
Case Study Library

Twelve product case studies. Company cases use only public, well-known facts; all numbers are labelled hypothetical unless sourced.



AUTHOR

Varun Kumar

Product Management Master Handbook

v0.1 · 23 October 2026

Built from the 55-lecture course study guide and extended to the 12-module Product Core curriculum. Hypothetical examples are labelled; no real company results are invented.

Case Study Library

Twelve product case studies. Company cases use only public, well-known facts; all numbers are labelled hypothetical unless sourced.

12 entries

Case 1 · Food delivery (Swiggy/Zomato-style): scheduled orders and checkout conversion

Hypothetical scenario

Swiggy and Zomato are large Indian food-delivery platforms; this case uses that *type* of product. The problem framing, evidence and every number below are **invented for teaching** — they are not data about either company.

Business background. A food-delivery marketplace sees demand spikes at lunch and dinner. Peak spikes strain courier supply and raise delivery times; off-peak capacity sits idle.

Target users. Office workers ordering lunch; families ordering dinner; restaurants; delivery partners.

Customer problem. Customers who know *when* they want food (e.g., 1 pm lunch) still order at the last minute and face long ETAs at peak; many abandon at checkout when fees and ETA appear.

Business objective. Raise completed orders and on-time rate at peak while improving checkout conversion.

Available evidence (hypothetical).

Signal	Value
Checkout completion (peak)	64% vs 72% off-peak
Share of abandoners who viewed ETA > 40 min	38%
Interviews (10 office workers)	7 plan lunch before 11 am; 6 said they'd schedule if reliable
Existing scheduled-order share	3% of orders (feature hidden in a menu)

Assumptions to test. (1) Customers will schedule if the slot is guaranteed; (2) restaurants can prepare to a slot; (3) smoothing demand improves on-time rate; (4) showing fee + ETA earlier reduces abandonment.

Product hypothesis. *Making scheduling visible in the cart for peak-hour orders, with a guaranteed slot, will raise peak checkout completion by ≥ 3 points and on-time delivery by ≥ 5 points, without reducing order volume.*



Candidate solutions. A) Surface scheduling in cart at peak; B) Early fee + ETA display; C) Peak pricing to shift demand; D) Office “group lunch” scheduled orders; E) Restaurant prep-time integration.

Prioritisation (RICE, hypothetical). B: $60,000 \times 2 \times 0.8 \div 2 = 48,000$; A: $40,000 \times 2 \times 0.5 \div 3 \approx 13,333$; D: $8,000 \times 3 \times 0.5 \div 5 = 2,400$; C excluded as first move (trust risk). Start with **B**, run **A** as a pilot in one city.

MVP. Early fee/ETA display (A/B) + scheduling banner in cart for orders started 11:00–12:30 in one city with 50 partner restaurants.

KPIs. Primary: peak checkout completion. Secondary: on-time rate at peak, scheduled-order share. Guardrails: total orders per user, cancellations, restaurant acceptance rate, complaint rate.

Risks and constraints. Missed scheduled slots destroy trust; restaurant capacity; courier allocation logic; cannibalisation (no net gain, just shifting).

Experiment plan. 1) A/B test early fee/ETA (Module 7 sizing). 2) City pilot for scheduling with holdout zones; 4 weeks; daily ops review.

Example metrics (hypothetical readout). Early fee/ETA: completion 64.0% $\rightarrow$ 66.4% (+2.4 pts, CI +1.1 to +3.7). Scheduling pilot: scheduled share 3% $\rightarrow$ 11%; on-time at peak 78% $\rightarrow$ 83%; total orders per user unchanged (no cannibalisation detected).

Recommendation. Ship early fee/ETA; expand scheduling to two more cities with restaurant prep-time integration first.

Potential outcomes. Best: smoother peaks, lower costs, happier customers. Worst: unreliable slots $\rightarrow$ complaints $\rightarrow$ roll back the guarantee.

Key PM lessons. Fix the obvious friction (fees, ETA) before building new behaviour; operational promises need operational proof; measure cannibalisation, not just adoption.

Case 2 · Ride-hailing (Uber/Ola-style): cancellation rates and marketplace dynamics

Hypothetical scenario

Uber and Ola are ride-hailing platforms operating in India. All framing and numbers here are **invented**; they do not describe either company's data or systems. (Your course's surge-pricing session, Lecture 15, is summarised in Module 8.)

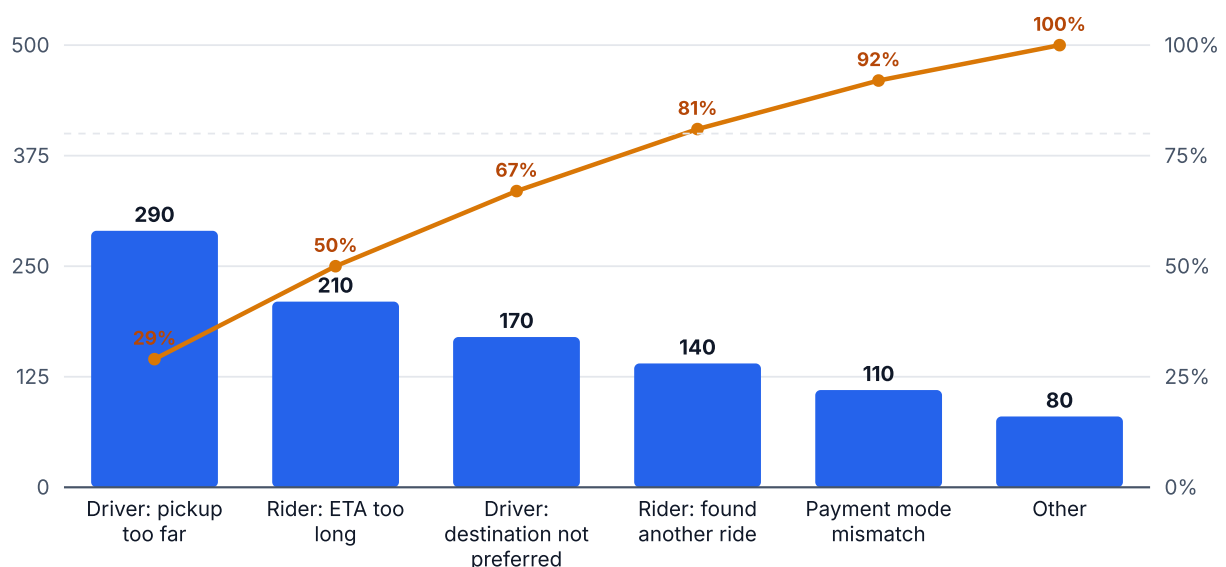
Business background. A two-sided marketplace matches riders and drivers. Cancellations waste time on both sides and damage trust.

