



IBM Data Analyst Capstone Project

Alejandro Velásquez

October 11th, 2022

OUTLINE



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

EXECUTIVE SUMMARY



- This report analyzes the current trends in the IT field, the most sought-after skills according to job postings on Github, and the better compensated languages
- According to the EDA applied to the data:
 - JavaScript, CSS and Python are the highest growing languages among professionals.
 - Python is the most sought-after programming language in job vacancies and one of the best remunerated.
 - SQL retains its spot as the preferred language for DBMSs.
- In the foreseeable future, JavaScript will still be the most used programming language
- Most of the IT workforce is comprised of highly educated young males.

INTRODUCTION



- The IT field rapidly grows each year. Thus, it's important to get insights about the technologies used by developers and their demographics.
- Companies and recruiters want and need to know the current trends to obtain the most apt workers in the IT field.
- Thanks to the Stack Overflow Developer Survey and Github job postings, we are able to gain an understanding about the on-demand skills and languages that are currently looked for, and their corresponding salary.
- Additionally, thanks to the survey we are able to understand important info about the respondents like:
 - Age, gender and location.
 - Formal education level and their main branch.

METHODOLOGY



- Stack Overflow's 2019 Developer Survey, Github's job postings and IBM's website were used as sources. The survey was stored as a CSV file, the job postings were obtained with a Python request through Github's API, and the IBM website was web scrapped.
- The data was cleaned and prepared with data wrangling techniques for it's subsequent analysis.
- EDA was performed on the prepared data to gain an understanding of the data, find correlations and obtain insights
- The key findings were plotted and graphed with several tools:
 - Python's Matplotlib library
 - IBM Cognos Analytics

RESULTS

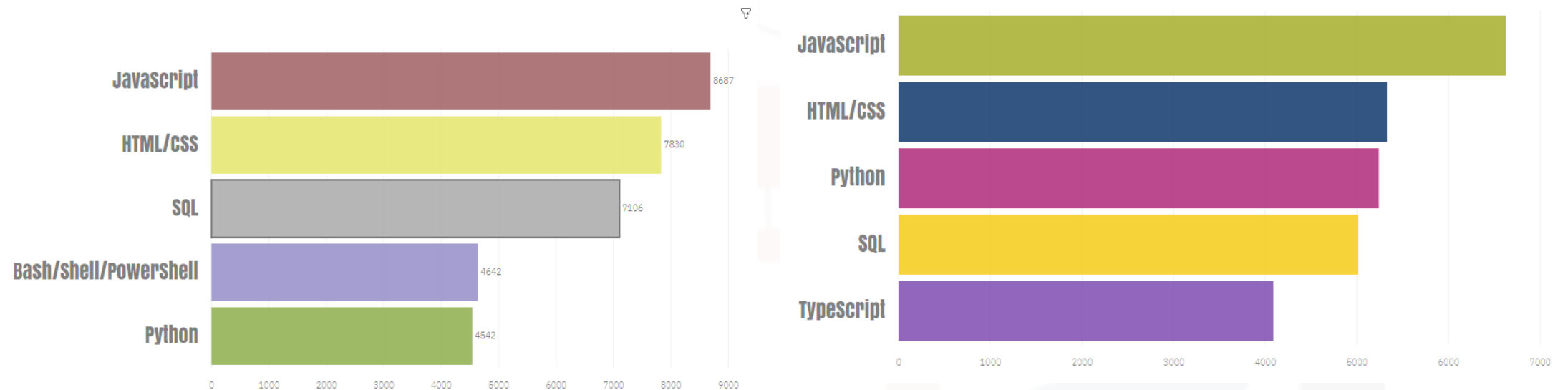


- Thanks to the EDA performed on the data, there were several indicators obtained, including correlations between age and yearly compensation, and the trends for technology adoptions.
- The job posting information was divided by type of technology, the current and future use of said technologies, and average salaries.
- The respondents were divided in age groups, formal education level and gender.

