IBM Data Analyst Capstone Project

By Mourlaye Traore

October 17, 2025

PROJECT OUTLINES

Introduction

Introduces the context of the analysis, data sources, and purpose of the study

Methodology

Covers the full analytical process—data collection, cleaning, visualization, and dashboard creation.

Results

Key outputs presented through visualizations, charts, and interactive dashboards.

Discussion

Interpretation of findings and their implications for data-driven decision-making.

Conclusion

Summary of overall insights, lessons learned, and final reflections.

Appendix

Supporting materials, additional visuals, and code samples.

EXECUTIVE SUMMARY

- This project applied end-to-end data analysis techniques using Python, IBM Cognos Analytics, and Google Looker Studio to analyze Stack Overflow survey data and related job market information. Through web scraping, API integration, and data wrangling, data was collected, cleaned, and standardized for analysis and visualization.

- The interactive dashboards revealed that HTML/CSS/JavaScript/TypeScript (291) are the most used languages, while C# (97) is among the least. PostgreSQL (994) and SQL Server (689) lead database use, and AWS (2,362) and Azure (1,649) dominate cloud platforms. Among frameworks, React, Spring Boot, and ASP.NET Core emerged as top preferences.

- Looking ahead, developers show growing interest in **Python**, **TypeScript**, and modern frameworks such as **FastAPI** and **React**, reflecting a move toward scalable and cloud-ready solutions. Demographic insights highlight a young, globally distributed, and highly educated developer community, mostly aged 25–34 with bachelor's or master's degrees.

- Overall, this project demonstrates how business intelligence tools and visual analytics can turn complex datasets into clear, data-driven insights that reflect real trends in technology adoption and workforce skills.

INTRODUCTION

- The purpose of this project was to demonstrate the complete data analysis workflow, from data collection to dashboard creation, using real-world survey data.

- The analysis focused on uncovering insights about developer demographics, technology usage, and future trends.

- The target audience includes **data analysts, business decision-makers, and technical stakeholders** seeking to understand technology adoption patterns.

- The project's value lies in showing how **data-driven insights** can support **strategic decisions** in workforce development, technology investment, and product planning.

By integrating Python, IBM Cognos Analytics, and Google Looker Studio, the project bridges technical analysis with **visual storytelling** for practical business applications.

METHODOLOGY

- The primary data source for this project was the **Stack Overflow Developer Survey dataset**, provided through the IBM Data Analyst Capstone Project.

- A brief data collection practice using web scraping and **API requests** was completed to simulate real-world data retrieval, though the main analysis focused on the Stack Overflow data.

- **Data wrangling** included removing duplicates, imputing missing values, and normalizing key numeric columns to prepare the dataset for analysis.

- **Exploratory Data Analysis** (EDA) was performed to examine data distribution, identify outliers, and study relationships between demographics, experience, and compensation.

- Finally, the cleaned dataset was visualized through interactive dashboards built in **IBM Cognos Analytics** and **Google Looker Studio**, presenting clear insights on technology usage, trends, and demographics.
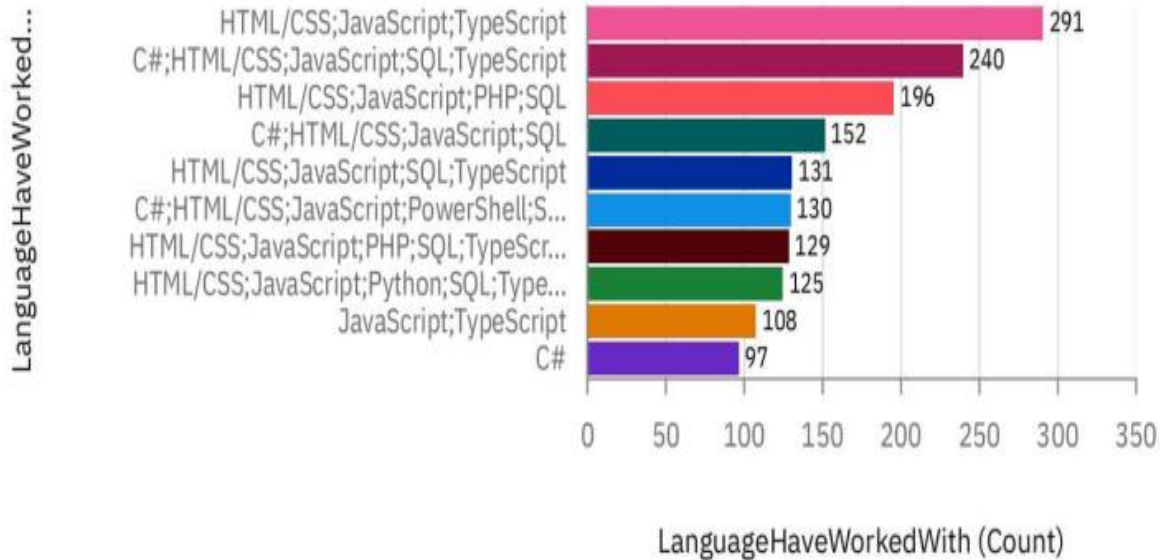
PROGRAMMING LANGUAGE TRENDS

Current Year

Next Year
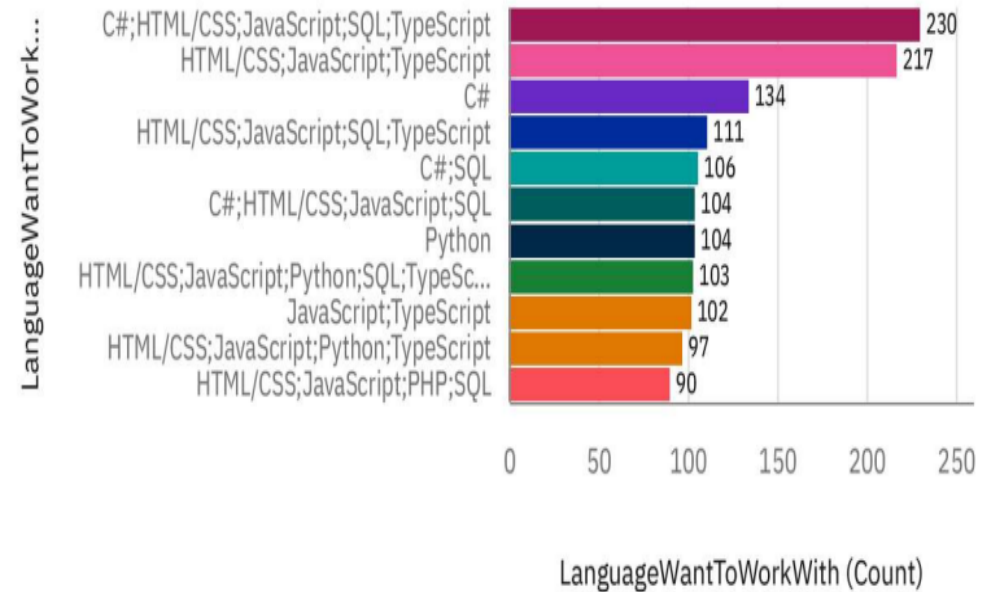
PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

- The most commonly used language combination among respondents is **HTML/CSS, JavaScript, and TypeScript**, with **291** mentions, showing their dominance in front-end web development.

