

# **Data Science: Jobs, Salary, and Education**

Ameya Kshirsagar, Rushali Bhudeka and Prachi Yadav

Department of iSchool, University of Wisconsin-Madison, Madison, Wisconsin.

## **Abstract**

With shifting dynamics, corporations and institutions are now looking at Data Science as a union of multiple talents and credentials rather than as a one-qualification profession. As a result, Data Science positions are being subdivided into smaller, more particular titles that need certain qualifications and expertise. As a result, this study seeks to investigate and comprehend current market dynamics and educational institute courses to answer the following questions: Companies' desired and used skills, as well as the highest paying Data Science roles/titles. We try to uncover capabilities that are a great fit for a certain Data Science Role in the work industry. The dataset used in the investigation was obtained from two sources: Kaggle and the ai-jobs.net Salaries website. The data set consists of a publicly filled survey on the internet. There were several notable findings about the distribution of data scientist skill sets and salary. The average compensation for data science positions is roughly \$120,000. Moreover, 40% of people in the sector feel that independent projects may demonstrate knowledge better than academic success. As a result, this study may be utilized by enthusiasts for additional analysis about the construction of the Data Science curriculum based on in-demand skills and more. Hence this study can be used by enthusiasts for further analysis regarding designing the Data Science curriculum based on skills in demand and more.

## **Introduction**

There was a time when technology was regarded as a luxury, and merely the upper class could utilize it. Solely the upper class could make a phone call to a distant country. Nevertheless, the times have changed, and it would not be inaccurate to say that technology today is essential for survival in society. Technology's impact has revolutionized businesses, governmental operations, institutions, and ordinary men's and women's everyday lives. This growth is visible to us in various forms. However, there are specific technical aspects that we need to be made aware of or ignorant of.

One such aspect is the collection of information by different technology platforms and devices that we are surrounded by. Every day, computerized judgments are made using this information. The data is gathered from waking up until retiring to bed at night. As an illustration, when we open the music app to listen to music, the app analyzes the theme we select and the music we listen to and then makes recommendations based on that analysis. We do not have to decide what to watch because Netflix will do that for us. Watching history is used to generate suggestions for us. We open Instagram, browse the feed, watch a few videos, and then close the program (Saltz & Dewar, 2019).

When we open the app again after a while, Instagram will exhibit similar reels as it has been keeping track of the activity and saving the information to display reels or posts that users may be

interested in, encouraging them to stay on the app longer (Rettberg, 2020). These programs continuously save users' data, which is then used to generate automatic judgments. The advertisements we see are personalized, and with each click, more and more data is being produced. With millions of clicks co-occurring worldwide, we can only imagine generating and collecting the numbers each second.

Thus, Big Data is a new way of monopolization. If we think about it, data is just numbers. The question is, who is working behind the scenes to achieve the desired derivatives? How is this information used to make decisions? Data is not helpful until processed; it is meaningless until it is handled or interpreted. The rise in demand to analyze the data and make significance out of it has resulted in the origin of the term Data Science.

Simply put, Data Science is a contemplation that makes sense of large numbers. Humans create algorithms or models that process gigantic amounts of data, generate output, and make the information meaningful. Moreover, Data Science studies data and how to analyze, process, and perpetrate various operations on it to develop output and create a business worth (Davenport & Patil, 2012b) Data Scientists are professionals with the specialized knowledge and skills to handle data and answer questions to improve business values. Regardless, a decade into data science, there is still no exact checkboxes, measures, or definition defining "Data Scientist."

The word Data Science has stood around for quite a while. It was first cited by John Wilder Turkey, an American mathematician, in 1962 in The Annals of Mathematical Statistics journal. (Apoorva Komarraju, 2021). However, the term gained popularity with exponentially expanding data and larger organizations realizing the workforce shortage to manage it. The Covid-19 Pandemic is another significant reason that brought so much attention to Data Science jobs. It pushed people to adapt to specific technologies that they never relied on, resulting in additional data supply. Even then, the IT need is redirecting in such a manner that it has given rise to thousands of Data Science positions. We can see Data Science as a set with multiple different sub-sets bisecting each other. The rise of so many Data Science Jobs has caught the attention of institutions, universities, governments, and the working force around the world.

People from all backgrounds, such as Computer Science, Information Science, Psychologists, Statistics, Economics, and Journalism, are diverting toward data science roles and jobs. (Davenport & Patil, 2012a) As a result, Data Science has been one of the highest-paying careers in the last century. At the peak of its attention, companies and institutions paid heavy salaries to people who fit specific criteria and had the skills to be Data Scientists.

Other than the job market, the one sector that was favorably influenced and impacted by the growth in Data Science is Educational Institutions. With thousands of academic and non-academic programs being presented, professionals and students are investing in them to enter the highest-paid workforce. Not just well-known Universities and Schools but many online outlets are qualifying professionals from every background to be prospective Data Scientists.

After gaining popularity, Data science jobs face retrogression as organizations realize that the domain is not just about analyzing numbers and representing them; any analyzing tool in the market could do it for them. Most AI and data science companies fail to follow ethics and audits,

resulting in being acquired by more giant sharks in the market. Even academic and non-academic institutions struggle to provide a single program that could include all the knacks required to be a data scientist. (Davenport & Patil, 2012a) With changing dynamics, companies, and institutions are now looking at something other than Data Science as a one-qualification job but as a union of different skills and qualifications. As a result, Data Science jobs are now being divided into smaller and more specific titles that meet certain qualifications and skills.

Therefore, this paper attempts to research and comprehend the current market dynamics and courses offered in educational institutes to answer the following.

1. Skills preferred and used by the companies.
2. Highest Paying Data Science Role/Titles.

We attempt to find the skills that fit perfectly with a particular Data Science Role in the job sector.

## **Literature Review**

(Davenport & Patil, 2012a) The term data scientist has been around for many years, but no particular skill set still describes them. Data scientists are not only those with robust data and computational skills. The paper describes data scientists as someone who can understand customer needs and business requirements and provide solutions to meet business needs. One of the essential skills that hiring professionals seek in data scientists is how effectively they can communicate the data to the stakeholders so they would understand. The paper mentions that the most desired skill in a data scientist is "curiosity," curiosity to dig deeper into the problem and provide the solutions to meet the business requirements. Most data scientists today have degrees in computer science, economics, or statistics, but they can also work in environments with a solid computational foundation and a data concentration. When hiring data scientists, keeping the word scientist in mind is essential, as the word "data" can be easily misleading. With the abundance of data, there are many professionals whose work and skills differ from data scientists. For example, a data management expert's role is to generate and organize the data, but he may need to learn how to use the unstructured data, analyze it and communicate it. The paper emphasizes the importance of social skills for a good data scientist. The dynamics of hiring have changed, and several hiring professionals are clear about what they are looking for while hiring data scientists. Primarily, they look into the candidate's skill sets, and secondary, they look for the degrees.

