Lawrence Menegus

CPSC 380

25 April 2023

The Popularity of Programming Languages Over Time

**Introduction**

Over time, the popularity of programming languages has been a topic of great interest among developers, students, and researchers in the field of Computer Science. The objective of this paper is to analyze the popularity of programming languages using Python Data Science techniques. This project will explore trends and patterns in programming language popularity across different data sources and time periods and identify factors that may contribute to changes in popularity. The main data sources for this project are the Stack Overflow Annual Developer Survey and the TIOBE Index, which provide comprehensive information on the programming languages used by developers and their popularity based on search engine results. Using Python analysis tools such as Pandas, NumPy, Matplotlib, Seaborn and other techniques, this paper will conduct a statistical analysis to identify correlations, trends, and variations within the data. The outcomes of this project will provide insights f active computer science students and individuals in the computer science field, helping to shed light on the popularity of programming languages and its implications for the industry.

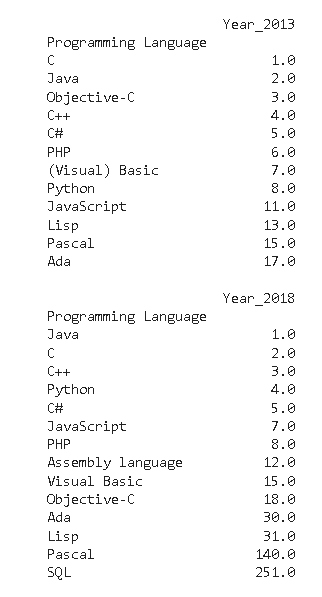
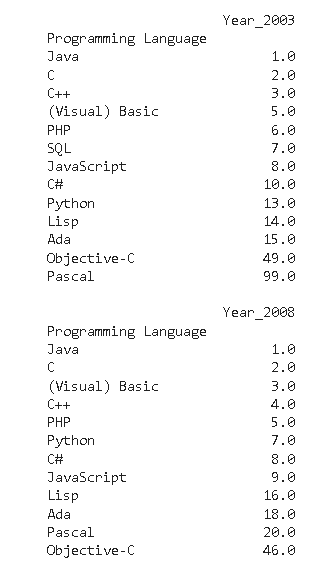
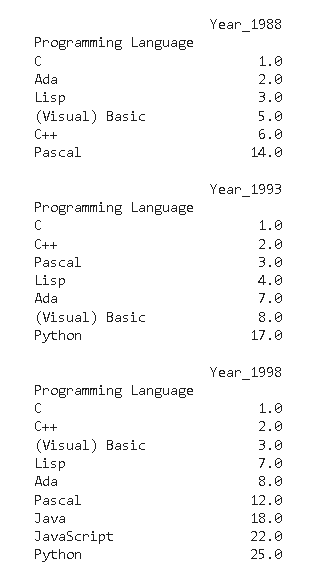
**Data**

For this project, I utilized a total of seven datasets. Two of the datasets were from The Tiobe Index while the remaining five were obtained from the Stack Overflow annual survey. The Stack Overflow datasets provided an abundance of valuable information, consisting of over hundreds of thousands of responses to a survey conducted between 2019 and 2022. These datasets contained insights into programming languages from individuals in the Computer Science industry and were comprised of a maximum of 84 columns and at least 90,000 rows per dataset.

To extract the necessary information for my analysis, I employed a series of techniques including data cleaning, joining, concatenating, and selecting specific columns such as Employment Status, years of coding experience, Education Level, Country, Language users Worked With, Language users Desired, Developer Type, and total Compensation for an Individual . I discarded any incomplete rows containing NaN and dropped irrelevant columns to streamline the data processing. This enabled me to extract thousands of rows and create a range of data visualizations such as Pie graphs, Scatter plot graphs, Line graphs, Horizontal Bar graphs, and Bar graphs. These visualizations were generated using techniques such as grouping, merging, masking, exploding, and splitting the data.