- **C#** and **HTML/CSS, JavaScript, SQL, TypeScript** follow closely (240 and 196 respectively), highlighting a strong presence of both Microsoft and full-stack ecosystems

- In contrast, when asked about future preferences, the top desired combinations remain similar — HTML/CSS, JavaScript, SQL, TypeScript (**230**) and **HTML/CSS, JavaScript, TypeScript (217)** — but standalone **C#** rises in popularity (134), indicating continued interest in established backend languages.

This comparison reveals a clear movement from traditional frameworks (C#, PHP, SQL) to emerging and performance-oriented technologies that support scalability, data processing, and automation.
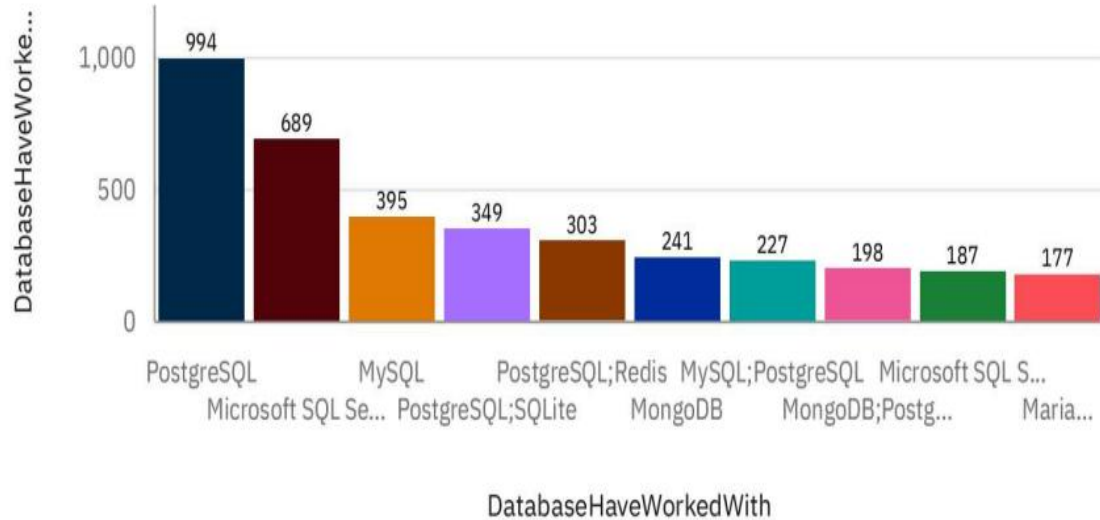
- The overlap between current and desired technologies suggests developers plan to **deepen their expertise** in existing languages rather than switching entirely to new ones.

- The consistent demand for **JavaScript-based stacks** (HTML/CSS, TypeScript, SQL) implies these skills remain central to modern full-stack and web application development.

- Organizations can leverage this insight by **investing in training and projects** that strengthen these ecosystems while maintaining flexibility for developers proficient in **C# and SQL-based** environments.
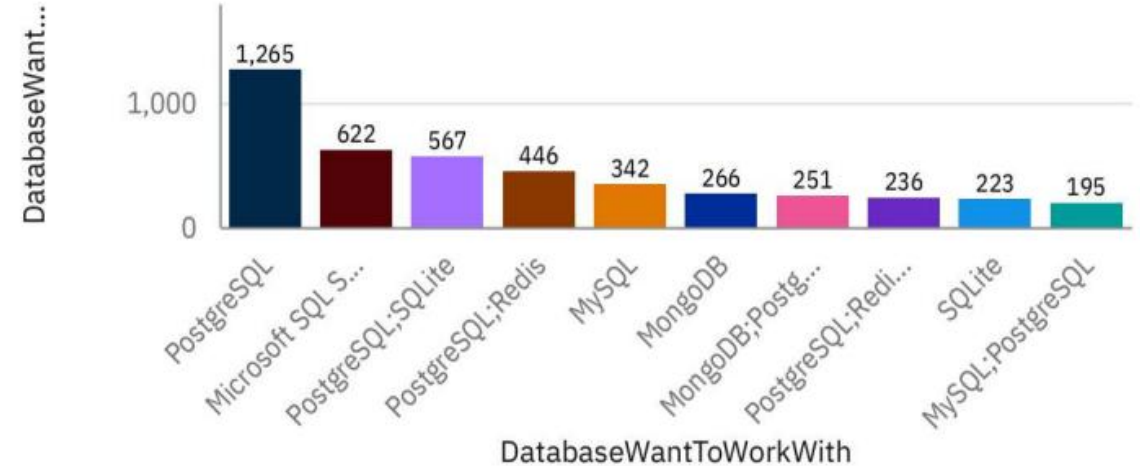
DATABASE TRENDS

