oxylabs®

The Growing Importance of Alternative Data in the Finance Industry

Introduction

Today's global financial markets are more volatile, dynamic and fast-moving than ever before. In such an environment, gaining an edge against the competition requires finding a unique approach. Thus, an increasing number of financial services companies are recognising the value of new, non-traditional and alternative sources of business and market data.

For the financial services market, alternative data sources refer to previously undiscovered data beyond the traditional company filings, financial statements, PR or management information, and other legacy datasets. Additionally, the insights gleaned from alternative data may be more resistant to value decay than ones gained from traditional sets.

Alternative data is, in general, any type of data that has not traditionally been used by financial services businesses. Data that is of value to the financial sector includes app download and usage info, social media sentiment data, website traffic information, search data and other non-traditional data sources.

Thus, investors and money managers are looking for new sources of information to generate higher returns for their clients, whilst retaining current risk management practices. Therefore, it's not entirely surprising that the global alternative data market was valued at $1.06 billion back in 2019. It's expected to grow at a compound annual growth rate (CAGR) of 40.1% from 2020 to 2027, according to a report published in late 2020 by Grand View Research.[1] Web scraping and financial transactions are the most common sources of alternative data for investors and financial services firms.

[1] https://www.grandviewresearch.com/industry-analysis/alternative-data-market

Harnessing alternative data yields valuable, previously unseen insights into business performance, market trends and future investment opportunities. Put simply, as financial services companies are able to generate, process and rapidly analyse real-time alternative data into actionable business insights, their effectiveness rises in tandem. Highly data-driven organisations are three times more likely to report significant improvements in decision-making, according to a recent PwC survey of more than 1,000 senior executives.[2]

This report, based on an Oxylabs survey of 252 senior data decision-makers from across the UK's thriving financial services sector, reveals the current trends of the industry with a focus on data management practices.

Julius Černiauskas
Chief Executive Officer at Oxylabs.io

" **Alternative data has been touted as revolutionary. To see how businesses at large understand the value of information, we decided to look deeper into the use and trends of alternative data in this industry in order to provide actionable insights to our readers. Our goal has been to categorise the value brought by alternative data and, at the same time, understand why some companies might be avoiding its implementation.**

---

[2] https://www.pwc.com/us/en/services/consulting/analytics.html

Survey methodology

Oxylabs, in cooperation with Censuswide, conducted a survey of 252 senior data decision-makers from leading UK-based financial services companies to gather valuable insights into the finance sector landscape when it comes to alternative data.

Developing faster, more efficient ways of gathering and analysing real-time data is a necessity for financial services businesses. Web scraping has thus become increasingly important for the sector and, for many companies, is now the primary method to obtain alternative data for analysis and insight.

At Oxylabs, we have seen a surge in inquiries from companies within the financial sphere. According to our calculations, the amount of inquiries has increased almost threefold over the past year, indicating a quickly rising interest in alternative data acquisition.

With this survey, we wanted to uncover insights into the current state of the financial sector, the most valuable data sources, budgets allocated to data acquisition, and whether COVID or Brexit had any impact. Additionally, we looked into the major challenges and struggles that companies face when dealing with data acquisition practices. We also wanted to look into why some companies still choose not to take part in it, hopefully addressing a number of misconceptions and stigmas around alternative data procurement in the financial services market.
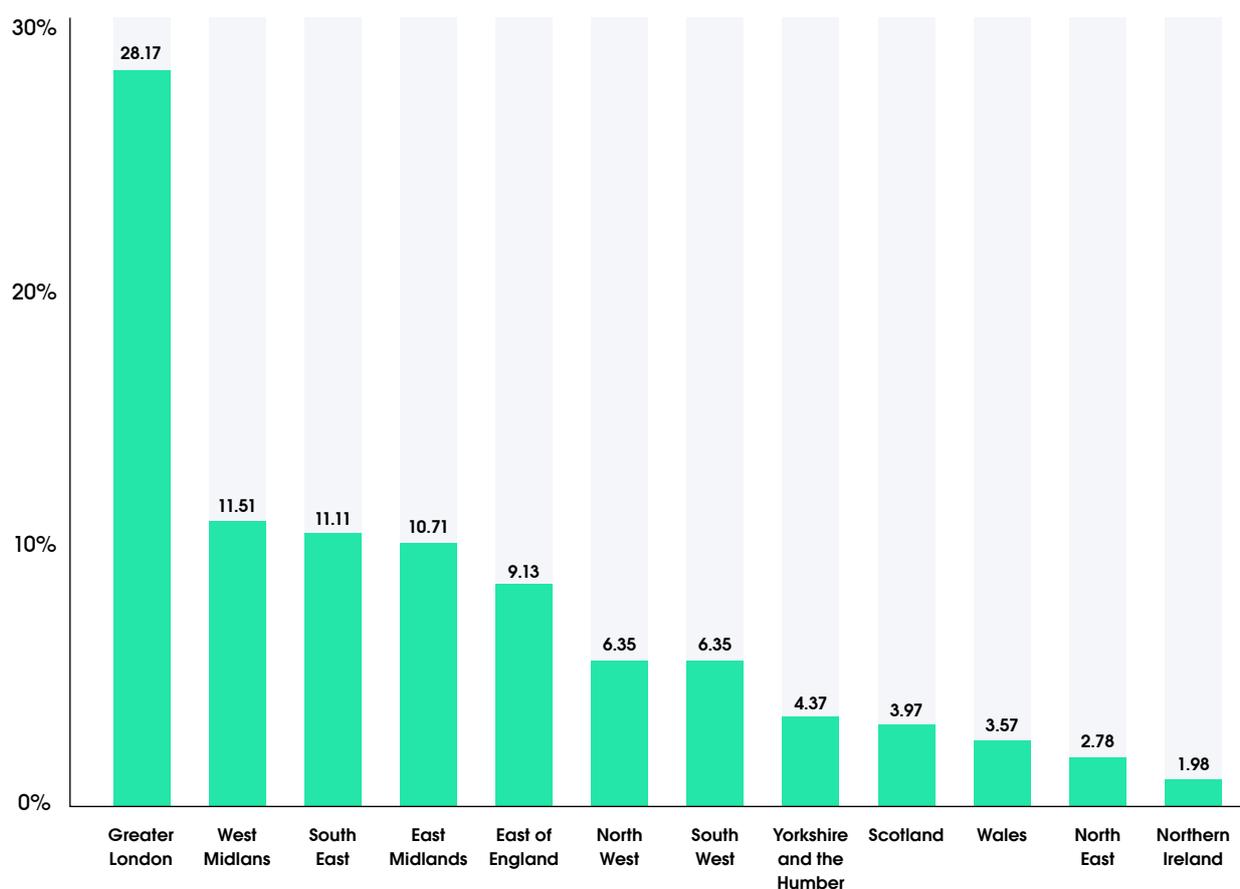
This report presents the key findings from our survey. These findings will help you gain a much clearer understanding of the key alternative data challenges ahead for today's Directors of Engineering, CTOs, VPs of Engineering, Chief Data Officers, Heads of Business Intelligence and Heads of Data & Analytics across the UK's finance sector.

