

# DATA ANALYST CAPSTONE PROJECT

LAWRENCE SHEED  
29 July 2021

# OUTLINE



Executive Summary

Introduction

Methodology

Results

- Visualization – Charts
- Dashboard

Discussion

- Findings & Implications

Conclusion

Appendix

# EXECUTIVE SUMMARY



The online survey from Stack Overflow is used to examine programming trends

Most popular language - JS

Fastest upcoming language - Python

Gender Gap (95% male / 5% female)

Age group 25-30

# INTRODUCTION



Analyzing software & developer trends

Stack Overflows annual developer survey is the largest and most comprehensive survey of programmers in the world, with almost 90,000 participants

Analysis Purpose:

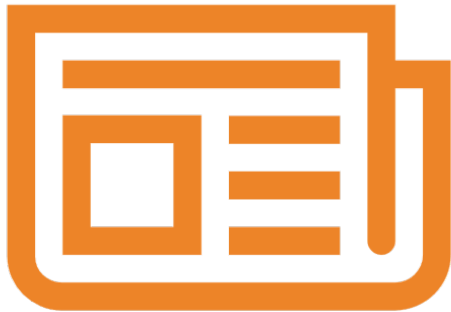
What languages are in demand?

What skills are needed in the future

What database skills are in demand

Audience: HR, and IT

# METHODOLOGY



## Data Collection Sources

- Stack Overflow Developer Survey 2019
  - based on a survey of 88,883 developers from 179 countries.
- GitHub Job Postings (Deprecated!)

## Data Exploration

## Data Cleaning

## Data Visualization

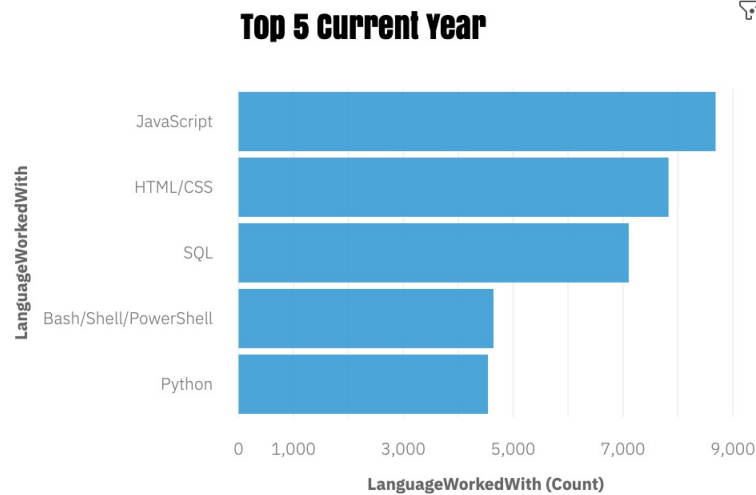
## Presentation

↑↓	Respondent	LanguageWorkedWith	LanguageD...eNextYear	DatabaseWorkedWith	DatabaseD...eNextYear	PlatformWorkedWith	PlatformD...eNextYear	WebF
	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	
	Null	Null	Null	Null	Null	Null	Null	Null
	4	C	C	MySQL	MySQL	Linux	Linux	Null
	4	C++	C#	SQLite	SQLite	Windows	Windows	Null
	4	C#	JavaScript	Null	Null	Null	Null	Null
	4	Python	SQL	Null	Null	Null	Null	Null

