



# IBM Data Analyst Capstone Project

Manikyala Rao Cherukuri  
Jan 2023

# OUTLINE

---



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

# EXECUTIVE SUMMARY

---

The following slides will summarize the data collected as part of **2019 stack overflow developers survey**.



This Data Analysis will have the insights regarding the following:

- Current Technology Usage (i.e., popular Languages, Databases, Platforms & Web Frames)
- Future Technology Usage (i.e., Desired languages, Databases, Platforms, Web Frames)
- Demographics (i.e., Gender, Age & Education levels between the Developers along with developers count per country)

These findings will help to understand what are the popular & desired technologies which are mostly liked by the software developers including country wise.

# INTRODUCTION

---



- Since 2011, the online programming knowledge sharing platform, Stack Overflow, has been conducting the Overflow Annual Developer Survey.
- Stack The primary objective of the annual surveys is to gather data regarding technology usage and trends among developers.
- In this analysis, a subset of the 2019 dataset was examined (present dataset:  $N = 11,552$  ; original dataset  $N \approx 90,000$ ).
- Audience: Developers (current and aspiring), HR professionals, educators, policy makers

# METHODOLOGY

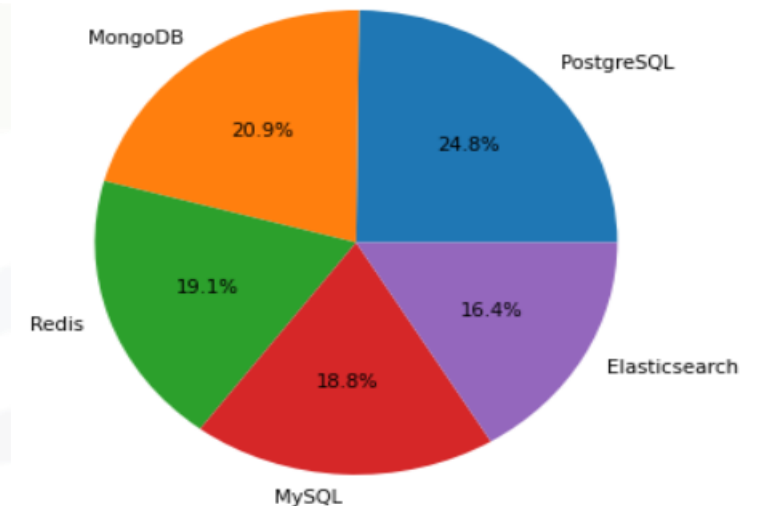
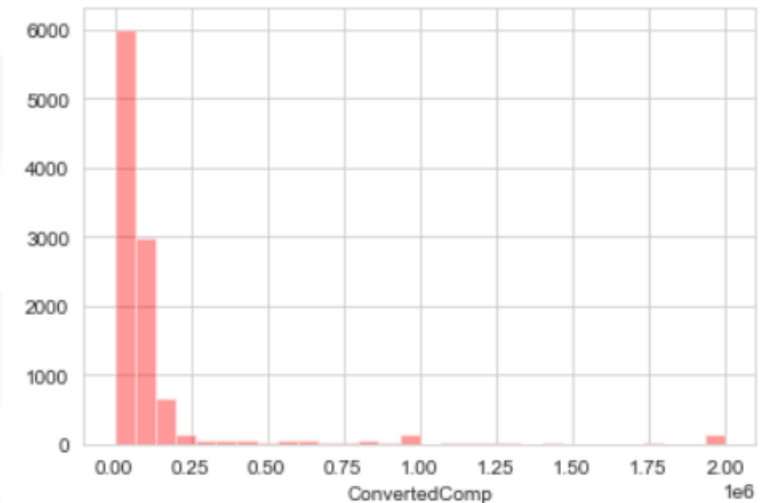
---



- **Data Source** : 2019 Stack Overflow Developer Survey
  - Here is the original data source [\[link\]](#)
- **Data Wrangling:** Here, we used the portion of (dataset provided by IBM) and that is cleaned using Python panda's library.
  - **Cleaning steps involved:** Duplicates removal, replacing the missing values, normalizing the column values.
  - Here is the dataset [\[link\]](#)
- **Exploratory Data Analysis:** It has been used to analyze the data [\[link\]](#) and plot the kernel density estimation curve & histogram for desired columns along with boxplot to identify the outliers.
- **Data Visualizations:** Here we have used Sqlite3 along with Python panda's library to analyze the visualize the data and used IBM Cognos to create a dashboard to observe the patterns in the data.