Our market research survey has addressed these burning questions within the financial sector:
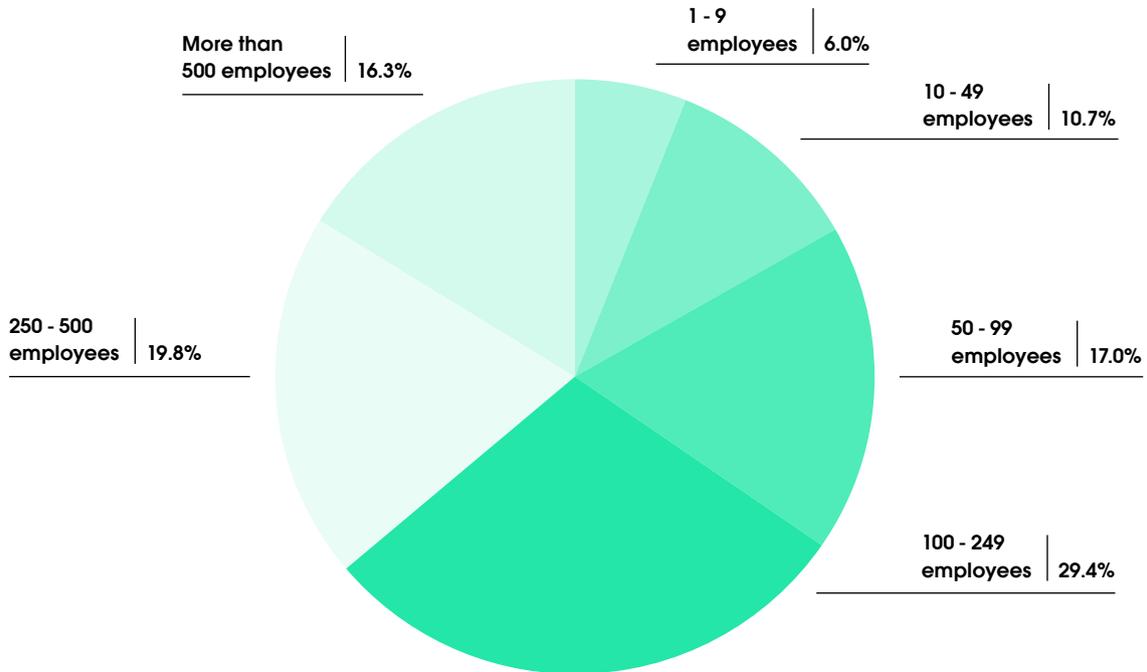
- What alternative data sources are bringing the most value?
- How reliant are businesses in the UK's financial sector on data?
- What methods (or combinations of methods) are used by the most prominent finance companies in the UK to acquire the alternative data they need to aid and improve their decision making?
- Taking into account today's turbulent economic times, what is the pulse regarding dedicated data budgets?

Survey sampling

Regions

Company size

More than 500 employees	16.3%
1 - 9 employees	6.0%
10 - 49 employees	10.7%
50 - 99 employees	17.0%
100 - 249 employees	29.4%
250 - 500 employees	19.8%

Key trends and insights

- **63%** of respondents have begun using **alternative data sources** as a way to improve decision making.

- **26%** of respondents have noted that their **data needs have increased** by a significant margin in the **past 12 months.**

- **38%** of respondents have already **increased data department budgets.** Nearly half of all respondents are **planning to increase their allotted budgets** while the other half is planning to at least keep them identical.

- **60%** of companies use integrations with **third-party databases.**

- There is an equal split among those who **outsource** (36%) and have **in-house scraping teams (36%).** In addition, **27%** do both.

- Of those who have implemented automated web scraping, **42%** of respondents **struggle** to ensure a **consistent flow of data**, making it the primary challenge. For those outsourcing data acquisition, finding **reliable partners and tools** (**36%** and **35%**, respectively) is one of the most common **challenges**.

- Most businesses would see significant **negative impact if data flow were interrupted for up to 2 days.** A minority **(4%)** would start seeing negative effects within **1 hour.**

- **36%** of respondents list high **risks of legal complications** as their primary concern.

- Of those who have not implemented automated web scraping, a significant amount **(39%)** of respondents state the primary reason they have yet to implement web scraping are **budget constraints.**

Julius Černiauskas
Chief Executive Officer at Oxylabs.io

" **The landscape of data has changed. While traditional sources maintain their relevance, nowadays few businesses rely exclusively on them. As external data acquisition becomes increasingly important, especially in the wake of the global pandemic, businesses are expecting to increase their scraping-related budgets. However, we do see that web scraping retains some of its negative light, such as the unclear legal status and high costs associated with data acquisition.**

Alternative data is vital

Traditional sources of data such as official public data (57%) and third-party data (44%) remain highly relevant and useful for the financial sector. However, respondents from our survey indicate that the majority (63%) are now using alternative data when making or helping clients make investment decisions.

Alternative data can be used to acquire previously unavailable signals for market trends and price changes. In-depth analysis through several different data points can reveal additional support for investment hypotheses. In its essence, alternative data grants the finance sector another powerful tool for predictive analytics.


Gediminas Rickevičius
Director of Strategic Partnerships at Oxylabs.io

Alternative data has revealed an interesting way to interpret the Efficient Market Hypothesis. We see that although all alternative data is public, markets don't adjust to the information effectively, making consistent alpha generation possible. It might be due to the fact that the signals are not as straightforward as those from traditional data sources. Thus, traditional sources are more susceptible to the decay of value than alternative ones."

