

Insights and Trends in Data Science: 2017-2021 Kaggle Survey Analysis

March 23, 2024
Created by: Ya Min Thu

Agenda

- Introduction to Kaggle Data Science Surveys
- Key Findings Summary
- Growth of Data Science Over Time
- Demographic Changes in Data Science
- Emerging Trends in Data Science
- Popular Tools and Technologies
- Future Predictions for Data Science
- Conclusion and Recommendations

Introduction to Kaggle Data Science Surveys

Overview of Kaggle Surveys

- The Kaggle data science surveys from 2017 to 2021 provide valuable insights into the trends and developments in the data science field.
- These surveys offer a comprehensive view of the tools, technologies, and demographics shaping the data science community over the years.
- Analyzing the survey data can help businesses and professionals stay informed about the latest trends and make data-driven decisions.
- The surveys play a crucial role in understanding the evolving landscape of data science and anticipating future directions in the field.

The Kaggle logo, consisting of the word "kaggle" in a lowercase, blue, sans-serif font.

Key Findings Summary

01

Analysis of demographic changes within the data science community.

02

Insights into the growth and popularity of data science from 2017 to 2021.

03

Identification of trends in tools and technologies used in data science over the years.

04

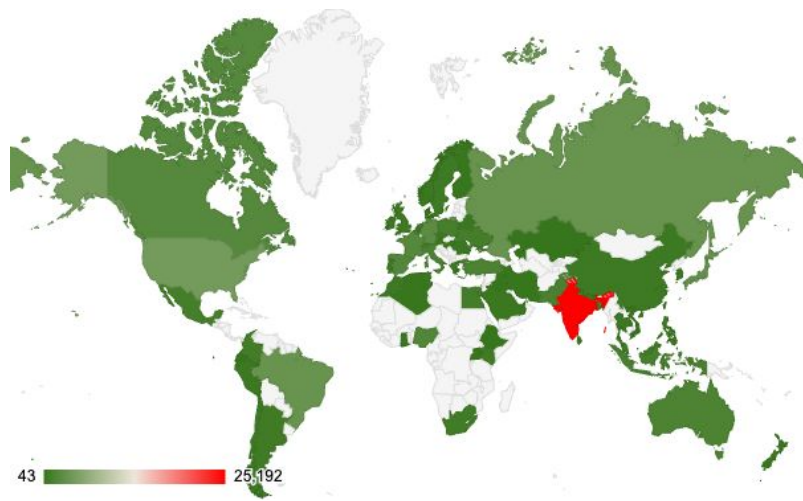
Understanding the evolution of data science based on the survey data.

01

Analysis of demographic changes within the data science community.

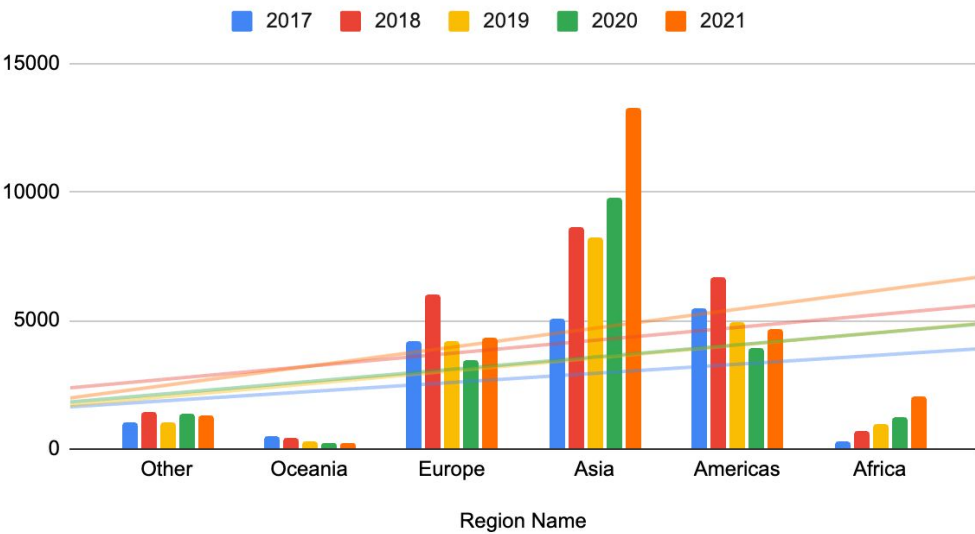
Growth of Data Science Over Time

Number of Respondents to Kaggle Data Science Surveys (2017-2021)



