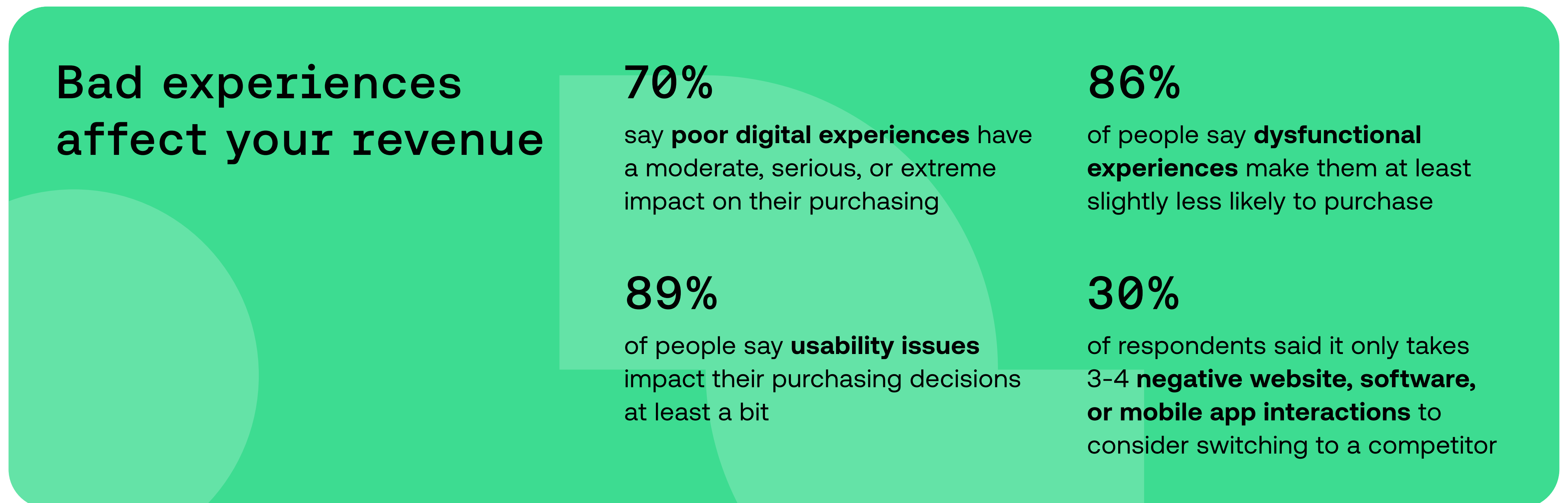


# Sauce Labs Continuous Testing Benchmark Report 2024

# The Impact of Bad Digital Experiences & The Case For Continuous Testing

In today's digital-first world, every experience counts. In fact, 86 percent of people say dysfunctional experiences make them less likely to purchase.



We've all experienced frustrating app malfunctions—like broken banking apps that make it impossible to check your account balance or e-commerce sites that keep crashing when you try to checkout. Or other more critical software failures that prevent cars from starting, keep planes grounded, and cause medical systems to break down.

In a recent global IT outage, more than 1500 flights were cancelled in the US and over 1000 delayed. 911 services were disrupted overnight in Alaska and Ohio and hospitals in Germany canceled non-emergency operations due to system failures.

These issues highlight the negative impact that software can have on people's lives.

Delivering quality apps means thinking about quality across the entire software development lifecycle—not just in the middle like traditional testing practices. This methodology, Continuous Testing (at the heart of Continuous Quality), allows the engineers and QA teams to get the feedback they need throughout the development process, so they can find bugs faster and fix them more efficiently.

## Want to learn more about the role of continuous testing?

Check out our eBook

[Platform for Test - Integrate Continuous Quality Across the SDLC to Make Every Experience Count.](#)



1) [InvGate](#), 8 CrowdStrike IT Outage Stats To understand How it Affected the World, July 23, 2024.

# The Continuous Testing Benchmark Report

Given the undeniable need for continuous testing across the world of software, we wanted to understand how organizations are approaching it and how it impacts their software delivery and business goals.

The goal of Continuous Testing Benchmark report is to help you better understand where you are at on your journey to continuous quality by looking at three core continuous testing metrics:

- Test Quality
- Test Runtime
- Test Platform Coverage

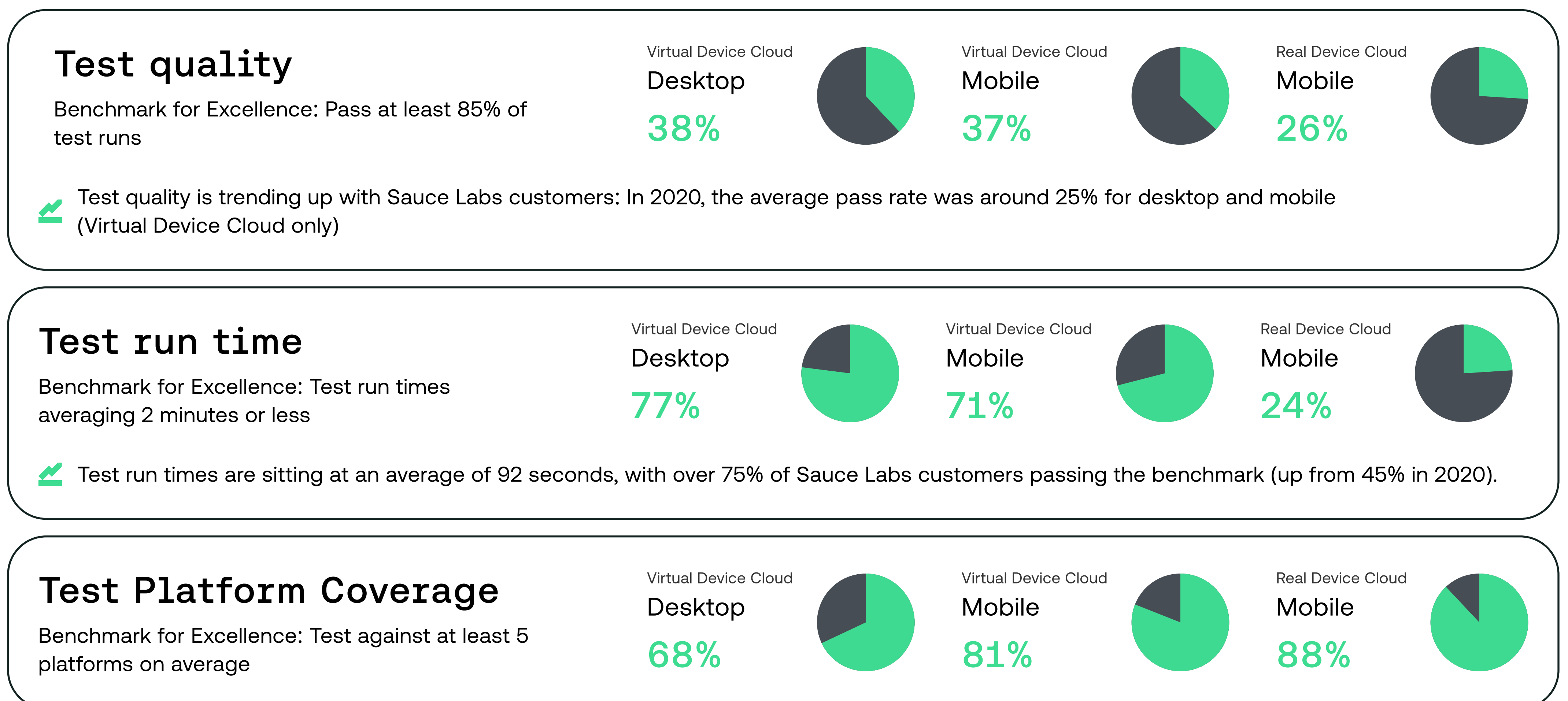
This report serves as a benchmark for assessing and improving test performance across these metrics, using actual test data:

- It leverages our expertise as a leading provider of test infrastructure. We've analyzed and anonymized data over a six month period (January - June 2024), during which time we executed over 20 billion tests (both mobile and desktop) per month across 100,000 customers.
- And, while test volumes vary significantly across different organizations, each one's influence on the overall metric performance is weighted equally.