Target users. Riders (commuters, airport trips); drivers (full-time and part-time).

Customer problem. Riders wait, then the driver cancels; drivers accept, then riders cancel or are far away.

Business objective. Reduce total cancellation rate while keeping completed trips and driver earnings stable.

Available evidence (hypothetical).



Cancellation reasons (hypothetical sample of 1,000 cancellations). Hypothetical. Driver-side reasons (pickup distance, destination) account for nearly half.

Assumptions. Showing destination before acceptance changes driver behaviour; better pickup-distance matching reduces driver cancellations; honest ETAs reduce rider cancellations.

Product hypothesis. *Capping pickup distance per trip type and showing a reliable ETA with a confidence band will cut cancellations by $\geq 15\%$ without lowering completed trips.*

User journey. Request → match → accept → drive to pickup → trip → pay → rate. Cancellations cluster at accept→pickup.

Candidate solutions. A) Smarter matching (distance caps, batched matching); B) Destination preview for drivers (with fairness safeguards); C) Cancellation fees/penalties; D) Rider ETA confidence bands; E) Incentives for long pickups.

Prioritisation. Weighted scoring (impact 40%, trust 25%, feasibility 20%, cost 15%) favours **A** and **D**; **C** risks trust and fairness; **B** risks discrimination by destination → require safeguards and policy review.

MVP. Distance caps + ETA bands in one city, randomised by geography/time (switchback design to limit interference).

KPIs. Primary: cancellation rate per request. Secondary: completed trips per hour, time to pickup. Guardrails: driver earnings per online hour, rider wait time, complaint rate, fairness (cancellations by area).

Risks. Marketplace interference (treatment affects control through shared drivers) — use switchback or geo tests; reduced matches in sparse areas; regulatory pricing rules.

Experiment plan. Switchback test (alternating hourly treatment) for 3 weeks; analyse by area and time.

Example metrics (hypothetical). Cancellation 19.0% → 16.1% (−15%); completed trips/hour unchanged; driver earnings/online hour −1% (within guardrail); sparse areas saw longer waits → adjust caps there.

Recommendation. Roll out with area-specific caps; explore destination preview only with fairness controls.

Key PM lessons. Marketplace experiments need interference-aware designs; optimise the system, not one side; guardrails for fairness.

Case 3 · Amazon-style e-commerce: discovery,

conversion and recommendations

Hypothetical scenario

Amazon is widely known for features such as product recommendations ("customers who bought this also bought") and customer reviews. This case is a **hypothetical scenario** for a large e-commerce marketplace; numbers are invented.

Business background. A marketplace with millions of products; search and recommendations drive most discovery.

Target users. Shoppers on mobile; sellers.

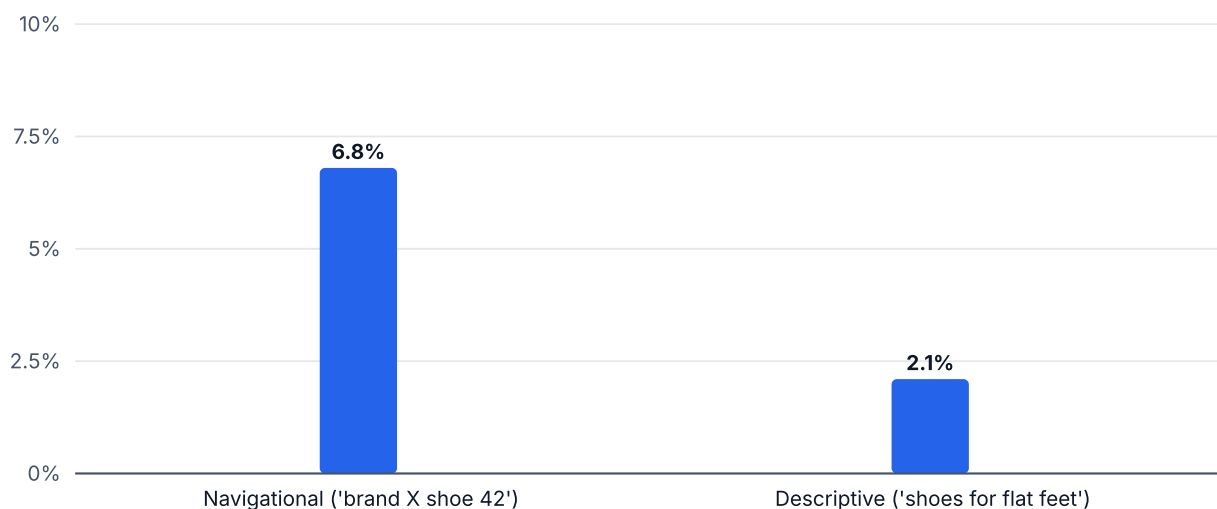
Customer problem. Search results for vague queries ("running shoes for flat feet") are noisy; shoppers bounce between filters and reviews.

Business objective. Increase search-to-purchase conversion and reduce time to purchase.

Available evidence (hypothetical). 41% of searches are "descriptive" (needs, not product names); their conversion is 2.1% vs 6.8% for navigational searches; session replays show repeated filter changes; review mining shows confusion about fit.

Assumptions. Semantic understanding of needs improves relevance; review summaries reduce decision time; personalised ranking helps without creating filter bubbles.

Product hypothesis. *Semantic search with need-based facets and AI-generated review summaries (with citations to reviews) will raise conversion of descriptive searches by $\geq 15\%$ relative.*



Conversion by search type (hypothetical baseline). Hypothetical. The gap identifies descriptive searches as the opportunity.

Candidate solutions. A) Hybrid semantic search; B) Need-based facets ("arch support"); C) Review summaries with "what buyers say about fit"; D) Personalised ranking; E) Guided Q&A shopping assistant.

Prioritisation. RICE favours **A** and **C** (high reach, moderate effort); **E** has low confidence and higher risk (hallucinated product claims).

MVP. Hybrid search for one category (footwear) + review summaries with links to source reviews.

KPIs. Primary: descriptive-search conversion. Secondary: time to add-to-cart, filter changes per session. Guardrails: return rate, complaints about inaccurate summaries, seller fairness.

Risks. Hallucinated summaries (must ground in reviews), review manipulation, privacy in personalisation, seller ranking fairness.

Experiment plan. A/B by user for 3 weeks in footwear; human evaluation of 300 summaries for accuracy before launch.

Example metrics (hypothetical). Descriptive conversion 2.1% → 2.5% (+19% relative); returns unchanged; 1.3% of summaries flagged as inaccurate in review → fix grounding before expansion.

Recommendation. Expand semantic search; hold summaries until inaccuracy < 0.5% on evaluation.

Key PM lessons. Segment the funnel by intent; ground AI in source content; measure returns and trust, not just conversion.