(Davenport & Patil, 2012b) Ten years ago, Harvard Business Review posted the article "Data Scientist: The Sexiest Job of the 21st Century" Does that still stand true? The paper covers the four major points of how data scientist roles and jobs have evolved since then. With the evolution of big data and more data collection, companies realized they needed to hire more people skilled in analyzing, programming, and communicating the data. Before 2012, this requirement was limited to the San Francisco Bay Area, and coastal cities and startups wanted to hire all the data scientists they could get. This evolution of big data has created more jobs, so the demand for data scientists has increased. At that time, HBR defined data scientists as "high-ranking professionals with the training and curiosity to make discoveries in the world of big data." As no particular career path was defined for people who could analyze, clean, and program, data scientists came from different backgrounds. Most people working in this field had Ph.D. in some scientific field and were

exceptionally good at solving complex problems and code. After a decade, data scientists are more in demand than ever, and every startup, small or big tech firms, is actively looking to hire them. According to the paper, data scientists' requirement by 2019, increased by 256% and is still projected to see more growth than any other industry. There were no data science programs earlier (2012), and data science professionals were hired from different backgrounds, such as statisticians, for the data scientist job. Changes in technology that data scientists are doing are also one of the main reasons why the data science job is still changing or vague. A decade ago, coding was one of the primary skills needed for a data scientist, but now it has been shifted to other jobs or is increasingly automated. Ethics is the most significant change in the field of data science job. People are more focused on ethics in data science, whereas earlier, it was not much considered. Ethics and data science go hand in hand. Some companies and courses have already involved ethics courses, along with data science, to make people more aware of ethical issues. With technology change, data science skills and requirements will continue to change, and so will the demand in the market. However, the question remains, is data science still the sexiest job?

(Pryor & Donnelly, 2009) The authors of this paper, Graham Pryor and Martin Donnelly discuss the roles and responsibilities of the "data workforce" and provide their insights from the DCC Research Data Management Forum and ways to improve capability, capacity, and status. While the authors tried to differentiate between four roles: data creator, data scientist, data manager, and data librarian, they admitted that no such terminology exists in the data community. Data creators can be described as experts in handling and manipulating data. Data scientists work with data creators to analyze and make sense of the raw data. Data managers are responsible for storing, accessing, and preserving data. Data librarians are the only ones officially trained in organizing, presenting, and archiving data. The paper further talks about the skill sets of data librarians. Data librarians can possess various skills, from "librarian skills" to "statistics skills." The paper gives details about various data-related educational programs in UK universities. Macdonald and Martinez-Urbe's observation indicates that the university information service could provide excellent opportunities for data professionals to develop a clear career goal.

]

(R. Geiger et al., 2018) The paper presents a report based on the survey of members and associates of data science majors of three different institutes in the US. This report identifies and addresses the situations concerning data science in academia, and focuses primarily on concerns related to data science careers. The paper further suggests roles and responsibilities related to data and computation fields across academics. Data science is in demand across the public and private sectors. Many universities have offered data science courses focused on building a successful career. The definition of "data scientist" is still unclear, though data science umbrellas several titles that come under the role of the data scientist, primarily to support or are adjacent to the data science role.

(Louis Columbus, 2017) The paper predicts the job growth of data scientists by 2020. The paper mentions the insights and numbers related to data science skills and jobs. Jobs that require a candidate to be skilled in machine learning are willing to pay an average of \$114,000. Data scientists pay an average of \$105,000. Most Data Science and Analytics jobs (59%) are in the IT, Finance, Professional Services, and Insurance sectors. The paper also predicted that data science and analytics jobs would grow by 15% by 2020. The most in-demand skills in analytics include machine learning, Apache Pig, Apache Hive, and Apache Hadoop.

(Ismail & Zainal Abidin, 2016) Decision-making is one of the most essential aspects for a business to grow and gain profit and helps stakeholders take decisions that benefit the organization. Stakeholders take their decisions based on precise, complete, and accurate data. Even if Big Data Analytics has a wide range of tools that can help decision-making, it is still important to employ people with the knowledge and abilities to do so. The paper aims to outline the most in-demand skills for data scientists. The paper mostly focuses on the top 5 skills, which are Statistics, Machine Learning, business, communication, and analysis. As per the paper, the data scientist first analyzes the data using analytics and data viz tools to find patterns in the data. The data scientist will then perform data migration, cleaning, integration, and analysis, and deliver the outcome. A data scientist should write code in R, Python, Java, Ruby, etc., and be familiar with NLP. One of the most essential skills a data scientist should have is communication skills to understand and cater to business needs. Data scientist acts as a bridge between the business domain and the technical domain. The paper further mentions that data scientists should also possess expertise in data mining, visualization, modeling, and machine learning to predict the future based on the past and present so that the existing business process could be improved (Ismail & Zainal Abidin, 2016). The paper concludes that even though there are so many skills required to be an expert data scientist, the skills vary from organization to organization. Communication and statistics skills are the most important, which are demanded by every other organization to fulfill business needs.

(Smaldone et al., 2022) With the increase in data, the demand for skilled data scientists have also increased. A data scientist's job is to make sense of such vast and complex data. This paper aims to list the skills that are in demand by today's professionals who seek to hire data scientists with relevant skills. The paper collects the relevant data on the skills required from the job advertisements of the Data Scientist role published on US employment websites with the highest recruitment traffic rate. The paper tries to map out the skills, experience, and qualifications of skilled data scientists to bridge the skill gaps and mismatches. According to the International Labour Organization, the gap between applicants' skills and desired skills by the entrepreneur is increasing (2018), which emphasizes the need to list the skills for data scientist roles. The paper divides employability skills into two categories: hard skills and soft skills. Hard skills are "specific capabilities to perform a specific job". These skills are learned over time with practice and study, hence depending on training and experience. Moreover, hard skills can be measured, quantified, and evaluated. Hard skills requirements can differ from organization to organization, but they might have the use of some equipment or systematic way of assessing or working with the data, or the use of relevant software, programming language, statistical knowledge, and machine learning. Soft skills are also one of the important skills required for the data scientist job role. It can be seen as complementary to hard skills, which help communicate important information to different domains for the business to flourish. With the increasing amount of data, business professionals seek an applicant who can get the data, clean it, analyze it and convert the data into useful information. The qualifications of a data scientist include statistical analysis, machine learning, and programming. The data scientist also requires soft skills, such as analytical problem solving, effective communication skills, technological aptitude, team player, etc.

## **Research Methodology**

Exploratory data analysis is a methodology adapted to generate a statistical and visual analysis for the targeted dataset before working on the focused research query to understand better the dataset and the behavioral aspect of its columns and rows concerning each other. Before diving deeper into data science-related jobs, salaries, and titles, as mentioned afore that the start of data science was in the 1960s as a subset of computer science whose primary emphasis was on data analysis and management; it took nearly data science nearly 40 years to gain its popularity in 2000s as its growing ability of statistical application of artificial intelligence, deep learning, and machine learning models which soon resulted in the rise of data science related jobs and its commercial need in 2010s.