The benchmark test data and insights in the report are designed to aid in strategizing and decision-making across DevOps and QA resources, leading to improvements in efficiency, velocity, risk management, and ultimately superior digital experiences.

Our last report was published in 2020 and used by global brands like Walmart, Salesforce, Verizon, Doordash, and Visa, looking to understand how they measured up in their continuous testing strategy. In this 2024 version, you will see comparisons to some of the data and insights we found back then, and get a glimpse into what has changed across the continuous testing landscape today.

## Benchmark Components



This year, we placed a huge emphasis on mobile as the mobile application landscape continues to explode. For the first time ever, we've included Real Device Cloud (RDC) data in this report.

We're also sharing metrics by industry for each of the components this year, and showcasing stories about industry leaders in testing – so you can gain a deeper understanding of how you match up to the benchmark and your industry peers.

# Benchmark Metric #1: Test Quality

## Why it matters

To ensure your QA practice sends out reliable signals, you must start with high-quality data. And to achieve that, you need high-quality tests. Test quality is fundamental in continuous testing — it serves as the primary indicator of its effectiveness and the foundation upon which other testing practices are built.

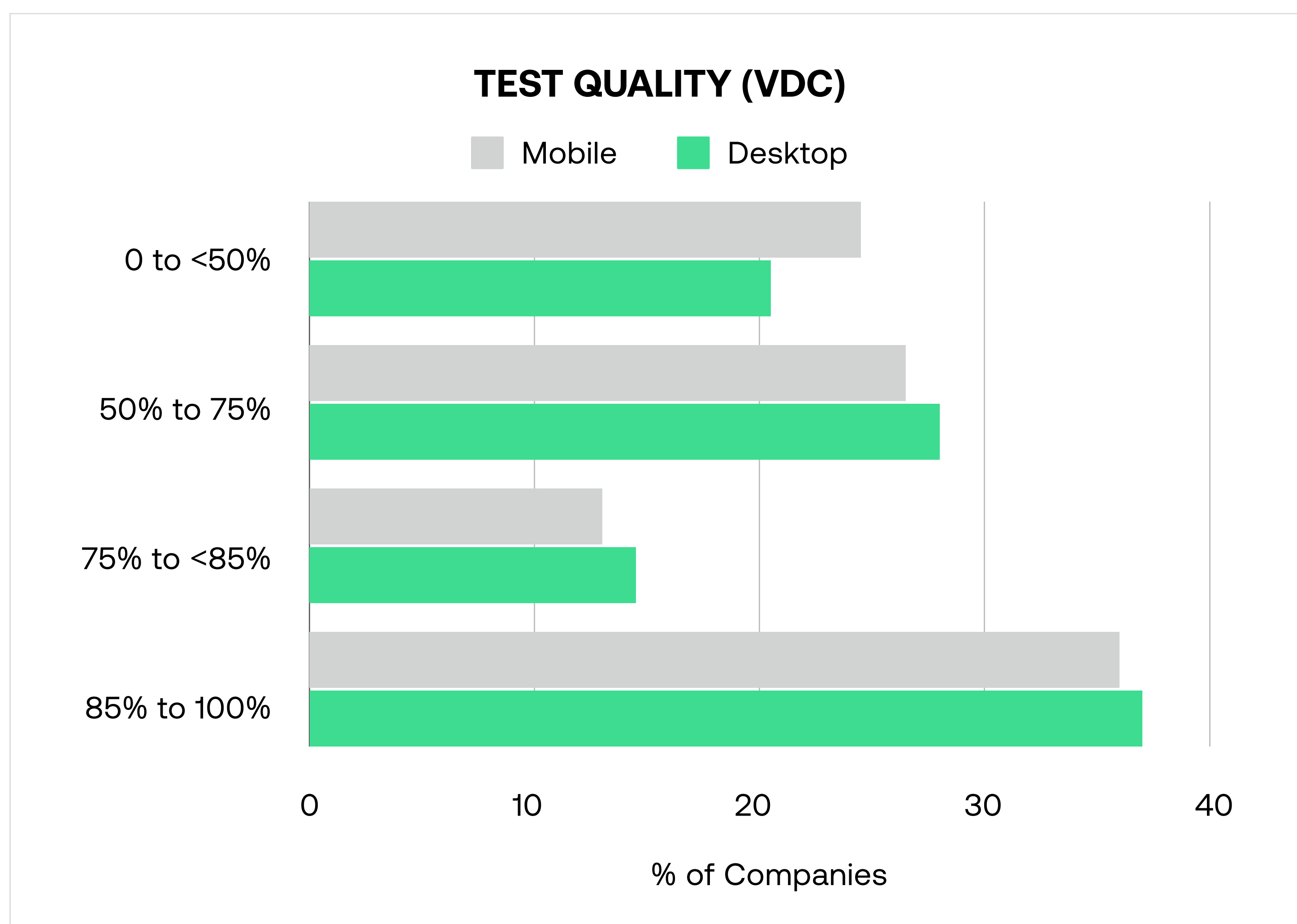
Automated testing aims to speed up development and reduce release cycles—ensuring that tests are reliable and consistent, requiring a high test pass rate. When tests fail, it's vital for developers to swiftly identify and fix the problem.

A high pass rate indicates that when a test fails, it likely points to an actual issue rather than an inconsistency in the test suite. High failure rates, often caused by test "flakiness" (tests that produce both passing and failing results despite no changes to the code or test), can disrupt this process. This necessitates manual intervention, slowing down the delivery of quality software. A low failure rate ensures that each failed test indicates a real issue rather than just inconsistencies in the test suite.

The Benchmark for Excellence: The CTB evaluates organizations based on the typical pass rate of their tests, highlighting the importance of not just the pass rate but also an organization's capacity to handle manual follow-ups on failures. In other words, your failure rate should not exceed your ability to analyze the failures. To achieve a test quality score of 100, symbolizing excellence, the average pass rate should be at least 85% across both desktop and mobile tests.

## How Organizations Performed:

- Desktop: 38% of organizations pass the test quality benchmark
- Mobile: 37% of organizations pass the test quality benchmark



2) Abstracta. What Percentage of Functional Testing Should be Automated?, September 18, 2023

## Test Quality: Virtual Device Cloud (VDC)

### Benchmark performance

37%

of organizations passed the benchmark

### Average pass rate

70%

In 2020, the average was around 25% for desktop and mobile on Virtual Device Cloud.

#### What this means:

Improvement in test quality over the last few years is due to advancements in artificial intelligence capabilities, shift left practices (catching bugs earlier in the SLDC), improvements in continuous integration (CI), and more investment into test automation as more companies see its value.

