

Exploring the Skills Needed by the Data Scientist

Kirk Paul Lafler; @sasNerd

Abstract

As 2.5 quintillion bytes (1 with 18 zeros) of new data are created each and every day, the age of big data has taken on new meaning. More and more organizations across industries are embracing Data Science / Computer Research Scientist skills resulting in an emerging demand for qualified and experienced talent. According to the Bureau of Labor Statistics (BLS) the number of data science jobs is expected to grow 19 percent over the next two decades – nearly three times as fast as the average growth rate for all jobs. Energized by this employment outlook, students and professionals across job functions are preparing for tomorrow's growing data science / analytic demands by acquiring a comprehensive skill set. To prepare for this growing demand, many colleges, junior colleges, Universities, and vocational training organizations offer comprehensive degrees and certificate programs to fulfill the increasing demand for analytical skills. This paper and presentation explores the skills needed by the Data Scientist / Analytics professional including non-technical skills such as critical thinking; business acumen and verbal/written communications; and technical skills such as data access; data wrangling; statistics; use of statistical programming languages like Python, R and SAS®; Structured Query Language (SQL); Microsoft Excel; and data visualization.

Introduction

Wikipedia describes data science, "as a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data." Wikipedia further describes data science, "as the profession to unify statistics, data analysis, mathematics, computer programming, operations research, machine learning and related methods in order to understand and analyze actual phenomena with data and enhance decision making."

Professionals with skills in data science and analytics are a "hot" commodity in the employment marketplace. The Bureau of Labor Statistics (BLS) (accessed on April 9th, 2021) projects that Employment of computer and information research scientists is projected to grow 15 percent from 2019 to 2029, much faster than the average for all occupations. One thing is certain, with the rapid increase in data collection activities by organizations that this will lead to an increased need for services related to data mining, or processing data sets to identify patterns and relationships to solve business problems.

Brian Holak, a site editor at Search Business Analytics, (January 31st, 2019) predicts, "The demand for data scientists continues to grow sharply with more employers than ever looking to hire data scientists. Data Scientists will experience a 29% increase in demand year over year, a 344% increase since 2013, with the supply of "qualified" data scientists lagging demand." Merrimack College projects, "The Federal Government will experience a 19% growth rate in the field of computer and information research scientists by 2026." Also, many employment / career websites (August 2022) offer students, junior professionals, and seasoned professionals with the organizations who are seeking the right candidates to hire amazing employment, contract, and internship opportunities. The exciting world of the Data Scientist is one that is experiencing an exponential rate of growth. But while demand is huge the supply of skilled applicants is still growing at a much slower pace.

Graduate Degree Programs

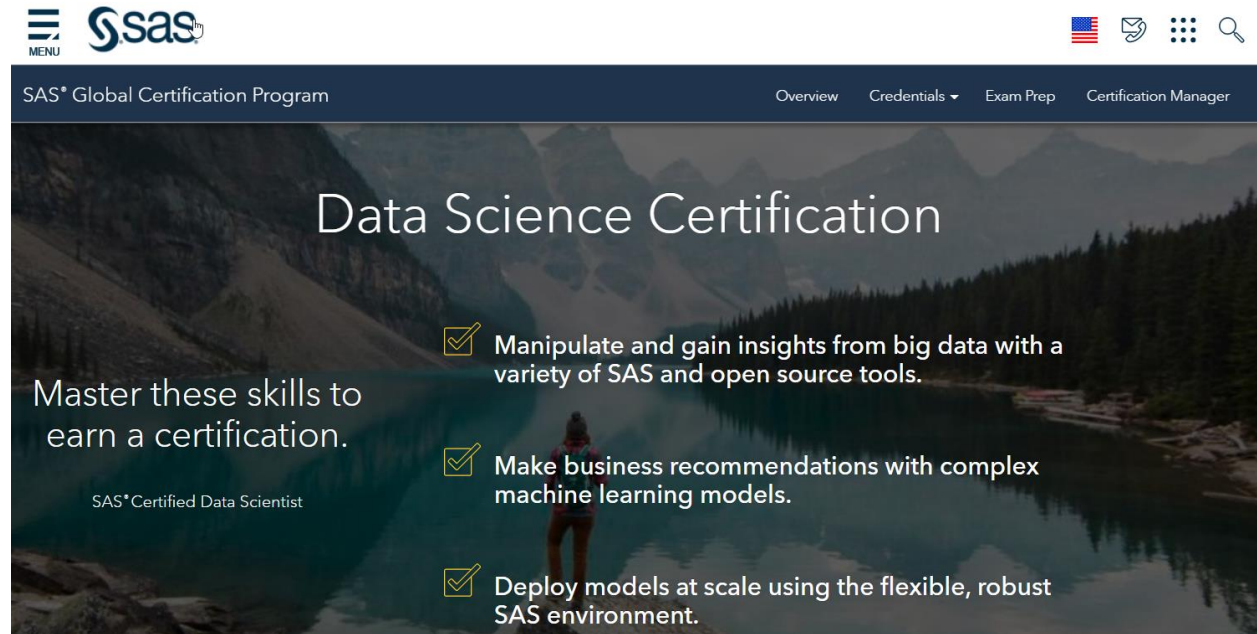
The majority of Data Scientists hold an undergraduate degree in a quantitative field such as Mathematics, Statistics, Decision Sciences, Computer Science, Management Information Systems, and Economics to name a few. Many also hold graduate degrees and/or certificates in the field of data science. The next table represents a starting point for professionals seeking Colleges and Universities offering graduate degrees in Data Analytics, Big Data Analytics and Data Science. Although this table identifies many academic institutions, it most definitely does not provide a complete list of all the academic institutions offering these prestigious graduate degrees. Readers are encouraged to perform your own search using one or more search engine(s) for a more complete list of all the colleges and universities offering these popular degree programs.