DATABASE TRENDS - FINDINGS & IMPLICATIONS

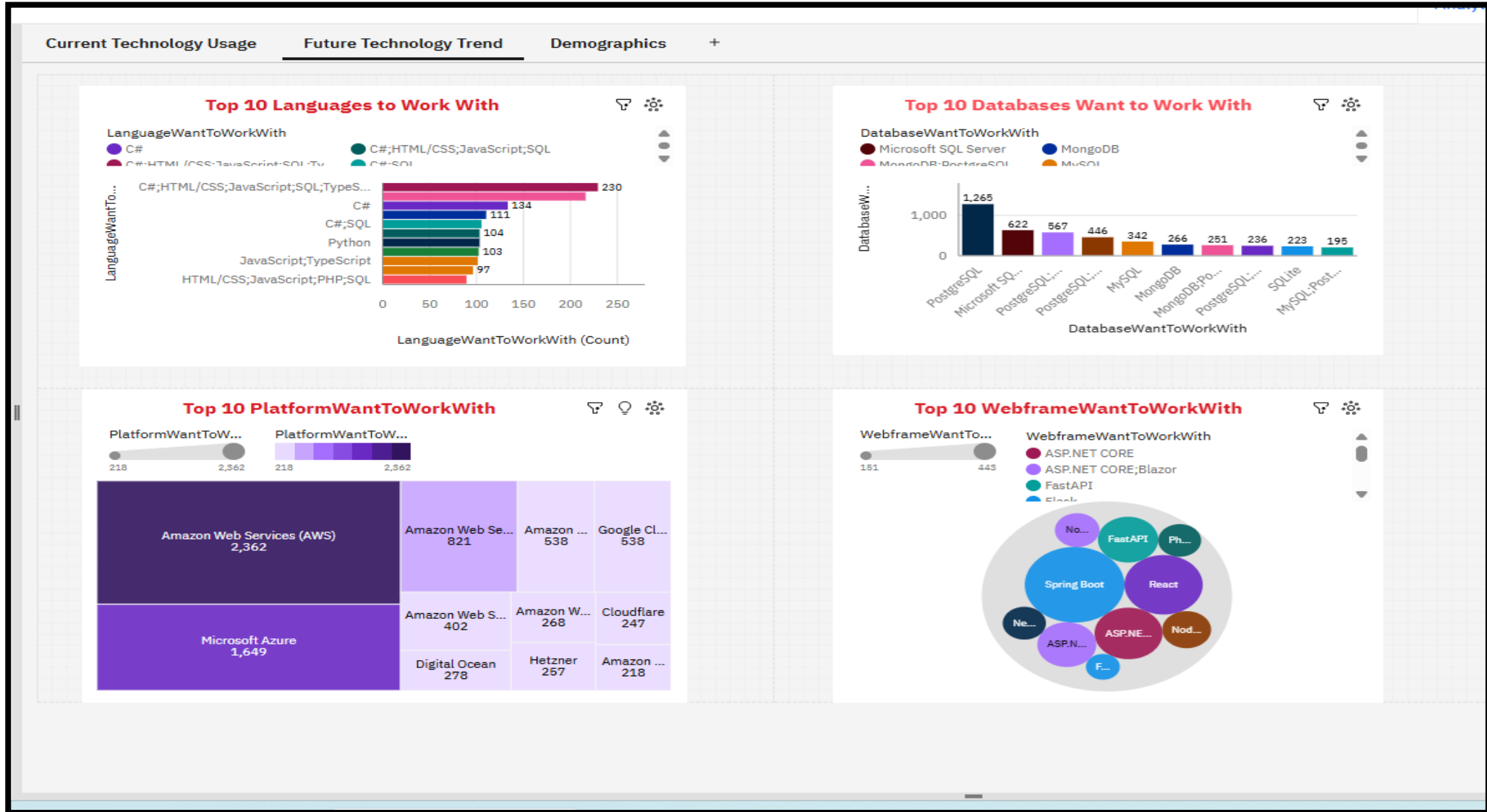- In current usage, PostgreSQL (994) and Microsoft SQL Server (689) lead as the most widely used databases, followed by MySQL (395) and PostgreSQL-SQLite combinations (349).

-
For future interest, PostgreSQL shows a significant jump to 1,265 responses, maintaining its top position and widening the gap with Microsoft SQL Server (622) and SQLite (567).

- Databases like **SQLite (567)** and **Redis (446)** also appear among the top desired technologies, reflecting developers' increasing interest in lightweight and high-speed data solutions.

- The consistent dominance and rising preference for **PostgreSQL** suggests a long-term shift toward open-source, scalable, and cloud-friendly database systems

- The growth in interest for **Redis** and **SQLite** implies developers are seeking **simpler and faster** alternatives for modern applications, especially in microservice and mobile environments.

- Companies should align infrastructure strategies with these trends, investing in **PostgreSQL expertise** and optimizing systems to support hybrid data solutions that combine relational and in-memory technologies.

