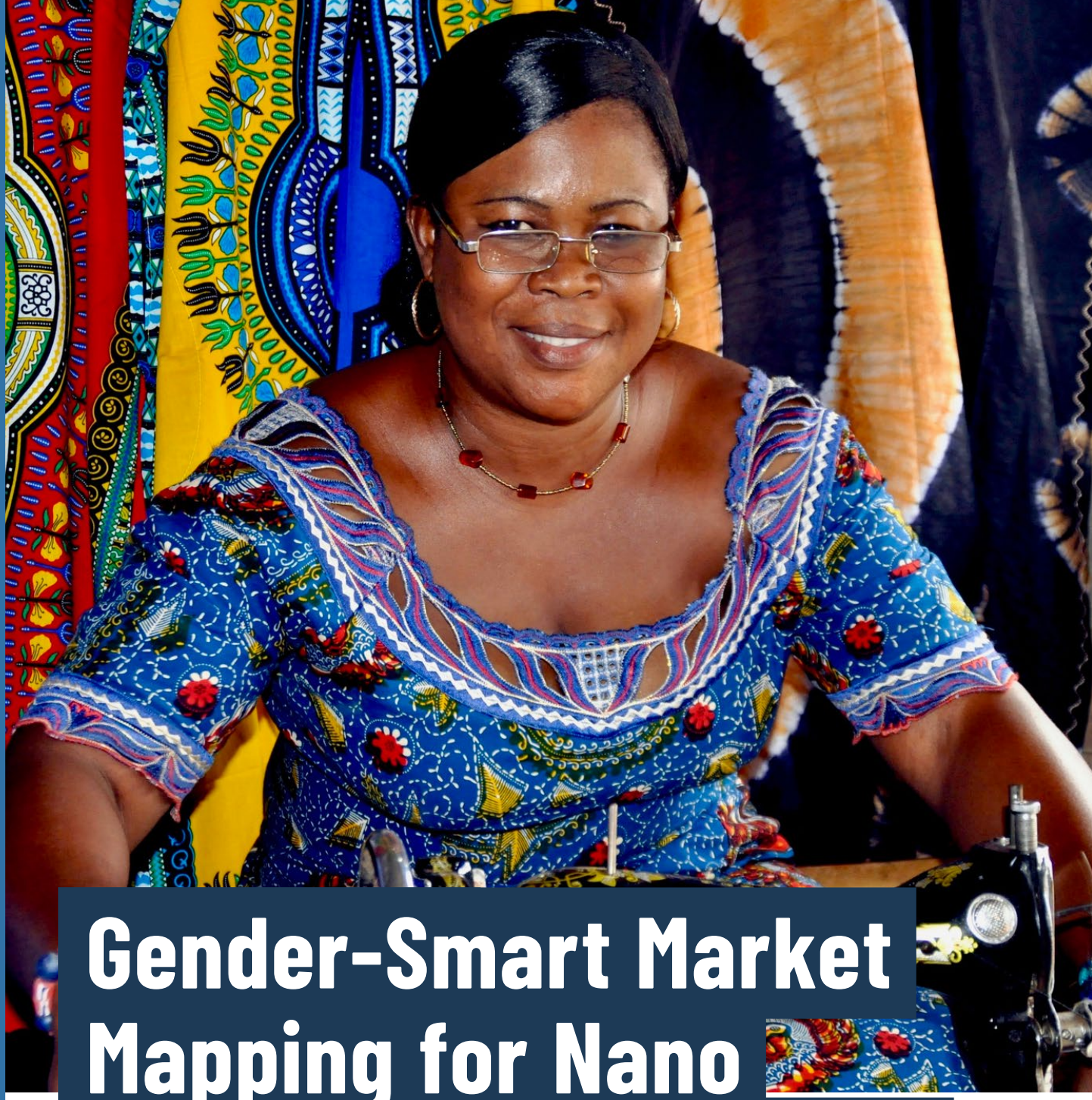
TECHNICAL GUIDE



Gender-Smart Market Mapping for Nano and Micro Enterprises: A Practical Guide

March 2025 • **Salome Wanjiku Kimani, Swati Sawhney, and Alexander Giorgios Sotiriou**

CGAP

Acknowledgments

The authors are grateful to David Kruijff of CGAP, who contributed significantly to this Technical Guide as a researcher and reviewer. Xavier Faz of CGAP provided significant technical direction and critical input to the team throughout the process.

The authors would like to thank internal CGAP peer reviewers Anaar Kara, Estelle Marie Lahaye, and Juan Carlos Izaguirre for their time and effort reviewing the research and providing direction. Finally, we are grateful to Jahda Swanborough, Feven Getachew, and Simrin Makhija for communications and editorial support.

As part of the larger research effort, CGAP solicited the guidance of Advisory Councils in each of the research countries: India, Kenya, and Uganda. We would like to thank the councils' members for their invaluable advice and guidance.

This Technical Guide is the result of extensive primary research conducted by EDI Global (in Kenya and Uganda) and Grant Thornton (in India). Cenfri led coordination of the overall research and data compilation. A special thank you to Lezanne Anderson, Liebe Burger, Christine Hougaard, and Yuvir Naidoo—the lead authors from Cenfri who are responsible for a majority of the guide's content.

CGAP

1818 H Street, NW, MSN F3K-306
Washington, DC 20433
Website: www.cgap.org
Email: cgap@worldbank.org
Telephone: +1 202 473 9594

Cover photo by iStock/Credit:Renate Wefers.

Contents

List of Figures

List of Tables

List of Boxes

Acronyms

FGD	focus group discussion
FSP	financial services provider
GDP	gross domestic product
KPI	key performance indicator
MFI	microfinance institution
MSE	micro and small enterprise
MSME	micro, small, and medium enterprise
NGO	non-governmental organization
NME	nano and micro enterprise
S-GDD	supply-side gender-disaggregated data
SME	small and medium enterprise
WNME	women-led nano and micro enterprise

Executive Summary

NANO AND MICRO ENTERPRISES (NMES) are small businesses that play a crucial role in economic growth, job creation, and community resilience. Nano enterprises are often informal, home-based, or self-employment ventures operating with minimal resources. Despite their significance, NMEs face persistent barriers to financial services access, which limits their ability to grow and thrive. Finding a way to address these gaps has the potential to strengthen livelihoods and enhance the economic contributions of these NMEs. It can also unlock new market opportunities for financial services providers (FSPs).

A clear understanding of the NME landscape—through market segmentation and sizing—is essential for designing targeted products, policies, and programs that effectively support these enterprises. This Technical Guide offers practical insights into these two processes, also outlining a gender-smart approach aimed at maximizing positive outcomes for women.

A gender-smart approach enables users of this guide to not only consider NMEs as a whole but also give particular attention to women-led NMEs (WNMEs). Women entrepreneurs face unique challenges, such as limited access to finance, networks, and business support. This approach provides FSPs and policymakers with an opportunity to better understand and support WNMEs, ultimately fostering a more inclusive and dynamic NME ecosystem. By doing so, NMEs as a whole—and WNMEs in particular—will have the resources they need to succeed.

SUMMARY. **A guide to segmenting and sizing NMEs**

THREE COMPONENTS

This guide outlines a three-pronged approach to NME segmentation and market sizing, with a specific focus on women-led NMEs:

COMPONENT 1

Establish clear objective(s) and gather relevant data. Set a focused direction and a robust foundation for the segmentation and market sizing exercise.

Outcome: Defined goal, tailored data approach, and a custom dataset.

COMPONENT 2

Develop a segmentation model to identify enterprise types with shared needs.

Outcome: Clear market segments highlighting nuanced needs and realities.

COMPONENT 3

Evaluate and size market segments to identify those with the highest potential and feasibility.

Outcome: A defined priority group for targeted action.

AUDIENCE

The guide is primarily designed for researchers, consultants, donors, and FSPs.

WHY THIS GUIDE

Segmentation and market sizing are essential for effectively allocating development efforts, mobilizing private sector investment, guiding public-good interventions, and identifying policy gaps to drive impact in the MSE finance sector—especially for nano, micro, and women-led enterprises. This guide provides a gender-smart, data-driven approach to support sustainable market development. Conducting these exercises will enable researchers, consultants, donors, and FSPs to:

GAIN A HOLISTIC UNDERSTANDING OF NME NEEDS

Follow a structured approach to understand the operations, challenges, and financial and nonfinancial needs of NMEs, including WNMEs.

LEVERAGE DATA-DRIVEN TOOLS FOR STRATEGIC ACTION AND IMPACT

Support data-based decision-making for targeted solutions, appropriate outreach strategies, and fit-for-purpose targets.

APPLY A GENDER-AWARE APPROACH TO NME DEVELOPMENT

Entrench a gender-aware and gender-focused approach for building a thriving NME ecosystem.

INTRODUCTION

Why segmentation and market sizing are essential

NANO AND MICRO ENTERPRISES (NMEs), as defined in Box 1, share many of the challenges that broader micro, small, and medium enterprises (MSMEs) face. Like their larger counterparts, NMEs—many of them women-led—are vital for driving growth**,** creating jobs, and supporting marginalized communities, especially in rural areas (Hewitt 2024)[1]. However, they remain disproportionately constrained by limited access to financial services, markets, and capacity-building opportunities. Addressing these challenges is not only essential for strengthening livelihoods and enhancing the economic contributions of NMEs but also presents both development and commercial opportunities**.** For donors and researchers, it can drive economic resilience, gender equality, and national development. For FSPs, it presents an opportunity to better serve an underserved market, fostering financial inclusion and long-term customer relationships while contributing to broader economic and social development.

Focusing on women-led NMEs (WNMEs) adds another dimension of impact. Women-led enterprises often face even greater barriers due to systemic gender inequalities, limited access to resources, and entrenched socioeconomic norms. Yet empowering these enterprises can have a transformative effect—not only on the women entrepreneurs themselves but also on their families, communities, and the broader economy. Prioritizing WNMEs enables a more inclusive and equitable NME ecosystem, unlocking opportunities for women and driving sustainable, impactful growth.

Effective segmentation and market sizing are critical tools for driving meaningful impact in the MSE finance sector, particularly for NMEs, and, even more specifically for WNMEs. Without a clear understanding of market dynamics, development efforts risk inefficient allocation, diluting their effectiveness. A gender-smart approach to segmentation enables stakeholders to better target interventions, ensuring resources reach the most underserved segments. It also plays a key role in mobilizing private sector engagement by providing clearer insights into viable business opportunities and underserved markets. Further, market diagnostics help identify where public-good interventions—such as capacity building, infrastructure, or financial inclusion policies—can

1 Women own 28 percent of MSMEs globally and are more likely to hire other women and engage with women suppliers, driving economic empowerment across communities and supply chains (Taghiyeva 2023). Estimates indicate that achieving gender parity in business growth could increase global gross domestic product (GDP) by 2 to 3 percent and create up to 433 million jobs (Citigroup 2022). Female entrepreneurship, therefore, plays a critical role in empowering women while delivering significant social and economic benefits (Schuber 2022, Women's World Banking 2024). See the report, "Diverse Paths: Finance for Women's Nano and Micro Enterprises (Sawhney, Kimani and Sotiriou 2025)," for a more detailed discussion of the value of WNMEs.

create system-wide change. Finally, segmentation and sizing expose gaps on the supply side, informing policy reforms that enhance financial access and support more inclusive economic growth. This Technical Guide serves as a practical tool to bridge these knowledge gaps for stakeholders that are interested in investing in NMEs, particularly WNMEs. It aims to enable stakeholders to develop data-driven and tailored products, policies, and programs that strengthen these enterprises while creating new opportunities for growth and inclusion.

Realizing these opportunities requires a targeted approach to prioritize segments with the greatest potential impact and that are feasible to support. This entails understanding the needs and realities of NMEs and WNMEs, categorizing them, and assessing the market size they represent. To this end, this provides details on the following:

- **Market segmentation** is the process of dividing a broad customer base into distinct groups that share similar characteristics, needs, or behaviors. This allows for more targeted strategies to improve services, customer experiences, and development outcomes (FinDev Gateway 2024). For NMEs, segmentation may consider factors such as business motivations, size, sector, and use of digital tools to better understand their specific challenges and opportunities.
- **Market sizing** then estimates the number, economic value, and growth potential of each market segment within a given country. It quantifies the potential demand, revenue opportunities, and overall impact of serving different customer groups, helping stakeholders to effectively allocate resources and identify areas for investment and policy action.

Market segmentation and sizing are often dismissed as costly, complex, or academic, overlooking their practical value. See Box 2 below for some common misconceptions in this regard.

Since 2023, CGAP has been examining the financial and nonfinancial needs of WNMEs, starting with a literature review of the finance markets in India, Kenya, and Uganda. The review was followed by segmentation and market sizing exercises based on primary research. Building on these findings, this guide offers a tested methodology for segmentation and market sizing. The guide has been developed to help donors, researchers, and private sector stakeholders better understand NMEs, particularly WNMEs, in an effort to create more targeted solutions. Box 3 outlines the potential benefits of this guidance for various stakeholders.

The guide is organized into three main stages, as shown in Figure 1. The three stages outline how to set objectives, collect data, and conduct a segmentation and market sizing exercise respectively. At each stage,

BOX 1. **Defining nano enterprises and micro enterprises**

For the purposes of this report, CGAP uses the term ***nano enterprises*** to refer to the smallest, often informal businesses run by individuals with minimal or irregular income (e.g., street vendors, barbers), typically businesses without any paid workers. Known globally by various names (e.g., sole proprietorships, own-account enterprises, self-employed), nano enterprises differ from what CGAP here calls ***micro enterprises*:** slightly larger businesses, such as small shops, with regularly paid workers and more stable revenue. Most definitions of micro enterprises miss the critical distinction between self-employed firms and those with up to nine workers by lumping them together in the same micro enterprise category.

By drawing a sharp line between firms with and without regularly paid workers, the unique and often overlooked challenges nano enterprises face can be captured. This guide highlights the importance of NMEs, along with a focus on WNMEs, which are defined as enterprises with at least 51 percent female ownership and/or women in key decision-making roles (International Finance Corporation n.d.). While WNMEs face unique barriers, including limited access to finance, markets, and networks, they also offer significant opportunities for economic growth and resilience. This definition aligns with existing literature and informs the methodology in this Technical Guide.

BOX 2. **Debunking misconceptions about market segmentation and market sizing**

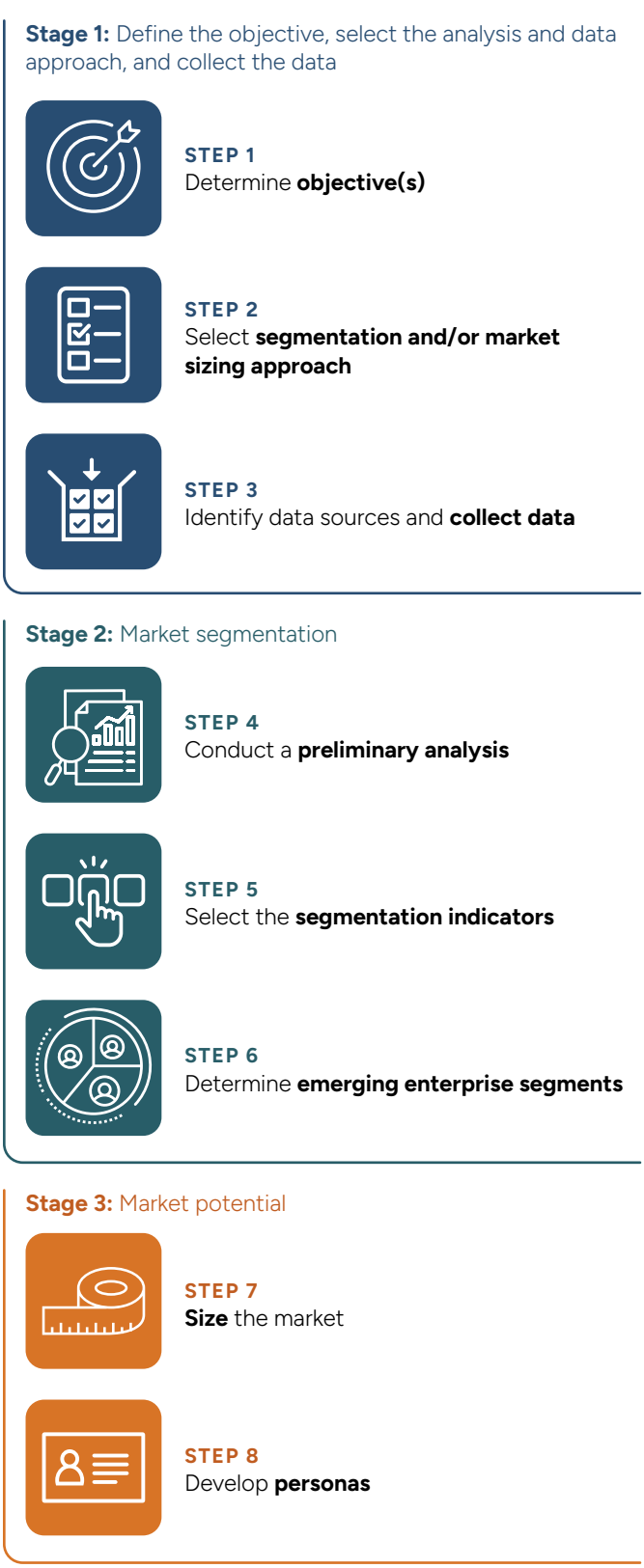
Market segmentation and market sizing are often perceived as costly, overly complex, or purely academic exercises. As CGAP's customer segmentation toolkit (CGAP 2016) highlights, these misconceptions can diminish their true value. In reality, effective segmentation and sizing do not necessarily require extensive resources. Meaningful insights can be derived even with limited budgets and publicly available data, such as national statistics. The key to success lies in aligning these efforts with clear business or development objectives and ensuring stakeholder engagement. When properly executed, segmentation and sizing can drive tailored solutions that effectively address both NME and WNME needs.

the guide offers practical, step-by-step instructions, along with examples and tips for easy application.

While the eight steps outlined in Figure 1 are sequenced to cover both segmentation and market sizing, each of the three main stages can also stand alone and is independently actionable. For example, readers that are only interested in segmentation (Stage 2) can skip the market potential section (Stage 3), and vice versa.

