

# IBM Data Analyst Capstone Project

---



## 2019 Stack Overflow Developer Survey Analysis

Erdoğan Taşel

25/05/2023

# OUTLINE

---



- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

# EXECUTIVE SUMMARY

---



- Stack Overflow developer survey data in 2019 has been analyzed at this project.
- "Annual Developer Survey examines all aspects of the developer experience from learning to code to their favorite technologies to version control and the workplace experience of professional developers." <https://insights.stackoverflow.com/survey/>
- Data is analyzed for
  - Current situation of popular languages, databases by Exploratory Data Analysis
  - Future Situation of languages and databases popularity
  - Demographic Analysis
  - Additional queries have been executed
- General findings are
  - Web development still has a higher trend because JavaScript and HTML/CSS are still preferable.
  - Python, SQL are the popular languages
  - MySQL has higher user but PostgreSQL might be next year.
  - Gender gap is high. There is no equality.

# INTRODUCTION

---



- The nature of the analysis is to gather the data among developers and examine the characteristics of them.
- The questions are
  - What is the current technology usage?
  - What is the future technology trends?
  - What is the Demographic examination of the respondents?
  - Are there any relationships among some data?
- What is the composition of data?
- What is the comparison of data?

# METHODOLOGY

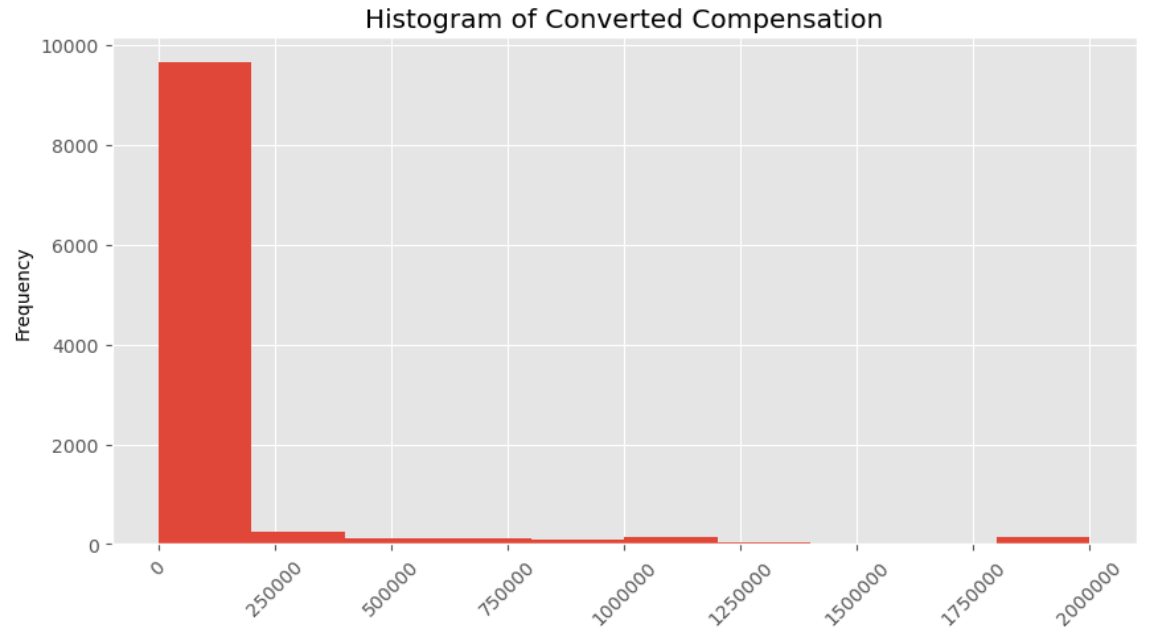
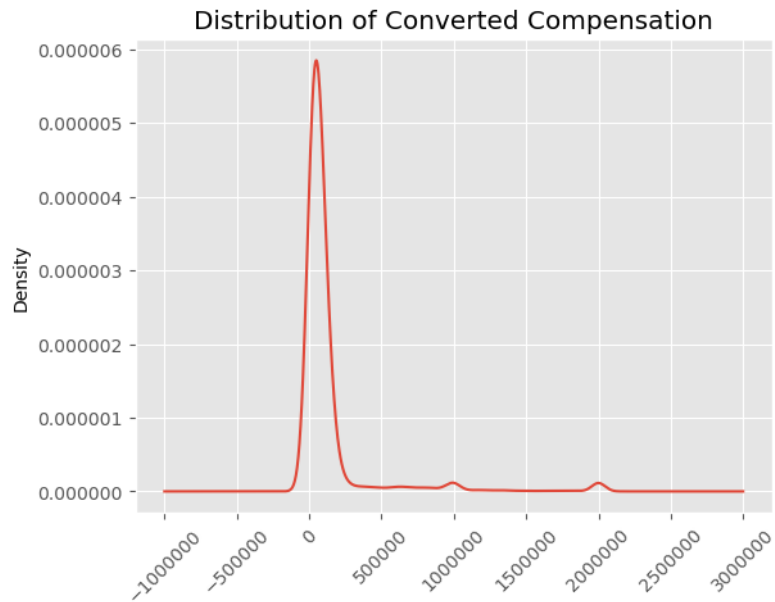
---



- For this analysis, subset of Stack Overflow Annual Developer Survey has been used and it was provided by IBM,
- The survey was fielded from January 23 to February 14, 2019,
- The column `ConvertedComp` contains Salary converted to annual USD salaries using the exchange rate on 2019-02-01. This assumes 12 working months and 50 working weeks.
- Data was loaded and cleaned by using Pandas,
- Data Wrangling and Exploratory Analysis have been performed by using Pandas library in Jupyter Lab,
- Data Visualization has been performed by executing SQL queries and Python codes,
- Data is analyzed and visualized by IBM Cognos for
  - Current situation of popular languages, databases by Exploratory Data Analysis
  - Future Situation of languages and databases popularity
  - Demographic Analysis (gender, etc.)
- Additional queries have been executed to find answers of some questions like number of users that only works for MySQL Database.

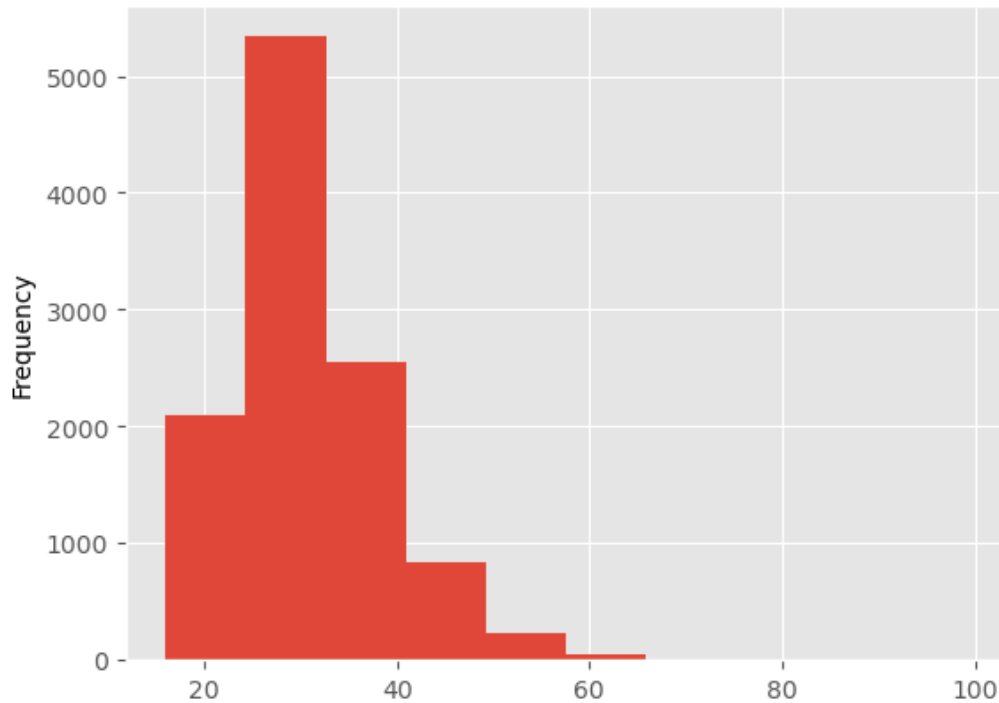
# RESULTS

- First descriptive statistics has been performed to describe the distribution of the Converted Compensation
- Sample size is 11398 and after removing outliers the sample size became 10519. It has 85 attributes.