Case 4 · Netflix-style streaming: personalisation and retention

Hypothetical scenario

Netflix publicly describes using personalised recommendations and has written about personalising artwork on its technology blog. This case is a **hypothetical** streaming-service scenario; numbers are invented.

Business background. Subscription streaming; content costs are high; churn is the main threat to revenue.

Target users. Members on TV and mobile; households with multiple profiles.

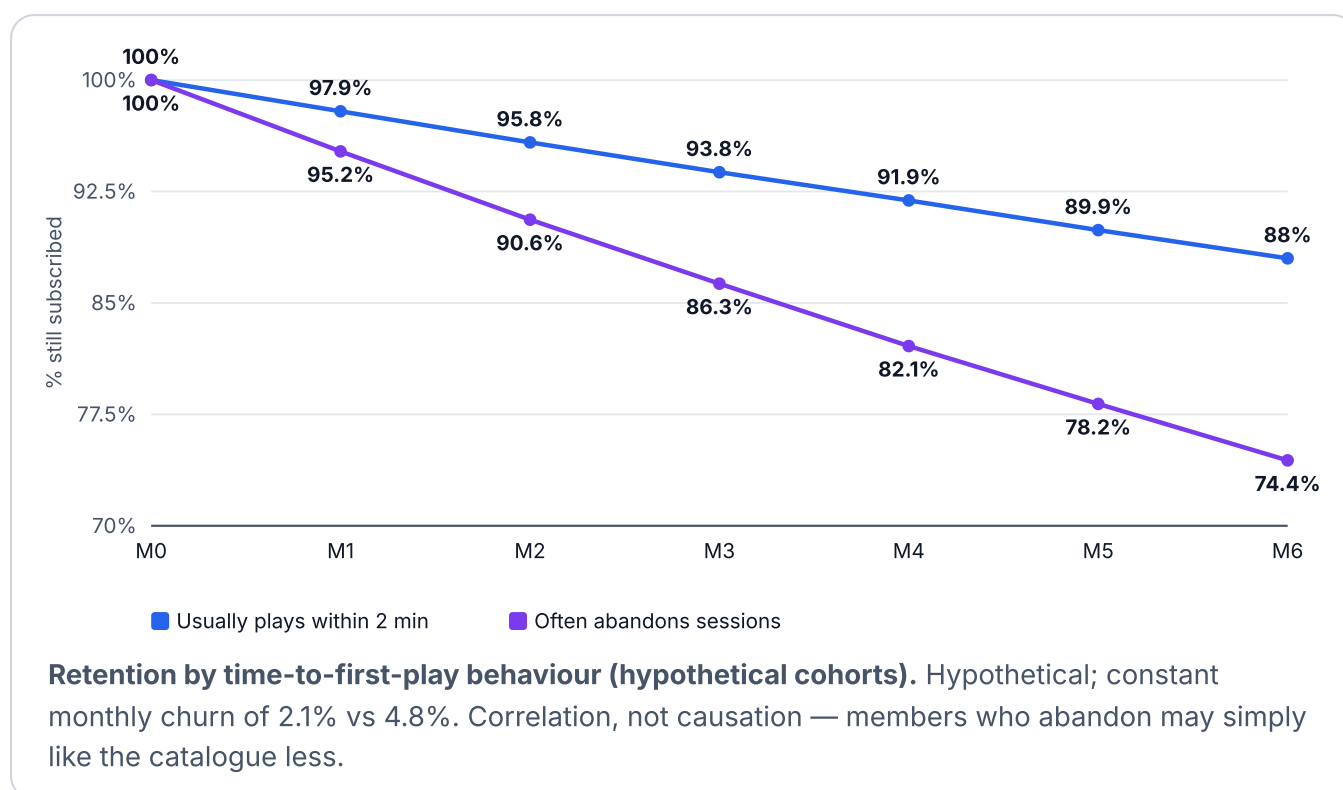
Customer problem. Members spend minutes browsing without choosing ("decision fatigue") and may conclude "there's nothing to watch," a churn risk.

Business objective. Reduce monthly churn and increase hours of satisfied viewing.

Available evidence (hypothetical). Members who start a title within 2 minutes of opening the app churn at 2.1%/month vs 4.8% for those who often abandon sessions; surveys cite “too much choice”; 30% of sessions end without playback.

Assumptions. Faster time-to-play reduces churn (correlation now — needs a causal test); better rows/artwork help choice; a “continue watching” improvement matters most for returning members.

Product hypothesis. *Personalised row ordering plus a “quick pick” for short sessions will reduce sessions-without-play by 15% and, over 3 months, reduce churn by 0.3 points.*



Candidate solutions. A) Personalised row ordering; B) Quick pick / shuffle; C) Better “continue watching” management; D) Artwork personalisation; E) Content-gap acquisition (not a product fix).

Prioritisation. A and C (high reach, medium effort, good confidence); B as a cheap test.

MVP. Row-ordering model change for 10% of members; quick-pick button test.

KPIs. Primary: sessions-without-play rate (leading); churn (lagging, 3-month holdout). Guardrails: completion rate (are people enjoying what they start?), complaints about repetitive recommendations, diversity of genres watched.

Risks. Optimising clicks over satisfaction; filter bubbles; long time to measure churn.

Experiment plan. A/B on row ordering 6 weeks with a long-term holdout for churn.

Example metrics (hypothetical). Sessions-without-play 30% → 26.5% (−12%); completion rate flat; 3-month churn delta −0.2 pts (CI includes 0) — continue holdout.

Recommendation. Ship row ordering; keep a holdout for churn; iterate quick pick.

Key PM lessons. Use leading indicators with long-term holdouts; separate correlation from causation; guard against engagement-at-all-costs.

Case 5 · Spotify-style music streaming: engagement and recommendations

Hypothetical scenario

Spotify is publicly known for personalised playlists such as Discover Weekly (launched 2015). This is a **hypothetical** music-streaming scenario; numbers are invented.

Business background. Freemium music streaming: ads on free tier, subscriptions on premium. Engagement predicts conversion and retention.

Target users. New free users; lapsed listeners; premium subscribers.