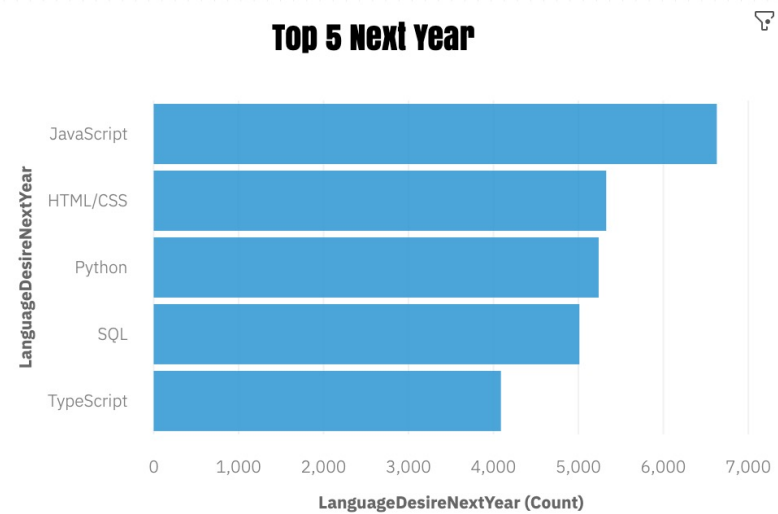
# RESULTS

# PROGRAMMING LANGUAGE TRENDS

Current Year



Next Year



## PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

### Findings

JS, HTML/CSS, SQL top 3

Python steadily improving in rankings

BASH/Powershell becoming less in demand

### Implications

Web frameworks in demand

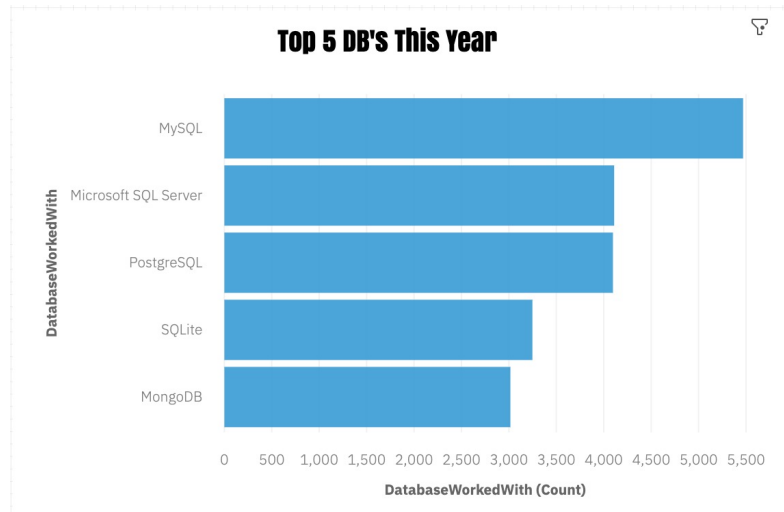
Python data processing for ML/ AI becoming more important

Asbtraction becoming more important –  
i.e. Docker / Kubernetes.

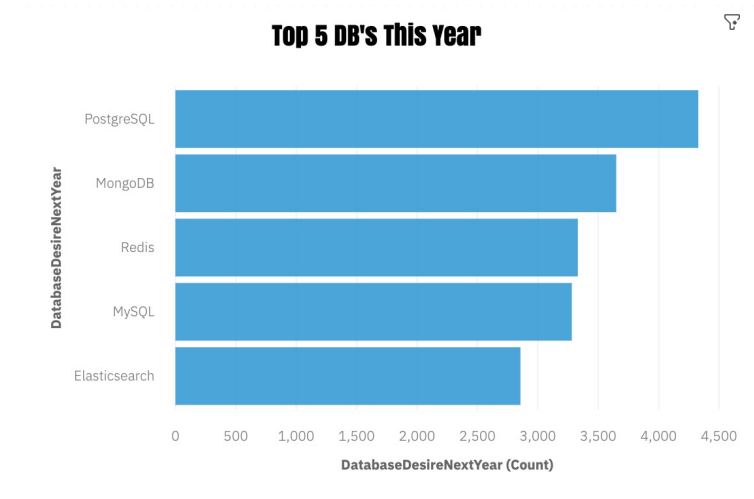


# DATABASE TRENDS

## Current Year



## Next Year



# DATABASE TRENDS - FINDINGS & IMPLICATIONS

## Findings

MySQL most popular, Postgres in close second

MSSQL second with Postgres  
Mongo and SQLite "databases" increasing in usage.

