



IBM Data Analyst Capstone Project

Valentine Golden Ghanem
January 03, 2024

OUTLINE



- Executive Summary
- Introduction
- Methodology
- Results
 - Visualization – Charts
 - Dashboard
- Discussion
 - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY



Insights from the Survey

- Technological Landscape: Python dominance, diverse database, and platform preferences.
- Demographics: Gender disparity, younger developer concentration, varied global distribution.
- Industry Implications: Guidance for education, inclusivity initiatives, equitable opportunities.
- Ethical Considerations: Data privacy paramount, continued commitment to ethical practices.

Conclusion

- Guiding the Tech Industry: Insights as a compass towards innovation, equity, and growth.

INTRODUCTION



- Overview of the Stack Overflow Developer Survey 2019 as a comprehensive dataset capturing developer insights.
- Objective of the report to extract trends in technology usage, anticipate future trends, and explore demographic distributions.
- Three primary facets of analysis: Current Technology Usage, Future Technology Trends, and Demographics.
- Point4
 - Utilization of IBM Watson Studio for data exploratory analysis and IBM Cognos Analytics for visualization and insights.
 - Utilization of Microsoft Powerpoint For presentation of report.

INTRODUCTION



Current Technology Usage

- Unveiling the prevalent landscape of programming languages, databases, platforms, and web frameworks used by developers.
- Exploring the data on the current technology preferences.
- Breakdown of programming language usage.
- Distribution of database and platform usage.

INTRODUCTION



Future Technology Usage

- Anticipating the future trajectory of technology adoption based on developer aspirations.
- Examining the projected technology preferences for the upcoming year
- Aspired programming languages.
- Desired databases, platforms, and web frameworks.

INTRODUCTION



Demographics

- Understanding the diverse characteristics within the developer community.
- Exploring demographic data such as gender distribution, geographic spread, and age groups.
- Correlation between education levels and gender representation.
- Other demographic insights from the survey.

INTRODUCTION



Analytical Tools

- Description of IBM Watson Studio's role in conducting robust data exploratory analysis.
- Utilization of IBM Cognos Analytics for visualization and deriving insights.

METHODOLOGY



Data Collection and Preparation

- Data Sources
- Preparation Process
- Python Libraries Utilized
- Data Loading in Python
 - Loading into Pandas
 - Utilizing IBM Watson Studio for EDA

METHODOLOGY



Data Cleaning and Preprocessing

- Cleaning Procedure
- Ensuring Data Accuracy
- Handling Anomalies and Outliers
- Preparation for Analysis
 - Missing Values and Duplicates
 - Standardizing Column Names

METHODOLOGY



Current Technology Usage Analysis

- Focus of Analysis
- Identifying Top Technologies
- Respondent Usage Patterns
- Dataset Source
 - Top 10 Languages
 - Databases, Platforms, and Frameworks

METHODOLOGY



Future Technology Trends Analysis

- Anticipated Developer Preferences
- Survey-Based Insights
- Visualization Approach
- Top Technologies for Upcoming Year
 - Desired Languages
 - Anticipated Platforms and Frameworks

METHODOLOGY



Demographics Analysis

- Understanding Respondent Profiles
- Demographic Visualization
- Gender, Country, and Age Insights
- Education-Level Analysis
 - Respondent Counts
 - Gender-Specific Education Levels

METHODOLOGY



Data Visualization and Interpretation

- Utilizing IBM Cognos Analytics
- Visual Representation Types
- Creating Insights
- Actionable Interpretations
 - Bar Charts and Histograms
 - Interpreting Visual Insights

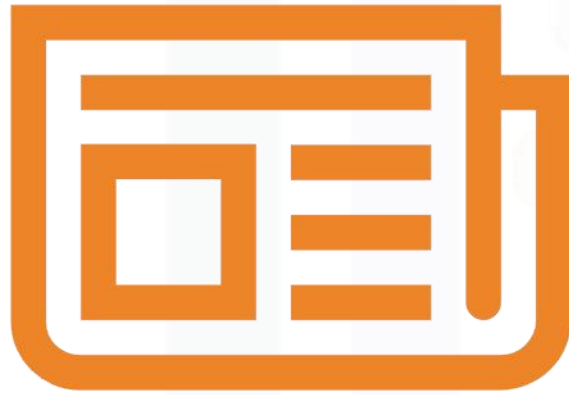
METHODOLOGY



Limitations and Ethical Considerations

- Analysis Limitations
- Survey Data Biases
- Sample Representativeness
- Ethical Data Handling
 - Dataset Inherent Limitations
 - Maintaining Respondent Anonymity

