

Trends and Predictions in Programming Language Popularity

An analysis of programming language trends over 2 decades

By Maaz Khalil

For Problem Solving With Data Taught By Dr. Kiran Garimella

17:610:562:90

Spring 2024

Background

Context and Significance

In the ever changing landscape of tech, programming languages play a huge role in shaping innovations along with career plans. As new languages emerge and existing ones evolve, understanding these trends becomes crucial for educators, industry titans, and aspiring developers

Historical Context:

Programming languages have seen varied waves of popularity driven by technological advancement, community support, hype trains, and shifts in industry. The significance of tracking these changes helps understand what industry participants should look towards when deciding what programming language to use

Background Cont.

Prior Research as a Reference

Previous studies, notably by Orłowska et al. (2021), analyzed programming languages based on usage data from GitHub and StackOverflow, highlighting the association between language popularity and community engagement (“Programming Languages: A Usage-based Statistical Analysis and Visualization”, Aleksandra Orłowska, Christos Chrysoulas, Zakwan Jaroucheh, and Xiaodong Liu, 2021).

Data Sources from Kaggle and Tools

Most Popular Programming Languages since 2004

Dataset that includes monthly popularity percentages for various programming languages from 2004 to 2023, offering a long data format view of trends in the tech industry

Detailed Programming Language Data

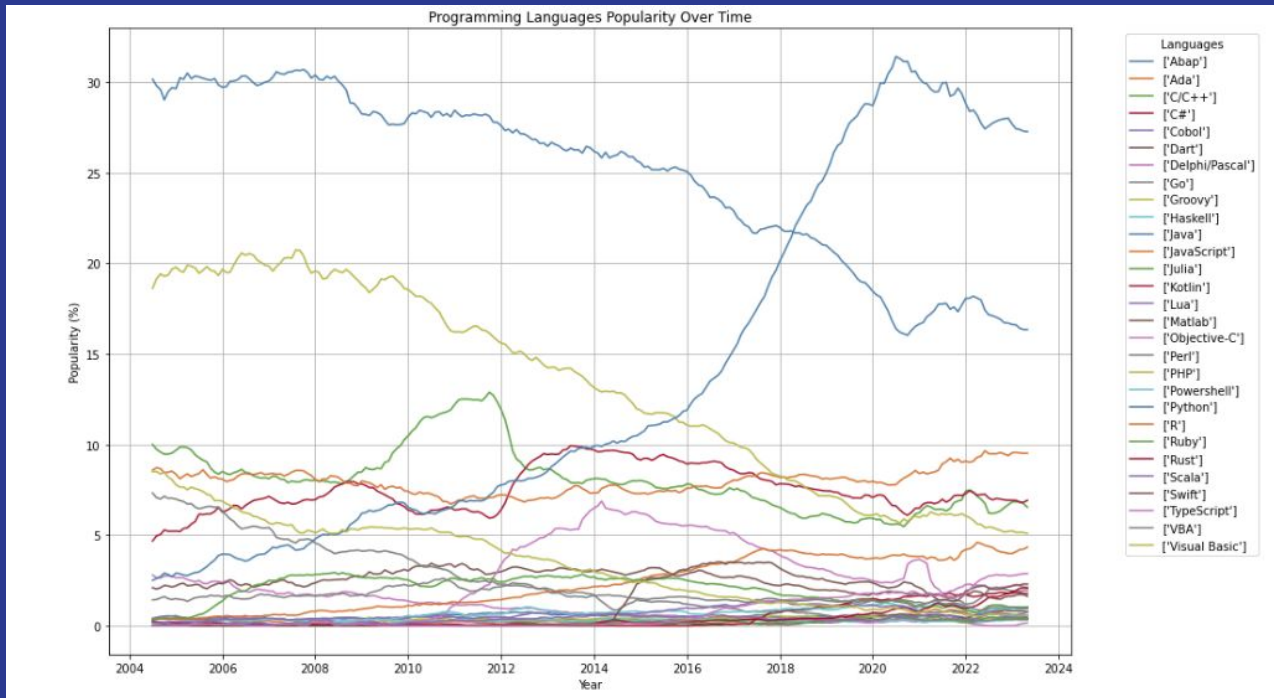
Supplemental dataset provides current data on Github activity, job market demand, and other relevant metrics for over 4000 languages (only 30 were used for analysis then shortened to 12 for the purpose of the study).