### Both mobile and desktop have increased since our 2020 report from 25%

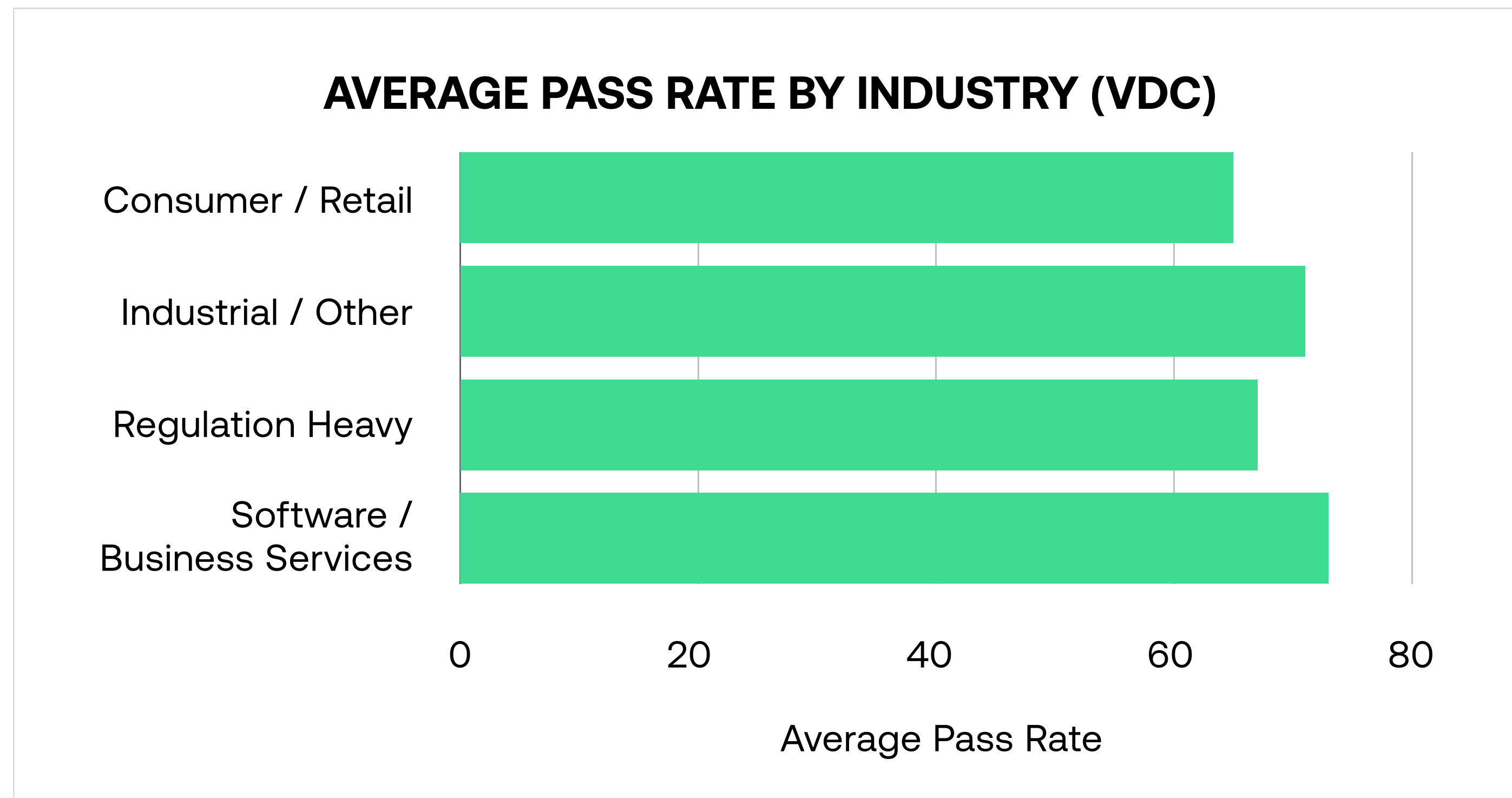
38%

Of desktop (Windows, Mac, Linux) pass the benchmark (24% in 2020)

37%

Of mobile (emulators, simulators) pass the benchmark (25% in 2020)

## Test Quality: VDC By Industry



The software industry had the highest pass rate across all industries: 41% of companies passed the benchmark.

But, why? We believe this is because software companies are the most familiar with automated test strategies and rely heavily on innovative technology support across the SDLC. In addition, digital experiences are the sole underpinning for their entire business model. Simply put, if their software doesn't work—they lose money.

## Excellence in Continuous Testing with Verizon Media

Benchmark pass rate: 100%

[Check out our case study](#)

[Verizon Media Accelerates Millions of Tests Every Month Using Open Source Technology and Sauce Labs](#)

## Test Quality: Real Device Cloud (RDC)

Given the massive increase in mobile usage since 2020—95% of adults in the US have a smartphone and 4 out of 10 of them say they are online almost constantly—we wanted to take a deeper look at the data and insights from our Real Device Cloud.

How organizations performed:

- Benchmark performance: 26% of organizations passed the benchmark
- Average pass rate: 65%

### Test Quality: RDC By Industry

The Industrial industry achieved the highest pass rate across all industries, with 76% of companies passing the benchmark. Due to the sector's reliance on stable and repeatable processes, they typically run specialized hardware that reduces the need to adapt to frequent device changes.

The consumer retail industry has the lowest pass rate, likely due to the complexities of supporting a wide variety of devices and user interfaces. This challenge is particularly apparent during Black Friday, when 54% of online sales were made on mobile. So that means that retail companies have to ensure that their digital experience works across many different types of devices.



“We were looking for complete end-to-end testing for Toyota's web and mobile experiences. Sauce Labs uniquely allows us to test both!”

## The Takeaway

Generally, virtual devices may have higher pass rates due to the controlled and consistent environment, which reduces the likelihood of encountering unexpected issues. While virtual devices might show higher pass rates, real devices provide more reliable and realistic test results. Taking a balanced approach to using both virtual and real devices is recommended to leverage the benefits of both and achieve comprehensive test coverage.

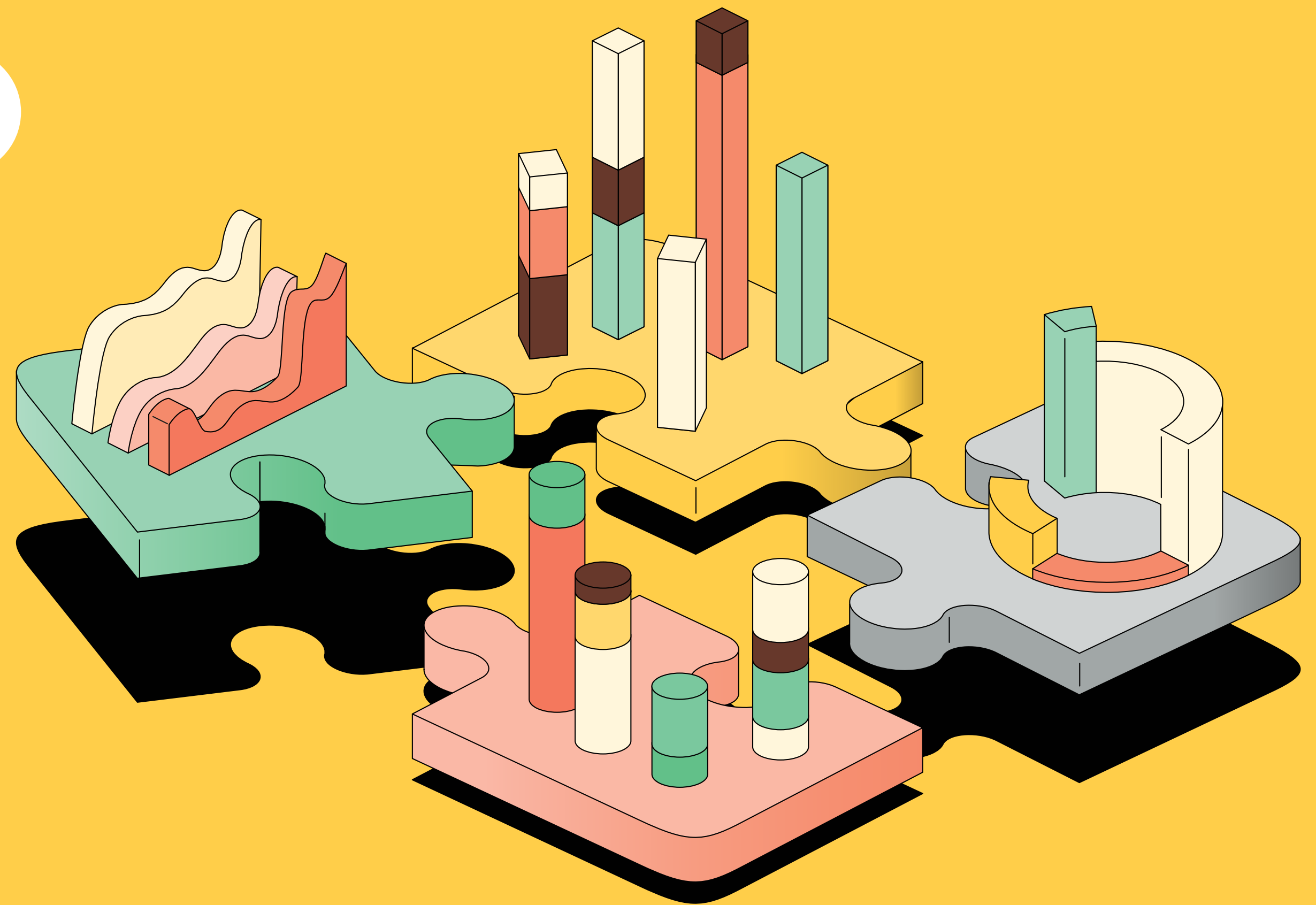
Overall, there's an urgent need for improvement in test quality across the board — especially as organizations increase the speed of development and scale their test automation. If you don't understand why tests fail, you can't facilitate continuous improvement. Failure analysis stands as an innovative approach to combat poor test quality, leveraging ML to analyze pass/fail data and enabling developers to quickly address issues and improve test quality.