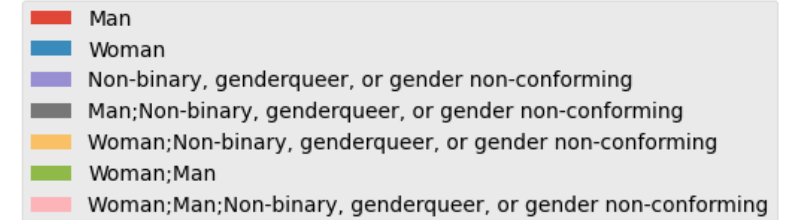
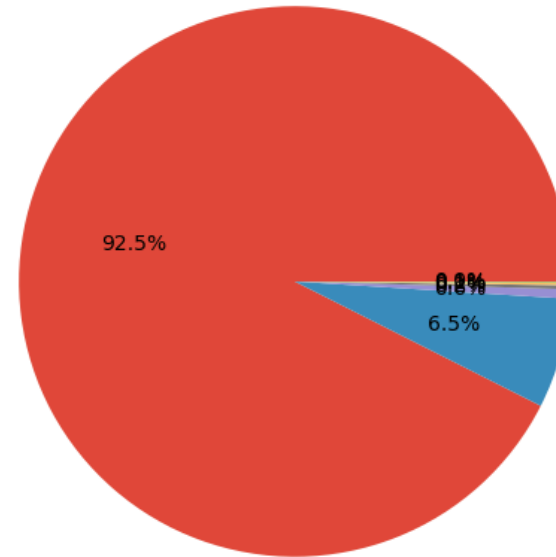


# RESULTS

Age Histogram

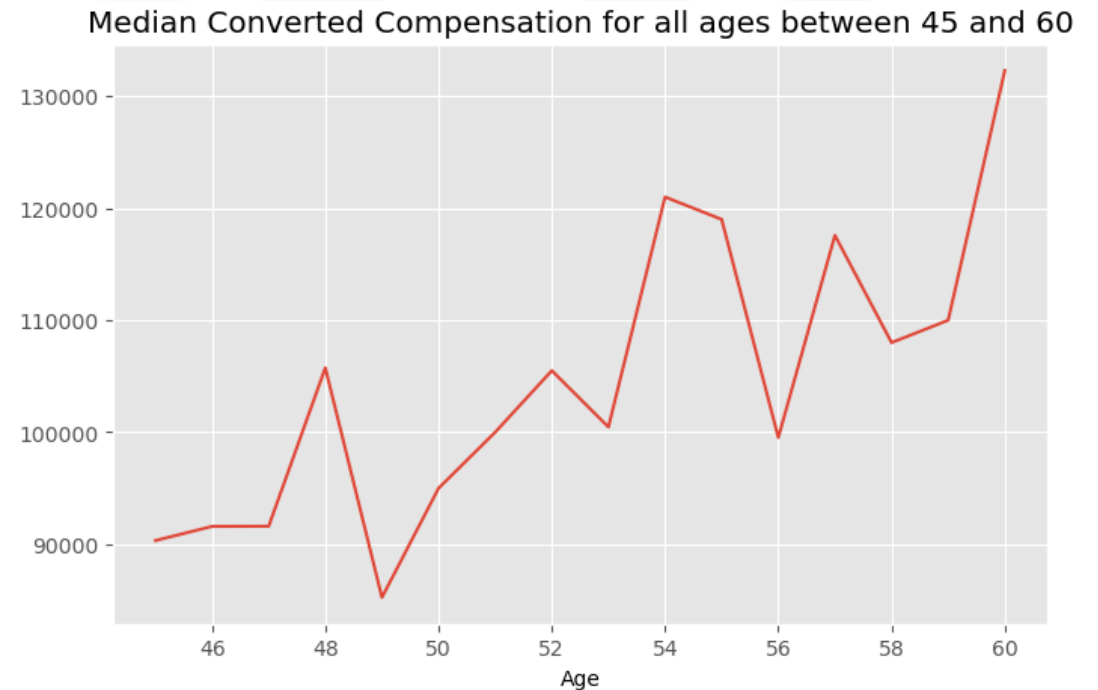
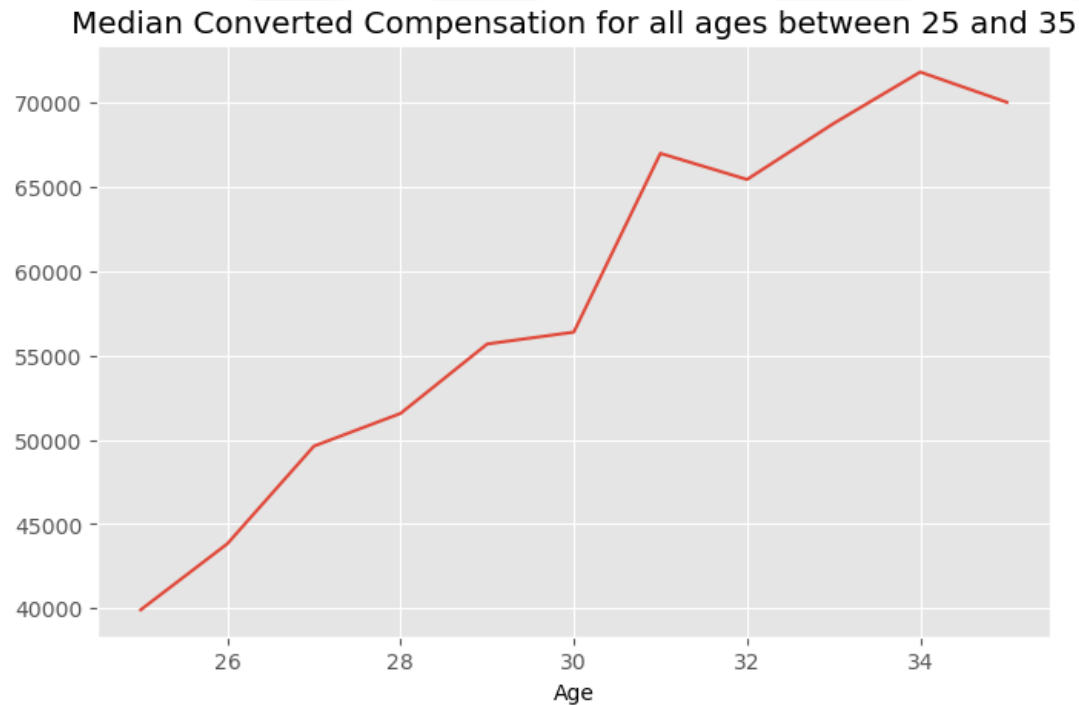


- Most of the respondents are between 25 to 35 years old.



