Let’s build our project on the most valued data science skills by aggregating the findings from the following studies and websites:

**From the web:**

* **Harvard Business School (** Online’s Business Insights Blog: 03 AUG 2021)

Source: <https://online.hbs.edu/blog/post/data-science-skills-for-beginners>

LIST:  **9 data science skills for beginners**:

1. Basic Data Literacy
2. Domain Fluency
3. Data Generation and Collection
4. Data Manipulation
5. Analytical Skills
6. Data Ethics
7. Critical Thinking Abilities
8. Communication Skills
9. Mathematical and Programming Skills

* **Coursera’s blog** post lists **7 essential skills for a data scientist**.

Source:<https://www.bing.com/search?q=+Coursera%E2%80%99s+blog+post+lists+7+essential+skills+for+a+data+scientist.&qs=n&form=QBRE&sp=->

**LIST : 7 essential skills for a data scientist**

1. Programming Programming languages, such as Python or R, are necessary for data scientists to sort, analyze, and manage large amounts of data (commonly referred to as “big data”). ...

2. Statistics and probability...

3. Data wrangling and database management...

4. Machine learning and deep learning...

5. Data visualization...

6. Cloud computing...

7. Interpersonal skills...

* **Indeed.com** provides a list of **15 top skills that a data scientist may possess**.

**Source:** [**https://www.indeed.com/career-advice/resumes-cover-letters/skills-for-a-data-scientist**](https://www.indeed.com/career-advice/resumes-cover-letters/skills-for-a-data-scientist)

**LIST :**

1. Cloud computing
2. 2. Statistics and probability
3. Advanced mathematics
4. Machine learning
5. Data visualization tools
6. Query languages
7. Database management
8. Python coding
9. Microsoft Excel
10. R programming
11. Data wrangling

Soft skills:

1. Independence
2. Communication
3. Project management
4. Analytical thinking

* **Free CodeCamp :**

**Source:** [**https://www.codecademy.com/resources/blog/machine-learning-programming-languages/**](https://www.codecademy.com/resources/blog/machine-learning-programming-languages/)

[**https://www.freecodecamp.org/news/first-steps-to-learn-data-science-or-ml-after-the-roadmap/**](https://www.freecodecamp.org/news/first-steps-to-learn-data-science-or-ml-after-the-roadmap/)

List:

1. Python

2. R

3. Java and 4. JavaScript

5. C++

6. Shell

7. Go (Golang)

Main Areas of Focus:

1. Deep learning
2. Machine learning
3. Applied statistics
4. 4 EDA & storing telling
5. Data extraction and wrangling
6. Programming

* **KAGGLE**

**Source :** [**https://www.shiksha.com/online-courses/articles/kaggle-datasets-for-practice/#Data-Scientists-to-shine-in-Kaggle**](https://www.shiksha.com/online-courses/articles/kaggle-datasets-for-practice/#Data-Scientists-to-shine-in-Kaggle)

**LIST :**

1. Machine Learning
2. Statistics & Mathematics
3. Business Acumen & Critical Thinking
4. Storytelling & Communication
5. Programming Skills

* **Udacity.com**

**Source :** [**https://www.udacity.com/blog/2020/11/skills-to-launch-data-science-career.html#:~:text=8%20Must-Have%20Skills%20for%20Data%20Scientists%201%20%231.,%237.%20Communication%20...%208%20%238.%20Business%20Acumen%20**](https://www.udacity.com/blog/2020/11/skills-to-launch-data-science-career.html#:~:text=8%20Must-Have%20Skills%20for%20Data%20Scientists%201%20%231.,%237.%20Communication%20...%208%20%238.%20Business%20Acumen%20)

**LIST :**

1. Math and Statistics
2. Analytics and Modeling
3. Machine Learning Methods
4. Programming
5. Data Visualization
6. Intellectual Curiosity
7. Communication
8. Communication

* **Intellspot**

Source: [35 Data Scientist Qualifications & Skills You Need To Succeed (intellspot.com)](https://www.intellspot.com/data-scientist-qualifications/)

List:

1. R
2. Python
3. Java
4. SQL
5. SAS
6. MATLAB
7. Scala
8. Julia
9. Discrete vs. Continuous Data
10. Binomial Distribution
11. Regression
12. Hypothesis Testing
13. Bayesian Thinking & Modeling
14. Machine Learning
15. Markov Chains
16. KNIME
17. Weka
18. RapidMiner
19. Apache Hadoop
20. Apache Spark
21. Data Melt
22. Apache Storm
23. TensorFlow
24. Tableau
25. Google chart
26. D3.js
27. JupyteR
28. Orange
29. Communication Skills
30. Data-driven Decision Making
31. Domain Knowledge And Business Acumen
32. Teamwork
33. Intellectual Curiosity And Passion For Work
34. Good Data Intuition
35. Project Management Skills

**FROM ACADEMIA**: Academic Journal Articles:

**FIRST STUDY:**

Wilbur W Stanton & Angela A Stanton (2020). Helping Business Students Acquire the Skills Needed for a Career in Analytics: A Comprehensive Industry Assessment of Entry-Level Requirements. *Decision Sciences Journal of Innovative Education. Volume 18, Issue1.* Retrieved from: <https://onlinelibrary-wiley-com.ez.lib.jjay.cuny.edu/share/N9Y2QSCNFDAFZ3J6YUCB?target=10.1111/dsji.12199>

**Study Research Question (RQ):** What are the credentials and skillsets in greatest demand for data science and analytics professionals?

**Study’s method:** Conducted an analysis of job postings on LinkedIn, based on U.S.-based position

**Study’s findings:**

**Figure 1. Analytics and data science job announcements. (Note: Data collected on LinkedIn on 20 April 2019)**

**Exhibit 1.**Typical LinkedIn analytics and data science entry-level job postings

|  |
| --- |
| **Data Science**: An example of the qualifications for an entry-level job for a large IT consulting company  *Requirements*   * Work experience in addition to degree: 3-5 years of data engineering/science-related activities with overall 10+ years’ experience. * Graduation from a 4-year college or university with a degree in statistics, physics, mathematics, engineering, computer science, or management of information systems. * Expert knowledge of SAS, Python Machine Learning, or R * Working knowledge of statistics, programming, and predictive modeling. * Working knowledge of code writing * Big Data/Hadoop/NoSQL and experience with large datasets * Working knowledge of SQL, QlikView, and Data Architecture |
| **Data Analytics**: An example of the qualifications for an entry-level job for a Large Pharmaceuticals Company  *Requirements*   * BS/BA required. Advanced degree in a quantitative subject (e.g., Statistics, Econometrics, Mathematics) preferred * Minimum of 3 years of analytical experience in pharmaceutical or biotechnology industries. * Commercial business intelligence tool report creation experience required (e.g., QlikView, Tableau, Oracle BI, Microstrategy, Cognos, Spotfire) * SAS code-writing skills (BASE, STAT, SQL, Macros) preferred, SQL query writing required. * Advanced proficiency in Excel, and PowerPoint required * Presentation skills to take sophisticated analysis and apply a business context for nonanalytical associates required |
| **Business Analytics**: An example the qualifications for an entry-level job for a Big Box Retailer **Note**: Data collected on LinkedIn on 20 April 2019  *Minimum Qualifications*   * Bachelor's degree in business, analytics, statistics, or related field and 2 years of experience in data analytics, project management, business, or related area OR 4 years' experience in data analytics, project management, business, or related area.   *Additional Preferred Qualifications*   * 3 years of experience with statistical programming languages. * Master's degree in business, analytics, or related field. * Project management certification. * Six Sigma certification. |