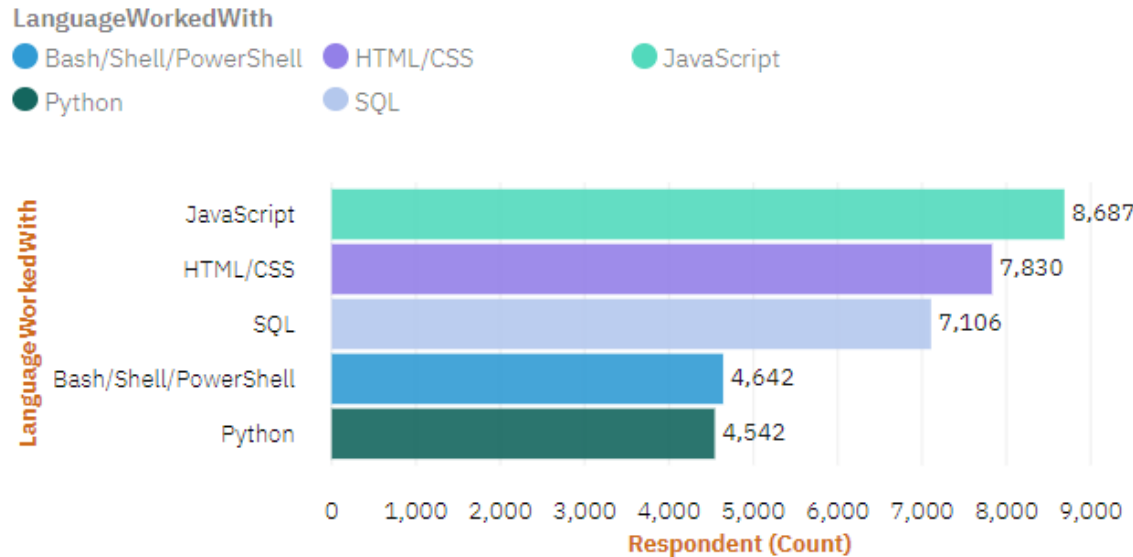
# RESULTS

- In **data wrangling** after removing the duplicates and replacing the null values, I have calculated the **NormalizedAnnualCompensation** for the total number of records 11398.
- In **Exploratory Data Analysis** have made following observations:
  - I have drawn histogram and kde for the converted annual compensation column and median compensation here is 57745\$.
  - And then calculated the median annual compensation for women i.e., 57708\$ after filtered the gender column.
  - Later drawn the boxplot for converted compensation column to understand and then filter out the outliers. After removing the outliers 10519 would be the final records count.
  - Here mean & median annual compensations are 59883\$ and 52704\$ respectively.
- In **Data Visualization** I have drawn a pie chat for top 5 databases, where respondents wish to learn next year.

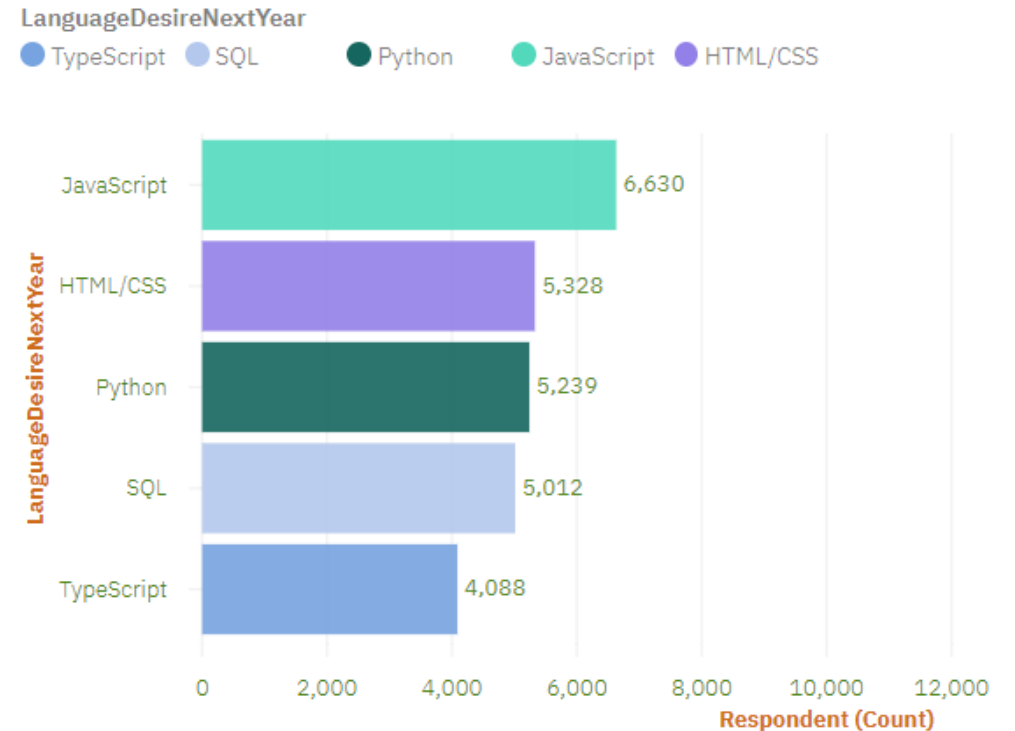


# PROGRAMMING LANGUAGE TRENDS

Bar chart for Top 5 Languages worked with



Bar chart for Top 5 LanguageDesireNextYear



# PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

---

## Findings

- JavaScript were the most popular in 2019 and will likely remain so the following year.
- SQL was popular in 2019 and will likely remain so.
- Increase interests in Python and TypeScript.
- Decreasing interest in PowerShell/ Bash.

