

Jobs Survey Data Analysis



Lakpa Sherpa
January 15, 2024

OUTLINE



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

EXECUTIVE SUMMARY



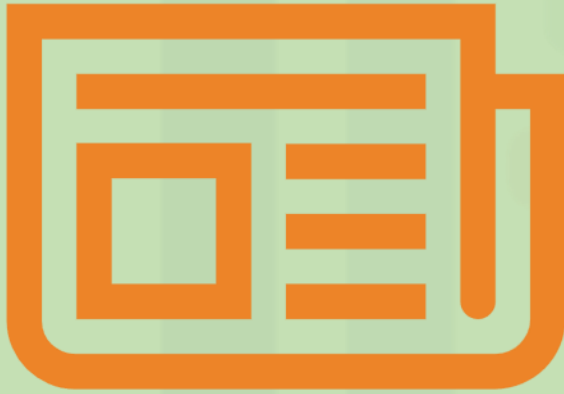
- Analyzes data to identify current emerging skills and future skills requirements.
- Methodology
 - Data Collection or Gathering
 - Data Cleaning and Analysis
 - Data Visualizations
- This report includes different types of charts to show the current and future top 10 languages, database, platform, and web frame developers responded in the survey.

INTRODUCTION



- Welcome to our insightful presentation on the evolving skills in the software industry. We've synthesized data from Stack Overflow's extensive survey, which reflects the experiences of nearly **90,000 software professionals** globally.
- Our organization is committed to staying ahead of technological advancements, and through rigorous analysis of various data sources—including job postings, training portals, and surveys—we've identified this year's key trends in programming skills.
- Join us as we reveal the most sought-after programming languages and database skills, for future skill development.

METHODOLOGY



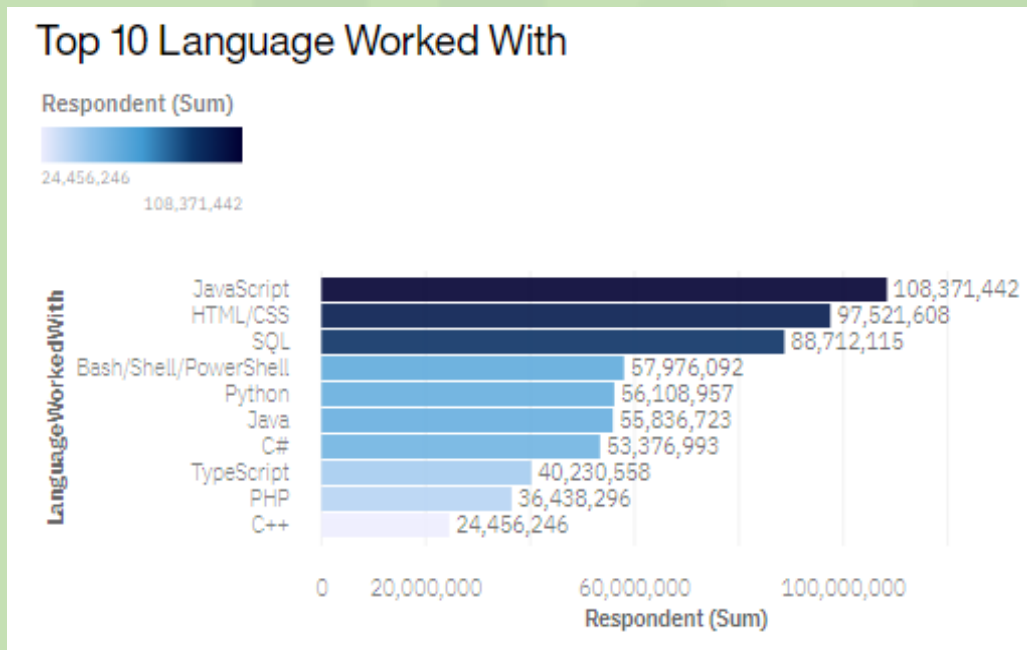
- Data Collection:
 - Using APIs and Web Scrapping
- Data Wrangling:
 - Finding and Removing Duplicated values and Missing Values
 - Replacing Missing Values with Other Value
 - Normalizing Data
- Exploratory Data Analysis:
 - Find the distribution of the data with different charts,
 - Find Outliers and Correlation between different fields.
- Data Visualization:
 - Create a Dashboard to present important charts in one place
 - Create different charts to answer questions raised by the stakeholders
 - Create different charts to find useful insights from the data

