Mining the Trend of the Most Frequently Used Programming Languages and Tools of the Last Decade

Ning Chih Chang
College of Engineering and Applied Science
University of Colorado Boulder
Boulder, CO, USA
nich1985@colorado.edu

ABSTRACT

This is a preliminary exploration of the most frequently used programming languages and development tools of the past decade. A trend analysis is conducted on 10 sets of the Stack Overflow Developer Survey, which ranges from year 2015 to year 2024. The most frequently used programming languages have been analyzed by occupation, years of coding, country and years. Moreover, the frequent patterns for each survey have been identified. The finding shows that JavaScript is the most frequently used language in the past 10 years, and it has associations with JavaScript frameworks and libraries as well as other web development languages and tools.

1 Introduction

As a computer science student, knowing the programming languages and development tools that have been frequently used in industry for the past years is important because it would help me understand what skillsets I would need to gain to prepare myself for the workforce. I am interested to know the trend of the most used programming languages and tools in the past 10 years. Is there a particular collection of programming languages and tools that have been frequently used by a certain demographic? Knowing the trend of programming languages and tools would not only benefit students and professionals to keep track of the latest technology for career development, but also would create opportunities for businesses and educational institutions in the software and programming field.

2 Related Work

There are some well-established indicators for the most popular languages on the web that are free to the public. The TIOBE index gives the most popular 100 programming languages which is updated monthly and is based on the number of hits on various websites [9, 14]. The IEEE Spectrum provides a similar indicator that is generated from the popularity among the IEEE members and developers, the employer demands and the current trend [2, 3]. However, both indicators currently do not cover programming tools such as development environment or frameworks.

Studies have been done in finding trends in programming tools using Stack Overflow posts. Approximately 11 million user question and answer posts of Stack Overflow have been used to find the most popular languages, tools and topic trends between 2014 and 2015 by topic modeling [10]. Trends of NoSQL database from 2008 to 2017 have been investigated by using the normal interest score of Stack Overflow posts [6]. Moreover, posts on Stack Overflow that are related to C, Java and Python over a 14-year period have been studied and interestingly the user's country and reputation are included in trend analysis [11].

Surveys have also been used to find trends. The Stack Overflow Developer Survey has been used in investigating gender insights [12] and finding the most used programming languages and tools for 23 different IT roles [5]. Interestingly, a study that collected data from surveys and multiple sources from schools and industries in 1993, 1998 and 2003 formed a regression model that illustrated the trend [4].

I would like to explore the trends of the most frequently used programming languages and tools with demographic data to see if there are any interesting patterns. The Apriori algorithm can be used to mine frequent itemsets [7] and has been applied to find trends in major selection among university students [1] and frequent patterns in human drug addiction behavior [8]. This would be an exciting tool to use in this study to find hidden patterns.

3 Data Set

The data set of this study is retrieved from Stack Overflow Annual Developer Survey webpage [13] and surveys from 2015 to 2024 are included, which is 10 sets in total.

The number of responses for each survey from 2015 to 2024 are 26086, 56030, 51392, 98855, 88863, 64461, 83439, 73268, 89184 and 65437, respectively. The number of attributes for each survey from 2015 to 2024 are 45, 65, 153, 128, 84, 60, 40, 70, 68 and 87, respectively.

The attributes among these surveys are not exactly the same, but there are five shared concepts, which are occupation/devtype(nominal), country(nominal), years of coding/programming experience(ordinal/ratio), education(nominal), languages worked in the past year(nominal), and tools worked in the past year(nominal).

Even though the 10 surveys also share languages and tools that one wants to work in the future, these attributes are not included in the study because we would like to focus on the most used languages and technologies, instead of the languages and technologies that people most want to work with.

Another shared attribute that is not included in the dataset is compensation/salary. This is because not all 10 surveys explicitly asked the respondents to enter their salary in a specific currency. Even though some surveys asked about the daily currency used, this does not guarantee that the salary entered is in the same currency. Furthermore, more than half of the surveys have over 30% of Nan values in salary, which implies that there is a certain amount of data that cannot be interpreted. Therefore, considering the issue raised above, compensation/salary is not included in the final attributes of the merged dataset.

The preferred operating system and preferred development environment have been retrieved as two separate attributes. This is because the data cleaning process started from the 2015 survey which has these two attributes separated from the collection of languages and technologies attributes. Moreover, they are listed as two independent attributes in most of the surveys.

4 Main Techniques Applied

This study is similar to the research that analyzed the most used languages and technologies to the 2020 Stack Overflow Developer Survey [5], yet this study is different because this study includes data across 10 years as well as exploring frequent patterns.

Since we are combining 10 sets of surveys, we need to make sure that only the shared attributes are included, and they should be consistent among surveys. Data preprocessing is crucial and going to be challenging in our case.

4.1 Data Cleaning

The rule for dropping Nan value for each survey is when any row/respondent has Nan values for all tool and languages related attributes (excluding the preferred operating system and the preferred development environment). The reason for dropping data is because the goal of our study is to understand the most used languages and tools in the past 10 years, and it would not make sense to include entries with Nan values in all languages and tools related attributes. The drop percentage per survey ranges from 0.81% to 28.13% among these 10 surveys, which leaves a total of 634,598 valid data for analysis.