Institutions Offering Graduate Degree Programs in Data Science, Data Analytics and Big Data Analytics (Alphabetical Sampling)

<u>Binghamton University</u>	<u>Louisiana State University</u>	<u>University of Illinois at Urbana-Champaign</u>
<u>Boston College</u>	<u>Michigan State University</u>	<u>University of Kentucky</u>
<u>Boston University</u>	<u>New Mexico State University</u>	<u>University of Massachusetts at Dartmouth</u>
<u>California State University</u>	<u>New York University</u>	<u>University of Miami</u>
<u>Cal Poly</u>	<u>North Carolina Agriculture and Technical State University</u>	<u>University of Minnesota</u>
<u>Carnegie Mellon University</u>	<u>North Carolina State University</u>	<u>University of Michigan at Ann Arbor</u>
<u>Clarkson University</u>	<u>Oakland University</u>	<u>University of Nevada, Reno</u>
<u>Clemson University</u>	<u>Ohio University</u>	<u>University of North Carolina at Charlotte</u>
<u>Columbia University – New York</u>	<u>Oklahoma State University</u>	<u>University of North Carolina at Greensboro</u>
<u>Cornell University</u>	<u>Old Dominion University</u>	<u>University of North Carolina at Wilmington</u>
<u>DePaul University</u>	<u>Pace University</u>	<u>University of Notre Dame</u>
<u>Dominican University of California</u>	<u>Penn State Great Valley</u>	<u>University of San Diego</u>
<u>Drake University</u>	<u>Purdue University</u>	<u>University of San Francisco</u>
<u>Drexel University</u>	<u>Rice University</u>	<u>University of South Dakota</u>
<u>Duke University</u>	<u>San Diego State University</u>	<u>University of Southern California</u>
<u>Elon University</u>	<u>San Jose State University</u>	<u>University of Southern Maine</u>
<u>Florida State University</u>	<u>Seton Hall University</u>	<u>University of Texas at Austin</u>
<u>Franklin University</u>	<u>Stanford University</u>	<u>University of Virginia</u>
<u>Georgetown University</u>	<u>Texas A&M University</u>	<u>University of Washington</u>
<u>George Washington University</u>	<u>Tulane University</u>	<u>University of Wisconsin Madison</u>
<u>Georgia State University</u>	<u>University of Alabama at Birmingham</u>	<u>Utah State University</u>
<u>Georgia Tech</u>	<u>University of Arizona</u>	<u>Vanderbilt University</u>
<u>Harvard University, Extension School</u>	<u>University of Bridgeport</u>	<u>Villanova University</u>
<u>Hofstra University</u>	<u>University of California – Berkeley</u>	<u>Walden University</u>
<u>Indiana University</u>	<u>University of California, Irvine</u>	<u>Wharton University</u>
<u>Johns Hopkins University</u>	<u>University of California San Diego MSBA</u>	<u>Willamette University</u>
<u>Kansas State University</u>	<u>University of California San Diego DS</u>	<u>Worcester Polytechnic Institute</u>
<u>Kennesaw State University</u>	<u>University of Chicago</u>	
<u>Lehigh University</u>	<u>University of Colorado, Boulder</u>	

Globally Recognized SAS Certifications

SAS Institute offer users with a globally recognized certification process that can help boost your prospects for success and validate your skills to employers. SAS certification testing is available in many areas and particularly for those who are interested in a Data Science career with a combination of credentials including: Data Curation, Advanced Analytics, and AI & Machine Learning. Click [here](#) for more information about [SAS Institute's Data Science certification](#).



The screenshot shows the SAS Global Certification Program website. At the top left is the SAS logo and a menu icon. At the top right are icons for a flag, a mail envelope, a grid, and a search magnifying glass. Below the navigation bar, the page title is "SAS® Global Certification Program". The main content area features a large background image of a person standing on a rocky shore looking at a calm lake reflecting mountains. The text "Data Science Certification" is prominently displayed. To the left, it says "Master these skills to earn a certification." and "SAS® Certified Data Scientist". To the right, there are three bullet points, each with a checkmark icon:

- Manipulate and gain insights from big data with a variety of SAS and open source tools.
- Make business recommendations with complex machine learning models.
- Deploy models at scale using the flexible, robust SAS environment.

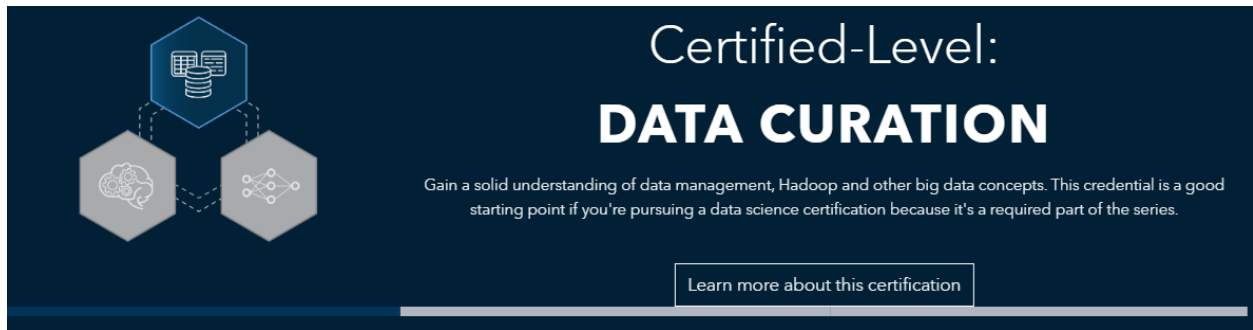
Source: Image captured from SAS Institute's website, copyright 2022 by SAS Institute Inc.