Tools Used

Python used for data manipulation, analysis, and visualization.

Libraries: Pandas for data frames, Matplotlib for plotting, Scikit-Learn for machine learning models, and Seaborn for advanced visualizations

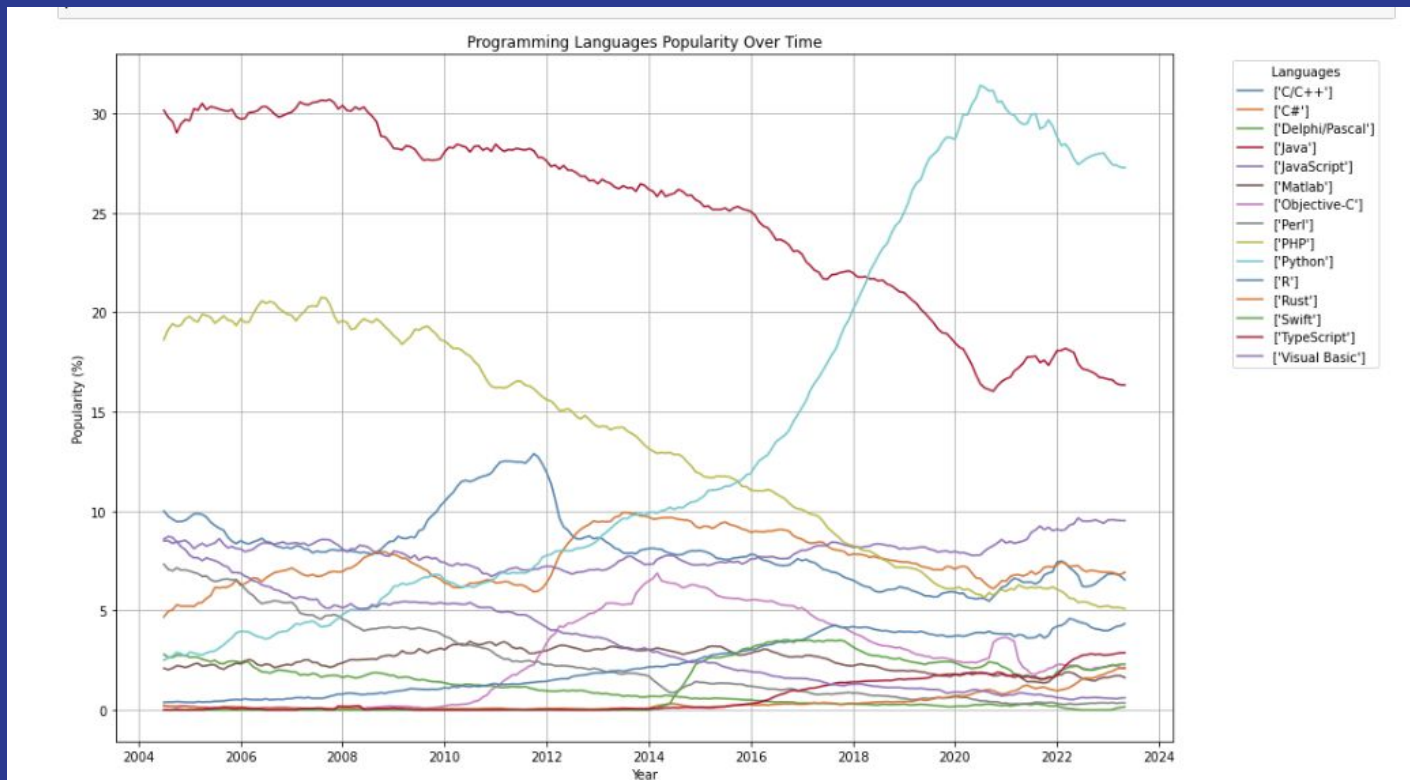
Initial View of Programming Language Trends



