



Data Analysis (*Capstone Project*)

Abid Hasan

Feb 27, 2022

OUTLINE



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

EXECUTIVE SUMMARY



- Javascript, HTML/CSS and Python are top ranked programming languages and will persist next year as well
- PostgreSQL, MongoDB and MySQL are top ranked databases and will persist next year as well

INTRODUCTION



- This report is result of Exploratory Data Analysis on a dataset which was generated by a survey about the technical skills mostly used by the developers and the skills that will persist in future
- Students, Developers, Programmers and IT firms can be benefitted from this report
- Based on the report, one can prepare oneself to learn programming languages and databases that are highly used and will be demanded in the next year

METHODOLOGY



- Data sources:
 - Stack Overflow Developer 2019 Survey
 - GitHub Job Postings
 - Programming Languages Annual Salary
- Techniques:
 - Exploratory data analysis using Python libraries and SQL
 - Reading SQL, CSV and JSON files
- Visualizations:
 - Pandas charts: barplot, histogram, lineplot, scatter plot
 - IBM Cognos Dashboards for interactive visuals

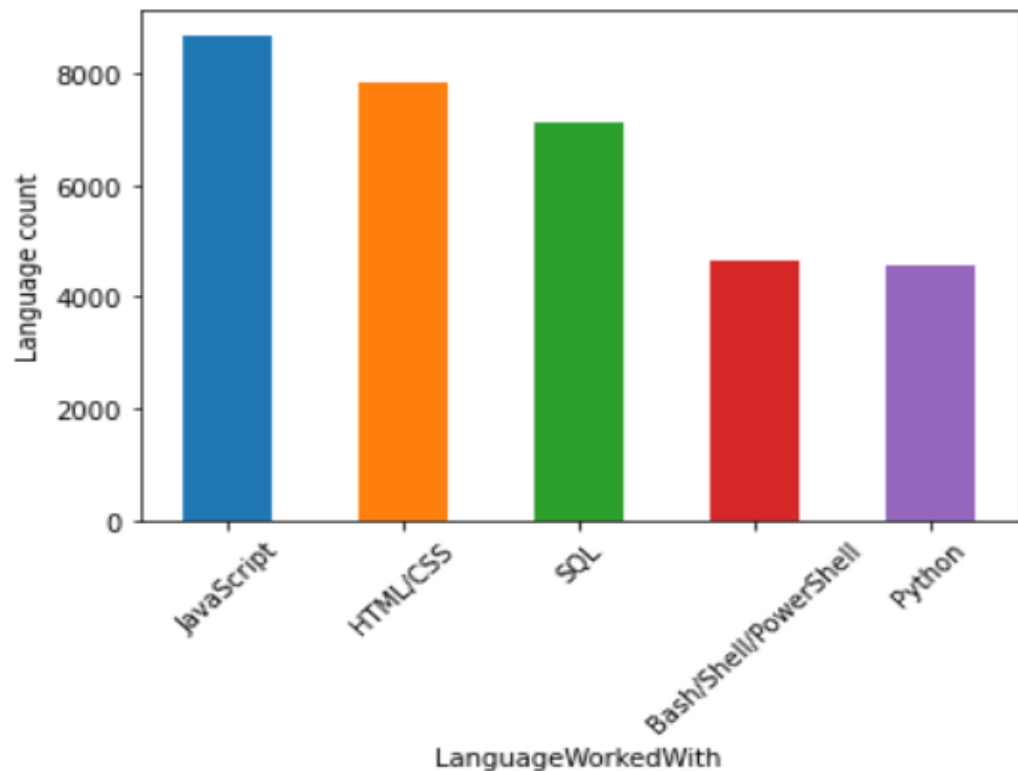
RESULTS



PROGRAMMING LANGUAGE TRENDS

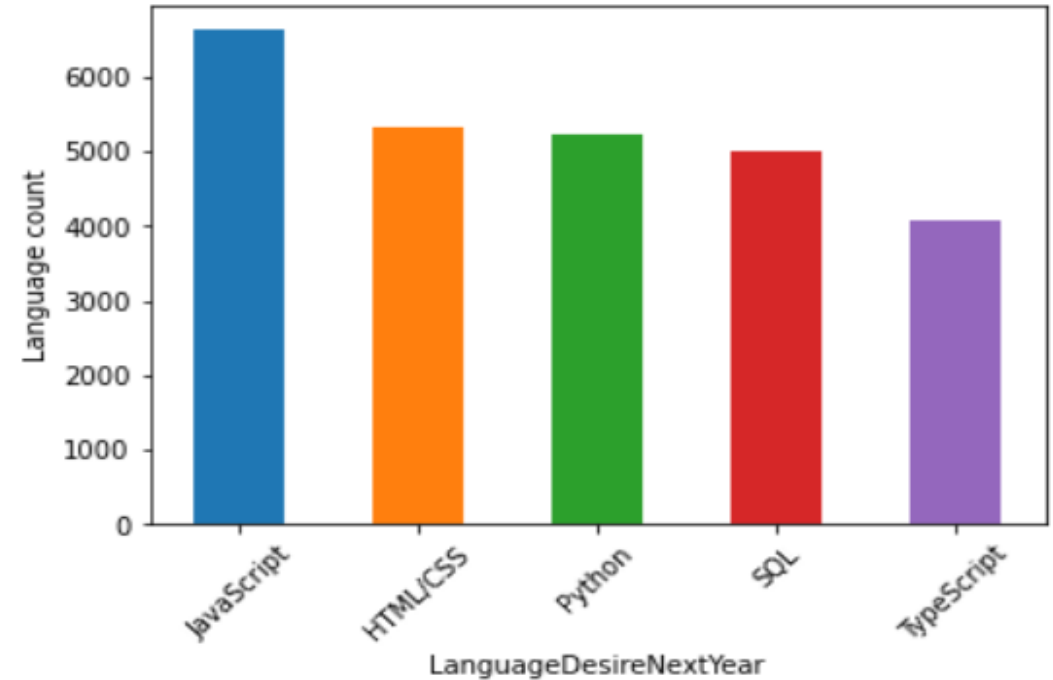
Current Year

Top 5 Languages worked with



Next Year

Top 5 Languages desired next year



PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

- Javascript and HTML/CSS are ranked as top 2 currently and will rank the same next year
- Python ranked 3rd in the demanded languages
- Bash/Powershell has lost its rank from the top 5 languages

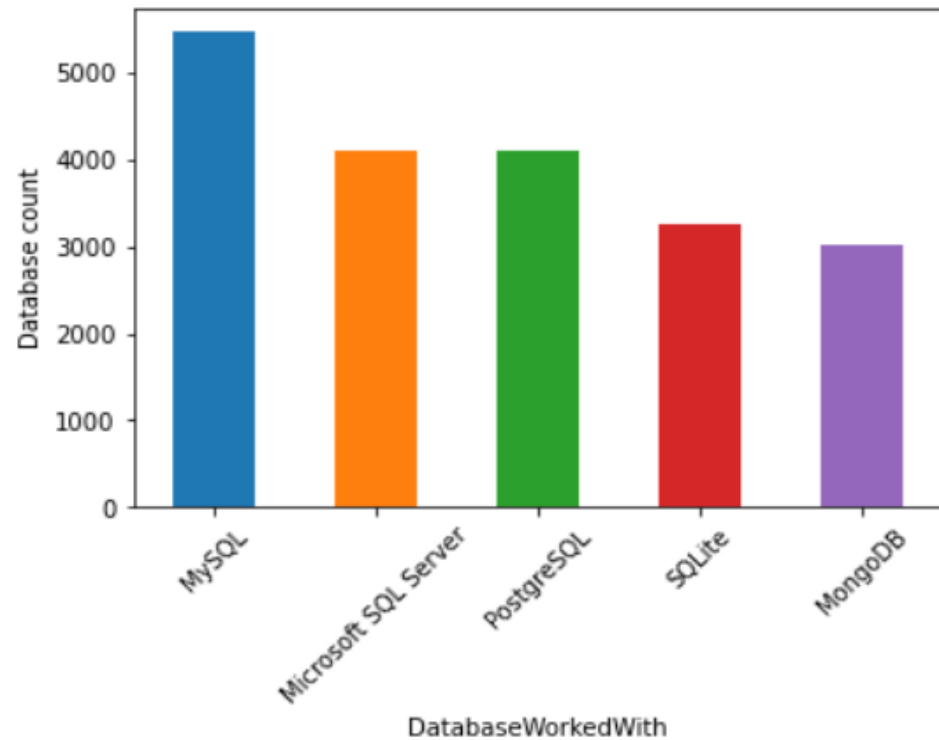
Implications

- Javascript and HTML/CSS are famous and must be learnt as they will be demanded next year in top5 as well
- Python is also recommended as ranks in the top languages for next year