To address areas requiring a deeper or more technical understanding, the guide includes several focused **Appendices.** Each of these is self-contained, and offers targeted insights on critical topics, such as data collection techniques and market sizing, specifically for credit products. The appendices are strategically referenced throughout to guide the reader and provide additional detail and support to those interested.

FIGURE 1. **Key stages and steps in a segmentation and market sizing exercise**



Source: Authors

BOX 3. **Who is this guide for?**

- **Researchers and consultants** can use the guide to conduct research on NMEs, identify financing gaps, and design targeted interventions such as financial education and business support. The findings can provide insights on existing regulatory barriers and inform possible policy advocacy for reforms.
- **Donors** can use the data collected as a result of the segmentation and market sizing exercises to define their strategies and design interventions like resilience-building, business development programs, and funding mechanisms. They can also use the data collected through the exercises to refine monitoring and evaluation indicators to track impact.
- **FSP**s can use the guide to identify market opportunities, design customer-centric products, and develop strategies for distribution and partnerships. Insights from segmentation can help FSPs secure internal support, improve marketing activities, and iteratively adapt products and services to better serve NMEs and WNMEs.

STAGE 1

Define objective, select analysis approach, and collect data

THE FOUNDATION OF ANY MARKET segmentation and sizing exercise is to clearly define its purpose and objectives. These objectives will serve as guiding principles throughout the process, which involves three key steps (please refer to Figure 1).

Key Steps for Stage 1

STEP 1: DETERMINE OBJECTIVE(S)

- Review organizational strategies, goals, and key performance indicators (KPIs) to ensure objectives align with broader goals.
- Set metrics for short-, medium-, and long-term success.

STEP 2: SELECT SEGMENTATION AND/OR MARKET SIZING APPROACH

- Determine the segmentation and market sizing approach (i.e., data-based, rules-based, combined).
- Choose dimensions and indicators to use for market segmentation and sizing.

STEP 3: IDENTIFY DATA SOURCES AND COLLECT DATA

- Determine data collection approach and sources.
- Start collecting data.

Step 1. Determine objective(s)

To set objectives, begin by reviewing existing organizational strategies, goals, and KPIs to determine how the market segmentation and sizing exercise can support them. Consider the following when establishing objectives:

- Focus on the desired **impact.**
- Identify the market **context** and target group—whether conducting a nationwide analysis or focusing on a specific segment, such as NMEs or WNMEs in a particular sector.
- Assess current **capabilities** and **resources** (time, team capacity, decision-makers, outsourcing).
- Ensure objectives are supported by clear, measurable **indicators of success** (short-, medium-, and long-term).
- Allow **flexibility** for adjustments based on evolving market insights.
- As relevant, incorporate a **gender lens** into the objectives and desired impact.

Table 1 provides examples of objectives for different stakeholders.

Step 2. Select segmentation and/or market sizing approach

A market segmentation and sizing exercise can be **data-driven, rules-based,** or a **combination** of the two (Vidal, Caire & Baron 2019):

- A **data-driven** approach, using techniques like cluster analysis, helps reveal hidden patterns across multiple dimensions, leading to deeper insights and more informed decision-making. This can enhance targeting, efficiency, and impact. However, it also requires specialized skills and consistent, high-quality data, which can be resource-intensive and challenging to maintain.
- A **rules-based** approach relies on predefined criteria or dimensions (e.g., livelihood, income level, geographic location) to segment the market. While this approach provides a clear and straightforward framework, it may not capture the full complexity of the market, potentially overlooking deeper insights or more nuanced patterns that could be critical for developing tailored solutions.
- A **combined approach** integrates data-driven and rules-based approaches. It starts with a rules-based structure using proven dimensions, and then moves on to refining these groups through data-driven techniques to provide a more nuanced and tailored understanding of the segments.

All specific tasks for this step and all subsequent steps are designed for use with the combined approach.

When using the combined approach, two tasks need be completed in this step before moving on to identification of data sources. The first is the selection of dimensions and indicators that will form the basis of the segmentation and market sizing exercise. The second is allocation of indicators to each dimension. Each of these tasks are described in further detail below.

TASK 1: SELECT THE INITIAL DIMENSIONS

A segmentation dimension refers to a specific variable or characteristic used to categorize a population or market into distinct groups. Dimensions are selected based on their relevance to understanding the varying needs, preferences, or behaviors of individuals or

TABLE 1. **Example objectives by stakeholder group**

Stakeholder group	Example objectives
Researchers and consultants	Identify distinct NME segments to better understand their needs, behaviors, and challenges. Develop evidence-based recommendations to inform tailored NME interventions, program designs, policies, and strategies.
Donors	Identify and understand the needs of different NME market segments to design targeted programs that meet their specific needs. Support the creation of public-private partnerships aimed at fostering sustainable NME growth. Promote good practices (e.g., gender inclusivity, customer centricity, sustainability and resilience) for organizations that work directly with NMEs. Use the above objectives to catalyze innovation and empowerment for NMEs and, specifically, WNMEs.
FSPs	Innovate on products to serve untapped NME market segments. Inform targeted marketing strategies. Increase market share among NMEs and, particularly, WNMEs.

Source: Authors

organizations within the market. While traditional MSME segmentation often uses broad dimensions like business sector and size, a deeper understanding of the NME market, particularly WNMEs, requires a more nuanced approach—especially in areas where research is limited. on a global literature review conducted in 2024, CGAP has selected eight key dimensions (see Table 2) for segmenting NMEs; however, other

TABLE 2. **Segmentation dimensions used for the validation exercise**

Dimension	Rationale for choosing the dimension
Business owner characteristics	
Mindset	The motivations behind starting a business, along with the entrepreneur's growth plans, mindset, and aspirations, significantly shape financial behavior and needs. A business owner's mindset also plays a critical role in securing financial support, utilizing credit effectively, and ensuring long-term sustainability.
Entrepreneurial ability	Entrepreneurial ability encompasses factors such as education, business experience, managerial competence, digital skills, and financial literacy. These elements are essential for accessing financial services, mitigating risks, and ensuring business sustainability. ***This dimension is particularly relevant for WNMEs.***
Role of business owner in the household	Family structure, including marital status, head-of-household responsibilities, and dependents, is key to understanding the unique contexts of enterprises. Recognizing these factors allows stakeholders to design tailored, supportive products and programs that align with the time and resources entrepreneurs can dedicate to their businesses, fostering inclusion rather than creating barriers to finance. ***This dimension is particularly relevant for WNMEs.***
Business and business owner characteristics	
Access to networks	Access to networks is a key dimension and includes memberships in associations, family or friends as financing sources, and support from sponsors or training programs. These networks enhance credibility, expand financing options, and provide mentorship, market access, and skills development opportunities. For WNMEs, access to informal savings groups or lending circles is also a crucial financial buffer. ***This dimension is particularly relevant for WNMEs.***
Business characteristics	
Level of business sophistication	Business sophistication encompasses factors such as registration status, premises type, recordkeeping, permits, technology use, business planning, and market differentiation. These characteristics influence business stability, risk assessment, and creditworthiness.
Sector	The sector in which an enterprise operates affects its access to financial services. Industries vary in risk and opportunity. Technology or service-oriented sectors may have more access to credit than seasonal or volatile industries.
Location	Geographic location (rural, urban, or peri-urban) plays a major role in financial services access. Urban and peri-urban areas benefit from better infrastructure and market access, while rural businesses may face challenges such as limited banking services and underdeveloped markets.
Size	Business size, including employee count, turnover, and profit, is essential for determining financial services access. These metrics help lenders assess financial stability, operational capacity, and repayment potential.

Source: Adapted from (CGAP 2024a).

dimensions may also exist. Please refer to **Appendix 1: Developing dimensions and indicators** for further details on how to develop dimensions.

TASK 2: ALLOCATE INDICATORS TO THE DIMENSIONS

For a dimension to be useful, it must be measurable. After identifying the key dimensions, the next step is to assign relevant, purposeful, and easily collectible indicators. These indicators should be clearly defined, directly relevant to their respective dimensions, and leverage available data for comparability across segments. Table 3 lists potential indicators for NME segmentation, with a focus on those most relevant to WNMEs.

The chosen dimensions and their corresponding indicators will directly influence both the data collection sample and the market segments that emerge. Therefore, it is essential to select dimensions and indicators that closely align with the intended objective. For example, if the goal is to develop a mobile payment solution for NMEs in Uganda, key dimensions should include indicators related to digital skills, business sophistication, and geographic location to ensure the solution is effectively tailored to the target market.

The dimensions outlined in Table 2 serve as a foundation for any NME segmentation and market sizing exercise. However, specific contexts and objectives may require adding or removing dimensions. For example, for the purposes of CGAP's research, type of ownership was included as an additional dimension in Uganda while risk tolerance was considered in Kenya. Any additional dimensions should be accompanied by relevant indicators to ensure meaningful analysis. For further guidance on how to develop alternative dimensions and indicators, please refer to Appendix 1: Developing dimensions and indicators.

TABLE 3. **Example indicators**

Dimension	Potential indicator	Rationale for indicator
Mindset	Reason for starting a business	Reflects the entrepreneur's motivation—whether driven by necessity, passion, or opportunity—which shapes business strategy, risk tolerance, and growth aspirations.
	Business plan (planning for expansion vs. planning for stability)	Highlights the entrepreneur's mindset toward growth, distinguishing between those focused on business expansion and those prioritizing stability for self-sustenance.
	Reinvestment into the business	Indicates growth orientation, as entrepreneurs who reinvest profits into their businesses tend to pursue expansion, whereas those using profits primarily for household expenses are more livelihood oriented.
Entrepreneurial ability	Level of education	Serves as an indicator of overall business capabilities, assessing the entrepreneur's skills, experience, and ability to sustain and grow their enterprise.
	Skills acquisition	Shows how business management is influenced by the entrepreneur's skills, whether achieved through formal education, informal training, or hands-on experience.
	Digital skills	Provides insights into digital engagement, assessing the entrepreneur's ability to adopt digital tools and participate in an increasingly digitalized economy.
Role of business owner in household	Marital status	Provides insight into financial pressures or support systems, which influence risk tolerance and business decisions.
	Head of household/financial decision-making power	Demonstrates the entrepreneur's level of autonomy in household decision-making, their ability to secure financial support, and possible attitudes toward reinvesting resources in the business.

TABLE 3. **Example indicators** (continued)

Dimension	Potential indicator	Rationale for indicator
Access to networks	Number of networks	Indicates the breadth of networks an enterprise can access, reflecting its ability to leverage support from different groups.
	Type of networks	Highlights the impact of network type on business support, differentiating between formal vs. informal and vertical vs. horizontal networks.
Level of business sophistication	Recordkeeping	Demonstrates operational efficiency and financial management, as strong recordkeeping practices enhance transparency and improve readiness for credit applications.
	Registration status	Signifies access to formal financial services and government support, as business registration facilitates legal protections, funding opportunities, and risk mitigation.
Sector	Detailed economic activity	Captures sector-specific risks, challenges, and growth potential, offering insights into the business environment and its role in the broader economy.
Location	Urban-rural distinction	Illustrates the influence of geographic location, highlighting differences in market access, infrastructure, and support services that shape business operations and growth opportunities. Where relevant, this can include peri-urban or semi-urban classifications.
	Province/state	Reflects the role of regional ecosystems in business operations, as factors such as provincial or state regulations, registration processes, and access to local government support shape the business landscape.
Size	Number of paid workers	Indicates business scale and employment capacity, influencing operational needs, financial requirements, and overall sustainability.
	Annual profit or turnover	Assesses revenue-generating ability, critical for evaluating financial health, creditworthiness, and long-term growth potential.
	Asset value	Highlights asset ownership as a key financial indicator, particularly relevant for insurance (asset protection) and credit (collateral).

Source: Adapted from (CGAP 2024a).

Step 3. Identify data sources and collect data

Once indicators have been selected, the next step is to determine the data required to populate them. Since the collected data will support both market segmentation and market sizing, careful consideration must be given to the following:

- **Type of data: Quantitative, qualitative, or a combination?**
  Quantitative data, expressed numerically, enables comparability, trend identification, and generalizable conclusions. Qualitative data, in contrast, provides deeper contextual insights. A mixed-methods approach leverages the strengths of both but also requires additional resources.
- **Data sources: Primary, secondary, or both?**
  Primary data—collected directly from NMEs—offers current, tailored insights but can be costly and time-intensive. Secondary data, sourced from existing reports and datasets, is often more cost-effective but may be outdated or lack the granularity required for specific segmentation objectives.

Choosing the right data approach is critical as data forms the foundation for segmentation, market sizing, and strategic decision-making. The process should begin by assessing the availability and quality of secondary data before identifying gaps that require primary data collection. A combined approach may be ideal, depending on the research objectives and available resources.

At a minimum, collection of qualitative and quantitative data will be required. The tasks outlined below describe the various data sources that can be considered and outline how to decide when to use primary versus secondary data. Task 1 outlines the process and sources for collection of quantitative data, while Task 2 delves further into qualitative data.

TASK 1: COLLECT QUANTITATIVE DATA

Collect relevant secondary data from credible sources such as industry reports, government databases, and market research studies. Evaluate the reliability and relevance of each source to ensure accurate segmentation and market sizing. There are various secondary datasets to consider:

1. **Internal data**
   Organizations with an established customer base can leverage existing demographic, transactional, and risk assessment data. FSPs may focus on financial behavior, while development agencies and governments may incorporate socioeconomic and geographic data. From a gender perspective, however, many organizations lack sex-disaggregated data, limiting their ability to integrate a gender perspective into segmentation efforts. Internal data may also fail to capture the unique challenges WNMEs face (see Key Considerations at the end of this section for some of the challenges associated with sex-disaggregated data).
2. **Publicly available data**
   National surveys, GDP data, and industry reports provide macroeconomic insights, helping to identify underserved regions and broad market trends. While cost-effective and accessible, public datasets may lack the detail needed for gender-responsive segmentation. See Box 4 for data sources that have been identified by CGAP in a related effort.
3. **Collaborating with external partners**
   Partnering with fintechs, government bodies, or research institutions can provide complementary insights into business performance and market dynamics. Such collaborations enable more refined segmentation and the development of tailored financial solutions, although they may face data quality and privacy constraints.

BOX 4. **Additional data sources for NME segmentation**

The literature review (CGAP 2024a) that informed this guide also includes a data library which catalogs available datasets and publications relevant to NME and WNME segmentation and finance, such as survey data, secondary literature, and administrative datasets. For each source, it provides details on data type, key variables, limitations, and insights provided.

When to collect primary data? If existing data is misaligned with objectives or contains gaps—such as missing gender information or insufficient details on business owner characteristics—it is essential to complement or replace it with primary data. Common methods include surveys, structured interviews, and observational studies. Digital tools, such as online questionnaires and mobile data collection apps, can improve efficiency, particularly for reaching dispersed or hard-to-access groups. While primary quantitative data is critical for a comprehensive market understanding, it requires careful planning to ensure data quality, manage response rates, and control costs.

TASK 2: INCORPORATING QUALITATIVE DATA

For deeper insights, qualitative data collection can supplement quantitative findings. Key approaches include:

1. **Qualitative survey questions.** Open-ended questions in surveys provide richer

Fieldwork across study countries highlights the need for tailoring data collection methods to local contexts. For detailed guidance on quantitative data sources and primary data collection techniques—including sampling strategies, questionnaire design, and best practices—see **Appendix 2: Best practice for data collection and lessons from the field.**

insights into motivations, behaviors, and challenges, ideal for quick, targeted analysis.

2. **Interviews and focus group discussions (FGDs).** These methods capture detailed customer perspectives, identify trends, and validate quantitative findings, although they require trained facilitators.
3. **Digital tools.** Platforms like chatbots and WhatsApp enable rapid, cost-effective engagement, particularly for informal feedback and surveys.

Key Considerations for Stage 1

Steps 1, 2, and 3 lay the foundation for market segmentation and sizing. Ensuring a robust and targeted dataset requires careful attention to the following:

DATA QUALITY

Ensuring data accuracy is critical. Sampling errors should be minimized by cross-checking against alternative sources (e.g., census data). Self-reported income data should be validated with supplementary sources like tax records and qualitative insights. Sensitive questions should be carefully structured to enhance response reliability. See **Appendix 2: Best practice for data collection and lessons from the field** for guidance on designing effective data collection tools.

CONTEXT ENABLERS AND LIMITATIONS

Data collection tools should align with both organizational needs and market realities. Digital surveys, while efficient, may be less effective in populations with low digital literacy or for complex inquiries. For instance, informal enterprises, particularly in grocery and textile sectors, are often overlooked by formal data channels. Snowball sampling[a], as demonstrated in the dipstick study[b] outlined in **Appendix 2: Best practice for data collection and lessons from the field** can ensure effective access to these enterprises through community networks and word of mouth.

TAKING A GENDER-SMART APPROACH

Primary research presents an opportunity to incorporate a gender lens, as secondary datasets often lack sex-disaggregated data. Gender-sensitive sampling, such as the 80 percent women, 20 percent men approach used in CGAP's survey, ensures sufficient representation for comparative and in-depth analysis. Additional strategies include prioritizing gender-focused data sources, incorporating qualitative research with women, and designing surveys that address gender-specific considerations.

OVERCOMING CHALLENGES IN SEX-DISAGGREGATED DATA

Accessing and utilizing sex-disaggregated data remains a challenge, particularly in the financial sector, where data is often collected at the individual rather than the enterprise level. Varying definitions (e.g., "women-led" vs. "women-owned") further complicate consistency. Despite these challenges, efforts like Supply-Side Gender Disaggregated Data (S-GDD) initiatives (Alonso and Dezso 2024) offer solutions by guiding FSPs on data collection and integration. Strengthening these practices can help identify gender gaps and tailor financial solutions to better meet women's needs. By addressing these considerations, organizations can improve data accuracy, deepen market insights, and develop more inclusive, data-driven strategies.

a The snowball approach involves finding research participants by leveraging existing connections. It starts with a small group of initial participants who then refer others from their networks, thus creating a "snowball" effect.
b A dipstick survey is a quick, small-scale survey used to gather initial insights or test assumptions before a more detailed study is conducted.