Q1. Which type(s) of data, if any, do you currently utilize when making or helping your clients make investment decisions? (Tick all that apply)
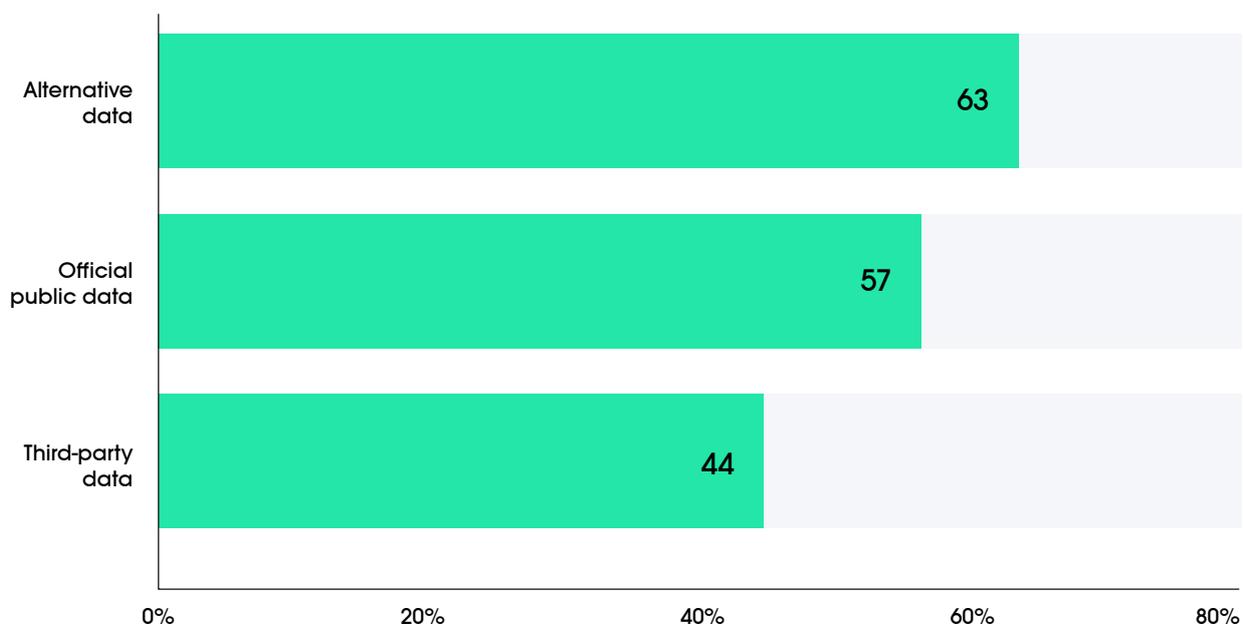
Alternative data - non-traditional data sources (e.g. senior staff hirings, corporate flight activity, supply chain data, etc.)

Official public data - traditional data provided by official filings, government entities, etc.

Third-party data - 3rd party data aggregators, market research & statistical databases (e.g. Bloomberg, Refinitiv, S&P Capital IQ platform, etc.)

Finance doubles down on data

Alternative real-time data is crucial for financial companies to make investment decisions. As we can see from our survey results, over a quarter (26%) of respondents in the financial sector note their data needs have increased significantly over the last year.

Clearly, the financial sector has doubled down on its data needs throughout the course of the last 12 months. There are numerous reasons for this, however, the most important we believe is the rapidly increasing amount of useful information scattered across the web. Proper implementations of data acquisition tools can help financial companies make greater decisions faster and more efficiently than ever before.

Tomas Montvilas
Chief Commercial Officer at Oxylabs.io

COVID-19 has brought with it a business migration into the digital landscape. As more day-to-day processes have become digitised, more actionable and important business-related data becomes publicly available. Implementing ways to gather such data provides previously inaccessible intelligence. Such data is twice as important for financial businesses - any information left by companies can be potentially used to generate ROI.

Q2. Have your data needs increased, or decreased over past 12 months?
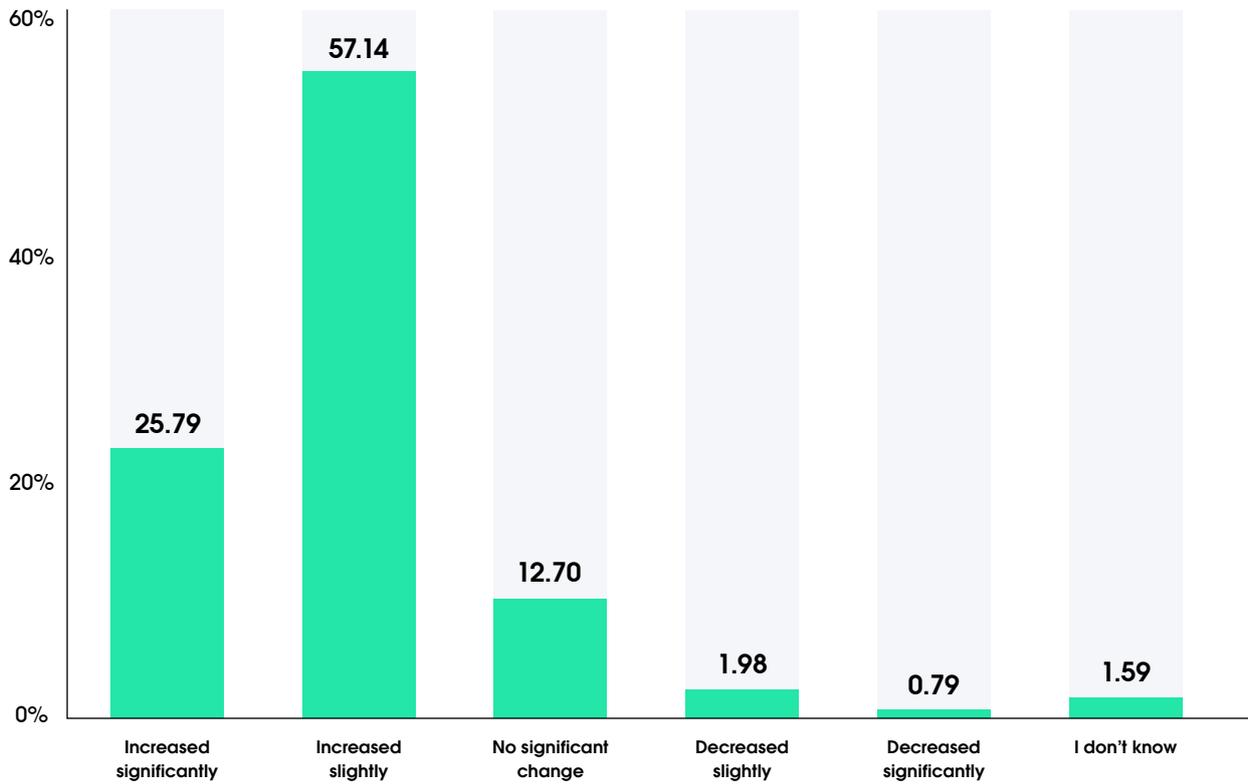
Alternative data needs of companies in the finance sector are clearly growing, a trend that is borne out by the increase in data collection budgets allocated by companies over the last year. Regarding financial commitments, a significant proportion of respondents in the finance sector (38%) have increased those budgets over the past 12 months.