DATABASE TRENDS

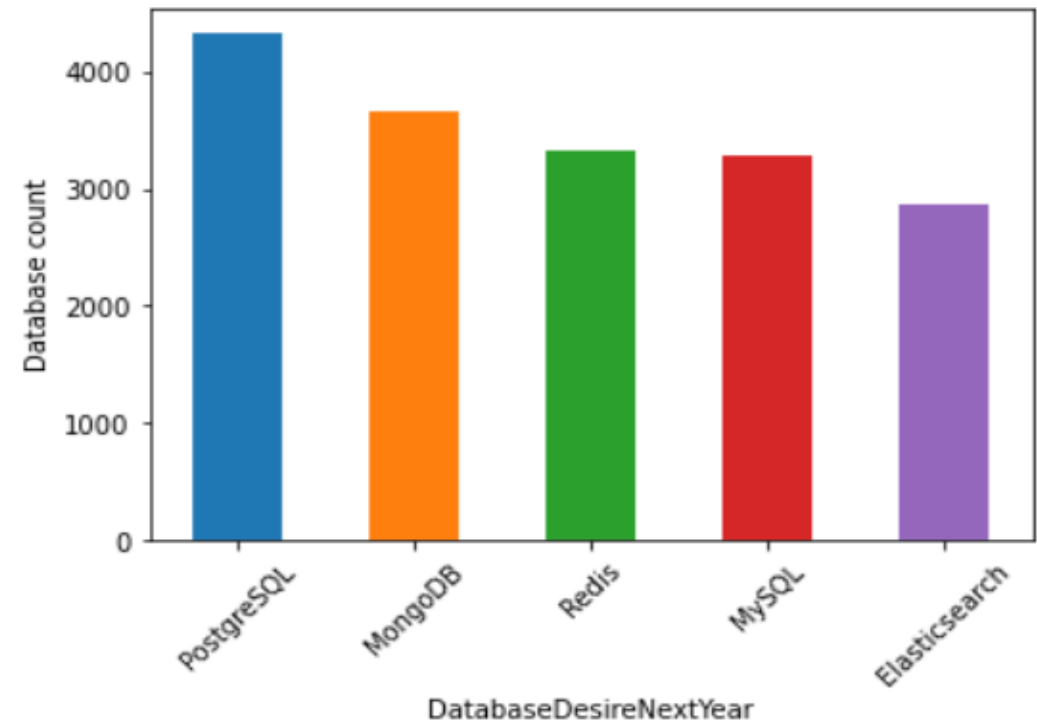
Current Year

Top 5 Databases worked with



Next Year

Top 5 Databases desired next year



DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- MySQL, PostgreSQL, MongoDB rank in the top 5 demanded databases next year as well
- MS SQL Server, SQL Lite lost there rank from the top 5 languages
- Redis and ElasticSearch will be demanded as two of the top 5 databases next year
- PostgreSQL will be the highest demanded language next year

Implications

- PostgreSQL is a database for future, it must be learnt
- Redis and ElasticSearch are also recommended as they will rank high in the top demanded databases for next year

DASHBOARD

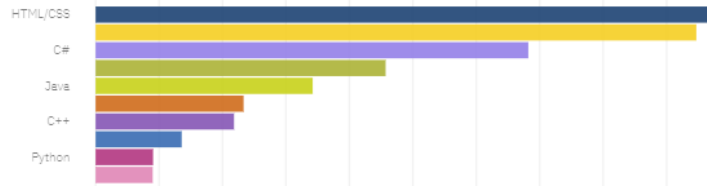


[IBM COGNOS: Dashboard link](#)

DASHBOARD TAB 1(Current Technology Trends)

Top 10 Languages Used

LanguageWorkedWith



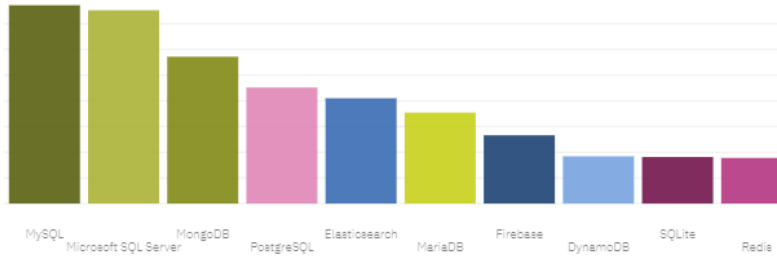
Platforms Used Frequently

Respondent (Sum)



