A Beginner’s Guide to SQL, Python, and Machine Learning

Discover how these pillars propel modern business through data and beyond.

By Michael Larner
Table of Contents

Introduction
Why You and Your Business Need to Understand Data Science 4

SQL: The Data Manager
SQL in Everyday Life 7

Python: Efficient Programming
Python in Everyday Life 10

Python and Machine Learning:
Seeing the Future
How SQL, Python, and Machine Learning Work Together 14
Machine Learning in Everyday Life 15

Where Do We Go From Here?
Boost Your Career 17
Drive Your Business Forward 18
Develop a Data-Driven Team in 3 Steps 18

About General Assembly 19
Introduction
In the coming pages, you’ll learn:

- What SQL, Python, and machine learning do.
- How these powerful technologies help solve real-world challenges.
- What types of job roles leverage SQL, Python, and machine learning.
- Why you and your organization need to be data literate in order to stay competitive.
- How to forge a path by learning these skills.

INTRODUCTION

In 2012, IBM revealed that 2.5 quintillion bytes of data were being created per day — an enormous sum of data that humankind had never known before. Since then, the volume of the world’s data has not only continued to increase, but it’s arriving at a faster and faster pace.

However, data by itself doesn’t have much value. After all, a pile of numbers and data files is just that: a pile of numbers and data files. The real value of data comes from making sense of the abundance of information. That’s why businesses and organizations across countless industries are investing in forward-thinking data talent — to leverage data’s predictive power, craft smart business strategies, and drive informed decision-making.

The sharp and strategic people who do this job are data scientists, data analysts, machine learning engineers, and business intelligence analysts — among other titles — and these data professionals are in high demand. In 2018, the jobs platform Glassdoor ranked data scientists as the Best Job in America for the third year in a row, with a median salary of $110,000 and more than 4,500 available jobs. Additionally, five other data- and analytics-related roles made the list of the top 50 jobs, ranked by number of openings in the field, salary, and overall job satisfaction.

Companies are quickly recognizing the vital need for data knowledge, impacting a vast array of industries including eCommerce, health care, finance, and sales, just to name a few. In order to stay competitive and grow their businesses, leaders are investing in their future by strategically training and hiring talent to ensure proficiency in key skills.

Why You and Your Business Need to Understand Data Science

There’s often a lot of confusion as to what exactly data science is. That’s partly because so many different companies and organizations use the term to mean slightly different things, and partly because the definition of data science is incredibly broad.

In general, data science is the practice of acquiring, organizing, and delivering
INTRODUCTION

complex data; discovering relationships and anomalies among variables; building scalable machine learning software; and turning data into valuable information.

You may find that at some companies, a person with the title of data scientist focuses more on the statistics and mathematical modeling aspects of the field, while at another company they could be more focused on delivering business value.

On a high level, data professionals collect, process, clean up, and verify the integrity of data. They apply engineering, modeling, and statistical skills to build end-to-end machine learning systems that predict consumer behavior, identify customer segments, and much more. They constantly monitor the performance of those systems and make improvements wherever possible.

Looking at the field as a whole, there’s a wide array of tools available to help data experts perform tasks ranging from gathering their own data to transforming it into something that’s usable for their needs.

In this paper, we break down three of the most prevalent technologies that are transforming how we understand and use data: SQL, Python, and machine learning. The first two are programming languages used to gather, organize, and make sense of data. The last is a specific field in which data scientists and machine learning engineers, using Python and other technologies, enable computers to learn how to make predictions without needing to program every potential scenario.

These skills have surprising uses beyond data, bringing delight, efficiency, and innovation to countless industries. They empower people to drive businesses forward with a speed and precision that was previously unknown.
SQL:
The Data Manager
In today’s digital age, we’re constantly bombarded with information about new apps, hot technologies, and the latest, greatest artificial intelligence system. While these technologies may serve very different purposes in our lives, many of them have one essential thing in common: They rely on data. More specifically, they use databases to capture, store, retrieve, and aggregate data.

