oxylabs

Report on

JOBS DATA EDA

Most popular programming languages

Prepared by

Laima Lukoseviciute

January 30, 2026

Table of contents

Introduction

This report presents an exploratory data analysis (EDA) of job posting data collected throughout 2025. The primary goal is to identify the most required programming languages in the current labor market and understand how these requirements shift across different professional specializations. By analyzing the intersection of job titles and required programming languages, I will aim to provide some insights into the most popular programming languages defining the industry in the year 2025.

The initial raw data was gained from a large-scale coresignal jobs data, totaling approximately 60 million global job postings. The analyzed data set used in this report consists of 34 features and around 600 000 observations, ranging from January 1st, 2025, to December 19, 2025. For a detailed breakdown of the features, please refer to the Table 1.

Table 1: The description of variables for data.

Variable Name	Type	Description
title	STRING	The professional title of the job listing.
[manager, engineer, analyst, scientist, developer]	BOOL	A flag indicating if the role is categorized as a management, engineering, analyst, scientist, developer role.
company_name	STRING	The name of the hiring organization.
company_industry	STRING	The sector the company operates in (e.g., Tech, Finance).
state	STRING	The US state of the job location.
created_at	TIMESTAMP	The date when the job listing was added to the database.
[Python, SQL, ...]	TIMESTAMP	Individual flags for 22 languages (e.g., Python, SQL, ...) indicating if the role listing ask experience in that role.

Data Preparation and Cleaning

For data preparation and cleaning please refer to "Appendix 1: Preparing the data set." It will give the detailed look into how the data frame that I will analyse here was prepared.

A preview of the analysed dataset is presented below in Table 2, Table 3, Table 4.

```
(581120, 33)
```

Table 2: Raw data pre-view first 5 rows(part 1)

	title	company_name	company_industry
0	Senior Manager, Capacity Engineering, North Am...	Amazon	Software Developr
1	Business Intelligence Engineer	IntelliSavvy	IT Services and IT (
2	HAZARDOUS SUBSTANCES ENGINEER	California Department of Toxic Substances Control	Environmental Ser
3	Senior ASIC Synthesis Engineer	NVIDIA	Computer Hardwa
4	Product Security Engineer	Grammarly	Software Developr

Table 3: Raw data pre-view first 5 rows(part 2)

	title	manager	engineer	analyst	scientist	developer
0	Senior Manager, Capacity Engineering, North Am...	1	1	0	0	0
1	Business Intelligence Engineer	0	1	0	0	0
2	HAZARDOUS SUBSTANCES ENGINEER	0	1	0	0	0
3	Senior ASIC Synthesis Engineer	0	1	0	0	0
4	Product Security Engineer	0	1	0	0	0

Table 4: Raw data pre-view first 5 rows(part 3)

	title	Python	SQL	JavaScript	Java	C#	C++
0	Senior Manager, Capacity Engineering, North Am...	1	1	0	0	0	0
1	Business Intelligence Engineer	1	1	0	0	0	0
2	HAZARDOUS SUBSTANCES ENGINEER	0	0	0	0	0	0
3	Senior ASIC Synthesis Engineer	1	0	0	0	0	0
4	Product Security Engineer	1	0	1	1	0	0

Observations and Features

This section begins the detailed exploration of the dataset's structure. I will examine the characteristics of each column to ensure data integrity and understand the available information.

Specifically, for categorical features (like "title", "company_name", "company_industry", "state"), I'll identify the distinct categories present and count the number of unique observations in each. This step confirms the data types and prepares us for subsequent analysis. Below you can see the breakdown of each type.

```
Programming languages:  ['Python', 'SQL', 'Java', 'JavaScript', 'TypeScript', 'C++', 'C#', 'Objective-C
Categorical features:  ['title', 'company_name', 'company_industry', 'state', 'broad_industry_group']
Date features:  ['created_at']
Job titles:  ['manager', 'engineer', 'analyst', 'scientist', 'developer']
```

In total 22 programing languages are present in the data set. The categotical information about the jobs data is title, company_name, company_industry, state, broad_industry_group. The date when the job posting was created is in the created_at column. And if the particular job posting is searching for a manager, engineer, analyst, scientist, developer.

> NOTE: Keep in mind that job posting can be in multiple job title categories, like scientist and analyst, as well as single job posting can ask experience in multiple programming languages.

Outliers

> NOTE:
>
> Outliers were mainly handaled in Appendix 1. Please refer to the corresponding notebook/report for more details. The `extract_programming_languages` function utilizes the Ollama Large Language Model (LLM) to identify the meaning of the word "Go", "C", "R", and "Assembly". This means the categorization is not completely accurate. This was nessesessary because word "Go" could either be an english verb or the programming lanuage, also some programming languages might have some false positives, due to the typos in the job descriptions like "C omplety" or "R espoucible".

From the tables Table 5 and tbl-description-date we see there is not so much additional improvements we could do, we perhaps could fix some typos in the data but that will not be so crucial for this analyses.

Table 5: Description of the categorical data

	title	company_name	company_industry	state	broad_industry_group
count	581120	575457	519357	435424	581120
unique	269843	56935	384	54	11
top	Software Engineer	Jobs via Dice	Software Development	California	Tech, Data & Telecom
freq	4146	32099	112922	79214	256058

From this we can see that there are 54 states in US, which includes states and US territories.

Table 6: Description of the date when job posting was created data.

	created_at
count	581120
mean	2025-07-16 18:30:15.418261248+00:00
min	2025-01-01 00:14:36+00:00
25%	2025-05-06 01:13:53+00:00

Table 6: Description of the date when job posting was created data.

	created_at
50%	2025-07-29 07:03:16.500000+00:00
75%	2025-10-08 02:06:08.401540352+00:00
max	2025-12-19 02:45:18.817379+00:00

Exploratory Data Analysis

This section presents the Exploratory Data Analysis (EDA) of the 2025 US job market for tech jobs. The goal is to identify patterns and trends within the job postings. The analysis begins with a general overview of the dataset, including the total number of observations and key features. This initial phase identifies the most common programming languages, the primary industries hiring for tech roles, and the organizations with the highest volume of postings. We also examine the timing of these posts to identify potential hiring seasons and pinpoint which geographic states are leading in tech employment.

Distribution of Features

Understanding the foundational characteristics of the data is the first step. This part of the report covers:

- Dataset Volume: A summary of the total jobs analyzed.

- Language Popularity: An absolute and percentage-based ranking of the 24 identified programming languages.

- Industry and Company Presence: Identification of the sectors and specific employers driving the most activity.

- Temporal and Geographic Trends: A look at hiring cycles throughout 2025 and the states with the highest density of tech opportunities.

Distribution of the Programming languages

Let's see what languages were the most popular in 2025. We can see that Python, SQL, Java, JavaScript, and C++ takes the top 5, for the full list see the Table 7.

Table 7: The most mentioned programming languages in 2025.

	count	percentage
Python	270961	46.63
SQL	257769	44.36
Java	118985	20.48
JavaScript	107389	18.48
C++	67077	11.54
Bash	64001	11.01
C#	52780	9.08
TypeScript	42748	7.36
R	34793	5.99
Go	27765	4.78
C	22369	3.85
Matlab	21628	3.72
Scala	14251	2.45
Ruby	12801	2.20
Swift	12735	2.19
Perl	12343	2.12
Kotlin	11522	1.98
Rust	10646	1.83
PHP	8893	1.53
Assembly	6074	1.05
Objective-C	3180	0.55
Dart	1107	0.19

```
(<Figure size 1000x600 with 1 Axes>,
 <Axes: title={'center': 'In-Demand Programming Languages: 2025 Market Share'}, xlabel='Programming Lan
```

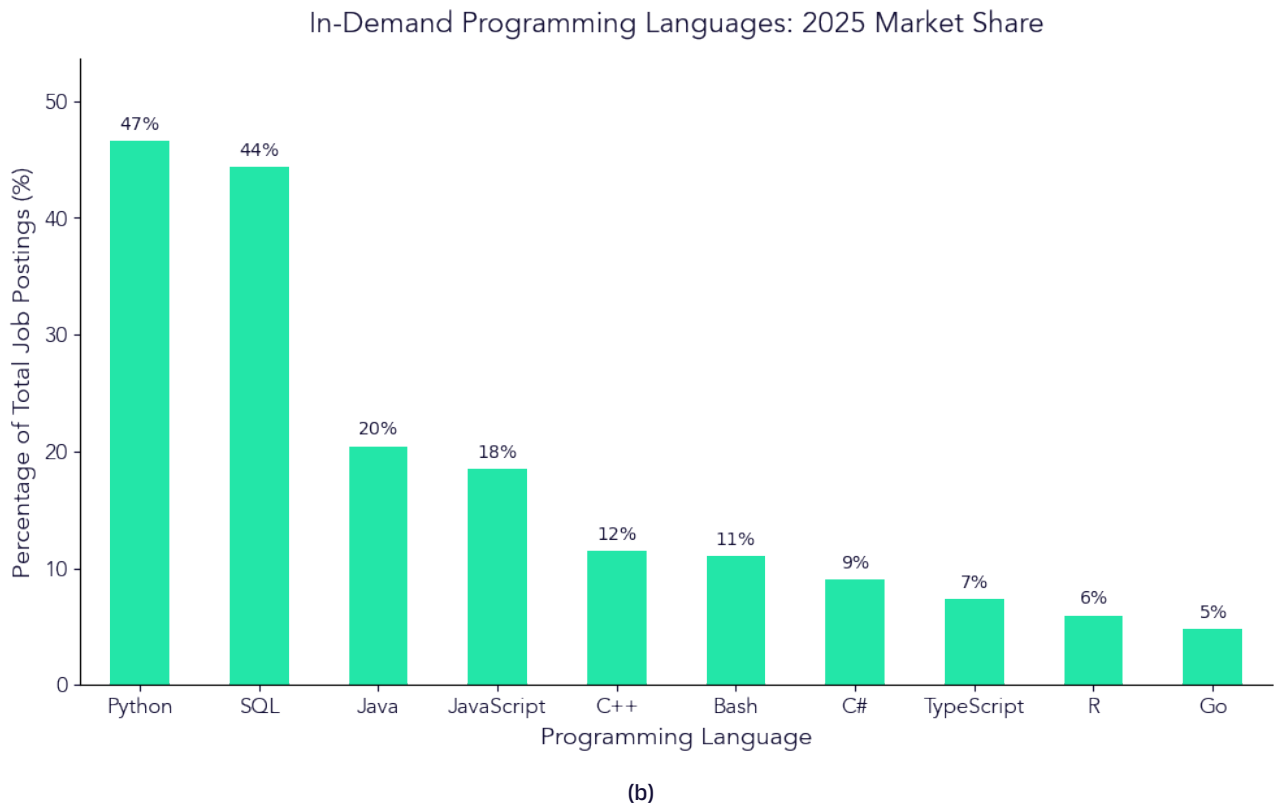(a) The most mentioned programming languages in 2025.