Top 10 Databases Used

DatabaseWorkedWith

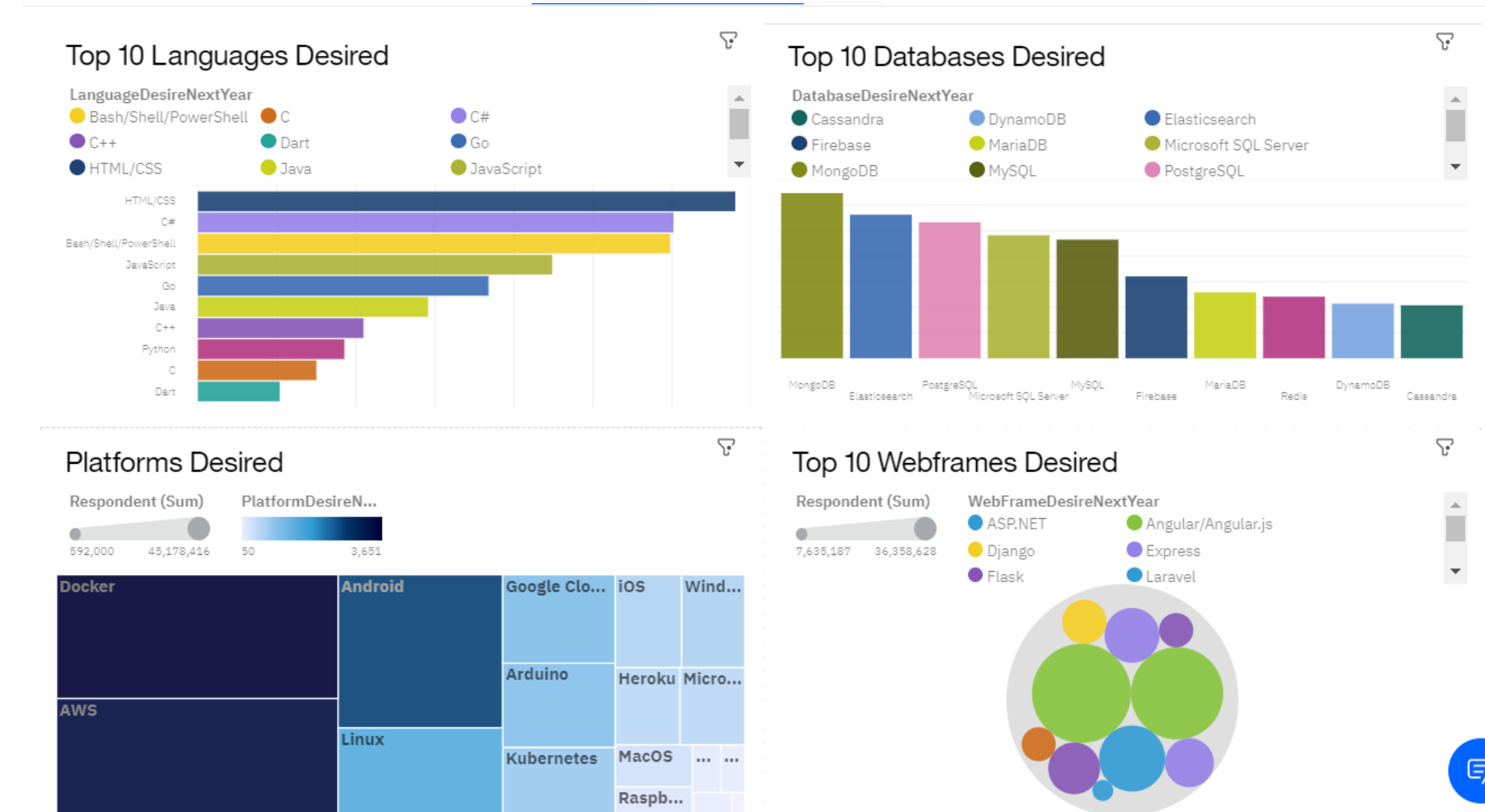


Top 10 Webframes Used

Respondent (Sum)



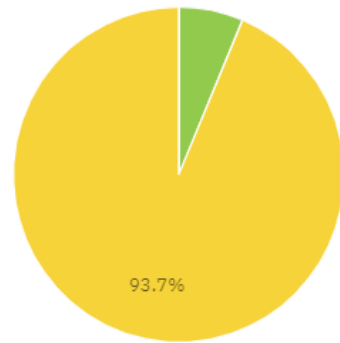
DASHBOARD TAB 2(Future Technology Trends)



DASHBOARD TAB 3 (Demographics)

