

WHITEPAPER

The Comprehensive Guide To Alternative Data Quality Assessment

A horizontal bar at the bottom of the page, divided into four colored segments: purple, pink, lime green, and orange.

TABLE OF CONTENTS

01 All you need to know about Alternative Data

02 All data differs: Different types of data

03 Role of Credolab: Diving into behavioural data

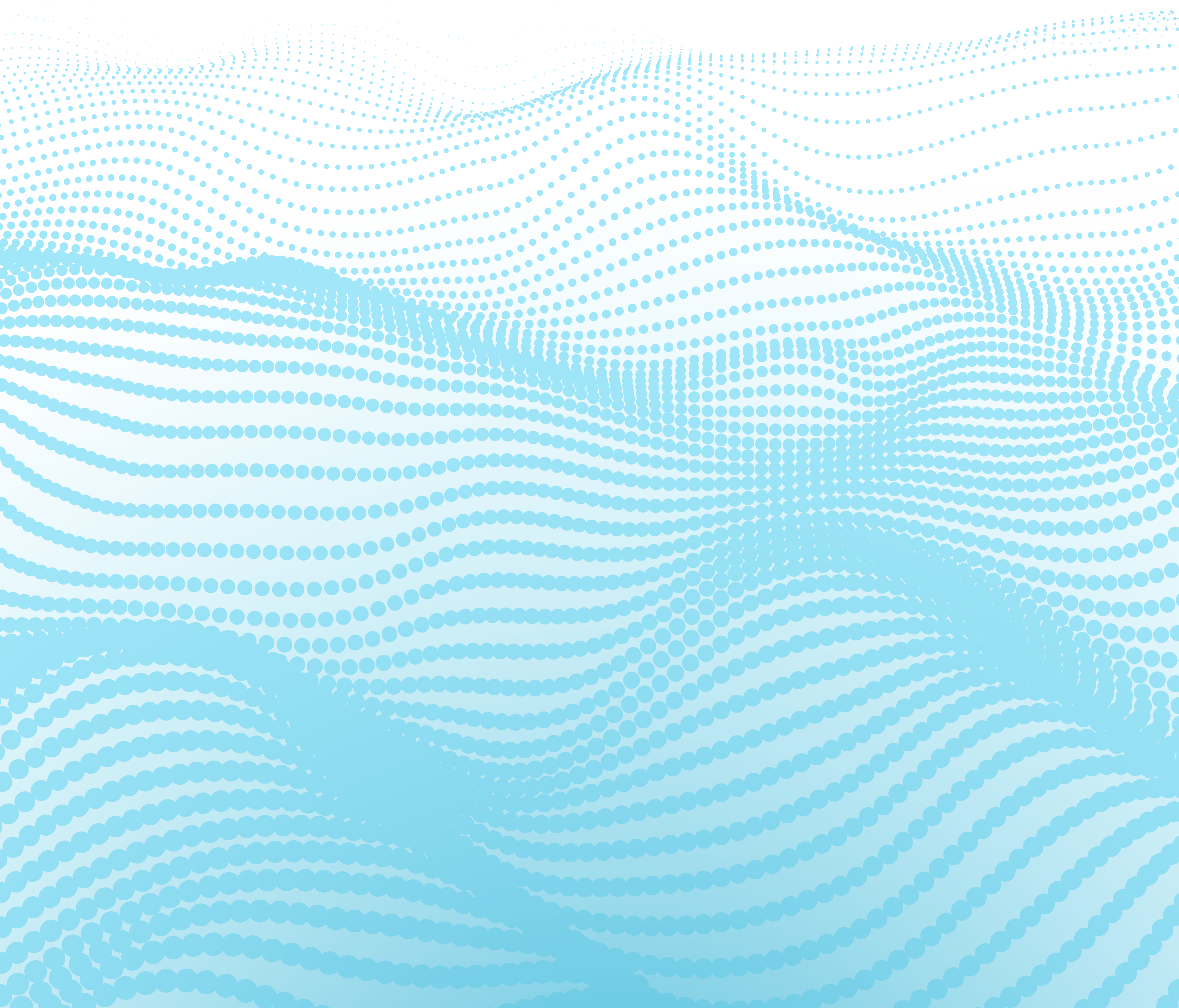
04 Assessing data quality Part 1: What makes a great alternative data source?

05 Assessing data quality Part 2: How is Credolab a good alternative data source?



INTRODUCTION

All you need to know about
Alternative Data



In the evolving landscape of finance and technology, 'alternative data' has emerged as a critical and transformative concept.

Alternative data reshapes how the financial landscape approaches data interpretation and customer understanding.

RESEARCH

This rapid growth underscores the increasing importance of alternative data in financial decision-making processes.

According to a report by [Grand View Research](#),

\$7.20 BN

is the value (in USD) of the global alternative data market size in 2023.

50.6%

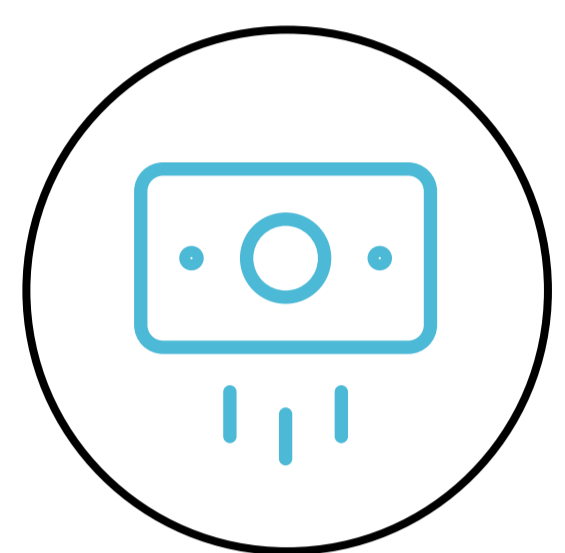
is projected to grow at a compound annual growth rate (CAGR) from 2024 to 2030.

Over the past years, the digital revolution has enabled unprecedented insights into user behaviours and creditworthiness by analysing digital footprints. The emergence of alternative data generates a more nuanced understanding of user behaviour, including one's ability or willingness to repay, ushering in a new era of more inclusive and accurate credit models.

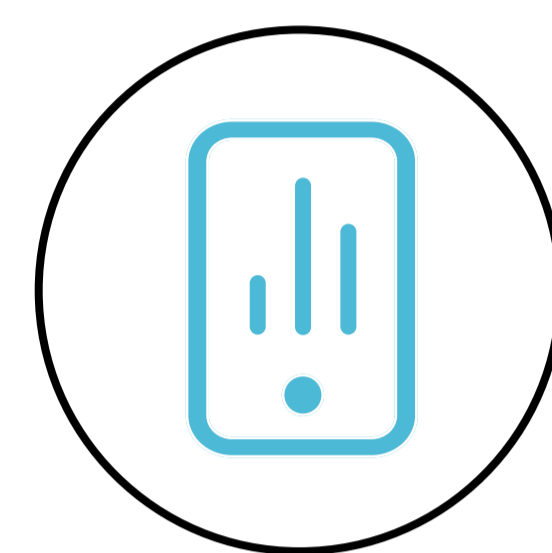
The rise of alternative data addresses the need to build more inclusive portfolios, especially for those New-To-Credit (NTC) or thin-files, predominantly in emerging markets. These individuals, often underserved or outright unbanked, possess limited credit histories, resulting in a lack of traditional data.

By harnessing diverse digital sources, interactions, and behaviours, alternative data transcends traditional risk assessment frameworks, offering a comprehensive picture of a user behaviour.