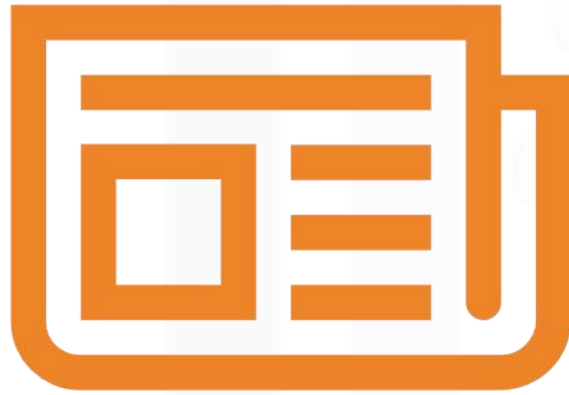
RESULTS



Current Technology Usage

- Top 10 Languages
Python, JavaScript, HTML/CSS, SQL, Bash etc leading usage.
- Top 10 Databases
MySQL, SQL Server, MongoDB, and PostgreSQL prominence.
- Platforms
Windows, Linux, Docker, Android, and AWS prevalent.
- Top 10 Web Frameworks
jQuery, AngularJS, React.js, ASP.NET usage highlights.

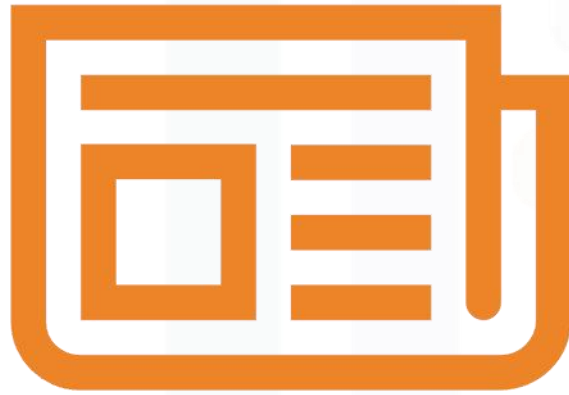
RESULTS



Future Technology Trends

- Desired Languages for Next Year
Python remains popular; TypeScript, Go, Kotlin rising.
- Desired Databases for Next Year
MongoDB, PostgreSQL, Redis, Elasticsearch growth.
- Desired Platforms for Next Year
AWS, Docker consistent; Kubernetes, Raspberry Pi interest.
- Desired Web Frameworks for Next Year
React.js, AngularJS, Vue.js, Express anticipation.

RESULTS



Demographics Overview

- Gender Distribution

Male dominance, significant gender imbalance.

- Respondent Distribution by Countries

Variation across countries; significant responses from the US, India, UK.

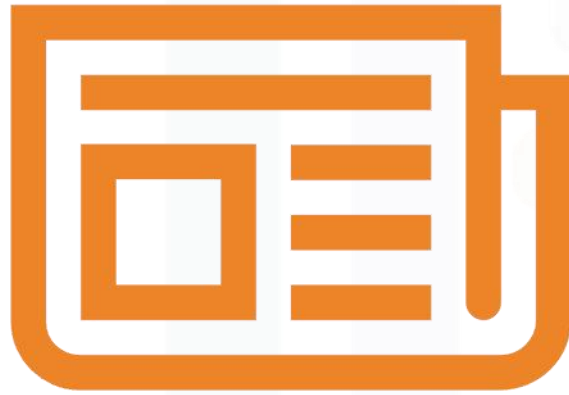
- Respondent Distribution by Age

Concentration in mid-20s to mid-30s, indicating a younger demographic.

- Gender-Based Education Level

Male dominance across education levels; potential gender gaps.