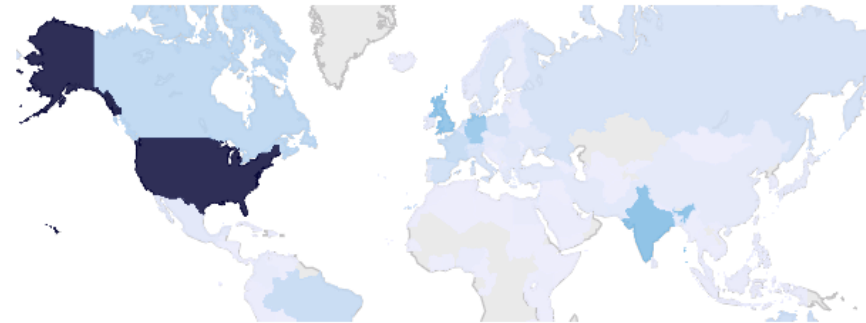
Respondent Classified by Gender

Gender
● Woman ● Man

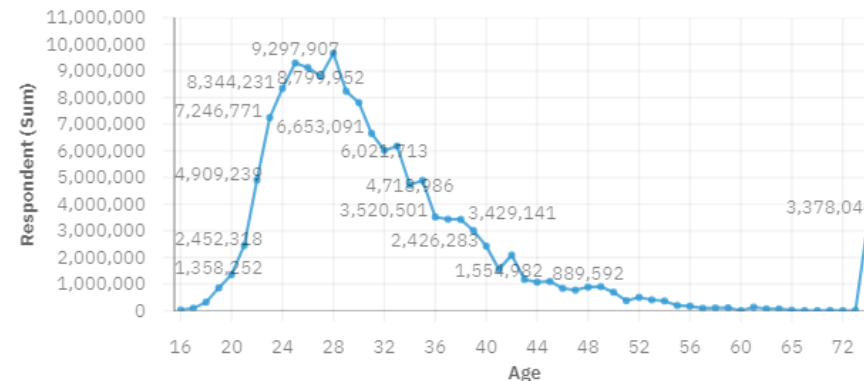


Respondent Count by Country

Respondent (Sum)
865 38,170,293

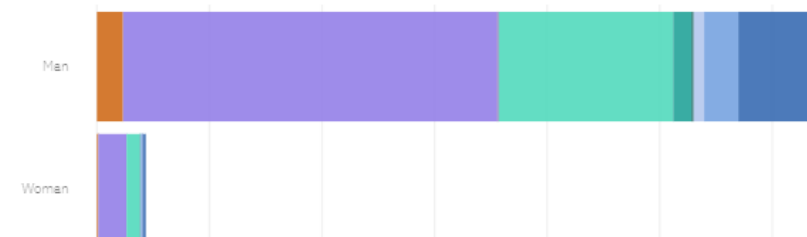


Respondent Count by Age



Respondent Count by Gender, classified by Formal Education Level

EdLevel
● Associate degree ● Bachelor's degree (BA, BS, B.E...
● I never completed any formal ... ● Master's degree (MA, MS, M.E...
● Other doctoral degree (Ph.D., Ed.D., ... ● Primary/elementary school



DISCUSSION



OVERALL FINDINGS & IMPLICATIONS

Findings

- Javascript is top programming language with which the developers work and will continue to persist in the next year at the top
- Python is the emerging language for the next year
- PostgreSQL is a the most demanded database for tomorrow

Implications

- Mastering the multiple languages will prove to be helpful, trends change year to year
- One must learn Python as it is emerging language for the next year for developers and analysts
- Database administrators must get hands-on training in PostgreSQL and MySQL database to remain up-to-date with the skills of future