RESULTS

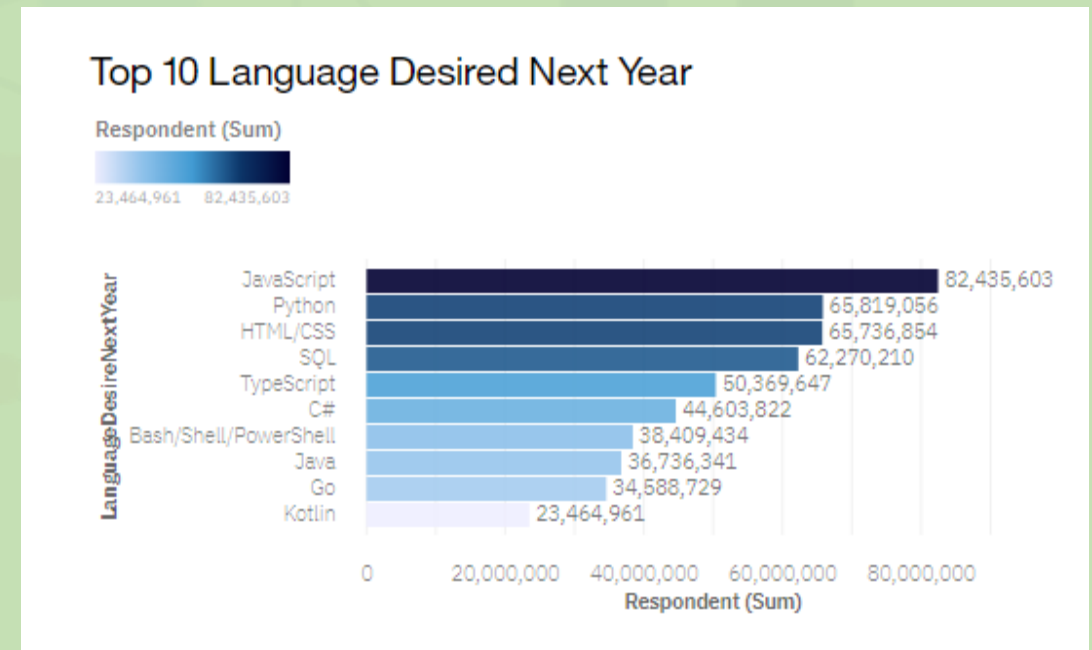
- I collected the random sample subset data from the original data for the purpose of the project.
- Then I used different techniques to clean the data and make it ready for the analysis.
- I looked for any duplicate and missing records from the dataset, then removed the duplicates and replaced the missing values from the dataset.
- I have compared Current year and Next year charts side by side to find out insights and trends about Top 10 programming language and database.
- Finally, I created a dashboard to present multiple charts in a single view to show other important findings from the dataset.