RESULTS



Technology Usage, Trends, and Demographics Insights

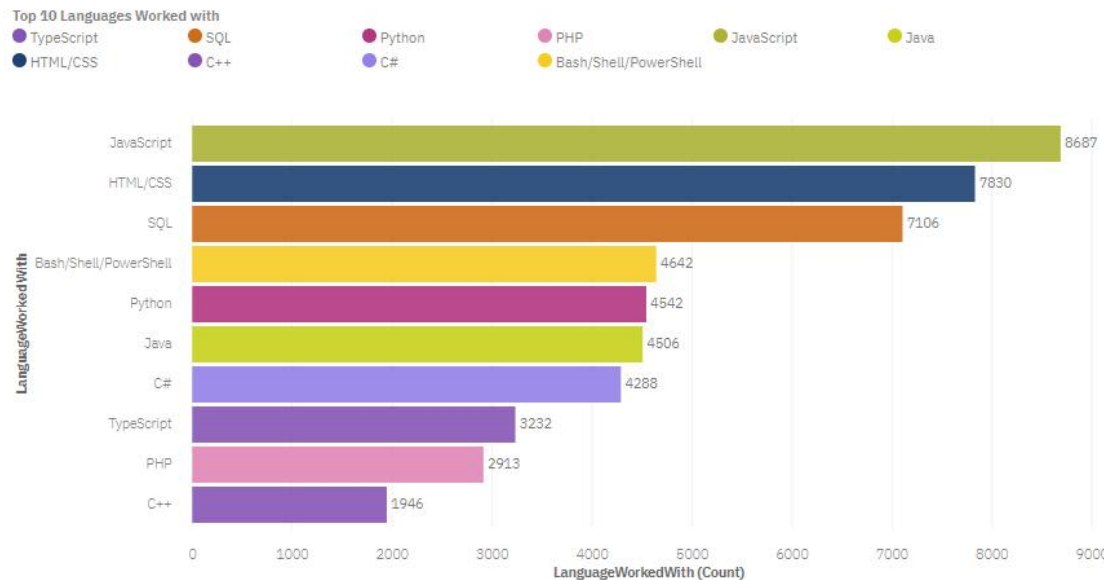
- Key Insights from Technology Usage
 - Python's Dominance
 - Diverse Database and Platform Usage
- Future Technology Trends Insights
 - Consistent Language Preferences
 - Emerging Interest in New Technologies
- Demographics Insights
 - Gender Disparity and Male Dominance
 - Youthful Respondent population between mid 20's and mid 30's
 - Massive imbalance in Formal Education with Male Dominance
- Gender-Based Education Level
 - Male dominance across education levels; potential gender gaps.

