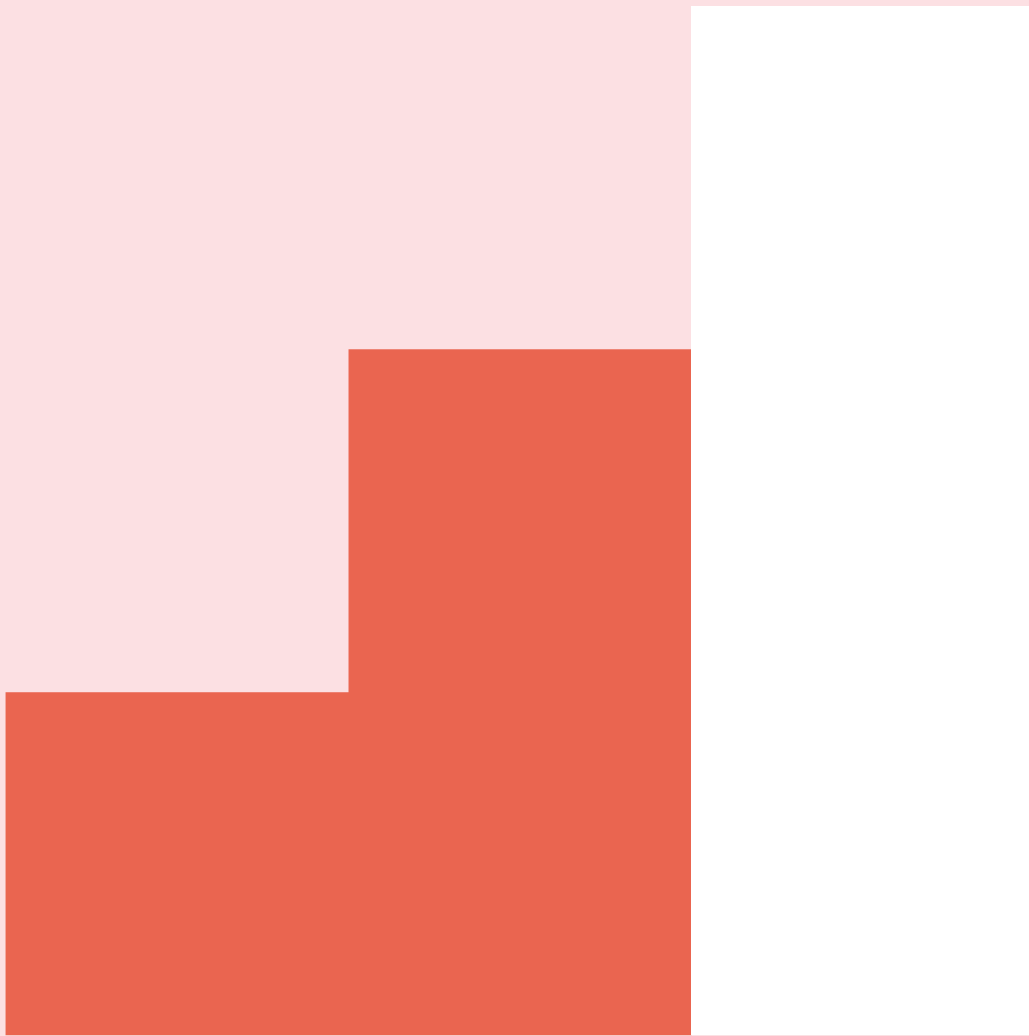


Zendesk Research

# Operational Benchmarking



## Abstract

### In focus: Operational benchmarking

Benchmarking has a long history in business: It's natural for companies to want to compare themselves to other businesses to give context and meaning to their own performance—and to understand where opportunities for growth and improvement exist. But frequently, similarities among companies in the same industry are merely superficial and don't offer a meaningful point of comparison for benchmarking customer service performance.

In this report, we address this shortfall through a cluster analysis that revealed four types of customer service operations—and within those, 12 distinct clusters, each characterized by a unique operational pattern. Each cluster is defined by workload, strategy, and resources, as well as a typical level of support performance that the component companies can be expected to achieve.

Looking at the four general types of support operations, first, we found the high-touch, consultative teams whose low ticket volume enables each agent to provide attentive and personable support. Next, we saw teams making large investments in support operations to mitigate a workload defined by very complex tickets. We also discovered teams with operational approaches that are still in flux—these teams could benefit from augmenting their strategy with one or more additional tactics. Finally, we found very operationally advanced teams, whose primary challenge is providing quality support at a high scale.

