IBM DATA ANALYTICS CAPSTONE

KAROL JOHNSTON

10 DECEMBER 2022

IBM Developer

SKILLS NETWORK

OUTLINE

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

EXECUTIVE SUMMARY

- The following slides summarize key findings from an analysis of data collected as part of the 2019 Stack Overflow Developer Survey.
- The analysis yielded insights regarding the following:
  - Most popular languages, databases, and other technologies (at the time of data collection)
  - Attitudes reflecting which technologies will become most popular in the future
  - Demographics (e.g., the gender gap among developers)

These findings are relevant particularly to current and aspiring developers, recruiters, educators, and policy makers.

INTRODUCTION

- Since 2011, the online programming knowledge sharing platform, Stack Overflow, has been conducting the Stack Overflow Annual Developer Survey.
- 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, 398; original dataset N ≈ 90,000).
- Audience: Developers (current and aspiring), HR professionals, educators, policy makers

METHODOLOGY

- **Data Source:** 2019 Stack Overflow Developer Survey
  - Link to Stack Overflow's annual survey data and results (Ctrl + Click)

• Data Wrangling: A portion of the dataset (provided by IBM) was loaded and cleaned using SQL and Python's pandas library.

• Cleaning procedure: Duplicates removal, data imputation, data normalization

• Link to dataset provided by IBM (Ctrl + Click)

• Analysis & Visualization: EDA and data visualization were conducted using various Python libraries and Cognos. Specifically, the following measures were examined:

• Technologies (i.e., languages, databases, platforms, and web frames) used in 2019

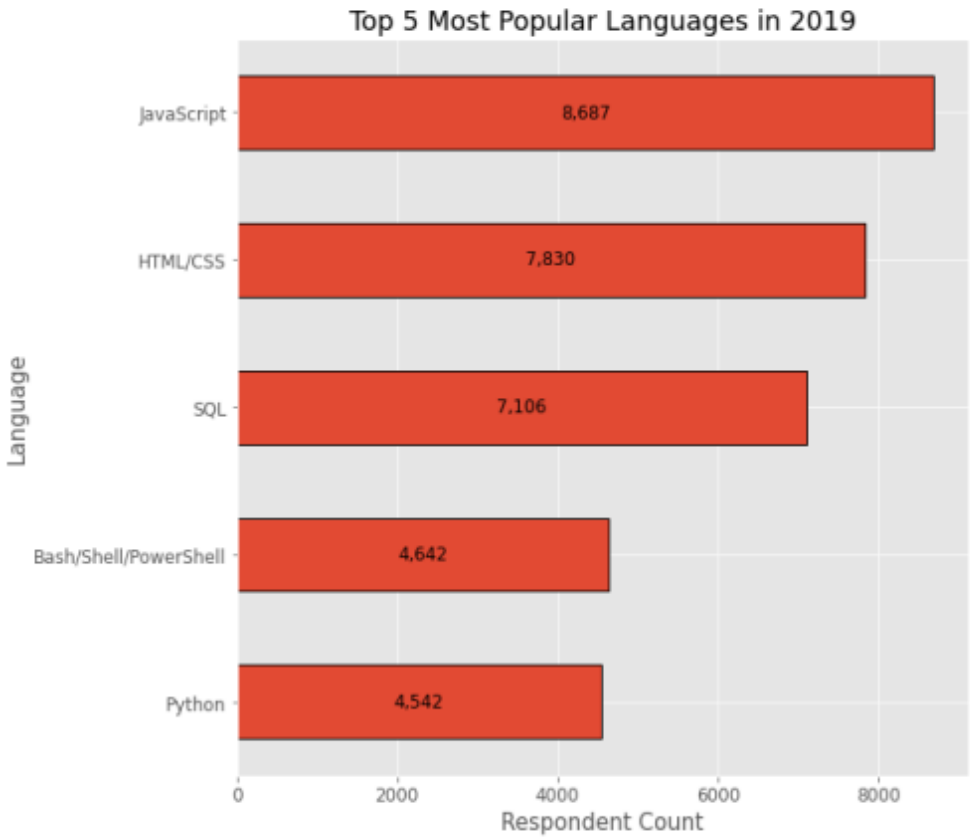
• Technologies most desired for the next year