Initial View of Programming Language Trends - Analysis

- This messy line graph shows the popularity of 30 programming languages over nearly two decades, showing changes over time.
- The graph shows a broad array of languages, with many languages showing very little popularity (under 2%) they are negligent for the purpose of our project

Updated View of Programming Language Trends



Updated View of Programming Language Trends - Analysis

- Refined graph shows the popularity trends of the top 15 languages, making it clearer to note trends.
- Some languages show a clear upward trajectory like Python, while others decline like Perl or remain the same (C#) over time.
- Shifts in popularity correlate with technological advancements and industry demands, such as the rise of Data Science boosting Python

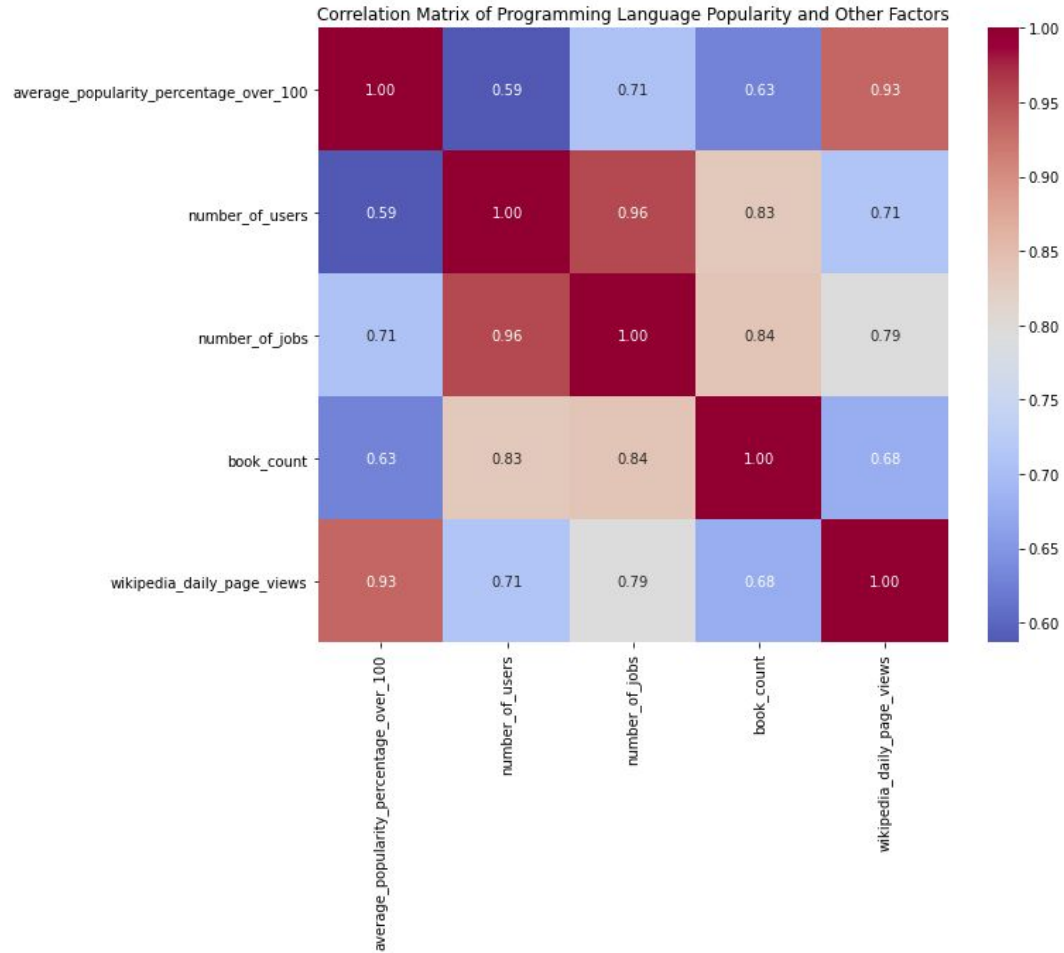
Notable Trends

- **Python:** Remarkable growth, becoming the most popular due to its versatility in web development, data science, and machine learning.
- **Java:** Consistently high popularity, favored for its applicability in enterprise environments
- **JavaScript:** Essential for web development, maintains strong popularity as the backbone of the internet