- The mean age of the respondents is 30.8
- The median age is 29
- The respondents are mostly male, 92.5%

# RESULTS:Converted Compensation Trend

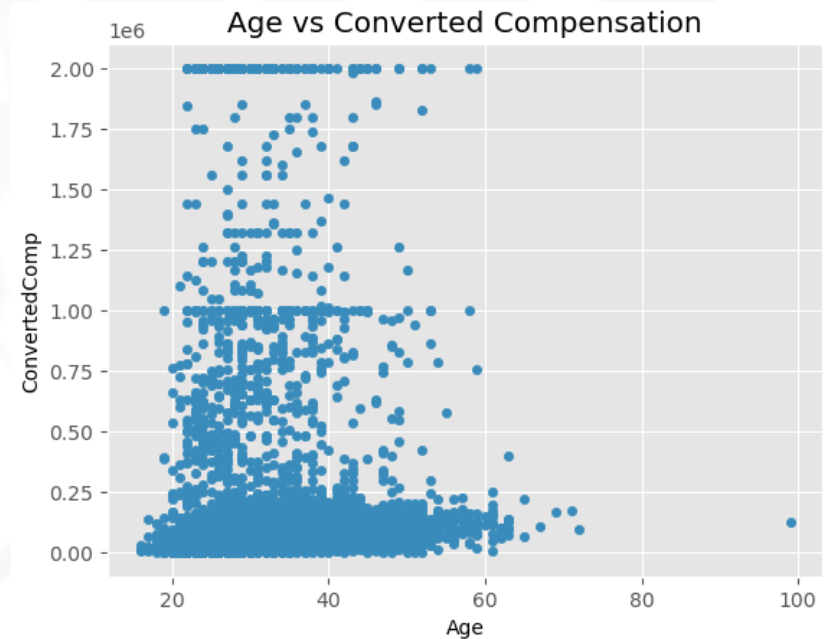
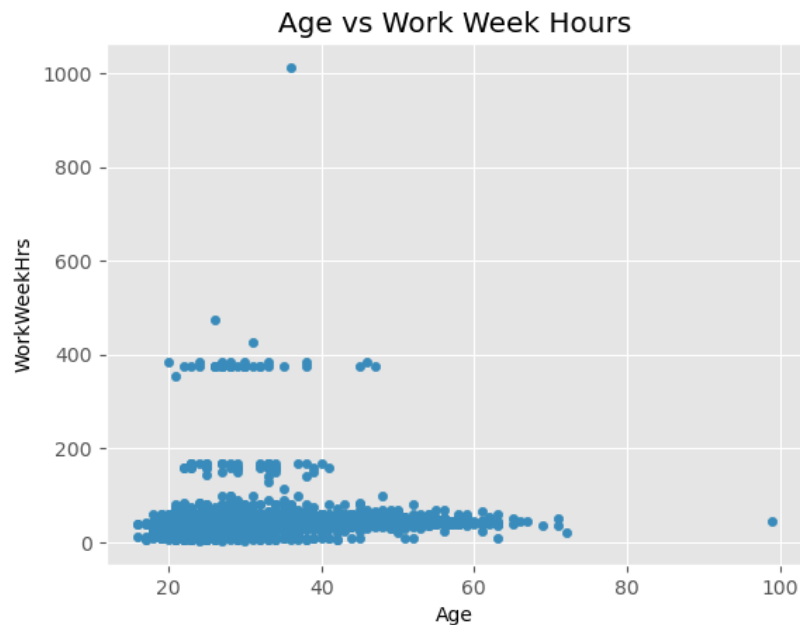


Compensation is increasing highly between 25 and 35 years old people. Between 45 and 60, Compensation increase is changing however the trend is getting higher.



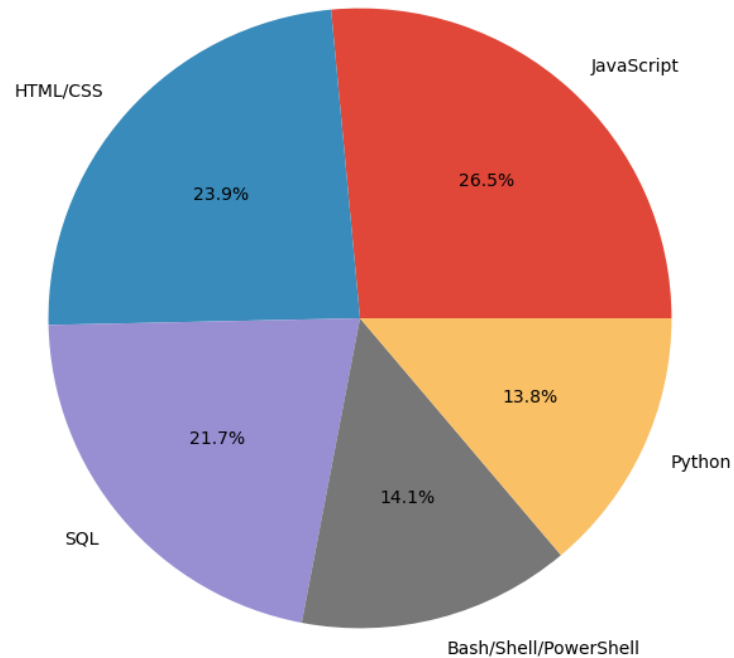
# CORRELATION

- There is no distinct correlation between Age and Work week hours.  $R = 0.037$
- Age and Converted Compensation are positively correlated.  $R = 0.40$

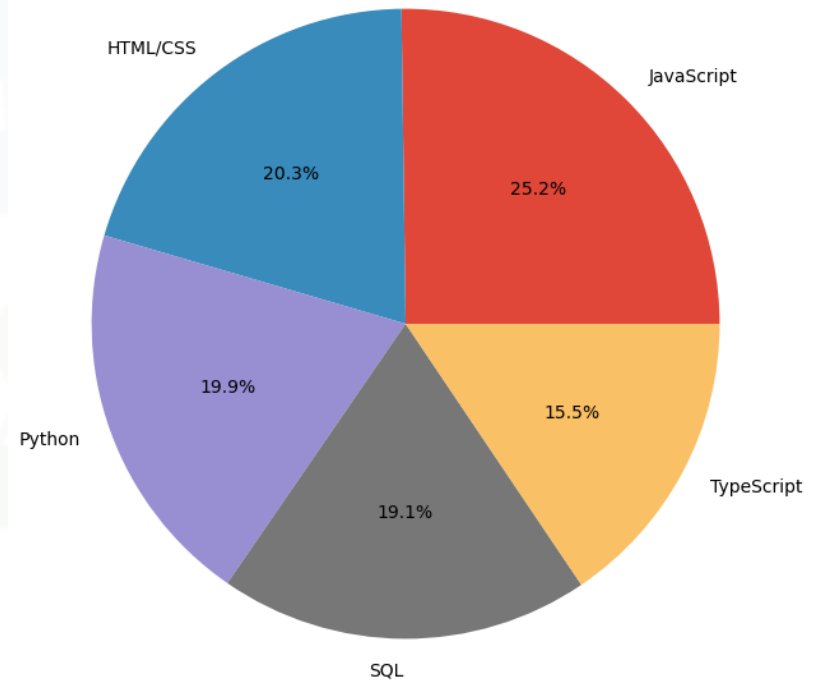


# PROGRAMMING LANGUAGE TRENDS

Top 5 Languages that Respondents Work With



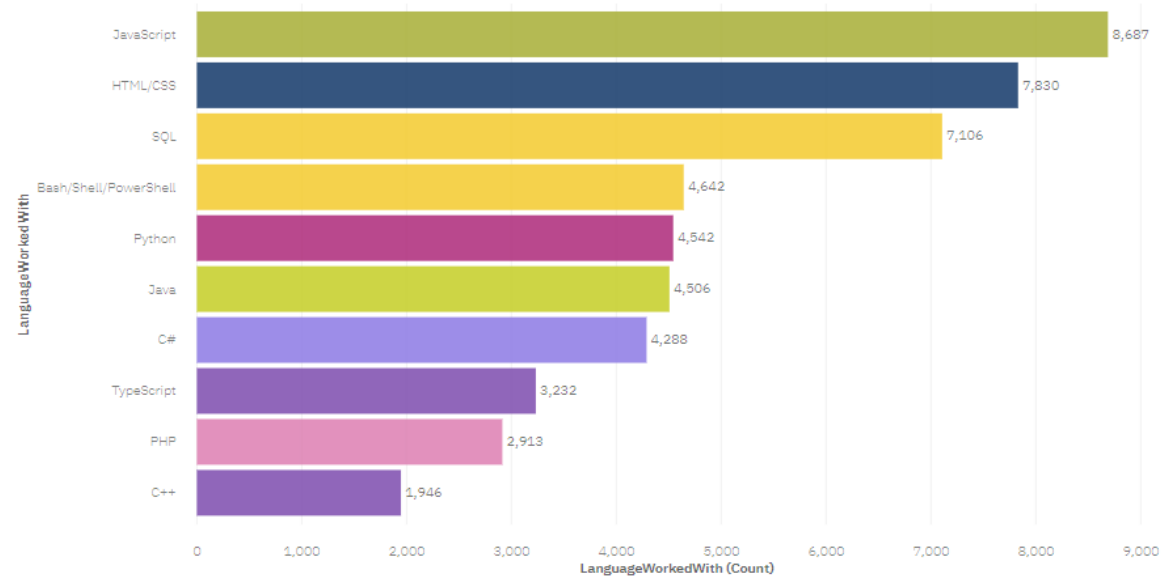
Top 5 Languages that Respondents Wish to Learn Next Year