This begs the question: How do we actually interact with databases to accomplish all of this? The answer: We use Structured Query Language, or SQL (pronounced “sequel” or “ess-que-el”).

Put simply, SQL is the language of data — it’s a programming language that allows us to efficiently create, alter, request, and aggregate data from databases. It gives us the ability to make connections between different pieces of information, even when dealing with huge data sets.

Modern applications can use SQL to deliver valuable pieces of information that would otherwise be difficult for humans to keep track of independently. In fact, pretty much every app that stores any sort of information uses a database. This ubiquity means that developers can use SQL to log, record, alter, and present data within the application, while analysts can use SQL to interrogate that same data set in order to find deeper insights.

SQL at a Glance

Pronounced: “sequel” or “ess-que-el”

A programming language used to: Create, alter, request, and aggregate data from databases.

Data scientists and programmers use it to: Make connections between different pieces of information, even when dealing with huge data sets.

SQL in Everyday Life: Real-World Examples

We’re constantly interacting with data in our lives, which means that, behind the scenes, SQL is probably helping to deliver that information to us. Here are a few examples:

**Extracting Data**

At its most basic, SQL is about accessing
Why SQL Matters

“Imagine you try to open an Excel file with thousands of rows of data — you know how slow it is just to open that file, let alone do any analysis in it.

“Using SQL (with your data in a queryable database) allows you to pull just the data, or aggregated and joined versions of it, that you need. Then you can either run simple summary statistics and logical functions on that data in SQL, or export the summarized versions of the data into Excel to continue manipulating or visualizing it.”

SQL

data locked away in databases. Think about the last time you received a report about how your company or team is performing. This probably had some key metrics like sales figures, conversion rates, or profit margins based on data stored in a system like a customer relationship management (CRM) or eCommerce platform.

A developer or analyst, or maybe even you, used SQL in order to access the data needed to produce that report.

WEB APPLICATIONS

Think about the last time you looked up the name of a movie on IMDb, the Internet Movie Database. Perhaps you quickly noticed an actress in the cast list and thought something like, “I didn’t realize she was in that,” then clicked a link to read her bio.

As you were navigating through that site, SQL may have been responsible for returning the information you “requested” each time you clicked a link.

SYNTHESIZING DATA TO MAKE BUSINESS DECISIONS

With SQL, you can combine and synthesize data from different sources, then use it to influence business choices.

For example, if you work at a real estate investment firm and are trying to find the next up-and-coming neighborhood, you could use SQL to combine city permit, business, and census data to identify areas that are undergoing a lot of construction, have high populations,
Python: Efficient Programming
Python at a Glance

A programming language used to:
- Automate processes.
- Build applications’ functionality.
- Delve into machine learning.

Data professionals and software engineers use it to:
- Clean data.
- Build AI models.
- Create web apps, and more.

Simplicity in Code

Here’s a cool example of just how simple Python is:
Below is code that tells the computer to print the words “Hello World”:

In Python:
```
print ("Hello World")
```

Yup, that’s really all it takes! For context, let’s compare that to another popular programming language, Java, which has a steeper learning curve (though is still a highly desirable skill set in the job market).

Python

Python is an immensely popular programming language used by data analysts, data scientists, and software engineers to automate processes, build the functionality of applications, and delve into machine learning. Companies like Google, SpaceX, and Instagram use it to clean data, build AI models and web apps, and more. It stands out for being very simple to read and write, while offering extreme flexibility and having an active community. This makes it a great language for new programmers to learn for a broad range of applications in data science, web development, and beyond.

Python in Everyday Life: Real-World Examples

Here are some fascinating examples of how Python is shaping the world we live in:

ARTIFICIAL INTELLIGENCE

Python is especially prevalent in the AI community, again for its ease of use and flexibility. For example, in just a few hours, a business could build a basic chatbot that answers some of the most common questions from its customers. To do this, programmers could use Python to scrape the contents of all of the email exchanges with the company’s customers, identify common themes in these exchanges with visualizations, and then build a predictive model that can be used by the chatbot application to give appropriate responses.

