



Desirable Data Science Skills

Malcomb C Brown

Sept. 10, 2021

OUTLINE



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

EXECUTIVE SUMMARY



- Scaling up our Data Processing
- The need to build a Data Science Team
 - Identify current data infrastructure
 - Establish Database type and structure
 - Access relevant skills required
- Database Knowledge
- Relevant Programming Languages
- Platforms & Web Frameworks

INTRODUCTION



- The technological advances in data, require us to change the way we utilize our data.
- In order to maximize our leverage, we need to building out a Data Science Team
- Identify the prominent skill sets will allow Management to better evaluate candidates.
- The 4 Foundational Pillars
 - Programming language
 - Database
 - Platform
 - Framework

METHODOLOGY



- Identify widely known & used Programming Languages from Survey Data
 - What are the top languages now ?
 - What is the future trend?
- Compare the various Databases from Survey Data
 - Which Databases offer scalability?
- Compare Technology Platforms for used by Survey Respondents
- Get sample job posting information from Github
 - Where are the majority for?
 - Developing an expected salary range

The background features a large, faint graphic of two hands, palms up, holding a network of interconnected nodes. The nodes are represented by small, rounded squares in various colors (pink, orange, green, blue) connected by thin grey lines, forming a complex web-like structure. The word "RESULTS" is centered over this graphic.

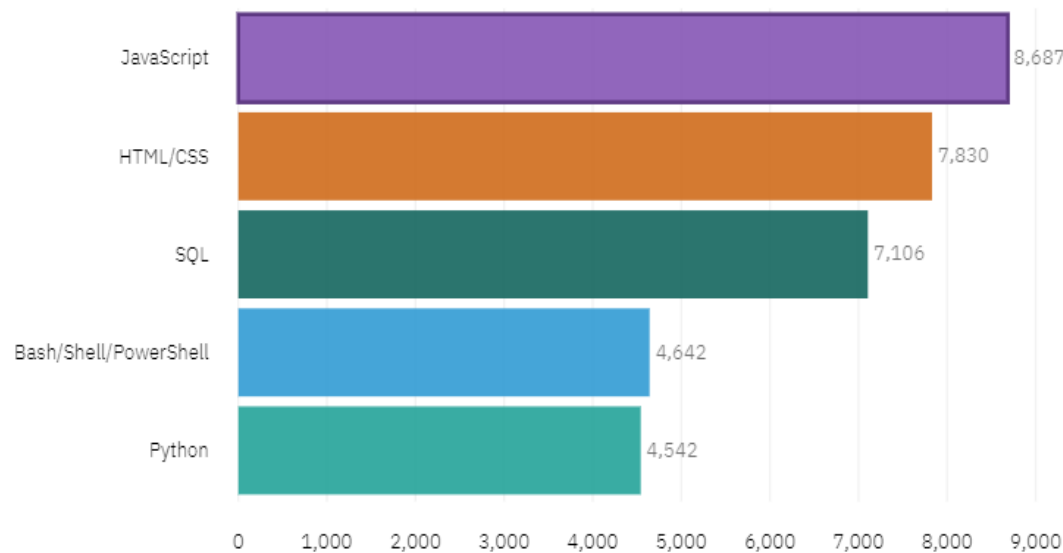
RESULTS

© IBM Corporation. All rights reserved.

PROGRAMMING LANGUAGE TRENDS

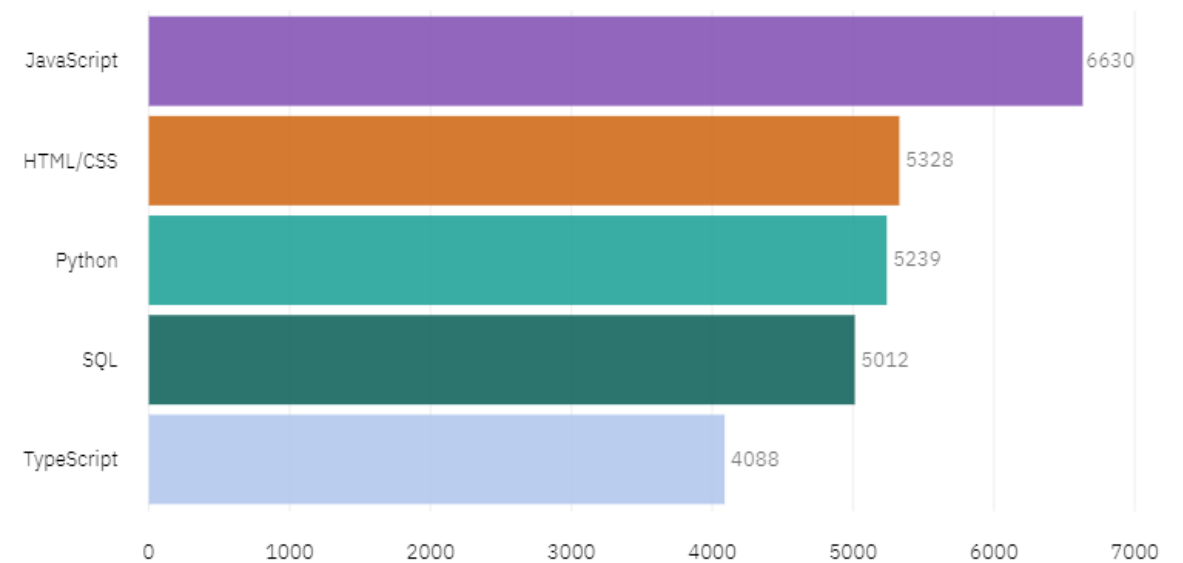
Current Year

Top 5 Language Worked With



Next Year

Top 5 Language Desired Next Year



PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

- JavaScript & HTML/CSS are the most popular.
- Python is growing in popularity
- SQL is essential for database management language

