



The State of Data Management Report

A Global Survey of Data & Analytics Leaders

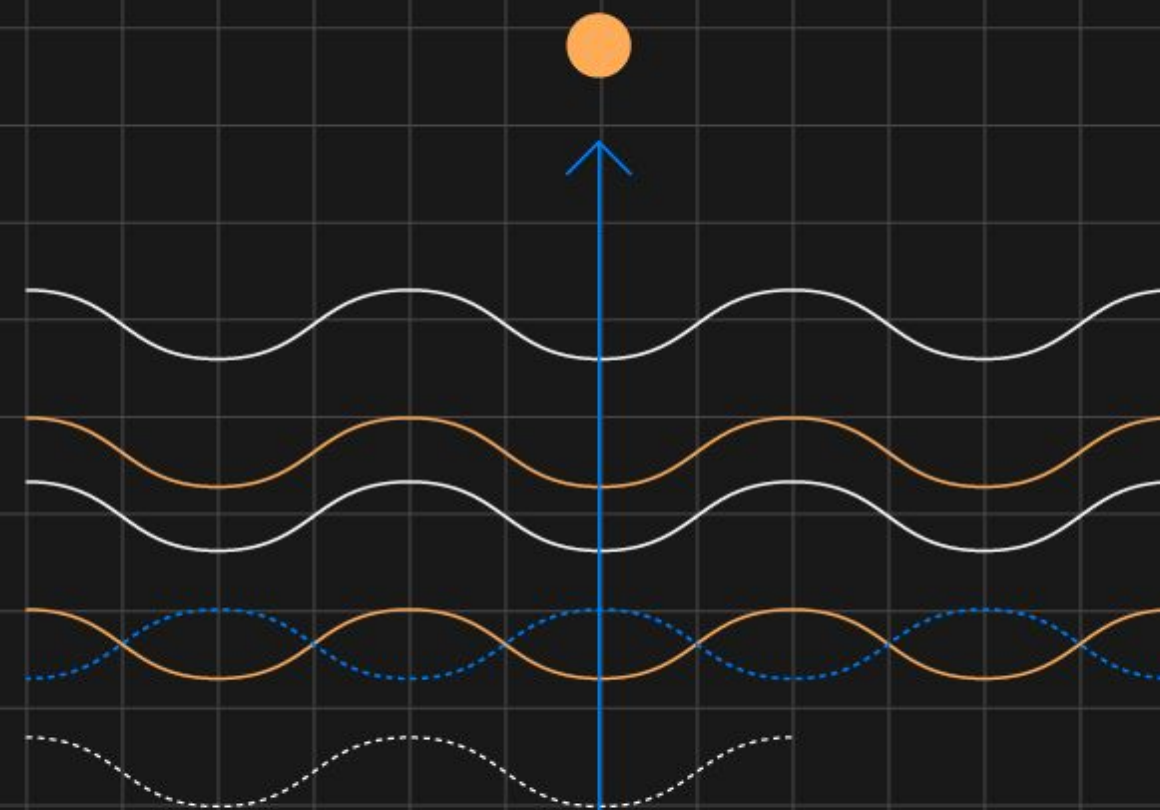
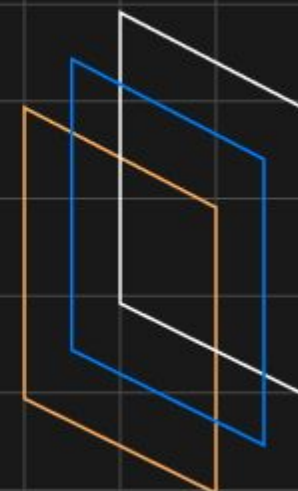
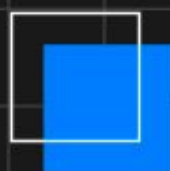


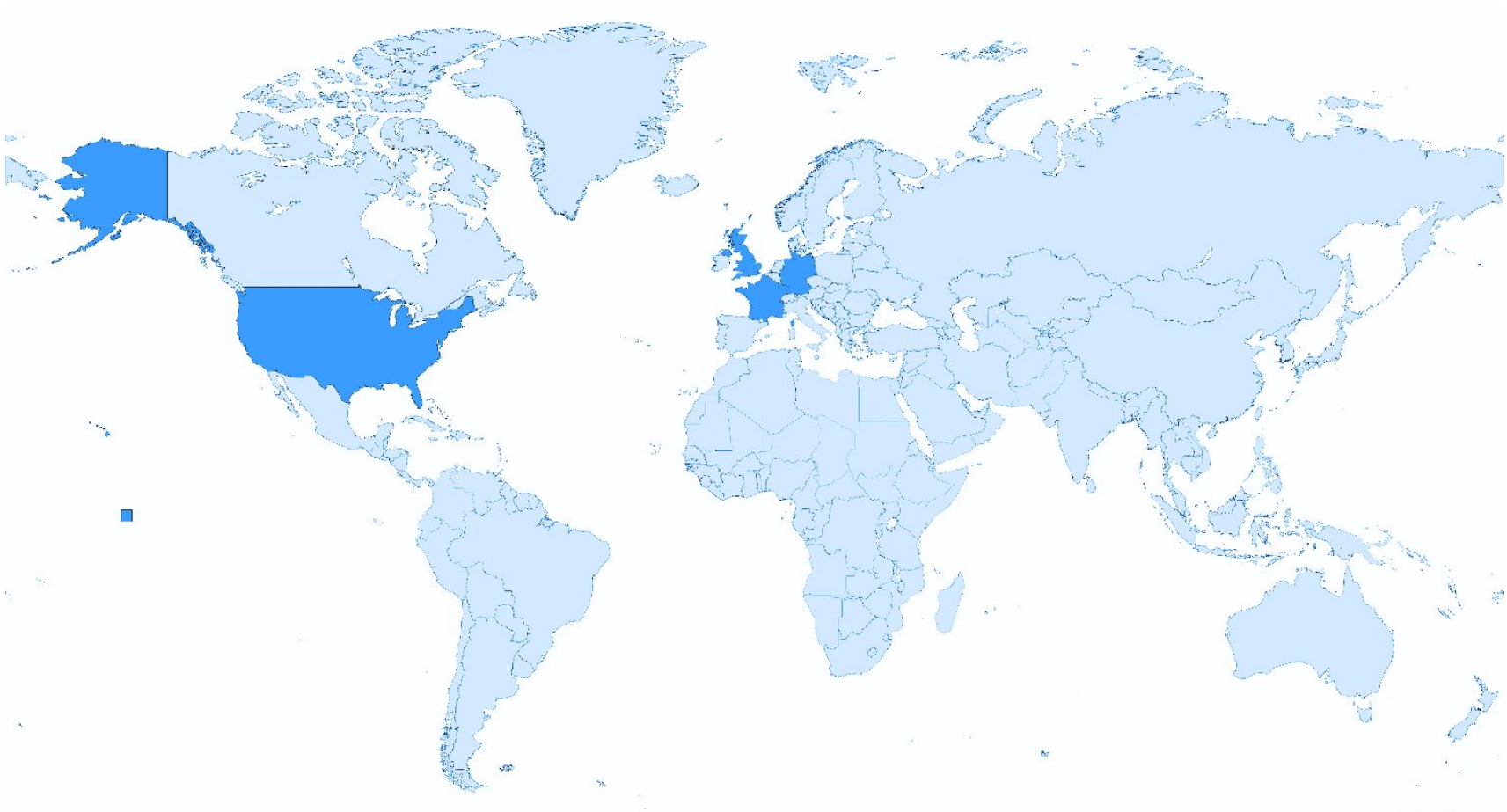
Table of contents

Section	Slide number
Research Methodology	3
Key Findings	4
Detailed Research Findings	8
- Cost of Building Data Pipelines	9
- Waste of Time & Talent	13
- Bad Business outcomes	20



Research Methodology

Wakefield Research conducted a quantitative research study between September 27th and October 12th, 2021, among 300 Data and Analytics Leaders at Businesses with 2,500+ Employees VP+, in the US, UK, Germany and France, at companies with \$100m+ in annual revenue, a minimum of 100 employees, and who are familiar with the data strategy/data use at their organization.