The infographic has a dark blue background. At the top, it says "How to Become a SAS® Certified Data Scientist". Below that, in white text, are the three credential areas: "DATA CURATION • ADVANCED ANALYTICS" and "AI & MACHINE LEARNING". In the center, there is a diagram with three hexagonal icons connected by dashed lines. The top icon is blue and shows a database cylinder and a document. The bottom-left icon is teal and shows a brain with gears. The bottom-right icon is purple and shows a neural network diagram. At the bottom, there is a paragraph of text:

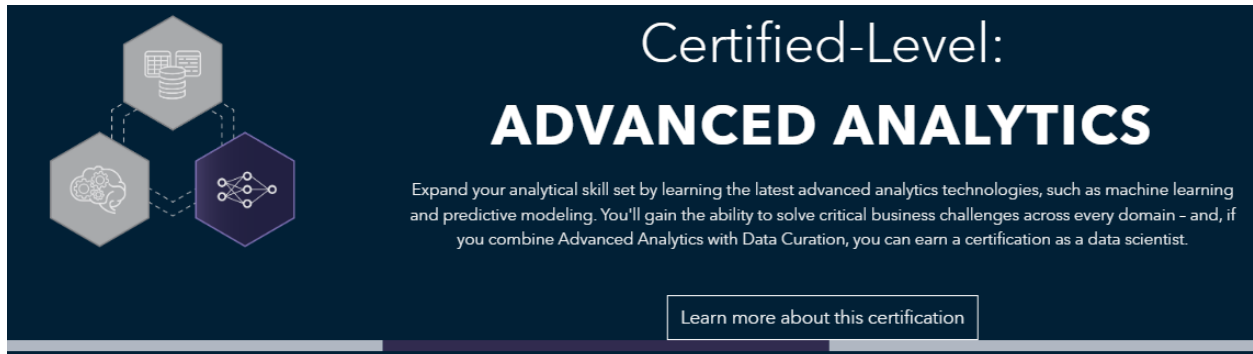
If you're interested in a data science career, our academy offers three professional-level credentials to boost your résumé. Earning one credential can launch a career – but a combination helps you earn a credential that could transform your future.

Source: Image captured from SAS Institute's website, copyright 2022 by SAS Institute Inc.



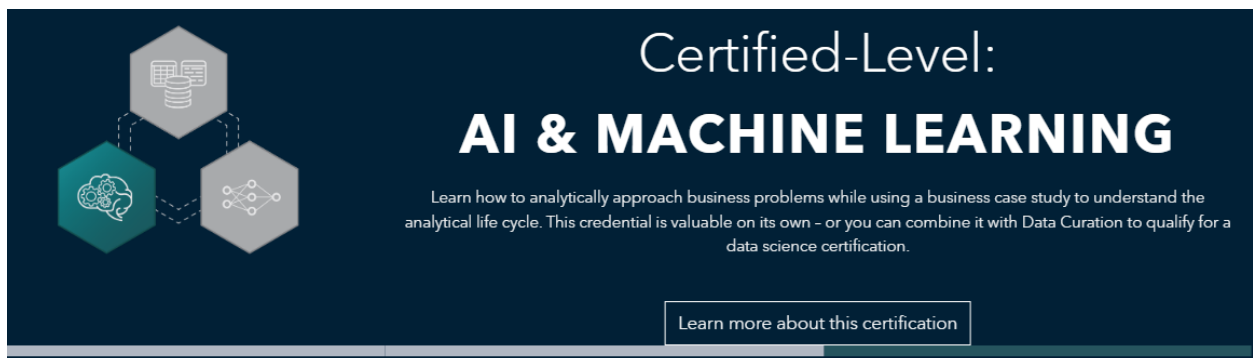
The banner features a dark blue background with a white icon on the left showing three interconnected hexagons: a top one with a database symbol, a bottom-left one with a brain, and a bottom-right one with a neural network. The text on the right reads 'Certified-Level: DATA CURATION' in large white font, followed by a paragraph: 'Gain a solid understanding of data management, Hadoop and other big data concepts. This credential is a good starting point if you're pursuing a data science certification because it's a required part of the series.' At the bottom right is a white button with the text 'Learn more about this certification'.

Source: Image captured from SAS Institute's website, copyright 2022 by SAS Institute Inc.



The banner features a dark blue background with a white icon on the left showing three interconnected hexagons: a top one with a database symbol, a bottom-left one with a brain, and a bottom-right one with a neural network. The text on the right reads 'Certified-Level: ADVANCED ANALYTICS' in large white font, followed by a paragraph: 'Expand your analytical skill set by learning the latest advanced analytics technologies, such as machine learning and predictive modeling. You'll gain the ability to solve critical business challenges across every domain - and, if you combine Advanced Analytics with Data Curation, you can earn a certification as a data scientist.' At the bottom right is a white button with the text 'Learn more about this certification'.

Source: Image captured from SAS Institute's website, copyright 2022 by SAS Institute Inc.



The banner features a dark blue background with a white icon on the left showing three interconnected hexagons: a top one with a database symbol, a bottom-left one with a brain, and a bottom-right one with a neural network. The text on the right reads 'Certified-Level: AI & MACHINE LEARNING' in large white font, followed by a paragraph: 'Learn how to analytically approach business problems while using a business case study to understand the analytical life cycle. This credential is valuable on its own - or you can combine it with Data Curation to qualify for a data science certification.' At the bottom right is a white button with the text 'Learn more about this certification'.

Source: Image captured from SAS Institute's website, copyright 2022 by SAS Institute Inc.

Projections, Outlook, and Findings

Employment of Data Scientists is projected to grow 36 percent from 2021 to 2031, much faster than the average for all occupations. The most in-demand technical skills for data science careers are Python and SQL. The average data scientist salary in the U.S. is \$125,242 / year.

