Report

Cracking the Credit Code: Alternative Data *and* AI for Financial Inclusion

MAY 2026

IFC International Finance Corporation
WORLD BANK GROUP

About IFC

IFC—a member of the World Bank Group—is the largest global development institution focused on the private sector in emerging markets. We work in more than 100 countries, using our capital, expertise, and influence to create markets and opportunities in developing countries. In fiscal year 2025, IFC committed a record $71.7 billion to private companies and financial institutions in developing countries, leveraging private sector solutions and mobilizing private capital to create a world free of poverty on a livable planet. For more information, visit www.ifc.org.

Acknowledgments

This report was developed by Karan Singhal and Jaylan ElShazly, under the direction of Principal Economist Namita Datta. The project team also includes Natnael Nigatu, Mahima Kataria, Sendy Jasmine Karunia Hadi, and Stephanie An Ni Wu. It was developed in close collaboration with Montserrat Ganuza, Anushe Khan, Jose Felix Etchegoyen, Hector Shibata Salazar, and Carlos Alberto Sanchez Reboiro.

Valuable assistance was provided by IFC and World Bank Group colleagues Alexis Geaneotes, Anna Fruterro, Arisha Salman, Camila Quevedo Vega, Collen Mansuda, Fangfang Jiang, Gayatri Vikram Murthy, Getrude Misango, Jiayi Ma, Maria Carolina Salazar Pardo, Maria Fernandez Vidal, Marie-Sarah Chataing, Natalia Kaur Bhatia, Nora Katherine Brown, Oliver Rowntree, Partha Guha Thakurta, Placide Bakala, Sheirin Iravantchi, and Urvashi Chopra. The team extends special thanks to Anna Polatschek, Arisha Salman, Camila Quevedo Vega, Diego Javier Ubfal, and Maria Fernandez Vidal for their detailed comments and engagement with the internal draft, which greatly strengthened the final version, as well as to Gina Wilkinson for editing support and Beatriz Calvo Garro for graphic design.

The team is grateful to the firms and individuals who generously shared their time and insights, which helped to shape the findings and analysis in this report. These include Carrie Ruh and Jared Miller from Accial Capital, Ramdhan Anggakaradibrata from Amartha, Ebby Gatamu and Kibe John from Cladfy, Frank Adibi and Stephany Lawson from CrownCity Technology, Siegfried Eisenmeier from Didi, Shrikant Patil from DigiAlly, Chilufya Mutale from Eshandi, Lisa Smith, Alfred Yeboah, and Laya Hess-Skinner from the Grameen Foundation, Panos Loukos and Ahmed Javed from GSMA, Ifedayo Durosinmi-Etti from Herconomy, Kunal Upadhyay from IIMA Ventures, Mohamed Adel Khalil and Abdellatif Olama from Jumia, Natasha Jethnandani from Kaleidofin, Munir Duri and Hayat Abdulmalik from Kifiya, Josefina Cafferata Leloir from Manda/UILS, Christian Ruehmer from Q-Lana, Sasidhar Thumuluri from SubK, Daniel Maldonado from TransUnion, Liz Dank and Mariya Shappo from Uber, Cinthia Merlos from Vexi, and Angela Ngo Ndouga from Yellow Factoring. The team extends special thanks to the teams at Vexi in Mexico and Eshandi in Zambia for sharing anonymized customer-survey and borrower-level data, respectively, which enriched the analysis with additional empirical insights.

Contents

Foreword

Women entrepreneurs are key drivers of global economic growth, job creation, and income generation. While progress has been made in narrowing gaps in account ownership between men and women and expanding women's access to mobile money and digital payments, large gaps persist. According to the World Bank's *Global Findex Database 2025*, nearly 1.3 billion adults worldwide still lack a bank account, and far more remain without the credit histories needed to qualify for formal loans. Many women and small enterprises continue to operate outside the visibility of conventional financial systems, even though they have strong repayment capacity and entrepreneurial potential.

This report, *Cracking the Credit Code: Alternative Data and AI for Financial Inclusion*, builds new evidence on how we can help close these gaps. It shows that alternative data and artificial intelligence (AI) models are expanding access for first-time borrowers and underserved groups. It also reveals that if given the opportunity, women borrowers often outperform men and progress to higher loan amounts or better credit terms over time, signaling growing trust and repayment consistency. Firms that actively design products for women or embed inclusion principles into their models benefit too, by reaching new customer segments and increasing revenues and profitability.

The World Bank Group's Gender Strategy 2024-2030 is centered on expanding solutions to help women enter and stay in the labor force, which includes enabling women-led businesses to grow in number and size and improve productivity. Our targets are ambitious: by 2030, the World Bank Group has committed to provide 80 million more women and women-led businesses with capital, addressing a critical constraint to entrepreneurship growth.

I invite financial institutions, investors, policy makers, and regulators to use the insights in this report to help unlock the unique opportunities presented by alternative data and AI-driven lending, and build a more inclusive financial ecosystem that enables women entrepreneurs and small businesses to grow, create jobs, and contribute fully to their economies.

Robin Mearns
Director for Gender,
World Bank Group

Foreword

Expanding access to responsible finance remains one of the most powerful ways to drive inclusive and sustainable growth. Across emerging markets, the financing gap for women-led micro, small, and medium enterprises (WMSMEs) is estimated to amount to hundreds of billions of dollars at a minimum. Bridging this gap requires innovation that extends the reach of finance while maintaining trust, transparency, and accountability.

Cracking the Credit Code: Alternative Data and AI for Financial Inclusion provides timely insights into how new data sources and technologies are reshaping credit markets. Drawing on interviews and a global mapping of fintech firms, as well as borrower-level data from partners in Africa, Asia, and Latin America, the report shows how alternative data—from mobile money records to digital-platform transactions—can improve risk assessment and expand access for previously excluded borrowers. It also illustrates how these innovations are strengthening the infrastructure of financial ecosystems, creating new pathways for inclusion, and enabling lenders to better serve micro and small enterprises at scale. This is closely aligned with research from CGAP which shows that using granular, gender-disaggregated data enables lenders to better assess risk and expand lending to women without increasing portfolio risk.

The World Bank Group is committed to providing 80 million women and women-led businesses with access to capital from 2025 to 2030. Collectively, the World Bank and IFC are boosting access to finance for millions of individuals and businesses, strengthening financial infrastructure, and supporting new tools that deliver impactful financial products and services. We also partner with innovative banks, fintechs, and digital platforms to better serve micro, small and medium enterprises, women, and underserved groups.

Through programs like Banking on Women, IFC has already mobilized and invested more than $12.5 billion in 346 financial institutions in 84 emerging markets to catalyze financial services for women and women-owned enterprises. In addition, the World Bank has provided technical assistance and financing to support over 50 countries to establish digital infrastructure, interoperable fast payment systems, open finance schemes, and regulatory reforms that support new business models and new market entrants.

The message from this work is clear: alternative data and AI are already reshaping credit access, with particularly strong gains for women and underserved borrowers. The next frontier is intentionality: embedding fairness, sex disaggregated monitoring, robust data-governance and fairness frameworks, and cross sector partnerships. This, in turn, can help financial institutions reach untapped markets, improve portfolio performance, and unlock opportunities for millions of women entrepreneurs and small businesses to grow, create jobs, and contribute to shared prosperity.

Tarek Himmo
WBG Director,
Financial Services Sector

Sophie Sirtaine
WBG Director,
Financial Services Sector, and
Chief Executive Officer of CGAP


Photo: © Arne Hoel / World Bank

Executive Summary

Traditional credit scoring systems often exclude women and underserved borrowers—especially those without formal financial histories or collateral—and this limits their access to credit and ability to grow their businesses and improve their livelihoods. Demand for financing among micro, small, and medium enterprises (MSMEs) far exceeds supply, especially among those run and led by women (WMSMEs), as shown in Figure A (IFC 2024, Findex 2025, FICO 2024).

In response, a new generation of credit scoring models powered by alternative data and artificial intelligence (AI) is redefining credit scoring for borrowers who lack formal credit histories. Yet, the benefits of this shift are uneven—women, informal workers, and microentrepreneurs remain underrepresented in credit portfolios.

FIGURE A

How the financing gap disproportionally excludes women

5.7 trillion
The financing gap for MSMEs rose by $1.3 trillion from 2015 to 2019.

1.9 trillion
The financing gap for WMSMEs continues to widen.

3 billion
Men and women lacking adequate credit histories.

1.3 billion
Men and women without bank accounts.

Women are less likely to apply for credit.

More likely to have applications denied.

Receive smaller amounts or poorer loan terms.

This report aims to inform practitioners and policymakers about ways in which innovations in credit scoring can advance financial inclusion while ensuring fairness and protecting consumers. It was guided by three central issues:

1. **Market Trends:** How are alternative data and artificial intelligence used in credit scoring, particularly across emerging markets?
2. **Gender Inclusion:** To what extent do these models expand women's access to formal credit?
3. **Data and Model Design:** What types of data and modeling approaches underpin these systems, and how might they differ for women borrowers?

This analysis draws on interviews with more than 30 experts in fintech and credit scoring and a global mapping of 448 alternative-credit scoring firms, as detailed in Figure B. It also includes a literature review and borrower-and customer-level insights from the fintech companies Eshandi in Zambia and Vexi in Mexico. Combined, these capture diverse data use, business models, and regional operations, while highlighting emerging partnerships, regulatory developments, and challenges around fairness and transparency.

FIGURE B

Key dimensions used to map the credit scoring ecosystem

Region of Operation

Business Models & Services
Direct lending, embedded finance, credit scoring, and risk and loan management tools, for example.

Inclusion
Do firms refer to women, the underserved, or excluded borrowers in their vision or mission statement?

Funding Stage & Recent Raises

Loan Types Supported
Financing for MSMEs, personal loans, and buy now, pay later credit, for example.

Gender Representation in Leadership
Are founders or senior executives women?

Alternative Data Types & Sources
This covers data types, such as transaction patterns, and data sources, such as mobile wallets.

Five Major Findings

1. Alternative credit scoring is expanding and evolving

Alternative data-driven scoring models aim to broaden access to credit by going beyond traditional metrics such as repayment history and formal income documentation. Instead, they draw on behavioral, transactional, and digital footprint data, ranging from utility bill payments to gig work records and telecommunication indicators, such as call frequency, recharging of airtime, and mobile money activity, as shown in Figure C.

FIGURE C

Examples of alternative data

Call/ SMS metadata, app usage, operating system, geolocation

Digital transactions, utility payment, shopping behaviors

Social media connections, login methods, app usage

Artificial Intelligence plays a growing and varied role in these credit scoring models, as shown in Figure D. While some firms fully automate score generation using machine learning (ML), others limit AI to specific tasks such as transaction categorization or fraud detection, often due to regulatory, cost, or explainability constraints. The frontier is dynamic, and credit scoring is no longer a one-size-fits-all approach. It is becoming more modular, context-specific, and embedded into broader digital financial ecosystems.

FIGURE D

Varying levels of sophistication in models that analyze data

More analytical sophistication

Rule-based System

For example, in Uganda, the Grameen Foundation uses data from group savings ledgers and applies rule-based methods to assess eligibility.

Hybrid

This includes machine learning-based classification combined with manual adjustments. For example, Manda in Argentina uses natural language processing methods to parse transactions into categories, then applies rule-based thresholds.

Fully Automated

Fully automated pipelines include those used by Eshandi in Zambia that auto-approve nano loans if proprietary machine learning based scores from mobile money data are greater than 0.9.

Together, these models enable visibility for women and other underserved borrowers

Reduce discrimination in credit decision-making.

Offer new pathways for women and informal borrowers.

Increase consistency in driving data-based credit decisions.

2. Alternative scoring is taking off but growth is uneven across regions and market segments

Most mapped fintech firms operate in Europe and North America, but significant traction is visible across East Asia and the Pacific, Africa, and South Asia. Loan segments range from individuals borrowing $10 to $20 in emergency microcredit to full-fledged small and medium-sized enterprise (SME) financing of $50,000 or more.

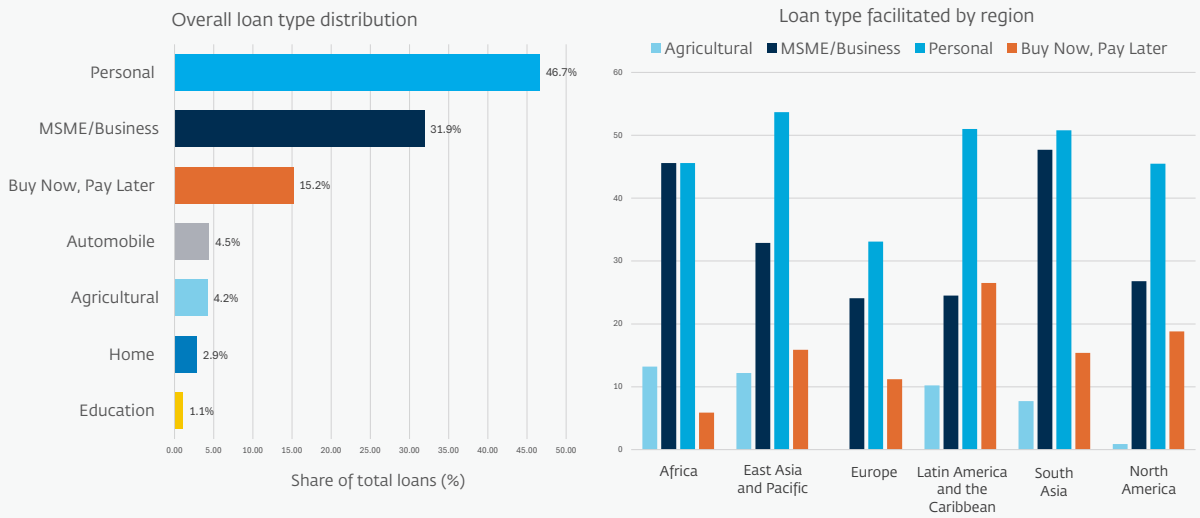
Transaction data drive most alternative-credit models, with AI used selectively rather than end-to-end. Transaction data—from mobile money, digital wallets, point-of-sale (POS) systems, and bank statements—is the most common input for non-traditional credit scoring models. In regions like Africa and South Asia, mobile money and digital wallet data are especially prevalent. Other examples of alternative data include utility payments, geolocation, device usage patterns such as phone metadata, and psychometric assessments. While most firms incorporate some form of AI or modeling sophistication—particularly to classify transactions or generate scores—fully autonomous models remain rare. Many providers continue to rely on rule-based or hybrid scoring approaches.

Context is key—transaction data may seem traditional in one setting and 'alternative' in another. The value lies in not only where data come from, but how they are processed, segmented, and deployed. Some firms rely on basic financial signals and reframe them through AI models to assess first-time borrowers. Others combine fragmented records and behavioral data into sophisticated scoring systems. Even relatively structured sources—such as bank statements or basic business records—can offer new insights to help evaluate borrowers with limited formal histories, thereby uncovering credit potential that conventional models often miss.

Personal and MSME loans dominate the product mix among mapped lenders, with buy now, pay later and agri-lending emerging as niches. As shown in Figure E, personal loans and MSME credit account for 47 percent and 32 percent of offerings, respectively. Just over one third of firms that develop their own credit scores also act as direct lenders, while 60 percent offer scoring and lending software to financial partners as part of a wider suite of products, such as fraud detection, loan management platforms, and financial advisory services.

FIGURE E

Loans offered by mapped fintechs, overall (left) and by region (right)

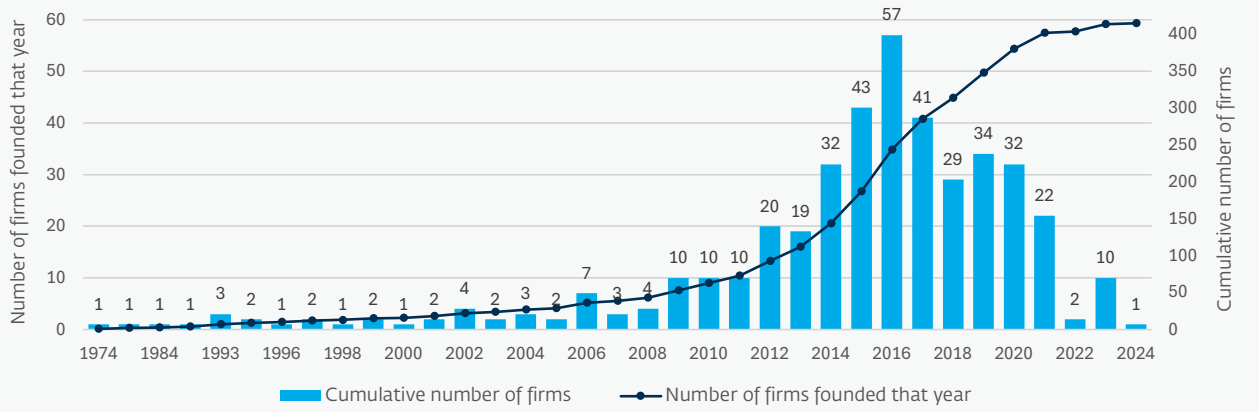


Source: Firms mapped in alternative credit scoring landscape analysis. N=448 firms
Note: Firms can appear in multiple categories as classifications are not mutually exclusive.

The ecosystem is young and dynamic, with most mapped firms concentrated at early-venture stages. As shown in Figure F, over 75 percent of mapped firms were founded in the last decade, and over 70 percent of funded firms raised capital in the three years prior to this report's publication. Series A–C is the most common funding stage. MSME lenders tend to attract more funding. Regionally, North America and Europe dominate mature and well-funded segments, while Africa has more unfunded firms.

FIGURE F

Year in which mapped fintech firms were founded



Source: N=415 mapped firms with available founding year data.
Note: Bars represent the share of newly-founded firms in that year. The dotted line reflects the total number of firms over time.

Alternative-credit scoring methods complement—but do not replace—traditional methods by helping borrowers gain access to greater amounts of credit. Alternative data and AI-based scoring models increasingly serve as bridges rather than substitutes for traditional lending. For many first-time borrowers—especially women and informal workers—they offer a pathway into the formal credit system by lowering entry barriers and establishing a verifiable repayment history. For those already in the system, these models can help unlock higher loan amounts, better terms, or faster approval cycles by capturing additional indicators of reliability and cash flow. In doing so, they enable lenders to identify creditworthy borrowers who may have been previously overlooked—turning exclusion into opportunity and inclusion into growth.

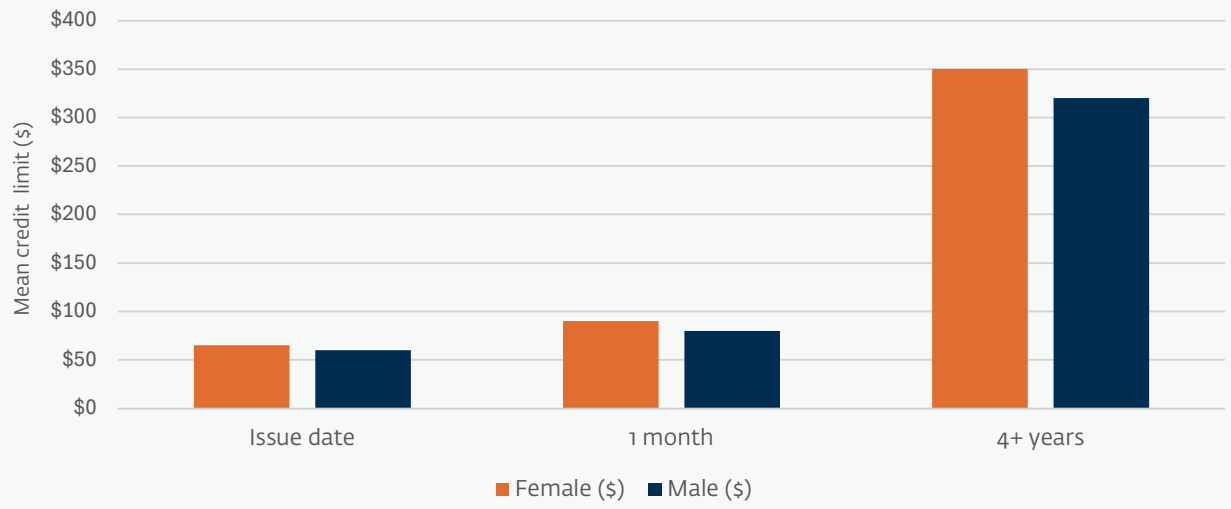
3. Gender-inclusive potential is real

When designed and deployed inclusively, alternative data and AI-driven models can meaningfully expand women's access to credit. This finding applied to almost every firm interviewed for this report as well as those highlighted in case studies. In many cases, women benefit when new data sources redefine what is considered 'creditworthy,' especially when these models capture cash flows or behavioral indicators that traditional systems overlook. Female borrowers are often overrepresented among successful repayment cases, receive better loan terms, and are already being reached at scale.

Evidence from multiple markets shows women's reliable repayment habits and growing credit confidence. In India, Kaleidofin has facilitated over seven million loans, largely to women. Eshandi, a fintech in Sub-Saharan Africa, has disbursed nearly one million loans to women. Borrower-level analysis of Eshandi's mobile money–based scoring model shows that women score slightly higher on machine learning–generated credit scores and are more likely to receive repeat loans, reflecting their consistent repayments and stable financial behavior. In Cameroon, Yellow Factoring reports higher repayment reliability and lower interest rates among women borrowers. In Mexico, survey data from digital credit card provider Vexi revealed that women use digital credit for business expenses more often than men and, over time, receive higher average credit limits, as shown in Figure G. This suggests that fintech-driven products can strengthen both access and confidence in using formal credit.

FIGURE G

Mean credit card limit, by gender and tenure

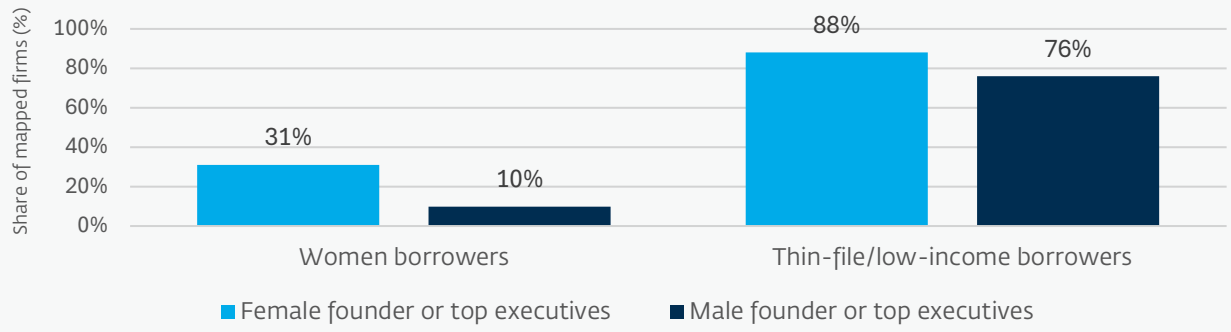


Source: Customer survey data of over 3000 Vexi borrowers in Mexico
Note: $1=18.5 Mexican pesos

Despite reliable repayment habits and other outcomes, most models are gender-neutral by design—and inclusion gains are largely incidental. Gender is rarely used as a predictive variable, and only 12 percent of firms publicly reference women on their website, in their outreach, mission or vision statement, or model framing, as shown in Figure H. While exclusion can sometimes stem from fairness concerns or lean model design, it also reflects missed opportunities. The outcomes suggest that well-designed inclusive scoring can be both effective and commercially viable.

FIGURE H

Share of mapped firms that publicly reference women or low-income borrowers



Source: Mapped firms' public communications. N=448 firms
Note: Thin-file borrowers have little or no formal credit history and include first-time borrowers, informal sector workers, and those without regular access to banking services.


Photo: © Daniella Van Leggelo-Padilla / World Bank

4. Pathways for women's inclusion are emerging but they require intentional support

While AI and alternative data can improve access for woman and underserved groups, they also introduce new risks of bias and exclusion. Biases in training data—such as gaps in women's digital footprints, underrepresentation in credit histories, or the mismeasurement of informal income—can reinforce exclusion if left unaddressed. Proxy variables such as education level, geography, or device type may unintentionally entrench discrimination, even when gender is omitted from models. Explainability and fairness checks are therefore critical to ensure that algorithms expand access.

Privacy and consent risks are particularly pronounced where women rely on shared phones, have limited digital literacy, or operate in environments with weak data-protection norms. Without safeguards, models that depend on sensitive or behavioral data may expose women to harm, misclassification, or loss of agency over their information.

Recent advances in gender-intentional design show that it can deliver both equity and commercial returns. Fintechs such as Kaleidofin and Eshandi are beginning to test fairness metrics and sex-disaggregated reporting, and the results often reveal that women's repayment performance is as strong or stronger—even where data appear 'thin.' Centering gender in model design and oversight can help maximize benefits for both firms and consumers, but scaling inclusion will require coordinated action among firms, regulators, and data partners.