C-Level	33%
Executive	67%

Results of any sample are subject to sampling variation. The magnitude of the variation is measurable and is affected by the number of interviews and the level of the percentages expressing the results. For the interviews conducted in this particular study, the chances are 95 in 100 that a survey result does not vary, plus or minus, by more than 5.7 percentage points from the result that would be obtained if interviews had been conducted with all persons in the universe represented by the sample.

Key Findings

Key Findings

Companies are overpaying—to the tune of over a half a million dollars a year—for data integration solutions, due to their obsolete reliance on the manual building and management of data pipelines.

Companies spend, on average, \$520K every year paying data engineers to manually build and maintain data pipelines. This is based on a median of 12 data engineers on a team, who earn an average salary of \$98K a year and spend 44% of their time on average on this obsolete task alone.

That cost doesn't even factor inefficiencies, as 4 in 5 data leaders (80%) say they at least sometimes have to rebuild data pipelines after deployment; this happens often or even *all the time* at 39% of companies. This means that, as jaw-dropping as that \$520K is, the actual cost of data integration that relies on this manual process could be an even larger line item on the balance sheet than most companies suspect.

Key Findings

Beyond eye-popping cost, the manual approach to data pipelines is a waste of time and talent.

In this white-hot labor market, companies are spending big on top talent to try to make the most of their data—yet most data & analytics leaders say their team’s time is being wasted on manual oversight of data pipelines. Time sunk into this takes away from engineers’ ability to provide more advanced analysis, as data leaders say this is the top area their engineers could contribute to if they weren't manually building data pipelines.

Not only are they wasting their top talents time, but they’re stuck in a quagmire of an unscalable process. Most data leaders (83%) say they are locked into their current approach, and 83% also agree that any scaling up would require more staff to manage data as a major component—and the top concern cited with such a process would be the cost and time it takes to train up new talent.

Key Findings

Companies are paying huge sums only to achieve bad outcomes. For all the time and money spent on data pipelines, data is still not fresh, leading to old and error-prone information that is costing companies money.

Just 13% of companies report being able to derive value from newly collected data within minute or hours. For most companies (76%), it takes several days or up to a week. As a result, most data leaders report that their end data users make business decisions with old or error-prone data at least some of the time—in fact, 36% say this happens often or all of the time.

This is a problem that goes all the way to the top: two-thirds of data leaders (66%) suspect their C-suite leaders wouldn't even know if they were making decisions with old or error-prone data. As a result, 85% of data leaders admit their companies have lost money because of business decisions made with this old or error-prone data.

Detailed Research Findings

Cost of Building Data Pipelines

The Math Doesn't Lie: Companies Pay Data Engineers More Than Half a Million Dollars to Build And Maintain Data Pipelines Manually

This shocking number comes from a straightforward calculation of the median number of engineers on a team times their average salary, multiplied by the average percentage of time these engineers spend on building and maintaining data pipelines.

$$\begin{array}{lcl} 12 & \times & \$98,400 \\ \text{Median number of data engineers} & & \text{Average salary for data engineers} \\ & \times & 44\% \\ & & \text{Average amount of time spend manually} \\ & & \text{building \& maintaining data pipelines} \end{array} \left. \vphantom{\begin{array}{lcl} 12 & \times & \$98,400 \\ & \times & 44\% \end{array}} \right\} \$519,552 / \text{year}$$

Thinking about your data team, how many employees on your data team are responsible for building and maintaining data pipelines? / What is the average salary of those on your data team who build and maintain data pipelines? /

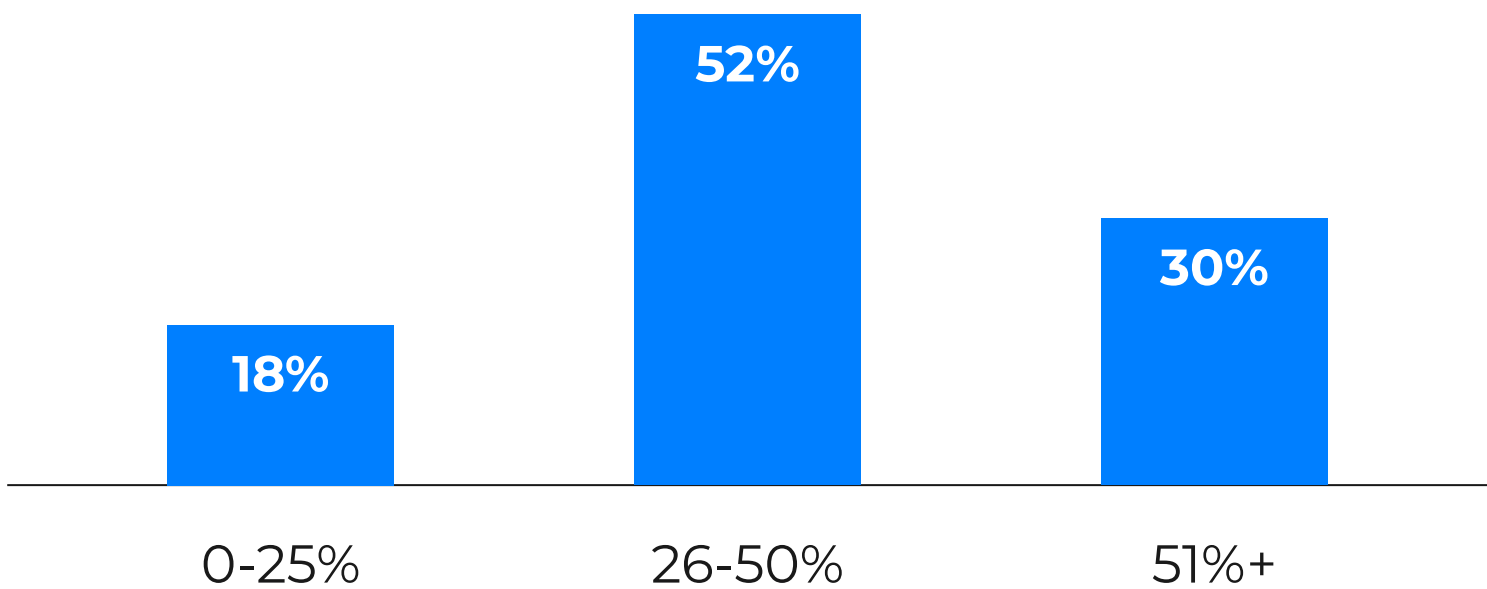
For those who build and maintain data pipelines, what percentage of their time is spent manually building and managing data pipelines?