STAGE 2

Market segmentation

ONCE OBJECTIVES HAVE BEEN defined, relevant dimensions and indicators set, and data collected, the market segmentation process can begin. Segmentation is crucial for estimating market size and potential and ultimately shaping strategic decisions. The process involves three key steps (following on from Steps 1, 2 and 3 in Stage 1 above):

Key Steps for Stage 2

STEP 4: CONDUCT A PRELIMINARY ANALYSIS

- Review data and conduct high-level analysis on key demographic variables and characteristics.

STEP 5: SELECT THE SEGMENTATION INDICATORS

- Conduct indicator and dimension correlation analyses.
- Determine which indicators or dimensions can be dropped and which are valuable for segmentation and should be kept.

STEP 6: DETERMINE EMERGING ENTERPRISE SEGMENTS

- Apply dimensions to data to segment data into unique enterprise segments.
- Based on the outcomes, choose enterprise segments with the largest representation in the sample.

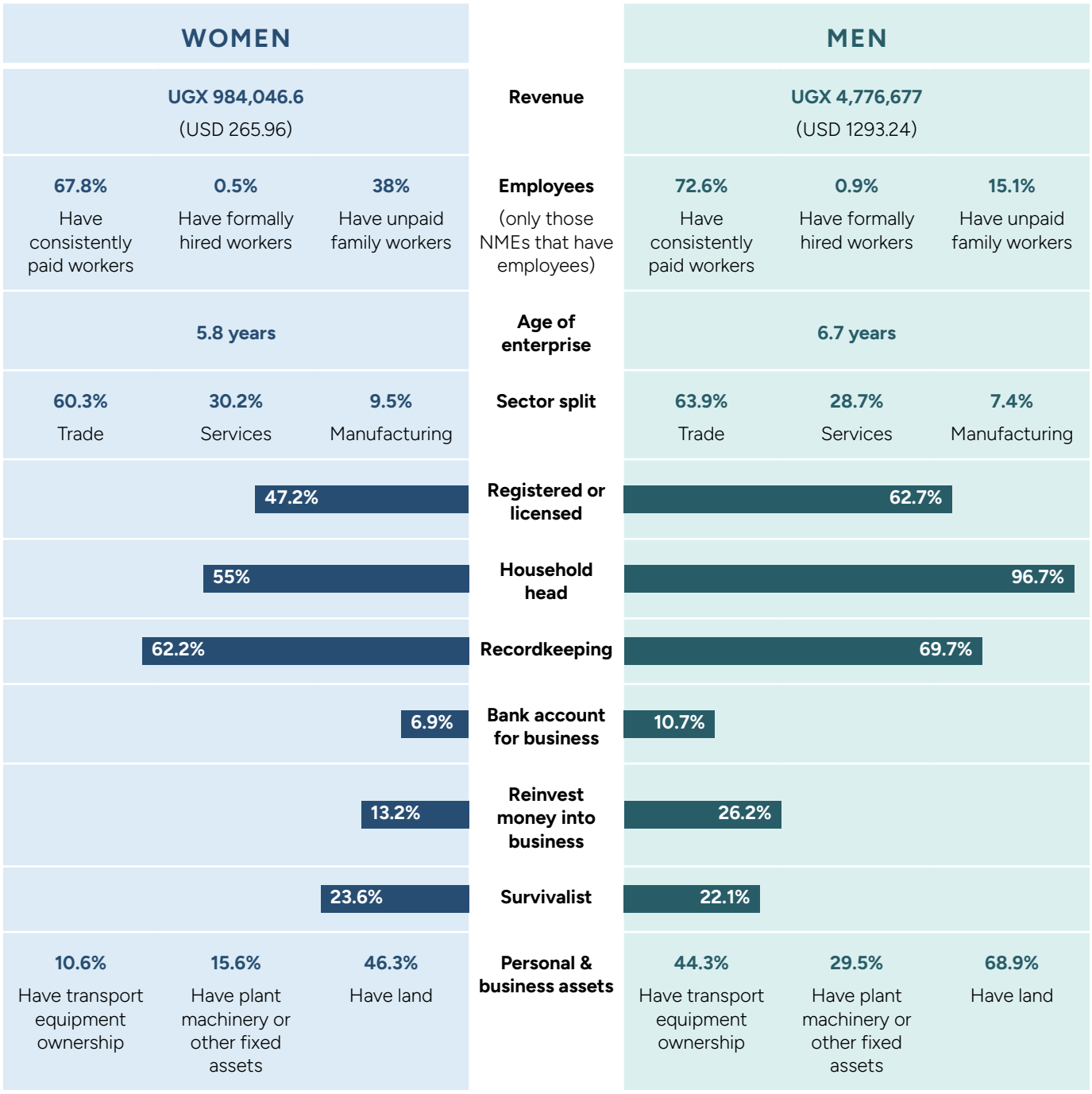
Step 4. Conduct a preliminary analysis

Before initiating the segmentation exercise, it is useful to analyze the dataset or sample to extract high-level insights on the socioeconomic profiles of enterprises, including variations by gender and size. Visualizations, such as bar or pie charts, can help compare distributions across groups (e.g., gender or urban vs. rural business ownership), highlighting key differences (e.g., investment approaches among male- and female-owned enterprises).

Figure 2 illustrates this approach using data from a representative sample of NMEs in Uganda (76 percent female-led enterprises and 24 percent male-led enterprises). For detailed findings, please refer to the report (Sawhney, Kimani and Sotiriou 2025).

Analyzing demographic variables (e.g., age, gender, business size) alongside characteristics like business motivations and management capacity helps reveal patterns and develop hypotheses for market segmentation. This step also tests whether the selected indicators effectively capture the intended dimensions.

FIGURE 2. **Key differences by gender, based on data from NMEs in Uganda**



Source: Author's calculation based on the CGAP nano and micro enterprise survey 2024

If an assigned indicator does not provide meaningful or relevant insights for the dimension it is assigned to, it indicates a misalignment between the indicator and the dimension it is intended to measure. This could mean that:

1. The indicator itself is not well-suited and should be replaced with a more relevant one.
2. The dimension is less relevant than initially assumed, requiring a broader reassessment of the segmentation framework.

In such cases, revisiting Step 2 may be necessary to refine the dimensions, select better-suited indicators, or adjust the overall approach to ensure meaningful segmentation. This is particularly important when additional dimensions or complementary indicators have been introduced, as their relevance and alignment with existing data must also be evaluated. For further detail, refer to **Appendix 1: Developing dimensions and indicators.**

Step 5. Select the segmentation indicators

The next step is to eliminate indicators that do not contribute to identifying meaningful market segments, ensuring the segmentation model remains focused and manageable. The intention with this exercise is to reach a point where each dimension is represented by a single impactful indicator to guide the segmentation process in the next step. This process is carried out through three sequential tasks. The first is to eliminate indicators with low variability, followed by the identification of independent and interdependent indicators. Finally, the few remaining indicators are further vetted with the aim of ideally having one indicator to represent each dimension. These tasks are outlined below:

TASK 1: ELIMINATE INDICATORS

In cases where a large concentration of respondents shares the same value for a particular indicator (as a proxy for the related dimension), that indicator adds minimal insight into differentiating between segments. For example, digital literacy can be used as an indicator for business sophistication. If the data indicates that all businesses have the same level of digital literacy, it can be removed as an indicator representing business sophistication because it is not a valuable means by which to segment the sample or has limited segmentation value. Eliminating indicators with low variability ensures that each remaining indicator contributes to distinguishing between customer segments, thereby improving the clarity of the segmentation analysis.

The only exception to this rule is when indicators (e.g., gender, marital status) may be particularly salient for donors or FSPs in designing interventions. Leaving these indicators in offers an additional tool through which to build in a gender- or other-focused lens. For instance, including gender or marital status in the data collection process allows for the subsequent assessment of how these indicators relate to other dimensions, such as age of the business, resilience to shocks, and related factors to extract key information (further explored in Task 1 of Step 6: Conducting a regression analysis). This step is for all indicators across all dimensions.

TASK 2: IDENTIFY INDEPENDENT AND INTERDEPENDENT INDICATORS

Some of the indicators that remain after completion of Task 1 in Step 5 (across multiple dimensions) may still represent similar underlying traits. To reduce potential redundant information in the underlying model and improve interpretability of the results it is important to assess the correlations among indicators to identify whether they are independent or interdependent (as Table 4 illustrates). Conducting a correlation analysis entails the use of statistical techniques to measure the degree of correlation between the indicators (see **Appendix 3: Analyzing indicators and dimensions** for more on relevant statistical techniques for correlation analysis). The purpose of this exercise is to examine indicators within the same dimension to identify and

TABLE 4. **Independent vs. interdependent indicators**

Type of indicator	Description	Impact on the sample
Independent indicators	Standalone traits (e.g., age, geographic location) that are unaffected by other variables.	These indicators have low or no correlation to others. They uniquely segment the dataset and can all be retained for analysis.
Interdependent indicators	Indicators that interact with others (e.g., income level correlating with business size), providing nuanced insights about customer behaviors and needs.	These indicators have medium or high correlations. Actions: (i) dropping one of the highly correlated indicators to avoid redundancy, and (ii) exploring analytical insights from the relationship between interdependent indicators.

Source: Authors

eliminate redundant or less valuable ones that do not significantly represent that dimension (i.e. those with medium – high interdependency).

TASK 3: DECIDE WHICH INDICATORS TO RETAIN, MERGE, OR DROP

Begin by reading the correlation outcomes. A high correlation (above 0.7) suggests a strong interdependency, while a low correlation indicates independence.[2] When indicators within the same dimension are highly correlated, remove the one least aligned with the objectives. While still useful for analysis, it may add less value to segmentation. Alternatively, merge correlated indicators into a single, new indicator, taking care to ensure that it remains interpretable.[3] For medium correlations, decide whether to keep or merge indicators based on the objectives, market context, and initial dimension analysis.

Remember: By the end of this exercise, each dimension should ideally be represented by a single impactful indicator to guide the segmentation process in the next step.

Step 6. Determine emerging enterprise segments

With the final set of dimension indicators (one dimension represented by one indicator), the dataset can be segmented to identify distinct enterprise segments. These segments represent groups of enterprises that exhibit similar characteristics and behavioral patterns based on the selected dimensions. Figure 3 illustrates the process of transforming the sample into enterprise segments using these dimensions.

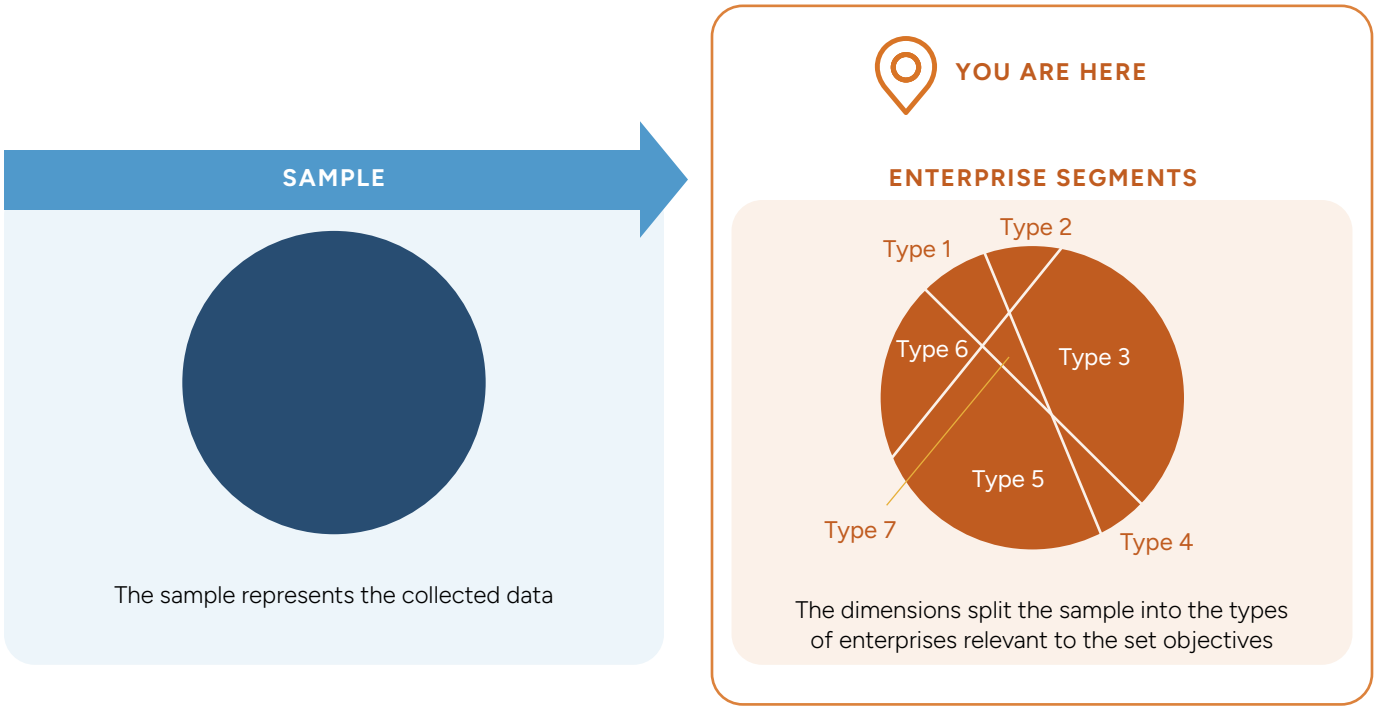
In the illustrative example in Figure 3, the sample is segmented by three intersecting white lines, each representing a segmentation dimension.[4] Enterprises are clustered based on similarities, forming distinct groups labeled as types 1–7. The figure also highlights the importance of sample representation. While segmentation reveals multiple enterprise groups, some may be too small to analyze further. For instance, in the visual, types 4 and 7 may each comprise less than 10 percent of the sample, limiting their significance for deeper analysis.

2 Where all indicators are independent, consider promoting those less representative of the dimensions to be standalone dimensions.

3 Principal component analysis can help combine indicators by constructing a composite indicator which summarizes the essence of those indicators, as long as the resulting measure remains clearly interpretable and actionable. For more on the methodology for merging indicators, see The Ten Step Pocket Guide to Composite Indicators and Scoreboards (European Commission 2024) and The Handbook on Constructing Composite Indicators (OECD/European Union/EC-JRC 2008).

4 Figure 3 provides a hypothetical visual illustration of how dimensions could segment the sample. It makes several assumptions that may not be true for real sample sets. For example, it assumes that each dimension divides the sample into mutually exclusive groups, and that each segment is homogenous based on the dimension used for segmentation.

FIGURE 3. **From sample to enterprise segments**



Source: Authors

There are two tasks involved in identifying meaningful segments, namely selection of final dimensions and developing segments using shortlisted dimensions. Details of these tasks are as follows:

TASK 1: SELECT THE FINAL DIMENSIONS

Using too many segmentation dimensions can create segments that are too small to be meaningful. To prevent this, it is essential to refine the dimensions beforehand, prioritizing those most relevant and valuable. Ideally, limiting segmentation to two to five dimensions ensures clarity and interpretability. Excluded dimensions can still be retained for broader analysis to generate additional insights. For example, although business sophistication (excluded as a segmentation dimension in Task 1 of Step 5) is not used for segmentation, it remains valuable for assessing business maturity, growth potential, resilience, and barriers to further sophistication. There are four sub-steps to reducing dimensions:

1. **Correlation analysis between segmentation dimensions**
   Step 5 used correlation analysis to identify the most meaningful indicators within each dimension. Here, it is again applied to assess relationships between the selected segmentation dimensions, ensuring they are sufficiently independent. The indicator chosen for each dimension in Step 5 will be used. If two segmentation dimensions are highly correlated, only one should be retained for further analysis to avoid redundancy.
2. **Regression analysis against core outcome(s)**
   Regression analysis evaluates the explanatory power of the selected segmentation dimensions in relation to key outcome indicators. These outcomes reflect critical stakeholder interests or study objectives, such as credit behavior (e.g., loan size), resilience (e.g., recovery from shocks), or digital usage (e.g., smartphone penetration). The analysis determines how strongly each dimension influences these outcomes. For example, if the goal is to understand digital transformation, smartphone penetration may serve as the outcome indicator. If regression results show that a dimension, such as ability, significantly affects smartphone penetration, it becomes a strong candidate for segmentation. This ensures that segmentation

is based on dimensions with the highest impact, leading to more actionable insights.

3. **Comparing segments within each dimension**
   Comparing segments within each dimension allows examination of how subgroups within each dimension differ across key indicators and how this influences segmentation. Each dimension is associated with an indicator, which categorizes enterprises into distinct groups. For instance, if the indicator is recordkeeping, businesses may fall into three groups: those with no records, those using paper-based records, and those using digital records. The segmentation process considers sample size constraints; if the digital recordkeeping group is too small, it may be combined with the paper-based group to form a more meaningful category. This approach ensures that segmentation remains statistically robust and analytically relevant.

4. **Interpreting results and finalizing core dimensions**
   After completing the regression and subgroup analysis, findings are synthesized, and the final set of segmentation dimensions is determined. This involves evaluating the pros and cons of retaining or removing each dimension to create a refined shortlist. Quantitative analysis alone may not always provide definitive answers on which dimensions offer the greatest explanatory power. Therefore, it is essential to incorporate qualitative insights, such as fieldwork observations (if available), client interactions, or other relevant contextual knowledge, to support decision-making. For a structured approach, a pros and cons table can help weigh the strengths and limitations of each dimension (as illustrated in **Appendix 3: Analyzing indicators and dimensions**).

TASK 2: DEVELOP SEGMENTS USING THE SHORTLISTED DIMENSIONS

The four sub-steps above refine the segmentation process by first ensuring dimensions are independent (correlation analysis), then assessing their explanatory power for key outcomes (regression analysis), followed by examining how subgroups within each dimension impact segmentation. Finally, the findings are synthesized, incorporating both quantitative and qualitative insights, to select the most meaningful segmentation dimensions. With the final dimensions in hand, the next step is to segment the dataset. There are two options to cluster entities according to the chosen dimensions:

Option 1: Establishing thresholds for segmentation dimensions

To ensure segment viability, establish thresholds for each selected segmentation dimension using cross-tabulation analysis. This method involves analyzing possible combinations of segmentation dimensions to determine a practical threshold, such as requiring at least seven entities per segment. This approach is most effective when using a maximum of three segmentation variables to maintain clarity and manageability.

