Data Analytics Capstone Project

Kacper Szafraniec

04.09.2024

OUTLINE

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

EXECUTIVE SUMMARY

- **Objective**: Analyze technology usage, trends, and demographics from global developer survey data.
- **Data Collection**: Gathered job postings, scraped programming salaries, and processed survey data.
- **Key Insights**: Identified top programming languages, databases, and emerging tech trends; visualized demographic profiles.
- **Tools**: Utilized Python for data processing, Matplotlib for initial plots, and IBM Cognos Analytics for interactive dashboards.
- **Deliverables**: Created comprehensive dashboards, exported insights, and prepared for presentation on skill trends and future needs.

INTRODUCTION

- **Context**: As technology evolves, identifying in-demand skills is crucial for staying competitive in the IT and consulting sectors.
- **Goal**: Analyze and visualize current and future technology trends, including programming languages, databases, and demographics from a global survey.
- **Approach**: Collected and processed data from job postings, web scraping, and surveys to identify key trends and insights.
- **Tools Used**: Employed Python for data analysis and visualization, and IBM Cognos Analytics for creating interactive and comprehensive dashboards.
- **Outcome**: Provided actionable insights and visualizations to guide strategic decisions on skill development and technology focus.

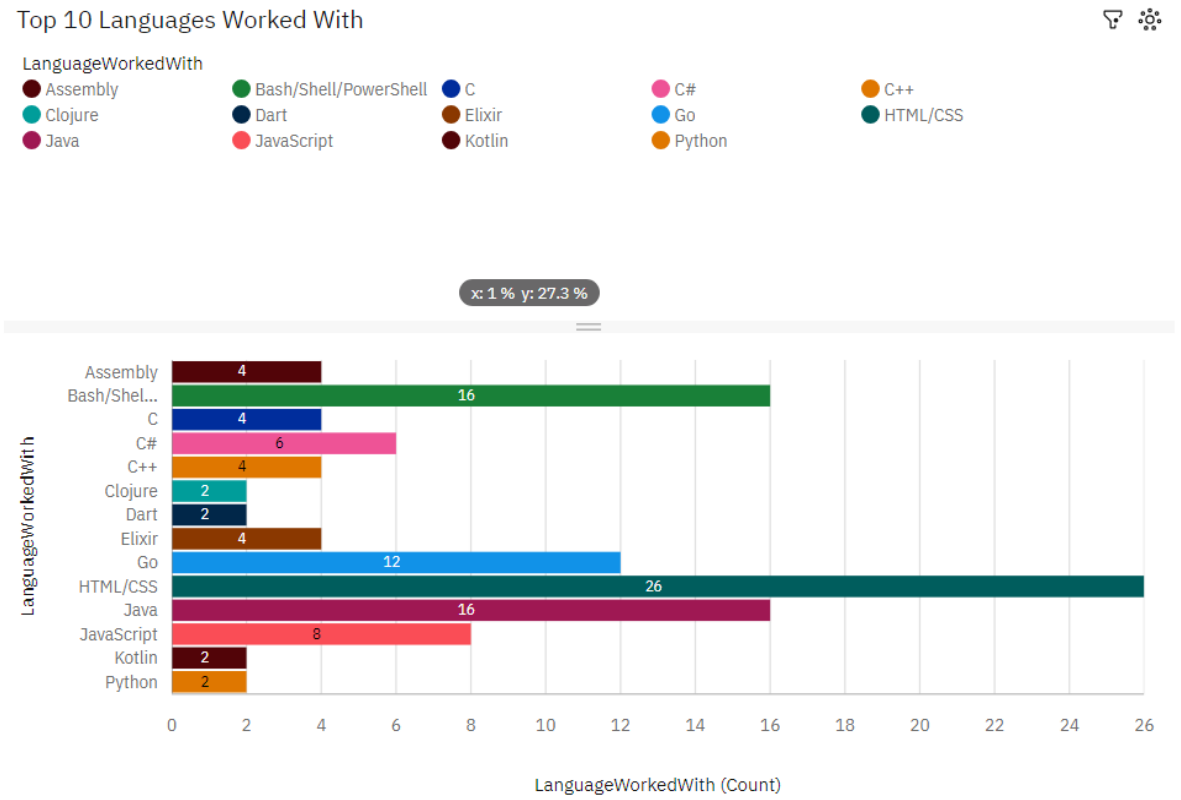
METHODOLOGY

- **Data Collection**: Used APIs and web scraping to gather job and salary data; analyzed survey responses on technology use.
- **Data Wrangling**: Cleaned data by removing duplicates, handling missing values, and normalizing compensation.
- **Visualization**: Created charts to analyze data distribution, relationships, and comparisons.
- **Dashboard Creation**: Developed interactive dashboards in IBM Cognos Analytics with focused tabs on technology usage, trends, and demographics.
- **Reporting**: Compiled findings into a PDF report and shared via GitHub.