**Table 1.**Top 20 Credential requirements for an entry-level position

| **Data Science** |  | **Data Analytics** |  | **Business Analytics** |  |
| --- | --- | --- | --- | --- | --- |
| **Position requirements** | **%** | **Position requirements** | **%** | **Position requirements** | **%** |
| Prior experience | 80.7% | Prior experience | 76.1% | Prior experience | 78.4% |
| Degree in computer science | 38.2% | Degree in business | 44.3% | Degree in business | 67.1% |
| Degree in management | 33.5% | Degree in management | 42.3% | Degree in management | 49.3% |
| Degree in engineering | 31.3% | Degree in engineering | 26.4% | Bachelor's degree | 27.1% |
| Degree in business | 29.2% | 5+ years of experience | 25.7% | 5+ years of experience | 23.8% |
| Bachelor's degree | 25.9% | Bachelor's degree | 25.6% | Degree in engineering | 20.5% |
| 5+ years of experience | 20.7% | Degree in computer science | 23.3% | Degree in computer science | 19.9% |
| Degree in information systems | 13.7% | Certifications | 15.9% | Degree in marketing | 19.4% |
| Master's degree | 11.1% | Degree in marketing | 14.3% | Degree in business intelligence | 15.3% |
| Degree in information science | 10.3% | Degree in statistics | 12.9% | Certifications | 14.9% |
| Degree in IT | 10.0% | Degree in information Systems | 12.2% | Degree in finance | 13.9% |
| Degree in mathematics | 9.8% | Master's degree | 11.5% | Degree in information systems | 13.1% |
| Degree in statistics | 9.0% | Degree in mathematics | 10.1% | Quantitative degree | 13.0% |
| Degree in decision science | 6.8% | Degree in IT | 10.0% | Degree in statistics | 12.2% |
| Quantitative degree | 6.4% | Degree in business intelligence | 8.8% | Degree in IT | 11.2% |
| Degree in MIS | 6.0% | Quantitative degree | 8.8% | Master's degree | 9.7% |
| 1-3 Years of experience | 5.8% | Degree in finance | 8.4% | Microsoft certifications | 8.6% |
| Degree in marketing | 4.7% | 1-3 Years of experience | 7.8% | Degree in operations mgnt | 8.5% |
| Degree in business intelligence | 4.6% | Degree in operations mgnt | 7.3% | Degree in MIS | 7.5% |
| Degree in operations mgnt | 3.7% | Degree in economics | 6.4% | Degree in accounting | 7.5% |

**Hard skills**

**Table 2.**Top 20 hard skill requirements for an entry-level position

| **Data Science** |  | **Data Analytics** |  | **Business Analytics** |  |
| --- | --- | --- | --- | --- | --- |
| **Position requirements** | **%** | **Position requirements** | **%** | **Position requirements** | **%** |
| Data analysis | 27.5% | Data analysis | 35.8% | Data analysis | 30.3% |
| Programming | 22.2% | Programming | 21.7% | Modeling | 17.9% |
| AI (artificial intelligence) | 19.6% | Modeling | 18.8% | Business strategy | 17.8% |
| Dashboards | 18.5% | Machine learning | 17.7% | Programming | 15.7% |
| Machine learning | 16.3% | Advanced analytics | 15.1% | Financial analytics | 15.1% |
| Modeling | 14.5% | Information technology | 13.3% | Information technology | 14.0% |
| Data processing | 12.0% | Data visualization | 13.2% | Web analytics | 13.8% |
| Developing algorithms | 11.5% | Data processing | 12.1% | Google analytics | 12.9% |
| Information technology | 11.2% | Optimization | 11.2% | Process management | 12.8% |
| Data architecture | 8.0% | Digital analytics | 10.3% | Financial services analytics | 12.3% |
| Data visualization | 7.8% | Operations management | 9.9% | Advanced analytics | 12.0% |
| Database development | 7.1% | Algorithms | 9.9% | Machine learning | 11.7% |
| Optimization | 6.9% | Data architecture | 9.8% | Operations management | 11.2% |
| New product development | 6.0% | Google analytics | 9.5% | Digital analytics | 10.9% |
| Data warehouses | 5.4% | Business strategy | 9.0% | Data visualization | 10.8% |
| Data mining | 5.2% | Web analytics | 8.9% | Business metrics | 9.9% |
| Process management | 4.7% | Data warehouses | 8.7% | Dashboards | 9.8% |
| Operations management | 4.4% | Process management | 8.6% | Digital marketing | 9.4% |
| Enterprise systems | 4.3% | Data mining | 8.6% | Marketing research | 9.3% |