5. Partnerships and regulation are shaping growth

Alternative credit scoring firms are partnering with banks, credit bureaus, mobile network operators, platforms, and others, to scale access, share risk, and legitimize new scoring approaches. For example, dominant players like Equifax and TransUnion have acquired niche analytics providers, while banks and mobile operators are using data partnerships to develop embedded credit tools. In India, Kaleidofin has teamed up with established lenders like Federal Bank to scale microloans, while telecom players MTB in Uganda and Safaricom in Kenya have launched credit products using their own data. Still, collaboration remains uneven. During interviews, fintechs noted that traditional institutions, especially those that are larger or more risk-averse, have been slow to adopt or scale alternative models due to concerns ranging from regulatory compliance and data privacy to the opaque nature of AI-based scoring methods.

Regulatory and operational challenges remain significant, and data privacy, explainability, and gaming risks must be addressed. Rapid innovation in AI and data use brings its own risks. For example, while some alternative credit score providers partner with licensed lenders to share compliance burdens, others rebrand their tools as analytic services rather than credit scoring to sidestep regulatory scrutiny. Inconsistent data privacy rules, limited consumer consent mechanisms, and varying explainability requirements across markets constrain how models are built and deployed. Without adequate safeguards, model integrity and borrower trust can be undermined by opaque algorithms, unverified data sources, or potential 'gaming' of digital signals. These challenges underscore the need for clear guidance that balances innovation with consumer protection and responsible data use.

Recommendations:
Enabling Responsible Innovation and Inclusive Credit at Scale

Alternative data and AI-based credit models have proven potential to expand financial access—especially for women—but scaling them responsibly requires deliberate policy, partnerships, and design choices. Regulators, financial institutions, and industry actors are encouraged to tackle the following priorities:

1. Create regulatory sandboxes and AI-testing environments to build confidence and oversight. These regulator-supervised frameworks enable companies to test new algorithmic models in real world conditions—evaluating accuracy, loss estimation, and fairness before rolling them out in the open marketplace. Successful pilots in markets such as Kenya and the Philippines show that empirical evidence from sandboxes can accelerate adoption while also protecting consumers.

2. Embed fairness testing and bias audits across model lifecycles. Encourage all financial institutions and fintechs to embed fairness checks—such as approval-rate parity or model explainability benchmarks—into model development and retraining. These processes help detect hidden biases arising from proxy variables, such as location or education, that may disadvantage women or low-income borrowers.

3. Encourage sex-disaggregated and intersectional reporting to inform and improve model design. Encourage lenders and credit scorers to collect and report gender-disaggregated outcomes across the credit lifecycle, including applications, approvals, loan size, and repayments. Linking this data with other demographic factors such as geography or employment type can uncover structural barriers and support evidence-based interventions that strengthen fairness.

4. Encourage traditional lenders to integrate alternative data responsibly and transparently. Banks and credit bureaus could partner with fintech innovators to layer alternative data—such as transaction or mobile-money records—onto traditional scoring frameworks. Doing so can expand reach without compromising prudential standards, as long as explainability, consent, and audit trails are also in place.

5. Promote cross-sector data partnerships to expand the visibility of underserved borrowers. Data-rich entities—such as mobile network operators, utilities, and e-commerce platforms—can play a catalytic role by enabling data-sharing with users' consent through secure application programming interfaces (APIs) that allow different software applications to communicate with each other. Structured partnerships between fintechs, incumbents, and regulators can reveal data-rich but invisible borrower segments, especially women and informal entrepreneurs. Figure I illustrates how open-finance frameworks and responsible AI can enable more inclusive, consent-based credit ecosystems.

6. Build consumers' digital credit skills and data-literacy to ensure inclusion translates into empowerment. Complement financial innovation with user awareness programs on data rights, consent, and digital credit management, especially in low-trust or shared-device contexts where women face heightened risks from data misuse or over-indebtedness.

7. Support open-finance frameworks that combine interoperability, access, and consent. When paired with algorithmic learning, interoperable data-sharing systems can allow lenders to assess borrowers more holistically while preserving consent and privacy. Clarifying how sensitive attributes such as gender can be shared between platforms, lenders, and regulators—while balancing anti-discrimination rules and fairness-testing—is essential for sustainable and responsible innovation, and will enable AI to learn from more representative datasets.

Photo: © AdobeStock


Photo: © Dominic Chavez / World Bank

1. Introduction

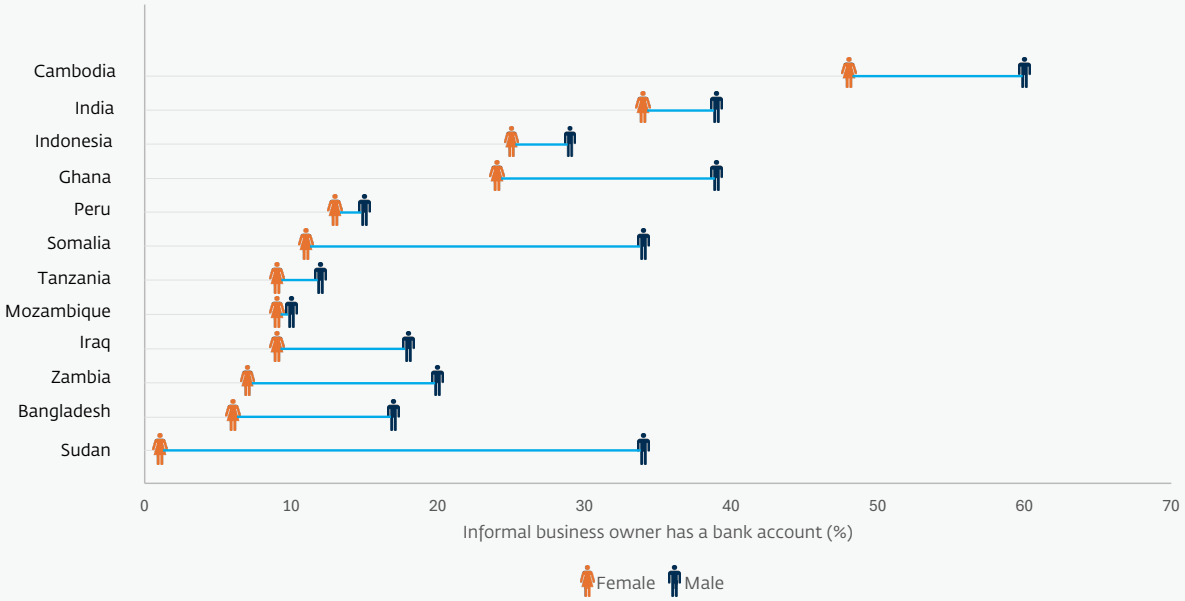
CREDIT ACCESS, GAPS, AND THE PROMISE OF ALTERNATIVE CREDIT SCORING

Credit is essential for inclusive economic development, yet despite major advances in digital access, approximately 1.3 billion adults remain unbanked globally while three billion people lack sufficient credit history to access financing**—and women are overrepresented in these groups** (Demiguruc-Kunt et al. 2022, Klapper et al. 2025). These gaps persist across the entire credit lifecycle—women are less likely to apply for credit, they are more likely to be denied when they do apply, and even when approved they tend to receive smaller loans with less favorable terms and shorter periods (Brock and De Haas 2023, Montoya et al. 2024, IFC 2024). This limited access to credit restricts women's capacity to invest in businesses, absorb shocks, and build financial resilience.

Traditional credit scoring systems systematically exclude women—whether they be individuals, entrepreneurs, or informal workers—especially in emerging markets. Longstanding credit scoring models rely on formal financial histories, collateral, and stable incomes, but these criteria disproportionately exclude women as they are less likely to own property that can be used as collateral or have formal employment (Klapper 2024, Demirguc-Kunt et al. 2022). They are also less like to participate in formal banking systems. As illustrated in Figure 1, bank account ownership is low among all owners of informal enterprises, and in most countries, women business owners are even less likely to have a bank account for that business. The depth of these gender gaps varies by country. For example, there is a larger gender gap in reliance on informal funding that favors men in Bangladesh, Mozambique, Sudan, and Zambia.

FIGURE 1

Gender gaps in bank account ownership among owners of informal enterprises

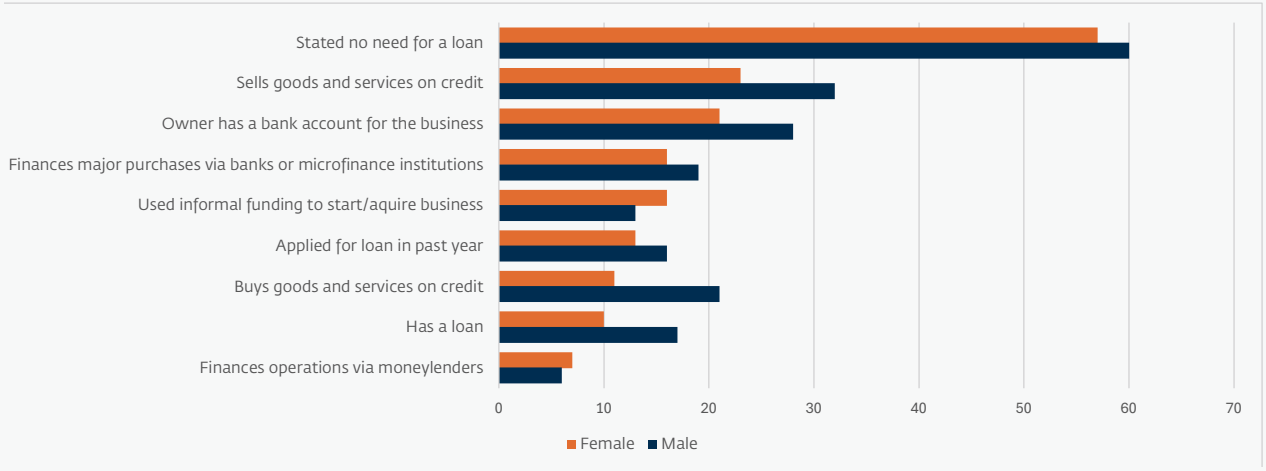


Source: Based on data from the World Bank's Informal Enterprise Survey (2022), covering selected emerging market economies.

Beyond account ownership, women informal entrepreneurs also differ from men in how they engage with credit, financial services, and business financing, as shown in Figure 2. These patterns point to both demand- and supply-side constraints shaping women's financial behavior and opportunities.

FIGURE 2

Experience with banking and business finance among informal enterprises, by gender of owner



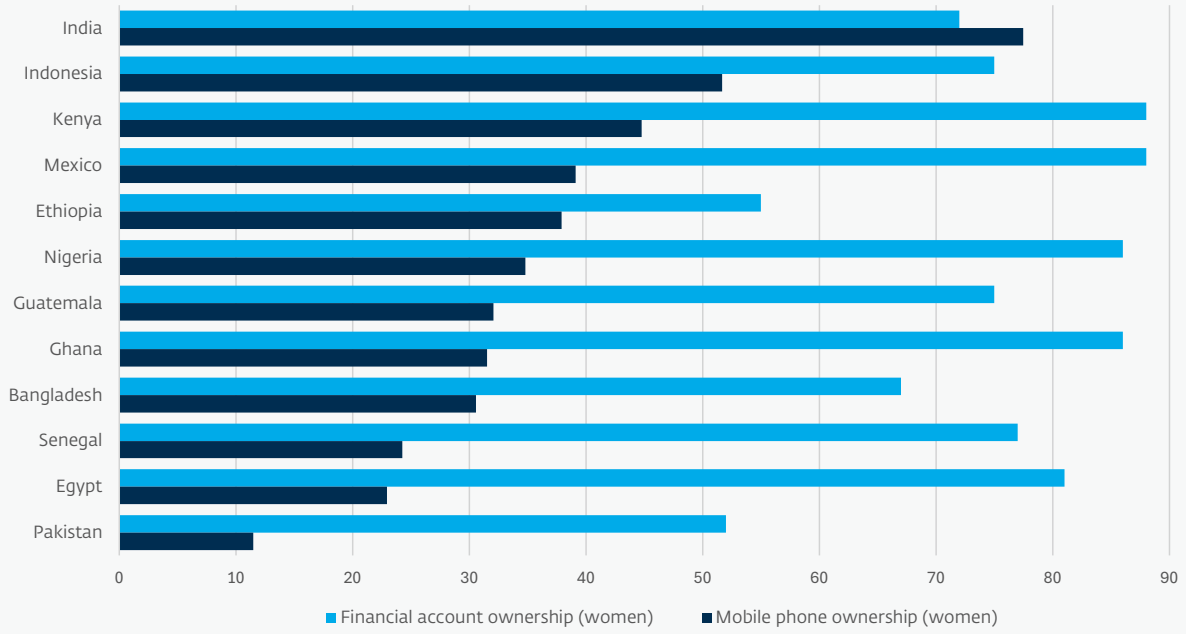
Source: Based on data from the World Bank's Informal Enterprise Survey (2022), covering the emerging market economies of Bangladesh, Cambodia, Ghana, India, Indonesia, Iraq, Mozambique, Peru, Somalia, Sudan, Tanzania, Zambia and Zimbabwe.

Exclusion from credit is not only an issue of fairness, it is also a missed business opportunity. Women borrowers often demonstrate better repayment behavior than men, and lenders frequently cite women's stronger repayment discipline and financial caution as key drivers of portfolio stability (IFC 2024). Despite this, women-owned and run micro, small and medium enterprises face a financing gap of almost $2 trillion (IFC 2025).[1] Gig workers, smallholder farmers, and low-income women also struggle to access credit (Deshpande 2020, Miller et al. 2020).

Women's fast-growing digital footprints offer untapped opportunities for alternative credit models. The last decade has seen a dramatic rise in mobile phone ownership, digital payments, and platform usage, even in low-income and rural settings (Jeffrie 2024). For example, mobile money penetration in Sub-Saharan Africa has reached 33 percent of the adult population (Demiguruc-Kunt et al. 2022). In most countries, digital usage and mobile phone ownership among women significantly exceeds formal financial activity or bank account ownership, as illustrated in Figure 3. In the Middle East and North Africa, 65 million women are unbanked but own mobile phones, while in South Asia and West Africa, women engage extensively with e-commerce, ride-hailing, and informal digital marketplaces (Jeffrie 2024, Shahid 2022). This offers great potential to leverage mobile-based data to expand access to credit.

FIGURE 3

Mobile phone and financial account ownership among women



Sources: World Bank Global Findex (2021), GSMA's Mobile Gender Gap Report (2023).

[1] IFC defines micro enterprises as those with fewer than 10 employees and under $100,000 in assets, while small businesses have 10-49 employees and $100,000-$3 million in assets for small firms, and medium-sized enterprises have up to 50-300 employees with assets between $3-15 million.

Photo: © Vincent Tremeau / World Bank

Alternative data—ranging from mobile call patterns, airtime top-ups, and app usage, to utility payments, e-commerce activity, psychometric responses, and geolocation metadata—offers new pathways for inclusion (Jagtiani and Lemieux 2018, Berg et al. 2020, World Bank 2025)**.** Alternative data refers to any non-traditional, non-bureau information used to assess creditworthiness, and when leveraged to capture real-time economic behavior and alternative financial footprints, it can provide a more accurate and comprehensive assessment than traditional metrics. This creates visibility for borrowers traditionally excluded from formal credit systems, especially the growing number of women and underserved groups who use digital channels.

Artificial intelligence and machine learning (ML) enhance the value of alternative data. These technologies, defined in Box 1, can analyze large, complex data sources at scale to generate dynamic scores, detect risk and fraud patterns, and personalize loan offers. Artificial intelligence and machine learning models outperform traditional scores in predicting default risk, especially for so-called thin-file borrowers (Björkegren and Grissen 2020).[2] Compared to manual underwriting—where gender biases in approval rates, loan terms, and collateral demands are well documented—algorithmic tools can offer more consistent and data-driven decisions.

Box 1. Artificial Intelligence and Machine Learning

In this report, the term 'artificial intelligence' is used primarily in connection with credit scoring decision systems that rely on machine learning. Most of the models described in this report use machine learning techniques such as random forests, gradient boosting, or regularized logistic regression to identify nonlinear relationships between alternative data and repayment outcomes. A smaller share use deep learning methods, while generative or agentic AI is currently limited to ancillary applications like data extraction or customer engagement. Throughout the report, 'AI-driven' encompasses a spectrum from simple algorithmic automation to advanced machine learning.

2 Thin-file borrowers are individuals with little or no formal credit history, making it difficult for traditional credit scoring systems to assess their creditworthiness. This includes first-time borrowers, informal sector workers, and those without regular access to banking services.

Together, alternative data and AI can be used to design and scale alternative credit models that reshape access to credit, but these benefits are not guaranteed. As discussed in the final section of this report, poorly designed models or biased data can reproduce or even amplify existing disparities. However, when used responsibly, these tools can reduce bias and expand access without compromising performance, as shown in Figure 4.

FIGURE 4

How AI and alternative data can benefit both lenders and borrowers

Value for Lenders

Stronger risk management
Improves credit risk prediction, detects fraud, and reduces human error.

Greater operational efficiency
Streamlines applications and underwriting, lowering costs and enabling scale.

Faster, tailored products
Real-time scoring enables quicker decisions and more personalized credit offers.

Value for Borrowers

Broader access to credit
Women and underserved groups face fewer rejections, less collateral, and better-fit products.

Less discrimination
Well-designed models reduce bias and improve fairness for marginalized borrowers.

Faster, more customized credit
Borrowers receive quicker approvals and credit aligned with their needs.

Objectives and Research Questions

Alternative data and artificial intelligence are transforming credit markets, yet their role in expanding access for underserved borrowers—especially women—remains underexamined. This report seeks to fill that gap by analyzing how alternative data and AI-based scoring models are being designed and deployed to advance inclusive finance in emerging markets. The report addresses three key areas:

1. Market Trends: What are the current market trends in the use of alternative data and AI for credit scoring, particularly in emerging economies?

2. Gender Inclusion: How do these models influence women's access to and experiences with formal credit?

3. Data and Model Design: What types of alternative data are currently used to predict credit approvals and repayment, and how might these differ for women borrowers?

The report aimss to synthesize emerging evidence across literature, firm-level data, and market practice to inform both practitioners and policymakers on how innovation in credit scoring can advance financial inclusion, while maintaining fairness, transparency, and consumer protection.

Methodology and Sources

This report combines three complementary sources of evidence—a mapping of the global alternative credit scoring landscape, firm-level insights, and literature and market reviews.

Global Landscape Mapping: The team created a database of over 400 firms that develop credit-scoring models using non-traditional data—either for their own lending or as tools for banks and other financial service providers. To capture variations in data types, geographies, and business models, the mapping drew on multiple public and proprietary sources, including market intelligence database Tracxn, the Cambridge Centre for Alternative Finance, SME Finance Forum rosters, and web-based searches.

Firm-level Insights: Analysis draws on interviews with over 30 stakeholders, including 15 fintechs, along with case studies and two new datasets that provided evidence and perspectives at borrower and customer levels. In Zambia, the fintech Eshandi shared anonymized data from over one million microloan applications scored using mobile-money data and machine-learning models. In Mexico, Vexi shared survey data from over seven thousand fintech customers, highlighting gender patterns in credit use, business financing, and confidence in applying for future credit.

More information on the report's methodology and sources is available in the appendices.

Scope and Limitations

This report focuses on digital credit-scoring models that use non-traditional data to extend credit to individuals and MSMEs in emerging markets. While the landscape mapping offers a broad snapshot of this ecosystem, it reflects firms with greater public visibility and documentation. Smaller or early-stage players may therefore be under-represented. The analysis does not evaluate model performance. Firm-level information is drawn from publicly available sources, meaning that the extent or intensity of alternative data use within each model cannot always be verified. Instead, it combines market intelligence, firm interviews, and selected case studies and borrower-level data to illustrate how alternative data and AI-based credit models are evolving in practice and influencing women's access to finance.

Photo: © AdobeStock


Photo: © AdobeStock

2. WHAT ARE ALTERNATIVE DATA-DRIVEN CREDIT SCORING MODELS?

As noted, alternative data refers to non-traditional sources that capture individuals' economic behaviors or digital footprints and go beyond standard credit bureau metrics to assess creditworthiness. Traditional credit scoring models are typically effective in developed markets where borrowers have formal credit records, including repayment history, amounts owed, length of credit history, new credit, and their credit mix.[3] However, in emerging markets these models can systematically exclude individuals without formal financial footprints, especially women and informal workers (Jagtiani and Lemieux 2018, Chen and Faz 2015, Anand 2022, Vidal and Barbon 2019).[4] Alternative data can narrow that gap by leveraging a wide range of real-world signals beyond traditional metrics, as shown in Figure 5.

FIGURE 5

Examples of alternative data

Call/ SMS metadata, app usage, operating system, geolocation

Digital transactions, utility payment, shopping behaviors

Regional economic indicators

Psychometric credit assessments

Social media connections, login methods, app usage

Business location, inventory management, local market activity

Farm characteristics, weather, satellite, soil conditions, crop type

3 For example, a typical FICO credit score is calculated based on the following approximate weightings: payment history (35 percent), amounts owed (30 percent), length of credit history (15 percent), new credit inquiries (10 percent), and credit mix (10 percent). These inputs rely on formal financial data, such as credit card usage, loan repayment records, and credit account age. Data is often unavailable or incomplete for individuals without access to traditional financial services.

4 Also, see Appendix B for a comparison between traditional and alternative data-driven credit scoring.

These data may be grouped into two broad and overlapping categories: behavioral data and transactional data. *Behavioral data* captures how individuals interact with digital tools and services—for example, mobile phone usage patterns such as call, text, and message frequency and top-up behavior, app activity, geolocation trails, social media engagement, or responses to psychometric surveys. These signals are often used to infer traits like reliability, routine, or intent to repay. *Transactional data* relates more directly to economic behavior and includes mobile money transactions, digital wallet balances, utility bill payments, airtime spending, or purchasing histories on e-commerce, ride-hailing, and delivery platforms. These proxies can offer insight into income flows, consumption stability, and repayment capacity.

In practice, these are not the only types of data, and the distinction may not always be clear. The same underlying source—such as mobile-money or e-commerce data—can yield both transactional indicators of frequency or volume and behavioral indicators, such as consistency, volatility, and seasonality, as shown in Table 1. For example, frequency of airtime top-ups can reflect behavioral consistency but also signal disposable income. Firms typically report data in ways that reflect this fluidity rather than categorizing them rigidly.

TABLE 1

Examples of behavioral and transactional data

	Data Source	Typical Use
Transactional	Mobile-money inflows and outflows, card or digital wallet payments, airtime purchase history, utility bills.	Proxy for cashflow, income stability, and/or repayment capacity.
Behavioral	Phone-use patterns, app login frequency, rating/review scores; psychometric test results, on-time bill payments.	Proxy for reliability, discipline, and/or business continuity.
Overlapping	Merchant transaction history (frequency=behavioral; volume=transactional), mobile-money usage (amounts=transactional; consistency=behavioral).	Context-dependent.

As mobile phone and internet penetration increase across emerging markets, these alternative signals are becoming more widely available and better suited to assess creditworthiness in underbanked populations.

Different market environments also shape how traditional and alternative data are combined. In practice, most fintechs operate along a spectrum—from data-scarce markets that rely more on alternative sources to mature systems where alternative data complement traditional credit records, as Shown in Table 2. It should be noted that mixed systems are becoming more common.

TABLE 2

How different markets use traditional and alternative data in credit scoring

Market Context	Typical Data Mix
Traditional Data-dominant	Traditional credit bureau or bank records are the foundation of credit scoring. Alternative data, such as app activity, transactions, and mobile money, provide behavioral insights or help assess thin file borrowers but do not replace traditional methods.
Data-scarce Environments	Alternative data serve as the primary inputs for credit scoring. Where credit history is unavailable, mobile money, telecommunications metadata, or platform data act as proxies for repayment capacity.
Mixed Systems	Firms dynamically integrate both sources, using bureau data where available and alternative data where needed. In many models, bureau data serve as verification flags or default risk filters, while other data drive scoring for clients without established credit histories. The goal is not to replace traditional data, instead it aims to strengthen inclusion by not penalizing borrowers who lack formal records.

Alternative data models that combine multiple data sources often show higher predictive performance than traditional scores alone, especially for borrowers with limited credit histories. For example, one large study in Mexico showed that integrating telecommunications and app-usage data improved default prediction or area under the curve (AUC) by 12 percentage points over the baseline (Di Maggio et al. 2022).[5] Similarly, metadata from telecommunications companies has been found to outperform traditional models in predicting defaults and has a higher area under the curve (Björkegren and Grissen 2020, IPA 2024). Psychometric scoring tools in Ethiopia perform well for micro and small business owners and have similar or better repayment predictive power than traditional scores (Alibhai et al. 2022). In Germany, digital footprint variables such as email address structure and device type were found to outperform traditional bureau data (Berg et al. 2020). Digital transaction data from delivery apps reliably predicted creditworthiness among first-time borrowers in Mexico (Chioda et al. 2024). Finally, grocery purchase data helped to expand access for individuals without formal credit scores, offering predictive value that is especially strong for new-to-credit borrowers (Lee et al. 2024).