SAUCE INSIGHTS

**With Sauce Labs advantage, utilize AI-powered failure analysis and data across the SDLC to gain insights and improve test reliability.**

[GET FAILURE ANALYSIS NOW](#)



# Benchmark Metric #2: Test Run Time

Test time is defined as Session start time to end time.

## Why it matters

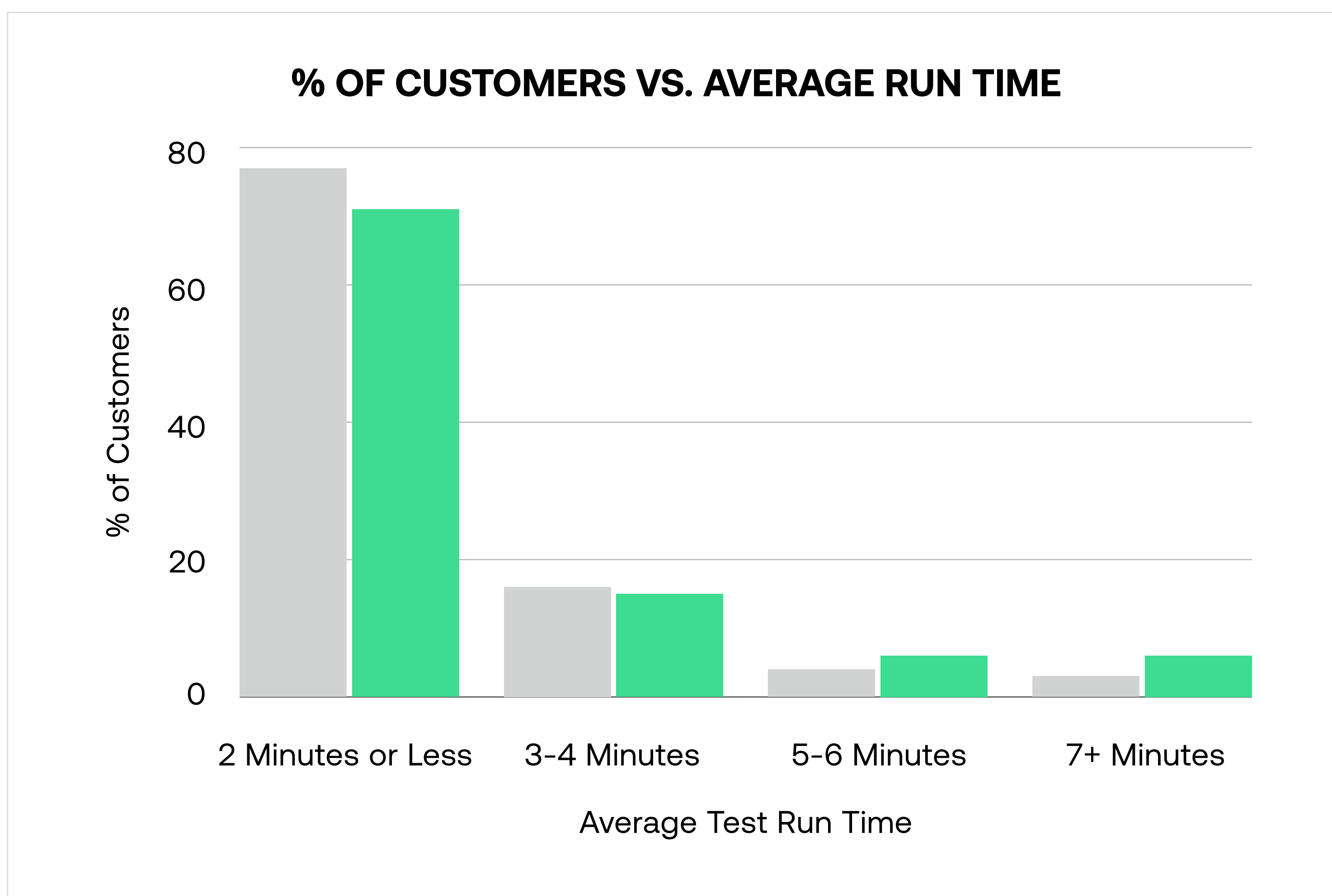
How velocity and automation work together: Velocity of innovation is the main goal for organizations that want to get to market quickly. By leveraging test automation, companies can enable development teams to test more efficiently, iterate faster, and release new updates, features, and products more frequently.

Why test speed is important: Slow, lengthy tests can hinder development by creating slower feedback loops for developers and delaying releases to end users.

The benefits of keeping test run times short extend beyond just speed alone. The data confirms that longer tests lead to poor test quality, as tests that complete in two minutes or less are nearly twice as likely to pass. Simply put, the longer a test takes to run, the more likely it is to fail. This is often because longer tests cover more parts of the application, increasing the chances of encountering issues. Additionally, longer tests are more challenging to troubleshoot. They often attempt to assess multiple functionalities in one script, making it harder to pinpoint the source of failure when they don't pass.

## The Benchmark for Excellence

The CTB Test Run Time metric looks at the average run time of the tests run on Sauce Labs. To achieve a score of 100 signifying excellence, the average run time of an organization's tests should be two minutes or less.



How Organizations Performed:

- Desktop: 77% of organizations complete in an average of two minutes or less
- Mobile: 71% of organizations complete in an average of two minutes or less



## Test Time: Virtual Device Cloud

### Benchmark performance

The pass rate for test time is 45% higher than in our 2020 report.

In 2024, the average test speed is 92 seconds.