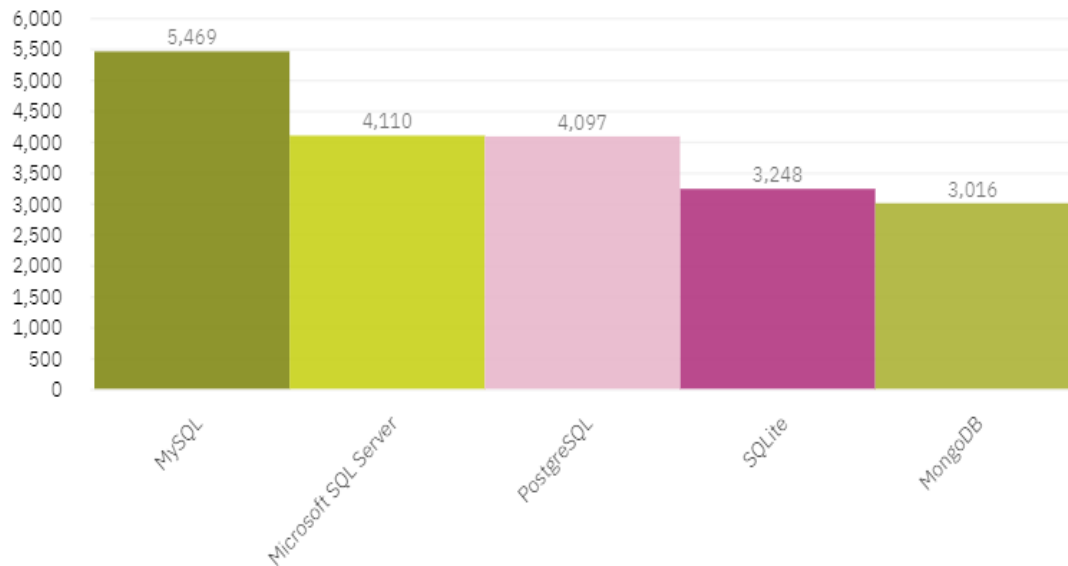
Implications

- Highest potential as a listed skill by an applicant
- Knowledge in several languages is a must
- SQL is a definite need

DATABASE TRENDS

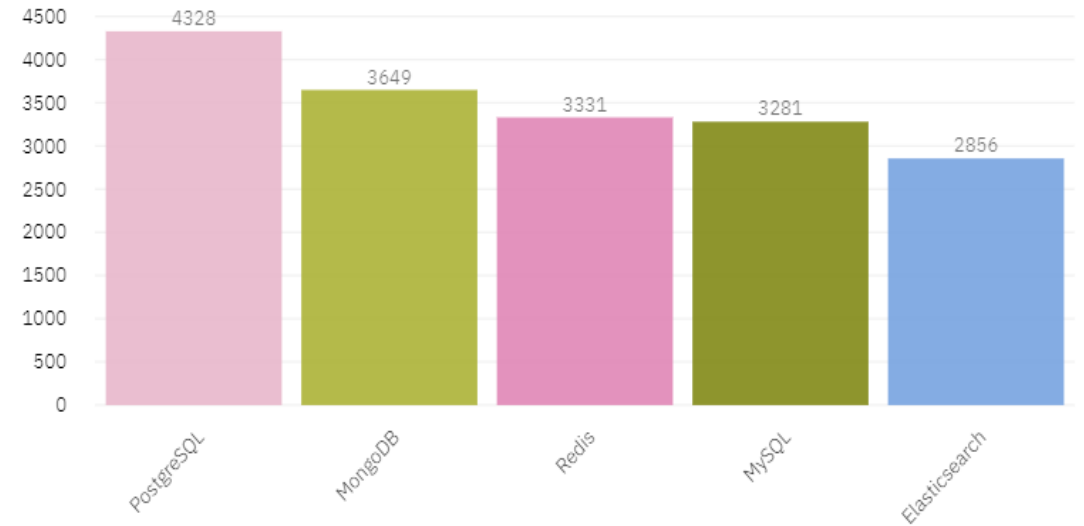
Current Year

Top 5 Database Worked With



Next Year

Top 5 Database Desired Next Year



DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- MySQL and PostgreSQL are among the most used databases
- MongoDB is on the rise
- Knowledge of multiple databases

Implications

- Knowledge of relational databases is prevalent
- Database will depend on the type of data to be stored
- The ability to work with structured and unstructured data

DASHBOARD

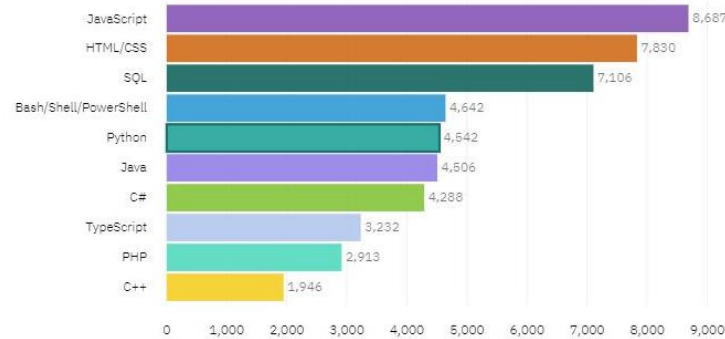


Access the full interactive dashboard by click link below.

