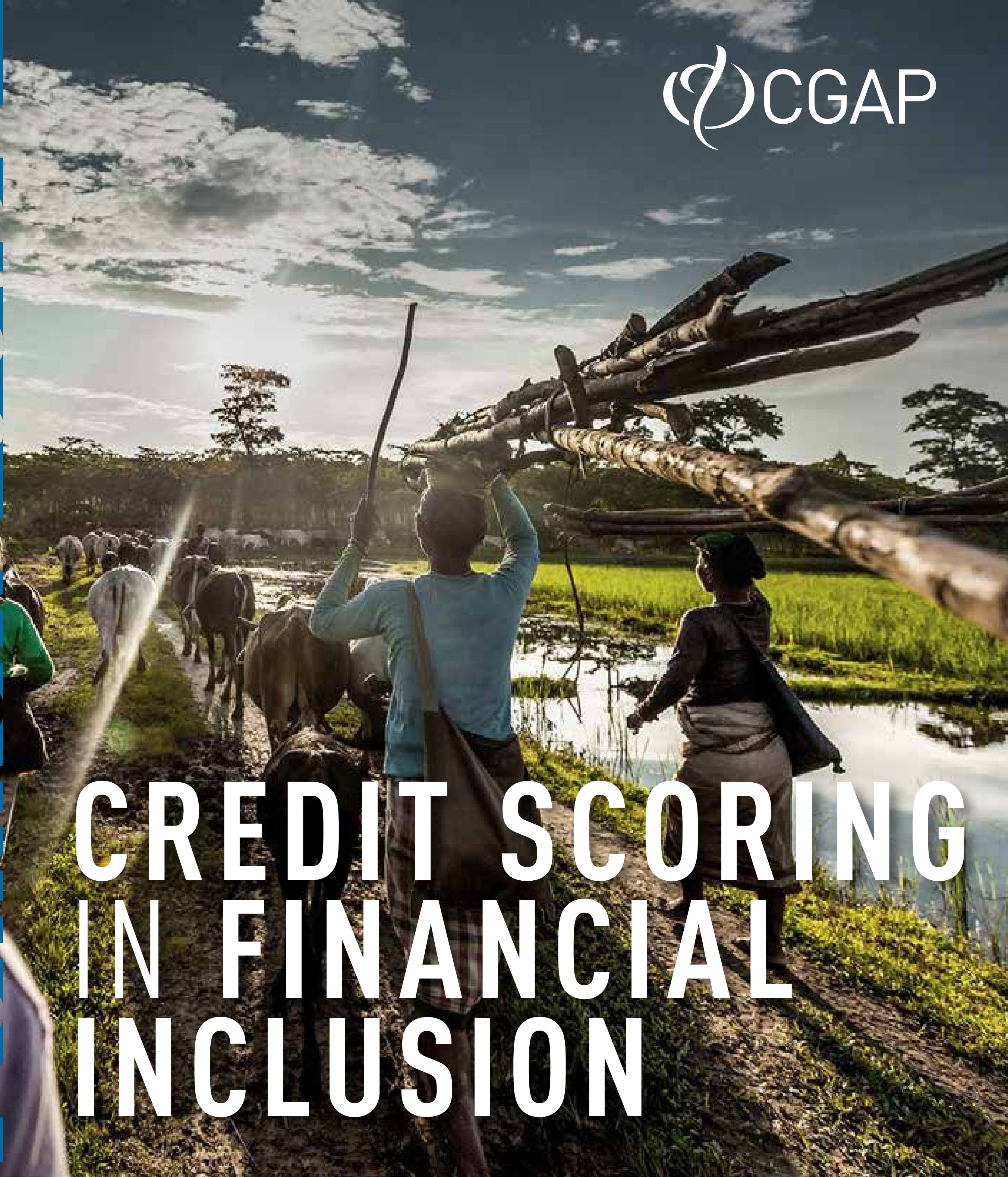
TECHNICAL GUIDE

CGAP



CREDIT SCORING IN FINANCIAL INCLUSION

How to use advanced analytics to build credit-scoring models that increase access

July 2019

Maria Fernandez Vidal and Fernando Barbon

Consultative Group to Assist the Poor

1818 H Street NW, MSN F3K-306

Washington DC 20433

Internet: www.cgap.org

Email: cgap@worldbank.org

Telephone: +1 202 473 9594

Cover photo by Sujan Sarkar, India.

CONTENTS

EXECUTIVE SUMMARY

STATISTICAL MODELS CAN HELP LENDERS IN EMERGING markets standardize and improve their lending decisions. These models define customer scoring based on a statistical analysis of past borrowers' characteristics instead of relying on the subjective judgments of loan officers. Evidence shows that statistical models improve the accuracy of credit decisions and make lending more cost-efficient. They also help companies make key decisions throughout the customer lifecycle.

Lenders sometimes assume that statistical credit scoring is too costly or difficult or that they do not have the kind of data needed to implement it. However, the primary input needed for this type of modelling is something many providers already possess: customers' repayment histories. This guide explains what types of data lenders can leverage for statistical credit scoring and the ways in which it can be used.

Furthermore, different statistical models can be used for building credit scores. Lenders who are new to data analytics can start with a simple model and tailor it over time to meet their needs. In this guide, readers will find a step-by-step approach to building, testing, fine-tuning, and applying a statistical model for lending decisions based on a company's growth goals and risk appetite.

This guide emphasizes that the effectiveness of data analytics approaches often involves building a broader data-driven corporate culture.

INTRODUCTION

THIS IS A STEP-BY-STEP GUIDE TO the methodologies, processes, and data that financial services providers can use to develop new credit scoring models. It is particularly relevant for markets that have limited credit bureau coverage and for providers who want to target customers who are traditionally excluded from formal credit. The Guide will show you how to conduct a scoring model project with limited external data and will provide real-life insights about opportunities and potential pitfalls from experience in the field. The Guide applies statistical theory to real credit scoring situations. [1]

Besides providers, others who work in financial services would find this Guide to be useful. These include loan officers, risk managers, and data scientists. Chief financial officers and chief executive officers can use this Guide to help them make decisions about a new loan product or lending process reform. The Guide is written from the perspective of a project manager because project managers often need to ensure that the business side of a company understands the technical and statistical work and that technical staff understand the company's business needs.

The techniques described here are meant to help organizations become more efficient and effective in providing financial services to their customers. They offer a simple, yet effective, credit scoring methodology and guidance around processes and decisions, including the knowledge, skills, tools, and data sources, needed when developing and deploying a new credit scoring project using internal and some limited external data sources.

This Guide addresses the following:

- How credit scoring works.
- Benefits of data-driven credit scoring methodologies.
- How to use data analysis in different scenarios, depending on access to data and data quality.
- How to deploy a credit scoring project and the resources and processes needed.
- Commonly used analytical techniques.
- How to use the data produced to create new and better credit products.

The Guide concludes with an illustrative case study of a microfinance organization.

1 See Anderson (2007) for more information.

BENEFITS OF CREDIT SCORING

CREDIT SCORING CAN HELP financial institutions grow their portfolios by lowering the cost of serving low-income customers and increasing the quality of service and customer satisfaction.

A credit scoring model is a risk management tool that assesses the credit worthiness of a loan applicant by estimating her probability of default based on historical data. It uses numerical tools to rank order cases using data integrated into a single value that attempts to measure risk or credit worthiness.

The decision-making process for credit scoring can be either subjective or statistical (Schreiner 2003).

Subjective scoring relies on the input of an expert, the loan officer, and the organization to produce a qualitative judgment.

Statistical scoring, on the other hand, relies on quantified characteristics of the prospect's portfolio history recorded in a database. It uses a set of rules and statistical techniques to forecast risk as a probability.

The two approaches complement each other and bring different benefits and challenges, as shown in Table 1. In this Guide, "scoring" refers to statistical scoring.

Statistical scoring models are:

- **Empirical.** Based on a rigorous statistical analysis that derives empirical ways to distinguish between more and less creditworthy consumers using data from applicants within a reasonable preceding period.
- **Statistically valid.** Developed and validated based on generally accepted statistical practices and methodologies.

TABLE 1. **Comparison of Subjective and Statistical Scoring**

Dimension	Subjective Scoring	Statistical Scoring
Source of knowledge	Experience of loan officer and organization	Quantified portfolio history in database
Consistency of process	Varies by loan officer and day-to-day	Identical loans scored identically
Explicitness of process	Evaluation guidelines in office; sixth sense/gut feeling by loan officers in field	Mathematical rules or formulae relate quantified characteristics to risk
Process and product	Qualitative classification as loan officer gets to know each client as an individual	Quantitative probability as scorecard relates quantitative characteristics to risk
Ease of acceptance	Already used, known to work well; MIS and evaluation process already in place	Cultural change, not yet known to work well; changes MIS and evaluation process
Process of implementation	Lengthy training and apprenticeships for loan officers	Lengthy training and follow-up for all stakeholders
Vulnerability to abuse	Personal prejudices, daily moods, or simple human mistakes	Cooked data, not used, underused, or overused
Flexibility	Wide application, as adjusted by intelligent managers	Single application, forecasting new type of risk in new context requires new scorecard
Knowledge of trade-offs and "what would have happened"	Based on experience or assumed	Derived from tests with repaid loans used to construct scorecard

Source: Schreiner 2003

- **Periodically revalidated.** Re-evaluated for statistical soundness from time to time and adjusted, as necessary, to maintain or increase its predictive power.

These models are especially useful in lending situations where the lender must manage a large volume of credit assessments for loan amounts that are relatively low and generally for retail credit for individuals and small businesses.

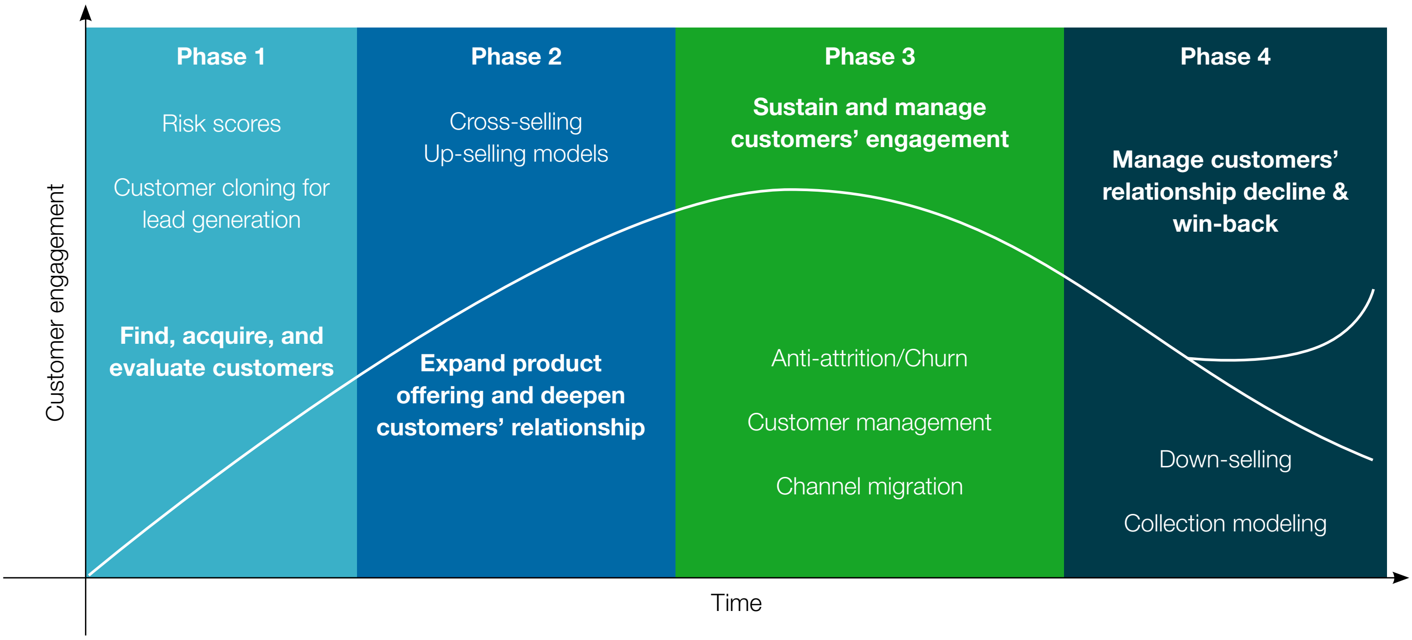
Even though the most common application of credit scoring is to assess credit worthiness, financial institutions (FIs) also use it to help them make decisions at other stages of their customers' life cycle. For each stage of the customer life cycle, there is a different scoring type based on specific data. Figure 1 illustrates this dynamic (CGAP 2016).

An application scoring model focuses on selecting the borrowers to approve from the applicant pool. Using an automated application credit scoring solution has several benefits, including:

- Operational efficiency gains
  - Reduce cost and time from manual risk assessment
  - Reduce customer turnaround with fewer in-person interactions
  - Lower administrative costs per unit
  - Lower the number of in-person interactions with prospective borrowers
- Improved accuracy of credit decisions (targeted lending based on default probability)
  - Minimize rejection of creditworthy applicants
  - Maximize rejection of high-risk applicants
- Establishment of an objective and standardized data-driven decision-making culture
  - Apply base objective and consistent decision making on empirical evidence
  - Standardize criteria for decision making
  - Minimize room for human error or bias

FIGURE 1. **The borrowing process customer lifecycle**

FIGURE 2. **Enhance the customer relationship throughout the lifecycle using scoring models**



- Loan uptake effectiveness
  - Target loan products to increase conversion rate
- Customer satisfaction improvement (faster, more targeted loans)
  - Reduce customer turnaround
  - Pre-approve specific products, upgrades, and cross-sells

Credit scores can also enhance the relationship with customers at the different stages in their lifecycle (see Figure 2):

- Assign credit limits based on risk level
- Offer applicants additional savings products as collateral to become eligible
- Price loan products based on risk level
- Adjust terms and shorten repayment schedules
- Offer additional products and upgrades
- Offer automated renewals for specific products
- Prequalify customers for new products or cross-sell products
- Redirect delinquent accounts to collection companies

Implementing a credit scoring model also has its challenges. For example, implementation can be complicated; it might require significant investment, depending on system capability; and it will require specific technical skills that are not always readily available. Additionally, because models are based on historical data, they are backwards looking, and they assume that the future will look like the past. Moreover, working with a limited amount of data significantly increases the risk of developing a biased model that predicts well only when it is applied to the original sample. Sections 2.5, 3.5, 3.6, and 5 address this issue in more detail.

The first step in credit scoring is to develop a scorecard. To manage large amounts of data, categorize the sources of information based on their broadest distinction: internal or external sources. In developing economies or when targeting low-income and first-time borrowers, external credit history sources may be scarce and less robust than those in developed countries. Despite this, institutions already have access to one of the strongest predictors of default—the repayment history of their own customers.

SECTION **1**

DATA FOR AUTOMATED CREDIT SCORING

WHEN CREATING A CREDIT-scoring service, the scoring algorithm, methodology, and processes you use depend on your organization's objectives. Several important questions need to be answered, for example:

- Is the scoring solution for a new or for an existing loan product?
- Is the scoring solution for a new or an existing customer?
- Are previous loan performance data available? Are these data reliable, updated, and accessible?
- Are there external credit data sources?
- Are there nonfinancial external data sources?

Although many of the processes, methods, and tools covered in this Guide can be a good fit for several different credit scoring projects, from a loan performance data standpoint, the following two scoring scenarios are considered:

1. Scoring solution with performance data for comparable products available (see Section 2, "Setting up a credit scoring project").
2. Scoring solution without performance data for comparable products available:
    2.1. Scoring solution with external data for comparable products for comparable customers available (see Section 1.1, "Using bureau data to develop a credit scoring model")
    2.2. Judgmental scorecard solution for new products when no data for comparable products are available (see Section 1.2, "Piloting to generate repayment data")

1.1 Using Bureau Data to Develop a Credit Scoring Model

Many financial institutions are exploring new and different lending models, either to offer new product features (fully digital services, much smaller amounts, much shorter terms, etc.) or to target new customer segments. They often discover that they do not have the historical performance data they need to create a scoring model. If external credit bureau data are available for their customers, most organizations can use the bureau's score as an input for their underwriting process. These scores aren't as predictive as tailored ones, which are developed ad-hoc using specific customer data. However, some credit bureaus offer different levels of customization of their scores. If the external credit bureau does not have robust information on the organization's customer base, the organizations may turn to customer cloning or profiling to provide a score for their other loan applicants.