4.2 Data Transformation

Data transformation needs to be performed on the shared attributes before data integration. Although all the surveys have the years of coding/ programming experience attribute, some of the surveys ask for

numeric response yet some provide a selection of ranges (i.e., ordinal) for respondents. To have a consistent attribute type for the integrated data set, data discretization is performed.

Moreover, even if the share attribute is ordinal, the selections may not be the same size. For example, the answer selections of the years of coding/programming experience attribute are "less than 1 year", "1-2 years", "2-5 years", "6-10 years" and "11+ years" in the 2015's survey. However, in the 2018's survey, the available answers are "0-2 years", "3-5 years", "6-8 years" and continue every two years until "30+ years". Therefore, data transformation is needed to ensure the interval labels of the same attribute are consistent among surveys.

Multiple choice answers have been converted into lists to facilitate data analysis process. String format answers that contain multiple values separated by semicolons are extracted and transformed into a Python list. For example, Linux-based;Windows" would be transferred into ["Linux-based","Windows"].

There are multiple attributes (e.g., framework and version control system) in each survey that have been joined into a new attribute called Tools for frequent pattern analysis.

The final eight attributes of this study are Occupation, Country, Education, Years_Coding, OS (i.e., Operating system), Dev_Env (i.e., Development environment), Year and Tools. The types of these attributes are nominal, nominal, nominal, ordinal, nominal, nominal, numeric-interval and nominal, respectively. The final attributes and the corresponding attributes that have been used in each survey are shown in Figure 1.

	2015 Survey	2016 Survey
Occupation	Occupation	occupation
Country	Country	country
Education	Training & Education*	education
Years_Coding	Years IT / Programming Experience	experience_range
OS	Desktop Operating System	desktop_os
Dev_Env	Preferred text editor	dev_environment
Year	Self-created	
Tools	Current Lang & Tech*, Source control used*	tech_do

Training & Education* is a collection of attributes that represent different categorical choices, such as Training & Education: No formal training and Training & Education: BS in CS.
Current Lang & Tech* is a collection of attributes that represent different categorical choices, such as Current Lang & Tech: Android and Current Lang & Tech: Hadoop.
Source control used* is a collection of attributes that represent different categorical choices, such as Source control used: Git and Source control used: Bitkeeper.

	2017 Survey	2018 Survey
Occupation	DeveloperType, WebDeveloperType, MobileDeveloperType, NonDeveloperType	DevType
Country	Country	Country
Education	FormalEducation, MajorUndergrad, EducationTypes	FormalEducation, UndergradMajor, EducationTypes
Years_Coding	YearsProgram	YearsCoding
OS	--	OperatingSystem
Dev_Env	IDE	IDE
Year	Self-created	Self-created
Tools	HaveWorkedLanguage, HaveWorkedDatabase, HaveWorkedFramework, HaveWorkedPlatform	LanguageWorkedWith, DatabaseWorkedWith, PlatformWorkedWith, FrameworkWorkedWith

	2019 Survey	2020 Survey
Occupation	DevType	DevType
Country	Country	Country

Education	EdLevel, UndergradMajor, EduOther	EdLevel, UndergradMajor
Years_Coding	YearsCode	YearsCode
OS	OpSys	OpSys
Dev_Env	DevEnviron	--
Year	Self-created	Self-created
Tools	LanguageWorkedWith, DatabaseWorkedWith, PlatformWorkedWith, WebFrameWorkedWith, MiscTechWorkedWith	LanguageWorkedWith, DatabaseWorkedWith, PlatformWorkedWith, WebframeWorkedWith, MiscTechWorkedWith, NEWCollabToolsWorkedWith

	2021 Survey	2022 Survey
Occupation	DevType	DevType
Country	Country	Country
Education	EdLevel, LearnCode	EdLevel, LearnCode
Years_Coding	YearsCode	YearsCode
OS	OpSys	OpSysProfessional use, OpSysPersonal use
Dev_Env	NEWCollabToolsHaveWorkedWith	NEWCollabToolsHaveWorkedWith
Year	Self-created	Self-created
Tools	LanguagesHaveWorkedWith, DatabaseHaveWorkedWith, PlatformHaveWorkedWith, WebframeHaveWorkedWith, MiscTechHaveWorkedWith, ToolsTechHaveWorkedWith	LanguageHaveWorkedWith, DatabaseHaveWorkedWith, PlatformHaveWorkedWith, WebframeHaveWorkedWith, MiscTechHaveWorkedWith, ToolsTechHaveWorkedWith, OfficeStackAsyncHaveWorkedWith, OfficeStackSyncHaveWorkedWith

	2023 Survey	2024 Survey
Occupation	DevType	DevType
Country	Country	Country
Education	EdLevel, LearnCode	EdLevel, LearnCode
Years_Coding	YearsCode	YearsCode
OS	OpSysProfessional use, OpSysPersonal use	OpSysProfessional use, OpSysPersonal use
Dev_Env	NEWCollabToolsHaveWorkedWith	NEWCollabToolsHaveWorkedWith
Year	Self-created	Self-created
Tools	LanguageHaveWorkedWith, DatabaseHaveWorkedWith, PlatformHaveWorkedWith, WebframeHaveWorkedWith, MiscTechHaveWorkedWith, ToolsTechHaveWorkedWith, OfficeStackAsyncHaveWorkedWith, OfficeStackSyncHaveWorkedWith, AISearchHaveWorkedWith, AIDevHaveWorkedWith	LanguageHaveWorkedWith, DatabaseHaveWorkedWith, PlatformHaveWorkedWith, WebframeHaveWorkedWith, EmbeddedHaveWorkedWith, MiscTechHaveWorkedWith, ToolsTechHaveWorkedWith, OfficeStackAsyncHaveWorkedWith, OfficeStackSyncHaveWorkedWith, AISearchDevHaveWorkedWith

Figure 1: Attributes of the merged data and their corresponding attributes in each survey.