75%

of organizations passed the benchmark

### Both mobile and desktop have increased since our 2020 report as well

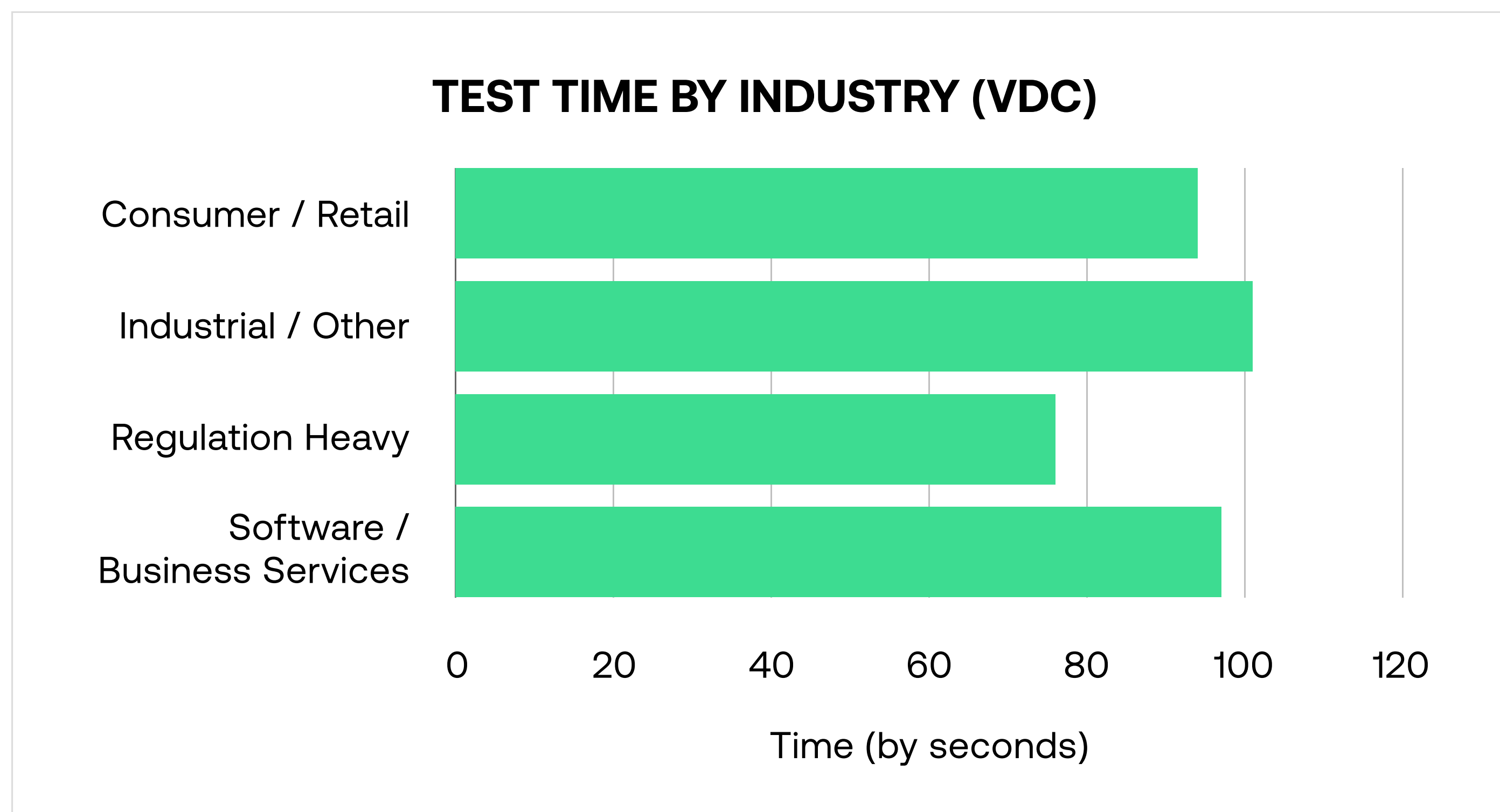
77%

Of desktop (Windows, Mac, Linux) pass the benchmark (24% in 2020)

71%

Of mobile (emulators, simulators) pass the benchmark (25% in 2020)

## Test Time: VDC By Industry



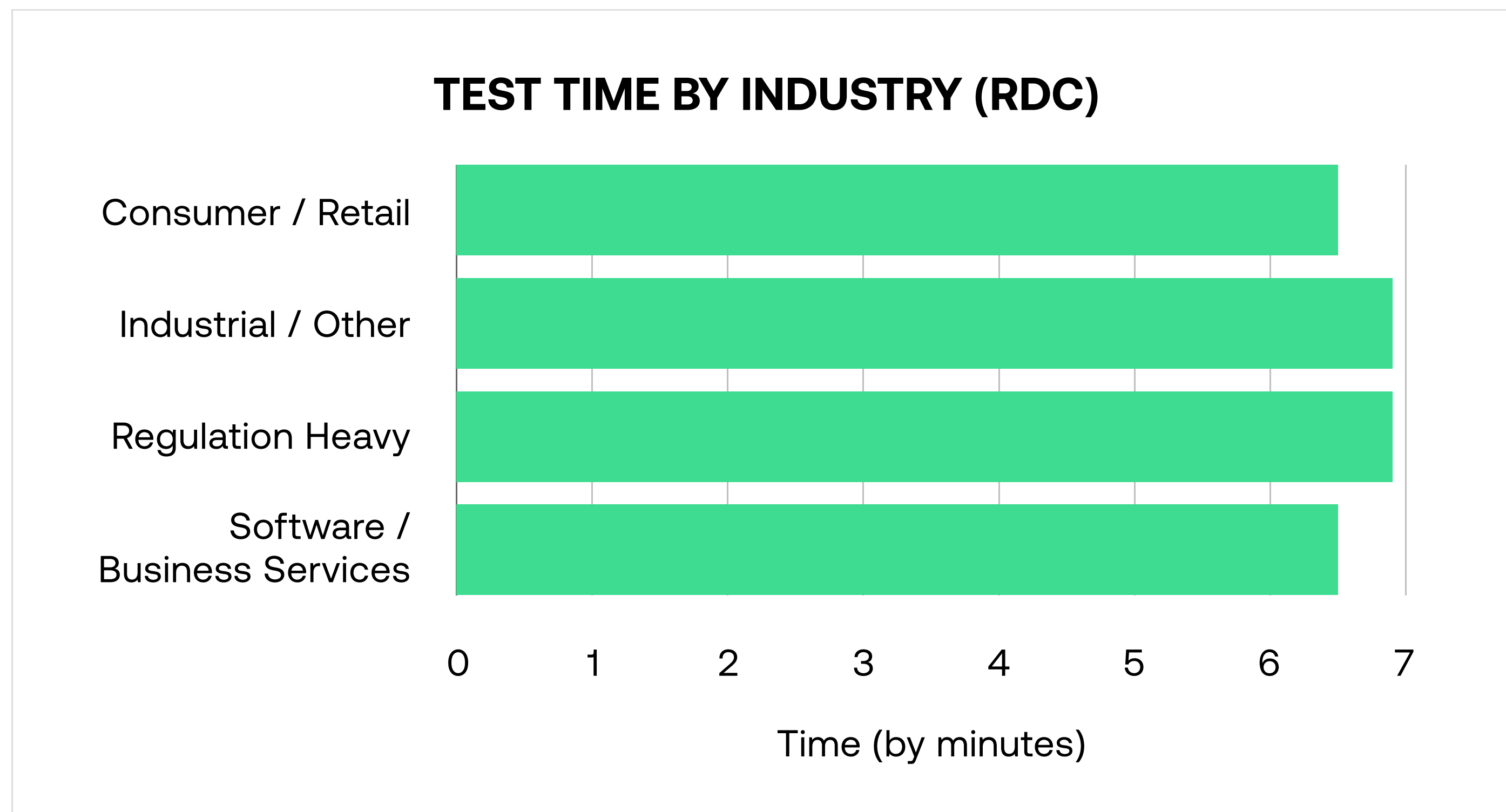
Regulation-heavy industries (like banking) had the highest number of organizations to pass the test run time benchmark of 2 minutes or less—coming in at 78%.

Banking and financial services organizations prioritize test speed due to the need to meet regulatory compliance and audits (updates have to be made in a timely manner). In addition to being compliant, online banking also stands as a core component to their business strategy, so the need to push code to keep online banking applications up to date is critical to beating our competitors, meeting customer expectations, and contributing to the bottom line.