Alternative scoring models vary widely in analytical sophistication—from rule-based systems to advanced AI/ML pipelines. At one end of the spectrum, rule-based systems apply expert-defined thresholds, for example, approving a loan if airtime top-ups exceed $5 per month for six months. These are common among micro-lenders and early-stage platforms. At the other end are machine learning-based models—such as random forests, gradient boosting, or neural networks—that use high-dimensional alternative data inputs to predict default risk. Hybrid systems lie in between, combining algorithmic insights with human or rule-based judgment.

A firm's choice of model architecture often reflects regulatory constraints, data availability, and technical capacity as much as pure predictive performance. Table 3 reflects the range of systems used by firms studied in this report.

[5] The area under the curve (AUC) is a common metric used to evaluate the performance of credit scoring models. It measures how well a model distinguishes between 'good' and 'bad' borrowers. A higher AUC indicates better predictive accuracy, with a score of 1.0 representing perfect prediction, while 0.5 is no better than random guessing.

TABLE 3

Examples of alternative scoring models

Model/System	Description	Example
Rule-based	Applies expert-defined thresholds or rules set manually, such as requiring a minimum level of mobile-wallet activity. Neither data processing nor scoring involves machine learning. This is common among smaller lenders and financial service providers using pre-set eligibility conditions.	The Grameen Foundation's financial service partners use rule-based models, where data are collected and scored manually without machine learning methods.
Hybrid	Elements of human or rule-based oversight are combined with analytical modeling. Machine learning techniques may support risk segmentation, feature generation, or behavioral scoring, but final approval decisions involve manual checks or policy thresholds.	In Latin America, Manda employs a hybrid model where algorithmic insights inform rules applied by partner lenders.
AI/ML-driven	Uses machine learning, such as gradient boosting, random forests, or neural networks, to predict repayment risk from large volumes of behavioral and/or transactional data. These models rely on continuous feedback loops to refine accuracy over time. They may incorporate traditional and alternative variables but operate autonomously for some types of lending decisions.	In Zambia, Eshandi uses adaptive machine learning-based scores that update daily based on mobile-money transactions and repayment outcomes.

Alternative Data Spotlight

School fees as credit signals: Regular, documented school-fee payments can reveal household income stability and prioritization of financial commitments, serving as a proxy for repayment capacity.

Photo: © AdobeStock


Photo: © Maria Fleischmann / World Bank

3. MAPPING THE ALTERNATIVE CREDIT LANDSCAPE: MARKET TRENDS, GROWTH, AND CASE STUDIES

Alternative credit scoring is a fast-growing segment within the broader credit scoring market, with fintechs and digital lenders driving momentum. While alternative credit scoring is a distinct and evolving segment, its growth is tightly interlinked with broader trends in AI-driven credit assessment, fintech innovation, and the shifting dynamics of digital lending. Together, these trends underscore the expanding market opportunity and the urgency of inclusive, data-driven solutions.

Currently, growth is largely focused in North America, Europe, and developed regions where AI infrastructure and regulatory clarity are stronger, but emerging markets are beginning to catch up. In Africa and the Middle East, for example, credit-scoring spend is expected to grow by 117 percent to approximately $3.7 billion through 2028 (Juniper Research 2023). While headline growth figures are global, emerging markets represent the fastest-growing segment and offer significant potential for inclusion.

The global credit scoring market is projected to grow by 67 percent to $44 billion by 2028.

AI-driven credit scoring is expected to have a 26 percent compound annual growth rate through 2031.

The value of AI in the fintech market is forecast to expand from $14.2 billion in 2024 to $76.2 billion by 2033.

Fintech lending and big-tech credit grew 37 percent from 2020-2021.[6]

Alternative Data Spotlight

Satellite imagery and geo-tagged data: Data such as road quality, electricity access, or local business activity can reveal a borrower's socio-economic environment. These environmental indicators help lenders understand community-level resilience and the stability of livelihoods.

6 Infographic sources: Malone 2023, InsightAce Analytic 2024, Market.US 2024, RiskSeal 2024, Gambacorta et al. 2024

Mainstream credit bureaus and financial institutions are slowly incorporating alternative data into their models. Leading bureaus such as Experian, TransUnion, and Equifax have introduced new scoring products that integrate non-traditional data (World Bank 2025). Experian Boost, for example, allows borrowers to incorporate utility and telecom payment histories into their credit files. TransUnion's Credit Vision suite leverages data on mobile usage, rent payments, employment, geolocation, and trade credit. Equifax's collaboration with FICO developed FICO Score XD, which uses utility and telecom data to score previously unscorable individuals. Despite this progress, industry stakeholders caution that many financial institutions continue to underutilize alternative data sources. A survey of 125 lending decision-makers in the United States found that while 90 percent believed that alternative data could help them approve more qualified borrowers, only 43 percent were actually using alternative data (Nova Credit 2024).

Beyond credit scoring, the broader alternative lending segment has also gained significant momentum, including digital-first models for personal loans, trade finance, invoice factoring, and mortgage lending. According to Tracxn, this segment includes over 17,000 firms globally, of which nearly 5,000 have received funding.[7] These firms raised a combined $12.5 billion in the last two years. Alternative credit scoring models often underpin lending decisions across this broader ecosystem, and this underscores their growing relevance in expanding access to finance. While headline funding and growth figures are skewed toward mature markets, the alternative lending segment is rapidly expanding in key emerging markets such as India, China, Nigeria, Indonesia, Mexico, Kenya, and the Philippines.

Alternative Data Spotlight

Typing and form-completion patterns: The ways in which applicants fill out digital forms can reveal confidence and reliability. Time spent, number of corrections, or hesitation patterns on specific fields help flag whether responses seem consistent or uncertain. For example, steady, deliberate typing can signal truthful reporting, while frequent edits or rushed entries can point to data inconsistencies.

Landscape Analysis and Insights from Stakeholders on the Evolving Ecosystem

To understand how these innovations are playing out in practice, this section draws on a landscape analysis of over 400 firms using or offering credit scoring tools based on alternative or non-traditional data, complemented by stakeholder interviews and firm-level case studies. It documents the diversity of business models, regional footprints, use of alternative data and AI, and the extent to which inclusion is integrated into scoring and lending practices, particularly in relation to women and underserved borrowers. As shown in Figure 6, the analysis identifies key firm types across geographies, maturity levels, and alternative data sources, and illustrates these through case studies of selected providers.

7 Tracxn (2025) last accessed on 7th July 2025

FIGURE 6

Key dimensions of landscape mapping

Region of Operation

Business Models & Services
Direct lending, embedded finance, credit scoring, and risk and loan management tools, for example.

Inclusion
Do firms refer to women, the underserved, or excluded borrowers in their vision or mission statement?

Funding Stage & Recent Raises

Loan Types Supported
Financing for MSMEs, personal loans, and buy now, pay later credit, for example.

Gender Representation in Leadership
Are founders or senior executives women?

Alternative Data Types & Sources
This covers data types, such as transaction patterns, and data sources, such as mobile wallets.

The dataset is drawn from multiple lists, including lists created by snowballing through manual web-based searches and searches based on large language models (LLM). For each firm, information was gathered from publicly available sources such as company websites, LinkedIn pages, and news articles, as well as proprietary market intelligence platforms. Firms were included if they incorporated non-bureau data such as mobile money, social media, or platform activity into their credit scoring models or lending decisions.[8] Many firms spanned multiple categories, combining several types of data, operating across regions, or adopting hybrid approaches to scoring and lending. While the resulting database covers more than 400 firms, the exercise should be viewed as a broad snapshot rather than an exhaustive census. Visibility is influenced by the availability of public data and online footprints, meaning that smaller firms—particularly early-stage players or those without dedicated websites—may be underrepresented.[9]

Consultations and interviews were conducted with fintech lenders and scoring solution providers operating in over 20 countries across all emerging market regions. Together, they offer a grounded understanding of models used, populations served, and challenges faced in operationalizing alternative credit scoring. As noted previously, it also brings together **micro-level evidence from two datasets**: borrower-level data from the fintech Eshandi in Zambia and customer-survey data from Vexi in Mexico. These provide complementary insights into how alternative data and AI models affect credit access and outcomes, especially for women borrowers.

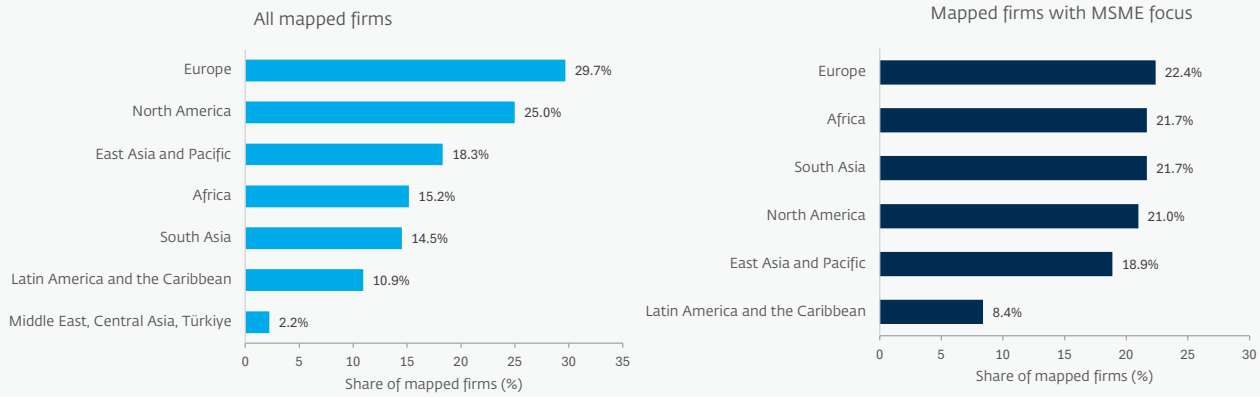
Alternative data-based credit scoring firms are concentrated in high-income markets but also show

8 It is important to note that characteristics such as region of operation, alternative data sources, business models, and inclusion strategies are not mutually exclusive.
9 Firms were included if they directly used non-traditional data in their own credit scoring or underwriting, or if they offered such models as a service to other financial institutions. Firms that only used alternative data-based scores developed by others, such as banks, fintechs, or financial service providers acting solely as distribution or lending partners, were excluded, as were large credit bureaus or banks unless they operated distinct subsidiaries focused specifically on alternative data-driven products. More details about the methodology are available in Note A3 in the Appendix.

a strong and growing presence across emerging economies. In Figure 7, the graph on the left provides a regional breakdown of mapped firms scoring or lending using non-traditional data, although it should be noted that firms may operate in multiple regions. Of these, almost 30 percent operate in Europe and a quarter operate in North America, while firms operating in Latin America and the Caribbean and the Middle East and Central Asia amount to 11 percent and 2 percent respectively. The graph on the right shows the region of operation for mapped firms that focus specifically on lending to the MSME segment.

FIGURE 7

Regional distribution of mapped firms



Source: Firms mapped in alternative credit scoring landscape analysis. N=448 firms.
Note: Firms can appear in multiple categories, as classifications are not mutually exclusive.

Alternative Data Types and Sources

Firms use a range of alternative data types tailored to their operating contexts. For example, Eshandi uses school fee payments as a proxy for responsible behavior and gambling history as a red flag. Vexi uses typing patterns for online forms to infer behavioral integrity. Agrifi in India, and Akello Banker in Uganda are agri-focused platforms that use satellite imagery, remote sensing, and transaction data to generate farm-level risk assessments. In Nigeria, Gigmile builds credit histories using employment and earnings data from gig platforms, while KarmaLife in India uses work histories and client feedback from blue-collar employment to create a proprietary KarmaScore for payday lending. AwanTunai in Indonesia, ChapChap in Uganda, and KompAI in Ecuador draw on a combination of invoice data, geolocation patterns, and psychometric assessments. In Colombia, the IFC-supported Brilla program is leveraging utility payment histories to offer loans to its customers. In India, Kaleidofin uses socio-economic or environmental variables like access to sanitation and basic infrastructure, as discussed in Box 2.

BOX 2

Kaleidofin: Expanding credit for women-owned businesses in India

Model	Founded	Active Markets	Funding/Metrics	Data Sources
Lending  Lends using alternative data **Scoring**  Provides scores to banks and financial service providers	2017	India	Raised $42 million in equity investments, including $5.3 million in 2025.	Transactions, business seasonality, location, access to infrastructure such as sanitation services.

Kaleidofin provides AI-powered credit decision-making and monitoring solutions to financial institutions in India and across emerging markets, with the aim of improving access for **informal workers, women borrowers, and microenterprises**. In addition, Kaleidofin runs a lending arm to test and showcase its model's effectiveness to other financial service providers.

Its core tool, known as **ki Score**, combines traditional financial metrics with **alternative indicators**, including savings and digital wallet data and socio-economic markers such as access to healthcare, hygiene conditions, and education levels. It also leverages behavioral patterns and exposure to disasters and climate risks such as floods, cyclones, landslides, and dust storms.

To date, **over 7.5 million loans** have been disbursed using Ki Score. Around **64 percent of its borrowers are rural** and **12 percent are first-time borrowers**. Borrowers typically access loan sizes from **$100 to $5,000,** with an average of **$675**. The system supports responsible lending at scale, achieving **85-90 percent loan approval rates** and enabling many borrowers to transition from informal to formal credit and from group loans to individual credit.

Kaleidofin emphasizes fairness and accuracy in its models. While formal gender fairness tests are less practical given that **women comprise 97 percent of its borrower base**, the company actively monitors for **bias across other dimensions such as religion and caste**.

Kaleidofin's infrastructure is used by over **50 banks, microfinance institutions, and non-bank financial institutions across India**, typically through **origination and monitoring partnerships, data-sharing agreements, technical/API integrations, and revenue-sharing or subscription-based models**. Kaleidofin also engages in **co-lending and financial partnerships with microfinance and non-bank financial institutions.**

Source: Interview and questionnaire responses from Kaleidofin (2025).

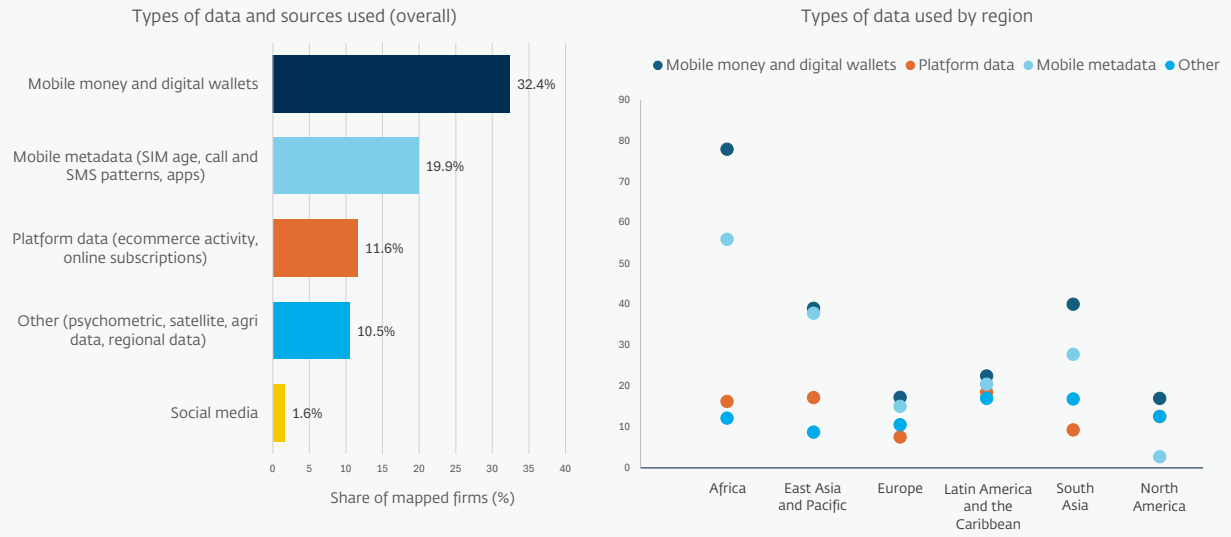
Transaction data is the most popular data type used in alternative scoring models. About 82 percent of firms reference some form of transaction data, ranging from mobile money flows to parsed bank statements, point-of-sale payments, and digital wallet transfers. While these data may originate from relatively formal sources, they may be considered 'alternative' as traditional credit scores do not analyze granular cash flow behaviors or real-world spending patterns. In contrast, alternative scoring models often extract new signals from transactional data, such as utility payments, school fees, airtime top-ups, or irregular earning cycles, which are especially relevant for thin-file borrowers.

As noted earlier, the distinction between 'transactional' and 'behavioral' data is often blurred, and the terminology varies across firms. In this report, these terms are used as they are reported by mapped companies—reflecting the information disclosed on their websites or marketing materials—without independently verifying their internal classification. Please refer to the earlier section on data types for a detailed discussion of these categories.

About a third of mapped firms say they use mobile money and wallet-based data, particularly those in Africa and South Asia. Other data sources, such as behavioral signals, platform work data, and social media, are less common, with very few firms noting social media as a core data source, as shown in Figure 8. Among MSME-focused firms, 87 percent cited a reliance on transactions and 39 percent used mobile money data.

FIGURE 8

Types of data and sources used to score and/or lend, overall and by region



Source: Based on publicly available information for mapped firms. N=448 firms.
Note: These categories are not mutually exclusive, and firms may appear in multiple columns. In many cases, firms do not clearly distinguish between data type and source, so both are retained to reflect how alternative data use is typically reported.

Data that qualify as alternative can vary by context. Some stakeholders noted that business transactions or bank statements, which are common in many models, may not be perceived as 'alternative' in the strictest sense. It may refer to *how* the data is sourced, for example mobile money versus bank statements, or *who* it covers, such as thin-file borrowers. It may also vary according to *how* it is used, for example, assessing cash flow consistency versus formal creditworthiness.

The data itself may be drawn from familiar environments, but its application, particularly for scoring the underserved, reflects a non-traditional and often innovative use case. Many scoring inputs, such as accounting software or corporate data, may appear more sophisticated or formal, but are still categorized as 'alternative' because they introduce signals that were previously absent from credit models (World Bank 2025).[10] In practice, the boundaries blur: for example, use of a bank statement alone may not be new—but using the school fee payments or digital top-ups within it to assess repayment potential is more innovative. Similarly, balance sheet information may be considered traditional, but when repurposed through machine learning or combined with contextual signals, it can become a powerful tool for lenders to reduce risk and enable more inclusive decision-making.

Firms undertake a structured translation process to convert raw alternative signals into decision-ready metrics. Transactional data—such as mobile money, digital wallet, or point-of-service payments—are aggregated into features representing cash inflows, outflows, volatility, and seasonality that serve as proxies for cashflow consistency and repayment capacity. Behavioral and device-level data require additional transformation. For example, phone top-up regularity, SIM card age, and handset stability may be used as persistence or income-stability indicators. Geolocation frequency and travel radius can proxy market access and work frequency. Platform ratings and task completion rates signal business reliability and customer satisfaction.

These engineered features are then mapped to historical repayment or default outcomes using supervised machine learning or rule-based systems. This establishes a predictive relationship between behavioral traits and financial performance. This translation from digital footprint to creditworthiness is central to how alternative data scoring makes invisible economic activity visible to formal lenders.

Alternative Data Spotlight

Digital wallet activity: How often users send, receive, or withdraw funds can serve as a real-time record of cash flows. Patterns in transaction frequency and balance stability help lenders assess a borrower's short-term liquidity. For example, regular small transfers or consistent wallet balances can indicate steady income, even for borrowers without formal bank accounts.

10 See Vidal and Caire (2024) on examples of how transactional data can be leveraged for lending for small and micro enterprises.

How Alternative Data and AI Are Used in Practice

Insights from interviews and case studies reveal that while many firms position themselves as alternative data-led, actual models and inputs vary widely. All models incorporated data that are non-traditional compared to formal bureau or collateral-based scoring, but the degree of innovation varied. For example, firms like Eshandi have built models that continuously update based on behavioral and transactional signals, whereas some firms anchor decisions in traditional indicators like invoices but apply more sophisticated analytics. They may use alternative data-driven risk proxies where formal documents are missing.

In practice, the use of alternative data is rarely exclusive. Fintechs and lenders typically blend traditional and non-traditional information sources depending on data availability. Where elements of open data ecosystems—such as credit bureaus, tax filings, or digital wallet transaction histories—are accessible, firms often use hybrid models that layer machine learning techniques on top of conventional bureau-based scoring.

Traditional credit scores still play a role in some contexts. These are often used as red-flag checks for defaults or used as variables in a larger model, even when alternative data forms the core of the decision model. Conversely, in data-sparse environments alternative data becomes the primary underwriting tool, particularly for loans to MSMEs or informal workers. These hybrid approaches illustrate how alternative data complements rather than replaces traditional credit assessment, extending visibility into segments previously excluded from formal systems.

The choice of alternative data provides insights into firms' business models, partnerships, and target borrowers. Direct lenders were found to rely on mobile phone or transaction data, as seen in Eshandi's use of mobile-money and unstructured supplementary service data (USSD) patterns.[11] Firms that only provide scores to other financial service providers, like the Grameen Foundation's LedgerLink, extracted insights from group savings and loan histories for rural clients. Other firms were found to leverage platform data. For example, Gigmile uses earnings and digital usage data to underwrite loans to gig workers , while firms such as Zhima Credit in China harness e-commerce and social-data footprints. Variation in data type and intensity is shaped by borrower segments—gig workers, small merchants, MSMEs, or rural entrepreneurs—and the legal or technical ability to collect data ethically and effectively. Kifiya similarly leverages diverse alternative data—ranging from mobile transactions and digital payment histories to contextual agricultural indicators—to enable partner banks to extend uncollateralized credit to MSMEs and first-time borrowers in Ethiopia. See Boxes 3 and 4 for case study examples.

11 Unstructured Supplementary Service Data is a real-time, text-based communication protocol that works over cellular networks without requiring internet access. It is commonly used on basic mobile phones for services like mobile money or balance checks.

BOX 3

Kifiya: Building digital infrastructure for inclusive MSME finance

Model	Founded	Active Markets	Funding/Metrics	Data Sources
Scoring Provides scores and financial infrastructure to banks and other financial service providers	2010	Ethiopia	In partnership with the Mastercard Foundation, Kifiya reached over 500,000 MSMEs owned or run by women.	Mobile money, utility payments, ecommerce transactions, agricultural productivity indicators.

Only 1.9 percent of small enterprises and 6 percent of microenterprises in Ethiopia have access to formal loans or lines of credit. Even within that narrow segment, lending is concentrated among a handful of larger and more established businesses. However, Kifiya's unique approach to uncollateralized credit is expanding financial access for MSMEs, small holder farmers, youth, and women entrepreneurs traditionally excluded from formal channels.

Kifiya builds digital infrastructure that enables partner banks to deploy modular, whitelabeled digital lending platforms. Along with data-decisioning infrastructure, this enables banks to offer credit to informal and first-time borrowers using AIdriven scores derived from multiple alternative data sources. Platforms powered by Kifiya include **Michu**, **Efoyta**, **Abol**, **Malefiya**, **Ansar,** and **Lewedaje**.

With minimal initial data, these platforms start borrowers on lower-ticket, uncollateralized credit of up to $2500. Early repayment behavior is tracked and used to build digital credit profiles, enabling progressive increases in credit limits. This adaptive model enables access even to borrowers lacking formal histories.

Kifiya's scoring models incorporate a diverse set of alternative data. These range from mobile transaction behavior, utility or digital payment histories, and platform usage patterns, to contextual data such as local market trends, weather, and soil indicators when assessing the creditworthiness of farmers. This data informs credit decisions via AI techniques such as random forests, and neural networks.

Kifiya intentionally targets women borrowers as part of its financial inclusion mandate. To date, it has unlocked **$340 million in credit for 713,000 MSME borrowers** along with $5.8 million in agriloans to small holder farmers. A total of **$175 million reached 544,000 female borrowers** through the Sustainable Access to Finance for Enabling Entrepreneurship program, which is backed by the Mastercard Foundation. Kifiya's internal analysis indicates that **women score as well as men** in its credit assessment models, demonstrating equity in underlying capabilities.

Source: Interview and questionnaire responses from Kifiya (2025).

Most firms embed some AI or automation into parts of the workflow, such as transaction classification, but a few rely on more automatic score generation. Hybrid or rule-based models remain common, often due to regulatory requirements, explainability concerns, or resource constraints. For example, Grameen Foundation's LedgerLink model largely uses a manual scoring process, and as shown in Box 4, this reflects the limited infrastructure and compliance conditions in the markets it serves. In contrast, firms like Eshandi, Kaleidofin, or Kenya's Cladfy deploy more advanced, data-driven models, including some that automatically approve loans for applicants falling within a defined score threshold.

BOX 4

The Grameen Foundation's LedgerLink platform

Model	Founded	Active Markets	Examples of Data Sources
Scoring Score provider to microfinance institutions and other financial service providers	1997	Uganda, Ghana	Meeting attendance, savings history, loan repayment, savings group performance, occupational history

The Grameen Foundation's LedgerLink platform enables **village savings and loan associations** in Uganda and Ghana to digitize financial behavior by capturing **data on savings, loan performance, and meeting attendance**. Its structured app works even when off-line and generates alternative credit profiles that financial service providers, particularly microfinance institutions, can use for underwriting.

