IBM Data Analyst Capstone Project: Analysis on Emerging Technology Skills and Trends

OUTLINE

Executive Summary

Introduction

Methodology

Results — Visualization – Charts / Dashboard

Discussion — Findings & Implications

Conclusion

Appendix

EXECUTIVE SUMMARY

▶ To be competitive in the global IT sector, it's essential to keep up with the ever-changing technologies. This report uses data analytics to highlight current and projected trends in the need for skills related to programming languages, databases and other technologies. It also studies the demographics of professionals in the technology sector.

▶ Data was gathered from a Stack overflow survey, IBM site, and Github job postings. It was collected, cleaned, subjected to exploratory analysis, and visualized on dashboards.

▶ The findings showed that Javascript is currently the most popular programming language and is anticipated to be so in the future. MySQL has the highest database usage at the moment but Postgre SQL is projected to have more demand in the future.

▶ Furthermore, majority of the survey respondents are males, are from the USA and are 28 years of age.

INTRODUCTION

- This presentation report uses data analytics to highlight current and projected trends in the need for skills related to programming languages, databases, platforms and web frames.

- The following inquiries were investigated using the data:
  - 1. Which programming languages are most in demand today?
  - 2. What are the most in-demand database skills?
  - 3. What popular IDEs or Web frames are there?

- The target audience for this research are IT professionals, HR managers, and anybody else with an interest in the IT sector who wants to learn about the top on-demand IT skills in their respective sectors that will also still be relevant in the future.

METHODOLOGY

Data in several formats, such as the number of jobs currently available for different technologies and for different places, were gathered using the Github jobs API on python.

To obtain the names of the programming languages and their yearly wages, the IBM website was scraped. The dataset from a 2019 Stack Overflow developer survey was downloaded and saved.

Python was used to clean and analyze the data. To assess the distribution of data, the presence of outliers, and the correlation between various columns in the dataset, an exploratory data analysis was carried out.