Q3. Did budgets allocated for data collection & management increase or decrease in the past 12 months?
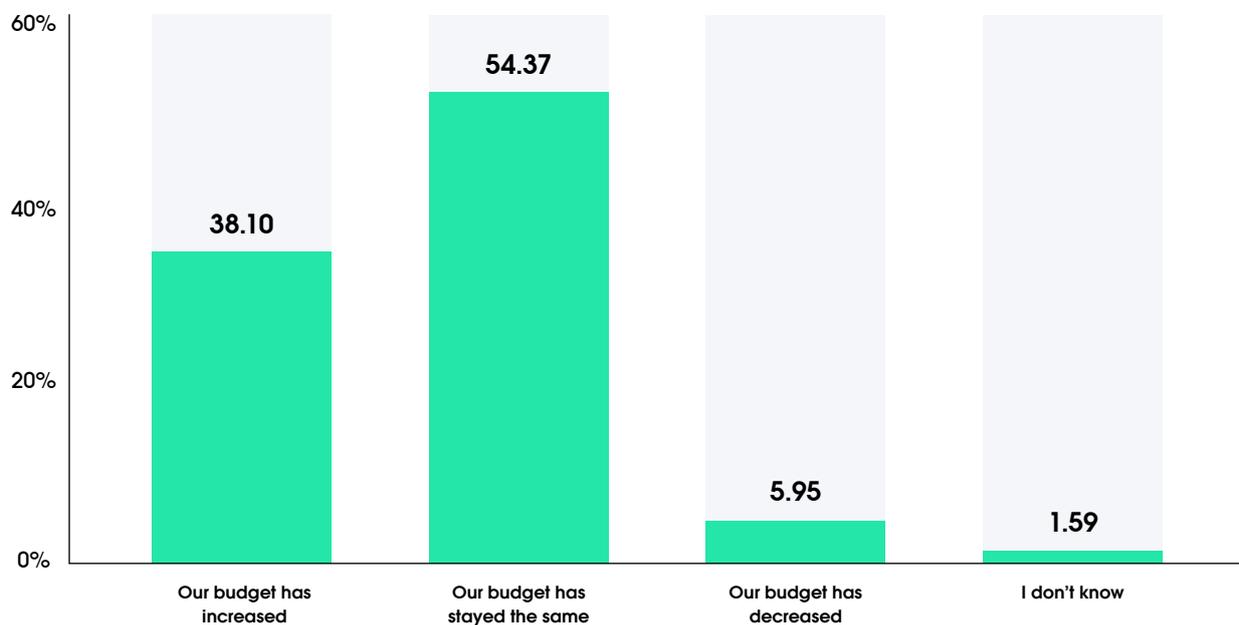
Additionally, the vast majority of respondents are either planning to increase their budgets (43%) or at least keep the same budgets (45%) over the next 12 months. The trend is clear: companies are going to be increasingly reliant on data to help them make business decisions.

Q4. Are you planning to increase or decrease budgets data collection & management in the next 12 moths?
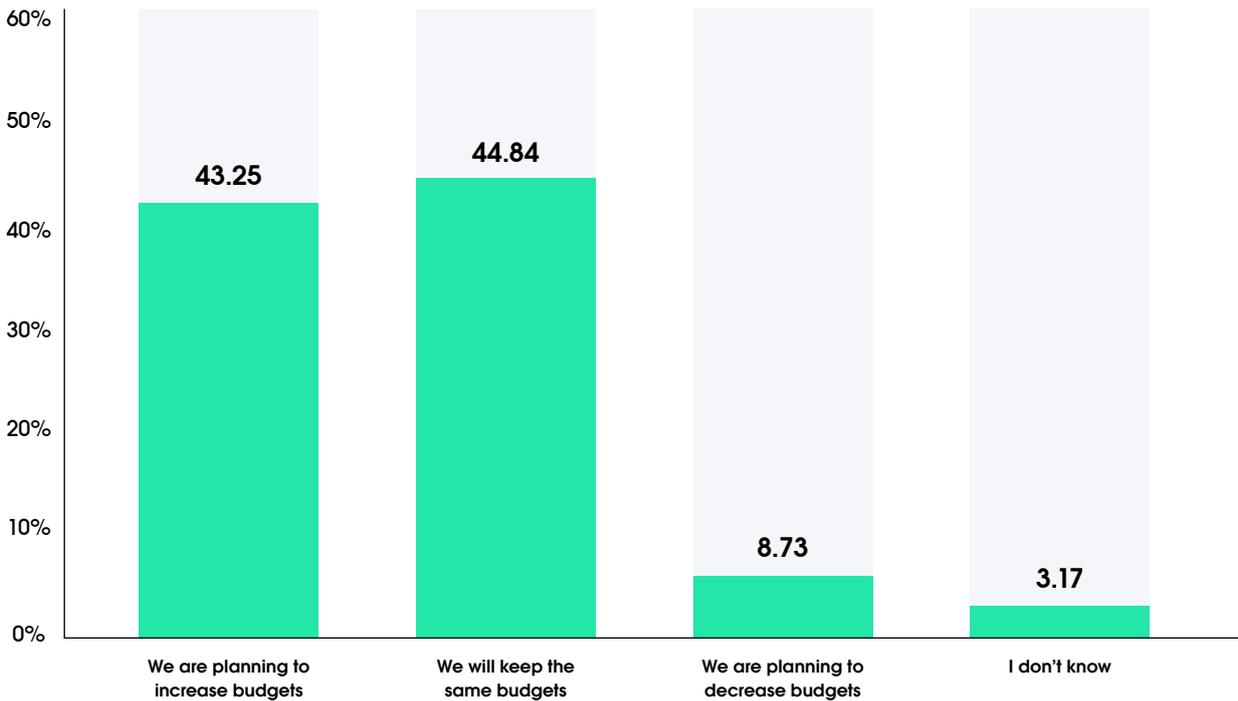
Almost 9% of respondents indicated they are planning to decrease their budgets allocated to data collection and management. However, that is most likely due to post-pandemic economic pressures. For these companies it's going to be far more difficult to remain competitive with many other companies in the finance sector strengthening their commitment to data.

