at the observed base rate.

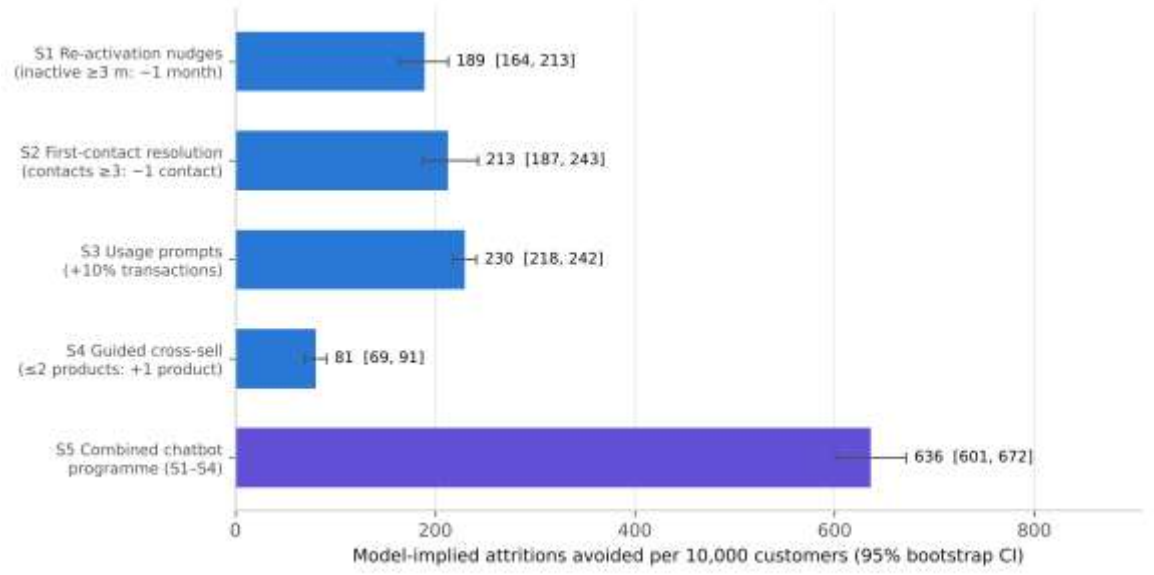


Figure 13. Model-implied attritions avoided per 10,000 customers under five chatbot scenarios, with 95% bootstrap confidence intervals (300 resamples), Dataset A.

These figures are upper-bound planning estimates rather than causal forecasts. They assume that a chatbot can shift behaviour by the stated amount and that the observed associations hold under intervention. The contact scenario in particular is valid only if contacts fall because problems are resolved; merely making the bank harder to reach would reduce contacts while increasing friction. The cross-selling scenario must also respect the ceiling shown in Dataset B.

4.9 Chatbot Retention Priority Map

The final step separates who should act. Each customer receives a profile-risk score, predicted from demographics, card tier and credit profile, and an engagement-risk score, predicted from the five levers alone (AUC = 0.880, against 0.693 for profile risk). Splitting both axes at the base rate yields four segments (Figure 14).