These include, but are not limited to:



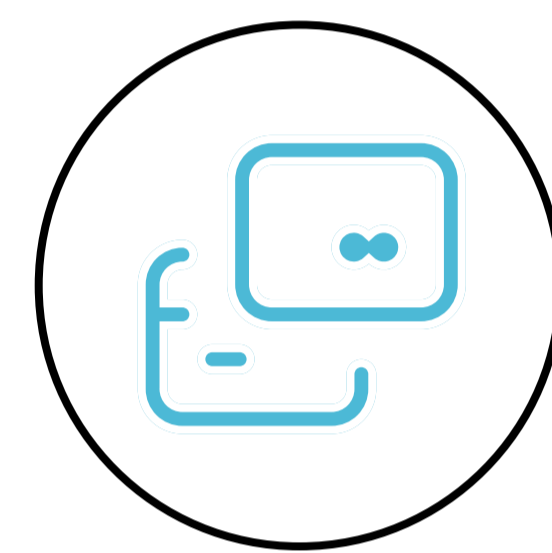
Rental Payments



Social Media Activity



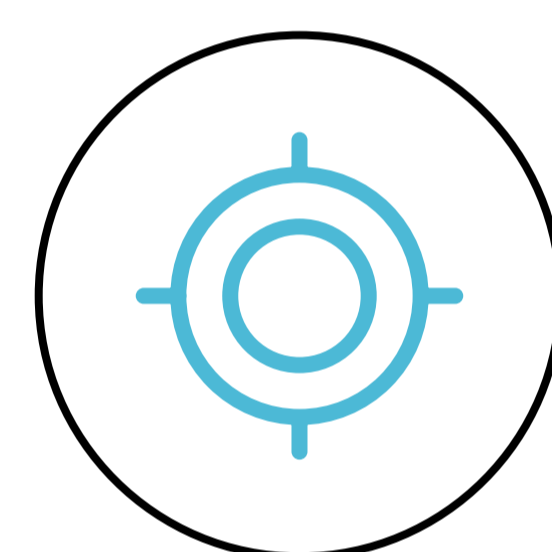
Subscription Models



Financial Transactions



Email Engagement



Geolocation Data

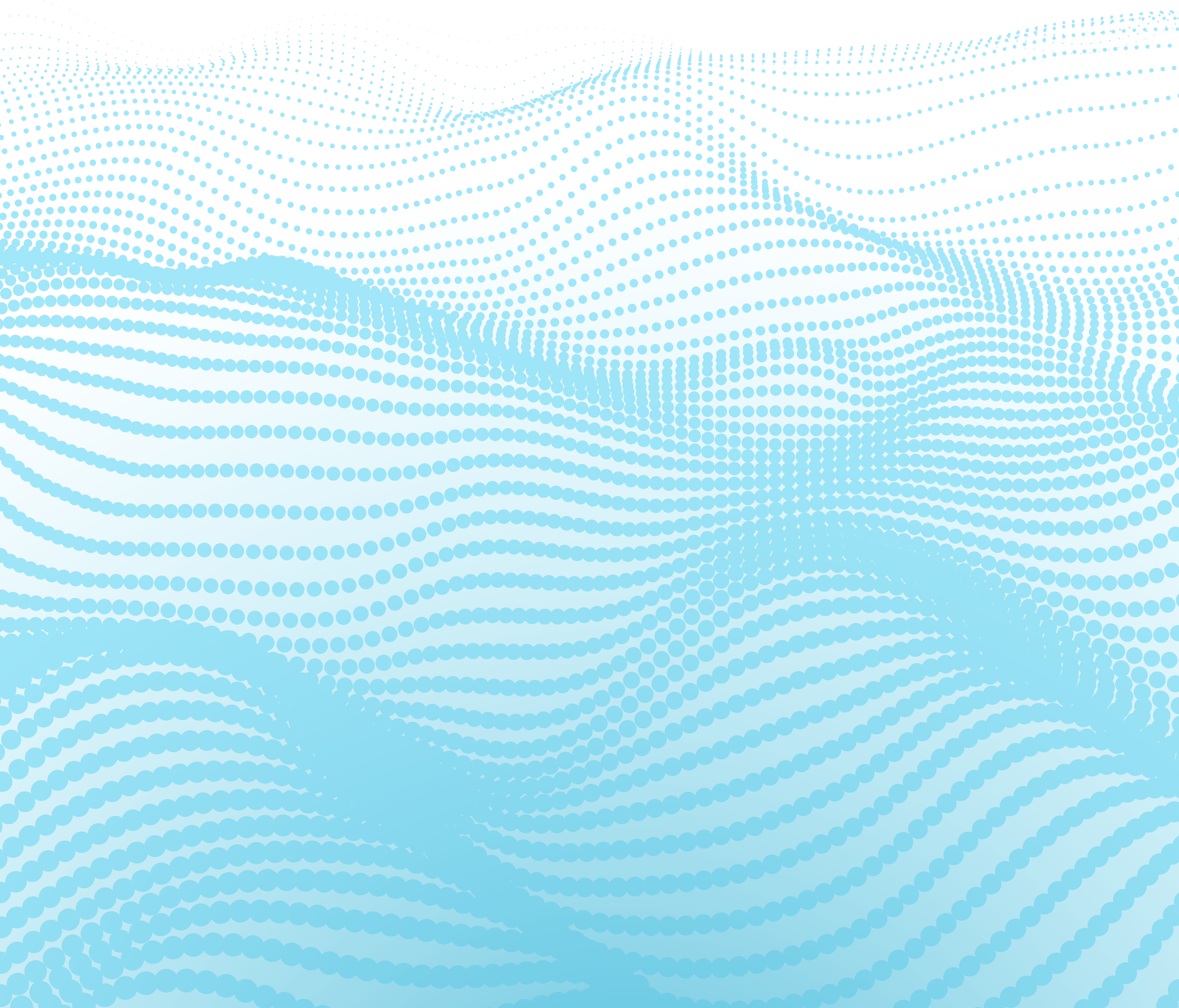
Furthermore, integrating Artificial Intelligence (AI) and Machine Learning (ML) algorithms with alternative data marks a pivotal shift in credit assessment methodologies, enabling a deeper and more accurate analysis of user behaviours and preferences. This powerful combination fosters personalised financial products and services, enhancing customer relations and driving profitability by accessing new and untapped market segments.

With advancements in AI and ML propelling the rise of alternative data, businesses face a unique opportunity to refine strategic approaches and achieve significant growth. Adopting a data-driven mindset leads to a wealth of new information and securing a competitive edge in the fast-evolving financial sector.

Exploring the capabilities and applications of alternative data, this ebook unlocks the potential of alternative data and how to assess data quality, spotlighting Credolab's role in this process. Become equipped with the knowledge and tools to convert real-time data and insights into actionable intelligence, fostering financial inclusion and innovation.

ALL DATA DIFFERS

Different Types of Alternative Data



At the heart of digital transformation is alternative data reshaping the financial landscape.

Behaviours

Habits

Interests

Transactions

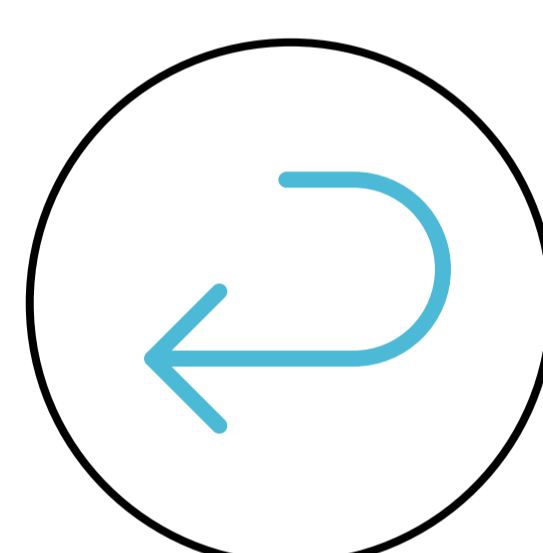
This innovative data stream captures new insights from non-traditional sources into individuals'

Unlike traditional data sources, alternative data draws from a wide array of sources to unveil a richer, more nuanced perspective on consumer behaviour, highlighting the practical applications of alternative data.

However, the power of alternative data lies not only in its diversity but also in its practical applications. For example, businesses who want to leverage alternative data insights for fraud detection and identity theft prevention can harness alternative data to enhance:



Predictive Accuracy



Optimise Returns



Mitigate Unforeseen Risk

This approach extends beyond mere data aggregation, offering a secure, insightful method to understand users and customers better. However, it is paramount to acknowledge that **not all alternative data is equal**.

