

HYPOTHETICAL CASE STUDY

Tarangini Bank: Knowing the Customer Behind the Balance

How a mid-sized Indian private bank set out to use k-means clustering on demographic and spending-behaviour data to rebuild its customer segmentation

This case is fictional and was written for classroom discussion. Tarangini Bank and all individuals named are invented; any resemblance to actual institutions or persons is coincidental. Industry trends are described at a general level, and figures should be verified against current sources before being cited.

1. Background of the Bank

On a humid Monday morning in July 2026, Ananya Deshpande, Chief Digital and Analytics Officer of Tarangini Bank, stared at a slide her team had prepared for the board. It showed two customers, both 34, both living in Pune, both holding a “Privilege” savings account with an average balance of about ₹2.8 lakh. One spent most of her money through UPI at kirana stores and school-fee portals, rarely visited a branch, and ran a mutual fund SIP with a competitor. The other withdrew cash every week, used his debit card mainly at fuel pumps, and had walked into his branch eleven times in the past year. In the bank’s systems, they were the same customer. “We are treating them as twins,” she told her team, “when they barely share a pulse.”

Tarangini Bank was founded in 1946 in Mangaluru as a community bank serving traders and coffee planters along the Karnataka coast. Over eight decades it grew into a mid-sized private sector bank with roughly 1,150 branches, 2,400 ATMs, and about 14 million customers across 22 states, although nearly 60% of its business remained concentrated in South and West India. In FY2026 its balance sheet stood at about ₹2.9 lakh crore. Retail loans made up 54% of advances, and its CASA ratio had slipped from 38% to 31% over four years.

The bank had a reputation for stability, conservative underwriting and personal service. Branch managers knew families across generations; many customers had opened their first account as students and later taken home and gold loans from the same branch. But the intimacy that defined Tarangini also made it slow to change. Decisions about products and campaigns were driven largely by branch feedback and product-led sales targets.

In 2024 the board appointed a new MD and CEO, Rohan Kulkarni, a consumer banker from a large private bank. He launched “Tarangini Next,” a five-year strategy to become “the most customer-understanding bank in India.” It set three goals: raise the digital share of transactions to 90%, lift products per customer from 1.9 to 2.6, and cut attrition in the first three years of a relationship by a third. Central to the strategy was the use of AI to understand customers better. Deshpande, recruited from a fintech lender in 2025, was given the mandate and a budget of ₹180 crore over three years. Her first priority was to rebuild how the bank segmented its customers.

2. Trends in Retail Banking in India

The Indian retail banking landscape that Tarangini faced in 2026 bore little resemblance to the one in which most of its senior managers had built their careers. Several forces were reshaping how customers saved, spent, borrowed and chose their banks.

The payments revolution. The single largest shift was the rise of the Unified Payments Interface (UPI), launched by the National Payments Corporation of India in 2016. By 2025 UPI was processing upward of 18 billion transactions a month and had become the default way Indians paid for everything from vegetables to rent. For banks, UPI was a double-edged sword. It moved customers' everyday spending onto digital rails, generating a stream of behavioural data that had never existed when such purchases were made in cash. But the customer-facing experience was dominated by third-party apps such as PhonePe and Google Pay, which together accounted for the large majority of volumes. Many customers interacted with their bank's brand only when they checked a balance or read a debit SMS. The bank held the account; the app held the relationship.

Digital public infrastructure. UPI sat within a broader "India Stack": Aadhaar-based eKYC, the Jan Dhan financial inclusion programme with more than 55 crore basic accounts, DigiLocker, and the Account Aggregator framework, which went live in 2021 and allowed customers to share financial data across institutions with consent. The RBI's pilot of a Unified Lending Interface in 2024 promised to reduce the friction of credit assessment further. Collectively these lowered the barriers to both switching and entry. A fintech with a good app and an Account Aggregator connection could now underwrite a customer it had never met.

The deposit squeeze. Indian households were rapidly financialising their savings, but not necessarily into bank deposits. Monthly inflows into mutual fund systematic investment plans crossed ₹25,000 crore in 2025 as young investors embraced equity markets through apps. Across the system, credit growth outpaced deposit growth for long stretches, forcing banks to compete for deposits with higher rates and pushing CASA ratios down. For mid-sized banks without the brand pull of the largest players, retaining low-cost savings balances had become a strategic concern.