(b)

Figure 1

From the analysis of the 2025 job postings, we can see that Python and SQL are the most prominent requirements, mentioned in 47% and 44% of listings, respectively. This highlights a continued industry-wide focus on automation and data management.

Java and JavaScript follow in the third and fourth positions, appearing in 20% and 18% of roles, indicating their stable role in enterprise and web development. The mid-tier languages show a competitive landscape, with C++ at 12% and Bash at 11%.

The remaining top languages—C# (9%), TypeScript (7%), R (6%), and Go (5%)—represent specialized yet significant segments of the hiring market. These results demonstrate that while the "Big Two" dominate, a diverse set of technical skills remains necessary across the IT sector.

What people might miss here is C, but it takes 4% of share and scores the 11th place.

Distribution of Job Titles by Job Type

In this section, we examine the distribution of job postings across primary professional categories. Based on the data visualized in Figure 2, there is a clear and substantial demand for technical roles, particularly within the engineering domain.

Engineers represent the majority of the market share at 59%, significantly outpacing other roles. Developers and Analysts follow with a combined market share of 32% (at 16% each). Specialized roles such as Scientists and Managers constitute the remaining 5% and 4% of the postings, respectively. These findings suggest that for the 2025 hiring landscape, companies are prioritizing the foundational technical infrastructure provided by engineering talent.

```
engineer     343827
developer     95714
analyst       92908
scientist     28481
manager       22698
dtype: int64
```
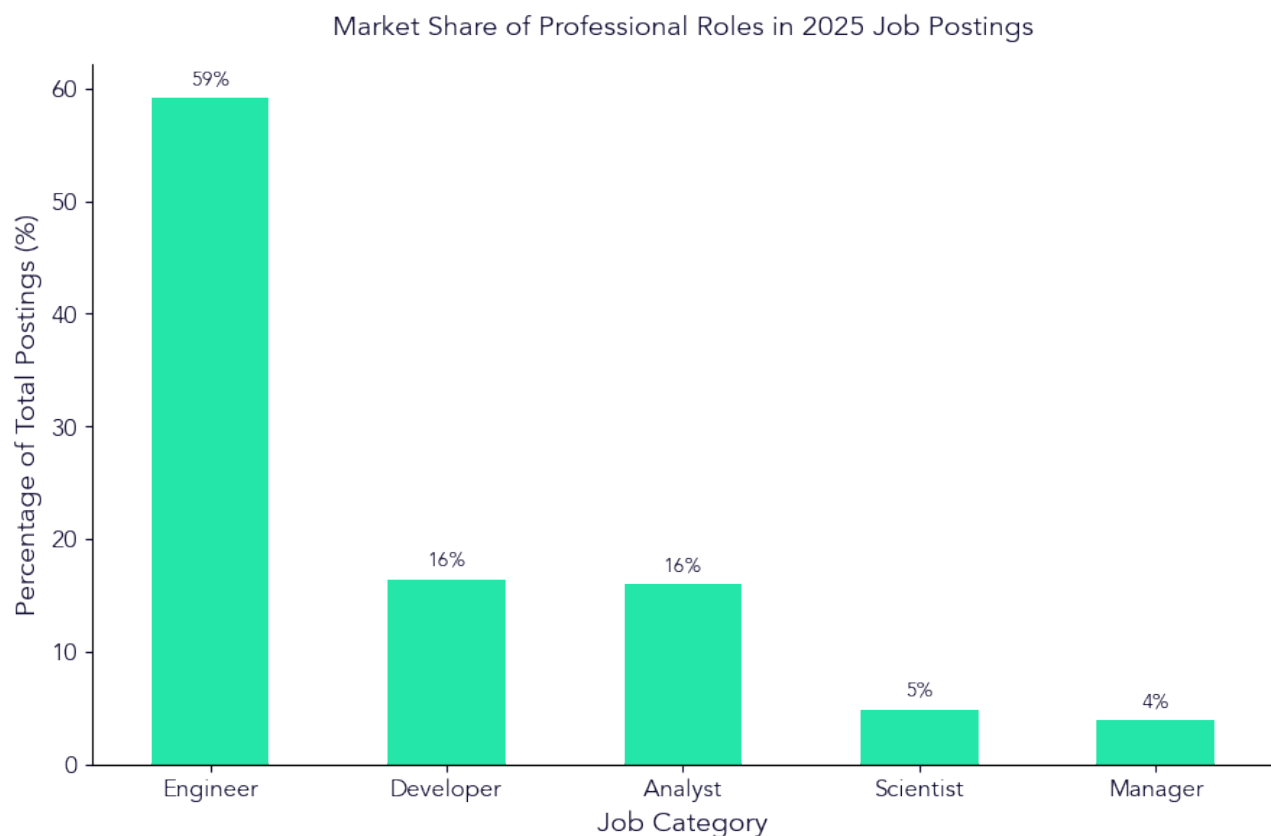
Market Share of Professional Roles in 2025 Job Postings

Figure 2: Distribution of job titles in 2025 in job postings.

Distribution of Company Names

In this section, we examine the leading organizations driving the demand for technical talent in the United States. While the tech landscape is vast and diverse, a small group of industry giants and specialized platforms accounts for a significant portion of total hiring activity.

From Figure 3, we can observe that Job via Dice maintains a commanding presence, representing 6% of the entire tech job market in the USA. This high volume highlights the platform's role as a primary aggregator for technical specialized roles.

When analyzing the "Big Tech" sector specifically, Amazon, Microsoft, Google, and Apple all appear within the top 20 hiring entities. However, only Amazon and Google secure spots within the top 10, each accounting for approximately 1% of the total market share. This data suggests that while these tech behemoths are influential, the market remains highly fragmented, with a significant amount of hiring distributed across thousands of mid-sized firms and diverse industry sectors.

Distribution of Industries

In this section, we analyze the distribution of technology-related roles across various economic sectors. Understanding where demand originates provides critical context for the current hiring landscape and identifies the primary drivers of technical growth.

Based on the data visualized in Figure 4, the Tech, Data & Telecom industry remains the dominant force, accounting for 44% of all professional requirements. This sector continues to outpace all others, reinforcing its role as the primary engine for tech employment.

Outside of the core technology sector, the demand is significantly more distributed. Professional, Legal & Business Services follow with a 14% market share, while Manufacturing, Industrial & Defense represents 10% of the postings. The Finance, Insurance & Real Estate (FinTech) sector accounts for 8%, highlighting a stable need for technical expertise in modernizing financial infrastructure. Notably, all other industries each represent less than 3% of the market share, indicating a high concentration of tech talent within the top four sectors.

Below @dist-company-industries you can see the raw distribution of the top 20 industries that are not cleaned, just for the sake of understanding the raw data.
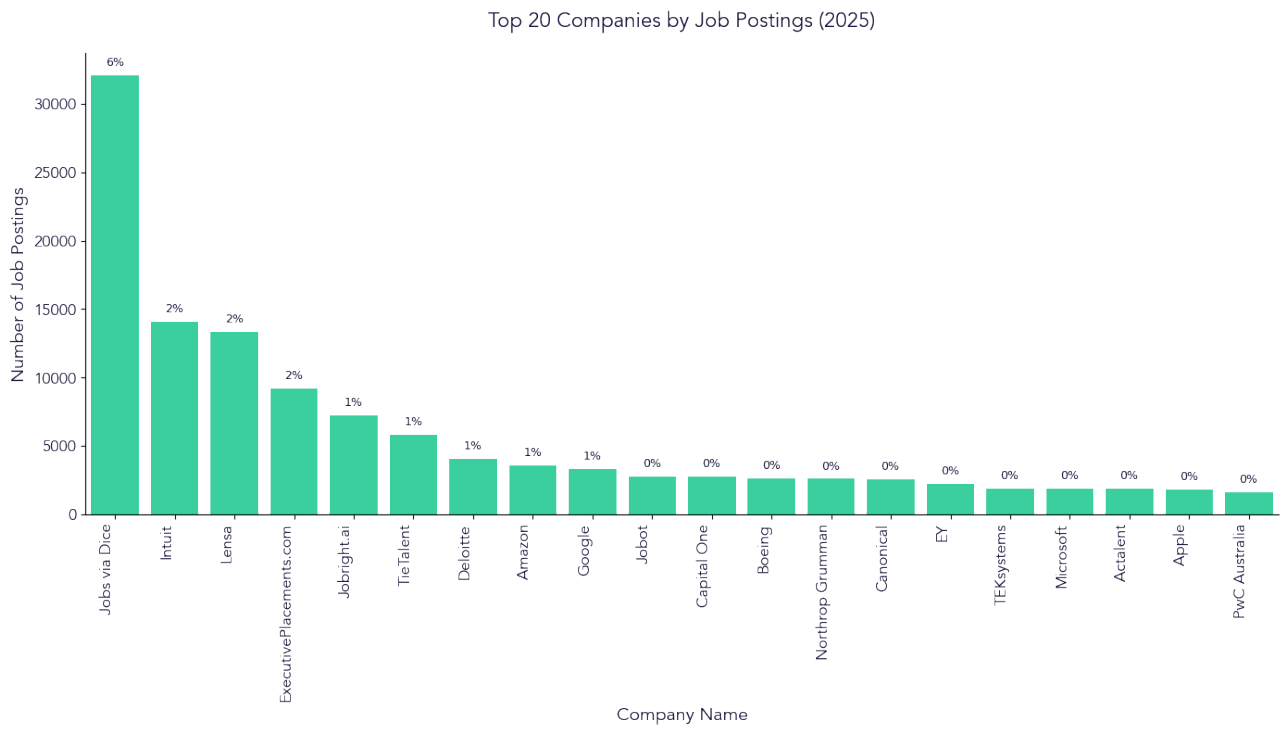
Figure 3: Barplot of the most mentioned companies in 2025 in job postings.

