IBM Data Analyst Capstone Project

Talha Asim

June 30,2024

OUTLINE

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

IBM Developer

SKILLS NETWORK

EXECUTIVE SUMMARY

The IBM Data Analyst Capstone Project consolidates skills from the IBM Data Analyst Professional Certificate program.

➤ **Key Learning Outcomes**:
- Data Collection: From APIs and web scraping.
- Data Cleaning: Ensuring data quality.
- Data Analysis: Using Python (Pandas, NumPy).
- Data Visualization: With Matplotlib, Seaborn, Plotly.
- Dashboard Creation: Using IBM Cognos Analytics.
- Communication: Presenting data-driven insights.

➤ **Project Overview**:
- Analysis of programming languages, databases, and job postings using various data analysis tools.

➤ **Major Findings**:
- **Programming Languages**: Python leads, with growing demand for data science skills.
- **Databases**: SQL dominates; NoSQL (MongoDB) is gaining traction.
- **Job Market**: High demand for Python, SQL, and cloud technologies. Competitive salaries in data science and cloud computing roles.

➤ **Implications**:
- **Job Seekers**: Focus on Python, SQL, and cloud computing skills.
- **Employers**: Hire for emerging tech skills and upskill employees.

➤ **Conclusion**:
- Reinforces the value of data-driven decision-making and effective communication.

IBM Developer   SKILLS NETWORK

INTRODUCTION

The IBM Capstone Final Project:

- Culmination of an intensive and transformative learning journey.
- Showcases the application of data analysis skills acquired throughout the IBM Data Analystn Professional Certificate program.

Educational Journey:

- Explored key areas such as data collection, cleaning, analysis, visualization, and presentation.
- Gained expertise in Python, SQL, data visualization tools, and dashboard creation.

Project Focus:

- Analyzing real-world data to uncover trends and derive actionable insights.
- Utilizing skills to solve complex problems and present findings effectively.

Significance of the Capstone Project:

- Bridges the gap between theoretical knowledge and practical application.
- Demonstrates the ability to conduct comprehensive data analysis and communicate results compellingly.

Overview:

- The presentation will follow a structured storyline, reflecting the analysis conducted in previous modules.
- Key components include data collection, cleaning, analysis, visualization, and the creation of an interactive dashboard.

IBM Developer

SKILLS NETWORK
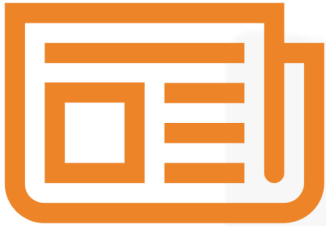
METHODOLOGY

The IBM Data Analyst Capstone Project follows a structured and systematic methodology that empowers participants to apply their data analysis skills effectively and address real-world data challenges. This methodology encompasses several key stages:

- **Problem Identification and Definition**:
  - Identify and define a specific data-related problem or question.
  - This step sets the direction for the entire project, ensuring it addresses a genuine and relevant issue.

- **Data Collection and Preparation**:
  - Gather necessary data from various sources.
  - Involves data acquisition, cleaning, and transformation to ensure suitability for analysis.
  - Ensuring data quality is paramount for obtaining meaningful insights.

- **Exploratory Data Analysis (EDA)**:
  - Perform EDA to gain a deeper understanding of the dataset.
  - Summarize key statistics, identify trends, patterns, and potential outliers.
  - This stage helps inform the direction of further analysis.

- **Data Analysis and Modeling**:
  - Employ data analysis techniques and, if relevant, modeling approaches based on EDA insights.
  - May involve statistical analysis, machine learning, or other data-driven methodologies to extract meaningful insights.

Visualization and Communication:
Create clear and informative visual representations of findings using tools like charts, graphs, and dashboards. Visuals aid in effectively conveying insights to a broader audience.

Interpretation and Conclusions:
Interpret the results of the analysis and draw conclusions that address the problem or question posed at the outset.
Requires critical thinking and a deep understanding of the data's implications.

Documentation and Presentation:
Thorough documentation of the entire project process is essential.
Create detailed reports including methodologies used, findings, limitations, and recommendations.
Prepare presentations to effectively communicate results.