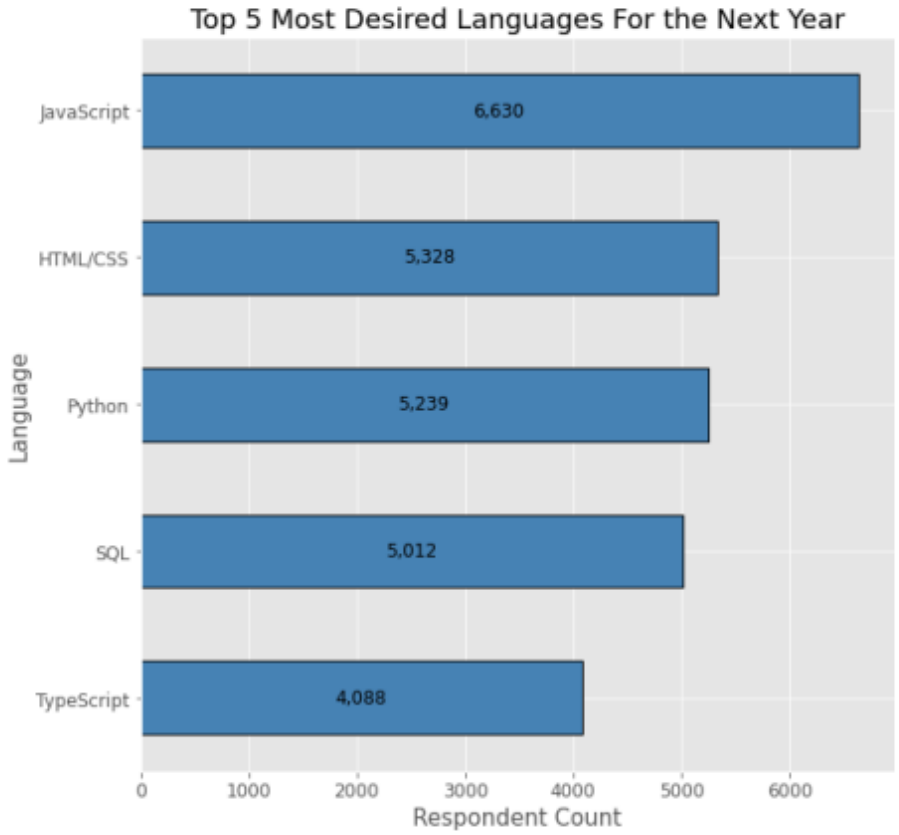
• Demographics (i.e., gender, country, age, and education).

RESULTS

- • First, descriptive statistics regarding demographics and total annual compensation were computed.
- • Sample size after compensation outlier removal: N = 10,519 (vs. before removal: N = 11,398).
- • The respondents had a median age of 29, and were predominantly male (i.e., 93.5% male vs. 6.5% female).
- • Median compensation: $52,704 USD per year.
- • Compensation and Age were positively correlated: r = .40.
- • Median compensation was higher for women: $54,956 than men: $52,339.