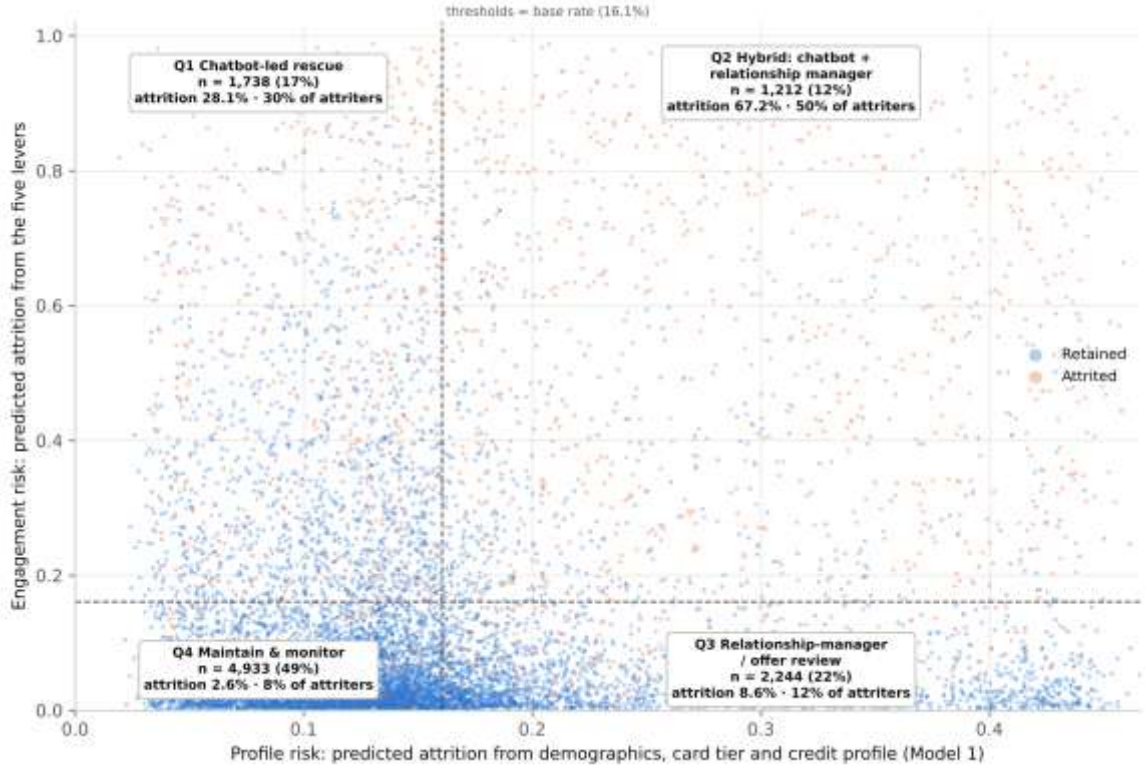


Figure 14. Chatbot Retention Priority Map: profile risk (Model 1) against engagement risk (levers-only model), Dataset A. Dashed lines mark the base rate; labels give segment size, observed attrition and share of all attriters.

Segment Q1, *chatbot-led rescue*, combines low profile risk with high engagement risk: 1,738 customers (17%) with 28.1% attrition, holding 30% of all attriters. Their risk is behavioural and conversational interventions are the natural first response. Segment Q2, *hybrid chatbot plus relationship manager*, combines both risks: 1,212 customers (12%) with 67.2% attrition and 50% of all attriters. Here the chatbot should detect and triage, while a human handles the conversation. Segment Q3, *relationship-manager or offer review*, carries profile risk with healthy engagement: 2,244 customers (22%), 8.6% attrition and 12% of attriters, for whom pricing or product fit matters more than service experience. Segment Q4, *maintain and monitor*, holds 4,933 customers (49%) with 2.6% attrition and 8% of attriters. The two engagement-risk segments therefore contain 29% of customers and 80% of attriters, the clearest single argument for engagement-centred chatbot design.

4.10 Hypothesis decisions and discussion

Table 4 summarises the decisions. All five hypotheses are supported, with the breadth hypothesis qualified by a ceiling.

Table 4. Summary of hypothesis decisions and implied chatbot capabilities

Hypothesis	Statement	Key evidence	Decision	Chatbot capability implied
H1a	Inactivity raises attrition	4.5% at 1 idle month vs 29.9% at 4; OR = 1.61 per SD	Supported	Proactive reactivation nudges after two idle months
H1b	Service friction raises attrition	7.2% at 1 contact vs 33.5% at 5; OR = 1.69 per SD	Supported	First-contact resolution, case tracking, warm hand-off
H1c	Usage frequency and momentum lower attrition	OR = 0.23 (frequency) and 0.52 (momentum) per SD	Supported	Usage prompts, spending insights, bill reminders
H1d	Relationship breadth lowers attrition	A: OR = 0.44 per SD; 25.6% (1 product) vs 10.5% (6). B: OR = 0.21 for 2 products, 16.1 for 3–4	Supported, with a ceiling	Needs-based, consent-driven guidance; no push selling
H1e	The link replicates in another bank	Churn 14.3% active vs 26.9% inactive; OR = 0.33; AUC 0.752 → 0.834	Supported	Activity status as a universal chatbot trigger

The findings extend the survey literature in three ways. First, they confirm in behavioural data what satisfaction and continuance studies infer from attitudes [10–15]: the experience a customer has with the bank's service channels is tied to whether the customer stays. The size of the effect is notable; behaviour explained several times more variance than demographics in both banks. Second, they show that the strongest chatbot-relevant risk signals are friction and disengagement rather than the customer's profile, which supports the argument that retention effort should target addressable rather than merely high risk [7]. Third, they show that friction and disengagement compound, so a chatbot that only answers questions without resolving them, or that waits for the customer to return, will miss the customers who matter most.