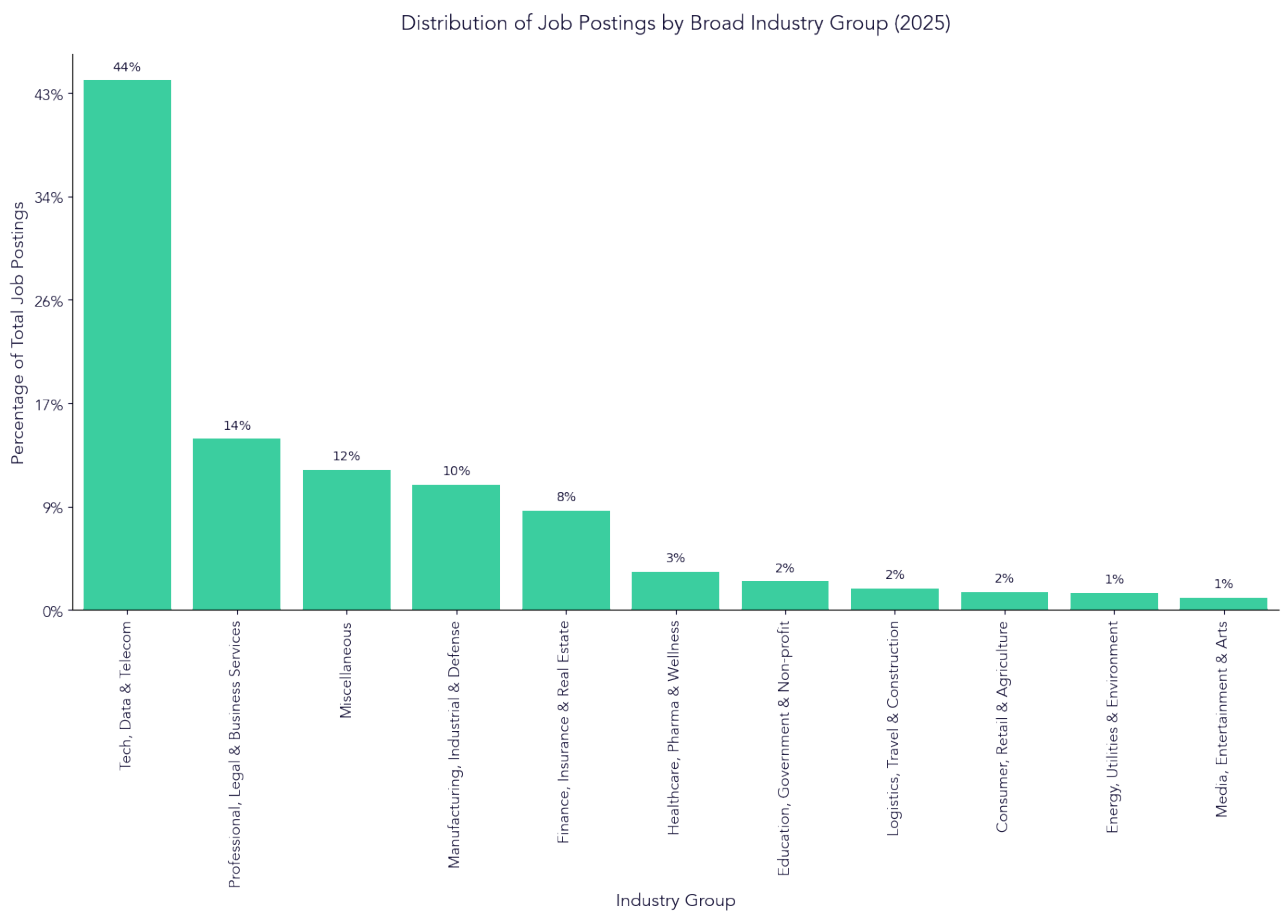Distribution of Job Postings by Broad Industry Group (2025)

Figure 4: Barplot of the most mentioned broad industries in 2025 in job postings.
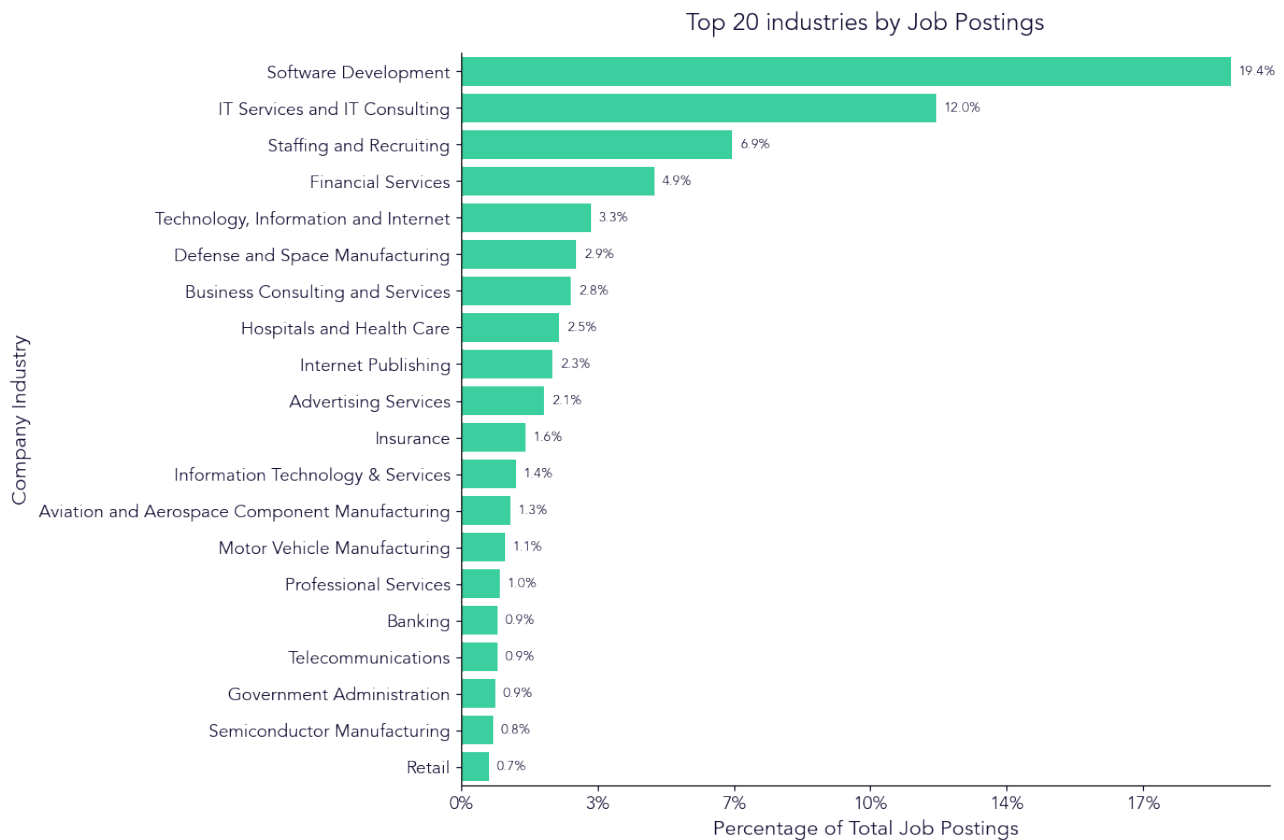
Figure 5: Barplot of the most mentioned raw company industries in 2025 in job postings.

Distribution of States

In this section, we analyze the geographic distribution of tech talent demand across the United States. Identifying these "hiring hubs" allows us to understand the regional concentrations of the digital economy and where companies are focusing their recruitment efforts.

```
Unable to display output for mime type(s): text/html
```

```
Unable to display output for mime type(s): text/html
```

```
Unable to display output for mime type(s): text/html
```

Based on the data visualized in Figure 6 and Figure 7, California remains the primary driver of technical employment, accounting for 14% of all job postings. Texas follows closely as the second-largest market, reinforcing its status as a significant and growing center for technology and innovation. New York has the third-largest market taking up 5% of all postings. Together, these three states represent a substantial portion of the national demand.

For a more granular view of regional trends, an interactive heatmap is available in the project files at `outputs/figures/us_hiring`. This interactive asset allows for state-by-state comparisons and provides specific market share details for each territory.

Distribution of Job Posting Times

In this section, we analyze the seasonal variations in hiring activity throughout 2025. Understanding these temporal peaks is essential for the PR team to align media pitches with the periods of highest market engagement.

Based on the data visualized in Figure 8, the hiring landscape exhibits a distinct "double peak" pattern. The highest volume of activity occurs in August and November, with each month accounting for 13% of the total annual postings. These surges likely correspond to the conclusion of the summer period and the finalization of year-end technical projects, respectively. Identifying these windows allows our department to release targeted content when the recruitment market is at its most active.