The dataset selected for the analysis is taken from two sources, i.e., Kaggle and ai-jobs.net Salaries website. The data set is a publicly filled survey on the website. The collected dataset is dated from 2020 to 2022. Moreover, the data set is updated weekly with the data entries processed from the salary survey, and the entire dataset is available in the public domain under the CC0 license (ImageSuggest, 2021). The data set, for now, consists of 1037 rows and 11 columns. The dataset consists of following attributes; work\_year, experience\_level, employment\_type, job\_title, salary, salary\_currency, salary\_in\_usd, employee\_residence, remote\_ratio, company\_location and company\_size. Where the attributes like experience\_level (Entry level, Mid, Senior, and executive level), employment\_type (Part-time, contract, freelance, and full-time), remote\_ratio (fully remote, partially remote, and on-site), and company\_size (small, mid and large scale industries) are further categorized into specifics (<https://salaries.ai-jobs.net/>).

After acquiring the dataset, we checked for null values, resulting in no null values. And some of the attributes required tweaking as they were in string format which needed to be changed to categorical data like experience\_level, employment\_type, remote\_ratio, and company\_size. Additionally, the columns “salary” and “salary\_currency” were dropped as “salary\_in\_usd” plays a better role in streamlining the analysis by making comparison metrics more uniform.

For analysis of Data science related skills, we again got the dataset from Kaggle via its annual Machine Learning & Data Science survey, and we analyzed the dataset from 2018 and 2019’s surveys as they are dated right before the data science job dataset distribution giving a better idea of what skills people are taking into the succeeding year. To clean this huge dataset, we made a new CSV file containing all the questions from each of the surveys and passed the survey dataset setting the questionnaire dataset as a filter mechanism to retrieve a new subset dataset to analyze, as shown in figure 1.

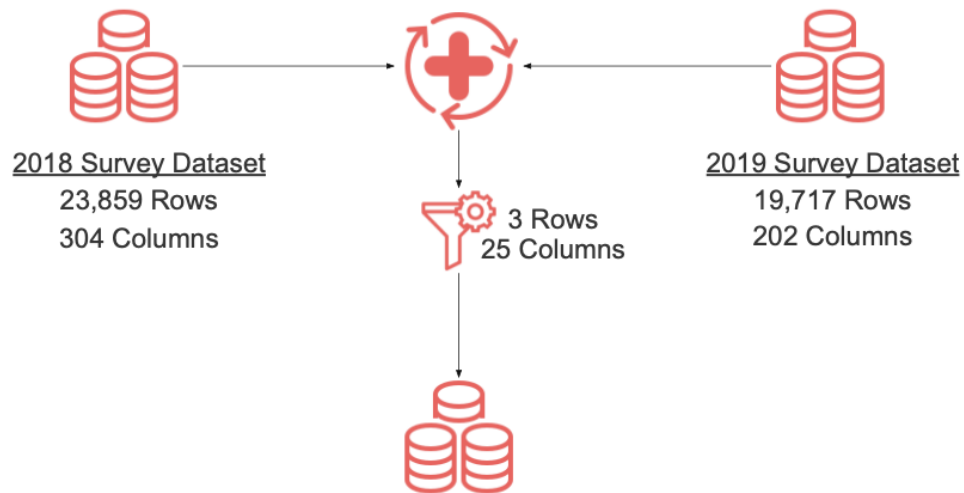


Figure 1: Kaggle survey dataset cleaning process.

## EDA Workflow

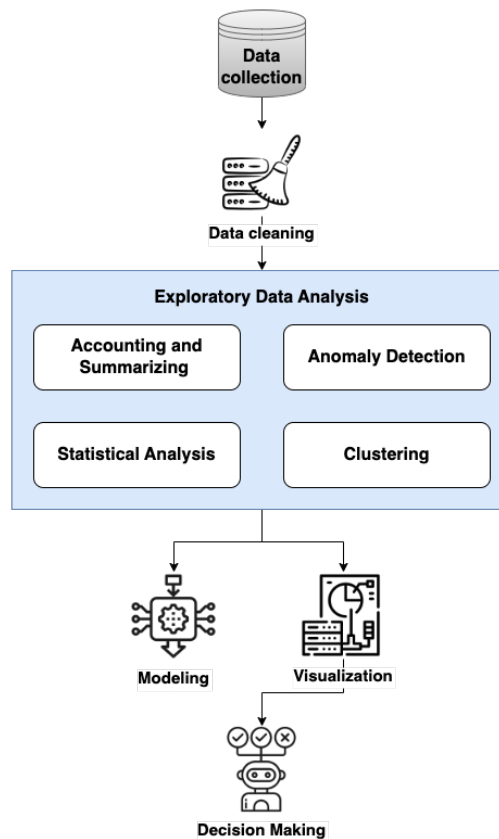


Figure 2: Workflow of EDA and research methodology adopted.

EDA involves summarizing data numerically and graphically and preparing data for more formal modeling techniques, as shown in figure 2 (Ameya Kshirsagar & Parth Sanghavi, 2022). By

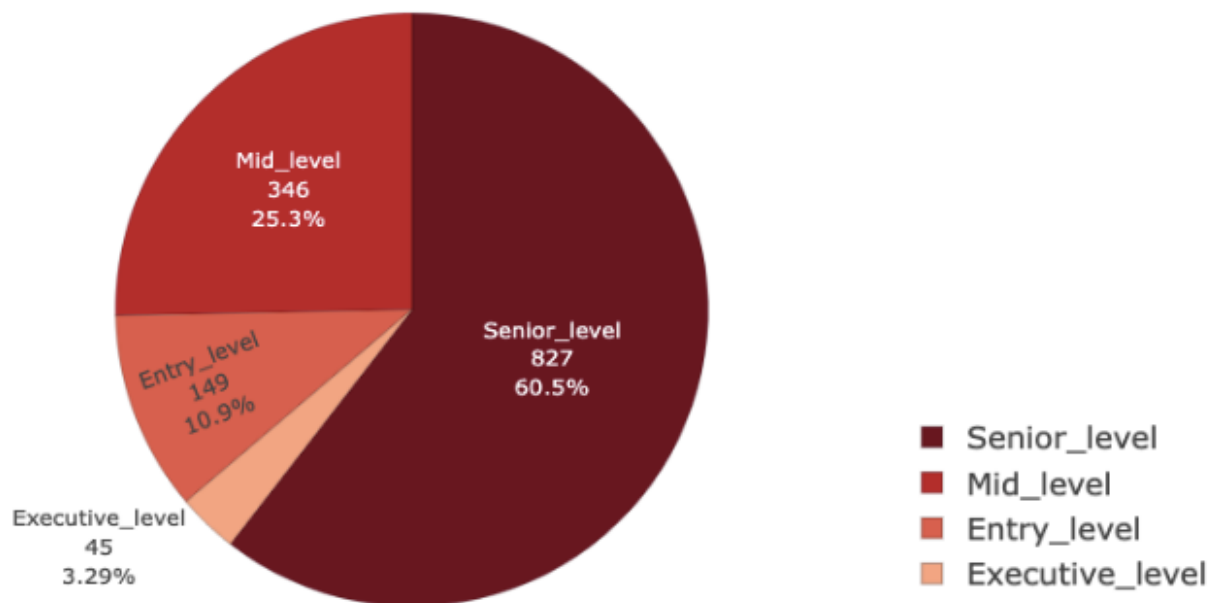
summarizing and accounting data, EDA may promptly reveal useful information, recognize trends, and reveal holistic links that can lead to additional study and potentialize its results. Moreover, figure 3 shows the descriptive statistic of the dataset adopted for the 4 potential attributes.