DASHBOARD TAB 1
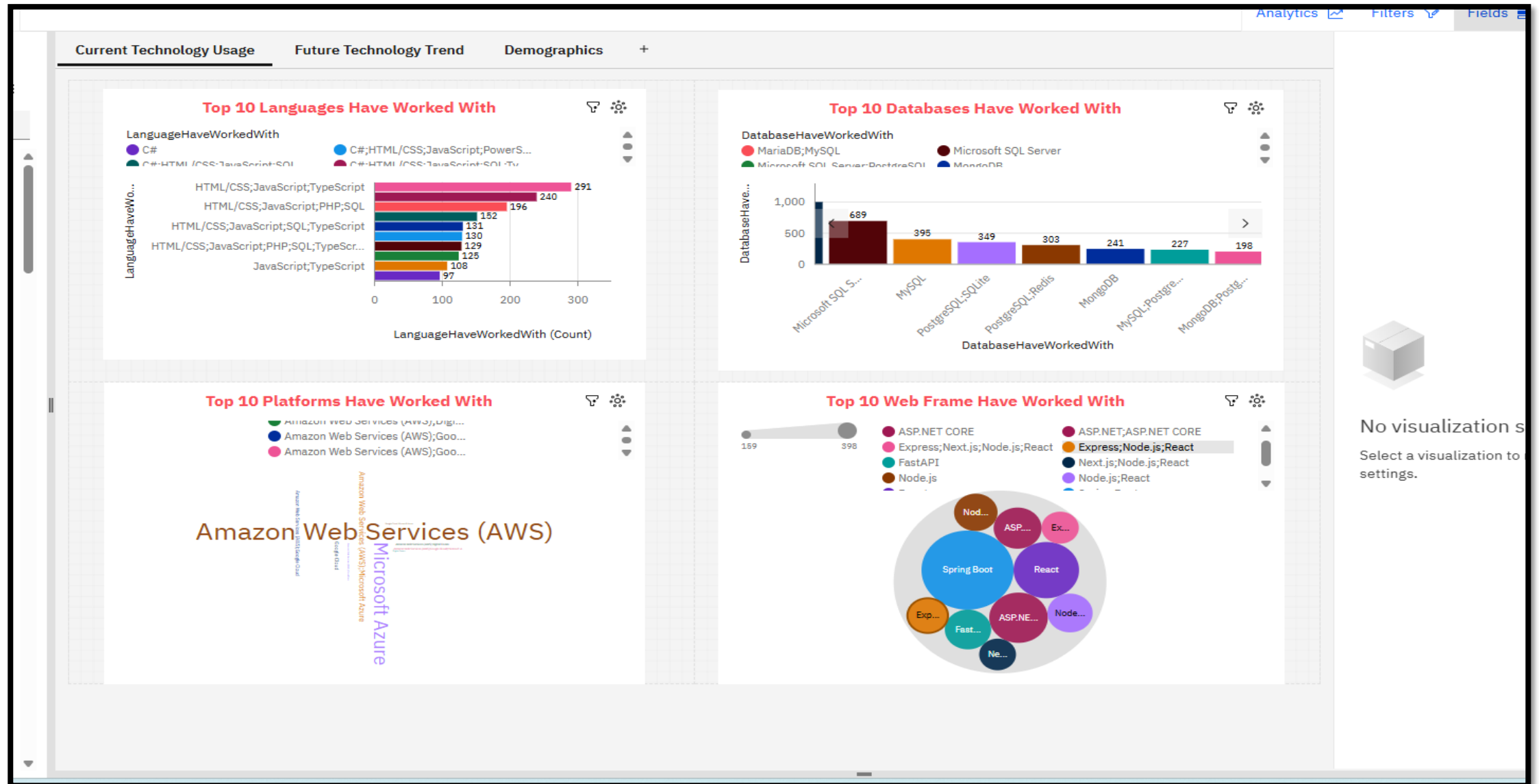
DASHBOARD TAB 2

Current Technology Usage | **Future Technology Trend** | Demographics | +
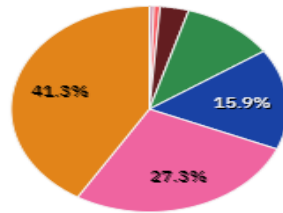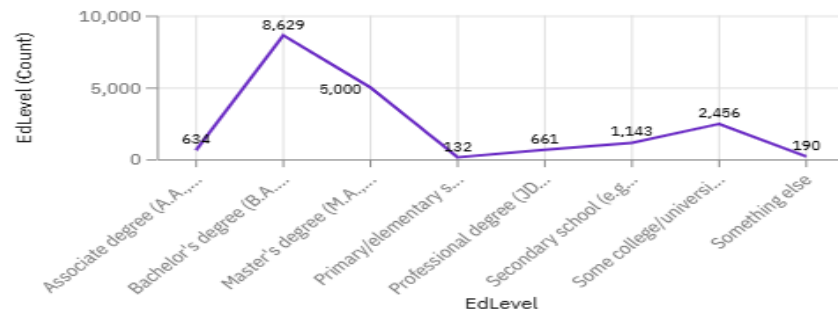
Top 10 Languages to Work With

LanguageWantToWorkWith
- C#
- C#;HTML/CSS;JavaScript;SQL;Ty...
- C#;HTML/CSS;JavaScript;SQL
- C#;SQL

Language	Count
C#;HTML/CSS;JavaScript;SQL;TypeS...	230
C#	134
C#;SQL	111
Python	104
JavaScript;TypeScript	103
HTML/CSS;JavaScript;PHP;SQL	97

LanguageWantToWorkWith (Count)

Top 10 Databases Want to Work With

DatabaseWantToWorkWith
- Microsoft SQL Server
- MongoDB
- MongoDB;PostgreSQL
- MySQL

Database	Count
PostgreSQL	1,265
Microsoft SQ...	622
PostgreSQL;...	567
PostgreSQL;...	446
MySQL	342
MongoDB	266
MongoDB;Po...	251
PostgreSQL;...	236
SQLite	223
MySQL;Post...	195

DatabaseWantToWorkWith

Top 10 PlatformWantToWorkWith

PlatformWantToW... 218 — 2,362
PlatformWantToW... 218 — 2,362

Amazon Web Services (AWS) 2,362	Amazon Web Se... 821	Amazon ... 538	Google Cl... 538
Microsoft Azure 1,649	Amazon Web S... 402	Amazon W... 268	Cloudflare 247
	Digital Ocean 278	Hetzner 257	Amazon ... 218

Top 10 WebframeWantToWorkWith

WebframeWantTo... 151 — 443

WebframeWantToWorkWith
- ASP.NET CORE
- ASP.NET CORE;Blazor
- FastAPI
- Flask

Spring Boot, React, FastAPI, Ph..., No..., Ne..., ASP.N..., ASP.NE..., Nod..., F...

DASHBOARD TAB 3

Discussion

Current Technology Usage

- Developers mostly use HTML/CSS/JavaScript/TypeScript (291) showing strong full-stack skills.
- PostgreSQL (994) and SQL Server (689) lead current database use.
- Amazon Web Services (AWS)2,301 usersMicrosoft Azure (1,931).
- React, Spring Boot, and ASP.NET Core dominate framework use.

Future Technology Trend

- Developers want to learn Elixir (**230**), HTML/CSS/JS/SQL/TypeScript (217), and Python (134) — showing interest in scalable, data-driven tools.
- PostgreSQL (**1,265**) stays the top desired database.
- AWS (**2,362**) and Azure (**1,649**) remain top cloud choices.
- React, FastAPI, and Phoenix show rising framework interest.

Demographics

- Most respondents are 25–34 years old (41.3%), mid-career professionals.
- Bachelor's (**8,62**9) and Master's (5,000) degrees dominate.
- Strong participation from the U.S., India, and Europe shows global reach.

OVERALL FINDINGS & IMPLICATIONS

- Developers primarily use HTML/CSS, JavaScript, and TypeScript (291), with strong reliance on PostgreSQL (994) and cloud platforms such as **AWS** (2,301) and **Azure** (1,931).

- Future preferences remain focused on HTML/CSS, JavaScript, SQL, and TypeScript (230–217), showing developers aim to deepen existing full-stack skills rather than shift to new languages.

- The developer community is young, globally distributed, and highly educated, mostly aged **25–44** with bachelor's or master's degrees, reflecting a skilled and diverse workforce.

- Organizations should invest in training and recruitment around dominant and enduring technologies such as JavaScript, TypeScript, PostgreSQL, and major cloud platforms (AWS, Azure).