PROGRAMMING LANGUAGE TRENDS

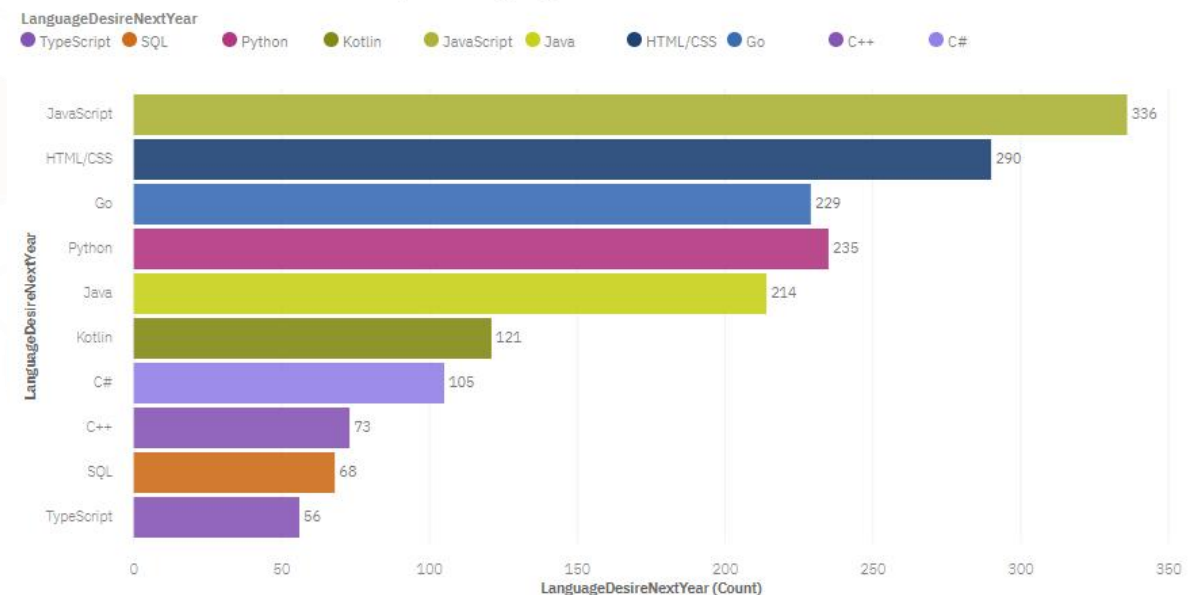
Current Year

Next Year

Top 10 Languages Tools Worked With



Top 10 Languages Desired Next Year



PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

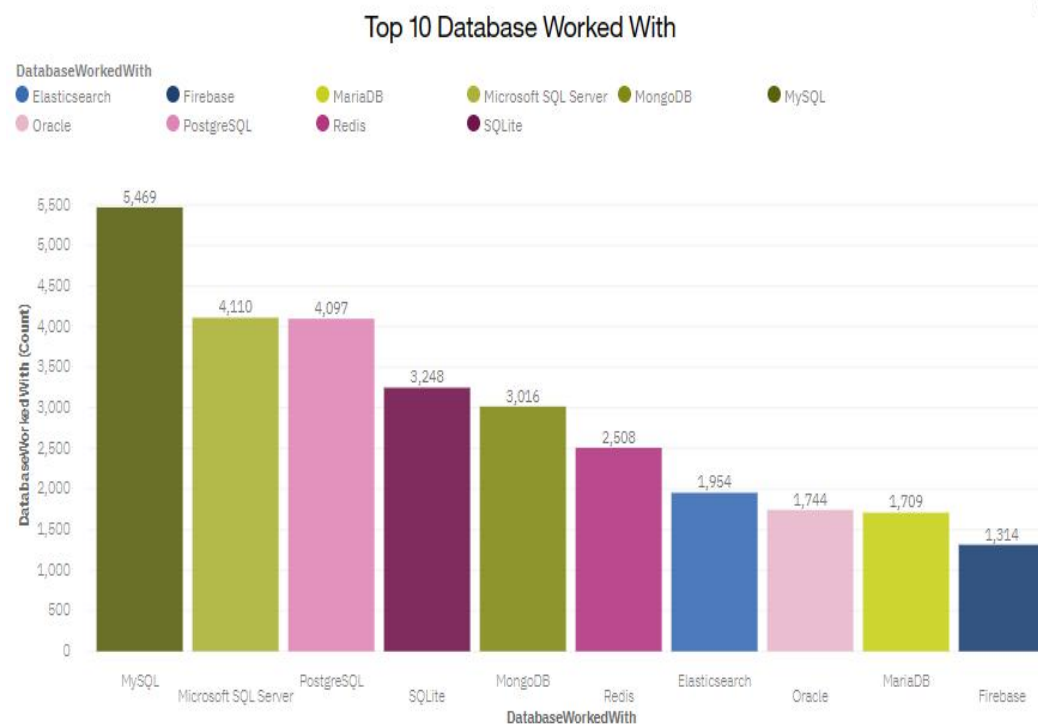
- Python, JavaScript, HTML/CSS, SQL, Bash etc. were amongst the widely used programming languages
- Python emerged as one of the most popular language.

Implications

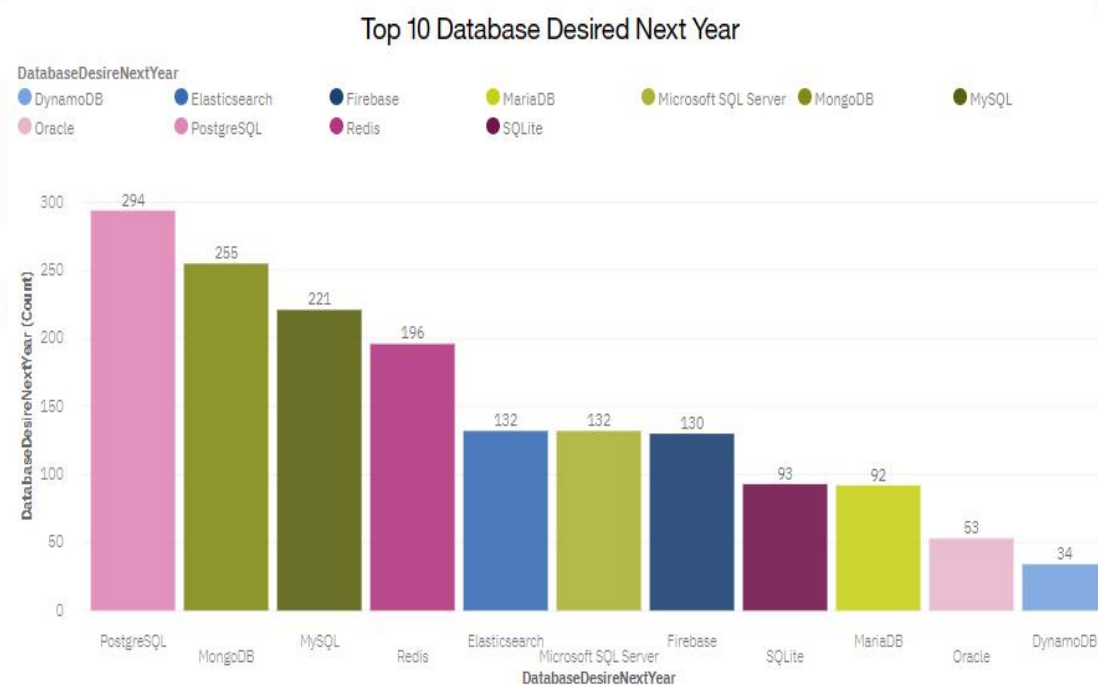
- Its widespread use indicates its versatility, ease of use, and popularity among developers, potentially