PROGRAMMING LANGUAGE TRENDS





PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

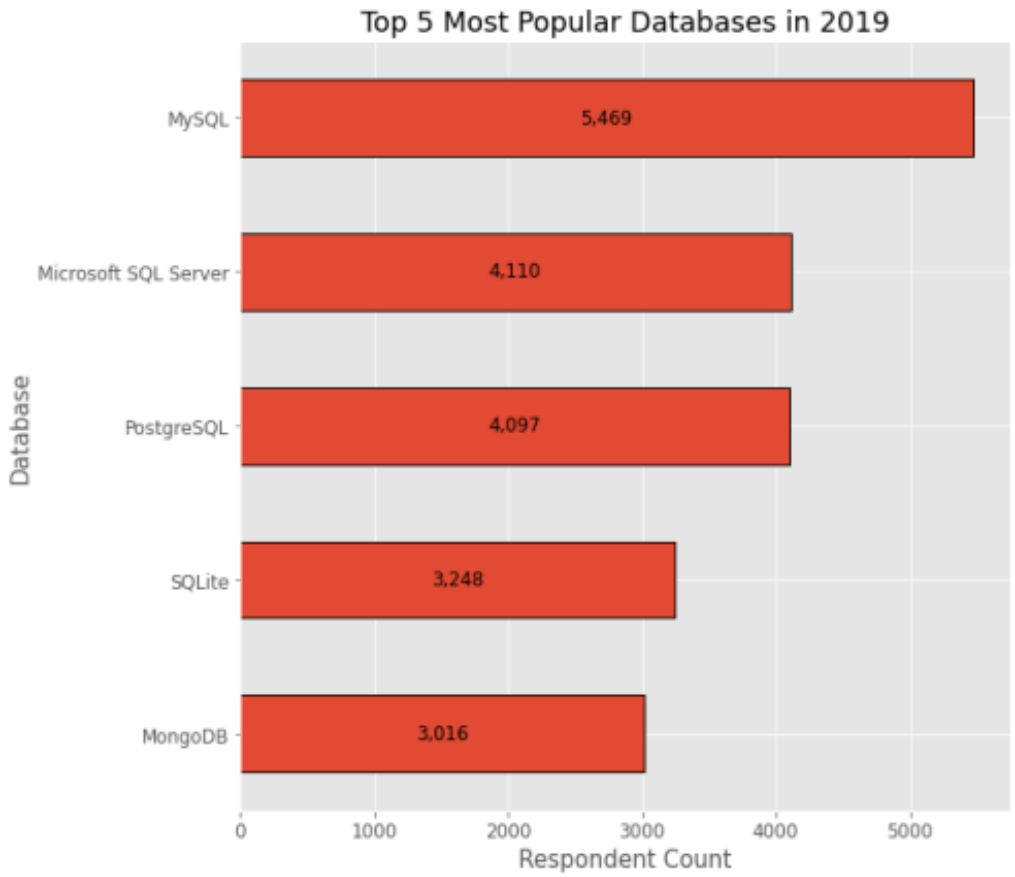
Findings

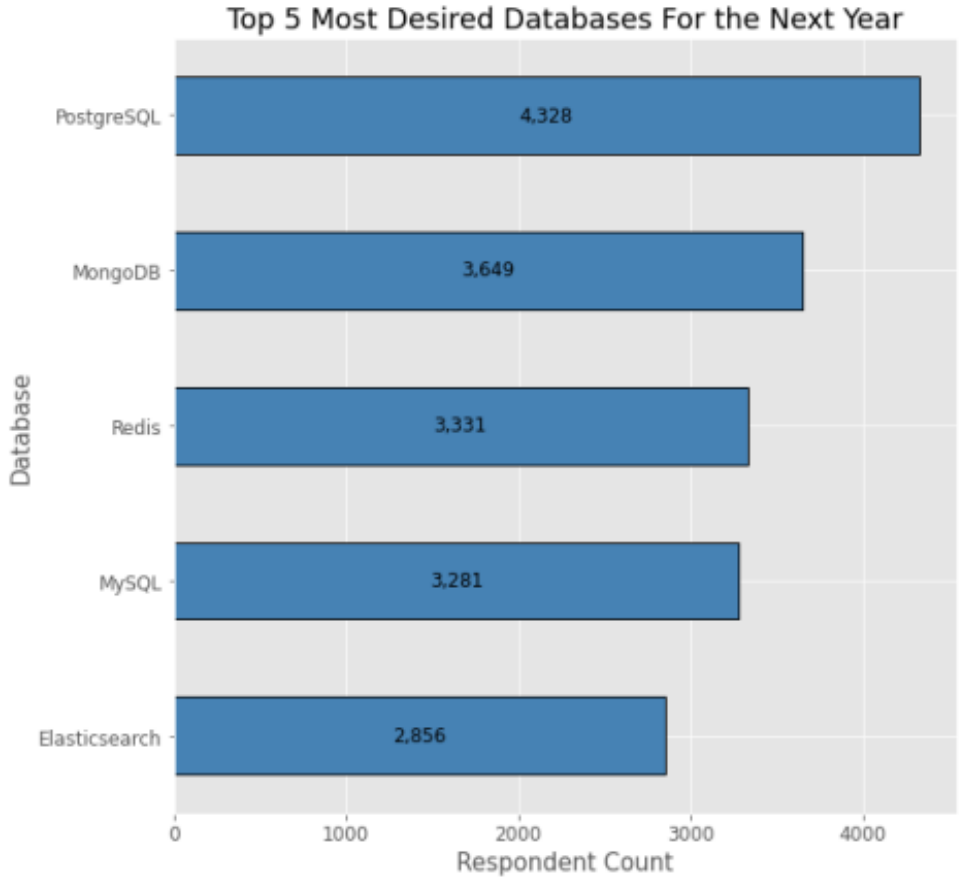
- *Javascript* and *HTML/CSS* were the most popular in 2019 and will likely remain so the following year.
- *SQL* was popular in 2019 and will likely remain so.
- Increasing interests in *Python* and *TypeScript*.
- Decreasing interest in *PowerShell/Bash*.