- Companies relying on outdated or single-stack frameworks should consider modernizing their tech environments to remain competitive and attract skilled developers..

- The continued rise of cloud-based, full-stack, and open-source ecosystems reinforces the need for adaptability, cross-platform expertise, and continuous upskilling within the tech workforce..

Conclusion

- The dashboards revealed clear patterns in both **current and future technology preferences**, offering a data-driven view of developer behavior and emerging trends.

- Web technologies such as HTML, CSS, JavaScript, and TypeScript, along with PostgreSQL and cloud platforms like AWS (2,301) and Azure (1,931), remain the dominant tools across the developer community.

- Future preferences indicate a continued focus on modern full-stack ecosystems, where developers aim to deepen expertise in existing frameworks such as React, Spring Boot, and ASP.NET Core, emphasizing scalability and integration with cloud environments.

- The project demonstrates the power of IBM Cognos Analytics and Google Looker Studio in transforming complex survey data into interactive dashboards that drive clear, actionable insights..
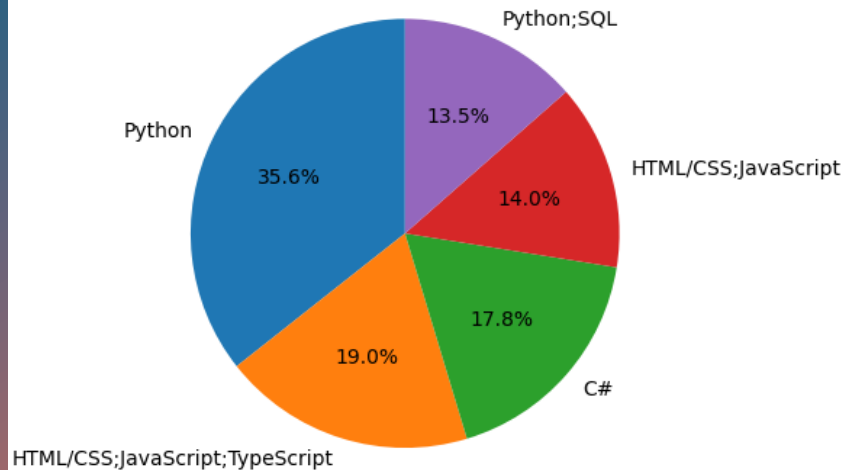
APPENDIX

Top 5 Most Desired Databases


Top 5 Languages Admired

Most Admired Languages and Databases
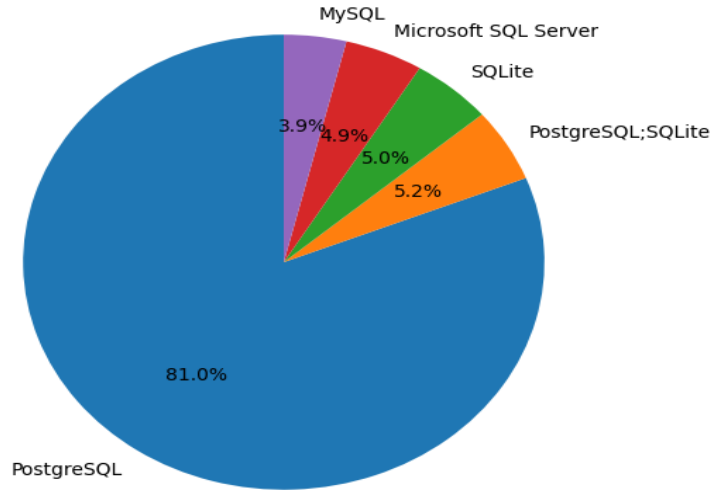
- PostgreSQL overwhelmingly leads as the most desired database, capturing **81%** of developer interest, far ahead of SQLite and MySQL.
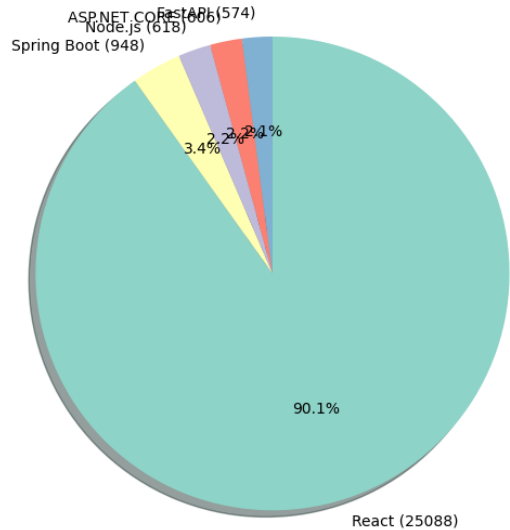
- **Python** stands out as the most admired language (**35.6%**), followed by **HTML/CSS/JavaScript/TypeScript (19%)** and **C# (17.8%)**.

- Together, these trends reflect developers' strong preference for open-source, scalable technologies that emphasize simplicity, flexibility, and performance.