### 4 types of customer service organizations

#### **Type 1: Relationship builders**

These small teams provide a personal customer experience that customers love.

#### **Type 2: Masters of complexity**

Driven by detailed support requests, these companies have sophisticated customer service operations.

#### **Type 3: Late bloomers**

With an unbalanced approach to support, these companies have yet to realize their full potential.

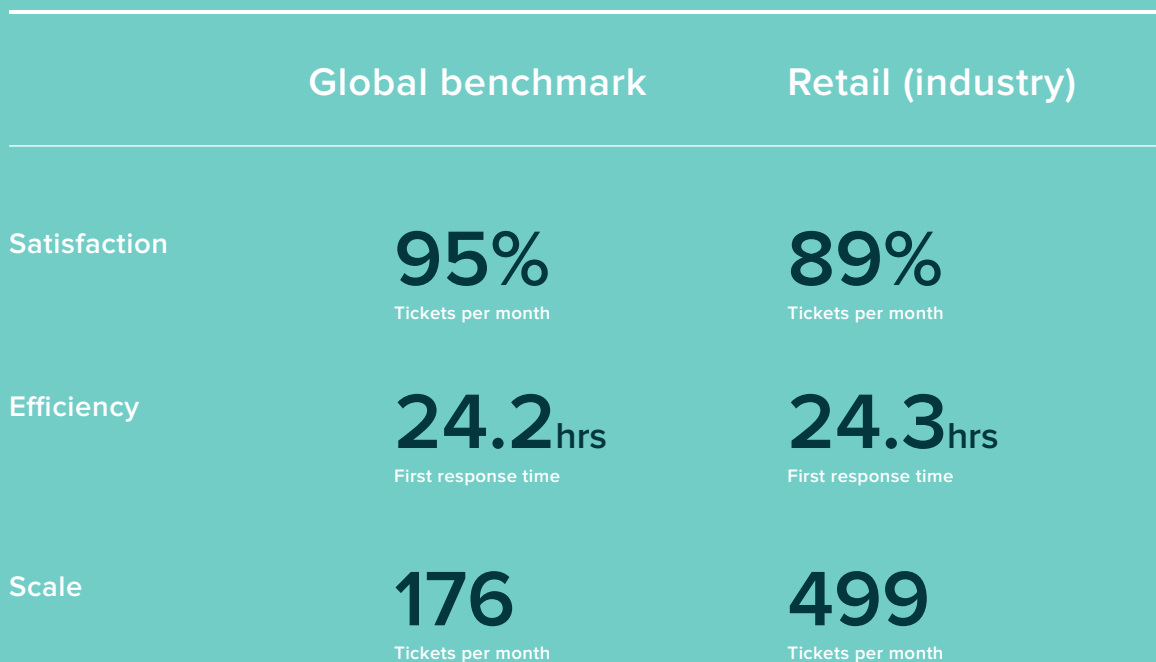
#### **Type 4: Captains of scale**

These teams set the gold standard for customer support operations.

## In search of better customer segments

In 2010, we started building the Zendesk Benchmark to give our customers a real, tangible way to not just measure customer service performance, but put it in a context that helps companies compare their performance to their peers. Unlike a survey or expert opinion, the Zendesk Benchmark is based on actual customer service and support data and focuses on three key performance indicators: 1) customer satisfaction, 2) first reply time, and 3) ticket volume (Fig. 1). Using these three performance indicators, Zendesk customers can compare their organizations to others that report to be in the same industry, of the same company size, or with the same target audience (all of which are self-declared by each organization).

**FIGURE 1: Global and industry-specific benchmarks aggregated from 25,000+ Zendesk customers**



While it's certainly useful to know how companies in similar verticals perform, today's diverse ecosystem of businesses makes these industry selections less black and white. For example, a foreign language learning software, like Babbel, helps millions of people around the world learn new languages. But what industry should Babbel self-select for the Zendesk Benchmark? Education? Software? Web applications?

This was the dilemma Stephan Seyfarth, Director of Customer Service at Babbel, faced with the Zendesk Benchmark. "The Zendesk Benchmark metrics were not remotely comparable to what our customer service organization has to deal with on a day-to-day basis," Seyfarth said. "With 100 employees and 50 part-time support agents, we serve millions of customers. The numbers that the Zendesk Benchmark provide based on our selected industry are far away from our numbers." Seyfarth's dilemma was that despite being in the education industry, the company's support operations had to handle over 2,000 tickets per day—well beyond the education industry's benchmark.