The inherent variability in the value of alternative data and the quality of insights can fluctuate significantly based on the source. This demonstrates the importance of selecting and aggregating data carefully to ensure the data's contributory value towards achieving objectives such as:



Improved credit risk assessments



Increased approval rates



Enhanced fraud prevention



Enhanced customer satisfaction, experience

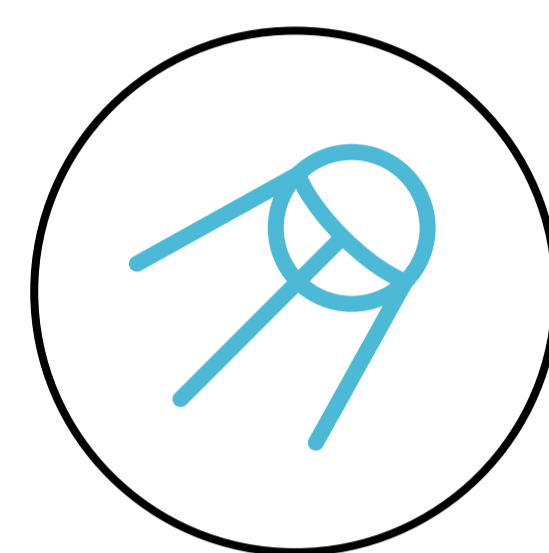
MULTIPLE DATA SOURCES

Furthermore, combining multiple data sources helps gather comprehensive insights that reduce information asymmetry and enrich the overall credit assessment.

Some examples of alternative data sources include:



Social media platforms



Satellite imagery



Purchase receipts tucked away in emails



Credit card transactions



Sensor outputs

Telco data, for instance, can be beneficial for preventing fraud. Many telco data providers offer fraud detection flags by comparing geolocation data with the information that the applicant declared during the onboarding process.

Telecom operators provide telco data, including:

Mobile phone usage

Server logs, SIM card swap

Call and SMS records

Billing and top-ups

Network status

Roaming data

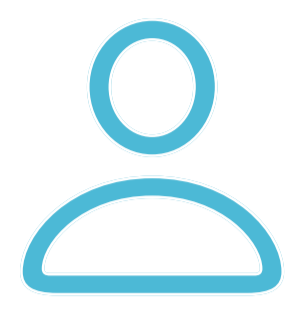


EXAMPLE

For example, a mismatch between the home and work address with the geolocation address obtained from the user's cell phone tower could be a warning for suspicious behaviour.

Utility companies

provide utility bill data from historical payment data, including:



Name



Statement date, due date and amount due



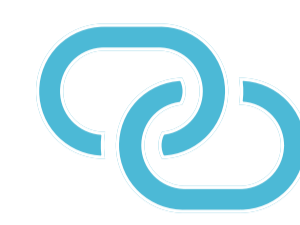
Address



Recent payment amount and recent payment date

E-commerce platforms

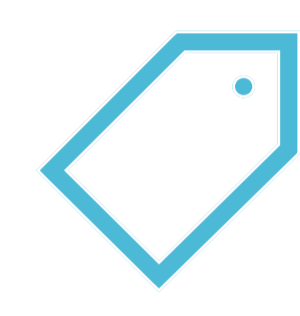
provide e-commerce data including:



Purchase patterns



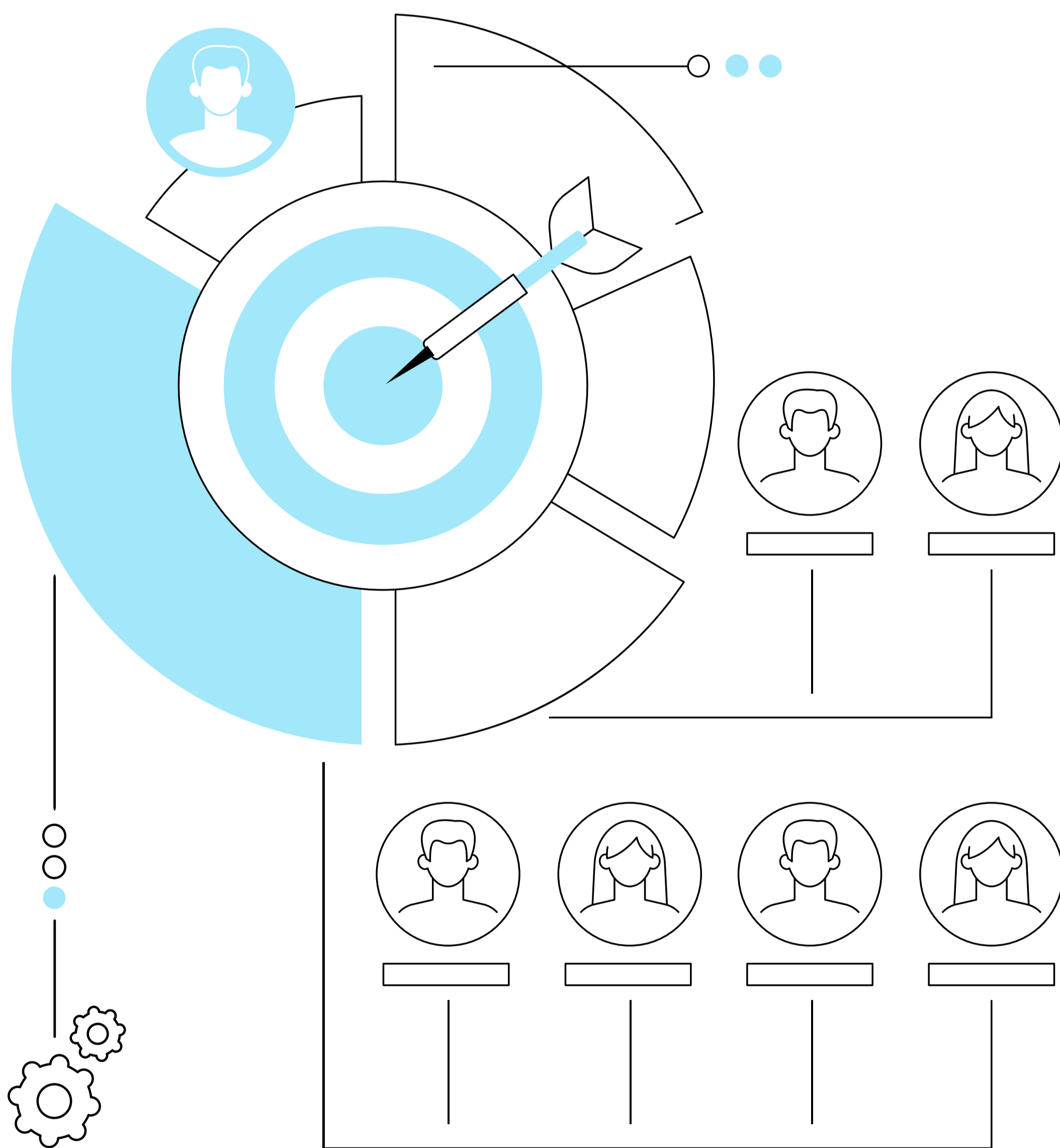
Products shopped



Pricing



Sales performance and other data about buyers and sellers



Platforms studying psychometrics that provide psychometric data, including

Psychological abilities

Personality characteristics

Behavioural styles

Required cognitive abilities

Typical Psychometric Credit Scoring

Tests are designed to evaluate an applicant's personality traits such as

self-control

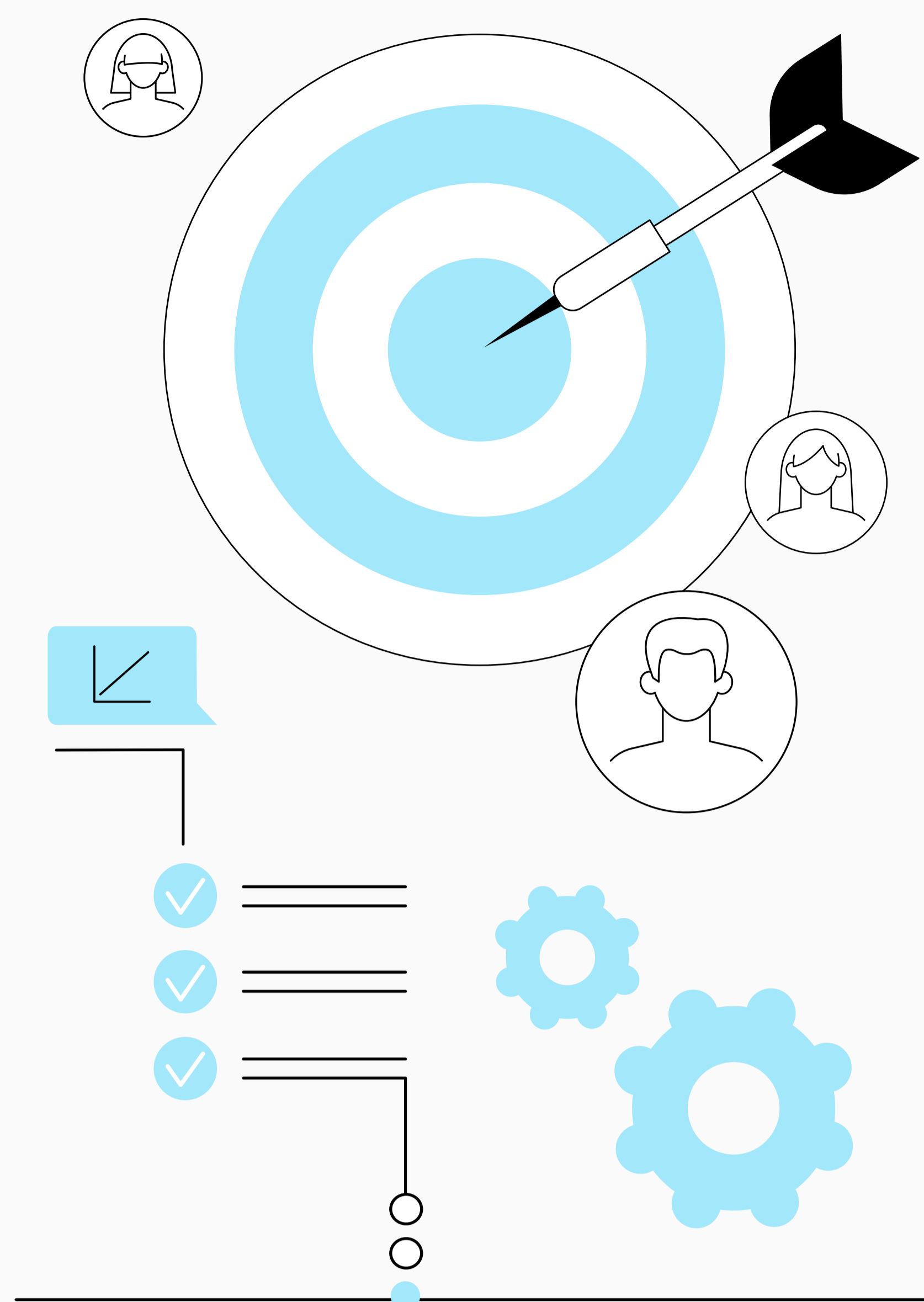
confidence

conscientiousness

responsibility

attitudes towards getting credit

attitudes towards repaying credit

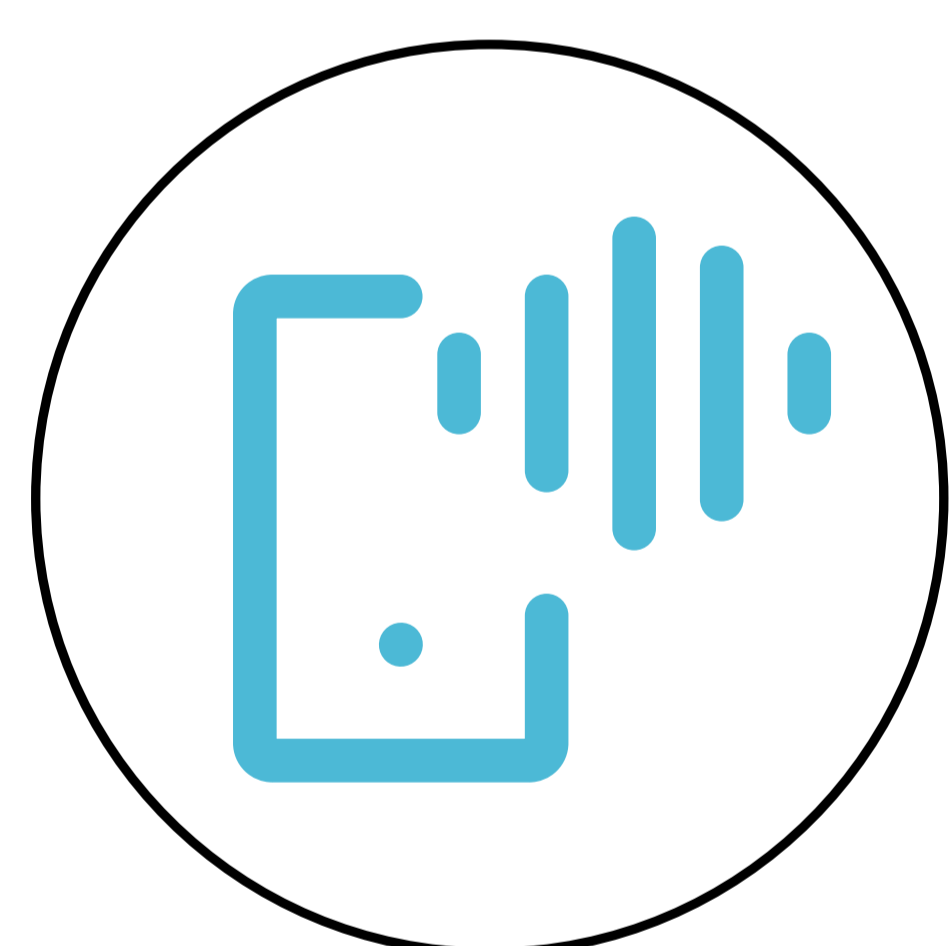


EXAMPLE

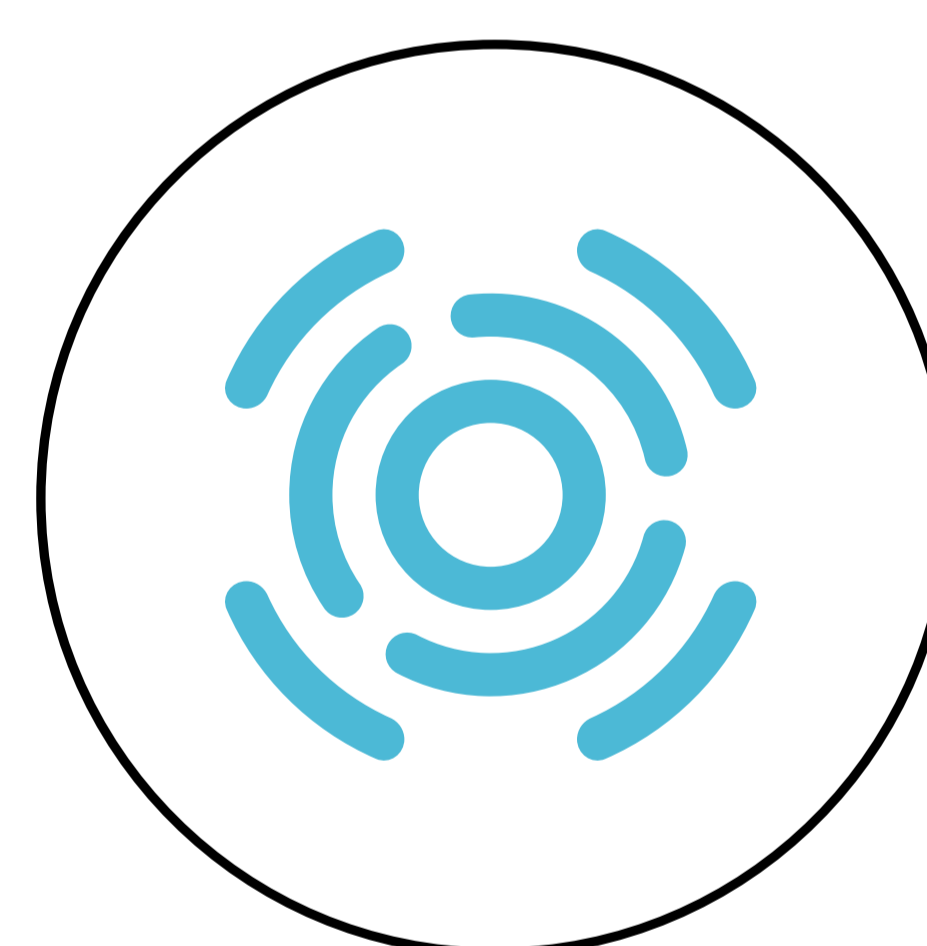
For example, using a subfield of psychometrics such as psychometric credit scoring to leverage non-financial data to capture crucial aspects of a borrower's character.