DATABASE TRENDS

Current Year



Next Year



DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- Diverse Database Usage: MySQL, SQL Server, MongoDB, and PostgreSQL were prominently used.

Implications

- Developers employ a range of databases, indicating varied project requirements and preferences, potentially affecting data storage and handling decisions.

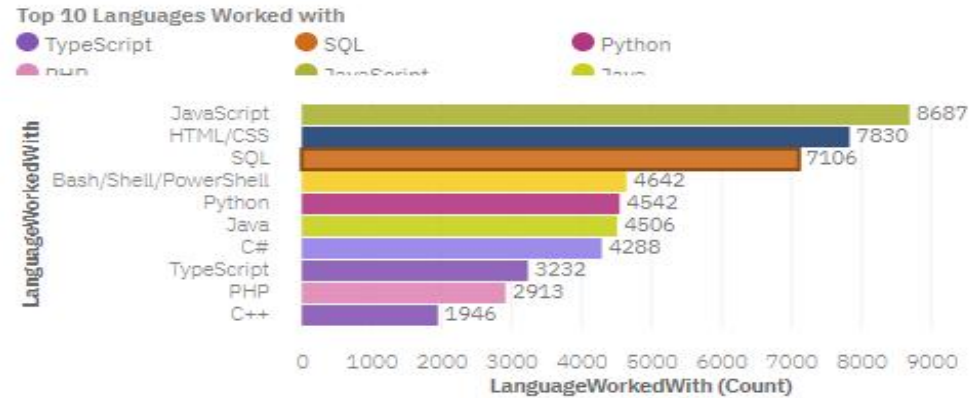
DASHBOARD



<https://eu-de.dataplatform.cloud.ibm.com/dashboards/6534426d-80f7-4f69-98cf-d8bcc7cfbd84/view/0711c32d1c9029e776d6f6e407917d0578372d5ce0bbd65683d77b490b367397a96e15c7c82d1a598f475463f6e4410ccc>