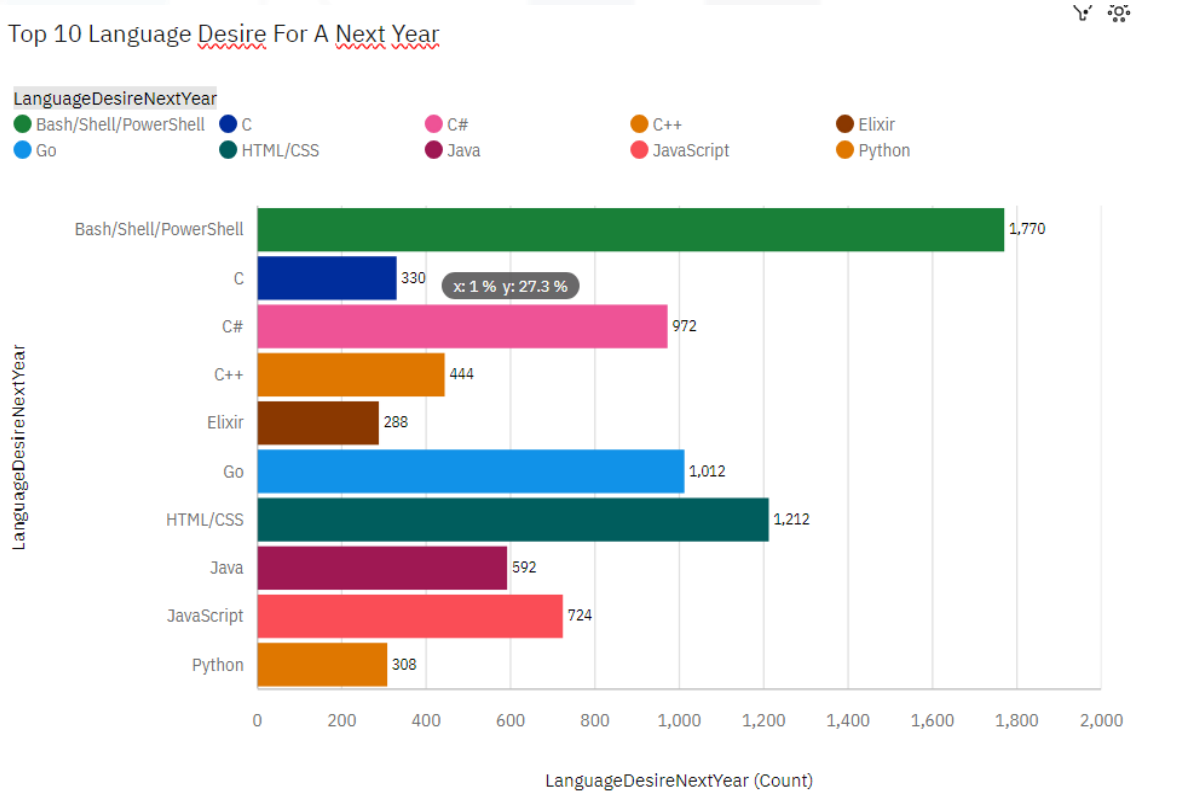
RESULTS

PROGRAMMING LANGUAGE TRENDS

Current Year



Next Year



IBM Developer

SKILLS NETWORK

PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

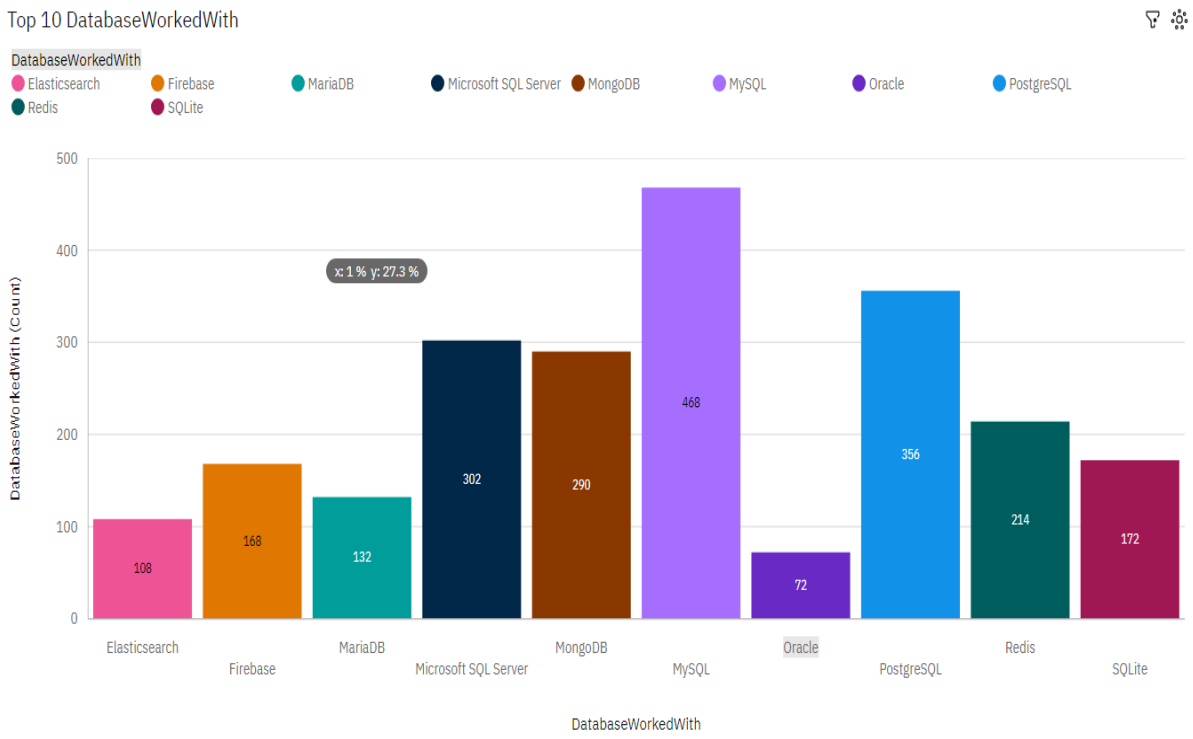
- **Consistency in Core Technologies**: HTML/CSS, Bash/Shell/PowerShell, and Go are consistently ranked among the top languages both in current use and in those developers want to learn, indicating their foundational importance across projects.
- **Emerging Interest in Go**: Go has a strong presence in both the current usage and learning interest lists, suggesting it is gaining traction as a preferred language for future projects.
- **Legacy Languages Maintaining Relevance**: C, C++, and Java are among the top languages worked with and remain high on the list of languages developers want to learn, demonstrating the continued importance of these languages in the industry.