In the first step in customer cloning, the institution uses the bureau's score as a proxy for "good" borrowers and identifies the lowest score that it wants to consider lending to. Then, it should model the probability of having a good score against the information available for all customers (sociodemographic, business data, transactional data for a different product, etc.).

1.2 Piloting to Generate Repayment Data

Financial institutions may want to launch a new credit product. They may need to develop a scoring solution, but they may not have loan repayment information for comparable products and customers. Or more commonly, their internal data are not reliable, and there are no external sources of credit data. Organizations can overcome this challenge by creating internal repayment data. The first step is to launch a controlled lending pilot to learn about the repayment behavior of your customers. After piloting a few loan cycles, you will have some default data and can develop a statistical model based on the customers' newly gathered "historical performance data."

PRODUCT FEATURE DEFINITION

The first step is to define the features of the product to be piloted. The product should be designed bearing in mind the needs of the customers and the competitive strengths of the organization. Consider the following when defining these features:

- To collect performance data, the loans need to mature. Therefore, the shorter the term, the sooner the data will be available. However, if the end goal is to score customers for a two-year productive loan, the information collected from two-week loans may not be directly relevant, and the customer profile might be too different for it to be valuable.
- The lower the amount, the more loans can be disbursed for a given risk investment. The same caveat applies: although it is generally prudent to start with smaller loans than those expected in the longer term, the amount must be similar enough to yield relevant data.

To determine the right product features, you will need to know as much as possible about the target customers. For example, here are some questions you'll need to answer:

- What amount is significant to them?
- Is their economic activity or the purpose of the loan seasonal?
- What is the customers' repayment capacity?
- Do they care more about availability or about interest?

You can gather this information through qualitative studies such as surveys, focus groups, and interviews.

EXPERT SCORECARD DEVELOPMENT

Most organizations that want to generate repayment data develop a "judgmental" or "expert" scorecard to decide which customers they should lend to for the product they want to launch. The expert scorecard is meant to predict the likelihood of default based on certain characteristics of the customers or "discriminants." There is a tradeoff between the risk an organization is willing to take and the bias that the expert scorecard will introduce in the sample. The cleanest sample possible is when all applicants are accepted or when applicants are accepted or declined randomly. This allows the model to be developed for a sample that perfectly aligns with the pool of applicants that will be screened in the future. However, this approach may lead to high loses in the initial cycle, especially for loans of relatively large amounts. Most organizations choose to start with a simple version of an expert scorecard that has a minimum set of rules or conditions that need to be met. This can mean setting up knock-off rules (e.g., no applicant under 25 is accepted, no applicant who is unemployed is accepted, etc.) and using a scorecard that assigns points to each characteristic, which allows a user to "compensate" for a low score in one factor with a higher score in another.

Loan experts need to work to together to identify the customer characteristics they believe are associated with default. Examples of discriminants include the following:

- Age
- Residence (own, rent)
- Number of years at residence
- Occupation
- Phone ownership
- Income
- Years in current job
- Previous employment
- Years in previous job
- Number of dependents
- Credit history
- Banking account ownership
- Credit outstanding
- Outstanding debts
- Purpose of loan
- Type of loan

For each discriminant, the loan experts should identify risk groups and segment customers they believe exhibit different risk behavior. For example, for the discriminant "age," you may consider that risk decreases as age increases and that the age "18–23" group repayment performance is likely to be significantly worse than for the "23–28" group.

The team assigns points for each risk group that reflect differences in risk behavior. The risk relationships should be simple and straightforward and intuitively make sense.

A judgmental scorecard must be as simple as possible. It needs to ensure that the discriminants can be easily and accurately measured, and it should confirm that the risk relationships identified make sense to the business.

PILOT IMPLEMENTATION

Once the product is defined and the judgmental scorecard developed, the servicing strategy needs to be defined. The organization should choose the quickest, most profitable channel available to the target customers. The idea is to set up a process comparable to the one the organization intends to use in the long run. Bearing in mind the features of the product chosen, the number of loan disbursements per wave should be maximized as much as the capital available allows. Ideally, the product being piloted should have some of the following characteristics:

- Low cost of service
- Short turnaround
- Small amounts
- Short terms
- Promptly collected and accurate data
- Ability to scale quickly

In the pilot, the product will be offered to preselected groups of customers based on the expert scorecard. The goal is to learn about the repayment behavior of your customers. The ultimate objective is to gather performance data to develop a statistically significant credit-scoring model.

Keep in mind that it is necessary to have a significant sample of customers that default on their obligations or at least pay past the initial term. Otherwise, the model cannot identify the predictors of default. A less conservative cut-off for approvals will provide more data in a shorter time frame, but this implies bigger losses from defaults. A more conservative approach, where restrictions are gradually lowered as the scorecard is updated and developed to be more predictive, will take more time to produce a relevant sample but should result in lower default losses. Organizations should decide their approach based on their risk tolerance and their desired time to market.

A conservative strategy starts by offering credit to the less risky customers (those with the best scores) to test the scorecard at the lowest risk possible. However, if the scorecard is accurate, it may be that very few customers default. This means that it may take additional waves of loans to collect enough defaults to be able to develop a statistical model.

An intermediate approach is to allocate a considerable percentage of the loans (e.g., 60–70 percent) to low-risk customers and the remainder to medium- or high-risk customers. This approach will allow you to test the scorecard with a wider range of discriminants. And, you will be able to collect default data faster, thereby avoiding the need for several loan cycles to reach a critical mass of defaults.

A risker and quicker tactic is to lend randomly or with few knock-off rules. This approach can accommodate tight timelines by accepting a considerable but contained risk during the pilot stage. Some companies choose not to develop the expert scorecard and simply lend to collect the default data as soon as possible.

Regardless of the route taken, to maximize the chances of the pilot being a success, you will need to implement a rigorous, effective process to capture and maintain all the data gathered.

How many defaults does a reliable model need, or how many loans must be disbursed before there is enough data to develop a statistical scorecard? There's no magic number but, as a rule of thumb, with good quality data, organizations can start developing a reliable model with around 100 defaults. Remember that this is an iterative, continuous process. If a model is initially developed with a small sample, the organization will need to update it relatively quickly.

SECTION **2**

SETTING UP A CREDIT SCORING PROJECT

THIS SECTION DESCRIBES THE infrastructure required for a financial institution to undertake a project to start a credit-scoring service. It covers the human resource requirements, including the implications of developing in-house expertise versus outsourcing. The impact on the existing organization IT infrastructure is examined, as well as the operational procedures and business processes. It concludes with a review of the characteristics of the data produced.

2.1 Characteristics of a Data-Driven Organization

Organizations need to be mature enough to adopt a data-driven decision-making approach. Systems may need to be upgraded, and operational procedures will need to be revised. Management and employees will need to learn new skills, and more important, **they will need to understand the benefits that credit scoring can bring to both the organization and themselves**. It is likely that you will need to hire or contract technical specialists. The magnitude of organizational disruption should not be underestimated nor should the resulting improvements in efficiency and new business opportunities.

The business transformation has six characteristics, or areas of focus (see Figure 3):

- Systems
- Data capture and collection
- IT/analytics resource
- Data analysis

FIGURE 3. **Transforming into a data-driven organization: Six focus areas**

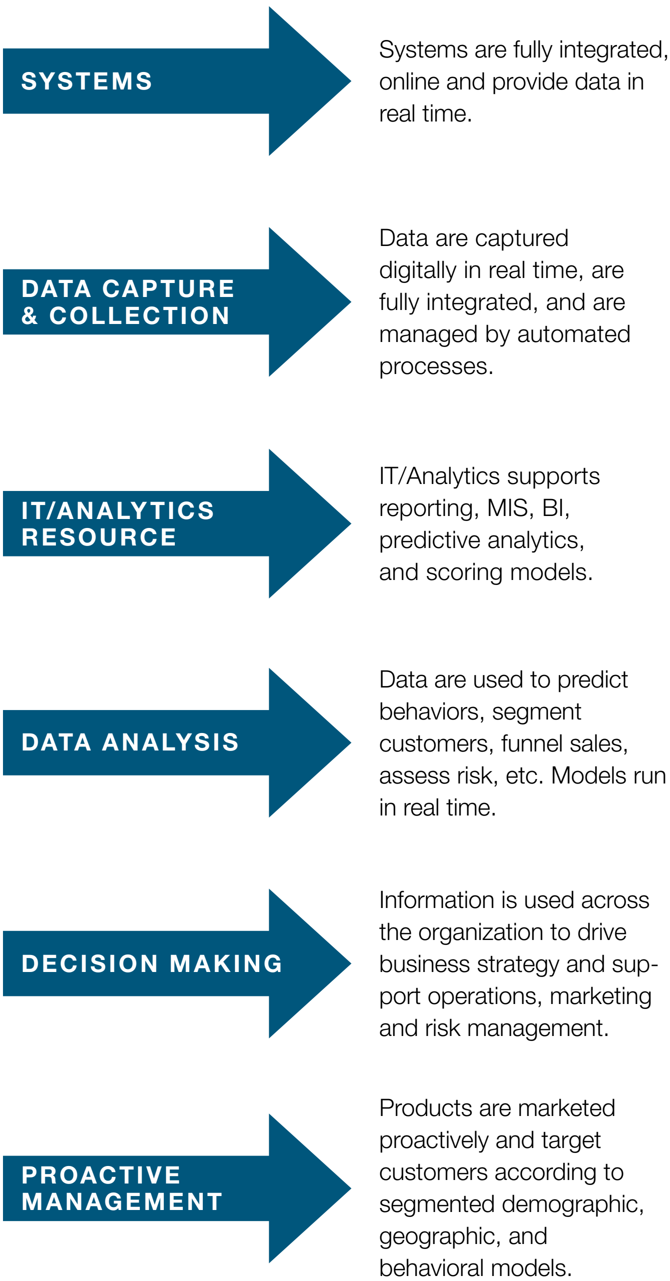
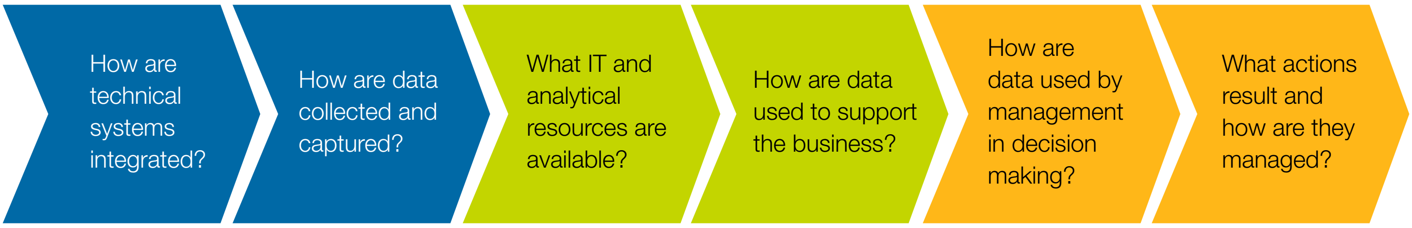


FIGURE 4. **Operational areas affected by using data to provide credit scores**



- Decision making
- Proactive management

The digital transformation will be determined by a set of business processes that specify how each task is done, both internally and when providing products and services to customers. Figure 4 shows a high-level checklist of the types of procedures required that can be linked to a credit scoring project from reporting to prediction.

An organization that wants to become a data-driven business needs to audit its current resources and identify any gaps in infrastructure and skills.

2.2 Key Roles and Responsibilities

Depending on the company's organizational structure, the team will probably already have some of the necessary skills and expertise. However, staff may need to take on some new roles or be trained for the credit-scoring initiative. The following are some recommended functional roles and responsibilities for credit scoring.

Credit analyst. Credit analysts or loan officers are responsible for describing existing processes and practices in detail, providing feedback, and using the experience to validate findings about customers' data. They are usually the closest to the customer, and their perspective is valuable. Loan officers will face big changes in the way they appraise their customers, so they need to trust and support the scoring. They should be involved early on as feedback providers.

Risk manager. Risk managers bring an aggregated portfolio risk perspective and will be changing the underwriting policies through time. Their role is imperative to properly place the data set in a business context before conclusions are drawn. Risk managers are also the source of the product specifications evolution in time. They play a key role in determining the overall risk appetite for the scoring model implementation.

Country manager/COO/CEO. Country managers have two different roles, operating at both a functional and an organizational level. Functionally, they set project objectives, oversee milestones, and validate key decisions and findings, such as good and bad definition, data sample timeframe, data sample default rate, current rejection rate and target, predictors validation, and so forth. They approve investments and, with their teams, define the appropriate risk-volume level of the cutoff point for the scoring model implementation. At an organizational level, they set the tone for operational change and communicate its benefits to each department and functional role. They generally are the sponsor of the project and can greatly influence scoring model adoption by getting involved in the project and expressing their support.

Data analyst. Data analysts are the sources of information, and they function as a liaison between IT data sources and the commercial business. Their main responsibility is to transform data into valuable and manageable information. Before the project and depending on the size and structure of the organization, data analysts, if there are any at all, may work at corporate headquarters or in-country. Data analysts need the statistical modeling, analytics, and risk assessment knowledge needed to develop the scoring engine. If analysts do not have the required knowledge, an external consultant can be hired to teach them the skills they need. To minimize the learning curve, data analysts need to be familiar with statistical modeling tools and analytics software packages as well as risk assessment methodologies. Once the scoring model is implemented, the data analysts will be responsible for the technical aspects of its continuous improvement. They are responsible for collecting the relevant data from the new lending process and ensuring its quality; generating the follow-up management reports; and, after some full cycles of loans, updating the model using the newly generated repayment data. If the organization uses a data-mart solution to manage its data, the analyst will be responsible for leveraging such a solution. This role, along with that of project manager, is instrumental in getting business and technical people to understand each other's perspectives, circumstances, and objectives for a successful implementation.

Database manager. Throughout the project, the database manager will be responsible for navigating the database systems, sharing insight on how data are captured and stored, extracting and formatting raw data, and investigating data anomalies. The manager may also be involved in statistical analysis of the data, and if the project requires IT developments, the manager will be responsible for translating the scoring project needs into IT technical specifications and requirements. IT database managers will be heavily involved at the beginning of the project. They will also be needed during and after project implementation, but not as frequently.

2.3 Acquisition of Credit Scoring Capabilities

A key decision in a credit-scoring project is how to acquire the scoring capabilities needed for development. Organizations typically followed one of three different approaches: create a data analytics department; use an external consultant to lead in-house development; use an outsourced solution. Each has clear advantages and disadvantages that may fit different institutions at different times, depending on circumstances. Moreover, these alternatives can be combined in different sequences as a long-term process to meet the organization's needs. Before reviewing the alternatives, it is important to understand some of the circumstances that may influence an organization's choice:

- Time to market: When does the company intend to launch the scoring model?
- HR availability: How feasible is it to find the specific skills in the marketplace?
- Cost: How much money is the organization willing to invest in the development?
- Data analytics: How critical is it in terms of strategic competitive advantage? Is this a one-off development or a long-term endeavor?
- Flexibility: How important is it for the organization to know the statistical details of the solution and to be able to modify them after implementation?
- Outsourcing options: Is there a trusted provider available to work in the region at the time?

Each of the alternatives can be assessed based on the answers to these questions:

- **Create a data analytics department**. To create a data analytics department, you will need to hire and develop talent. As the department matures, it will be able to develop, update, and improve not only the scoring model, but it will also be able to support the company's other data analytic needs. This approach is best suited for institutions that depend heavily on data analytics. It will take time to develop the department, available qualified staff, and enough demand to leverage the analytics resources at scale.
- **Use an external consultant to lead in-house development.** This approach involves allocating the internal resources described in the previous section to the project and hiring an external consultant that has relevant statistical or scoring expertise. This allows an organization to take its first steps into developing a data analytics team or simply to get the project done pragmatically using mainly internal resources. This approach may be appropriate when there are valuable internal resources, enough time to develop internally, availability of a consultant willing to share scoring knowledge with the organization, and commitment to develop data analytics capabilities within the organization. First, the organization needs to ask: Is it ready to learn from an external consultant and is the external consultant willing share knowledge with the organization?
- **Use an outsourced solution.** In this case, the organization fully outsources the development of the scoring model. It will provide the raw data and agree on all the required definitions, but it is not involved in developing the solution and does not know the details of the model. This approach should be considered when time is scarce, there are no internal resources available to execute the project, or there are no repayment history data. This approach is useful especially when the solution uses nontraditional data or when a quick solution is needed to launch a pilot to generate repayment history data. A company that uses this approach needs to carefully plan the integration of the scoring solution with the lending processes before starting any analytic work. Sound technical solutions have been known to fail in the implementation phase because of poor planning.