	work_year	salary_in_usd	remote_ratio
count	1367.000000	1367.000000	1367.000000
mean	2021.723482	123996.93782	63.862473
std	0.556787	66028.86410	45.313340
min	2020.000000	2324.000000	0.000000
25%	2022.000000	76895.50000	0.000000
50%	2022.000000	120221.00000	100.000000
75%	2022.000000	165000.00000	100.000000
max	2022.000000	600000.00000	100.000000

**Figure 3:** Descriptive statistics of the dataset.

## Data Visualizations and Discussion

- Experience Level distribution

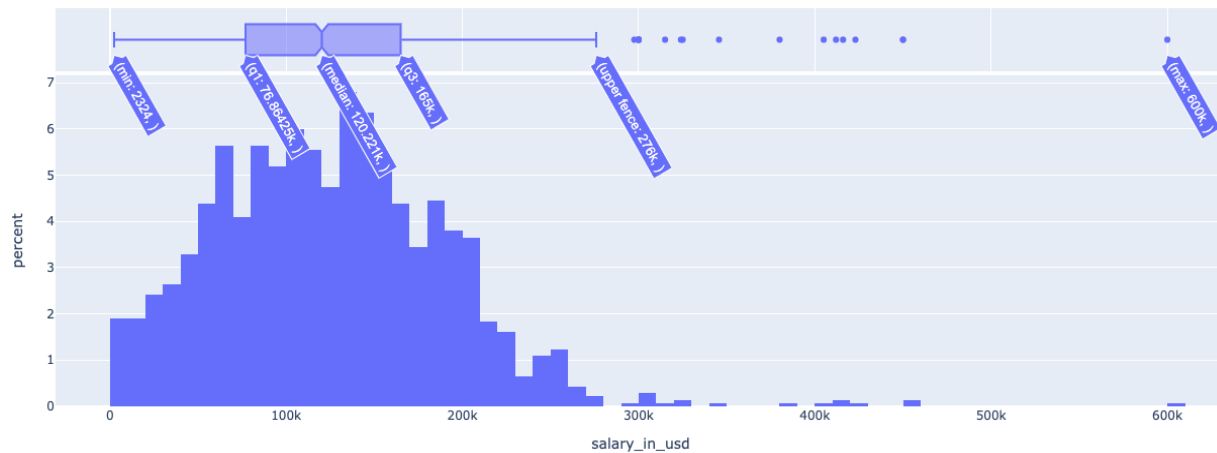


**Figure 4:** Experience Level Distribution

The above figure 4 shows the distribution of how the experience level is distributed across the given dataset. Furthermore, one surprising thing was that according to the company's traditional hierarchy, the percentage of senior-level employees (60.5%) was the highest compared to all other levels.

- Data Science related jobs and salary distribution

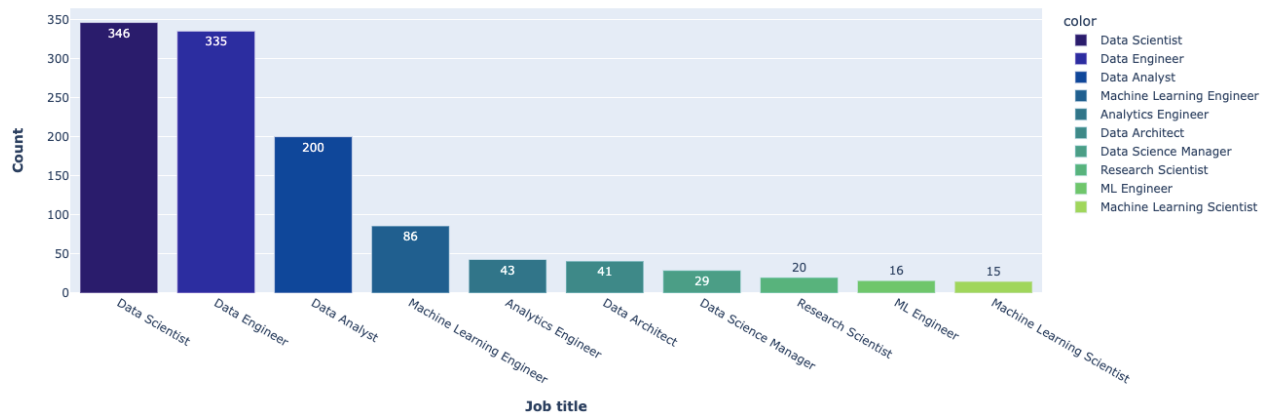




**Figure 5:** Data Science jobs' salary distribution

The above figure 5 shows the histogram of data science jobs' salary distribution irrespective of the seniority level of the job. However, excluding the outliers and just considering the bell curve only we can fairly say that the salaries are normally distributed, there might be several reasons for the distribution. One of them might be the seniority level of an individual, another might the scale of the company they are serving in, as big techs can provide higher monetary motivation to the employees. Moreover, as we can see, the median salary is ~\$120.22k which almost matches the past studies' predictions as aforementioned in the literature review.

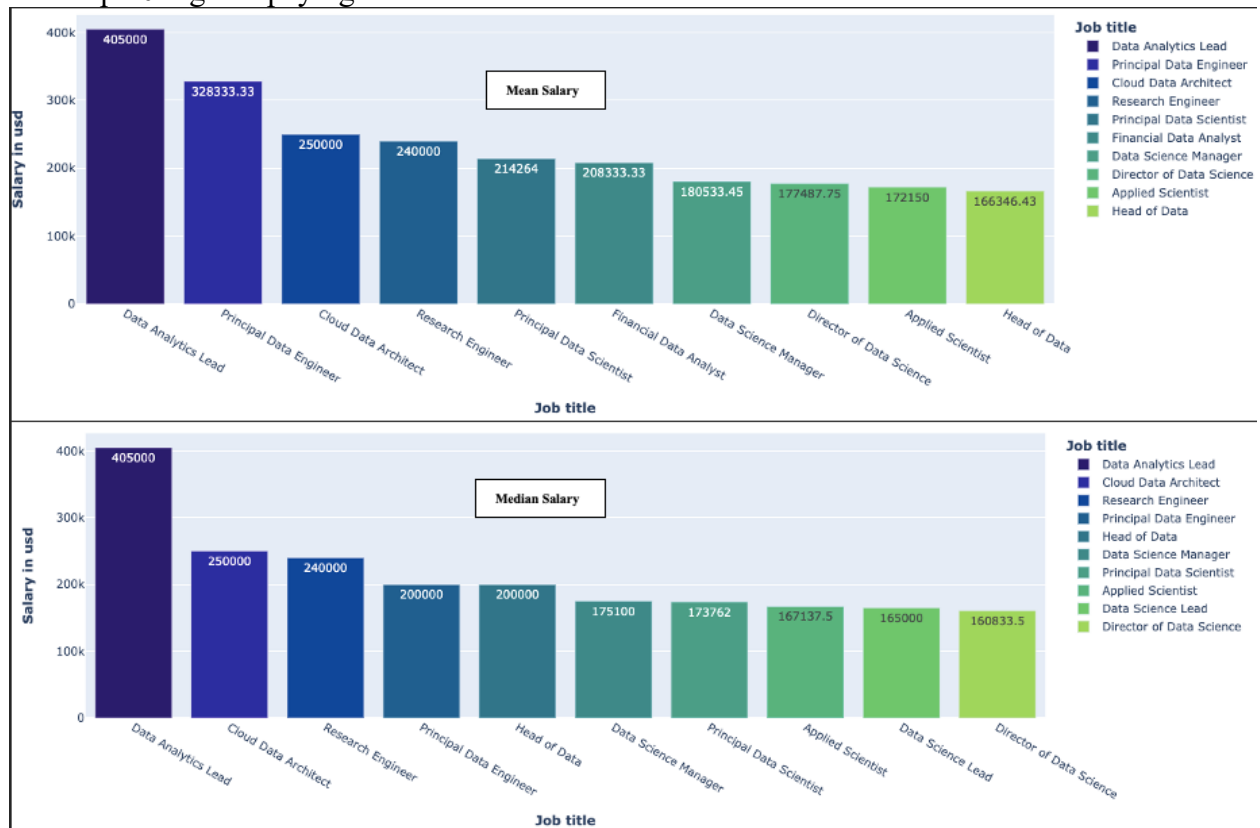
#### - Top 10 Data Science related job titles