DASHBOARD TAB 1

Top 10 Languages Tools Worked With



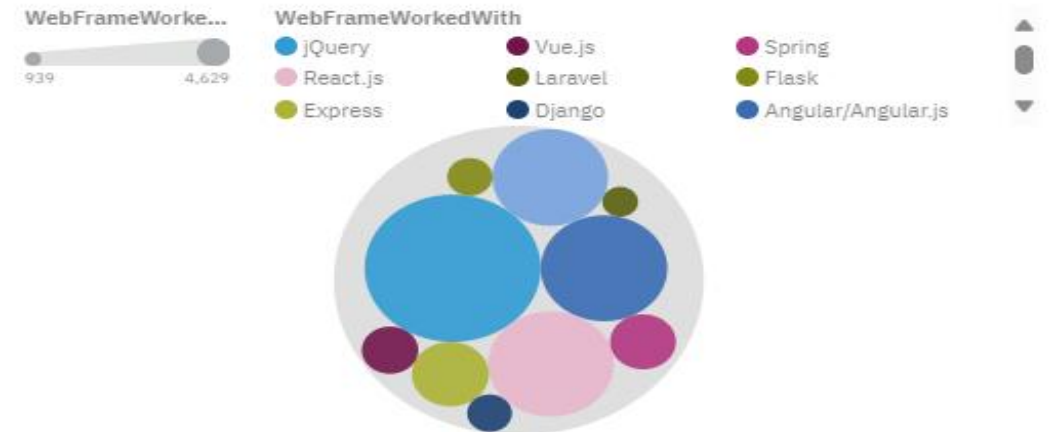
Top 10 Database Worked With



Platform Worked With

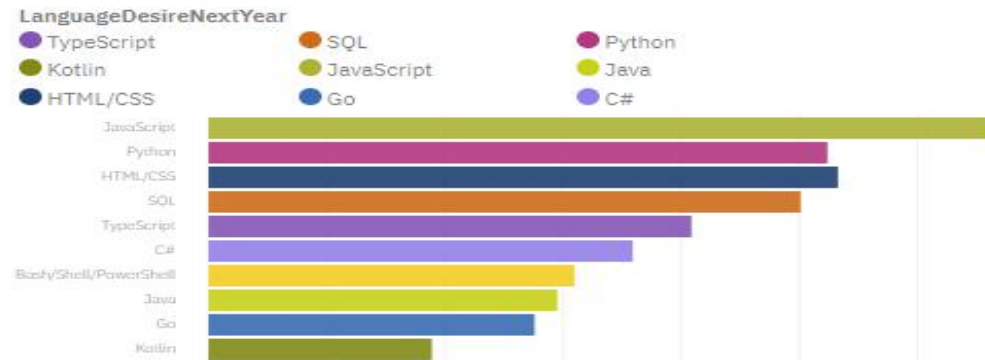


Top 10 Web Frame Worked With

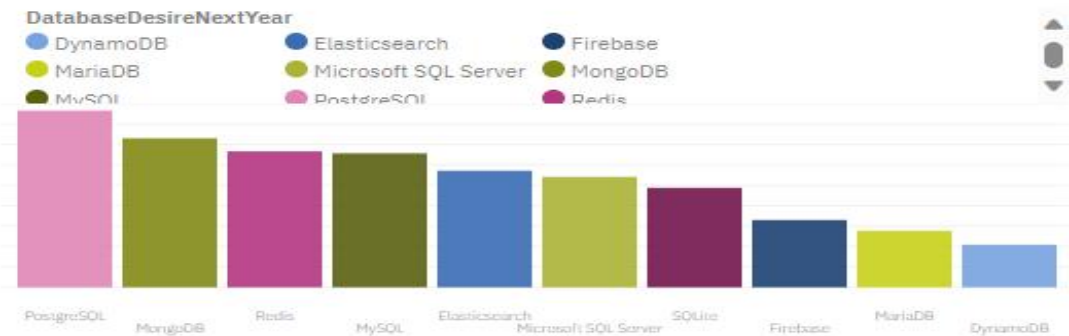


DASHBOARD TAB 2

Top 10 Languages Desired Next Year



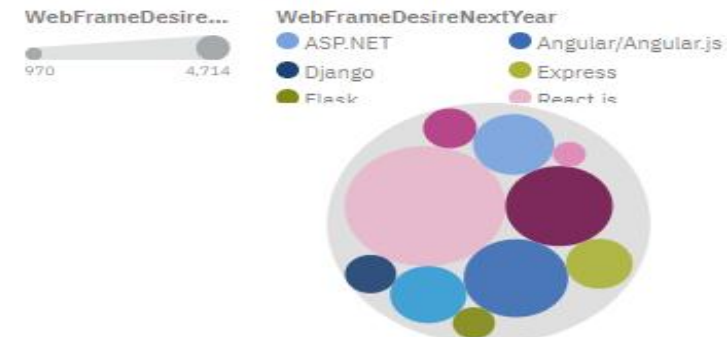
Top 10 Database Desired Next Year



Platform Desired Next Year



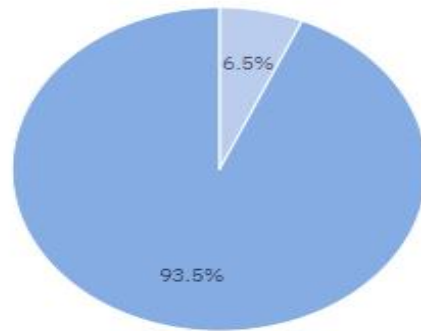
Top 10 Web Frame Desired Next Year



DASHBOARD TAB 3

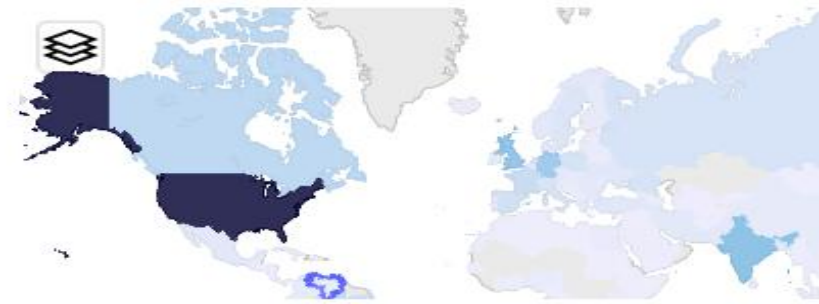
Respondent classified by Gender

Gender
● Woman ● Man

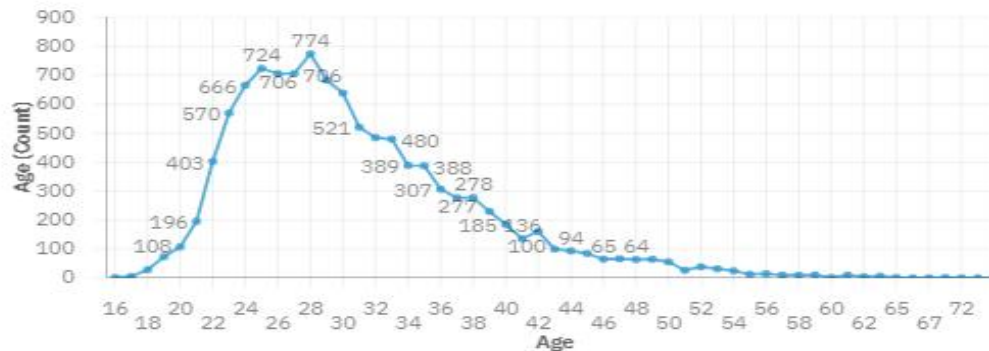


Respondent Count for Countries

Country (Count)



Respondent Count by Age



Respondent Count by Gender, classified by Formal Education Level

EdLevel

● Associate degree ● Bachelor's degree (BA, BS, B.E...
● Master's degree (MA, MS, M.E... ● Other doctoral degree (Ph.D, E...
● Professional degree (JD, MD, ...



DISCUSSION



- Insights and Implications

Current Technology Usage and Implications

- Python's Dominance: Reflects versatile adoption and potential influence on future trends.
- Diverse Technology Usage: Highlights complex project requirements and the need for compatibility optimization.

Future Technology Trends and Significance

- Persistent Desire for Python: Signals stability and guides educational priorities.
- Emerging Language Interest: Indicates evolving preferences and potential new paradigms.

DISCUSSION



Demographics and Industry Dynamics

- **Gender Disparity:** Highlights ongoing challenges for a more inclusive tech community.
- **Geographical and Age Distribution:** Emphasizes younger developer concentrations and implications for talent strategies.
- **Gender-Based Education Disparities:** Underlines the need for targeted efforts in tech education inclusivity.

DISCUSSION



Limitations and Ethical Considerations

- Acknowledging Survey Limitations: Biases and representativeness impacting findings' generalizability.
- Ethical Adherence: Rigorous privacy maintenance for respondent confidentiality.

Conclusion and Recommendations