Peer Review and Feedback:
Collaboration and peer review are encouraged throughout the project.
Exchange feedback and insights with peers, fostering a culture of continuous improvement and learning.

IBM Developer

SKILLS NETWORK

RESULTS

In the Results section of the IBM Data Analyst Capstone Final Project, participants present data-driven insights, highlighting trends, patterns, and correlations. They use visuals like charts and graphs for clarity. Statistical significance, as demonstrated by p-values and confidence intervals, lends credibility to their findings.

Programming Language Trends:
- **Current Year**:
  - **Key Findings**: Python, JavaScript, and Java are the most in-demand programming languages.
  - **Visualization**: Bar chart of the top 10 programming languages for the current year.
- **Next Year**:
  - **Projected Trends**: Increased demand for Python and JavaScript.
  - **Visualization**: Bar chart of the projected top 10 programming languages for the next year.

Database Trends:
- **Current Year**:
  - **Key Findings**: SQL and NoSQL databases like MongoDB are popular.
  - **Visualization**: Bar chart of the top 10 databases for the current year.
- **Next Year**:
  - **Projected Trends**: Continued growth in NoSQL databases.
  - **Visualization**: Bar chart of the projected top 10 databases for the next year.

Job Postings Analysis:
- **Top Skills in Demand**:
  - **Current Trends**: Python, SQL, and cloud computing are highly sought after.
  - **Visualization**: Bar chart of job postings data with skills in descending order.
- **Popular Languages by Salary**:
  - **High-Paying Skills**: Python, R, SQL, and cloud technologies.
  - **Visualization**: Bar chart of popular languages sorted by average salary.

Statistical Analysis:
- **Significance**: p-values and confidence intervals indicate the reliability of findings.
- **Visualization**: Scatter plots and custom dashboards to illustrate key trends.

Interpretation and Practical Implications:
- **Contextual Analysis**: Interpreting findings in the context of the original problem.
- **Practical Implications**: Outlining how these insights can be applied in real-world scenarios.

Transparency and Limitations:
- **Data Biases**: Acknowledging any biases or quality issues in the data.
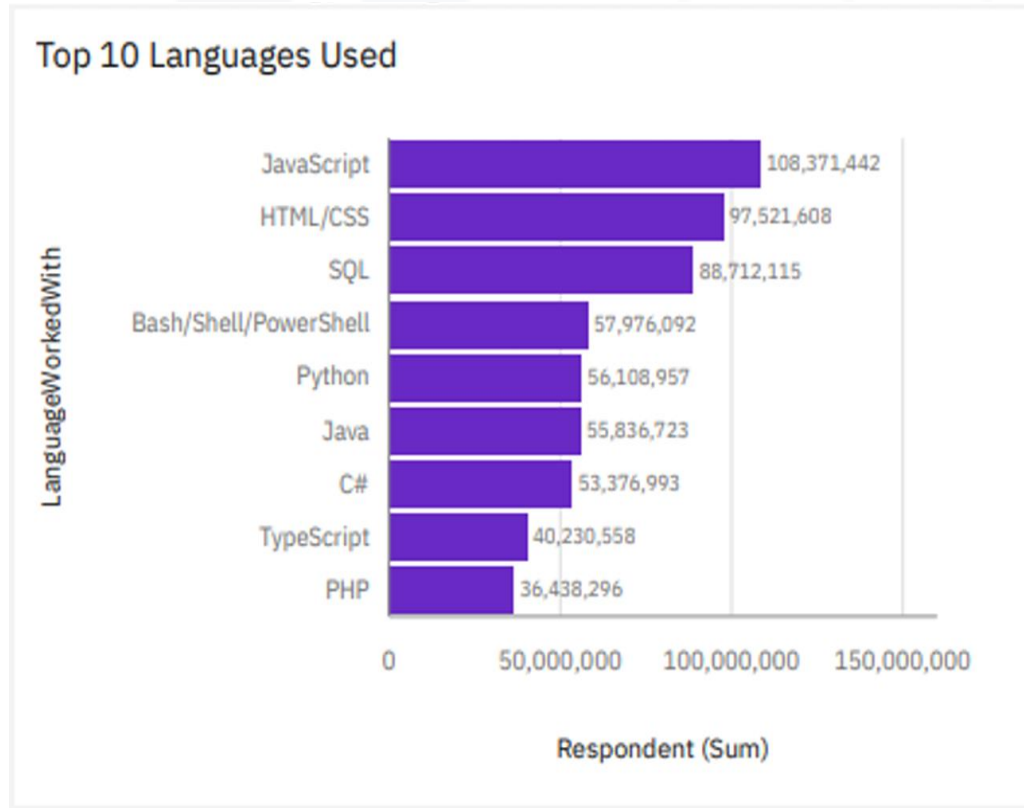- **Recommendations**: Providing suggestions for stakeholders or future research.