FILE-SHARING APPLICATIONS

When the file-storage platform Dropbox was created in 2007, it used Python to build the desktop applications and server infrastructure responsible for actually sharing the files. Python is still powering the company’s desktop applications today. In other words, Dropbox was able to write a single application for both Macs and PCs that still works after more than a decade!
Why Python Matters

“Python is an amazing tool for improving your speed on basic analytics tasks when you’re working with multiple variables.

“For example, using a simple line of Python code can help you create scatterplots that visualize correlations between all combinations of your variables in 30 seconds instead of manually doing that for each combination in Excel. And that’s just basic visualization in Python! It’s also a critical language in developing algorithms and systems for machine learning.”

WEB APPLICATIONS

Python is used to run various parts of some of today’s most trafficked websites, including Pinterest, Instagram, Spotify, and YouTube. In fact, the visual bookmarking platform Pinterest has used Python in some form since it was founded (e.g., to power its web app, build and maintain data pipelines, and perform analyses).

HOLLYWOOD SPECIAL EFFECTS

Remember that summer blockbuster with the huge explosions? A lot of companies, including Lucasfilm’s Industrial Light & Magic (ILM), use Python to help program those awesome special effects. By using Python, companies like ILM have been able to develop standard toolkits that they can reuse across productions, while still retaining the flexibility to build custom effects in less time than ever before.

PYTHON

Ariana Dugan
Education Product Manager
Data Courses, General Assembly

Why Python Matters

“Python is an amazing tool for improving your speed on basic analytics tasks when you’re working with multiple variables.

“For example, using a simple line of Python code can help you create scatterplots that visualize correlations between all combinations of your variables in 30 seconds instead of manually doing that for each combination in Excel. And that’s just basic visualization in Python! It’s also a critical language in developing algorithms and systems for machine learning.”

WEB APPLICATIONS

Python is used to run various parts of some of today’s most trafficked websites, including Pinterest, Instagram, Spotify, and YouTube. In fact, the visual bookmarking platform Pinterest has used Python in some form since it was founded (e.g., to power its web app, build and maintain data pipelines, and perform analyses).

HOLLYWOOD SPECIAL EFFECTS

Remember that summer blockbuster with the huge explosions? A lot of companies, including Lucasfilm’s Industrial Light & Magic (ILM), use Python to help program those awesome special effects. By using Python, companies like ILM have been able to develop standard toolkits that they can reuse across productions, while still retaining the flexibility to build custom effects in less time than ever before.
Python and Machine Learning: Seeing the Future
EVER wonder how apps, websites, and machines seem to be able to predict the future? Like how Amazon knows what your next purchase may be, or how self-driving cars can safely navigate a complex road situation?

The answer lies in machine learning. Machine learning is a branch of AI that often leverages Python to build systems that can learn from and make decisions based on data. Instead of explicitly programming the machine to solve the problem, we show it how it was solved in the past and the machine learns the key steps that are required to do the same task on its own.

Machine learning is revolutionizing every industry by bringing greater value to companies’ years of saved data. Leveraging machine learning enables organizations to make more precise decisions instead of following intuition.

There’s an explosive amount of innovation around machine learning being used within organizations, especially given that the technology is still in its early days. Many companies have invested heavily in building recommendation and personalization engines for their customers. But machine learning is also being applied in a huge variety of back-office use cases as well, like to forecast sales, identify production bottlenecks, build efficient traffic routing systems, and more. Machine learning algorithms fall into two categories, supervised and unsupervised learning.
MACHINE LEARNING

Types of Machine Learning

SUPERVISED LEARNING
Predicts future values by relying on past data.
Common uses include:
• Recommendation engines
• Fraud detection
• Image recognition

UNSUPERVISED LEARNING
UnCOVERS hidden structures from data.
Common uses include:
• Market segmentation
• Identifying suspicious user behavior

Ratings as training data to the model and then predicts your rating for unseen movies.