In Aleksandra Yosifova's April 6th, 2023 report, The Data Scientist Job Outlook in 2023 – Research on 1,000 LinkedIn Job Postings, researched 1,000 LinkedIn job postings to answer questions about the Data Scientist job outlook in 2023. The report attempts to answer the following questions:

- Is data science still the sexiest job of the 21st century?
- Where do data scientists work?
- How much do they earn?
- What skills are required for a successful data science career?
- What are the next in-demand tools and techniques in the field?

Data Science / Analytics Annual Base Salaries

With massive quantities of data being collected at astounding rates by organizations worldwide, the demand for Data Scientists and Analytics professionals to collect, analyze and interpret all this data is growing rapidly. The annual base salaries for all this talent are impressive as well. In a July 28, 2022 Indeed.com article, the highest paying cities for Data Scientists appear, below.

Highest Paying Salaries by City (Indeed.com)			
Los Angeles, CA	\$167,921 / year	San Francisco, CA	\$131,425 / year
Houston, TX	\$159,289 / year	Austin, TX	\$127,883 / year
New York, NY	\$153,603 / year	Atlanta, GA	\$120,049 / year
Washington, DC	\$139,768 / year	San Diego, CA	\$118,342 / year
Chicago, IL	\$135,593 / year		

Salary Information from [www.Indeed.com](https://www.indeed.com/career/data-scientist/salaries) (Source: <https://www.indeed.com/career/data-scientist/salaries>)
 In Rick Chen's (April 7, 2022; TeamBlind.com) article the average salaries by city for Data Scientists are shown, below.

Annual Average Salaries by City (TeamBlind.com)					
Los Gatos, CA	\$447,916	New York, NY	\$143,926	St. Louis, MO	\$122,671
Los Angeles, CA	\$196,892	Bellevue, WA	\$143,097	Bentonville, AR	\$122,285
Sunnyvale, CA	\$173,178	Redmond, WA	\$142,438	Dallas, TX	\$122,017
Cupertino, CA	\$166,712	Seattle, WA	\$141,741	Chicago, IL	\$121,975
Menlo Park, CA	\$165,761	Washington, DC	\$141,306	Wilmington, DE	\$120,800
Santa Monica, CA	\$164,272	Newark, NJ	\$139,311	Hartford, CT	\$120,491
San Francisco, CA	\$163,670	Boston, MA	\$137,567	McLean, VA	\$120,066
San Jose, CA	\$161,090	Palo Alto, CA	\$137,327	Atlanta, GA	\$118,977
Mountain View, CA	\$157,767	Portland, OR	\$136,992	Raleigh, NC	\$117,774
Cambridge, MA	\$151,500	Kansas City, MO	\$135,363	Minneapolis, MN	\$115,918
Santa Clara, CA	\$150,943	San Diego, CA	\$133,190	Columbus, OH	\$115,781
San Mateo, CA	\$149,046	Redwood City, CA	\$131,791	Philadelphia, PA	\$114,963
Boulder, CO	\$148,989	Irvine, CA	\$129,872	Tampa, FL	\$114,750
Fort Lauderdale, FL	\$148,400	Denver, CO	\$128,348	Phoenix, AZ	\$113,975
Miami, FL	\$148,250	Jersey City, NJ	\$124,638	San Antonio, TX	\$112,571
Charlotte, NC	\$145,885	Pittsburgh, PA	\$123,855	Baltimore, MD	\$110,857
Oakland, CA	\$144,625	Austin, TX	\$123,167		

Average Salaries (Source: <https://www.teamblind.com/blog/index.php/2022/04/07/best-paying-cities-data-scientist-2022/>)

Data Scientist / Analytics Employment / Career Opportunities

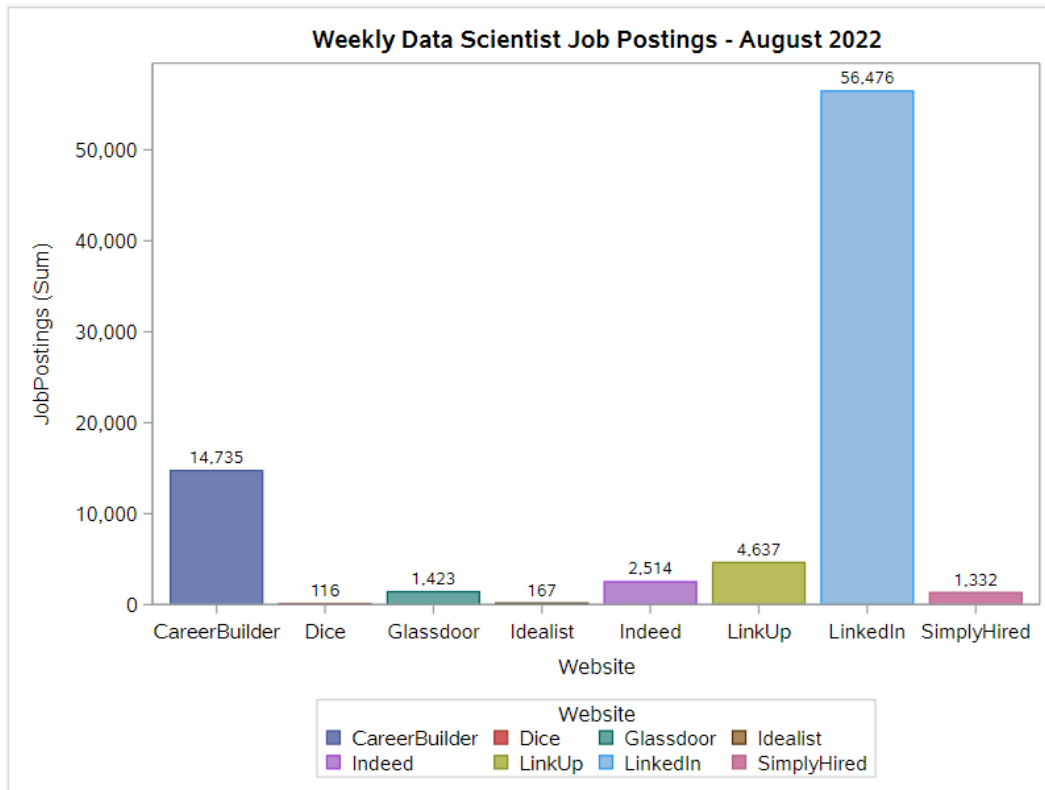
There appears to be considerable demand for "Data Scientists" when searching the various employment / career websites. Although cross-postings do exist, an exact match search for data scientist jobs was performed to ensure the results were relevant and the search keyword was processed similarly. The following VBAR chart shows the number of data scientist job postings by job listing website for the week of August 8th, 2022.