**Figure 6:** Top 10 DS job titles

As shown in figure 6, we can see the most frequent job title with their count and observe that the role “Data Scientist” is amongst the top role in the analyzed dataset with a count of 346 people, additionally, there are 50 distinct roles in the dataset all of which are related to data science in some manner. There still exists a nuanced confusion considering the job titles in the field of “Data Science” due to which there is a higher chance of grouping distinct types of roles under one umbrella of “Data scientist” by the number of companies leading to a much higher count of the most generic title “Data scientist”.

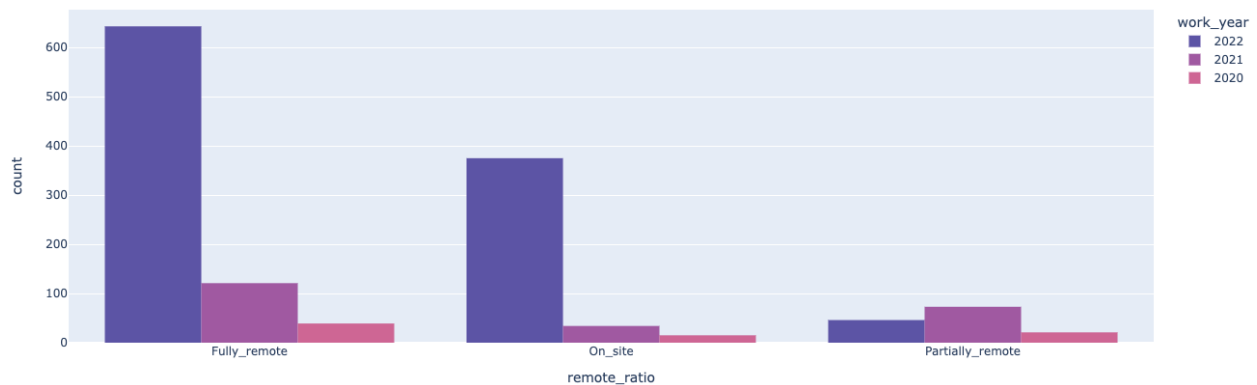
- Top 10 highest-paying Data science related roles



**Figure 7:** Top 10 Avg. and median highest paying roles in data science-related roles

As we can see from the above graph (figure 7) that the topmost data science-related title, "Data Analyst Lead," gets the highest mean and median pay of approximately \$405k irrespective of employees' location globally, but on checking the counts of "Data Analyst Lead" in the dataset it was rectified that there was only one instance of it. To best rectify the highest paying job role, it is best to compare the graphs and rectify the common ones.

- Growth of Data science related jobs across years



**Figure 8:** Growth of DS jobs (During COVID-19)

It is evident from bar graph 8 that we can see significant and consistent growth in data science-related jobs even during COVID-19 and lockdowns. Moreover, it is also apparent that there is a jump in the number of remote jobs during the COVID-19 year. Moreover, we can also observe that even after assuaging COVID-19 situations, there is a significant rise in remote jobs in 2022 indicating that more and more companies are adapting to the new work model as it might have resulted in better revenues and employees' better work-life balance.

#### - Data science job yearly growth

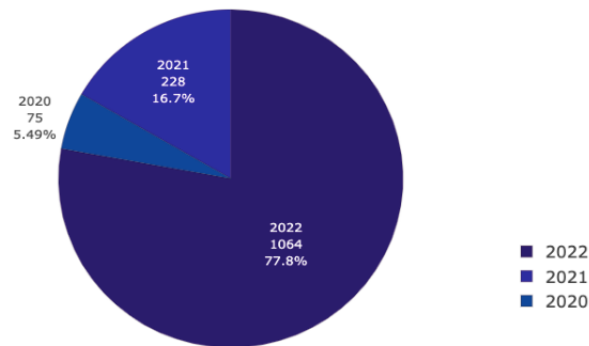


Figure 9: Year V/s employee percentage/count

Figure 9 shows the distribution of the data science-related job employees segregated annually. Moreover, we can see that, in line with figure 7, there is a significant and consistent growth in data science-related jobs, with 5.49% in 2020, 16.7% in 2021 i.e., a rise in jobs even during the times of COVID-19, and a whopping 77.8% in the year 2022 as more and more companies adapted the work from home culture.

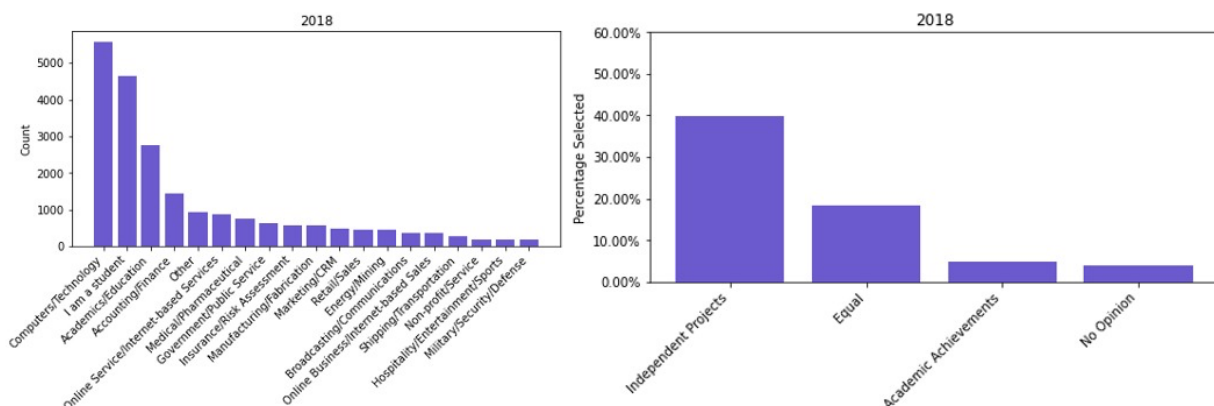


Figure 10: Employer's industry and factor of contributing expertise in data science

In figure 10 we can see that the major survey takers are IT companies, or students or Academicians out of which most of them have voted for either "Independent projects" or "Equal" (Meaning both

independent projects and academic achievements), hence it can be said that “independent projects” plays a significant role in gaining expertise in data science related roles as more than half survey takers votes for it.