Data Engineers Are Wasting Their Time on Inefficient Processes

Companies spend a lot of effort to bring in top data talent, then waste nearly half of their time on manually building and maintaining pipelines. More than 4 in 5 data leaders (82%) say their engineers spend more than a quarter of their time on this.

Percentage of Time Spent Manually Building and Managing Data Pipelines

N=300



44%

The average amount of data engineers' time that data leaders report they spend manually building and managing data pipelines.

Insights by Company:

Just 13% of larger companies (more than 3,500 employees) have their data engineers spend 25% or less of their time on manually building data pipelines.

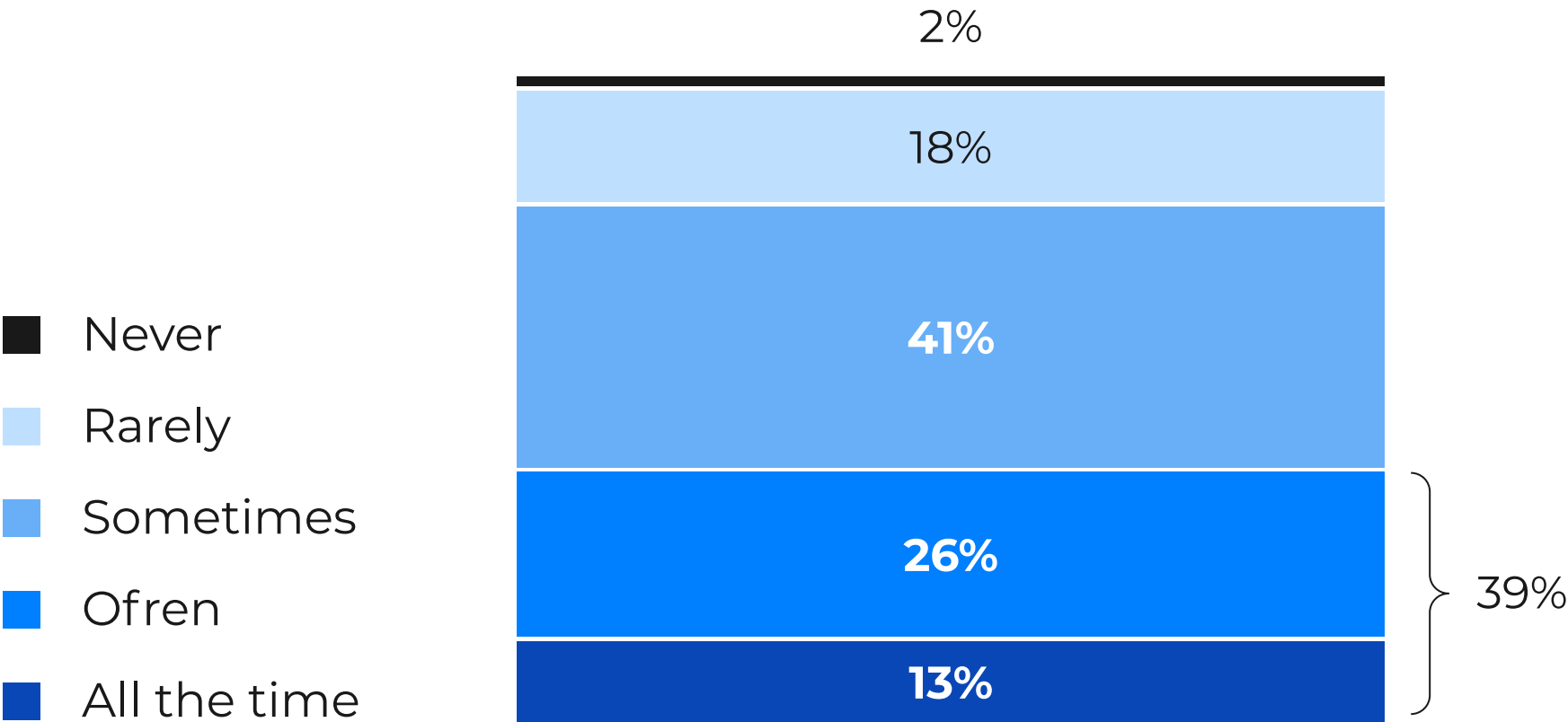
For those who build and maintain data pipelines, what percentage of their time is spent manually building and managing data pipelines?

Rebuilding Pipelines After Deployment is Common

In this white-hot labor market, companies are spending big on top talent to try to make the most of their data—yet nearly 3 in 4 data & analytics leaders (72%) say their team’s time is being wasted on manual oversight of data pipelines.

How Often Data Pipelines Have to be Rebuilt After Being Deployed

N=300



Insights by Company:

- This happens often or all of the time at **46%** of firms with <\$500M in revenue.
- Younger companies (30 years or less) are nearly twice as likely to say this happens often or all of the time, **52%** to **28%**.

How often do data pipelines at your company have to be rebuilt after being deployed?