**Soft skills**

**Table 3.**Soft skill requirements for an entry-level position

| **Data Science** |  | **Data Analytics** |  | **Business Analytics** |  |
| --- | --- | --- | --- | --- | --- |
| **Position requirements** | **%** | **Position requirements** | **%** | **Position requirements** | **%** |
| Data driven | 29.2% | Analytical skills | 29.2% | Analytical skills | 45.7% |
| Analytical skills | 21.8% | Problem-solving skills | 26.1% | Problem-solving skills | 35.0% |
| Written communication skills | 19.3% | Big data experience | 25.8% | Written communication skills | 30.0% |
| Mathematics (skills/abilities) | 18.3% | Ability to work on teams | 25.5% | Ability to work on teams | 27.3% |
| Ability to work on teams | 16.2% | Written communication skills | 25.5% | Verbal communication skills | 26.5% |
| Management science skills | 16.0% | Innovative | 22.7% | Leadership ability | 25.4% |
| Verbal communication skills | 15.5% | Verbal communication skills | 21.8% | Innovative | 21.2% |
| Innovative | 14.7% | Leadership ability | 20.6% | Skills in business intelligence | 21.1% |
| Big data experience | 13.5% | Statistics (skills/abilities) | 16.4% | Interpersonal skills | 16.0% |
| Problem-solving skills | 12.9% | Data driven | 16.0% | Quantitative skills | 15.7% |
| Skills in information science | 12.1% | Ability to work cross-functionally | 13.2% | Big data experience | 15.6% |
| Statistics (skills/abilities) | 12.0% | Skills in business intelligence | 13.0% | Ability to work cross-functionally | 14.3% |
| Leadership ability | 12.0% | Mathematics (skills/abilities) | 12.0% | Statistics (skills/abilities) | 14.3% |
| Ability to work collaboratively | 11.6% | Interpersonal skills | 11.6% | Consulting ability | 13.8% |
| Ability to work cross-functionally | 7.5% | Quantitative skills | 10.0% | Data driven | 12.9% |
| Quantitative skills | 6.5% | Ability to work collaboratively | 15.3% | Management science skills | 12.8% |
| Skills in business intelligence | 6.1% | Management science skills | 9.0% | Ability to work collaboratively | 19.7% |
| Consulting ability | 4.1% | Consulting ability | 8.7% | Mathematics (skills/abilities) | 10.5% |
| Presentation skills/abilities | 3.5% | Presentation skills/abilities | 7.0% | Provide business insights | 8.9% |
| Ability to work independently | 3.4% | Skills in information science | 6.3% | Presentation skills/abilities | 8.5% |

**Software skills**

**Table 4.**Software skills requirements for an entry-level position

| **Data Science** |  | **Data Analytics** |  | **Business Analytics** |  |
| --- | --- | --- | --- | --- | --- |
| **Position requirements** | **%** | **Position requirements** | **%** | **Position requirements** | **%** |
| Python | 25.7% | SQL | 31.8% | SQL | 29.9% |
| SQL | 23.5% | Python | 28.1% | Python | 24.1% |
| Java | 17.8% | Java | 16.5% | Java | 18.0% |
| R | 11.4% | R | 16.3% | Microsoft Excel | 13.3% |
| AWS | 10.5% | AWS | 12.6% | R | 12.8% |
| Hadoop | 9.2% | Hadoop | 12.4% | PowerPoint | 12.4% |
| Oracle | 6.7% | Tableau | 12.0% | Tableau | 12.3% |
| Tableau | 5.5% | Microsoft Excel | 9.1% | Experience in MS products | 9.5% |
| Scala | 5.3% | SAS | 8.5% | SAS | 8.2% |
| C/C++ | 5.0% | Oracle | 8.5% | Oracle | 8.1% |
| SAS | 4.3% | PowerPoint | 8.0% | Hadoop | 7.9% |
| Azure | 4.2% | Scala | 6.0% | AWS | 7.7% |
| Microsoft Excel | 4.2% | Azure | 5.9% | Microsoft Access | 5.6% |
| Matlab | 3.9% | Experience in MS products | 5.8% | Azure | 4.8% |
| PowerPoint | 3.4% | Microsoft Word | 5.1% | SAP | 4.7% |
| Experience in MS products | 3.3% | Microsoft Access | 5.0% | Microsoft Word | 4.4% |
| Perl | 3.3% | C/C++ | 3.3% | Scala | 3.5% |
| Microsoft Access | 3.2% | Perl | 2.9% | SPSS | 2.6% |
| Microsoft Word | 3.2% | Redshift | 2.8% | Redshift | 2.4% |
| Redshift | 2.0% | Matlab | 2.7% | Perl | 2.3% |