# PROGRAMMING LANGUAGE TRENDS

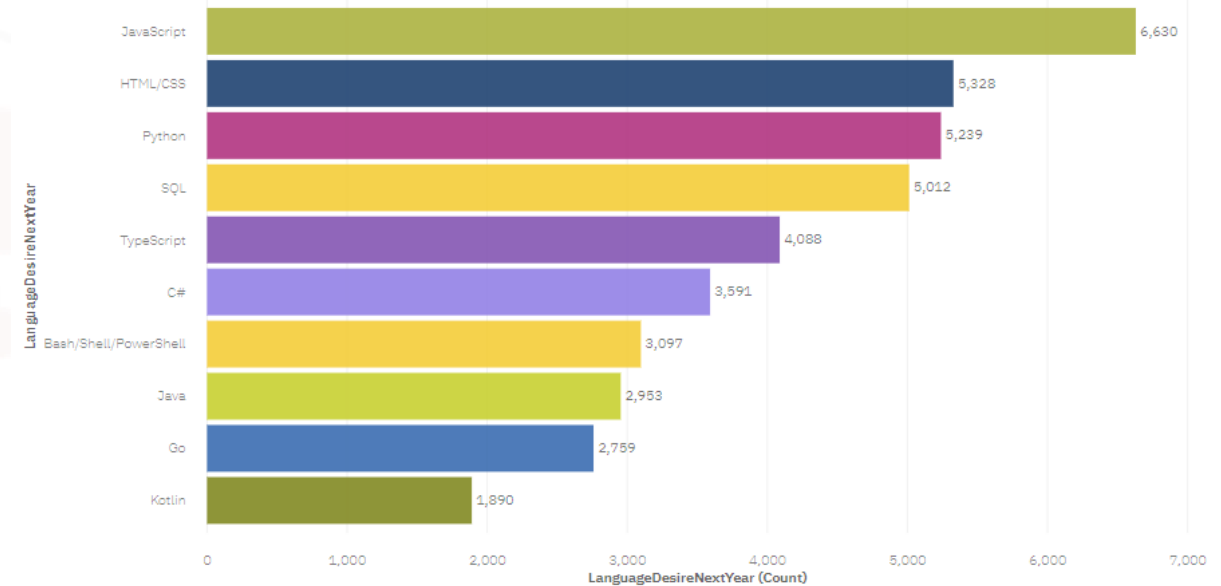
## 2019

Top 10 Languages Worked With



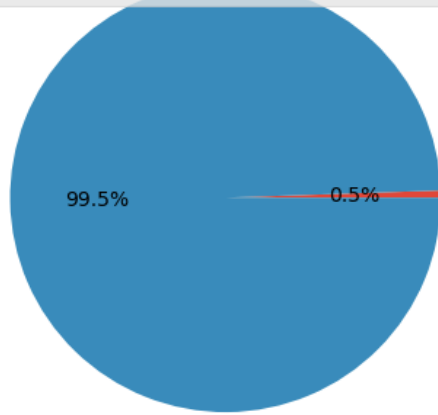
## Next Year

Top 10 Language Desire Next Year

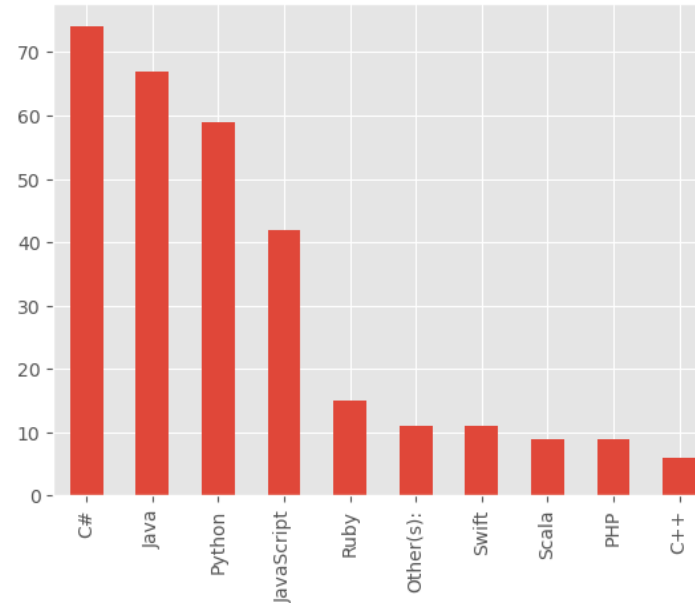


# PROGRAMMING LANGUAGE TRENDS

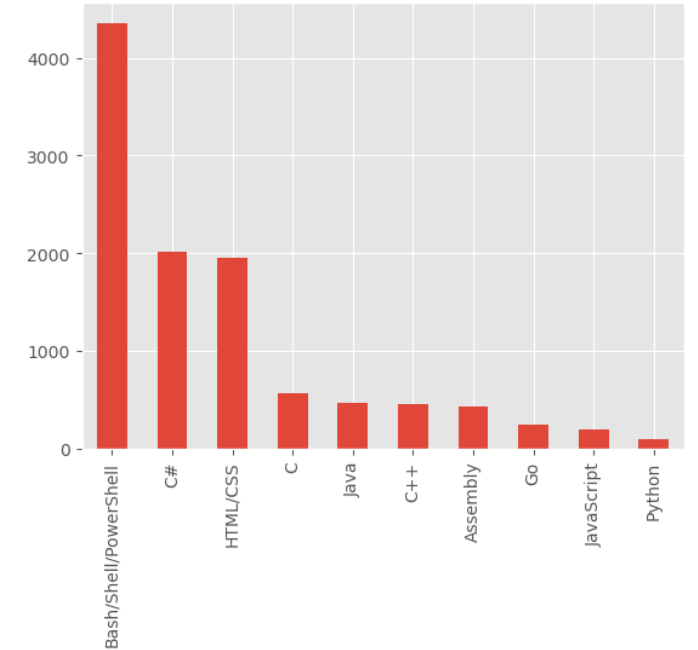
Number of Respondent that knows only one language  
Number of Respondent that knows more than one language



Top 10 Languages that Respondent Knows Only One of Them



Top 10 Languages that Respondent Knows More Than One of Them



99.5% of the respondents know multiple programming languages and almost all of them knows Bash/Shell/PowerShell because ,it is the majority of the operating systems

# PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

---

## Findings

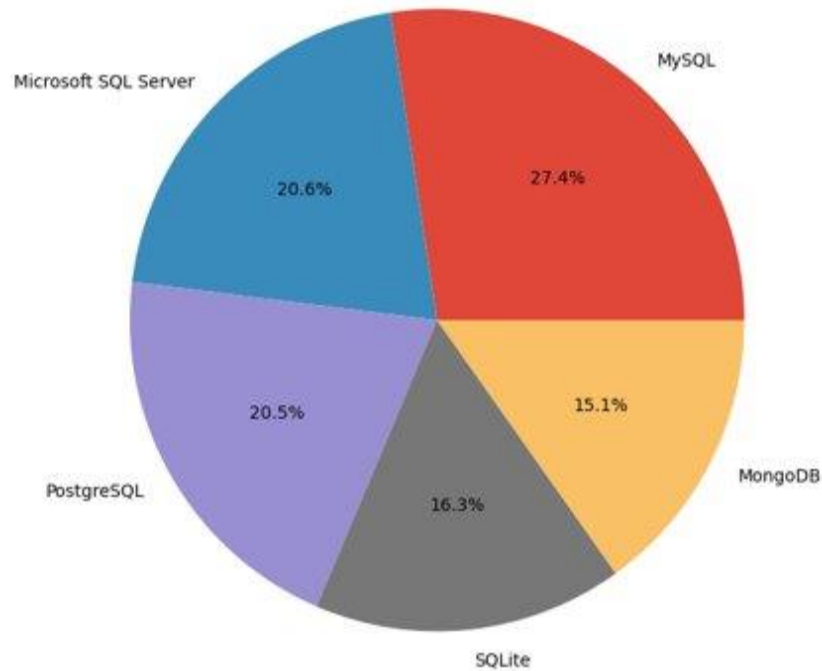
- 26.5% of the respondents worked with JavaScript
- 25.2% of the respondents want to learn JavaScript next year
- Interest to Python is increasing
- SQL is still popular and useful
- JavaScript has always high interest among programmers
- Almost all of the programmers knows more than one language.