Code:

```
data Weekly_DataScientist_Jobs ;
  input @1 Website $15.
        @16 JobPostings comma8. ;
  format JobPostings comma8. ;
datalines ;
LinkedIn      56,476
Indeed        2,514
Dice          116
Glassdoor    1,423
SimplyHired   1,332
CareerBuilder 14,735
Idealist      167
LinkUp        4,637
;
run ;

proc sgplot data=Weekly_DataScientist_Jobs ;
  title1 "Weekly Data Scientist Job Postings - August 2022" ;
  vbar Website / response=JobPostings group=Website datalabel ;
run ;
```

Results:



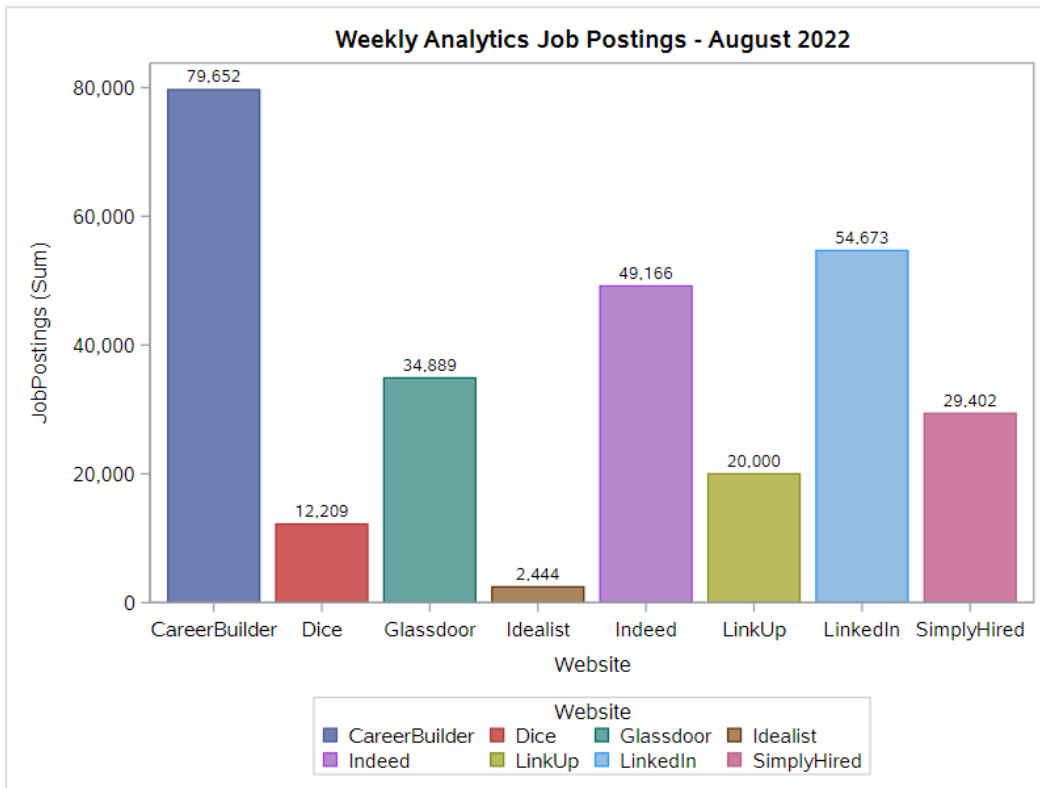
Using an exact match search, job posting results are displayed using a VBAR chart for “analytics” to show the number of analytics job postings for the week of August 8th, 2022.

Code:

```
data Weekly_Analytics_Jobs ;
  input @1 Website $15.
```

```
@16 JobPostings comma8. ;  
format JobPostings comma8. ;  
datalines ;  
LinkedIn      54,673  
Indeed        49,166  
Dice          12,209  
Glassdoor     34,889  
SimplyHired   29,402  
CareerBuilder 79,652  
Idealist       2,444  
LinkUp        20,000  
;  
run ;  
  
proc sgplot data=Weekly_Analytics_Jobs ;  
title "Weekly Analytics Job Postings - August 2022" ;  
vbar Website / response=JobPostings group=Website datalabel ;  
run ;
```

Results:



Data Scientist Job Postings on Employment / Career Websites

Job postings on employment / career websites for "Data Scientist" are displayed, below, for the week of August 8th, 2022.