The example in Table 5 cross-tabulates enterprises by motivation (reason for starting a business) and sector, highlighting which segment combinations represent a significant portion of the sample. Based on these dimensions, four meaningful segments could be constructed:

- **Entrepreneurs that show entrepreneurial skills or apply unique skills -** Tailoring or grocery
- **Entrepreneurs that only work near home**: Textiles and handicrafts
- **Entrepreneurs with no other alternative**: Grocery and refreshments
- **Entrepreneurs with no other alternative**: Garments / tailoring and Textiles

While these clusters provide clear segmentation, they also result in exclusion of some enterprises if they do not fit within the defined thresholds. This trade-off ensures that the identified segments are distinct and analytically meaningful, but it may limit the inclusivity of the analysis.

This process should be repeated across all selected dimensions to systematically identify and extract contextually relevant and analytically robust enterprise

TABLE 5. **Cross-tabulation analysis - type of business and motivation for starting the business**

	To show entrepreneurship skills	Only work near home	Applying unique skills to business	No other alternative
Garments and tailoring	9	1	11	9
Grocery and general trade	9	4	7	10
Handicraft	1	10	0	4
Refreshments and food	2	3	3	11
Services	1	0	3	1
Textiles	2	12	2	11

Source: Authors

segments. By refining segmentation dimensions, establishing thresholds, and ensuring statistical and contextual relevance, this process creates meaningful enterprise clusters that can drive targeted insights. These insights can help stakeholders better understand distinct business needs, tailor interventions, and design strategies that effectively support different enterprise segments.

Option 2: Using confirmatory cluster analysis

Option 2 involves using *confirmatory cluster analysis* to group data into clusters based on predefined segmentation dimensions.[5] This method tests whether the data supports certain assumptions about how the groups should be organized. Unlike *exploratory cluster analysis*, which looks for patterns without prior expectations, confirmatory cluster analysis checks whether the expected groups exist, based on what is already known. It uses statistical tools to verify whether the data fits the anticipated structure, helping confirm that the groupings are valid.[6]

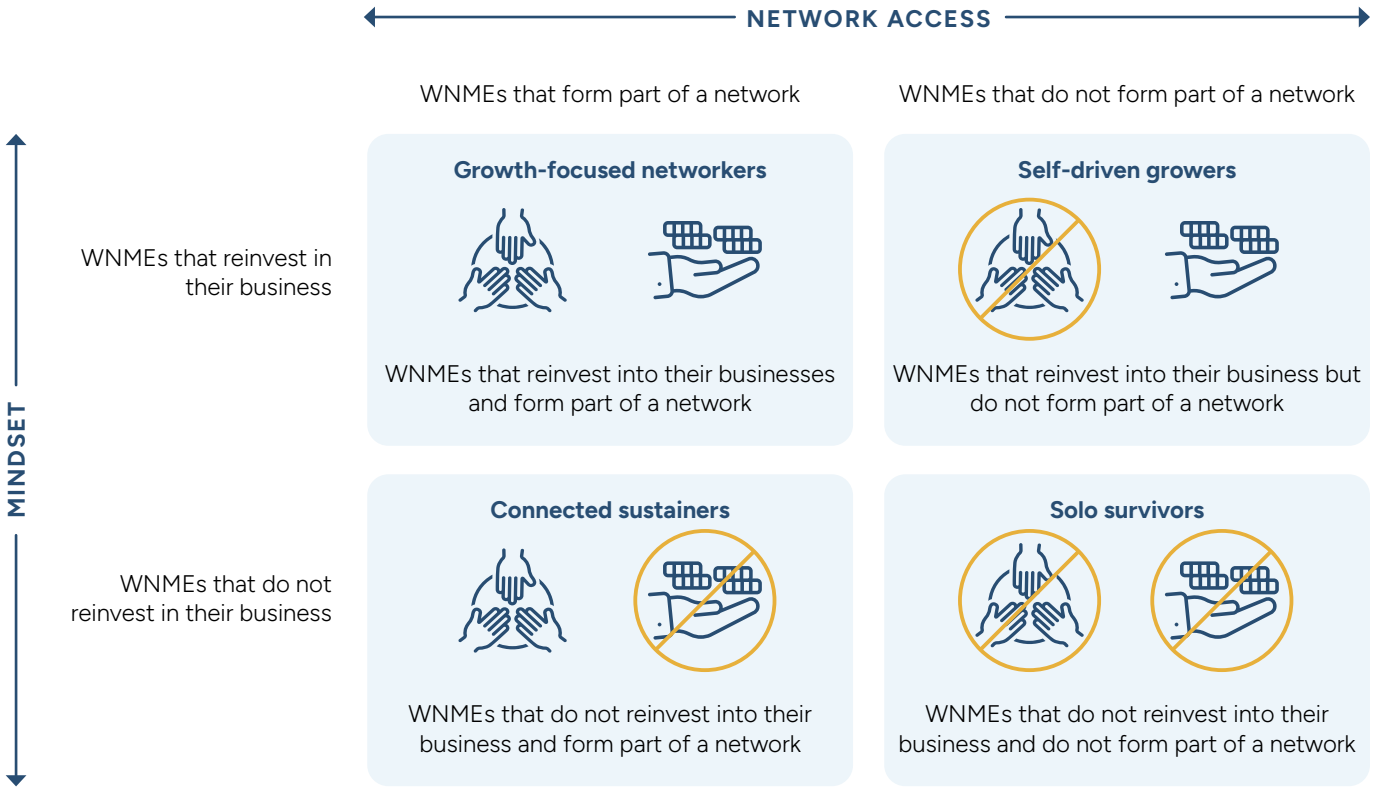
This approach is particularly useful when there is a larger number of shortlisted segmentation dimensions. Instead of using rigid thresholds (as in Option 1), confirmatory cluster analysis identifies trends and proportions within segments, allowing for some variation. For example, rather than defining a segment strictly by entities whose members only have primary education, a segment might be identified where 78 percent of members have primary education, thus accommodating variation. This flexibility prevents the exclusion of data due to strict criteria. However, it is crucial that the resulting segments remain interpretable and useful for decision-making.

By applying either approach, clear enterprise types aligned with the overall objective (as identified in Step 1) can be extracted. In India, for example, "network access" and "mindset" were chosen as final dimensions, which helped identify the four enterprise types illustrated in Figure 4.

5 A cluster analysis is a statistical technique used to group similar data points into clusters based on their characteristics or features. It aims to maximize the similarity within each cluster while minimizing the similarity between different clusters, helping to identify patterns or structures in the data.

6 The Dunn Index (Dunn 1956) measures how well clusters are formed by comparing the distance between different clusters to the spread within a single cluster. A higher Dunn Index means better-defined clusters. Similarly, the Silhouette Score (Shahapure and Nicholas 2020) checks how well data points fit within their assigned cluster compared to others, indicating the quality of the grouping. Confirmatory Factor Analysis (Statistics Solutions 2025) in turn, tests whether the observed data fits expected or hypothesized patterns or groupings, ensuring that the clusters align with the intended structure.

FIGURE 4. **Illustrative example: Enterprise types identified in India**



Source: Authors

The combination of indicator analysis and dimension analysis creates a customized segmentation tool tailored to the objectives defined in Step 1.

Appendix 3: Analyzing indicators and dimensions provides further details on the methodology for the segmentation exercise, including:

Examples from Kenya to illustrate how correlation and regression analyses can help refine indicators and dimensions for segmentation. It also provides a practical example of creating a "pros and cons" table to finalize dimension selection.

Guidance on advanced techniques that help connect segmentation dimensions to outcomes like revenue growth and resilience.

The success of the segmentation approach depends on the specific objective and dataset used, and results may not always be immediately clear. If the suggested methods do not yield viable enterprise segments, refer to **Appendix 4: Alternative segmentation approaches** for other approaches to consider, such as exploratory cluster analysis.

Key Considerations for Stage 2

Steps 4, 5, and 6 focus on segmenting the dataset in a way that aligns with the objective defined in Step 1. To ensure the final segmentation is both accurate and meaningful, consider the following points:

CHOICE OF METHODS

Consider statistical models and methodologies that align with your team's internal expertise and experience. The models recommended in this guide, used in India, Kenya, and Uganda, were the best available options at the time the research was conducted, and suited the capabilities of the implementing agencies. However, alternative approaches may also be worth exploring. Institutional teams or consultants should evaluate the benefits and resource requirements of the recommended methods and consider other options based on their own expertise and experience.

USING QUALITATIVE DATA

Do not overlook qualitative data in the segmentation process. While the indicator and dimension analyses apply a quantitative approach to refine the segmentation model—focusing on dimensions that are most representative and independent—important qualitative insights could be missed. It is crucial to incorporate qualitative information to capture nuances, particularly in the different realities faced by WNMEs. In Kenya and Uganda, for example, the data showed that women with access to two or more informal networks were more likely to consider formal networks for accessing finance. This underscores the importance of integrating both qualitative and quantitative insights. Understanding the role of informal networks, as in Kenya and Uganda, can help donors and stakeholders identify key factors influencing women-led enterprises' engagement with formal financial services, enabling more effective program design and targeting.

A GENDER-SMART APPROACH

Consider allowing gender to guide the insight collection and dimension selection process. When adopting a gender-smart approach, prioritize generating insights that are particularly relevant to women-led entrepreneurs and the overall objective. This involves gathering gender-specific initial insights and using gender as a key criterion when selecting segmentation indicators and defining the segments. Doing so ensures that the segmentation process captures the unique needs and challenges faced by women-led enterprises, ultimately leading to more tailored and effective strategies.

STAGE 3

Market potential

IN THIS STAGE, FOCUS SHIFTS FROM identifying segments to understanding the market potential of each enterprise segment. It includes two core components:

1. **Market sizing.** This component quantifies the economic and strategic value of each segment, including metrics such as the number of potential enterprises, aggregate demand, and projected growth. These insights are crucial for designing effective strategies, policies, and products to better serve NMEs and WNMEs by providing a clear picture of the target audience's size and the viability of engaging with them.
2. **Persona development.** This component enables development of archetypes to represent the typical characteristics, behaviors, and needs of target segments, offering a humanized view of how to effectively approach these groups. This step is an integration of market segmentation, sizing, and qualitative insights.

Together, market sizing and persona development provide a comprehensive view of the market, identifying segments that are not only aligned with the objectives in Step 1 but also feasible to serve, and with the greatest potential impact. This enables organizations to prioritize resources and strategies based on the scale of opportunities and the specific needs of each segment. This section explores Steps 7 and 8 in detail.

Key Steps for Stage 3

STEP 7: SIZE THE MARKET

- Conduct market sizing to estimate the total, unmet, and addressable market for each enterprise segment.

STEP 8: DEVELOP PERSONAS

- Build detailed profiles for each prioritized enterprise segment using demand-side data.

Step 7. Size the market

The market segmentation exercise, as outlined in previous steps, defines specific enterprise segments within the sample. But how many enterprises of each type exist in the total market of a particular country or sector?

Determining the size of each enterprise segment is the first step in understanding market potential. Using the segments identified earlier, the market sizing exercise serves three core purposes: (i) determining the number of enterprises within each segment; (ii) assessing the aggregate demand for a program, financial product, or other type of support (based on the objective in Step 1); and (iii) identifying which enterprises can feasibly be reached.

There are two possible approaches to market sizing, as noted in Table 6:

1. A top-down approach, which provides a broad market overview.
2. A bottom-up approach, which delivers detailed, segment-specific insights.

The choice between the two sizing methods depends on the required level of detail, time available, and the availability and quality of data. A top-down approach may be sufficient if granular, nationally representative data covering most relevant indicators is available. This involves calculating the number of enterprises that meet the criteria for each segment based on those indicators. However, such comprehensive and up-to-date datasets may not always be available, especially for dimensions like the business owner's motivation, their ability, and access to networks, which are harder to find. Thus, in many cases, a bottom-up approach will be needed to fill the gaps. For more guidance, refer to **Appendix 2: Best practice for data collection and lessons from the field**—a comprehensive and practical manual enriched with examples and ready-to-use data collection templates.

Combining the top-down and bottom-up approaches allows practitioners to leverage the strengths of both methods, resulting in a more comprehensive and reliable market size estimate. Cross-checking estimates from different sources reduces bias and enhances the accuracy of the conclusions.

A market sizing exercise has three underlying tasks. Depending on the specific objectives and organizational goals, it is possible to conduct only one or more of these tasks that align with the objective set in Step 1. However, since the tasks build upon each other to provide a comprehensive view of market potential, sequentially conducting all three is recommended.

TASK 1: CALCULATE THE NUMBER OF NMEs (OR WNMEs) IN EACH SEGMENT

This step utilizes the top-down approach, starting with national datasets, to identify the number of NMEs operating within each segment identified in Stage 2. The approach is informed by a combination of data from the listing survey and a comprehensive business survey conducted by CGAP in 2024, both of which were used for this specific research. Follow the detailed steps outlined in Table 7 to calculate the representation of the selected segments. An example from Kenya is

TABLE 6. **Overview of market-sizing approaches**

	Top-down approach	Bottom-up approach
Description	Starts with macro-level data (e.g., overall market size, total revenue, population), then works down to estimate specific segments or niche markets using a representative dataset.	Begins with detailed, micro-level data (e.g., specific enterprise characteristics) collected directly from surveys or interviews and extrapolates to understand specific characteristics of the target population. These individual-level insights are then aggregated to derive broader market estimates.
Data needs	Aggregated, nationally representative data, such as census data, industry reports, or large-scale enterprise surveys.	Micro-level data, typically from primary sources like surveys or interviews, which are often more granular but resource-intensive to collect.
Objectives that can be achieved	Fast and cost-effective market sizing approach to get a general sense of the overall market potential.	More granular data and information with greater accuracy but more cost-and time-intensive.

Source: Authors

provided to clarify and illustrate the process. The data used in the example below is from a listing survey and a comprehensive business survey conducted by CGAP in 2024.

TASK 2: CONDUCT A GAP ANALYSIS TO IDENTIFY THE PORTION WITH UNMET NEEDS

Once the segments are sized, the next key question is: what portion of NMEs (or WNMEs) in each segment faces the largest gap between demand and supply? This market gap sizing exercise estimates the unmet demand for a specific product, service, or intervention within a target market. It quantifies the gap between current supply and existing demand, revealing how much demand remains unsatisfied. Demand and supply are defined here by the objective set in Step 1. For example, if the objective is to develop a program supporting WNMEs with business training, the focus would be on identifying the number of WNMEs that require and are interested in such training. Similarly, if the goal is to create a financial product, demand would be the number of NMEs seeking and needing that specific product.

To identify the unmet demand, the process involves calculating the total potential demand and determining which portion is currently being served. For example, using the previous case of business training, this could be done by comparing the proportion of NMEs (or WNMEs) within a selected segment that currently has access to business training with the proportion that expressed a need for business training based on the survey data. These data points can then be extrapolated using the segment population size estimates to determine which segments are facing the largest gap in demand.

Market gap sizing often requires a bottom-up approach, as the specific data points needed to

TABLE 7. **Calculating number of NMEs in each segment**

Step		Example from Kenya
1	Begin by identifying target population estimates using secondary data to estimate the total market for NMEs within the broader MSME segment. If no secondary data exists, leverage survey data in line with the bottom-up approach.	Kenya's overall MSME population is 7.4 million enterprises (KNBS 2016).
2	Next, calculate the size of the NME market.	98 percent of MSMEs are nano enterprises (0 employees) or micro enterprises (1–9 employees) (Ministry of industrialization, trade and enterprise development 2020). Multiply total MSME segment population (7.4 million enterprises) by 98 percent. = 7.25 million MSMEs are NMEs.
3	If taking a gender-smart approach, calculate the overall WNME portion and segment population sizes.	Number of women-led NMEs: 57 percent. Multiply this by number of NMEs = 4.13 million.
4	Use key breakdowns, such as business size, location, and sector, to map out specific subgroups and relevant enterprise segments as identified in the market segmentation exercise.	*Example of a segment defined as "WNMEs that are registered and reinvest profits back into the business"*: Female NME owners = 4.13 million, multiplied by 51 percent (registered) and then multiplied by 78 percent (reinvest profits back into the business) = 1.64 million.
5	Read the result	**An estimated 1.64 million WNMEs are registered and also reinvest profits back into their businesses.**

Source: Authors using data from CGAP listing and comprehensive business survey conducted in 2024

estimate demand and unmet demand are typically not included in large-scale datasets. When using primary data, analyze the data to calculate averages and key metrics for the shortlisted indicators for each enterprise segment and the overall target group. These indicators can include financial needs (e.g., average loan size, working capital requirements) and nonfinancial needs (e.g., access to training, business development services) that WNMEs experience, as well as how they currently address them. Table 8 provides guidance on how to conduct a gap analysis, illustrated by an example from Kenya. The data used is from a listing survey and comprehensive business survey conducted by CGAP in 2024.