The solution is **built for women-led groups with low-literacy, including refugees and rural communities**. It emphasizes usability and transparency. Due to infrastructure constraints and explainability concerns, it **does not use any AI models**.

Financial institutions using LedgerLink data have extended credit to women who previously lacked access to affordable loans by digitizing informal financial behavior, the Grameen Foundation is helping close the gap between savings group participation and formal financial inclusion. Its approach shows that **community-driven data** can be a viable alternative to traditional credit scoring, especially in hard-to-reach markets.

Source: Interview and questionnaire responses from the Grameen Foundation (2025).

Alternative Data Spotlight

Digital sales and point-of-sale data: This includes records from mobile payment terminals, QR code payments, or digital merchant apps that capture both revenue patterns and business seasonality for small retailers and micro-enterprises. For example, a salon or corner shop that accepts regular mobile point-of-service payments can demonstrate steady cashflow even without formal bookkeeping, helping lenders estimate repayment capacity more accurately.

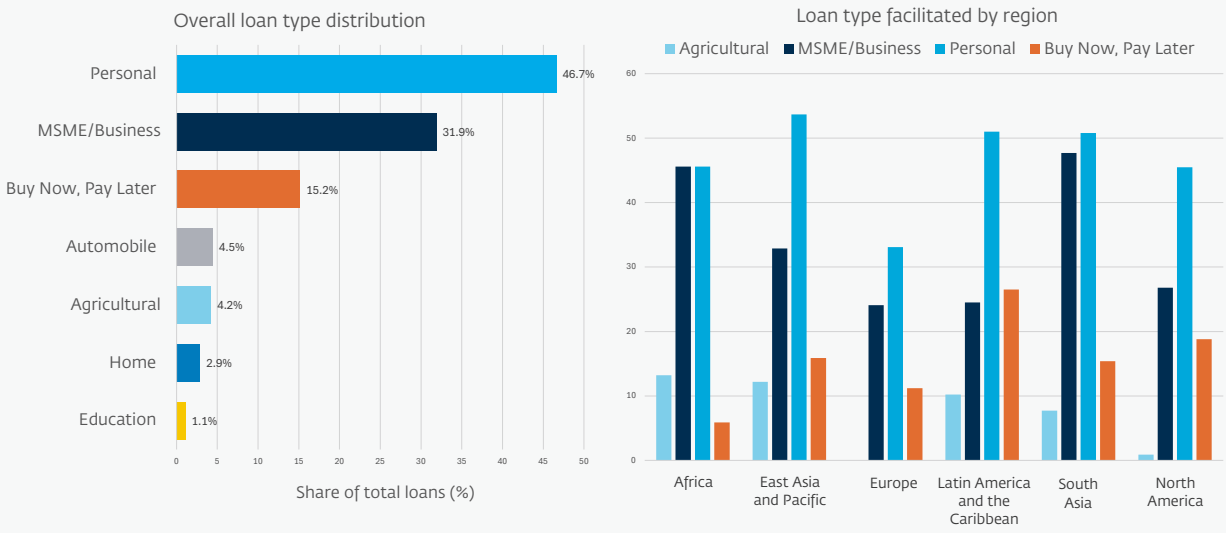
Loan Products and Customer Segments

Personal and MSME loans dominate the product mix, with buy now, pay later and agri-lending emerging as niches. Personal loans and MSME credit are the most common offerings disbursed or facilitated by mapped fintechs, at 47 percent and 32 percent respectively. Within the business segment, mapped firms enable a wide range of products designed to smooth cashflows for small traders, service providers, or platform-based sellers, such as working-capital loans, invoice discounting, merchant cash advances, and SME credit lines. A smaller subset of firms focuses on asset financing, such as vehicle or equipment loans, or supply-chain finance where repayment is linked to receivables from larger buyers.

As shown in Figure 9, buy now, pay later is a growing segment, representing 15 percent of offerings, while agri-lending represents a niche segment. Auto loans and revenue-based financing, where repayments are tied to a share of borrower revenues, are also present but less dominant. Many firms facilitate multiple loan types, as shown in Figure 9.

FIGURE 9

Types of loans facilitated or disbursed, overall and by region

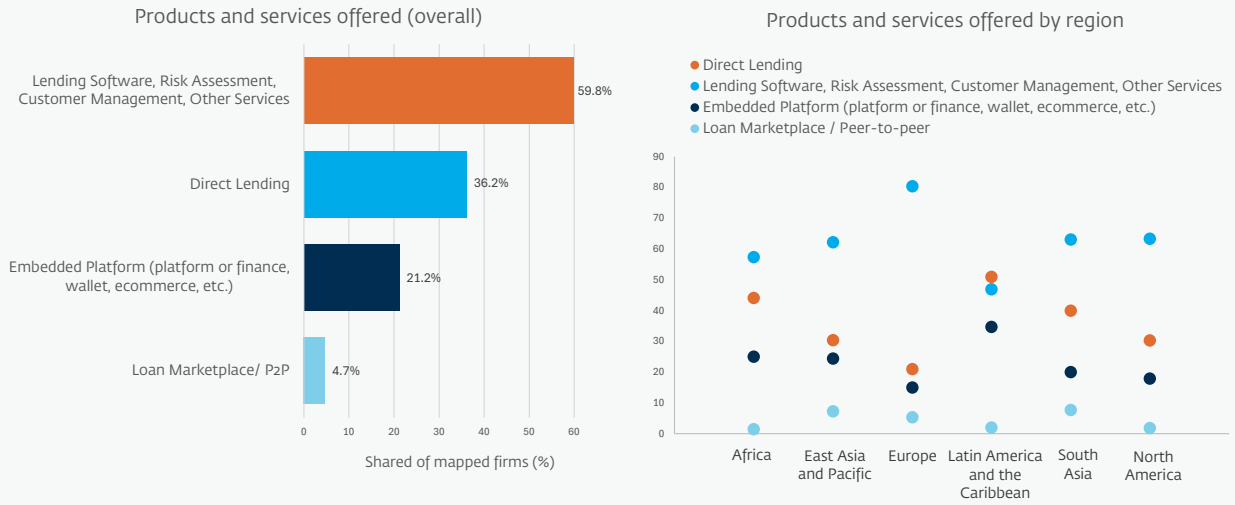


Note: Firms can appear in multiple categories, as classifications are not mutually exclusive. N=448 firms

Many firms go beyond scoring to offer lending, other software-as-a-service (SaaS) products, or services related to embedded finance. About 36 percent of mapped firms that develop their own credit scores also act as direct lenders, while 60 percent offer scoring and lending software to financial partners as part of a wider suite of products such as fraud detection, loan management platforms, and financial advisory services. Embedded finance and marketplace models are also emerging. A majority of firms, especially in Europe, South Asia, and North America, offer scoring tools as part of a broader suite of services, such as risk assessment solutions or lending infrastructure. Direct lending is most common in Latin America and the Caribbean and Africa at 51 and 44 percent respectively, as shown in Figure 10.

FIGURE 10

Products and services offered by mapped firms, overall and by region



Source: Landscape analysis. Africa (N=68), East Asia & Pacific (N=82), Europe (N=133), Latin America & Caribbean (N=49), South Asia (N=65), and North America (N=112).
Note: Firms may be active in multiple regions, and classifications are not mutually exclusive. The Middle East, Central Asia, and Türkiye region is excluded due to a small sample size (N=10).

Mapped firms often partner with financial service providers who incorporate their scores into lending workflows. For example, Grameen Foundation's LedgerLink scores are used by microfinance institutions across Uganda and Ghana. Eshandi combines mobile money and telecommunications data to score borrowers in Zambia and is now partnering with banks in new markets to license the model. Kaleidofin in India initially disbursed loans directly to validate its model but now works with microfinance institutions and banks to scale lending. Others, like QLana, embed credit scoring within a broader suite of asset and loan management tools for banks and asset managers. However, it is difficult to determine the extent to which these models influence final lending decisions.

Banks and other financial service providers use these scores as one of several inputs. For example, In Bangladesh, a partnership between City Bank and BKash explores alternative data-based lending, but such initiatives remain a small share of their overall portfolios. In contrast, Sub-K—an established financial intermediary in India serving bottom-of-the-pyramid clients—is now integrating AI-enabled tools to digitize onboarding, automate know-your-customer (KYC) processes, and streamline field-level credit appraisals. See Box 5 for a case study. This marks a shift from traditional microfinance workflows toward layered automation and alternative data–driven support for decision-making.

BOX 5

SubK: Layering AI and automation in last-mile microfinance

Model	Founded	Active Markets	Funding/Stage/Metrics
Lending Lends using alternative data	2010	India	Over $20 million raised $12 million annual revenue (2025) 350,000 women borrowers served

Sub-K IMPACT Solutions is an India-based financial services provider that partners with banks and non-bank finance companies to extend credit and payments at the last mile. Its portfolio has reached **over 350,000 women borrowers running microbusinesses**, more than **50 percent of whom are first-time borrowers**, with an **average loan size of $350**. Around **30 percent of borrowers are rural** and all have bank accounts.

Building on a traditional microfinance model, Sub-K has begun to **integrate machine learning and generative AI into its credit origination and approval process**. Its **AI/ML risk tool** combines **socio-economic, customer-centric, field officer, and geo-spatial variables** to identify lower-risk borrowers, allowing them to streamline approvals. This is coupled with remote credit appraisals or video-based 'house visits' conducted through Sub-K's **VISyst platform**. More than **150,000 video appraisals** have been conducted so far, representing about **80 percent of all assessments**.

Sub-K is also piloting more advanced approaches, such as **AI-powered risk assessments using unstructured conversational data**. Through recorded borrower interactions, the system applies **computer vision, speech-to-text transcription,** and **sentiment analysis** to extract features such as **asset ownership, housing conditions, emotional tone, and repayment capacity** and generate a composite repayment risk score with confidence levels.

While these AI-based pilots are still being refined, they mark a shift toward **alternative data–driven scoring** in the microfinance segment. The firm's innovations are particularly notable given its borrower base of **low-income women, many of whom are accessing credit for the first time**. By combining traditional microfinance lending with AI and alternative data enabled tools, Sub-K seeks to reduce the costs and risks of serving last-mile borrowers.

Source: Interview and questionnaire responses from Sub-K IMPACT Solutions (2025).

Alternative Data Spotlight

Borrower interviews and digital interactions: This information can reveal behavioral traits linked to creditworthiness. Using speech-to-text tools, lenders can analyze tone, word choice, or emotional cues, also known as *sentiment analysis*, to infer qualities such as confidence, honesty, or planning ability. For example, brief spoken or written assessments have been used to score micro-entrepreneurs with no credit history, helping to expand lending to first-time borrowers in low-data environments.


Photo: © Vincent Tremeau / World Bank

Alternative data models increasingly intersect with the digital-platform economy, although in distinct ways. Around 20 percent of firms in the sample engage with platform-linked data, such as gig work and ecommerce records that capture earnings consistency, transaction volume, and customer ratings—signals that help underbanked digital workers and merchants access credit.

A growing subset uses embedded-finance models that integrate credit products within non-financial platforms or banking apps via an application programming interface (API). This approach allows users to access loans seamlessly while transacting or managing payments, thereby expanding reach without requiring standalone lending apps. Embedded-finance activity is particularly visible in Latin America and the Caribbean, where about 35 percent of firms in the sample describe partnerships or APIs that deliver credit within existing platforms. Together, these trends illustrate both new data sources and new distribution models—but they should not be conflated.

Text analysis of websites and other public-facing information suggests some regional differences in how alternative credit models are positioned. As illustrated in Figure 11, firms in Africa more often refer to 'savings,' 'farmers,' and 'mobile money,' while those in East Asia and the Pacific mention 'AI,' 'machine learning,' and 'embedded' more frequently. These variations likely reflect regional market structures, with greater telecommunications and mobile-money integration in Africa, and more analytics-driven ecosystems in East Asia and the Pacific. The patterns are based on how companies describe themselves rather than standardized taxonomies, and they offer only a broad indication of how firms present their focus areas. Similar contrasts emerged in several interviews, which pointed to similar regional differences in business models, data strategies, and partnerships.

Alternative Data Spotlight

Ride-hailing and delivery-app data: Information on trip frequency, distance, and customer ratings can provide a real-time picture of gig workers' activity and earnings. These indicators help lenders assess the income stability and repayment capacity of platform-based workers who lack formal pay slips.

FIGURE 11

Word cloud from firm descriptions, by region

Africa

East Asia and the Pacific

Latin America and the Caribbean

South Asia

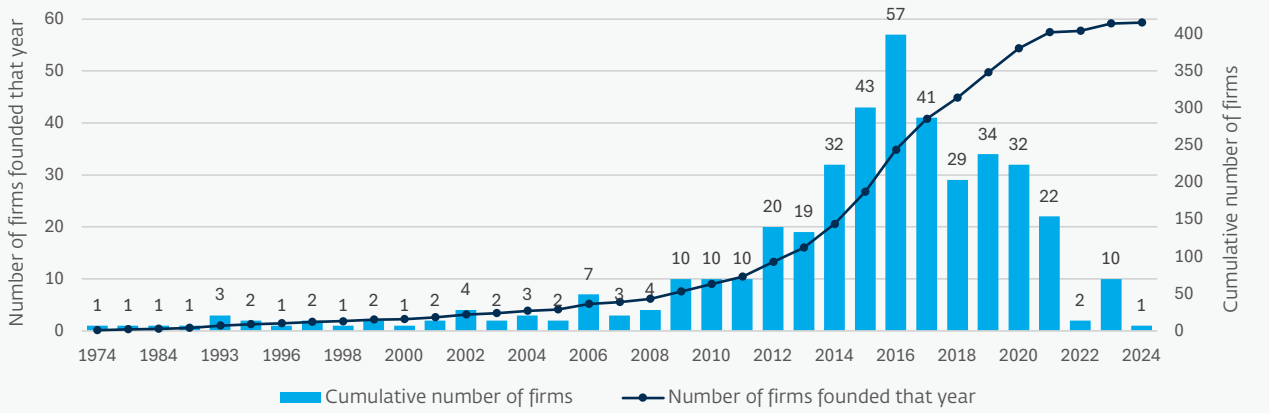
Source: Firm descriptions from Tracxn. Africa (N=54), East Asia & Pacific (N=70), South Asia (N=58), Latin America & Caribbean (N=45).
Note: The sample is restricted to firms operating in emerging markets, excluding the Middle East, Central Asia, and Türkiye region due to limited observations. Firms may operate across multiple regions.

Funding and Maturity Landscape

Most alternative credit scoring firms are relatively young. Over 75 percent of firms were founded after 2014, reflecting a surge driven by expansion in mobile infrastructure, platform economies, and digitized payments. Some older firms have more recently pivoted into alternative credit. A drop-off after 2020, shown in Figure 12, may reflect a lag in visibility for newly-founded firms. Early-stage players often take time to gain traction, secure funding, or receive media coverage, which can delay their appearance in data sources. Future iterations may pick up more firms as they scale.

FIGURE 12

Year in which mapped fintech firms were founded

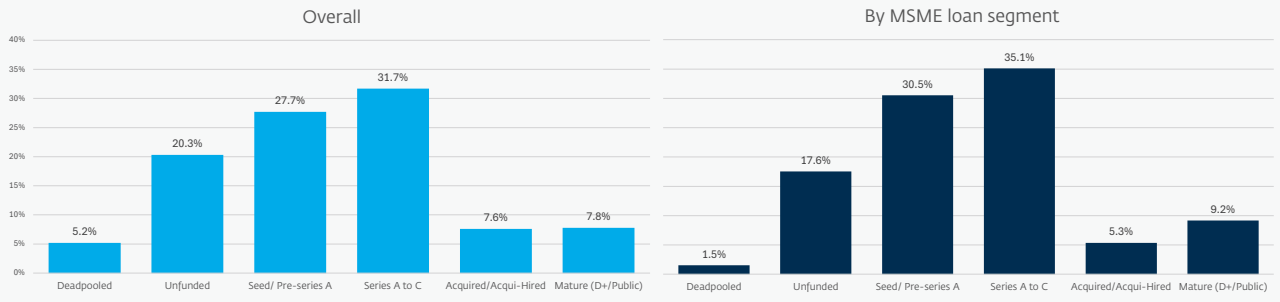


Source: N=415 mapped firms with available founding year data.
Note: Bars represent the share of newly-founded firms in that year. The dotted line indicate the total number of firms over time.

Funding activity reflects a young and fast-evolving ecosystem, with notable early-stage momentum. Firms offering MSME scoring or lending have higher rates of funding and maturity than the full sample, as seen in Figure 13. Over 30 percent of mapped firms are venture-backed at Series A–C stage. Sixteen percent are at seed, and 15 percent have matured at Series D or higher, or they have exited. Seventy-two percent of funded firms raised capital in the three years prior to publication of this report. However, only 5 percent were 'deadpooled'—a classification for firms that have shut down or no longer operate. This is relatively low, likely reflecting the fact that the sample is drawn from firms with a visible online presence or some degree of traction.[12]

FIGURE 13

Funding and maturity stages of mapped firms, overall and with an MSME focus



Source: Mapped firms with available funding data. N=415.
Notes: Acqui-hired typically refers to a company acquired mainly for its employees to strengthen the acquiring firm's workforce. Unfunded firms are tagged as 'early stage' in Tracxn. Mature firms include Series D, E, F, I and beyond, public firms, and those tagged as 'late stage.'

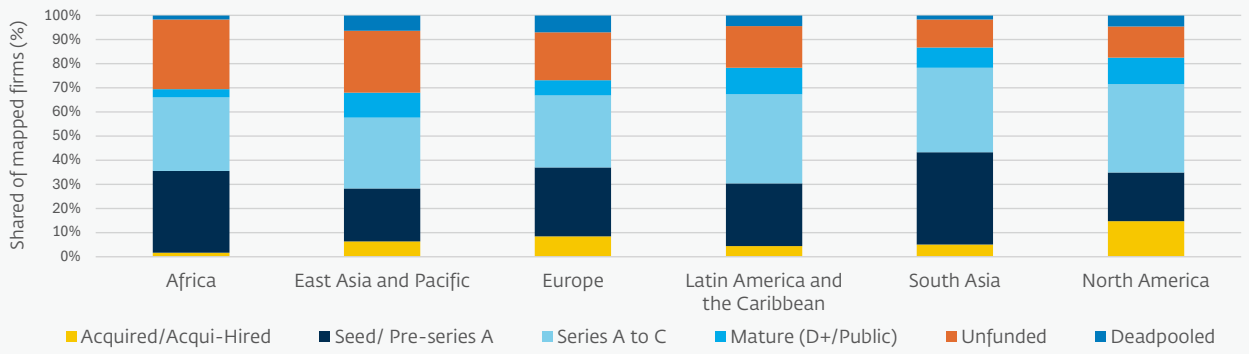
12 In contrast, a narrower set of firms historically tagged in Tracxn as 'Alternative Credit Scoring' firms suggest a higher deadpooled share of roughly 24 percent. While that data may capture older firms that are no longer active, it may still underrepresent the smallest or unfunded firms with limited public information.

Africa has the highest share of unfunded firms at 29 percent, while North America and Europe have a larger share in the growth or exit stage, including public and acquired firms.

As illustrated in Figure 14, South Asia stands out for its strong share of seed-stage firms, while seed, venture, and mature firms are distributed relatively evenly in East Asia and the Pacific.

FIGURE 14

Funding and maturity stage of mapped firms by region of operation

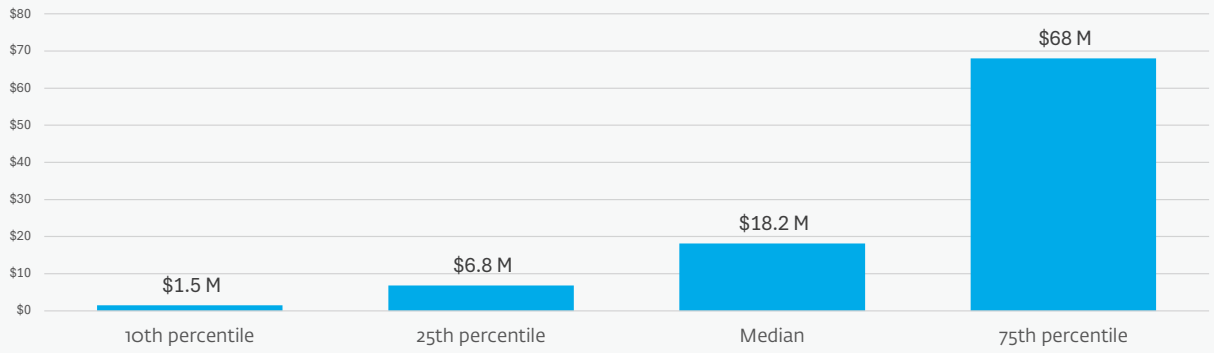


Source: Traxcn. Africa (N=59), East Asia & Pacific (N=78), Europe (N=130), Latin America & the Caribbean (N=46), South Asia (N=60), and North America (N=109)
Note: Firms may be active in multiple regions. Classifications are not mutually exclusive. The Middle East, Central Asia, and Türkiye region is excluded due to its small sample size.

The distribution of total capital raised by funded firms varies widely across the landscape. As shown in Figure 15, among the subset of funded firms with available data, the median amount of total capital raised is roughly $18 million. A smaller group has attracted substantially larger investments, reflecting the presence of well-capitalized or more established players, including larger entrants in the sector, alongside many early-stage or modestly-funded providers. This heterogeneity in funding levels is important for interpreting regional patterns of capital flows.

FIGURE 15

Capital raised by funded firms



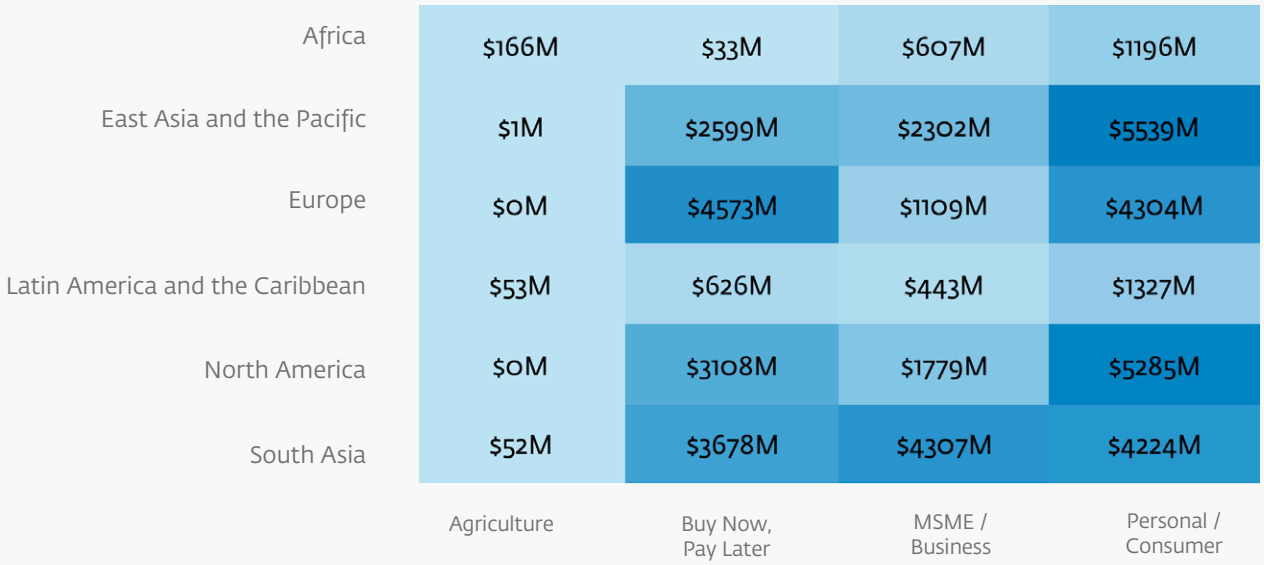
Source: Mapped firms with available funding data from Tracxn. N=238
Note: Percentiles show the distribution of capital raised among funded firms in the sample; for example, firms at the 75th percentile have raised about $68M or more.

North America, Europe, and South Asia dominate funding volumes, but traction is visible across regions. In North America and Europe, firms that focus on personal and consumer loans have the highest share in total raised capital. In South Asia, firms focused on MSME lending have attracted significant investment, with models operating in India leading in terms of funding flows. In Africa, firms focused on personal and consumer loans and MSME segments raised more capital than others operating in the region.

It is important to note that these amounts are indicative rather than definitive, and they are drawn from a subset of funded firms available in Tracxn. Not all funding rounds are disclosed, and firms operating across multiple regions or loan segments may be counted twice. However, the trends shown in Figure 16 offer valuable insights into capital flows, such as the concentration of funding in MSME and personal loan models and the geographic spread of investor activity. These trends help identify high-interest sectors and regions, and with more granular and complete datasets future iterations could refine these estimates. Box 6 contains a case study of the role investors can play in shaping the emerging market lending ecosystem.