Implications

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

DATABASE TRENDS





PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

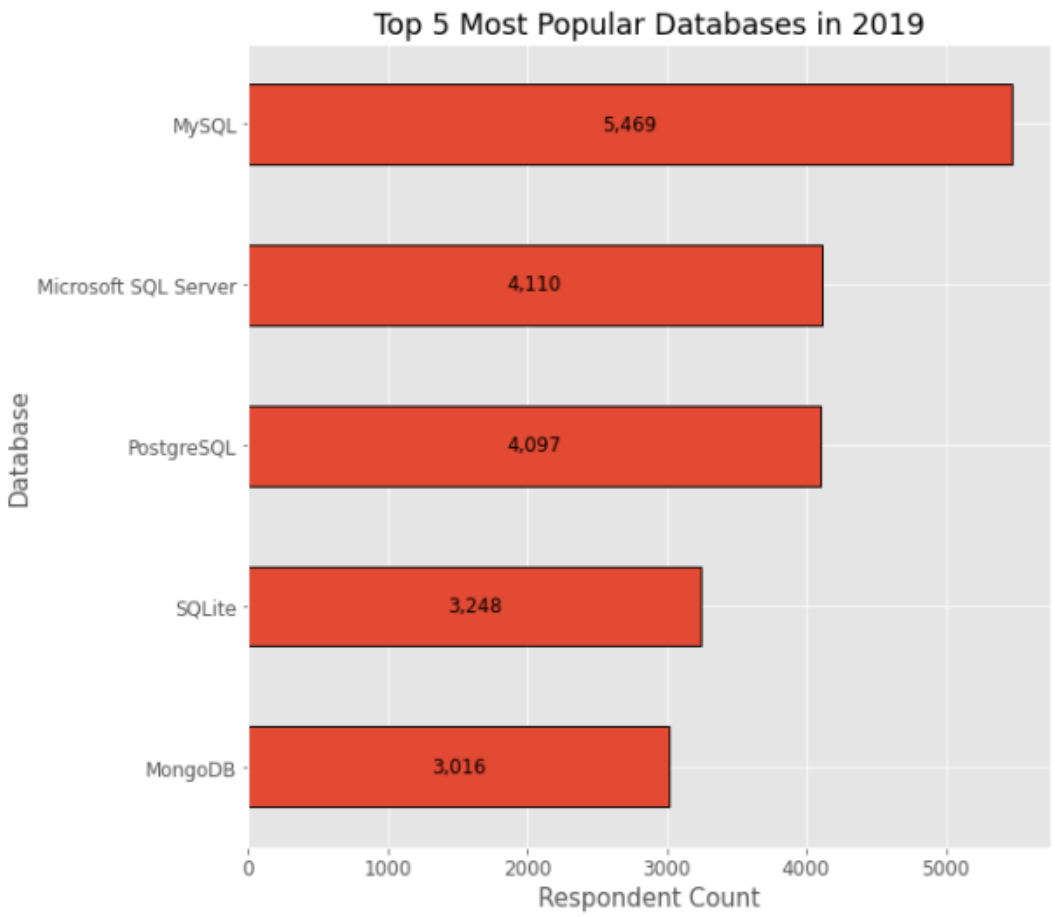
Findings

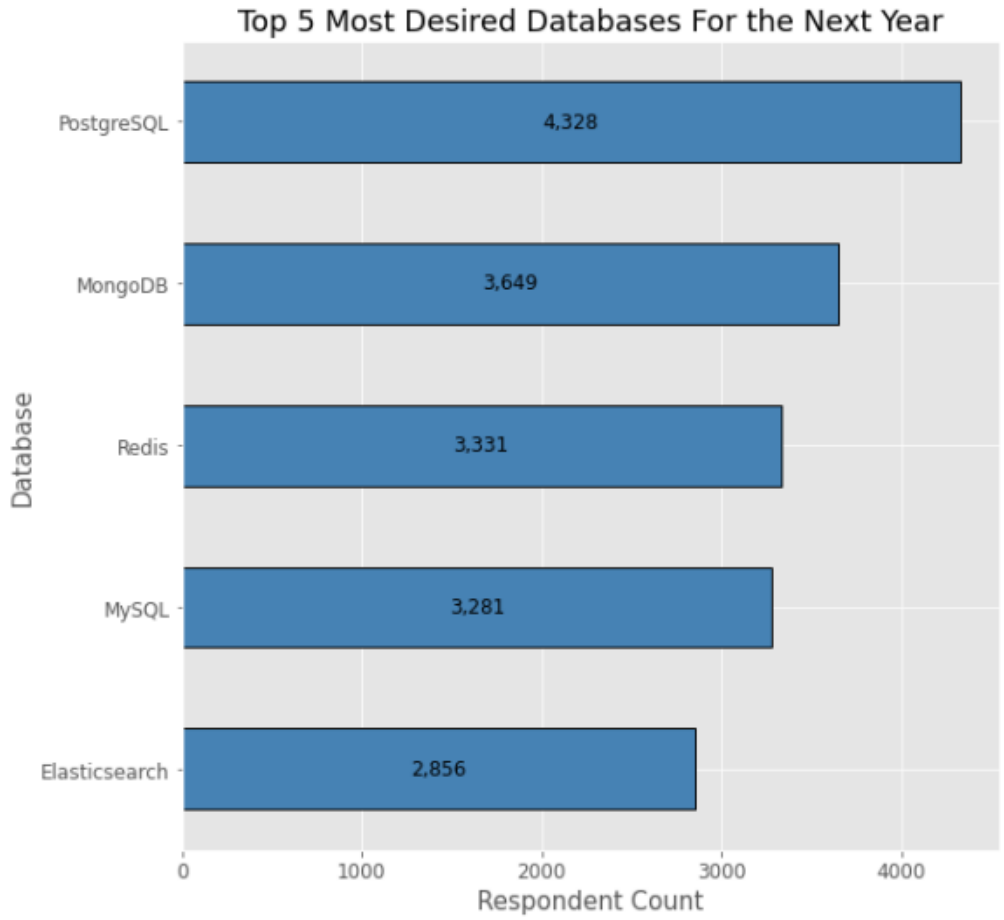
- *Javascript* and *HTML/CSS* were the most popular in 2019 and will likely remain so the following year.
- *SQL* was popular in 2019 and will likely remain so.
- Increasing interests in *Python* and *TypeScript*.
- Decreasing interest in *PowerShell/Bash*.

Implications

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

RESULTS: DATABASE TRENDS





DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

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

Implications

- There appears to be increasing developer preference toward 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.