## Implications

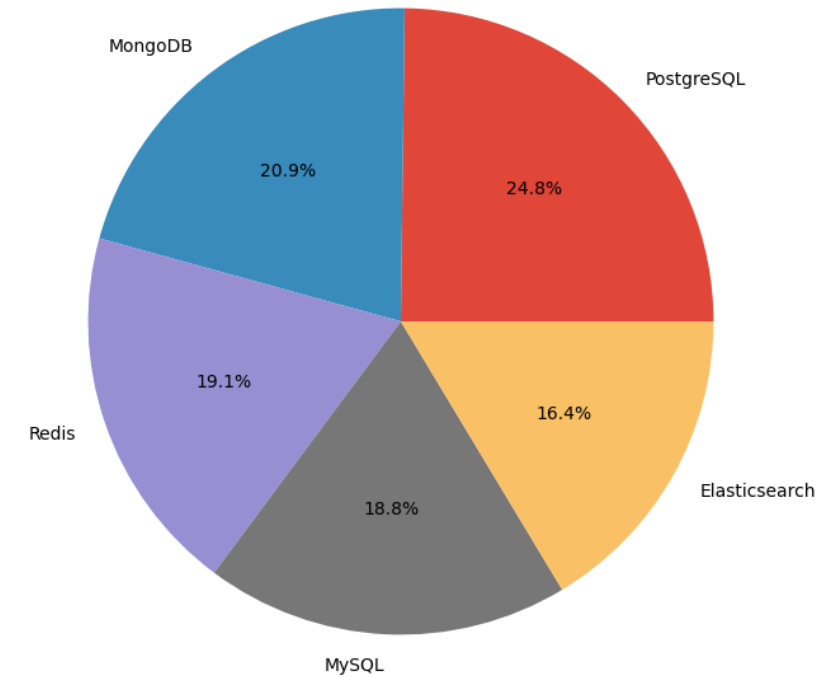
- C++ interest is decreasing in years
- SQL and Python is popular because of Data Science/Analyst, ML and AI popularity
- Typescript is seemed as new trend
- JavaScript and HTML/CSS shows that front end developing is popular
- Kotlin is in the top 10 list of language that are wished to learn
- Almost all of programmers knows Bash/Shell/PowerShell it is the majority of the operating systems

# DATABASE TRENDS

Top 5 Databases that Respondents Work With



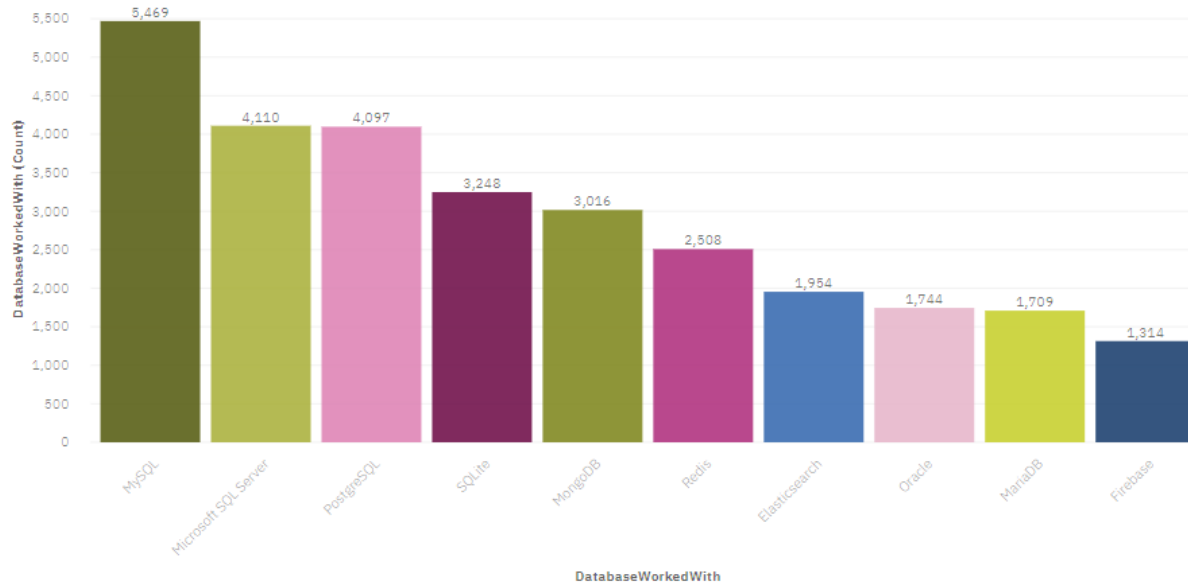
Top 5 Databases that Respondents Wish to Learn Next Year