PROGRAMMING LANGUAGE TRENDS

Current Year

Next Year



PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

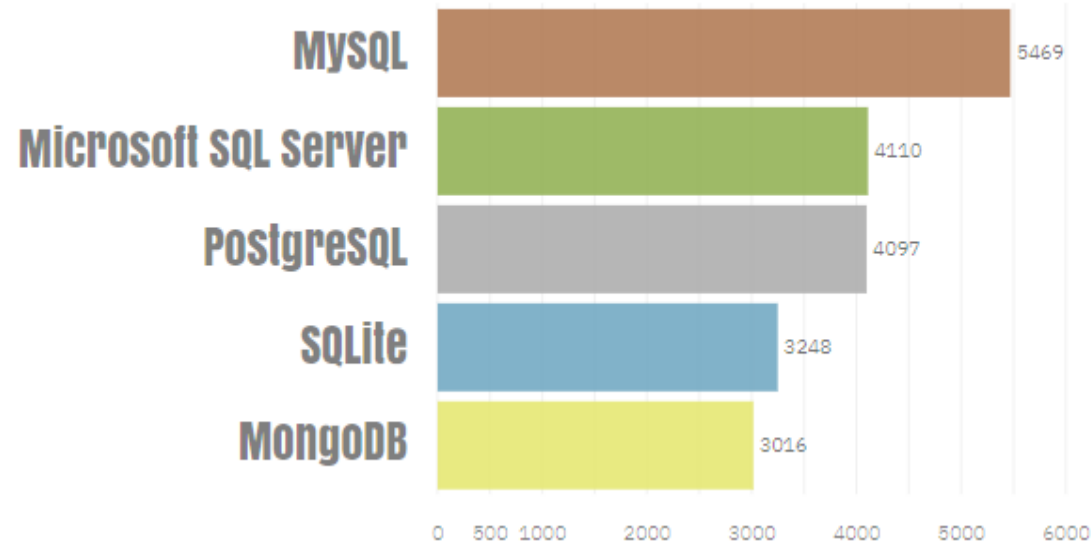
- Web interactive languages are the most in demand.
- SQL is the preferred language for DBMSs.
- Python isn't as used as other languages, but is among the most desired to learn by professionals.

Implications

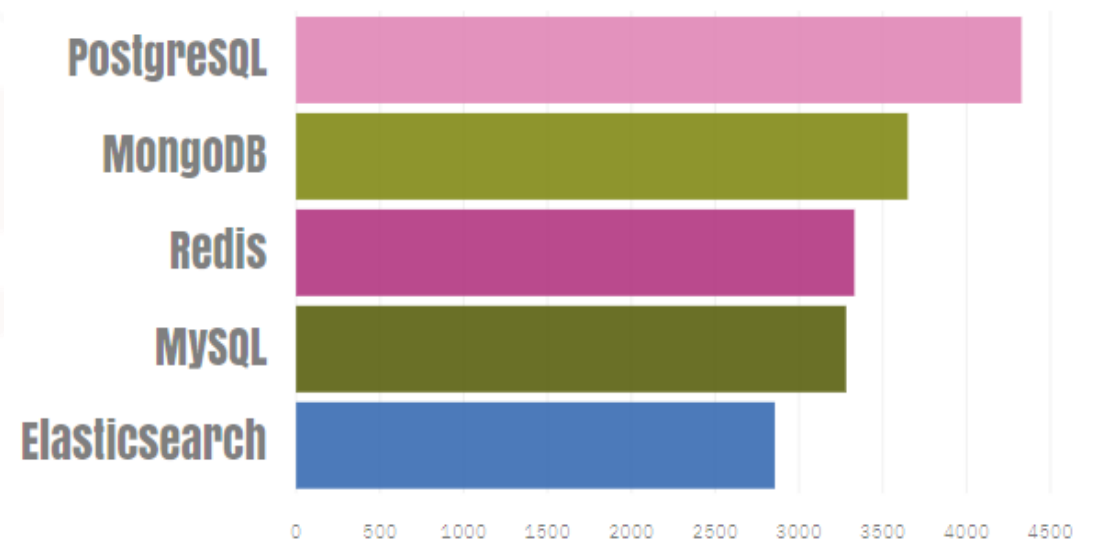
- Web interactive languages like JavaScript and CSS are the preferred languages to learn by professionals.
- Python will become one of the most used programming languages in the future
- SQL will continue to be relevant in the near future.

DATABASE TRENDS

Current Year



Next Year



DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- Relational and SQL based databases are the most currently used.
- NoSQL DBs are among the most looked to learn by professionals.
- MongoDB is the most popular NoSQL database.

Implications

- NoSQL databases like MongoDB and Redis will be among the most used in the future.
- PostgreSQL will become the most popular RDBMS.
- ElasticSearch will continue to be the most widely-used enterprise search engine.