ElasticSearch newcomer

## Implications

Open Source is still king.

MSSQL losing traction.

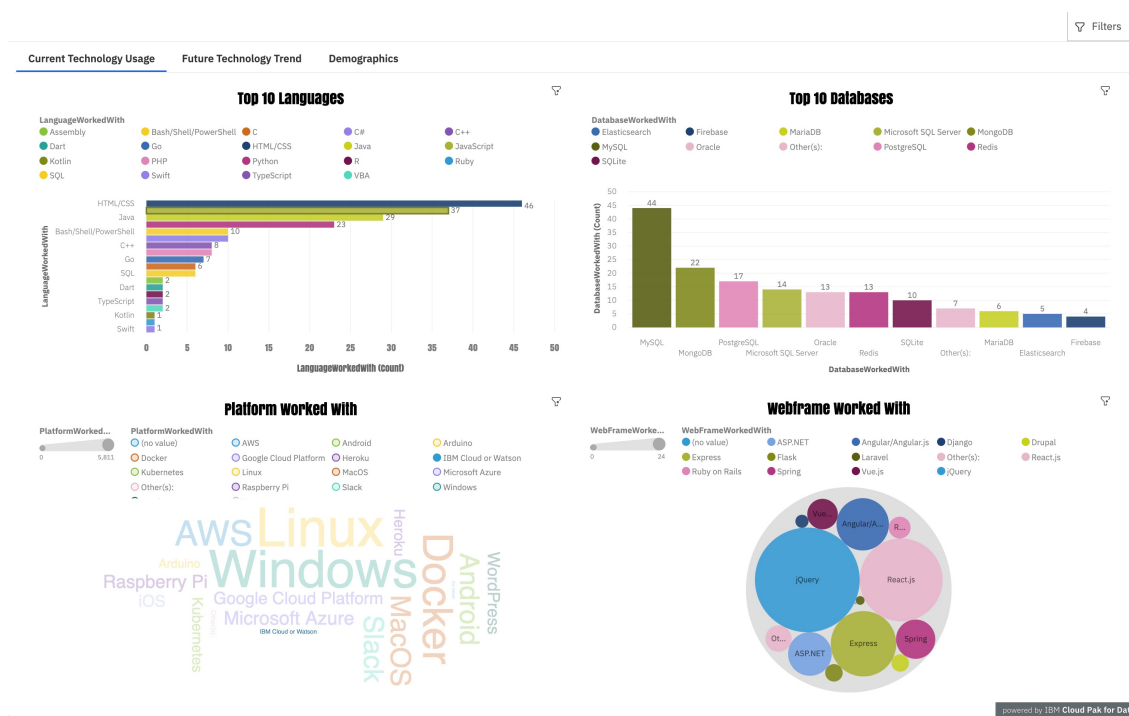
DB 'Lite' lite db vs transactional DB's becoming more prevalent – cloud use again popping its head in.