From Figure 1, we can see that the 2017 Survey does not have a question regarding the preferred operating system (i.e., the OS attribute) and the 2020 Survey does not ask about the preferred development environment (i.e., the Dev_Env attribute). Therefore, both OS and Dev_Env are not used in this study.

4.3 Data Integration

After ensuring that the shared attributes are consistent across 10 surveys, the shared attributes are combined into one large data set and the rest of the attributes are discarded.

4.4 Derive The Most Used Languages

To see the trend of the most used languages in the past 10 years, we can retrieve the programming languages with the highest frequency from the merged dataset. Furthermore, we can dig deeper by looking at different occupation/devtype, years of coding/programming, country and year.

4.5 Derive Frequent Patterns with Apriori

Each respondent's most worked with languages and tools are combined into a single attribute called 'Tools'. By using the Apriori algorithm, the frequent itemsets for each year are identified.

4.6 Result Validation

There will be limited validation methods for this study because the aim of this study is not establishing a prediction model, which means it would not be logical to reserve partial data for validation. The frequent patterns that we will find using Apriori may overlap with the most frequent languages and tools using descriptive statistics, which may serve some degree of validation. Also, frequency graphs can be generated to check the accuracy of the result.

5 Key Results

The Python language and Pandas, NumPy and Matplotlib libraries are used to manage and perform statistical analysis on the merged dataset as well as to provide visualization of results. The most used languages by different attributes and the frequent itemset per year are discovered.

5.1 The Most Used Languages

The result of the top 10 most used languages among 10 surveys is shown in Figure 2. We observe that JavaScript is the first most used language, which has the count number of 408,183 and takes approximately 64% of the total dataset. Interestingly, the second most used language is HTML/CSS. Along with TypeScript and PHP, there are several popular languages that are related to web application development.

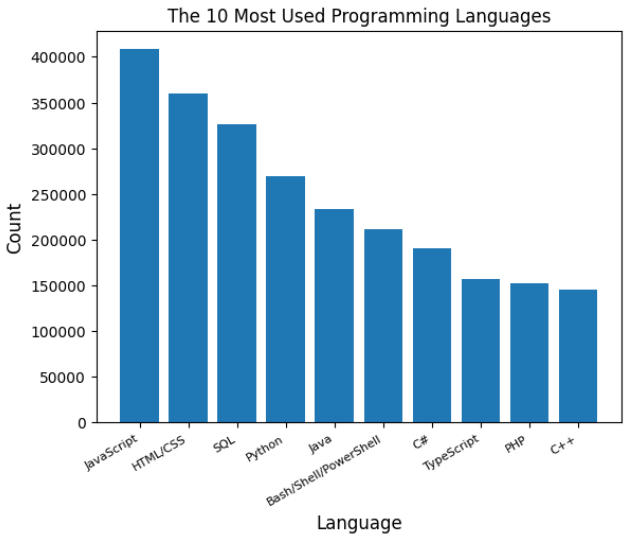
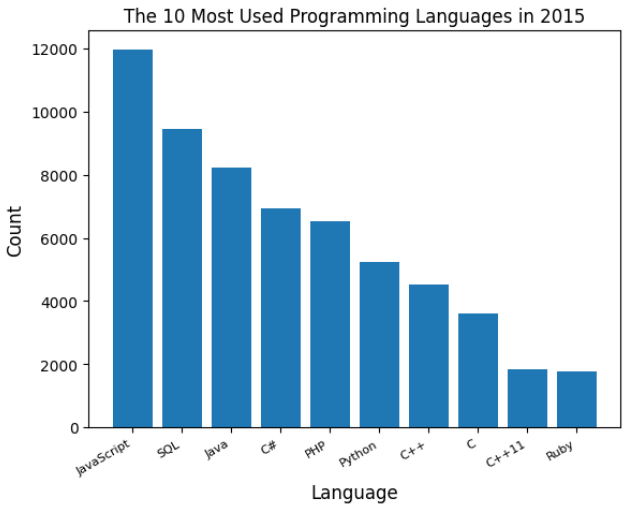


Figure 2: The 10 Most Used Programming Languages.

5.1.1 The Most Used Languages Every Five Years. There has been a change in the frequently used languages every five years in the past decade. In Figure 3, we can see the first most used language in 2015, 2019 and 2024 has been JavaScript, which shows the strong popularity of this language throughout the last 10 years. We also observed that HTML/CSS and Python have replaced Java in the top three most used languages. It is exciting to see how the popularity of Python and TypeScript has increased in the last 10 years.



