

Essential Skills for Citizen Data Scientists

Published 13 March 2020 - ID G00716448 - 40 min read

Initiatives:

- [Analytics and Artificial Intelligence for Technical Professionals](#)

Shareable Summary

View & download slides



Citizen data scientists can catalyze advanced analytics initiatives in your organization. Data and analytics technical professionals can cultivate citizen data scientists by focusing on 19 essential skills, such as business subject matter expertise, advanced data literacy and ML model development.

More on This Topic

This is part of an in-depth collection of research. See the collection:

- [Roadmap for Data Literacy and Data-Driven Business Transformation: A Gartner Trend Insight Report](#)

KEY INSIGHTFULL DOCUMENT

Overview

Key Findings

- A citizen data scientist is an experienced business intelligence (BI) analyst who extracts predictive and prescriptive insights from data while not being as technically skilled as an expert data scientist. Citizen data scientists occupy a hybrid, boundary-spanning role, not a distinct stand-alone position in the organization.

- Citizen data scientists exploit modern analytics and business intelligence (A&BI) platforms, which increasingly have augmented analytics capabilities and integrations with data science and data preparation tools.
- Technical professionals hoping to revamp their careers have the opportunity to grow into the citizen data scientist roles via independent learning, intuitive augmented analytics platforms and the support of a community of practice.
- Technical skills are not enough. Citizen data scientists must develop deep business subject matter expertise, collaboration skills, intellectual curiosity and an ability to learn independently.

Recommendations

Data and analytics technical professionals seeking to cultivate citizen data scientists in their organization should:

- Identify experienced self-service BI users in the organization and invest in upskilling them into a citizen data scientist role.
- Extend an existing A&BI community of practice to include advanced analytics initiatives in the organization and support access to additional training.
- Invest in robust modern A&BI platforms and data preparation tools, and encourage would-be citizen data scientists to master the extensive capabilities of these products.
- Delineate the boundary between self-service BI and data science work and position citizen data scientists to work at the frontier of self-service BI to assist and collaborate with expert data scientists.

Skills Analysis

As organizations mature their data and analytics programs, they strive for advanced analytics capabilities with data science and machine learning (ML). Often, organizations face an evolutionary progression from traditional reporting, to self-service BI, to advanced analytics, and moving from descriptive to diagnostic to predictive and prescriptive insights with their data. A maturing organization may at some stage decide to hire expert data scientists to explore new advanced analytic terrain, but in the lead-up to that point, a small group of sophisticated self-service BI users typically fills a bridging role. These sophisticated BI users are called citizen data scientists. Depending on the technology used, they engage in some ML model development activities in an attempt to derive advanced insights from data beyond what typical BI consumers encounter in their day-to-day work. A typical self-service BI user is involved in data preparation, analysis and visualization to answer important business questions. A citizen data scientist, however, does all of this plus takes part in creating ML models and executing advanced statistical analysis to accelerate and automate data insights.

Citizen data science is defined as an emerging set of capabilities and practices that allows users to extract predictive and prescriptive insights from data while not requiring them to be as skilled and technically sophisticated as expert data scientists.

You will not find organizations posting jobs for citizen data scientists, because it is not a defined position in an organization. Rather, it is a role — a kind of hybrid set of duties that extend the capabilities traditionally possessed by BI users in order to help build and support advanced analytics capabilities in the organization. If we consider a spectrum where the typical self-service BI user engages in moderately sophisticated analytics and an expert data scientist engages in highly sophisticated analytics, the work of a citizen data scientist would lie somewhere in between.

Given its newness and hybridity, the exact scope of duties for the role of the citizen data scientist are the subject of some debate. Much of this debate can be chalked up to an organization's size and maturity. In smaller organizations (or any size organization with relatively small analytics programs or small IT teams), people have to wear multiple hats to begin with. A citizen data scientist in a smaller organization, then, would likely be involved in an end-to-end advanced analytics process, from data preparation on through to ML model development and deployment. In a larger organization (or any size organization with relatively large analytics programs or large IT teams), a citizen data scientist may focus more exclusively on a few tasks — such as ML model prototyping. Others on a multidisciplinary team will then handle other aspects of the workflow.

Analytic maturity is another variable in the debate about a citizen data scientist's duties. Organizations with relatively immature analytics programs are less likely to have expert data scientists on staff. This means those serving in citizen data scientist roles are the most advanced analytics users in the organization and must take on a broader scope of work when it comes to developing and deploying statistical models. However, in highly analytically mature organizations, which may have a sizable team of expert data scientists on staff, citizen data scientists play more of a supporting role to a more robust advanced analytics workflow. In the larger or more analytically mature organization, citizen data scientists are likely an important cog in an advanced analytics machine. But in the smaller or less analytically mature organization, the citizen data scientist is the advanced analytics machine.