# DASHBOARD

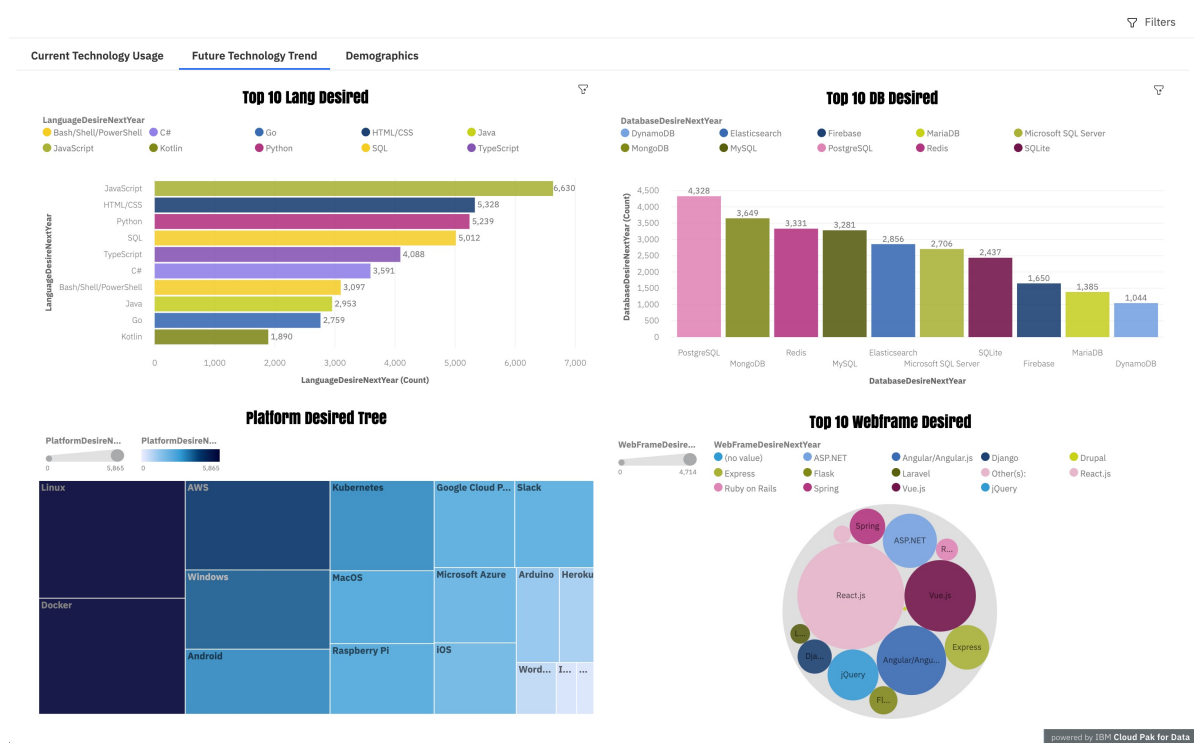


<https://Dash.Board> ← Click me

# DASHBOARD TAB #1



# DASHBOARD TAB #2



# DASHBOARD TAB #3

All tabs

No filters have been applied.

This tab

Gender  
Man, Woman

Current Technology Usage

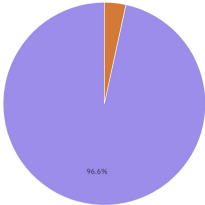
Future Technology Trend

**Demographics**

**Respondent by Gender**

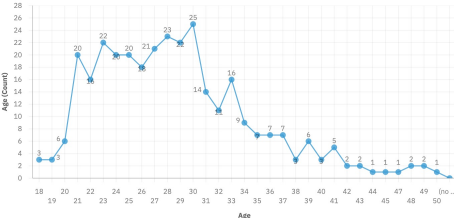
Gender

Woman Man



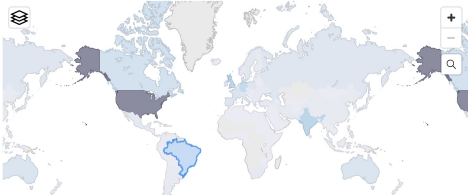
**Respondent Count by Age**

Age (Count)



**Respondent by country**

Respondent (Cou...)

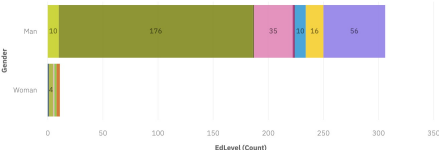


**Respondent count by Gender, classified by Formal Education Level**

EdLevel - Gender (Count)

Man

Woman



powered by IBM Cloud Pak for Data

# DISCUSSION



Technology Trends this year, and next year

Training and Reskilling Requirements

Gender Bias in Technology Jobs

Future of MS SQL

Will Python take on the top dog role in future?