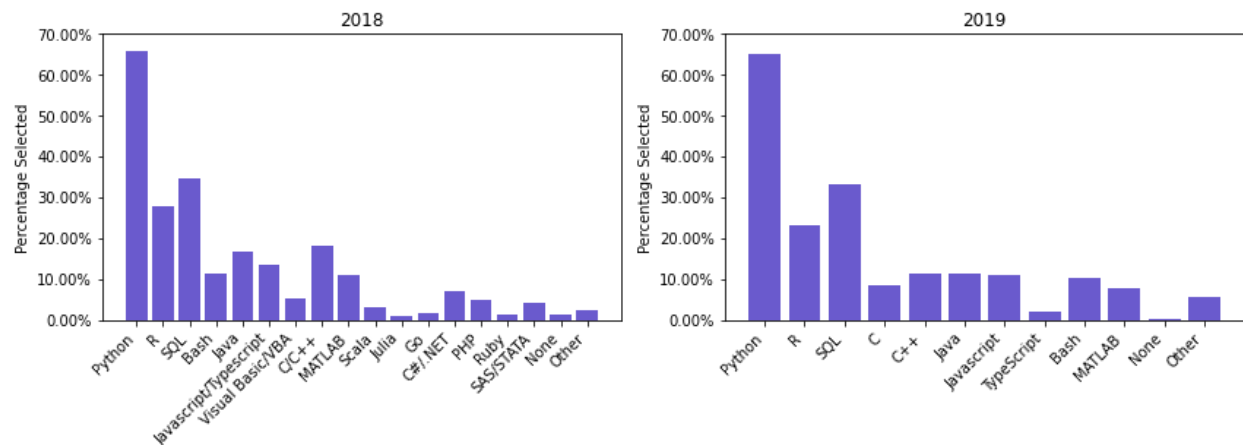


Figure 11: Most used programming language by Data science practitioners

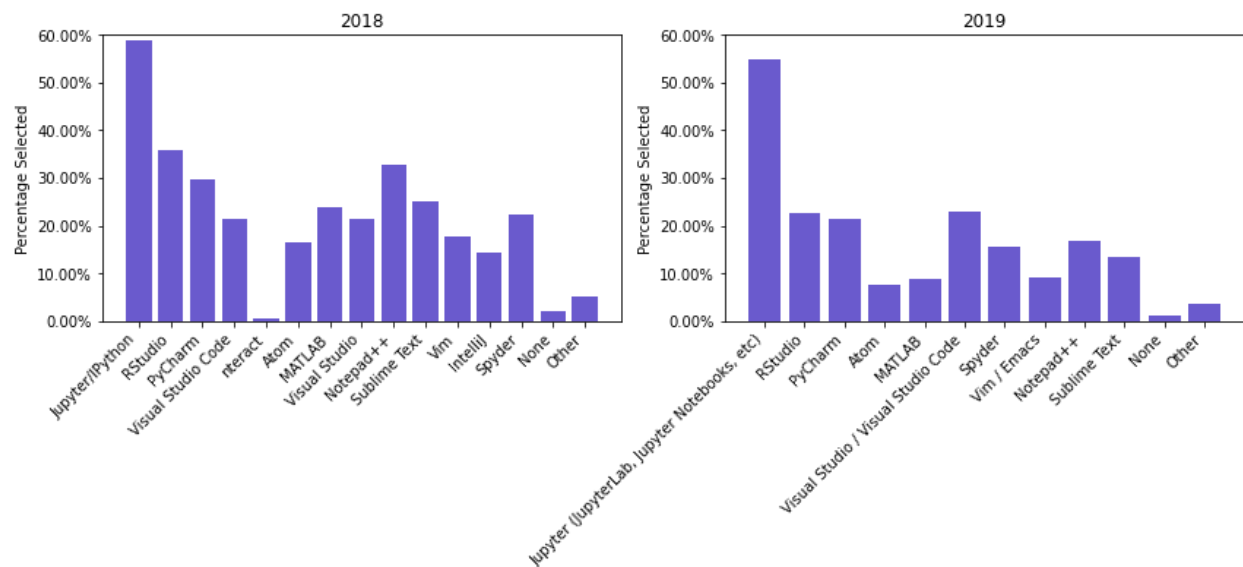


Figure 12: Most used IDEs by Data science practitioners

As shown in figure 12 it is evident that Jupyter is undoubtedly the most used IDE for python and data analysis-related tasks followed by RStudio (As R is also one of the most used languages for data analysis tasks, as shown in figure 11); but one thing to notice is that there is a significant drop in the usage of Notepad++ from the year 2018 to 2019 this is due to the rise in several python libraries to ease the work of data analysis. Moreover, Notepad is majorly considered the jack of all, but when it comes to heavy analysis and library support it cannot execute code from its IDE and does not correct the errors in code efficiently.

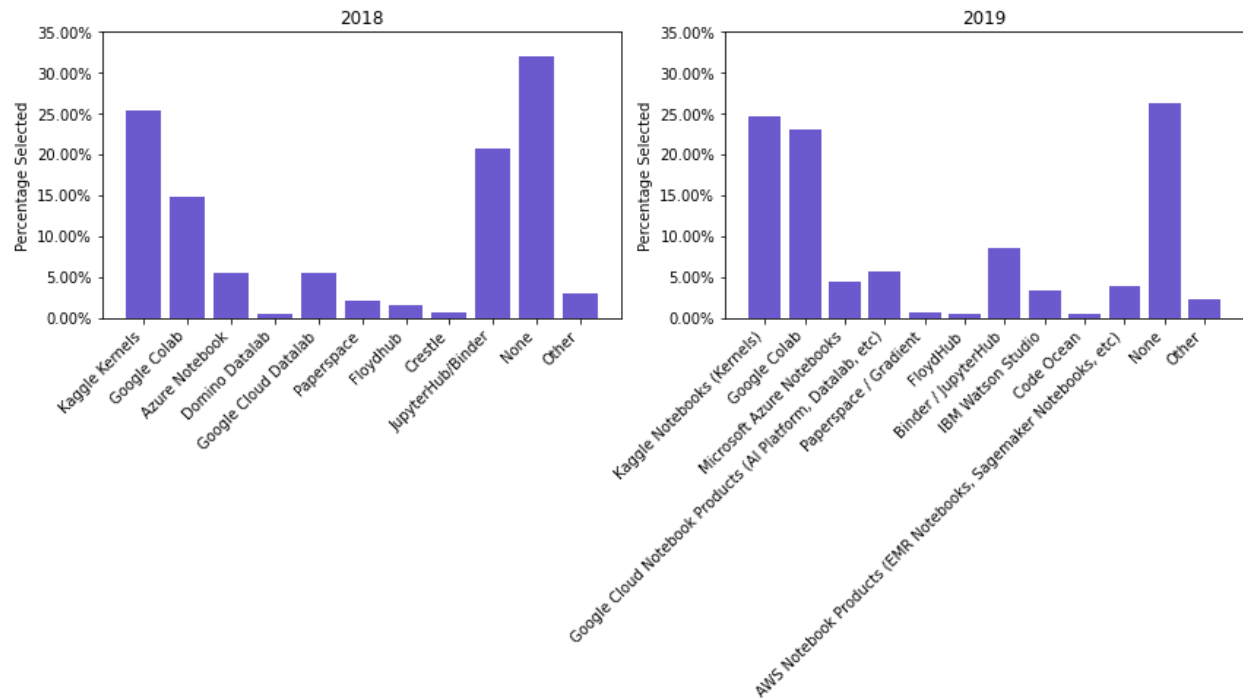


Figure 13: Most use notebook hosting services/products.