TASK 3: DETERMINE THE PORTION OF THE SEGMENT WITH UNMET DEMAND THAT CAN FEASIBLY BE SERVED

This step involves estimating the "addressable" portion of the segment's demand: the portion that can realistically be reached and served by a product, service, or intervention. Within the context of determining the addressable market for credit, the focus extends beyond assessing reach to also evaluating creditworthiness, ensuring financial sustainability, and preventing over-indebtedness. It is critical that interventions are designed with responsible finance principles in mind, ensuring that offerings are not only accessible but also sustainable. This approach guarantees that the market demand reflects both the feasibility of access and the capacity to responsibly

TABLE 8. **Conducting a gap analysis to identify portion with unmet needs**

Step		Example from Kenya
1	Set the objective for the gap analysis (ensure it is aligned with the overall objective)	Objective set: Identify the credit gap among all WNMEs in Kenya.
2	Identify the segment population (can be taken from the previous step, if conducted)	Segment population: Women-led nano enterprises. 4.13 million women-led NMEs, 83 percent of which are nano enterprises = 3.44 million.
3	Determine the size of the demand and what the demand was for	Credit demand in Kenya: The segment population multiplied by the total demand (those who applied for a loan plus those who did not but needed one), then multiplied by the median loan amount applied for (to determine the amount of the population that has had its needs met had applied for). 3.44 million x (the 39 percent that applied for a loan + 37 percent that did not apply but needed one) = 2.6 million Multiply demand with the median loan amount applied for (6,667 Kenyan shilling [Ksh]) = 17.3 billion Ksh.
4	Determine the demand that has been met	Credit demand met = the segment population multiplied by the percentage that has received a loan and then further multiplied by the median loan value. 3.44 million x 37% percent that received a loan * median loan value received (6,667 Ksh) = 8.5 billion Ksh
5	Determine the unmet demand	Unmet credit demand = Credit demand – met credit demand 17.3 billion Ksh – 8.5 billion Ksh = 8.8 billion Ksh
6	Read the answer	**8.8 billion Ksh is the estimated value of unmet credit needs for women-led nano enterprises. It is also estimated that in total, 2.6 million women-led nano enterprises need credit.**

Source: Authors using data from CGAP listing and comprehensive business survey conducted in 2024

engage with the product, safeguarding against financial harm and promoting long-term viability. Box 5 below provides detail on responsible finance.

Identifying the addressable portion of the segment through a responsible finance lens helps determine who is likely to benefit from the intervention, product, or service—and who may be excluded from it. "Excluded" refers to individuals or enterprises within the segment who, due to factors like financial inaccessibility, creditworthiness, and other barriers (e.g., limited collateral, gender biases), are not able to access or benefit from the offering. This process requires a bottom-up approach to data collection where insights are directly gathered from the target group to ensure a clear understanding of their needs and constraints. Table 9 provides a guide, using data from a listing survey conducted by CGAP in 2024, with an example drawn from fieldwork conducted in Kenya to illustrate the process.

TABLE 9. **Calculating portion of target group with unmet demand that can feasibly be served**

Step		Example from Kenya
1	To derive the proportion of the target segment that fulfils the specific criteria relevant to the objective set in Step 1, start by identifying these criteria (e.g., ownership at least 51 percent women-led, registration status, network reliance, stage of business development, sector focus, geographic focus, digital adoption, credit history).	The objective is to determine the number of women-led nano enterprises that have smartphones and thus could be reached with digital solutions.
2	Calculate averages for the selected criteria and multiply by the segment size to identify the potential servable size.	3.44 million women-led nano enterprises, 58 percent of which use a smartphone = 2 million
3	Read the result	An estimated 2 million women-led nano enterprises use smartphones and could be reached with digital solutions.

Source: Authors using data from CGAP listing survey conducted in 2024

BOX 5. **What is responsible finance?**

Responsible finance refers to the provision of financial products, including credit, in a manner that is fair, transparent, and aligned with consumer needs and financial capacities. It ensures that borrowers can manage their obligations without falling into over-indebtedness, which fosters trust in the financial system and promotes overall financial stability (McKee, Lahaye and Koning 2011). This is critical for enhancing financial inclusion and ensuring that underserved markets can access appropriate financial products.

When taking a gender-smart approach, responsible finance becomes even more essential. Women entrepreneurs often face unique challenges such as limited collateral, gender biases, and unequal access to networks. Designing financial products that address these specific barriers can empower women-led businesses, promote gender equality, and contribute to broader economic development in women's entrepreneurship.

Several resources are available to guide the adoption and/or practice of responsible finance. These include, but are not limited to the Principles for Responsible Banking (UNEPFI n.d.), Responsible consumer credit lending (Directorate-General for Financial Stability, Financial Services and Capital Markets Union 2019), Responsible lending: overview of regulatory tools (World Bank Group 2013), Responsible Finance (Community Development Finance Institutions n.d.), and Responsible Finance: Putting Principles to Work (McKee, Lahaye and Koning 2011).

Estimating the market size for credit is more complex than for other financial products because it involves not just demand but also the creditworthiness of potential borrowers. Factors such as income, debt servicing capacity, and eligibility influence whether credit can be accessed, making it harder to accurately measure potential demand. Unlike other financial products or support interventions like skills training, where demand is more straightforward, credit requires a more nuanced approach that considers both supply-side and demand-side factors. For more detailed information on how to conduct a market sizing analysis, see **Appendix 5: Credit market sizing.**

The three steps outlined in Table 9 provide a clear quantitative perspective on the enterprise segments identified in Stage 2. Specifically, they help determine (i) the number of NMEs or WNMEs within each segment, (ii) the size of the demand based on the objective set in Stage 1, and (iii) the portion of each segment that can realistically be served.

Step 8: Develop Personas

Personas represent the culmination of the segmentation and sizing process (see Box 6). Once market segments are defined and their quantitative dimensions mapped, personas bring these segments to life by transforming data into multidimensional profiles. Personas go beyond raw demographics, capturing motivations, challenges, and aspirations to offer a more complete picture of an archetypical customer type.

While not essential to the market sizing process, creating personas helps uncover core challenges and unmet needs that go a long way toward ensuring long-term sustainability of support. Without this deeper insight, FSPs may risk offering products (e.g., credit) that are not aligned with their customers' realities, potentially leading to high default rates and negative impacts on financial wellbeing. Similarly, donors or policymakers may create solutions that overlook the complexities of the target market. Personas ensure that strategies are better tailored to meet the diverse needs of WNMEs.

BOX 6. **What are personas?**

Personas are carefully crafted profiles that represent key segments of a target market, combining data-driven insights with qualitative understanding. They distill complex segmentation dimensions into archetypes that capture the motivations, challenges, and behaviors of typical customer types. By integrating various segmentation attributes, personas provide a richer, more complete view of the audience that helps donors, researchers, and FSPs to connect with their core markets.

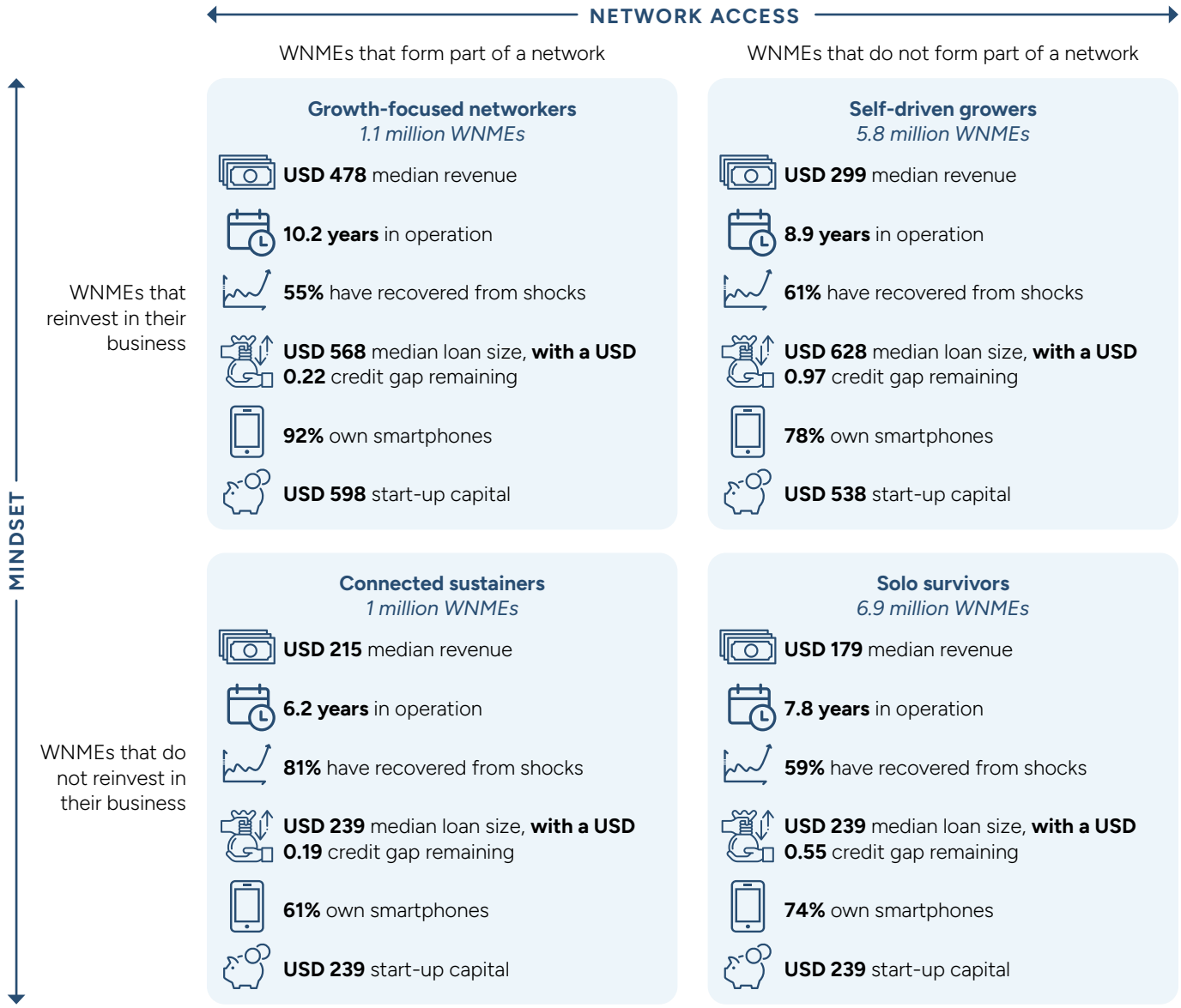
Each persona serves as a strategic tool, guiding the design of marketing strategies, messaging, and tailored solutions that align with the unique needs and aspirations of the people it represents. Personas help stakeholders move from abstract data to concrete actions, improving the chances of success in meeting market demands.

Building enterprise segments into personas involves three interconnected tasks, including analyzing the data, refining enterprise profiles and using that information to develop enterprise personas. Details on these tasks are below:

TASK 1: ANALYZE DATA FROM MARKET SEGMENTATION AND SIZING

Personas are crafted by analyzing the data from the market segmentation and sizing exercises to uncover the unique needs, behaviors, and characteristics of each segment. This process essentially adds depth and detail to the enterprise segments by incorporating key quantitative and qualitative insights. First, it is necessary to form a high-level picture of the features of each enterprise type. For example, using the enterprise types previously outlined in Figure 4, Figure 5 then demonstrates how the data collected through market research and surveys in India can be used to further define these segments, highlighting their specific traits, needs, and behaviors. This enables the creation of more accurate and actionable personas that reflect the full spectrum of the target audience's characteristics.

FIGURE 5. **Further defining enterprise types in India**



Source: Authors using data from CGAP listing and comprehensive business survey conducted in 2024

TASK 2: REFINE ENTERPRISE PROFILES

Next, build out the information for each enterprise type by incorporating details that are specifically relevant to the objective outlined in Step 1. This involves using qualitative research quotes and specific findings from the quantitative data analysis to provide a comprehensive view of each enterprise type. It is essential not only to focus on demographic insights but also to capture the realities, perceptions, and unique needs of NMEs/WNMEs within each segment. This deeper understanding ensures the personas are grounded in the lived experiences of the target audience.

TASK 3: DEVELOP CLEAR ENTERPRISE PERSONAS

Finally, synthesize this key information to transform enterprise segments into well-defined personas. These personas can take different forms, such as case studies or concise write-ups, which are particularly useful for guiding strategic initiatives. Alternatively, they can be structured as data in Excel sheets to assist with product or service development. Figure 6, for example, presents a persona for the "self-driven grower" enterprise type, based on the data and insights discussed in Figure 5.

FIGURE 6. **Illustrative persona for a *"self-driven grower"***

Self-driven grower

Represents 5.8 million WNMEs operational, representing 39% of the total market size in India

KEY DEMOGRAPHICS

GENDER

Predominantly female-owned (at least 51% female ownership)

LOCATION

Mostly located in rural and peri-urban areas

AGE

Average age of 40 years

BUSINESS AGE

On average, 8.9 years

INFORMATION KEY TO THE PROGRAMME

BUSINESS CHARACTERISTICS

- Ownership and Reliance: Majority rely on husbands or male relatives for startup capital and decision-making support.
- Revenue: Average of USD 299 per month.
- Resilience: 61% have recovered from past shocks.

CREDIT AND FINANCE

- Credit Access: Credit gap of USD 0.97 billion, despite high loan sizes reinvested into businesses.
- Borrowing Behavior: 75% are not attempting to borrow due to perceived lack of need. Of those that do, 45% borrow to expand their business.
- Savings Behavior: 77% save their business income .

POTENTIAL FOR GROWTH

- Markers for Growth: High potential due to good use of formal credit sources and healthy savings habits.
- Challenges: Struggle with credit access due to low revenue and lack of awareness of the benefits of borrowing.

OTHER KEY CONSIDERATIONS

SMARTPHONE PENETRATION

High, at 78%, enabling potential use of digital channels such as WhatsApp for communication, client engagement, and supplier interaction.

NETWORK AND SUPPORT

Often operate independently without formal network memberships.

DIGITAL LITERACY

Low

Uses basic digital tools for communication and inventory tracking.

FINANCIAL LITERACY

Low

Requires support in debt management and structured repayment options.

NETWORK ACCESS

Limited

Not part of any formal business network, restricting access to peer learning and market linkages.

Source: Authors using qualitative from CGAP surveys conducted in 2024

Key Considerations for Stage 3

While planning the market sizing exercise and personal development, it is important to consider the following:

OBJECTIVE OF MARKET SIZING IS KEY

The **level of detail** required for a market sizing exercise depends on its intended purpose. The approach described offers a high-level methodology to identify areas where interventions could be most impactful. For example, a donor organization may use market sizing to create a longitudinal study to track market changes over time and inform strategy. An FSP may focus on segments that are both profitable and serviceable to optimize resource allocation. Meanwhile, a government statistics bureau may use a more precise, census-like method to capture a comprehensive market view. Each approach must balance precision and scope based on the goals, stakeholders' needs, and available resources.

ADDITIONAL DATA FOR MARKET SIZING

Market sizing often requires additional data beyond what was collected for segmentation. It involves estimating the total market size, unmet demand, and addressable market, which can be complex. The dataset from Step 3 may not be enough for this task. Unless the secondary data is exceptionally robust and granular, primary data collection becomes essential to supplementing and enhancing the existing secondary data. For guidance on primary data collection methodologies, refer to **Appendix 2: Best practice for data collection and lessons from the field,** which outlines approaches to address these data requirements. Additional data could include:

- **Data for total target market sizing:** Aggregate data to determine the overall number of enterprises in the target market, with enough detail to split by segment dimensions. If secondary data is lacking, this may require a national business survey or a primary survey.
- **Data for market gap sizing:** Detailed data on demand patterns, such as the proportion of WNMEs accessing services and their unmet needs (e.g., businesses unable to access required services). If secondary data is insufficient, this can come from national surveys, FSP data, or primary data collection.
- **Data for addressable target market sizing:** Data to assess the potential addressable market can include device ownership, ownership of relevant documentation (e.g., registration documents, business records), and assets size. This data would likely be collated across secondary datasets and primary data collection. It is crucial to incorporate responsible finance principles by including creditworthiness assessments (e.g., current debt levels, borrowing history) to ensure sustainable market engagement.
- **Data for market potential assessment:** In addition to basic indicators for market sizing and segmentation, additional data on business

Key Considerations for Stage 3 (continued)

characteristics and owner profiles can further refine market potential. These characteristics provide a deeper understanding of the target segments and help tailor interventions or products. While not used for segmentation, this information is valuable for assessing market potential within selected segments. A comprehensive list of these characteristics is included in **Appendix 1: Developing dimensions and indicators.**

Note: *The accuracy of market-sizing estimates heavily depends on the quality of the data. This includes the accuracy of responses, the appropriateness of selected indicators, and the representativeness of the sample used.*

LEVERAGING SUPPLY-SIDE INSIGHTS

A supply-side analysis complements the demand-side approach by evaluating the existing financial services landscape to identify gaps, opportunities, and barriers in service delivery. For private sector stakeholders, it uncovers competitive niches and potential partnerships; for the public sector and donors, it highlights market failures and guides targeted interventions. By integrating supply-side insights with demand-side findings, stakeholders can develop more effective, scalable solutions tailored to WNME needs.

DEVELOPING USEFUL PERSONAS

Although not intrinsic to the market sizing process, persona development can be valuable for ensuring that strategies and solutions align with the realities of NMEs/WNMEs. A few key considerations when developing personas include integrating both qualitative and quantitative insights, capturing not just demographics but also motivations and challenges, and grounding personas in the lived experiences of the target audience. Additionally, personas can take various forms—such as case studies, narrative profiles, or structured datasets—depending on their intended use, whether for strategic planning or product development.

Conclusion

THIS GUIDE UNDERSCORES THE POWER of a tailored, evidence-based approach to understanding NMEs, particularly WNMEs. It provides actionable guidance on how to design and implement market segmentation and sizing exercises. In addition, it guides the reader on how to develop personas that bring the realities and needs of these segments to life. Equipped with these insights, stakeholders will be positioned to develop targeted strategies and interventions that directly address the needs of these segments. In so doing, it is possible to tap into previously underserved markets and unlock new opportunities for impact and growth.

Key takeaways

- **Segmentation and sizing are crucial.** These exercises are important for identifying and prioritizing customer groups with the highest potential for impact and feasibility. By understanding the distinct needs of NME segments, businesses can develop targeted strategies to enhance services, customer experiences, and development outcomes. Segmentation categorizes businesses based on shared characteristics while market sizing quantifies growth, revenue, and impact potential—guiding strategic decisions and resource allocation. Ultimately, this data-driven approach fosters inclusive, strategic, and sustainable development across the NME ecosystem.
- **Principles for success.**
  There are three key principles that are critical for success:
  - Set clear objectives to ensure alignment with organizational goals and KPIs.
  - Collect the right data, ensuring the quality of secondary data and filling gaps with primary data when necessary.
  - Incorporate responsible finance principles, including creditworthiness assessment, to promote sustainable, inclusive market participation and impactful results.
- **Make the process dynamic.** Regularly leveraging this Technical Guide will ensure that interventions remain responsive to the changing market landscape. It should be used not only for initial design but also to track impact and iterate over time, keeping the needs of NMEs central to ongoing solutions.