The Myth of Consistently New Languages

- Most languages that were in the top stayed there throughout the years with few exceptions. Highlighting that there are some favored languages throughout the years
- The top 15 languages account for 88% share of the programming landscape, highlighting the utility of these few languages over thousands of others
- **Decline of PHP and Perl:** Despite their early popularity, these languages have seen a decline, overshadowed by newer, more versatile languages

Correlation Insights



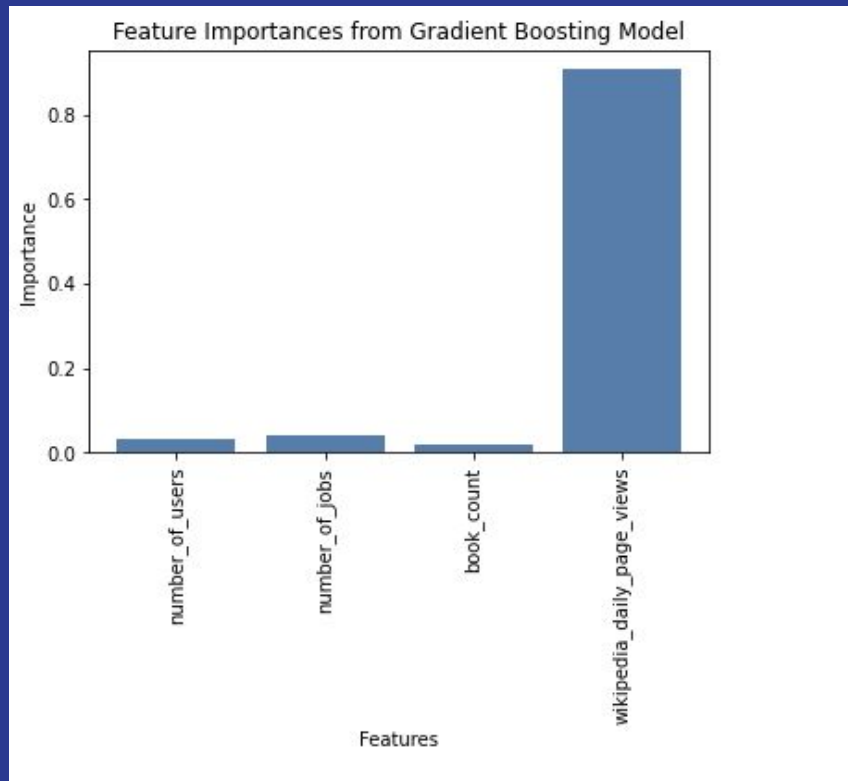
Correlation Insights Analysis

- **High Correlation between Popularity and Wikipedia Page Views (.93)** shows us that languages that have more educational interest or curiosity are more widely used in the programming world
- **Strong Relationship Between Number of Jobs and Popularity (.71)** shows us that languages with higher job availability are more popular, showing us that employment demand drives the adoption of languages

Machine Learning Insights

- Initially tried Linear Regression and Random Forest for prediction, but these were unsuccessful due to unpredictable nature of data. Highlighting the complex nature of this study.
- Gradient Boosting Model proved successful showing an R-squared of .36. Ultimately emphasizing Wikipedia page daily views were the most influential factor, followed by number of jobs.

Gradient Boosting Visualization



Conclusions and Takeaways

- **Staying Power:** Languages like Python, Java, and Javascript have been growing consistently over time.
- **Future Outlook:** Continuous monitoring of technology trends and developer preferences is essential as the landscape evolves.
- **Impact on Stakeholders:** Educators, developers, and aspiring developers should focus on the languages that are not only popular now but have the ability to stay relevant.

References

Scholarly Articles

Orlowska, A., Chrysoulas, C., Jaroucheh, Z., & Liu, X.
(2021). *Programming Languages: A Usage-based
Statistical Analysis and Visualization*. In 2021 The 4th
International Conference on Information Science and
Systems (ICISS 2021), March 17–19, Edinburgh, United
Kingdom. ACM, New York, NY, USA, 6 pages.
<https://doi.org/10.1145/3459955.3460614>