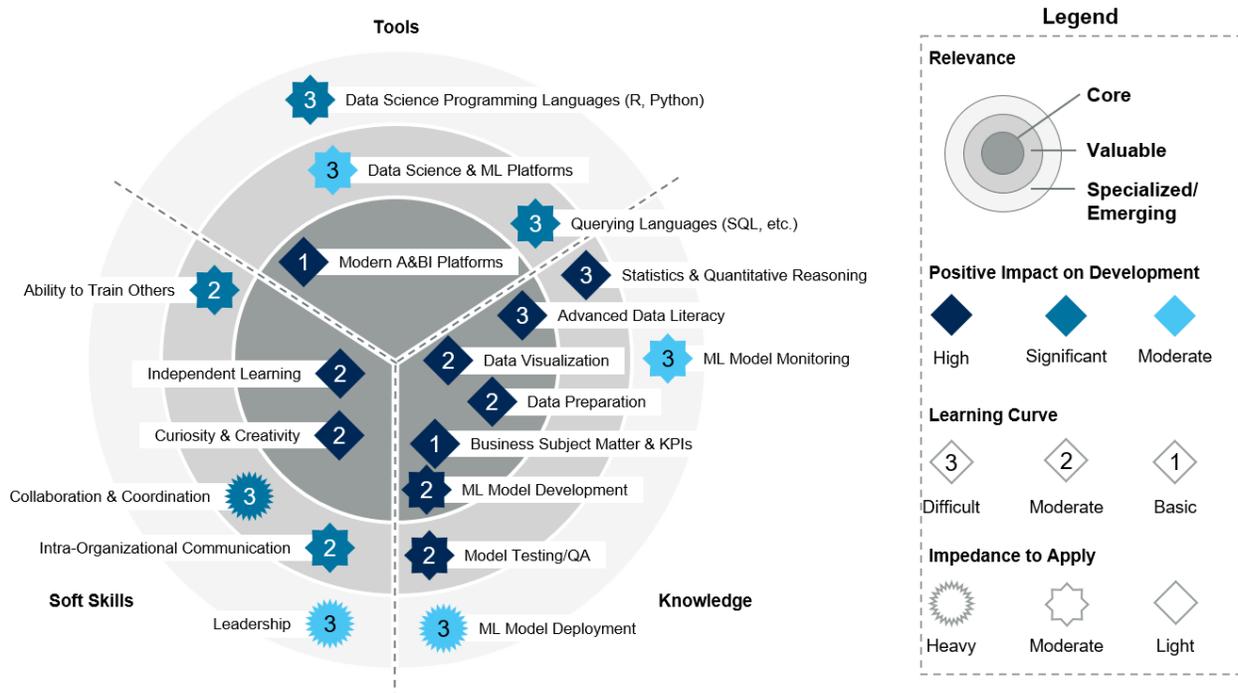
Despite the variance in duties across different organizations, there is still a common base of skills and knowledge a citizen data scientist must possess. This research thus answers the following question:

What are the essential skills for citizen data scientists?

Citizen Data Scientist Skills Scope

Figure 1 shows the 19 essential skills for citizen data science. Each skill is classified by relevance as core, valuable or specialized/emerging. Then, its essential characteristics are assessed. Each skill is rated in three dimensions: positive impact, learning curve and impedance to apply.

Figure 1: Citizen Data Scientist Skills Scope



Source: Gartner
ID: 716448



The 19 skills depicted in Figure 1 are divided into three focus areas: tools, knowledge and soft skills. A skill’s relevance to citizen data science is depicted by its location on one of three concentric circles. The inner circle of the graphic represents core citizen data science skills that should be learned and put into action immediately. The middle ring represents valuable skills to begin learning, and the outer ring is for skills that may be necessary to learn in the future or for specialized situations relevant to a particular organization.

The shading of the skill points in the graphic represent the level of positive impact the skill will have on a citizen data scientist — the darker the shade the more positive the impact. The number inside the skill point represents the level of difficulty that may be encountered when trying to learn a particular skill — the higher the number, the steeper the learning curve. Finally, the number of points on the skill point shape represents the impedance to apply a particular skill in one’s current organization or data and analytics architecture. The more points on the shape, the more resistance you will encounter to begin applying the skill successfully.

Essential Skills for Citizen Data Scientists

Citizen data science requires a lot of skills. At its essence, a citizen data scientist is an experienced self-service analytics platform user with an added overlay of ML skills and knowledge. Indeed, this is the common — though not the only — path to becoming a citizen data scientist. The typical pathway involves the most advanced users of an organization’s self-service BI platforms endeavoring to develop ML models to achieve more sophisticated data analysis. Other positions in the organization can grow into the citizen data science role — say, a database administrator or a data engineer or an application developer — and they may bring their own unique skill sets to the table. But they will ultimately need to understand the core competencies of analytics and business intelligence to truly excel as a citizen data scientist. Thus, peeling back some of the ML-specific skills from this slate of 19 essential skills for citizen data scientists would leave a good picture of what a typical modern BI analyst would look like. Below, we examine the skills one at time, beginning with the most crucial skills.

Core Skills

Table 1 lists the eight core skills for citizen data science. Citizen data scientists need to have these core skills for baseline competence in this role, since mastery of this particular mix of tools, knowledge and soft skills allows them to accomplish the day-to-day work of citizen data science.

Table 1: Core Skills for Citizen Data Scientists

Enlarge Table

Skill	Focus Area	Impact	Learning Curve	Impedance
Modern A&BI Platforms	Tools	High	Basic	Light
Business Subject Matter and Key Performance Indicators (KPIs)	Knowledge	High	Basic	Light
Data Visualization	Knowledge	High	Moderate	Light
Data Preparation	Knowledge	High	Moderate	Light
ML Model Development	Knowledge	High	Moderate	Moderate
Advanced Data Literacy	Knowledge	High	Difficult	Light

Independent Learning	Soft Skills	High	Moderate	Light
Curiosity and Creativity	Soft Skills	High	Moderate	Light

Gartner (March 2020)

Modern A&BI Platforms

A citizen data scientist must intimately know at least one modern A&BI platform. Modern A&BI platforms, such as Microsoft Power BI, Tableau, Qlik, and ThoughtSpot, emphasize self-service analytics with easy-to-use interfaces and administrative capabilities to enable governance. These A&BI platforms enable typical business users to access data sources; generate their own analyses, reports and dashboards; and share that user-generated content with colleagues.