Table 2 compares the main characteristics of each alternative.

TABLE 2. **The characteristics of the three strategies for acquiring credit-scoring capabilities**

Strategy	Creation of a department	In-house with consultant	Outsourced solution
Time to market	Long time to acquire and develop talent. Less suitable when there is a tight deadline.	Medium-short time depending on the in-house resources that can be allocated to the project.	Shortest time.
HR availability	Availability of statistical talent in the region could be a limiting factor. In medium or large organization this department could be centralized.	Mainly conducted with trusted internal resources who could potentially own the solution in the future.	Mostly required for project set-up and implementation.
Investment	High	Medium	Medium-Low
Data analytics	Department as a source of competitiveness and profit.	May or may not involve some level of commitment.	Usually in an early stage of data analysis.
Flexibility	Most flexible. Organization knows all the details and can change or update them if needed.	Can be quite flexible if agreed on with consultant. Can be updated or upgraded internally or externally.	Low flexibility. Usually a closed solution. Contract may or may not include maintenance and update of the model.
Offer	Not applicable	If organization intends to learn data analysis, the scoring expert should be willing to share methodology.	It's key to find the right provider paying especial attention to the implementation phase.

2.4 IT Infrastructure

In addition to the core banking system and any loan management solution an organization uses to run its business, two key solutions will support the implementation of a credit scoring model: a data mart and the scoring engine.

DATA MART

A data mart (DM) is a subset of the enterprise-wide data warehouse specifically oriented to store and retrieve the data used each time a customer is scored. At a secondary level, the DM serves two main purposes regarding the scorecard:

- **Scorecard monitoring**: The DM can compare the actual repayment data of a loan with the score assigned to it at the time of application. This makes it possible to monitor how well the scorecard predicts the probability of default on an account and the current rejection rate of the process.
- **Analytics continuity**: The DM allows an organization to manage and continuously improve its scorecard implementation. The predictive power of a model tends to weaken in time because of changes to the business context, customers, or even the lending processes. Between 12 and 36 months after implementation, the DM can provide all the data necessary to update the algorithms or to redevelop the model to compensate for changes.

The DM manages two types of data: application and financial performance.

Application data include the sociodemographic and behavioral information of the customer at the time of application. Typical application data mirrored in the DM include the following:

- Customer profiling information captured at the time of application review
- Loan product information required (date, amount, duration, interest rate, etc.)
- Scoring variables used in the scorecard
- Credit bureau data, if available
- History of a customer at the time of application (customer tenure, previous loan cycles, savings products, insurance, etc.)

Financial performance data are preselected information queried from the core banking system or loan management system to help understand repayment behavior and correlate it with application data. These snapshots should be taken at regular intervals in time matching the repayment frequency of the loan (weekly, monthly, etc.)

Performance data include core data and other elements. The following core data are expected in the DM:

- Delinquency counters (e.g., number of days overdue)
- Amount of repayment made during the period
- Loan status: active, closed, written off
- Write-off amount

The additional data elements are related to other products or data sources, including, for example:

- Average savings balance
- Maximum savings balance
- Minimum savings balance
- Number of deposits
- Average deposit size
- Number of withdrawals
- Average withdrawal size
- Mobile money use data
- Psychometric data

Ideally, the DM should cover five years of history and no less than three. Data that become obsolete should be archived, never deleted, because they may be useful in future as, yet unknown, algorithms.

SCORING ENGINE

Independently of the statistical tool used to build the scoring model, the algorithm used to calculate the score should be parameterized into the organization's IT systems to be available to the front office when customers apply for a loan. If an immediate decision on credit is needed on scorecards developed for new customers, the score computation should be available online. The scoring engine should always be owned by the lending institution regardless of the scorecard development strategy.

2.5 Business Processes

A credit scoring model provides the expected probability of default of an applicant for a loan. For the scorecard to become an actionable business decision tool, it must be combined with other underwriting criteria and decision rules through either manual or automated processes. This means that **the underwriting criteria and decision rules are as important as the model in predicting the performance of a loan**.

The credit process comprises four stages:

1. Acquisition or prospection
2. Origination or underwriting
3. Servicing or customer management
4. Collections

The implementation of an application scoring model is an important part in the underwriting stage, but the performance of a loan depends on the whole process. Therefore, when implementing a scoring model, review all the stages in the credit process and make sure they align with the business strategy. There are several ways in which the different stages could affect loan performance regardless of the scoring model. For example, if an organization implements an excellent scoring model but fails to identify and attract the adequate profile of prospective customers (or if the scorecard and cut-off values are not calibrated to the pool of customers), the scoring model will not behave as expected. It may reject most of the prospects. Similarly, if a model is developed using a biased sample different from the customers it will be assessing, the model will not be able to predict their repayment behavior accurately.

Even though a scoring model is empirically derived people will use it, and they will need to trust it. Loan officers need to be involved to maximize their buy-in, particularly if the process will continue to involve loan officers for all or some of the customers. In addition, they must be reassured that the new service will not undermine their role in the organization.[2]

A successful credit-scoring project depends not only on the model's performance, but also on operational measures, including the following:

- Proper integration of model in the whole underwriting process
- Training, early involvement, and empowerment of loan officers
- C-level sponsors of an objective data-driven decision-making culture

2 For more detail, see IFC's Field Note 8: Changing Change management: Adapting internal and external culture in times of digital transformation (https://www.ifc.org/wps/wcm/connect/93567f5c-6eb5-4e23-bc13-ee0c55f3eabc/IFC+MCF+Field+Note+8_DFS+Change+Management+MCF.pdf?MOD=AJPERES)

- Continuous improvement of the model: monitoring its performance and update or redevelop it when necessary

PROJECT GOVERNANCE

The general ongoing organization of the project is also a success factor. Project governance must be in place. It needs to ensure that progress is being made and roadblocks and other issues are identified early on. Recommended infrastructure should include the following:

- A dedicated team that has clear functions and responsibilities
- A detailed workplan with tasks, key milestones, and decision-making meetings
- A weekly call to share progress and tackle operational issues
- A clear owner of the outcome of the project who is responsible for its implementation and ongoing improvement

2.6 Data Sources, Coverage, and Quality

In financial institutions, data can be collected in many ways, so it is essential to know where each piece of data comes from and when and how it is captured. Some data fields are likely to be more reliable and consistently collected than others. Table 3 ranks some types of data commonly collected listed in order of expected reliability for predictive analytics.

In addition to the reliability, there are three other important aspects of data to consider when developing a scorecard: data sources, data coverage, and data quality.

Data Sources. Organizations tend to store data in several different systems, tables, databases, etc. Therefore, you will need help from an IT analyst to know what data are available and where they are stored.

Data Coverage. How consistently the organization collects and stores data across its customer base is critical because the model will be using variables that should be available not only during the development phase, but also on a recurring basis for scoring purposes.

Data Quality. The accuracy of data sets and the ability to use them to analyze and create actionable insights for other purposes is fundamental. Key elements of high-quality data are people, processes, and technology. Data quality can be assessed through six dimensions:

- **Timeliness**. The degree to which data from a specific point in time represents the current conditions.
- **Completeness**. The proportion of stored data that is complete vs missing for customers in the system.
- **Uniqueness**. Each data point is either recorded once or consistent across sources. (If the same information is stored in more than one place, it is important to agree on a single source of truth as the default.)

TABLE 3. **Reliability of the different types of data**

Type of Data	Reliability	
Transactional	HIGH	If stored consistently, transactional data (e.g., account deposits and withdrawals, loan payments, bill payments) are usually very reliable. These data provide a historic and objective record of a customer's actual behavior or economic activity.
Documentary	HIGH	Identity and basic demographic data are often taken from (or verified by) official documents (e.g., national ID cards).
Collected from devices	HIGH	Device data can be just as reliable as financial transactional data.
Psychometric	ABOVE AVERAGE	Despite being self-reported, there are different techniques in psychometric questionnaires to validate candidates' answers, such as using similar questions in different sections of the test. However, data reliability depends on the quality of the tests and how they are administered.
Collected by Staff	AVERAGE	These data may be influenced by the judgment, work style, or experience of the person collecting it.
Self-reported	BELOW AVERAGE	Data reported by a customer (such as on an application or survey) can be less accurate because people have different styles of communicating information. Customers may tailor responses to maximize their chances of being approved.

- **Validity**. Data are valid if they conform to the syntax (format, type, range) of their definition.
- **Consistency**. The absence of difference, when comparing two or more representations of a thing against a definition (e.g., Revenue in $ has to be 1000 x Revenue in thousands of $, number of 30-day delinquencies has to be equal or greater than number of 90-day delinquencies).
- **Accuracy**. The degree to which data correctly describe the event or characteristic presented (i.e., is the information correct?).

2.7 Sample Structure

The data set used to build a scoring model needs to be structured in a specific way for statistical software to work properly.

- The data set needs to have a "wide format," which means that each row represents an individual customer or individual loan and each column represents a variable. If the database is created at the loan level, the customer data associated to that loan needs to reflect information as of the time of application to that loan, and not the latest loan (see Section 3.2, "Extract Historical Data").
- If a customer has had several loans with the organization a "loan cycle" variable should be created to take account of this and should be tested. In some cases, repeat customers can have a very different risk profile, or there may be much more available information, so that that two different models can be created for new and repeat customers.

2.8 Types of Data Variables

Different types of variables should be considered when creating a data set for analysis.

Raw variables. These are extracted from the systems "as is" without any transformation.[3] Most variables are raw variables (e.g., customer ID, gender, date of birth, educational level, years as customer, etc.). Raw variables need to be scored in one of two ways, depending on the nature of the information they contain:

- **Categorical.** A variable that can take on one of a limited, and usually fixed number, of possible values; it assigns individuals to a group or nominal category based on a qualitative property (e.g., gender, nationality, type of ID, or level of education).
- **Numerical**. A variable that has a numeric value that has a numeric meaning (e.g., distances, income, or age).

Derived variables. Derived variables are created by transforming a raw variable. The main reason for doing this is usually mathematical or statistical. Among the derived variables there are different types of transformations:

- **Target or dependent variable**. This is the most important variable of the data set because the model will try to predict this variable based on the independent variables. In a scoring model, the different values for this variable are good (0) or bad (1).
- **Tags or flags.** These are variables that could be transformed from the data source at the time of extraction. The IT analyst should be consulted and involved in the extraction to learn which flags or tags are available in the systems. For example, when a customer is enrolled strictly as a guarantor, a "Guarantor" flag could be added that assigns 1 to these customers and a 0 to the rest.

3 In data analysis, transformation is the replacement of a variable by a function of that variable to change the shape of its distribution or relationship.

- **Ratios.** Sometimes it is statistically or mathematically easier or better to work with ratios of raw variables (e.g., "number of loans taken/years as customer" to annualize loan uptake).
- **Rankings.** Starting from a raw variable, each customer can be assigned a measure of position from top to bottom in regard to the raw variable.
- **Feature scaling.** In some cases, when working with variables in different units or with very different value ranges, it may be convenient to standardize the variables to make them comparable. Also, when using discriminant analysis, which measures distances between data points to identify important individual features and good and bad loans in the data, scaling or standardizing values prevent large values from dominating the analysis.
- **Categorizing.** When raw data fall into certain well-defined categories, it can be useful to categorize the variable. For example, if an organization is lending to small and medium enterprises (SMEs) and data on the number of employees of each SME are available, that data can be used to create a variable called "size of business" that defines them as "small," "medium," or "large."
- **Dummy variables**. Sometimes categorical variables don't work well because they imply an order or ranking. Replacing small, medium, and large with 1, 2 and, 3 correctly implies that 3 is more different from 1 than it is from 2. For categories that are not ranked (e.g., type of industry), dummy variables can be more accurate. Dummy variables take the value 0 or 1 to reflect the presence or absence of the categorical effect. For example, instead of naming business's industry, one could create a set of variables, one for each industry and for each customer; assign 1 to those in the industry and 0 to all the rest of the industry variables. Statistical software or simple formulas in Excel can create dummy variables—it does not have to be a manual process.

Later in this Guide, other transformations are introduced that could be applied to raw variables to improve the predictive power of the model.

SECTION 3

PREPARING THE PROJECT DATA SET

THIS SECTION GOES OVER A STEP-by-step process for getting the data set ready to develop an automated credit-scoring model. First, historical data need to be available and appropriate resources need to be in place as described earlier in this Guide.

3.1 Assess the Current Lending Process

A benchmark details the current lending process and assesses its strengths and weaknesses. To develop a benchmark, start by reviewing the lending process from acquisition to collection, from the perspective of the key players. Do the following:

- Map the end-to-end process. Interview the chief resource officer, loan officers, and branch managers. Review the application documentation.
- Look for inefficiencies in how resources are used (e.g., officer time, amount and relevance of data collected, data cross-checks/validations, etc.).
- Analyze how decisions are made and what information was used for the decision. Review the time and resources needed to capture the information. Assess whether relevant cases are being reviewed by the right structure (credit committees, branch manager, loan officers, etc.) and identify any unnecessary bureaucracy.
- Focus on sources of data error (e.g., manual input, lack of cross-checks, unverifiable information [usually input incorrectly], too many fields or too little space to write in the application form, applications storage/availability, etc.).
- Review the lending policies for each segment. Track their changes over time. Plot the business strategies in a timeline. Look for growth periods; these could show a less strict lending policy and changes in repayment behavior.

3.2 Extract Historical Data

You need to understand how data are acquired, stored, and updated throughout the lending process. The objective is to separate reliable information from less reliable sources, generally by how data are captured, stored, or updated (see Section 2.5, "Business Processes").

- Map all systems and sources of data to a data architecture to understand what information is available and from what source and to get a sense of what to include in a first historic data set.
- Identify which variables are relevant for a credit scoring model and assess the feasibility of extracting them. In a loan process, there are different types of data involved; the most common are application/profile data and historic financial performance data. When extracting the data, the date for each of these variables should be different. Previous loan repayment history data should be as recent as possible, while application variables should be extracted on or just before the date of application itself—this will be different for each customer. The model predicts repayment behavior of future customers based on the profile attributes of those who have already repaid a loan, so those attributes must describe the current customer's profile at the time of application.

Note: It important that no data generated after the loan was disbursed are included in connection with that loan. For example, if you are using the credit bureau score as a predictor, it must be the score available before the loan was disbursed. Mixing newer variables in the model would make it look artificially predictive and would not hold after implementation.

The IT analyst who will run the queries should be included in the extraction discussions to weigh in on feasibility. A commercial representative also should be present to provide business context.

3.3 Produce Counts of Main Performance Variables

After historic loan performance data are extracted, the loan performance variables evolution over time must be analyzed. This has a three-fold objective:

- Gain a detailed understanding of the loan performance to ensure the data make sense
- Identify the best window of time to capture the application data sample on which to base the model development. (Covered in in the next step)
- Quickly check that there are enough defaulted loans to create a model that will have reliable predictive power.

Which **loan performance variables** to use in developing the model depends on the variables each organization has. In established institutions, most of the variables will be centralized in a core banking system. In others, the loan management system may be handled separately. Regardless of where the performance variables are stored, the most common variables include the following:

- Number of accepted and rejected applications
- Number of defaulted loans
- Number of late payments
- Total number of days of late payments
- Number of consecutive overdue payments
- Number of loans that are 30, 60, and 90 days in arrears
- Various delinquency flags on previous loans

3.4 Define "Bad"

Organizations need to define the specific set of rules that classifies a loan as "bad." The definition should be easy to interpret and allow for the performance of the accounts to be tracked over time. Several tools can be used to do this. Which tools you select should be based on the data available, experience from loan officers and risk managers, external bureau definition/guidelines, and correlation between early signs and eventual default, among others.

Financial institutions commonly use PAR90. However, many organizations also use PAR60 or even PAR30, which gives them an earlier trigger for collections. The Basel Committee on Banking Supervision defines default essentially as a delinquency stage of 90 days or more.[4] Other definitions include the number of consecutive overdue payments, normally two or three. Some organizations distinguish between different levels of delinquency related to pricing and profitability. It is important to decide on the best definition for the project because it directly affects who the model will propose to accept or reject.