Geographic and Work-Mode Distribution of Developers
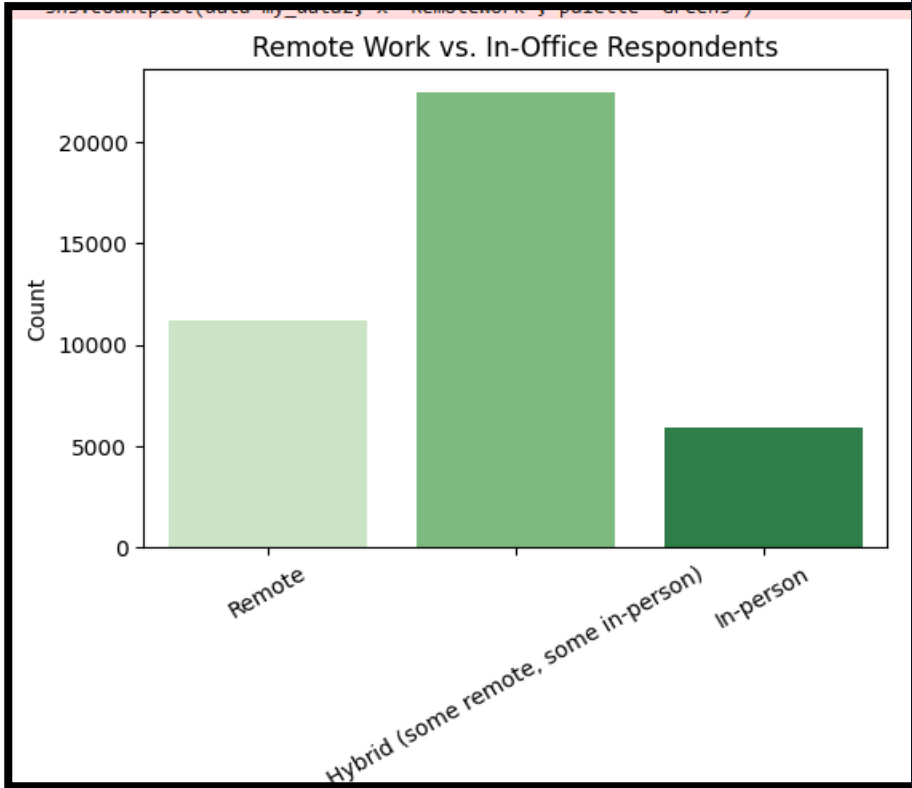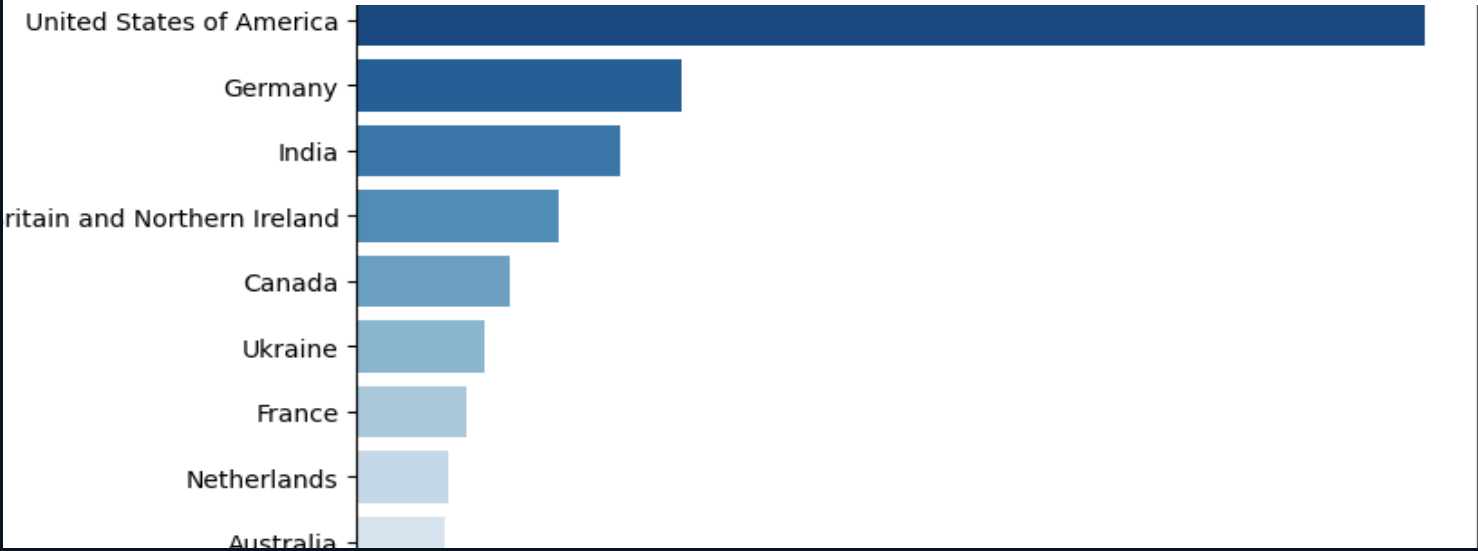
- The United States leads in developer representation, followed by Germany, India, and the United Kingdom, showing strong participation from major global tech hubs.

- A majority of respondents work in hybrid environments, while fully remote work remains more common than fully in-person roles.

- These results highlight an increasingly global and flexible workforce, where distributed teams and cross-border collaboration are the new norm.

Top 5 Web Frameworks Developers Want to Work With

ASP.NET CORE (1006) FastAPI (574)
Node.js (618)
Spring Boot (948)

3.4% 2.2% 2.2% 2.4%

90.1%

React (25088)

Top 5 Developer Types

Developer, desktop or enterprise applications

Developer, front-end

5.9%

7.9%

Student

12.0%

50.6% Developer, full-stack

23.5%

Developer, back-end

# Developer Roles and Preferred Web Frameworks

- React overwhelmingly dominates future framework preferences, with 90.1% of developers choosing it, far ahead of Spring Boot (3.4%) and Node.js (2.2%).

- Over half of respondents identify as full-stack developers (50.6%), followed by back-end developers (23.5%) and students (12%).

- These findings emphasize that the developer community is strongly oriented toward versatile, full-stack roles and modern JavaScript frameworks driving front-end innovation.
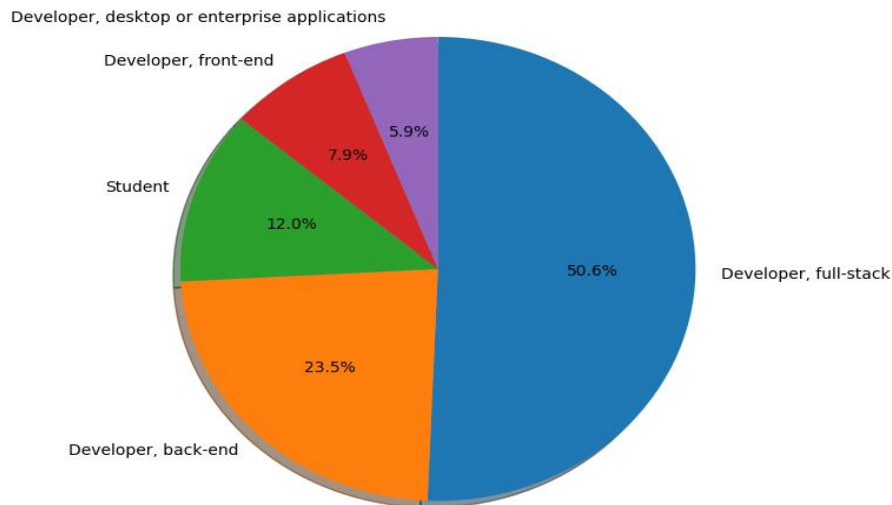
Stack Overflow Participation Patterns

- Most developers engage with Stack Overflow infrequently, with the largest group participating less than once per month.