Increasingly, modern A&BI platforms have developed augmented analytics capabilities that give business users the ability to discover data and produce sophisticated visualizations through natural-language Q&A, chatbot, search and other capabilities. Many modern A&BI platforms have rolled out automated machine learning (autoML) capabilities that allow for the creation, training and testing of simple ML models through easy-to-use visual interfaces without deep technical expertise. Significant data literacy and knowledge about business subject matter is still necessary to make sense of autoML outputs (more on this later), but the actual work of creating ML models is fairly simple with today's modern A&BI platforms.

Advanced analytics and ML model development using the built-in, user-friendly augmented analytics capabilities of mainstream A&BI tools enable citizen data science. Thus, it is important that a citizen data scientist understand at least one modern A&BI platform extremely well and know how to exploit the more advanced features of these platforms in order to do more sophisticated analysis. Modern A&BI platforms are relatively affordable and widespread across most businesses today. Mastering a modern A&BI platform is a high-impact move you can make without encountering many barriers or a steep learning curve. For more information, see:

- [“Demystifying the Analytics and BI Space”](#)
- [“Demystifying the Data Science Lab and AI Hub for Business Analytics”](#)
- [“Using Augmented Analytics to Boost A&BI”](#)
- [“The Evolving Capabilities of Analytics and Business Intelligence Platforms”](#)

Business Subject Matter and KPIs

Expert data scientists often find themselves juggling a variety of data science projects from various parts of the business. They dwell in the more abstract statistical level of a particular project without an expert understanding of a particular domain, such as HR or finance or

marketing. But because citizen data scientists typically emerge from self-service BI consumer roles located down in particular lines of business, they bring with them a deep understanding of specific business subject matter and the KPIs associated with that domain. They acquire a special kind of data literacy relevant to domain-specific business problems, both what the KPIs are and where to find relevant data to measure those KPIs (more on data literacy below). Data scientists need thorough knowledge of their particular subject matter in order to intelligently discover, prepare and analyze data and to build ML models that answer questions relevant to a line of business' KPIs.

This is precisely where citizen data scientists fill a gap in the knowledge of an analytically mature organization's expert data science team. In the larger and more mature organization, the citizen data scientist can serve as a subject matter expert on a larger data science team. They can help to translate domain-specific business knowledge to expert data scientists to inform models and help to prototype models and prepare training data.

Business subject matter expertise is a requirement for a citizen data scientist. But, at least for the typical self-service BI user aspiring to become a citizen data scientist, this expertise comes from the experience gained on the job as an analyst. Thus, it is rated here as having a basic learning curve with light impedance. However, those striving for a citizen data science role without a business intelligence background should note that additional learning and experience will be necessary to master a particular business subject area and a company's important relevant KPIs. Independent research, immersion in trade publications, participation in industry associations, informational interviews with subject matter experts, and possibly more formal training/schooling will be necessary to make up for the missed knowledge that comes from years of experience working as a business analyst in a particular domain.

Data Visualization

Understanding data visualization often goes hand-in-hand with mastering a modern A&BI platform, but the two skills are in fact distinct, and access to a robust A&BI platform is not necessarily a requirement to understanding data visualization. Data visualization is about the ability to present insights from data in a visual way, as elegantly and clearly as possible, to suit an audience and use case. Mastering data visualization is about understanding how data analysis marries with aesthetics and usability to help communicate something notable about data (see Figure 2).

Figure 2: Data Visualization Links Audience, Data Type and Visualization Type

Steps for Selecting the Right Data Visualizations



Source: Gartner 2019
ID: 383886

Gartner

Modern A&BI platforms do much of this work for you, generating beautiful graphs and tables from data with relative ease. But citizen data scientists should also master more sophisticated forms of data storytelling. They should know how to build single custom visualizations that tell complete stories about a business question or how to construct a series of single visualizations that collectively advance a storyline to answer a business question. A connected skill here is advanced data literacy (more on this below). Citizen data scientists should have a deep enough understanding of the contours of the data and its relevance to a particular business domain to be able to decipher what an automatically generated visualization truly means. They should be able to determine whether the generated visualization is actually the best way to express the insight visually. And they should be able to fix it or build a custom alternative if necessary. Mastering data visualization can have a high impact on one's ability to become a citizen data scientist, and there are few hurdles standing in the way of gaining this knowledge. But the learning curve can be difficult for a business analyst who may not have much experience customizing a visualization or for a nonbusiness analyst who lacks deep knowledge of a particular business domain. Mastering the ability to communicate visually about data may also compensate for shortcomings one may have in written or spoken communication abilities. Find inspiration and freely available tools and lessons online, such as at [D3.js](#). For more information, see:

- [“Communicate Effectively With the Right Data Visualizations”](#)
- [“How to Get More Value From Data Visualization”](#)
- [“Beyond BI Reporting: Engaging Decision Makers Through Data Storytelling”](#)
- [“Building Data Visualization Into Your Modern Business Intelligence Platform”](#)

Data Preparation

Like data visualization, data preparation is a core skill that a typical self-service BI user would be expected to have. Data preparation saves time and improves the ability for analysts to generate insights from data (see Figure 3).

Figure 3: Data Preparation Reduces the Time to Insight for Analytics

What Is Data Preparation?