Figure 13 shows the distribution of the most used notebook hosting services, in which Kaggle stays consistently at the top (if ignoring the “None”) followed by JupyterHub/Binder for 2018 which significantly dropped its customer retention in 2019 when Google Colab came in, as Google colab gives its user the liberty to access the GPU (NVIDIA P100 or T4) with RAM size of 32 GB for the standard version and 52 GB RAM for the pro+ version.

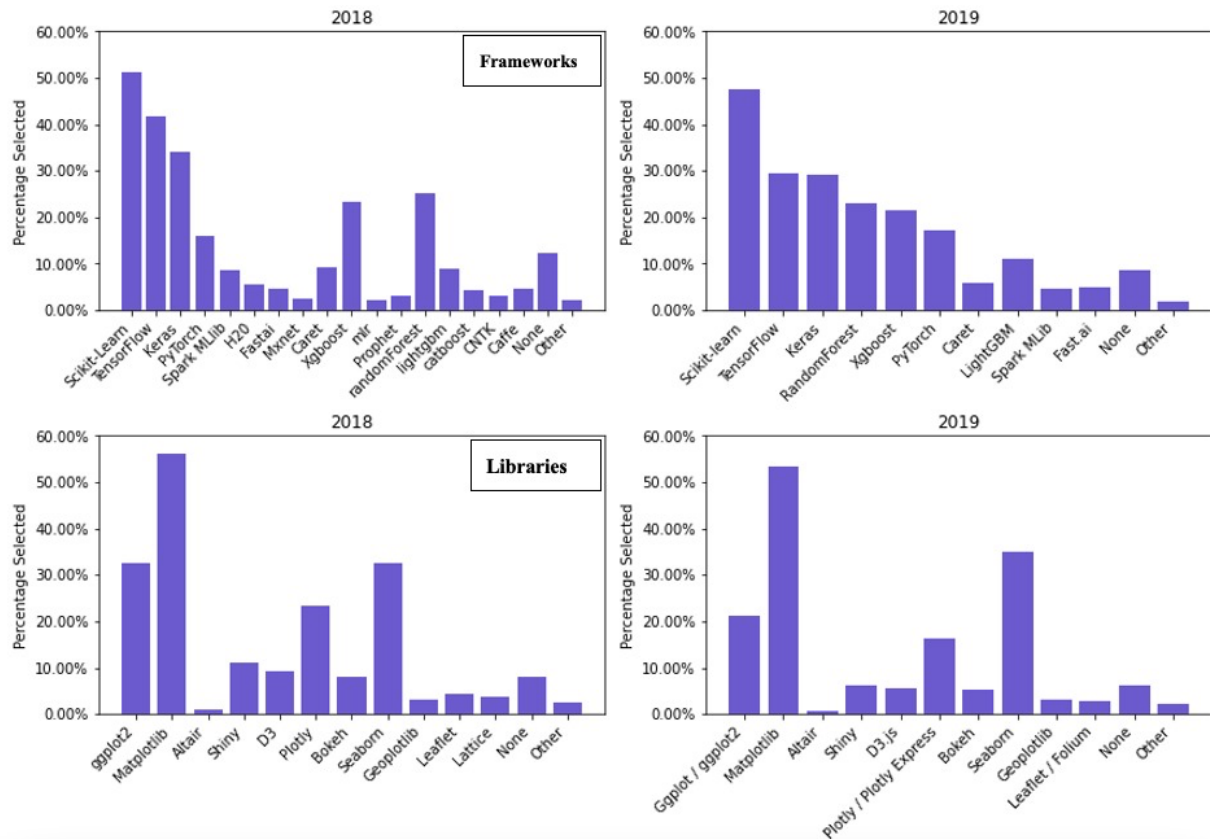


Figure 14: Most used Machine learning frameworks and visualization libraries.

From figure 14 we can see that the use of frameworks was quite sparse in 2018 as the spectrum of analysis possibilities was quite broad, but as time passed by people started using more and more efficient frameworks in the succeeding years resulting in a more fine-tuned number of frameworks used repeatedly to generalize the code for better understanding, readability, and analysis. Moreover, we can see a consistent and same trend for the usage visualization libraries (Matplotlib, seaborn, ggplot, and plotly) as these 4 majorly cover the required visualization for the analysis and communication.

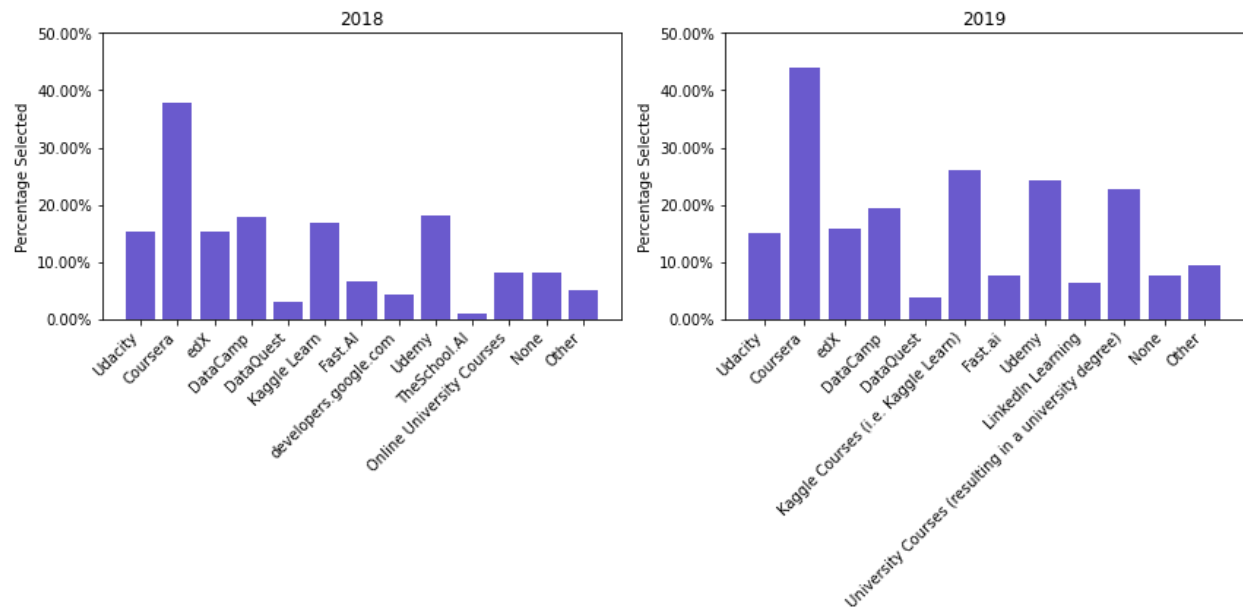


Figure 15: Most used learning platform by Data science and Machine learning practitioners

We can see from figure 15 that Coursera gained popularity in the succeeding year and was consistently the most used learning platform for data science enthusiasts, even though most of the well-curated courses are highly charged, this is possibly due to the ability to audit the whole course for free and gaining the required knowledge; additionally, Coursera also provides free access to the course to the learner with financial difficulties.