The definition also affects the sample size. The more restrictive the definition (e.g., 30 days in arrears), the more delinquent accounts there will be. By contrast, the more lenient the definition (e.g., 90 days and two consecutive overdue payments), the lower the count of bads there will be in the sample. The model needs to have statistical significance. As a rule of thumb, a scoring model with a good predictive power is based on a development sample that has good profile and performance data for a significant number of bads. At least 1500 bads should be included in the sample, if available. The optimum sample size depends on each organization's unique situation, but generally, you would want at least 500 bad customers. However, models with as few as 50–100 defaults can be developed if there is no other option (e.g., for a new product), but the model should be revised regularly as more information becomes available.[5]

Once the definition of "bad" is established, all the loan disbursed accounts in the data set should be classified in mutually exclusive categories related to the performance of the loan.

Bads. In hindsight, these are the loans that an organization would not have disbursed. Each organization should define the criteria to classify a loan as bad based on its risk appetite. Another aspect to consider when defining a bad loan is the profitability of the customer over time. A conservative approach usually means significant loss of profit by rejecting profitable customers.

Goods. These are the loans that an organization would be happy to repeat. Usually, these customers haven't had significant delays in payments and exhibit good repayment behavior.

4 https://www.bis.org/bcbs/qis/qis3qa_f.htm

5 Later in the Guide, how to measure the predictive power of the model is discussed.

3.5 Define the Sampling Window

There are two reasons to choose an appropriate sampling window from which to extract data:

- **Changes over time.** In the lending process, many things change over time, such as aggregated repayment behavior of customers, assessment and lending policies and procedures, commercial objectives, market share, competition, external economic factors, and so forth.
- **Data management.** The profile data for each customer must be collected at the time of or before the application, which requires a thorough collection effort from the different data sources. This could be relatively easy if all the application data are in one database. However, in many cases, application information is stored in different places. For example, profile data may be stored in the core banking system, while the customer's financial information, officer's appraisal, and committee decision may be stored in several separate databases. Therefore, instead of using all the accounts in the data set of historic data, you should select a smaller, highly relevant and more manageable sample of accounts from which to retrieve the application data.

There are several important considerations when selecting the time frame for the sample:

- The more recent the data, the better they will reflect future repayment behavior.
- Loans that are not fully repaid must have passed their maturity date by at least X days, where X is your definition of default. This gives the last installment the time to be X days in arrears.
- There must be enough bad counts. In a healthy portfolio, bad counts are scarce, and for an account to be valid, it must have application and performance data. There is a trade-off between having more recent cases and having more cases. The lower the number of bads, the further in time you will have to go to ensure a large enough sample.

To ensure an out-of-sample prediction, the sample should represent the applicants of whom the model is intended to predict repayment behavior. There could be different prerequisites applied to the applicants before their applications are accepted and analyzed by the model.

3.6 Data Collection

Now that you have categorized the accounts and chosen a time window, the next step is to define which data fields to extract for the sample, using the mapped systems and data sources you created earlier. You may want to create a data dictionary to make this process more efficient**.** A data dictionary summarizes all the specific data variables (application and performance), their definitions, and feasibility of extraction.

You should pull only standardized fields and avoid open text, qualitative, and subjective judgmental fields. Keep in mind that application information, which is not formally checked, could be gathered less rigorously and thus be less reliable. So, when selecting the variables, it is important to know the details of the processes from which the data have been generated. (See sections 2.5 and 2.6.)

After you have defined which variables to extract, ensure that:

- Performance data are extracted on the same date for all accounts.
- Profile and sociodemographic data are extracted at the time of each application.

Depending on the different sources of data, this could be a complex process and it always requires detailed verification steps.

3.7 Data Accuracy Check and Transformations

Once the data set has been created with the raw variables, each of the variables must be thoroughly analyzed to assess its accuracy and to spot issues in the data.

Doing a descriptive statistical analysis is a quick way to spot issues with the data. The analysis quantitatively describes features of a data set in terms of "central tendency and variability or dispersion." Measures of the central tendency include mean, median, and mode. Measures of variability include variance and minimum and maximum values and quantiles. Two common issues in data sets that can seriously affect analysis are outliers and missing values. Table 4 presents elements to check for in a new data set.

TABLE 4. **Common descriptive statistics**

Name	Helps you to know
Type	What type of data to expect (numeric or text)
Number of unique values	Variations in the population
Number of missing values	How well the data have been collected
Mean, median, and mode	What the average client looks like
Histograms	How the population is distributed
5 highest and lowest values	What the outliers are or which values are likely to be errors

OUTLIERS

Outliers are extreme values that don't reflect the reality in the data (Vidal, Caire, and Barbon 2019). They are often the result of errant data entry or capture and can seriously distort the results of many types of statistical analysis.

It is usually good practice to remove extreme outliers from a data set. If you have a lot of data and relatively few outliers, the easiest way to do this is to delete the rows with outliers from the data set. If data are limited, you could replace the outlier with the mean (average) value or with a capped value (e.g., all values over 90 are recoded to be 90).[6]

A simple way to systematically identify and handle outliers in a large data set is to use the Interquartile Range Rule (IQR). IQR sets boundaries for the highest and lowest values expected in a data set; values that fall outside of this are considered outliers. See Box 1.

BOX 1. **Interquartile range rule example**[a]

To identify outliers through the IQR rule:

- Sort the data column from lowest to highest value.
- Identify the first quartile threshold, *Q1*, as the value one-quarter of the way through the sorted data column (i.e., 25 percent of values are less than this number and 75 percent of the values are greater than it).
- Identify the third quartile threshold, *Q3*, as the value three-quarters through the sorted data column.
- Calculate the interquartile range (IQR), which is IQR = *Q3* – *Q1*.
- Multiply the IQR by 1.5 (as a rule of thumb or experiment with other values).
- Add IQR x 1.5 to *Q1*. Any value below this is an outlier.
- Add IQR x 1.5 to *Q3*. Any value above this is an outlier.
- Numerical examples of outliers are given in the CGAP Data-driven segmentation in financial inclusion Guide.

a Adapted from https://www.thoughtco.com/what-is-the-interquartile-range-rule-3126244.

MISSING DATA

Data sets usually have missing values (Vidal, Caire, and Barbon 2019). When a significant share of data is missing for any given column in a data set, it is important to understand why. Some possible reasons for missing data include:

- The field is not applicable to all clients.
- The field is for optional information that clients are not required to provide.
- The information is not available because the client was not asked for it or the client was asked and did not have or did not provide the information.
- The information for the field was not collected before or after a specific date.

Depending on why information is missing, some data columns may not be appropriate for analysis or should be processed with techniques for working with missing variables. Table 5 presents some of these methods.[7]

6 See Advanced Handling of Missing Data in the Quick R online reference library at https://www.statmethods.net/input/missingdata.html

7 For more, see https://www.r-bloggers.com/missing-data-in-r/

TABLE 5. **Working with missing data**

Strategy	Use	Pros	Cons
Remove rows	• Data are plentiful • Relatively few missing records • Complete data will always be required in future	Uses complete records No assumptions	Loss of data
Replace with Unique Value (such as "-1" or "999999999")	• High percentage of missing values across different columns • Missing values may be possible for same columns in future	Preserves data Can accommodate missing data in future	May introduce bias if past reasons for missing data do not persist
Replace with an Average Value	• Data are limited • Relatively few missing values • Missing values will not be possible in the future	Preserves data	Assumes past missing cases were no different than average cases
Replace with Predicted Value	• Data are limited • Relatively few missing values	Preserves data May be more realistic/accurate than using average values	Increases complexity of model

DATA TRANSFORMATIONS

The statistical methods used for scoring model development are sensitive to the way the data are input. Data transformation generates a different representation of the data that improves the predictive power of a model. Data transformation is typically used for four objectives: linearization, standardization, conversion, and transformation.

Linearization. The statistical methods used in scoring models are linear procedures, which means that they fit straight lines to the input data. However, the relationship between input variables and loan default might not always be linear. So, if one tries to fit a linear or logistic regression to these variables, the model will provide the best fitting straight line, which may not be a good representation of the relationship.

Standardization. When the range of one predictor is 1,000 to 10,000 and of another is 0.01 to 1, then the parameter coefficients, the model weights, will be very different even when the two variables contribute equally to the model. It is good practice to transform numeric variables so that they are on the same scale. For example, a common standardization technique for numeric variables is to subtract the mean and then divide by the standard deviation.

Conversion. Some model construction techniques require data to be in a numerical format (numerical or ordinal). Categorical data need be converted to numeric to be used for model development.

Transformation. Variables may be replaced by a function to address a nonlinear relationship. For example, being at the current job one year or five years might be significantly different in terms of income stability, while being at a job 20 or 24 years might not be relevant since both are equally stable. In this case you might chose to replace any tenure above 10 years with 10, to effectively capture that the length of tenure is only relevant for short/medium tenures.

SECTION **4**

SCORING MODEL DEVELOPMENT

ONCE THE DATA ARE READY FOR score development, a random part of the total sample available (typically 30 percent) should be set aside as the testing sample. The testing sample will be used to assess the predictive power of the model after it is developed.

The following are the fundamental assumptions in predictive modeling:

- Predictive models assume the future will be like the past: it is assumed that relationships found in past data will persist into the future. However…
- Past relationships between products, client characteristics, and behaviors can be expected to last only for as long as other things stay the same—if policies, processes, and/or targeted client segments change, relationships observed in the past may no longer be valid.
- Associations in the data may not reveal causality. For example, microfinance data often show that loans with two or more guarantors experience delinquency more than loans with only one guarantor. Intuitively, we might expect the opposite—that two or more guarantors would reduce risk of nonrepayment. But lenders typically ask for additional guarantors only when they already sense a borrower is "riskier" than average. When the delinquency pattern in the data is observed, it simply confirms the initial belief that those borrowers were riskier. The delinquency is not due to the loans having two guarantors.

For a deeper understanding of the statistical detail, refer to the "Predictive Analytics" appendix.

4.1 Classification Methodologies

Most credit scoring models are developed using proven classification models or methodologies that use specific data field "predictors" to estimate the probability of default. Two common techniques are binary logistic regression and classification or decision trees.

BINARY LOGISTIC REGRESSION

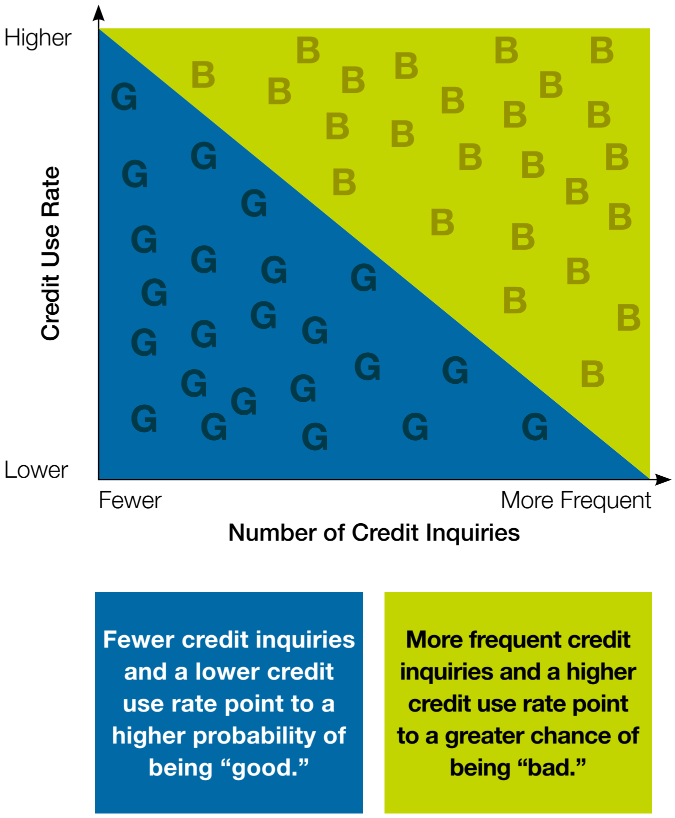
Binary logistic regression is the most common method. It is a classification algorithm that is used to estimate the probability of a binary response based on one or more predictor variables or features. More formally, a logistic or logit model is one where the log-odds of the probability of an event is a linear combination of independent predictor variables. A binary logistic model allows the calculation that the presence of a risk factor increases the odds of a given outcome by a specific factor.

The example in Figure 5 plots borrowers against two variables and reveals a strong discriminatory predictor. Those with high credit inquiries and high credit use are more likely to be bad. The two variables correlate in data sets with a higher probability of default.

When building a multifactor logistic regression model for credit scoring, the following steps are likely to improve its transparency and understanding:

1. Screen each piece of data that may have a meaningful relationship to the target variable or default. For data with such relationships, build single-variable models that can later be entered into a multivariable model.
2. Check for correlation of candidate variables. If one or more variables are highly correlated (i.e., correlation coefficient of > 0.80), test the model, with one or the other for the multivariable model. If the model with both variables

FIGURE 5. **Example of binary regression for credit scoring against two variables**



performs in a similar way as the one with the single variable, keep only the variable that performs better by itself.

3. Build the multivariable model one variable at a time, observing how all the model estimates change as you add each additional variable.

CLASSIFICATION OR DECISION TREES

Classification trees help to better identify groups, discover relationships between them, and predict future events based on target variables. Trees can help you understand the relationship between variables and how they classify goods and bads.

They work by repeatedly splitting data into contrasting groups to identify groups with a larger proportion of the target variable. They test all the possible combinations among input variables to identify how the best variables combine to explain the outcome and display those relationships graphically in nodes. Trees show the nodes or branches in order of their discriminant power. They start at the top with the "root" or the outcome variable and then open repeated branch levels until the next branch does not explain anything further. The nodes at the bottom of the branch are called final nodes, and they include the risk clusters in which the model ranks the customer base. As opposed to logistic regression, trees do not require that the data have a normal distribution. Because tree-based algorithms can handle data of any type, they can be a relatively easy way to put in all the data at once and discover what factors and combinations of factors have been associated with default. Figure 6 provides an example of a classification tree.

In this tree, the root or the outcome the model is trying to predict is "Credit Rating." The first node is given by "Income Level," which based on its values, opens in three branches with different proliferations:

- **Low.** For customers at the low income level, the tree did not find any relationship with any input variables. This could be because low level concentrates a very high percentage (~82) of bad customers in and of itself, without other conditions.
- **Low, medium**. For customers in this category, the model found that the "number of credit cards" that a customer has determines significantly whether those customers would turn bad or not. By comparing the variable "numbers of credit cards" to the likelihood of a customer turning bad, the model identified a significant difference between having fewer than five cards and five or more. For those with fewer than five, there was no further relationship to explain the result, but among those with five credit cards or more, the younger customers (younger than 28 years old) had a higher propensity to turn bad.
- **Medium**. For these customers, the only additional variable that predicted outcome was the number of credit cards.

This demonstrates how the model identifies the relationship between the most predictive input variables and uses them to estimate the likelihood of a bad. In this example, three risk drivers were identified:

- "Income level" (the higher the income level, the lower the risk)
- "Number of credit cards" (the more cards, the higher the risk)
- "Age" (the older the customer, the lower the risk)

The classification concludes by placing customers into different risk groups (nodes) in descending order of concentration of bad. The output of a scoring model for this example ranks the risk groups as shown in Table 6.