**SECOND STUDY:**

Smaldone et al. (2022). Employability skills: Profiling data scientists in the digital labour market. *European Management Journal.* Volume 40, *Issue 5, October 2022, Pages 671-684*

[*https://doi.org/10.1016/j.emj.2022.05.005*](https://doi.org/10.1016/j.emj.2022.05.005)

**METHOD : *“*** job advertisements were extracted from job search portals via web scraping to collect data for processing (Blázquez Soriano et al., 2012). Web scraping is defined as “a technique of digital data extraction from a website through a software parsing in real-time instances wide-spreading hyperlinks reality” (Munzert et al., 2014, p. 11). To extract information from employment websites, a two-step process was conducted: (1) detecting specific page URLs and (2) extracting information and target data by completing the pre-designed software task. Octoparse software (https://www.octoparse.com/) was used for the parsing processes (Slamet et al., 2016). After choosing the population, the existence of “data scientist” as a pre-defined keyword for the opened job vacancies was verified to avoid a self-selection bias. Then, the sampling method to perform was defined. According to the literature (Iachan, 1982; Madow, 1949; Madow & Madow, 1944), systematic sampling was employed, choosing the first sample according to random sampling and then keeping a fixed interval of

, representing the value of the sampling interval (

) and thus ensuring reliability. The choice of the US context was motivated by the fact that this market offers around 5000 more job listings with the keyword “data scientist” than does the European counterpart (6000 observations in the US market against 3000 observations in the UK market, 2500 observations in the French market and just over 800 for the Italian market).

Performing web scraping, a sample of 1395 job advertisements from the US market was obtained, recording 12,872 words in 1383 documents after excluding invalid cases. One of the chosen portals is classified as the third best job search website in the USA, as well as the most accessible website for web scraping; thus, adequacy was ensured through the screen-scraping interface (Liu et al., 2010; Robert Half, 2021). Extracted data were analysed using R software (https://www.r-project.org/). Five dimensions were extracted from each job advertisement: firm, location, open position, job description, and recommendations. The rationale for the breakdown of the dimensions was that multiple variables were required for analytic purposes, and thus the most common item labelling already present in the data source *was followed****”***

**STUDY FINDINGS:**

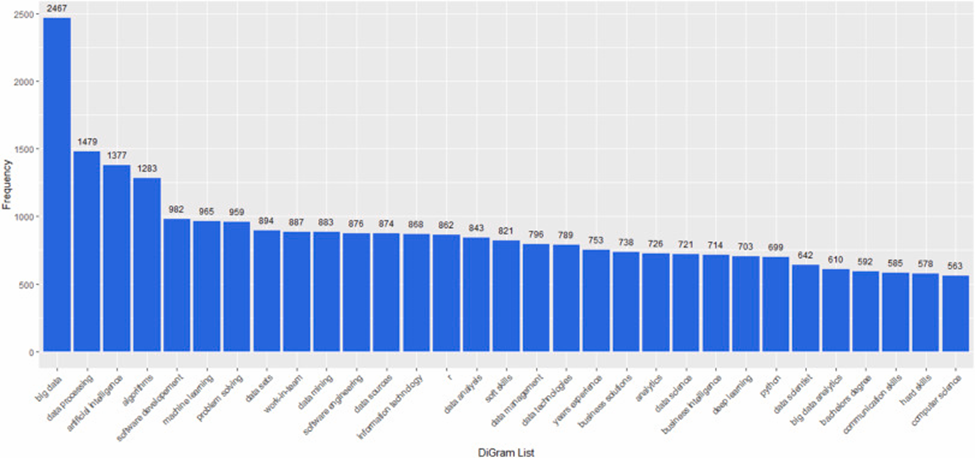
Fig. 1. Bar plot with frequency count. 



Fig. 2. Word cloud after sparsity reduction.

Fig. 3. Topic modelling of employability skills.