Charts, graphs, and dashboards were created using Python and Cognos analytics to visualize the data. All the python analyses were carried out on Jupyter notebook through visual studio.
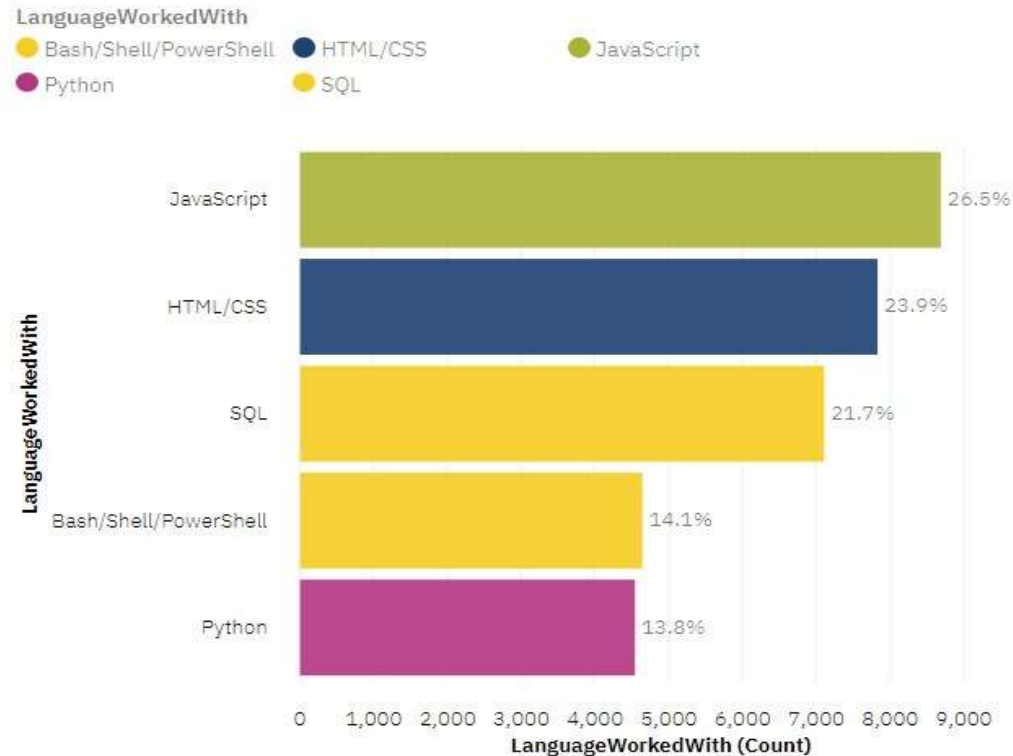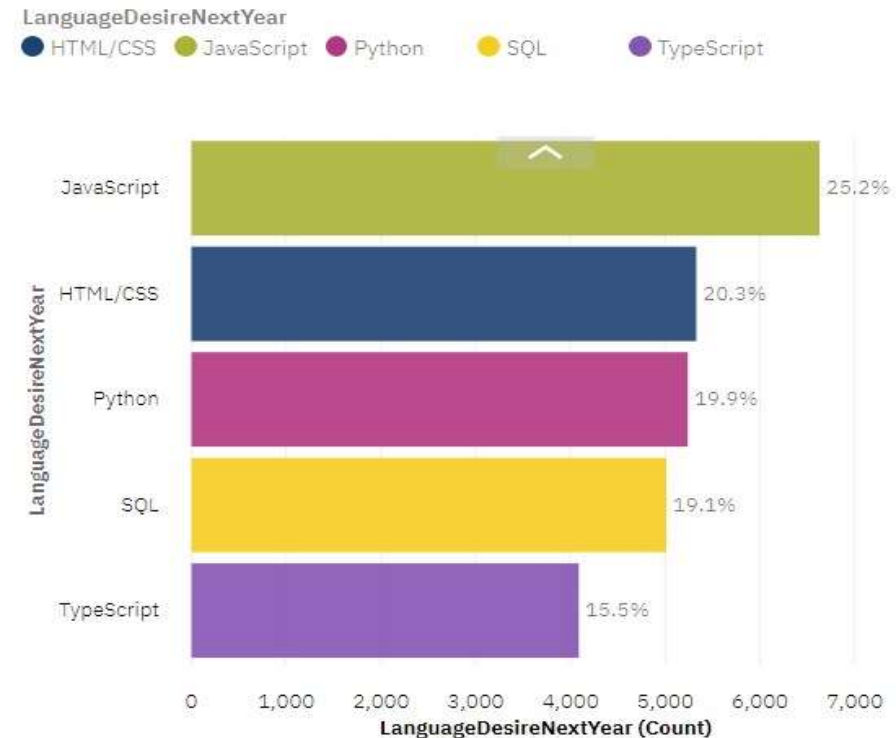
Results
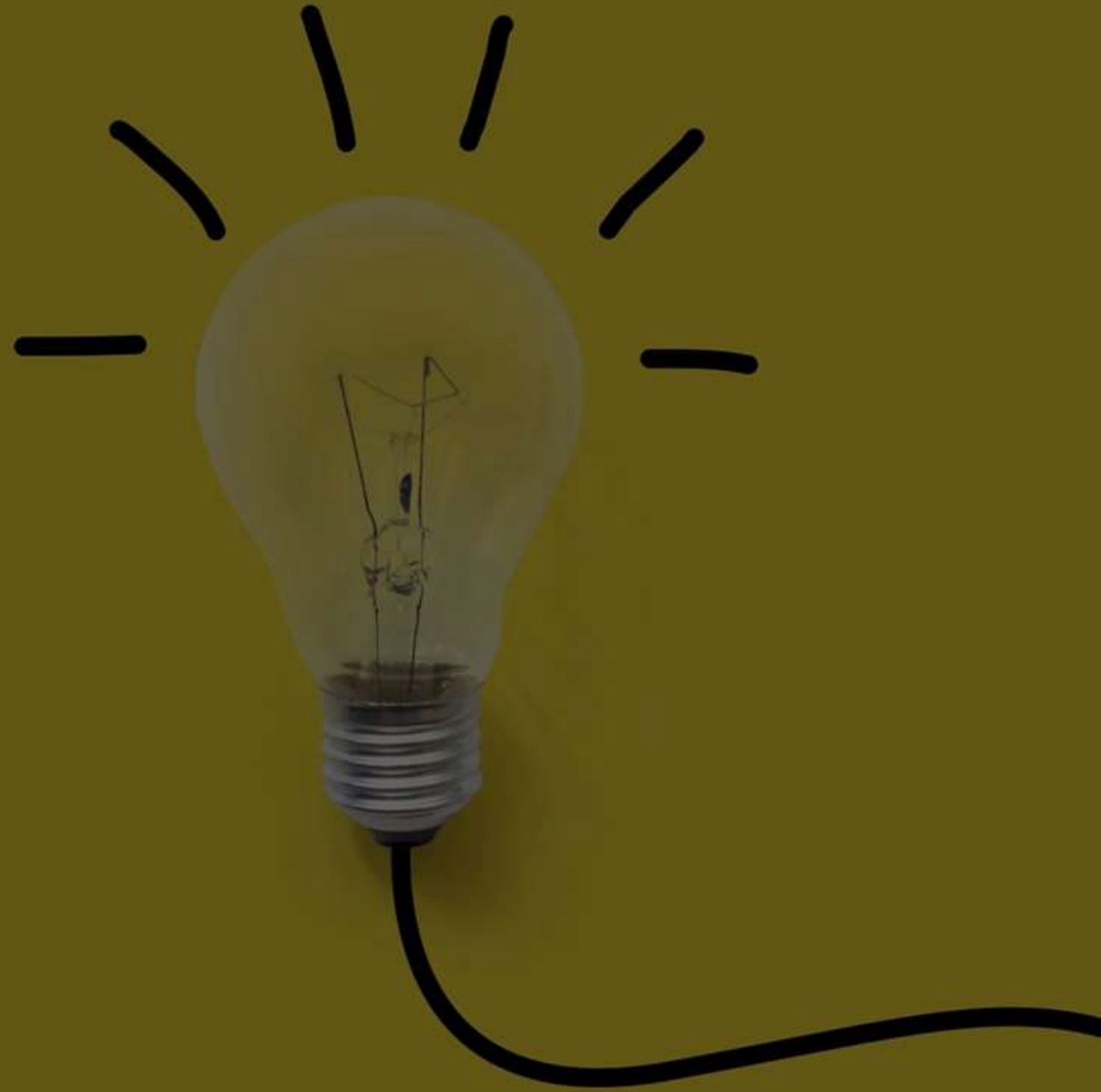
PROGRAMMING LANGUAGE TRENDS