# OVERALL FINDINGS & IMPLICATIONS

## Findings

Technology Changes year on year

Most respondents are male; Gender Gaps in Developer Jobs

Cloud Computing/ Virtualization Related jobs are growing

Python!

## Implications

Companies need to keep up to date with the latest tech

Hiring and Training with a Gender focus

Larger shift to Cloud Services in future

AI / ML becoming increasingly important!



# CONCLUSION



Technology Trends for this and next year

Swift and Python developers attract higher salaries.

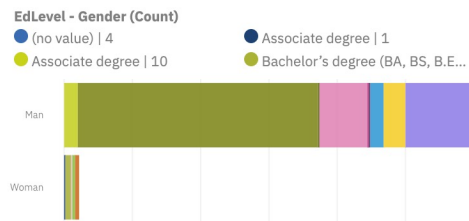
Demographic Overview – young workforce mainly 25-30

Actions to be taken

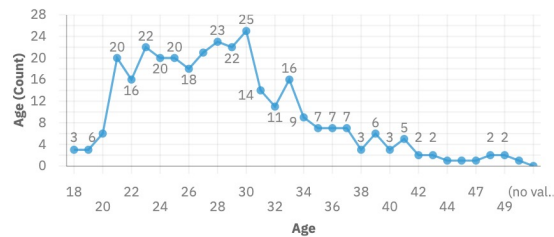
Incorporate ML to predict trends and provide further analysis in future