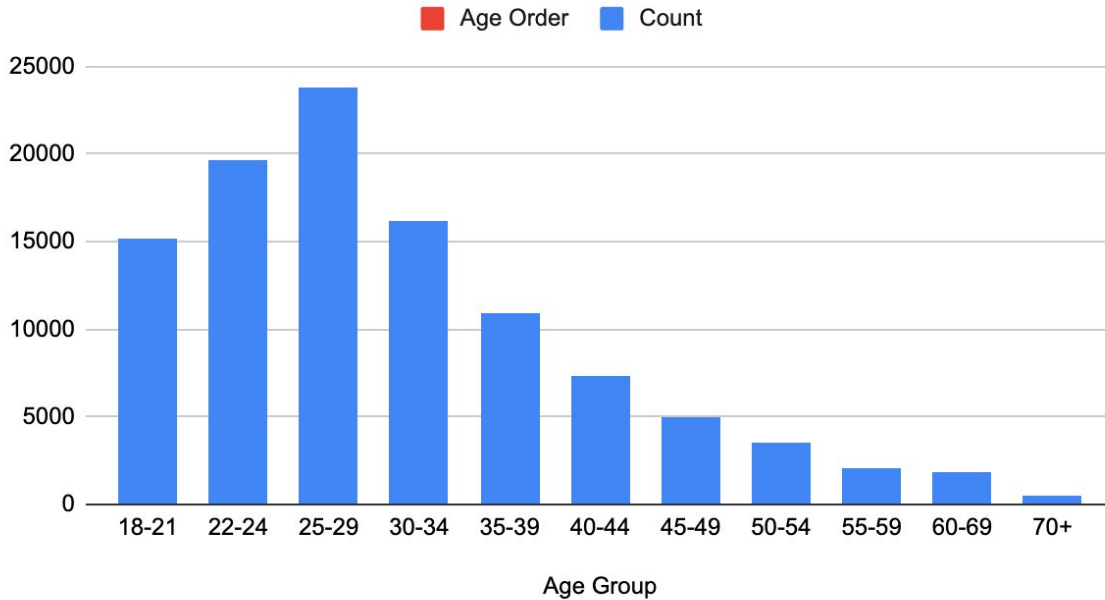
The chart illustrates a significant increase in the number of survey respondents year over year, indicating a growing interest and engagement in the field of data science.

Distribution of Survey Respondents from Various Regions



Demographic Changes in Data Science

Distribution of Age Group (Sorted)