Seyfarth was not the only customer who voiced this concern, when we asked customers for feedback on the Zendesk Benchmark. We wanted to better understand what other performance indicators our customers wanted to see, as well as identify any improvements. We received 262 direct replies with feedback.

The top three critiques we heard were:

1. My operations are much different than others in my industry. How should I benchmark myself?
2. The metrics are simply not relevant to me, as I'm not a "traditional" customer service team. For example, I'm an HR team. I don't want to compare myself to companies in my company's industry, but rather other HR teams regardless of industry.
3. The benchmark metrics do not take into account my organization's business hours.

In response, we began exploring how to address these points of feedback. The biggest questions we asked ourselves were:

- Is there a better way of benchmarking that is not a strict comparison based on self-selected verticals but rather on how an organization or team uses Zendesk and behaves operationally?
- If so, then what are the common operational patterns for companies or departments that use Zendesk?

The goal was to create a way for teams using Zendesk, like the customer service group at Babbel, to compare themselves to other teams that use Zendesk in a similar way, even if they're not in the same industry (or aren't the same size or don't support the same type of customers).

This report is our first step towards discovering how a new usage-based clustering methodology can help us better understand the common denominators in the operational patterns of our customers, so we can curate benchmarks that:

1. Suggest more accurate comparisons and provide more relevant guidance to businesses that are at the intersections of multiple industries.
2. Build a more relevant means of comparison for teams that do not operate in a traditional customer support structure, like the entrepreneur at a one-man shop who wears all the hats or the internal team serving other employees within a company.
3. Deliver meaningful metrics that are rooted in how organizations support and interact with their customers.

It all starts with a sophisticated analysis of key customer service performance metrics to identify patterns among Zendesk customers.

## The building blocks of Zendesk's operational clusters

Cluster analysis is a machine-learning technique that allows us to group Zendesk Benchmark customers together based on operational metrics. First, we gathered data for each customer in our sample on seven different operational variables or dimensions. Then we combined the seven dimensions into two. This allowed us to plot all the operational data for each account onto a two-dimensional scatterplot and quickly identify which accounts cluster together (for more details on our analysis, see the methodology section of the appendix).

Clustering helps us understand which data points are similar in a complicated, high-dimensional space. By using this method to look at product usage data for companies and teams that use Zendesk, we can group our customers into clusters of companies that operate in a similar way. We can then use these clusters to benchmark for first reply time, customer satisfaction, and other performance metrics that are relevant to how a customer service organizations operate.

In creating our usage-based clusters across our customer base, we looked at similar operational patterns based on three cornerstones of customer service interactions:

1. How prepared a company is to interact with its customers
2. How a company plans and manages its support resources
3. How mature a company is in optimizing for support efficiencies

From these three cornerstones, we developed seven measures of performance to evaluate and compare our customers.

## Operational patterns

## Performance metrics

---

How prepared a company is to interact with its customers

**Number of new tickets** is the metric for the average number of inbound customer inquiries received by a customer support team on a monthly basis. Not only is this metric a measure of demand for support, it's also a big influencer on the key performance indicator of customer satisfaction (Fig. 2). As ticket volume rises, customer satisfaction tends to dip.

**Tickets per active agent** is the metric for the average number of tickets an individual agent is able to solve on a daily basis. This metric shows how much supply is available to meet the demand for support. Similar to number of new tickets, an increase in tickets per active agent also has a negative correlation with customer satisfaction.

**Agent comments per ticket** is the metric for the average number of agent comments on each ticket, which measures the difficulty and complexity of support inquiries. The complexity of support interactions can certainly vary depending on the type of business and can heavily influence how efficiently agents can support customers.

---

How a company plans and manages its support resources

**Weekday hours** is the metric for the median number of hours support agents worked on tickets during weekdays. This metric measures the number of hours that support was available to customers and correlates strongly with performance. We see that increasing weekday working hours correlates strongly with decreasing first reply time (Fig. 3).

**Weekend hours** is the metric for the median number of hours support agents worked on tickets during weekends. Being more available to customers is certainly a unique strategy that can vary between businesses and how they choose to run their support operations.

---

How mature a company is in optimizing for support efficiencies

**Number of business rules** is the metric for the average number of business rules (e.g., triggers and automations) that run on tickets. Optimizing workflows is a more sophisticated approach that mature support teams can employ to reduce response times and drive greater efficiencies—thereby improving customer satisfaction (Fig. 4).