A global financial service organization

[Read More →](#)

“Sauce Labs has helped us increase both quality and deployment frequency (4x more frequent) thanks to 60x faster test automation.”



For RDC test time across the board, 24% passed the benchmark with an average run time of around six minutes.

Software companies lead the way in meeting the test run time benchmark across industries. Since their main focus is software production, they invest heavily in technology and prioritize streamlining their development process. This enables them to drive effective test strategies that improve speed and overall efficiency.

When it comes to test run time, there are some important things to remember. In terms of efficiency, virtual devices operate in a more controlled environment, meaning they can be finely tuned for performance which reduces variability in test times and increases the predictability of outcomes.

Whereas real devices provide the most accurate reflection of user interactions and environments, which is critical for the final verification stages before release. However, these real-world scenarios introduce variables—like actual hardware performance under certain network conditions—that extend test times.

In order to leverage the strength of both VDC and RDC environments, organizations need to adopt a balanced testing strategy that utilizes both real and virtual devices. Implementing private device clouds can mitigate availability issues, ensuring devices are always accessible for testing.

It's important that companies stay focused on continuous improvement as well, reviewing and optimizing test scripts, investing in test automation tools, and adapting parallel testing to improve overall test times.

# Benchmark Metric #3: Test Platform Coverage

## Why it matters

### The need for comprehensive platform coverage

As the variety of mobile and laptop devices, operating systems, and browsers available to consumers continues to grow rapidly, the need for comprehensive test coverage has become crucial. It's essential to ensure that an application looks, functions, and performs as intended across as many platforms and user scenarios as possible. Comprehensive platform coverage is vital to digital confidence and keeping customers satisfied.

### Customers expect your apps to work

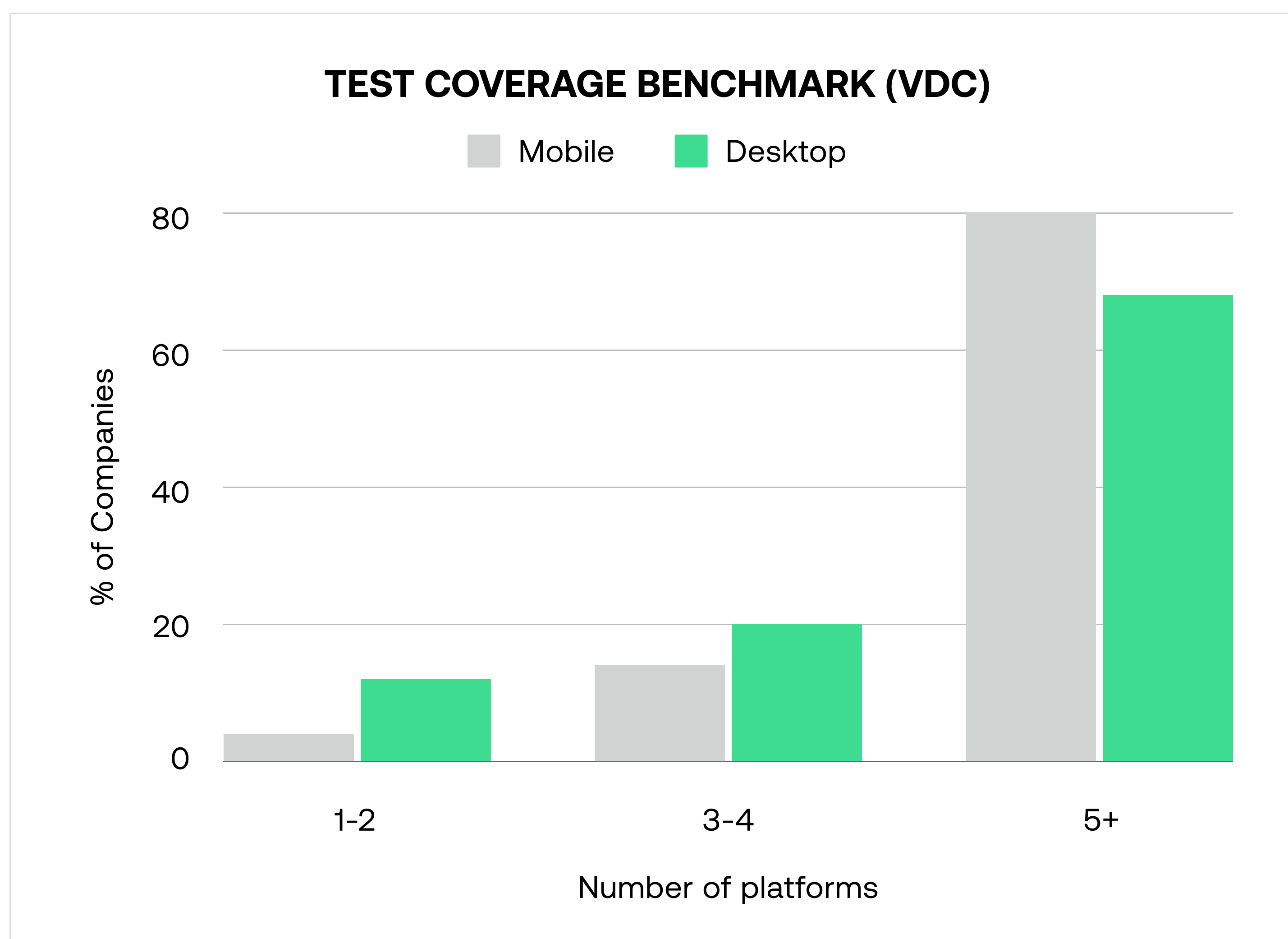
Every business is now a digital business, and web and mobile applications are the new storefront. The more devices, operating systems, and browsers on which a business can "open" their store, the more opportunities they have to connect with customers. Consumers expect your apps to work across all of their devices, from phones to laptops to desktops irrespective of the browser or the operating system they are using.

### Maintaining platform coverage efficiently

Maintaining a huge range of physical devices in-house is clunky and expensive. Utilizing cloud-based solutions like Sauce Labs VDC and RDC significantly reduces these expenses by providing access to platforms and devices on-demand. You can even take your test strategy one step further by collecting data about which devices, versions and operating systems your end-users engage with most often, and build a coverage plan, rooted in data.