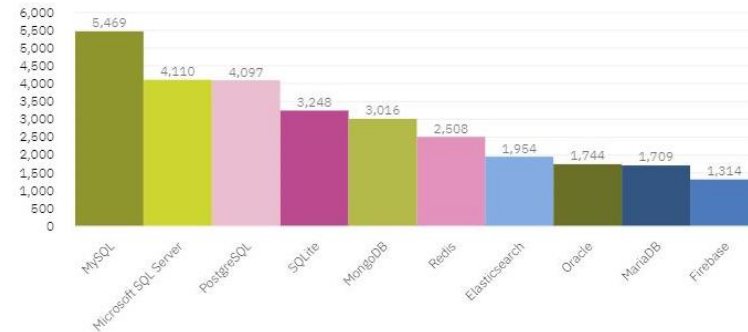
<https://dataplatform.cloud.ibm.com/dashboards/a12574d1-952b-4208-a00a-033018b34b43/view/551cd23a67ac3cea7fe7c8e407cd2f0f7a332d08b1bb8b0084d37b490f367897f36c15c7c87d1e5cd2120464f0ee4758cc>

Current Technologies Being Used

Top 10 Language Worked With



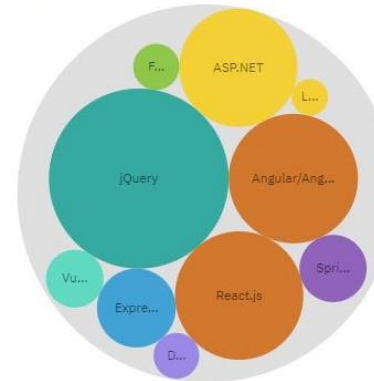
Top 10 Database Worked With



Platform Worked With Word Cloud

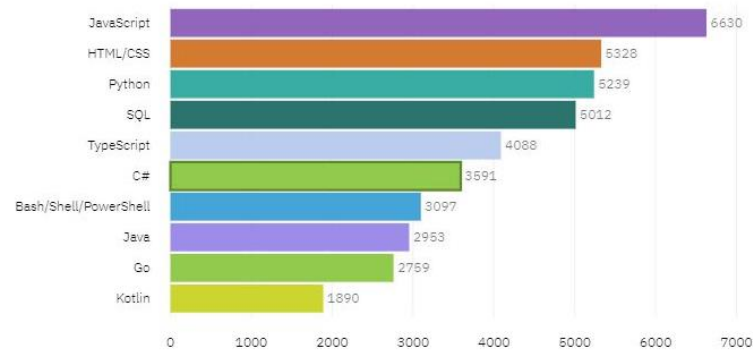


Top 10 WebFrame Worked With

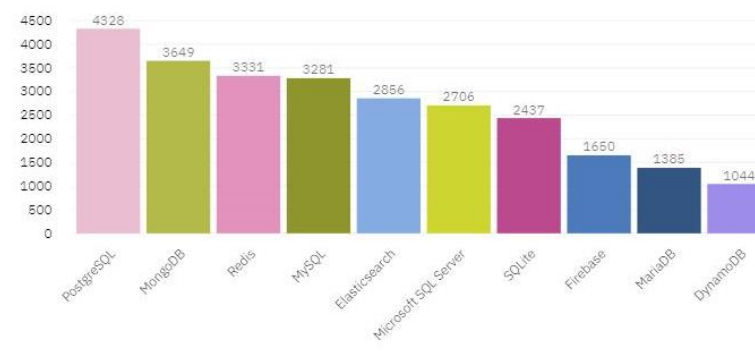


Technologies Desired in the Future

Top 10 Language Desired Next Year



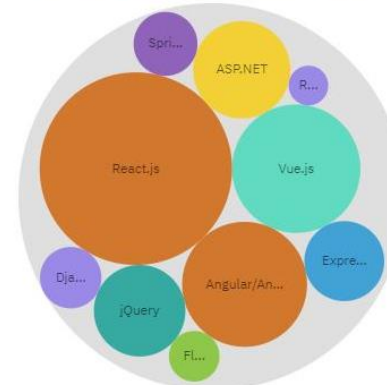
Top 10 Database Desired Next Year



Platforms Desired Next Year



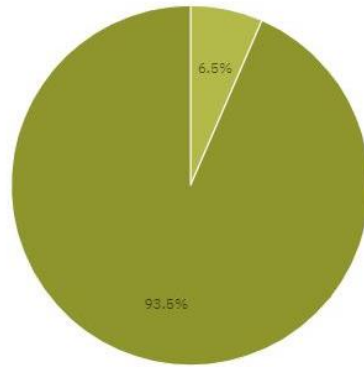
Top 10 WebFrame Desired Next Year



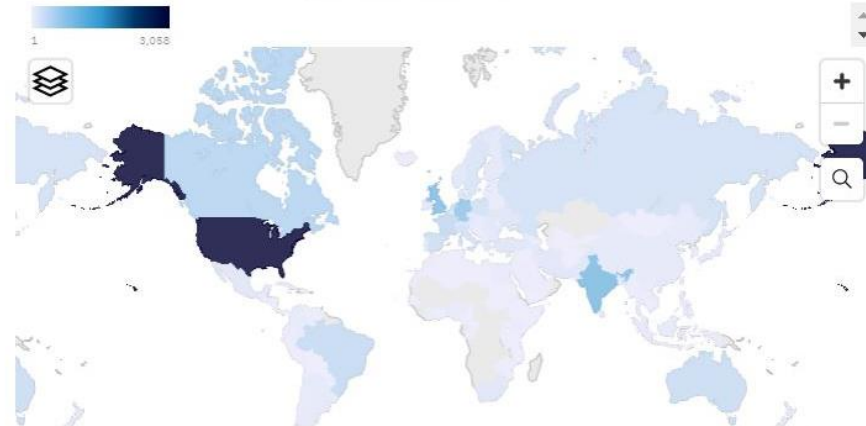
Demographics (Male & Female)