Language Demand by Category

After establishing the baseline, the analysis investigates how language requirements vary across different contexts. This section explores the relationship between programming languages and:
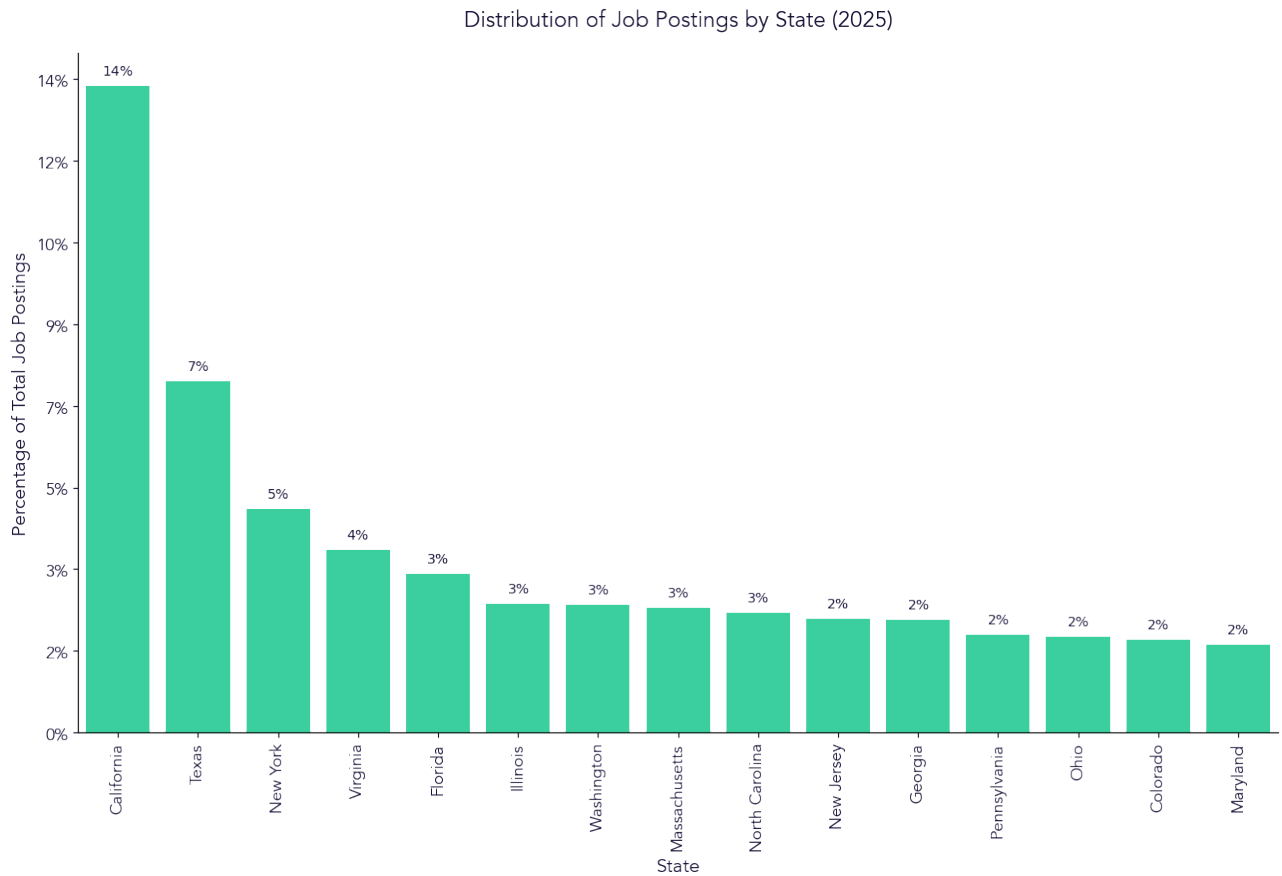
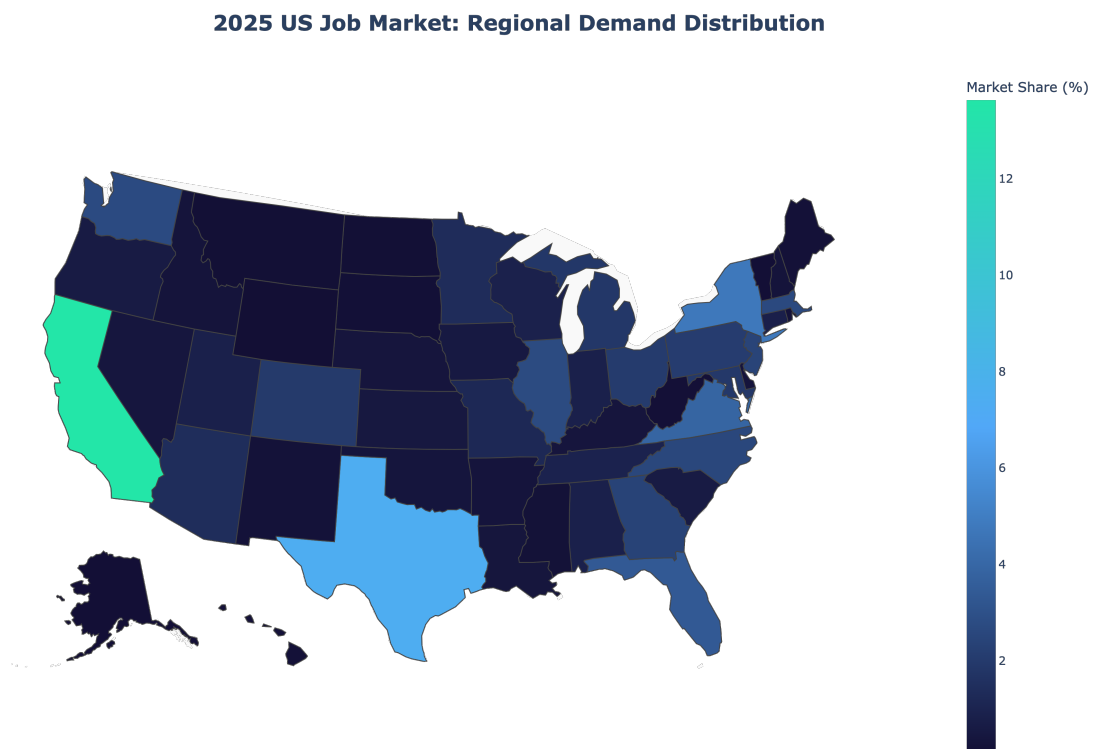Figure 6: Distribution of job postings by state in 2025.

Figure 7: 2025 US Job Market. Regional Demand Distribution

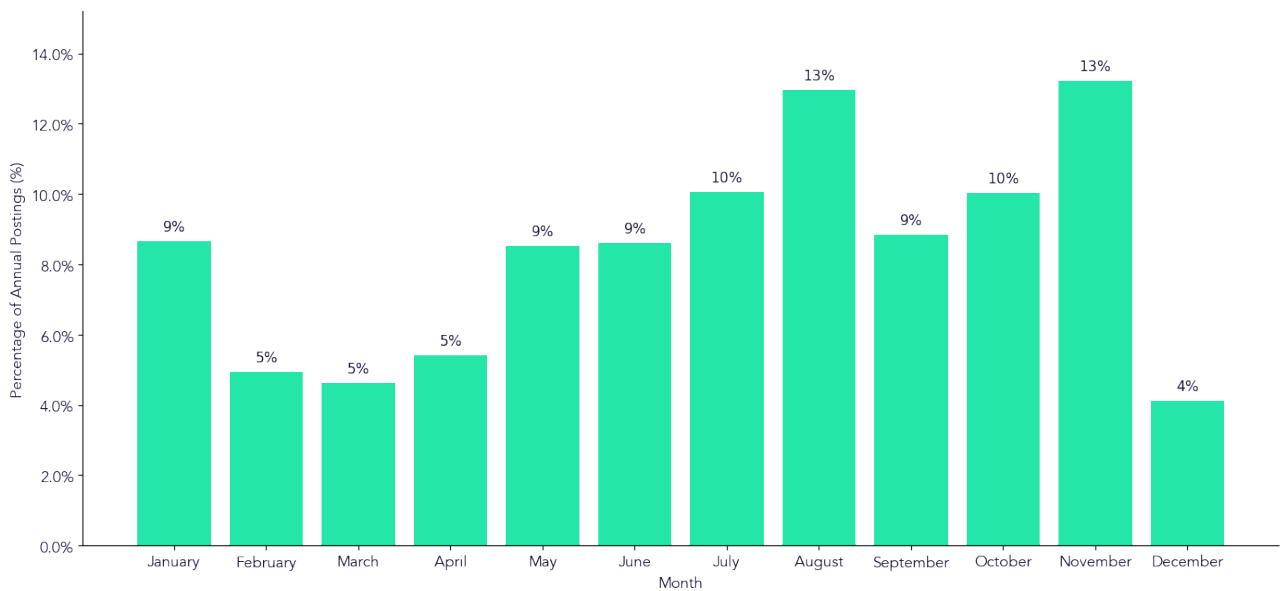**2025 Monthly Job Posting Distribution**

Figure 8: Distribution of job postings by month in 2025.

- Job Titles: Which languages are synonymous with roles like "Data Scientist" or "Software Engineer."
- Languages Duos: Which languages are going together in the jobs descriptions.
- Industries: How the tech stack in "Finance" differs from "Healthcare" or "Manufacturing."
- Geography: Regional preferences for specific technologies across the United States.

Language Demand by Job Title

In this section, we analyze the specific technical requirements for leadership roles. Understanding the linguistic preferences for management-level positions provides insight into the tools required to oversee modern technical teams and data-driven projects.

- For the Manager category, the data reveals a high reliance on data processing and automation tools. The top two languages are Python, mentioned in 51% of job postings, and SQL, mentioned in 49%. This nearly equal split suggests that managers are increasingly expected to possess both scripting capabilities and direct database proficiency. Secondary requirements include Java (23%) and JavaScript (17%), which reflect the oversight of enterprise applications and web-based platforms. For a more detailed breakdown of these trends, please see Figure 9.