Exploring the Skills Needed by the Data Scientist, continued

CareerBuilder Jobs Upload/Build Resume Salaries & Advice Recommended Jobs Post a Job Sign In Sign Up

14,735 Data Scientist Jobs

Job Type 7 days Any Easy Apply Only Work from Home/Remote

Create Job Alert. Get similar jobs sent to your email Save

List of Jobs

Sort by: Relevancy Date

- P** 3 DAYS AGO **Lead Data Scientist** Pocket FM Corp | Work From Home | Full-Time Easy Apply
- C** 5 DAYS AGO **Sr. Data Scientist, eCommerce An...** Caterpillar, Inc. | Peoria, IL | Full-Time
- F** 5 DAYS AGO **Principal Data Scientist - People In...** F. Hoffmann-La Roche Ltd | California Hot Springs, CA | Full-Time

JOIN US

LEAD DATA SCIENTIST

Pocket FM Corp | Work From Home, GA | Full-Time **Apply Now**

Job Details

About Pocket FM

Pocket FM is India's largest audio OTT platform offering a vast collection of podcasts and audiobooks across a wide variety of genres and languages and making our entry to the global audience. It's an amazing time to join Pocket FM as we continue shaping the future of audio entertainment both in India and globally. We are India's leading audio OTT platform with the largest collection of podcasts and audiobooks across a wide variety of genres and languages, and making our entry to the global audience.

Source: Image captured from CareerBuilder.com's website.

Dice Tech Careers Insights For Employers Login/Register

Search: Data Scientist Search Location Search Jobs

Job title, Keywords, Company City, Country, Postal code

120 Jobs

116 Data Scientist Jobs

Email me jobs for this search We receive over 1,000 new job postings a day. Data Scientist Email Jobs

Data Scientist (Business Intelligence)

Black Knight Inc Remote Save

Full-time Posted 18 hours ago - Updated moments ago

Black Knight is the premier provider of integrated technology, services, data and analytics that lenders and servicers look to first to help successfully manage the entire loan life cycle. Our deep understanding of

NLP Data Scientist

ABN TECH CORP San Francisco, CA, USA Save

Contract, Third Party Posted 11 hours ago

Title: NLP Data Scientist Location: San Francisco, CA (onsite after few weeks) Duration: 6+ months contract/ Full time Job Description: 9+ years of experience required Examples of the types of tasks that this

Easy Apply

Remote Options: Remote Only, Exclude Remote, Work From Home Available

Posted Date: Any Date, Today, Last 3 Days, Last 7 Days

Source: Image captured from Dice.com's website.

Exploring the Skills Needed by the Data Scientist, continued

The screenshot shows the Glassdoor.com interface. At the top, the search bar contains 'data scientist' and 'Location'. Below the search bar, there are navigation tabs for 'Jobs', 'Companies', 'Salaries', and 'Careers'. The main content area is divided into two columns. The left column shows a list of job listings with details such as company name, job title, location, salary range, and date posted. The right column shows a detailed view of a 'Principal Data Science' job at Discover, including a job description and a 'Job Description' section.

1,423 Data Science Jobs

Discover 3.9★
Principal Data Science
Riverwoods, IL
\$106K - \$139K (Glassdoor est.)
New 24h

AthenPollo LLC
Data Scientist, Cell Line Development
Boston, MA
\$90K - \$120K (Employer est.)
Easy Apply 4d

Leidos 3.9★
Data Scientist
Reston, VA
\$94K - \$196K (Employer est.)
New 3d

PeritusHub LLC
Jr. Data Scientist
Lake Forest, IL
\$80K (Employer est.)
Easy Apply 24h

The Home Depot 3.8★
Data Scientist, Marketing & Online (Remote)
Atlanta, GA
\$90K - \$160K (Employer est.)
24h

GTT, LLC
Data Scientist

Bring Curiosity Shine Bright

Discover 3.9★
Principal Data Science
Riverwoods, IL
\$106K - \$139K (Glassdoor est.)
New Be one of the first to Apply!

Discover. A brighter future.

With us, you'll do meaningful work from Day 1. Our collaborative culture is built on three core behaviors: We Play to Win, We Get Better Every Day & We Succeed Together. And we mean it – we want you to grow and make a difference at one of the world's leading digital banking and payments companies. We value what makes you unique so that you have an opportunity to shine.

Come build your future, while being the reason millions of people find a brighter financial future with Discover.

Job Description

Responsible for working closely with management to execute analytical initiatives. Responsible for providing thought leadership & strategic thinking to solve business problems by leveraging techniques such as

Show More

Source: Image captured from Glassdoor.com's website.

The screenshot shows the Indeed.com interface. At the top, the search bar contains 'Data Scientist' and 'Where remote'. Below the search bar, there are navigation tabs for 'Find jobs', 'Company reviews', 'Find salaries', and 'Employers / Post Job'. The main content area is divided into two columns. The left column shows a list of job listings with details such as company name, job title, location, salary range, and date posted. The right column shows a detailed view of a 'Senior Data Analyst' job at System1, including a job description and a 'Resume Insights' section.

2,514 Data Scientist Jobs

new
Senior Data Analyst
System1 4.7★
Remote
Estimated \$61.7K - \$78.1K a year
Full-time

- Strong **data** visualization skills and ability to break down complex **data** sets for technical and non-technical audiences.
- What You Can Bring to the Team: *

Just posted

new
Sr. Data Analyst
arreglo consulting
Remote
Estimated \$95.5K - \$121K a year
Contract 8 hour shift

Senior Data Analyst
System1 ★★★★★ 3 reviews
Remote • Remote
Full-time
Apply on company site

Resume Insights
Here's how your resume aligns with the job description

Experience & Skills

SQL R AWS

+ show more

Education & Certificates

Source: Image captured from Indeed.com's website.