The other two datasets from the Tiobe Index were equally valuable, with the first providing information on the most popular programming languages as of March 2023. Employing the same data cleaning techniques, I extracted the ranking of each language and the percentage of people using each language compared to the others. The second dataset displayed the rankings of programming languages from 1988-2018, with Figure 1.0 illustrating the rankings for each of the years. To analyze this data, I separated the datasets for each year and removed NaN values from programming languages that did not exist.

**Figure 1.0**

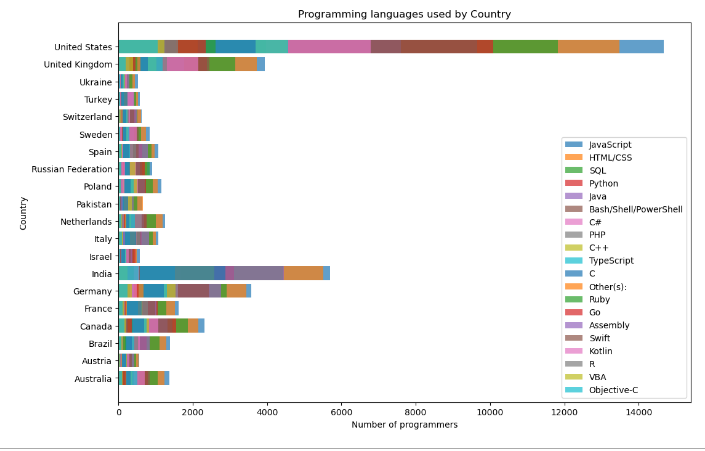


**Results**

This project has uncovered fascinating insights into the changing popularity of programming languages over time. It is evident that the most popular programming languages have varied over different time periods. For instance, in the 1980s, C and Pascal were the leading languages, while in the 1990s, C++, Java, and Perl emerged as the most widely used languages. In the 2000s, Python, Ruby, and PHP gained popularity, whereas in the 2010s, JavaScript, Java, and Python were the dominant languages.

My analysis revealed a strong correlation between the country of the programmer and the popular programming languages used. However through my analysis I found that the programming languages varied from country to country. However I did notice that the major of individuals who did take this survey were from the United States, I created Figure 1.1, which displays the Programming Languages Used based on Country. The visualization highlights the variations in the number of programmers and the countries they are from.

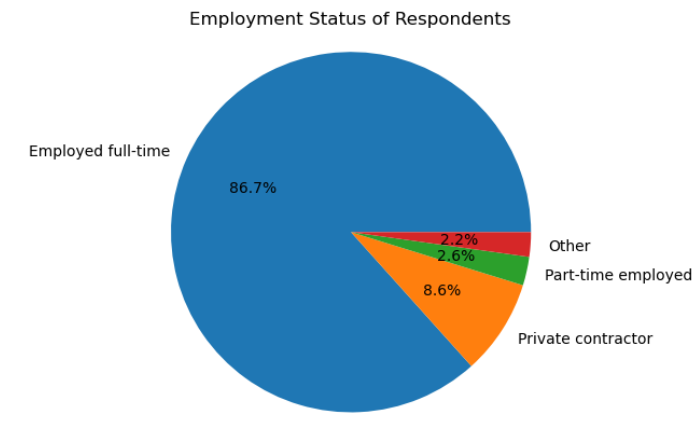
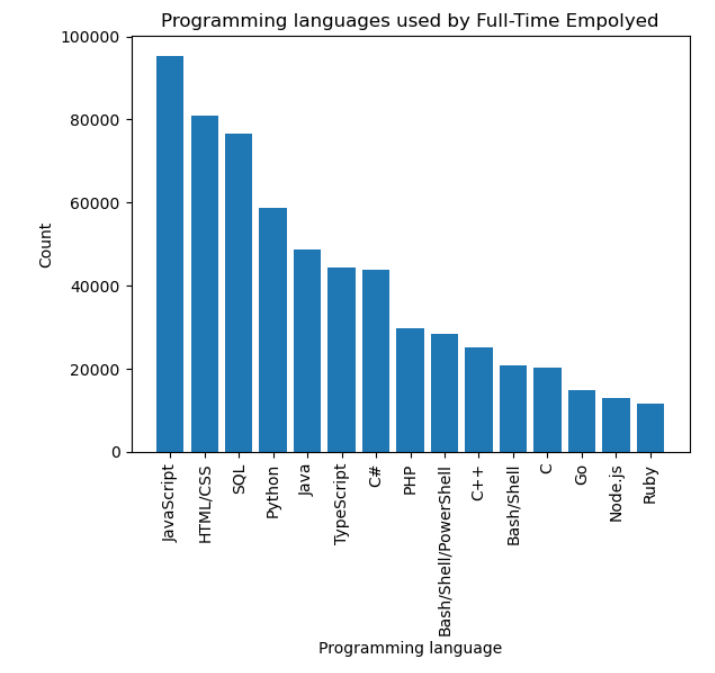
**Figure 1.1**