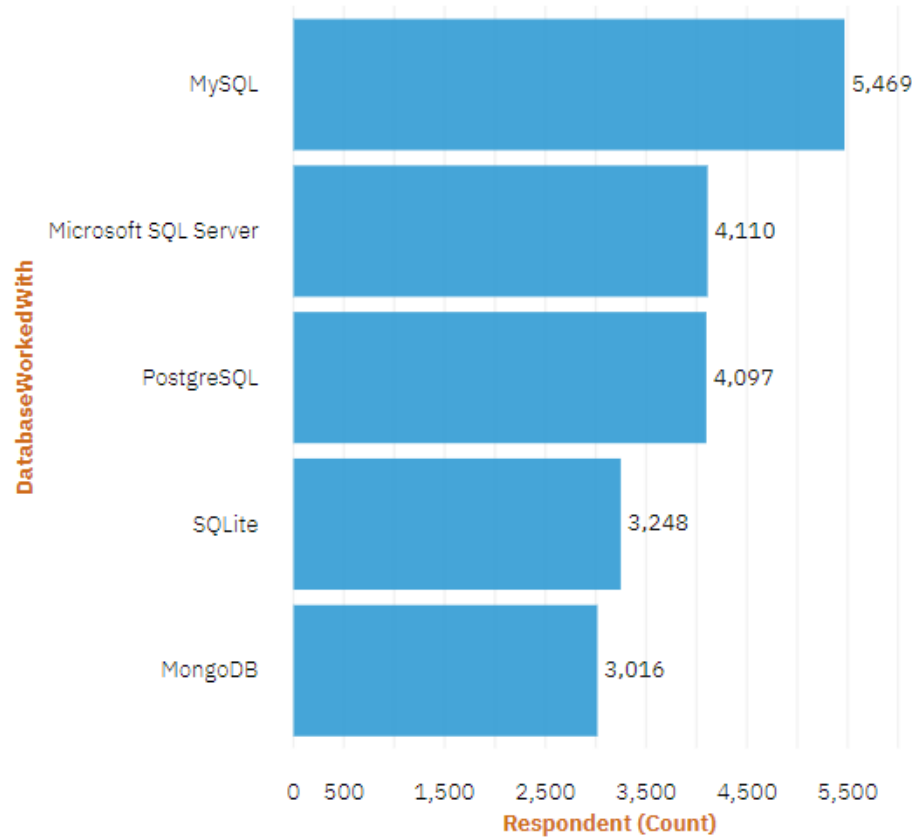
## Implications

- Web development is still in high demand, and JavaScript languages and — HTML/CSS however, remain the dominant TypeScript the future.
- SQL remains the preferred language for big data storage and querying and it's not likely go away anytime soon.
- Python rising popularity likely reflects the growth of AI and ML work.

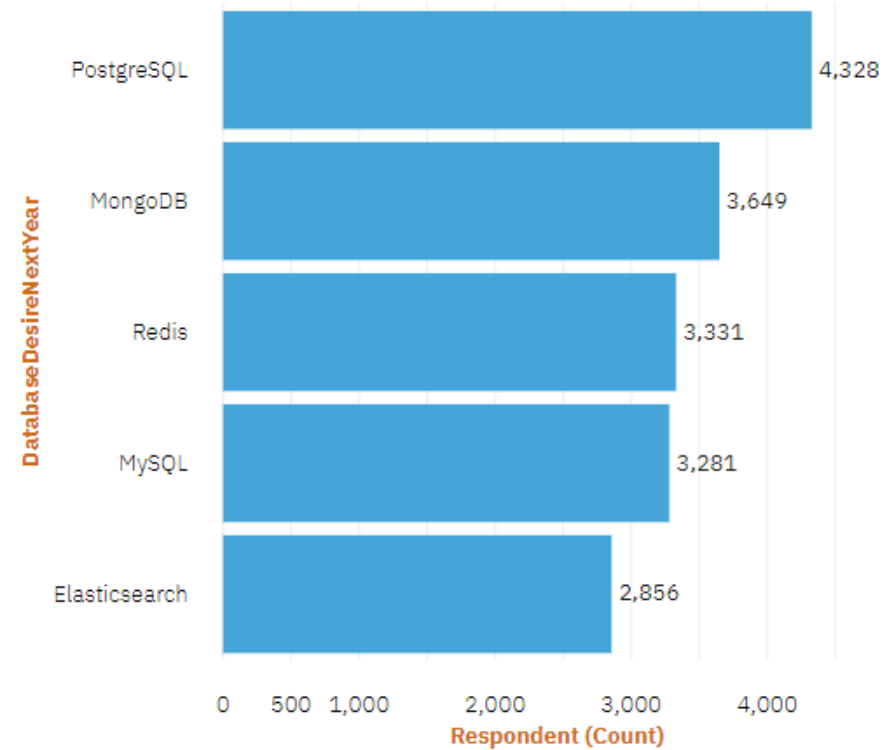


# DATABASE TRENDS

Bar chat for top 5 DatabaseWorkedWith



Bar chat for top 5 DatabaseDesireNextyear



# DATABASE TRENDS - FINDINGS & IMPLICATIONS

---

## Findings

- SQL database programs are the most popular in 2019. MySQL is the lead among them.
- PostgreSQL is gaining popularity over other SQL database programs, and it was the overall most desired database for the next year.
- MongoDB popular in 2019 and gaining interest.
- Increasing interest in Elasticsearch.

## Implications

- There appears to be increasing developer preference towards open - source database programs.
- NoSQL database programs are gaining popularity, which likely reflects a growing need to handle non relational and unstructured data.
- Current and aspiring data analysts should develop competence in NoSQL in addition to SQL database programs.

# DASHBOARD

---



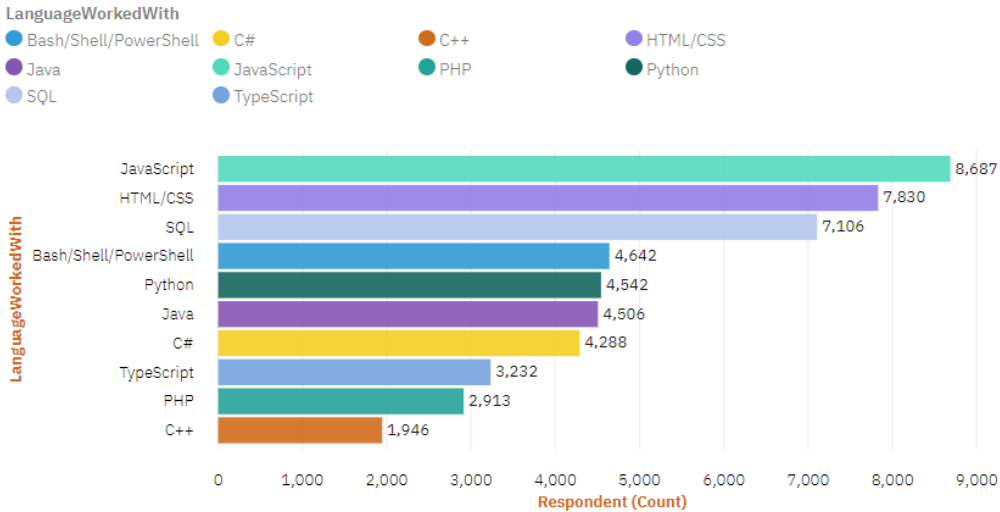
The following link contains the full, interactive dashboard summarizing (a) Cognos current technology use future technology trend , and (c) demographics , (b) of the survey respondents:

**[Click this link to navigate to the dashboard](#)** [Ctrl+click]

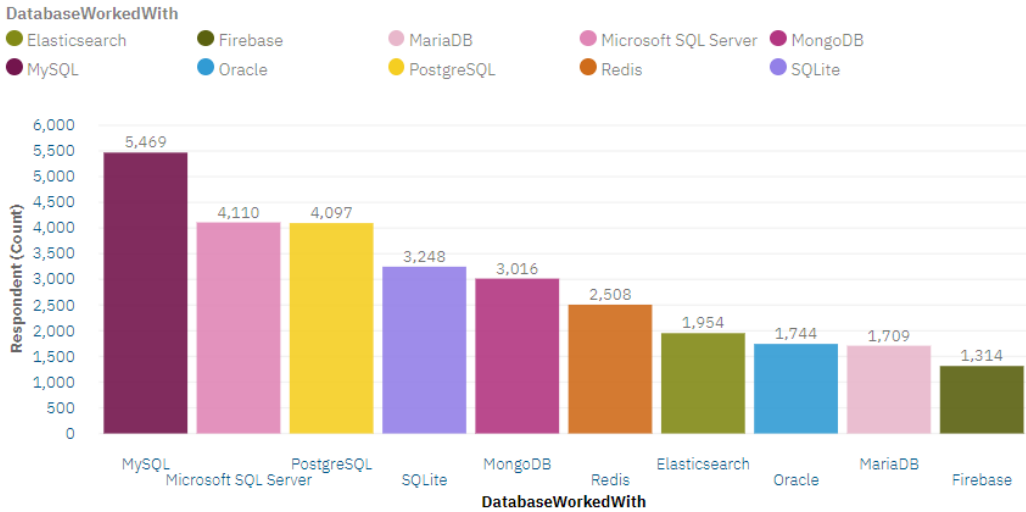
Static screenshots of the dashboard are shown in the next three slides.

# DASHBOARD A: Current Technology Usage:

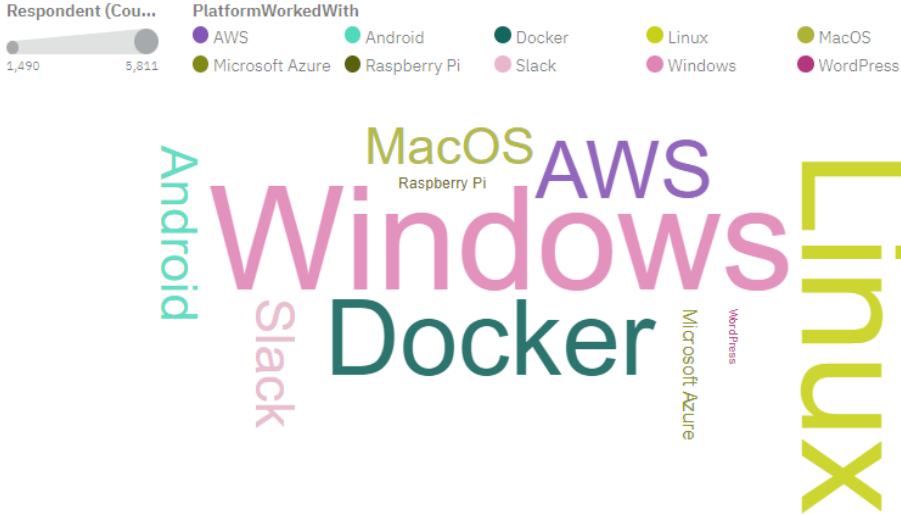
Bar chat for Top 10 Languages worked with



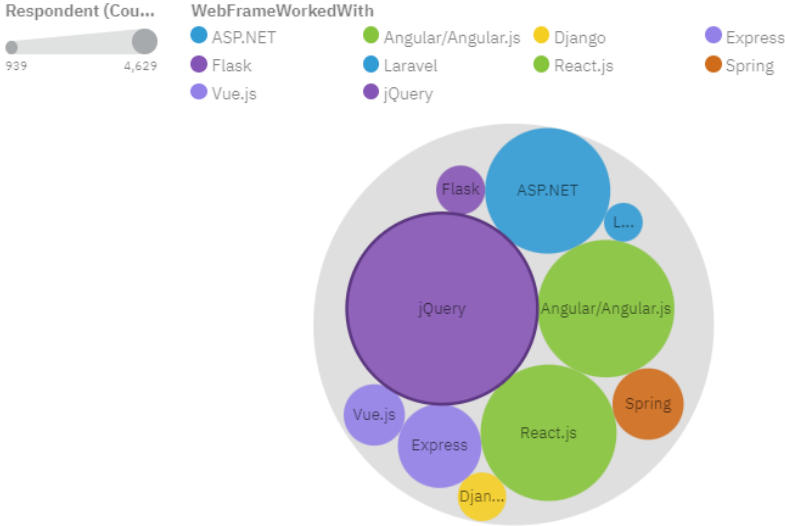
Column Chat for Top 10 Databases worked with