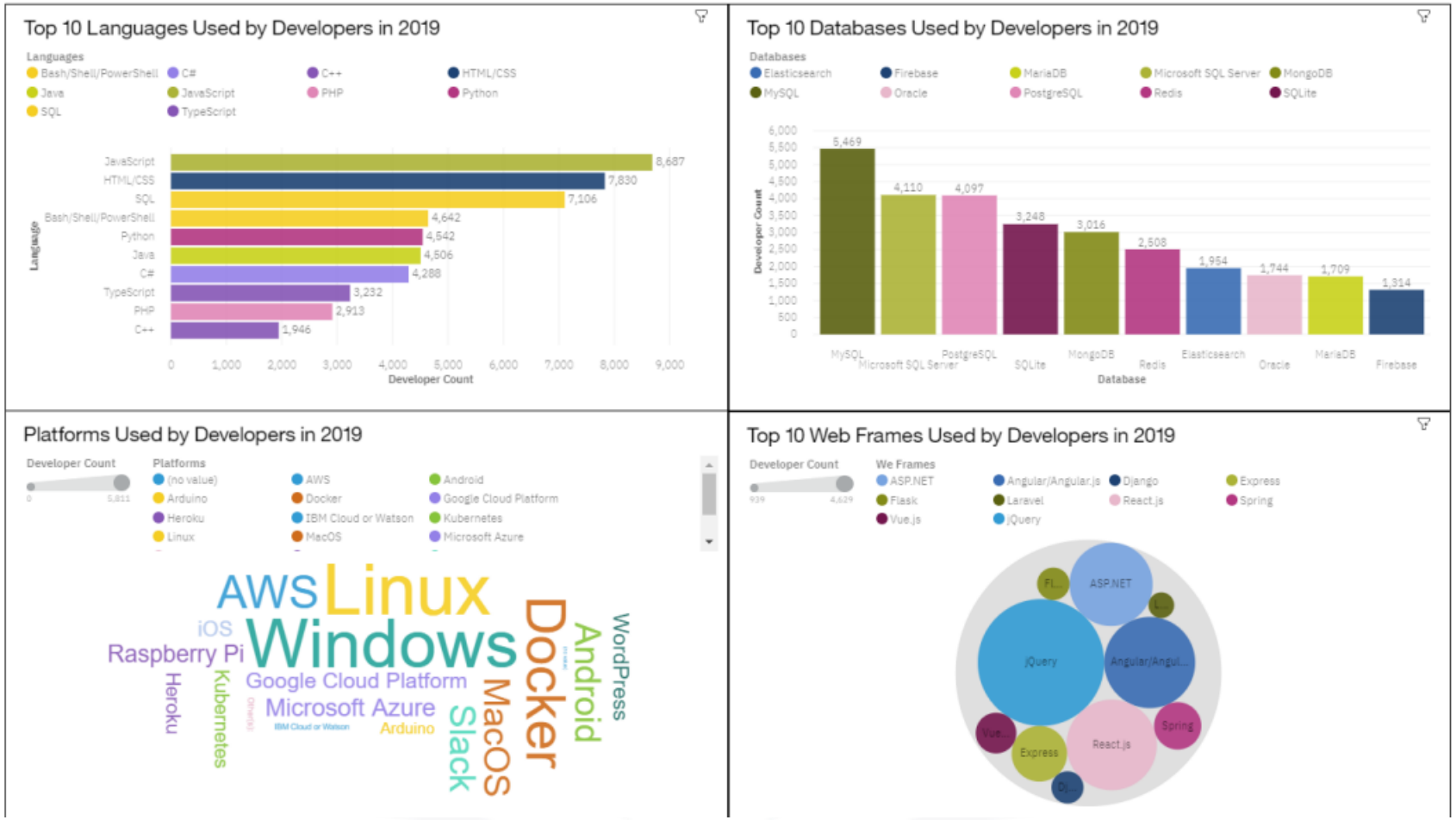
DASHBOARD

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

Click here to open the dashboard (Ctrl+Click)

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

DASHBOARD A: CURRENT TECHNOLOGY USE



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 *Javascript* and *HTML/CSS* remain high usage. There's also increasing interest in *Typescript*.
- 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.
- A severe gender representation gap (in favor of men), despite median compensation being slightly higher for women.
- 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.