**Self-service ratio** is the metric that measures Help Center views per new ticket. This metric is an indicator for how much a company has invested in helping customers find their own answers with useful knowledge base articles or community tips, which can help deflect the number of 1:1 customer interactions agents need to support.

FIGURE 2: An increasing number of new tickets tends to drive down customer satisfaction

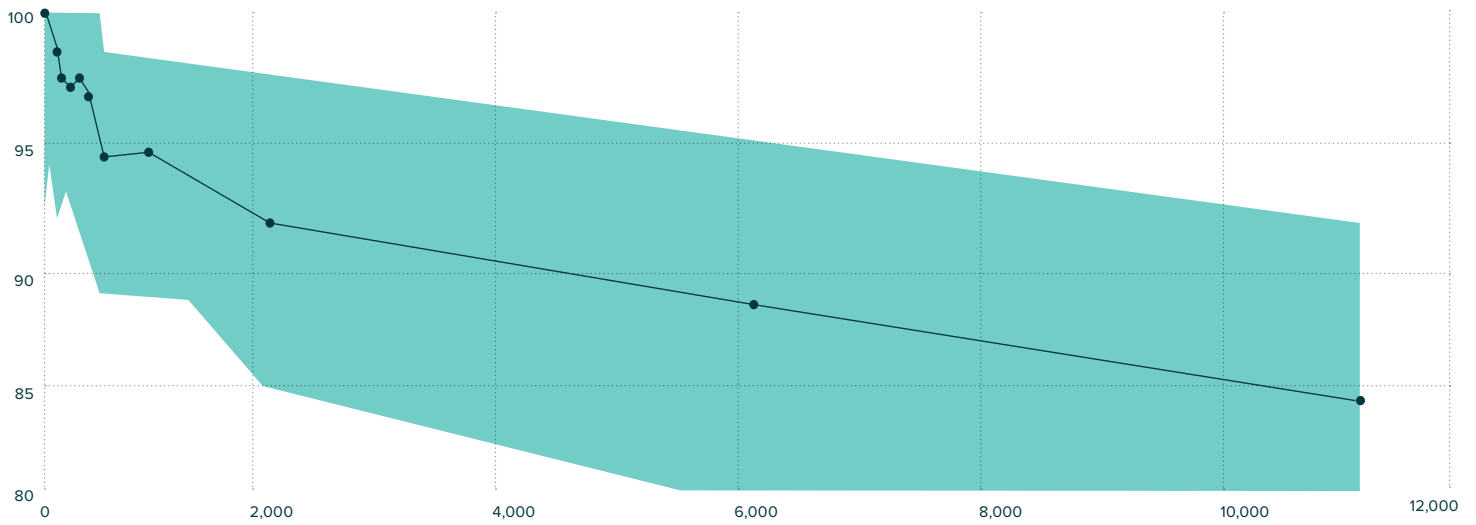


FIGURE 3: Increasing support hours during the week tends to decrease the time to first reply on tickets