Classification trees do have some limitations:

- Special expertise may be needed to get results that best fit tactical/strategic needs.
- Variables and cut points are sensitive to the math used to grow trees, and there are many methods available. If

FIGURE 6. **Example of using a classification tree to predict credit rating**

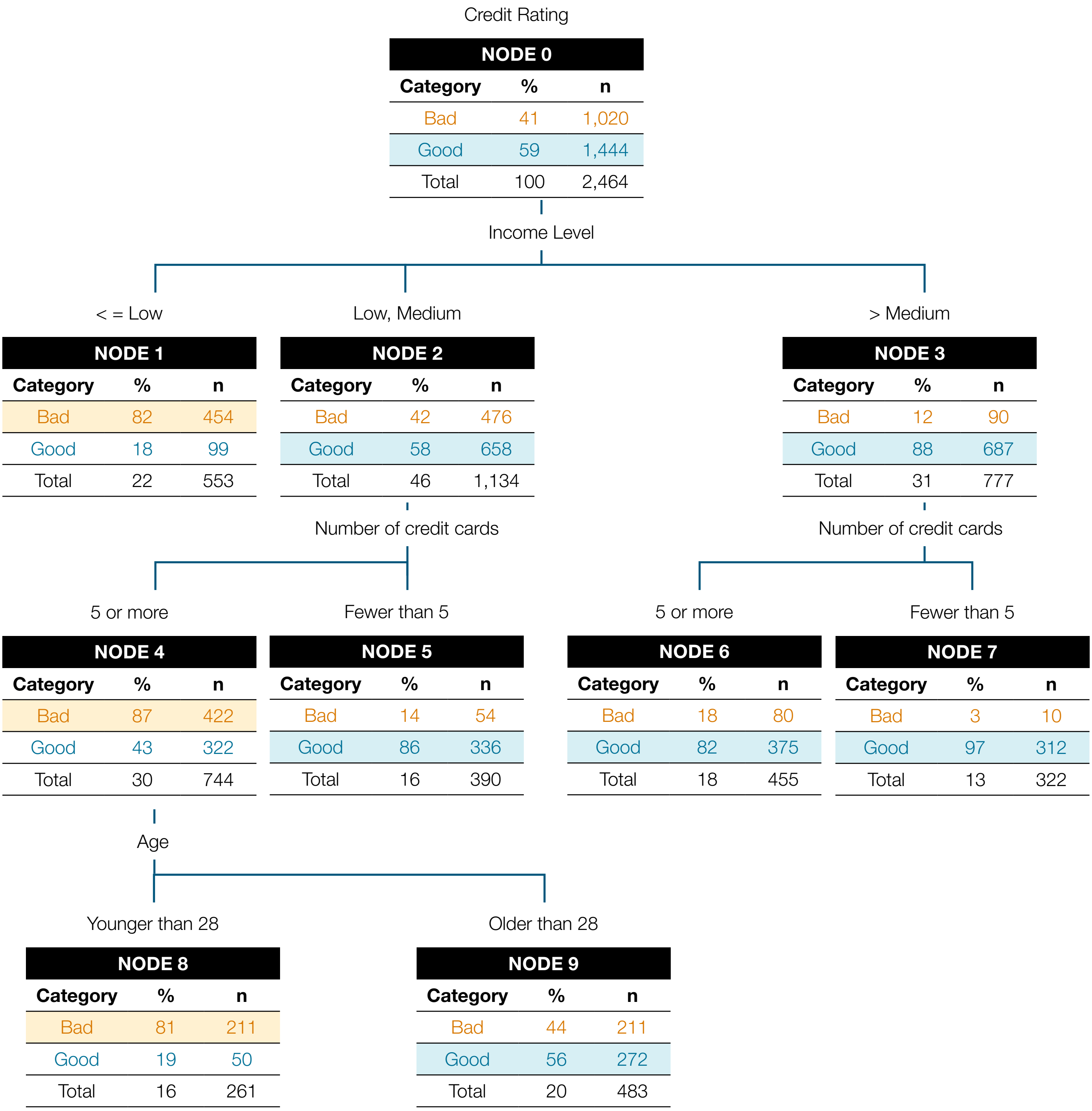


TABLE 6. **Use of nodes to classify risk profiles**

Terminal Node Identifier	Bad	Good	Total	Bad (%)	% of customers	cumulative % of customers
Node 1	454	99	553	82	22	22
Node 8	211	50	261	81	11	33
Node 9	211	272	483	44	20	53
Node 6	80	375	455	18	18	71
Node 5	54	336	390	14	16	87
Node 7	10	312	322	3	13	100

expertise is not available, try the "default" settings meant to be most applicable to the general case, and test and compare several competing methods.

- No relationships will be revealed for certain data points that do not meet the "significance" criteria for making splits in the tree, so other methods (such as single-variable regression or cross-tables) must be tried to investigate other variables of interest.
- The use of artificial intelligence techniques, like neural networks, are increasing as the amounts of available data increases and the cost of computational power decreases. However, these models are expensive to implement and maintain, they are opaque and hard to interpret, making it challenging to validate them against common business sense, and they prone to overfitting.[8]

4.2 Univariate Analysis

Once the sample represents the population that the model will analyze, the next step is to conduct a unidimensional analysis of the variables to understand how good they are at predicting future repayment performance. This method consists of creating histograms for both categorical and numerical variables and then, given the default rates (bads/accepted) among the different groups within a variable, estimating the predictive power of that variable alone. To estimate the predictive power of the variable, analyze its correlation to the "default rate."

Variables can be numeric, like income, or categorical, like marital status or type of employment. Ranges of a numeric variable or categories in a categorical variable can be combined to simplify and increase stability. The following are some important considerations:

- Size of each group is meaningful as a percentage of total sample (5 percent is a recommended minimum).
- Categories being combined show similar risk behavior.
- Risk trend of the group within a variable is monotonic and/or makes sense from a business or market standpoint. A monotonic relationship does one of the following: (1) as the value of one variable increases, so does the value of the other variable or (2) as the value of one variable increases, the other variable value decreases.

This is called coarse classing, and it is slightly different than the approach between categorical and numerical variables.

4.3 Categorical Variables

This approach leverages the organization's knowledge of the market. In this approach, the different values for each variable are consolidated into groups that exhibit similar risk behavior among themselves and behavior that is as different as possible to that of other groups. Each group or category comprises the values that show similar repayment behavior.

The example in Table 7 shows the grouping of the variable "residential status" and the "default" of each category when there are very few cases in a group. In this example, repayment follows whether the customer owns a house or not. Customers who rented and those living at the company's facilities showed similar repayment behavior, and therefore, they were grouped into one category. (Note that this is regardless of the default rate [39.1 percent], because it is not statically stable or reliable given the small group size.)

4.4 Numerical Variables

For numerical variables, the grouping is done by selecting ranges of the numerical values that comply with the categorical variables. To extend the previous example, as well as the three criteria mentioned in Section 4.2, there is the additional consideration that the percentage of bads should be monotonic. As a rule of thumb, up to 10 classifications (bins) can be created but typically much fewer are needed.

Table 8 shows an example of the sociodemographic numerical variable "customer age." Some observations on this example include:

- The size of the categories is comparable (7–20 percent).
- The trend of the default is monotonic, and it follows the logic that the older a customer, the better the repayment behavior.

4.5 Develop a Known Scorecard

If there are too many prospective variables, a quick way to shortlist the most important ones is to calculate the correlation between each variable and default. The Excel function "CORREL" can be used to estimate the correlation coefficients and then compare their magnitude to preselect predictors. Repeating this procedure for all the input variables results in shortlisting those with most

8 The term "overfitting" refers to creating a model that works well on the specific cases it was developed on but not on others, making it ineffective once implemented.

TABLE 7. **Example of coarse classing where one group is small**

Before grouping

Residential_status_ungrouped					
Category	**Number of Goods**	**Number of Bads**	**Total N**	**% of Bad/Total**	**% Total N**
Own	301	98	399	24.6	12
Rent	1,413	873	2,286	38.2	71
Company	14	9	23	39.1	1
Family	343	149	492	30.3	15
Total	**2,071**	**1,129**	**3,200**	**35.3**	**100**

After grouping

Residential_status_grouped					
Category	**Number of Goods**	**Number of Bads**	**Total N**	**% of Bad/Total**	**% Total N**
Own	301	98	399	24.6	12
Rent + Company	1,427	882	2,309	38.2	72
Family	343	149	492	30.3	15
Total	**2,071**	**1,129**	**3,200**	**35.3**	**100**

TABLE 8. **Addition of numerical variables to the categorical classification**

Customer_age					
Category	**Number of Goods**	**Number of Bads**	**Total N**	**% of Bad/Total**	**% Total N**
From >17 to <=28	132	107	239	44.8	7
From >28 to <=32	282	202	484	41.7	15
From >32 to <=36	401	249	650	38.3	20
From >36 to <=39	297	172	469	36.7	15
From >39 to <=44	428	201	629	32.0	20
From >44 to <=49	274	113	387	29.2	12
From >49 to <=66	257	85	342	24.9	11
Total	**2,071**	**1,129**	**3,200**	**35.3**	**100**

univariate predictive power. Table 9 provides guidance on how to interpret correlation coefficients.

Note that the weak variable might still add predictive power to a model if it is not correlated to the rest of the variables in the model, so it should not be automatically discarded.

Next, run a logistic regression to find the coefficients or weights of each predictor. This should be done either with statistical software or with third-party excel add-ins.

The output of the logistic regression includes two important figures: the p-value and the coefficients.

TABLE 9. **Interpreting correlation coefficients**

Coefficient of correlation	
Value of r	**Strength of relationship**
-1.0 to -0.5 or 1.0 to 0.5	Strong
-0.5 to -0.3 or 0.3 to 0.5	Moderate
-0.3 to -0.1 or 0.1 to 0.3	Weak
-0.1 to 0.1	None or very weak

TABLE 10. **Example of a scaled table of variables**

Score	Goods	Bads	Total	Bad Rate %	% Goods	% Bads	Accum % Goods	Accum % Bads
1–100	100	150	250	60	1	19	1	19
101–200	200	135	335	40	2	17	3	37
201–300	350	100	450	22	3	13	5	50
301–400	750	95	845	11	6	12	12	62
401–500	900	82	982	8	8	11	19	73
501–600	1,250	70	1,320	5	10	9	30	82
601–700	1,350	55	1,405	4	11	7	41	89
701–800	1,750	35	1,785	2	15	5	56	94
801–900	2,500	30	2,530	1	21	4	77	97
901–999	2,780	20	2,800	1	23	3	100	100
Total	**11,930**	**772**	**12,702**	**6**	**100**	**100**		

- The p-value validates the significance of the relationship between a predictor and default. You need to analyze p-values to check if the correlation among predictors may have affected them from the univariate analysis. So, for each predictor, analyze the magnitude of the p-value. Basically, a p-value of X% means that there is X% probability that the event (i.e., default) is not explained by the predictor. Different thresholds of p-value can be used to consider a predictor relevant or significant. Some use 1 percent (e.g., in clinical tests); the most commonly used is 5 percent. Nevertheless, in business settings, 10 percent is accepted as a significant predictor if the model uses out-of-sample or testing sample data.
  - For values of p-value <=0.01, 0.05, or 0.1, the variable is relevant and should be kept in the model.
  - For values of p-value >0.01, 0.05, or 0.1, the null hypothesis remains valid, and the predictor should be discarded.

  P-values are a quick check. If there is some doubt, evaluate the predictive power of the model with and without the predictor to decide whether to include it or not. A low p-value does not necessarily mean that the variable makes no contribution to the predictive power of the model.
- Coefficients determine the "weight" and the "direction" in which each predictor will affect the probability of default. A negative sign indicates an inverse relationship between the predictor and default; a positive sign indicates a direct relationship between the predictor and risk. These signs, or the nature of the relationship, should make sense considering business knowledge and experience. For example, customer age would normally have a negative coefficient sign because the older applicants are, the lower their risk of default.

Once the coefficients are checked and the predictors work, it is possible to calculate the score for each account using the equation of the regression.

Generally, several iterations, where different variable combinations are tested, will be needed until the appropriate model is found. The model that shows the higher predictive power out-of-sample should be selected, provided that all coefficients and variables make business sense.

4.6 Model Scaling and Validation

Once the score for each account has been calculated, each row will have a predicted value between 0 and 1 (statistical packages calculate this automatically). To simplify it for better intuitive understanding, you should "calibrate" the model, which means adjusting the scale of the scores to ones that are frequently used and widely accepted. Ranges of scores commonly used are 1–100, 300–850, and 1–999.

The score distribution or "performance table" can be developed using descending ranges of the score that distribute the sample in deciles, for example. The score ranges can also be distributed normally, which could result in more separation for the lowest and highest risk groups. For every score range, tabulate the distribution of goods, bads, and default rates. The resulting table can be used to evaluate the model and to convert it to business decision rules for underwriting (see Table 10 for an example).

SECTION **5**

EVALUATING A SCORING MODEL

WHEN DEVELOPING A MODEL, you need to evaluate the different alternatives so that you are able to choose the one that yields the highest predictive power. You will evaluate a model to see how well it fits a different data sample and captures real relationships between variables versus the specifics of cases in the development sample. This testing is generally done on a hold-out sample—a subsample of about 30 percent of the total data available that is separated up front and not used in the development. Testing can also be done on an out-of-time sample, which is generally data from the period right after the sample period. In both cases, the objective is to validate that the predictive power of the model in the testing sample is close to that in the developing sample. If this is not the case, the variables that present the biggest difference should be evaluated and adjusted. You may need to reduce the number of categories, eliminate outliers, or eliminate variables with a high proportion of missing values.

5.1 Confusion Matrix

A confusion or error matrix is a table layout that helps you evaluate the performance of an algorithm (see example in Table 11). Each row of the matrix represents the instances in the predicted probabilities of default; each column represents instances in actual defaults at an aggregate level. The matrix shows two errors of different natures and consequences for the business:

- **False Negative**. Predicting a bad loan when it was a good one. The impact would be a potential loss of profits.
- **False Positive**. Predicting a good loan when it was a bad one. The consequence could be a potential loss of interests and principal and even the addition of recovery costs.

The accuracy of the model evaluated in this example is 90 percent, thus, the error or misclassified classes account for 10 percent.

A predictive model should be at least 50 percent accurate because a model that gets more than half of the cases wrong does not perform better than deciding at random.

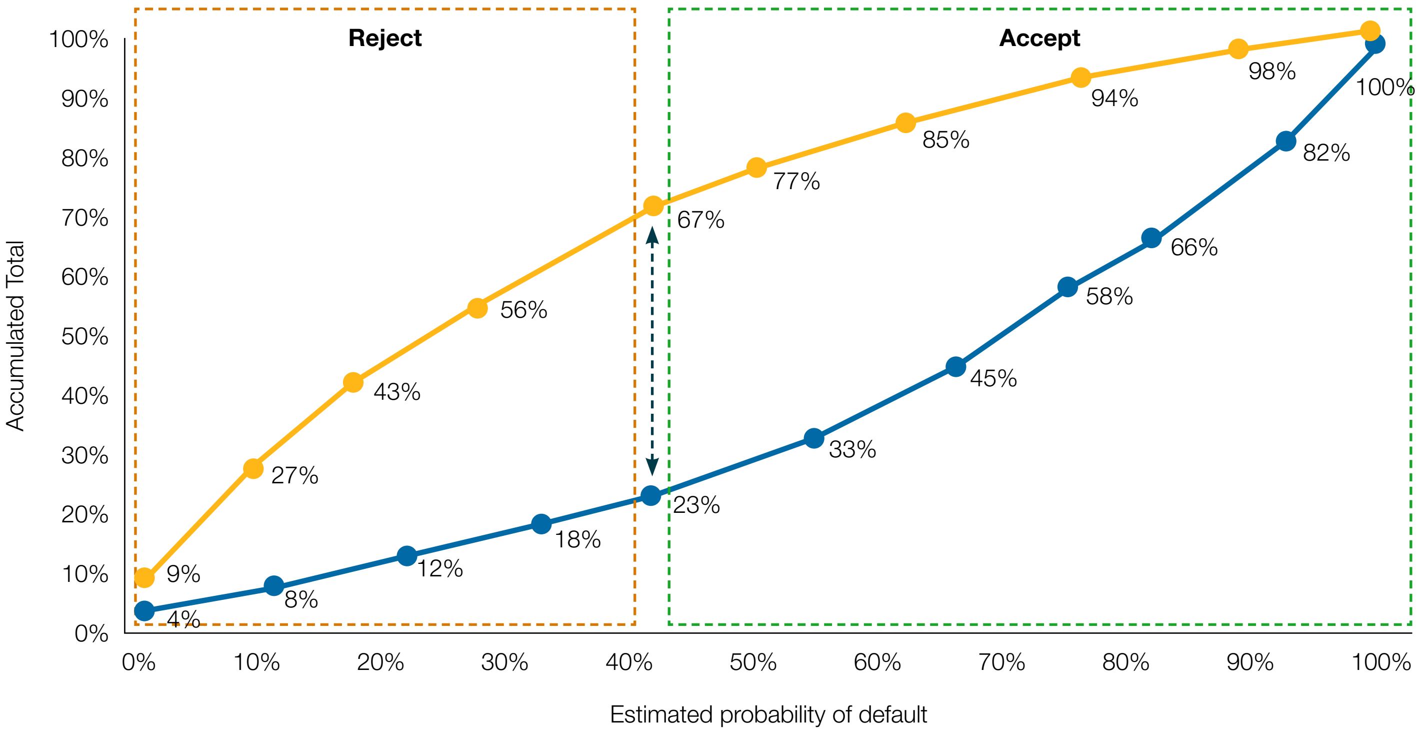
5.2 Kolmogorov-Smimov Test

The Kolmogorov-Smimov (KS) test measures the maximum vertical separation between two cumulative distributions (good and bad) in a credit scorecard. The higher the separation between the two lines, the higher the KS, which translates into a more accurate scorecard. In the example shown in Figure 7, the maximum difference in the accumulated rates happens at 67 percent of bad and 23 percent of good. Thus, the difference or the KS is 44 percent. After the KS point, the separation between classes

TABLE 11. **Example of a confusion matrix**

		Actual Performance	
		Low credit quality	High credit quality
Model prediction	Low credit quality	True negative (TN) Correct prediction (46%)	False negative (FN) Error (6%)
	High credit quality	False positive (FP) Error (4%)	True positive (FP) Correct prediction (44%)

FIGURE 7. **Example of a Kolmogorov-Smimov curve**



starts to shrink again. By accepting applicants at this point, an organization would be accepting 77 percent of goods and 33 percent of bads.