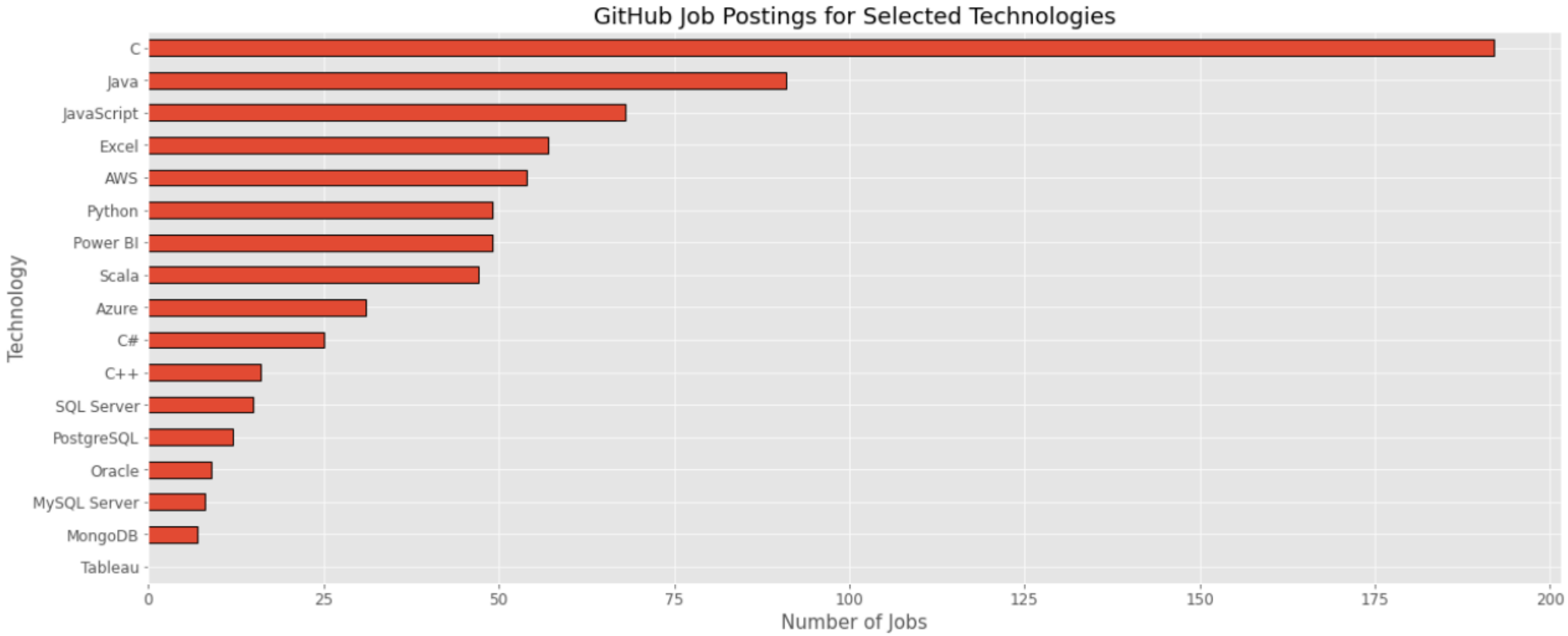
CONCLUSION

A subset of data collected as part of the *2019 Stack Overflow Developer Survey* was examined.

The findings yielded numerous insights into the technologies most used and desired by developers in addition to the developer demographic.