- Technical Engineers show a strong preference for backend and automation stability. Python leads at 51%, serving as the primary language for automation and system integration. SQL follows at 31%, while Java (23%), JavaScript (18%), and C++ (16%) round out the core stack required for building scalable infrastructure. For a more detailed breakdown of these trends, please see Figure 10.

- For Analysts, SQL is the undisputed leader, appearing in 73% of job postings. This underscores the fundamental requirement for direct data extraction and manipulation. Python (35%) and R (12%) serve as the secondary toolkit for advanced statistical modeling and visualization, while other languages represent a marginal share of the market. For a more detailed breakdown of these trends, please see Figure 11.

- Data Scientists exhibit the most concentrated language requirements. Python dominates this field at 86%, confirming its status as the industry standard for machine learning and research. SQL remains vital at 50% for data retrieval, while R retains a significant foothold at 23% for specialized statistical analysis. For a more detailed breakdown of these trends, please see Figure 12.

- The Developer profile differs significantly, emphasizing application logic and database interaction. SQL takes the lead at 55%, indicating that database management is a core developer responsibility. JavaScript (42%) and Java (33%) highlight the focus on front-end and enterprise development. Notably, Python appears in only 29% of these roles, while C# maintains a strong presence at 20%. For a more detailed breakdown of these trends, please see Figure 13.
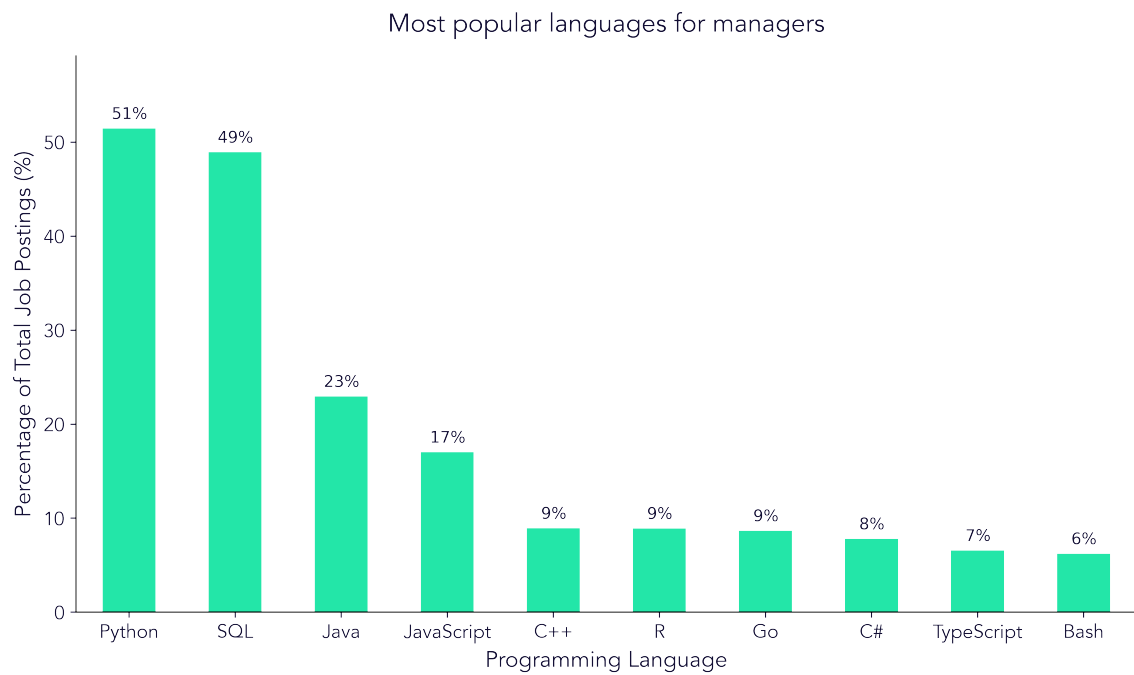
10

Most popular languages for managers

Figure 9: Programming languages demand for managers distribution

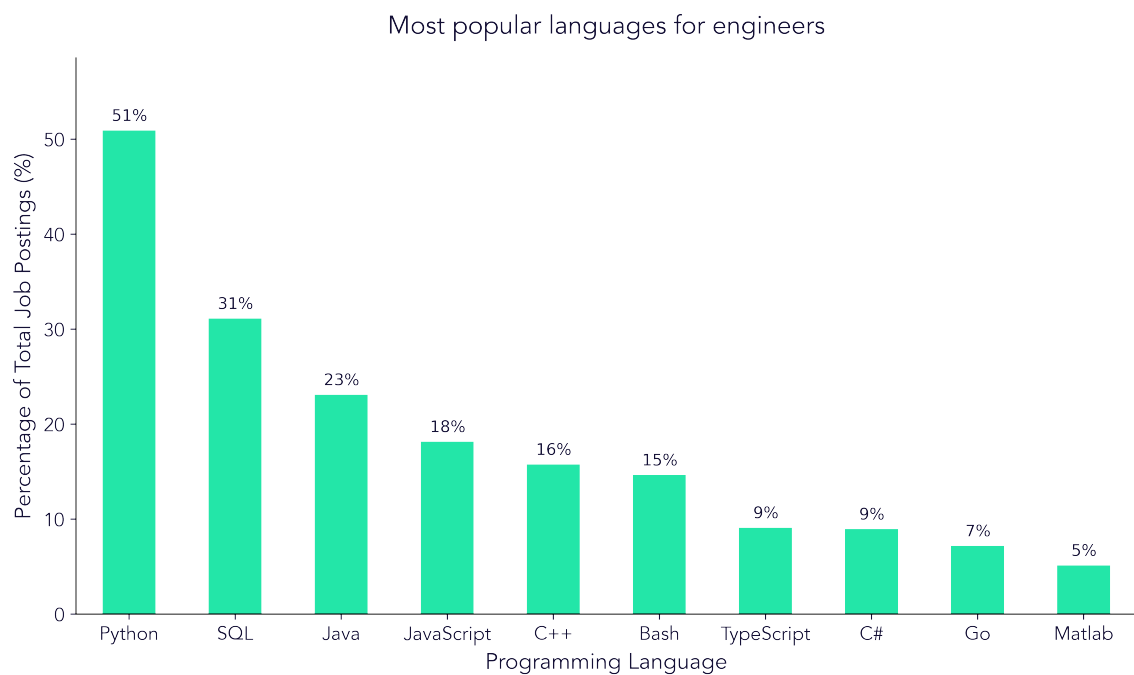## Most popular languages for engineers

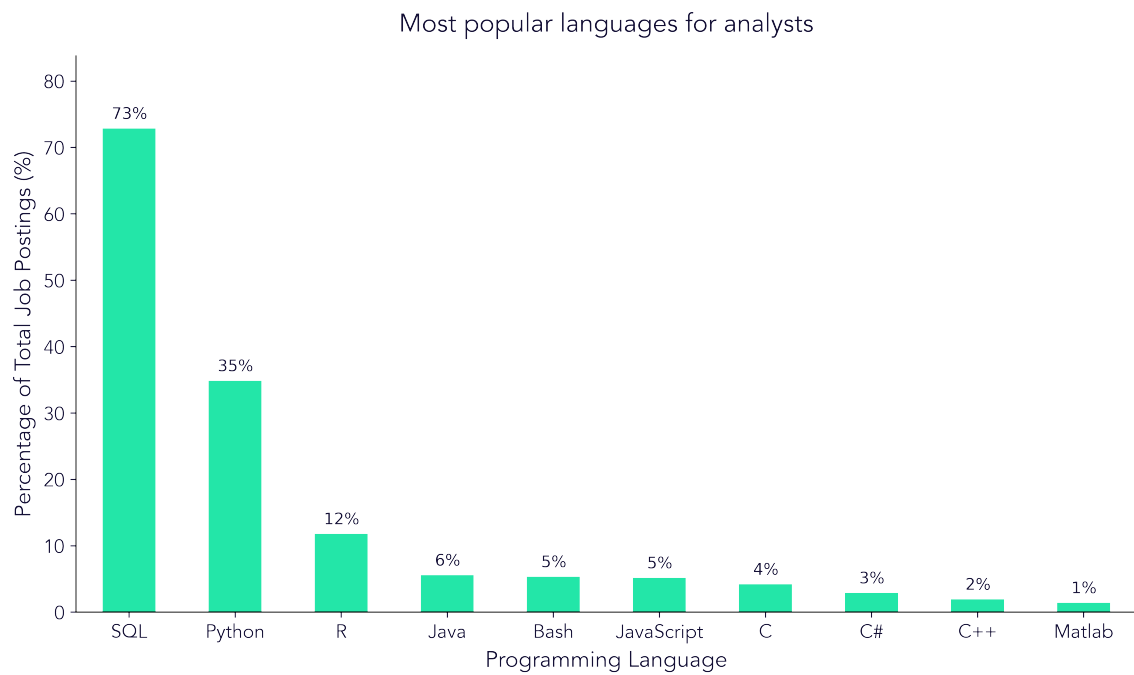Figure 10: Programming languages demand for engineers distribution