The effectiveness of the KS tool and the other tools discussed later in this Guide depend on the veracity of the sample and its characteristics. The results from a heterogeneous group (where the proportion of goods and bads is close to 50-50) will be higher or more accurate than when a homogeneous group is assessed (where most of the population is classified as good). This means that the tools are effective when comparing different models during their development (as they use the same sample) but it can be misleading when comparing models being applied to different products or on different samples.

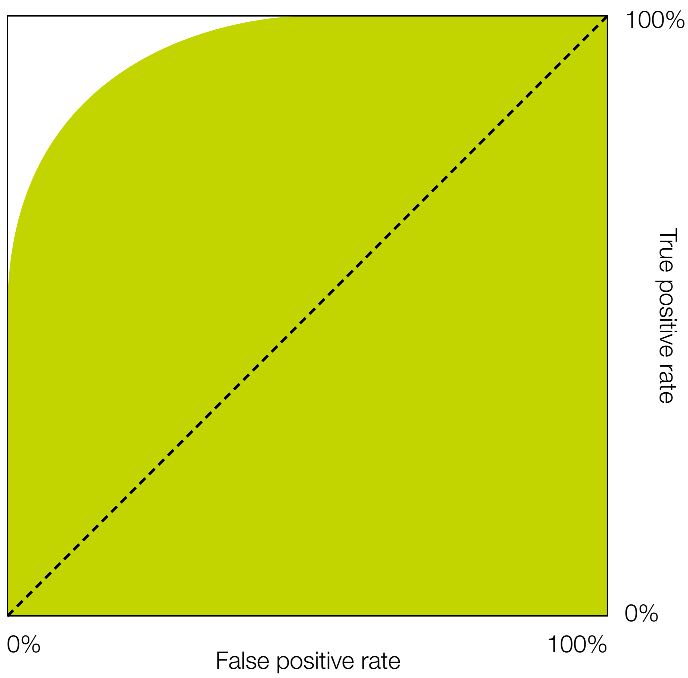
5.3 Receiving Operating Characteristic Curve, Area under the Curve and Gini

The Receiving Operating Characteristic (ROC) curve is a graphical plot of the True Positive Rate (percentage of bads rejected) versus the False Positive Rate (percentage of goods rejected) for every threshold or cut-off.

The diagonal dotted line in Figure 8 shows where the ROC curve would be if the model was as good as random at predicting default.

The Area under the Curve (AUC) measures the percentage of the box that is under this curve (in green). The higher the percentage, the more accurate the scorecard. It is calculated

FIGURE 8. **The ROC curve**



by integrating the ROC curve lower bounded by the diagonal dotted line at 0.5.

The Gini coefficient is a scale of predictive power; it is a linear transformation of AUC.

$$Gini = 2 * AUC - 1$$

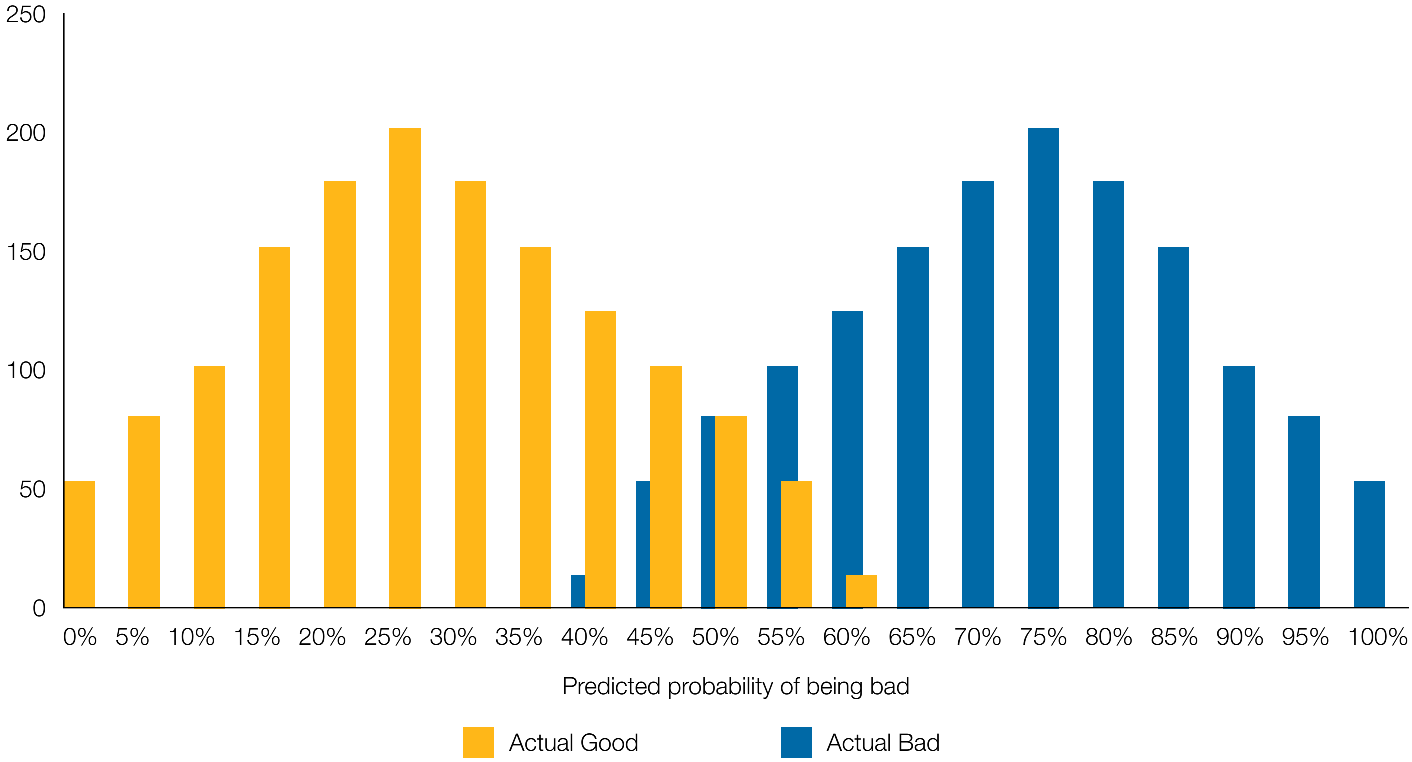
As the formula indicates, AUC goes from 0.5 to 1, and Gini goes from 0 to 1.

A Gini of 0- and AUC of 0.50- is a random prediction, while a score with a Gini of 1- and AUC of 1- is perfectly predictive.

5.4 Graphic Distribution

Another way to visually evaluate the performance of a model is to use a graphic distribution of the actual performance (good and bad) across the predicted probabilities of default. The more separated these distributions are, the more accurate the model. In the example in Figure 9, there are no actual bads below a predicted default of 40 percent. Likewise, there aren't any actual goods beyond a probability of default of 60 percent.

FIGURE 9. **Distribution of actual goods and bads across the predicted probabilities of being bad**



SECTION 6

HOW TO USE THE SCORING MODEL

ONCE THE MODEL HAS BEEN validated, it can be converted into a business decision-making tool. You will need to add few calculations to the performance table (Table 10) shown in Section 4.6. These calculations are as follows (see Table 12):

- Difference between cumulative goods and bads (KS)
- Cumulative percentage of the sample (estimated acceptance)
- Estimated default rate

To convert the statistical model into a decision-making tool, the organization must decide on the minimum score requirement for accepting customers. The minimum score reflects the level of risk the organization wants to take—the further down the table and the more customers receiving loans, the higher the default rate. In the end, the trade-off is portfolio volume and profit versus risk.

The cut-off should reflect the business objectives of the organization—for example, maximizing profit, reducing losses, or gaining market share.

Volume and risk are usually well evaluated, but when setting the cut-off many organizations fail to estimate the profitability of a defaulted customer and of a good customer. This is crucial if the main objective of the business is to increase profit. There are two situations in which organizations may not assess profitability properly:

- The actual loss in profit from a default (the average defaulter would probably have partially paid the loan)
- The potential for future profits from repeat loans to good customers

TABLE 12. **Example of a scoring model based only on "good" and "bad"**

Score	Goods	Bads	Total	Bad Rate %	% Goods	% Bads	KS	Estimated acceptance	Estimated default
901–999	2,780	20	2,800	1	23	3	20.7%	22%	1%
801–900	2,500	30	2,530	1	21	4	37.8%	42%	1%
701–800	1,750	35	1,785	2	15	5	47.9%	56%	1%
601–700	1,350	55	1,405	4	11	7	52.1%	67%	2%
501–600	1,250	70	1,320	5	10	9	53.5%	77%	2%
401–500	900	82	982	8	8	11	50.4%	85%	3%
301–400	750	95	845	11	6	12	44.4%	92%	3%
201–300	350	100	450	22	3	13	34.4%	95%	4%
101–200	200	135	335	40	2	17	18.6%	98%	5%
1–100	100	150	250	60	1	19	0.0%	100%	6%
Total	**11,930**	**772**	**12,702**	**6**	**100**	**100**			

A trade-off chart offers a visual of the decision (see Figure 10) (CGAP 2016). In this example, a cut-off of 600+ means an approval rate of 45 percent (left x-axis) with a bad rate of 6 percent (right x-axis). A cutoff of 400 would increase approvals to 65 percent and the bad rate to 7.5 percent.

Keep in mind that the cut-off does not necessarily mean accept or reject. Figure 11 shows an example where an organization may choose to accept customers with scores above 600; refer to manual scrutiny score between 600–400; and reject scores below 400. For referral cases, many organizations do a personalized assessment and ask for additional supporting documentation or collateral to mitigate risk.

Once the cut-off point has been agreed on, the model is ready to be used. The organization needs the infrastructure and resources to provide operational support to the new credit scoring service and to implement its "go to market" plan. Experience shows that it is prudent to start small with a few applicants and to regularly iterate and update the model over the first few months as real-life data become available. As confidence in the model grows, it can be rolled out to larger numbers of applicants.

FIGURE 10. **Example of a trade-off chart**

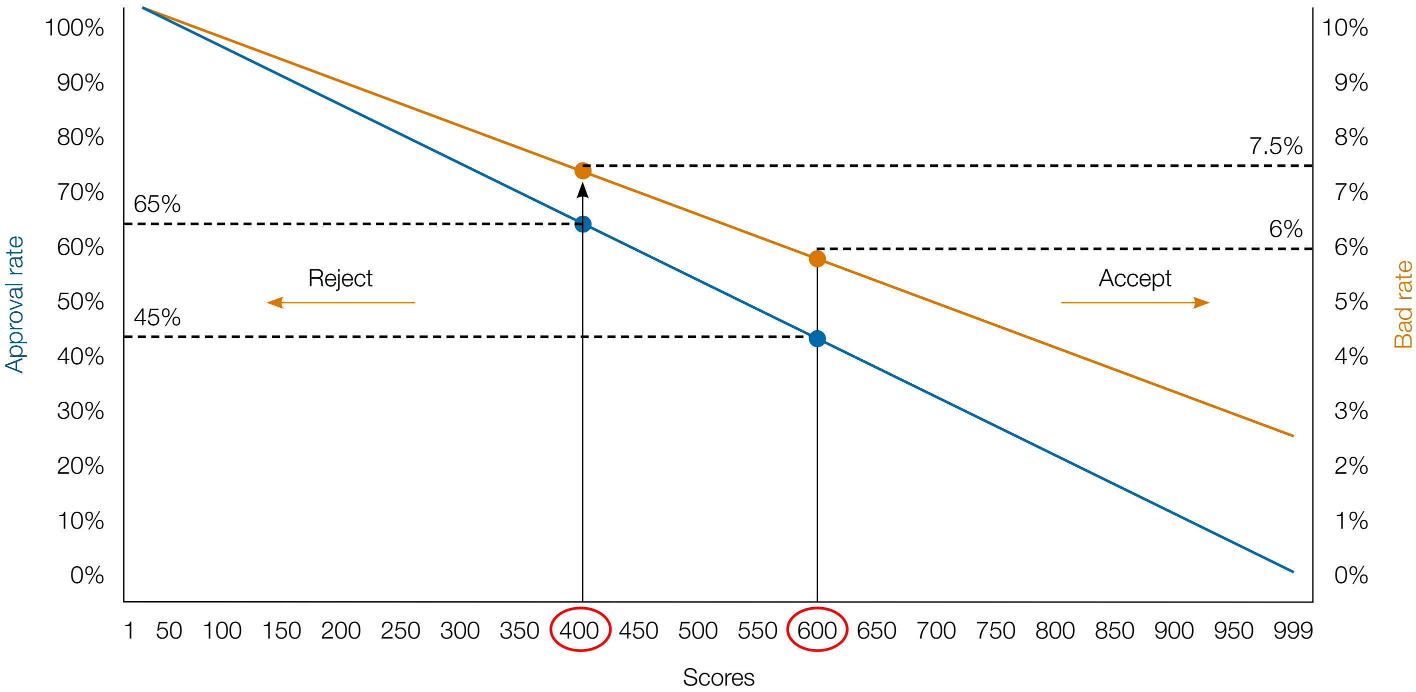
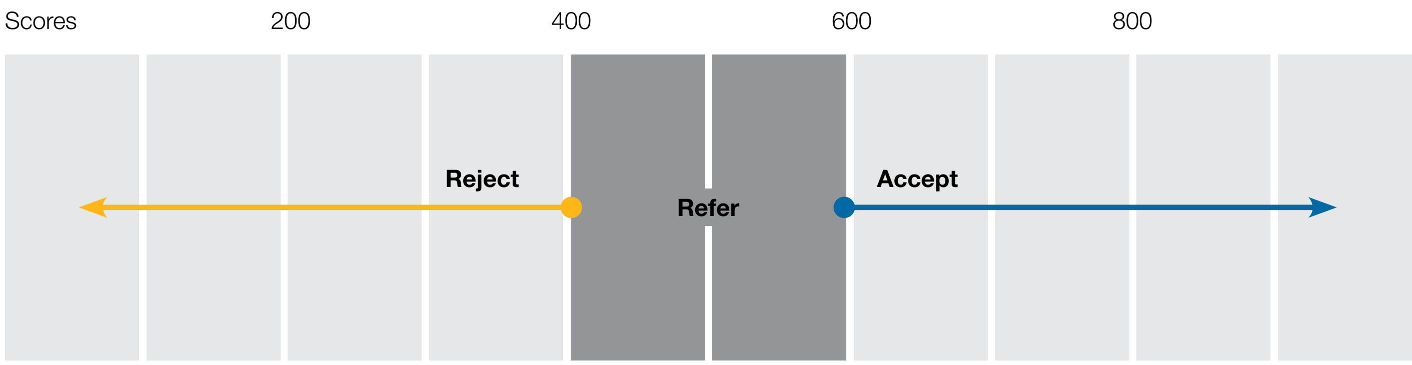


FIGURE 11. **There is a grey area where applicants may be referred for further scrutiny**



SECTION **7**

ADVANS CÔTE D'IVOIRE CASE STUDY

ADVANS IS AN INTERNATIONAL microfinance organization that targets developing markets where clients lack adequate access to financial services. It is in nine countries: Cambodia, Cameroon, Ghana, Democratic Republic of Congo, Côte d'Ivoire, Pakistan, Nigeria, Tunisia, and Myanmar.

7.1 Developing a New Credit Product

Education expenses can be a significant financial burden on low-income families, especially smallholder families that have seasonal incomes. For cocoa farmers in Côte d'Ivoire—who have, on average, four kids in school—school costs come due in August or September, several weeks before their biggest harvest. This creates a liquidity problem that Advans Côte d'Ivoire is trying to solve with a new digital credit product. By leveraging partnerships with farmers' cooperatives and telecom operator MTN, in a recent pilot Advans achieved 100 percent on-time repayment and notably increased school attendance among borrowers' children. This innovative work with cocoa farmers earned Advans the 2018 European Microfinance Award for inclusive finance through technology.