The unsecured credit cycle. Personal loans, credit cards and small-ticket digital loans grew rapidly after the pandemic. In November 2023 the RBI raised risk weights on unsecured consumer credit, signalling discomfort with the pace of growth. Lenders responded by tightening underwriting, which placed a premium on knowing which customers had genuine repayment capacity and which were stretching. Behavioural data — how a customer spends, saves and pays — was increasingly seen as a valuable complement to a bureau score.

Regulation of data and AI. The Digital Personal Data Protection Act, 2023, and its subsequent rules introduced consent-based processing, purpose limitation and significant penalties for misuse. The RBI's 2022 digital lending guidelines restricted how lenders and their partners could collect and use customer data, and in 2025 an RBI-appointed committee published a framework for the responsible and ethical use of AI in the financial sector, emphasising fairness, explainability and board-level accountability. Banks could no longer treat customer data as a free resource to be mined; they had to justify each use.

The new face of the customer. India's banking customer was getting younger, more digital and more geographically dispersed. Much of the growth in new accounts, UPI users and credit-card issuance was coming from Tier 2 and Tier 3 cities, where customers often preferred vernacular interfaces and mixed digital and cash behaviour in ways that metro-centric products did not anticipate. At the same time, a large cohort of older customers — many of them high-balance depositors — remained loyal to their branches. Banks needed to serve both groups without alienating either.

Fraud and trust. The digitisation of payments brought a surge in digital fraud, mule accounts and social-engineering scams. The RBI and its innovation hub promoted AI-based tools to identify mule accounts, and banks invested heavily in transaction monitoring. Customer trust, built painstakingly over decades, could be dented by a single viral incident.

The AI moment. Finally, the rise of generative AI after 2023 triggered a rush of pilots across Indian banks: chatbots in regional languages, AI-assisted relationship managers, automated document processing. Yet many executives privately admitted that their AI investments had produced impressive demonstrations and few measurable outcomes. The underlying problem was rarely the model. It was that banks did not understand their customers well enough to know what to personalise, for whom and through which channel.

For Tarangini, these trends converged on a single conclusion. The bank could not outspend the largest private banks on technology, nor out-innovate the fintechs on user experience. Its potential edge lay in combining the deep, long-standing relationships of its branch network with a far sharper, data-driven understanding of who its customers were and how they actually behaved.

3. The Customer Segmentation Problem

Tarangini's existing segmentation was a product of its history. Retail customers were placed into four tiers — Classic, Value, Privilege and Wealth — based almost entirely on average quarterly balance in savings and deposit accounts. A second, informal layer of segmentation lived in the heads of branch managers, who flagged "important" customers on the basis of personal knowledge. Marketing was planned product by product: the cards team pulled a list of Privilege customers under 40; the home loans team targeted customers with salary credits above a threshold.

The results were predictable. Response rates for most digital campaigns had fallen below 1%. Customers complained of irrelevant offers — a 68-year-old retiree receiving a pitch for a travel credit card, a young professional being called by a branch about a fixed deposit. An internal analysis showed that 41% of customers who closed their accounts in FY2025 had been in the Privilege or Wealth tiers, and most had shown declining transaction activity for six months or more before leaving — signals the balance-based segmentation never captured.

Deshpande's diagnosis was that balance measured what customers had, not who they were. Two customers with the same balance could have entirely different needs, aspirations and preferred ways of engaging. She wanted a segmentation that combined demographics — age, gender, city tier, occupation, income band, relationship vintage — with behavioural signals that could serve as proxies for psychographics: how customers actually spent their money, and through which channels.

The bank's transaction data was rich in such signals. Card spends carried merchant category codes; UPI merchant payments could be mapped to categories; bill payments revealed school fees, insurance premiums and utility usage; channel logs showed whether a customer preferred the app, internet banking, ATMs or the branch. From these the team could infer whether a customer was experience-oriented (travel, dining, entertainment), family-oriented (education, groceries, healthcare), aspirational (electronics, fashion, EMI purchases) or thrift-oriented (high savings rate, low discretionary spend).

