Courses   How it works   Mentors

Students   Blog

When career is right for you?   All Categories

Blog > Data Science > **22 APIs Every Data Scientist Should Know**

22 APIs Every Data Scientist Should Know

Sakshi Gupta Sakshi Gupta | 13 minute read | July 8, 2020

In this article

What is an API?

22 APIs every data scientist should learn

How do you learn APIs?

APIs are the essential building blocks for data science. They provide key data sources and enable data integration and visualization. In this blog, we look at the most important APIs and how to leverage them.

But how do APIs apply to data science and which APIs should be part of your data science toolkit? Let's start by defining what an API is.

What is an API?

An Application Programming Interface (API) allows pieces of code to interact with one another. Developers use APIs to build their websites with specific features, like a Google Maps interface, instead of having to write code from scratch. Some may be open-source, while others charge a fee for implementation. You typically need to register a developer account or have some other means of authentication for APIs.

An API typically has three core elements:

- **Access:** Who is the user accessing the service?
- **Request:** What is the service or data being requested? This includes both methods (what questions do you need answered with the data or service solicited?) and parameters (supplementary details).
- **Response:** How does the system respond to the request?

Representational State Transfer (REST) provides a method for communication over web services, with commands like GET, PUT, POST, and DELETE.

APIs in R normally use the HTTR package, while Python users will want to become familiar with the Requests HTTP library. APIs function like web applications but send data-exchange outputs in formats like JSON, XML, etc., instead of HTML.

Here are some tutorials for a deep-dive into APIs for data science: CareerCon, DataQuest, and Towards Data Science.

One way to think about an API is as a structured manner for you to obtain a permitted set of data from an application owner.

A particular API allows you to build on top of the rich datasets large, well-resourced organizations are building, either building live interactive web applications (for example, an app that uses the Spotify API to try to determine what instruments they might like from their liked songs on the

Courses   How it works   Mentors
Students   Blog

When career is right for you   All Categories

22 APIs every data scientist should learn

APIs can be useful for many parts of the data science process, but have particular applications for machine learning. Many large tech companies and machine learning specialized startups provide ready-to-use frameworks for analysis.

Here are some of the most popular APIs in data science:

- **Amazon Machine Learning API**

Built on the AWS cloud platform with a user-friendly interface, Amazon helps with prediction models, generates useful visualizations, and facilitates statistical analysis.

Amazon Machine Learning API is great for customer awareness. You can predict customer conversion, purchasing habits, and lifetime value, based on type and number of orders. Amazon's API can also recognize human activity through sensor data and detect fraudulent users from analyzing web history.

Documentation: Amazon Machine Learning Documentation

Tutorial: AWS Machine Learning Samples. Examples of how you can use the Amazon Machine Learning API to build simple applications.

- **IBM Watson Discovery API**

IBM Watson allows you to sift through online search content and find patterns in enterprise data. It is all about applying cognitive skills to machines and studying how humans interact with applications.

This API makes it easier to translate text into language (speech to text and text to speech), determine how a message resonated with a particular audience, model users based on specific social characteristics, and answer frequently asked questions in real-time.

Documentation: IBM Cloud API

Tutorial: Create and query a data collection in IBM Watson Discovery, an official post that goes through to the timeline and process of running and getting actionable insights into IBM Watson.

- **Google API**

All Categories

Documentation: Google Cloud APIs

Tutorial: Google has a bunch of API and tutorials all compiled into one place.

- **Twilio API**

The Twilio API allows you to transfer your programmatic skills to applications that have to do with texts and calls. Create apps that can text your users or use their phones to communicate critical data. Manage and generate phone numbers programmatically through the Twilio API. You can also tap into WhatsApp programmatically as well as verify phone number ownership to reduce account signup fraud.

Documentation: Twilio Docs

Tutorial: How to build a chatbot, an example of how to use the Twilio console to build something from end-to-end with a way to then integrate it into SMS or Whatsapp programmatically.