The assessment process is brought to a scientific and data-driven level by recognising the importance of traits such as responsibility, trustworthiness, and financial habits to predict a borrower's ability to repay loans. However, using psychometrics adds friction to the onboarding journey as applicants would need to answer 15-20 questions before a decision could be made

Software companies, like Credolab, provide behavioural data, including



**Behavioural
biometric
metadata**



**Device
fingerprint
metadata**

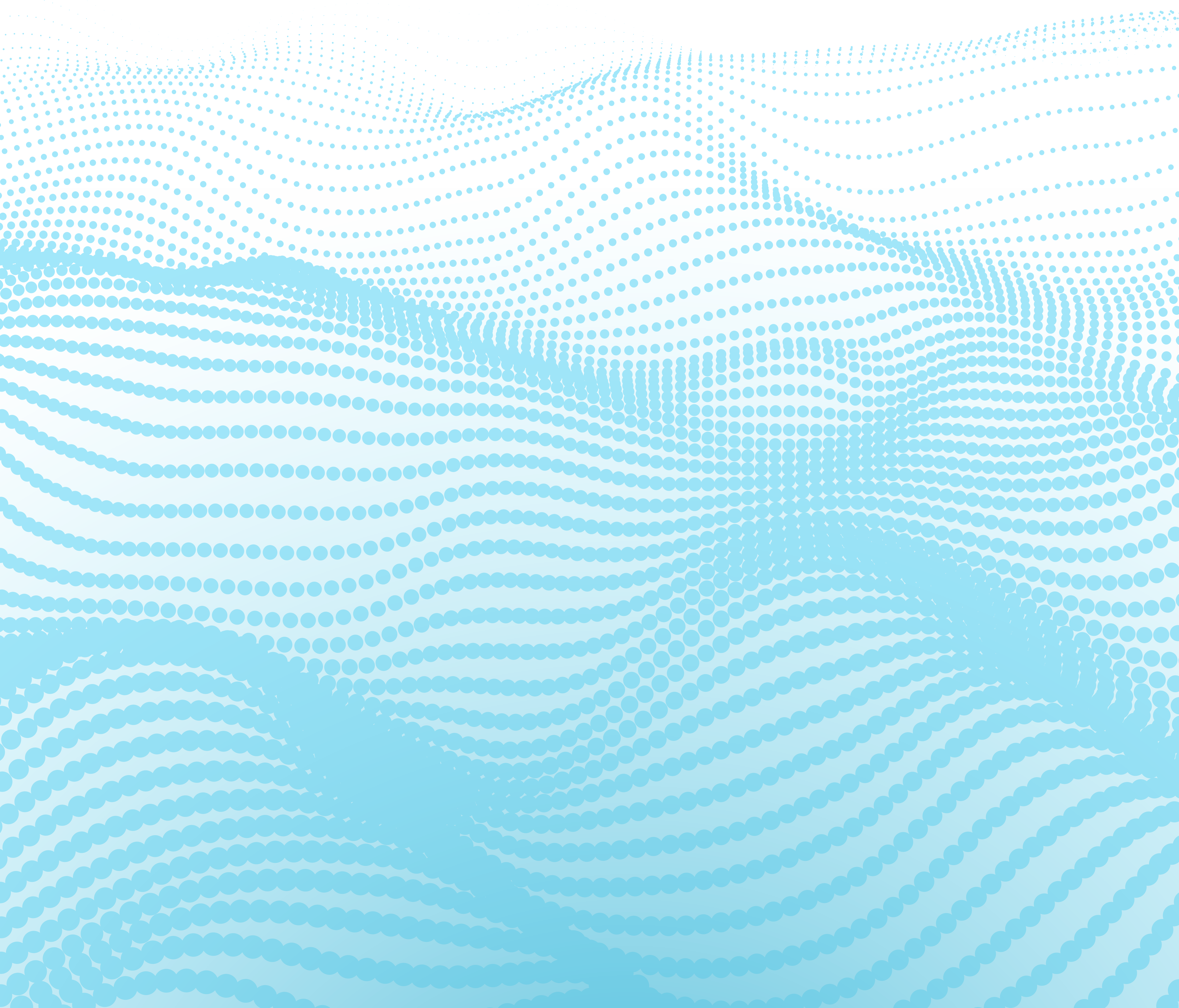
Behavioural data is another type of alternative data to understand a person's creditworthiness generated by, or in response to, a customer's engagement.

Behavioural analysis can identify through users' actions whether a person is who they say they are. As users interact with forms and applications, they leave behavioural traces that reflect their intent.

Analysing behaviour identifies patterns that can be used to detect suspicious behaviour and evaluate each user's digital experience.

ROLE OF CREDOLAB

Diving into behavioural data



Credolab is a powerful ally in alternative data, offering a complementary behavioural data and analytics solution that adds another layer of depth to gathering information and understanding users.

Specialising in transforming metadata into actionable insights and scores

up to

80k

data points

since January 2016

8 years

experience in understanding behavioural data

2.7k

scorecards developed in 2023 alone

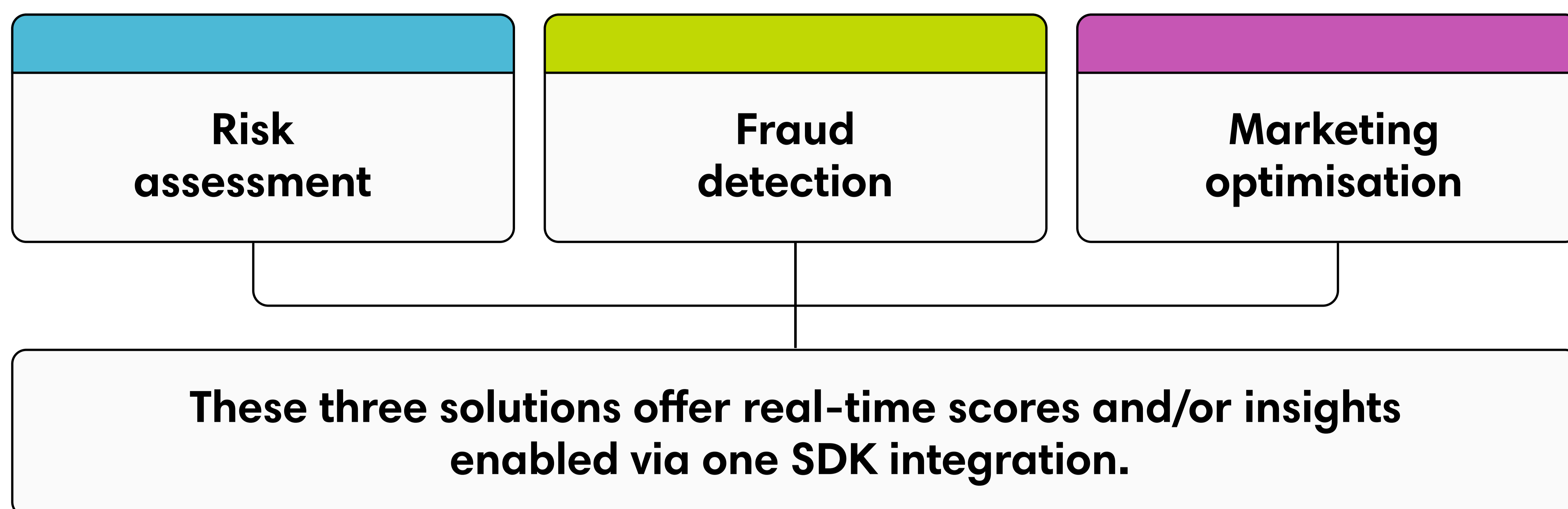
350 million

insights derived from analysing more than 350 million datasets globally

Using smartphone and web metadata, Credolab focuses specifically on **behavioural biometrics metadata and device fingerprint metadata**

Credolab utilises privacy-consented and permissioned data to deliver a holistic approach to credit risk assessment, fraud detection, and marketing optimisation, ensuring inclusivity and precision in diverse global markets. Our platform provides actionable insights and scores that offer contextual information about every mobile app and website user for improved data-driven decision-making.