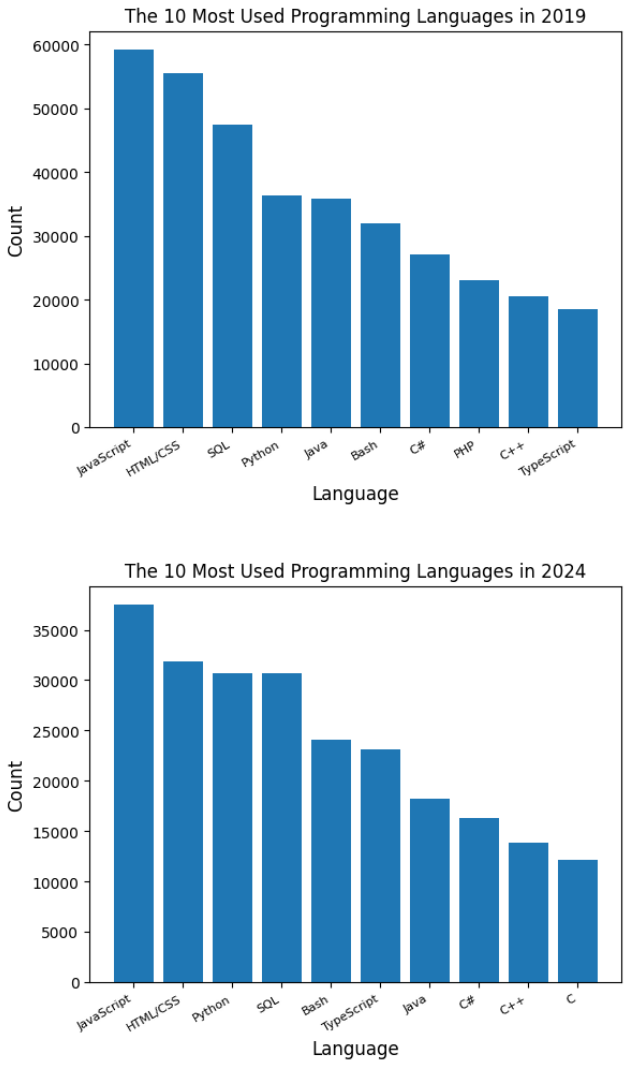


Figure 3: The 10 Most Used Programming Languages in 2015, 2019 and 2024.

5.1.2 The Most Used Languages by Years of Coding. When we investigate the number of years of coding from Figure 4, the first most used language is JavaScript for both beginners (i.e., 0 – 2 years of coding experience) and experts (i.e., over 11 years of coding experience). However, Bash is not as frequently used in beginners compared to experts.

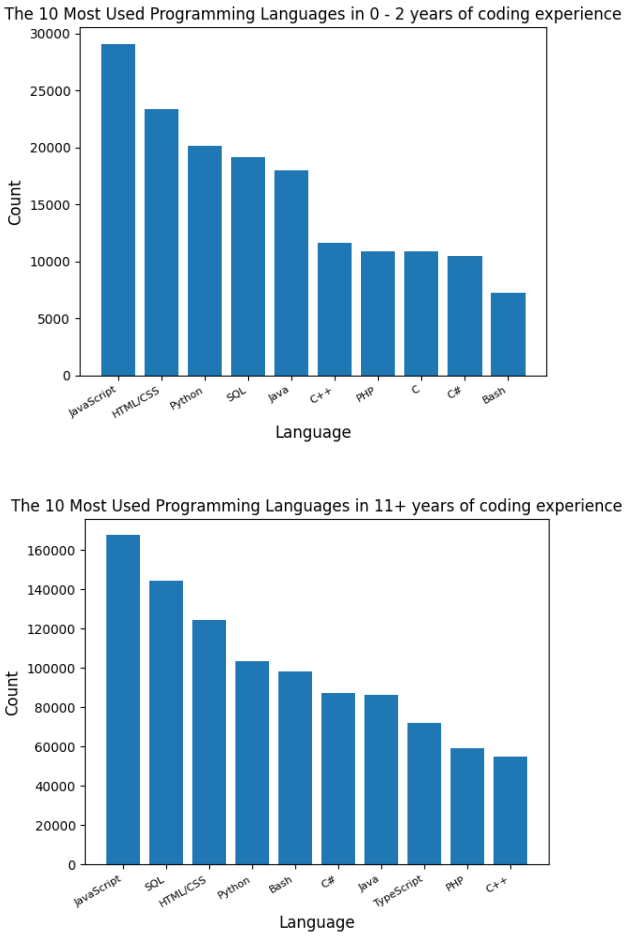
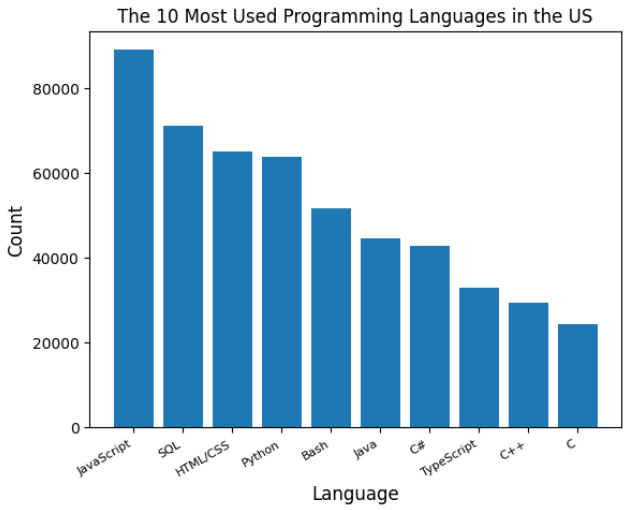


Figure 4: The 10 Most Used Programming Languages in 0-2 years and 11+ years of coding experience.