Respondents Classified by Gender

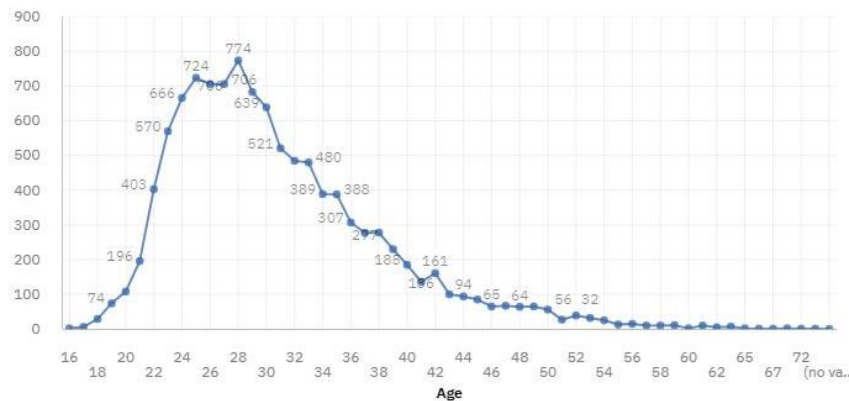
Gender
● Woman ● Man



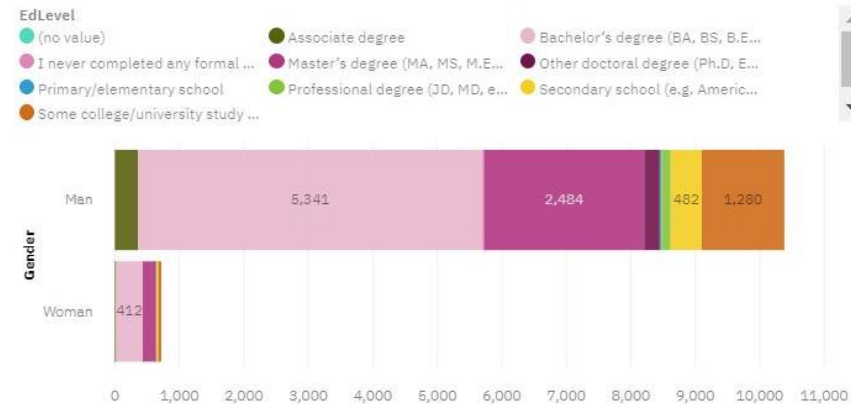
Respondents by Country



Respondent Count by Age



Formal Education Level by Gender



DISCUSSION



- What type of Database is best for us?
- Can we stay up to date with trends?
- Is the Language important?
- Can we build our own platform?

OVERALL FINDINGS & IMPLICATIONS

Findings

- SQL is one of the most popular languages
- Python is rapidly growing in use and popularity
- MySQL & PostgreSQL are 2 of the most popular Relational Databases in use.

Implications

- SQL is required to work with relational databases & should be a prerequisite
- Python's wider adaption and ease of use will allow for easier adoption
- Having the ability to utilize both SQL and another language(i.e. Python) will be increasingly necessary

CONCLUSION



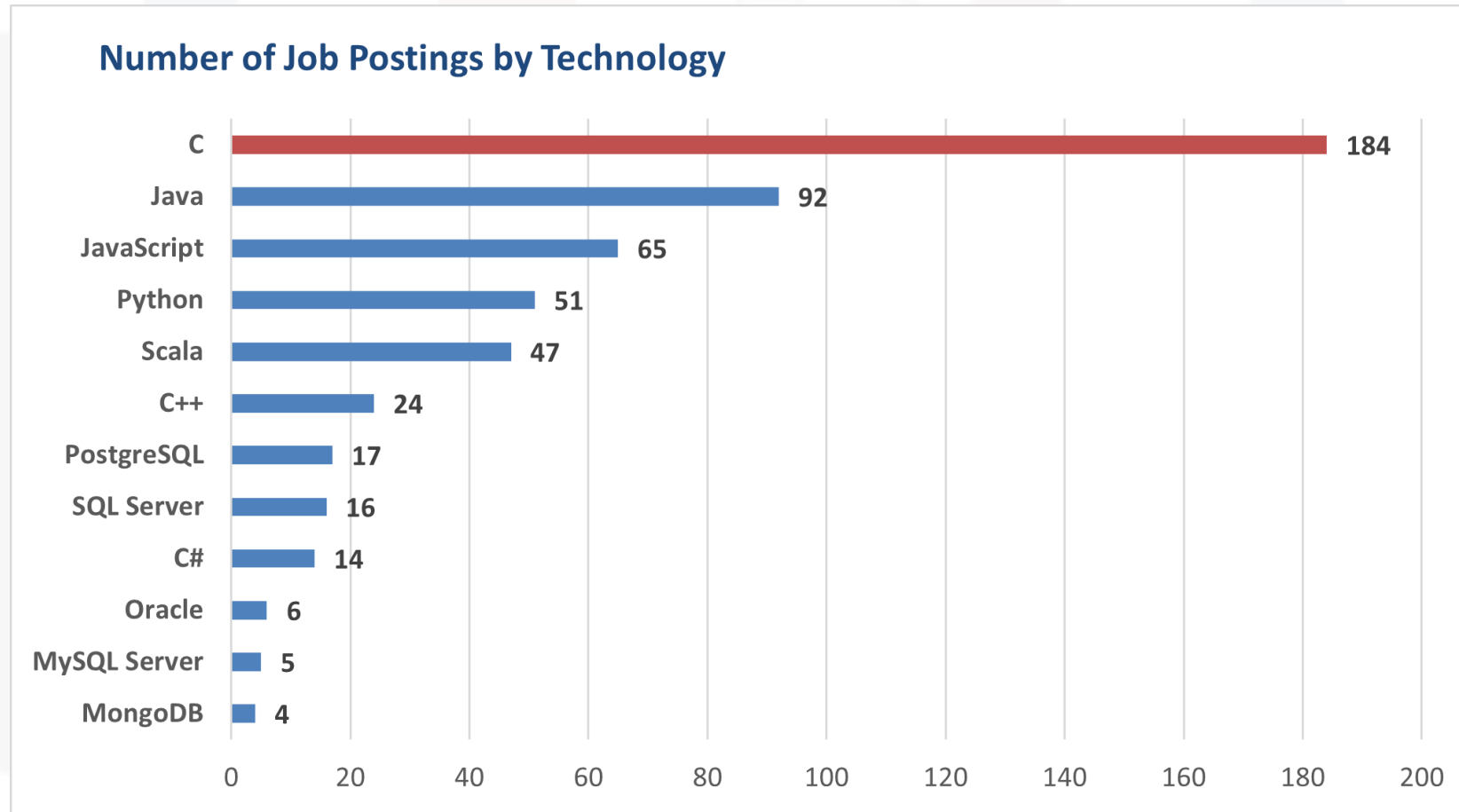
- SQL is a requires language skill, as it interacts with the database
- JavaScript, HTML/CSS, and Python should be highly sought after
- MySQL and PostgreSQL should be evaluated as a Database option
 - MongoDB, also, if Machine Learning is going to be a focus
- Establishing a solid team can be accomplished as each applicant will have previously acquired several of the prerequisite skillsets