Taking a **gender-smart approach** to segmentation and sizing involves, among other things, collecting sex-disaggregated data, identifying gender barriers, conducting gender sensitive interviews and surveys, and ensuring that the data is analyzed specifically for women. This adds value for stakeholders of all kinds—from researchers to FSPs. By addressing gender-specific barriers such as unequal access to finance and networks, interventions can create more inclusive markets where all businesses, not just women entrepreneurs, thrive. A gender-smart approach unlocks economic potential, broadens customer bases, fosters innovation, and strengthens the MSME ecosystem for sustainable growth.

APPENDIX 1

Developing dimensions and indicators

EFFECTIVE SEGMENTATION HINGES ON selecting dimensions and indicators that align with strategic objectives. The term *dimensions* refers to specific characteristics or attributes used to group or differentiate customer segments, such as demographics, behavior, or business performance. Dimensions serve as the foundation for differentiating customer groups, evaluated against criteria like relevance, scalability, and stability. These dimensions must link to measurable indicators, ensuring the segmentation is reliable and focused on intended outcomes.

The term *indicators* refers to measurable variables that reflect the state or level of a dimension, providing the data needed to analyze and evaluate segments. Carefully balancing these elements ensures robust and impactful market segmentation. This appendix provides guiding principles for establishing unique dimensions and indicators.

Activity 1: Establish unique dimensions

Establishing unique dimensions means identifying and selecting trait(s) that can segment the sample into meaningful target groups. The study that formed the basis for this guide relied on a comprehensive literature review to find and select dimensions. Conducting such a study through the lens of a specific objective set in Stage 1 (e.g., specifically focusing on digital skills) or relying on organizational knowledge could result in identifying more targeted or focused dimensions for consideration than those presented in Step 2. Other methods that could reveal interesting dimensions for consideration include analyzing existing datasets like administrative records, transactional data, or customer databases that provide insights into behavioral and demographic trends and/or engaging experts, stakeholders, and practitioners. Ultimately, the goal is to pinpoint characteristics that segment the market, ensuring they are relevant and applicable to the specific context and objectives.

Activity 2: Test the dimensions

Each dimension must be evaluated in the local context. The following six questions help to guide the decision on whether to add or remove a dimension:

1. *Does this dimension help achieve the primary objective related to understanding or targeting the market?* Ensuring alignment between the dimension and the overall objectives makes the segmentation exercise more relevant.
2. *Can this dimension be linked to measurable indicators?* Decision-makers need quantifiable or qualitative indicators to reliably track and evaluate the effectiveness of the dimension in their segmentation exercise.
3. *Is this dimension effective in identifying meaningful differences in customer needs, behaviors, or characteristics?* This helps ensure that the dimension provides real value in differentiating groups and that it is actionable and specific,

allowing decision-makers to tailor interventions and strategies to unique customer needs.

4. *Can the segments derived from this dimension be scaled for larger audiences or broader markets?* This helps decision-makers assess whether the segmentation is applicable on a broader level, making it more useful for long-term planning and resource allocation.
5. *Will this dimension remain relevant and stable over time, or will it need frequent updates as the market or environment changes?* Stability over time ensures that the dimension will not require constant adjustments, saving time and resources in the long run.
6. *Does this dimension support effective comparison across different segments, or can it assist in identifying overlaps or gaps between segments?* This helps ensure that decision-makers can draw insights not just within a single segment but also across multiple segments, making the segmentation more comprehensive.

In whittling down or adding to the initial list of eight dimensions in Table 2: (i) take care not to include too many dimensions, as this can create complex segments that are difficult to manage and draw actionable insights from—three or four is recommended; (ii) keep dimensions focused on the objectives; and (iii) do not choose dimensions solely based on current trends (e.g., COVID-19-specific behaviors) as they may quickly lose relevance.

Activity 3: Select suitable indicators

High-quality indicators tailored to specific dimensions enable decision-makers to track meaningful differences and align segmentation with broader goals, such as targeting digital skills or mindset-driven initiatives. While it is possible and even desirable to have multiple indicators for measuring one dimension, it is important to keep it reasonable: one to three indicators per dimension, with one being ideal.

Selecting suitable indicators begins by aligning with objectives. When choosing indicators, ensure they align with objectives, as objectives will always determine the relevant indicators. For example, if the goal is to develop a digital skills program for WNMEs, segmentation should focus on businesses with an aspirational mindset and some digital exposure. In this case, the reason for starting the business could indicate mindset while digital capability could be missing from the ability dimension. Adding an indicator for digital tool usage would help ensure the segmentation aligns with the program's goals. Similarly, if a new dimension is introduced, relevant indicators must be defined to match the exercise's objectives.

Ensuring indicator quality is crucial for reliable data and accurate segmentation. Apart from being relevant and aligned to the goal, indicators should be capable of producing consistent results across varied contexts. Measurability is also essential, with data ideally collected via both qualitative and quantitative methods. Indicators should be designed to highlight meaningful differences within the sample.

For examples of the characteristics for businesses and business owners which can be used to inform indicators, please refer to the CGAP Literature Review 2024 (CGAP 2024a). *For further guidance on selecting indicators, please refer to* **Appendix 2: Best practice for data collection and lessons from the field** *and* Table 10*, which provide examples of relevant indicators.*

APPENDIX 2

Best practice for data collection and lessons from the field

THE STUDY CONDUCTED IN INDIA, KENYA, and Uganda that informed this guide employed a combination of quantitative and qualitative data collection methods to enable a targeted WNME segmentation and market sizing exercise. Focusing on WNMEs, the study aimed to better understand the unique challenges and opportunities faced by these enterprises.

A gender lens was deliberately incorporated into the research approach to ensure that the data collection design prioritized insights for WNMEs. It included skewing the sample to capture roughly 80 percent of data from women, enabling comparison with male-owned businesses while allowing for deeper segmentation within the women's sample.

If time and resources are limited, organizations can skip Round 1 of the data collection process and proceed directly to Round 2, which offers a larger representative sample sufficient for segmentation and market sizing. However, wherever possible, an initial light-touch round of data collection to extract preliminary insights and hypotheses does add value.

The data collection process involved two rounds. First, a dipstick survey[7] was used to develop hypotheses. Second, a larger representative survey was used to test the hypotheses, refine segmentation, and inform market sizing. Both rounds included focus group discussions (FDGs) to contextualize findings, with gender-sensitive adaptations such as female enumerators and gender-specific FGDs to ensure trust and relevance to WNMEs. These techniques ensured that the data collection was both robust and reflective of the unique needs of WNMEs. A detailed overview of the approach follows.

Round 1. Small sample data collection to extract key hypotheses

The first round of data collection involves a dipstick study using a small, nonrepresentative sample designed to gather initial insights and develop hypotheses for further testing in Round 2. This phase includes both quantitative and qualitative components, conducted through a questionnaire and FGDs, respectively:

- **Component 1: Design and roll out the questionnaire.** A concise, well-aligned

7 A dipstick survey is a quick, small-scale survey used to gather initial insights or test assumptions before a more detailed study is conducted.

questionnaire should be developed based on key study objectives, with adjustments made for country-specific sensitivities, particularly around income and tax. Questions should be piloted and cognitively tested to improve clarity and response accuracy. After receiving feedback from the pilot, changes must be incorporated into the questionnaire and interviews should be conducted. A snowball recruitment approach[8] can be used to leverage existing networks for rapid data collection.

- **Component 2: Conduct FGDs.** Following the questionnaire design and initial data collection, the second step is to validate and refine the findings through FGDs (see Box 7 for sampling guidance), allowing for the correction of any inconsistencies and incorrect assumptions in collected data. The primary purpose of these discussions is to gain a deeper, more nuanced understanding of the insights that emerge from the quantitative surveys. For a gender-focused objective (as was the case in the CGAP project), FGDs should be conducted with groups of WNMEs, alongside a control group of male-owned NMEs for comparison. While the guiding questions for these discussions can remain concise, they should be focused on exploring key themes from the survey data.

BOX 7. **Sampling Guidance**

It is recommended to target roughly 100 enterprises per country for interviews, with at least five or six interviews per sector. Select the sample to ensure coverage across key regions with notable NME activity, including urban, peri-urban, and rural areas. For example, the study in Kenya and Uganda intentionally focused on peri-urban and rural locations as NMEs tend to operate in those areas.

FGDs should be conducted in each sample state or region, with six to eight participants per discussion to encourage participation. The sample size should be informed by factors such as population size, the number of regional divisions, population distribution per state and district, and geographic spread, as discussed in Round 2. Additionally, indicators like religion and income distribution can be considered in order to determine whether to over- or under-sample specific locations.

8 The snowball approach involves finding research participants by leveraging existing connections. It starts with a small group of initial participants who then refer others from their networks, thus creating a "snowball" effect.

TABLE 10. **Sample questions to test potential indicators**

Dimension	Indicator	Potential question	Rationale
Mindset	Reason for starting the business	Why was the business started?	Understanding the motivation behind starting the business helps distinguish necessity-driven entrepreneurship from opportunity-driven ventures.
	Reinvestment of profit	How do you use income from the business?	Provides insight into whether the business is aimed at sustaining survival/livelihood or if it is intended for growth.
	Ongoing motivation	What motivates you on a day-to-day basis to continue to operate the business?	Goes beyond understanding how the business was started to include intrinsic and external factors that influence long-term business sustainability and growth.
	Time invested in the business	How much time do you spend working on the business every week?	Helps assess the level of commitment and potential barriers to scaling up, including time constraints due to household responsibilities.
Role in the household	Marital status	What is your marital status?	Provides insights into how social roles and personal circumstances affect business operations and access to resources.
	Household head	Who is the household head?	Examines gender-based power dynamics in decision-making and financial control within the household, which can impact business growth, control over assets, etc.
	Household responsibilities	Who handles most of the household work?/How much time do you spend a week on household work?	Reveals how household responsibilities may interact with business activities, influencing time availability and growth potential. Marital status and household head indicators are generally more interchangeable and generally more available in data. The household responsibilities indicator offers greater granularity but is less often available in data.
	Business decision-making power	Who decides what happens with the business income?	When an enterprise owner controls how income is used (e.g., reinvestment, savings, everyday expenses), this typically indicates a high degree of economic autonomy and influence. This often translates into a more prominent role in business autonomy and the power to separate business and personal finances.

TABLE 10. **Sample questions to test potential indicators** (continued)

Dimension	Indicator	Potential question	Rationale
Ability	Educational level	What is the highest level of education you completed?	Determines how education impacts business management skills, decision-making, and the ability to adopt new technologies or financial practices. This indicator is less valuable in markets where the majority of enterprises are owned by individuals with limited to no education.
	Literacy level	Which of these are you able to do: 1. Read and write 2. Calculations for the business	Explores whether literacy and education are being leveraged for business growth, indicating gaps in skills utilization.
	Skills acquisition	How did you acquire the skills to run the business?	Examines whether entrepreneurs have formal or informal training and where skills gaps may exist for targeted interventions.
	Digital literacy	Which of these are you able to do: 1. Use apps on a phone 2. Send emails 3. Use social media (WhatsApp, Facebook) to talk to clients or suppliers 4. Use social media (Facebook, Instagram, X) to market the business 5. Use a search engine to find information to use for your business 6. Order something online for your business	Asking about digital literacy provides insight into an enterprise's capacity to effectively use more modern solutions for daily activities.

TABLE 10. **Sample questions to test potential indicators** (continued)

Dimension	Indicator	Potential question	Rationale
Business sophistication	Business registration	Do you have a business license or permit?	Determines the level of formality in business operations and potential barriers to legal recognition and compliance.
		Is your business registered?	Assesses the extent of business formalization, which impacts access to financing, government support, and market opportunities. This indicator is less relevant to understanding the maturity of a business; in many contexts, registration is a requirement. It can, however, reveal gaps if business registration is required for accessing certain types of support.
	Recordkeeping practices	Do you maintain business records/ accounts (e.g. sales, customers, profits, credit, inventory)?	Evaluates financial literacy, business management skills, and the ability to track performance over time.
		If yes, how do you maintain records?	Identifies the methods used for recordkeeping, which can affect access to credit and business growth.
	Use of digital tools	Are any electronic devices or methods used for business communication?	Measures the extent of digital adoption for communication, which can impact market access and operational efficiency.
		Do you use digital platforms for business transactions (i.e., cards, digital wallets, mobile money apps)?	Identifies digital financial tools in use, assessing the level of financial inclusion and potential for scaling digital services.
	Device ownership	Which of these do you own: 1. Point-of-sale device 2. Smartphone 3. Basic phone 4. Desktop/laptop/tablet	Asking about device ownership helps gauge an enterprise's technological capacity and readiness to adopt digital tools. The presence of modern devices, such as smartphones, computers, or specialized equipment, indicates a level of investment in technology that supports efficient operations, communication, and data management.
	Bank account for business	Do you have a bank account that you use only for your business?	Inquiring about bank account ownership provides insight into the enterprise's formal financial management practices. Having bank accounts suggests the business is integrated into the formal economy, adheres to standardized financial practices, and is more likely to have access to credit and other financial services.

TABLE 10. **Sample questions to test potential indicators** (continued)

Dimension	Indicator	Potential question	Rationale
Sector	Industry/sector classification	Which sector do you operate in?	Provides insights into the distribution of enterprises across sectors, allowing for sector-specific policy recommendations (especially for women-led enterprises).
		What types of activities does your business engage in?	This question is more nuanced than solely focused on a sector. It offers granularity, especially in heterogenous sectors.
Network	Membership in associations	Are you or the business members of any associations, cooperatives or savings groups? (Yes/No)	Examines how social and business networks provide support, knowledge-sharing, and financial access to women entrepreneurs.
		How many networks are you a part of?	Data from the CGAP survey indicates that businesses in more networks are often part of more mature networks and have a better support system to fall back on if they experience shocks.
		What types of networks are you a part of?	Provides insight not only into the maturity of the business but also its access to other stakeholders, such as suppliers.
Size	Number of employees	How many people are involved in your business overall?	Determines business size, workforce involvement, and potential employment opportunities created by women-led enterprises.
		How many people are regular workers whom you have paid monthly/consistently over the last six months?	Provides insights into job sustainability, regularity of employment, and the role of female entrepreneurs in job creation.
	Sales volume	What were total sales last month/ two weeks/week? (For seasonal businesses, what were total sales during your peak month?)	Tracks revenue trends, profitability, and financial health of the business, helping to identify financial challenges. This indicator could be challenging, based on data availability and reliability. In this case, number of employees would be a more appropriate indicator. However, number of employees also fluctuates.
	Asset size	What is the total size of your assets?	Measures capital investment trends and constraints women entrepreneurs face in securing funds for business expansion. This indicator could be challenging, based on data availability and reliability. In this case, number of employees would be a more appropriate indicator. However, number of employees also fluctuates.

TABLE 10. **Sample questions to test potential indicators** (continued)

Dimension	Indicator	Potential question	Rationale
Location	Geographic location	In which region/county is the interview being conducted?	Provides geographical context to business challenges, resource distribution, and market accessibility.
	Urban/peri-urban/rural classification	What is the location of the enterprise: urban, peri-urban, or rural?	Differentiates between urban, peri-urban and rural business environments, highlighting disparities in infrastructure, market access, and financial inclusion. This indicator is more relevant in countries with a significant difference between urban and rural. On the other hand, countries like India have greater regional differences, such as between states and Union Territories.

Source: Authors from CGAP surveys conducted in 2024

TIPS FOR ROUND 1:

- Table 10 below provides a number of sample questions that could be asked to test each potential indicator, as well as a rationale for the information each question could provide. It also offers different indicators per dimension to illustrate how certain indicators (and questions) may be more relevant to an institution and/or its objective rather than others.
- Carefully crafting survey questions is crucial since the way a question is asked can greatly influence the responses received. Clear questions ensure that respondents understand what is being asked, leading to more accurate and reliable responses. Conversely, poorly worded or leading questions can confuse participants or skew their answers, ultimately compromising the survey's integrity. Additionally, the local context, such as cultural and socioeconomic factors, can significantly impact how questions are interpreted; even subtle variations in phrasing can lead to different responses. Taking the time to design fit-for-purpose questions that take into account the factors mentioned above not only improves data quality but also enhances the overall effectiveness of the survey in driving informed decisions.

Round 2. Larger sample data collection to test hypotheses and inform segmentation and sizing of segments

The second round of data collection builds on the first by employing a larger, more representative sample. This phase tests the hypotheses developed in Round 1 and gathers insights to inform segmentation and sizing. As in Round 1, Round 2 includes both quantitative and qualitative components conducted through listing and business surveys, as well as FGDs. Box 8 provides guidance on sampling. The four components of Round 2 are detailed below.

1. **Component 1: Design and roll out a listing survey.** There are several ways to capture perspectives from NMEs. Due to their informality, many lack contact information or business registration with government authorities. One way to address this issue is to use a listing survey (i.e., a survey conducted to systematically identify, map, or "list" all units or entities within a given area or population). This serves two purposes: (i) it samples NMEs for the main business survey, and (ii) it generates estimates for market sizing. Good

practice would involve a concise listing survey that includes straightforward questions on business details, owner information, business performance, and borrowing behavior. It should take no longer than 10–15 minutes. *Table 10 above provides sample questions that can be used in this small sample data collection round.*

2. **Component 2: Design and roll out a comprehensive business survey.** To capture specific nuances about the target sample, a comprehensive business survey will need to be conducted. The business survey is typically more comprehensive and should capture a range of indicators for segmentation, market sizing, and market potential assessment. The indicators should cover aspects like age, education level, location, business literacy and training, household role, household size, primary income earner status, reasons for starting the business, business size and sector, decision-making in the household, and use of business income. It should also include indicators such as revenue, growth, resilience, and access to formal financing.
3. **Component 3: Conduct focus group discussions.** The same FGD approach used in Round 1 can be applied, with a larger sample size. To streamline data collection and analysis, discussions should be summarized for each question or probe area and disaggregated by group type where possible (e.g., by business size, sector). Include one or two verbatim quotes per summary to illustrate key points and streamline theme-based analysis. This approach enhances efficiency during the analysis phase, allowing content related to specific themes to be easily extracted from FGD transcripts. Ensure a balanced urban-rural split in community selection and, where relevant, conduct separate FGDs for WNMEs and MNMEs to capture gender-specific insights (as discussed in Box 8 above). Recruit six to eight diverse participants per session, considering business sector, enterprise size, and age, and engage local leaders as needed to identify participants. **For an example guide on conducting FGDs, please refer to Section 1.1 of the Annex: Examples of data collection instruments.**
4. **Component 4: With data sources in hand, the last element of the data collection exercise is collating the data across the sources used.** Conducting a meaningful segmentation analysis requires a single, collated dataset. Data collation is relatively easy where primary data was collected. After collecting the various quantitative and qualitative data, focus on cleaning, storing, and preparing the data for analysis. Key actions include removing duplicates and checking for errors and missed entries. Next, apply consistent formatting for uniformity. Once cleaned, the data can be stored and analyzed. The collation process is slightly more complex where multiple data sources were selected from either secondary sources or from a mix of primary and secondary sources. That requires more careful planning to ensure accuracy and relevance. Key issues to look for include:
    - Evaluating data consistency and comparability and highlighting key differences in terms of definitions used for enterprise size, timeframes, units of measurement, and samples (sizes, locations, etc.).
    - Awareness of source reliability and bias, as the data collected is often influenced by each organization's objective and agendas.