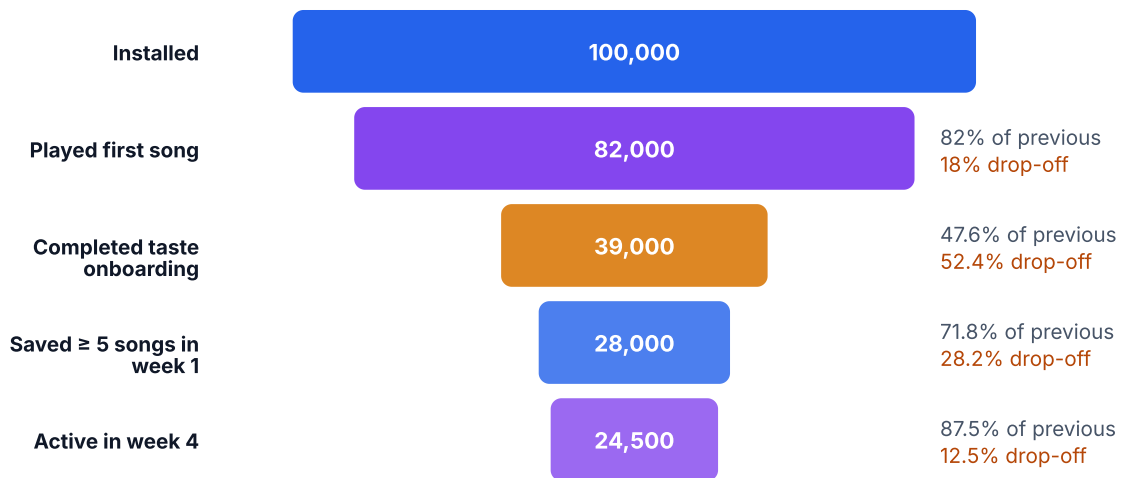
Customer problem. New users don't know what to play beyond a few favourite songs; personalisation is weak until the system learns their taste (cold start).

Business objective. Increase week-4 retention of new users and free→premium conversion.

Available evidence (hypothetical). New users who save ≥ 5 songs in week 1 retain at 46% at week 4 vs 18% for others (correlation); onboarding taste picker skipped by 52%; interviews: "the app plays the same things."

Assumptions. Better cold-start recommendations increase saving behaviour; saving causes retention (to test); an onboarding taste quiz with fewer, smarter steps won't be skipped.

Product hypothesis. *A 3-step taste onboarding (genres → artists → mood) plus a "first mix" playlist will raise week-1 saves ≥ 5 from 28% to 35% and week-4 retention by 3 points.*



New-user engagement funnel (hypothetical cohort of 100,000). Hypothetical. The onboarding step loses more than half — the main lever.

Candidate solutions. A) Shorter taste onboarding; B) “First mix” playlist; C) Social import (follow friends); D) Push notifications for new releases; E) Free-tier on-demand trial.

Prioritisation (ICE). A $8 \times 7 \times 8 = 448$; B $7 \times 6 \times 6 = 252$; D $5 \times 6 \times 9 = 270$ but risk of fatigue → guardrail; start with **A + B** together as one test.

MVP. New onboarding + first mix for 20% of new users in two markets.

KPIs. Primary: week-4 retention. Leading: week-1 saves ≥ 5 . Guardrails: skip rate in first mix, notification opt-outs, streams per user quality (not just volume).

Risks. Taste quiz biases toward popular music; artists' fairness; measuring causality for “saves”.

Experiment plan. A/B for 5 weeks (to observe week 4); stratify by market.

Example metrics (hypothetical). Onboarding completion 39% → 61%; saves ≥ 5 : 28% → 33%; week-4 retention 24.5% → 26.4% (+1.9 pts, CI +0.8 to +3.0); skip rate unchanged.

Recommendation. Ship; next test social import.

Key PM lessons. Fix cold start early; validate “aha” metrics causally; watch quality of engagement.

Case 6 · WhatsApp-style messaging: onboarding and

communication workflows

Hypothetical scenario

WhatsApp is widely known for phone-number-based sign-up with verification codes. Your course mapped WhatsApp onboarding as a user-flow exercise (Lecture 5). This case is **hypothetical**; numbers are invented.

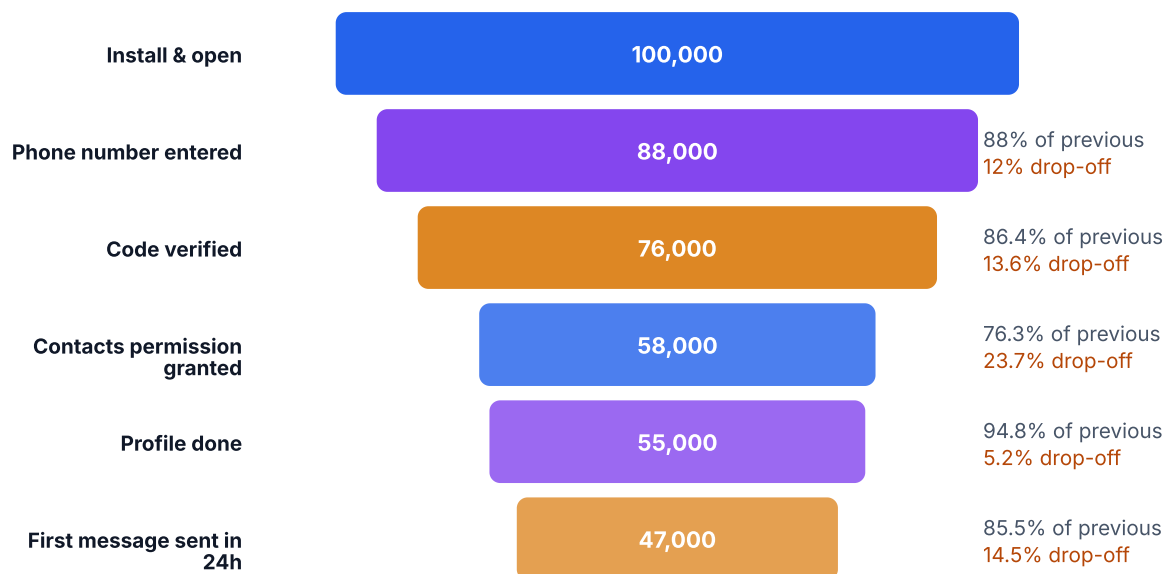
Business background. A messaging app's value depends on contacts already using it (network effects).

Target users. First-time smartphone users; users switching phones; small businesses.

Customer problem. New users drop during verification (SMS delays), contact-permission prompts, and profile setup; some never send a first message.

Business objective. Increase the share of installs that send a first message within 24 hours.

Available evidence (hypothetical).



Onboarding funnel (hypothetical, 100,000 installs). Hypothetical. Verification (~12k) and contact permission (~18k) are the largest drops.

Assumptions. Auto-reading the verification code and voice-call fallback reduce failures; explaining *why* contacts are needed (just-in-time) increases grants; users without permission can still be productive.

Product hypothesis. *Verification fallback + just-in-time contact permission with a clear explanation + a "message a contact by number" path will raise first-message-in-24h from 47% to 53%.*

User flow. Open → enter number → receive code (SMS or call) → verify → permission ask (in context) → profile → chat list with suggestions → first message.