Data preparation reduces the time to insight for analytics and operational use cases



Data preparation is an iterative, agile process for *finding, combining, cleaning and transforming raw data into curated datasets* for self-service data integration, analytics/BI and data science use cases.

Source: Gartner (April 2019)
ID: 386354

Gartner

Also like data visualization, many leading A&BI platforms and specialty data preparation tools have evolved to make this kind of work simpler, more visual and more intuitive. It is more important for an aspiring citizen data scientist to understand data preparation from a conceptual and procedural standpoint than to master any one data preparation tool, however. That is, quite sophisticated data preparation can be done with even the simplest, ubiquitous tools — such as Microsoft Excel. This knowledge can translate to advanced analytic work as a citizen data scientist just as well as having mastery of a dedicated data preparation tool like Alteryx or Tableau Prep.

Citizen data scientists must understand at a deep level how to find, combine, clean, and transform raw data into curated datasets for analysis. Data preparation involves a number of techniques, including data exploration, profiling, cleansing and augmentation. They can gain experience with this process through the normal end-to-end work of a self-service BI user (connecting to data sources, combining and transforming data, and generating reports and dashboards from the data). They can also gain this experience by familiarizing themselves with data preparation capabilities native to modern A&BI platforms and with stand-alone data preparation tools.

For more information, see:

- [“Drive Data Scientists’ Productivity With Data Preprocessing Techniques”](#)
- [“Market Guide for Data Preparation Tools”](#)
- [“Peer Lessons Learned: Implementing Data Preparation Tools”](#)

ML Model Development

An ML model is the result of training an algorithm on a set of data. The model is a mathematical description of a real business process or question that finds patterns and

significant insights in data. Engineering and executing ML models is a key work product that sets a citizen data scientist apart from a typical self-service BI user. The level of sophistication in the ML models citizen data scientists create may not be on par with what an expert data scientist can do. However, many advances can be made with some of the most common ML model types. Learning to build a model and train it is a core skill to learn.

Increasingly, augmented analytics capabilities in modern A&BI platforms are democratizing access to the ML model development process to non-technically-savvy business users. These autoML capabilities allow users to walk through a series of steps to configure a model, select a model type and choose predictive signals to train a model. There are many guardrails and assumptions built in to the autoML features that come with modern A&BI platforms, meaning that it works well with some of the most common ML model types and use cases. More innovative, groundbreaking ML model engineering would require more advanced tools for sure, but the accessibility of creating a functional ML model with an A&BI platform makes these platforms appealing go-to toolsets for citizen data scientists. Citizen data scientists should become proficient in the built-in autoML capabilities of the modern A&BI platforms they already use in their organization. They should also learn more about the general concepts behind building and training an ML model.

For more information, see:

- [“Augment Data Science Initiatives With AutoML”](#)
- [“A Guidance Framework for Operationalizing Machine Learning for AI”](#)
- [“Market Guide for Augmented Analytics Tools”](#)

Advanced Data Literacy

Data literacy is a broad concept. At its most basic level, data literacy can be seen as getting an organization to embrace data-driven decision-making in the first place. This requires a shift from operating purely on intuition, tradition and experience to seeking out data to explain what is happening in the business and to justify future actions. An intermediate level of data literacy can be seen as a more creative and critical approach to data at hand and a desire to acquire and analyze more data to pursue “what if” questions and test hypotheses. Getting individuals in an organization to intermediate levels of data literacy is itself a challenge.

Advanced data literacy, however, is an ability to deconstruct and question the lineage and underpinnings of data and analyses. It is about intimately knowing the data itself, the way the data is modeled, the kinds of analysis being performed on the data and the business context (the subject matter and KPIs) surrounding the data to know if something is right or wrong. Functionally, advanced data literacy would mean that someone using augmented analytics tools would know enough about the data and the context to know if something did not look right in the resulting autoML model or visualization. It means knowing enough about what is “under the hood” to be able to diagnose a problem in the data and analytics workflow just by seeing the outcome.

The learning curve for developing advanced data literacy is steep. Attaining advanced data literacy requires not only subject matter expertise to understand the business context for a dataset and analysis, but also a finely tuned critical thinking mind. Critical thinking can be cultivated through a skeptic posture toward new information, a desire to understand the inner workings of complex systems, practice with problem solving, attention to detail and a commitment to precision. Much of this is developed through on-the-job experience, tracing data lineage in complex data visualizations and a commitment to ongoing reflection and self-improvement.

For more information, see:

- [“Build a Data-Driven Enterprise”](#)
- [“Toolkit: Enabling Data Literacy and Information as a Second Language”](#)
- [“Toolkit: Curriculum for Data Literacy Training Programs”](#)

Independent Learning

Citizen data scientists are frequently at the frontiers of data and tool knowledge relative to their self-service BI user peers. They explore the advanced features of modern A&BI platforms and data preparation tools and rarely hesitate to teach themselves new skills — such as a querying language — if they need to do something more advanced. And they may be the only one or part of just a small group of other BI users in an organization pushing at these boundaries. As such, the ability to acquire new knowledge and skills independently is paramount.

Many of the essential skills chronicled in this research can be acquired by an eager independent learner, making use of many free and low-cost resources online and from A&BI vendors. Independent learning requires curiosity (more on that next), a proactive mentality and discipline. Citizen data scientists can not only self-diagnose the knowledge and skills they lack, but they know how to find educational resources to fill the gaps. They also know how to gain the moral and sometimes financial support from managers for additional training. Naturally, too, they know how to stick with a learning plan long enough to extract actionable insights from it.