PROGRAMMING LANGUAGE TRENDS

Current Year

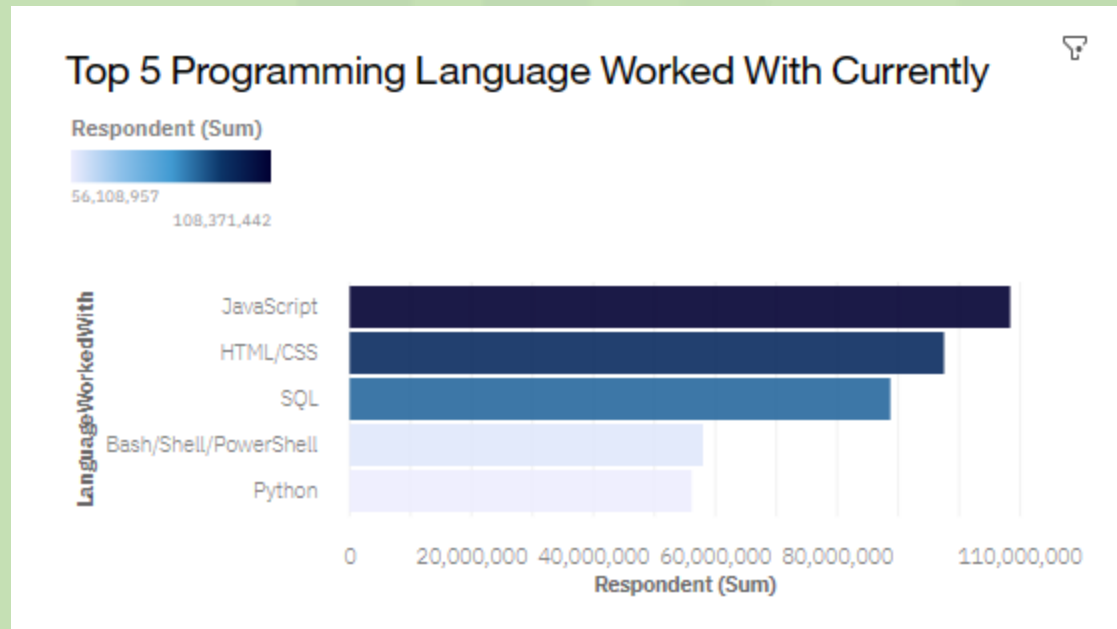


Next Year

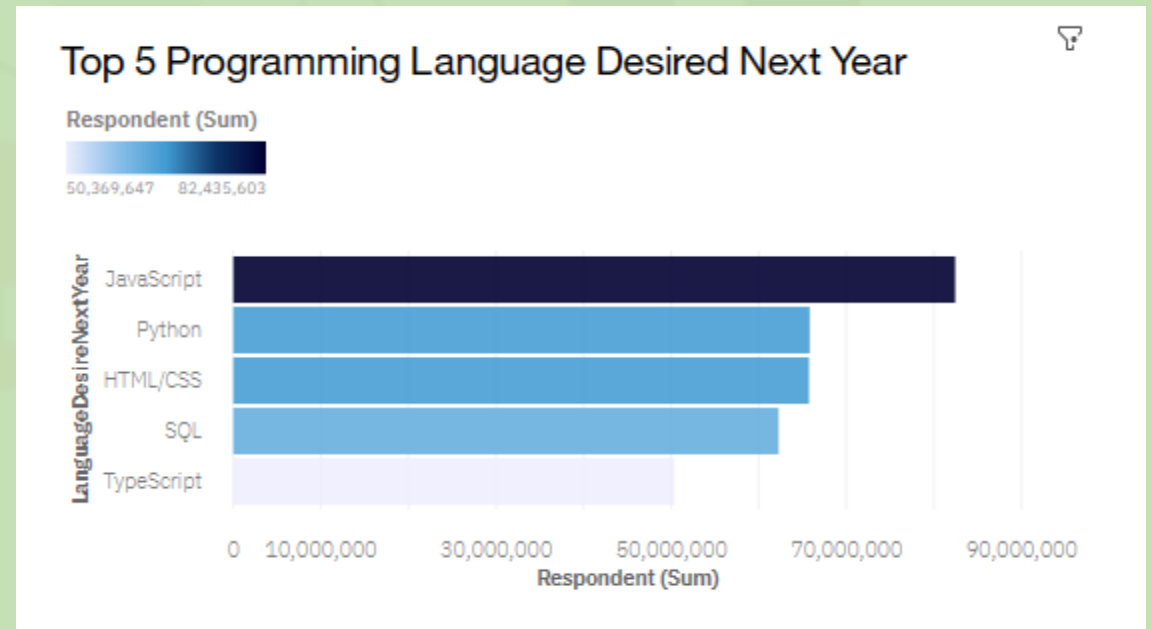


PROGRAMMING LANGUAGE TRENDS

Current Year



Next Year



PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

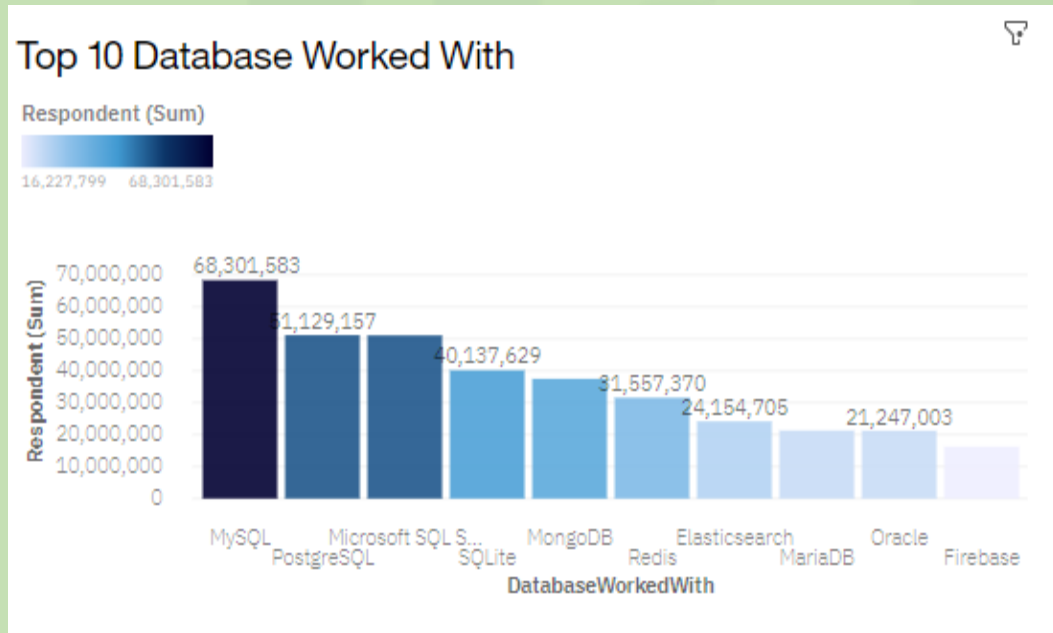
- JavaScript, HTML/CSS, and SQL is the top three preferred language for current year