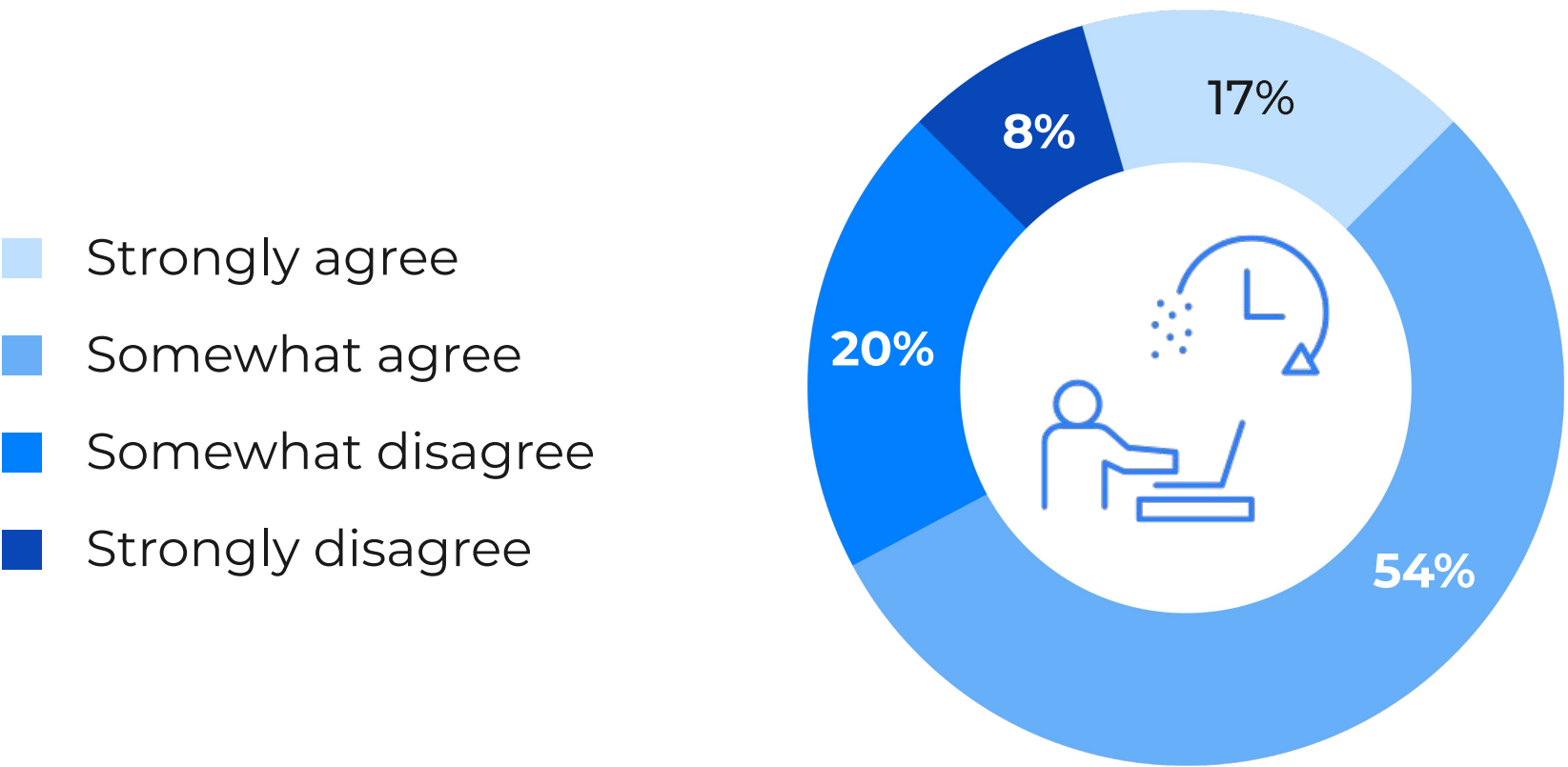
Waste of Time & Talent

Data Leaders Agree: This is a Waste of Time and Talent

It would be one thing if the processes companies used for manually building and managing pipelines were optimized, but most data leaders (80%) admit they have to rebuild data pipelines after deployment from changing APIs, for example. For 39%, they say this happens often or all of the time.

Company Puts Effort Into Hiring Top Data Talent only to Waste Their Time With Manual Oversight of Data Pipelines

N=300



Insights by Company:

- A strong majority of data leaders at companies, both large and small, agree that their data engineers’ time is being wasted.
- Data leaders at younger companies (30 years or less) are more likely to feel this way, **81%** to **63%**.

How often do data pipelines at your company have to be rebuilt after being deployed?

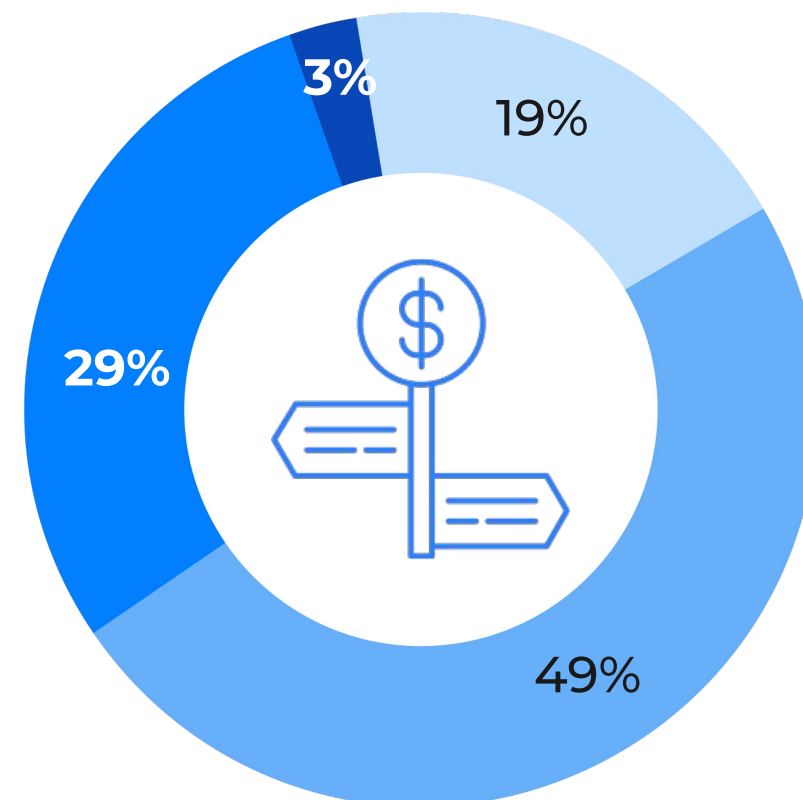
Companies Are Leaving Money On the Table With This Approach

Time spent manually building (and re-building) data pipelines is time not spent contributing to business decisions—which nearly all data leaders (97%) agree would improve business outcomes if they were able to.

How Improved Business Outcomes Would Be If Data Team Was Able to Contribute More to Business Decisions

N=300

- Significantly improved
- Somewhat improved
- A little improved
- Not improved at all



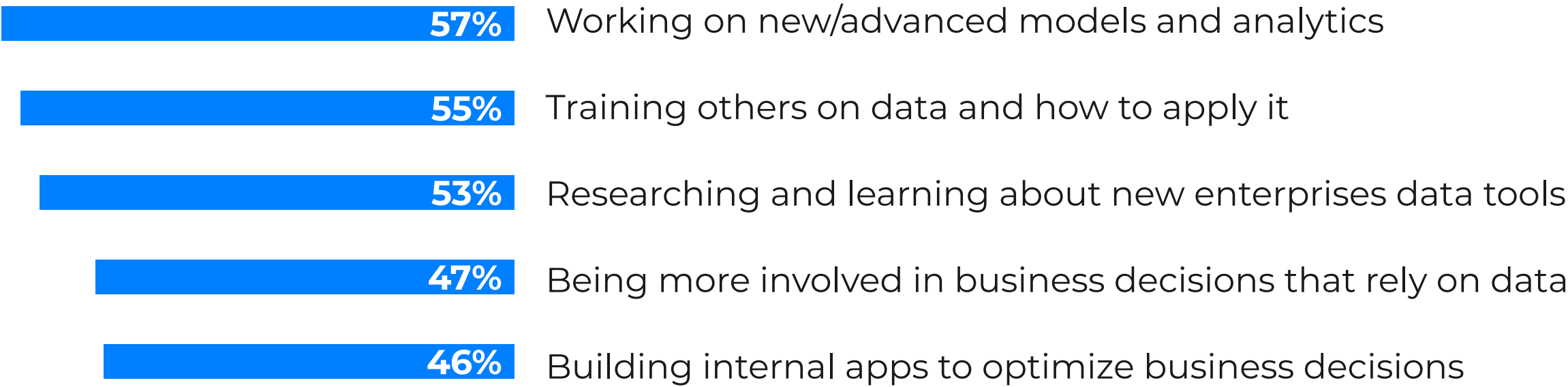
How improved would business outcomes at your company be if your data team was able to contribute more to business decisions?