# DATABASE TRENDS

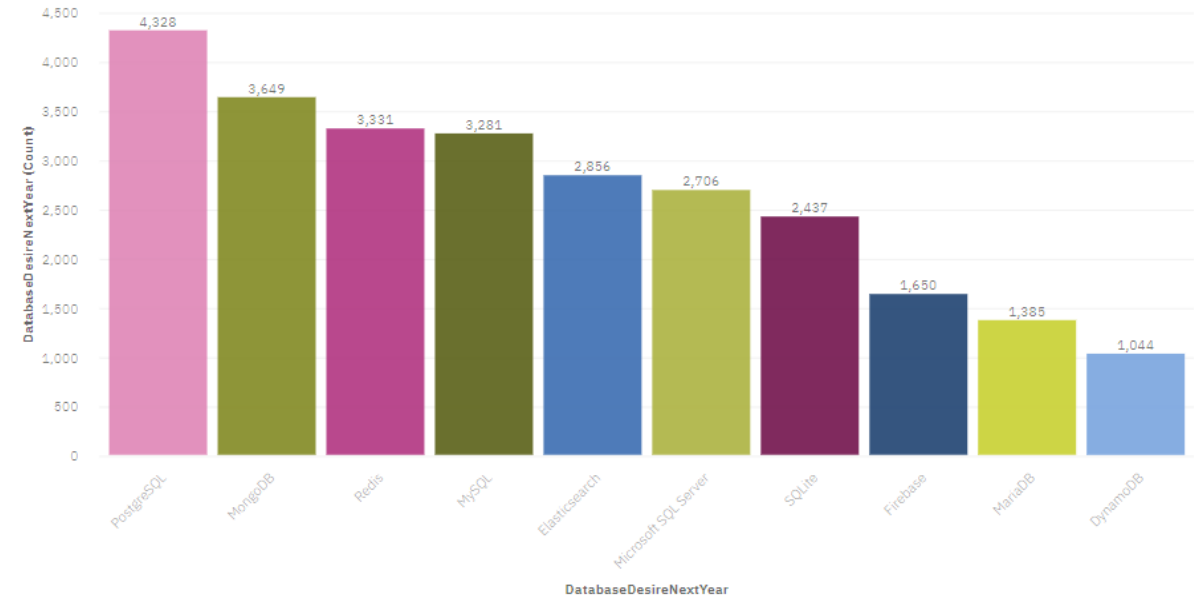
## 2019

Top 10 Database Worked With



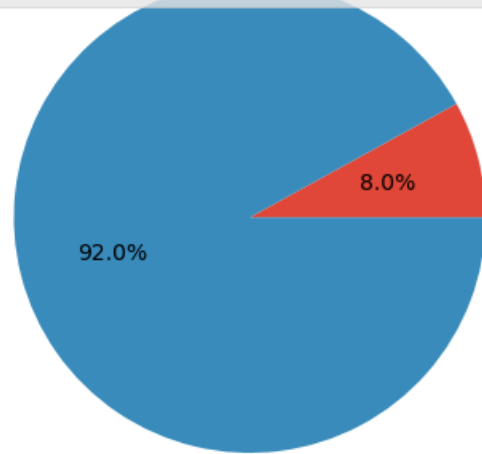
## Next Year

Top 10 Database Desire Next Year

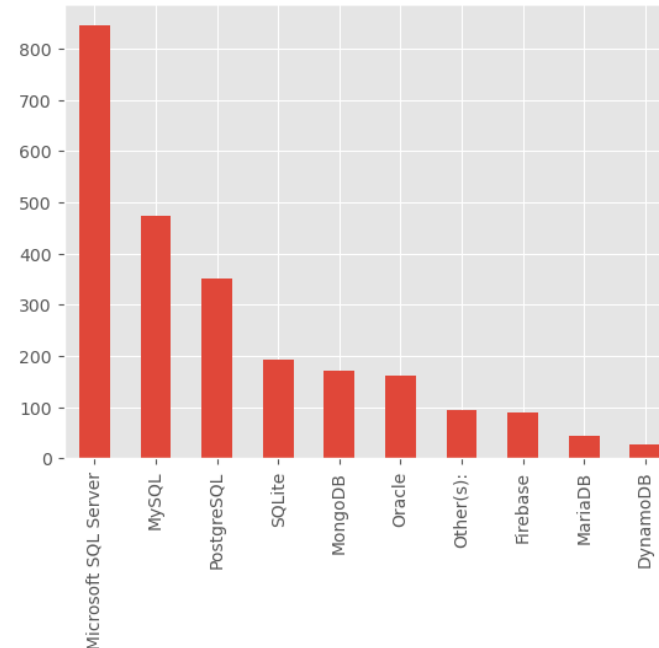


# DATABASE TRENDS

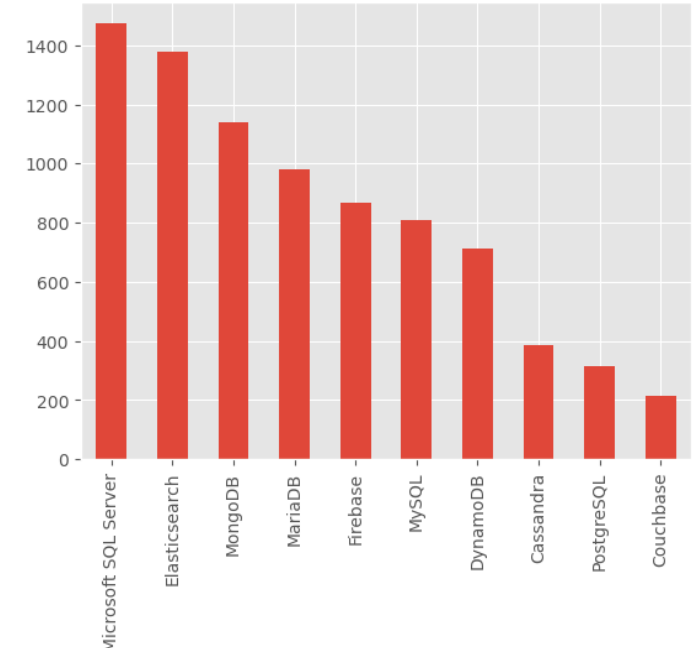
Number of Respondent that knows only one database  
Number of Respondent that knows more than one database



Top 10 Databases that Respondent Knows Only One of Them



Top 10 Databases that Respondent Knows More Than One of Them



More than 90% of the respondents knows more than one DBMS. They know mostly Microsoft SQL Server and after that they know Elasticsearch, MongoDB and MariaDB

The people that know only one database use mostly Microsoft SQL server.



# DATABASE TRENDS - FINDINGS & IMPLICATIONS

---

## Findings

- 24.8% of the respondents want to learn PostgreSQL next year
- 27.4% of the respondents worked with MySQL
- PostgreSQL may take the lead from MySQL
- MongoDB interest is increasing
- Elasticsearch is new interest
- More than 90% of the respondents knows more than one database programs, Microsoft SQL Server, Elasticsearch, MongoDB and MariaDB respectively

## Implications

- MS SQLServer popularity is decreasing so developers prefer mostly open source DBMS
- Open source DBMS performance, support and freedom might be one reason to be preferred.
- NoSQL database usage is increasing
- SQLite popularity is getting low because it is file-based database, not server
- MS SQL server, MySQL, PostgreSQL are known by almost everyone because they are basics of DBMS
- Elasticsearch and MongoDB are related with big data and NoSQL databases so AI, data science and ML are trend topic

# DASHBOARD

---



You can access the dashboard from the link below. It is interactive and you can analyze the data.

[Cognos dashboard](#)