The results also explain why the barrier literature matters for retention. If repeated contact is a leading signal of exit, then a chatbot that fails to understand a problem or that hides the route to a human adds a contact without resolving anything. The importance customers place on responsiveness, expertise and security [16], the role of trust in disclosure [17] and the penalty that can follow a poorly handled bot disclosure [18] all point in the same direction: a retention chatbot must be judged by resolution, not containment. Finally, the interpretability of the logistic model and the DEI responds to the call for churn analytics that managers can act on [19], and the strong role of quarter-on-quarter momentum confirms the value of time-varying behavioural signals [20].

5. Conclusion

5.1 Summary of findings

This study asked which customer behaviours a banking chatbot can influence, how much they matter for retention, and how a bank should act on them. In two independent retail banks, behaviour explained far more attrition than demographics. Service friction and inactivity raised the odds of leaving by 69% and 61% per standard deviation and amplified each other; usage frequency, momentum and moderate relationship breadth protected the relationship. The five levers raised explained variance almost six-fold in the primary bank and more than doubled it in the replication bank. A transparent index of the same levers placed 77% of attriters in the least-engaged 30% of customers, and a

combined chatbot programme implied about 40% fewer attritions under stated assumptions. All five sub-hypotheses of the programme's retention hypothesis were supported, with breadth showing a ceiling.

5.2 Theoretical contributions

The study moves the chatbot–retention discussion from intention to behaviour. It shows that the mechanisms emphasised by satisfaction and continuance models, namely competent, responsive and trustworthy service, correspond to measurable behaviours that precede exit. It also introduces the distinction between addressable and non-addressable attrition risk into chatbot research, operationalised through nested models, a formative engagement index and a two-axis priority map. Finally, the finding that friction and disengagement compound refines engagement theory for automated service: an automated channel that answers without resolving can deepen disengagement rather than reverse it.

5.3 Managerial implications for UAE banks

The results suggest a staged roadmap that UAE retail banks can calibrate on their own records.

1. **Instrument the levers.** Record inactivity, contacts by reason and channel, resolution status, transaction momentum and product holdings for every customer; compute the DEI monthly; apply the Friction–Disengagement trigger (three or more idle months together with three or more contacts).
2. **Build a resolution-first chatbot.** Measure the chatbot by first-contact resolution and repeat-contact rate rather than by containment. Support Arabic, English and the code-mixed messages common among the UAE's multilingual customers, disclose the bot clearly, and offer a one-tap route to a human with the conversation history attached.
3. **Add proactive, consent-based engagement.** Use reactivation nudges after two idle months, spending insights and bill reminders to sustain momentum, and needs-based product guidance that stops short of pushing additional products. Personalisation should operate within UAE data-protection and Central Bank consumer-protection requirements.
4. **Orchestrate with people.** Route Q1 customers to chatbot-led rescue journeys, Q2 customers to a joint chatbot-and-relationship-manager journey, and Q3 customers to pricing or product review. Test every intervention with randomised control groups so that budgets follow measured uplift rather than predicted risk.

5.4 Limitations and future research

Four limitations frame the conclusions. First, neither dataset comes from the UAE, so the effect sizes are benchmarks to be re-estimated on UAE data; the planned stratified survey of 200 customers and 100 banking professionals will test perceptions, satisfaction and implementation challenges that behavioural records cannot capture. Second, the design is cross-sectional and associational: the scenarios estimate potential, not proven effects, and randomised field experiments are needed to establish causal uplift. Third, the datasets contain no direct record of chatbot use, so the levers represent behaviours a chatbot can address rather than chatbot exposure itself; future work should link conversation logs to account outcomes. Fourth, the contact variable records neither channel nor reason, and at its extreme it may partly capture the exit process. Future studies should also examine how multilingual and code-mixed conversations, common in the Gulf, affect resolution and retention.

Declarations

Data availability. Both datasets are publicly available. Dataset A (Credit Card Customers, file BankChurners.csv) is archived on Zenodo, record 4322342; the analysis notebook reads a GitHub mirror of the same records. Dataset B (Bank Customer Churn, file Churn_Modelling.csv) is read from a public GitHub copy.

Funding. None.

Competing interests. None declared.