Other Improvements Are Also Being Left On the Table

Nearly 3 in 5 data & analytics leaders (57%) say more advanced analysis is a top area where engineers could contribute if they weren't manually building data pipelines. Other top areas where they could contribute include training others on data (55%) and how to research and learn about new enterprise data tools (53%).

Most Beneficial Things Data Engineers Could Be Doing

Among those whose data engineers manually build pipelines
N=300



Insights by Company:

At companies with less than \$500M in revenue, data leaders see the greatest benefit as training others in the company on how to use data (61%) and researching new enterprise data tools (60%).

Asked among those who spent time manually building and managing data pipelines: What are the most beneficial things data engineers at your company could be doing if they were not manually building and managing data pipelines?



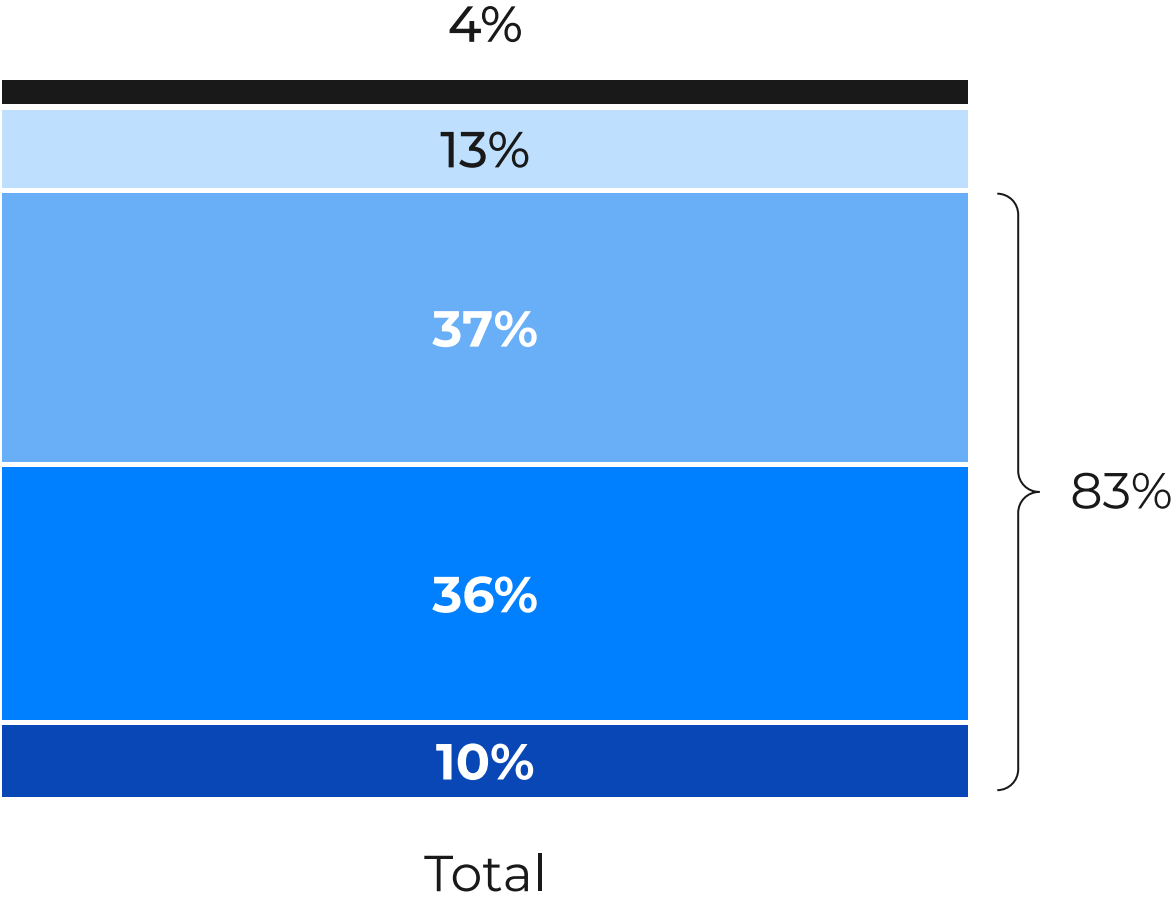
Companies Are Finding Themselves “Locked in” to Their Obsolete, Inefficient, and Wasteful Processes

More than 4 in 5 data leaders (83%) say they are locked into their current data integration setup, leaving them incapable of adopting newer and better solutions even when they present themselves.

Locked Into Current Data Integration Set-Up

N=300

- Not locked in at all
- Not too locked in
- Somewhat locked in
- Mostly locked in
- Completely locked in



Based on your company's investment into its data processes and tools, how locked into your current data integration set-up would you say your company is?



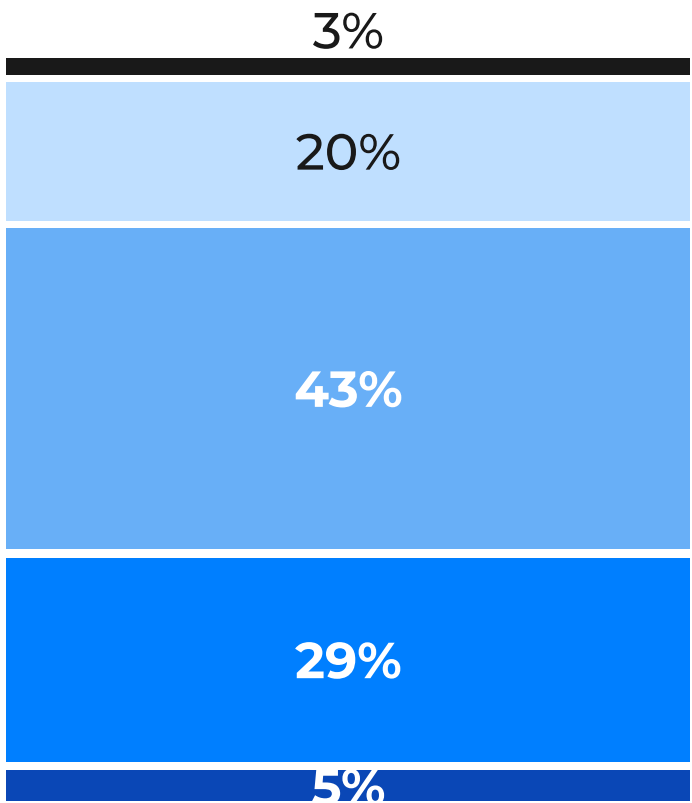
The High Cost Of Manually Building Data Pipelines Extends to Scaling Up As Well

Because companies are so reliant on human effort to manually build and maintain data pipelines, most (83%) agree that hiring more staff would be a requirement of scaling up. This all makes the process expensive, as 77% of data leaders admit.