Value of automating data collection

Q5. What methods, if any, do you currently use to collect data? (Tick all that apply)

Method	Value
Integrations eith 3rd party databases	59.52
Automated ETL processes, web scraping pipelines	52.38
Manual data collection & cleaning	51.59
Don't know	1.59
We don't currently use any methods to collect data	1.19
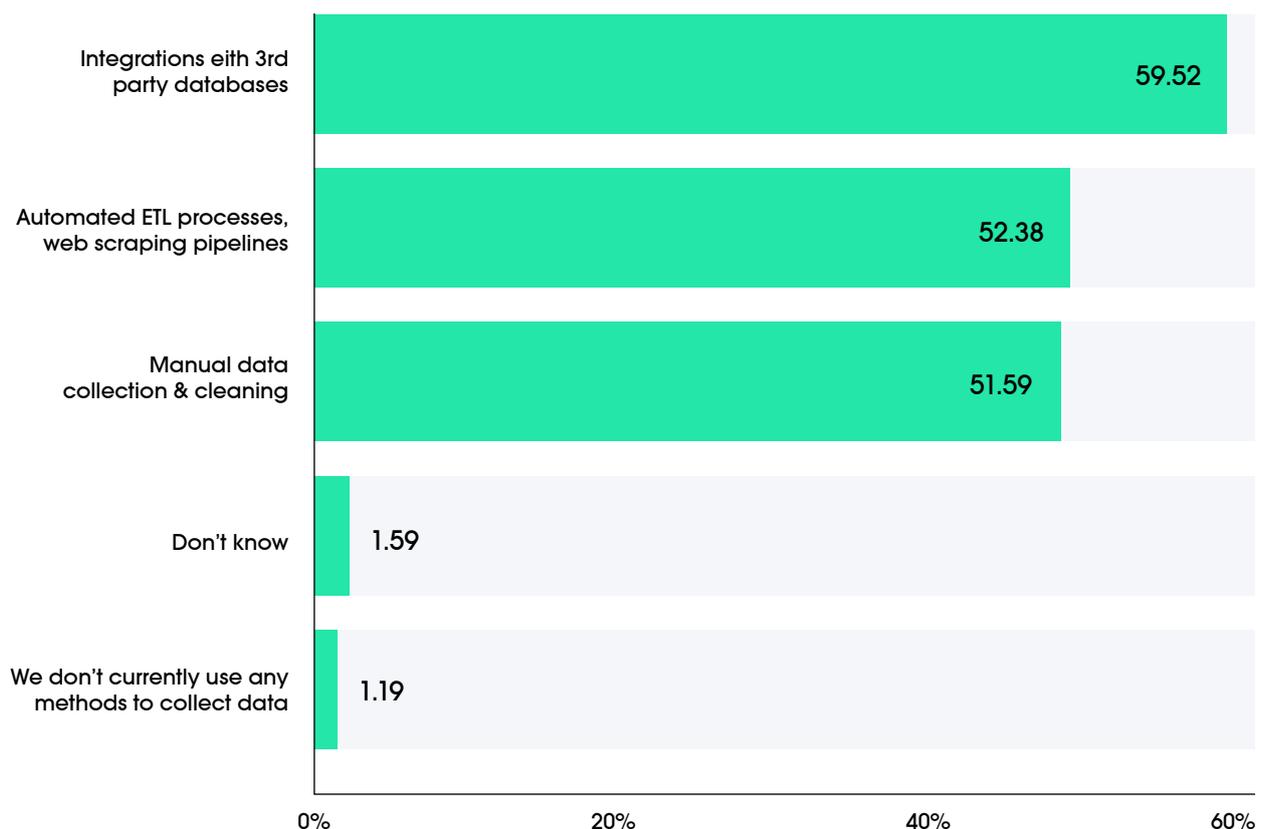
Companies in the finance sector are using numerous ways to collect data. We investigated what combinations of methods were the most popular. A majority of respondents (60%) said they use integrations with third-party databases as one of the primary sources.

In addition, over half (52%) of all respondents are reliant on an automated Extract, Transform, Load (ETL) framework. ETL is a process that takes raw data, extracts the information required for analysis, transforms it into a format that can serve business needs, and loads it into a data warehouse. This typically summarises data to reduce its size and improve performance for specific types of analysis.

However, a high proportion of respondents (52%) are still looking into manual data collection and cleaning as one of their methods. Manual data collection and cleaning is extremely ineffective when the ease and simplicity of automation is taken into account. Put bluntly, in today's digital economy, it's a waste of resources and labour to not use the latest web-scraping tools and not to automate data collection processes.

Survey results highlight that those companies not yet utilising ETL processes or web scraping pipelines are lagging behind the industry as a majority have begun using data-driven decision making. Thus, we highly recommend considering implementing data automation and governance processes.

Tomas Montvilas
Chief Commercial Officer at Oxylabs.io

"

Data is only as good as the actions taken from the signals it created. It's not enough to simply analyse data and create some insights that might or might not be profitable. These insights and their grounding must fall into the hands of the right people at the right time. Proper data governance strategy execution ensures that no information is wasted.