People who are not naturally inclined toward independent learning may find the citizen data scientist’s need to constantly learn difficult. There are some strategies for cultivating a strong independence in one’s learning pursuits, however. Ongoing self-assessment (“what knowledge am I missing to get to the next level?”) can be put into action as a written set of learning goals and steps toward success. Then, commit to your written learning plan by dedicating specific time daily or weekly for learning and practice. Becoming a better independent learner is basically a self-improvement plan, and the well-worn techniques we know for other aspects of self-improvement in our lives — sticking to New Year’s resolutions, weight loss plans, learning to play an instrument — really do work here.

Curiosity and Creativity

The best citizen data scientists are driven by curiosity and creativity, and it is probably their innate or cultivated curiosity and creativity that led to their becoming citizen data scientists in the first place. Power users of modern A&BI platforms usually become identified as power users because they proactively make use of more — and more advanced — capabilities native to the tools. They are curious about some of the features of the platforms that their colleagues may not be using. And they seek more creative ways to combine and analyze data to come up with fresher insights. Curiosity goes hand-in-hand with the development of advanced data literacy as well: improving your data literacy requires curiosity about where data comes from, what data means and how to generate new insights, while catalyzing more creative investigation into data.

Sure, we see some people as just innately curious or born with a creative eye, but these are skills that can be cultivated or sharpened with some deliberate training. Brainstorming exercises, roleplaying games and design thinking workshops can help people see problems from new perspectives, consider information they may have ignored all along, and think programmatically about how to identify and solve challenges with the tools they have at hand. Interaction with colleagues from different departments or exposure to different industries and use cases can help people translate new ideas into their work and change their points of view. From a problem-solving standpoint, we each have our own go-to heuristics — or problem solving strategies — for tackling challenges, and these heuristics are ingrained through a lifetime of our unique experiences. Seek opportunities to mix with people who have different heuristics — who come from different backgrounds than you, think about different topics on a day-to-day basis than you and deploy tools in their work that are unfamiliar to you. This exposure can help you expand your problem-solving toolbox and learn to view data and business problems in a different way. You can find some of these opportunities in professional associations outside of work, in social activities or through interdisciplinary working groups in your company. In the end, this deliberate exposure to new people and new ideas can develop creativity in how you work and spur ongoing curiosity for learning.

For more information, see:

- [“Exercises for Personal and Team Development: Creativity”](#)
- [“Ignition Guide to Conducting a Design Thinking Workshop”](#)
- [“Presentation: Increase Insight Productivity by Encouraging Creativity From the Insights Team”](#)

Valuable Skills

Table 2 lists the seven valuable skills for citizen data science. Valuable skills are skills that are worth beginning to learn or worth deepening your knowledge of.

Table 2: Valuable Skills for Citizen Data Scientists

Enlarge Table

Skill	Focus Area	Impact	Learning Curve	Impedance
Querying Languages (SQL, etc.)	Tools	Significant	Difficult	Moderate
Data Science and ML Platforms	Tools	Moderate	Difficult	Moderate
Model Testing/QA	Knowledge	High	Moderate	Moderate
Statistics and Quantitative Reasoning	Knowledge	High	Difficult	Light
Ability to Train Others	Soft Skills	Significant	Moderate	Moderate
Intraorganizational Communication	Soft Skills	Significant	Moderate	Moderate
Collaboration and Coordination	Soft Skills	Significant	Difficult	Heavy

Gartner (March 2020)

Querying Languages

In the course of an A&BI platform user's career, even with tools that use visual drag-and-drop interfaces to minimize the user's exposure to code, they no doubt will find themselves having to understand and interact with querying languages. Knowledge of a querying language is valuable for citizen data scientists, too, and SQL is really the dominant language you would need to know to navigate this work. Querying languages allow you to pull data from a database and manipulate it for analysis. Much of this work is done behind the scenes in the query editor capabilities of modern A&BI platforms, and connecting to data sources is often quite straightforward in these tools. But a deeper knowledge of SQL will give you the ability to interrogate a dataset more intimately. It will help you see how values are distributed and what the outliers are, how to sort and filter the data, and generally better understand the structure of the data you are working with. As you pursue more advanced analysis and want to bring in new, raw and diverse data sources to play with, knowledge of SQL will come in handy and help you peek under the hood of your modern A&BI platform or data preparation tool. It will help you to better understand — or alter — how the tool connects to data sources.

SQL is a fairly straightforward programming language with a syntax novices find approachable and experienced coders find easy to master. Teach yourself the basics of SQL with one of many free resources available online, such as:

- [Intro to SQL: Querying and Managing Data](#), from Khan Academy
- [SQL Tutorial](#), from W3Schools
- [SQLTutorial.org](#)

Data Science and ML Platforms

Beyond the ML capabilities native to modern A&BI platforms, it is valuable for a citizen data scientist to begin learning about some of the leading platforms specifically designed for data science and ML. Much of this is about understanding how expert data scientists work. This allows citizen data scientists — especially in larger and more analytically mature organizations — to become streamlined with and conversant in the tools experts are likely to use in their day-to-day work. But it is also about understanding the advanced data science capabilities of these platforms, the flexibility these platforms afford data scientists in developing more creative solutions and the ways these platforms shepherd ML models through a full MLOps life cycle. A citizen data scientist can get by initially with a solid comprehension of modern A&BI platforms, data visualization, data preparation and ML model development skills. But as citizen data scientists grow, they will benefit from a deeper understanding of the tools the experts use.