5.1.3 The Most Used Languages by Country. Most of the survey applicants are from the United States, and about 66% of the US respondents have used JavaScript. Undoubtedly, JavaScript is the first most used language for the respondents in the United States as well as for the respondents in India, which is the second largest country group in the dataset. Figure 5 shows the most popular languages of these two countries and the top three results are similar. We noticed that Java is at the fourth place of the 10 most used languages in India, which is at a much higher position than the results of the United States.



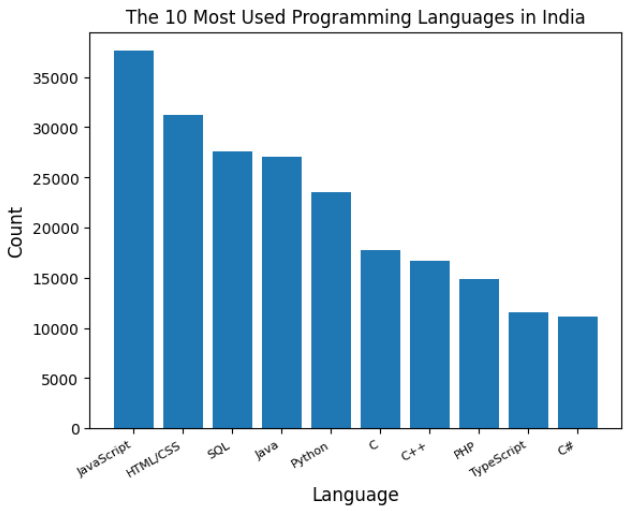
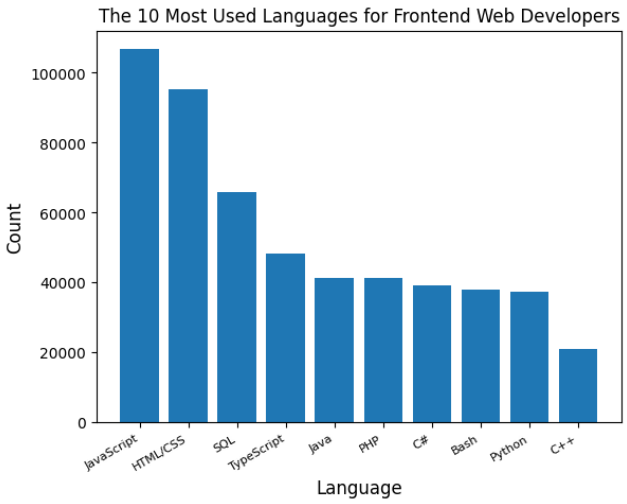
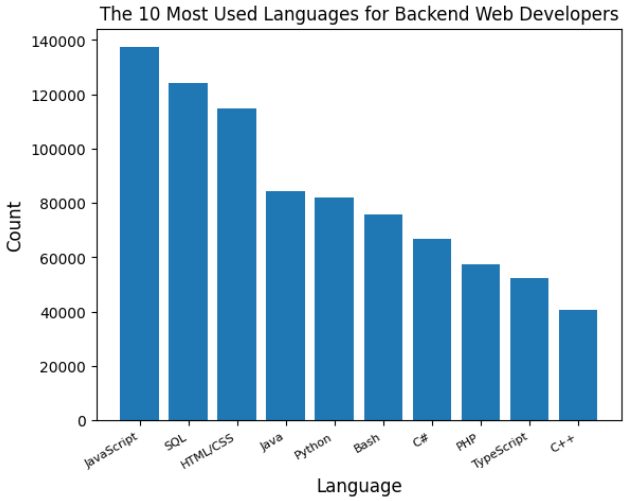


Figure 5: The 10 Most Used Programming Languages in the United States and India.

5.1.4 The Most Used Languages by Occupation. Moreover, the most used languages differ among various occupations. For web developers that work in the front-end, back-end and/or full-stack, the top 3 most used languages are still JavaScript, HTML/CSS and SQL (Figure 6).





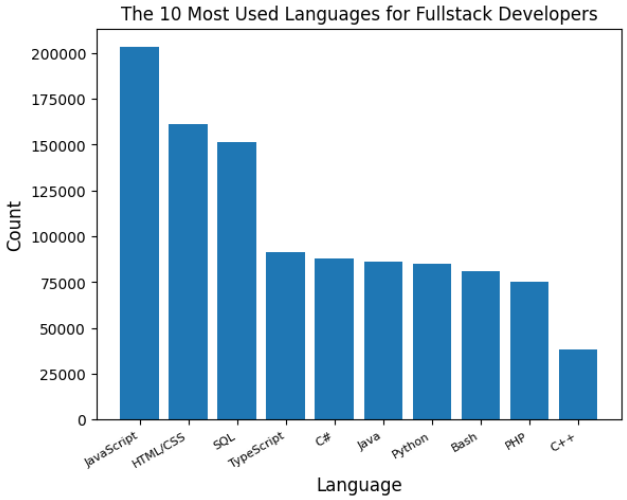


Figure 6: The 10 Most Used Programming Languages for Web Developers.