APPENDIX

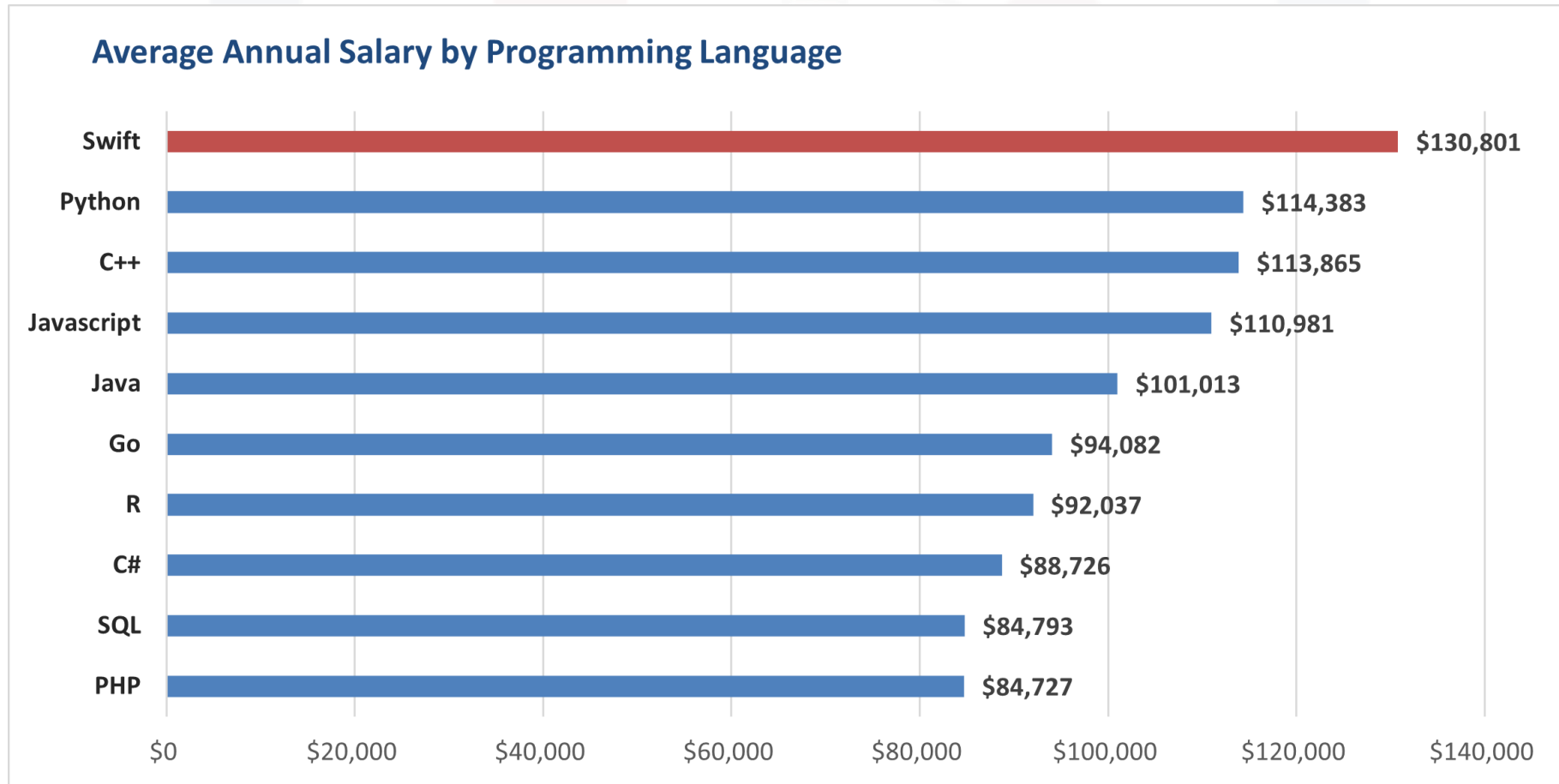


- GitHub Job Posting Bar Chart
- GitHub Avg. Annual Salaries base on Programming Language
- Language Comparison(Current Yr. vs. Next Yr.)
- Database Comparison(Current Yr. vs. Next Yr.)