Divided into three core solutions, Credolab empowers:



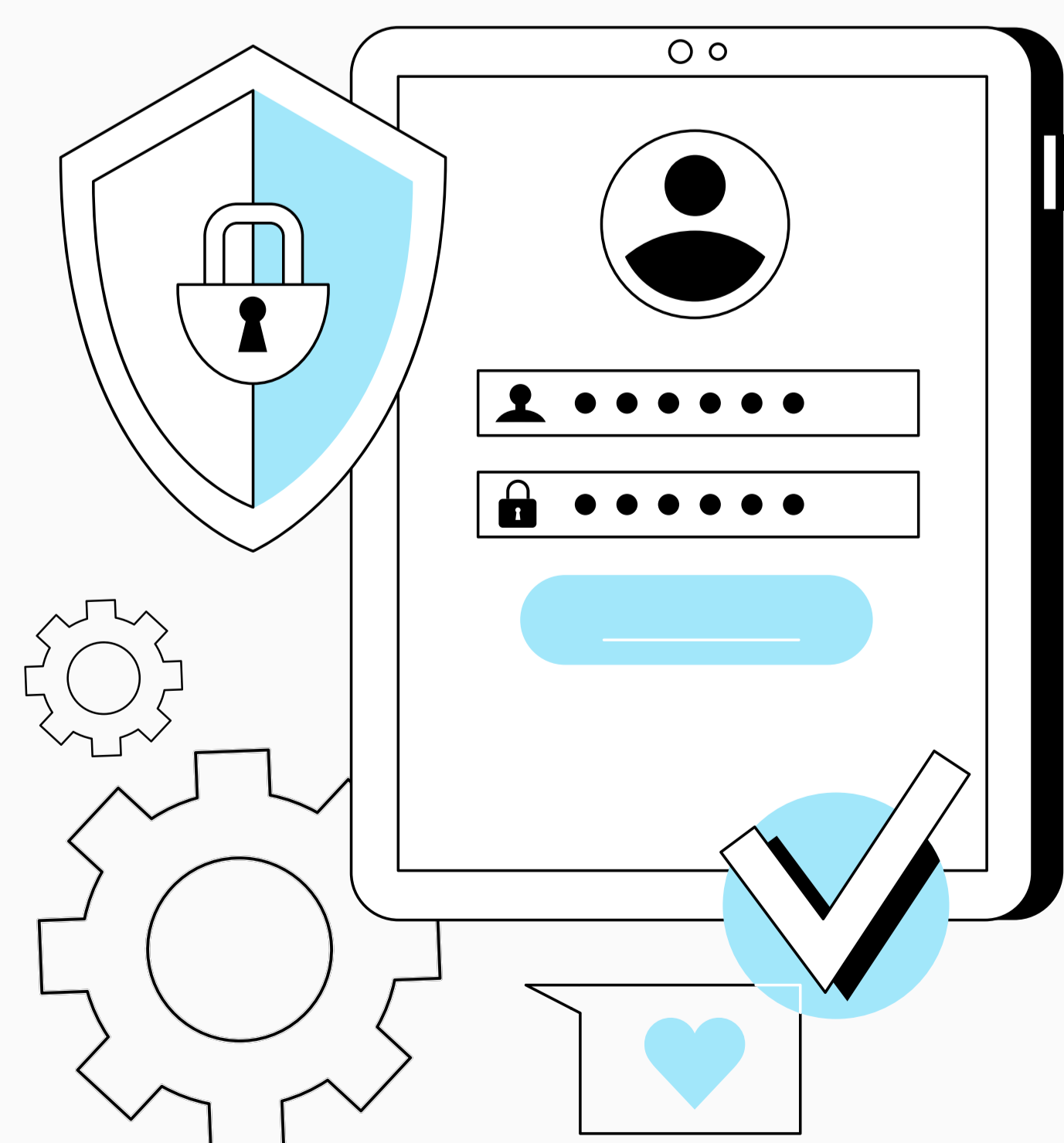
These insights from smartphone and web metadata can:

Detect repeat loan applications using the same device, browser, or IP address.

Dig up how often applicants change their key information like address or income.

Discover if the phone is locked or open to a global network.

Predict the lifespan of the mobile device and check how active its owner is based on the volume of media created and/or contacts saved over time.

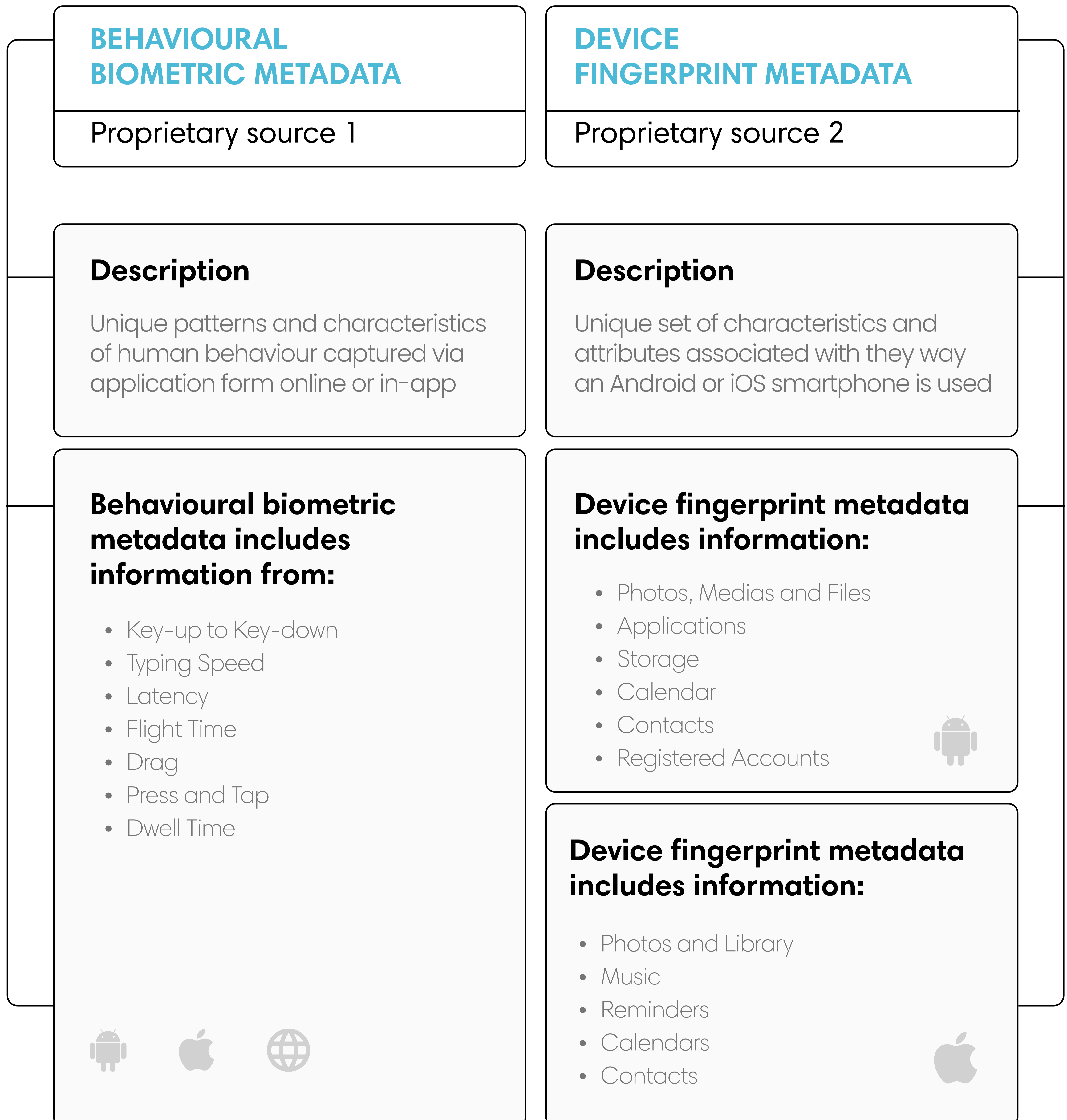


EXAMPLE

For example, Credolab can analyse and suggest specific information to perform checks for their underwriting strategy through behavioural data, creating a secure and effective user verification while exposing possible fraudulent behaviours.