Once these components are complete, the opportunity arises to investigate differences and their significance on the study, and then adjust for comparability. This includes normalizing values (especially units of measurement), recategorizing enterprise types based on the selected definition, and aligning time periods and location. The approach reduces the risk associated with the differences. Finally, it is necessary to acknowledge the new scope of the differences by communicating any uncertainties and flagging appropriate disclaimers in the findings.

BOX 8. **Sampling guidance for Round 2**

Part 1: Guidance on how to select a sample

- As in Round 1, careful consideration should be given to determining the sample size and parameters for data collection. One approach is to review regional-level micro enterprise population data and focus on the top three regions with the highest density of micro enterprises. From there, the most populous county in each region can be chosen, followed by a random selection of sub-counties and wards or districts that ensures a mix of urban and rural areas. Use of population density can assist in determining whether a ward or district is predominantly urban or rural. It is recommended to list NMEs that cover at least 50 percent of the villages, towns, or economic hubs within a ward or district. In the Kenya and Uganda studies, the approach resulted in approximately 3,840 NMEs listed in each country respectively.
- For the comprehensive business survey, a diverse and representative sample is essential for capturing the unique needs and challenges of NMEs or WNMEs (the CGAP study specifically focused on WNMEs*). The sampling methodology can achieve this by targeting key indicators such as gender, urban-rural distribution, and business size categories (NMEs). Additionally, representation across major business sectors and geographic areas is critical to ensure a well-rounded dataset that supports deeper analysis. While targeting 500 enterprises per country is recommended, sample size adjustments can be made based on the population size and distribution of enterprises within the country of interest.

Illustratively, CGAP's research in India adopted a multistage sampling approach to ensure national representativity. It resulted in a robust sample of 2,055 enterprises. The process included the following four steps:

1. **State-level sampling.** The total sample was divided equally among six states, ensuring balanced representation across rural and urban areas and a representation of states with diverse income levels. Approximately 300–400 enterprises were sampled from each state.
2. **District-level sampling.** Three districts per state were selected based on enterprise distribution data (e.g., the Annual Survey of Unincorporated Sector Enterprises in India). Districts with a high concentration of WNMEs were prioritized to ensure that the sample reflected areas with significant NME activity.
3. **Local area sampling.** Within each district, a mix of urban and rural areas were selected to capture diverse enterprise dynamics. The CGAP listing survey (2024) was used to identify women-led enterprise clusters and nano enterprises across various sectors. For example, 35 percent of the sampled enterprises came from manufacturing, 37 percent came from trade, and 28 percent came from food, accommodation, education, and health services.
4. **Survey execution.** The final sample of enterprises, selected via the stages above, was surveyed to gather comprehensive data on business characteristics, challenges, and opportunities.

**Importantly, when designing a gender-smart survey, the sample should be selected to achieve roughly 80 percent female representation in the total sample, as interviewing a mix of men-owned and women-led enterprises provides insight into the key characteristics, trends, drivers, competition, opportunities, and challenges through a gender lens.*

Part 2: Determining the content of the survey

Questions for surveys can be structured according to dimension to gain a better understanding of each dimension. Example questions can include:

Mindset: *What was the main reason for starting this business?* (e.g., I learned skills at another job; I could not find a job; to improve my status in community, etc.)

Role in the household: *Who is considered to be the head of the household?* (e.g., myself; my spouse; another male or female household member; we hold equal responsibility)

Entrepreneurial ability: *What is the highest level of education you have completed?* (e.g., primary education; secondary education; vocational training; etc.)

BOX 8. **Sampling guidance for Round 2** (continued)

Level of business sophistication: *How do you keep records of the business?* (e.g., physical books; use computer software; keep records by memory; etc.)

Sector: *What does your business do?* (e.g., retail trade; manufacturing; repairs; etc.)

Access to networks: *Are you a member of any of the groups or associations in your community?* (e.g., If yes: trade/business association; self-help group; savings group; religious group; etc.)

Size: *Does anyone work in the business with you? How many people are paid or unpaid, full-time or part-time?* (e.g., If yes: members of the family; paid workers; seasonal workers.)

Location: *In which region is your business located?* (e.g., If urban/peri-urban/rural; specific state or district.)

For the full questionnaire used during the comprehensive business survey, refer to Section 2.2 of Annex: Examples of data collection instruments.

BOX 9. **Process lessons from the field**

- **Conducting pilot surveys to refine data collection.** Conducting a pilot survey helps refine questions, test participant understanding, and address logistical challenges. For example, a pilot in the study revealed that business owners needed more time to respond due to customer commitments. Keeping questionnaires concise minimized dropout rates and improved response accuracy.
- **Leveraging networks to reach home-based businesses.** Many WNMEs, especially in the groceries or textiles sectors, operate informally from home and are often overlooked by formal channels. As the dipstick study demonstrated, using snowball sampling can ensure effective access to enterprises through community networks and word of mouth.
- **Remaining adaptable to country-specific nuances**. Adapting sampling approaches to account for country-specific nuances, such as enterprise size definitions and sample sizes, ensures more accurate, reliable, and granular insights. This approach provides a clearer picture of the challenges enterprises face, offering more actionable data than generalized national studies.
- **Building trust to encourage honest responses.** Trust is crucial when collecting sensitive information from WNMEs. Employing female enumerators and conducting gender-specific FGDs created a supportive environment for women, encouraging openness. This can be particularly effective in generating reliable data.
- **Remaining cognizant of the respondent's time.** It is important to allow enough time to interview and survey business owners, accounting for the fact that they may have to attend to customers during their interview if they do not have support at their business. To prevent dropout and manage respondent fatigue, ensure that interactions are kept short (approximately 45–60 minutes for interviews and 1.5–2 hours maximum for FGDs).
- **Factoring in cultural and permissions requirements.** Being mindful of cultural sensitivities ensures smoother implementation and fosters trust within the community.
- **Securing necessary permissions.** Obtaining approvals from local officials, as was the case in Uganda, helps streamline processes and mitigate potential delays.
- **Prioritizing confidentiality to alleviate concerns.** Ensuring confidentiality and explaining data usage fosters trust. In India, Kenya, and Uganda, respondents were wary of sharing financial details due to fears of taxation. Disclaimers emphasizing anonymity and confidentiality reassured participants and encouraged more candid responses.

APPENDIX 3

Analyzing indicators and dimensions

THIS APPENDIX PROVIDES DETAILED guidance on analyzing and selecting indicators and dimensions in segmentation and market sizing exercises, with a focus on practical examples from Kenya. It outlines the process of conducting correlation analyses, emphasizing the importance of cleaning data, identifying interdependencies, and selecting the most relevant dimensions. By combining quantitative findings with qualitative insights like field observations, this appendix ensures that segmentation is both statistically robust and contextually grounded. It also introduces advanced steps like regression analysis to link dimensions to key business and behavioral outcomes. Business outcomes refer to measurable results that indicate the success or performance of a business. Specifically, they include revenue growth (the business's ability to increase its earnings over time) and resilience (the business's capacity to withstand financial or operational shocks and to sustain itself in changing conditions). Practical examples illustrate how to assess the relevance of indicators like registration, skills, and recordkeeping, ensuring the segmentation framework focuses on impactful indicators.

Statistical how-to guide for correlation analyses

A first step to analyzing the dimensions and indicators is to better understand the statistical techniques behind the correlation analyses, as discussed below. These statistical techniques build on a wealth of guidance in the literature on conducting these types of analyses. For more information, refer to the CGAP Customer Segmentation Toolkit (CGAP 2016), which references various other data sources, discusses considerations for data analysis, and provides tips for analyzing data.

A practical guide to conducting a correlation analysis

There are many ways to conduct a correlation exercise. This section offers one such approach: using statistical techniques to measure the correlation between indicators. A step-by-step guide is set out below, and Box 10 provides a practical example of what to work toward. The same approach is followed when conducting the correlation analysis for dimensions.

Step 1. Clean the data in the data file. Start by ensuring that the data is free from missing values, duplicates, or incorrect entries. Standardize all numerical values (e.g., a single currency for all financial data) and convert categorical data into numerical codes for analysis.

Step 2. Import the data into a data analysis program. Import the data into a program that can be used for data analysis, such as Excel, Python (*Pandas* and *NumPy* libraries), Stata, R (*dplyr* package), or SPSS. For complex analysis, consider using Python or R for flexibility. For simpler datasets, such as the exercise conducted in this study, Excel or Stata will suffice.

Step 3. Use statistical techniques to measure the degree of correlation between the indicators. Use the Pearson correlation coefficient for continuous data (Kremelberg 2011) or Spearman's rank correlation for ordinal/ranked data (Statstutor n.d.). The first level of correlation analysis should focus on indicators within the same dimension to identify and remove redundant or less valuable indicators that do not significantly contribute to representing the dimension. The second level of analysis should assess the correlation between all indicators across dimensions. This will help to identify indicators that may be of less value for segmentation overall. Consider creating a correlation table (see Box 10), to easily compare correlations between various indicators.

Step 4. Analyze and select indicators. Start by reading the correlation outcomes. Consider a high correlation (e.g., above 0.7) as suggestive of a strong interdependency, while a low correlation indicates independence between indicators. Then decide which indicators to keep for the segmentation dimension validation. With high correlations, remember that highly correlated indicators within the same dimension

BOX 10. **Indicator correlation analysis example**

To illustrate what the correlation analysis entails, Table 11 shows the correlation between various indicators of WNME owners' ability in Kenya specifically on indicators of experience, education, and skills.

The results in the table show a weak correlation between education and experience (-0.2522) and no correlation between skills and either experience or education (0.1850 and 0.2633, respectively). This suggests that these factors are largely independent measures of entrepreneurial ability and should be evaluated separately (as standalone dimensions). The p-value highlights statistical significance, with the weak correlation between education and experience (p=0.0171) being significant.

** Numbers outside the parentheses represent the correlation coefficient, with ranges categorized as follows: low correlation (0.0–0.2), medium correlation (0.2–0.5), and high correlation (>0.5). The numbers inside the parentheses indicate the sample size used for the correlation analysis, followed by the p-value, which measures statistical significance (e.g., [2, 0.034] indicates a sample size of two and a p-value of 0.034). A p-value below 0.05 indicates statistical significance, showing that the correlation is unlikely due to random chance.*

TABLE 11. **Correlation matrix for ability indicator analysis**

WNME owners' ability			
Indicator	**Experience**	**Education**	**Skills**
Experience	N/A	-0.2522 (0.0171) Weak correlation[a]	0.1850 (0.637) No correlation
Education	-0.2522 (0.0171) Weak correlation[b]	N/A	0.2633 (0.137) No correlation
Skills	0.1850 (0.637) No correlation	0.2633 (0.137) No correlation	N/A

a Note: These figures refer to the Spearman's rho reading of each correlation, which is used to measure the level of correlation.

b Note: These figures refer to the Spearman's rho reading of each correlation, which is used to measure the level of correlation.

means that one needs to be dropped. Drop or merge the indicator that least aligns with the objectives—keeping in mind the indicator will still be relevant for analysis and extracting insights but just less relevant for segmentation purposes. Where low correlations emerge, on the other hand, there is an opportunity to promote the most prominent indicator to become a standalone dimension. This helps simplify the segmentation, focusing on the most impactful indicators while ensuring each dimension contributes to the distinctiveness of the segments.

Step 5: Conducting series of regression analyses to shortlist the segmentation dimensions. A regression analysis enables further refinement of the segmentation process. It helps to determine the influence of different dimensions on (i) the other segmentation dimensions, and (ii) key variables of interest, such as revenue growth or market potential. By linking dimensions back to core objectives, regression analysis provides a clearer understanding of which factors have the most significant impact on core business characteristics and outcomes. This ensures that the segmentation is not only relevant but also actionable, helping stakeholders prioritize segments that will drive the most value. Box 11 provides an example of the regression analysis exercise.

BOX 11. **Regression analysis example**

A final step is to conduct a regression analysis of each remaining dimension, specifically on key outcome variables including business growth, formal borrowing, factors that indicate resilience, and the proportion of operational costs covered by borrowing. The regression analysis will help highlight which dimensions may have a stronger association and, hence, greater explanatory power when it comes to different outcomes. The regression analysis will have the following structure:

$Yi = \alpha+\beta iXi+\gamma Zi+\epsilon$

Yi represents the vector of outcome variables.

Xi represents the vector of independent indicators derived from the shortlisted segmentation dimensions.

βi are the coefficients that measure the impact of the studied dimension indicator on the outcome variable.

Zi is a control variable representing business type and gender of business owner to account for structural differences across these categories.

If the outcome variables are numerical (e.g., age of enterprise, revenue, number of employees), a linear regression model should be used. If the outcome variables are categories, logistic regressions are appropriate. For independent variables that are categories, it is best to turn each category into a separate dummy variable.

Table 12 shows an example of a results table for a regression analysis of the mindset and entrepreneurial ability dimensions.

Based on data from Kenya, Table 12 presents a regression analysis linking different segmentation dimensions and their single indicators to key business outcomes (e.g., revenue, employment, growth, borrowing, resilience). The highlighted cells indicate statistically significant relationships, and asterisks denote levels of significance.

Key findings include:

- Entrepreneurs driven by necessity, such as supporting their families, face greater challenges in achieving business resilience and growth compared to those motivated by passion or skills.
- Businesses run by sole income earners were 14 percent less likely to be resilient to shocks compared to businesses run by owners that were not their household's sole income earner. This could speak to sole income earners not having sufficient financial support when their business goes through difficulty because their family is comprised of dependents.
- Entrepreneurs who received formal training were 1.8 years younger than their business counterparts who had not received training, suggesting that

BOX 11. **Regression analysis example (continued)**

TABLE 12. **Regression analysis on outcome variables**

Segmentation dimension	Indicator	Age of enterprise	Business Revenue	Formal borrowing	Recovered from recent shock (Resilience)
Mindset	Survivalist mindset	-1.56 **	-8480.861	.03	-.12 **
	Sole income earner in household	1.27	-840.26	-.05	-.14 ***
Entrepreneurial ability	Education level	-.18 ***	519.53	.01 ***	0
	Lack of formal training	-1.81 **	-21406.05 ***	-.07	-.06

Source: Authors using data from CGAP comprehensive business survey, 2024

the desire for training is significantly higher for businesses in their early stages.

- Entrepreneurs who did not receive formal business training generated business revenue that was 21,406 Ksh less relative to entrepreneurs who had received formal business training.

Key considerations when interpreting results include:

1. Causality vs. correlation: The relationships shown do not necessarily indicate causation.
2. Context matters: Understanding the economic environment and financial landscape of the surveyed businesses is crucial for accurately interpreting these findings.
3. Magnitude vs. significance: A statistically significant result (denoted by asterisks) does not always mean a strong impact. The size of the coefficient is equally important in assessing the practical relevance of a finding.
4. Possible omitted variables: Other unmeasured factors (e.g., industry type, macroeconomic conditions) could influence these relationships.

APPENDIX 4

Alternative segmentation approaches

WHEN TRADITIONAL SEGMENTATION approaches fall short, alternative methods like exploratory cluster analysis can offer fresh insights on segments. This appendix delves into the underlying patterns in data, identifying clusters that may not align neatly with predefined categories but can serve as a valuable starting point for refinement. It is useful to note that although exploratory cluster analysis often produces complex, nonactionable results, techniques like cross-tabulation or confirmatory analysis can refine these into meaningful segments. This appendix highlights advanced methods that can be used to uncover nuanced patterns or distinct behavioral segments, enabling more tailored and effective strategies.

Exploratory cluster analysis entails identifying natural groupings or patterns within a dataset by grouping similar observations based on selected variables without prior assumptions (Kaufman and Rousseeuw 1990). Exploratory cluster analysis generally reveals clusters that are not that easy to interpret, which means that cross-tabulation or confirmatory cluster analysis methods are preferable. However, it can be used as an initial starting point from which to fine-tune the results. For instance, an exploratory cluster analysis might render the following two segments:

- Motivated, Growth plans, Non-primary earner, Solely responsible for household work, Married, Not household head, Low general education, No formal business training, Advanced business skills, Unregistered, Non-tax paying, Nano, Manufacturing Sector, Urban
- Motivated, Growth plans, Primary earner, Not solely responsible for household work, Married, Household head, High general education, No formal business training, Basic business skills, Registered, Tax paying, Nano, Manufacturing sector, Rural

These are not easy-to-interpret or meaningful clusters for strategy purposes. Thus, it is possible to consider crossing out all the characteristics that are similar in the two segments and focus on the differences:

- ~~Motivated, Growth plans~~, Non-primary earner, Solely responsible for household work, ~~Married~~, Not household head, Low general education, ~~No formal business training~~, Advanced business skills, Unregistered, Non-tax paying, ~~Nano, Manufacturing Sector~~, Urban.
- ~~Motivated, Growth plans~~, Primary earner, not solely responsible for household work, ~~Married,~~ Household head, High general education, ~~No formal business training~~, Basic business skills, Registered, Tax paying, Nano, ~~Manufacturing sector~~, Rural.