Word Count Chart for Top 10 Platforms Worked With

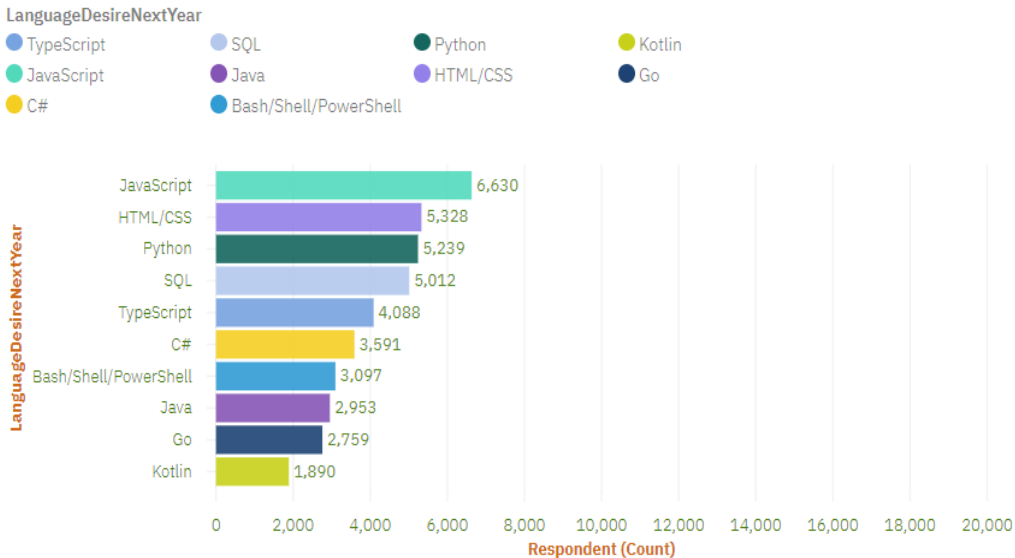


Hierarchy bubble chart for Top 10 Web Frames Worked With

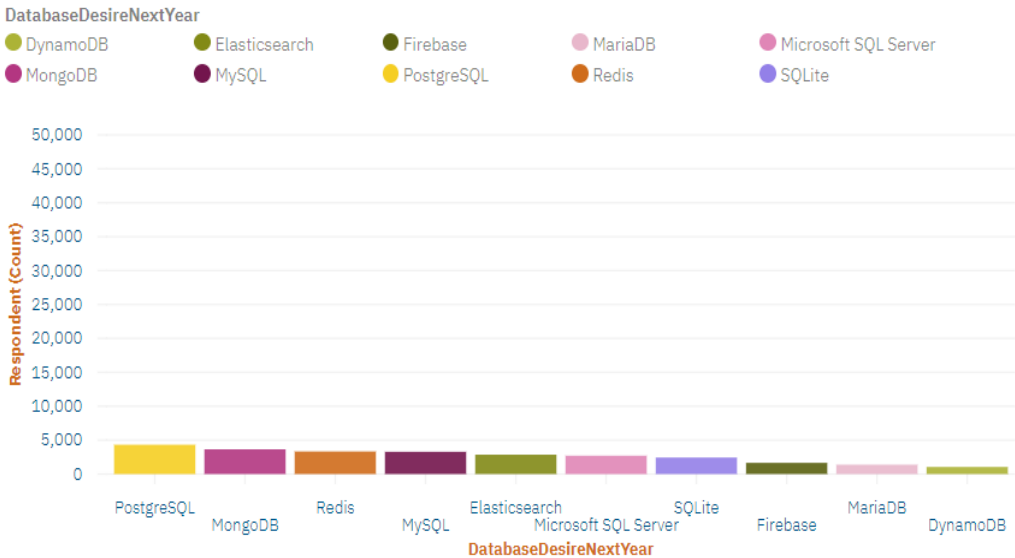


# DASHBOARD B: Future Technology Usage:

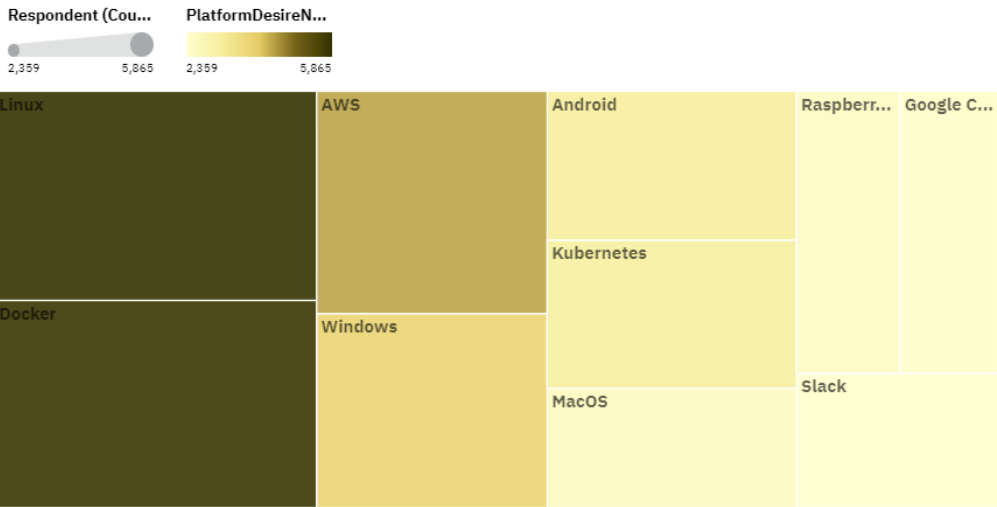
Bar chart for Top 10 LanguageDesireNextYear



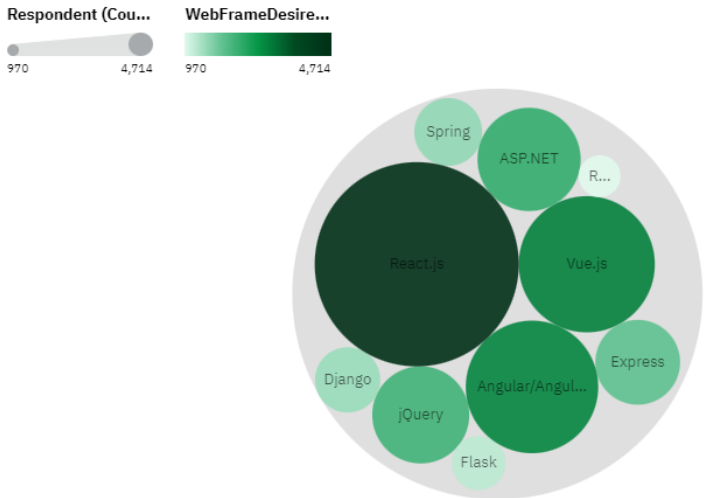
Column Chart for Top 10 DatabaseDesireNextYear