Supervised learning enjoys more commercial success than unsupervised learning. Some common use cases include fraud detection, image recognition, credit scoring, product recommendation, and malfunction prediction.

UNSUPERVISED LEARNING
Unsupervised learning is about uncovering hidden structures within data sets. It’s helpful in identifying segments or groups, especially when there is no prior information available about those segments. These algorithms are commonly used in market segmentation. They enable marketers to identify target segments in order to maximize revenue, create anomaly detection systems to identify suspicious user behavior, and more.

For instance, Netflix may know how many customers it has, but wants to understand what kind of groupings they fall into in order to offer services targeted to them. The streaming service may have 50 or more different customer types, aka segments, but its data team doesn’t know yet. If the company knows that most of its customers are in the “families with children” segment, it can invest in building specific programs to meet customer needs. But without that information, Netflix’s data experts can’t build a supervised machine learning system.

So, they build an unsupervised machine learning algorithm instead, which identifies and extracts various customer segments within the data and allows them to identify groups such as “families with children” or “working professionals.”

How Python, SQL, and Machine Learning Work Together

To understand how SQL, Python, and machine learning are related to one another, let’s think of them as a factory. As a concept, a factory can produce anything if it has the right tools. More often than not, the tools used in factories are pretty similar (e.g., hammers and screwdrivers).

What’s amazing is that there can be factories that use those same tools, but build completely different products (e.g., tables vs. chairs). The difference between these factories is not the tools, but rather how the factory workers use their expertise to leverage these tools to produce a different result.

In this case, our goal would be to create a machine learning model, and our tools would be SQL and Python. We can use SQL to extract data from a database and then use Python to shape the data and perform the analyses that ultimately produce a machine learning model. Your knowledge of machine learning will ultimately enable you to
**Why Machine Learning Matters**

“Machine learning will set you apart as a data worker. Being able to successfully execute machine learning systems means that you’ve been able to familiarize yourself with the data you’re working with, wrangle it into the state you need it to be in, and develop a system that provides truly meaningful, and often predictive, insights to a business.

“For example, machine learning can be used to detect fraud in finance and insurance, which has huge financial impacts for those companies.”

---

**MACHINE LEARNING**

achieve your goal.

To round out the analogy, an app developer with no understanding of machine learning might choose to use SQL and Python to build a web app. Again, the tools are the same, but the practitioner uses their expertise to apply these skills in a different way.

**Machine Learning in Everyday Life: Real-World Examples**

While machine learning-powered innovations like self-driving cars and voice-activated robots seem ultra-futuristic, the technology behind them is actually widely used today. Here are some great examples of how machine learning impacts you every day:

**RECOMMENDATION ENGINES**

Think about how Spotify makes music recommendations. The recommendation engine peeks at the songs and albums you’ve listened to in the past, as well as tracks listened to by users with similar tastes. It then starts to learn the factors that influence your music preferences and stores them in a database, recommending similar music that you haven’t listened to — all without writing any explicit rules!

**VOICE-RECOGNITION TECHNOLOGY**

We’ve seen the emergence of voice assistants like Amazon’s Alexa and Google’s Assistant. These interactive systems are based entirely on voice-recognition technology powered by machine learning models.

**RISK MITIGATION AND FRAUD PREVENTION**

Insurers and creditors use machine learning to make accurate predictions on fraudulent claims based on previous consumer behavior, rather than relying on traditional analysis or human judgement. They also can use these analyses to identify high-risk customers. Both of these analyses help companies process requests and claims more quickly and at a lower cost.

**PHOTO IDENTIFICATION VIA COMPUTER VISION**

Machine learning is common among photo-heavy services like Facebook and the home-improvement site Houzz. Each of these services use computer vision — an aspect of machine learning — to automatically tag objects in photos without human intervention. For Facebook, these tend to be faces, whereas Houzz seeks to identify individual objects and link to a place where users can purchase them.
Where Do We Go From Here?
“Dipping your toes in this uncharted water may seem daunting — but it shouldn’t!