Advans already had been lending to cocoa farmers cooperatives in Côte d'Ivoire and offering them a savings account where cooperatives can deposit cocoa payments through their phones. The company knew that many in this segment go through hardship due to the mismatch between their farming cashflows and timing of school expenses. So, with technical support from CGAP, it began to explore the potential of education loans for qualified customers.

Given that this was a new digital loan product, for which no previous loan performance information existed, a pilot was set up using a judgmental scorecard based on data available for farmers that had a savings account with Advans. The project had three objectives:

- **Establish a direct lending relationship** with cocoa farmers who do not have a borrowing history with Advans through a digital school loan. All applicants had an Advans savings account.
- **Create repayment history from farmers** and eventually graduate them to larger loans and continue learning about their repayment behavior.
- **Strengthen partnerships with cooperatives** by providing them concrete differentiating value that farmers appreciate.

Product. Loans were timed for the start of the school year, which is several weeks before the main cocoa harvest when funds are scarce. The funds were disbursed directly to farmers who could then transfer it to their MTN Mobile Money accounts. Farmers could pay the school fees digitally, and/or withdraw cash at agents to pay for various school-related expenses. Repayments were flexible and allowed farmers to repay on installments through the harvest months (next three months). The product was simple and easy to use. In addition, field agents provided extensive on-the-ground training for farmers to ensure that even the less literate understood the product and how to use it. The cooperatives played a major role in delivering the product by supporting farmer education and payment collection. Research showed that farmers strongly preferred to pay at the time of cocoa delivery with the help of a cooperative representative, as they had been doing for their input loans.

Data. The loan eligibility and amount were determined using information from the farmers' saving behavior and from other reported information on their agriculture assets and production. The information held by the individual cooperatives on their lending to farmers was not standardized and was not always digitized. Therefore, that information could not be directly used in the scorecard. Despite this, Advans leveraged the cooperatives' knowledge of their members by requesting a partial guarantee from the cooperatives—whereby the cooperatives could choose not to guarantee all farmers. The cooperatives had been borrowing from Advans to offer input loans to their farmers. Under these arrangements, farmers would repay their cooperatives, who would, in turn, repay Advans. Since Advans lacked insight into individual farmer's past repayment behaviors, the cooperatives agreed to cover 30 percent of any amount of new education loans its farmers did not repay. This allowed Advans to tap into the cooperatives' knowledge of which farmers had not been good borrowers in the past (based on the assumption that cooperatives would not guarantee farmers they knew to be bad repayors).

Results. Every one of the 242 loans offered during the pilot was repaid on time. The loans had an average amount of $165, which is smaller than the typical productive microfinance loan but significantly larger than most digital loans, especially for first-time borrowers. At the end of the cocoa season, Advans Côte d'Ivoire and CGAP surveyed 45 borrowers to assess customer satisfaction. All of them said they were satisfied with the product, and 96 percent said they would apply again next year. About 60 percent of the farmers were satisfied with their loan amount, while 40 percent said it was not enough to cover the costs for all their children. Most importantly, the surveyed farmers reported that the percentage of their children who started school at the beginning of the period increased from 49 to 73 percent.

Key lesson. The partnership with the cooperatives was a key success factor to ensure customer education and repayments. For the cooperatives who compete for farmers to grow and achieve scale, the ability to offer this additional service to their members was an attractive opportunity. Several issues needed to be overcome. These provided invaluable learning for the full implementation of small loans by Advans, particularly the need to streamline the way farmers' eligibility data was gathered by the cooperatives and collected by Advans.

Going forward. Advans is working to improve the eligibility data collection process. Because of the popularity (and low risk) of the pilot product, Advans has increased the number of farmers it is lending to for the next school year five-fold: the number of participating cooperatives increased from five to 20, and the number of education loans has gone from 242 to 1,118. Moreover, based on the solid repayment rates during the pilot and feedback that farmers with many children needed larger loans, Advans increased its average loan size from $165 to $190. The feasibility of providing a similar loan product to members of other value chains with different seasonality constraints is being analyzed. As the product expands and more repayment data become available, the initial judgmental model can be replaced by a statistical model.

7.2 Streamlining the Existing Underwriting Operation

Advans had an exhaustive and labor-intensive underwriting process for all its loans, which include at least one visit to urban customers (entrepreneurs) regardless of the loan amount. It agreed to develop a scoring model using credit repayment data from existing customers to streamline the underwriting process for loans of up to CFA 3 million (approximately US$5200).

After several months of collaborative analysis and development, Advans is set to implement a scorecard to automate the underwriting process for small loans. It is also planning to implement scoring models in other subsidiaries worldwide to reduce costs and improve the decision-making process.

By applying this model, Advans expects to automate a significant portion of its credit decisions, focusing loan offices on the more complex, larger, and risker loans and reducing the cost of originating the smaller, lower-risk loans.

Several key people from Advans were involved in the different steps of the project. The following explains the different roles these people played in this project.

Project manager. The project manager oversaw and coordinated the project's activities and processes. She defined the timeline and ensured deadlines were met. She coordinated and led the interactions between Advans and CGAP. She also managed Advans' resources (manhours) and assigned staff to each task.

Credit analyst. The credit analysts or loan officers provided the details of the underwriting process. They validated the main predictors found in the data against customers' profiles. It is important to involve these analysts throughout the project so they understand and support the new process.

Risk manager. The risk manager provided all the changes in the risk policies along the years, which helped the team to choose time windows of comparable data. He validated the definition of bad and the consequent default rate and the list of default predictors. The risk manager supported the CEO in determining the cut-off point in line with overall risk appetite of the company.

Country manager. The country manager set the main objectives of the project and validated all key decisions and findings. He approved investments and granted access to the project manager. He defined the adequate risk-volume level of the cut-off point for the scoring model implementation. As the sponsor of the project to the organization, he communicated and explained the benefits of the project to his high-level staff.

Corporate risk manager. The corporate risk manager had two roles. On the one hand, she brought risk expertise to and oversaw the risk decisions taken during the project. On the other hand, she learned all the details about the project so that she would be able to roll out the implementation in other countries.

Data and risk expert. The data and risk expert led and executed the technical aspects of the scoring model: from generating the data request to the IT/database manager to develop the scoring engine to defining data mart requirements for implementation. He brought many years of experience in developing scoring engines to the team, and he transferred all the necessary knowledge to the organization to update the model once it has new data.

Database manager. The database manager extracted and formatted the raw data samples and shared details on how each data point was captured and stored. He checked and validated the reasons for outliers and missing values. He developed a few programs to collect different sorts of decentralized data. Toward the implementation, he supported the team in generating the IT technical specifications and requirements for the scoring engine.

SECTION **8**

FINAL LESSONS

WHEN AN ORGANIZATION decides to become data driven, it should expect the implementation journey to be somewhat complicated and time consuming. There will be challenges along the way, but experience shows that these can be overcome and lead to significant business benefits. Organizations that have already made the journey, report significant improvements in business efficiency and cost savings; improvements in their ability to identify and capitalize on new opportunities; and better, customer-focused service.

Using data for credit scoring replaces a company's reliance on "gut feeling" with statistical probability and reduces the potential risks associated with personal relationships, personal judgement, and dependence on the individual employees that hold them. Loan applications can be processed much faster with less in-person interaction, which increases customer satisfaction, the potential size of the loan book, and the capacity of field operatives. Some small loans can even be assessed automatically and disbursed in minutes.

However, credit scoring is a sophisticated discipline that requires special resources, including staff with specialized expertise and a conducive infrastructure. The specialized expertise can be in-house or outsourced. If the data management infrastructure is not already in place, it must be acquired. Credit scoring may also require significant organizational change.

The following are recommendations for organizations seeking to implement credit scoring.

Do not underestimate the importance of keeping employees informed. It is essential to communicate with staff early and clearly about the changes and the benefits they can expect for themselves and the business. Don't forget to provide regular updates. If loan officers are involved in the new process, the sooner they are involved in the project, the greater the chances of its success.

Invest in setting up a data analytics infrastructure if it is not already in place. You may need to ensure your IT team has data analytics capabilities. Expertise in data analytics can be recruited into the organization or outsourced to a third party. The cost and time taken to put this in place should be factored into the business case.

Create a business benchmark by assessing the existing lending process and **its strengths and weaknesses**. Be clear about how credit scoring is expected to improve the existing service and set key performance indicators to measure its effectiveness. The company may be aiming to reduce costs, improve portfolio quality, or do both.

Review and update the business processes associated with providing credit. A key benefit of credit scoring is that it is more efficient than previous practices, so if the business processes do not take account of efficiency improvements, the benefits may be lost. For example, the model might be able to automate 50 percent of credit decisions but if analysts still review all applications anyway, the efficiency of the model will not be captured.

When building a scoring model, **consider the customer sample measured**, the appropriate **time window**, the **sources of data**, and data **quality**. Missing and marginal (outlying) data must be handled with by an approved technique so that they do not skew the results unduly.

Test **the model rigorously using historical data that were not used to develop it** to avoid a biased model.

Launch credit scoring with a small number of applicants, possibly even performing "old-style" rating in parallel while the model is iterated and improved in the early days. As confidence in the model grows, larger numbers of applicants can be processed.

Test and improve the model continuously. Setting up a statistical credit model is not a one-time task. Model monitoring requires ongoing resources and reviews of model results against expected performance. Additionally, the model will need to be recalibrated periodically. Recalibration involves refreshing the model's parameters by incorporating new data into it. The smaller the initial sample used to develop the model, the sooner it will benefit from recalibration. The model may need to be redeveloped after several years. This is particularly important if the risk and customer profile change significantly or if the model monitoring identifies a decrease in model effectiveness that cannot be address though a simple recalibration.

REFERENCES

Anderson, Raymond. 2007. The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation. Oxford University Press. https://global.oup.com/academic/product/the-credit-scoring-toolkit-9780199226405?cc=es&lang=en&#

CGAP. 2016. An Introduction to Digital Credit: Resources to Plan a Deployment. CGAP. https://www.cgap.org/research/slide-deck/introduction-digital-credit-resources-plan-deployment

Fernandez Vidal, Maria, Dean Caire, and Fernando Barbon. 2019. "Data-Driven Segmentation in Financial Inclusion." Washington, D.C.: CGAP.

Davico, Gisela Paola, and Christian Rodriquez. 2018. "Changing Change Management: Adapting Internal and External Culture in Times of Digital Transformation." Field Note 8. Washington, D.C.: World Bank Group. http://documents.worldbank.org/curated/en/115081548846139465/Field-Note-8

Kabakoff, Robert I. 2017. "Quick R: Advanced Handling of Missing Data." DataCamp. https://www.statmethods.net/input/missingdata.html

Schreiner, Mark. 2003. "Scoring: The Next Breakthrough in Microcredit?" Occasional Paper 7. Washington, D.C.: CGAP. https://www.cgap.org/research/publication/scoring-next-breakthrough-microcredit

APPENDIX
BY DEAN CAIRE

Using R Software

Regressions and many of the analyses presented in this Guide can be performed using Excel. However, you may need specialized software to take on significant modeling tasks.

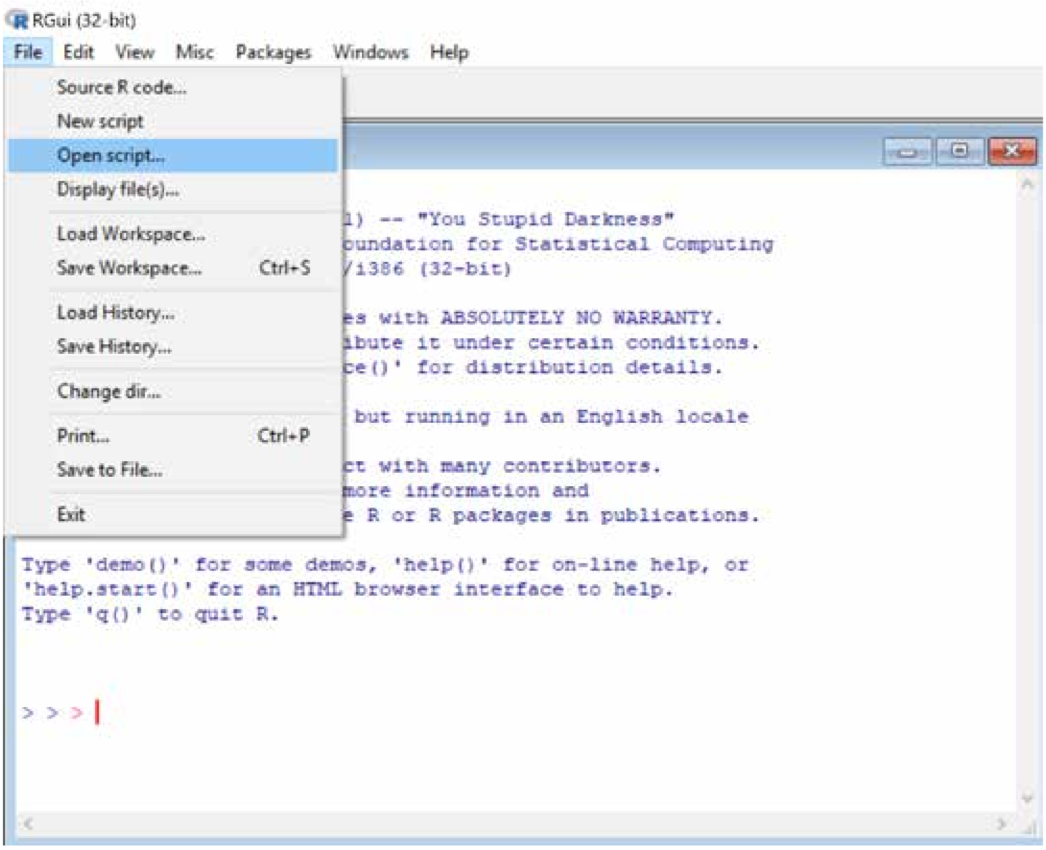
R is an open source, free software that has become increasingly popular among model developers. This appendix first walks you through how to set up R and offers a sample code (the sample code is also available as a text file) for regressions. The software has libraries that include standard statistical procedures and R codes for a wide variety of uses and tests. It is available for free at http://www.r-project.org/.

The rest of the appendix provides step-by-step guidance to run the code. In addition to following the online instructions on installing the program, you should read the basic introductory materials and/or watch one of the many introductory tutorial videos available on the internet.

IMPORTING DATA AND INSTALLING PACKAGES

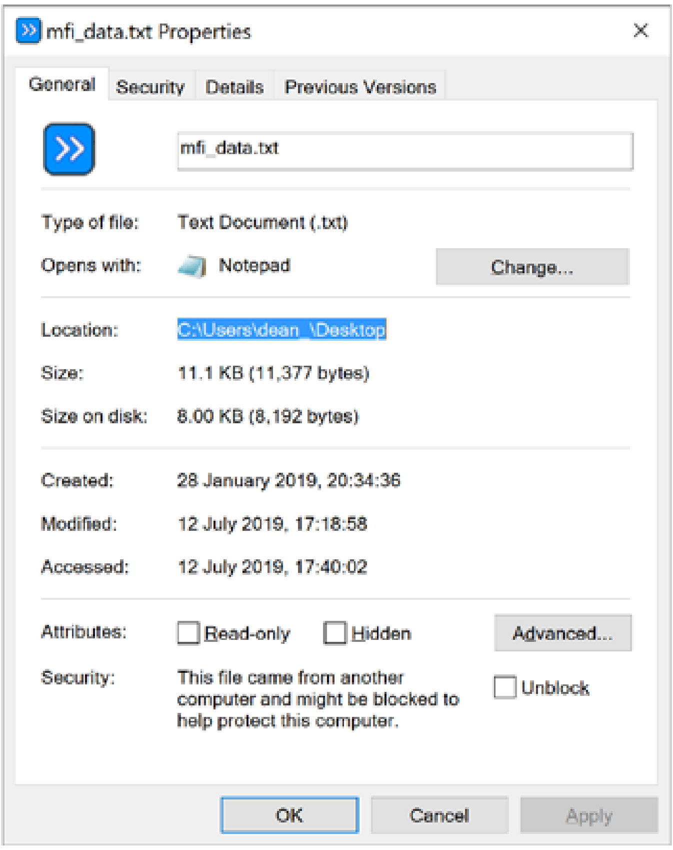
To follow the examples, download the sample data file.