Surprisingly, the first most used language for Data Scientistic/Machine Learning Specialist and AI Developers is Python rather than JavaScript (Figure 7). In other words, Python seems to be more popular in the field of data science, machine learning and AI. Moreover, Java seems to be a popular language for Mobile Developers (Figure 8) which is the second most used language in the group.

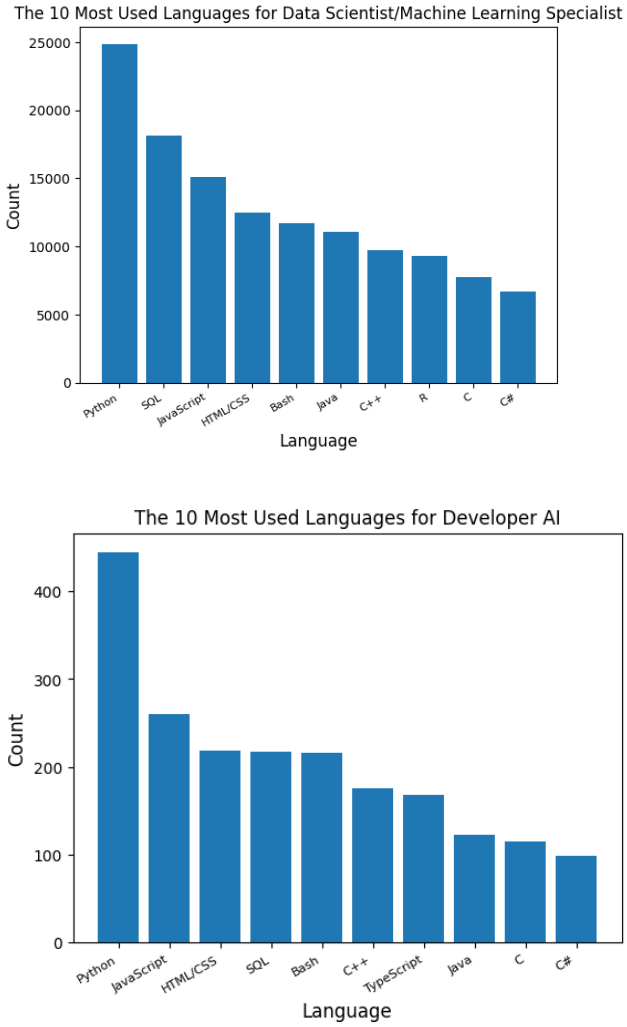


Figure 7: The 10 Most Used Programming Languages for Data Scientists/Machine Learning Specialists and AI Developers.

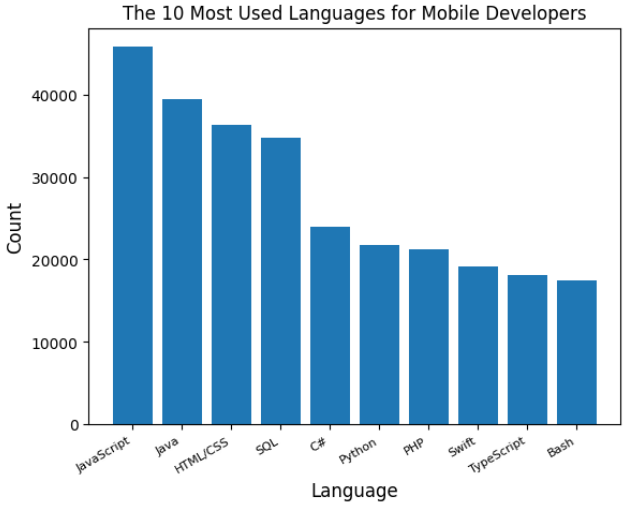


Figure 8: The 10 Most Used Programming Languages for Mobile Developers.

5.2 Frequent Patterns

The support level for finding the frequent pattern using Apriori is 20% and the confidence level is 60%. Interesting association rules and association rules with higher confidence are shown in Figure 9.

Most of the surveys result in frequent 2-itemsets and show association between JavaScript and its related framework and library (e.g., AngularJS and jQuery) as well as association between JavaScript and web application related languages and/or tools (e.g., PHP and HTML).