FIGURE 16

Total capital raised by region of operation and loan segments, in millions ($)



Source: Mapped firms with available funding data from Tracxn. N=238
Note: Firms may be active in multiple regions or loan segments, so their total funding may be counted in more than one category. As a result, funding totals across regions and segments are not mutually exclusive. Funding data was available only for a subset of firms, as many funded firms in Tracxn did not publicly disclose investment amounts.

BOX 6

Accial Capital: Data-driven debt financing to scale inclusive fintechs

Model	Founded	Active Markets	Funding/Stage/Metrics
Investor /Funder 	2018	East Asia, Latin America	Investor in Vexi (Mexico), FinMaq (Colombia & Mexico), Meddipay (Colombia) and others $3.5 billion in disbursed loans, including $1 billion to women 2.9 million loans and 500,000 borrowers, including 175,000 women

Accial Capital is an **impact investor** that provides structured debt financing to fintech lenders across emerging markets. Founded in 2018, it operates across Latin America, Southeast Asia, and Central Asia, and has supported over $3.5 billion in loans for underserved consumers and small businesses.

Through its proprietary **ORCA platform**, Accial monitors millions of **borrower-level data points** received directly from fintech investees via **APIs.** This provides visibility into **repayment patterns, portfolio dynamics, and borrower outcomes** well beyond the typical reporting dashboards used by its peers. Accial also complements its quantitative monitoring with qualitative insights. Beyond the datasets it receives from partners, the firm uses emergent AI technology to **understand borrower experiences** directly. For instance, in Mexico, Accial worked with investee Vexi to collect and process **voice-note responses** from customers to better understand how credit products were impacting users' lives and advancing Accial's goal of widespread financial wellness.

Nearly **half of its loan disbursements have reached women borrowers**, and a significant share has gone to **rural populations** often excluded from formal finance. Accial Capital illustrates the role **investors** can play in shaping the emerging market lending ecosystem.

Source: Interview and questionnaire responses from Accial (2025).

Partnerships, Incumbents, and the Regulatory Environment

Legacy financial players are acquiring alternative-credit scoring firms, forging partnerships, and building their internal capabilities. American Express acquired Kabbage, a digital lender for small businesses and TransUnion acquired FactorTrust, a firm focused on subprime analytics. Equifax owns DataX, a provider of non-prime credit reporting and risk analytics. Meanwhile, firms like RentReporters provide rent payment data to bureaus to improve renter's credit visibility. These partnerships help bureaus and banks access data sources that would otherwise be out of reach due to compliance barriers. They also allow legacy players to develop new offerings like credit-building tools tailored for thin-file consumers (World Bank 2025). Telecom providers and digital platforms are also entering the credit ecosystem by leveraging their own data assets, with MTN Uganda offering MoMoAdvance and Safaricom offering M-Shwari, for example.

While partnerships with incumbents are expanding, this is not without friction. Many alternative credit scoring firms are forging ties with established financial institutions, reflecting rising institutional interest in new data-driven approaches. Kaleidofin in India has partnered with larger financial players such as Federal Bank to scale microfinance loan offerings through embedded scoring and tailored lending products. Others collaborate with telecommunications companies, merchant networks, or mobile wallet providers on subscription or revenue-share models to access customer data and distribution channels. During interviews, fintechs noted that traditional institutions, especially those that are larger or more risk-averse, have been slow to adopt or scale alternative models due to concerns ranging from regulatory compliance and data privacy to the opaque nature of AI-based scoring methods. Incumbents seem to be experimenting with these models, but not yet fully mainstreaming alternative approaches.

Regulatory considerations strongly influence how firms position their scoring models and structure go-to-market strategies. Interviews revealed that some providers deliberately avoid labeling their tools as 'credit scores' to sidestep stricter regulation. Instead, they brand them as analytics or advisory services, placing the burden of decision-making on licensed financial institutions. Others act as lenders or officially licensed scorers and are therefore more directly bound by local credit regulations. Globally, digital lenders report that uncertainty around data privacy, explainability requirements, and the regulatory classification of scoring outputs significantly shape product design. Some firms seek clarity through regulatory sandboxes or AI frameworks, while others mitigate compliance burdens by embedding within bank partnerships.[13] In markets like India, digital lending guidelines from the central bank require transparent audit trails and explicit borrower consent for alternative data use. They also prohibit access to sensitive phone metadata like contact lists or call logs (World Bank 2025).

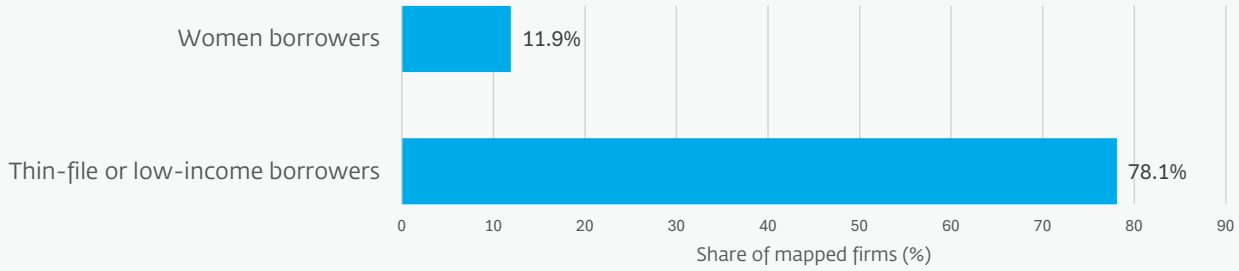
Inclusive Lending and Female-led firms

Most alternative credit-scoring firms frame themselves as inclusion-focused, although the emphasis is often broad. Around 80 percent of mapped firms use terms like 'underserved,' 'thin-file,' or 'first-time borrower' in connection with their services, as illustrated in Figure 17. This framing aligns with financial inclusion goals, but it is difficult to verify whether this reflects actual targeting or if is primarily reputational signaling. However, framing still matters as it can influence investor interest, partnership models, and perceptions about the legitimacy of using alternative data.

13 A regulatory sandbox is a controlled environment set up by financial or data regulators that allows firms to test innovative products, services, or business models with real users under relaxed regulatory requirements and close supervision. The goal is to encourage innovation while managing risks before wider rollout.

FIGURE 17

Target borrower segments mentioned on firms' websites or in other public information



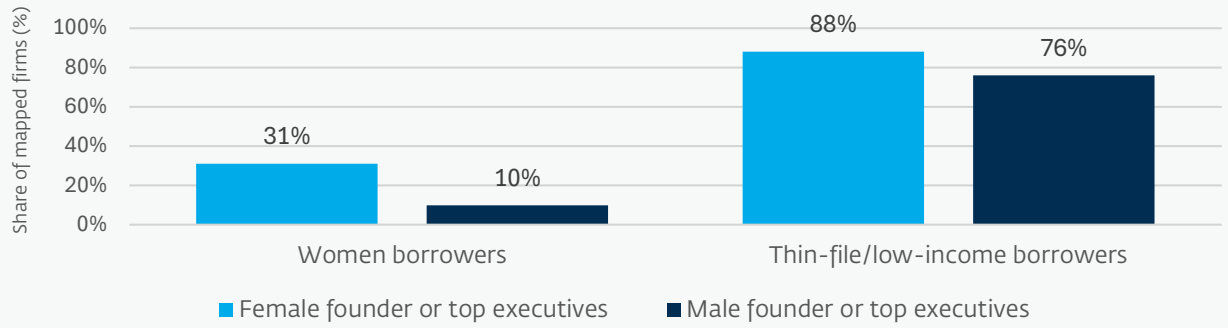
Source: Mapped firms=448

While women can benefit from many alternative credit scoring models, firms rarely target women explicitly. Only 12 percent of mapped firms refer to women or female borrowers on their websites and in their public-facing information, indicating limited intentionality in product or model design. However, interviews conducted for this report reveal Kaleidofin explicitly targets low-income women in India through gender-sensitive scoring and product features. In addition, Eshandi in Zambia initially served only women because they were perceived as more reliable borrowers, and even now women comprise almost 60 percent of its client base, as discussed later in this section in Box 7**.**

Female-led firms more often center inclusion in their mission statements, on their website, and in other public information. Firms with at least one woman founder, chief executive officer, or chief technical officer are more likely to refer to thin-file or low-income borrowers in their public communication and information channels, at 88 percent versus 76 percent for firms with all male leadership. About a third of firms with women leaders explicitly mention women borrowers versus 10 percent of firms with all male leadership, as shown in Figure 18.

FIGURE 18

Target segments mentioned by firms, by gender of founder or leadership

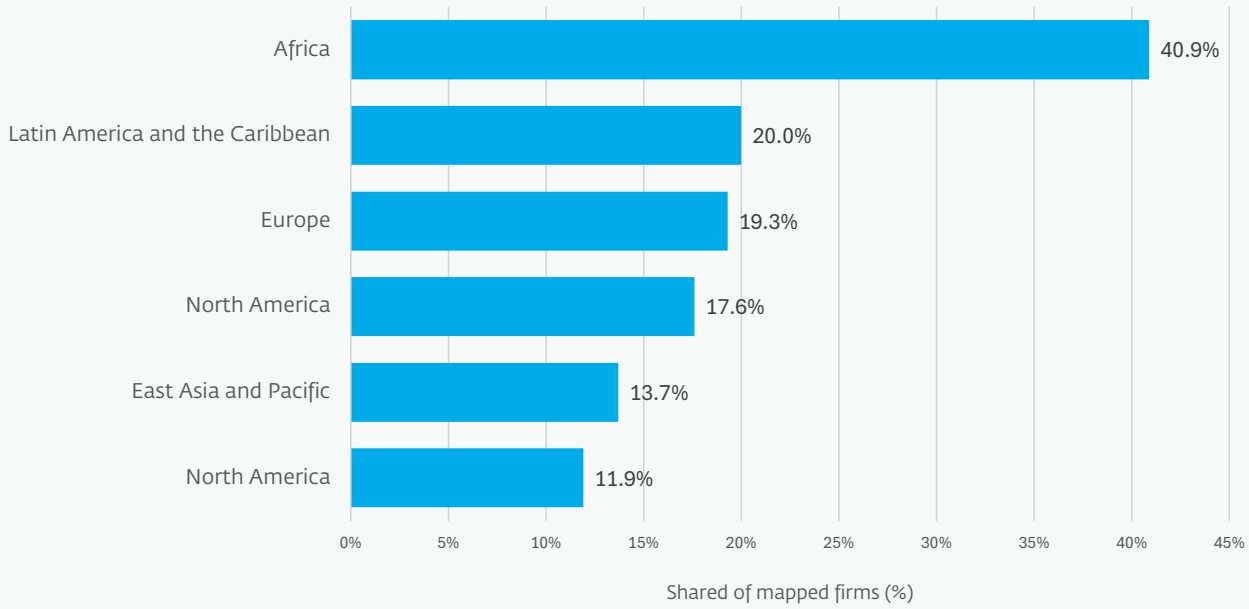


Source: Mapped firms with available information. N=297

Generally, female leadership in mapped firms is limited, however those serving African markets stand apart with 41 percent having a female co-founder, chief executive officer, or chief technical officer. Founder and leadership information was available for about 300 of the mapped firms, and 20 percent of these had a woman in a senior role. This increases to more than 30 percent among firms involved in MSME lending. A regional breakdown is contained in Figure 19.

FIGURE 19

Female leadership in mapped firms, by region



Source: Mapped firms. Africa (N=44), East Asia & Pacific (N=51), Europe (N=83), Latin America & Caribbean (N=30), South Asia (N=42), and North America (N=68).
Note: Firms may be active in multiple regions, and classifications are not mutually exclusive. The Middle East, Central Asia, Türkiye region is excluded due to its small sample size.

Female Borrowers

Borrower-level data provides quantitative evidence on how alternative data-driven scoring can improve women's access to credit. Analysis of over one million microloan applications from Eshandi in Zambia showed that female borrowers score slightly higher on machine learning–based credit scores than men, when controlled for behavioral and demographic factors.

BOX 7

Eshandi: AI-driven lending using mobile money and telecommunications data in Africa

Model	Founded	Active Markets	Funding/Metrics	Data Sources
Lending  Lends in Zambia using alternative data **Scoring**  Provides scores to financial service providers	2019	Zambia, Kenya, Zimbabwe, South Africa	Ranked 4th on the Financial Times' list of Africa's fastest-growing companies in 2025	Mobile money patterns, location, SIM card age, business documentation, formal histories if available

Eshandi is a fast-growing fintech lender and scorer headquartered in Zambia, with operations extending across Kenya, Zimbabwe, and South Africa focused on personal and MSME lending. It has disbursed over 1.5 million loans and serves more than **400,000 active borrowers, of whom 58 percent are women.**

To assess creditworthiness, the company draws on **mobile money transaction patterns via USSD, SIM card age, airtime usage, and behavioral markers,** such as school fee payments, betting activity, and fuel payments, along with formal documentation when available. Its proprietary AI-driven scoring engine updates risk profiles daily and flags high or low-risk borrowers based on behavioral segmentation. For instance, **school fee payments are predictive of responsible repayment while frequent betting activity flags higher risk**.

Eshandi observed **strong repayment rates, higher digital wallet balances, and greater reinvestment** among female clients, **particularly rural women**, and its peer-to-peer lending model encourages savings and community support among women borrowers and other underserved segments. The firm also **partners with mobile network operators**, giving it access to over 10 million mobile users.

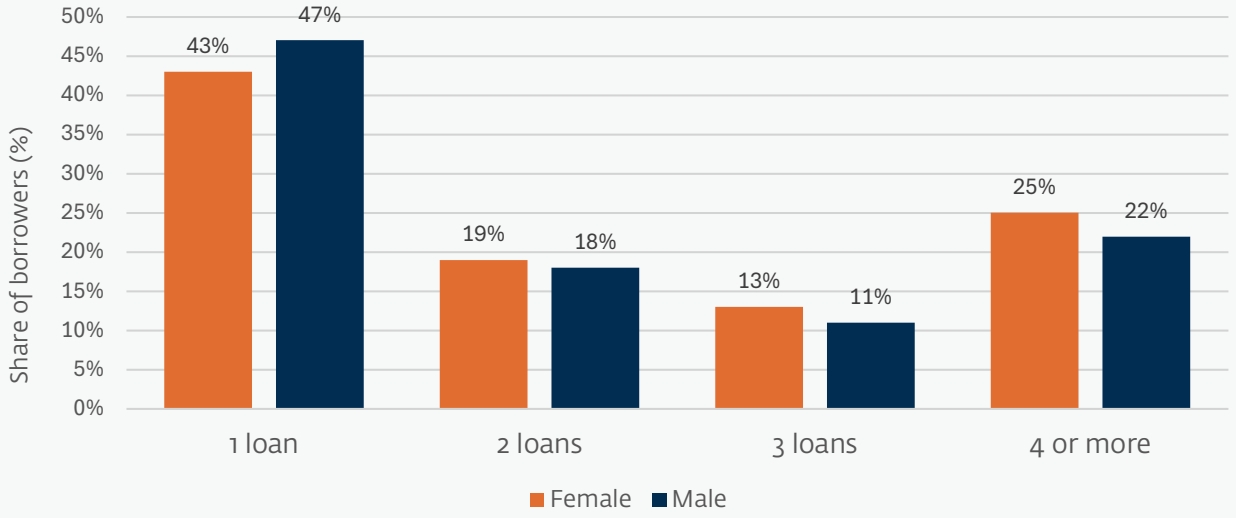
Eshandi operates as both a **direct lender** and a **credit scorer**, holding licenses in several countries and participating in Zambia's regulatory sandbox. Eshandi demonstrates how inclusive design and local data partnerships can scale financial access for underserved borrowers, particularly women. An extended case study on Eshandi is available in Appendix A and includes the underlying scoring model and a comparison of determinants across male and female borrowers.

Source: Interview and questionnaire responses and data from Eshandi (2025).

Women also receive repeat loans more frequently, as shown in Figure 20, and in larger average amounts, reflecting repayment reliability and consistent financial behavior. Rural women, in particular, benefit from mobile money-based scoring, which captures stable informal income flows often missed by traditional systems. Behavioral indicators, such as gambling activity, help differentiate genuine risk from informal financial activity—reducing bias rather than amplifying it. These findings underscore the inclusion potential of AI-enabled models in low-income markets.

FIGURE 20

Number of loans approved by Eshandi, by gender of borrower



Source: Eshandi data on 325,000 borrowers and 850,000 loan approvals in Zambia in 2024.

Despite positive outcomes for women, models that incorporate gender-segmentation and inclusion remain limited.[14] This report found that gender is often omitted from credit scoring models due to fairness concerns, regulatory caution, or the desire to maintain lean, explainable models. In interviews, one firm stated, "We have not yet performed a formal gender fairness audit. We plan to implement fairness metrics like approval rate parity and disparate impact ratio in upcoming evaluations." Some providers are beginning to pilot structured fairness checks. For example, Kaleidofin in India tests its models for performance across different customer characteristics, such as gender, income, and geography, using metrics such as area under the curve, precision, and recall to identify potential biases. The firm is also engaging early with regulators on emerging AI fairness frameworks and exploring responsible data-use practices with financial partners. Some firms use gender for segmentation or to glean market insights, but rarely as a variable in score generation. Still, even without explicit design, several models appear to serve women well. This suggests untapped potential to amplify these benefits through intentional gender-inclusion strategies.

14 See Vidal and Caire (2024) for a discussion on gender-intentional scoring

Photo: © Visual News Associates / World Bank

Loan characteristics highlight the inclusion potential of these models, especially for women borrowers. Among interviewed firms, loan sizes varied from microloans of as little as $10 to small business loans exceeding $50,000. These loans covered diverse needs such as working capital, asset purchases, personal finance, and emergency credit. Several firms demonstrate meaningful scale and outreach to women. For example, Kaleidofin has facilitated over six million loans to women and Eshandi has facilitated over one million loans to women in Zambia. Other firms say they offer better pricing and terms for female borrowers. Yellow Factoring in Cameroon offers women lower monthly interest rates of 1.8 percent versus 2.2 percent for men, and Cladfy in Kenya says that due to stronger repayment practices and records, it offers lower pricing to women.

Women borrowers are outperforming male counterparts across several metrics. For example, Yellow Factoring's experience in Cameroon showed that women had better onboarding and completion rates, at 88 percent versus 63 percent for men. While women applied for lower average loan amounts than men, they more often repaid on time. Similarly, another firm found that women had higher retention and better repayment post-onboarding. Complementary evidence from 7,000 customer surveys by digital credit card provider Vexi in Mexico further highlights how inclusive models can foster financial confidence and gradually build credit among women, as shown in Box 8.

BOX 8

Vexi: Alternative data-driven credit cards for first-time borrowers in Mexico

Model	Founded	Active Markets	Funding/Metrics	Data Sources
Lender and credit card provider	2016	Mexico	Raised $32 million in equity investments $15.3 million credit line raised in April 2025	Age, gender, location, education, income, behavioral patterns, geolocation

Vexi is a digital credit card provider in Mexico focused on **first-time and underserved borrowers**. To date, it has issued **over 1.4 million cards** and has around **200,000 active users**. Its customer base includes **gig workers, young adults, and small entrepreneurs.** About **65 percent of customers are first-time borrowers** and **30 percent run their own businesses**. Women make up roughly 24 percent of cardholders.

Vexi uses a proprietary AI/ML model that blends bureau data with **alternative data** such as device consistency, geolocation, and education. It also uses predictive features such as form-completion behavior, for example, the number of times a user hits backspace while entering sensitive fields like income. In those cases, women tended to exhibit **more deliberate and consistent input patterns,** which in turn **correlated with stronger repayment outcomes**.

These signals enable **scoring, fraud detection, and the setting of appropriate credit limits**, which typically start as low as $50. Fifty to 60 percent of its borrowers borrow less than $100, while 30 percent are approved for loans between $100 to $300. The median borrower income is approximately $1,000 per month.

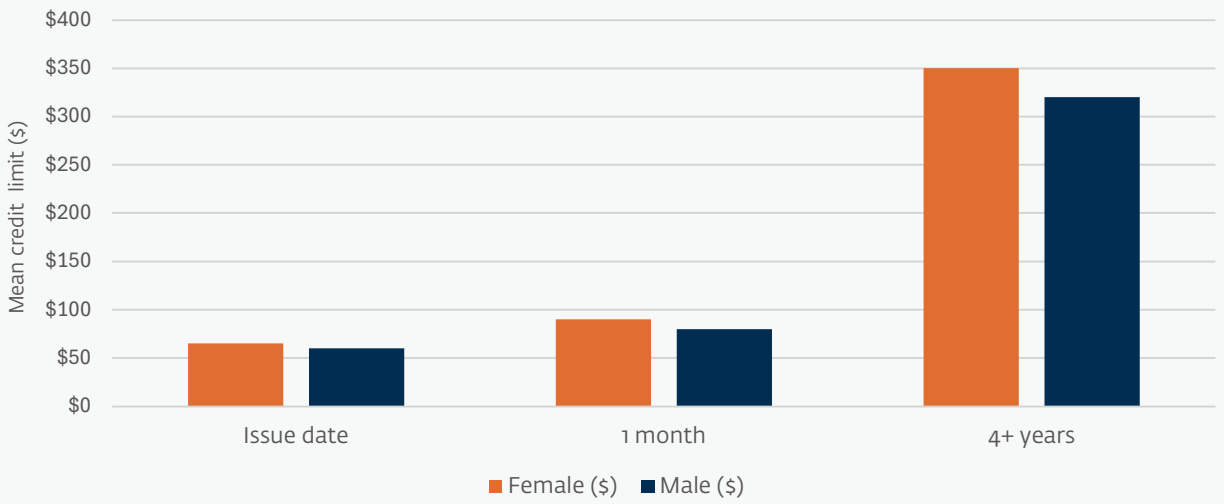
While approval rates are similar for men and women, **women apply at lower rates and drop off at higher rates** during completion and upfront fee payment. Interestingly, **women who are approved receive slightly higher average credit limits** of $100 compared to $90 for men. To address these gender gaps, Vexi piloted **AI-powered chatbots, clearer messaging, and gender-disaggregated tracking** to redesign user journeys and build trust. The firm also partners with a range of mobile network operators**,** banks, and fintechs to expand its reach and refine its credit scoring model. By combining behavioral data with traditional metrics, Vexi is developing a scalable model to extend credit responsibly to new and underserved borrowers in Mexico. See Appendix A for a deeper look at Vexi.

Source: Interview and questionnaire responses and data from Vexi (2025).

For women entrepreneurs, fintech credit cards are an important entry point into formal finance. Women reported greater use of credit cards for business-related expenses at 62 percent versus 44 percent among men. As shown in Figure 21, women's average credit limits gradually exceeded men's, suggesting responsible repayment behavior among women borrowers. Confidence in applying for future credit also rose more sharply for women, from 3.0 to 4.4 on a five-point scale over twelve months. Full results and gender-disaggregated figures are available in the extended case study in Appendix A.

FIGURE 21

Mean credit card limit, by gender and tenure



Source: Customer survey data of over 3000 Vexi borrowers in Mexico.
Note: $1= 18.5 Mexican pesos

Photo: © Simone D. McCourtie / World Bank

Photo: © Dominic Chavez/World Bank

4. CHALLENGES AND OPPORTUNITIES: RISK, FAIRNESS, AND INCLUSION IN AI-DRIVEN CREDIT SCORING

The expansion of alternative data and AI-based scoring offers the potential for greater inclusion but it also introduces risks and trade-offs. As alternative data and AI models rapidly scale, there are growing concerns around transparency, fairness, and privacy. Stakeholders must balance the promise of improved credit access for underserved borrowers with the risk that poorly designed or unregulated models could entrench existing inequalities or create new forms of exclusion. While many models are improving efficiency and risk prediction, critical questions remain: who is newly-included through these models and do some borrowers lose access they would otherwise have through traditional methods? These trade-offs warrant reflection, especially given firms' rational incentives to reduce default risks and expand customer segments through more inclusive products.[15]

Biases in data and model design can reinforce exclusion, particularly for women and marginalized groups. AI systems often reflect historical, measurement, and selection biases embedded in the data they are trained on. These include underrepresentation of women in credit histories, mismeasurement of informal income, or exclusion from digital transaction ecosystems. These biases risk encoding discriminatory patterns into model outcomes—whether intentionally or through proxy variables. Variables such as location, education, device type, or platform ratings that correlate with attributes such as gender and income can lead to *digital redlining*—for example, where automated systems systemically assign lower scores to borrowers from low-income neighborhoods or women using shared devices because of indirect data proxies (Federal Reserve 2021, Prove 2021). Firms and policymakers can adopt multi-level

15 Empirical findings point to substantial inclusion gains. Turner et al. (2012) show that among thin-file borrowers, 64 percent saw an improvement in credit standing when alternative data was introduced, compared to just 1 percent who experienced a decline. Among those already scoreable, 25 percent moved to a lower-risk tier, while only 6 percent were downgraded, suggesting that alternative data can unlock access without compromising risk accuracy. However, it is important to further examine whether the downgraded were fairly assessed, and whether they retain access through other channels.

strategies to mitigate algorithmic bias. Fairness optimization techniques across the AI pipeline can reduce inequitable outcomes. Pre-processing strategies can remove historical bias from training data. In-processing approaches can adjust model structure to improve parity, while post-processing methods can finetune outcomes to ensure equitable treatment (Kozodoi et al. 2022).