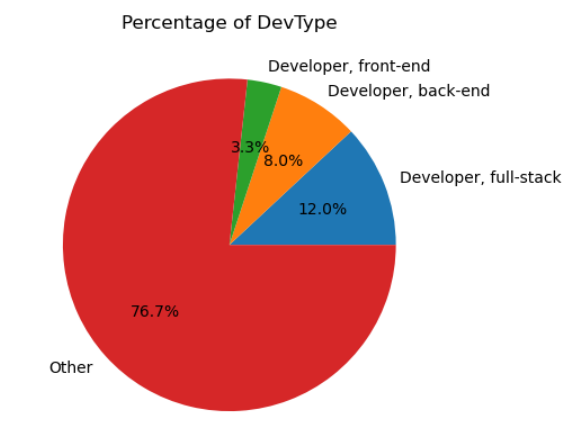
Drawing from the insightful survey datasets, my analysis revealed that full-time employed developers tended to prefer certain programming languages over others.\, with JavaScript, HTML/CSS, SQL, and Python emerging as the most used languages, accounting for a remarkable 86.7% of survey participants. To present these findings in a clear and visually appealing manner, I created Figure 1.2, which displays the distribution of employment statuses among the thousands of respondents in the survey.

Additionally, I wanted to identify the top ten most widely used programming languages among full-time employed individuals. Figure 1.3 presents these findings, allowing viewers to gain a quick understanding of the most popular languages in this category.