Impact and Relevance:
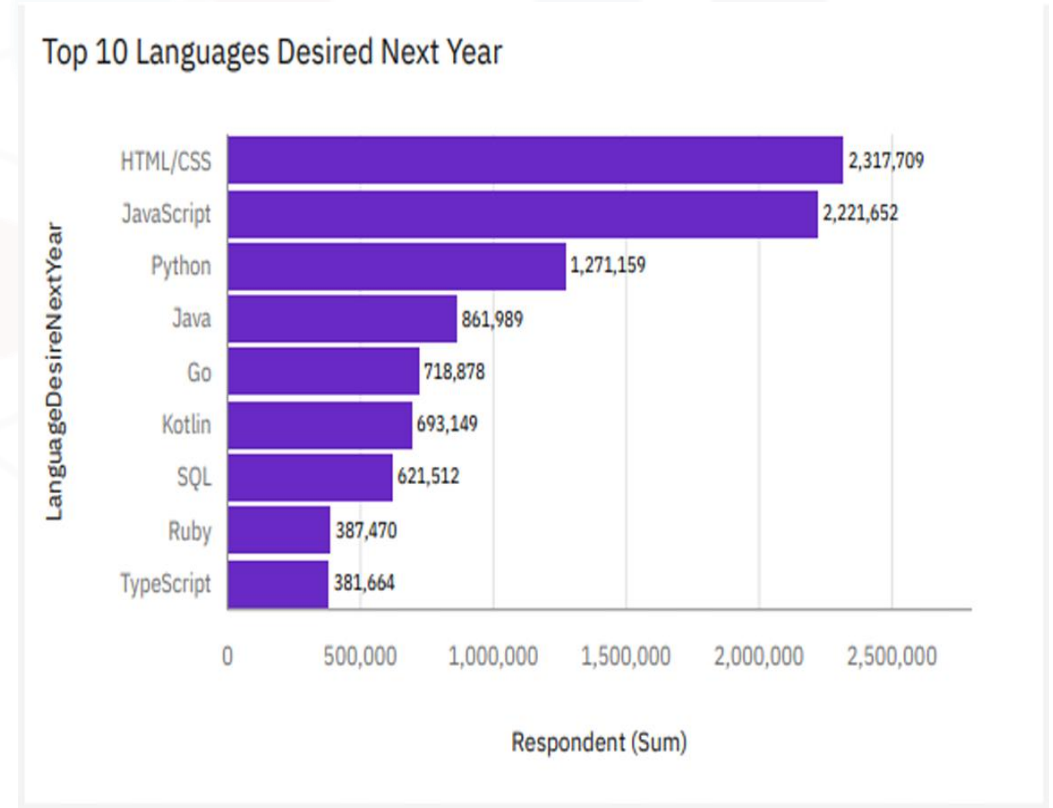- **Field Impact**: Discussing the relevance of findings within the industry.
- **Summary**: Concluding with a succinct summary of key findings and their significance.

PROGRAMMING LANGUAGE TRENDS

Current Year

Next Year

Top 10 Languages Used

Language	Respondent (Sum)
JavaScript	108,371,442
HTML/CSS	97,521,608
SQL	88,712,115
Bash/Shell/PowerShell	57,976,092
Python	56,108,957
Java	55,836,723
C#	53,376,993
TypeScript	40,230,558
PHP	36,438,296

Top 10 Languages Desired Next Year

Language	Respondent (Sum)
HTML/CSS	2,317,709
JavaScript	2,221,652
Python	1,271,159
Java	861,989
Go	718,878
Kotlin	693,149
SQL	621,512
Ruby	387,470
TypeScript	381,664

IBM Developer

SKILLS NETWORK

PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

- **Finding 1**: JavaScript remains the most used programming language, indicating its importance in web development.