Candidate solutions. A) Code auto-fill + call fallback; B) Just-in-time contact permission with explanation; C) Skip-able profile; D) Starter suggestions ("Say hi to Priya"); E) Offline/poor-network optimisation.

Prioritisation. RICE: B (reach 76k, impact 2, conf 80%, effort 1) = 121,600; A (reach 88k, impact 1, conf 80%, effort 2) = 35,200; D lower confidence. **B then A.**

MVP. Just-in-time permission explanation + "message by number" fallback.

KPIs. Primary: first message within 24h. Guardrails: spam reports, permission grant rate, privacy complaints, uninstalls.

Risks. Privacy perceptions around contacts; spam; SMS costs; regulatory consent rules.

Experiment plan. A/B on new installs in 3 countries for 2 weeks.

Example metrics (hypothetical). Permission grant 76% → 82% of verified; first-message 47% → 51.5%; spam reports unchanged.

Recommendation. Ship B; test A next; monitor privacy feedback.

Key PM lessons. Ask for permissions in context with a reason; always offer a path that works without permission; network products must reach first value fast.

Case 7 · Slack-style collaboration: channel creation and team UX

Hypothetical scenario

Slack organises work in channels. This case is a **hypothetical** team-collaboration product scenario (your project folders include earlier Slack channel-UX work you can use as a portfolio companion). Numbers are invented.

Business background. B2B collaboration SaaS sold per seat; value grows as teams adopt shared channels.

Target users. Team members, team leads, workspace admins.

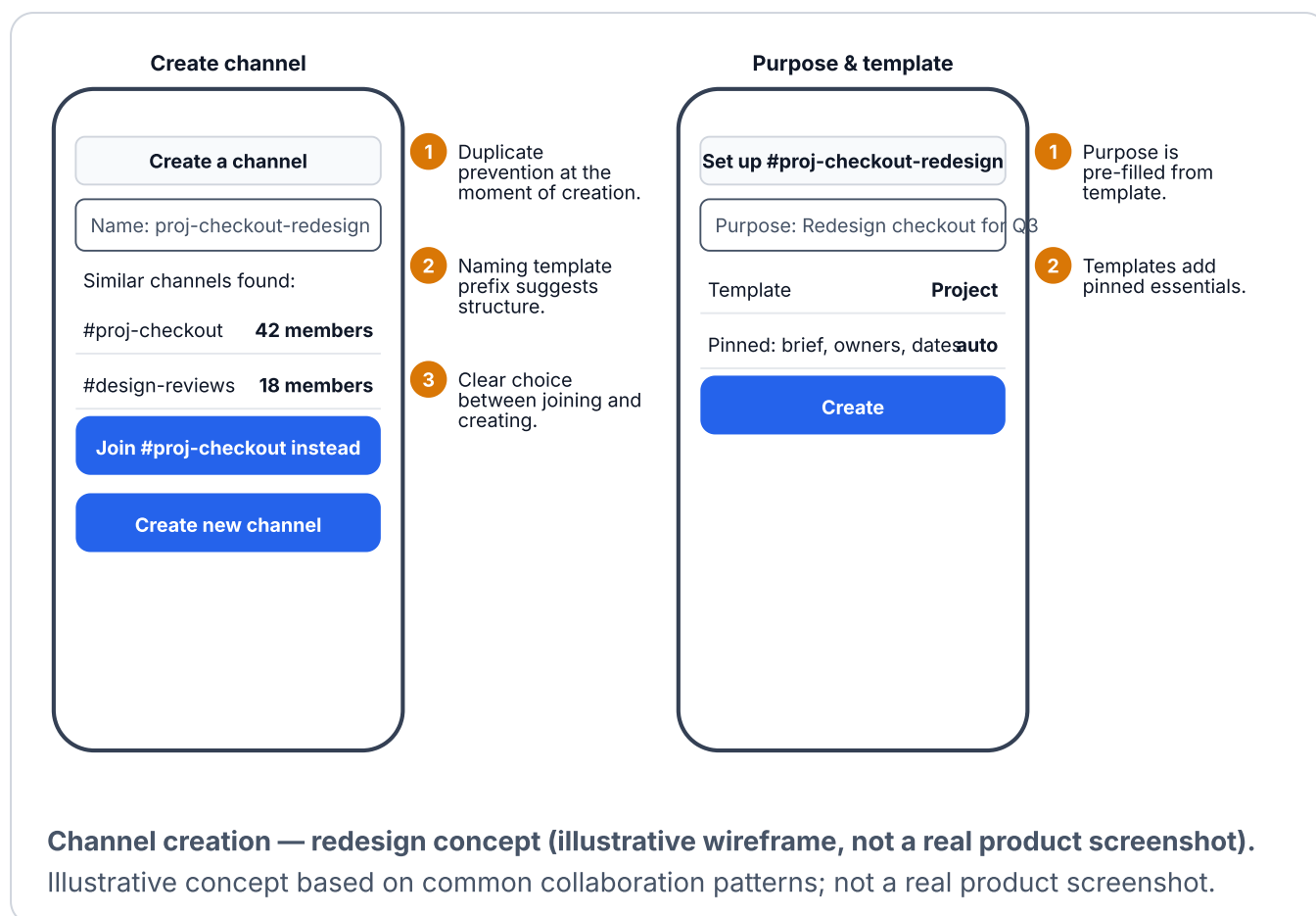
Customer problem. New teams create duplicate or poorly named channels; people can't find the right one and fall back to DMs, which hides knowledge.

Business objective. Increase share of messages in channels (vs DMs) and weekly active teams per account — leading indicators of expansion.

Available evidence (hypothetical). 22% of channels created in the last quarter duplicate an existing topic; 35% of new channels have no purpose set; tree tests show 4 of 8 participants failed to find the “design-review” channel.

Assumptions. Suggesting existing channels at creation prevents duplicates; templates with purpose and naming conventions improve findability; admins want light governance.

Product hypothesis. Showing “similar channels” and naming templates during creation will cut duplicate channels by 30% and raise channel-message share by 3 points.



Candidate solutions. A) Similar-channel suggestions; B) Naming templates; C) Channel directory with search improvements; D) Admin archiving of inactive channels; E) AI channel summaries.

Prioritisation. Value vs complexity: A and B quick wins; C medium; E later (needs eval and privacy review).

MVP. A + B for new channels in 10% of workspaces.

KPIs. Primary: duplicate-channel rate; channel message share. Guardrails: channel creation volume (don't block legitimate creation), time to create, admin complaints.

Risks. Friction slows urgent channel creation; suggestions leak private channel names (must respect permissions).

Experiment plan. Workspace-level randomisation (to avoid team interference) for 4 weeks; usability test of the creation flow first (6 participants).