▶ Current year

▶ Next year

PROGRAMMING LANGUAGE TRENDS – FINDINGS & IMPLICATIONS

Findings

▶ Javascript, HTML/CSS, SQL, Shell languages and Python are the most used languages currently.

▶ Javascript, HTML/CSS, Python, SQL, and Typescript will be the most used languages next year and future years.

▶ Python will have more demand than SQL next year.

Implications

▶ Javascript and HTML are used for web development which means that web development as a tech skill has the highest demand, especially as Typescript is getting viral.

▶ Python is gaining more and more traction due to the increase in demand for AI and ML skills.

▶ SQL is the still the most relevant language for data professionals. It is important for aspiring data analysts, scientists, business analysts etc to have SQL skills
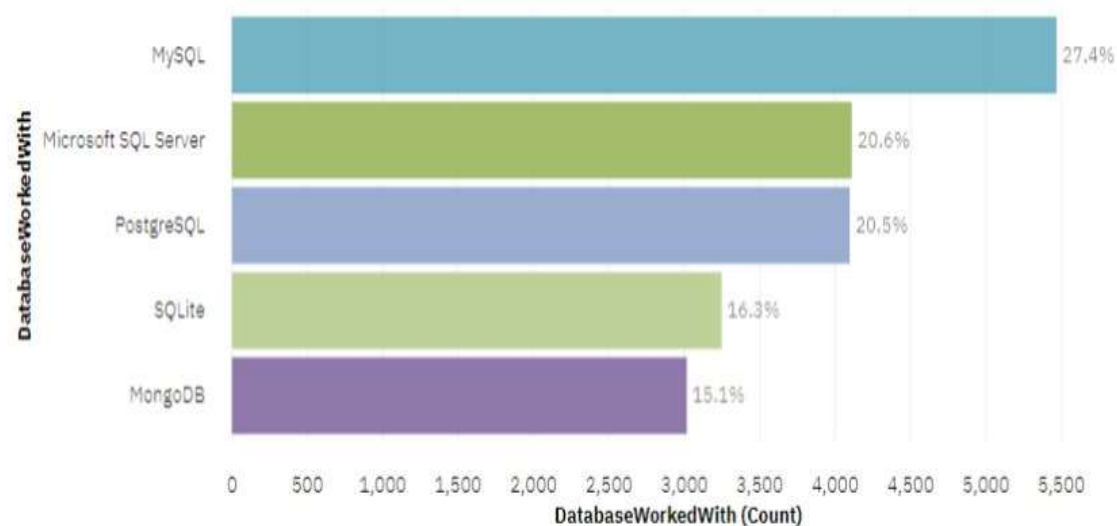
DATABASE TRENDS