1. Open a new script in R (File -> New script)

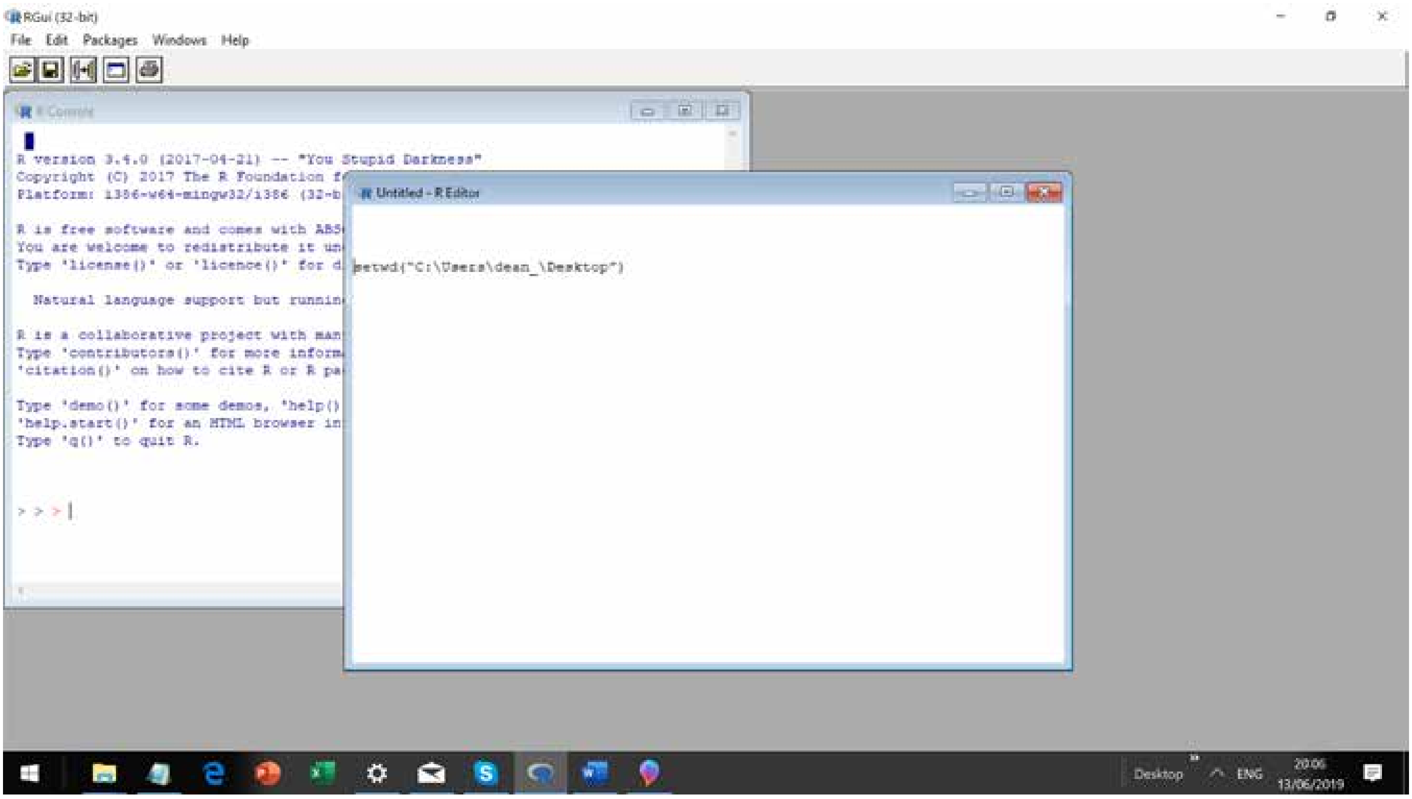


2. Set a path to the location on your computer where you have placed the data file you will analyze.

Navigate to the file and right click on it which launches a properties window:



Copy the path to the file location (outlined in blued) and paste it into the script editor as follows:



Double the forward slashes in the path and add the following text:

setwd("C:\\Users\\dean_\\Desktop")

3. Import the data

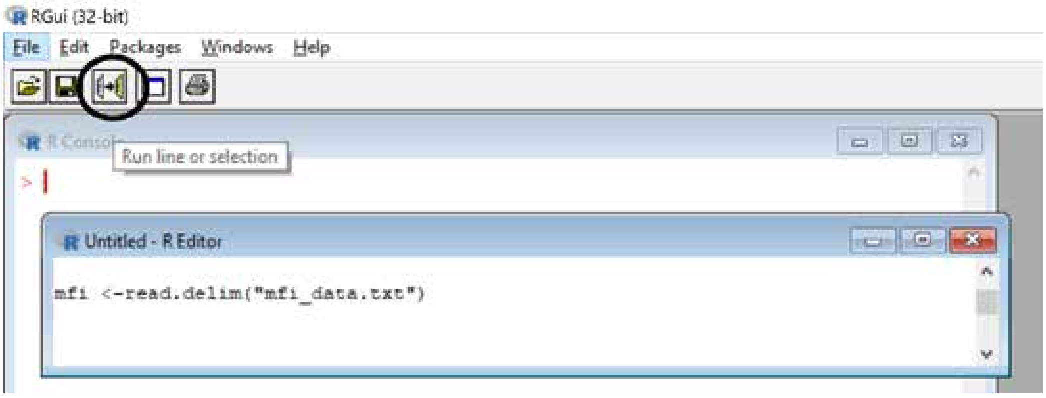
The sample data file is a tab-delimited text file with the following features:

- The first row contains column headings
- All subsequent rows contain data records, and there are no extra cells such as lines with totals or other single cells not relating to columns or rows of the table
- Correctly formatted data
- No missing values

The script to read in the data file is:

mfi <-read.delim("mfi_data.txt")

To run this code (and read in the data file), highlight this text in the R script editor and either push the button shown circled or the key combination "ctrl + R"

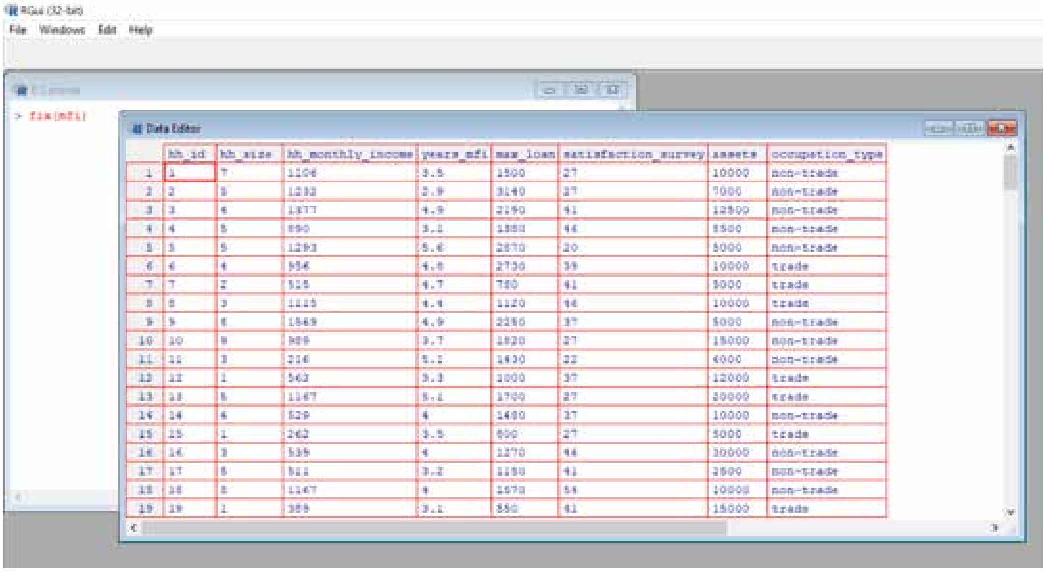


In this code, you created an object arbitrarily named "mfi" to which you will assign (using "<-") a data table that contains the data in the .txt file.

You can look at the imported data table in the R data editor using the "fix" command, which brings up a "spreadsheet" view of the data in R as shown below

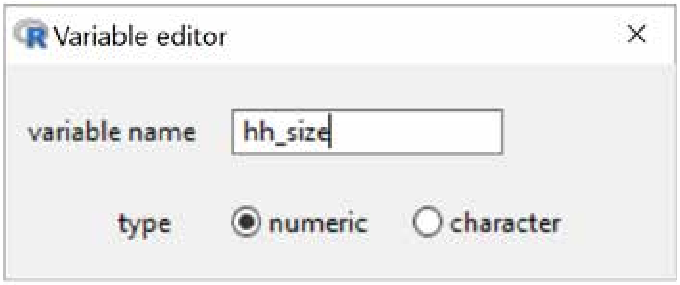
Highlight that code and press the button or use "ctrl + R".

DATA EDITOR IN R



	hh_id	hh_size	hh_monthly_income	years_mfi	max_loan	satisfaction_survey	assets	occupation_type
1	1	7	1106	3.5	1500	27	10000	non-trade
2	2	5	1232	2.9	3140	27	7000	non-trade
3	3	6	1377	4.9	2190	41	12500	non-trade
4	4	5	890	3.1	1380	46	8500	non-trade
5	5	5	1293	5.6	2870	20	5000	non-trade
6	6	4	956	4.8	2730	39	10000	trade
7	7	2	515	4.7	780	41	5000	trade
8	8	3	1115	4.4	1120	46	10000	trade
9	9	8	1569	4.9	2250	37	5000	non-trade
10	10	9	989	3.7	1820	27	15000	non-trade
11	11	3	214	5.1	1430	22	6000	non-trade
12	12	1	562	3.3	1000	37	12000	trade
13	13	5	1167	5.1	1700	27	20000	trade
14	14	6	529	4	1480	37	10000	non-trade
15	15	1	262	3.5	600	27	5000	trade
16	16	3	539	4	1270	46	30000	non-trade
17	17	8	511	3.2	1150	41	2500	non-trade
18	18	8	1167	4	1570	54	10000	non-trade
19	19	1	355	3.1	550	41	15000	trade

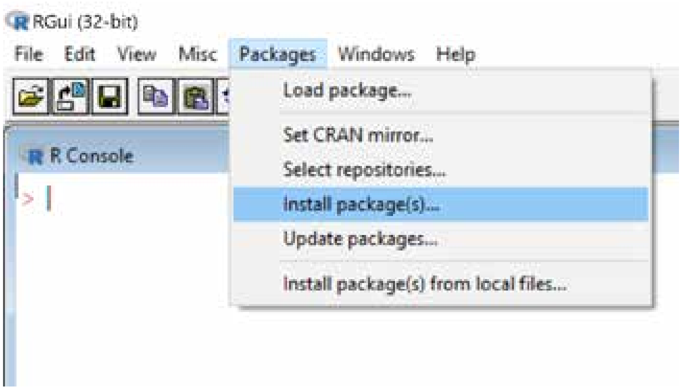
You can visually check your data in simple ways such as double-checking the number of rows in the table, checking that text (string) values are visible (particularly when you are working with various languages), checking that numeric data are formatted as numeric data, etc. You can see the variable name and data type by clicking on the column name in the data editor. This calls up a message box of the type:



4. Install the packages in R.

The example codes below call for specific libraries at the beginning of the code. Before you can run a code that calls for a library, you need to install the library into the software.

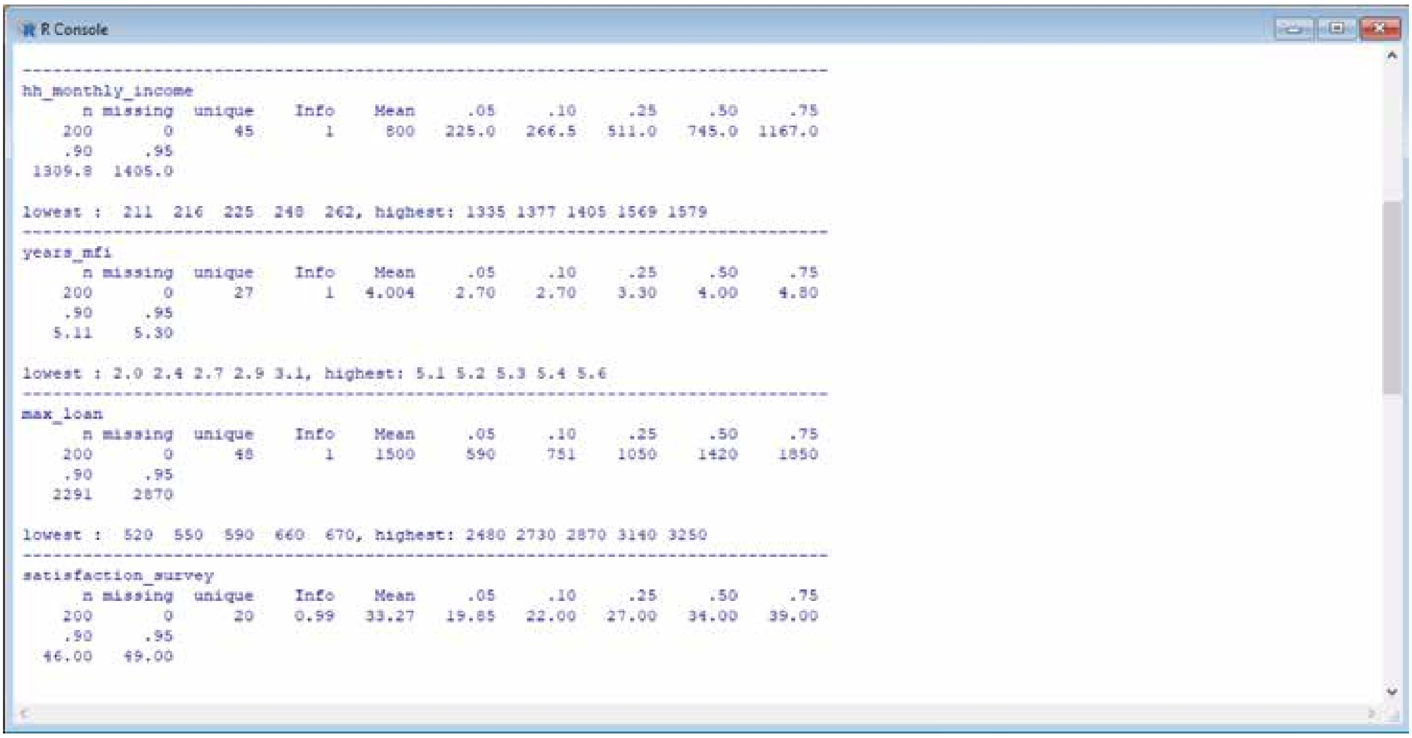
Navigate to Packages -> Install package(s)…



Select any of the online repositories (e.g., "0-Cloud[https]") and click "OK".

Find the required packages. Highlight them, one at a time, and then press "OK".

This downloads the necessary function libraries to the R directory on your computer. These need to be installed only once; they will be available for future use by running code to load the library into memory.



Statistical Techniques Used in the Guide

DESCRIPTIVE STATISTICS

```
str(mfi)
###looks like this
library (Hmisc)
###looks like this:
```

SIMPLE LINEAR REGRESSION

```
fit <- lm(mfi$hh_monthly_income ~ mfi$hh_size, data=mfi)
summary(fit) # show results
```

LOGISTIC REGRESSION

```
library(caTools)

mfi$has_deposit[mfi$deposit_balance > 0]      <- 1
mfi$has_deposit[mfi$deposit_balance <= 0]     <- 0

logit_model <- glm (mfi$has_deposit ~

mfi$occupation_type          + 1, binomial()) summary(logit_model)

mfi$logit_model =predict(logit_model,type = "response")

colAUC(mfi$logit_model,mfi$has_deposit, plotROC=TRUE)
```

CLASSIFICATION TREES

```
library(rpart)
library(party)
library(rpart.plot)

mfi$has_deposit_num = factor(mfi$has_deposit,labels=c("No Deposit","Has Deposit"))

binary.model <- rpart(has_deposit_num ~ hh_size +
hh_monthly_income          +
years_mfi                  +
max_loan                   +
satisfaction_survey        +
assets                     +
occupation_type            +
health_insurance           +
location                   +
rent_own,
method="class", data=mfi, cp =.01)

summary(binary.model) # detailed summary of splits
rpart.plot(binary.model,tweak=1.3)
```

CGAP
cgap.org