- JavaScript, Python, and HTML/CSS is the top three desired language for next year

Implications

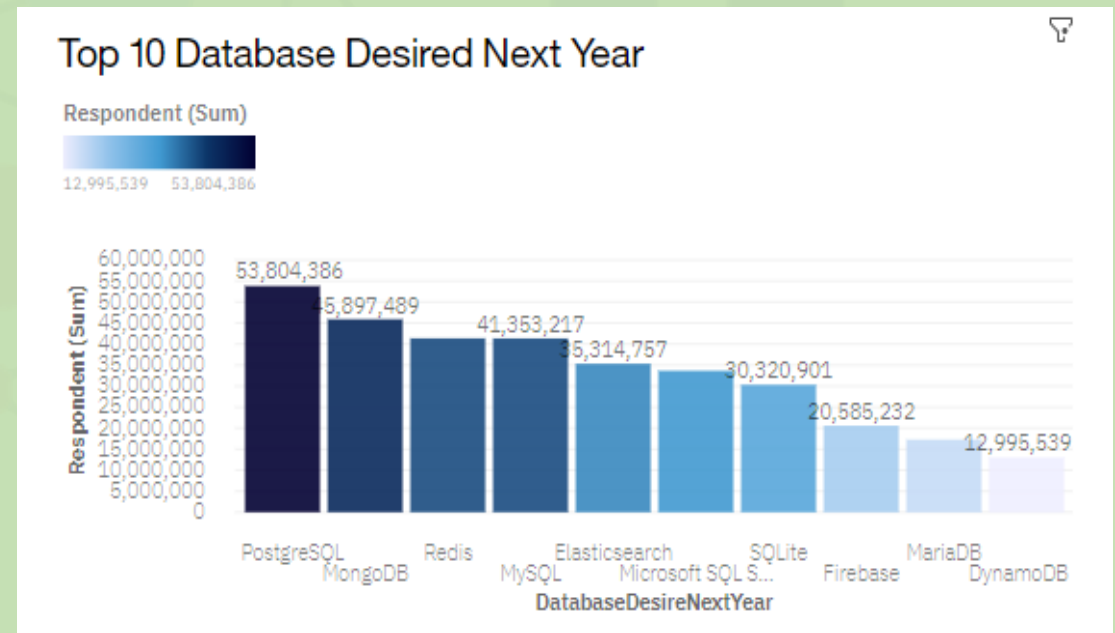
- We can understand these three programming language are commonly used by most of the developers at their work.
- Most of the developers wants to learn these languages for next year.