- **Census.gov API**

If you've ever felt that you've needed critical demographic and economic data from the U.S. government, this API helps you query that information and put together interesting applications and data projects built on one of the most reputable data collection agencies. You can aggregate historical data tied to a FIPS code that defines a certain Census geographic area.

Documentation: Census Data API User Guide

Tutorial: A Tutorial for the Census Data API is a short blog post that goes into how to get started with registering for an API key and then making some sample queries. You can use this website to figure out and get familiar with what data the census contains.

- **Spotify API**

Get the metadata associated with the most popular songs (or even the most obscure). You'll also get access to user data such as the songs related to the ones they like if you get authorization from them, allowing you to build rich applications driven by the aggregate collection of music Spotify hosts as well as the individual-level data it holds.

Tutorial: Getting Started with Spotify's API is a Medium article that contains all of the steps you need to get started with the Web API as well as Spotify, a Python library and API wrapper that can be used to quickly access Spotify data.

- **Yummly API**

This API feeds in information about different recipes and the foods that compose it. Use it to come up with new recipes and to analyze a current one. The API returns back JSON data, and supports both HTTP and HTTPS requests. You need to register for a developer account in order to get access, and there is rate limiting applied.

Documentation: Yummly API Documentation

Tutorial: Correct Yummly API call for a recipe – Stack Overflow is a Stack Overflow question that deals with how to properly structure a query for the Yummly API.

- **New York Times API**

Searching through the New York Times has never been easier with this API, which programmatically goes through different sections and articles all the way back to 1851. You can find articles through a query filter or facet.

The API returns a max of 10 results at a time, while the meta node allows you to paginate through your results up to 1,000. In practice, it's meant for fine tooth-combing of a selection of articles for a particular time and a particular subject: you might be interested, for example, in what the New York Times wrote about a particular historical figure sometime between a certain date range.

The API not only offers New York Times results but also has Reuters and Associated Press articles as well, which you can easily facet and search through.

Documentation: New York Times API Documentation

Tutorial: Scraping New York Times Articles with Python is a tutorial by UC Berkeley that shows the basics of constructing a query around the API. In this case, they go through and look for every Amnesty International mention between 1980 and 2004.

- **Reddit API**

different usernames.

The API is built to be able to allow for a wide scope of read and write applications, as well as to build datasets. Reddit is one of the world's largest repositories of text data and sentiment analysis (in the form of upvotes and downvotes on content and comments). This property leads it to be an immensely valuable trove of data — this has led to many language generation machine learning models being trained on Reddit data.

Documentation: Reddit API Documentation

Tutorial: PRAW Repository PRAW stands for Python Reddit API Wrapper. The quickstart tutorial gets you started on how to quickly work with Reddit's API for a variety of meaningful use cases.

- **Zillow API**

If you're looking for data on housing prices based on a variety of factors or a variety of online housing data, look no further than the collection of APIs offered by Zillow. You'll be able to stream real estate and mortgage data, allowing you to create websites that simulate real estate portals or do data analysis on real estate patterns.

Documentation: Zillow API Documentation

Tutorial: Zillow API to Google Sheets is an article with a third-party service (Apipheny) that makes it easy to import Zillow data quickly into Google Sheets.

- **Instagram API**

Use the Instagram API to query metadata and data about Instagram posts and users. You'll be able to get posts around different hashtags, information about users and their following and follower counts, and much more. You'll need to sign up to be approved as a developer but once you are, you'll be able to have access to a wealth of information and perhaps the largest image repository in the world.

The legacy API is now deprecated and has been split into Basic Display API and Graph API.Use it to look at the social network and the profiles of different Instagram users and how they're related to one another.

Documentation: Instagram Developer API Documentation

get access to forecasts and different weather apps live from an authoritative government source.

Documentation: Weather.gov API Documentation

Tutorial: The Github repository contains lots of code and community discussion for the US National Weather Service API.