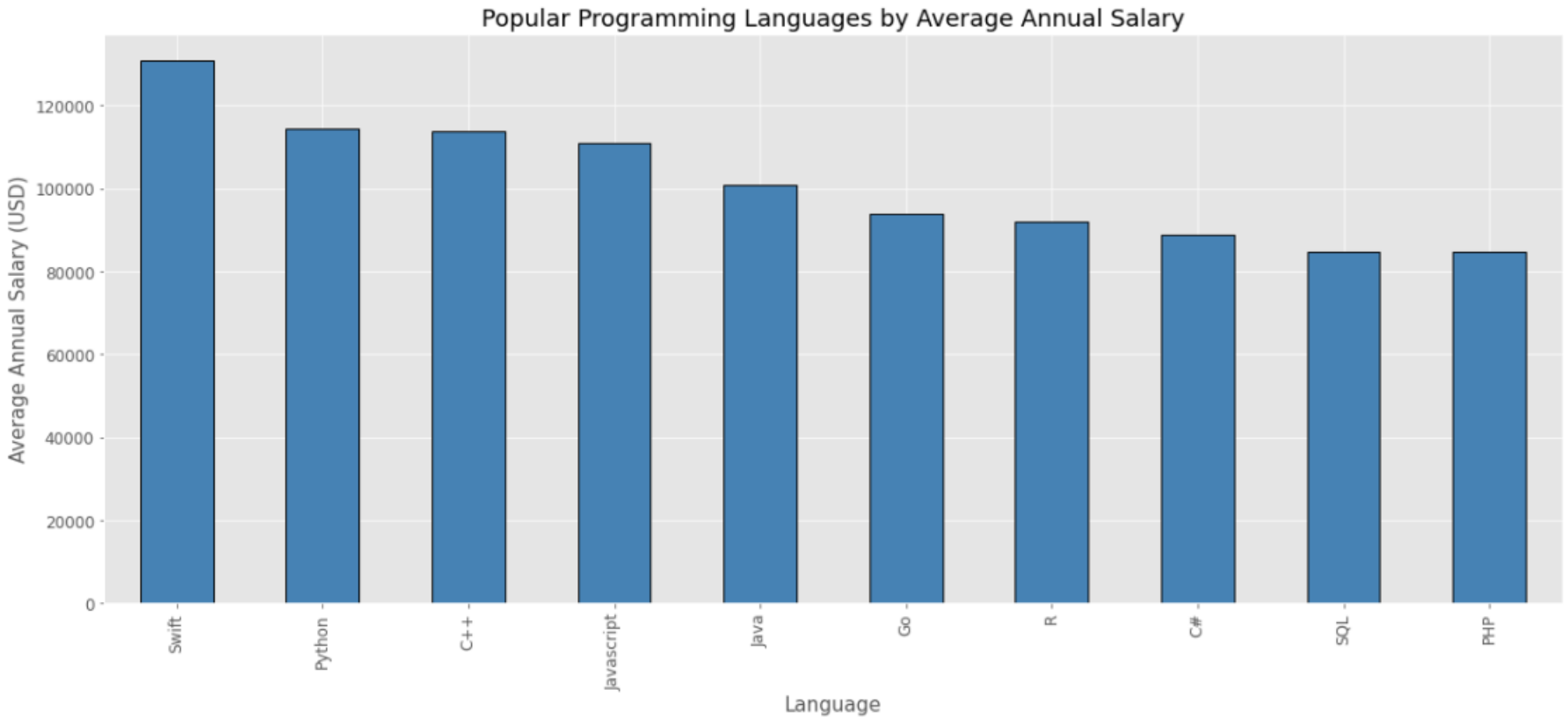
These insights should be particularly relevant for current and prospective developers aiming to remain competitive, businesses aiming to upskill their talent, educators in the field, and policy makers aiming to address gender and economic issues.

APPENDIX A: GITHUB JOB POSTINGS FOR SELECTED TECHNOLOGIES



Note: Data on *GitHub* job postings for the 15 selected technologies shown above were collected using the *GitHub Jobs* API on May 3, 2021.

APPENDIX B: POPULAR LANGUAGES BY SALARY



Note: The figure above is based on survey data regarding popular programming languages provided by IBM. Link: [Popular Programming Languages].