DATABASE TRENDS

Current Year



Next Year



DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- MySQL is the top database worked with for the current year
- PostgreSQL is the top database desired for the next year

Implications

- Most of the developers currently work with MySQL database at their job.
- Most of the developers wants to learn PostgreSQL for the next year.

DASHBOARD



- I created a dashboard using the IBM Watson. You can use the link to see the dashboard.
- The dashboard has 3 tabs, each tab contains 4 charts.

[Click Here](#)

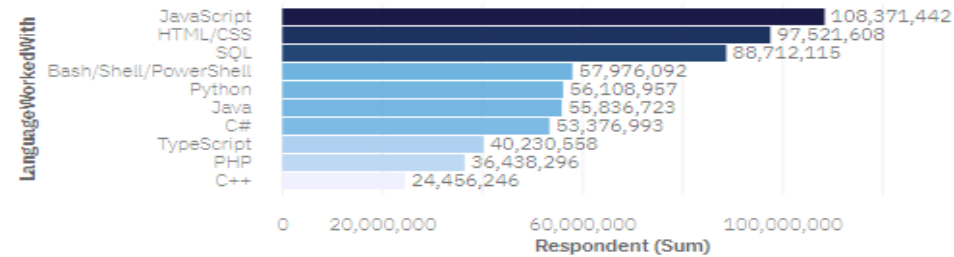
DASHBOARD TAB 1

Top 10 Language Worked With

Respondent (Sum)



24,456,246 108,371,442

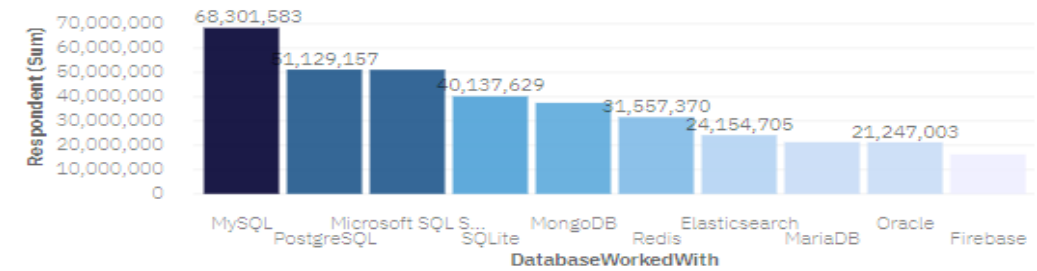


Top 10 Database Worked With

Respondent (Sum)



16,227,799 68,301,583



Platform Worked With

Respondent (Sum)



2,532,293 72,858,619

Respondent (Sum)



2,532,293 72,858,619



Top 10 Web Frame Worked With

Respondent (Sum)

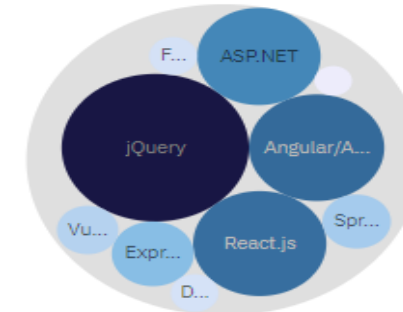


11,517,716 57,310,840

Respondent (Sum)