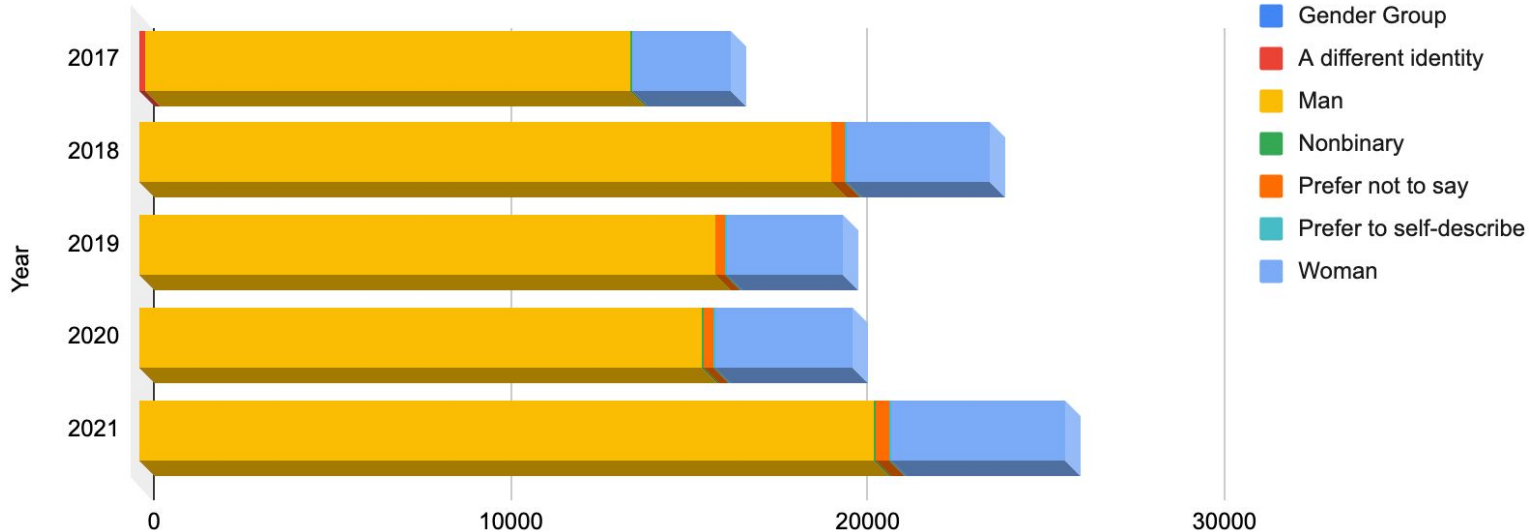


Age Distribution

- 106300 rows of data with 293 columns
- Most popular age groups are 25-29, 22-24 and 30-34.
- Age ranges from 18 to 70+

Increased Diversity in Gender Groups

Gender Distributions Over The Years



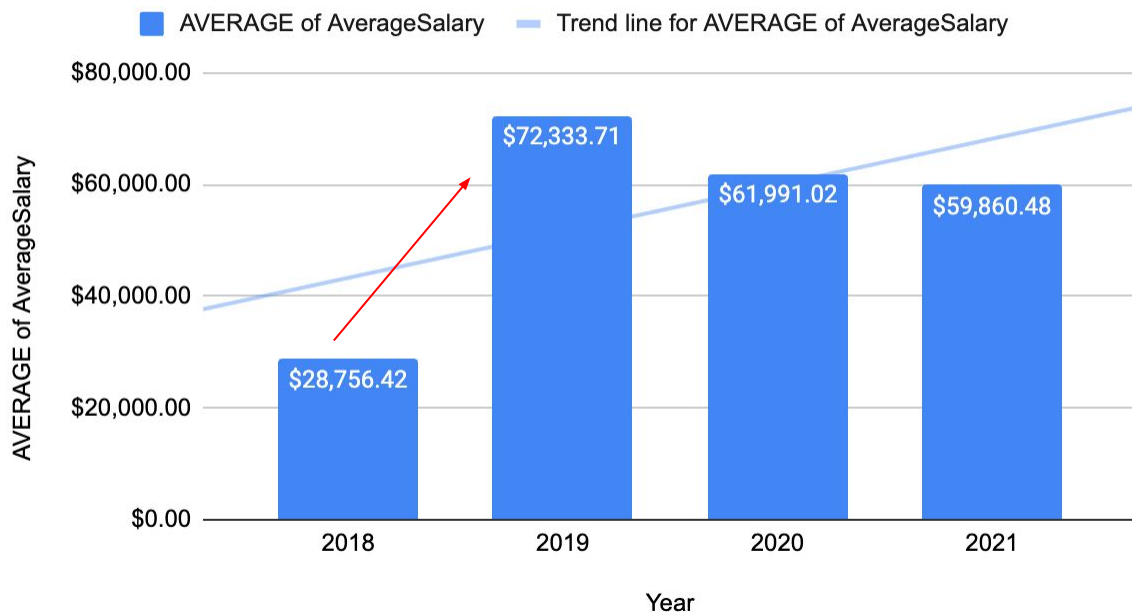
- **2017** : Consisted of majority Male followed by female
- **2021**: Female percentage grew from 2017, new groups introduced

02

Insights into the growth and popularity of data science from 2017 to 2021.

Average Yearly Salary Trends Over The Year

Cumulative Average Yearly Salaries(USD)