	Frequent Itemsets and Association Rules
2015	**[‘JavaScript, ‘PHP’]** Support (PHP => JavaScript) = 0.22 Confidence (PHP => JavaScript) = 0.77
2016	**[‘JavaScript, ‘PHP’]** Support (PHP => JavaScript) = 0.20 Confidence (PHP => JavaScript) = 0.77
2017	**[‘AngularJS’, ‘JavaScript’]** Support (AngularJS=>JavaScript) = 0.22 Confidence (AngularJS=>JavaScript) = 0.91 **[JavaScript, Node.js]** Support (Node.js => JavaScript) = 0.23 Confidence (Node.js => JavaScript) = 0.92
2018	**[‘Angular’, ‘CSS’, ‘HTML’]** Support (Angular, CSS => HTML) = 0.20 Confidence (Angular, CSS => HTML) = 0.98 Support (Angular, HTML => CSS) = 0.20 Confidence (Angular, HTML => CSS) = 0.96 **[‘Angular’, ‘HTML’, ‘JavaScript’]**

	Support (Angular, HTML => JavaScript) = 0.20 Confidence (Angular, HTML => JavaScript) = 0.95 Support (Angular, JavaScript => HTML) = 0.20 Confidence (Angular, JavaScript => HTML) = 0.90
2019	**['.NET','C#']** Support (.NET => C#) = 0.22 Confidence (.NET => C#) = 0.91 **['Angular/Angular.js', 'JavaScript']** Support (Angular/Angular.js => JavaScript) = 0.20 Confidence (Angular/Angular.js => JavaScript) = 0.90 **['JavaScript', 'Node.js']** Support (Node.js => JavaScript) = 0.30 Confidence (Node.js => JavaScript) = 0.93 **['JavaScript', 'React.js']** Support (React.js => JavaScript) = 0.21 Confidence (React.js => JavaScript) = 0.94 **['JavaScript', 'jQuery']** Support (jQuery => JavaScript) = 0.32 Confidence (jQuery => JavaScript) = 0.92 **['Microsoft SQL Server', 'SQL']** Support (Microsoft SQL Server => SQL) = 0.22 Confidence (Microsoft SQL Server => SQL) = 0.79
2020	**['Confluence', 'Jira']** Support (Confluence => Jira) = 0.27 Confidence (Confluence => Jira) = 0.92 **['JavaScript', 'Node.js']** Support (Node.js => JavaScript) = 0.33 Confidence (Node.js => JavaScript) = 0.93 **['JavaScript', 'React.js']** Support (React.js => JavaScript) = 0.24 Confidence (React.js => JavaScript) = 0.93 **['TypeScript', 'JavaScript']** Support (TypeScript => JavaScript) = 0.23 Confidence (TypeScript => JavaScript) = 0.92 **['JavaScript', 'jQuery']** Support (jQuery => JavaScript) = 0.29 Confidence (jQuery => JavaScript) = 0.93 **['Microsoft SQL Server', 'SQL']** Support (Microsoft SQL Server => SQL) = 0.22 Confidence (Microsoft SQL Server => SQL) = 0.79
2021	**['AWS', 'Docker', 'Git']** Support (AWS, Docker => Git) = 0.21 Confidence (AWS, Docker => Git) = 0.94 **['AWS', 'Git', 'JavaScript']** Support (AWS, JavaScript => Git) = 0.23 Confidence (AWS, JavaScript => Git) = 0.91
2022	**['JavaScript', 'Node.js']** Support (Node.js => JavaScript) = 0.33 Confidence (Node.js => JavaScript) = 0.91 **['JavaScript', 'React.js']** Support (React.js => JavaScript) = 0.29 Confidence (React.js => JavaScript) = 0.90 **['JavaScript', 'TypeScript']** Support (TypeScript => JavaScript) = 0.31 Confidence (TypeScript => JavaScript) = 0.90 **['JavaScript', 'jQuery']** Support (jQuery => JavaScript) = 0.20 Confidence (jQuery => JavaScript) = 0.92
2023	**['ChatGPT', 'GitHub Copilot']** Support (GitHub Copilot => ChatGPT) = 0.21 Confidence (GitHub Copilot => ChatGPT) = 0.86 **['HTML/CSS', 'JavaScript']** Support (HTML/CSS => JavaScript) = 0.46 Confidence (HTML/CSS => JavaScript) = 0.88 **['JavaScript', 'Node.js']** Support (Node.js => JavaScript) = 0.31 Confidence (Node.js => JavaScript) = 0.91 **['JavaScript', 'React']** Support (React => JavaScript) = 0.29 Confidence (React => JavaScript) = 0.89 **['Pip', 'Python']** Support (Pip => Python) = 0.23 Confidence (Pip => Python) = 0.89
2024	**['ChatGPT', 'GitHub Copilot']** Support (GitHub Copilot => ChatGPT) = 0.26 Confidence (GitHub Copilot => ChatGPT) = 0.82 **['HTML/CSS', 'JavaScript']** Support (HTML/CSS => JavaScript) = 0.46 Confidence (HTML/CSS => JavaScript) = 0.88 **['JavaScript', 'Node.js']** Support (Node.js => JavaScript) = 0.29 Confidence (Node.js => JavaScript) = 0.90 **['JavaScript', 'React']** Support (React => JavaScript) = 0.28 Confidence (React => JavaScript) = 0.88 **['Pip', 'Python']** Support (Pip => Python) = 0.25 Confidence (Pip => Python) = 0.88

Figure 9: The partial result of the frequent itemsets and association rules of the most frequently used language and tools in each survey

It is also worthwhile noting that ChatGPT and GitHub Copilot are associated based on the result generated from the Apriori algorithm. Interestingly, both tools share the same developer (i.e., OpenAI) [15].