### The Benchmark for Excellence

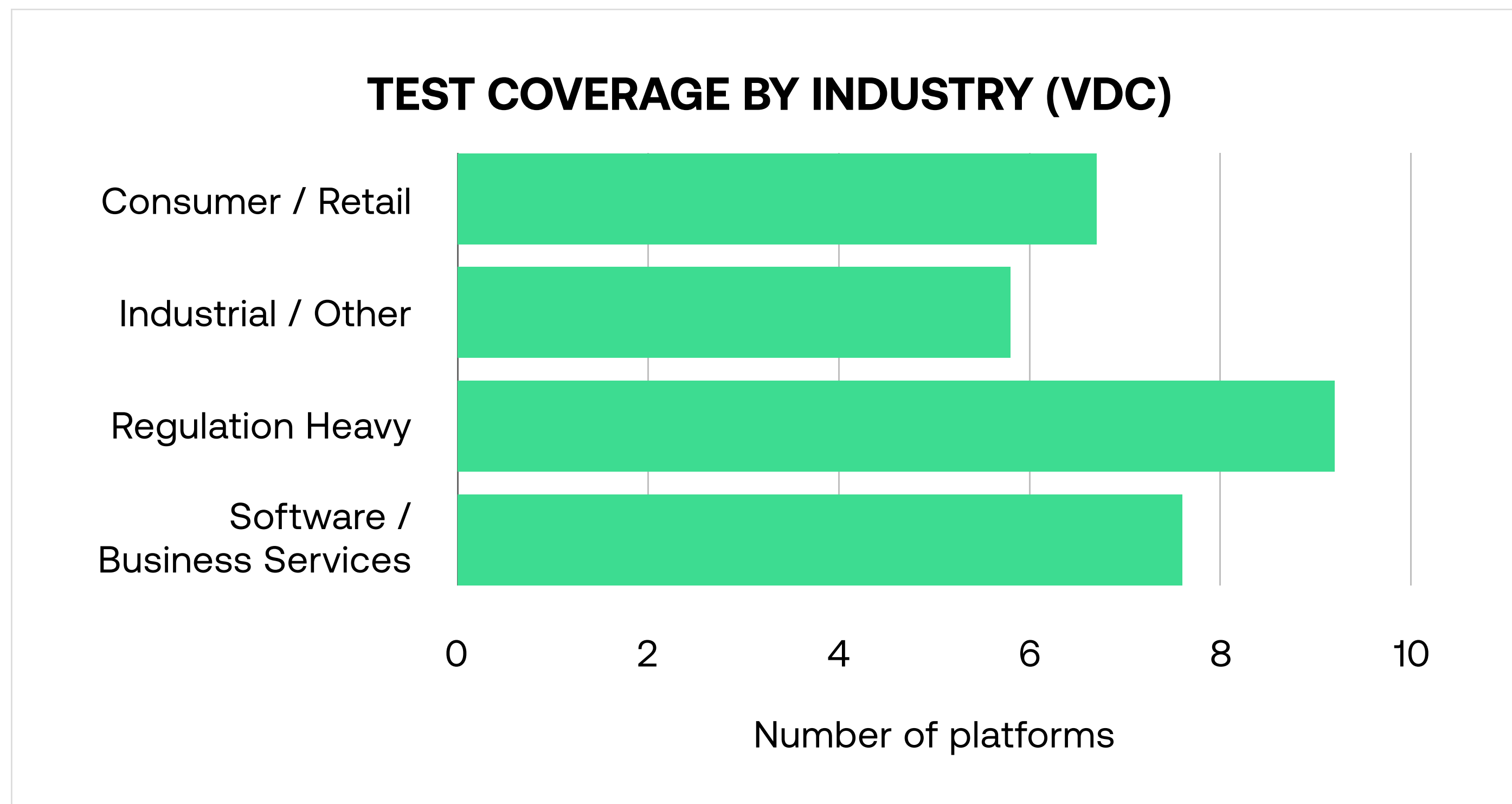
The Test Platform Coverage metric looks at the number of platforms against which an organization tests (device, browser, OS) and gives full marks if at least five platforms are included in their typical testing activity. For desktop tests, a platform is defined as any combination of an operating system and a browser. For mobile tests, a platform is defined as any device type of 5 real devices or greater.



#### How Organizations Performed:

- 72% of organizations passed the benchmark of 5 platforms or greater
- Average number of platforms for VDC: 7.67

## Test Platform Coverage: VDC By Industry

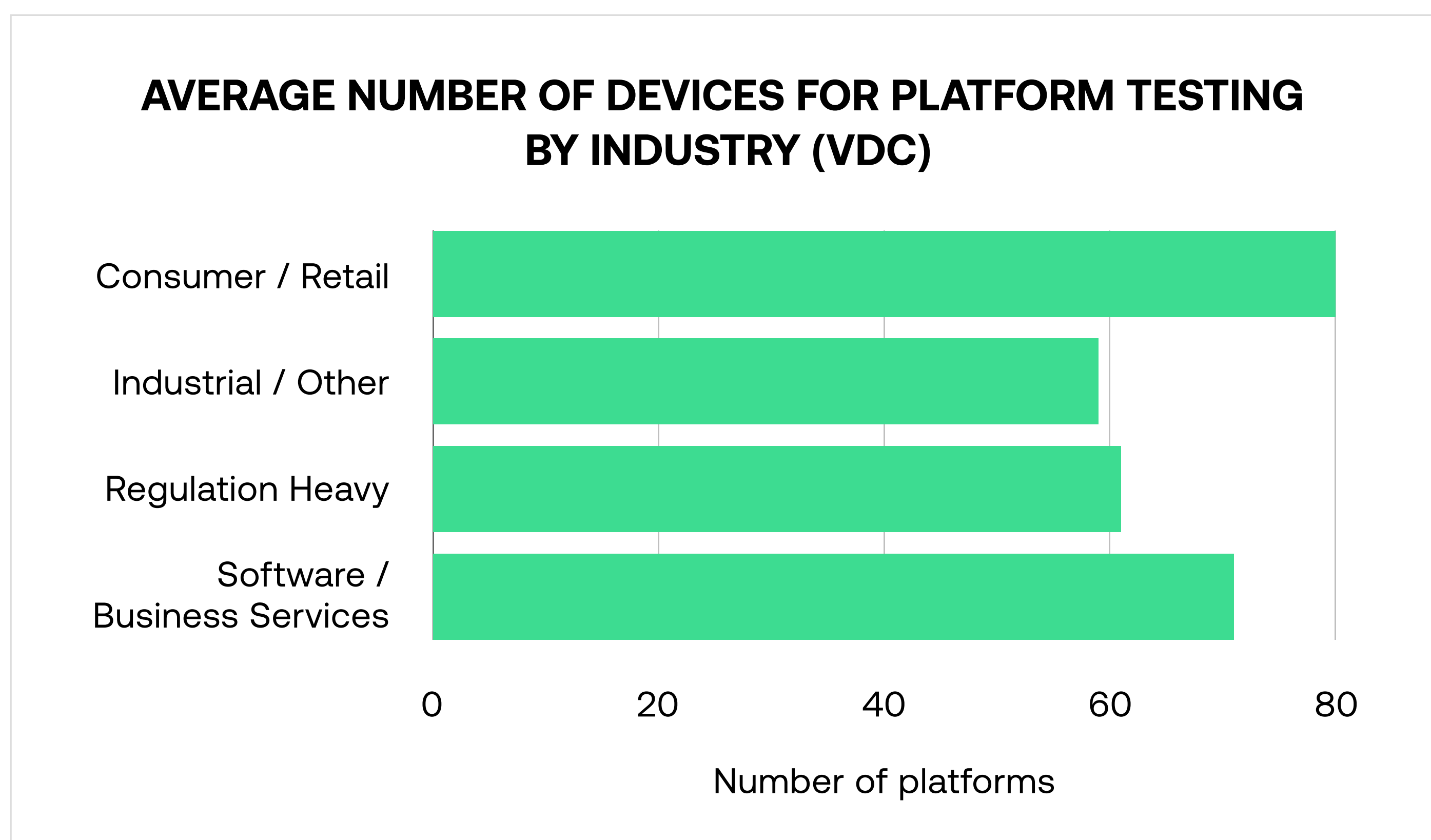


Banking and financial services organizations likely test for more user scenarios due to compliance requirements that make it necessary to have thorough testing across environments.



“Sauce Labs has helped us significantly improve quality metrics – increasing virtual device coverage by 37% while decreasing QA manual regression testing efforts across multiple browsers and devices.”

For test platform coverage on RDC, 79% of organizations reported testing on 10+ devices, with 59.5% testing on 30+.



The fact that consumer and retail organizations test on so many varieties of devices highlights the strong demand for consumer applications to work seamlessly across a wide range of platforms and user scenarios.



Walmart eliminates manual testing and infrastructure maintenance

“Thanks to Sauce Labs, we ensure our apps work on... over 300 unique real devices.”

## The Takeaway

Enterprises have clearly gotten the memo that user experience matters, and every experience counts. Surviving in the all-digital era means delivering software that works as intended whenever, wherever, and however customers wish to access it.

Four years ago, 14% of organizations tested on only two or fewer platforms. Today, that number has dropped to just 10%, with the majority now testing across at least three platforms. The percentage of organizations running desktop tests on five or more platforms has grown by 4% since 2020.