- A significant portion reported *never participating at all*, showing that many developers use the platform more as a reference than as active contributors.

- This trend highlights a passive engagement culture, where Stack Overflow serves primarily as a learning and troubleshooting resource rather than a collaborative forum.


Stack Overflow Participation Frequency

JOB POSTINGS

Job Postings by Technology colored by Technology
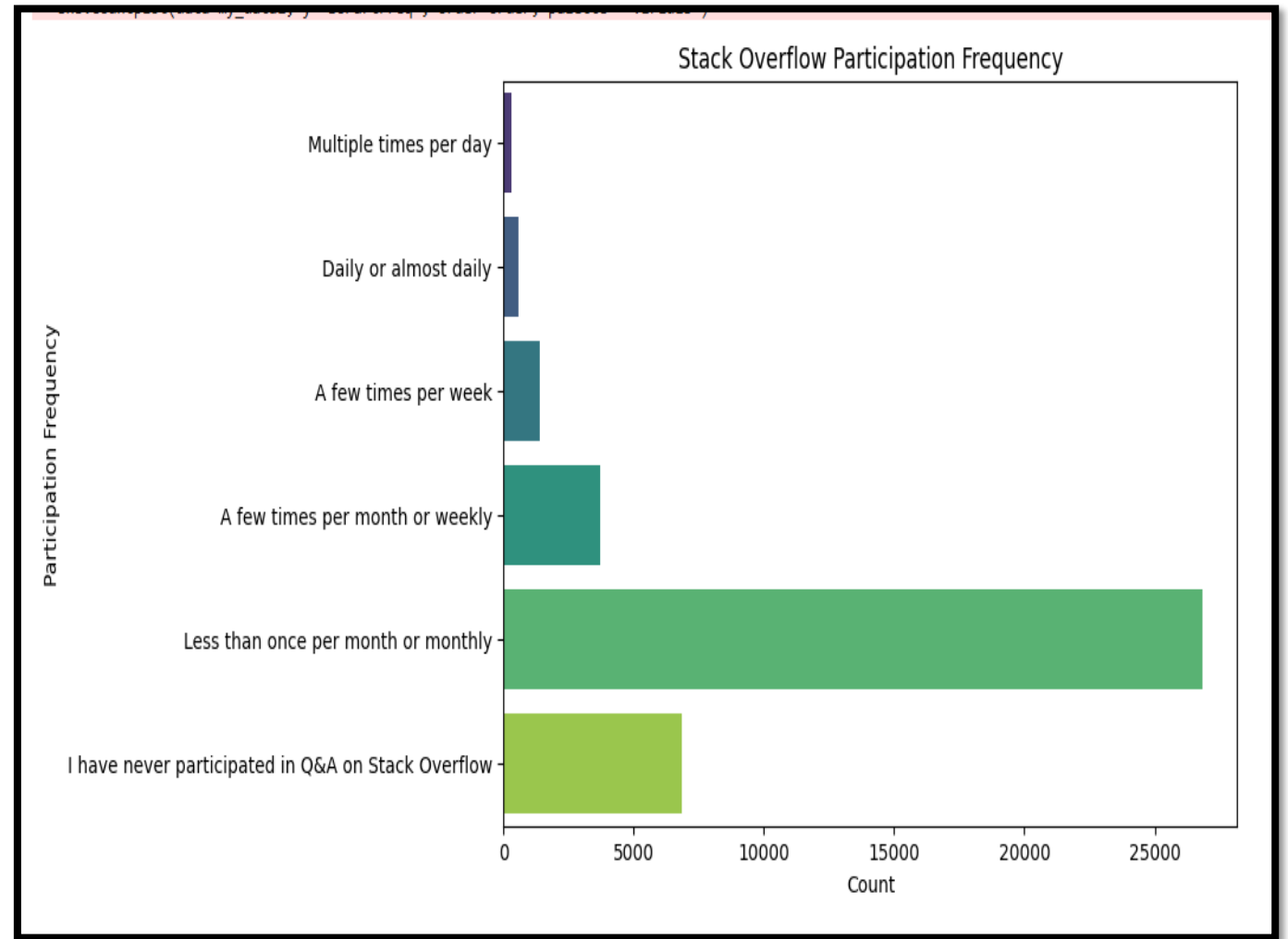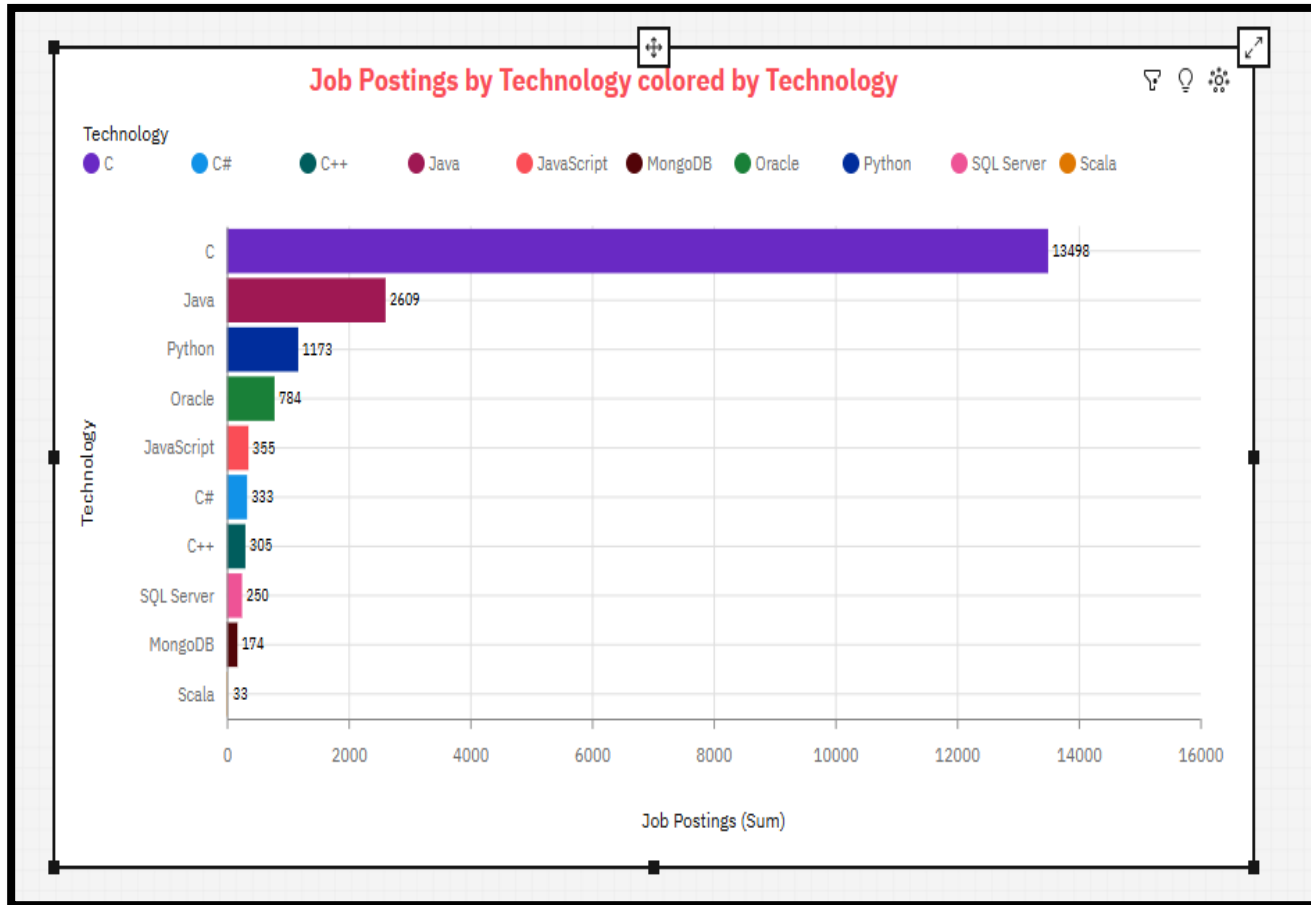
- **C** stands out as the most in-demand technology, with 13,498 job postings, reflecting its continued dominance in low-level and system programming roles.