DASHBOARD



Capstone Project Dashboard

DASHBOARD TAB 1

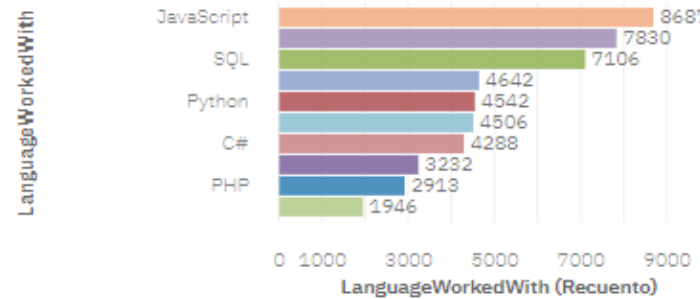
Current Technology Usage

Future Technology Trend

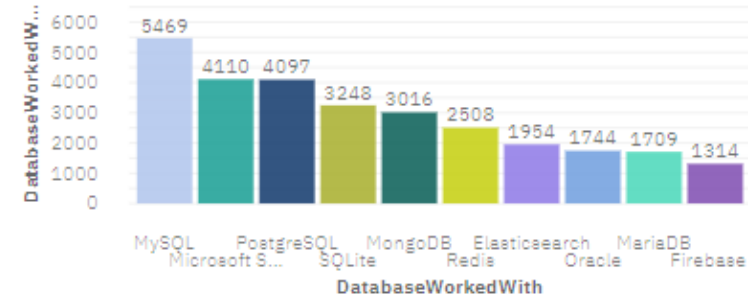
Demographics



Top 10 Language Worked With



Top 10 Database Worked With



Platforms most worked with

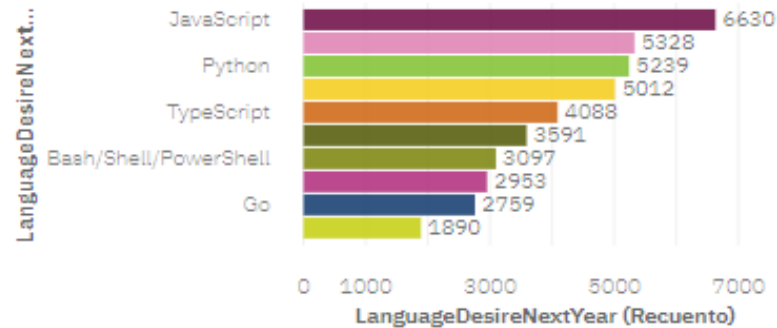


WebFrames Most Worked With

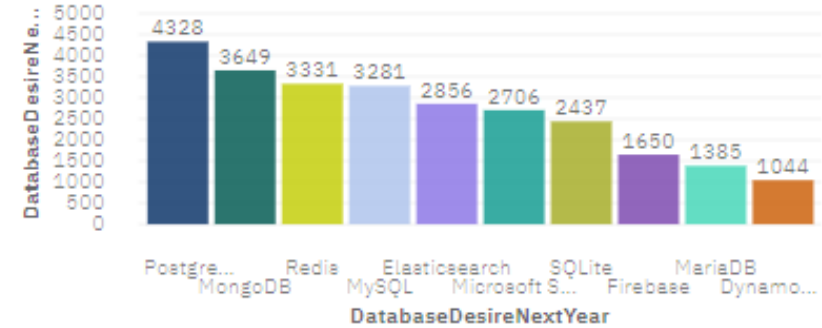


DASHBOARD TAB 2

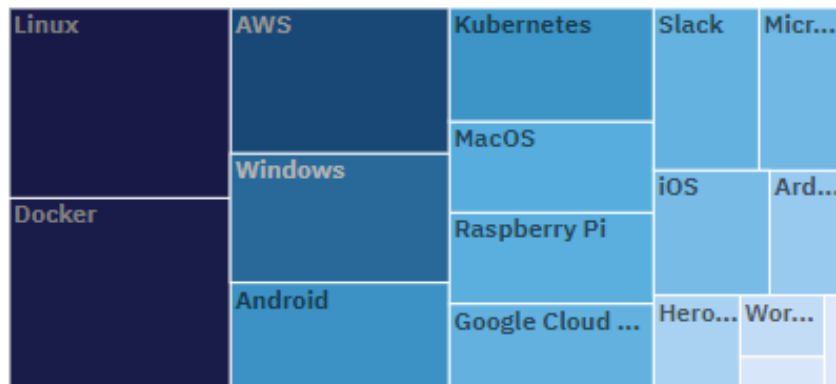