Outsourcing data collection

Q5a. Who is responsible for your web scraping activities in the organisation?
*Respondents who rely on automated ETL processes, web scraping pipelines

Category	Value
In-house team	36.36
Outsourced web scraping team or partners	36.36
Both	27.27
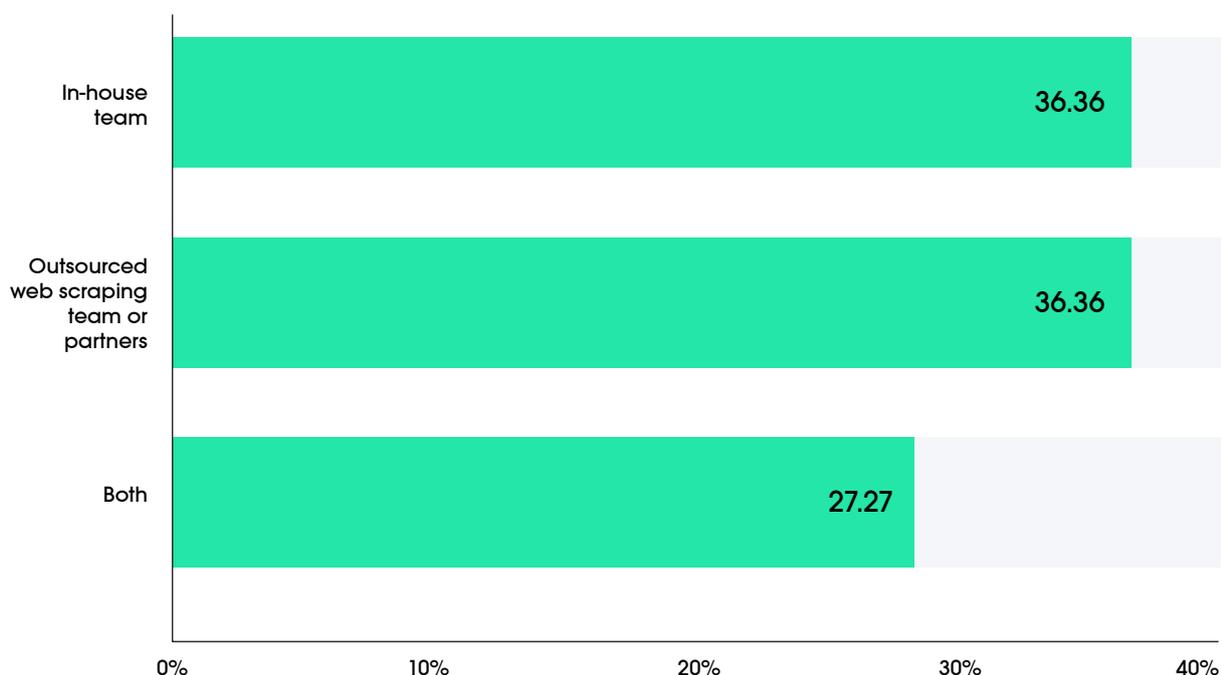
Our survey results reveal that there is an equal split among respondents outsourcing (36%) and doing in-house scraping (36%), with a smaller number (27%) using both methods. Larger companies (250+ employees) seem to have a slight preference for the latter. They likely already have their own developers, system administrators, data analysts, and other data specialists, easing the development process.

Larger companies collect data in-house due to the increased flexibility and adaptability. Those companies only require supporting tools, such as proxies, to perform their data gathering operations.

However, running data acquisition processes in-house has the drawback of being costly and difficult to maintain. Teams of developers need to be consistently dedicated to the technical aspect of data acquisition which is part of the reason why some companies decide to outsource their web scraping departments.

On the other hand, it's clear that nowadays, it is easier than ever to outsource ready-to-use solutions. Tools like Real-Time Crawler allow businesses to send a simple HTTP request and retrieve the information in several file formats. Using these solutions greatly reduces the initial costs and overall management burden.

Juras Juršėnas
Chief Operating Officer at Oxylabs.io

Data collection and integration processes produce significant amounts of value to any business that can utilise the additional information. However, higher entry costs, driven by building an in-house team might dissuade some from the pursuit of being data-driven. For those that are having doubts whether data can be of use, finding a third-party web scraping solution provider is the easier option. Comparatively, there are very little startup costs and data can be integrated nearly instantly.

Overcoming data collection challenges

Q5b. What are the challenges you face in your web scraping activities, if any? (Tick all that apply) *Respondents who rely on automated ETL processes, web scraping pipelines

Challenge	%
Ensuring consistent data flow	41.67
Managing and processing large datasets	38.64
Getting data in real-time	37.12
Finding reliable partners to outsource activities	35.61
Ensuring collected data quality	35.61
Finding most efficient tools or tech stack for data acquisition	34.85
Finding talent to develop and manage data collection processes	34.85
There are no challenges we face in our web scraping activities	0.76

Web scraping is tough. However, once everything sets into place, vast amounts of data insights become available and allow businesses to gain a competitive edge. Yet, despite all the benefits and hype surrounding data acquisition, all respondents in our study indicate they are still facing a wide range of issues and challenges with their web scraping activities. For example, 42% of respondents struggle to ensure a consistent data flow, an issue most often faced by in-house data teams.

Collecting data is only the first step. The next step is to clean the data, make it readable, and implement running algorithms to extract insights. According to our survey, 39% of respondents still find managing and processing large datasets to be challenging.

To make correct investments and respond in real-time to the fluctuations of the global stock market, companies must ensure they have a consistent real-time data flow. A third (37%) of respondents indicated that getting data in real-time was still a challenge they faced in their web scraping activities.

Finally, finding the best data talent is an ongoing issue for a third (35%) of financial businesses investing in web scraping activities. In addition, a third (36%) of respondents indicated that they found it challenging to find reliable partners to outsource web scraping activities to. While web scraping is more accessible than ever, it seems there is a lot left to be desired by the providers. Many of the challenges outlined by our respondents correlate with low service quality as they struggle with finding reliable partners and tools (36% and 35% respectively).