Citizen data scientists searching for an entry point to learning these platforms can look to autoML. While not a complete replacement for expert-level ML development, autoML lowers the barrier to entry to learn the skills of ML model development and the platforms that specialize in data science and ML. Many analytically mature organizations with expert data scientists may lean on autoML as a powerful complement to, but not full replacement for, expert data science knowledge. These organizations may enlist citizen data scientists in performing autoML tasks to help streamline a larger data science workflow.

For more information, see:

- [“Augment Data Science Initiatives With AutoML”](#)
- [“Demystifying the Data Science Lab and AI Hub for Business Analytics”](#)
- [“Magic Quadrant for Data Science and Machine Learning Platforms”](#)
- [“Solution Criteria for Data Science and Machine Learning Platforms”](#)

Model Testing/QA

Citizen data scientists should also learn about what ML model testing, tuning and quality assurance involves. After models are created and trained, citizen data scientists will play a part in testing the models and watching for issues that may arise, such as bias or variance in the data. Tuning will be necessary to bring the model on target and address variance or bias

in the model. Because citizen data scientists bring strong business domain knowledge to the table, they can also test models to ensure compliance with business objectives or even government and industry regulatory requirements.

Citizen data scientists in smaller or less analytically mature organizations may take on much of this work themselves. Conversely, in larger and more mature organizations, they may engage in this testing and QA work as part of a larger operation that feeds to a team of expert data scientists. In the latter instance, for example, seasoned citizen data scientists may even begin to assist expert data scientists with some feature engineering tasks such as normalization, binning and solving for missing values.

For more information, see:

- [“Drive Data Scientists’ Productivity With Data Preprocessing Techniques”](#)
- [“Improve the Machine Learning Trust Equation by Using Explainable AI Frameworks”](#)

Statistics and Quantitative Reasoning

Much of what separates expert data scientists from typical self-service BI users is a deep understanding of statistics. Many expert data scientists hold advanced degrees in mathematical disciplines or in business, scientific and social scientific fields where statistics is widespread. You and your colleagues may not have that level of knowledge of statistics and may never attain that level of knowledge without advanced degree training. But a citizen data scientist does need a deeper knowledge of statistics than the typical BI user to pursue advanced analytics initiatives. Typical BI users understand descriptive statistics, but citizen data scientists must also know inferential statistics, probability distributions, regressions and Bayesian concepts. They must also have a more precise understanding of statistical significance, hypothesis testing and the impact of outliers in creating variance and bias in a result.

Statistics learned from an academic textbook can be difficult for some. It’s dry stuff, for one. But textbook statistics lessons are also usually divorced from the particulars of your organization or from the realm of business in general. Many find they can learn statistics better by seeing it in action, and your modern A&BI platform and data preparation tools can come in handy here. Pore over the results and visualizations from an analysis over and over again, changing just one aspect of the dataset (or how you join the data) each time. Experiment with how the contours of the data — its distribution, for instance — affect the shape of a curve you get in a visualization, and play with different ways to analyze data. Some tools — such as DataRobot — have capabilities that encourage independent learning when it comes to statistics, too, because they document how a given result was reached, the variables used and how a model works. This experimentation with the business data you’re already familiar with, combined with some good free statistics lessons you can find online, can help you work through new quantitative concepts.

Your local library is sure to have a book on statistics you can check out to have handy on your desk. But you can also find a lot of free resources for learning statistics online. Here are just a few:

- [Statistics and Probability](#), from Khan Academy
- [Seeing Theory: A Visual Introduction to Probability and Statistics](#), by Daniel Kunin

Ability to Train Others

As the importance, velocity, variety and volume of data grows in the enterprise, so too does the desire for more sophisticated analytics to make sense of data. Upskilling self-service BI users into citizen data scientist roles will not be a one-time, one-off pursuit. Citizen data scientists need to share their knowledge with others in the organization and continue to lift up other experienced BI users into the citizen data scientist role alongside them. Their curiosity and ability to learn independently needs to be passed on. Thus, the ability to train others is a valuable skill to learn.

The pioneering first few citizen data scientists in an organization — especially in organizations that do not yet have expert data scientists on staff — will need to build the bridge behind them as they go. They should collect and curate learning resources to make available to colleagues in an internal knowledge base. They should put together working groups, brown bag lunch series and other opportunities to steer others toward the skills they need to attain in order to move into advanced analytic work. Citizen data scientists should make use of their organization’s existing technical systems (e.g., Microsoft SharePoint, social collaboration tools and company intranet) and cultural norms (e.g., “lunch and learn” programs and mentorship programs) to carve out opportunities for peer training and knowledge sharing.

Professional development and corporate training experts, often nested in the human resources department, are also great resources for ideas and materials for structuring training programs. Connect with these people early on in your journey as a citizen data scientist. Finally, if your organization has an analytics center of excellence or a data and analytics steering committee, seek their help and resources for formalizing a knowledge base and finding a way to pass on these lessons to others aspiring to become citizen data scientists.

Intraorganizational Communication

Not only is the citizen data scientist a hybrid, emerging role, but it is also a key pivot point between IT and business. In some organizations, data engineers (another hybrid, emerging role; see: [“Toolkit: Job Description for the Role of a Data Engineer”](#)) specialize in building data pipelines to aid the business in data analysis, acting as a boundary-spanning role between IT and business. But in organizations without data engineers, citizen data scientists may fill this role between business and IT.