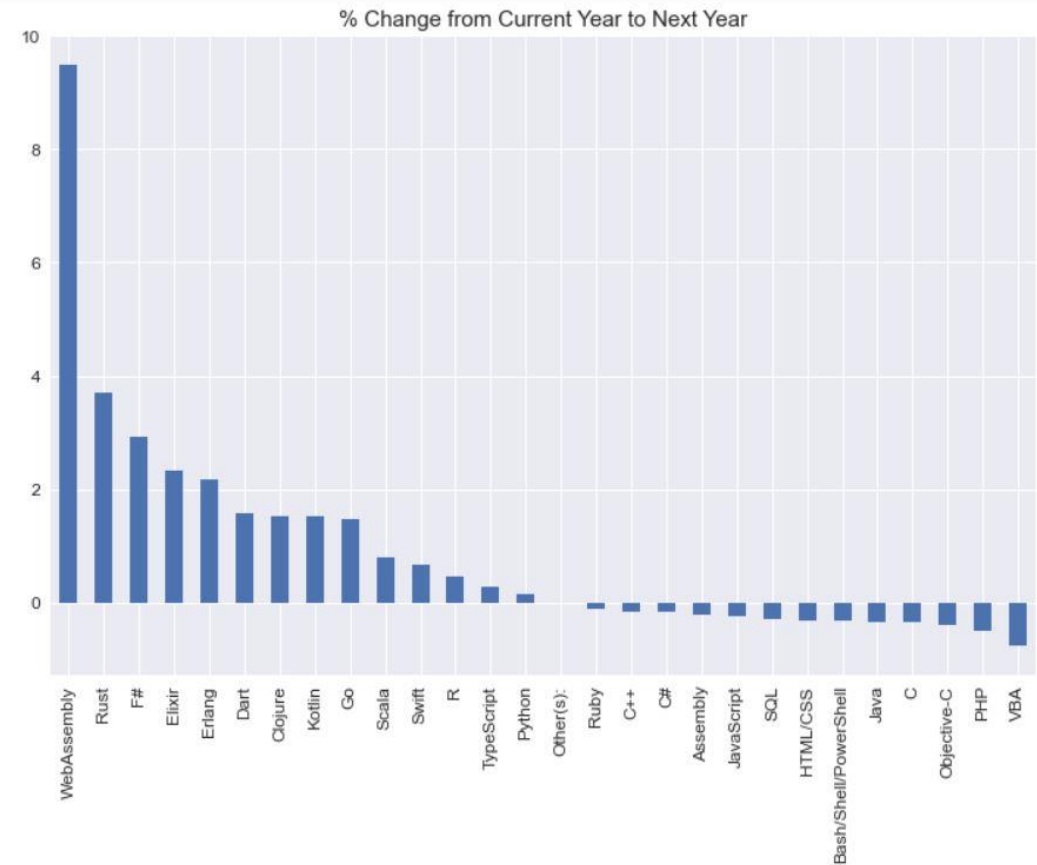
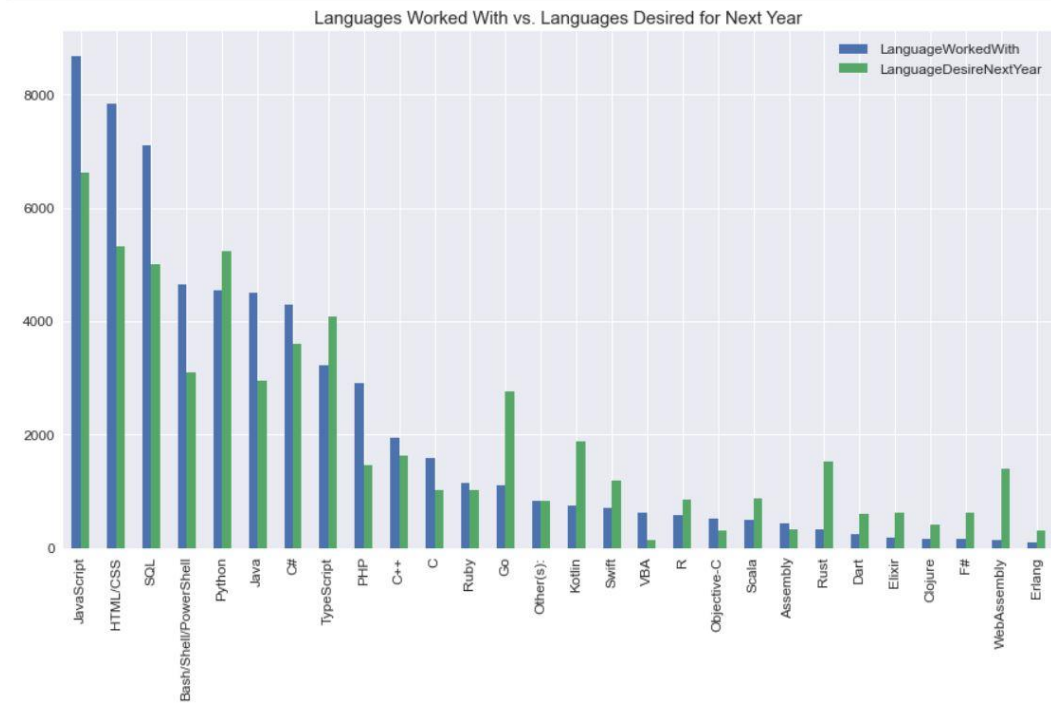
GITHUB JOB POSTINGS