The analytics team, led by Dr. Karthik Subramanian, proposed k-means clustering as the starting point. K-means is an unsupervised algorithm that partitions customers into k groups by repeatedly assigning each customer to the nearest cluster centroid and recomputing the centroids until the assignments stabilise. Its appeal lay in its simplicity, its speed on millions of records and its interpretability: each cluster could be described by the average profile of its members.

Subramanian was candid about the challenges. First, k-means measures similarity with Euclidean distance, so variables on large scales, such as spend in rupees, would swamp variables on small scales, such as the share of transactions made on mobile. Every feature would need to be standardised, and heavily skewed spend values log-transformed. Second, many demographic variables were categorical. One-hot encoding them could distort distances; an alternative, k-prototypes, handled mixed data but was unfamiliar to the business. Third, k itself was not given. The elbow method and silhouette scores would help, but the final number had to be one that marketers and branches could act on. Fourth, k-means was sensitive to outliers and to the initial placement of centroids: a few ultra-high-net-worth customers could drag centroids, and different random starts could yield different clusters. Finally, segments had to be stable enough to plan around yet refreshed often enough to reflect changing behaviour.

There were organisational questions too. Product heads worried that new segments would cut across their targets. Branch managers feared that “algorithms” would override their judgement about customers they had known for years. And the compliance team asked whether using spending patterns for segmentation was consistent with the purposes for which customers had shared their data.

4. The Data Infrastructure in the Bank

If segmentation was the question, data was the constraint. Tarangini's technology estate reflected decades of incremental growth. At its heart was a core banking system implemented in 2009, which held customer, account and transaction records. Around it sat a constellation of specialised systems: a card management platform run with a third-party processor, a UPI switch, separate platforms for internet and mobile banking, a loan origination system, an in-house gold-loan application, a call-centre CRM and a treasury system (see Exhibit 1).

In 2014 the bank had built an on-premise enterprise data warehouse, primarily to serve regulatory reporting and management information. Core banking data was loaded nightly; data from other systems arrived weekly or monthly, and some — such as mobile app clickstreams — was never loaded at all. The warehouse was designed around accounts, not customers. Producing a single view of a customer's relationship required joins across a dozen tables and, often, manual reconciliation.

The most stubborn problem was customer identity. Over the years the bank had created multiple customer identification files (CIFs) for the same person — one when he opened a savings account as a student, another for a home loan, a third through an acquired bank's branch. An internal audit estimated that about 12% of active customers had duplicate CIFs. Unless these were resolved, any clustering would split one customer's behaviour across several "people." The team planned to combine deterministic matching on PAN and other permissible identifiers with probabilistic matching on name, date of birth, address and mobile number.

Data quality was uneven. Demographic fields captured at account opening were often stale: occupation had not been updated for a third of customers in more than five years, and income was self-declared and frequently blank. Address data was inconsistent, complicating classification by city tier. Transaction data, by contrast, was rich and reliable but needed substantial enrichment. Card transactions carried merchant category codes, but UPI payments to merchants carried only a merchant identifier and free-text narration, which had to be classified into spending categories. Person-to-person UPI transfers — a large share of volumes — were especially hard to interpret: was a ₹15,000 transfer rent, a loan to a friend or a payment to a domestic helper?

In 2023 Tarangini began building a cloud-based data lake hosted in Indian data centres, in line with RBI requirements on the localisation of payment data. By mid-2026 card, UPI and core banking transactions were streaming into the lake in near real time, but other sources were still being migrated. Deshpande's team proposed a "customer 360" layer on top of the lake: a resolved customer identity, a feature store holding several hundred engineered variables per customer, and a model-management environment to train, version and monitor models.

For segmentation, the team drafted an initial feature set in five families (see Exhibit 2). Demographic features included age, gender, city tier, occupation category, declared income band and relationship vintage. Spend-mix features captured the share of monthly spend across about fifteen categories. Spend-intensity features covered total monthly spend, average ticket size, transaction frequency, the share of spend on weekends and late evenings, and festive-season uplift. Channel features included the shares of transactions made through UPI, debit card, credit card, net banking, ATM cash and the branch, along with monthly app logins and self-service usage. Financial-health features — average balance, savings rate, EMI-to-inflow ratio — rounded out the picture.