6 Applications

The findings of this study can be used to generate a study guide for Computer Science/Information Technology and related major students. From various results, we can see that JavaScript is the first most used language among many occupations, years of coding

and countries in the past decade. Additionally, SQL has appeared numerously in the top five most used languages of our analysis, which suggests that database management is extremely important.

Moreover, when exploring different job positions, such as Data Scientist or Mobile Developer, we understand that Python and Java are also important languages to learn, which allows us to customize the study guide to match different career paths of students. The findings would also give directions to developers who intend to switch career path and/or to strengthen their skillset.

Additionally, we observed that Python and TypeScript have become more popular in the past 10 years. This finding could help educators to design suitable courses. Businesses should also keep up with the latest technology in order to stay competitive in the industry. Moreover, this would also be a good opportunity for researchers who focus on new language development to investigate the cause of this shift.

Online surveys might be prone to sampling bias [16], which implies that the result of this study may not represent the population that we intend to understand. However, this is still a good preliminary trend analysis that may guide future research that incorporates random sampling techniques.

REFERENCES

[1] Almahdi Alshareef, Salem Ahmida, Azuraliza Abu Bakar, Abdul Razak Hamdan, and Mohammed Alweshah. 2015. Mining survey data on university students to determine trends in the selection of majors. In *2015 Science and Information Conference (SAI).* 586-590. https://doi.org/10.1109/SAI.2015.7237202

[2] Stephen Cass. 2024. The Top Programming Languages 2024: Typescript and Rust are among the rising stars. (August 2024). Retrieved October 14, 2024 from https://spectrum.ieee.org/top-programming-languages-2024

[3] Stephen Cass. 2024. The Top Programming Languages Methodology 2024: How we construct the rankings. (August 2024). Retrieved October 14, 2024 from https://spectrum.ieee.org/top-programming-languages-methodology-2024

[4] Yaofei Chen, Rose Dios, Ali Mili, Lan Wu and Kefei Wang. 2005. An empirical study of programming language trends. *IEEE Software* 22, 3 (June 2005), 72-79. https://doi.org/10.1109/MS.2005.55

[5] Oluwaseun Alexander Dada, George Obaido, Ismaila Temitayo Sanusi, Kehinde Aruleba, and Abdullahi Abubakar Yunusa. 2023. Hidden gold for IT professionals, educators, and students: Insights from stack overflow survey. *IEEE Transactions on Computational Social Systems* 10, 2 (April 2023), 795-806. https://doi.org/10.1109/TCSS.2022.3151130

[6] Harshit Gujral, Abhinav Sharma and Parmeet Kaur. 2018. Empirical investigation of trends in NoSQL-based big-data solutions in the last decade. In *2018 Eleventh International Conference on Contemporary Computing (IC3)*. 1-3. https://doi.org/10.1109/IC3.2018.8530582

[7] Jiawei Han, Micheline Kamber, and Jian Pei. 2012. *Data Mining: Concepts and Techniques* (3rd. ed.). Morgan Kaufmann Publishers, Waltham, MA.

[8] Md. Mehedi Hassan, Sadika Zaman, Swarnali Mollick, Md. Mahedi Hassan, M. Raihan, Chetna Kaushal, and Rajat Bhardwaj. 2023. An efficient Apriori algorithm for frequent pattern in human intoxication data. *Innov. Syst. Softw. Eng.* 19, 1 (March 2023), 61-69. https://doi.org/10.1007/s11334-022-00523-w

[9] Paul Jansen. 2024. TIOBE index for October 2024. (Oct. 2024). Retrieved October 14, 2024 from https://www.tiobe.com/tiobe-index/

[10] Vishal Johri and Srividya Bansal. 2018. Identifying trends in technologies and programming languages using topic modeling. In *2018 IEEE 12th International Conference on Semantic Computing (ICSC)*. 391-396. https://doi.org/10.1109/ICSC.2018.00078

[11] Yusuf Ade Putra Perdana and Yusuf Sulistyo Nugroho. 2024. An empirical study of the popularity of three programming languages on stack overflow: Trends, solutions, and users. *AIP Conf. Proc.* 2926, 1 (Jan. 2024), 020092-1-020092-8. https://doi.org/10.1063/5.0182877

[12] Karina Kohl Silveira, Soraia Musse, Isabel H. Manssour, Renata Vieira, and Rafael Prikladnicki. 2019. Confidence in programming skills: Gender insights from StackOverflow developers survey. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)*. 234-235. https://doi.org/10.1109/ICSE-Companion.2019.00091

[13] Stack Overflow. 2024. Stack Overflow Annual Developer Survey. Retrieved October 14, 2024 from https://survey.stackoverflow.co/

[14] TIOBE Software BV. 2024. TIOBE Programming Community Index Definition. Retrieved October 14, 2024 from https://www.tiobe.com/tiobe-index/programminglanguages_definition/

[15] Wikipedia. 2024. GitHub Copilot. Retrieved December 7, 2024 from https://en.wikipedia.org/w/index.php?title=GitHub_Copilot&oldid=1253874251

[16] Wikipedia. 2024. Sampling bias. Retrieved December 7, 2024 from https://en.wikipedia.org/w/index.php?title=Sampling_bias&oldid=1258039711