- **Finding 2**: Python shows significant usage and is highly desired for the next year, highlighting its growing influence in data science and machine learning.

- **Finding 3**: HTML/CSS is both highly used and desired, reflecting its fundamental role in web development.
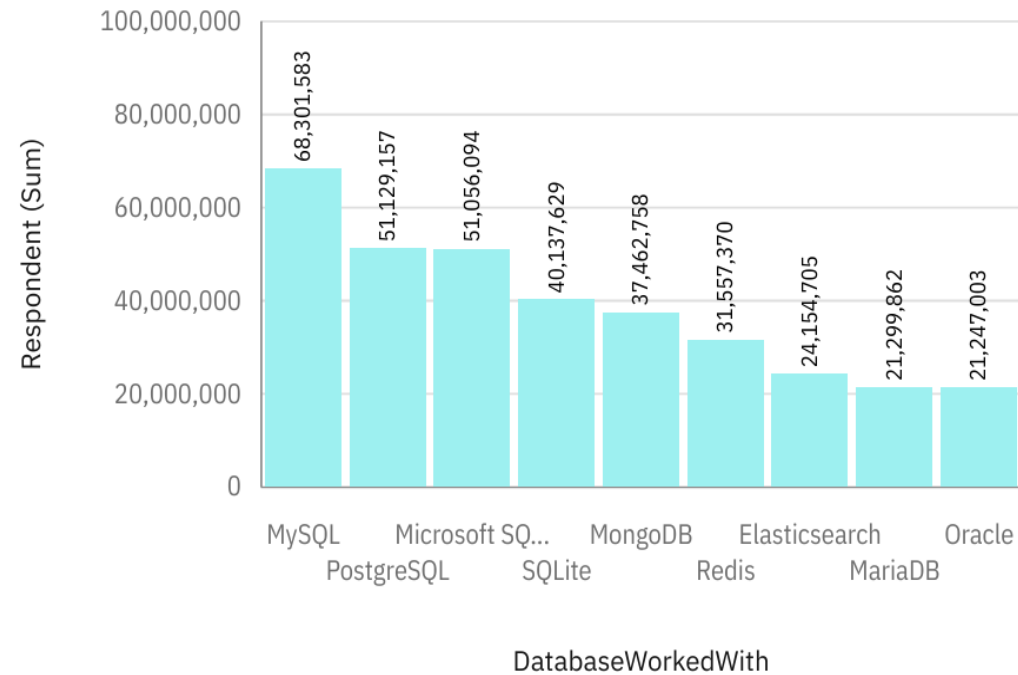
Implications

- **Implication 1**: The continuous demand for JavaScript and HTML/CSS suggests that web development skills remain crucial in the tech industry.

- **Implication 2**: The increasing popularity of Python emphasizes the need for expertise in data science, machine learning, and related fields.