Example metrics (hypothetical). Duplicates 22% → 15% (−32%); channel message share 61% → 63.5%; creation time +4 s median (acceptable).

Recommendation. Ship to all; add directory search next.

Key PM lessons. In B2B, randomise at the account/workspace level; respect permissions in suggestions; measure collaboration outcomes, not clicks.

Case 8 · Card issuer (American Express-style): digital onboarding and application drop-offs

Hypothetical scenario

American Express and other issuers offer digital card applications. This case is a **hypothetical** card-issuer scenario; it does not describe any issuer's actual process, data or results. Regulatory steps are general; real processes require compliance review.

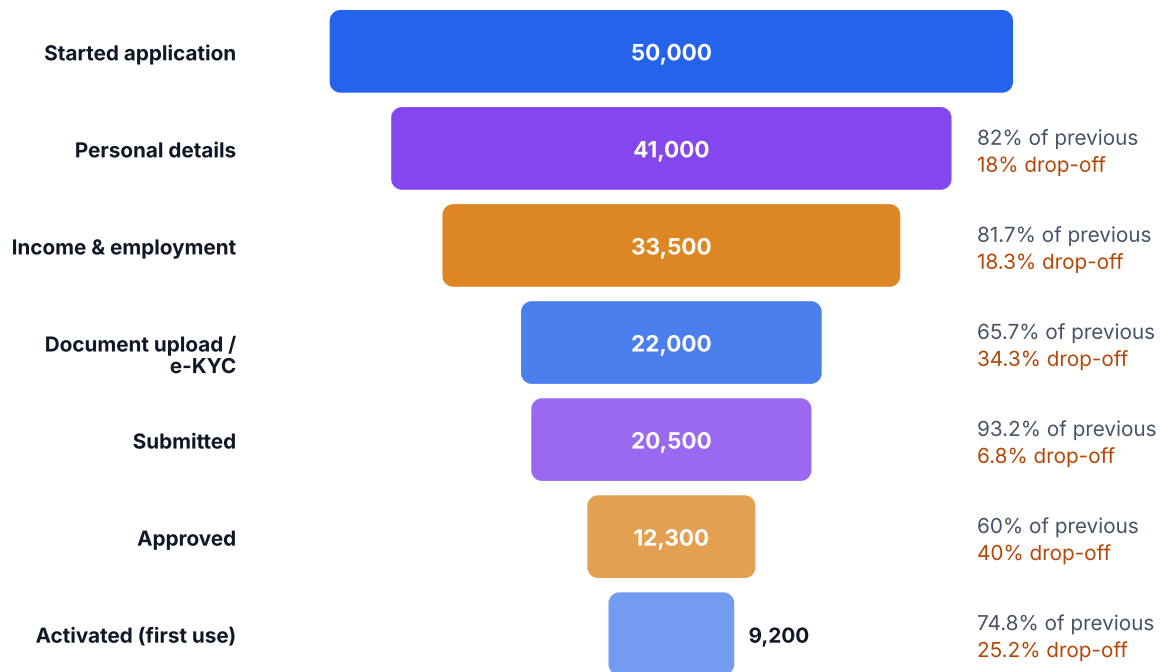
Business background. Digital applications for credit cards; acquisition costs are high, so each abandoned application is expensive.

Target users. Salaried professionals and self-employed applicants on mobile.

Customer problem. Long forms, document uploads that fail on mobile, unclear status after submission, and "will this affect my credit score?" anxiety.

Business objective. Increase submitted-application rate and approved-and-activated cards while keeping credit and fraud risk within policy.

Available evidence (hypothetical).



Application funnel (hypothetical month, 50,000 starts). Hypothetical. The largest drop is document upload / e-KYC (-11,500); approved-to-activated is also leaky.

Interviews (8): fear of credit-score impact; uploads failing on low light; no status after submitting.

Assumptions. Pre-fill and save-and-resume cut form abandonment; camera guidance improves uploads; a soft eligibility check reduces anxiety (where available and compliant); status tracking increases activation.

Product hypothesis. *Camera guidance + save-and-resume + clear status tracking will raise submit rate from 41% to 46% of starts with no increase in fraud or credit-loss indicators.*

Candidate solutions. A) Camera guidance and auto-capture; B) Save-and-resume with reminders (consented); C) Pre-fill from verified sources (with consent); D) Soft eligibility check messaging; E) Application status tracker; F) Activation nudges after approval.

Prioritisation (weighted, hypothetical): A 4.2, B 3.9, E 3.6, F 3.4, C 3.1 (higher compliance effort), D 2.8 (depends on bureau/legal). Start **A + B**, then **E + F**.

MVP. A + B for Android mobile web (largest drop segment).

KPIs. Primary: submit rate (starts → submitted). Secondary: upload success rate, time to complete, approved→activated. Guardrails: fraud rate, manual-review backlog, complaints, approval quality (early delinquency indicators).

Risks. Fraud exploits easier flows; consent and data-protection for pre-fill; accessibility; vendor SDK reliability.

Experiment plan. A/B on Android web for 4 weeks; fraud and risk review weekly; compliance sign-off before launch.

Example metrics (hypothetical). Upload success 67% → 81%; submit rate 41.0% → 44.6%; fraud indicators unchanged; activation unchanged (E/F next).

Recommendation. Ship A + B; build status tracker and activation journey next quarter.

Key PM lessons. In regulated onboarding, partner with risk and compliance from day one; target the biggest technical failure point; measure through to activation, not just submission. (*Practical Project 1 develops this end to end.*)

Case 9 · Payments gateway (Razorpay/Stripe-style): transaction reliability

Hypothetical scenario

Razorpay and Stripe are payment platforms used by businesses to accept online payments. This case is a **hypothetical** payments-platform scenario; numbers are invented and say nothing about either company.