Cost of scaling up data pipeline production

N=300

- Not expensive at all
- Only a little expensive
- Somewhat expensive
- Very expensive
- Extremely expensive

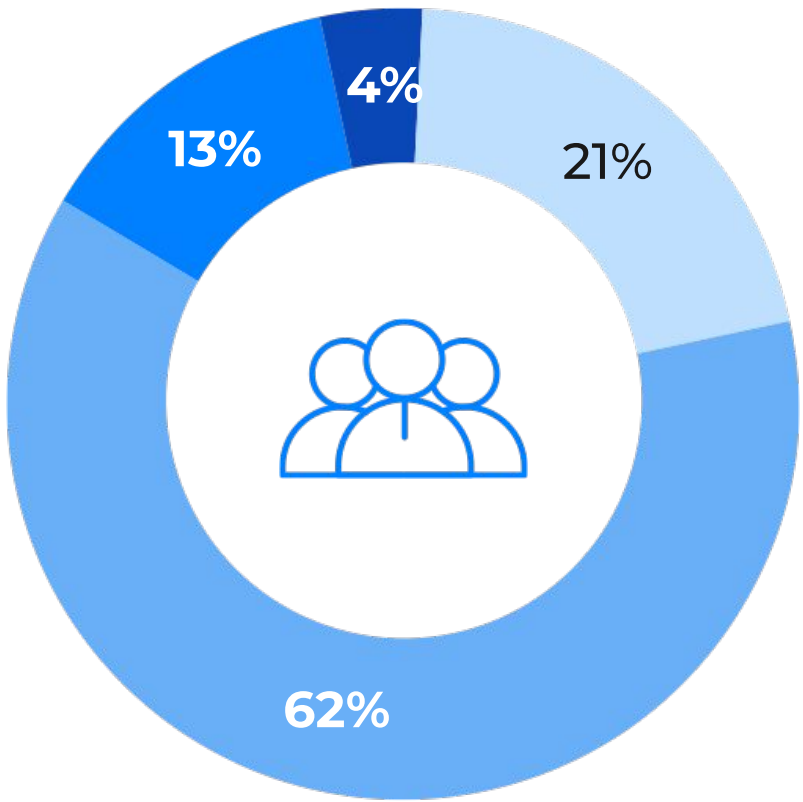


Total

Hiring more staff to manage data would be a major component

N=300

- Strongly agree
- Somewhat agree
- Somewhat disagree
- Strongly disagree



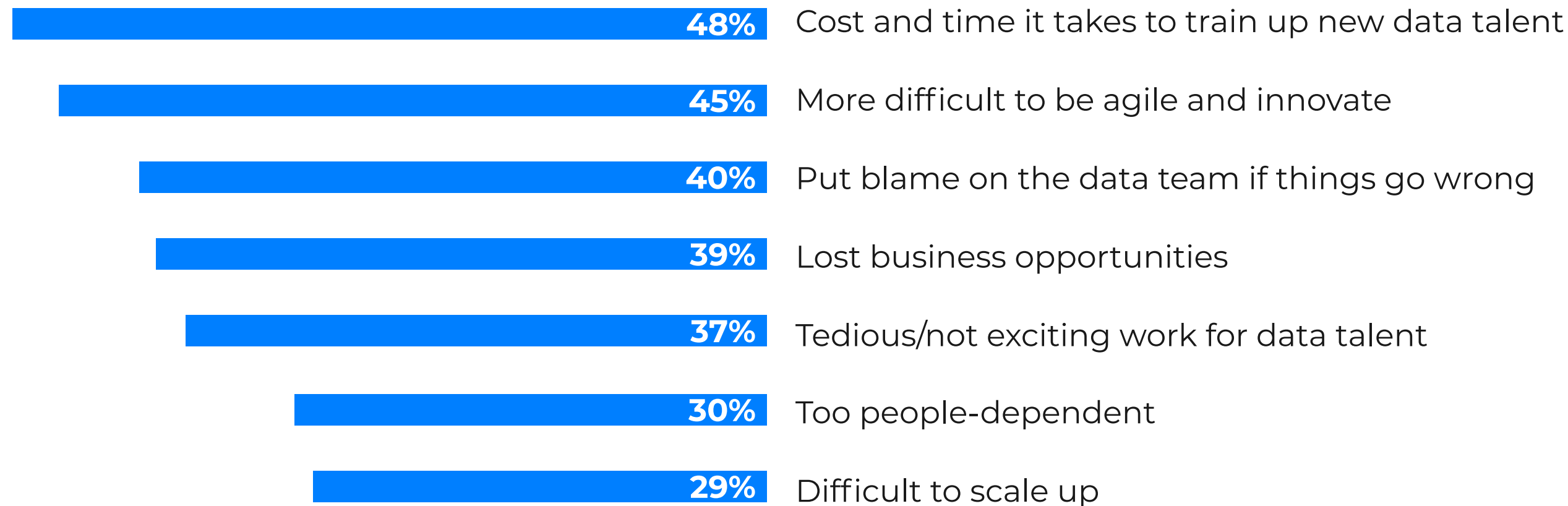
How expensive would it be to scale up production of your data pipelines? / How much do you agree or disagree with the following statement - If we scaled up our data management capacity, hiring more staff to manage data would be a major component.

The Costs of Scaling Up Includes Additional Time Sinks Like Training New Staff On Manual Processes

The cost and time it would take to train up new data talent is the top concern that data leaders cite with the manual approach to building and maintaining data pipelines. They also are concerned about the impact to the company, especially through missing out on innovations and lost business opportunities.

Top Concerns With Data Team Manually Managing Data

N=300



What are your top concerns you have with your data team manually managing data?

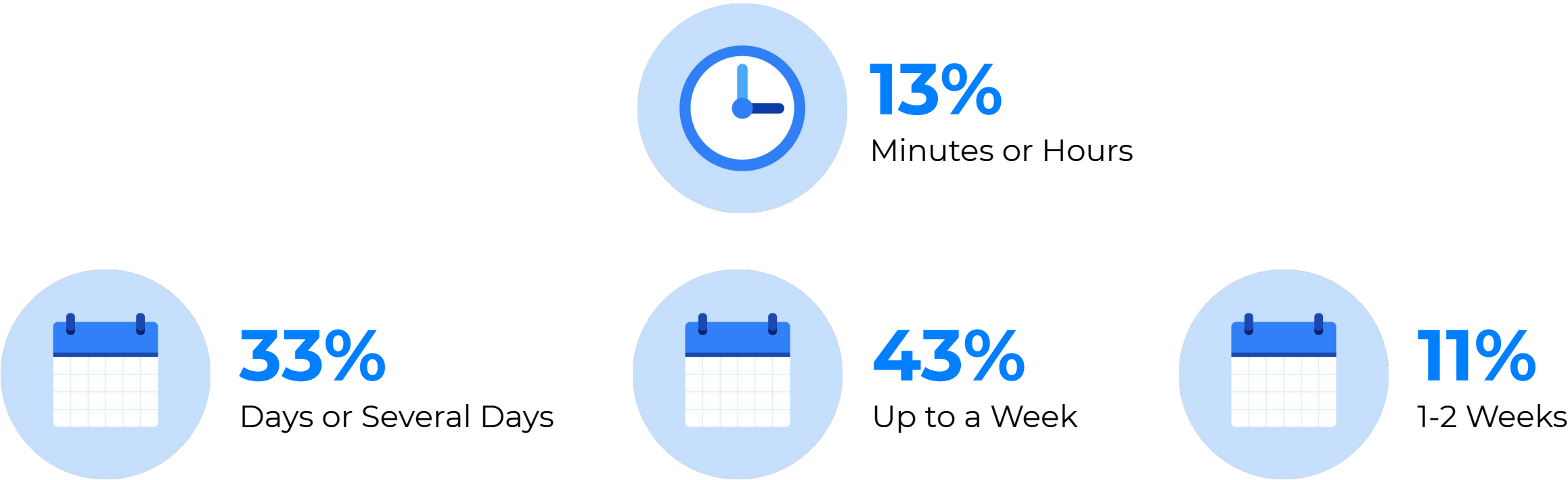
Bad Business Outcomes

Despite the Amount of Time and Effort Put Into Data Pipelines, Companies Still Derive Value From Data AT a Snail's Space

Just 13% of companies report being able to derive value from newly collected data within minutes or hours. For most companies (76%), it takes days or up to a week. Almost all (96%) agree that business outcomes would be improved if their data was fresher.