Figure 11: Programming languages demand for analysts distribution

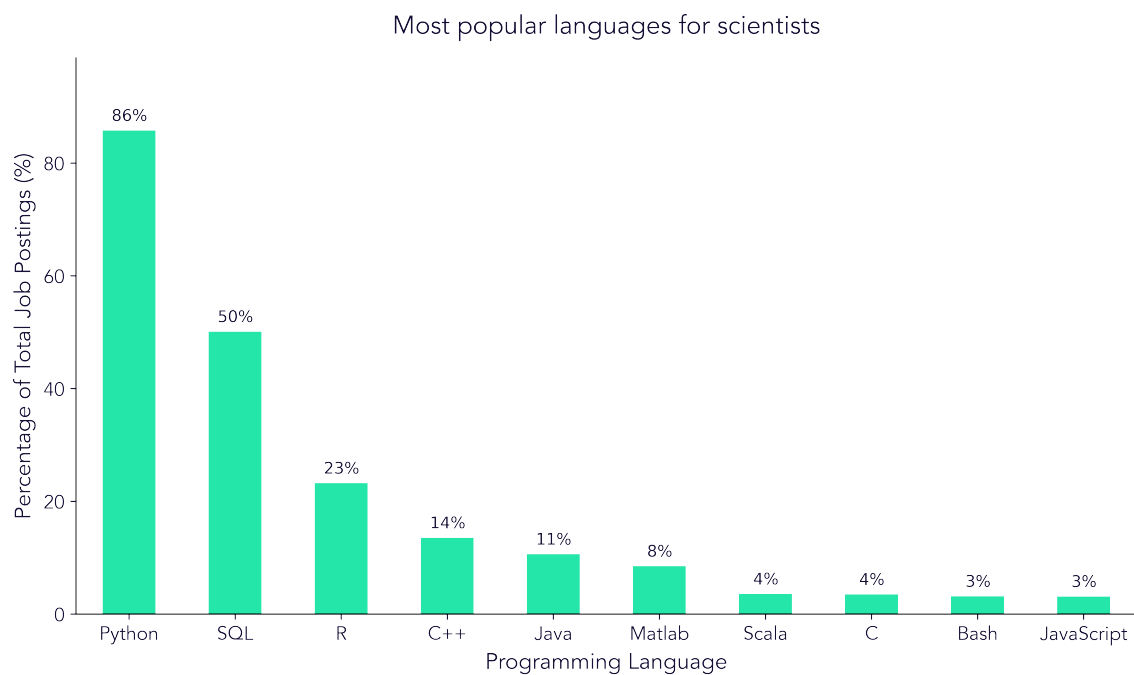**Figure 12: Programming languages demand for scientists distribution**
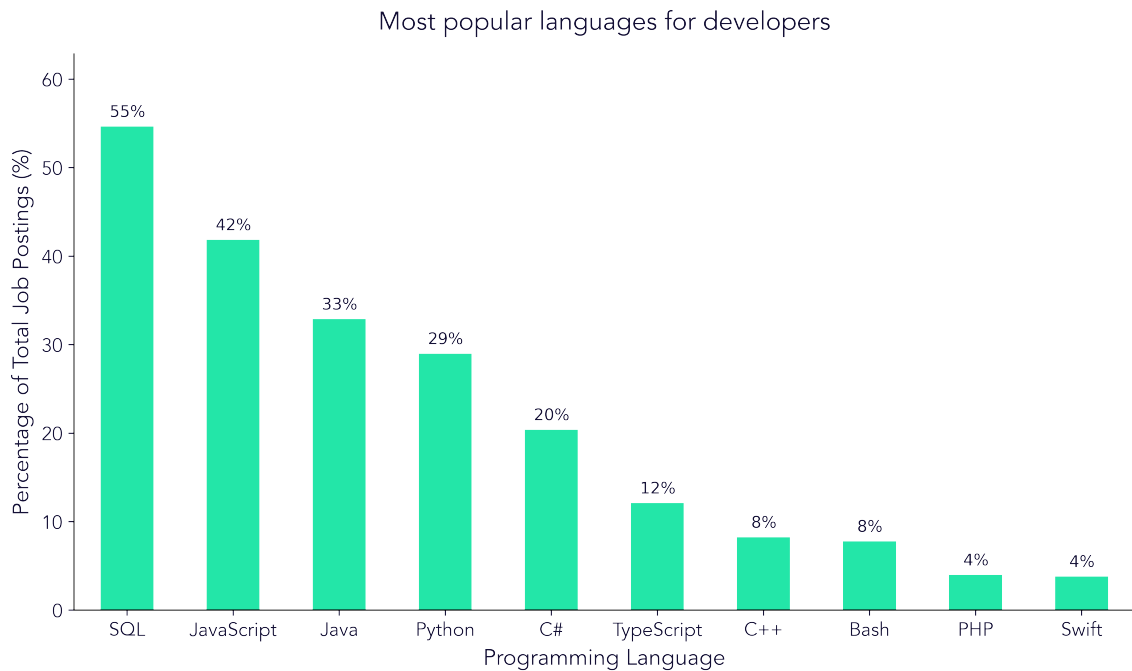
Most popular languages for developers

Figure 13: Programming languages demand for developers distribution

Language Co-occurrence Frequency

In this section, we explore the relationships between different programming languages by analyzing how frequently they are mentioned together within the same job postings. Understanding these pairings provides a clearer picture of the "tech stacks" currently favored by employers in 2025.

Based on the data visualized in Figure 14, the most significant synergy exists between SQL and Python, which appear together in 21% of all job postings. This highlights the standard industry expectation for professionals to combine data retrieval skills with automation and analytical scripting.

Other notable combinations include Java and Python, which co-occur in 11% of listings, often indicating roles involved in large-scale enterprise integration or backend infrastructure. Beyond these primary pairings, other language combinations remain relatively rare, suggesting that most roles are either highly specialized in a specific ecosystem or centered around the dominant Python-SQL core.

The Most in Demand Language by State

In this section, we explore which programming language holds the dominant position within individual states. By identifying the most frequently requested language in each region, we can map out the technical "territories" of the 2025 US job market.

According to the data visualized in Figure 15, the national landscape is characterized by a dual dominance of Python and SQL. While SQL is the most popular language in the majority of states—particularly across the Midwest and Southern regions—Python maintains a strong stronghold in specific high-tech corridors. Python is the leading language across the West Coast (California, Oregon, and Washington) and the Northeast/East Coast (New York, Vermont, New Hampshire, Massachusetts, Virginia, and Maryland). Interestingly, Python also takes the top spot in non-contiguous states and specific hubs such as Alabama, Hawaii, and Alaska. This geographic split suggests that while SQL remains the universal standard for data-heavy business operations, Python is the preferred tool for the innovation-heavy clusters on the coasts.

```
Unable to display output for mime type(s): text/html
```

Market Leader Deep Dive: FAANG and MANGO

The final part of the analysis focuses on the industry giants that often set the trend for the rest of the market. We compare two distinct groups:

FAANG (Meta, Apple, Amazon, Netflix, Google) These established leaders provide a benchmark for high-scale, mature tech environments. We analyze their core language requirements to see which traditional technologies remain dominant.
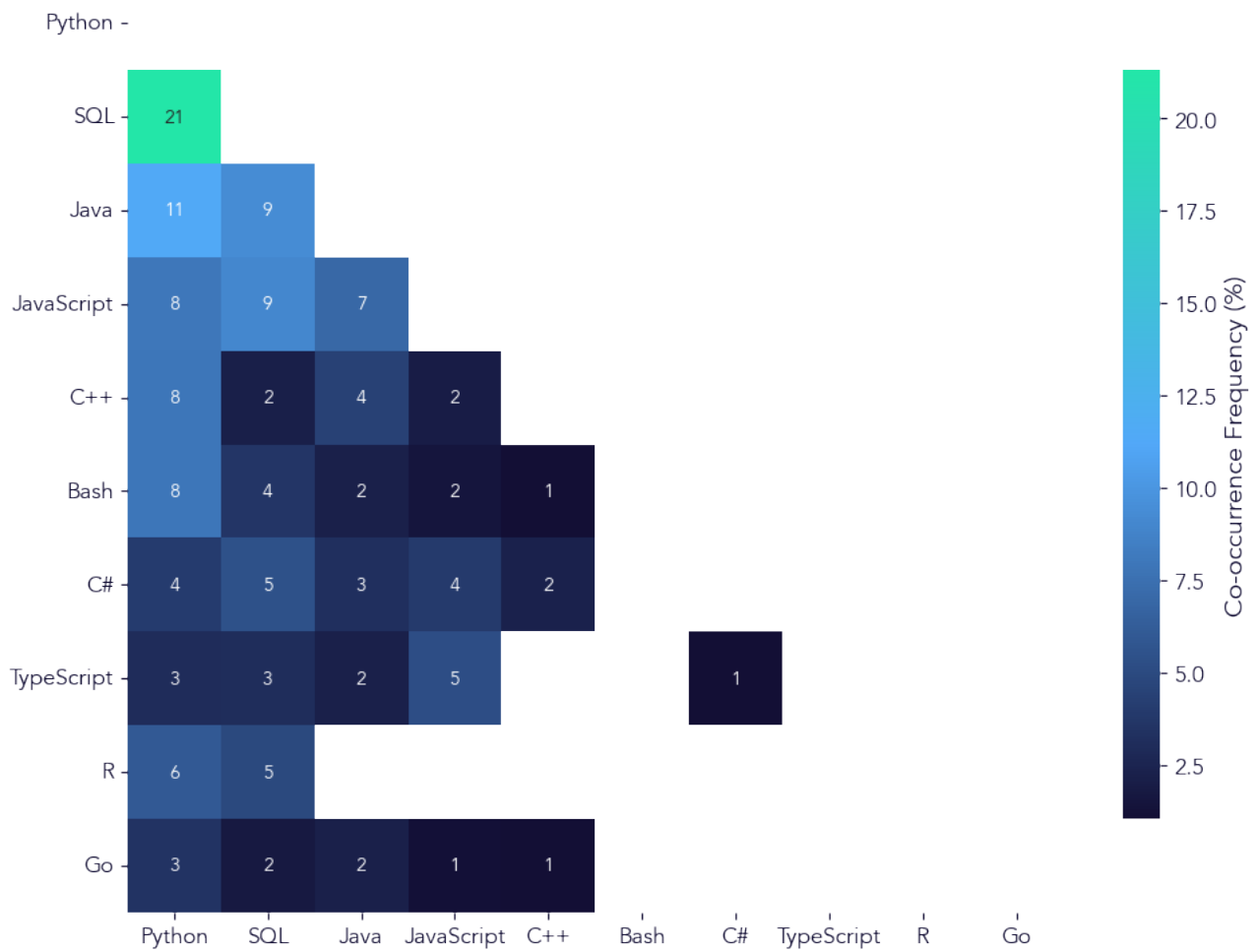
Figure 14: Language co-occurrence heatmap for 2025.