CONCLUSION



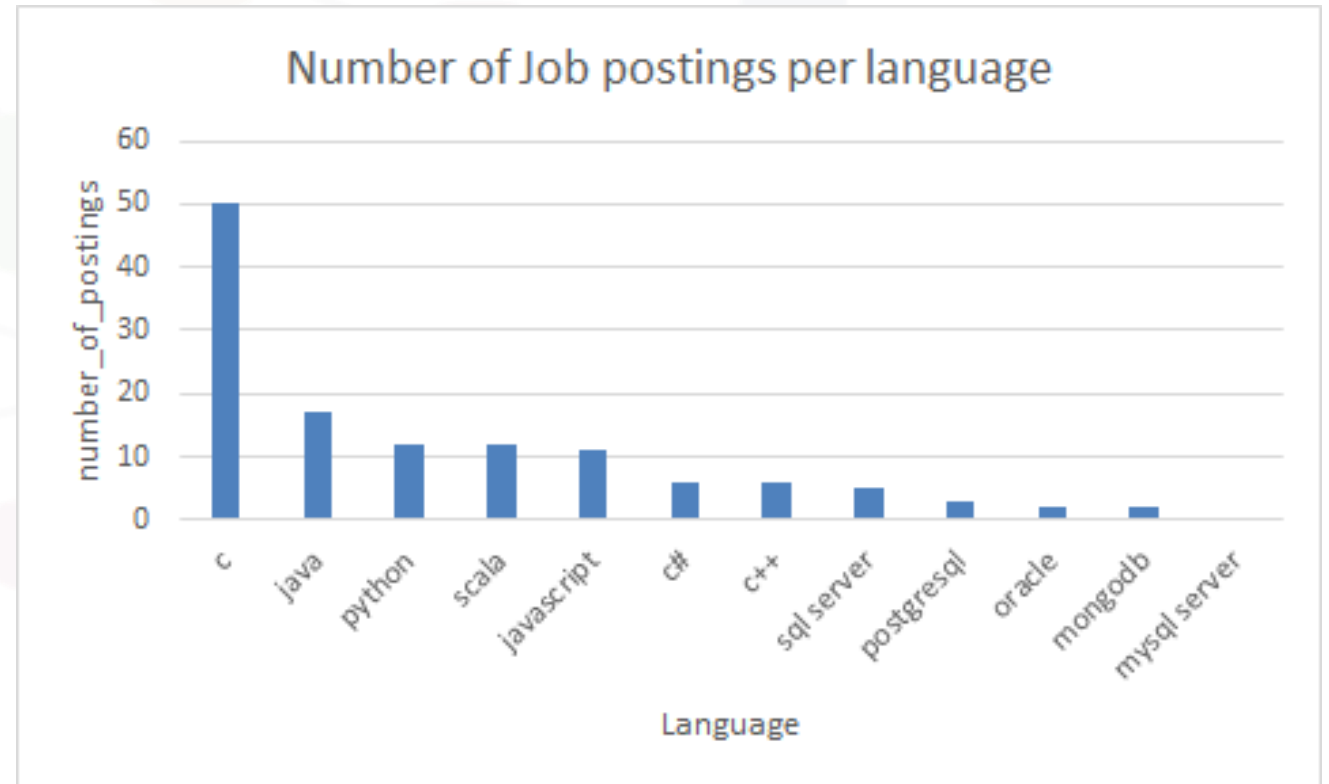
- Exploratory data analysis using Python and SQL helped to identify the trends in languages and databases being used and will be demanded most in the future
- IBM Cognos Dashboard provided visualizations on top 10 languages and databases
- Technology and skills required vary based on the market demand and business requirements
- Python and PostgreSQL are emerging and Javascript will remain at the top
- Updating and learning new skills based on market demand can be beneficial for one's career growth