This then leaves one with household role, ability, business sophistication, and location related segmentation dimensions to define the segments.

Two different clustering techniques are well-suited in this context: (i) K-means clustering, which divides the sample into K clusters based on similarity, and (ii) hierarchical agglomerative clustering, which builds

a tree-like structure to represent the relationships between data points and their clusters.[9] For both methods, the cluster analysis can be run with the segmentation dimensions or with the selected outcome indicators. The former allows identification of clusters of WNMEs with similar patterns across multiple dimensions, whereas the latter allows identification of distinct behavioral segments, such as "resilient savers" or "growth-oriented borrowers," uncovering patterns not captured by predefined dimensions.

9 K-means clustering splits data into a set number of groups (clusters) by finding similarities between data points. Each point is assigned to the group whose center it is closest to, creating clusters of data with similar traits. This method is useful for identifying groups within a dataset based on shared characteristics. Hierarchical agglomerative clustering (HAC) starts by treating every data point as its own cluster, then gradually merges the closest clusters together to form a hierarchy. The result is a tree-like diagram (dendrogram) that shows how clusters are related, making it useful for understanding relationships and groupings in the data. For more on these clustering techniques, refer to K-means clustering (Salazar 2023) and HAC (Ranjan 2020).

APPENDIX 5

Credit market sizing

ESTIMATING THE MARKET SIZE FOR credit presents distinct challenges compared to other products or services. In contrast to other financial products, where demand can often be gauged by simply counting the number of accounts created or by estimating consumers' willingness to take it up, credit is a dynamic service that involves not only willingness but also eligibility and repayment capacity. Access to credit depends not only on a consumer's desire for it but also on meeting specific requirements, and particularly demonstrating creditworthiness. This adds layers of complexity in understanding potential demand at any given time and assessing the size of the potential credit market. To navigate this complexity, a more nuanced approach may be required—one that combines multiple data sources and analytical methods. It includes leveraging financial and behavioral data to assess both willingness and eligibility, segmenting the market based on creditworthiness indicators, and incorporating proxy indicators such as income stability, employment trends, and historical repayment behavior. Additionally, partnerships with financial institutions and alternative data providers can help fill gaps in traditional credit assessments, enabling a more comprehensive and accurate estimation of market potential.

Notwithstanding the complexities outlined above, there are three core steps to assessing the size of the opportunity (i.e., the credit gap in the market):

1. **First, calculate demand for credit.** There are several methods for estimating the market size for credit, each with its own advantages and disadvantages. Select the approach that aligns best with (i) the objective set out in Stage 1, (ii) the granularity required (low-, medium-, or upper-bound credit estimates), and (iii) the available data. Some of the most popular methods and their benefits include:

   a. **Supply-side loan data that can reveal overall demand for credit**. This is the simplest approach as it only looks at enterprises that have applied for a loan. Therefore, it also gives the lowest bound credit estimate. The method involves using data gathered on the number of loans applied for, specifically how many loans were applied for, and the median (most common) amount applied for. It then looks at the supply-side outstanding loan data: how many applicants received loans and the median amount received. This does, however, fail to account for the potential demand. Many small businesses, for example, may not apply for credit due to the high cost or unavailability of suitable credit products.

   b. **Survey-based estimations that can reveal past demand and perceived demand.** This approach looks at the actual demand as identified through the supply-side loan approach, as well as the perceived need. It thus gives a medium bound estimate. This involves directly asking consumers about their interest in more credit, typically through surveys, and asking what type of credit they would be interested in. It can also be used to understand past credit behavior, which enables

the surveyor to assess the need for credit by looking at whether people applied for a loan and, if so, for what amount. While this method provides direct feedback from potential customers, it may not fully capture the complexities of creditworthiness. Consumers may express a desire for more credit yet not have the ability to qualify for it or repay it, thereby skewing the results.

c. **Debt-to-revenue ratio approach that can reveal potential demand.** This is one of the most widely used methods for estimating credit demand, particularly for MSMEs. It leverages more data and thus also provides an upper bound estimate. This ratio is used to determine how much debt a business can sustainably absorb based on its revenue generation capacity. Applying this ratio can generate an estimate of the amount of credit an enterprise could potentially require for business expansion, working capital, or other operational needs (IFC 2017). A good starting point is to use the values provided by the IFC as a benchmark. While these benchmark values are a good place to start, it is important to note that they were developed as an ideal and may thus lack significant country context realities which could result in a skewed outcome. The values may also not be as relevant for the smallest enterprises or NMEs.

   In the case of a small business, for example, suppose annual revenue is $500,000 and the industry average debt-to-revenue ratio is 0.25. These figures suggest that the business could absorb $125,000 in debt. If the total number of small businesses in the market was 10,000, total market demand could be estimated at $1.25 billion. However, this method requires accurate revenue data and assumes that all businesses are equally likely to require debt. The advantage of the approach is that it is relatively simple to apply and provides a more realistic picture of the amount of debt businesses can handle without becoming financially overburdened. However, it also has limitations. For instance, it may not account for differences in profitability, cash flow, or risk tolerance among businesses within the same industry, which could lead to variations in the amount of debt they can service. It is important to note that since the NME sector is only now being explored in greater detail, the applicability of a 0.25 ratio remains uncertain. This uncertainty is compounded by the intricate nature of household and business finances within the sector, making it a challenge to derive precise financial estimates.

2. **Next, assess credit supply.** Similarly to assessing demand, there are several methods for estimating supply. Some of the most popular methods include:

   a. **Survey approach.** Surveys can help identify which applicants received credit and whether they obtained the full amount requested. To understand credit supply, enterprises can be asked about their current outstanding loans or the financing they have accessed in the past. A concern with this approach is its reliance on self-reported data, which may lack accuracy due to recall biases. However, a key benefit is that the approach allows for the inclusion of both formal and informal credit sources. A critical consideration is whether to include certain sources of capital (e.g., informal loans from moneylenders) in these calculations. Although those sources may meet capital needs, it may not be in a sustainable way.

   b. **Leverage available/accessible supply-side data for a more reliable result.** Leveraging loan portfolio data from those supplying the loans (e.g., FSPs; government agencies; those consolidating borrower data, such as credit bureaus [less relevant for NMEs]) can yield more reliable insights into loan dynamics. However, getting reliable supply-side data on informal loan portfolios may be challenging. Additionally, there may be challenges with

disaggregating the supply-side data to gain insights on specific segments.

3. **Finally, calculate the size of the opportunity (the gap) between demand and supply.** This is simply done by subtracting the answer in Step 2 from the answer in Step 1 (demand – supply = gap). The calculation provides an indication of the credit market size, noting the importance of ensuring responsible finance as set out in Stage 3 of the main text.

While no estimation is perfect, this credit market sizing exercise helps to identify which segments are most in need of intervention and where market-driven solutions can be effectively implemented. Leveraging different approaches and data to estimate demand, supply, and the gap can help improve the accuracy of the results and get a better sense of the lower- and upper-bound of the credit market gap. As such, understanding market size is essential for determining where to direct resources and how to structure interventions to reduce market inefficiencies and foster inclusive economic growth.

Once market sizing is complete, the next step is to build a strategy to address each segment identified in Stage 2 of the main text. Understanding market sizes and gaps enables stakeholders (e.g., researchers, donors, FSPs) to identify which segments can be served by market-driven solutions and which may require additional intervention (e.g., donor or government support, reliance on informal providers). The various options for solutions are described below:

- **Market-driven solutions.** Some segments, such as well-established enterprises, may present opportunities for purely market-driven financial services. These segments are typically more creditworthy and have predictable cash flows, making them attractive to traditional FSPs.
- **Nudge-driven solutions.** For other segments, a small nudge, such as better risk assessments or more tailored financial products, may be sufficient to encourage FSPs to engage with underserved groups.
- **Government support.** In some instances, government support may be required to bridge gaps, particularly for larger scale infrastructure projects or for individuals and businesses that are not yet banked or creditworthy under current market conditions.

Key considerations for conducting a credit market sizing exercise

THE VALUE OF SUPPLY-SIDE DATA

In the case of credit market gap sizing, do not underestimate supply-side data. Supply-side data is crucial for credit market sizing because it provides insights into the availability and distribution of credit within the market. It helps identify how much credit is currently being offered, which is essential for assessing whether the market is meeting demand or if gaps exist (IFC 2018). This data also reveals the types of credit products available, and their terms and conditions, enabling an evaluation of whether these offerings align with the needs of potential borrowers. By understanding the existing credit supply, stakeholders can better gauge market saturation, forecast future trends, and identify areas where interventions are needed to address unmet demand or optimize product offerings.

CHALLENGES OF CREDIT DEMAND ESTIMATION

Be aware of the challenges of estimating total credit demand. Even with the average gross loan portfolio, it remains difficult to estimate the total potential revenue and profit available in the market. To achieve this, it is necessary to incorporate additional information, such as average interest rates, the cost of funds, transaction sizes, loan tenors, and operating expenses. This broader set of factors adds complexity but provides a more complete view of the financial ecosystem, enabling more accurate market sizing. A further challenge comes from distinguishing between productive credit, which is used for business growth, and consumer credit, which is used for consumption. This is particularly difficult with the smallest enterprises, where personal and business finances are often intertwined, complicating the process of accurately identifying the credit demand of each sector.

INCORPORATING RESPONSIBLE LENDING PRINCIPLES

Incorporating responsible lending and credit quality into market analysis is essential. Focusing solely on the demand-supply gap risks overlooking critical aspects of credit risk that may justify stricter lending policies. By examining indicators of credit quality alongside credit gaps, a more nuanced understanding of the lending landscape is gained, ensuring that financial products are designed to avoid over-indebtedness and promote sustainable lending practices. This is especially important in markets where a majority of borrowing occurs informally and many potential borrowers are outside the formal credit system, making accurate debt profile information difficult to obtain. Therefore, integrating responsible finance measures that account for these complexities is crucial to promoting financial inclusion without compromising borrower well-being. This ultimately ensures that credit facilitates positive outcomes and prevents debt traps.

References

Alonso, Tatiana, and Diana Dezso. 2024. *Supply-Side Gender Data is a Game Changer for Financial Inclusion*. February 27. https://www.cgap.org/blog/supply-side-gender-data-is-game-changer-for-financial-inclusion.

CGAP. 2016. *Customer Segmentation Toolkit*. https://www.cgap.org/research/publication/customer-segmentation-toolkit.

—. 2024a. "Literature Review: Micro and Small Enterprise (MSEs') Financing in Kenya, Uganda and India." https://www.findevgateway.org/paper/2024/08/literature-review-micro-and-small-enterprise-mses-financing-in-kenya-uganda-and-india.

—. 2024. *The Next Chapter for Women's Financial Inclusion: Moving Toward a Visible Step-Change*. https://www.cgap.org/research/publication/next-chapter-for-womens-financial-inclusion-moving-toward-visible-step-change.

Citigroup. 2022. "Women Entrepreneurs." https://www.citigroup.com/global/insights/women-entrepreneurs_20220304?utm_source=chatgpt.com.

Community Development Finance Institutions . n.d. *Responsible Finance*. https://responsiblefinance.org.uk/responsible-finance-providers/what-is-responsible-finance/.

Crittenden & Crittenden . 2021. "Empowering Women Through Micro-entrepreneurship." https://link.springer.com/referenceworkentry/10.1007/978-3-319-95687-9_156.

Directorate-General for Financial Stability, Financial Services and Capital Markets Union. 2019. *FSUG opinion and recommendations for the review of the consumer credit directive*. https://finance.ec.europa.eu/publications/fsug-opinions-2019_en.

Dunn, Edgar. 1956. *The Market Potential Concept and the Analysis of Location*. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1435-5597.1956.tb01564.x.

European Commission. 2024. *The Ten Step Pocket Guide to Composite Indicators and Scoreboards*. https://knowledge4policy.ec.europa.eu/composite-indicators/toolkit_en.

FinDev Gateway. 2024. *Segmentation of Women-Led Nano and Micro-enterprises: A Guide to Knowledge Resources*. https://www.findevgateway.org/guide/2024/09/segmentation-of-women-led-nano-and-micro-enterprises-guide-to-knowledge-resources.

Hewitt, Sarah. 2024. *Small business, big impact: The transformative power of women-led enterprises*. https://www.weforum.org/stories/2024/06/small-business-big-impact-the-transformative-power-of-female-entrepreneurs/.

IFC. 2018. "Financing India's MSMEs: Estimation of Debt Requirement of MSMEs in India." https://documents1.worldbank.org/curated/en/759261548828982149/pdf/134150-WP-IN-Financing-India-s-MSMEs-Estimation-of-Debt-Requirement-of-MSMEs-PUBLIC.pdf.

—. n.d. *IFC's Definitions of Targeted Sectors*. https://www.ifc.org/en/what-we-do/sector-expertise/financial-institutions/definitions-of-targeted-sectors?utm_source=chatgpt.com.

—. 2017. "MSME FINANCE GAP - ASSESSMENT OF THE SHORTFALLS AND OPPORTUNITIES IN FINANCING MICRO, SMALL AND MEDIUM ENTERPRISES IN EMERGING MARKETS." https://documents1.worldbank.org/curated/en/653831510568517947/pdf/121264-WP-PUBLIC-MSMEReportFINAL.pdf.

International Finance Corporation. n.d. https://www.ifc.org/en/what-we-do/sector-expertise/financial-institutions/definitions-of-targeted-sectors.

Kaufman, and Rousseeuw. 1990. "Finding Groups in Data: An Introduction To Cluster Analysis." https://www.researchgate.net/publication/220695963_Finding_Groups_in_Data_An_Introduction_To_Cluster_Analysis.

KNBS. 2016. "MSME Survey." https://statistics.knbs.or.ke/nada/index.php/catalog/69/download/236.

Kremelberg, David. 2011. *Pearson's r, Chi-Square, T-Test, and ANOVA*. Sage Publications, Inc. . https://us.sagepub.com/sites/default/files/upm-binaries/33663_Chapter4.pdf.

McKee, Kate, Estelle Lahaye, and Antonique Koning. 2011. *Responsible Finance: Putting Principles to Work.* https://www.cgap.org/research/publication/responsible-finance-putting-principles-to-work.

Ministry of industrialization, trade and enterprise development. 2020. "Kenya micro and small enterprises policy." http://www.parliament.go.ke/sites/default/files/2020-10/Sessional%20Paper%20on%20the%20Kenya%20Micro%20and%20Small%20Enterprises%20policy%20from%20the%20Ministry%20of%20Industrialization%2C%20Trade%20and%20Enterprise%20Development.pdf.

OECD/European Union/EC-JRC. 2008. *Handbook on Constructing Composite Indicators: Methodolgy and User Guide.* Paris: OECD Publishing. https://doi.org/10.1787/9789264043466-en.

—. 2008. *Handbook on Constructing Composite Indicators: Methodology and User Guide.* Paris: OECD. doi:https://doi.org/10.1787/9789264043466-en.

—. 2008. *Handbook on Constructing Composite Indicators: Methodology and User Guide.* Paris: OECD. https://www.oecd.org/en/publications/handbook-on-constructing-composite-indicators-methodology-and-user-guide_9789264043466-en.html.

Ranjan, ! 2020. *Hierarchical Clustering (Agglomerative).* https://medium.com/analytics-vidhya/hierarchical-clustering-agglomerative-f6906d440981.

Salazar, R. 2023. *Getting started with K-means clustering in Python.* https://domino.ai/blog/getting-started-with-k-means-clustering-in-python.

Sawhney, Swati, Salome Kimani, and Alexander Sotiriou. 2025. *Diverse Paths: Finance for Women's Nano and Micro Enterprises.* CGAP. https://www.cgap.org/research/publication/diverse-paths-finance-for-womens-nano-and-micro-enterprises

Schuber, Estrada and. 2022. "Women, SMEs and sustainable development – lessons learnt for the road ahead." https://www.unido.org/stories/women-smes-and-sustainable-development-lessons-learnt-road-ahead.

Shahapure, Ketan, and Charles K. Nicholas. 2020. *Cluster Quality Analysis Using Silhouette Score.* https://www.researchgate.net/publication/347085825_Cluster_Quality_Analysis_Using_Silhouette_Score.

Statistics Solutions. 2025. *Confirmatory Factor Analysis.* https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/confirmatory-factor-analysis/.

Statstutor. n.d. https://www.statstutor.ac.uk/resources/uploaded/spearmans.pdf.

Taghiyeva. 2023. "Empowering women micro-entrepreneurs through mobile." https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-for-development/wp-content/uploads/2023/02/Empowering-women-micro-entrepreneurs-through-mobile.pdf.

UNEPFI. n.d. *Principles for Responsible Banking: Shaping the Future of Banking.* Edited by United Nationals Environment Program Finance Initiative. https://www.unepfi.org/banking/bankingprinciples/.

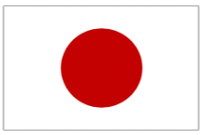
Vidal, Caire & Baron. 2019. "How financial services providers can use data analytics to better segment and serve customers." https://www.cgap.org/sites/default/files/publications/2019_07_Technical_Guide_Data_Driven_Segmentation_0.pdf.

Women's World Banking. 2024. "Advancing Women-Led MSMEs through Digital Financial Inclusion." https://www.womensworldbanking.org/wp-content/uploads/2024/06/WDFI-Advocacy-Hub_WhitePaper.pdf.

World Bank Group. 2013. *Responsible lending : overview of regulatory tools (English).* Washington, D.C. http://documents.worldbank.org/curated/en/596151468336064796.

World Bank. 2022. *Responsible finance and its role in improving financial resilience and well-being.* https://blogs.worldbank.org/en/developmenttalk/responsible-finance-and-its-role-improving-financial-resilience-and-well-being#:~:text=As%20the%20type%20of%20risks,to%20independently%20use%20financial%20products.

CGAP Members



CGAP members and strategic partners as of June 2024

CGAP Members (continued)

CGAP Strategic Partners

CGAP members and strategic partners as of June 2024

Transforming Lives with Financial Inclusion

cgap.org