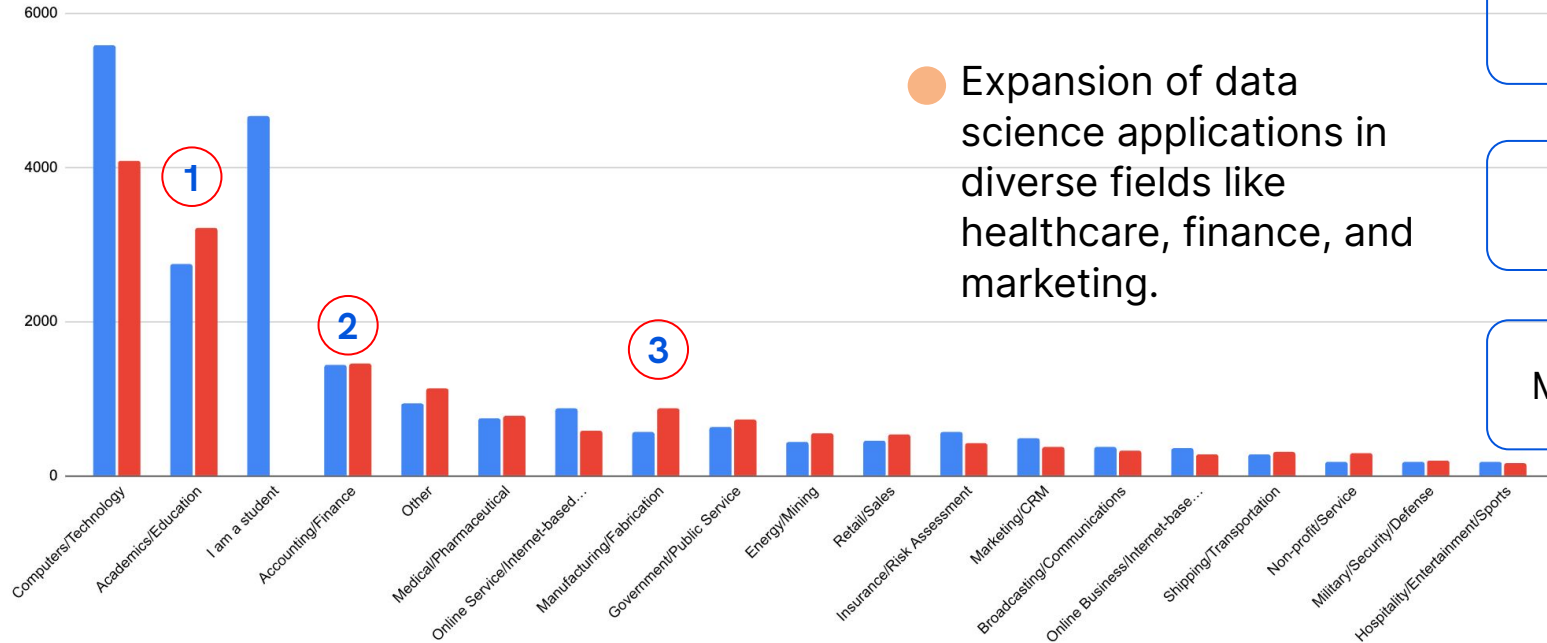
- Average salary increased from 2018-2019 for **60.24%**
- Likely due to global occurrences at that time that impacted the data science jobs
 - Eg. the global pandemic and created job opportunities in healthcare