- Summary of Insights: Valuable guidance for stakeholders in decision-making and fostering inclusivity.

OVERALL FINDINGS & IMPLICATIONS

Findings

- **Python Dominance:** Python emerged as the most popular language among developers in both current usage and future preferences.
- **Diverse Database Usage:** MySQL, SQL Server, MongoDB, and PostgreSQL were prominently used databases.
- **Platform Preferences:** Windows, Linux, Docker, Android, and AWS were commonly used platforms.

Implications

- Its widespread adoption signifies its pivotal role in the developer landscape, impacting tooling, education, and project choices
- The diversity in database usage reflects the need for adaptable data storage solutions, guiding decisions in data architecture and management.
- Understanding these preferences assists in aligning development environments and infrastructure choices for compatibility and efficiency.

OVERALL FINDINGS & IMPLICATIONS

Findings

- **Stability in Technology Trends:** Consistent desires for certain languages, databases, platforms, and web frameworks for the upcoming year.
- **Gender Disparity:** A significant gender imbalance was observed among respondents.
- **Geographic and Age Distribution:** Concentration of respondents in the mid-20s to mid-30s range with varied global distribution.

Implications

- **Stability in preferences** guides educational priorities, influences project planning, and supports strategic decision-making.
- **Highlighting existing gender gaps** emphasizes the need for diversity initiatives and inclusive practices within the tech industry.
- **Understanding the demographic trends** aids in talent acquisition, educational programs, and industry targeting strategies.

OVERALL FINDINGS & IMPLICATIONS

Findings

- **Education Disparities:** Male respondents showed higher distributions across various education levels compared to females.
- **Continued Preference for Established Technologies:** AWS, Docker, and established languages continued to be favored.
- **Emerging Interest in New Technologies:** TypeScript, Go, Kotlin, Redis, Elasticsearch, Kubernetes, Raspberry Pi, Vue.js, and Express showcased growing interest.