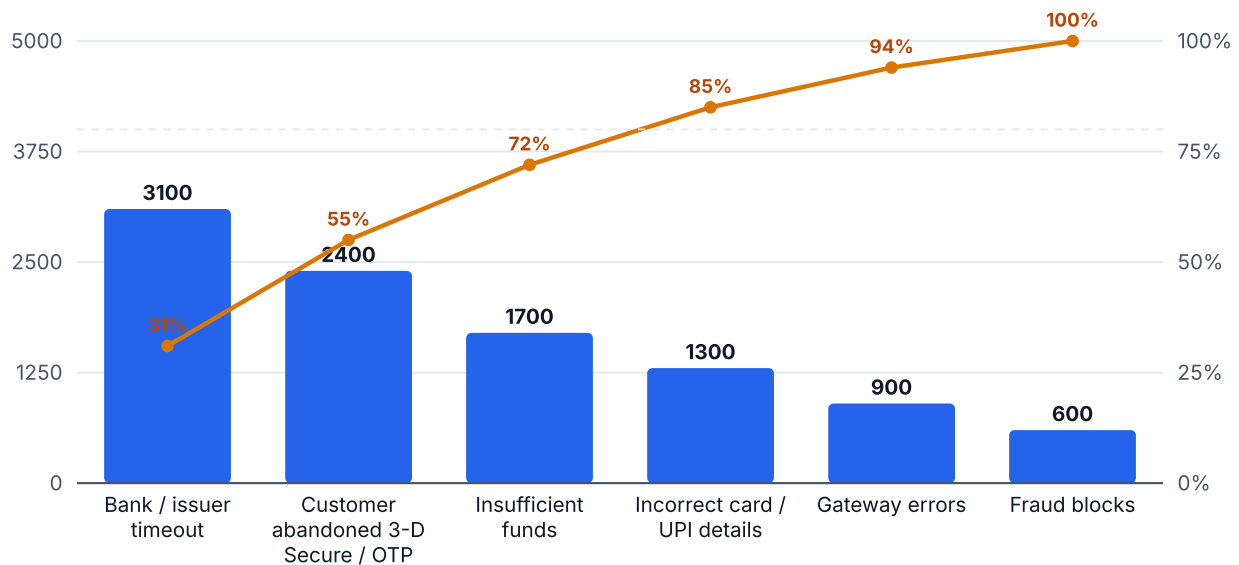
Business background. A payments platform earns per successful transaction; merchants choose providers largely on success rate, cost and developer experience.

Target users. Merchants (developers, finance teams) and their end customers paying online.

Customer problem. Merchants see unexplained payment failures at peak (sale days), lose revenue, and can't tell whether the cause is the bank, the customer or the gateway.

Business objective. Increase overall payment success rate and merchant retention; reduce support tickets about failures.

Available evidence (hypothetical).



Failure causes (hypothetical, 10,000 failed payments). Hypothetical. Some causes are addressable by product (retries, routing, UX), others are not (insufficient funds).

Assumptions. Smart routing to healthier bank/acquirer paths reduces timeouts; better authentication UX reduces abandonment; real-time failure reasons reduce merchant tickets.

Product hypothesis. *Health-based routing plus a merchant-facing failure dashboard will raise success rate by 1.5 points and cut failure-related tickets by 25%.*

User journey (merchant). Integrate API → go live → monitor dashboard → investigate failures → contact support → reconcile settlements.

Candidate solutions. A) Dynamic routing based on real-time issuer health; B) Retry with idempotency for transient errors; C) Authentication UX improvements (clear OTP screens, fallback methods); D) Merchant failure analytics with reason codes; E) Status page and proactive incident alerts.

Prioritisation. Cost of delay is high before the festive sale season → **A + E** first (WSJF), then D, then C.

MVP. Routing for the top 5 issuers by volume; incident alerts via webhook/email; reason-coded dashboard (beta) for 50 merchants.

KPIs. Primary: payment success rate (by method, issuer). Secondary: p95 authorisation latency; failure tickets per 1,000 transactions; merchant NPS/CSAT. Guardrails: duplicate charges (target 0), fraud rate, cost per transaction.

Risks. Routing changes increase cost; incorrect health signals; duplicate charges on retries (needs idempotency); regulatory requirements on authentication.

Experiment plan. Merchant-level or transaction-level randomisation with careful monitoring; staged rollout; incident review.

Example metrics (hypothetical). Success rate 92.0% → 93.4%; issuer timeouts -28%; tickets -22%; duplicate charges 0.

Recommendation. Expand routing to more issuers; make the failure dashboard GA with clear reason codes and recommended actions.

Key PM lessons. Reliability is the product for platforms; decompose failures and work on the controllable ones; developer- and finance-facing transparency reduces support load.

Case 10 · Duolingo-style learning app: gamification and retention

Hypothetical scenario

Duolingo is widely known for gamification such as streaks and leaderboards (leagues). This case is a **hypothetical** language-learning scenario; numbers are invented.

Business background. Freemium learning app; habit formation drives retention and premium conversion.

Target users. Casual learners, exam preparers, travellers.

Customer problem. Learners lose motivation after the first week; a broken streak can make them quit entirely.

Business objective. Raise day-30 retention and weekly learning minutes without increasing churn from frustration.

Available evidence (hypothetical). Users with a 7-day streak are 3× more likely to be active at day 30 (correlation); 40% of churned users broke a streak of ≥ 10 days in their last week; interviews: "I lost my streak, so why bother."

Assumptions. Protecting progress after a lapse (streak freeze/repair) prevents quitting; adaptive daily goals reduce burnout; social features help some segments but not all.

Product hypothesis. *Offering one free "streak repair" after a missed day and adaptive goals will raise day-30 retention by 2 points while keeping learning minutes stable.*



Habit loop for daily learning. The loop reinforces itself, but a broken streak can flip motivation to zero. Design recovery, not just pressure.

Candidate solutions. A) Streak repair; B) Adaptive daily goal; C) Friend quests; D) Leaderboards for all; E) Learning-outcome check-ins (mini tests).

Prioritisation. RICE favours A and B (broad reach, low effort); D risks demotivating low performers → opt-in only.

MVP. A + B for new users in one market.

KPIs. Primary: day-30 retention. Leading: 7-day streak rate. Guardrails: lesson quality (accuracy), notification opt-outs, reported anxiety/frustration in surveys, learning outcomes (test scores).

Risks. Gaming metrics (very easy lessons to keep streaks); dark-pattern perception; wellbeing.

Experiment plan. A/B for 6 weeks; survey sub-sample about motivation and pressure.

Example metrics (hypothetical). Day-30 retention 18.0% → 20.3%; learning minutes flat; opt-outs unchanged; self-reported pressure slightly lower.

Recommendation. Ship streak repair; refine adaptive goals; add learning-outcome measures to the North Star.

Key PM lessons. Gamification must serve the user's real goal (learning); design for lapses; pair engagement metrics with outcome and wellbeing guardrails.

Case 11 · Notion-style workspace: product-led growth

Hypothetical scenario

Notion is commonly cited as a product-led-growth example with a free plan and a large template gallery. This case is a **hypothetical** PLG workspace scenario; numbers are invented.

Business background. A flexible docs/wiki/database tool; individuals sign up free; revenue comes from team plans.

Target users. Individual knowledge workers; team leads; workspace admins.