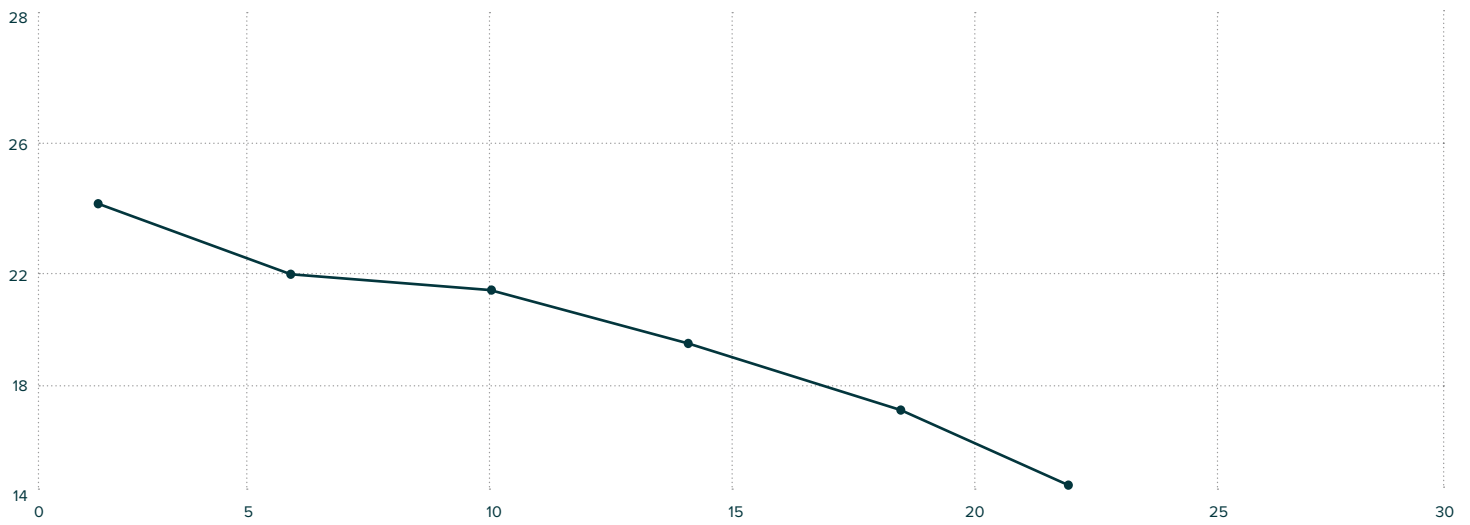
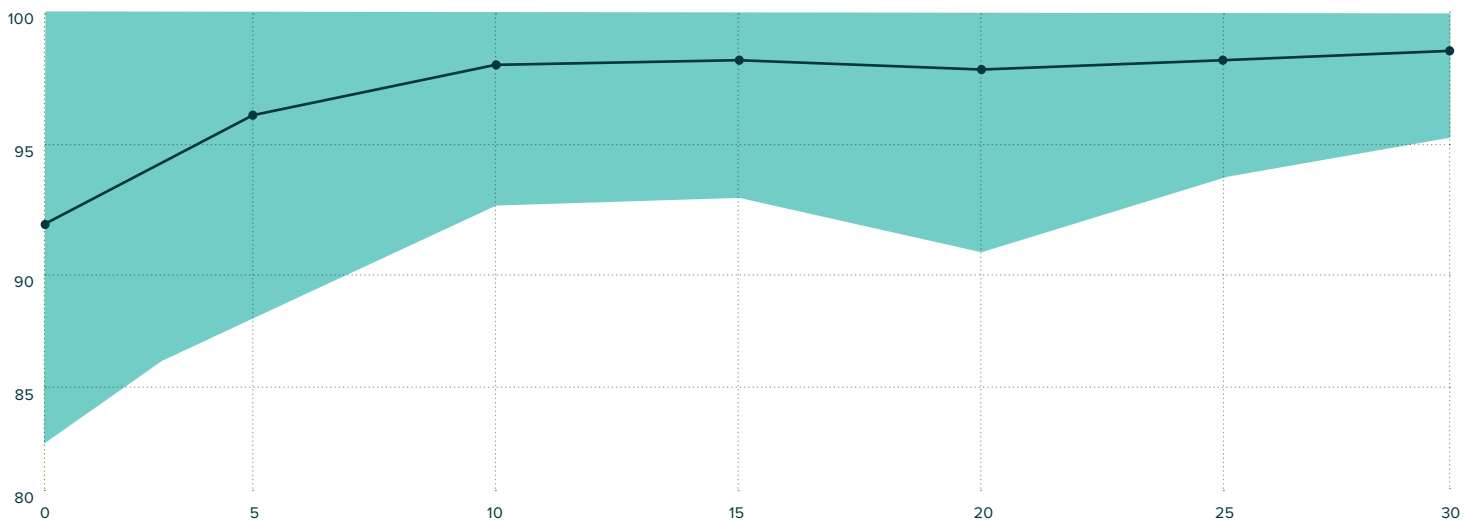


FIGURE 4: Increasing the number of business rules per ticket tends to improve customer satisfaction



## The 4 types of customer service organizations

Based on these seven performance metrics, we used clustering to identify which customers had similar support operational patterns and found 12 distinct clusters. We then grouped these clusters into four general types of customer service organizations, which we'll explore later in this section. Below you can see the cluster breakouts and how they compare across the seven dimensions.