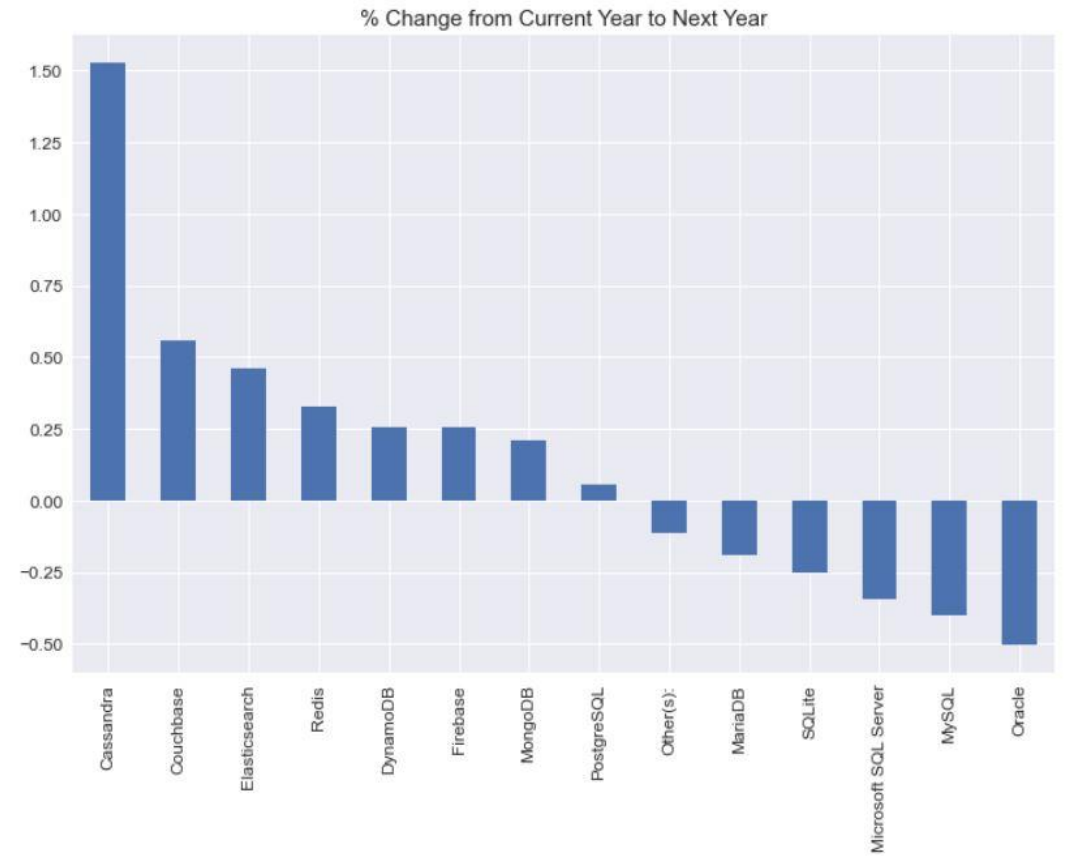
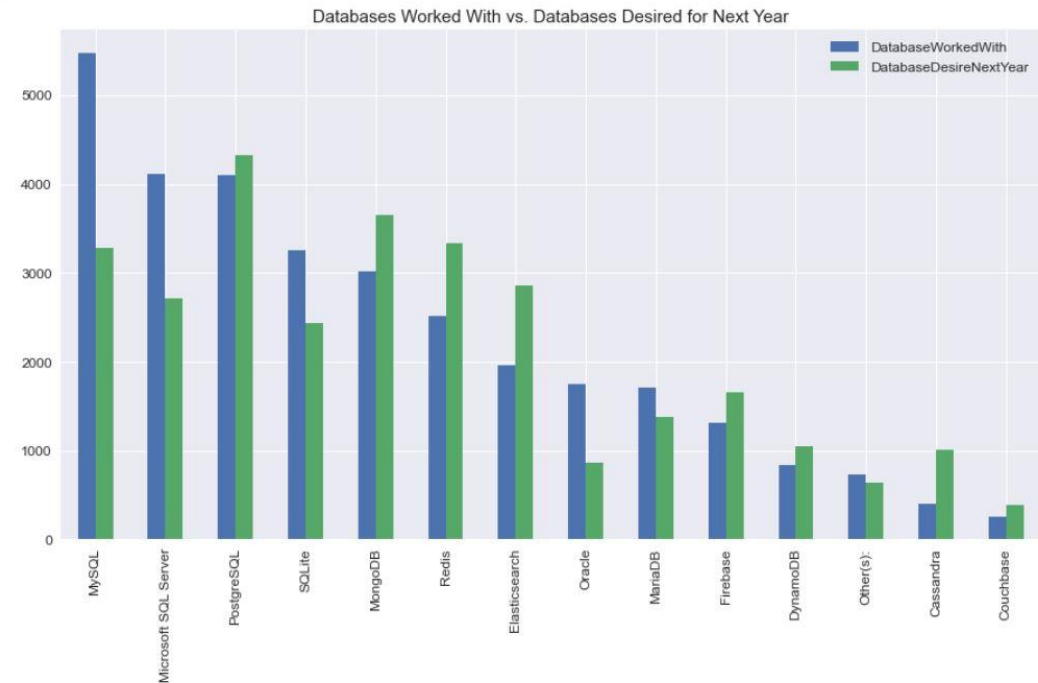
POPULAR LANGUAGES