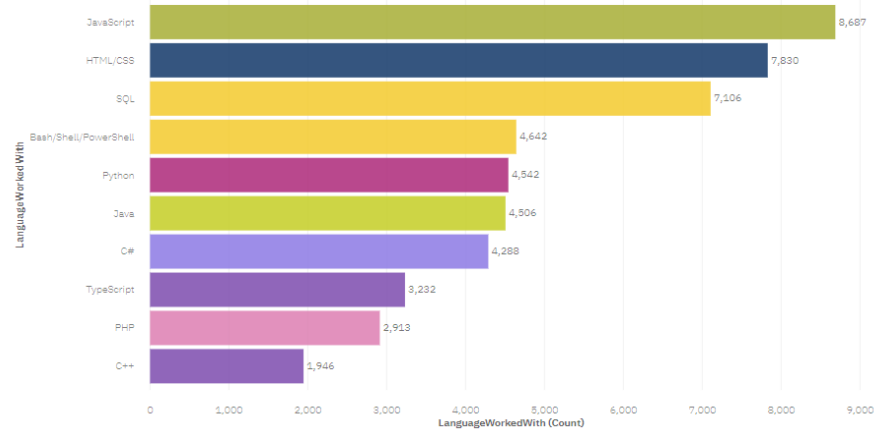
# DASHBOARD TAB 1

Current Technology Usage

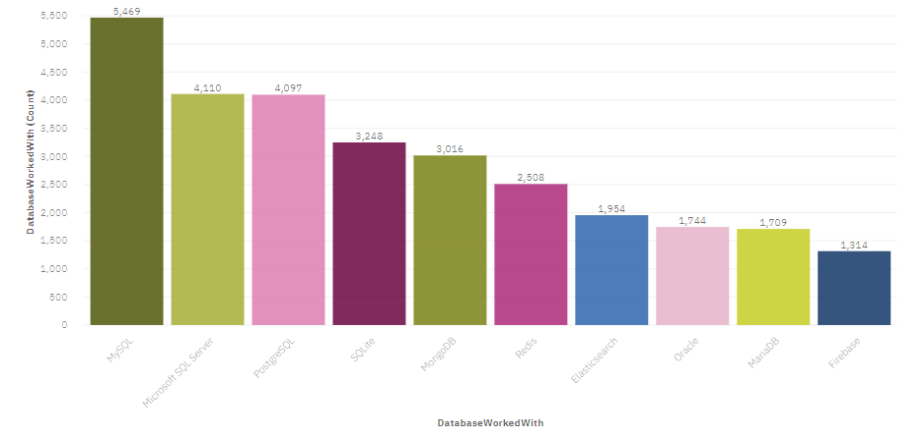
Future Technology Trend

Demographics

Top 10 Languages Worked With



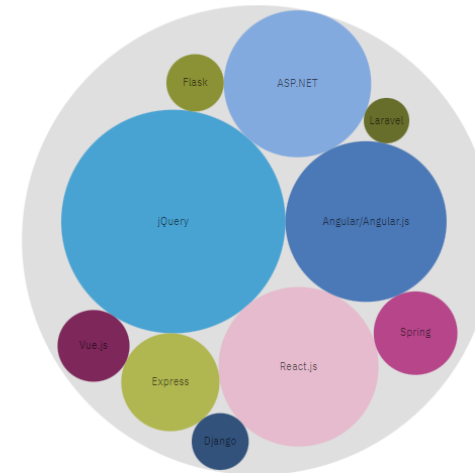
Top 10 Database Worked With



Platforms Worked With



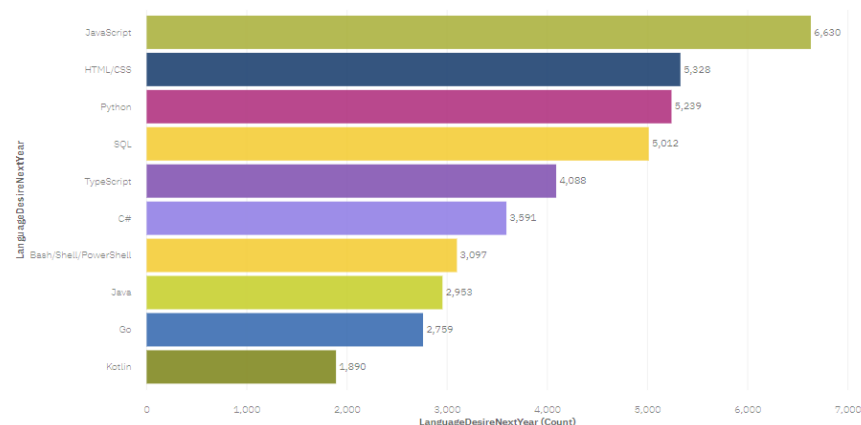
Top 10 Web Frame Worked With



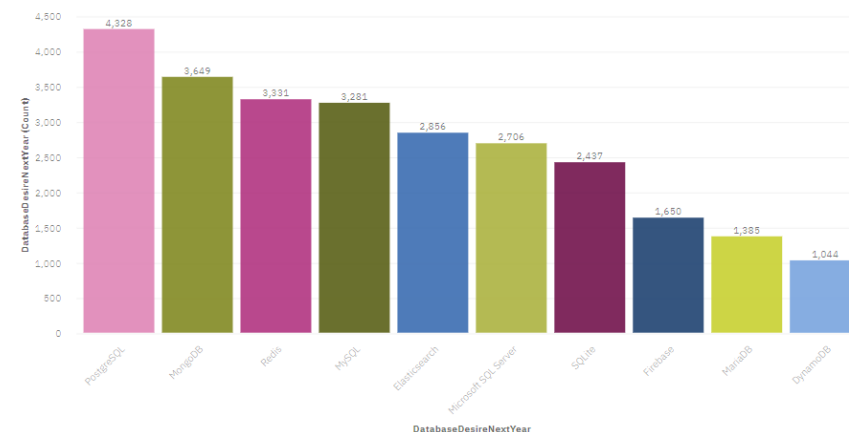
# DASHBOARD TAB 2

Current Technology Usage   **Future Technology Trend**   Demographics

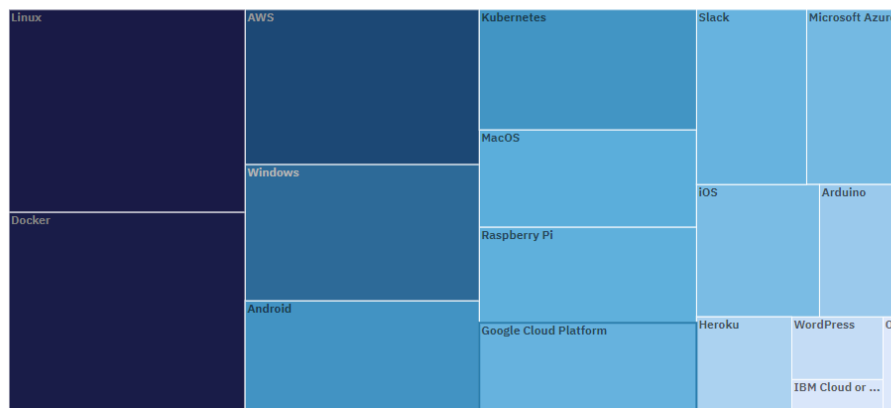
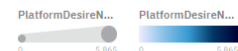
Top 10 Language Desire Next Year