**FIGURE 5: The 12 clusters based on our seven performance metrics—in order of ticket volume**

Number	New tickets	Comments/ticket	Ticket/agent	Weekday hours	Weekend hours	Biz rules/ticket	Self-service ratio
Cluster 6	Low	Medium	Very low	Low	No	Medium	Low
Cluster 11	Low	Very high	Very low	Medium	No	Very high	High
Cluster 4	Medium	Very high	Low	High	No	Very high	Medium
Cluster 7	Medium	Low	Medium	Low	No	Low	Very low
Cluster 0	Medium	Medium	Medium	Medium	No	High	Low
Cluster 3	High	Very low	Very high	Low	No	Very low	Very low
Cluster 1	High	Low	High	Medium	No	Low	High
Cluster 8	High	Medium	High	High	Yes	Medium	Medium
Cluster 2	High	Medium	High	High	No	Medium	Low
Cluster 9	High	Low	High	Medium	No	Low	Very low
Cluster 10	Very high	Very low	Very high	High	No	Low	Low
Cluster 5	Very high	Low	Very high	High	Yes	Medium	Medium

To better understand, let's look at a concrete example. At the beginning of this report, we talked about how Babbel does not fit the profile of the typical customer service operation in the education industry. So what customer service companies or teams should Babbel compare itself to if the comparison is based off the operational profile of the company?

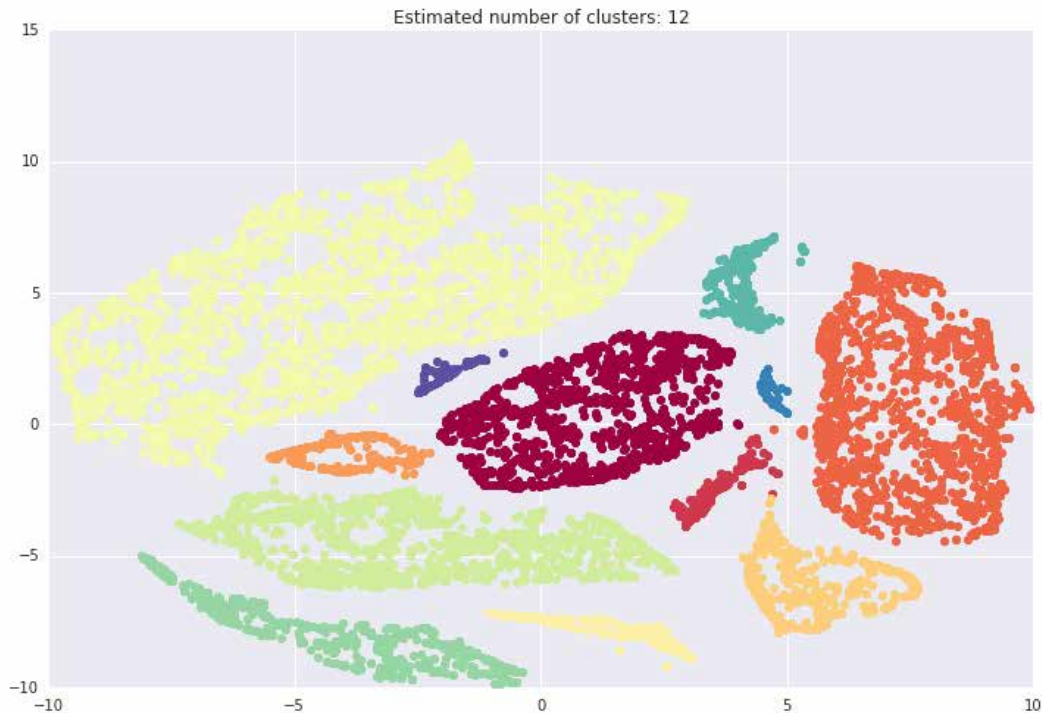
Falling into Cluster 5, Babbel handles a very high number of customer inquiries. Agents work long weekday hours and on the weekends. Each agent at Babbel has to handle many more tickets than their industry peers. But they are able to do so with very few touches, while deflecting some of the volume through workflow automations and self-service.

When comparing Babbel to other organizations in Cluster 5, we see that the benchmark metrics around customer satisfaction, first reply time, and ticket volume are much more relatable than those of the education industry (Fig. 6).

	Babbel	Cluster 5	Education industry
Customer satisfaction (50th percentile)	<b>91%</b>	<b>85.5%</b>	<b>98%</b>
First reply time (50th percentile)	<b>25.9</b> hrs	<b>18.8</b> hrs	<b>33</b> hrs
Tickets per month (50th percentile)	<b>61,537</b>	<b>9,332</b>	<b>79</b>

**FIGURE 6:** Babbel’s benchmark metrics compared to Cluster 5 and the education industry

“The new benchmarks from our cluster are more relatable and better reflect our business behavior,” Seyfarth said. “For us, it’s not really about the industry we’re in but rather about how we operate and support our customers. This is always our top priority.”



To better understand the characteristics of each cluster, the types of companies represented, and how they relate to one another, we grouped similar clusters together to create four high-level types of customer service operations.