Also, although expert data scientists bring superior knowledge in statistics and data to the table, they may lack deep subject matter expertise in a particular business domain. The citizen data scientist serves as a translator, then, helping expert data scientists tailor their work to the relevant KPIs of a line of business and plucking important projects from the business for advanced analysis. The citizen data scientist is thus a liaison, connecting disparate units in the organization to streamline or enhance operations.

As such, the citizen data scientist must possess the valuable skill of intraorganizational communication (or to put it another way, interdepartmental, interdisciplinary communication). Major A&BI initiatives, such as the transformation to self-service BI or the introduction of advanced analytics, are at their essence change management programs. A significant amount of translation has to occur between units in the business to ensure everyone is striving toward the same goals. The citizen data scientist is perfectly poised for this challenge given their hybrid positioning in the organization. Citizen data scientists can sharpen their intraorganizational skills by practicing empathy and perspective-taking, maintaining regular communication with all parts of the organization, and avoiding domain-specific and technical jargon. Speak and write plainly, communicate regularly and help to keep everyone on the same page regarding goals and timelines. Citizen data scientists should also strive for clear and concise email- and memo-writing, learn best practices for effective visual communication (e.g., through PowerPoint), and devote considerable time to developing their public speaking skills. On this latter point, public speaking organizations like Toastmasters International have chapters in cities around the world and welcome new members who want to hone their craft. An effective communicator, particularly one who can understand and appeal to different stakeholders across the organization, is an ideal citizen data scientist.

For more information, see:

- [“Four Must-Have Practices for Successful Organizational Change”](#)

Collaboration and Coordination

They may occupy a hybrid space, and there may only be one or two of them in an organization, but citizen data scientists are not lone wolves. The success, relevance and impact of their work depends on their ability to collaborate and coordinate with others in the organization. Citizen data scientists must shepherd the right resources — new data sources, for instance — through a data and analytics pipeline to create useful ML models. Doing this requires a commitment to teamwork and an ability to navigate bureaucracy, timelines, budgets and governance restrictions. Collaboration and coordination skills are co-emergent with strong intraorganizational communication skills.

For the analytically mature organization with an existing expert data scientist crew, the citizen data scientist’s coordination skills are especially valuable. In these organizations, the citizen data scientist can serve as a kind of assistant to larger-scale data science projects. They can help to triage early-stage or less sophisticated ML model development efforts before they scale up to more robust production with the expert data science team. Citizen data scientists, particularly with their intraorganizational communication skills, must coordinate a pipeline of tasks that feed the expert data scientist core.

For more information, see:

- [“Exercises for Personal and Team Development: Teamwork”](#)

Specialized/Emerging Skills

Table 3 lists the four specialized or emerging skills for citizen data science. The skills described below are either only emerging in their importance to the citizen data scientist role or they may be specialized for particular use cases or organizations at certain analytical maturity stages. Consider these skills as a list of next frontiers to conquer once you have mastered the core and valuable skills above.

Table 3: Specialized/Emerging Skills for Citizen Data Scientists

Enlarge Table

Skill	Focus Area	Impact	Learning Curve	Impedance
Data Science Programming Languages (R, Python)	Tools	Significant	Difficult	Moderate
ML Model Monitoring	Knowledge	Moderate	Difficult	Moderate
ML Model Deployment	Knowledge	Moderate	Difficult	Heavy
Leadership	Soft Skills	Moderate	Difficult	Heavy

Gartner (March 2020)

Data Science Programming Languages (R, Python)

Programming languages commonly used for data science include R and Python. R is primarily for statistical analysis, and Python is more of a general-purpose language for data science, including for production. Expert data scientists may use both in the course of their work. Citizen data scientists would be wise to become familiar with one or both of these languages to more deeply understand the nuances of data science and to work more closely with expert data scientists.

R and Python are both open-source software projects with robust user communities. The software is free to download, and ample documentation and video tutorials exist online. There are also many popular boot camps, certificate programs and other kinds of paid programs available from companies and universities. For more information, see:

- [The R Manuals](#), from the R Development Core Team
- [Python for Beginners](#), from the Python Software Foundation

ML Model Monitoring

ML model monitoring involves tracking the performance of models to determine which ones are worth building on, retraining and deploying into production. In small or analytically less mature organizations with no expert data scientists, citizen data scientists would necessarily involve themselves in monitoring and some deployment of models at small scale. In larger and more analytically mature organizations, however, citizen data scientists can aid expert data scientists in the full MLOps life cycle by assisting with ML model monitoring tasks. This is thus a specialized skill that may be worth cultivating depending on the nature and maturity of your organization.

Citizen data scientists can capture snapshots and data about model performance, run diagnostics, and generate analyses and visualizations about the models. Monitoring models could require piecing together many technologies to capture logs (such as data, compute cluster and model telemetry), analyze the data, and visualize it for performance evaluation and alerting. Familiarize yourself with the ML monitoring capabilities of major data science platforms or with specialty monitoring tools.

For more information, see:

- [“A Guidance Framework for Operationalizing Machine Learning for AI”](#)

ML Model Deployment

ML model deployment is where the results of the ML process (models, insights, etc.) are deployed to AI-based enterprise applications (e.g., image processing systems, NLP-based systems and conversational platforms) and embedded within analytical platforms. In other words, it is where the output extracted from an ML model is put to work. As with ML model monitoring, citizen data scientists may not be involved in much of this particular work, or they may be integral to a larger MLOps workflow with a team of expert data scientists, depending on the size and maturity of the organization’s analytics program. Citizen data scientists may find value in learning about the larger MLOps process and about the ins and outs of what it takes to deploy an ML model. They may also find value in getting familiar with some of the common tools and programming languages in the market for model deployment.