▶ Current year

▶ Next year

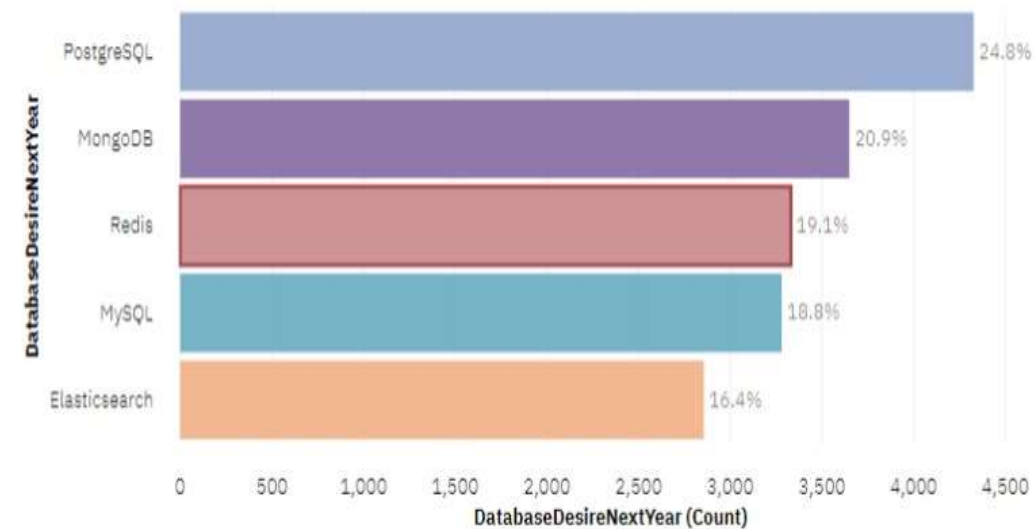## Top 5 Databases in the Current year

DatabaseWorkedWith
● Microsoft SQL Server ● MongoDB ● MySQL ● PostgreSQL ● SQLite

DatabaseWorkedWith	
MySQL	27.4%
Microsoft SQL Server	20.6%
PostgreSQL	20.5%
SQLite	16.3%
MongoDB	15.1%

0   500   1,000   1,500   2,000   2,500   3,000   3,500   4,000   4,500   5,000   5,500

DatabaseWorkedWith (Count)

Top 5 Databases for the Next Year

DatabaseDesireNextYear
● Elasticsearch ● MongoDB ● MySQL ● PostgreSQL ● Redis

DatabaseDesireNextYear	
PostgreSQL	24.8%
MongoDB	20.9%
Redis	19.1%
MySQL	18.8%
Elasticsearch	16.4%

0   500   1,000   1,500   2,000   2,500   3,000   3,500   4,000   4,500

DatabaseDesireNextYear (Count)

DATABASE TRENDS – FINDINGS & IMPLICATIONS

Findings

- MySQL, Microsoft SQL server, Postgre SQL, SQLite and MongoDB are the top 5 most used databases at the moment.

- However, Postgre SQL, MongoDB, Redis, MySQL and Elasticsense are projected to become more popular in the future.

- Redis and Elasticsense are relaatively new tools and are set to gain more traction in the IT space.
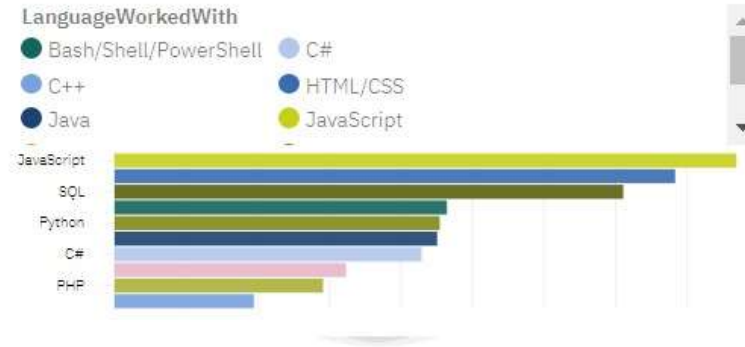
Implications

- SQL is still a top tool to watch out for in data specialists.