## Type 1: Relationship builders

These small teams provide a personal customer experience that customers love

Cluster ID	Description	Tickets per month	First reply time	Satisfaction
6	Low-ticket volume consultancies and services	16	22.8 hours	100%
0	Mid-ticket volume software and services	139	24.6 hours	98.8%

Despite this group's below average first reply times, these teams lead the pack in customer satisfaction. Agents on these teams focus on providing great customer service instead of trying to rapidly solve a ton of tickets.

**Cluster 6:** Given their leisurely reply time of 23 hours, you wouldn't expect over 50% of these teams to have 100% satisfaction. Why? It must be the personal touch, shown through high interactions in each request. These agents are part of small teams (three agents on average) with a shared inbox or small consultancies where everyone in the business is answering customer questions, at least part-time. They only need to answer a very low number of inquiries because their below-average ticket volume, at 16 requests per month, is very manageable. But it's not just that their customers don't need help: These are mature support organizations that invest as much in business rule automation and self-service as their high-volume counterparts. Their customers are driven by quality, not speed. This cluster is predominantly made up of businesses with 100 or less employees and fewer than 20 tickets per month.

**Cluster 0:** These companies are operationally similar to Cluster 6, except on a larger scale. Pulling in ten times as many tickets, their inquiries are also more complex to solve, and they're more likely to be mid-sized teams and businesses. They exhibit the same pattern of high customer satisfaction despite somewhat longer first reply times, indicating their customers' preference for support quality over speed.

## Type 2: Masters of complexity

Driven by detailed support requests, these companies have sophisticated customer service operations.

Cluster ID	Description	Tickets per month	First reply time	Satisfaction
11	Low-volume, high-complexity self-service leaders	42	21.5 hours	100%
4	High-complexity software, consulting, and energy firms	200	22.6 hours	98.2%

Who gets the most complex and highly interactive tickets? It's these companies. Mostly business vendors, software providers, and consulting firms, they work long weekday hours providing the friendly, personal kind of customer service that customers with complicated questions appreciate. They tackle the tangle of support requests with more business rule updates per ticket than any other cluster. Additionally, they don't skimp on staff: Keeping their tickets-per-agent ratio low allows room for customer service experiences that shine through even the most opaque questions.

**Cluster 11:** These small teams, with just six agents on average, are the absolute leaders in self-service with a ticket deflection strategy that uses content to keep each agent's workload low. Conversely, the ticket first reply times can be high. If your tickets demand many comments, and your team receives between 20 and 50 per month, emulate this cluster's well-rounded style. Create a great knowledge base and automate support processes where possible.

**Cluster 4:** This cluster contends with a much larger monthly ticket volume than Cluster 11. Though highly optimized, they don't quite match Cluster 11 when it comes to business rule automation or self-service. The software, consulting, and energy companies that fall into this cluster would do well to review their support processes and knowledge bases.

### Type 3: Late bloomers

With an unbalanced approach to support, these companies have yet to realize their full potential

Cluster ID	Description	Tickets per month	First reply time	Satisfaction
3	Mid-volume, low-complexity, low self-service	519	25 hours	93.1%
7	Mid-volume, low-speed, low working hours	91	28.4 hours	97.6%
10	Ultra-high-volume B2C vendors and retailers	5,773	26.2 hours	86.3%
9	High-volume B2C vendors and retailers	551	222.3 hours	92%

Companies of every shape and size can be found in these clusters. What they all have in common is a missing piece in their customer service strategy.

**Cluster 3:** Despite having the simplest tickets, companies in this group have neglected self-service. Instead, they choose to spread their workload of around 500 tickets per month across too few agents. Organizations with simple tickets can gain huge returns by investing in their Help Centers. The resulting lower ticket volume, lower ticket-to-agent ratios, and faster first reply time can all positively influence customer satisfaction.

**Cluster 7:** Like the companies in Cluster 3, agents on these teams are spread a little too thin. They have a low number of agents for their ticket volume, and therefore have a relatively high ticket-to-agent ratio. These teams could benefit from further investment in both business rule automation and self-service, both of which are quite low given the scale of tickets they receive each month.

**Cluster 10:** These teams receive a massive amount of tickets per month but still have some work to do on optimizing support operations. Their high tickets-per-agent ratio could be mitigated by more investment in business rule automation and self-service. Another tactic that would give companies in this cluster an edge: offering 24x7, rather than 9x5, customer service.