- In contrast, Scala shows the lowest demand with only 33 postings, indicating its limited use in current job markets compared to mainstream languages.

- Overall, the data suggests that employers prioritize widely adopted, versatile languages like C and Java (2,609) over niche or specialized technologies.

POPULAR LANGUAGES

Salary by language colored by language
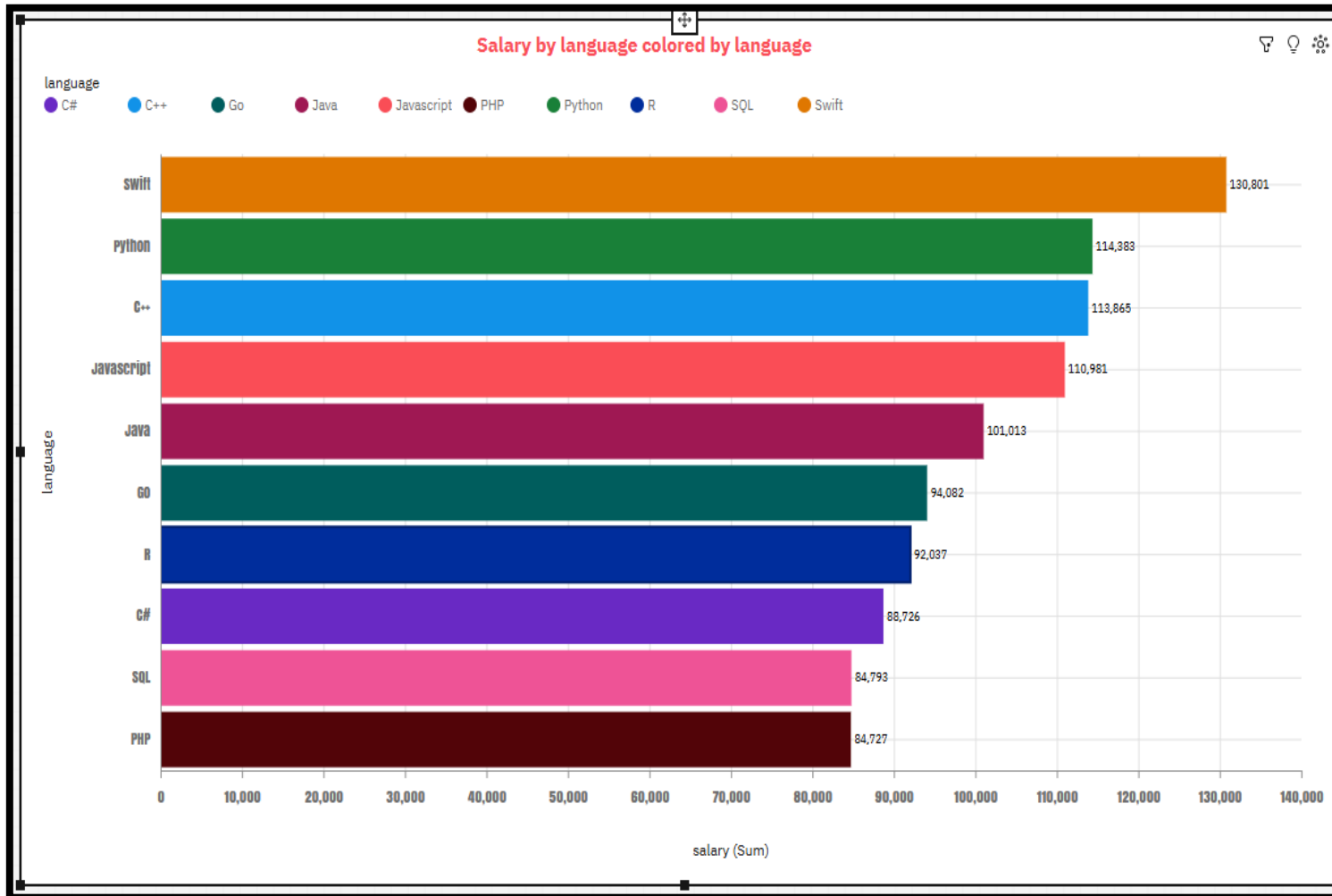
- Among the popular programming languages, Swift offers the highest average salary at $130,801, followed by Python ($114,383) and C++ ($113,865), reflecting their strong industry demand.

- Languages like PHP ($84,727) and SQL ($84,793) fall on the lower end, indicating their maturity and wider talent availability.

- Overall, the data shows that newer or specialized languages tend to command higher salaries, while legacy and widely used ones remain stable but less lucrative.

Thank you