- Companies still prefer Open source databases.

- Oracle SQL was not among the top 5. It is losing relevance as time passes.

DASHBOARD

The permanent link of the Cognos dashboard:
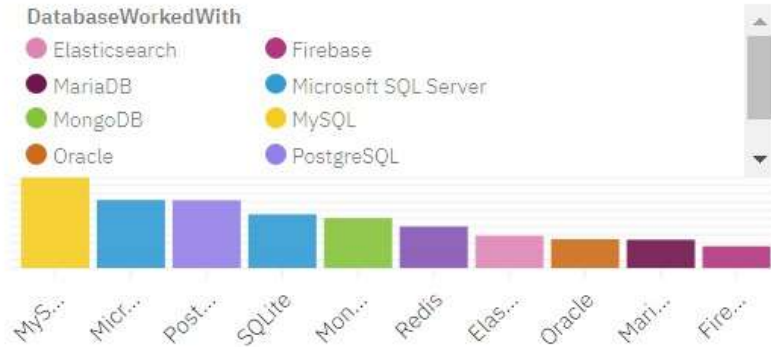
https://eu-gb.dataplatform.cloud.ibm.com/dashboards/eafbc83d-998f-4cd9-ae12-bfe6021457a7/view/4316dc2402af1bef1ff3bde4079e25002961205cb6bb8004d2d47b490e612397a93a4594c87d4f5b8b425330fabd15089a
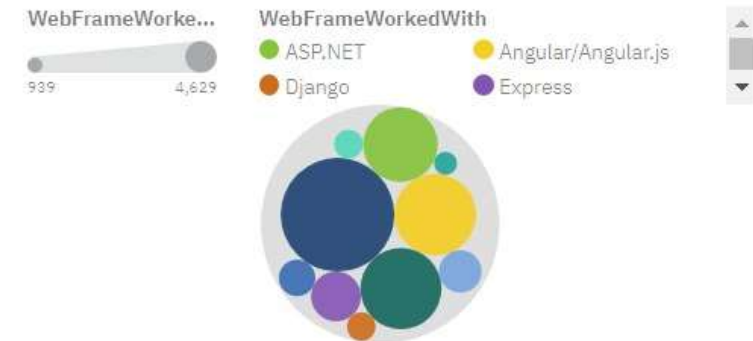
DASHBOARD TAB 1
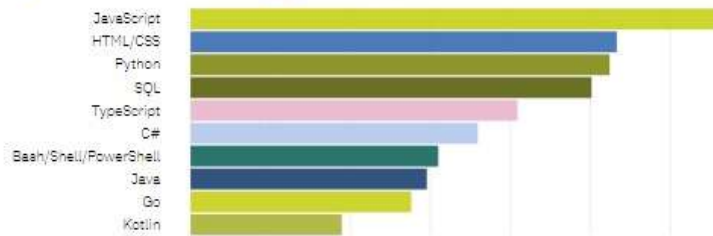
DASHBOARD TAB 2

DASHBOARD TAB 3
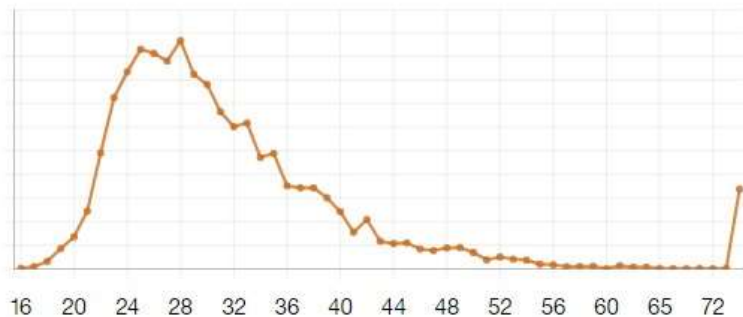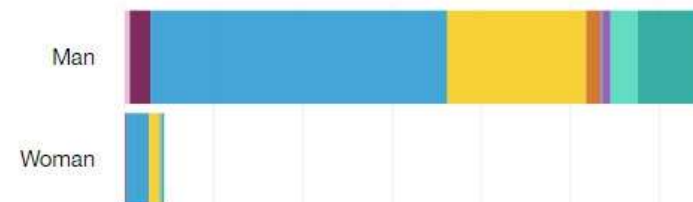
DISCUSSION

▶ Upskilling in the Technology sector.