- **Implication 3**: Emerging languages like Go and Kotlin gaining traction suggest a shift towards modern and efficient programming languages for developing scalable applications.
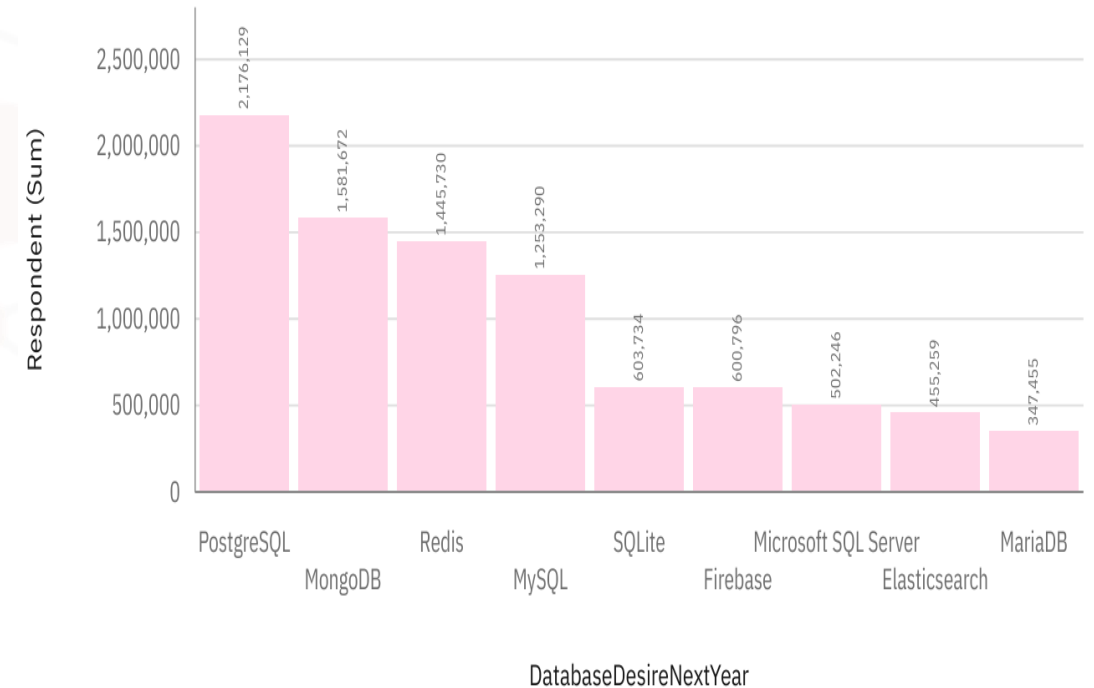
IBM Developer

SKILLS NETWORK

DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- **Finding 1**: MySQL is the most used database currently, reflecting its widespread adoption for web applications.

- **Finding 2**: PostgreSQL is highly desired for the next year, indicating a shift towards open-source, feature-rich databases.

- **Finding 3**: NoSQL databases like MongoDB and Redis are gaining popularity, highlighting the demand for flexible, scalable data solutions.

Implications

- **Implication 1**: The dominance of MySQL and PostgreSQL suggests that proficiency in these databases is crucial for database management roles.

- **Implication 2**: The increasing interest in NoSQL databases such as MongoDB and Redis emphasizes the need for skills in handling unstructured data and building scalable applications.

- **Implication 3**: The rising popularity of Firebase and Elasticsearch points to the growing demand for real-time database solutions and search engine capabilities, respectively.
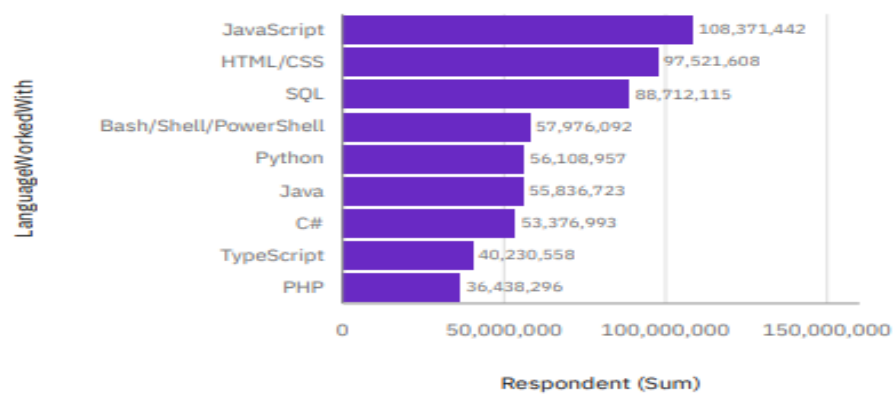
DASHBOARD