Tree Map Chart for Top 10 PlatformDesireNextYear



Hierarchy bubble chart Top 10 WebFrameDesireNextYear

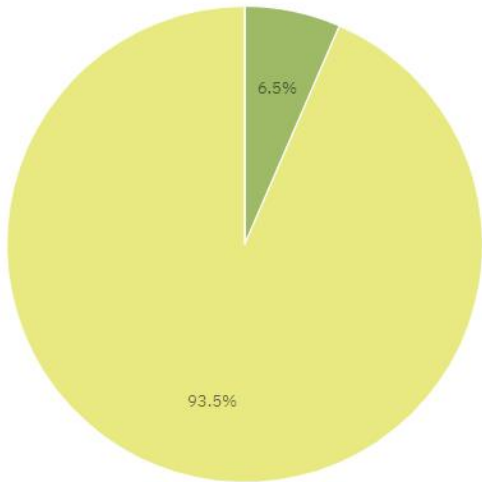


# DASHBOARD C: Demographics:

Pie Chart for Respondent classified by Gender

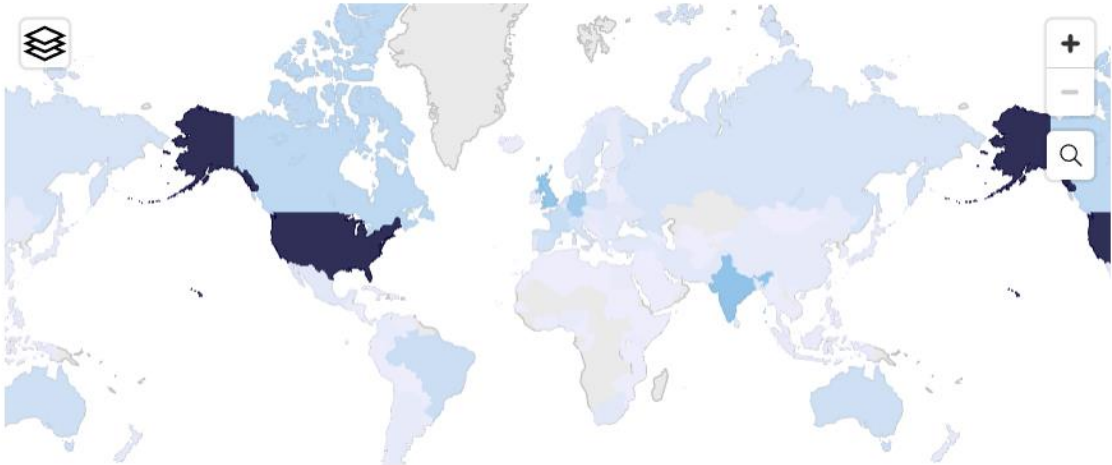
Gender

- Woman
- Man

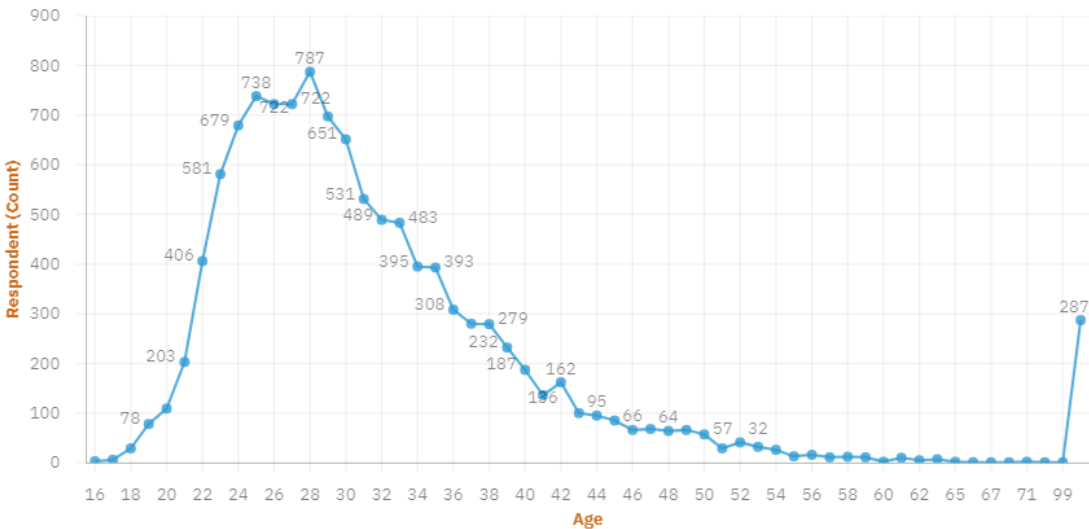


Map Chart for Respondent Count for Countries

Respondent (Cou...  
1 3,127



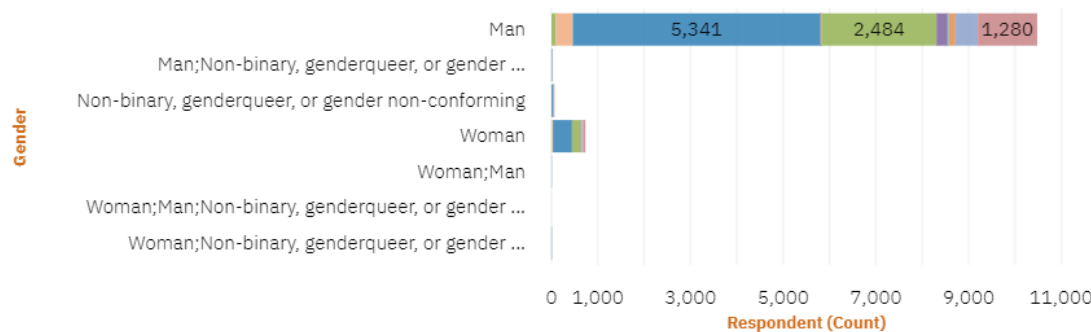
Line Chart for Respondent Count by Age



Stacked bar chart for Respondent Count by Gender, classified by Formal Education Level

EdLevel

- (no value)
- I never completed any formal ...
- Primary/elementary school
- Some college/university study ...
- Associate degree
- Master's degree (MA, MS, M.E...
- Professional degree (JD, MD, e...
- Bachelor's degree (BA, BS, B.E...
- Other doctoral degree (Ph.D, E...
- Secondary school (e.g. Americ...