APPENDIX



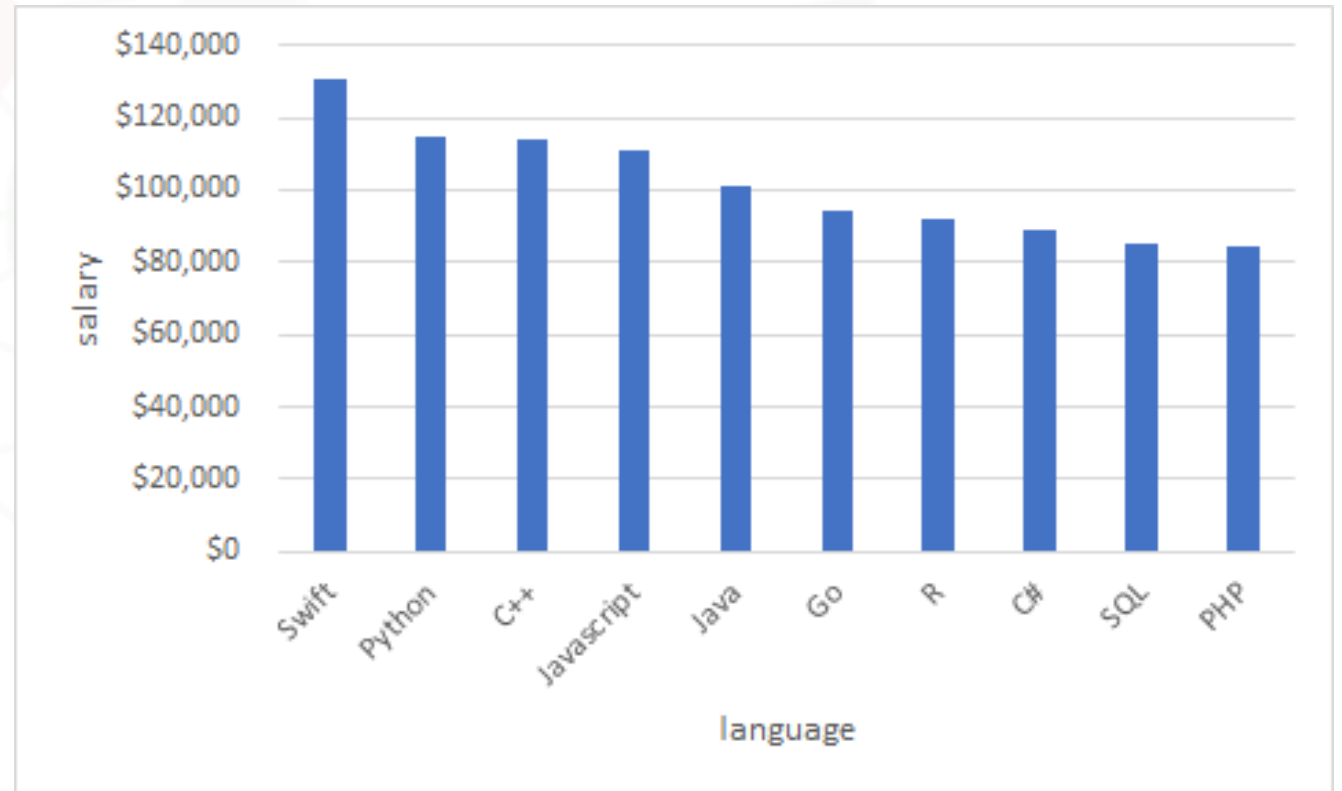
GITHUB JOB POSTINGS

- Collected the job postings data using GitHub API in a file named “github-job-postings.xlsx”
- Presented that data using a bar chart
- Ordered the bar chart in the descending order of number of job postings



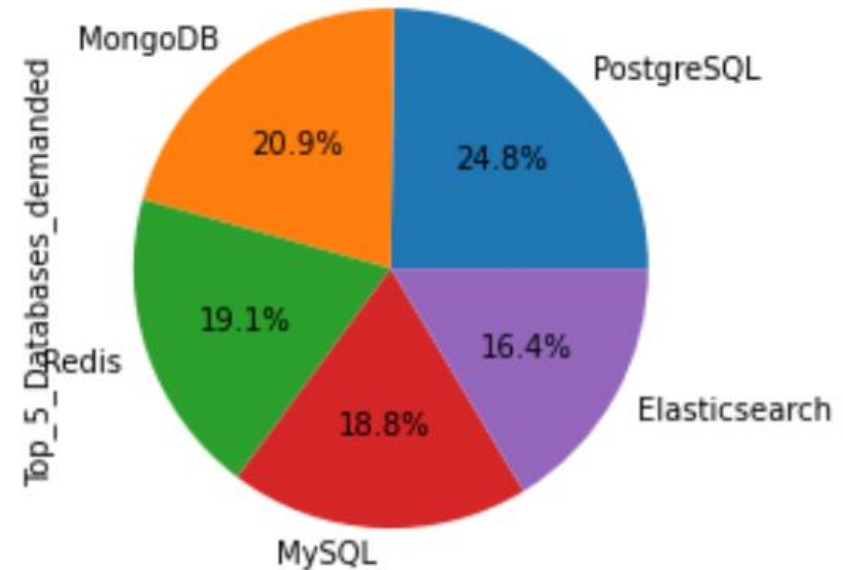
POPULAR LANGUAGES

- Collected the job postings data using web scraping in a file named “popular-languages.csv”
- Presented that data using a bar chart
- Ordered the bar chart in the descending order of salary