Exploring the Skills Needed by the Data Scientist, continued

The screenshot shows the LinkedIn job search interface. At the top, there's a search bar with 'Data Scientist' and 'United States' selected. Below the search bar are filters for 'Jobs', 'Past Week', 'Experience Level', 'Company', 'Job Type', 'On-site/Remote', 'Easy Apply', and 'All filters'. A prominent green banner indicates '56,476 Data Scientist Jobs'. The main content area is split into two columns. The left column lists several job postings: 'Data Scientist' at Net Pay Advance, Inc. (Wichita, KS), 'Senior Data Engineer' at VIDA (Edina, MN), 'Sr. Front End Software Developer' at DHL eCommerce Solutions (Weston, FL), 'Senior Data Migration Analyst' at Latham, The Pool Company (Latham, NY), and 'Junior SAS Programmer' at Leidos (Baltimore, MD). The right column provides detailed information for the 'Data Scientist' role at Net Pay Advance, Inc. It includes the job title, location (Wichita, KS), time ago (2 hours), and number of applicants (3). It also lists job details such as 'Part-time · Entry level', '51-200 employees · Financial Services', and 'Work Schedule Monday - Friday; 40 hours/week'. A description of the role is provided, stating that the Data Scientist will develop and implement advanced analytics for analyzing data problems. There are 'Apply' and 'Save' buttons for this job.

Source: Image captured from LinkedIn.com's website.

The screenshot shows the SimplyHired job search interface. At the top, there's a search bar with 'data scientist' and 'City, State, ZIP or "Remote"'. A purple 'Search Jobs' button is on the right. Below the search bar are filters for 'Relevance', 'Date', 'Job Type', 'Minimum Salary', and '7 days'. A prominent green banner indicates '1,332 Data Scientist Jobs'. The main content area is split into two columns. The left column lists several job postings: 'Data Scientist - 100% Remote' at Paramount Pictures (Remote), 'Jr. Data Scientist' at PeritusHub LLC (Lake Forest, IL), and 'SR Data Scientist, Revenue Science' at AREVO Group (Remote). The right column provides detailed information for the 'Data Scientist - 100% Remote' role at Paramount Pictures. It includes the job title, location (Remote), and a 'Quick Apply' button. It also lists job details such as 'Full-time', 'Estimated: \$140,000 - \$170,000 a year', and '6 days ago'. A 'Benefits' section lists 'Health savings account', 'Dental insurance', 'Employee discount', 'Health insurance', and 'Professional development assistance'. A 'Qualifications' section lists 'SQL', 'Data visualization', 'APIs', 'R', and 'Machine learning'.

Source: Image captured from SimplyHired.com's website.

Data Science / Analytics Skills

Data Science / Analytics skills are all encompassing and requires in-depth knowledge and experience in many technical and non-technical areas.



The table, below, presents “key” areas where skills are needed by the data science / analytics community of professionals.

Technical Skills	Non-Technical Skills
SAS, Python and R	Critical Thinking
SQL and RDBMS techniques	Intellectual Curiosity
Excel	Business Acumen
Cloud Computing Services (aka, SaaS)	Verbal / Written Communication
Statistics – Descriptive and Inferential	Storytelling
Artificial Intelligence (AI)	Ability to Work in a Team
Machine Learning (ML)	
Structured and Unstructured Data	
Data Wrangling techniques	
Analytical techniques	
Data Visualization techniques	

Software: Programming, Database, Statistical, Visualization, Spreadsheet, and Cloud Computing Services

SAS Software

SAS is a statistical software suite of products developed by SAS Institute Inc. for advanced analytics, multivariate analysis, business intelligence, data management, predictive analytics, and criminal investigation. SAS runs on all important platforms and supports object-oriented and structured programming along with other programming paradigms. Developed by Dr. James Goodnight, Anthony Barr, John Sall and Jane T. Helwig. The SAS software suite has more than 200 components including Base SAS, SAS/STAT, SAS/GRAPH, SAS/OR, SAS/ETS, SAS/IML, SAS/AF, SAS/QC, SAS/INSIGHT, SAS/PH, Enterprise Miner, Enterprise Guide, SAS/EBI, and SAS Grid Manager.

SQL

Structured Query Language (SQL) is a relational database language that is used in programming relational database management systems (RDBMS). It is specifically useful in handling structured data. SQL comprises many types of statements including a data query language (DQL), a data definition language (DDL), a data control language (DCL), and a data manipulation language (DML). There are several types of SQL implementations including SAS' PROC SQL, Microsoft's SQL-Server, Oracle, and IBM. SQL was originally developed by Edgar Frank "Ted" Codd, Donald D. Chamberlin, and Raymond F. Boyce in the early 1970s.

Python

Python is an open-source programming language that is available under a free software license. It supports object-oriented and structured programming along with other programming paradigms. Developed by Guido van Rossum in the late 1980s, Python is designed to be an "easy to read language" with numerous third-party modules to interact with other languages; extensive support libraries such as web service tools; text processing; string operations; internet protocols; a powerful scripting language; an extensive user community; and many other features.

R

R is a powerful open-source programming language and is used for statistical computing, graphics and data analysis. Available under a free software license, R runs on all important platforms and is used by statisticians, data miners and thousands of major corporations and institutions worldwide. Developed by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, their initial version of R was released in 1995 with a stable beta version in 2000. R boasts an extensive array of packages including data wrangling; data analysis; plotting; graphing; reporting; statistics; an extensive user community; and many other features.

Excel

Microsoft Excel is widely used spreadsheet software operating under Windows, macOS, Android and iOS platforms to allow users to format, organize, manipulate, and calculate data in spreadsheets. Common Excel uses include the collection and storage of data, business analysis, data analysis, statistical analysis, accounting and budgeting, account management, project management, performance reporting, administrative and managerial management, operations management, and office administration. Users can arrange data in a spreadsheet using graphical tools, formulas, and pivot tables to work with large quantities of data to identify sums, averages, percentages, unique values, minimum and maximum values, ranges, outliers, and other needs.