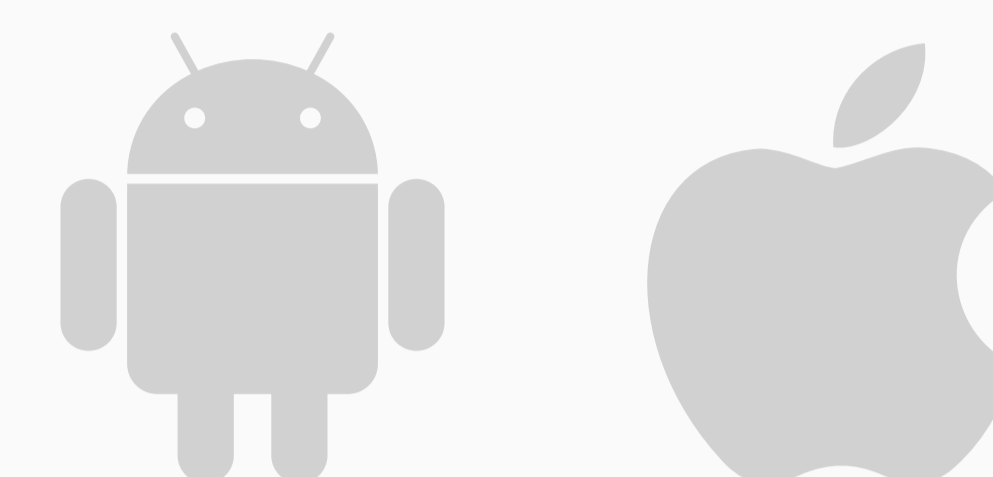
CREDOLAB'S PROPRIETARY SOURCES

Whether for risk assessment, fraud detection, or marketing segmentation, our SDKs collect only privacy-consented, anonymous, and depersonalised behavioural biometrics and device fingerprint metadata without any file or Personally Identifiable Information (P.I.I.) ever leaving a mobile app or website



Using these proprietary sources of metadata, Credolab evaluates risk, detects fraud and enriches marketing segmentations on:

iOS and Android



Credolab collects permissioned metadata across:



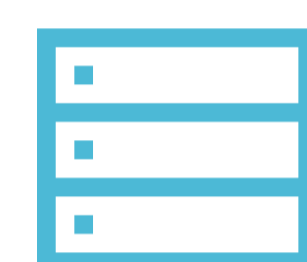
Device Information



External storage (media)



Registered Accounts



Internal storage (media)



Contacts



Application



Calendars



User Interface interactions

This also includes the total time spent applying for a loan or credit card, the time spent in the same position, latency and keystroke patterns.

Web



Credolab collects:



Keystroke Patterns



Browser Information



Language and Operating System metadata



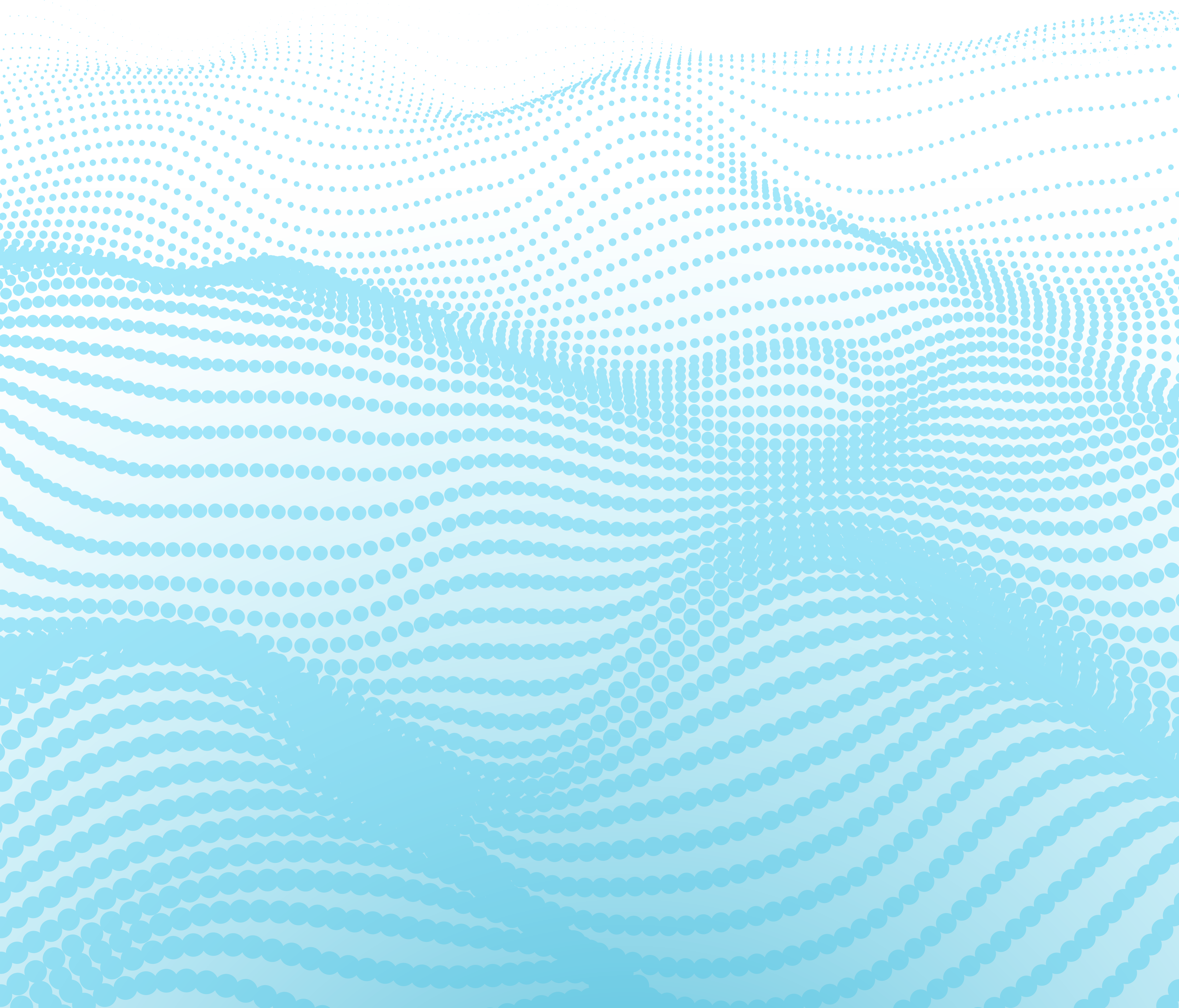
User Interface interactions

This also includes the total time spent applying for a loan or a credit card, the time spent in the same position, Latency, and Keystroke patterns.

PART 1

ASSESSING DATA QUALITY

What makes a great
alternative data source?



The value of alternative data varies by source. Hence, it is imperative not to rely on one type of alternative data.

The Introduction explained how alternative data transcends traditional risk assessment frameworks by weaving together a rich tapestry of digital interactions and behaviours. However, each of these sources varies in their own way.

EXAMPLE

For example, data like rental payments are predictive and may be available to many consumers with no credit file. However, the rental market is very fragmented, and data is not uniformly reported. Therefore, coverage could be higher.

Another set of data, such as public records data, is available to even more people than those with files at the credit bureaus, so their coverage is very wide.

However, since the information in public records is not explicitly focused on payments, it is less predictive than credit bureau data. Moreover, it also proves to be valuable in scores developed from combined data sources.

| Type of data | Rental Payments | Public Records Data | Credit Bureau data |
|--------------|---|---|--|
| Predictivity | High | Low | High |
| Availability | To many consumers with no credit file | To even more people than those with files at the credit bureaus | To people with files at the credit bureaus |
| Status | Fragmented market and varied data reports | Not explicitly focused on payments and valuable in scores | Explicitly focused on payments |
| Coverage | Limited | Wide | Limited |

Alternative data has a wide range of value, so to evaluate what constitutes a good alternative data source, here are five main characteristics based on Oliver Wyman's framework:



Coverage

A new or alternative data source will ideally have broad and consistent coverage.