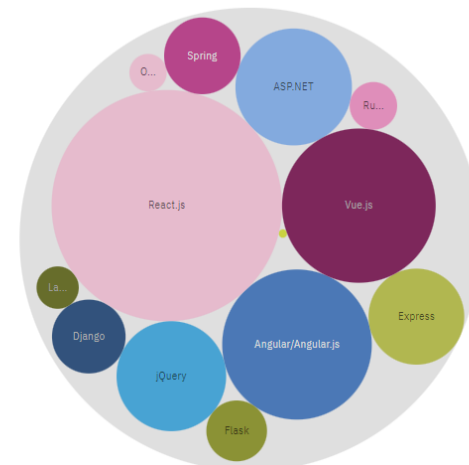
Top 10 Database Desire Next Year



Platform Desire Next Year



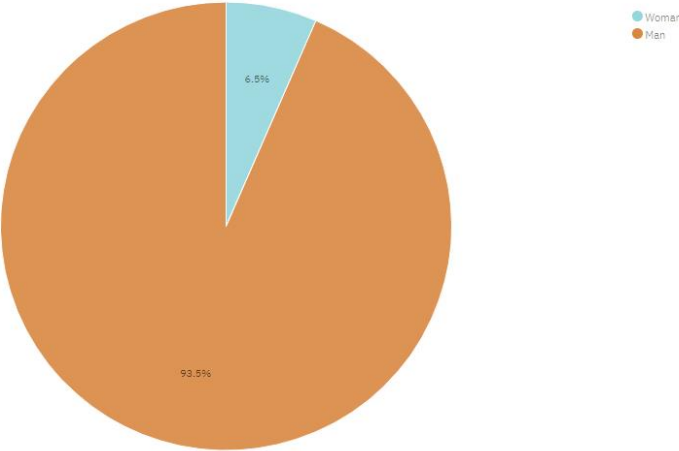
Top 10 Web Frame Desire Next Year



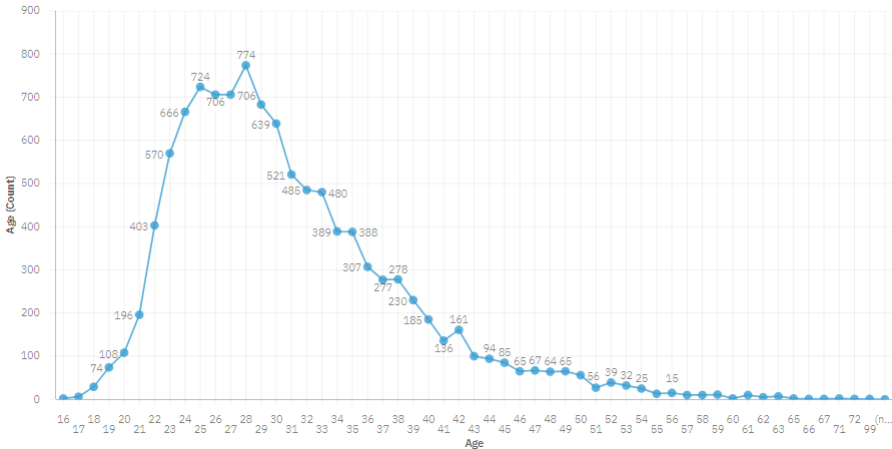
# DASHBOARD TAB 3

Current Technology Usage   Future Technology Trend   **Demographics**

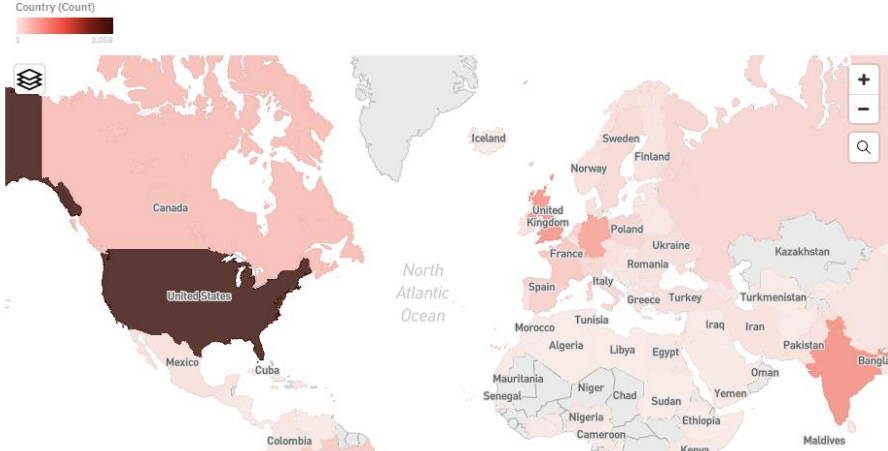
Respondent classified by gender



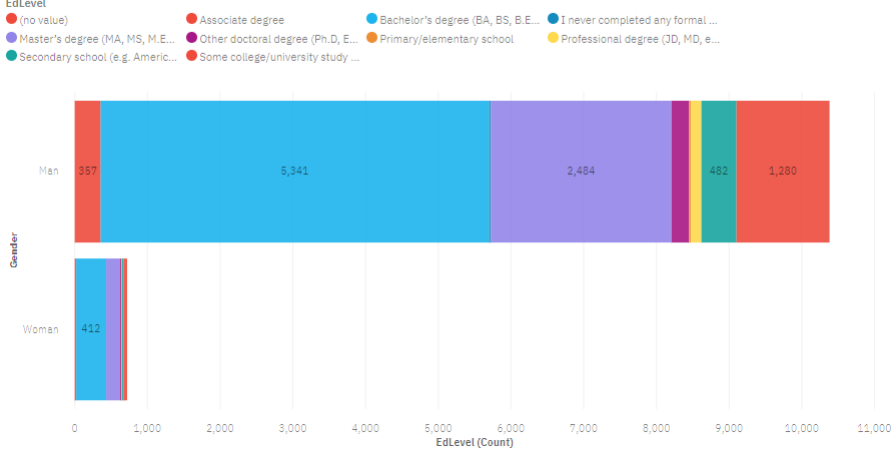
Respondent Count by Age



Respondent Count for Countries



Respondent Count by Gender, classified by Formal Education Level



# DISCUSSION

---



- Which technologies are high demand among developers?
- What is the current technology usage?
- What is the future trend of technology?
- What is the distribution and trend of converted compensation?
- What is the demographic analysis of the respondents? For example age, gender, geography, education level and compensation

# OVERALL FINDINGS & IMPLICATIONS

---

## Findings

- Programmers prefer Windows and Linux
- Interest to Python is increasing
- SQL is still popular and useful. PostgreSQL may take the lead from MySQL
- JavaScript has always high interest among programmers
- Mongo DB as a NoSQL database system was the most used in 2019
- Gender gap is too high.
- Most of the developers attended to the survey from USA
- Most of the respondents have Bachelor Degree
- Two or more database systems and languages are widely used among programmers

## Implications

- Next year Linux platform popularity will increase because of security and freedom
- Web development is still popular. React.js for web frame is becoming popular
- Because of developing of ML, AI and data science, SQL, Python and No SQL database programs are still popular and they will be for near future.
- Gender gap seems to be higher for the future. There should be some policies and opportunities for increasing the number of female developers
- Technology is rapidly changing. Developers should keep updated by themselves
- One database system and language are not enough

# CONCLUSION

---



- Web development still has a higher trend because JavaScript and HTML/CSS are still preferable.
- Python, SQL are the popular languages
- MySQL has higher user but PostgreSQL might be next year.
- Respondent that knows only one database system and language is lower than respondent that knows more than one. Because Web development and Data science, AI needs more than one database and language knowledge
- Most of the respondents knows two or more programming languages and database systems
- Gender gap is high. There is no equality between genders



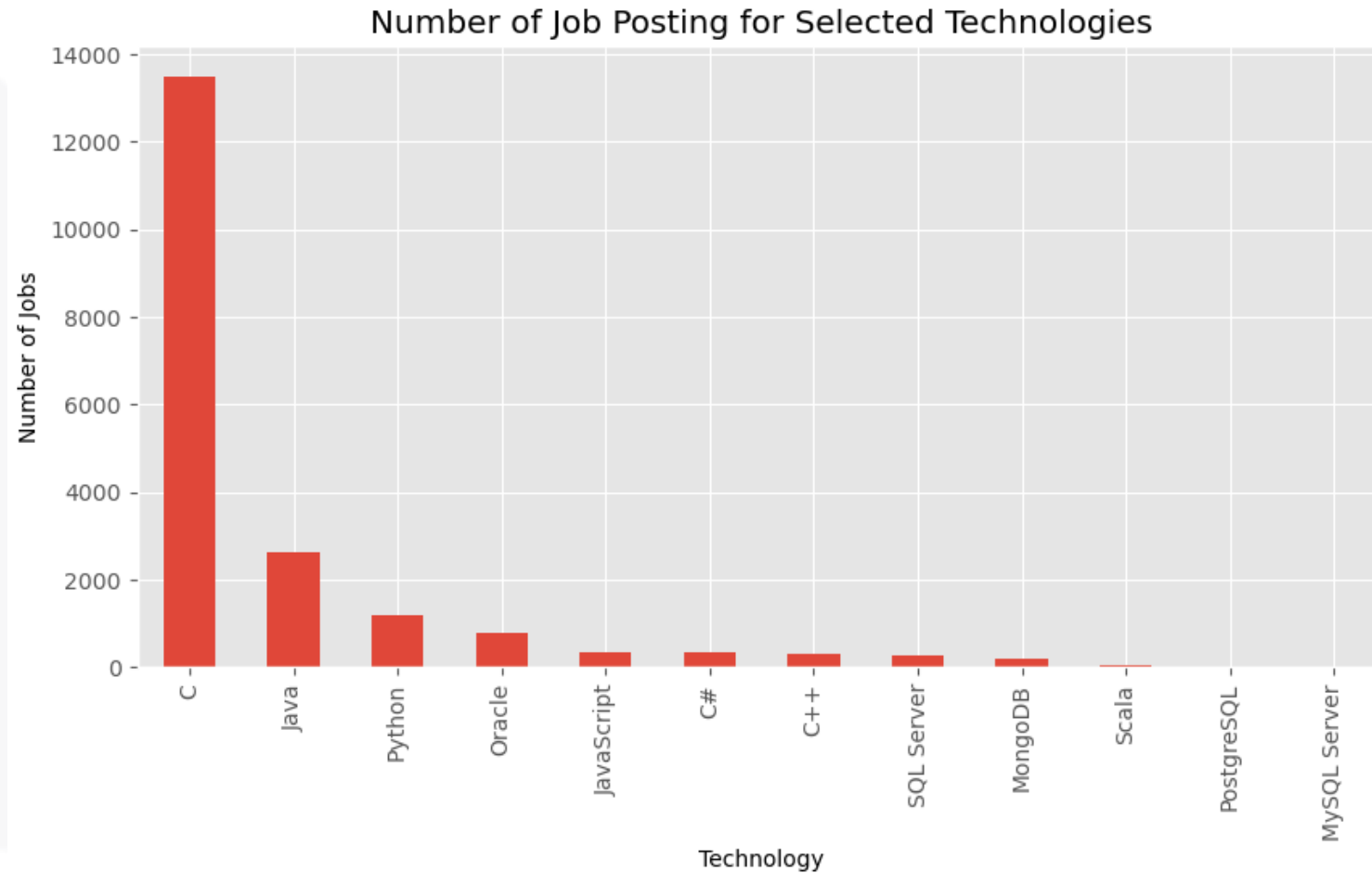
# APPENDIX

---



- Include any relevant additional charts, or tables that you may have created during the analysis phase.

# JOB POSTINGS



# POPULAR LANGUAGES