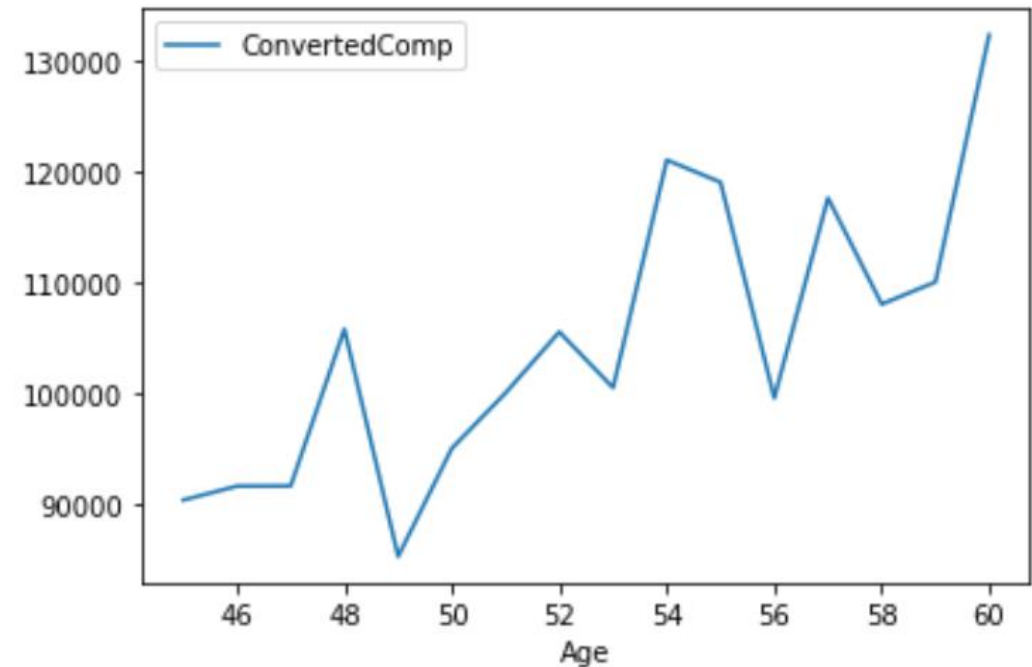
ADDITIONAL GRAPHS

- Visualized Survey data using pie chart
- Percentage demand of database for next year



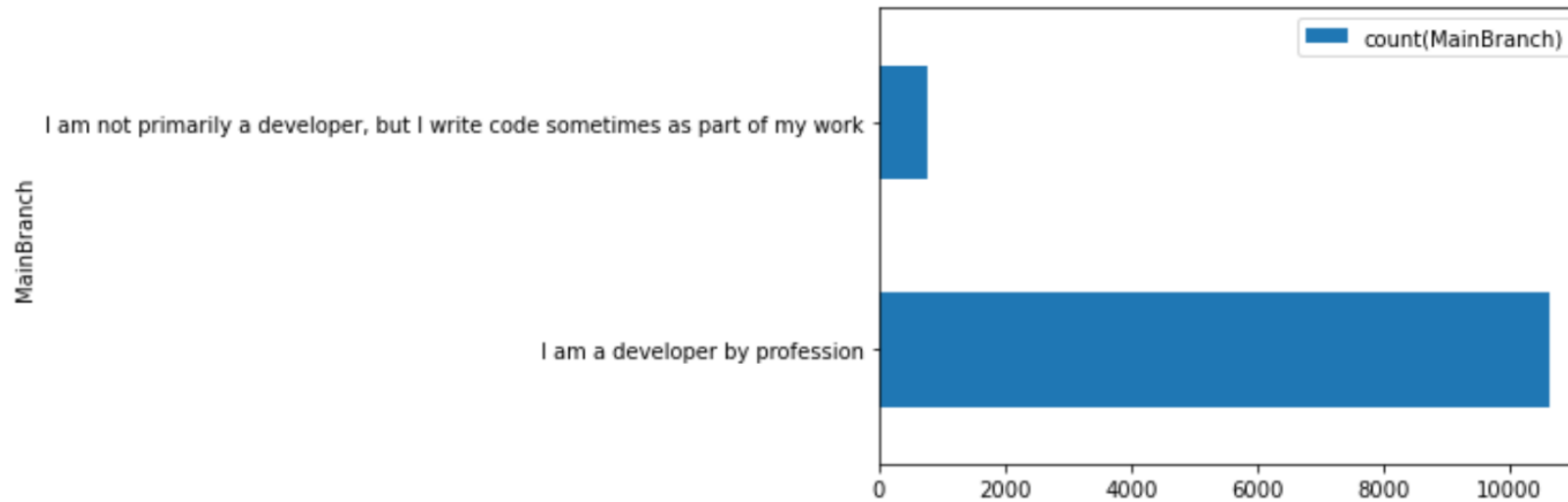
ADDITIONAL GRAPHS(contd.)

- Visualized high positive correlation between age and salary



ADDITIONAL GRAPHS(contd.)

- Horizontal bar chart to visualize the profession of respondents who took part in the survey



THANK YOU!