Cloud Computing Services

Cloud computing is the delivery of computing services (aka, SaaS) including software, databases, servers, storage, networking, analytics, and intelligence over the Internet to offer users improved and affordable computing speed, flexibility, and scale. From my own experience using a few cloud services and from reviewing an article on cloud service providers ([Peterson, Richard. July 26, 2022](#)) cloud services are offered by SAS Institute Inc., Amazon Web Services (AWS), Microsoft, IBM, Google, ServerSpace, Adobe, Kamatera, VMware, Rackspace, Red Hat, Salesforce, Oracle, SAP, Verizon, Linode, HostPapa, DigitalOcean, ScalaHosting, OVHcloud, LiquidWeb, Vultr, CloudSigma, LimeStone, Navisite, and Dropbox.

Conclusion

Professionals with skills in data science and analytics are a “hot” commodity in the employment marketplace. With the Bureau of Labor Statistics (BLS) employment projections estimating growth of 15 percent from 2019 to 2029, data science careers are poised to accelerate greatly. Wikipedia mentions that, as a multi-disciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data,” data science “is the profession to unify statistics, data analysis, mathematics, computer programming, operations research, machine learning and related methods to analyze and better understand actual phenomena with data and enhanced decision making.”

This paper and e-poster has attempted to share the many ways that students, junior professionals, and those who are already pursuing a career as a Data Scientist can acquire greater knowledge as well as the ability to hone your skills for this exciting profession. From obtaining an advanced degree in data science, analytics, or big data analytics from one of the many distinguished universities and organizations; pursuing a certification offered by SAS Institute and others; and enhancing essential skills using the “free” SAS OnDemand for Academics (ODA) software; anyone seeking to enhance their skills to become a “skilled” data scientist. We also explored the many employment / career websites (August 2022) that are available to students, junior professionals, and seasoned professionals with the many organizations who are seeking the right candidates to hire for amazing employment, contract, and internship opportunities. The exciting world of the Data Scientist is one that is experiencing an exponential rate of growth, but while demand is huge the supply of skilled applicants is still growing at a much slower pace.

References

- Bureau of Labor Statistics (BLS) (accessed on April 9th, 2021), [BLS Employment Projections for Computer and Information Research Scientists from 2019 to 2029](#).
- Chen, Rick. (April 7th, 2022). [The Top 50 Best-Paying Cities for Data Scientists in the United States](#).
- Granville, Vincent (August 25th, 2018), [“Demand for Data Scientists is Booming and will only Increase.”](#)
- Holak, Brian (January 31st, 2019). [“Demand for Data Scientists is Booming and will only Increase.”](#)
- Indeed.com. (July 28th, 2022). [Highest paying cities for Data Scientists in the United States](#).
- Lafler, Kirk Paul (2022), [“Exploring the Skills Needed by the Data Scientist”](#), Proceedings of the Western Users of SAS Software (WUSS) Conference – 2022.
- Lafler, Kirk Paul (2022), [“Exploring the Skills Needed by the Data Scientist”](#), Proceedings of the SouthEast SAS Users Group (SESUG) Conference – 2022.
- Lafler, Kirk Paul (2021), [“Exploring the Skills Needed by the Data Scientist”](#), Proceedings of the SouthEast SAS Users Group (SESUG) Conference – 2021.
- Lafler, Kirk Paul (2019), [“Exploring the Skills Needed by the Data Scientist”](#), Proceedings of the MidWest SAS Users Group (MWSUG) Conference – 2019.
- Lafler, Kirk Paul (2019). [PROC SQL: Beyond the Basics Using SAS, Third Edition](#), SAS Institute Inc., Cary, NC, USA.
- Lafler, Kirk Paul (2016), [“What’s Hot – Skills for SAS Professionals”](#), Proceedings of the SouthEast SAS Users Group (SESUG) Conference – 2016.
- Lafler, Kirk Paul (2016), [“What’s Hot – Skills for SAS Professionals”](#), Proceedings of the Pharmaceutical SAS Users Group (PharmaSUG) Conference – 2016.
- Peterson, Richard (July 26, 2022 Update). [“17 Top Cloud Service Providers \(2022 Updated\)”](#), guru99.com.

Acknowledgments

The author wishes to thank the 2023 IowaSUG 2023 Conference Committee for accepting my paper; the 2023 IowaSUG 2023 Executive Committee for organizing and supporting a “live” conference event; SAS Institute Inc. for providing SAS users with wonderful software; and SAS users everywhere for being the nicest people anywhere!

Trademark Citations

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

Author Information

Kirk Paul Lafler is an educator, developer, programmer, consultant, and data analyst; and works as an adjunct professor at San Diego State University; an advisor and adjunct professor at the University of California San Diego Extension; and an educator teaching SAS, SQL, Python, Excel, and cloud-based technology courses, workshops, and webinars to users around the world. Kirk has nearly 50 years of programming experience and specializes in SAS software, SQL, RDBMS technologies (Oracle, SQL-Server, Teradata, DB2), Python, and other languages and productivity tools. As an author, Kirk's most recent and/or upcoming books include Exploratory Data Analysis (EDA) By Example (PB&J Press. Summer 2023), Python ETL By Example (PB&J Press, Spring 2024), PROC SQL: Beyond the Basics Using SAS, Third Edition (SAS Press. 2019), and other books along with numerous papers and articles on a variety of SAS, SQL, and Python topics. Kirk is actively involved with international, regional, and local SAS, SQL, and Python user groups and conferences as an Invited speaker, educator, keynote, and section leader; and is the recipient of 27 "Best" contributed paper, hands-on workshop (HOW), and poster awards.

Comments and suggestions can be sent to:

Kirk Paul Lafler, sasNerd

SAS® / SQL / Python Consultant, Application Developer, Programmer, Data Analyst, Educator and Author

E-mail: KirkLafler@cs.com

LinkedIn: <https://www.linkedin.com/in/KirkPaulLafler/>

Twitter: @sasNerd