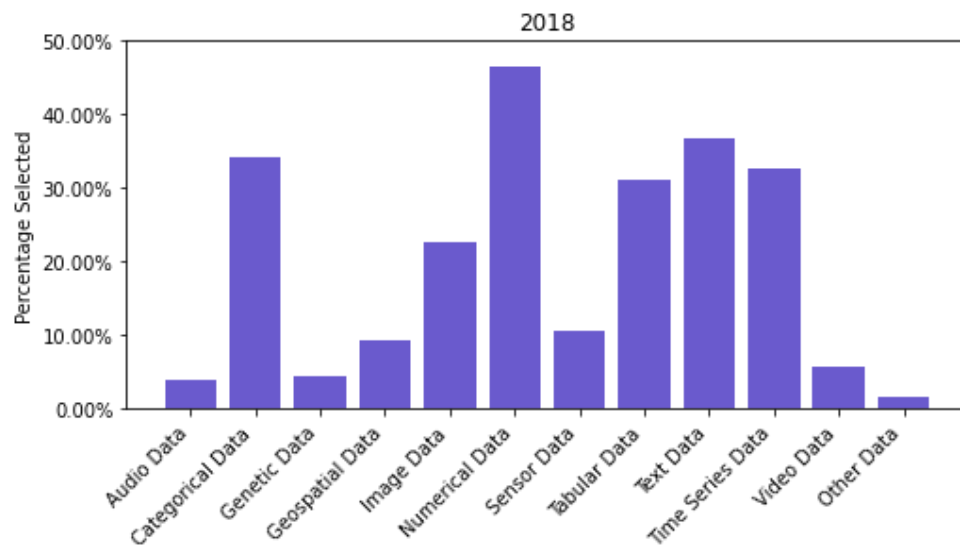


Figure 16: Most used types of data sets

Figure 16 shows the most used type of datasets by data science practitioners, where numerical data is the most used followed by text, graphical, and time series datasets. The reason major reason behind the usage of numerical and text datasets is their ease of availability to the general public and their compatibility with most statistical techniques and libraries.

## Initial Inferences

- The median Salary for data science-related jobs approximates around \$120.22k.
- It is also evident from the dataset that it is biased towards US-based employers and employees.
- From the initial EDA, we observe significant growth in data science-related jobs (Past three years).
- Parallel to the rise in data science-related jobs, there is also a significant rise in remote types of jobs.
- Based on the 2018 Kaggle Survey, about 40% of people in the industry believe that independent projects can demonstrate expertise better than academic achievement.
- Numerical, Categorical, text, and time series data are amongst the most used type of datasets by professionals.
- Coursera is consistently amongst the top learning platform for Data Science and Machine Learning users/learners.
- Python is the topmost used language followed by SQL and then R.

## Future scope

- Analyzing and designing the Data Science curriculum based on skills in demand.
- Analyzing the hiring companies' skills requirements as per the Job description.
- It's unlikely that a single program can inculcate all the skills, so breaking down the program into subsets of skillsets to further analyze.
- Data science-related online-course analysis and how they help data science practitioners.

## References:

- Ameya Kshirsagar, & Parth Sanghavi. (2022). Geothermal, oil and gas well subsurface temperature prediction employing machine learning. *47 Th Workshop on Geothermal Reservoir Engineering*, 1–11.  
<https://pangea.stanford.edu/ERE/db/GeoConf/papers/SGW/2022/Kshirsagar.pdf>
- Apoorva Komaraju. (2021). *Data Scientists, This is the History Behind your Data Science Jobs*. Analytics Insight. <https://www.analyticsinsight.net/data-scientists-this-is-the-history-behind-your-data-science-jobs/>
- Davenport, T. H., & Patil, D. J. (2012a). Data scientist: The sexiest job of the 21st century. *Harvard Business Review*, 90(10), 5. <https://hbr.org/2022/07/is-data-scientist-still-the-sexiest-job-of-the-21st-century>
- Davenport, T. H., & Patil, D. J. (2012b). Data scientist: The sexiest job of the 21st century. *Harvard Business Review*, 90(10), 5. <http://blogs.sun.ac.za/open-day/files/2022/03/Data-Scientist-Harvard-review.pdf>
- ImageSuggest. (2021). *What Is a Creative Commons Zero (CC0) License?*  
<https://imagesuggest.com/blog/what-is-a-creative-commons-zero-cc0-license/>
- Ismail, N. A., & Zainal Abidin, W. (2016). Data Scientist Skills. *IOSR Journal of Mobile Computing & Application*, 03(04), 52–61. <https://doi.org/10.9790/0050-03045261>
- Louis Columbus. (2017, May). *IBM Predicts Demand For Data Scientists Will Soar 28% By*



2020. Forbes. <https://www.forbes.com/sites/louiscolombus/2017/05/13/ibm-predicts-demand-for-data-scientists-will-soar-28-by-2020/?sh=68f6e1787e3b>
- Pryor, G., & Donnelly, M. (2009). Skilling Up to Do Data: Whose Role, Whose Responsibility, Whose Career? *International Journal of Digital Curation*, 4(2), 158–170. <https://doi.org/10.2218/ijdc.v4i2.105>
- R. Geiger, Charlotte Mazel-Cabasse, Chihoko Cullens, Laura Norén, Brittany Fiore-Gartland, Diya Das, & Henry Brady. (2018). Career Paths and Prospects in Academic Data Science: Report of the Moore-Sloan Data Science Environments Survey. *SocArXiv*. <https://doi.org/10.31235/osf.io/xe823>
- Rettberg, J. W. (2020). Situated data analysis: a new method for analysing encoded power relationships in social media platforms and apps. *Humanities and Social Sciences Communications*, 7(1), 1–13. <https://doi.org/10.1057/s41599-020-0495-3>
- Saltz, J. S., & Dewar, N. (2019). Data science ethical considerations: a systematic literature review and proposed project framework. *Ethics and Information Technology*, 21(3), 197–208. <https://doi.org/10.1007/s10676-019-09502-5>
- Smaldone, F., Ippolito, A., Lagger, J., & Pellicano, M. (2022). Employability skills: Profiling data scientists in the digital labour market. *European Management Journal*, 40(5), 671–684. <https://doi.org/10.1016/j.emj.2022.05.005>