Regulatory guidance needs to balance anti discrimination safeguards with the potential of inclusive AI design. This is particularly relevant in the use of gender indicators and sex-disaggregated data in credit scoring**.** In many jurisdictions, legislation such as the US Equal Credit Opportunity Act and Fair Credit Reporting Act restricts the use of protected attributes like gender, religion, or age in credit decision-making. These rules aim to prevent discriminatory outcomes but can inadvertently constrain analysis of gender gaps, as discussed in Box 9. In contrast, sex-disaggregated monitoring and fairness testing can improve model accuracy and highlight inclusion gaps, provided gender data are used for auditing rather than decision-making. Regulators increasingly recognize this distinction, allowing segmentation and fairness testing while keeping protected attributes out of production models.

BOX 9

Sex-Disaggregated Data: From oversight to innovation

Credit reporting systems in many markets are largely 'sex-blind.' In other words, they exclude demographic data in an attempt to reduce sex discrimination in credit markets. However, this represents a missed opportunity. Sex-disaggregated data can in fact strengthen credit reporting and scoring systems by revealing how models perform across borrower groups. It can enable lenders and regulators to identify gender gaps, detect bias, refine credit-risk models and assessments, and enhance overall lending.

A 2025 review by the International Committee on Credit Reporting found that although most credit registries and bureaus collect sex-disaggregated data, only a few use it in credit risk assessment. In most cases, data on sex is used for data matching, customer identification, or compliance reporting. However, when sex-disaggregated data is incorporated into model validation and portfolio analysis, it can enable policy makers and lenders to test whether scoring models perform equally well for men and women and identify hidden patterns that may distort credit decisions.

Evidence from pilots in emerging markets indicates that such analysis can improve predictive accuracy and expand lending to women without increasing portfolio risk. Given that women are traditionally underserved, leveraging alternative data can further enhance model performance and support the development of more transparent, accurate, and inclusive credit ecosystems. However, effective integration requires a context-sensitive approach, consistent reporting templates, and clear data-sharing protocols across lenders and credit bureaus, as well as safeguards against potential misuse.

Source: Sex-Disaggregated Data in Credit Reporting and Scoring: A Global Analysis (World Bank Group 2025)

Gaming of AI models and short-term behavioral shifts could undermine model reliability. There is some concern that prospective borrowers may learn to manipulate inputs to achieve better scores, and that individuals may adjust behavior, such as pruning social media connections, if they suspect these influence scores. Overreliance on one-off data snapshots or highly sensitive behavioral signals without triangulation may expose models to gaming and reduce predictive robustness.

Data privacy and ethical use must be paramount, especially in contexts with weak protections or gendered digital divides. Women often face greater risks from data misuse, especially in regions with low digital literacy, shared phone usage, or weak consent frameworks. Emerging credit applications that access SMS logs, geolocation, or photos without robust protection mechanisms can expose users to exploitation (Kelly and Mirpourian 2021).Moreover, gender differences in willingness to share data can be influenced by trust, security, or socio-cultural norms, and this can limit women's participation in AI-enabled lending (Berg et al. 2020, Chen et al. 2023). The use of shared devices, the influence of social norms, or the underutilization of smartphone features can all lead to lower scores despite strong underlying creditworthiness. In such cases, 'gender-neutral' models may not eliminate bias, and instead risk obscuring structural differences that result in lower scores for women.

The risks of exclusion can increase when characteristics such as gender, race, residence, age, or displacement overlap. For instance, rural women working in the informal economy may face the overlapping barriers of weaker mobile connectivity, fewer verifiable transactions, and limited digital literacy. Recognizing these intersectional layers is key for inclusive AI design. Segmenting fairness metrics or model diagnostics by intersecting attributes, for example gender and location, can reveal hidden disparities and guide more targeted mitigation.

Data availability and data visibility are not the same. Many women and informal workers are 'data-rich' but invisible. In other words, their digital activities exist but remain inaccessible to formal lenders due to siloed systems, consent barriers, or a lack of standardized formats. Making these data visible is essential to equitable AI-driven credit.

Photo: © IFC

Photo: © Binyam Teshome / World Bank

5. UNLOCKING INCLUSIVE CREDIT AT SCALE: WHAT'S NEXT?

Alternative credit models are already expanding financial access, especially for women, and this is sometimes by design and at other times incidental. Interviews and case studies reveal that several alternative scoring approaches have reached women borrowers at scale, offered innovative and tailored products at more favorable terms. This is the result of both intentional efforts to expand inclusion, such as Kaleidofin's women-focused design, and from incidental outcomes like women outperforming men on repayment metrics, as observed by Eshandi, Vexi, and Yellow Factoring. These patterns underscore that inclusion is not only equitable, it is also commercially beneficial.

However, systemic barriers continue to limit scale and reach, especially among mainstream financial players. While startups and mission-oriented fintechs are using alternative data to innovate, large institutions like banks often remain hesitant. Interviews revealed concerns around the reliability and consistency of alternative data, particularly from fragmented or informal sources. In addition, traditional risk frameworks, regulatory conservatism, and a lack of familiarity with AI-driven tools can impede wider adoption. As a result, high-potential models face challenges scaling beyond niche or pilot stages.

Gender-segmented AI models offer one pathway to increase inclusion in existing models. Gender-specific model design enhances performance without increasing default risk (Vidal and Caire 2024).[16] These models allow for fairness audits and accommodate gendered differences in borrowing behavior, informal income, or financial documentation. Despite the potential, a global survey by IFC found only about 60 percent of fintechs collect sex-disaggregated data and far fewer use it in modeling (IFC 2024).

Momentum is growing in emerging markets, but it remains uneven. Landscape analysis reveals that

many firms operate in developed markets, but activity is increasing in emerging markets such as India, Kenya, Nigeria, Indonesia, and Mexico, with notable growth in the MSME and personal loan segments. The diversity and volume of firms operating in these geographies point to growing adoption.

Forward-looking regulatory approaches are critical to the responsible and inclusive use of AI and alternative data in lending. Several countries are making noteworthy progress in this area (World Bank 2025). For example, Brazil, China, Tajikistan, and Uganda have updated their credit reporting frameworks to explicitly accommodate alternative data. Others are issuing broader guidance on responsible data use in credit markets, such as India's Account Aggregator framework and the European Union's Consumer Credit Directive.

In parallel, flexible tools like regulatory sandboxes and pilot initiatives provide safe environments to test innovations before wider adoption. They enable regulators and industry to test whether alternative data and AI models can meet prudential and consumer-protection benchmarks, such as accurate estimation of expected losses, explainability, and non-discrimination. In countries like Kenya, and the Philippines, sandbox tests have accelerated regulator comfort with machine learning-based scoring by generating empirical evidence on loss prediction and fairness. Sandbox pilots help regulators understand model behavior and design proportionate frameworks that authorize data-driven credit scoring. Embedding sandbox learnings into permanent guidance can help overcome regulatory conservatism and promote responsible AI adoption at scale. Beyond data access and experimentation, policies that support gender-disaggregated reporting, borrower control over data sharing, and investments in foundational digital infrastructure are essential to building inclusive and trusted credit ecosystems.[17]

Open finance and data-sharing initiatives offer structural opportunities. The next wave of inclusion may depend on enabling data ecosystems and regulatory support for ethical data-sharing. Open finance frameworks have already been piloted in countries like Kenya and Brazil, providing lenders real-time access to borrower-permissioned data from digital wallets, utilities, or savings platforms (World Bank 2025, SME Finance Forum 2024).[18] In addition to regulatory sandboxes, supportive data governance structures will be key to scaling innovations safely and inclusively. Investing in gender-disaggregated monitoring, fairness metrics, and model audits is equally important to ensure AI supports, rather than undermines, equity (World Bank 2025, IFC 2024).

Open-finance infrastructure and AI complement each other.[19] While open-finance frameworks enable consent for the sharing of data across financial institutions, AI systems convert diverse data streams, such as banking, mobile money, utility, and platform activity into credit-relevant insights. With proper safeguards, this interoperability allows lenders to tailor products, price risk more accurately, and expand inclusion for thin-file borrowers. For example, AI can detect repayment capacity trends from utility and merchant transactions once APIs make them available. Conversely, open-finance standards improve AI explainability by ensuring structured, auditable data inputs. Together, they can lay the foundation for affordable, gender-responsive financial products built on trustworthy digital infrastructure

Targeted partnerships are essential to scale impact. Fintechs, mobile network operators, platforms, and traditional lenders must collaborate to enable data-sharing, distribution, and lending at scale. For example, telecommunications operators hold unique behavioral data that can power scoring models, while incumbent banks possess capital and customer bases. Development finance institutions can play a catalytic role in facilitating these cross-sector linkages and de-risking experimentation.

16 An ongoing evaluation (JPAL 2025) is also testing the impact of such models in the Dominican Republic
17 For a more detailed discussion of regulatory practices and recommendations, see Section 5 of The Use of Alternative Data in Credit Risk Assessment: Opportunities, Challenges, and Policy Recommendations (World Bank 2025).
18 See Salman (2021), Vidal and Sirtaine (2024), and Vidal and Plaitakis (2022) for a rich discussion on open finance and data sharing for inclusion.
19 See Kumaraswamy and Salman (2025) for a discussion on policy levers for adopting open-finance.

Photo: © AdobeStock

APPENDIX A:
TWO EXTENDED CASE STUDIES

1. Eshandi's Automated Microloan Scoring in Zambia: Gender and behavioral insights from AI and alternative data

Company Overview and Context

This case study draws on borrower-level data from Eshandi, a digital lender and credit scoring platform operating primarily in Zambia. The dataset covers over one million microloan applications processed in 2024, representing approximately 300,000 unique borrowers. Approximately 850,000 loans were approved with an average size of about $20**.** Each loan application was fully automated and evaluated through machine-learning models built on mobile money transaction histories, including airtime usage, digital wallet balances, and repayment patterns. This enables rapid scoring and approval decisions, particularly for first-time borrowers and microentrepreneurs who typically lack collateral or formal credit records.

This case study explores gender-related patterns in machine learning-based credit scores, loan approvals, and repayment-related indicators, using borrower-level data to understand how alternative data and algorithmic scoring models influence credit access and gender inclusion.

Loan approvals are determined algorithmically based on mobile money data, allowing credit access even for borrowers without formal financial records. Each borrower record contains a machine learning-based score and associated behavioral indicators derived from mobile money data, including:

- **Repayment behavior:** Measures such as *repayment hygiene* summarize a borrower's on-time payment record while *delinquency severity* covers the frequency or duration of overdue balances.
- **Balances and credit use:** Indicators such as *average balance*, *credit utilization*, and *balance stability* reflect liquidity patterns, volatility in digital wallet balances, and how actively borrowers use available credit.
- **Income and spending:** Variables including *income stability* and *spend ratio* capture the regularity of incoming transfers and the proportion of income that is spent or retained, serving as proxies for financial resilience.
- **Behavioral indicators:** Metrics like the *gambling score* and *gambling multiplier* identify engagement with betting or gaming platforms and the intensity of related spending relative to income. A higher multiplier typically signals riskier expenditure behavior. *Entertainment spending ratio* reflects discretionary spending on nonessential activities.

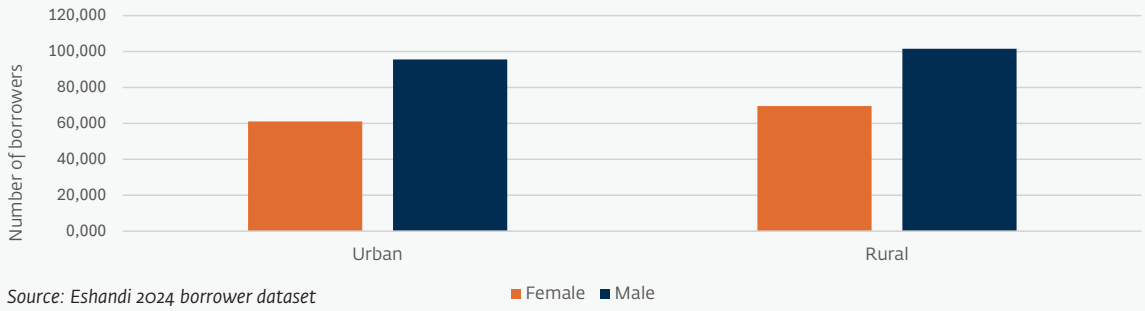
Together, these indicators feed into Eshandi's composite machine learning-based credit score, which predicts repayment risk and approval likelihood. The underlying credit-scoring engine combines 'transactional embeddings' with adaptive machine learning techniques to capture borrower behavior over time. Mobile-money and payment transaction histories are transformed into structured behavioral indicators, and analyzed using recurrent neural networks to detect spending and repayment patterns. These behavioral profiles are then grouped into borrower 'archetypes' through probabilistic clustering, enabling more nuanced segmentation than rule-based models. Finally, an ensemble classifier refines default-risk predictions through continuous feedback from repayment outcomes, allowing the model to adapt dynamically to changing borrower behavior and market conditions.

Findings

Women represent around 40 percent of all unique borrowers and over 45 percent of individual loan disbursements. The gender split varies across geography: rural borrowers form a particularly large share of Eshandi's client base, reflecting the accessibility of mobile money credit in areas underserved by traditional finance, as shown in Figure A1.

FIGURE A1

Borrowers by location and gender



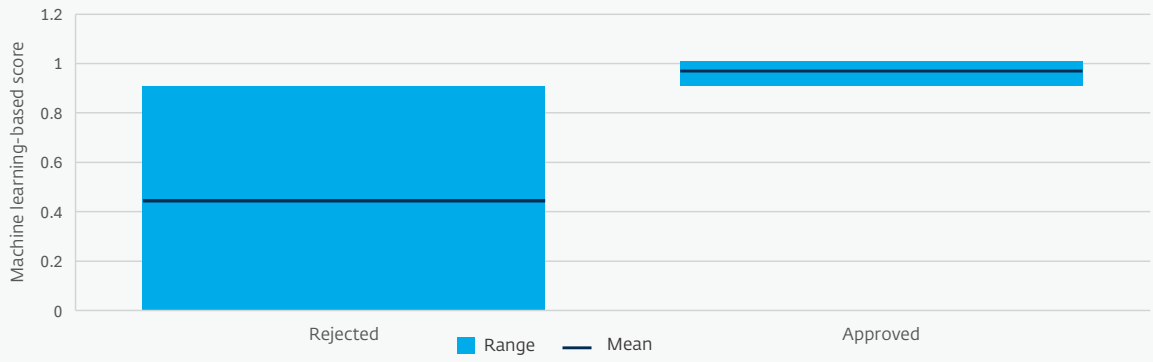
Source: Eshandi 2024 borrower dataset

Eshandi's disbursement process for microloans relies on its proprietary machine learning score, which serves as the decisive variable in approval decisions. Applicants receive scores between 0 and 1, and those above 0.9 are automatically approved.

Loan approvals are tightly linked to the machine learning-based score by design, with a clear threshold effect at 0.9. The distribution confirms that borrowers above this cutoff are automatically approved, while those below it are typically rejected—validating the model's internal consistency and automation logic. Figure A2 illustrates the strong threshold effect: approved borrowers cluster around machine-learning scores near 1.0, while rejected applicants average below 0.5

FIGURE A2

Machine-learning score distribution by approval status

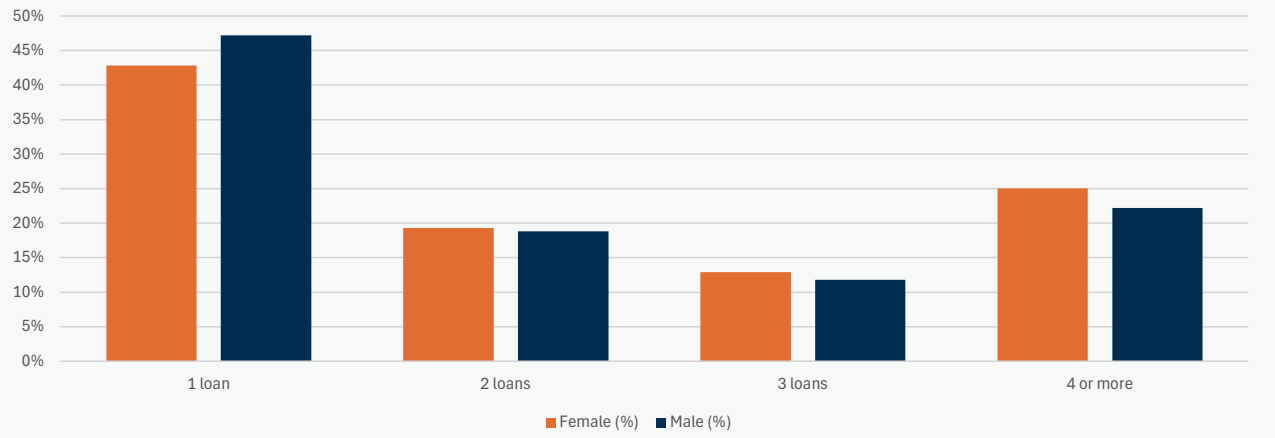


Source: Eshandi 2024 borrower dataset

Women and rural borrowers show stronger retention, with higher-scoring clients continue borrowing and building credit histories over time. As shown in Figure A3, women are more likely to have multiple approved loans than men: 25 percent of women borrowers had received four or more loans compared to 22 percent of men. Conversely, a slightly higher share of men had taken only one loan, suggesting that female borrowers tend to get higher approvals with Eshandi over time.

FIGURE A3

Number of loans approved by gender



Source: Eshandi 2024 borrower dataset

Higher machine-learning scores are associated with stronger client retention and repeat borrowing. Borrowers with higher scores are more likely to receive multiple loans over time, suggesting that Eshandi's scoring framework effectively identifies clients who continue to perform well after their first loan, as shown in regression estimates in Table A1.[20] While scores drive approvals by design, this pattern highlights how behavioral reliability—captured through transaction and repayment indicators—translates into sustained access to credit.

Women access significantly more repeat loans than men at about 0.15 additional loans on average, controlling for age, location, and the machine-learning score. Rural borrowers also display higher loan frequency, indicating that mobile-money-based scoring captures informal yet stable financial behaviors often invisible to traditional systems.

20 Eshandi's model is proprietary, and the specific variables used in score computation are not publicly disclosed. The analysis therefore relies on regressions that control for borrower characteristics such as age and location to interpret observed gender differences, rather than implying these variables are directly included in the model.

Table A1

Determinants of loan frequency/number of loans by gender and other characteristics

Characteristic	Number of Loans
Machine Learning-based Score	2.215***
	(0.101)
Female	0.150***
	(0.00821)
Rural	0.353***
	(0.00811)
Age	0.0265***
	(0.000389)
Other Machine Learning-based Scores (Controls)	YES
Constant	-0.709***
	(0.0975)
N	325993

Source: Eshandi 2024 borrower dataset
*Note: Results estimated using an ordinary least squares (OLS) regression model. Standard errors in parentheses. *p<0.1, ** p<0.05, ***p<0.01*

Among approved borrowers, higher machine learning scores are linked to larger disbursement amounts—especially for women. Conditional on approval, higher machine learning scores predict larger loan sizes, indicating that the model differentiates not only *who* gets credit but also *how much*. Regression results in Table A2 show that this relationship is stronger for women: each incremental increase in score is associated with a larger increase in average loan amount for female borrowers than for males. This suggests that the scoring framework is likely to be associated with women's stronger repayment behavior and converts it into more favorable credit terms.

Table A2

Loan Amount Regression Results by Gender

	1	2	3	4
Characteristic	**Disbursed Amount**	**Disbursed Amount**	**Disbursed Amount**	**Disbursed Amount**
	MALES	FEMALES	MALES	FEMALES
Machine Learning Combined Score	10.15 (9.489)	31.32*** (11.39)		
Gambling Activity			-9.693*** (2.735)	-3.947 (3.304)
Rural/Urban	YES	YES	YES	YES
Age	YES	YES	YES	YES
Constant	137.0*** (9.165)	116.3*** (11.00)	147.3*** (0.541)	146.7*** (0.637)
N	497529	349341	497529	349341

*Note: Results estimated using an ordinary least squares (OLS) regression model. Standard errors in parentheses. All monetary values are in Zambian kwacha (ZMW); approximately 22 ZMW = 1 USD (October 2025). *p<0.1, ** p<0.05, ***p<0.01*

Gambling activity is used as a proxy for risky behavior, and significantly reduces loan amounts for men but has no effect for women,. This suggests that the model identifies risk-taking patterns more frequently among male borrowers, while rewarding reliability and consistency more often observed among female borrowers.

Women score slightly higher on machine learning-based credit scores, with the strongest advantage among rural borrowers. Alternative data-based scores capture women's repayment reliability—especially in rural markets where traditional systems fall short. In Table A3, regression results examining the determinants of Eshandi's credit scores show that female borrowers score slightly higher overall, even after controlling for age and location**.** The gender coefficient remains positive and significant, indicating that the model captures stronger or more stable financial behaviors among women.

Breaking the score into its constituent sub-indices of income stability, balance stability, delinquency, and gambling activity reveals women perform comparably to men on other metrics such as balance, income, and delinquency metrics.

Table A3

Regression results of determinants of machine learning-based scores for rural borrowers only

	1	2	3	4	5
	Machine Learning-based Total Score	**Income Stability**	**Balance Stability**	**Delinquency Score**	**Gambling Multiplier**
Female	0.00061***	0.000254	-0.0002	-0.0001	-0.0002
	(0.0001)	(0.000181)	(0.0002)	(0.0002)	(0.0002)
Rural/Urban	YES	YES	YES	YES	YES
Age	YES	YES	YES	YES	YES
Constant	0.962***	0.0535***	0.0535***	0.0529***	0.053***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
N	325993	325993	325993	325993	325993

*Note: Results estimated using an ordinary least squares (OLS) regression model. Standard errors in parentheses. *p<0.1, ** p<0.05, ***p<0.01*

Regression results in Table A4 show that among rural borrowers, women's advantage is more pronounced. They score higher on the overall machine learning-based credit score and income stability, with no evidence of higher delinquency or risk behavior compared to men. These findings reinforce the idea that alternative data can better recognize women's repayment reliability, rather than penalizing informality or smaller transaction volumes.

Table A4

Regression results of determinants of machine learning-based scores for rural borrowers only

	1	2	3	4	5
	Machine Learning-based Total Score	**Income Stability**	**Balance Stability**	**Delinquency Score**	**Gambling Multiplier**
Female	0.0012***	0.00047*	-0.0001	0.00007	-0.0001
	(0.0003)	(0.00025)	(0.00025)	(0.00025)	(0.00025)
Age	YES	YES	YES	YES	YES
Constant	0.957***	0.053***	0.0531***	0.0532***	0.0527***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
N	170465	170465	170465	170465	170465

Source: Eshandi 2024 borrower dataset
*Note: Results estimated using an ordinary least squares (OLS) regression model. Standard errors in parentheses. *p<0.1, ** p<0.05, ***p<0.01*

Key Insights and Implications

1. **Alternative data can reveal women's financial reliability:** The model consistently rates women, especially rural women, slightly higher on machine-learning based scores, capturing repayment discipline, income stability, and low-risk spending patterns that traditional systems overlook.
2. **Microloans can serve as a steppingstone toward formal finance:** Although individual loans remain small, repeat borrowing and high repayment rates point to a gradual 'graduation' effect. Borrowers who perform well under these models can establish digital credit histories that open doors to larger and more conventional credit products over time.
3. **Behavioral indicators can promote fairness when well-designed:** Data-driven measures of financial behavior, such as income regularity and gambling activity, help differentiate between genuine credit risk and informal financial behavior. While gambling scores penalize men more frequently, this reflects behavioral heterogeneity rather than algorithmic bias—showing that alternative data can improve rather than distort inclusion outcomes.

These results provide early empirical support for the potential of AI and alternative data-driven credit systems to advance women's financial inclusion, especially in emerging markets with widespread mobile money use and limited traditional credit infrastructure.

2. Vexi Customer Insights: Gender patterns in credit access, business use, and confidence in borrowing

Company Overview and Context

Vexi is a Mexico-based fintech offering credit cards designed to expand formal credit for underserved populations, particularly young and first-time borrowers. Vexi's use of alternative data for credit scoring and a fully digital onboarding process have helped it grow rapidly in recent years, and provide a unique window into how inclusive fintech models can bridge access gaps in credit markets.