Each choice involved trade-offs. Twelve months of history gave stable patterns but missed recent life events; three months was more responsive but noisier. Including many correlated variables — total spend alongside transaction count, for instance — would give some behaviours outsized weight in the distance calculation. The team considered principal component analysis to reduce dimensionality, at some cost to interpretability.

Governance added another layer. Under the DPDP Act the bank needed a lawful basis and a clear purpose for processing personal data. Legal counsel advised that internal analytics to improve services could be justified, but that some uses, such as sharing segment-level data with partners, would require explicit consent. Sensitive inferences — about specific health conditions or religious practice, for example — were to be avoided, and healthcare spending would be used only as a single aggregated category. Training data would be pseudonymised and access restricted on a need-to-know basis.

Finally, there was talent. Subramanian's team had 22 data scientists and engineers, many recently hired; few had banking experience, and business teams had limited fluency in analytics. Deshpande knew that success would depend as much on "translators" — people who could turn clusters into campaigns and branch actions — as on the algorithm itself.

5. Possible Use Cases After Improved Customer Segmentation

To build support, Subramanian's team ran a pilot in spring 2026 on a random sample of 500,000 active customers, using 34 standardised features and k-means with k-means++ initialisation. The elbow curve and silhouette scores pointed to between five and eight clusters; the team chose six, which business stakeholders found the most interpretable (see Exhibit 3). Even this early result reframed internal debates: the largest cluster, "Digital Family Builders," contained customers from all four balance tiers. The leadership team identified several use cases that a robust segmentation could unlock.

Personalised next-best offers. Instead of product teams pushing campaigns to broad lists, segment profiles could guide which products and messages were most relevant — a lifestyle card for experience seekers, education loans and child plans for family builders, recurring deposits for young savers. Segments would form the first layer, with propensity models within each segment refining individual targeting.

Channel strategy and branch transformation. Digital-first segments could be served through the app and video banking, while cash-reliant segments could be helped towards digital adoption with assisted service in branches. Branches in catchments dominated by traditional customers might be staffed and designed differently from those in digitally savvy neighbourhoods.

Deposit retention. By tracking customers' movement between segments over time, the bank could spot early warnings — a customer drifting towards a dormant cluster, or showing rising transfers to investment platforms — and trigger timely retention actions.

Credit and risk. Behavioural segments could sharpen pre-approved credit offers, particularly for new-to-credit customers in smaller cities. Risk teams cautioned that segments must complement, not replace, credit models and be tested for unfair outcomes.

Product design and partnerships. Clusters could reveal unmet needs. The pilot suggested that many small-business owners were running their trade through personal savings accounts, pointing to an opportunity for simple current accounts, merchant QR codes and small working-capital lines. Spend patterns in the experience-seeker cluster could guide the choice of co-branding partners for a new card.

Personalised digital experience. The mobile app's home screen, notifications and vernacular content could be tailored to each segment, giving customers a reason to open the bank's app rather than a third-party payments app. Marketing spend could also be reallocated from mass campaigns to the segments with the highest expected response.

Deshpande insisted that each pilot be run against control groups and measured on campaign conversion, products per customer, digital adoption, attrition and cost-to-serve, so the business could see impact rather than activity.

6. Going Forward

As Deshpande prepared for the board meeting, she weighed the decisions ahead. Should the bank roll out the six-cluster segmentation to all 14 million customers now, accepting imperfect data, or first spend a year fixing customer identity and demographic records? Was k-means the right tool, or should the team move to k-prototypes, Gaussian mixture models or hierarchical methods that might capture more nuance but would be harder to explain? How often should segments be refreshed, and who should own them — analytics, marketing or a new customer-strategy function?

There were deeper questions too. How would branch managers, whose judgement had long defined Tarangini's relationships, be brought on board? How could the bank use spending data to serve customers better without making them feel watched? And how would it know whether the new segmentation was truly improving customer orientation, rather than merely making campaigns more efficient?