https://github.com/expertoclock/IBM-Data-Analyst-Capstone-Project/blob/main/IBM%20Cognos%20Analytics%20and%20Google_Looker_studio%20Dashboards.pdf
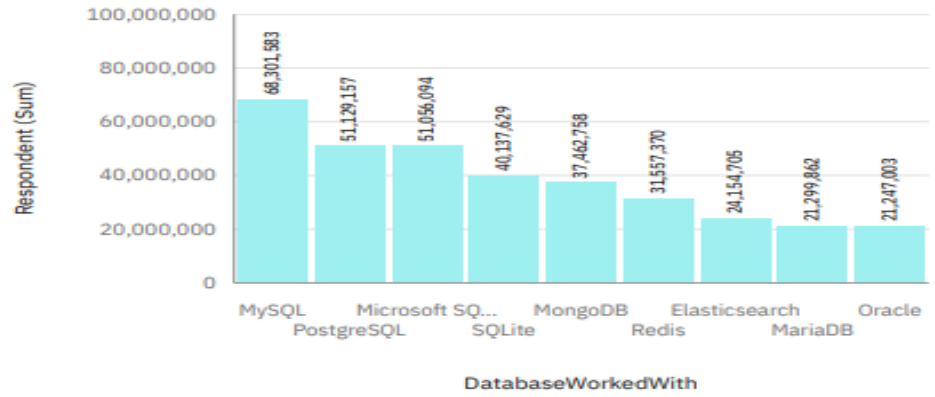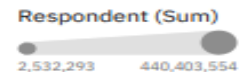
DASHBOARD TAB 1

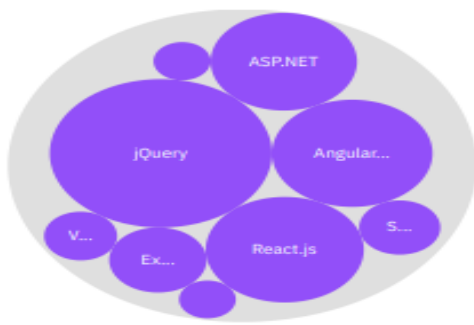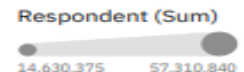## Current Technology Usage

Top 10 Languages Used

Language	Respondent (Sum)
JavaScript	108,371,442
HTML/CSS	97,521,608
SQL	88,712,115
Bash/Shell/PowerShell	57,976,092
Python	56,108,957
Java	55,836,723
C#	53,376,993
TypeScript	40,230,558
PHP	36,438,296

LanguageWorkedWith
Respondent (Sum)

Top 10 Databases Used

Database	Respondent (Sum)
MySQL	68,301,583
PostgreSQL	51,129,157
Microsoft SQ...	51,056,094
SQLite	40,137,629
MongoDB	37,462,758
Redis	31,557,370
Elasticsearch	24,154,705
MariaDB	21,299,862
Oracle	21,247,003

DatabaseWorkedWith

Platforms Worked With

Respondent (Sum)

2,532,293   440,403,554

(no value)

Linux AWS
Microsoft Azure
Google Cloud Platform
Docker
Windows

Top 10 Web Frameworks Used

Respondent (Sum)

14,630,375   57,310,840

ASP.NET
jQuery   Angular...
V...   React.js   S...
Ex...

IBM Developer

SKILLS NETWORK

DASHBOARD TAB 2

Future Technology Trend

Top 10 Languages Desired Next Year

LanguageDesireNextYear	Respondent (Sum)
HTML/CSS	2,317,709
JavaScript	2,221,652
Python	1,271,159
Java	861,989
Go	718,878
Kotlin	693,149
SQL	621,512
Ruby	387,470
TypeScript	381,664

Top 10 Databases Desired Next Year

DatabaseDesireNextYear	Respondent (Sum)
PostgreSQL	2,176,129
MongoDB	1,581,672
Redis	1,445,730
MySQL	1,253,290
SQLite	603,734
Firebase	600,796
Microsoft SQL Server	502,246
Elasticsearch	455,259
MariaDB	347,455

Platforms Desired Next Year

Respondent (Sum) 3,752,980 — 377,308,989

Respondent (Sum) 3,752,980 — 377,308,989

(no value) | Linux | Windows | Android | Kubern... | MacOS
Docker
AWS | Raspberry Pi | Slack | Ar... | H...
Google Clo... | Microsoft Azure | Word...
iOS

Top 10 Web Frameworks Desired Next Year

Respondent (Sum) 450,235 — 1,961,768

React.js, Vue.js, Express, F..., Rub..., jQ..., S...