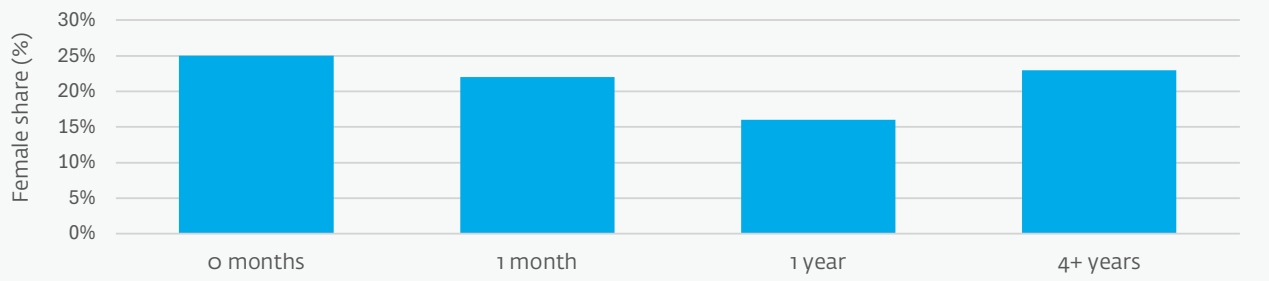
In July and August 2025, Vexi conducted a series of customer surveys across different tenures, ranging from newly-onboarded users to those who had held a card for more than four years. The surveys were designed to capture borrower experiences, satisfaction, and perceptions of credit use and access over time. Collectively, the surveys reached more than 7,000 respondents, enabling comparisons across gender and tenure cohorts. The August survey incorporated additional questions on credit scoring awareness, confidence in applying for future credit, and perceived impacts of using Vexi products on credit history and business outcomes.

Although the findings are based on self-reported survey data, they broadly align with Vexi's internal customer metrics—particularly for indicators such as average credit limits and gender composition. The consistency across rounds and tenures suggests that the observed gender patterns are not driven solely by selection effects or survey composition.

Women represent a stable, although smaller, share of Vexi's customer base. Across survey rounds, women represented roughly one-quarter of new customer respondents, with shares ranging between 15–20 percent across longer tenures. This distribution is comparable across survey rounds, indicating a stable share of female borrowers within Vexi's customer base, as shown in Figure A4.

FIGURE A4

Share of female respondents by customer tenure with Vexi



Source: Primary survey data collected in July and August 2025

Findings

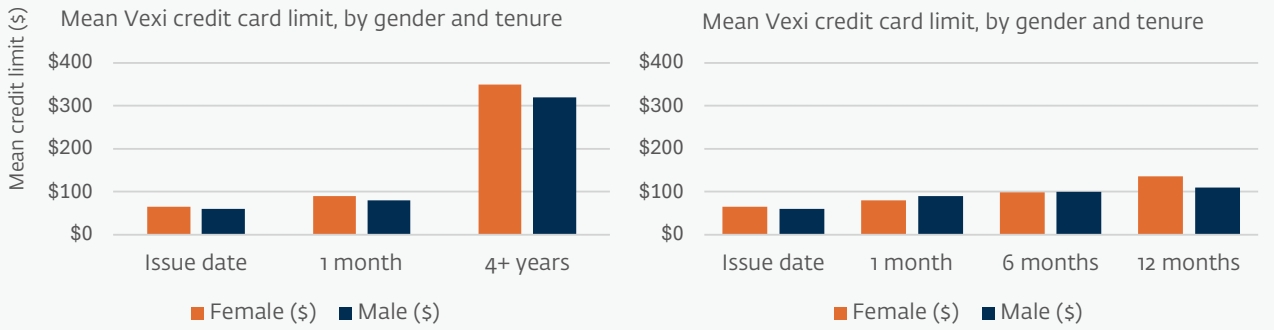
Credit Profiles and Gender Composition

When their accounts first open, women and men have comparable credit limits, but they rise faster for women over time. Both women and men started with credit limits of roughly $65 (MXN 1,200, but by the first month, the average limit for women begin to rise at a slightly faster rate. Among customers with four or more years of tenure, women's mean credit limit reaches roughly $350 (MXN 6,400) compared to $320 (MXN 5,900) for men, as shown in Figure A5. Similar patterns emerged in the August round, where women's average limit after 12 months exceeded that of men, at $134 versus $110 for men).

Higher limits over time suggest women's responsible repayment patterns and creditworthiness are initially underestimated. Median values follow a similar trajectory, underscoring that the observed gaps are not driven only by a small number of high-limit accounts.

FIGURE A5

Mean credit limits by gender and tenure



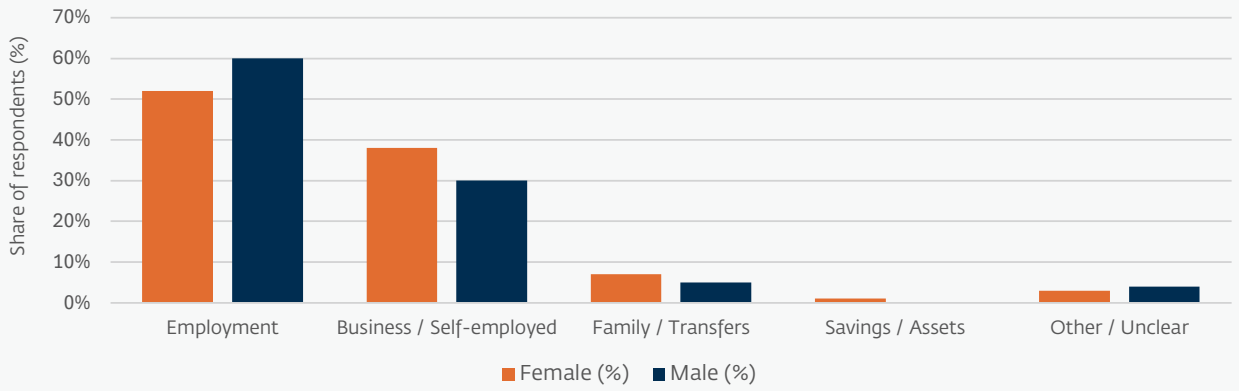
Sources: Left: Survey responses August 2025 (N=4213); Right: Vexi survey responses July 2025 (N=1395)

Higher limits among women align with broader evidence of lower default rates and stronger portfolio performance. The divergence in credit limits by tenure aligns with broader evidence that once given access, women borrowers in digital credit markets tend to exhibit lower default rates and improve their standing over time. For Vexi's customer base, these findings highlight the importance of retention and credit-building features for sustaining inclusive growth in credit access.

Women borrowers are more likely to be self-employed or entrepreneurs than men. Among new customers, men are more likely to report salaried employment as their main source of income, as shown in Figure A6. In contrast, a higher share of women—around 37 percent compared with 30 percent of men in July 2025—report that their primary income comes from self-employment or running a business. This gender gap in income type is consistent with Vexi's broader customer data and reflects a mix of necessity and opportunity entrepreneurship among women borrowers.

FIGURE A6

Main income source by gender

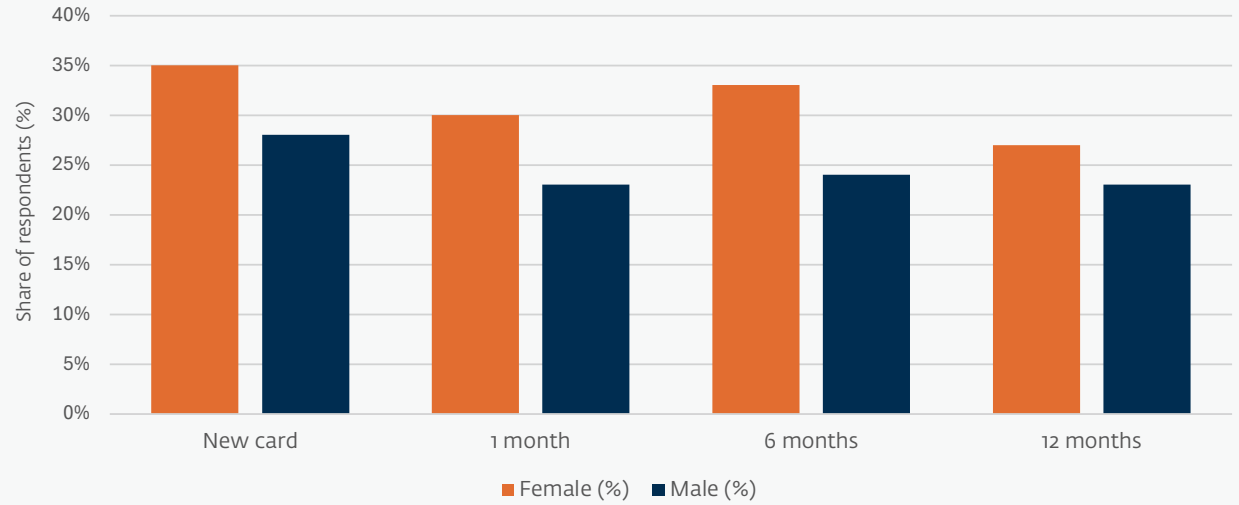


Source:: July 2025 survey of new Vexi users (N=907)

For many women, credit cards serve as working-capital tools rather than simply consumption instruments. The importance of business-related income remains evident even as customers mature, as shown in Figure A7. Across different tenures, women consistently report business activity as a key income source—suggesting both continued entrepreneurial engagement and a sustained reliance on credit cards for managing short-term business cash flows.

FIGURE A7

Share of customers reporting business income by tenure and gender



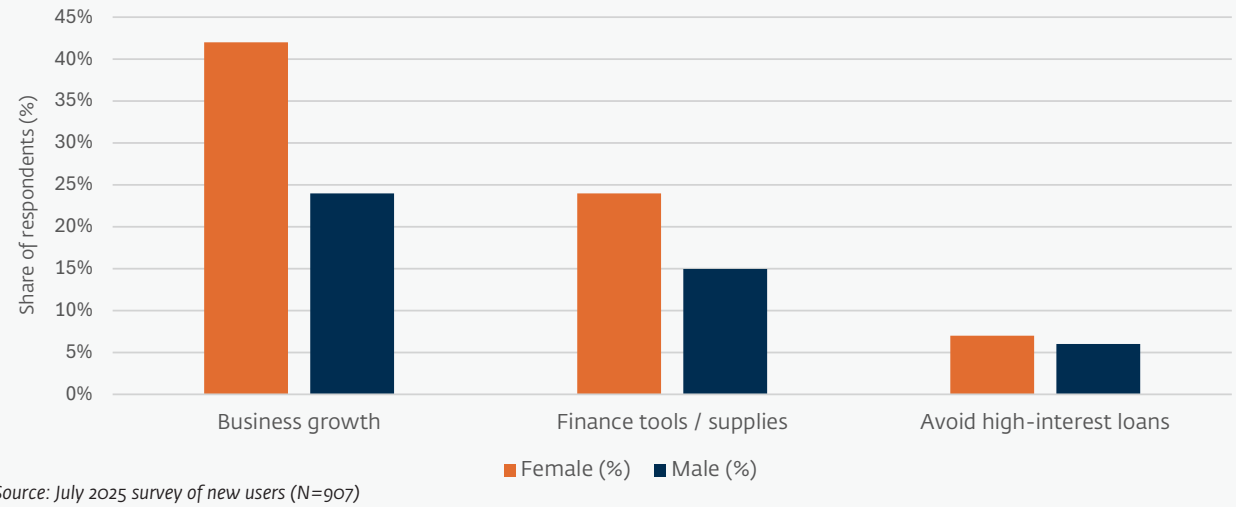
Source: August 2025 survey data (N=4213)

Women's use of credit cards for business persists over time, pointing to ongoing financing constraints. While men's reliance on business income remains stable at around one-quarter across tenures, women's share starts higher among new customers at 36 percent, and remains elevated even after one year of tenure. This persistence may also signal limited access to other credit instruments for women-owned businesses, making Vexi cards an important entry point into formal credit.

Women use Vexi cards for business far more than men, highlighting unmet financing needs. For many users—particularly women—the Vexi card plays a dual role as both a personal and business financing tool. Among those who report owning or having a family business, 62 percent of women indicated that they use their Vexi card for business-related expenses, compared to 44 percent of men. This pattern underscores both the opportunity and the financing gap faced by women entrepreneurs, who are often excluded from formal business credit channels and instead rely on consumer financial products to sustain or grow their operations.

FIGURE A8

Impact of Vexi cards on family businesses



Source: July 2025 survey of new users (N=907)

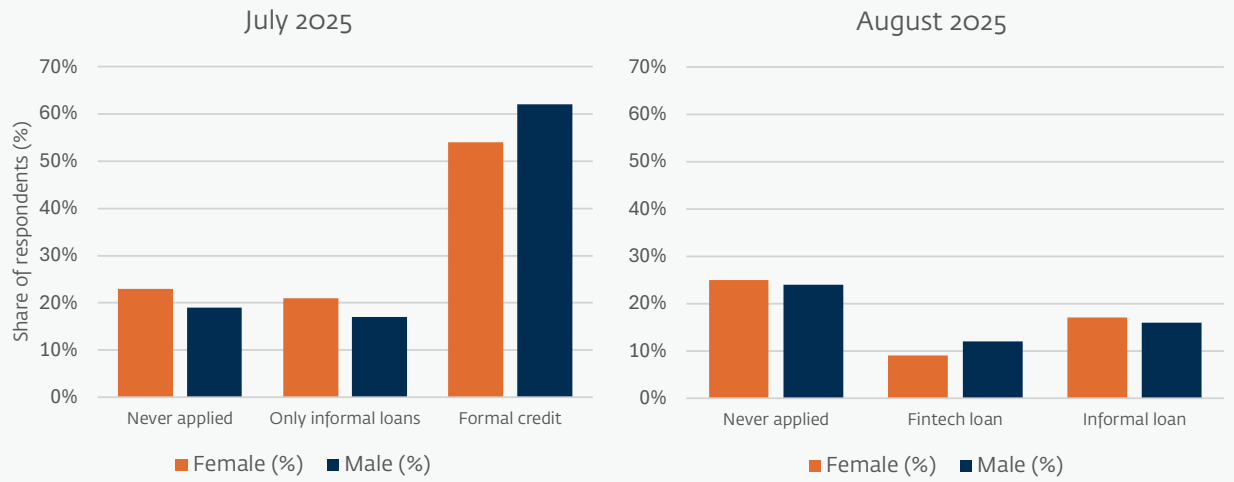
Vexi cards support women's business growth and investment in productive assets. Around 42 percent of women business owners stated that the card had contributed to business growth versus 24 percent of men, and nearly one in four women said it had helped them purchase business-related inputs, as shown in Figure A8.

Despite these gains, women remain reliant on high-cost or limited credit options for scaling. Only about 7 percent of women and 6 percent of men reported that the card had helped them avoid high-interest loans. This suggests that many women entrepreneurs continue to face constraints in accessing affordable financing for scaling their enterprises. Relatively higher business use and reported impact among women indicates untapped potential for gender-tailored credit products that better match the needs of women-owned microenterprises, such as working capital lines or business-specific cards.

Women borrowers were less likely to have accessed formal credit before joining Vexi. Among new customers, women were less likely than men to have ever applied for formal credit and more likely to have relied on informal sources of credit. Based on the survey conducted in July, more women depended on informal loans, while in August a similar share of women reported that they relied exclusively on informal borrowing, slightly higher than among men at 17 percent versus 16 percent. Meanwhile, only 9 percent of women had previously accessed fintech credit, compared to nearly 12 percent of men, as illustrated in Figure A9.

FIGURE A9

Customers who applied for credit prior to Vexi



Sources: Left: July 2025 data (N=907 new customers); Right: August 2025 data (N=2776 new customers)

Approval rates for women in formal banking remain lower than for men. Only 15 percent of women who had applied for bank loans were approved compared to 19 percent of men. This reflects persistent gender biases in traditional credit systems and higher entry barriers for women.

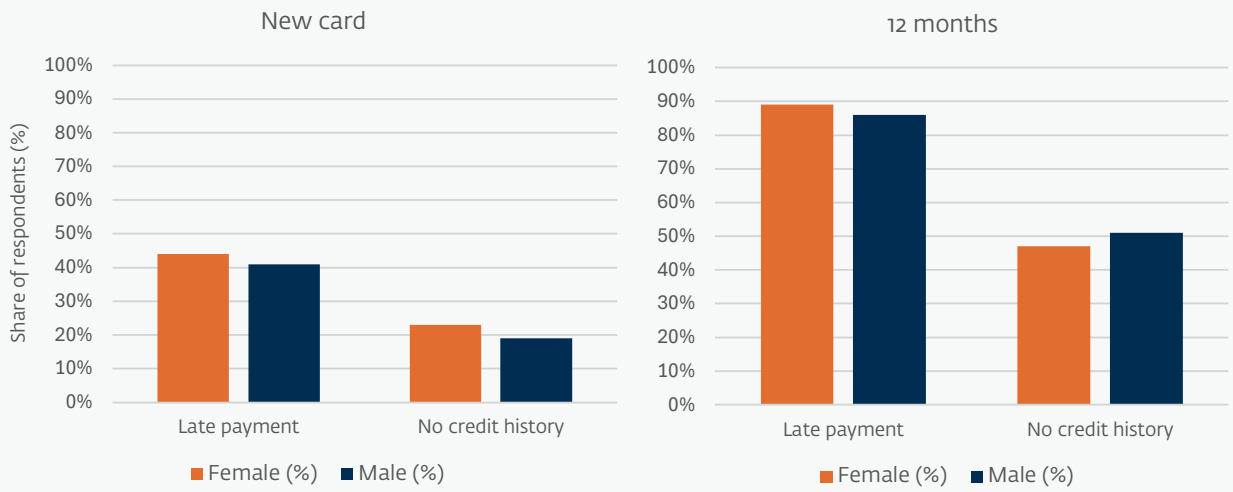
For many women, Vexi represents their first formal credit experience and a bridge to mainstream finance. The platform's fully digital onboarding process and alternative data-based credit assessments appear to be lowering entry barriers for borrowers who might otherwise be excluded from mainstream credit markets. This also underscores the potential for fintech models to complement rather than replace traditional lending channels, by providing initial access and helping borrowers build credit histories over time.

Engagement with Vexi improves customer understanding of how credit scores work. Among new customers, fewer than half of respondents recognized late payments as key to credit scores. Their awareness of other factors, like lack of prior history, was even lower.

Within one year, credit awareness nearly doubles for both women and men. After 12 months of having a credit card, nine in ten customers correctly identified late payments as a major score determinant, and nearly half understood the role of maintaining a credit history, as illustrated in Figure A10. These shifts highlight the value of embedding education and feedback mechanisms in fintech lending models, and how this helps new borrowers understand and navigate formal credit systems.

FIGURE A10

Perceptions of factors that influence credit scores, by tenure and gender

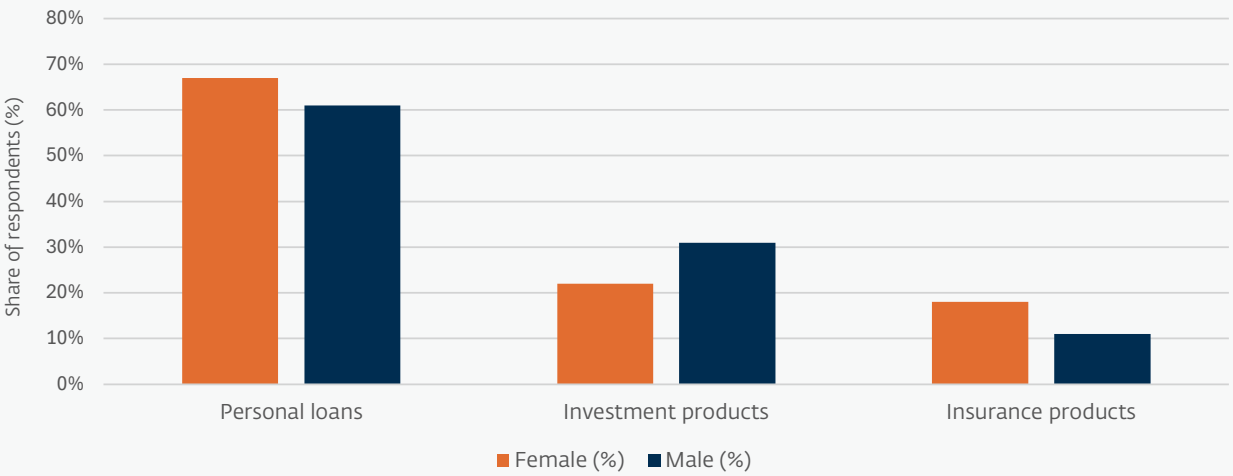


Source: August 2025 survey (Left: N=644; Right: N=389)

Women express greater demand for personal loans and insurance, while men favor investments. Roughly two-thirds of female respondents and 61 percent of male respondents listed personal loans as their most sought-after product, as shown in Figure A11.

FIGURE A11

Interest in additional services, by gender and product



Source: July 2025 survey (N=349)

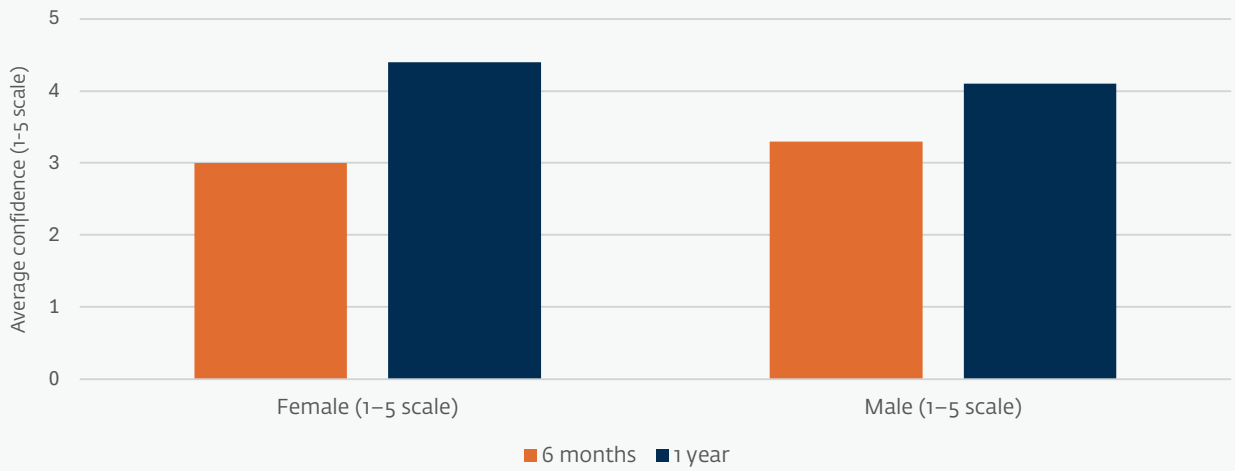
Gendered preferences reflect differing financial priorities and exposure to risk. Women's stronger interest in personal loans and insurance suggests a focus on liquidity and income security consistent with their higher representation among self-employed and business-owning respondents. Men's relatively greater preference for investment products may reflect higher disposable income and more confidence in engaging with longer-term financial instruments.

Strong demand for additional products also signals a financing gap beyond credit cards. For many, particularly women, credit cards represent an accessible entry point into formal finance but are insufficient to meet business or household credit needs. These insights point to the potential value of integrated product offerings that can support customers' progression along the credit lifecycle, such as personal loan upgrades, bundled insurance, or credit-for-business features.

Confidence in applying for future credit increases sharply with tenure, especially for women. Among respondents who had been with Vexi for six months, the average confidence score was 3.0 for women and 3.3 for men on a five-point scale. After one year, confidence levels increased markedly for both genders, to 4.4 for women and 4.1 for men, as shown in Figure A12. This suggests growing familiarity with formal credit systems and improved self-perceptions of creditworthiness.

FIGURE A12

Confidence in applying for future credit by gender and tenure



Source: August 2025 survey (N=793)

Women's confidence grows faster, driven by positive experiences and stronger credit histories. Eight-five percent of women report that Vexi helped them build or improve their credit history compared to 78 percent of men, pointing to the psychological and practical benefits of inclusive credit design.

Key Insights and Implications

1. **Digital Credit Can Strengthen Women's Financial Confidence.** Survey results show that women not only access digital credit at similar or higher rates than men but also report larger gains in confidence when applying for new credit. This suggests that fintech-enabled borrowing can play a formative role in women's perceived creditworthiness and financial capability.
2. **Credit Use Extends Beyond Consumption to Business Investment.** Over 60 percent of female borrowers report using credit cards for microenterprise expenses—possibly treating revolving credit as working capital. This challenges the perception of digital credit as primarily consumption-driven and highlights its role in small-business resilience.
3. **Responsible Design and Gradual Limit Increases Support Inclusion.** Women's higher average credit limits after one year indicate the value of incremental risk-based scaling. Vexi's model that uses behavioral and payment-history data to adjust limits illustrates how alternative data can reward consistent repayment and help women build stronger financial profiles.
4. **Lenders and Policymakers can Encourage Stronger, More Inclusive Data Collection.** Embedding gender-disaggregated tracking in customer analytics can uncover similar patterns elsewhere. Regulators and investors could encourage fintechs to monitor gender outcomes explicitly, ensuring that digital credit expansion translates into measurable empowerment rather than over-indebtedness.