How Long It Takes Data End Users to Derive Value From New Data That Is Collected

N=300



96%

of data leaders who take days or longer say that their companies would see improved business outcomes if data end users were able to more quickly derive value from new data; this includes 66% who say outcomes would be somewhat or significantly improved.

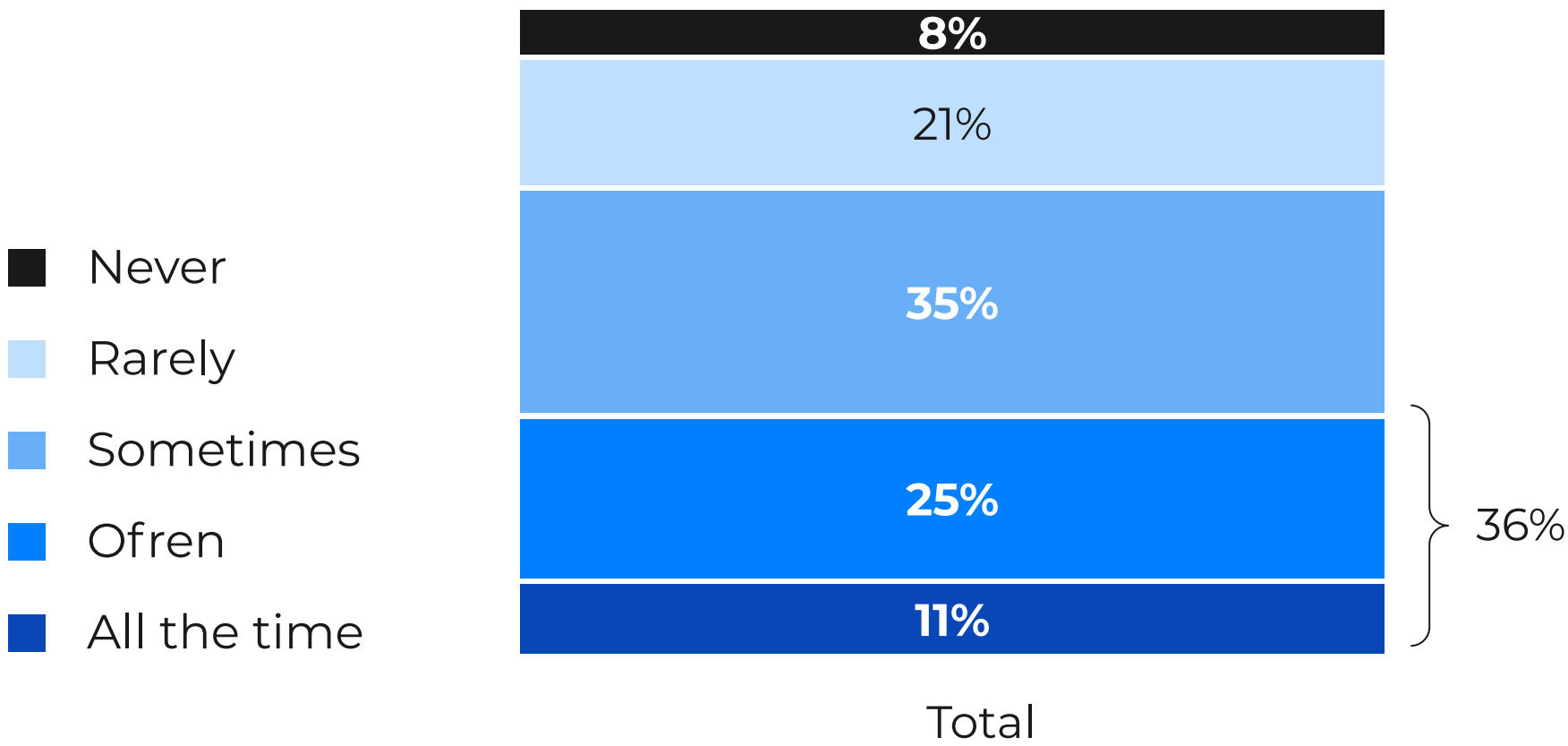
How quickly can data end users at your company derive value from new data that is collected? / How improved would business outcomes at your company be if data end users were able to more quickly derive value from new data that is collected?

Lack of Data Freshness Results in Old and Error-Prone Data

Slow and stale data is more than just a nuisance: it spells potential disaster for end users' business decisions. Most data leaders (71%) admit that their end data users make business decisions with old or error-prone data at least some of the time—including 36% who say this happens often or all the time

Data End Users Make Business Decisions With Old Or Error-Prone Data

N=300



Insights by Company:

- More than 2 in 5 (**44%**) larger companies (>3,500 employees) say these bad decisions happen often or all the time.
- Younger companies (30 years or less) are twice as likely to say this happens often or all of the time, **51%** to **24%**.