Implications

- Identifies potential gender gaps in educational opportunities, stressing the need for inclusivity in tech education initiatives.
- Consistent preferences influence infrastructure planning, technology adoption, and skill development initiatives.
- Emerging technologies may shape future trends, guiding skill development and project planning.

OVERALL FINDINGS & IMPLICATIONS

Findings

- Ethical Considerations: Adherence to data privacy and ethical standards throughout the analysis.

Implications

- Ensuring data confidentiality and privacy is crucial for maintaining trust and credibility in data-driven studies

CONCLUSION



Technological Landscape Insights:

- Python's Dominance: Influential role in development practices.
- Diverse Technology Usage: Complex project requirements and adaptability needs.

Demographic Dynamics and Industry Challenges:

- Gender Disparity and Age Concentration: Highlighting inclusivity needs.
- Education Disparities: Urgency for equitable opportunities in tech education.

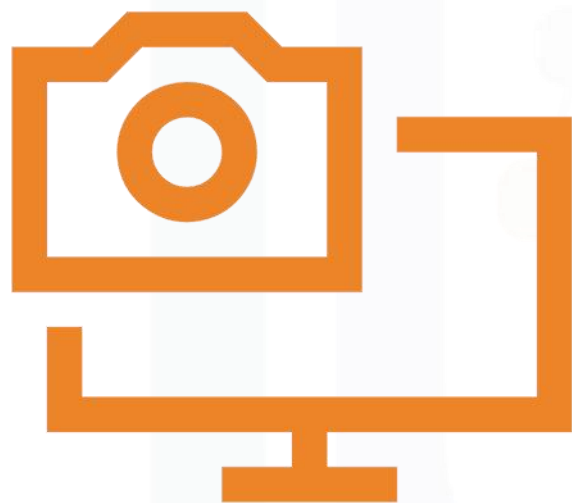
Implications for the Tech Industry

- Stable Technology Trends: Guiding educational priorities and infrastructure planning.
- Emerging Interests: Shaping future trends and skill development initiatives.

Ethical Considerations and Future Directions

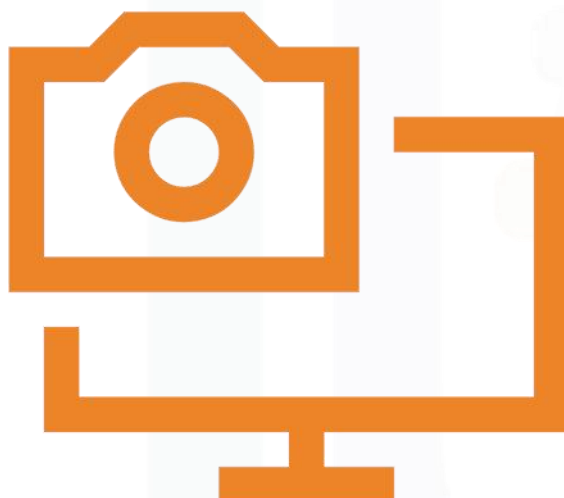
- Adherence to Ethical Standards: Crucial for maintaining data integrity.
- Inclusivity Initiatives: Imperative for a diverse, innovative tech ecosystem.

APPENDIX



Demographics Data								
Respondent	MainBranch	Hobbyist	OpenSourcer	OpenSource	Employment	Country	Student	EdLevel
4	I am a developer by profession	No	Never	The quality of OSS and closed source software ...	Employed full-time	United States	No	Bachelor's degree (BA, BS, B.Eng., etc.)
9	I am a developer by profession	Yes	Once a month or more often	The quality of OSS and closed source software ...	Employed full-time	New Zealand	No	Some college/university study without earning ...
13	I am a developer by profession	Yes	Less than once a month but more than once per ...	OSS is, on average, of HIGHER quality than pro...	Employed full-time	United States	No	Master's degree (MA, MS, M.Eng., MBA, etc.)
16	I am a developer by profession	Yes	Never	The quality of OSS and closed source software ...	Employed full-time	United Kingdom	No	Master's degree (MA, MS, M.Eng., MBA, etc.)

APPENDIX



Technology Data					
Respondent	LanguageWorkedWith	LanguageDesireNextYear	DatabaseWorkedWith	DatabaseDesireNextYear	PlatformWorkedWith
NaN	NaN	NaN	NaN	NaN	NaN
4	C	C	MySQL	MySQL	Linux
4	C++	C#	SQLite	SQLite	Windows
4	C#	JavaScript	NaN	NaN	NaN