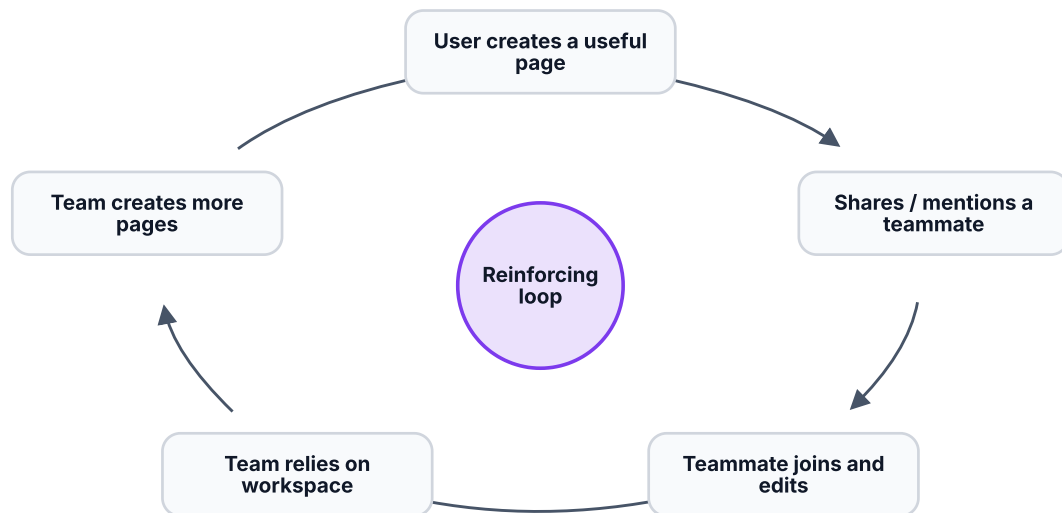
Customer problem. New users face a blank page and don't know what to build; individual users rarely invite teammates, so they never hit team value.

Business objective. Increase free→team conversion and expansion (seats per workspace).

Available evidence (hypothetical). Users who start from a template activate at 48% vs 21% from blank; workspaces with ≥ 3 active members convert at 9% vs 0.8% for single-user; template gallery used by 30% of new users.

Assumptions. Templates shorten time-to-value; collaboration triggers (mentions, shared pages) drive invites; upgrade prompts at limits convert better than generic prompts.

Product hypothesis. *Role-based template onboarding plus a “share with your team” step after first page creation will raise 14-day workspaces with ≥ 3 members from 12% to 16%.*



Collaboration growth loop. The loop turns individual value into team adoption — the moment monetisation becomes natural.

Candidate solutions. A) Role-based template onboarding; B) Share prompt after first useful page; C) Team trial when 3rd member joins; D) Usage-limit upgrade prompts; E) Sales-assist for PQL workspaces > 20 users.

Prioritisation. A and B (activation and loop entry) first; C/D next; E for larger accounts.

MVP. A + B for new sign-ups from work email domains.

KPIs. Primary: % new workspaces with ≥ 3 active members by day 14. Secondary: activation, invites per user, free $\rightarrow$ team conversion at day 60. Guardrails: spammy invites (bounce/complaint rate), user satisfaction, privacy of shared pages.

Risks. Over-prompting invites; templates that don't fit; enterprise security concerns.

Experiment plan. A/B by new workspace for 4 weeks, conversion read at day 60 with holdout.

Example metrics (hypothetical). Activation 34% $\rightarrow$ 41%; ≥ 3 -member workspaces 12% $\rightarrow$ 15.2%; invite complaint rate unchanged; day-60 conversion pending.

Recommendation. Ship; test team trial trigger next.

Key PM lessons. In PLG, activation and the collaboration loop are the growth engine; monetise at moments of team value; measure loop health.

Case 12 · AI-powered personal financial assistant: AI product design and responsible automation

Hypothetical scenario

A fictional bank app adds an AI assistant. Everything here is **hypothetical**. Financial advice and automated payments are regulated in most markets; real products need legal and compliance review. Module 12 contains a related full case on an AI *customer-support* assistant.

Business background. A digital bank wants deeper engagement and fewer support contacts by helping customers understand and manage their money.

Target users. Young professionals managing first salaries; households with recurring bills.

Customer problem. Customers don't know where their money goes, miss bill due dates, and find statements hard to interpret.

Business objective. Increase monthly active usage of money-management features and reduce late payments, while avoiding unsuitable financial advice.

Available evidence (hypothetical). 61% of surveyed customers can't name their top three spending categories; 14% paid at least one bill late in 6 months; support receives many "what is this charge?" queries.

Assumptions. Natural-language answers about the customer's own transactions are valuable; bill reminders reduce late payments; users want suggestions but not autonomous payments at first.

Product hypothesis. *An assistant that answers questions about the customer's own transactions with citations to those transactions, and drafts bill-payment reminders and payments for explicit approval, will reduce late payments by 20% among users and cut "what is this charge?" contacts by 25%.*



Autonomy plan by capability. Start with explanations and reminders; earn autonomy with evidence. No investment or credit advice in scope.

Candidate solutions. A) Spending insights Q&A (read-only); B) Bill detection and reminders; C) Payment drafting with approval; D) Budget coaching; E) Investment suggestions (excluded — regulated advice).

Prioritisation. Value vs risk: A and B high value, low risk; C medium risk with approvals; D later; E out of scope.

MVP. A + B in-app for 5% of customers; English and one regional language.

Design and controls. * Answers grounded in the customer's own transactions (read-only tools) and approved help content; **citations** to transactions. * **Refuse** investment, tax and credit advice; route to human advisers or general education content. * **Permissions** enforced in backend; no access to other customers' data. * **No external communication** by the assistant; payments only with explicit confirmation (later phase). * **Privacy:** minimise data in prompts; vendor terms prevent training on customer data; logs redacted. * **Evaluation:** test set of 300 questions (category math, merchant identification, refusals, injection attempts in merchant names/notes).

KPIs. Primary: late-payment rate among users; Secondary: weekly active users of insights, CSAT, "what is this charge?" contacts. Quality/safety: answer accuracy (sampled), unsuitable-advice incidents (target 0), privacy incidents (target 0), refusal correctness.

Risks and constraints. Miscategorised transactions leading to wrong insights; perceived advice; over-reliance; prompt injection via transaction descriptions; regulatory obligations.

Experiment plan. Offline eval gate → 5% randomised pilot for 8 weeks → compare late payments vs control.

Example metrics (hypothetical). Category answers accurate in 94% of sampled cases (target 95% → improve merchant mapping); late payments -17% vs control; charge-query contacts -21%; 0 advice incidents; 2 injection attempts blocked in red-team.

Recommendation. Improve categorisation accuracy to ≥ 95%, then expand to 25%; begin design for payment drafting with approvals.

Potential outcomes. Best: trusted daily money companion. Base: useful insights and reminders. Worst: accuracy or advice incidents → restrict to reminders only.

Key PM lessons. Scope out regulated advice explicitly; ground in the customer's own data with citations; autonomy is earned; evaluation and red-teaming are launch gates.