Increasing Number of Industries Adopting Data Science

Valuable Domain Knowledge

Industries Applying Data Science Over the Year

■ 2018 ■ 2021



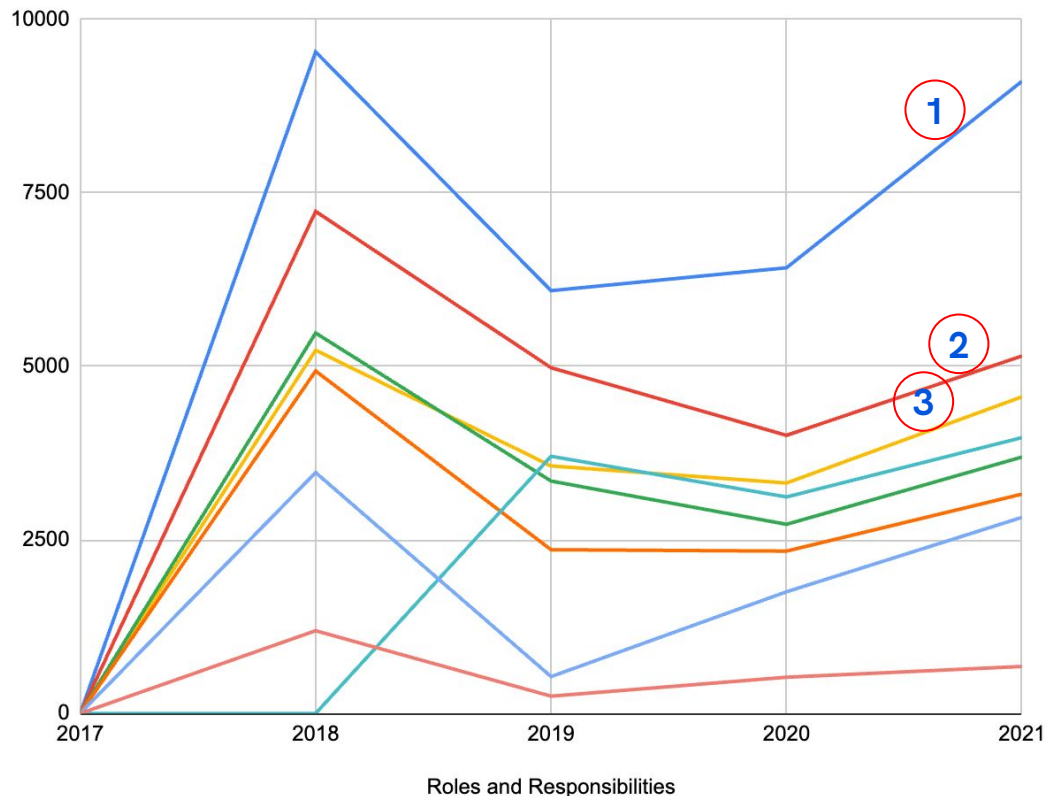
Academics

Finance

Manufacturing

Highly Important Tasks At Work Over The Years

Most Important Roles and Responsibilities Over the Years



Valuable Skill Sets

Data Analytics

Machine Learning

Data Infrastructure

03

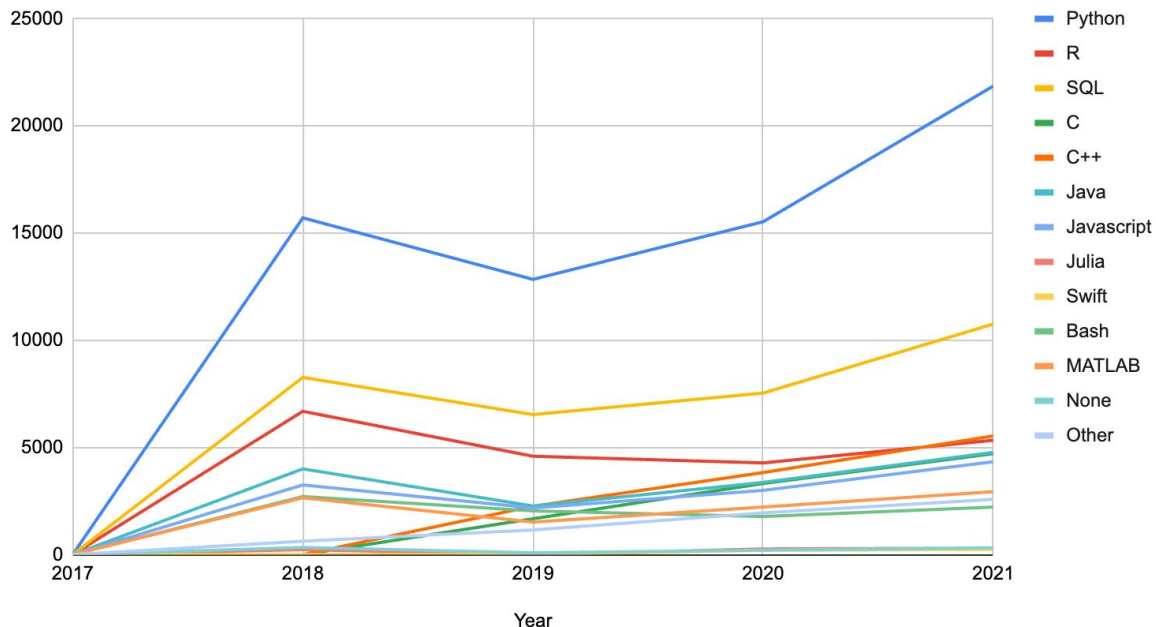
Identification of trends in tools and technologies used in data science over the years.