DASHBOARD TAB 3

Demographics
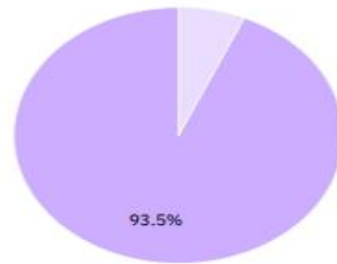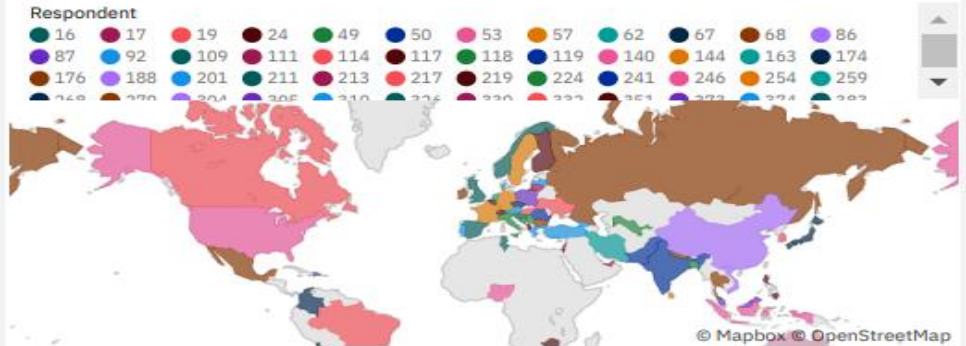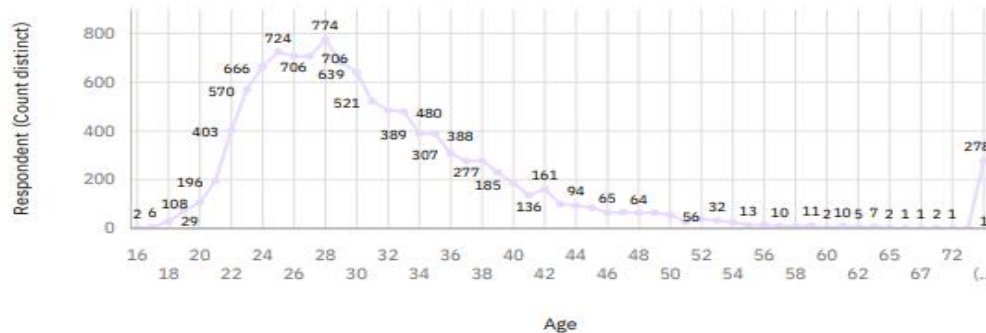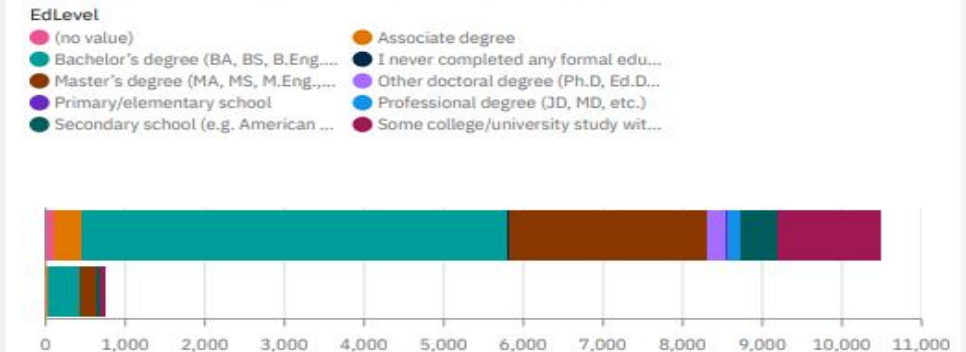
DISCUSSION

The analysis of the current and future trends in technology usage reveals several key insights and implications for professionals in the field. JavaScript, HTML/CSS, and SQL continue to dominate the programming landscape, underscoring their foundational role in building web applications and managing databases. The strong presence of PostgreSQL, MySQL, and MongoDB reflects a preference for open-source, scalable, and flexible database solutions that can handle a variety of data management needs.

The growing interest in containerization technologies like Docker and Kubernetes, as well as real-time databases like Firebase, suggests a shift towards modern, efficient, and scalable solutions that can support the increasing demands of today's applications. These trends indicate that professionals should focus on acquiring and refining skills in these areas to remain relevant and competitive.