2025 Most Mentioned Language by State
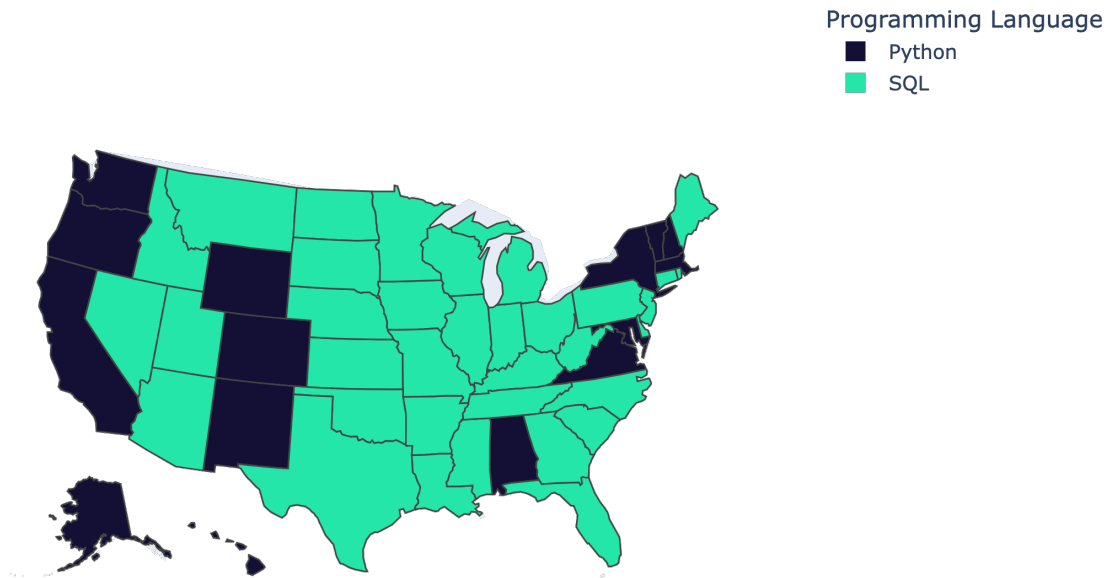
Programming Language
- Python
- SQL

Figure 15: Most popular programming language by state.

MANGO (Microsoft, Apple, Nvidia, Google, OpenAI/Anthropic) The MANGO index represents the architects of the current AI era. This section highlights the languages preferred by companies at the forefront of hardware innovation and Large Language Model development. Comparing these groups reveals whether AI-centric firms are pivoting toward newer languages faster than the broader market.

FAANG

In this section, we examine the hiring patterns of the world's most influential technology giants. These organizations often set the standard for technical requirements and recruitment volume across the global IT sector.

Based on the data in Figure 16, Amazon emerges as the clear leader in recruitment volume among the FAANG group, accounting for 39% of all job postings within this elite cohort. Google follows with 24%, while Microsoft represents 15% of the demand. Apple and Meta round out the group with 13% and 10% of the market share, respectively. This distribution highlights Amazon's aggressive expansion and its significant role as a primary employer for technical talent.

The technical requirements for these companies, as shown in Figure 17, reflect a heavy emphasis on versatility and systems performance. Python is the undisputed standard, required in 67% of all postings. There is also a notable demand for systems-level and enterprise languages, with C++ mentioned in 42% of roles and Java in 37%. Data management remains a core competency with SQL at 32%, while JavaScript represents 17% of the requirements, primarily for front-end and full-stack engineering positions.

MANGO

In this section, we pivot our focus to the "MANGO" cohort—Meta, Apple, NVIDIA, Google, and Oracle—representing a more hardware and AI-centric evolution of the traditional tech giants. Analyzing this group allows us to see how the demand for talent shifts when emphasizing cutting-edge infrastructure and silicon innovation.

According to Figure 18, Google stands as the dominant hiring force in this group, responsible for 34% of the job postings. Microsoft and Apple maintain a strong presence with 19% and 17% respectively, while NVIDIA—fueled by the ongoing AI hardware boom—accounts for 15% of the requirements. Meta occupies a smaller portion of the current hiring landscape at 3%. This distribution underscores a market where established cloud and hardware leaders are currently driving the bulk of new technical opportunities.

The technical stack preferred by these leaders, as seen in Figure 19, highlights a significant lean toward high-performance computing. Python remains the primary requirement, appearing in 69% of postings. However, the
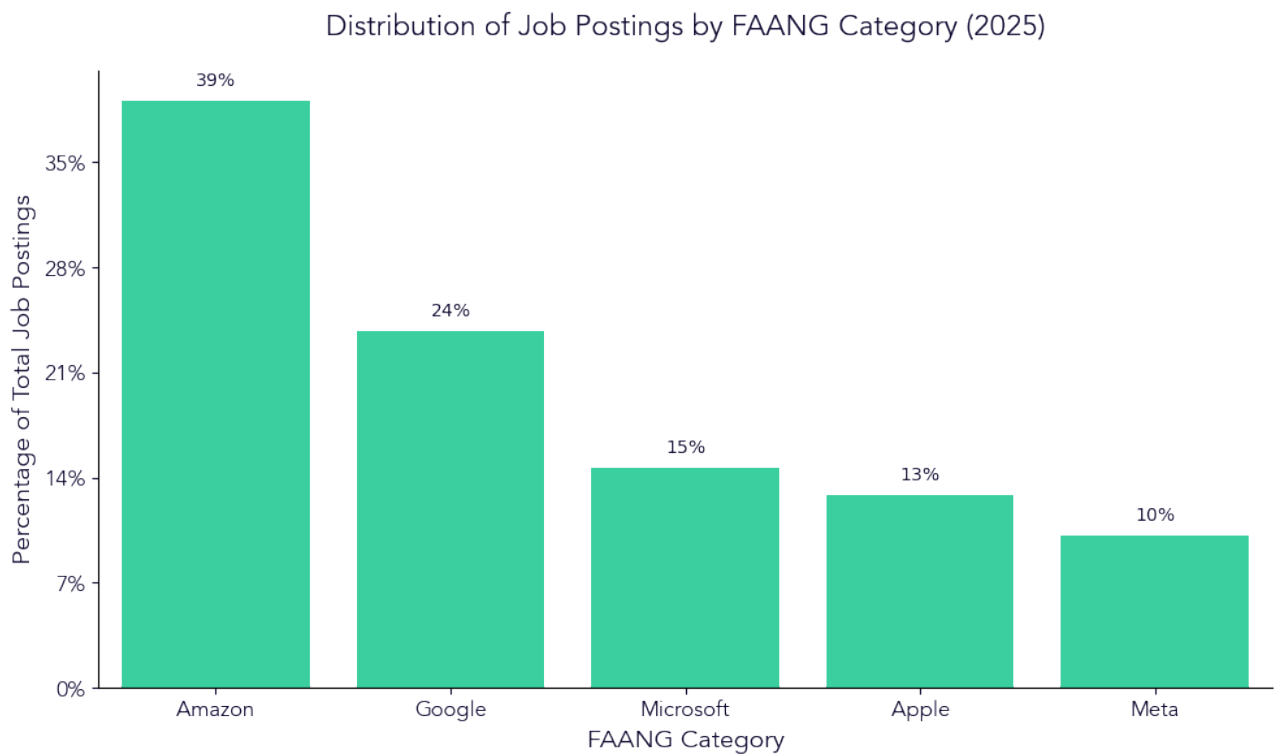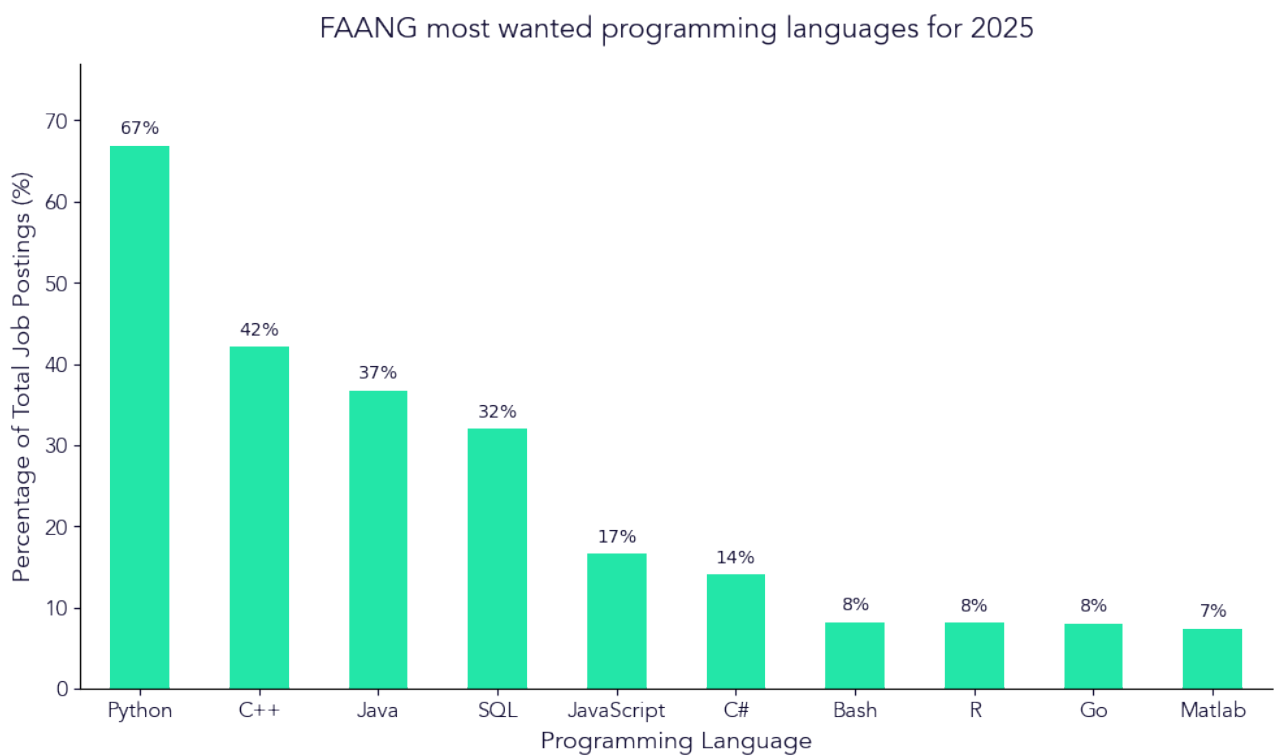
Figure 16: Distribution of job postings within the FAANG companies.

```
(<Figure size 1000x600 with 1 Axes>,
 <Axes: title={'center': 'FAANG most wanted programming languages for 2025'}, xlabel='Programming Langu
```

(a) Distribution of demanded languages within the FAANG companies.