# APPENDIX

**Respondent count by gender, classified by Formal Education Level**

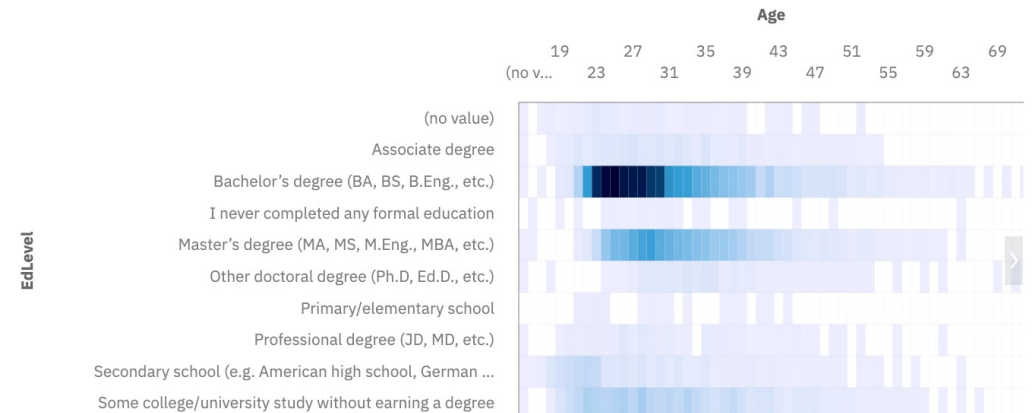


**Respondent count by Age**

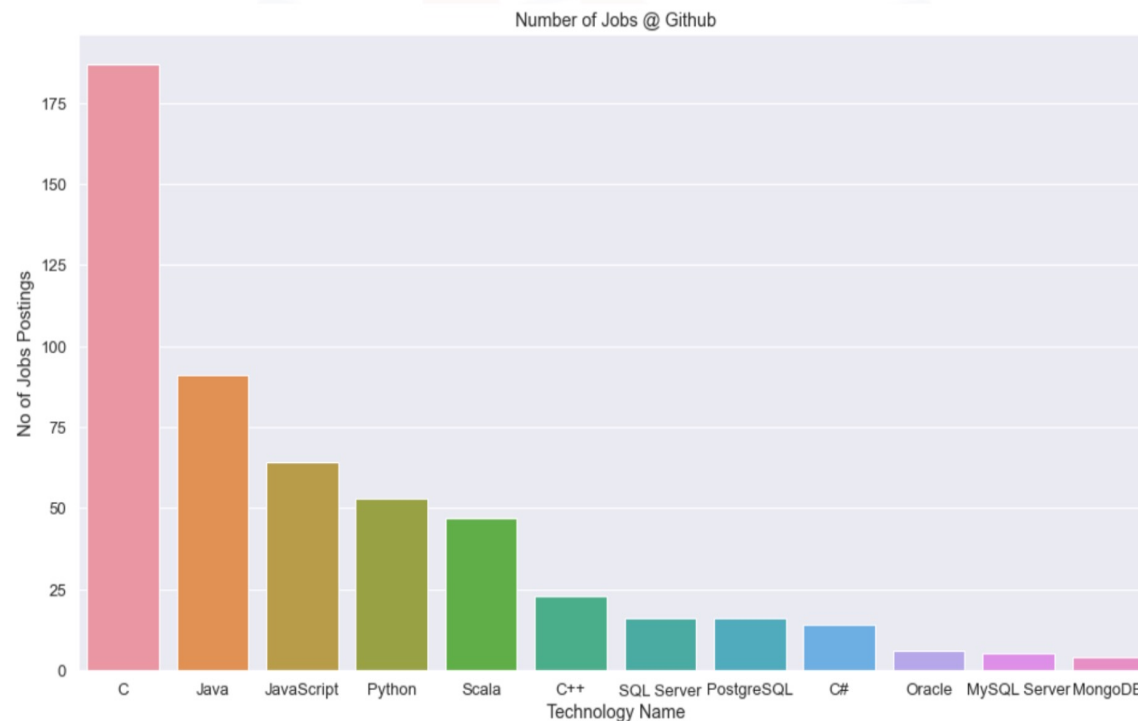


**Age (Count)**

0 474

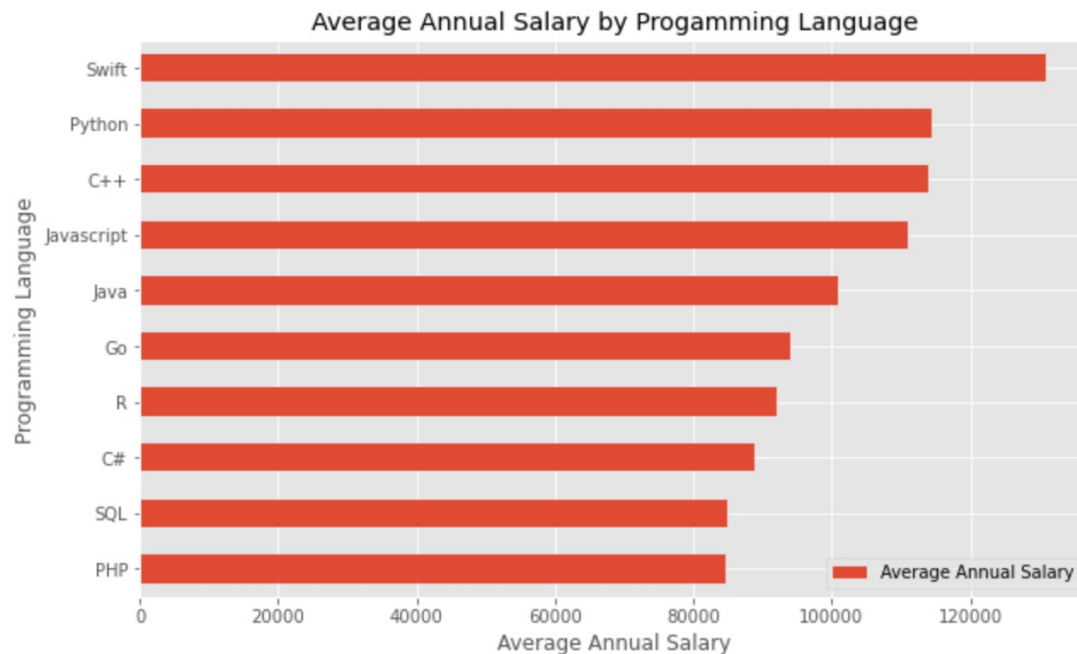


# GITHUB

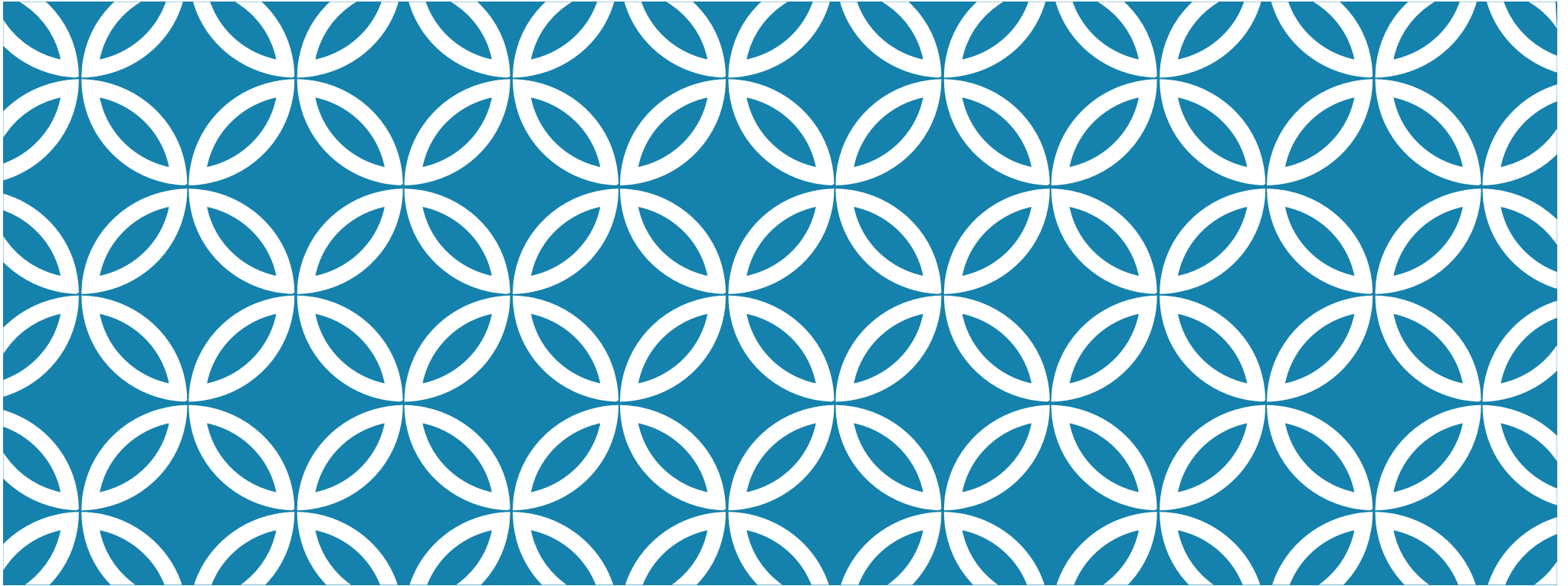


In Module 1 you have collected the job postings data using GitHub API in a file named “github-job-postings.xlsx”. Present that data using a bar chart here. Order the bar chart in the descending order of number of job postings.

# POPULAR LANGUAGES



In Module 1 you have collected the job postings data using web scraping in a file named “popular-languages.csv”. Present that data using a bar chart here. Order the bar chart in the descending order of salary.



**THE END**

LAWRENCE SHEED  
29 July 2021