Aleksandras Šulženko
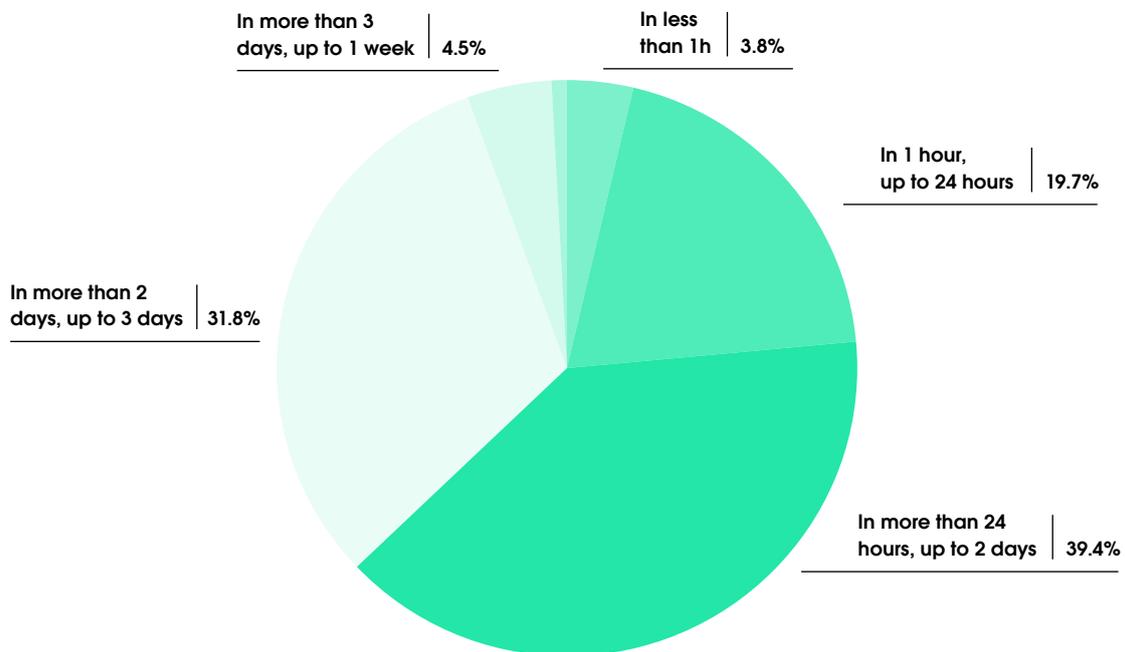Product Owner at Oxylabs.io

" **Maintaining functionality and delivering high quality data consistently takes a high degree of effort. Dedicating so many resources to in-house data acquisition may not be within the grasp of every business. Yet, those who can will outperform those who cannot. While outsourcing data acquisition reduces flexibility and adaptability, the lower upfront costs allow businesses of all sizes to derive insights from previously unattainable information. Finding the correct web scraping partner may be difficult, however, the benefits will always outweigh the risks in the long-run.**

Ensuring an uninterrupted flow of data

Q5c. In the event of a data flow interruption from your web scraping activities, how soon does this start impacting your core business processes, if at all?

*Respondents who rely on automated ETL processes, web scraping pipelines

In more than 3 days, up to 1 week | 4.5%

In less than 1h | 3.8%

In 1 hour, up to 24 hours | 19.7%

In more than 2 days, up to 3 days | 31.8%

In more than 24 hours, up to 2 days | 39.4%

How soon will an interruption in data flow impact on the core business operations of financial businesses? Real-time data flow is vital for a significant proportion of the financial companies in our study, with nearly a quarter (24%) noting that any interruption in data flow within a 24-hour period will rapidly impact their core business processes. A small amount (4%) of companies state that issues would begin to arise within one hour.

Tomas Montvilas
Chief Commercial Officer at Oxylabs.io

"**Real-time data has earned its place as one of the most valuable assets in any data-driven company. Ensuring an uninterrupted flow of data is the prerequisite for certain business and pricing intelligence use cases. Businesses in the financial sector rely on reacting faster than other market players in order to maximise ROI. Real-time data becomes an irreplaceable aid whenever any financial decision needs to be made.**

Perceptions of automated web scraping

While many surveyed financial companies already use automated data acquisition, others are yet to utilise it. There are several reasons for this. A significant amount of respondents (39%) indicated that budget constraints are the primary reason they have yet to implement automated web scraping and cleaning. While significant advancements are being made through the appearance of out-of-the-box-solutions, the upfront integration costs and lack of functionalities still remain above the bar for some.

Additionally, the stigma that web scraping carries a high risk of legal complications remains - 36% of respondents in the financial sector listing that as a major concern. Oxylabs have frequently attempted to educate burgeoning data-driven companies about the legal intricacies of web scraping, however, clearly the message needs to be spread wider and with more clarity.

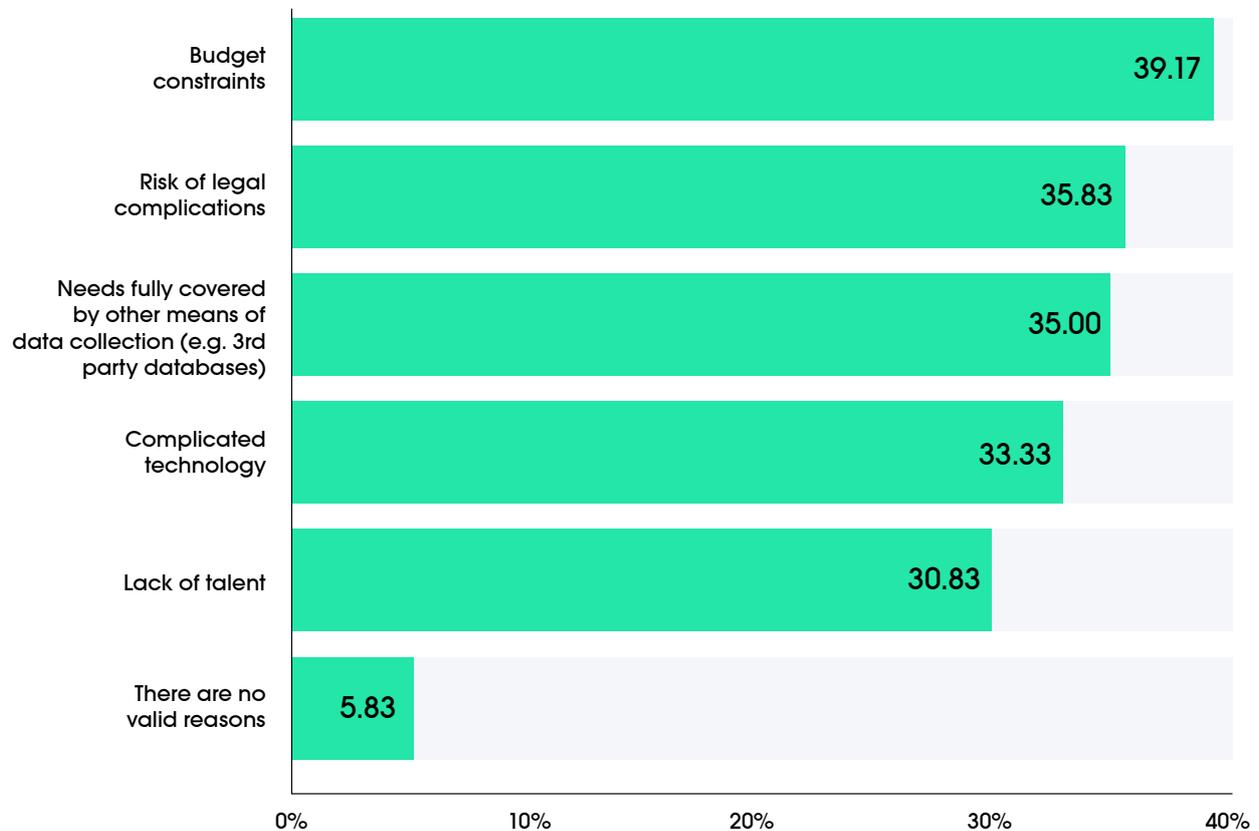**Denas Grybauskas**
Head of Legal at Oxylabs.io