Exponential Growth Of Programming Language usages



Advancements in Python, R, and SQL Adoption

Growth Of Usage of Programming Languages

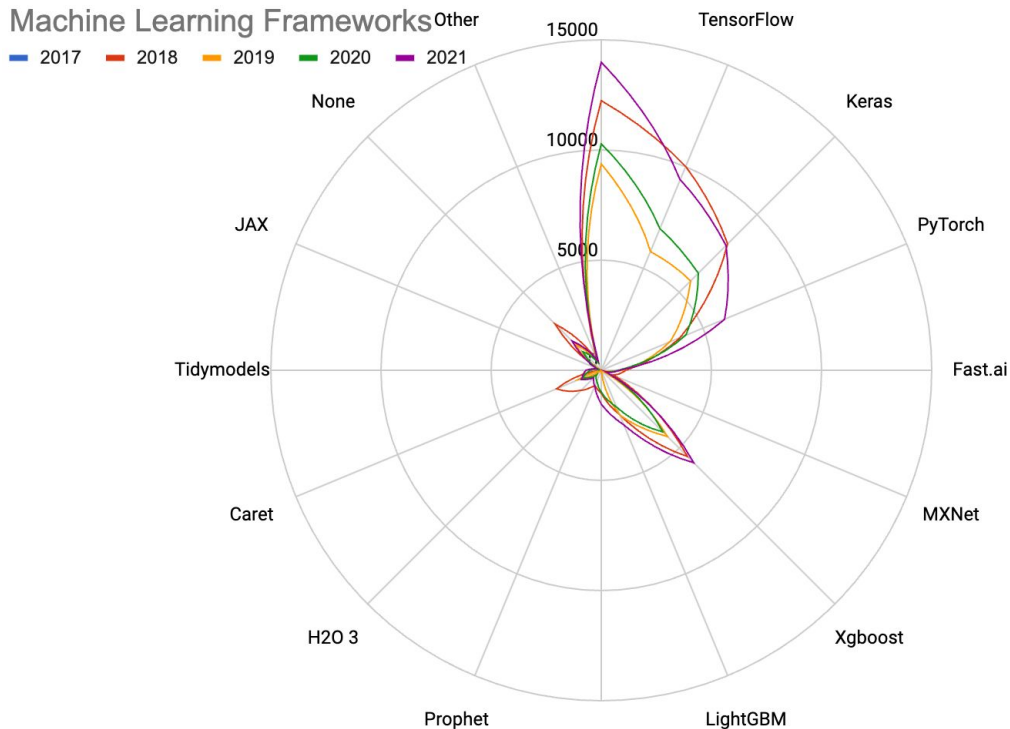


Python, R, and SQL have shown exponential growth in adoption for data analysis and machine learning tasks, indicating a shift towards these tools for advanced analytics.

Emerging Technologies in Machine learning



PyTorch, Scikit-learn, Keras, TensorFlow

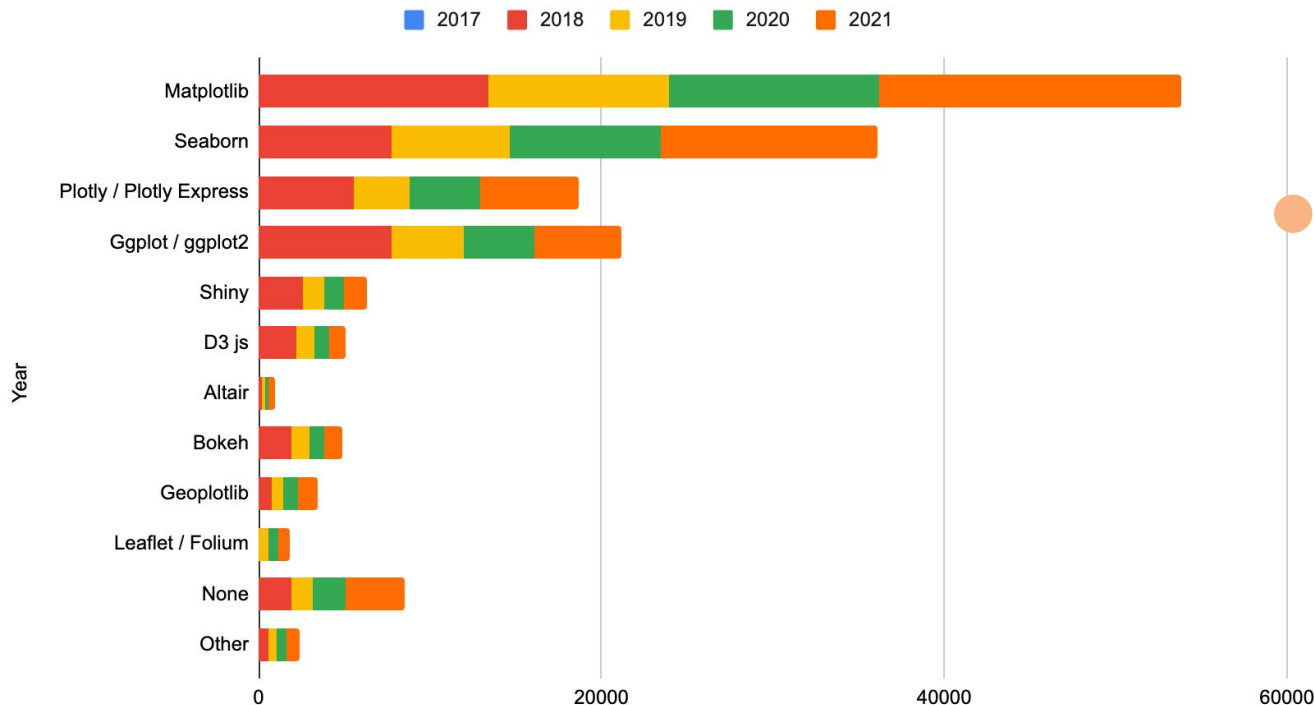


Increased adoption of machine learning techniques for data analysis and prediction.