11,517,716 57,310,840

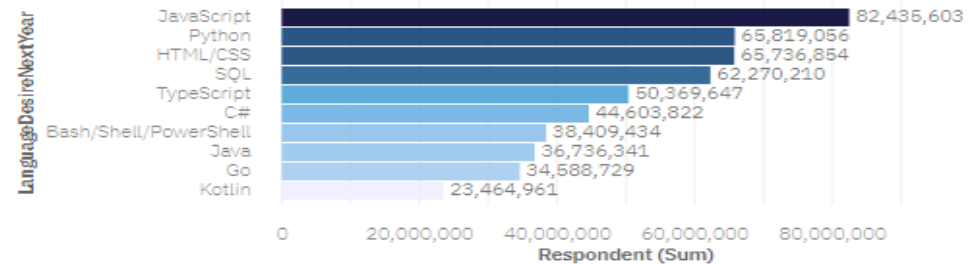


DASHBOARD TAB 2

Top 10 Language Desired Next Year

Respondent (Sum)

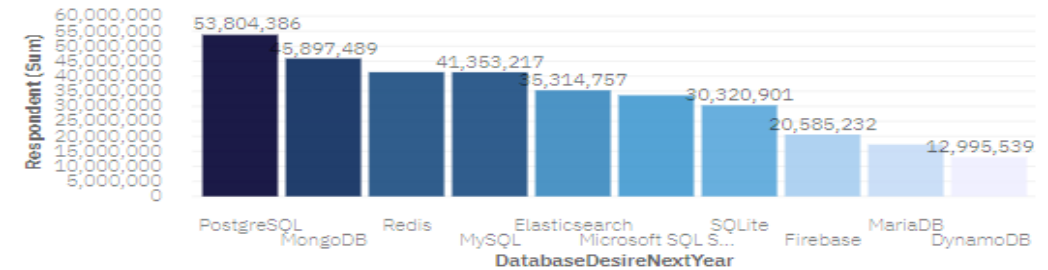
23,464,961 82,435,603



Top 10 Database Desired Next Year

Respondent (Sum)

12,995,539 53,804,386



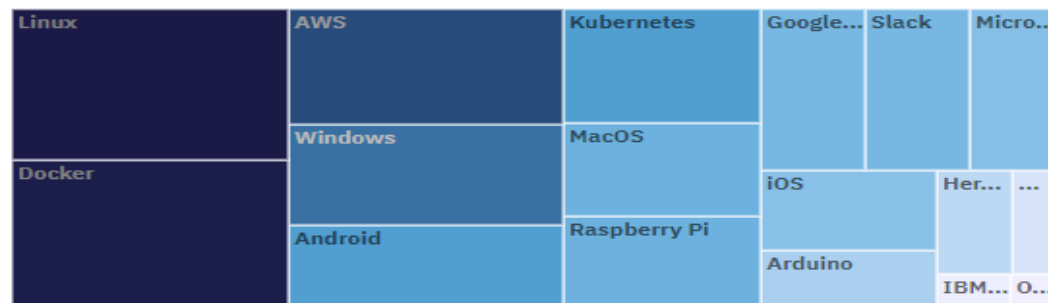
Platform Desired Next Year

Respondent (Sum)

3,752,980 73,486,491

Respondent (Sum)

3,752,980 73,486,491



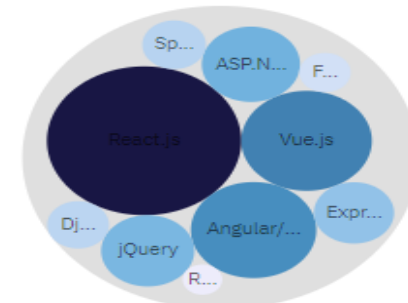
Top 10 Web Frame Desired Next Year

Respondent (Sum)

11,886,651 58,321,909

Respondent (Sum)