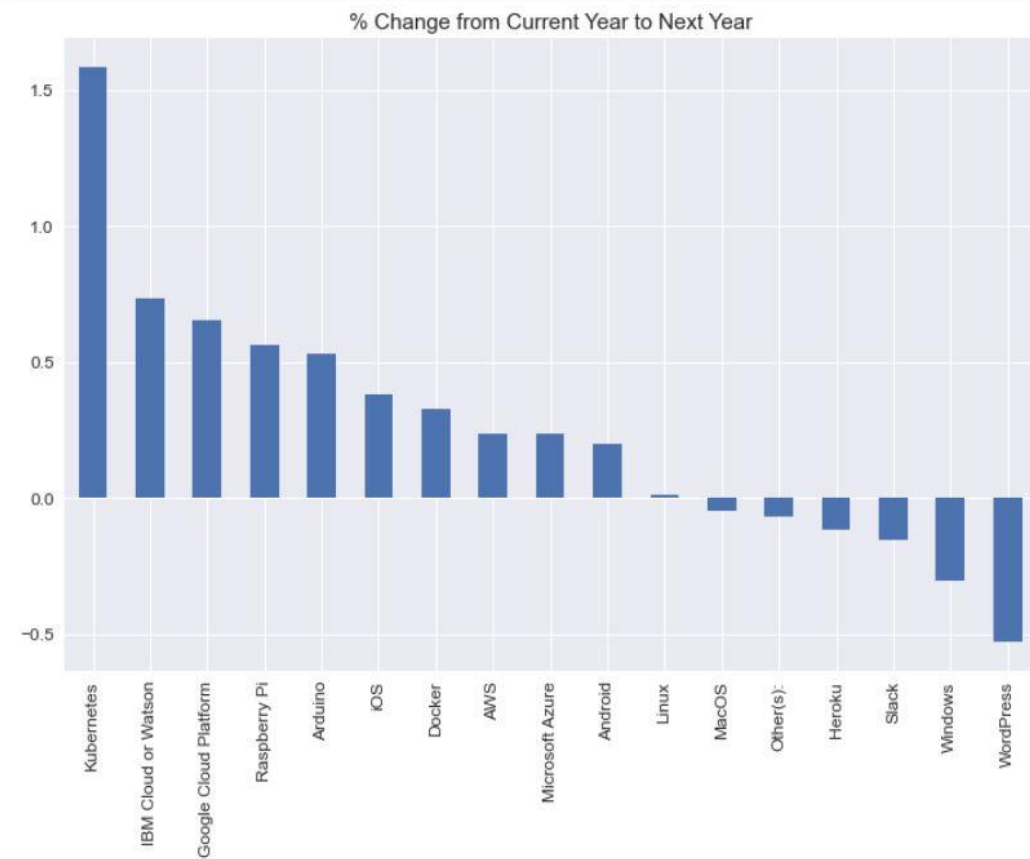
Language Comparison



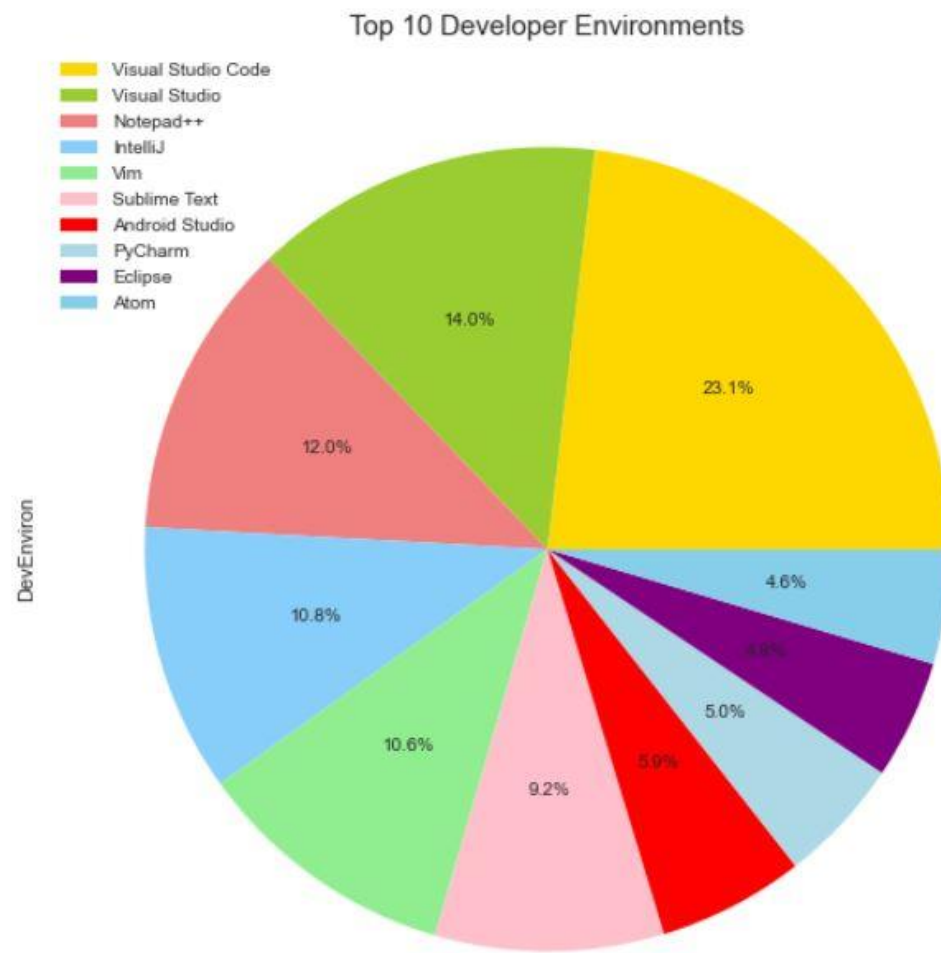
Database Comparison



Platform Comparison



Top 10 Developer Environments



Related Articles



<https://www.northeastern.edu/graduate/blog/data-analyst-skills/>

<https://bootcamp.cvn.columbia.edu/blog/data-analyst-skills/>

<https://www.indeed.com/career-advice/resumes-cover-letters/skills-for-data-analyst>

<https://www.stitchdata.com/resources/data-analysis-tools/>

<https://www.jigsawacademy.com/10-most-popular-analytic-tools-in-business/>

<https://www.proschoolonline.com/blog/top-10-data-analytics-tools>