▶ How do we close the wide gender gap in the Technology sector?

▶ Is completing a masters or doctorate degree really a requirement?

▶ The increasing demand for mobile development as Kotlin is getting popular.

▶ More tech education, access and development in less developed regions in South east Asia, South America, Africa and some parts of Europe.

▶ How relevant will Oracle SQL still be in the future?

OVERALL FINDINGS & IMPLICATIONS

Findings

- Most people in the IT field have a Bachelors' degree.

- Web development languages are the most popular and on-demand tools in the IT field currently.

- The Tech sector is filled with majorly young people under 40 years of age.

- Most respondents want to learn Postgre SQL and React JS next year.

Implications

- It is important for data professionals to develop proficiencies in NoSQL in addition to SQL databases.

- Web development is still a very lucrative skill.

- Less developed countries need more access to tech trainings and education.

CONCLUSION

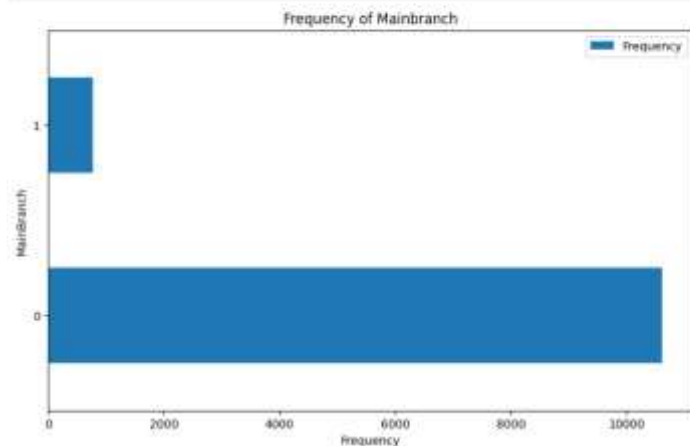IT IS EXPEDIENT TO STAY UPDATED IN THE TECH SECTOR AS THE TRENDS KEEP CHANGING OVER TIME.
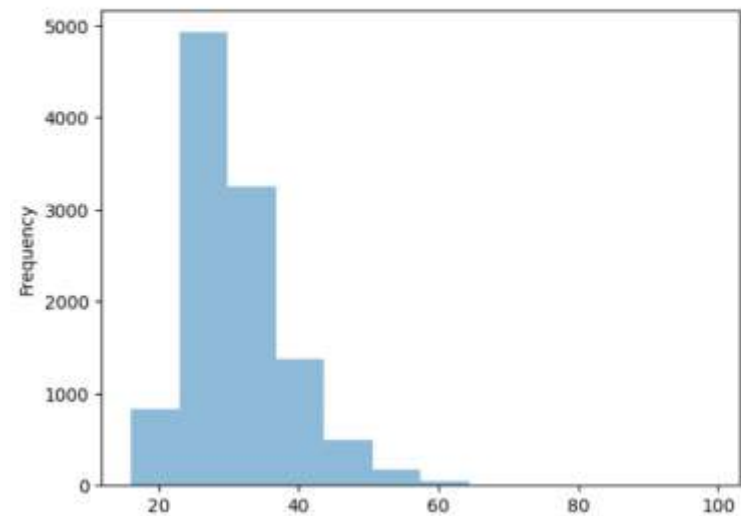
APPENDIX