**Figure 1.2**  **Figure 1.3**

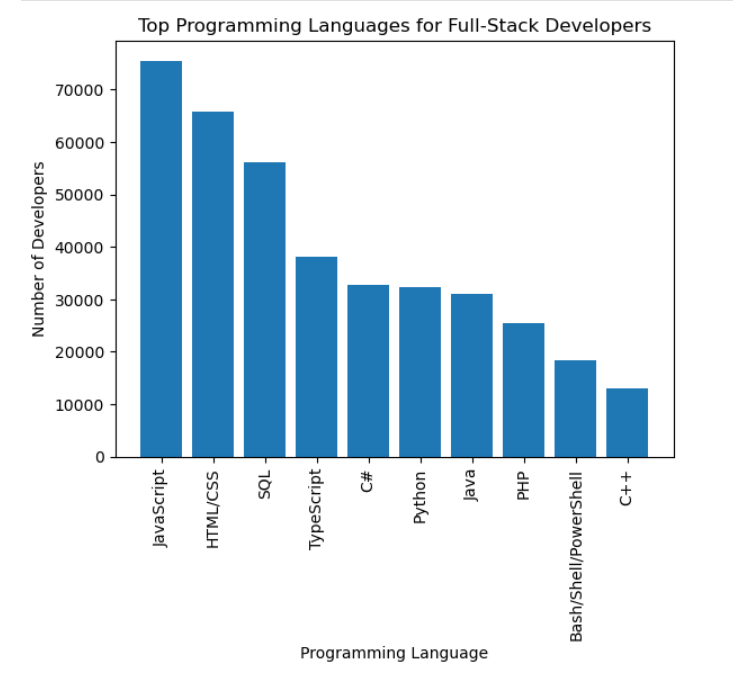
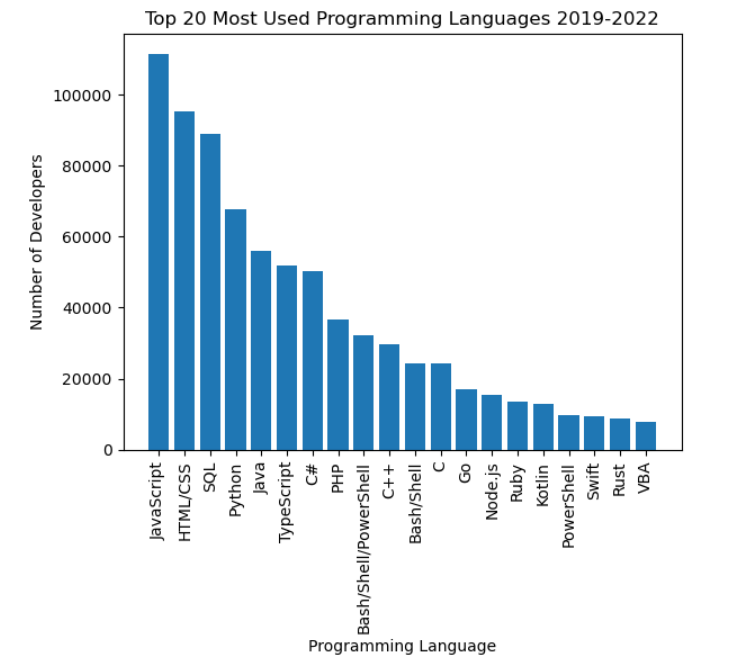


Developer type emerged as a crucial factor influencing the popularity of programming languages, with 76.7% of the respondents in the Annual Survey identifying as at least one of the 100 diverse types of developer occupations. For better visualization of these findings, we grouped any developer type percentage below 3% as "Other." Our analysis showed that Full-Stack Developers had the highest percentage at 12%, followed by Back-end Developers at 8%, and Front-end Developers at 3.3%. Figure 1.4 presents these percentages of Dev-type based on the respondents of the Annual Survey.

 My Analysis found that the top three programming languages used were the same among Full-Stack Developers. We represented the top twenty languages used between 2019-2020 in Figure 1.5, while Figure 1.6 illustrates the top ten languages used by Full-Stack Developers between 2019-2022. These visual aids enable viewers to quickly grasp the most used languages within these categories, facilitating a more comprehensive understanding of the relationship between programming languages and developer types.

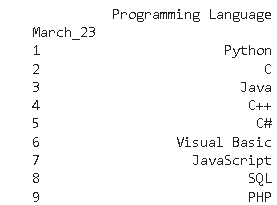
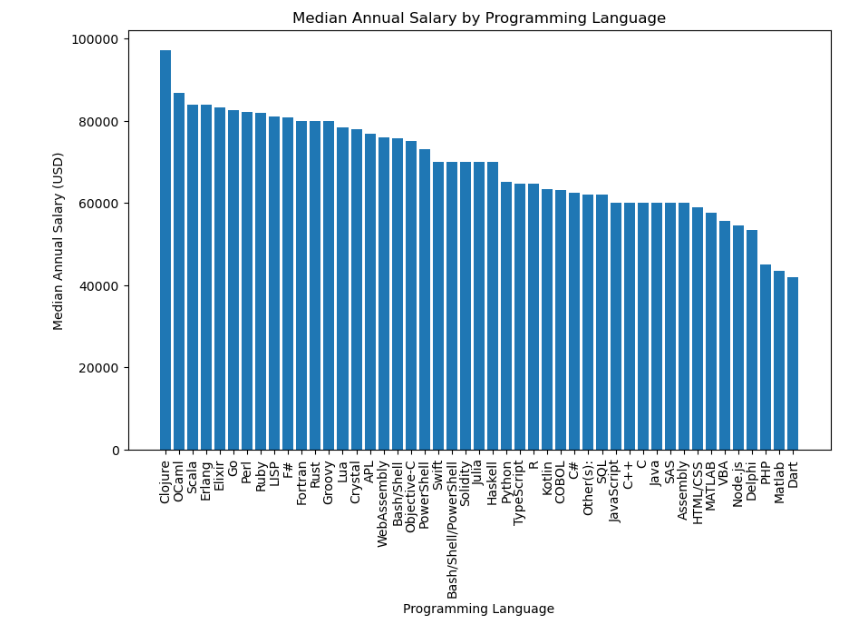
**Figure 1.4**