“There’s so much opportunity in the field for individuals and teams to grow.”

WHERE DO WE GO FROM HERE?

You can dive into SQL, Python, and machine learning without any formal background. However, each topic has a different set of fundamentals that you’ll need to understand as you progress in your learning. For example, Python will expose you to the world of object-oriented programming, while SQL will expose you to database design concepts. Machine learning will require a good understanding of data analysis.

Dipping your toes in this uncharted water may seem daunting — but it shouldn’t! There’s so much opportunity in the field for individuals and teams to grow. Here are actionable steps you can take — right now — to get started.

Boost Your Career

There are countless reasons to add SQL, Python, and machine learning skills to your toolkit, whether or not you’re seeking a full-time role in the data field.

No matter your role or industry, this knowledge can take your hireability to the next level. Here are just some of the things you can do with data expertise:

- **Become a skilled problem-solver.** Programming languages like SQL and Python teach you problem-solving skills that are applicable in many business scenarios you’ll encounter.
- **Be more cross-functional.** Having key programming and data skills under your belt makes it easier to work with teams across your organization. Being able to speak the same language as software engineers, business intelligence analysts, and data professionals helps streamline requests, bring clarity to the workflow, and provide insight into technical action items.
- **Build the technology of the future.** Data skills enable you to help build new, groundbreaking technologies, including web applications, machine learning models, chatbots, and much more.
- **Expand your career potential.** Glassdoor has named data scientists the No. 1 Job in America three years running, and, based on previous projections from the management consultancy McKinsey, IBM predicts that by 2020, the number of data science and analytics job listings will grow by nearly 364,000 listings, to about 2.72 million.
- **Improve communication.** Data professionals must communicate to nontechnical audiences — including stakeholders across the company — in a compelling way to highlight the business impact and opportunity. At the end of the day, those stakeholders have to act on and possibly make far-reaching decisions based on data findings.
Drive Your Business Forward

With the field of data science expanding so rapidly — and a huge shortage in skilled data professionals — many organizations are struggling to adapt their workforces at the same speed. To combat these challenges, business leaders need to be strategic when it comes to hiring and cultivating data talent in their organizations.

For a business to leverage data to its potential, leaders need to:

- **Know what you need.** Look to a data leader in your organization, or outside experts in the field, to help evaluate what specific skills and roles your company is missing. This will help you decide whether to invest resources in training team members who are ready to take on more advanced skills, or search for outside candidates.

- **Enable people to access data.** Data literacy empowers employees from any team to find potential insights and interpret data sets to make better decisions. Python and SQL are fantastic tools that are relatively easy to learn and offer tremendous power for people across an organization.

- **Tap into your own workforce.** Upskill or reskill the valuable talent you have, and train them to meet your business needs. Their industry knowledge serves as a strong foundation for learning new data tools and techniques that your company needs to grow. Think through training pipelines for various levels of experience, and continue developing your workforce’s skills as new technologies emerge.

As the world becomes increasingly data-driven, learning to leverage these key technologies will create endless possibilities for your career and your organization. There’s never been a better time to dive in, and we hope this guide helps get you on your journey.
General Assembly is a global education company on a mission to empower individuals and companies through dynamic training programs, exclusive thought-leader events, and high-impact networking opportunities. Our curricula focus on the in-demand skills every company today needs: data, coding, design, digital marketing, and product management.

In our data programs — including the full-time Data Science Immersive, part-time Data Science and Data Analytics courses, and short-form workshops and bootcamps — we give individuals and teams the tools to access this data in order to answer their own business questions in ways that were never possible before. To help guide people and employers toward the skills and tools driving success in the field, we work with industry leaders on our Data Science Standards Board to shape our course material and training.

GA also trains and sources data talent for businesses, whether you need a few data professionals or a few thousand. Our innovative assessments benchmark current employees’ skill levels in order to pinpoint areas of growth opportunity. Then, we customize learning experiences to exactly what your workforce needs to become indispensable thought leaders in data.