# DISCUSSION

---



Taken together, the findings yield insights into the following questions:

- What kinds of developer technologies are in top demand?
- Which technologies should prospective developers and data professionals be learning?
- Which technologies should educators place more emphasis on teaching in upcoming years?
- What does the distribution of annual compensation for developers look like?
- What is the developer demographic like? Is there a gender representation gap?

# OVERALL FINDINGS & IMPLICATIONS

---

## Findings:

- High usage and interest in Java Script & HTML/CSS is remain high usage.
- There's also increasing interest in Python.
- High usage and interest in SQL. MySQL had the highest usage in 2019, but PostgreSQL is gaining interest and was the overall most desired database program for the next year.
- NoSQL database programs gaining interest, of which MongoDB was the most used in 2019 and desired for the next year.
- Technology divide between countries.

## Implications:

- Web development is still in high demand. Current and prospective developers may consider picking up Typescript in addition to JavaScript and HTML/CSS.
- With the growing need to handle big data and perform AI and ML work, data professionals should continue to enhance SQL competence but also enhance competence with NoSQL database programs and Python.
- Businesses need to adapt to changing technology preferences, especially in terms of talent acquisition and development.
- Policy makers, educators, and organizations should work to minimize the gender representation gap in addition to the technology divide between countries.



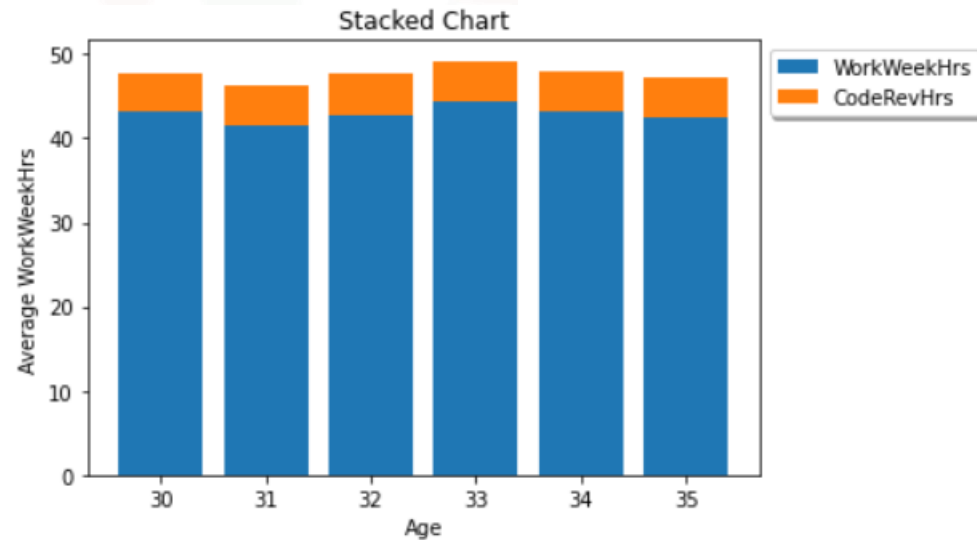
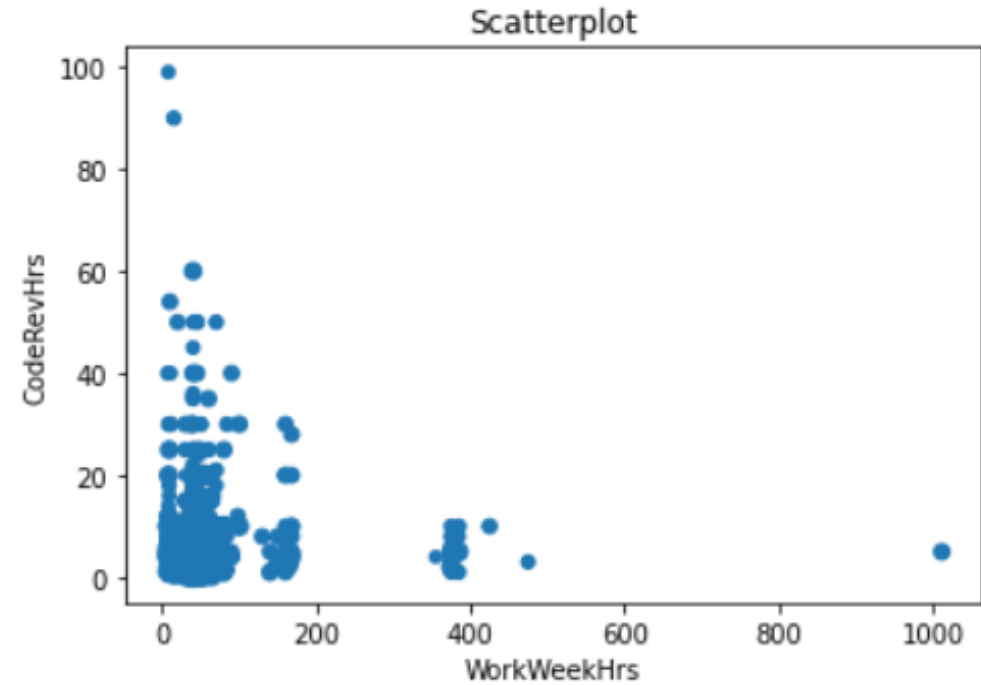
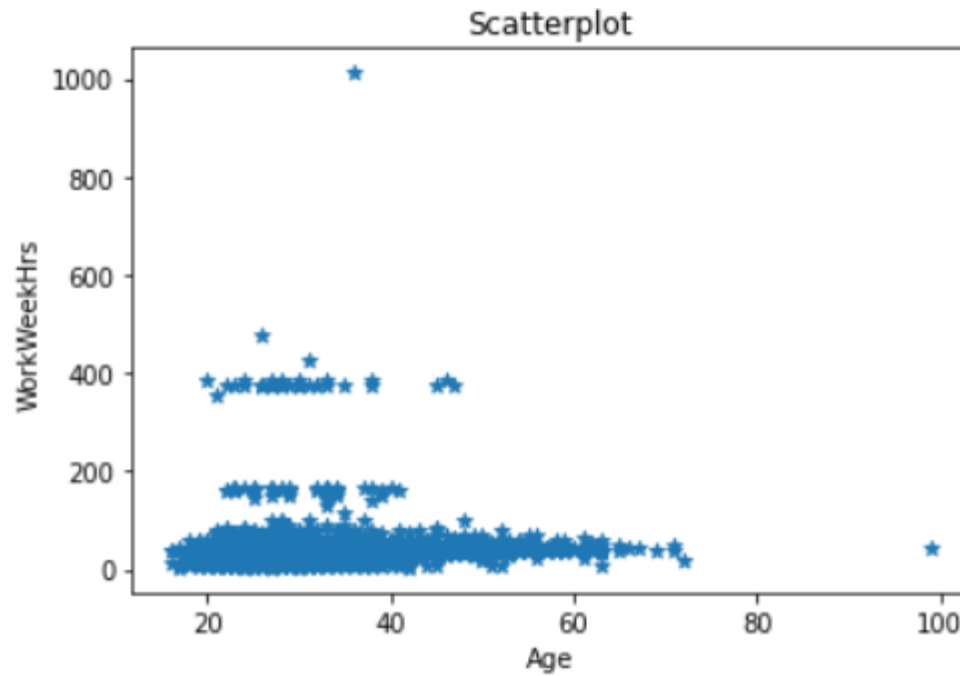
# CONCLUSION