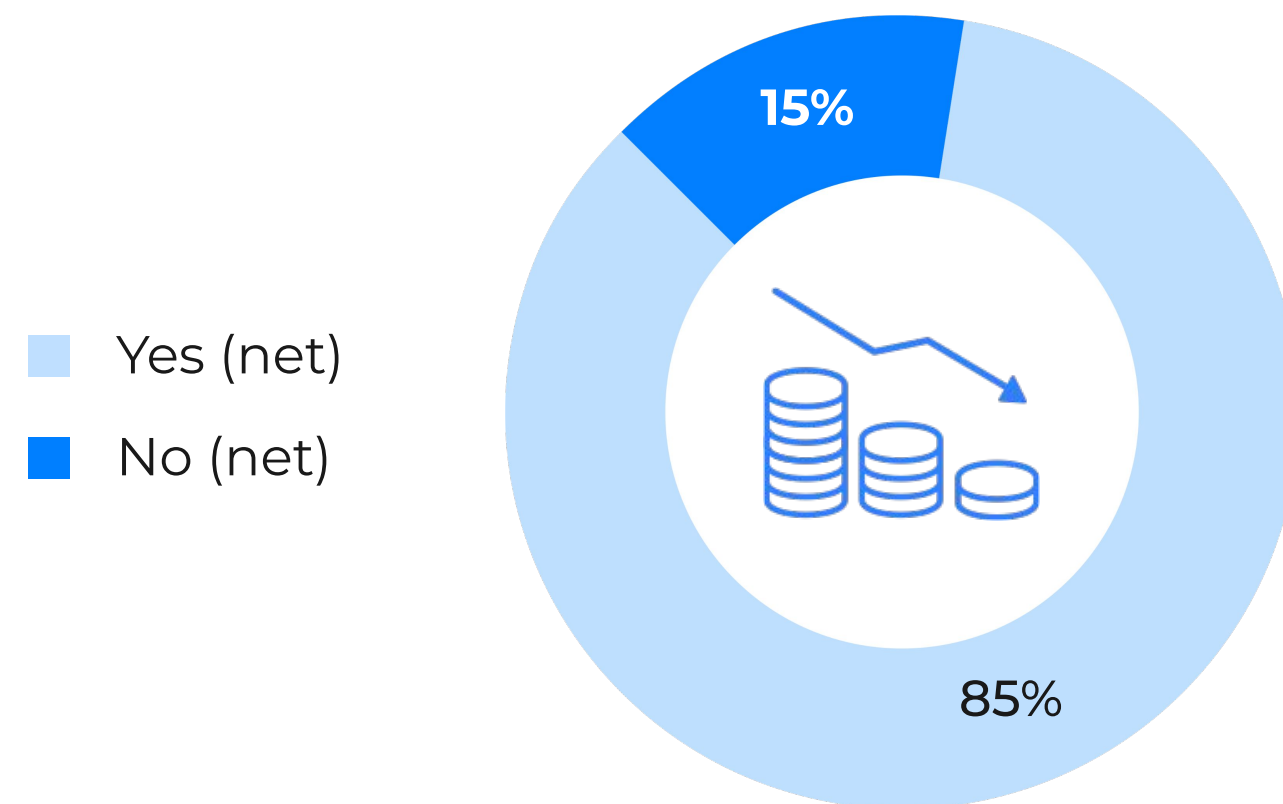
How often would you say that data end users make business decisions with old or error-prone data?

This Bad Data Is Costing Companies Money

The impact of making business decisions with old or error-prone data is hard to overstate: a massive 85% of data leaders say that it has cost their company money. This applies to companies both large and small.

Business Decisions Made Based On Old Or Error-Prone Data Have Cost Company Money?

Among those who say that data end users make business decisions with old or error-prone data
n=213



Insights by Company:

This is slightly more common at smaller companies (**91%**), but more than 4 of 5 data leaders at larger companies (**81%**) still report that business decisions based on bad data have cost their company money.

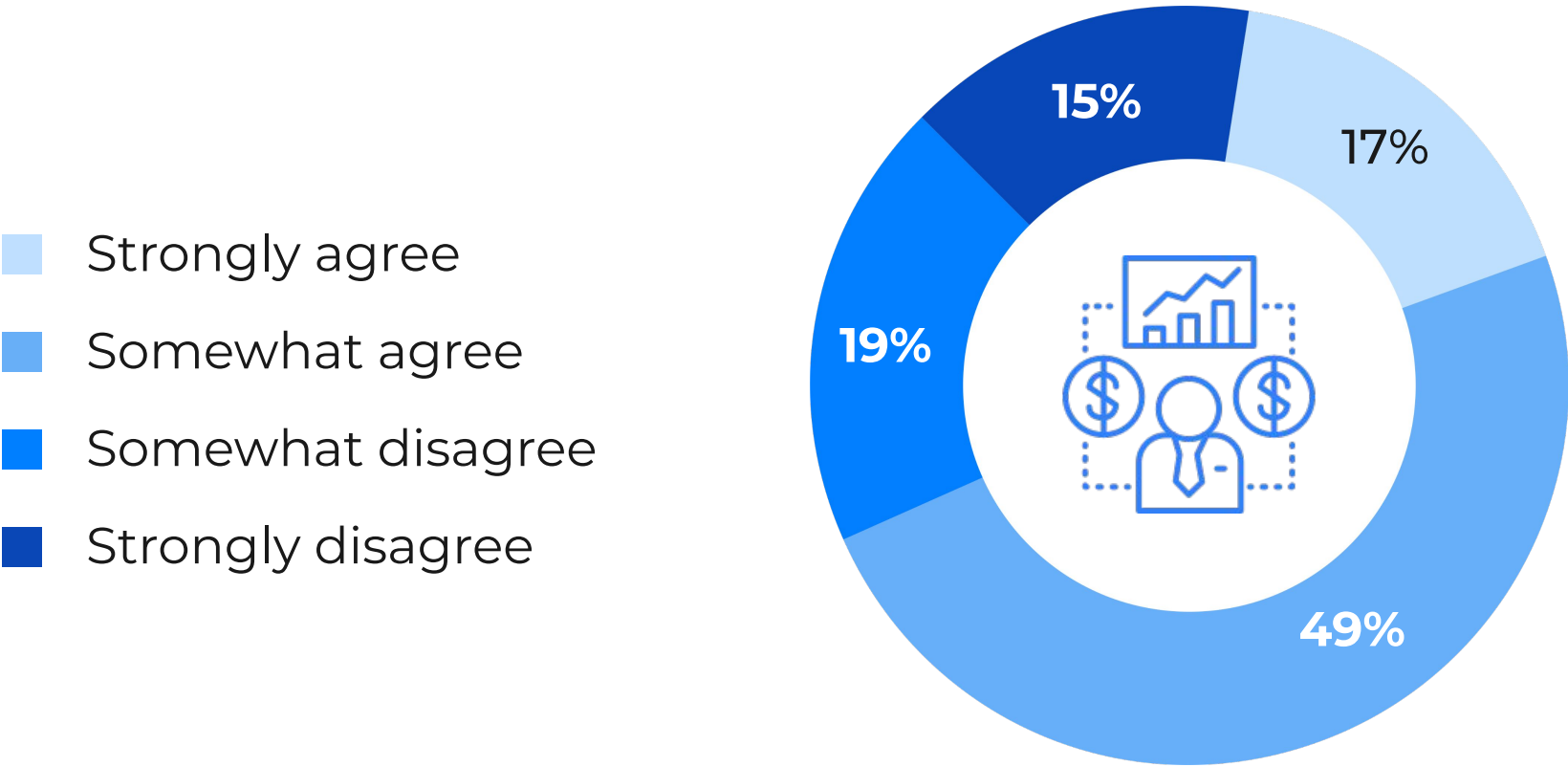
Have business decisions been made at your company based on old or error-prone data that have cost the company money?

C-Suites May Not Even Know This Is Happening

Two-thirds of data leaders (66%) agree that their C-suite leaders wouldn't even know if they were making decisions based on old or error-prone data—the types of decisions that are costing companies money, which could be huge sums when being made at the C-suite level.

C-Suite Leaders Wouldn't Know If They Were Making Decisions Based On Old Or Error-Prone Data

n=213



How much do you agree or disagree with the following statement - Our C-suite leaders wouldn't know if they were making decisions based on old or error-prone data.

WAKEFIELDRESEARCH.COM

Copyright ©2021 Wakefield Research. All rights reserved.
All information contained herein is confidential and
proprietary to Wakefield Research.