Specificity

A data source should contain detailed data elements about an individual—data elements that provide part of a full picture of the borrower.



Accuracy and timeliness

The data must be accurate, current, or frequently updated. A data source should have a system for ongoing data verification and management.



Predictive power ('signal')

The data should contain information relevant to the behaviour being predicted.



Orthogonality

Alternative credit data should complement traditional credit bureau data.



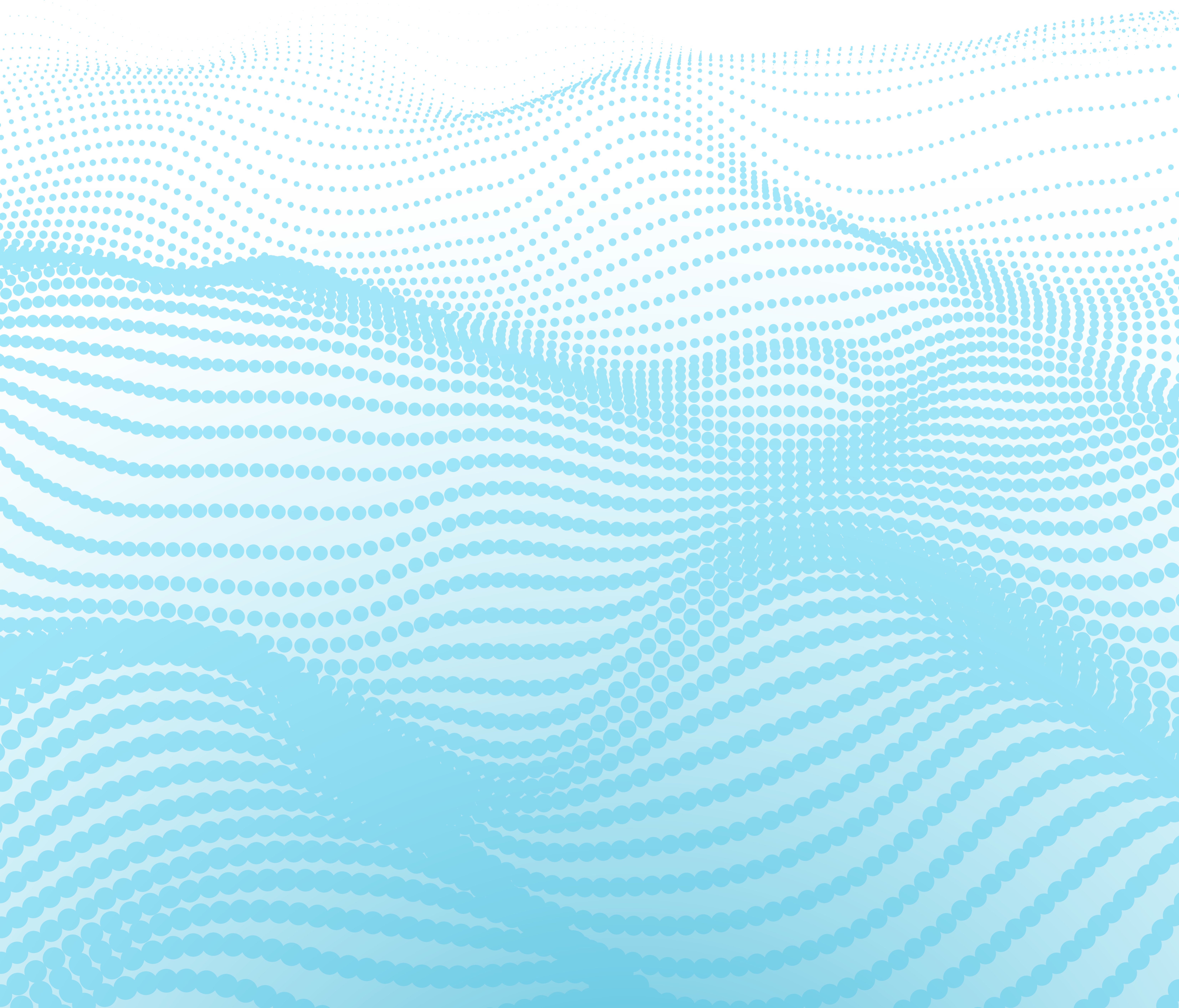
Regulatory compliance

Data sources must comply with existing financial regulations.

PART 2

ASSESSING DATA QUALITY

How is Credolab a good alternative data source?



Credolab leverages a combination of supervised and unsupervised machine learning algorithms, to analyse and predict consumer behaviour based on smartphone and web metadata.

Using the framework by Oliver Wyman, as explained in Part 1, here is a breakdown of the characteristics to assess the quality of Credolab as a good source of alternative data.



Coverage

Credolab allows the collection of rich insights for 100% of any app or web application user. We have developed our mobile SDKs to mirror the businesses' app's permissions on Google Play and the App Store and will put the metadata to work.



Specificity

Behavioural biometrics (e.g., typing speed/cadence, UI interactions, mouse/finger gestures, etc.) and device fingerprint metadata (e.g., app installation patterns, Operating System usage, audio/images/video creation patterns, etc.) while accessed in an anonymous and depersonalised way, is always and only about the individual user.



Accuracy and timeliness

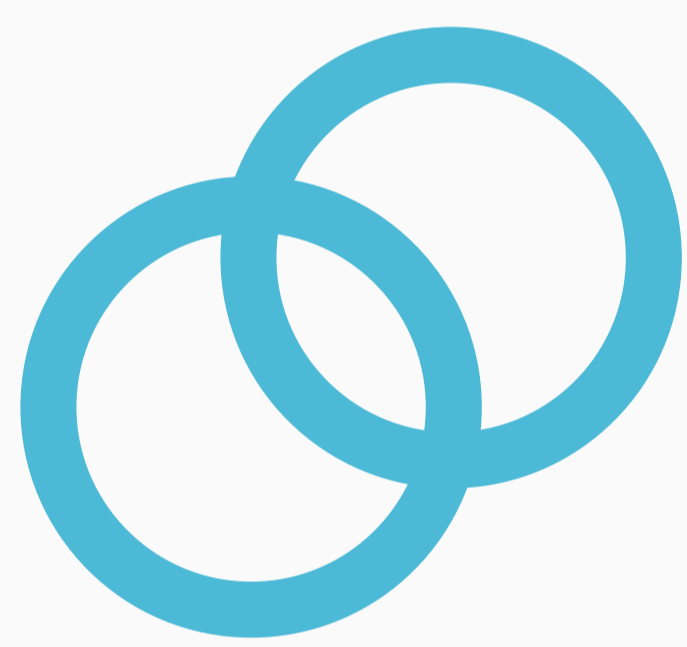
Credolab metadata is collected contextually, at the point of application, and automatically whenever the user applies for a loan or businesses themselves decide to trigger the SDK.



Predictive power

Credolab's digital scorecards are built on credit outcome data and tailor-made for each client. The performance of the Credolab expert scorecard is illustrated by region and the Gini Coefficient in the table below:

| Distribution of Gini Coefficient | Global | NAM | LATAM | EMEA | APAC |
|----------------------------------|--------|-------|-------|-------|-------|
| 0.00 - 0.10 | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| 0.10 - 0.15 | 5.8% | 7.1% | 8.8% | 7.2% | 0.0% |
| 0.15 - 0.20 | 19.4% | 14.3% | 18.4% | 21.6% | 23.2% |
| 0.20 - 0.25 | 21.1% | 23.1% | 19.2% | 20.4% | 21.8% |
| 0.25 - 0.30 | 29.0% | 34.2% | 25.5% | 30.1% | 26.1% |
| 0.30+ | 24.8% | 21.3% | 28.1% | 20.7% | 28.9% |



Orthogonality

There is a very low correlation between traditional credit bureau data (transactional in nature) and Credolab's metadata (behavioural in nature). Using both data sources, businesses can assess a customer's ability to repay a credit product and their willingness to repay.



Regulatory compliance

Credolab accesses only first party and privacy-consented behavioural biometrics and device fingerprint metadata. Users are requested to give consent to access data and made aware of exactly what data we access, why we access data, and how we process it (in an anonymous way; only businesses know the true identity of the user, not Credolab).