- **Imgflip API**

Ever wanted to tap into a stream of different memes? The Imgflip API allows you to do just that. With a simple query, you can return a JSON object that contains 100 or so memes ordered by how many times they were captioned in the last 30 days—an interesting feature that lets you get the most popular and trending memes.

You can then also generate memes programmatically with Imgflip's caption_image function. Images created with the API will be publicly accessible to anybody with the returned URL.

Documentation: Imgflip API Documentation

Tutorial: React.js Meme Generator Tutorial is a handy Youtube video that shows how you can integrate the Imgflip API with React.js to create an interactive web application dedicated to spawning memes.

- **The SportsDB API**

The SportsDB API compiles salaries and different salary caps across different teams in different sports. It will always remain free, but you can sign up on Patreon for additional features such as live scores.

You'll get JSON results back from players, teams and leagues around the world. For example, this is the profile for Mario Balotelli, a famous Italian striker. You'll be able to get positional data for players, detailed information about them, as well as their wage and height.

The SportsDB API is the perfect plug-and-play API if you want live sports scores and detailed player and league information.

Documentation: SportsDB API Documentation

Tutorial: The SportsDB forum has a section for third-party projects built on top of SportsDB.

the most definitive database of startups, their funding rounds, and who provided the funding. Use their API to get a pulse on the latest financing rounds—what industries are getting them, who the funders are, and what companies are securing the financing they need to power forward.

Access to the full API requires an Enterprise or Applications License with CrunchBase, but you can be authorized under the Open Data Map plan to join data from people and organization-level datasets to your own—you just need to be careful with attribution. The API is strictly read-only and is meant to take the incredible amount of data Crunchbase holds and put it towards applications and analyses that advance the startup ecosystem.

Documentation: Crunchbase API Documentation

Tutorial: How To Build An Automated Sales Pipeline With The Crunchbase API (Python)

- **SkyWatch API**

SkyWatch connects developers with satellite images around the world, allowing you to access to do interesting projects on images of Earth from space. They're partnered with NASA and different satellite providers so you can create datasets of maps and other views.

Documentation: SkyWatch API Documentation

Tutorial: How an Artificial Intelligence company uses satellite data is a case study built by the company behind SkyWatch.

- **USGS Earthquakes API**

Get real-time access to earthquake information as it comes in and is recorded around the world. Your application or data project will soon be able to ingest the pattern of the Earth's movement across decades. The API returns results in JSON, including the geographic location of earthquakes, and their depth and magnitude as well as other associated facts. It also returns which regions are currently under high alert for earthquakes.

Documentation: USGS Earthquake API Documentation

Tutorial: Earthquake Data API is a simple tutorial that shows how with Python's requests library (or any other http library), you can construct an open, unauthenticated query — and get back information on where earthquakes occurred, their magnitude, and whether or not they caused a tsunami.

This API uses the Johns Hopkins public data on COVID-19 incidence, and divides it by different regions and different reports, updating all of the time.

Documentation: The documentation for the API can be found here.

Tutorial: This article sums up different COVID-19 APIs.

- **Facebook Graph API**

Social media platforms have become integral to every consumer-focused business model. These APIs allow you to capture big data from posts, comments, and likes, which can be valuable for marketing and sales decisions. This aggregated data can facilitate opinion mining and sentiment analysis. Facebook's API allows you access to Facebook data.

Documentation: Different Facebook documentation documents exist for various products and APIs.

Tutorial: How to use Facebook Graph API and extract data using Python is a medium article on Towards Data Science that runs through how to use the Graph API to get lists of people who are attending an event and the admins of said event.

- **Twitter API**

The Twitter API allows you to access data about the world of tweets and allows you automated access to read and write data. You can create Tweets and Direct Messages automatically from it, though for data science purposes, the amount of data available can help with sentiment analysis and anything that requires a lot of text data around emerging or trending topics.