Top 10 Languages Desired to Learn



Top 10 Databases Desired to Work with



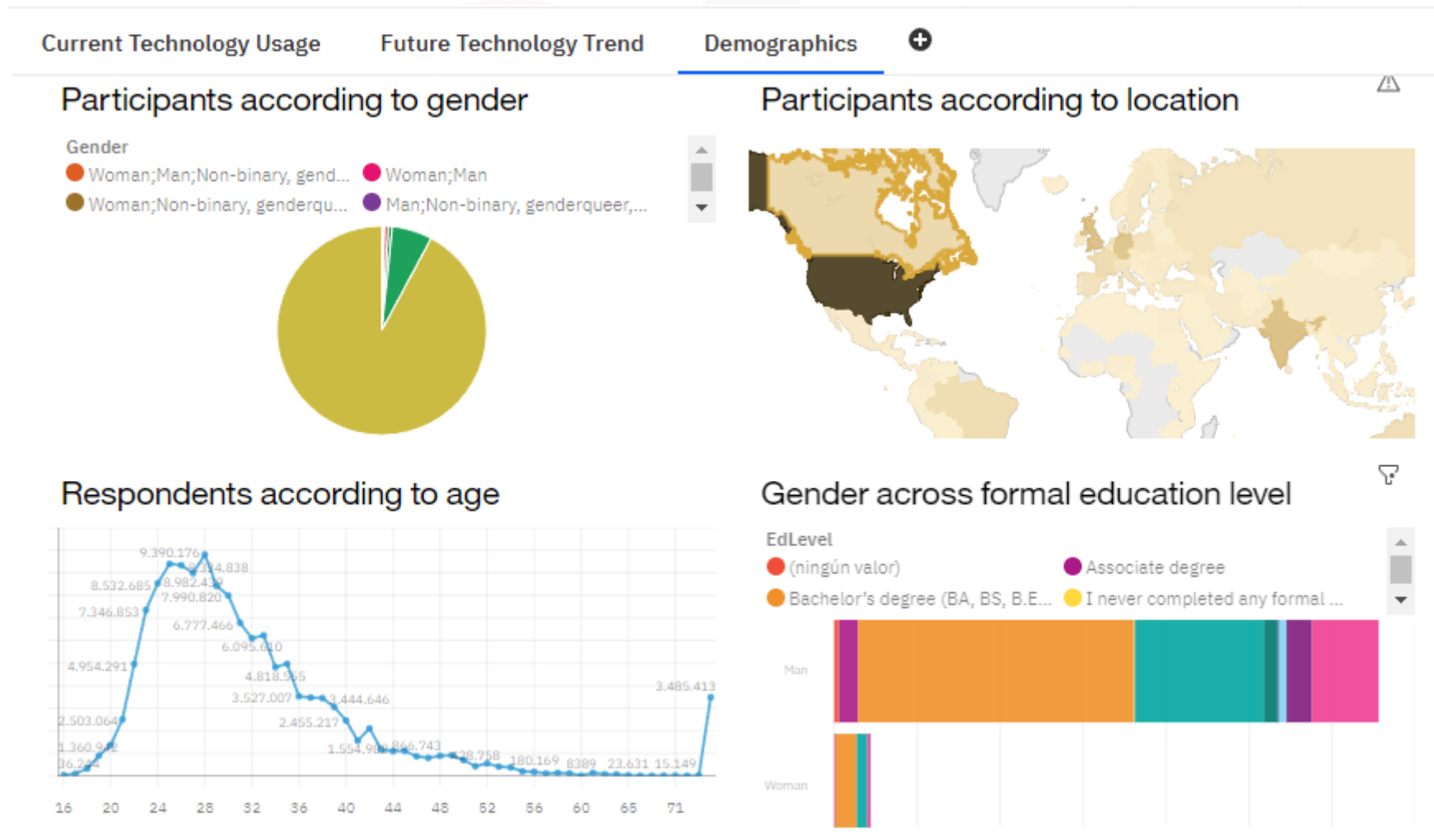
Desired Platforms Tree Map



Top Webframes Desired to Work With



DASHBOARD TAB 3



DISCUSSION



- There were 11398 participants in the survey, of which 92,2% identified themselves as males.
- The average respondent was between 20 and 40 years old. With the 24-28 age group being the most numerous.
- Virtually all respondents were professionals, and had a formal education degree.