Furthermore, the analysis highlights the importance of continuous learning and adaptation in the rapidly evolving tech landscape. By staying informed about the latest trends and technologies, professionals can better position themselves to meet the demands of the industry and contribute to innovative solutions that drive progress and success.

OVERALL FINDINGS & IMPLICATIONS

Findings

- **Finding 1: Current Technology Usage**
  - **Top 10 Languages Used:** JavaScript, HTML/CSS, SQL, Bash/Shell/PowerShell, Python, Java, C#, TypeScript, PHP are extensively utilized in the industry.
  - **Top 10 Databases Used:** MySQL, PostgreSQL, Microsoft SQL Server, SQLite, MongoDB, Redis, Elasticsearch, MariaDB, Oracle are prominent in database management.
  - **Platforms Worked With:** Various platforms like Google Cloud Platform, IBM Cloud or Watson, Microsoft Azure, Windows, Linux, Raspberry Pi, Docker, WordPress, Kubernetes, Android, MacOS, AWS, Arduino, Heroku, Slack, iOS are utilized across different projects.

- **Finding 2: Top 10 Web Frameworks Used**
  - **Frameworks:** jQuery, Angular, React.js, ASP.NET, Express, Spring, Vue.js, Django, Flask, Ruby on Rails are popular frameworks for web development.

- **Finding 3: Industry Trends**
  - **Emerging Technologies:** Trends indicate increasing adoption of cloud computing (AWS, Azure, Google Cloud), containerization (Docker, Kubernetes), and mobile development (Android, iOS).

Implications

- **Implication 1: Skill Development**
  - **Language and Framework Skills:** Professionals should focus on enhancing skills in JavaScript, Python, React.js, and Angular to align with current industry demands.
  - **Database Proficiency:** Knowledge of MySQL, PostgreSQL, and MongoDB is crucial for effective data management and analysis.

- **Implication 2: Platform and Cloud Adoption**
  - **Cloud Services:** Familiarity with AWS, Azure, and Google Cloud Platform is advantageous for scalability and flexibility in project deployment.
  - **Platform Expertise:** Experience with Linux, Docker, and Kubernetes is essential for efficient application deployment and management.

- **Implication 3: Future-readiness**
  - **Continuous Learning:** Staying updated with emerging technologies like serverless computing, AI/ML integration, and cybersecurity enhances career prospects.
  - **Adaptability:** The ability to adapt to evolving frameworks and platforms ensures relevance in a dynamic tech landscape.

IBM Developer

SKILLS NETWORK

CONCLUSION

- **Key Findings and Insights:**
  - Analysis revealed significant trends in programming languages, databases, and technology platforms.
  - Visualizations and statistical analysis provided actionable insights into industry preferences and future trends.

- **Implications for Data Analysis and Business Impact:**
  - Identified languages like JavaScript, Python, and SQL are critical for current and future data analysis needs.
  - Trends in database preferences such as PostgreSQL and MongoDB indicate shifting data management strategies.

- **Recommendations for Skill Development:**
  - Emphasize proficiency in emerging technologies like cloud computing (AWS, Azure) and data visualization tools (Tableau, Power BI).
  - Continuous learning in statistical analysis and machine learning enhances analytical capabilities.

- **Significance of the Capstone Project:**
  - Demonstrated the application of data analysis techniques to real-world scenarios.
  - Highlighted the importance of data-driven decision-making in business strategy.

- **Future Directions and Industry Impact:**
  - Anticipated trends in technology adoption underscore the need for ongoing adaptation and innovation.
  - Insights gained contribute to informed strategies for organizational growth and competitiveness.

APPENDIX

	Country	Record Count ▾
1.	United States	3,127
2.	India	897
3.	United Kingdom	821
4.	Germany	708
5.	Canada	432
6.	France	339
7.	Brazil	325
8.	Australia	285
9.	Netherlands	254
10.	Spain	251
11.	Russian Federation	208
12.	Poland	205
13.	Italy	188
14.	Sweden	162
15.	Switzerland	148
16.	Ukraine	108
17.	South Africa	102
18.	Israel	102
19.	Mexico	98

1 - 100 / 135

IBM Developer

SKILLS NETWORK