Documentation: Twitter API documentation

Tutorial: This page provides live use cases for the Twitter API and tutorials on how to build those use cases.

- **BigML**

For supervised and unsupervised machine learning, BigML is a RESTful API that enables real-time machine learning predictions using datasets, models, clusters, and anomaly detectors. Users can create a data source and then answer questions, like how many rooms will be used in buildings

---

Tutorial: The BigML Quick Start guide helps you get set up with a series of simple steps.

- **Data Science Toolkit**

Pete Warden, founder of the OpenHeatMap project and writer of the Data Source Handbook for O'Reilly, also has a collection of open data sets and open-source tools for data science on Github as part of his Data Science Toolkit API.

This API is based on a Linux distribution and comprises a REST/JSON API with command line, Python, and Javascript interfaces. It allows you to extract important insights, like geocodes, and execute data manipulations, like turning IP or street addresses to coordinates (latitude and longitude), files to text, HTML to text, text to dates/times, etc. To learn the R interface to use the toolkit, RDSTK is a helpful package for beginners.

Documentation: The Data Science Developer Docs are supposed to be here, but you might want to refer to the repository which has uptime as of writing.

Tutorial: The Data Science Toolkit repository has different documentation and examples.

How do you learn APIs?

As a data scientist, you may be required to build your own APIs that others can access, so you will need to think through different use cases. A common starting point is to practice adding a web API through Flask, a micro Python interface.

In order to build your data science toolkit and learn APIs, you will want to know querying languages like SQL; the Python programming language and libraries like scipy, TensorFlow, scikit, pandas, numpy, and matplotlib; frameworks like Spark and Scala; and tools for data visualization, regression, and high-performance data analysis.

As you master data science through programs like Springboard's data science bootcamp, you can find countless online resources to practice with APIs, such as:

- RapidAPI, the most extensive online API marketplace featuring 7500 APIs
- ProgrammableWeb, which has a huge searchable database of API tutorials and industry news and research

~~Which career is right for you?~~   All Categories

facial recognition and customer feedback synthesis.

By contributing to more open-source APIs, data scientists can expand our capacity to process big data by increasing data volume, giving faster access to data storage centers, and quickly developing new products and services through cognitive APIs, which can transform inputs into data formats readable by other systems through machine learning.

To fast-track your knowledge of APIs and preparation for a data science career, check out Springboard's data science, data engineer, and machine learning bootcamps.

Since you're here…

Are you a future data scientist? Investigate with our free step-by-step guide to getting started in the industry. When you're ready to build a CV that will make hiring managers melt, join our 4-week Data Science Prep Course or our Data Science Bootcamp—you'll get a job in data science or we'll refund your tuition.
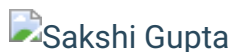
About Sakshi Gupta

Sakshi is a Senior Associate Editor at Springboard. She is a technology enthusiast who loves to read and write about emerging tech. She is a content marketer and has experience working in the Indian and US markets.

Download our guide to becoming a data scientist in six months

Learn how to land your dream data science job in just six months with in this comprehensive guide.

| Enter your email | Download now

Courses   How it works   Mentors
Students   Blog

DATA SCIENCE
Text Mining in R: A Tutorial

DATA SCIENCE
What Skills Do You Need to Become a Data Engineer?

CAREER TRACKS	RESOURCES	ABOUT US	GET SOCIAL
Data Analytics Bootcamp	Free Learning Paths	About the Company	
Data Science Bootcamp	E-books and Guides	Jobs	
Machine Learning Bootcamp	Blog	Contact Us	SCHOLARSHIPS
Software Engineering Bootcamp	Career Assessment Test	Become a Mentor	Student Beans
UI/UX Design Bootcamp	Student Outcomes	Hire Our Students	Inclusion Scholarships
		Affiliates	
		Partners	
		Community	

Which career is right for you?   All Categories

Bootcamp

Tech Sales
Bootcamp

Copyright 2022   Terms Privacy Conduct Security

Regulation