OVERALL FINDINGS & IMPLICATIONS

Findings

- JavaScript is the most used programming language. MySQL is the most common database. Windows and Linux are comparable in terms of usage, and jQuery is the most worked with web frame.
- Most professionals wish to migrate to PostgreSQL for their databases, Linux and Docker for their platforms, and ReactJs for their web frames.
- An overwhelming majority of developers are young males.

Implications

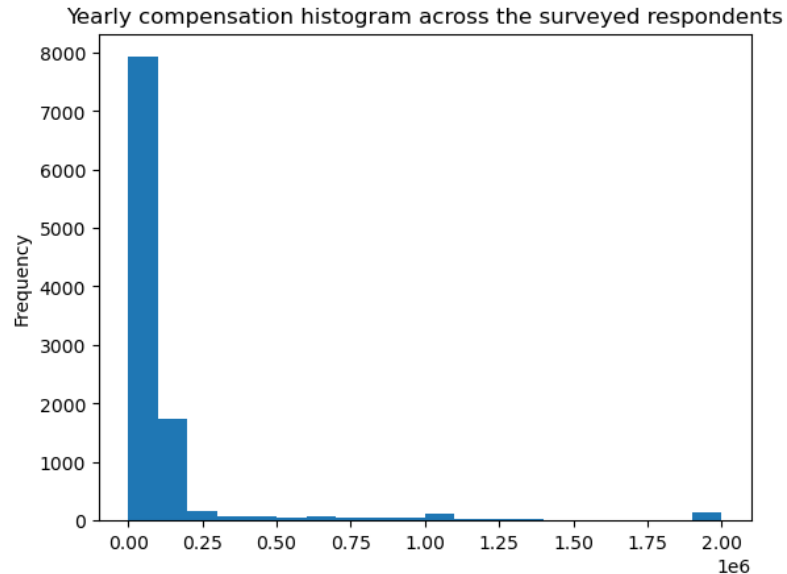
- In the foreseeable future, IT will still be a male-dominated field.
- JavaScript will keep its top spot as a programming language in the near future, but Python will become a close contender.
- JavaScript-based interfaces and webframes will remain in-demand for the time being.

CONCLUSION



- Webpage interactive languages are still in high demand, with companies offering plenty of jobs and adequate compensation.
- Python is a highly sought language, with high paying jobs and a growing utilization.
- The rise of remote work has had an impact on developer's preferred platforms, with Docker quickly becoming the choice of professionals.
- There's a need to incentivize gender diversity on the IT field.

APPENDIX



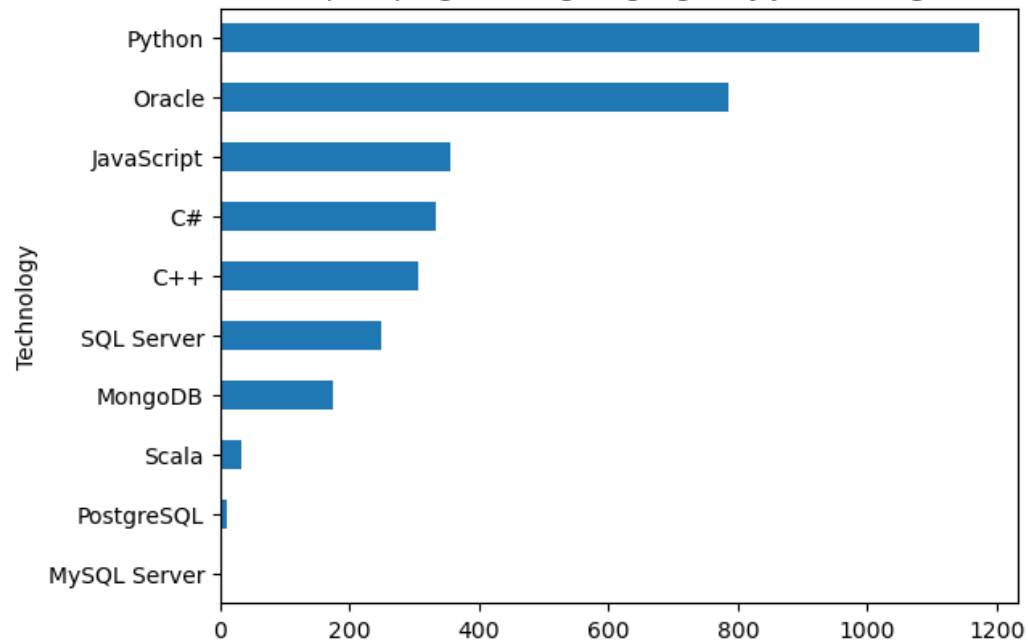
```
df.loc[df['CompFreq'] == 'Yearly', 'NormalizedAnnualCompensation'] = 1 * df['CompTotal']  
df.loc[df['CompFreq'] == 'Monthly', 'NormalizedAnnualCompensation'] = 12 * df['CompTotal']  
df.loc[df['CompFreq'] == 'Weekly', 'NormalizedAnnualCompensation'] = 52 * df['CompTotal']
```

```
df[['NormalizedAnnualCompensation']].median()
```

```
NormalizedAnnualCompensation    100000.0
```