References

[1] Huang, Ming-Hui, and Roland T. Rust. "Artificial Intelligence in Service." *Journal of Service Research*, vol. 21, no. 2, 2018, pp. 155–172.
[2] Verhoef, Peter C., et al. "Digital Transformation: A Multidisciplinary Reflection and Research Agenda." *Journal of Business Research*, vol. 122, 2021, pp. 889–901.
[3] Adamopoulou, Eleni, and Lefteris Moussiades. "Chatbots: History, Technology, and Applications." *Machine Learning with Applications*, vol. 2, 2020, article 100006.
[4] Rahman, Mahfuzur, et al. "Adoption of Artificial Intelligence in Banking Services: An Empirical Analysis." *International Journal of Emerging Markets*, vol. 18, no. 10, 2023, pp. 4270–4300.
[5] Qureshi, Omer Aftab, et al. "Chatbots: Can They Satisfy Customers in the Banking Sector?" *BUiD Doctoral Research Conference 2023*, edited by Khalid Al Marri et al., Lecture Notes in Civil Engineering, vol. 473, Springer, 2024, pp. 360–376.
[6] Jenneboer, Liss, et al. "The Impact of Chatbots on Customer Loyalty: A Systematic Literature Review." *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 17, no. 1, 2022, pp. 212–229.
[7] Ascarza, Eva. "Retention Futility: Targeting High-Risk Customers Might Be Ineffective." *Journal of Marketing Research*, vol. 55, no. 1, 2018, pp. 80–98.
[8] Hollebeek, Linda D., et al. "Rise of the Machines? Customer Engagement in Automated Service Interactions." *Journal of Service Research*, vol. 24, no. 1, 2021, pp. 3–8.
[9] Kumar, V., et al. "Understanding the Role of Artificial Intelligence in Personalized Engagement Marketing." *California Management Review*, vol. 61, no. 4, 2019, pp. 135–155.
[10] Eren, Berrin Arzu. "Determinants of Customer Satisfaction in Chatbot Use: Evidence from a Banking Application in Turkey." *International Journal of Bank Marketing*, vol. 39, no. 2, 2021, pp. 294–311.
[11] Nguyen, Dung Minh, et al. "Determinants of Continuance Intention towards Banks' Chatbot Services in Vietnam: A Necessity for Sustainable Development." *Sustainability*, vol. 13, no. 14, 2021, article 7625.
[12] Ashfaq, Muhammad, et al. "I, Chatbot: Modeling the Determinants of Users' Satisfaction and Continuance Intention of AI-Powered Service Agents." *Telematics and Informatics*, vol. 54, 2020, article 101473.
[13] Hsu, Chin-Lung, and Judy Chuan-Chuan Lin. "Understanding the User Satisfaction and Loyalty of Customer Service Chatbots." *Journal of Retailing and Consumer Services*, vol. 71, 2023, article 103211.
[14] Alnaser, Feras Mi, et al. "Does Artificial Intelligence (AI) Boost Digital Banking User Satisfaction? Integration of Expectation Confirmation Model and Antecedents of Artificial Intelligence Enabled Digital Banking." *Heliyon*, vol. 9, no. 8, 2023, article e18930.
[15] Alt, Mónika-Anetta, et al. "Banking with a Chatbot – A Study on Technology Acceptance." *Studia Universitatis Babeș-Bolyai Oeconomica*, vol. 66, no. 1, 2021, pp. 13–35.
[16] Mogaji, Emmanuel, et al. "Emerging-Market Consumers' Interactions with Banking Chatbots." *Telematics and Informatics*, vol. 65, 2021, article 101711.
[17] Lappeman, James, et al. "Trust and Digital Privacy: Willingness to Disclose Personal Information to Banking Chatbot Services." *Journal of Financial Services Marketing*, vol. 28, no. 2, 2023, pp. 337–357.
[18] Luo, Xueming, et al. "Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases." *Marketing Science*, vol. 38, no. 6, 2019, pp. 937–947.
[19] De Caigny, Arno, et al. "A New Hybrid Classification Algorithm for Customer Churn Prediction Based on Logistic Regression and Decision Trees." *European Journal of Operational Research*, vol. 269, no. 2, 2018, pp. 760–772.
[20] Gattermann-Itschert, Theresa, and Ulrich W. Thonemann. "How Training on Multiple Time Slices Improves Performance in Churn Prediction." *European Journal of Operational Research*, vol. 295, no. 2, 2021, pp. 664–674.