JOB POSTINGS

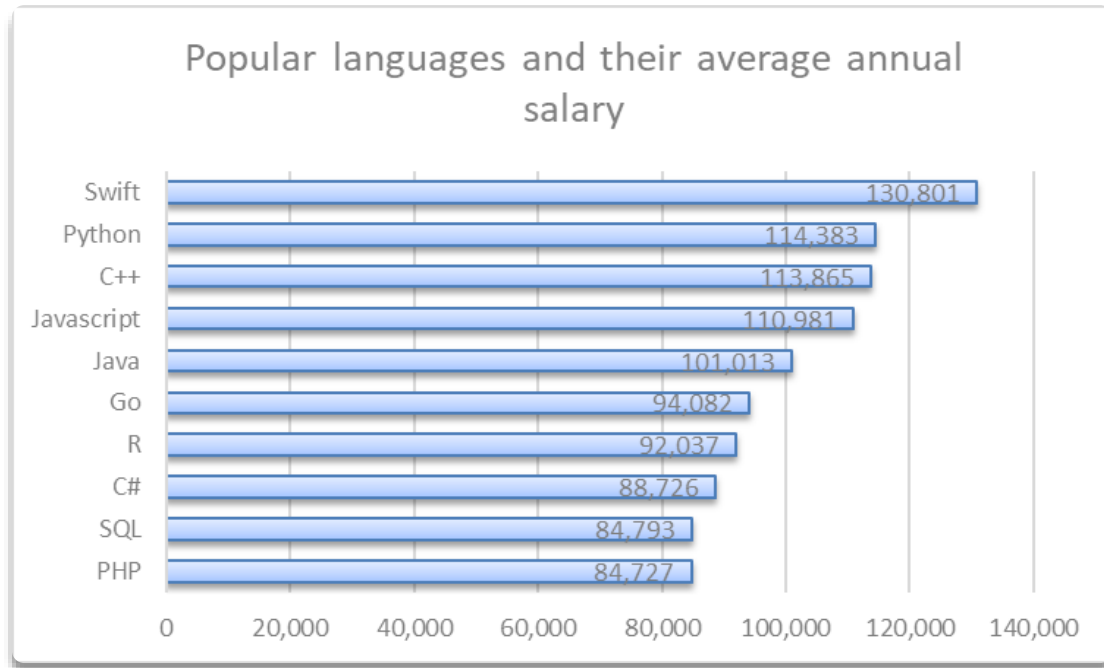Bar chart presenting the job posting data collected using Github Job API.

number of job posting

POPULAR LANGUAGES

Popular languages and their average annual salary

Language	Salary
Swift	130,801
Python	114,383
C++	113,865
Javascript	110,981
Java	101,013
Go	94,082
R	92,037
C#	88,726
SQL	84,793
PHP	84,727
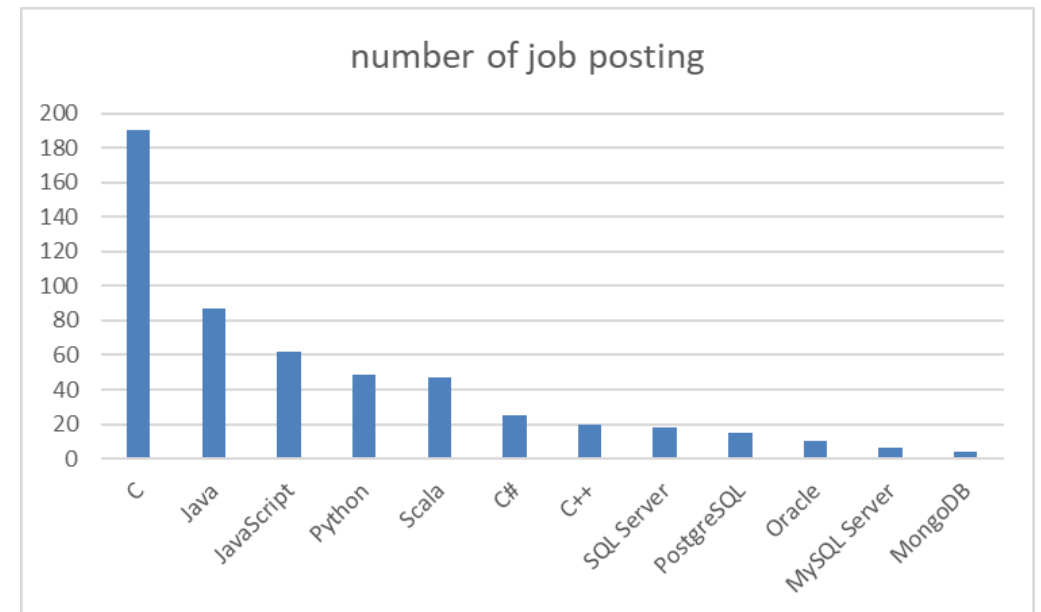
Bar chart displaying popular languages and their average annual salary. The data was collected through web scraping the Github jobs data and saved in a csv file.