This year marks the first year we are reporting on RDC data, and the scale of coverage is striking. Almost 90% of companies are testing across 5 real devices, with over half testing on more than 30. One organization even tests across 675 distinct devices.

Additionally businesses are becoming more strategic with their coverage by leveraging data from Google Analytics. By understanding which devices and systems their customers use most, companies can prioritize testing on specific devices - like the iPhone 15 or the Google Pixel 8. This ensures the best possible experience for the end users.

# Winning with Continuous Testing

From the data, it's clear that organizations are making big strides in implementing continuous testing across their SDLC. And it's paying off. From the previous report, the number of organizations achieving the standard for excellence grew 40% on average across the three benchmark metrics.

When it comes to balancing these metrics, each organization needs to make choices that reflect their business needs. Your industry, user expectations, and business goals should all shape your testing strategy. There's no one-size-fits-all approach.

For example, if delivering the best user experience is critical for your business, testing across a wider range of platforms may be your top priority, even if it takes more time. But if speed to market is more critical, you might prioritize faster releases while maintaining quality standards.

Ultimately, the goal is to develop a testing strategy tailored to your business needs. We recommend leveraging the CTB data and the Sauce Labs team to find the right balance, which ensures that your testing strategy drives the success for your business.



"With Sauce Labs [continuous testing], we can make changes and updates pretty quickly. If an issue arises, we can test it fast, fix it and deploy the next day. The next morning the problem disappears and everyone is happy."

Ultimately, the goal is to develop a testing strategy tailored to your business needs. We recommend leveraging the CTB data and the Sauce Labs team to find the right balance, which ensures that your testing strategy drives the success for your business.

## Continuous Testing Recommendations

- Balance testing across your SDLC, understanding when to leverage virtual devices, real devices, as well as [Visual testing](#), [Mobile App Distribution](#) and [Error Reporting](#)
- Run [atomic](#) tests to improve test quality, reduce test run times, and achieve parallelization.
- Leverage [failure analysis](#) to surface your most common test errors, prioritize how frequently they are occurring, and gain insight into where you can focus efforts to improve your pass rates.
- Be sure to send the status of your tests as [pass/fail data](#), in order to capture insights over time.
- Adopt a [mobile-first strategy](#) in which your investment in mobile testing and development is proportional to current and anticipated customer usage.
- Seek the help of [experts](#). Testing is hard. Experienced testing experts can help you navigate the challenges.

---

Data Note: Tests with a 0 pass rate, indicating they timed out, and tests with null times have been excluded from the analysis.

For our industry breakdown, we followed ZoomInfo's standardized industry definition.

- Retail: companies that sell directly to consumers, including clothing, electronics, and consumer products.
- Software: companies that develop apps, SaaS products, and cloud services.
- Regulation-Heavy: include healthcare, finance, and energy companies that must comply with government regulations.
- Industrial sector: companies that manufacture goods or provide services, like construction and machinery.



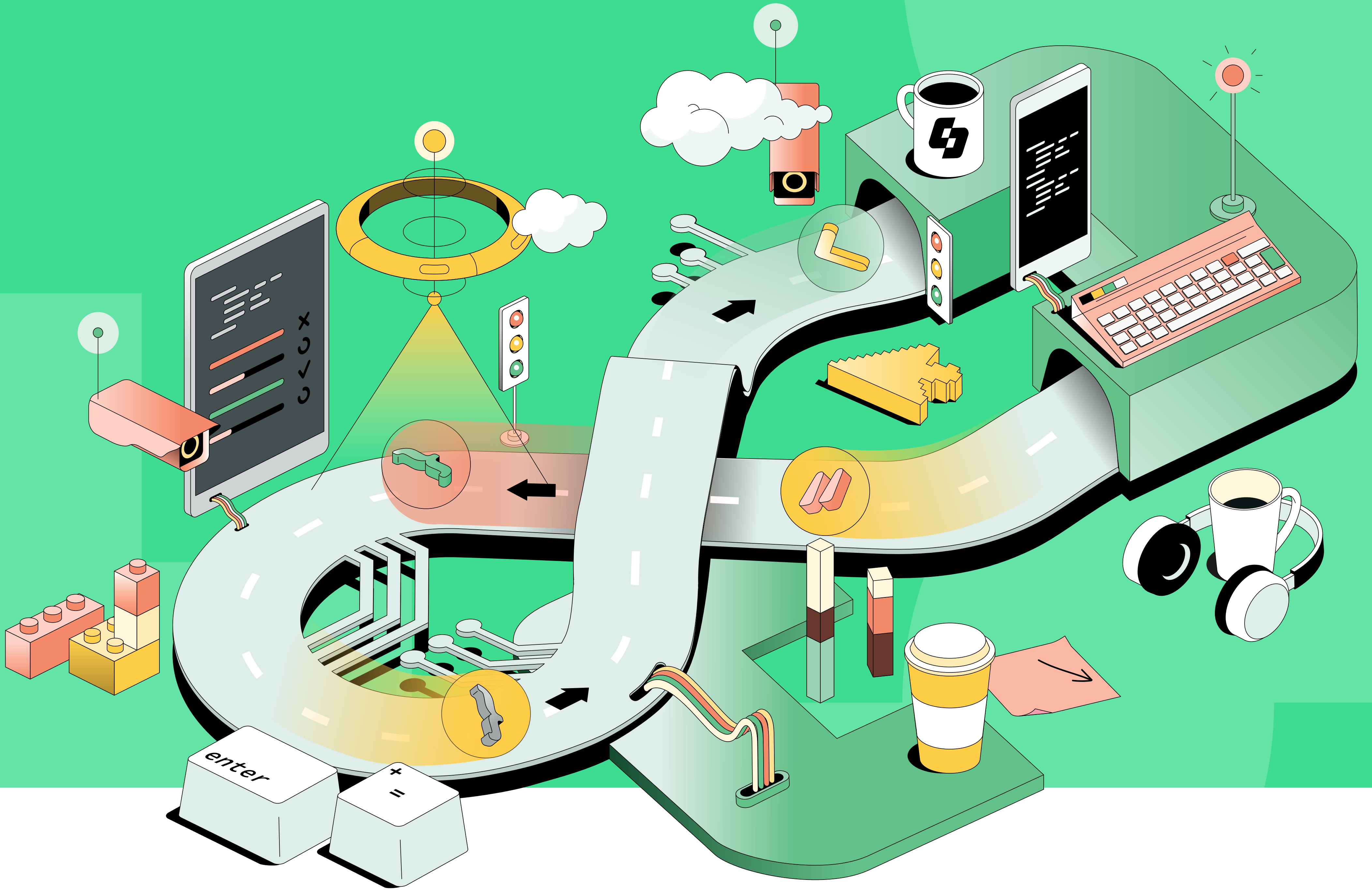
THE ROI OF SAUCE LABS VERSUS SELF-MANAGED TESTING

## Reduce maintenance and unlock innovation all while increasing customer satisfaction

Eliminate value-draining toll that self-managed test systems require. Calculate the ROI of a platform that gives teams more time to focus on delivering competitive advantage and customer loyalty.

[CALCULATE YOUR ROI](#)





# About Sauce Labs

Sauce Labs is the leading cloud-hosted platform for automated testing of web and mobile applications, enabling fast delivery of high-quality software across the development lifecycle. Founded by the creators of Selenium, Sauce Labs has been the testing leader for over 15 years and now runs over 1 billion tests annually. Trusted by Fortune 500 companies like Toyota, Walmart, Verizon, and Fidelity Investments, its scalable, secure platform supports testing across thousands of operating systems, browsers, and devices while meeting the highest compliance standards.

For more information, please visit

[saucelabs.com](https://saucelabs.com)