Increased Popularity of Visualisation Tools



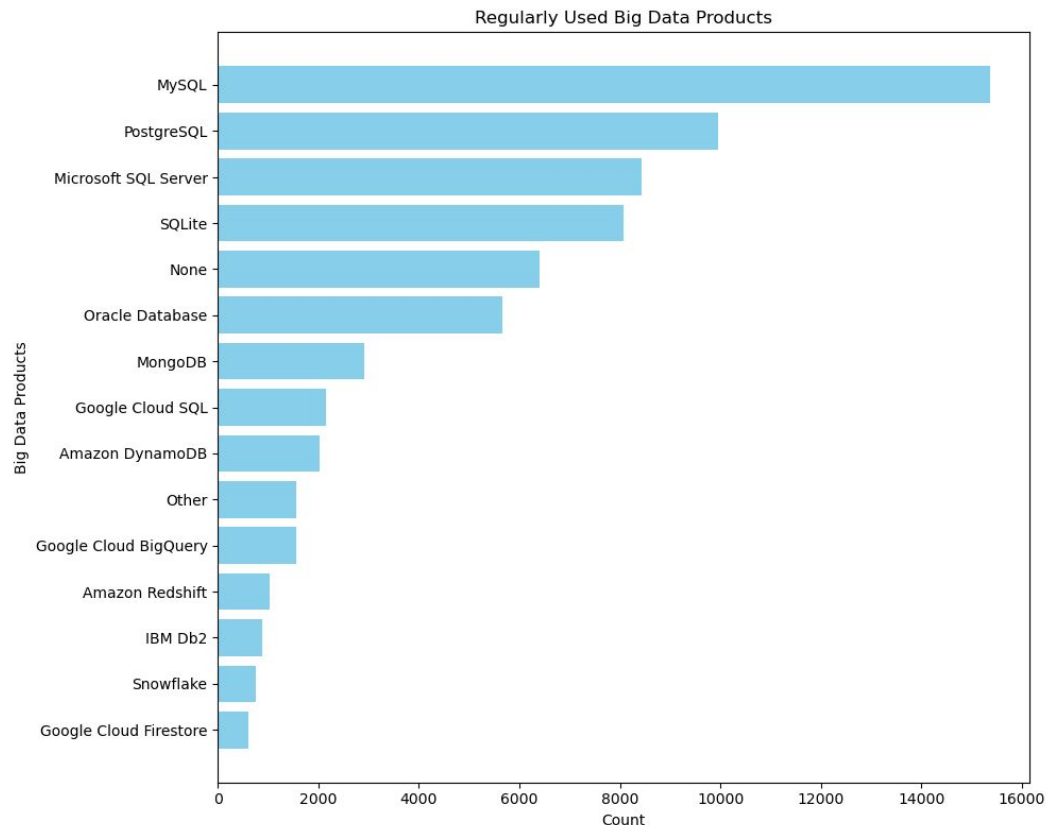
Matplotlib, Seaborn, Plotly



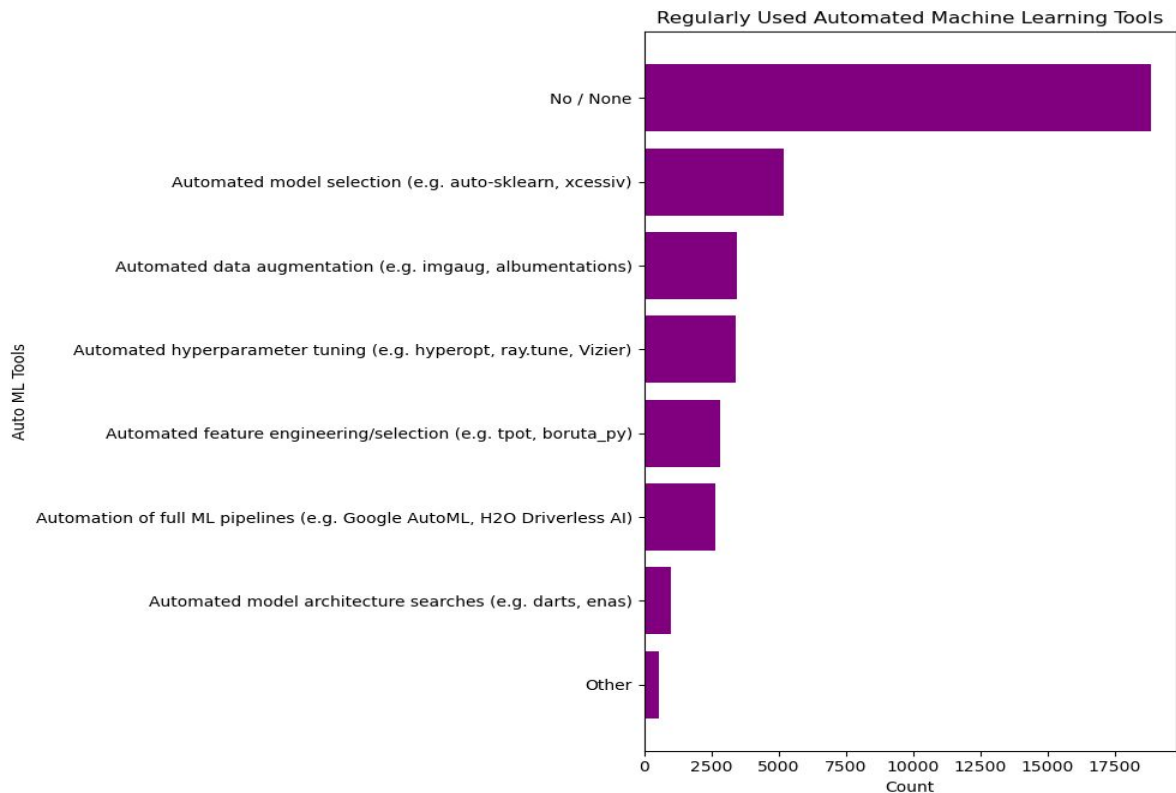
Visualization tools like Tableau, Matplotlib, and Seaborn simplify complex data, aiding comprehension and informed decision-making by presenting insights visually, enhancing patterns' recognition and facilitating communication.

Need for scalability with Cloud

- Rapid integration of cloud computing solutions in data science workflows for scalability and flexibility.



Increased Use of Automated Machine Learning Tools



Though majority is not in use, there has been a shift towards automated machine learning (AutoML) tools to streamline model development and deployment.

04

Understanding the evolution of data science based on the survey data.

Challenges Faced by Data Scientists

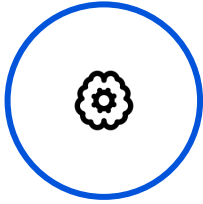
Technical Challenges

- Managing and analyzing large volumes of data efficiently.
- Keeping up with rapidly evolving technologies and tools in data science.
- Ensuring data quality and accuracy for reliable analysis.

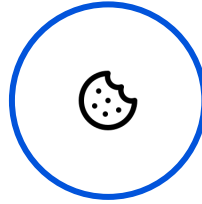
Career-related Challenges

- Limited access to quality datasets for analysis.
- Balancing technical expertise with business acumen and domain knowledge.
- Navigating job market competitiveness and salary expectations in the data science field.

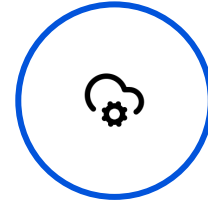
Future Predictions for Data Science



Rapid advancement of the AI field



More automation of repetitive actions in the future with the help of AI



Important to continuously keep up to date with advancements in AI to leverage it for each industry and organisation goal

Conclusion and Recommendations

Key Insights

Identify the top trends like the growth of machine learning, the impact of AI ethics, and the increasing adoption of cloud computing in data science.

Business Recommendations

Encourage businesses to invest in upskilling their workforce in data science, prioritize data ethics and privacy, and leverage cloud-based solutions for data analysis.

Professional Advice

Advise data professionals to stay updated on emerging technologies, continuously enhance their skills, and network within the data science community for career growth.

Future Outlook

Highlight the importance of preparing for the future by embracing automation, integrating AI responsibly, and fostering a culture of data-driven decision-making.

References

- Olteanu, Andrada. Kaggle Data Science Survey 2017-2021. Kaggle, 2021.
<https://www.kaggle.com/datasets/andradaolteanu/kaggle-data-science-survey-20172021>