(b)

Figure 17

influence of systems engineering is evident in the high demand for C++ (46%), which is considerably higher than the general market average. Java (28%) and SQL (26%) follow as essential enterprise tools, while JavaScript (19%) maintains its necessity for interface and application development.
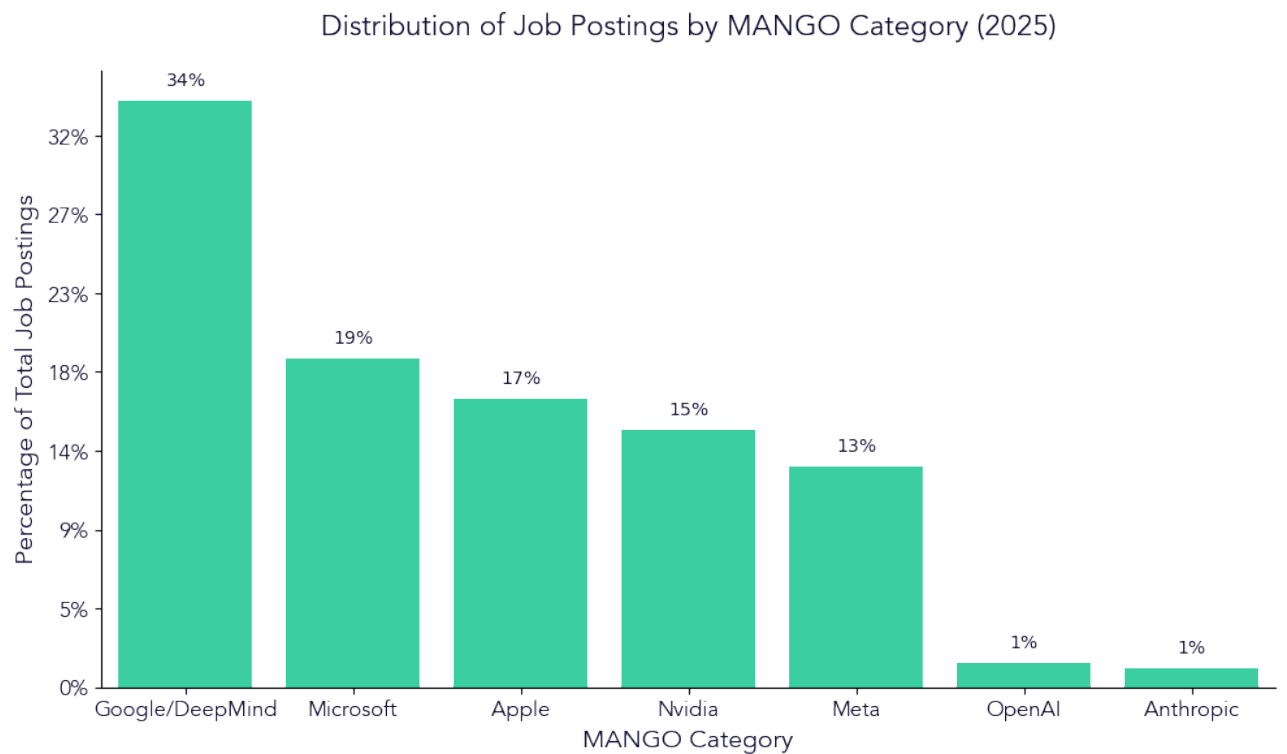

Figure 18: Distribution of job postings within the MANGO companies.

Key Insights

The Foundational Duo: Python (47%) and SQL (44%) are no longer specialized skills but are instead the baseline requirements for modern technical literacy. Their dominance across industries indicates that automation and data management are core priorities for organizations of all sizes.

Role-Specific Specialization: There is a clear divide in technology stacks between roles. Data Scientists are increasingly expected to master Python and R, whereas Developers and Engineers show a higher requirement for JavaScript, TypeScript, and Java. Analysts remain primarily focused on SQL for data retrieval.

The AI Tech Stack Pivot: The "MANGO" group (Microsoft, Apple, Nvidia, Google, OpenAI/Anthropic) shows a distinct preference for languages that support AI development and hardware optimization, such as Python, C++, and Go, compared to the broader market.

Enterprise Stability: Despite the rise of newer languages, Java and C++ maintain strong double-digit market shares (20.5% and 11.5% respectively), highlighting their continued importance in maintaining legacy systems and performance-critical applications.

Geographic Concentration: Approximately 25% of the analyzed US technical demand is concentrated in just two states—California and Texas—suggesting that despite the rise of remote work, regional tech ecosystems continue to drive the bulk of hiring volume.

Summary

The 2025 technical job market is characterized by a massive reliance on data processing, automation, and infrastructure maintenance. Through the analysis of over 600,000 job postings, it is evident that Python and SQL have established themselves as the "lingua franca" of the industry, appearing in nearly half of all technical job listings.

While traditional enterprise languages like Java and JavaScript remain vital for web and corporate development, there is a visible shift in the requirements of market-leading organizations. Companies at the forefront of the AI era are prioritizing languages like C++ and Python to build the next generation of hardware and software.

```
(<Figure size 1000x600 with 1 Axes>,
 <Axes: title={'center': 'MANGO most wanted programming languages for 2025'}, xlabel='Programming Langu
```

(a) Distribution of demanded languages within the MANGO companies.
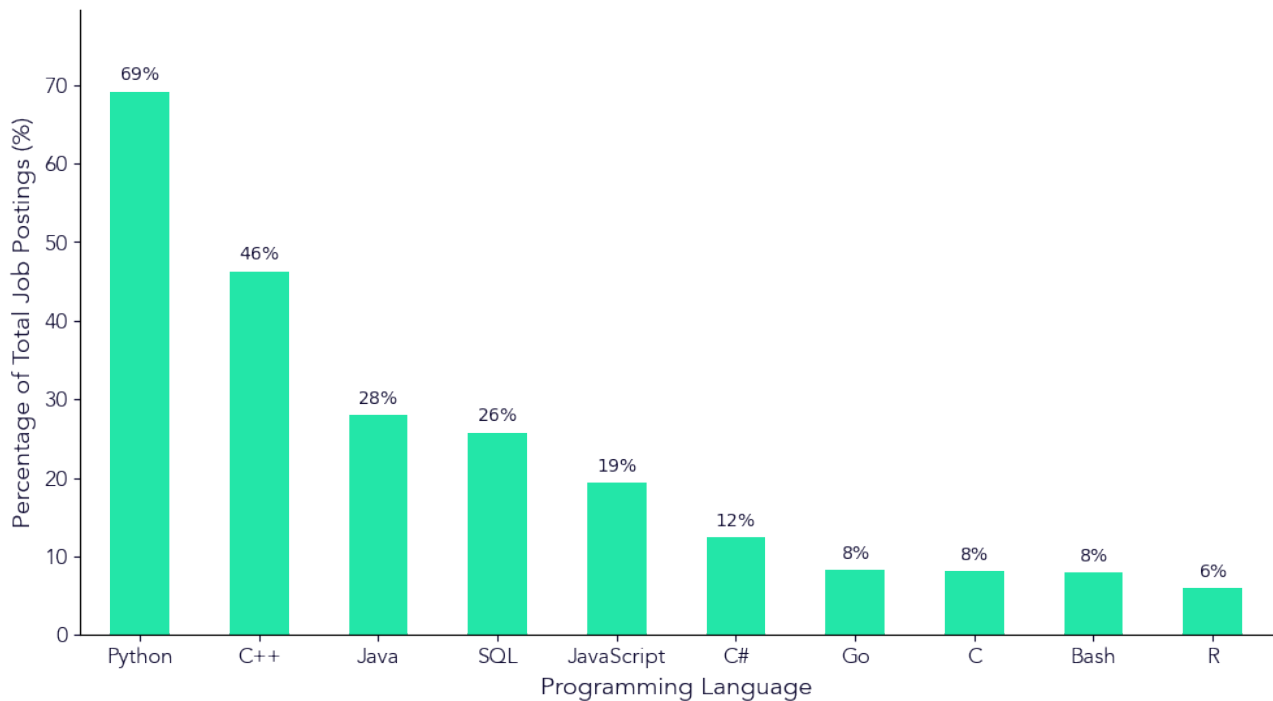
(b)

Figure 19

Hiring activity remains concentrated in established technology hubs, with California and Texas leading the demand in the United States. Across different specializations, the "Engineer" category remains the most dominant, representing 59% of the market share. This confirms that companies are prioritizing the foundational technical infrastructure and construction of systems over purely analytical or management roles for the 2025 calendar year.

Suggestions for Further Improvements

Based on the patterns observed and the performance metrics analyzed, this section outlines potential next steps for refining the data collection process, enhancing the proxy network infrastructure, or expanding the scope of the analysis. These suggestions are intended to maximize efficiency and further optimize the data scraping solutions.