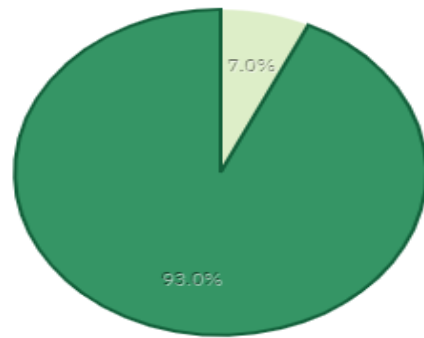
11,886,651 58,321,909



DASHBOARD TAB 3

Respondent Classified by Gender

Gender
● Woman ● Man

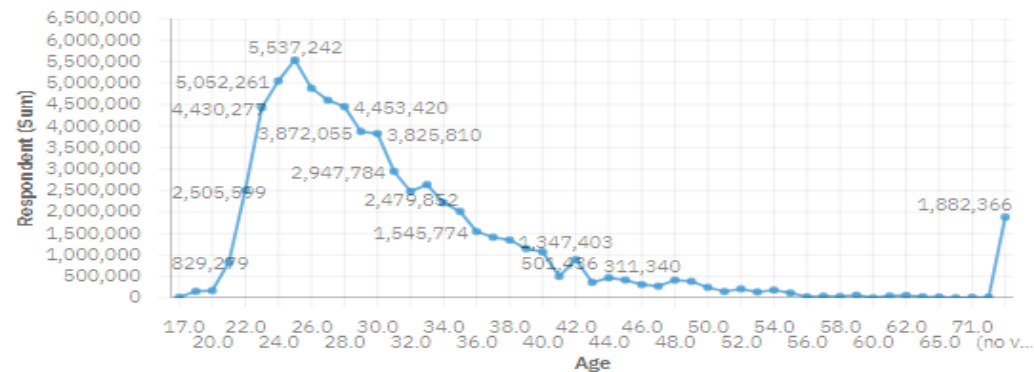


Respondent Count for Countries

Respondent (Sum)
749 20,839,703

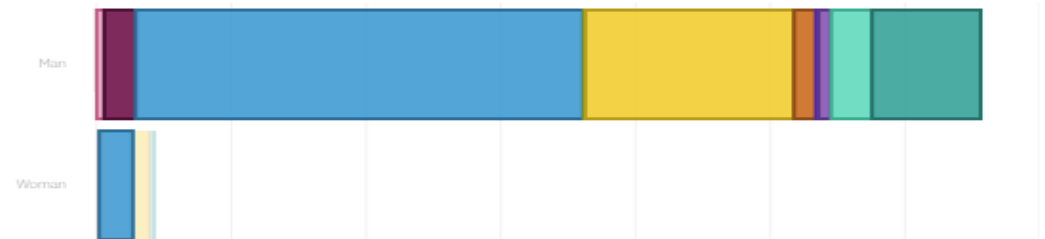


Respondent Count by Age



Respondent Count by Gender, Classified by Formal Education Level

EdLevel
● (no value) ● 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, E...



DISCUSSION

- Why is Java Script widely used by developers?

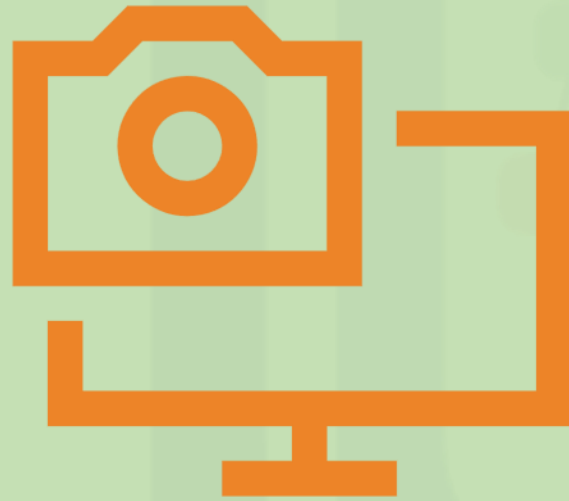


CONCLUSION



- Java Script is the widely used and desired language by most of the developers.
- Python is another desired language by most of the developers to learn for next year.
- Majority of the developers are man with a Bachelor's degree.