Citizen data scientists can also play an important role in connecting the full circle of ML model deployment by focusing on how business users will apply the results of a model to the original business requirement that served as the impetus for the ML journey. MLOps does not stop at deployment. Rather, it should continue on a trajectory that gets ML-enabled insights into the hands of business users to make better data-driven decisions and monitor the impact and success of the model on the business.

For more information, see:

- [“Solution Path for Building an Effective Technical AI Strategy”](#)
- [“Building a Framework for Managing Effective Machine Learning Workloads”](#)
- [“Preparing and Architecting for Machine Learning: 2018 Update”](#)

Leadership

In time, the citizen data scientists in an organization will be poised for leadership. Because they collaborate across boundaries and disciplines, they get to build a reputation with many key players in the organization. Because they are continually learning about how IT works and about bigger business priorities, they get a higher-level view of how the machinery of the organization works together. Because they focus on how they can train others and form a community of practice, they develop a knack for identifying, assessing and cultivating talent in others. And because they are invested in polishing their communication skills, they continue to cultivate a leading message on behalf of data and analytics innovation in the company. This gets the attention of senior leaders. The sheer number of opportunities for organizational exposure, cross-training, and professional development make citizen data scientists natural candidates for loftier leadership positions down the road.

Experience as a citizen data scientist can also open the door to a career as an expert data scientist. As this research has attempted to describe, citizen data scientists occupy a place on a continuum between everyday self-service BI user and expert data scientist. Just as they may upskill from a BI user into the citizen data scientist role by learning some key data science skills, so too can they continue the professional development path to become expert data scientists if they so desire. The gap between BI user and citizen data scientist may be a smaller chasm to jump than the one between citizen data scientist and expert data scientist, however. Expert data scientists often have formal training, advanced degrees, and deeper statistical knowledge that is most often learned in a formal curriculum. If a pathway to a more general leadership role in the business is not your cup of tea, consider opportunities for formal education in data science to make the move into an expert data scientist career.

Citizen data scientists can develop their broad leadership potential in much the same way aspiring business managers do: seize professional development resources (e.g., executive management training programs, executive MBAs), volunteer to take on special projects that impact the broader company, mentor colleagues, and communicate the value of their advanced analytics initiatives to supervisors in terms that relate to the bottom line. Demonstrate impact and take advantage of opportunities for growth.

For more information, see:

- [“Leading Upskilling Initiatives in Data Science and Machine Learning”](#)
- [“New Digital Age Leadership: Guiding People to Engage, Collaborate — and Do Their Best”](#)

Recommendations

Data and analytics technical professionals looking to cultivate citizen data scientists in their organization should:

- Start by learning the core skills: These skills are foundational. They represent the knowledge a citizen data scientist needs to fulfill the role. In addition to the slate of skills that a typical BI analyst would be expected to have (mastery of modern A&BI platforms,

business subject matter, data visualization and data preparation), core skills for the citizen data scientist also include:

- ML model development
- Advanced data literacy
- Curiosity and creativity
- Independent learning

The core skills are also the building blocks for other valuable and specialized skills.

- Beyond the core, prioritize high-impact, light-impedance skills: Once you have mastered the core skills for citizen data science, tackle valuable and specialized skills in a deliberate way, starting with those that will have the most impact and have the fewest barriers (light impedance).
- Survey the existing A&BI, data science and ML operations in the organization and take an inventory of individuals' skills: Because citizen data scientists often evolve from the sophisticated users of modern A&BI platforms, pinpoint the few experienced users in your organization who seem interested in and capable of ramping up their advanced analytic capabilities. Use this research to assess their current competencies, identify knowledge gaps and build a plan for growth into a citizen data scientist role. Use your survey of existing operations in the organization to design a robust analytic workflow.
- Support a community for peer development: Bring would-be citizen data scientists together and invest in the resources they need to start learning together and building a knowledge base. A regular meeting and a SharePoint site for the group can do wonders to kickstart the journey.
- Determine the relationship between self-service BI users, citizen data scientists and expert data scientists: Because citizen data scientists occupy a hybrid role, it is important to clarify how they will work as extensions of the typical BI user core and/or extensions of a larger expert data science workflow. The maturity and size of your organization may indicate the appropriate fit for citizen data scientists in the beginning, but their work may evolve as your analytic enterprise matures over time.

Is this content helpful to you?

YESNO

Recommended by the Author

- [Demystifying the Analytics and BI Space](#)
- [Demystifying the Data Science Lab and AI Hub for Business Analytics](#)
- [How Citizen Data Science Can Maximize Self-Service Analytics and Extend Data Science](#)

- [Leading Upskilling Initiatives in Data Science and Machine Learning](#)
- [Maximize the Value of Your Data Science Efforts by Empowering Citizen Data Scientists](#)
- [Build a Comprehensive Ecosystem for Citizen Data Science to Drive Impactful Analytics](#)
- [Using Augmented Analytics to Boost A&BI](#)
- [Solution Path for Building an Effective Technical AI Strategy](#)
- [Drive Data Scientists' Productivity With Data Preprocessing Techniques](#)
- [Augment Data Science Initiatives With AutoML](#)

Recommended For You

- [Essential Skills for AI Engineers](#)
- [Augment Data Science Initiatives With AutoML](#)
- [Essential Skills for Machine Learning Architects](#)
- [Comparing AI-/ML-Based Systems That Minimize Data Science Requirements](#)
- [Demystifying the Data Science Lab and AI Hub for Business Analytics](#)