**Figure 1.5** **Figure 1.6**



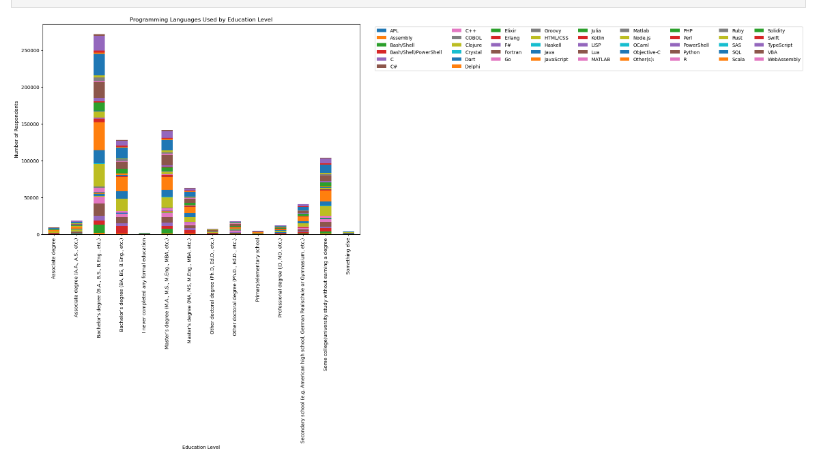
As part of my investigation into the Programming Languages, I also examined whether a correlation exists between salary and programming languages. Upon conducting my analysis, I discovered that the popularity of a programming language cannot be solely attributed to the salary of the respondents. In fact, those earning a salary of $80,000 or more utilized programming languages outside of the top nine most popular options. Referencing Figure 1.7, we can see the top nine programming languages that are widely used in March of 2023. Additionally, Figure 1.8 displays the programming languages that correspond to salary levels.

**Figure 1.7**  **Figure 1.8**



I also delved into the relationship between education level and the programming languages utilized. Through my analysis, it was revealed that individuals possessing a bachelor’s degree primarily utilized programming languages such as JavaScript, Java, C++, Python, PowerShell, Bash, and PHP, alongside a few other languages with minimal usage percentages. Remarkably, individuals holding a master’s degree and those with limited college education but without earning a degree also used the same set of languages. Moreover, respondents with an associate degree, those without any formal education, individuals who completed their doctorate, and even those who only attained a Grade School or High School education utilized the same set of programming languages as those with a bachelor’s or master’s degree. These findings are illustrated in Figure 1.9.

**Figure 1.9**



**Discussion**

Exploring the popularity of programming languages over time is a fascinating research topic. My analysis highlighted that programming languages are not static and can vary from year to year. I observed the growth trajectory of Python, which initially had the least popularity but gradually gained significant traction over the years. Through my analysis, I identified multiple factors that contributed to this progression. As illustrated in Figure 1.0 and Figure 1.7, Python's usage has surged over time, eventually becoming the most popular programming language by 2023.

This project has utilized Python data science techniques to analyze the popularity of programming languages over time. It analyzed multiple datasets from the Stack Overflow annual survey and The TIOBE Index to identify trends, patterns, and correlations in the data. The analysis revealed that the most popular programming languages have varied over different time periods, with factors such as country, employment status, and developer type influencing language preference. The findings of this project provide valuable insights for active computer science students and individuals in the computer science field, highlighting the implications of programming language popularity for the industry.