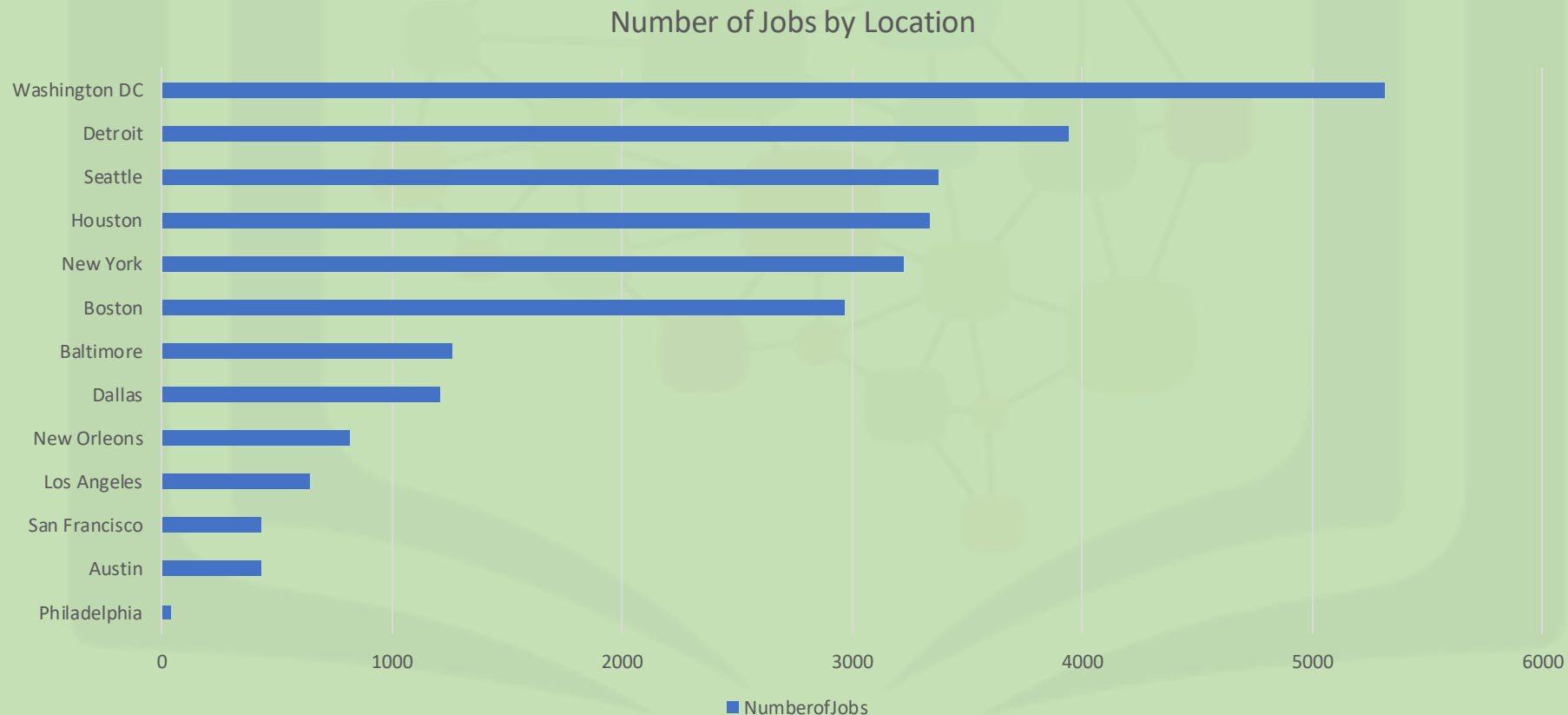
APPENDIX



- I would like to thank IBM Data Analyst course team members for helping me with necessary platform and learning materials to complete this project.
- The dataset used for this project is gathered from a comprehensive survey conducted by Stack Overflow.

JOB POSTINGS

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