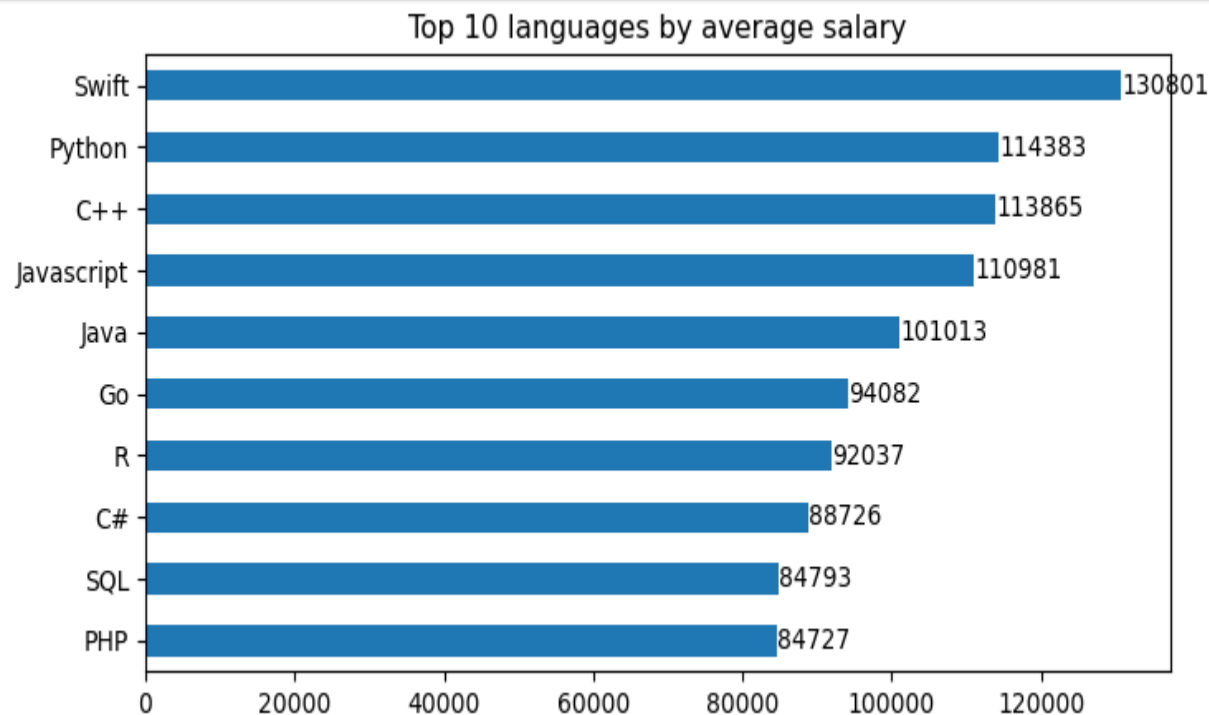
- Most respondents earn between USD 50,000 and USD 150,000. The median salary was USD 100,000.

JOB POSTINGS



- Job postings on Github, ordered by the required skill. Clearly showing Python as a highly-sought language.

POPULAR LANGUAGES



- Average salaries by programming language, Swift and Python take the top spots.