Kulkarni had been clear: the board wanted a recommendation, not a menu. Deshpande looked again at the slide of the two 34-year-olds from Pune. The algorithm had found the difference between them in minutes. Turning that insight into a bank that behaved differently would take far longer.

Exhibit 1: Key Customer Data Sources at Tarangini Bank (mid-2026)

System	Key data held	Status in data lake	Main challenges
Core banking system (2009)	Customer master, accounts, balances, transactions	Streaming, near real time	Duplicate CIFs; outdated demographic fields
Card management (third-party processor)	Debit and credit card transactions with merchant category codes	Streaming, near real time	Credit card history held partly with processor
UPI switch	P2P and P2M transactions, VPAs, narration text	Streaming, near real time	Free-text narration; purpose of P2P transfers unclear
Mobile and internet banking	Logins, feature usage, clickstream	Partly migrated (logins only)	Historical clickstream never captured
Loan origination and gold-loan systems	Applications, collateral, repayment history	Monthly batch	In-house gold-loan app lacks APIs
Branch and ATM systems	Cash withdrawals and deposits, branch transactions	Daily batch	Purpose of branch visits rarely recorded
Call-centre CRM	Complaints and service requests	Not yet migrated	Unstructured notes; inconsistent tagging

Source: Hypothetical, prepared for this case.

Exhibit 2: Candidate Feature Set for k-means Segmentation

Feature family	Example variables	Preparation for k-means
Demographic	Age, gender, city tier, occupation category, declared income band, relationship vintage	Bin age; one-hot encode categories with down-weighting, or use k-prototypes for mixed data
Spend mix (psychographic proxies)	Share of monthly spend on groceries, dining, travel, entertainment, education, healthcare (aggregated), fuel, fashion, electronics, utilities, EMIs, investments and others	Shares sum to one; standardise (z-score)
Spend intensity and timing	Total monthly spend, average ticket size, transactions per month, weekend and late-evening share, festive-season uplift	Log-transform skewed values; cap outliers at the 99th percentile; standardise
Channel behaviour	Share of transactions via UPI, debit card, credit card, net banking, ATM cash and branch; app logins per month; self-service usage	Standardise; drop or combine highly correlated variables
Financial health	Average balance, savings rate, EMI-to-inflow ratio, transfers to investment platforms	Log-transform; standardise; test runs with and without balance to avoid recreating old tiers

Source: Hypothetical, prepared for this case. Dimensionality reduction (e.g., PCA) was under consideration for the full roll-out.

Exhibit 3: Illustrative Pilot Results — Six Customer Clusters (k = 6)

Cluster	Share of sample	Defining characteristics (relative to average)	Potential actions
Digital Family Builders	24%	Age 30–45, metro and Tier 2; high share of spend on education, groceries and healthcare; over 80% of transactions via UPI and app; home-loan EMIs	Education loans, child plans, family health insurance, app-led offers

Cluster	Share of sample	Defining characteristics (relative to average)	Potential actions
Cash-Comfortable Traditionalists	21%	Age 50+, Tier 2 and 3; frequent ATM cash and branch visits; high fixed-deposit balances; low discretionary spend	Assisted digital adoption, senior-citizen deposits, fraud-awareness outreach
Drifting Dormants	19%	Mixed demographics; transaction activity falling over six months; salary credits stopped or rising outward transfers	Retention triggers, reactivation offers, RM call-backs
Young Urban Experience Seekers	17%	Age 22–32, metros; high travel, dining, entertainment and online shopping; heavy credit-card use; late-evening activity; low balances	Lifestyle cards, starter SIPs, responsible-credit nudges
Small-Business Hustlers	13%	Age 28–50, Tier 2 and 3; high-frequency UPI inflows and merchant payments; irregular balances	Current accounts, merchant QR, working-capital loans
Affluent Global Spenders	6%	Age 35–55, metros; high-ticket travel and forex spend; large investment transfers; multi-channel users	Wealth relationship manager, premium cards, forex services

Source: Hypothetical illustrative results from a pilot on 500,000 customers using 34 standardised features. k was selected using the elbow method, average silhouette scores and business review.