" **Web scraping is definitely fraught with a lot of legal and technical intricacies. As there are no industry-wide regulations and ethics agreements, it's easy to get into trouble. However, one of the most important pieces of advice, outside of always consulting with a legal professional, is to only extract publicly available data. All other types of data are bound to be protected by a swath of laws and regulations and scraping them without infringing upon someone's rights is nearly impossible.**

Some (35%) respondents claimed that their needs were already fully covered by other means of data collection, such as third-party databases. However, as other companies expand upon their data acquisition efforts, the usefulness of third-party databases might decrease as in-house web scraping allows for greater quality control, data exclusivity, and goal adaptability.

Finally, a third (33%) of respondents see automated web scraping technologies as too complicated. Often, the perception arises from attempts to build complete in-house scraping pipelines, which need to be set from the ground up. However, that indicates an immense opportunity for services that create out-of-the-box solutions. Simplifying architecture and improving ease of use could bring an incredible amount of new clients.

Q6. What are the reasons for not implementing automated web scraping & cleaning fot data collection? (Tick all that apply)

*Respondents who do not rely on automated ETL processes, web scraping pipelines

Category	Value
Budget constraints	39.17
Risk of legal complications	35.83
Needs fully covered by other means of data collection (e.g. 3rd party databases)	35.00
Complicated technology	33.33
Lack of talent	30.83
There are no valid reasons	5.83

Data strategy focus in 2021 & beyond

Finally, most respondents have a focused data strategy for the forthcoming 12 months. There is an even split from respondents in our study regarding the focus. Data quality assurance as one of the goals sits at the lowest rung of the ladder (38%) while others (e.g. infrastructure reliability, collecting more signals, and utilising more alternative data) are divided equally amongst themselves (46-48%).

Q7. What is your data strategy focus for the upcoming 12 months? (Tick all that apply)

Data collection and managemen infrastructure reliability	48
Collecting more real-time signals for faster investment decisions	47
Collecting and utilizing more of alternative data in data driven investment decisions	46
Data quality assurance	38

0%   10%   20%   30%   40%   50%
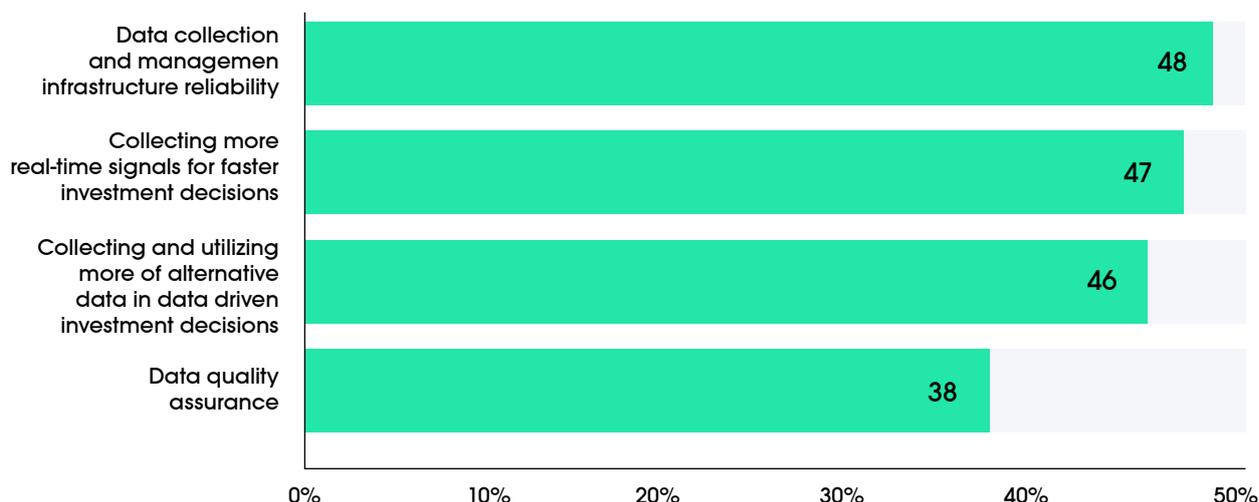
Thus, as we can see, alternative data is playing and will continue to play an increasingly important role in financial decision making. There are numerous financial incentives to do so, however, due to the complexity of external data acquisition, some companies are yet to implement these processes. Third-party web scraping services and tool providers have an incredible opportunity to enter the market and solve most of the quality assurance problems that are causing companies to avoid external data acquisition.

Julius Černiauskas
Chief Executive Officer at Oxylabs.io

" **Data has always been the primary driver of decision-making in the finance sector. However, we have shed some light on the new developments in data collection strategies. As I can see, there are ample opportunities for everyone in this field - from legal professionals to data scientists. Web scraping is still bristling with challenges and those who can solve them are sitting on a gold mine.**

oxylabs®

Want to Know More?

If you would like to know more about any of the topics mentioned in this white paper or **learn about our products**, please get in touch! Our team is ready to answer any of your questions and offer you the best solution for your business needs.

Get in touch with Oxylabs

Our Mission

Our mission is to share all the know-how that we collected over the years in the industry in order to create the future where big data is accessible to all businesses. We seek to create a healthy environment for everyone to grow and thrive in.

Our Values

As a leading company in the proxy and web scraping industry, we ensure that the highest standards of business ethics lead all our operations. Our core values guide us toward achieving our mission. **Learn more**