---



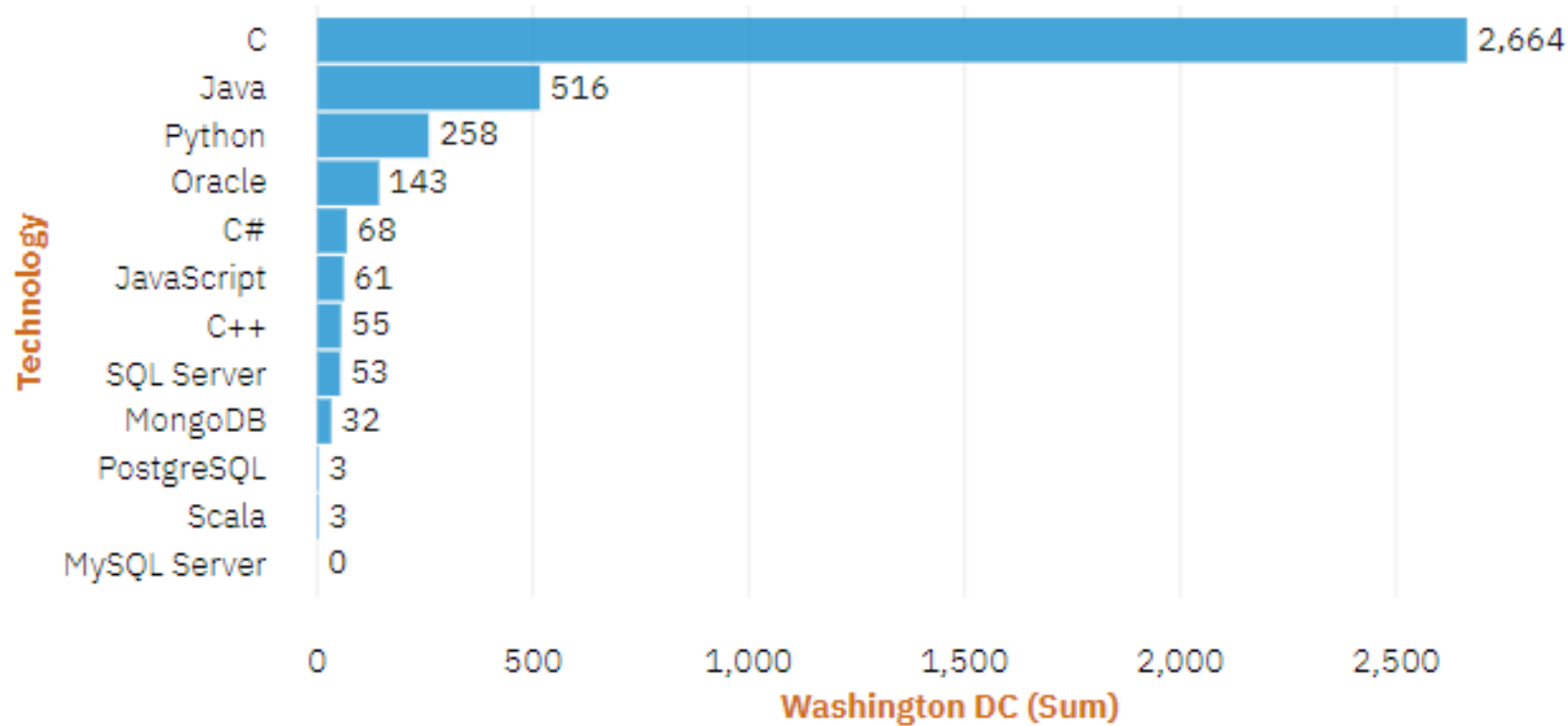
- A subset of data collected as part of the Overflow Developer Survey 2019 Stack was examined.
- The findings yielded numerous insights into the technologies most used and desired by developers in addition to the developer demographic.
- These insights should be particularly relevant for current and prospective developers aiming to remain competitive, businesses aiming to upskill their talent, educators in the field, and policy makers aiming to address gender and economic issues.

# APPENDIX



# JOB POSTINGS

Bar chart for Number of jobs VS Technology



# POPULAR LANGUAGES

Bar chart for Languages VS Annual Salary