APPENDIX B:
TRADITIONAL VERSUS ALTERNATIVE DATA-DRIVEN CREDIT SCORING MODELS

Categories	Traditional Data	Alternative Data
Data Sources and Variables	Credit history, credit mix such as loans and credit cards, utilization, repayment rates, assets, liabilities. Age, education, occupation (included in newer models).	Online/social media activity, utility and rental payments, grocery store behavior, app usage, GPS data, SMS logs, personality traits, and more.
Data Collection	Credit reports from bureaus that use standardized metrics like FICO score and VantageScore, such as Equifax, Experian, and TransUnion.	Modeled using existing firm data or from embedded platforms such as e-commerce or gig platforms, often supported by financial analytics firms.
Modeling and Accuracy	Established models/metrics based on historical data assess creditworthiness with rich financial history.	AI-based models offer better predictive capability, capturing real-time financial behaviors and emerging credit patterns.
Reach/Coverage	Rigid classifications exclude borrowers without formal documentation or credit history.	Can serve underserved populations, including thin-file and no-file applicants; Adapts to diverse borrower profiles, including gig workers, informal economy participants, and rural borrowers.
Advantages	Stable, well-understood metrics with established regulatory compliance.	Expands financial inclusion, provides real-time insights, enhances credit prediction for specific populations.
Concerns	Limited view of financial behavior, dependence on historical data, and potential failure to capture recent changes.	Privacy concerns, ethical considerations, often unclear regulatory standards, and the 'black-box' nature of some AI-driven models.
Costs and Efficiency	Lower costs due to standardized processes.	Higher initial costs and implementation time but reduced operational costs in the long term.
Data Granularity	Broad, standardized financial indicators.	Highly detailed, personalized behavioral and transactional data.
Data Frequency	Periodic updates, for example monthly or quarterly.	Possibility of real-time or near-real-time data collection and analysis, depending on the data source.
Regulatory Framework	Well-established regulations and compliance standards.	Evolving regulatory landscape, with potential for new guidelines.
Fraud Detection	Limited to patterns in formal financial transactions.	Enhanced ability to detect unusual behavioral patterns across diverse data sources.
Implementation Scalability	Easily integrates into legacy banking systems with minimal changes.	Requires advanced infrastructure and partnerships but offers highly customizable solutions.
Ethical and Social Considerations	Primarily financial in focus, with limited attention to ethical or social implications.	Raises concerns over data ownership and consent but also creates opportunities for environmental, social, and governance-aligned financial inclusion, particularly for female borrowers.

APPENDIX C:
SOURCES, EXAMPLES. AND USE CASES OF ALTERNATIVE DATA

Data Type	Typical Sources and Examples of Data Points	Relevance / Use Case Examples
Mobile Phone Metadata	Call/SMS logs (frequency, contacts), airtime top-ups, handset type, battery status.	Predicts repayment capacity from communication habits and spending consistency.
Mobile Money & Digital Wallet Transactions	Deposits, transfers, peer-to-peer payments, merchant payments.	Captures informal cash flows and payment behavior.
Utility & Bill Payment Data	Prepaid/postpaid electricity, water, gas, internet bills.	Signals reliability in recurring payments
Platform-based Behavior	Ride-hailing, e-commerce, delivery or freelancer platforms: frequency, ratings, cancellations.	Indicates work activity or business performance on digital platforms.
Psychometric Scores	Surveys testing traits like responsibility, planning, optimism, and honesty.	Used in low-data environments, especially for women micro-entrepreneurs.
Device & App-level Data	App usage patterns, screen time, geolocation, operating system version, contact list size.	Suggests smartphone literacy, digital integration, or use of specific apps, for example whether the app is for gaming or for educational purposes.
Social Graph/ Network Data	Size and structure of contact network, interaction frequency.	Used as a proxy for social capital, connectedness, and accountability.
Geospatial & Satellite Data	Farm boundaries, housing density, road access, lighting at night.	Used for MSMEs, agri-lending, and asset ownership estimation. Especially useful in rural settings.

APPENDIX D:
METHODOLOGY FOR STAKEHOLDER CONSULTATIONS AND LANDSCAPE MAPPING

Stakeholder Consultations

Over 30 stakeholders were interviewed, including fintech firms, alternative credit scoring providers, and specialists from IFC, IBRD, and other institutions. This included firms with a range of models, such as direct lenders, score-only providers, and embedded finance enablers, serving any emerging market region. Firms were selected to reflect variation in customer base, data sources, and levels of gender intentionality. These interviews provided insights into business models, innovation strategies, and regulatory or operational challenges that are not always documented in published literature.

Landscape Mapping

To map the global ecosystem of firms using non-traditional data in credit scoring, several sources were used. This included systematic searches for firms tagged as 'alternative credit scoring, 'embedded lending,' and 'alt data marketplace' in the market intelligence database Tracxn and the Cambridge Centre for Alternative Finance. The research team also identified firms that participated in SME Finance Forum sessions or featured on IFC project rosters. Large language machine-based searches were used to identify 'firms like X,' supplemented by snowballing from accelerator cohorts and partner references. The initial longlist comprised over 800 firms.

Firms were included if they met the following**:**

- **Developed or directly used a credit-scoring model that incorporates non-traditional data** such as mobile-money transactions, device metadata, or behavioral or platform footprints in their own underwriting or credit-assessment processes.
- **Offered these models or scores as a product or service** to other lenders or used them internally to issue loans through their own lending operations.

Firms that only partnered with scorers or those without sufficient public documentation of scoring logic were excluded. Similarly, large credit bureaus and banks were excluded unless they had distinct subsidiaries or entities offering alternative data-based scoring products.

After applying these criteria, 448 firms remained for which additional information could be obtained online. These were tagged across the following dimensions:

- **Business model:** For example, direct lender, embedded finance enabler, marketplace platform.
- **Types and sources of data used:** Mobile money, transaction history, telecommunications metadata, platform activity, or psychometric inputs.
- **Types of loans facilitated directly or supported through their scores:** This incudes personal, MSME, and buy now, pay later credit.
- **Market framing:** Whether the firm explicitly mentioned targeting thin-filed, underserved, or low-income borrowers, or referenced women and female borrowers in its public facing communications, such as its mission or vision statement.

Regions of operation were categorized based on IFC regional groupings. A single firm can operate in multiple regions. In some cases, its headquarters might not indicate region of operation. For example, Tala is based in the United States but its services are only offered

in Kenya, India, and other countries outside the US. When information was lacking, the headquarters was used to mark the region of operation. Where available, data on founding year, funding stage (unfunded, Seed, Series A–C, Series D+/exit, deadpooled), and recent funding activity, firm descriptions, and acquisition status were matched using Tracxn.

Tags were generated through a combination of ChatGPT-assisted extraction (using the o3 model) and manual verification. After this step, all the firms were manually checked. Deadpooled firms made up approximately 5 percent of the sample and were retained in the main analysis because the small sample size would not substantially affect trends and they held valuable information on business models.

This mapping offers a snapshot of alternative credit scorers in emerging markets but does not claim to be exhaustive and has some limitations:

The landscape likely reflects firms with greater public visibility, such as those with stronger search engine optimization presence, more mature websites, or some degree of market traction. As such, smaller, early-stage, or regionally-contained firms, especially those without an online footprint, may be underrepresented.

The share of unfunded or deadpooled firms in the sample is not high, likely reflecting the fact that the sample is drawn from firms that have some degree of traction already. In contrast, a narrower set of firms historically tagged in Tracxn as 'alternative credit scoring' firms suggest a higher deadpooled share of about 24 percent. While that data may capture older firms that are no longer active, it may still underrepresent the smallest firms or unfunded firms with limited public information.

The presence and type of alternative data sources, for example, mobile money, social media, platform activity, and the nature of firm operations were recorded based on public disclosures. Hence, the intensity or centrality of such data in the model design was often unclear. Some firms may use alternative data only marginally, and the degrees of reliance on it are not captured.

The analysis erred on the side of inclusion, retaining firms even when the scoring logic was opaque but likely to involve non-traditional data.

Glossary

AI-driven Scoring: Credit-decision systems where machine Learning-based algorithms analyze large datasets - often including alternative data - to generate credit scores or lending recommendations.

Agentic AI: AI systems that can take actions on their own—such as running tasks, making recommendations, or interacting with other software—without needing step-by-step human instructions. In finance, agentic AI could automate parts of loan processing or monitoring, while still operating within rules set by humans.

Alternative Data: Non-traditional information used to assess creditworthiness, such as mobile money transactions, utility bill payments, platform earnings, or behavioral signals, rather than credit-bureau records.

Application Programming Interface (API): A software connector enabling secure data exchange between systems - these are central to open-finance ecosystems.

Area Under the Curve (AUC): A model-performance metric (0–1) showing how well a score distinguishes good from bad borrowers; higher AUC indicates better predictive accuracy.

Bias: Algorithmic or data bias is systematic distortion in model outputs caused by unequal representation or historical discrimination in training data.

Buy Now Pay Later (BNPL): Short-term consumer credit allowing installment payments for purchases, often embedded in e-commerce or digital wallet platforms.

Credit Scoring Model: A statistical or algorithmic tool estimating the probability that a borrower will repay a loan. This is used to arrive at lending decisions.

Data Privacy / Consent Framework: Rules and systems ensuring that personal data are used only with informed consent and adequate protection.

Data Visibility vs. Data Availability: Availability refers to whether digital traces exist; visibility means lenders can legally and technically access and interpret them.

Deadpooled: Industry term for startups that have shut down or become inactive.

Deep Learning: An advanced form of machine learning using layered neural networks, often for processing text, voice, or image data. Rarely used in current credit-scoring models in emerging markets.

Default / Non-performing Loan (NPL): When a borrower fails to make payments as scheduled; NPLs are typically loans more than 90 days overdue.

Digital Redlining: A form of algorithmic exclusion where models unintentionally deny access to specific groups (e.g., women, rural borrowers) due to biased data patterns.

Embedded Finance: The integration of credit, insurance, or payments directly within non-financial platforms (e.g., ride-hailing or retail apps).

Explainability (XAI): The degree to which an AI system's reasoning can be understood and explained to users, regulators, or borrowers.

Fair-lending Regulation: Legal frameworks aimed at preventing discrimination in credit decisions, for example the US, Equal Credit Opportunity Act and Fair Credit Reporting Act.

Fairness Audit / Fairness Metrics: Analytical checks (e.g., approval-rate parity) to ensure AI models produce equitable outcomes across gender or other groups.

Feature Engineering: The process of transforming raw data (e.g., mobile transactions) into measurable indicators or features that can be used in a model.

Funding Round: Sequential stages of external investment that signal a firm's growth and valuation milestones, such as Seed, Series A–C, etc.

Gamification / Model Gaming: Borrower behavior that manipulates digital signals (e.g., artificial spending) to raise a credit score.

Gradient Boosting: A technique that builds decision trees one after another, with each new tree trying to fix the mistakes of the previous ones. This step-by-step learning makes it one of the most accurate methods for predicting the likelihood of repayment.

Sex-disaggregated Data: Data broken down by sex to identify differences in outcomes and support gender-inclusive analysis.

Gender-segmented Model: A credit model trained separately for men and women to test fairness and improve predictive accuracy.

Intersectionality: Recognition that individuals experience overlapping forms of disadvantage (e.g., women with a disability) that can compound exclusion.

Invoice-based Financing: Lending where cash advances are made against unpaid invoices; repayment comes from buyer payments.

Know Your Customer (KYC): Verification process financial institutions use to confirm customer identity and prevent fraud or money laundering.

Machine Learning (ML): A subset of AI that enables models to learn relationships from data and improve predictions automatically without explicit programming.

Neural Networks: A machine-learning technique inspired by how the human brain processes information. Neural networks learn patterns by passing data through layers of interconnected "neurons," each transforming the input slightly before sending it to the next layer. In credit scoring, they can uncover complex relationships—such as how combinations of spending habits, repayment timing, and mobile-money activity predict default—but they are often less transparent than simpler models.

Open Finance / Open Banking: Policy frameworks that let consumers share financial data securely across providers via standardized APIs to access tailored products.

Point-of-Sale (POS) Data: Records from card readers or digital-sales devices capturing merchant revenues and transaction frequency.

Precision: A metric that measures how accurate a model's positive predictions are. In credit scoring, precision answers the question: "Of all the borrowers the model predicted would repay, how many actually did?" High precision means the model makes few 'false positives' (e.g., approving borrowers who then default).

Proxy Variable: An indirect indicator used in place of a harder-to-measure concept, for example, phone recharge frequency as a proxy for income regularity.

Random Forests: A machine learning method that combines many small decision trees and lets them 'vote' on the final prediction. Because it averages across many trees, it tends to be stable and good at spotting patterns that are not obvious in traditional credit scoring.

Recall: A metric that measures how many of the true positive cases the model successfully identified. In credit scoring, recall answers: "Of all the borrowers who actually repaid, how many did the model correctly flag as good borrowers?" High recall means the model misses fewer eligible borrowers.

Regularized Logistic Regression: A simpler, more transparent model that estimates the probability a borrower will repay a loan. Regularization adds a light penalty to avoid overfitting, making the model more reliable when there are many indicators or noisy data.

Regulatory Sandbox: A controlled environment where regulators and fintechs test innovations under relaxed rules before full market approval.

Revenue-based Financing: Loan repayment structure where borrowers repay as a fixed percentage of revenue rather than a set installment.

Software-as-a-Service (SaaS): A business model where software is offered on a subscription basis via cloud platforms; common among scoring providers.

Supervised Learning: A type of ML method that trains algorithms on labeled data (e.g., "repaid" vs. "default") to predict outcomes for new cases.

Telco Data: Metadata from mobile-network operators, such as SIM age, call frequency, or recharge patterns, used as alternative credit indicators.

Thin-file Borrower: Individuals lacking sufficient credit-bureau history for traditional scoring, often women, youth, or informal-sector workers.

Unsupervised Learning: ML method that identifies patterns or clusters in data without labeled outcomes, sometimes used for segmenting borrowers or detecting fraud.

White-labeled (Digital Lending Platform): A modular software product developed by a fintech and rebranded by partner financial institutions as their own. The provider builds the underlying infrastructure (e.g., onboarding flows, scoring engines, loan management systems), while banks or MFIs customize the interface, branding, and workflows. This allows institutions to launch digital lending products quickly without developing the technology in-house.

References

Alibhai, Aly Salman, Rachel Cassidy, Markus P. Goldstein, and Sreelakshmi Papineni. 2023. *Evening the Credit Score? Impact of Psychometric Loan Appraisal for Women Entrepreneurs.* Policy Research Working Paper 10230. Washington, DC: World Bank. https://documents.worldbank.org/curated/en/099441511142232337.

Anand, Malika. 2022. *Can Platform Data Expand Access to Transformative Credit?* June 8. https://www.cgap.org/blog/can-platform-data-expand-access-to-transformative-credit.

Berg, Tobias, Valentin Burg, Ana Gombović, and Manju Puri. 2020. "On the Rise of FinTechs: Credit Scoring Using Digital Footprints." *The Review of Financial Studies* 33 (7): 2845–97. https://doi.org/10.1093/rfs/hhz099.

Björkegren, Daniel, and Darrell Grissen. 2020. "Behavior Revealed in Mobile Phone Usage Predicts Credit Repayment." *The World Bank Economic Review* 34 (3): 618–34. https://doi.org/10.1093/wber/lhz006.

Brock, J. Michelle, and Ralph De Haas. 2023. "Discriminatory Lending: Evidence from Bankers in the Lab." *American Economic Journal: Applied Economics* 15 (2): 31–68. https://doi.org/10.1257/app.20210180.

Chen, Gregory, and Xavier Faz. 2015. "The Potential of Digital Data: How Far Can It Advance Financial Inclusion?" *Focus Note* 100. Washington, DC: CGAP. https://www.cgap.org/sites/default/files/Focus-Note-The-Potential-of-Digital-Data-Jan-2015_1.pdf

Chen, S., S. Doerr, J. Frost, L. Gambacorta, and H. S. Shin. 2023. "The Fintech Gender Gap." *Journal of Financial Intermediation* 54 (April): 101026. https://doi.org/10.1016/j.jfi.2023.101026.

Chioda, Laura, Paul Gertler, Sean Higgins, and Paolina C. Medina. 2024. "FinTech Lending to Borrowers with No Credit History." Working Paper No. 33208. Working Paper Series. National Bureau of Economic Research, November. https://doi.org/10.3386/w33208.

Demirgüç-Kunt, Asli, Leora Klapper, Dorothe Singer, and Saniya Ansar. 2022. *The Global Findex Database 2021: Financial Inclusion, Digital Payments, and Resilience in the Age of COVID-19.* Washington, DC: World Bank. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099818107072234182

Deshpande, Rani. 2020. "Financial Services for Gig Workers: An Intersection of Needs in Kenya." CGAP, June 4. https://www.cgap.org/blog/financial-services-for-gig-workers-intersection-of-needs-in-kenya.

Di Maggio, Marco, Dimuthu Ratnadiwakara, and Don Carmichael. 2022. "Invisible Primes: Fintech Lending with Alternative Data." Working Paper No. 29840. Working Paper Series. National Bureau of Economic Research, March. https://doi.org/10.3386/w29840.

Federal Reserve. 2021. "Speech by Governor Brainard on Supporting Responsible Use of AI and Equitable Outcomes in Financial Services." https://www.federalreserve.gov/newsevents/speech/brainard20210112a.htm.

FICO. 2024. "How to Use Alternative Data in Credit Risk Analytics." November 4. https://www.fico.com/blogs/how-use-alternative-data-credit-risk-analytics.

Gambacorta, Leonardo, Yiping Huang, Han Qiu, and Jingyi Wang. 2024. "How Do Machine Learning and Non-Traditional Data Affect Credit Scoring? New Evidence from a Chinese Fintech Firm." *Journal of Financial Stability* 73 (August): 101284. https://doi.org/10.1016/j.jfs.2024.101284.

IFC. 2024. *Her Fintech Edge*. International Finance Corporation. https://www.ifc.org/en/insights-reports/2024/her-fintech-edge-market-insights-for-inclusive-growth.

IFC. 2025. *MSME Finance Gap*. International Finance Corporation. https://www.smefinanceforum.org/sites/default/files/Data%20Sites%20downloads/IFC%20Report_MAIN%20Final%203%2025.pdf.

InsightAce Analytic. 2024. "AI In The Credit-Scoring Market Latest Trends Analysis Report in 2024." https://www.insightaceanalytic.com/report/ai-in-the-credit-scoring-market/2578?utm_source=whatech&utm_medium=refferal&utm_campaign=shorturl&utm_content=whatech-com-902002.

IPA. 2024. *Designing Credit Products to Support Women's Economic Activity: Synthesis of Recent Literature | IPA*. Innovations for Poverty Action. https://poverty-action.org/designing-credit-products-support-womens-economic-activity-synthesis-recent-literature.

Jagtiani, Julapa, and Catharine Lemieux. 2018. "Do Fintech Lenders Penetrate Areas That Are Underserved by Traditional Banks?" *Journal of Economics and Business*, FinTech – Impact on Consumers, Banking and Regulatory Policy, vol. 100 (November): 43–54. https://doi.org/10.1016/j.jeconbus.2018.03.001.

Jeffrie, Nadia. 2024. "The Mobile Gender Gap Report 2023." *Mobile for Development*. https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-for-development/blog/the-mobile-gender-gap-report-2023/.

JPAL. 2025. "The Impact of a Women-Specific Credit Scoring Model on Womens Access to Credit in the Dominican Republic." The Abdul Latif Jameel Poverty Action Lab (J-PAL). https://www.povertyactionlab.org/evaluation/impact-women-specific-credit-scoring-model-womenaeutms-access-credit-dominican-republic.

Juniper Research. 2023. "Credit Scoring Market Report 2023-28: Size, Trends." https://www.juniperresearch.com/research/fintech-payments/identity/credit-scoring-research-report/.

Kelly, Sonja, and Mehrdad Mirpourian. 2021. *Algorithmic Bias, Financial Inclusion, and Gender*. Women's World Banking. https://www.findevgateway.org/paper/2021/02/algorithmic-bias-financial-inclusion-and-gender.

Klapper, Leora, Dorothe Singer, Laura Starita, and Alexandra Norris. 2025. *The Global Findex Database 2025: Connectivity and Financial Inclusion in the Digital Economy*. Washington, DC: World Bank. https://doi.org/10.1596/978-1-4648-2204-9.

Klapper, Leora. 2023. *Expanding Financial Inclusion through Digital Financial Services: A Literature Review*. Policy Research Working Paper. Washington, DC: World Bank. https://documents.worldbank.org/curated/en/099415112192435594.

Kozodoi, Nikita, Johannes Jacob, and Stefan Lessmann. 2022. "Fairness in Credit Scoring: Assessment, Implementation and Profit Implications." *European Journal of Operational Research* 297 (3): 1083–94. https://doi.org/10.1016/j.ejor.2021.06.023.

Kumaraswamy, Sai Krishna, and Arisha Salman. 2025. "Harnessing Open Finance for Competition and Financial Inclusion." *PYMNTS.Com*, August 26. https://www.pymnts.com/cpi-posts/harnessing-open-finance-for-competition-and-financial-inclusion/.

Lee, Jung Youn, Joonhyuk Yang, and Eric Anderson. 2024. "Using Grocery Data for Credit Decisions." SSRN Scholarly Paper No. 3868547. Social Science Research Network, July 3. https://doi.org/10.2139/ssrn.3868547.

Malone, Cara. 2023. *The Role of AI in Credit Scoring | Whitepapers*. Juniper Research. https://www.juniperresearch.com/resources/free-research/the-role-of-ai-in-credit-scoring/.

Market.US. 2024. "AI In Fintech Market." *Market.Us*. https://market.us/report/ai-in-fintech-market/.

Miller, Margaret, Klapper,Leora, Ghada Teima, and Matthew Gamser. 2020. "How Can Digital Financial Services Help a World Coping with COVID-19?" World Bank Blogs. https://blogs.worldbank.org/en/psd/how-can-digital-financial-services-help-world-coping-covid-19.

Montoya, Ana Maria, Eric Parrado, Alex Solis, and Raimundo Undurraga. 2024. "Bad Taste: Gender Discrimination in Consumer Lending." *Journal of Political Economy Microeconomics*, ahead of print, November 25. https://doi.org/10.1086/734144.

Nova Credit. 2024. "State of Alternative Data in Lending 2024 - Report." https://info.novacredit.com/report-state-of-alternative-data-in-lending-2024.

Prove. 2021. "2 Billion People Don't Have Access to Formal Financial Services." https://www.prove.com/blog/2-billion-people-dont-have-an-access-to-formal-financial-services.

RiskSeal. 2024. "AI Credit Scoring: Trends and Opportunities in Lending." https://riskseal.io/blog/how-ai-credit-scoring-is-changing-the-lending-industry.

Salman, Arisha. 2021. "Digitally Included, Yet Poor: What Do Their Data Trails Tell Us? | Blog | CGAP." *CGAP*, May 25. https://www.cgap.org/blog/digitally-included-yet-poor-what-do-their-data-trails-tell-us.

Shahid, Ehtesham. 2022. "How Tech Can Help Bank the Unbanked in the MENA Region." December 27. https://www.zawya.com/en/business/fintech/how-tech-can-help-bank-the-unbanked-in-the-mena-region-bhifbl1q.

SME Finance Forum. 2024. "Key Considerations for Open Finance." https://www.smefinanceforum.org/post/key-considerations-for-open-finance.

Tracxn. 2025. "Alternative Lending - 2025 Market & Investments Trends - Tracxn." October 23. https://tracxn.com/d/sectors/alternative-lending/__iDRJt6s6Ayy-CmMFwq3YPSCFWB_yrlgDFTq7fkRoLUQ.

Turner, Michael, Patrick Walker, Sukanya Chaundhuri, and Robin Varghese. 2012. *A New Pathway to Financial Inclusion: Alternative Data, Credit Building, and Responsible Lending in the Wake of the Great Recession.* Durham, NC: Policy and Economic Research Council. https://www.perc.net/wp-content/uploads/2013/09/WEB-file-ADI5-layout1.pdf

Vidal, Maria Fernandez, and Ariadne Plaitakis. 2022. *How Can Data Sharing Support Inclusion? | Blog | CGAP*. July 8. https://www.cgap.org/blog/how-can-data-sharing-support-inclusion.

Vidal, Maria Fernandez, and Dean Caire. 2024. "Leveraging Transactional Data for Micro and Small Enterprise Lending." March 20. https://www.cgap.org/research/publication/leveraging-transactional-data-for-micro-and-small-enterprise-lending.

Vidal, Maria Fernandez, and Fernando Barbon. 2019. "Credit Scoring in Financial Inclusion." Technical Guide. Washington, DC: CGAP. https://www.cgap.org/sites/default/files/publications/2019_07_Technical_Guide_CreditScore.pdf

Vidal, Maria Fernandez, and Sophie Sirtaine. 2024. "Open Finance Can Reduce Financial Inclusion Gaps: Here's How." *CGAP*, March 10. https://www.cgap.org/blog/open-finance-can-reduce-financial-inclusion-gaps-heres-how.

World Bank. 2022. "Informal Sector Enterprise Survey." https://microdata.worldbank.org/index.php/catalog/6490.

World Bank. 2025. *The Use of Alternative Data in Credit Risk Assessment: Opportunities, Risks, and Challenges.* Washington, DC: World Bank. http://hdl.handle.net/10986/42970.

Copyright & Disclaimer

2121 Pennsylvania Avenue, N.W.
Washington, D.C. 20433
www.ifc.org

IFC | International Finance Corporation
WORLD BANK GROUP