**Cluster 9:** In terms of operational inputs, these teams are almost identical to the mature support organizations in Cluster 1 (see Type 4 below). They receive roughly the same number of tickets per month, and their tickets have a similar level of complexity; they also achieve a very similar ticket-to-agent ratio. They even have identical average numbers of agents and groups. However, unlike the self-service leaders in Cluster 1, companies in this cluster have the lowest self-service ratios of all. Teams in this cluster can benefit from adopting the strategies used by more mature companies

#### BioIQ, Cluster 10

BioIQ is a web and mobile app that helps users manage their health and wellness through appointments, results, and proactive management solutions. Historically, this company was compared with other healthcare providers. However, the company's support operations operates much more like a high volume freemium product like WhatsApp, which is in the same segment (Cluster 10).

at their scale; namely, deflecting tickets through self-service and offering 24x7 customer service.

#### Type 4: Captains of scale

These teams set the gold standard of customer support operations.

Cluster ID	Description	Tickets per month	First reply time	Satisfaction
1	High-volume B2C vendors	540	23.4 hours	91.3%
2	High-volume, high-speed retail and service providers	774	21.8 hours	95.7%
8	High-volume B2C vendors and retailers	1398	14.5 hours	92.6%
5	Ultra-high-volume B2C vendors and retailers	9,322	18.8 hours	85.5%

These companies take advantage of multiple support operation strategies. They have large ticket volumes, large agent teams, and their tickets are not simplistic; so it comes as no surprise that they all rely heavily on business rule automation and self-service to mitigate their sizeable workloads.

**Cluster 1:** All things considered, these teams run very well-oiled support machines. Not only do they do a great deal of business rule automation; they are the absolute leaders in self-service.

**Cluster 2:** These companies sit at the intersection of size and complexity: Their tickets come in at high volume and present a significant level of difficulty. They have responded to this double challenge by developing sophisticated support operations that incorporate high levels of business rule automation with very effective self-service documentation.

**Cluster 8:** These companies are nearly operationally identical to Cluster 2, with one exception: They provide 24x7 support instead of 9x5. This makes them one of the most well-rounded clusters, comparable to Cluster 5 in terms of the sophistication of their customer service operations.

**Cluster 5:** Because these teams operate at the largest scale when it comes to ticket volume, they have a unique operational profile. They employ 24x7 support as a strategy to reduce first reply time, and their high ratio of tickets per agent reflects an extremely efficient support team. Although their customer satisfaction is the lowest of all the clusters, this is most likely driven by the immense scale of their ticket volume; as illustrated in Figure 2, customer satisfaction tends to drop as ticket volume increases.

#### Modular Mining, Cluster 8

Modular Mining provides companies all over the world with solutions and consulting for mining operations. Their self-selected industry is consultancy, which seems like an apt fit. But Zendesk customers in this group are largely dominated by very small IT and tech consultancies.

In the operational benchmarking, Modular Mining is grouped with a much more relevant segment of high-volume and mature support operations that have invested in organizational structure, processes for issue escalation, and streamlined integration into their other business systems. These companies have a high bar for customer service despite managing a high volume of complex customer issues.

## What's your type?

Industry benchmarks—as well as company size and audience type—are still useful points of comparison for many companies. But as we've discovered in this quarter's benchmark exploration, many companies will better relate to companies that have similar operations, regardless of industry. Comparing companies across seven operational dimensions, instead of just one or two based on company profile, results in a benchmark that in many cases better relates to the company's resources, constraints, and business processes.

## Research methodology

All of the analysis described in this report was performed using the Python analytics stack with a deserved special mention for Pandas and Scikit-Learn. After pulling the required data from our Hive server, the first step in the clustering process was to bin and then normalize each of the features. Features were binned using 100 percentile bins, each representing a single percentile range of the data and the binned representation was then normalized to a zero mean and standard deviation of one.

As cluster goodness metrics such as Silhouette Score expect Gaussian distributed clusters and there was no guarantee that our data should take this form, we decided to follow an approach of dimensionality reduction and visual inspection to discover suitable cluster boundaries.

We visually inspected pairs of features plotted together as well as the first two principal components of the dataset obtained using principal component analysis. As there was no obvious cluster formation from this stage, we resorted to using a more advanced dimensionality reduction technique called t-Distributed Stochastic Neighbor Embedding (t-SNE). Using t-SNE ensured that accounts which were close to each other in the original seven-dimensional space of our dataset remained neighbors when projected into two dimensions. Visual inspection of the resulting two-dimensional data showed 12 well-defined clusters. We used Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to formalize the cluster boundaries and tag our data. Each cluster was then analyzed for its defining traits.