Implications

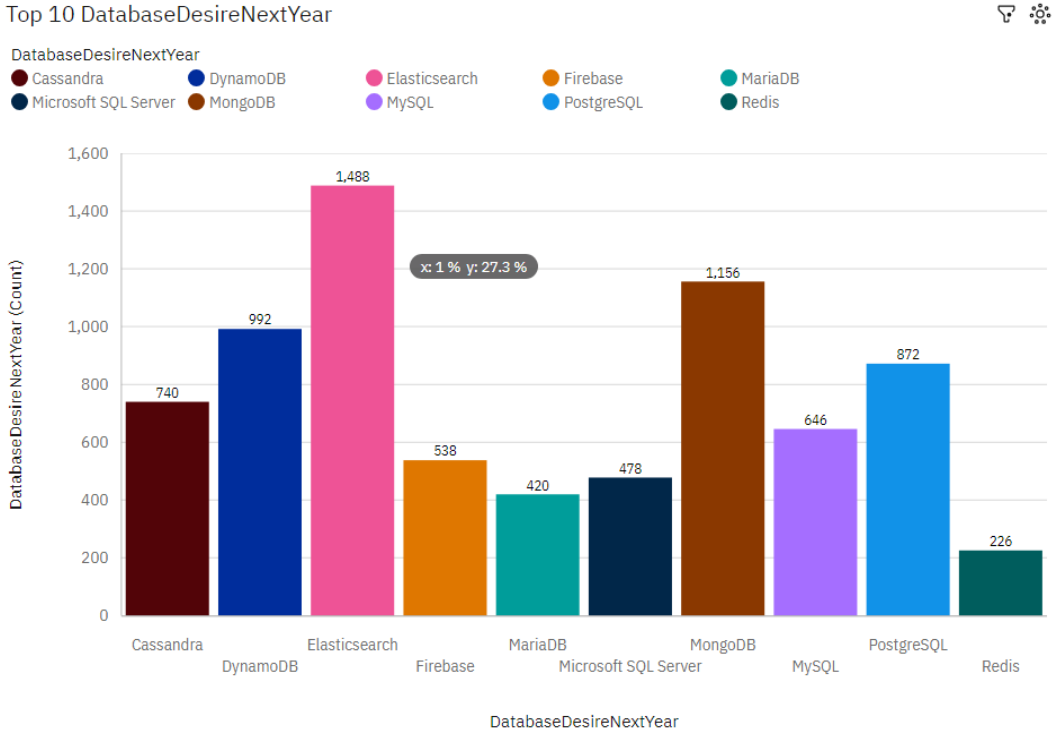
- **Invest in Training for Key Languages**: Organizations should focus on providing training in HTML/CSS, Bash/Shell/PowerShell, and Go, as these are not only widely used but also in high demand for future learning, ensuring that teams are equipped with the most relevant skills.
- **Adopt and Expand Go Usage**: The strong interest in learning Go suggests that companies should consider adopting or expanding its use in their technology stacks to stay ahead of industry trends and leverage its growing popularity.
- **Support for Legacy Systems and Projects**: The enduring popularity of C, C++, and Java implies that organizations must continue to support and develop expertise in these languages to maintain and optimize legacy systems, while also preparing for future growth and updates.

DATABASE TRENDS

Current Year



Next Year



IBM Developer

SKILLS NETWORK

DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- **Shift Toward NoSQL Databases**: MongoDB and Elasticsearch, both NoSQL databases, are highly ranked in both current use and future interest, with Elasticsearch moving to the top spot in desired learning.
- **Growing Interest in DynamoDB and Cassandra**: DynamoDB and Cassandra, both of which are NoSQL databases, are increasingly sought after for future learning, despite not being in the current top 10 usage list, indicating a trend towards distributed databases.
- **Declining Interest in Traditional RDBMS**: While MySQL and Microsoft SQL Server are still among the most used databases, their positions drop in the list of databases developers want to learn next year, suggesting a gradual shift away from traditional relational database management systems (RDBMS).

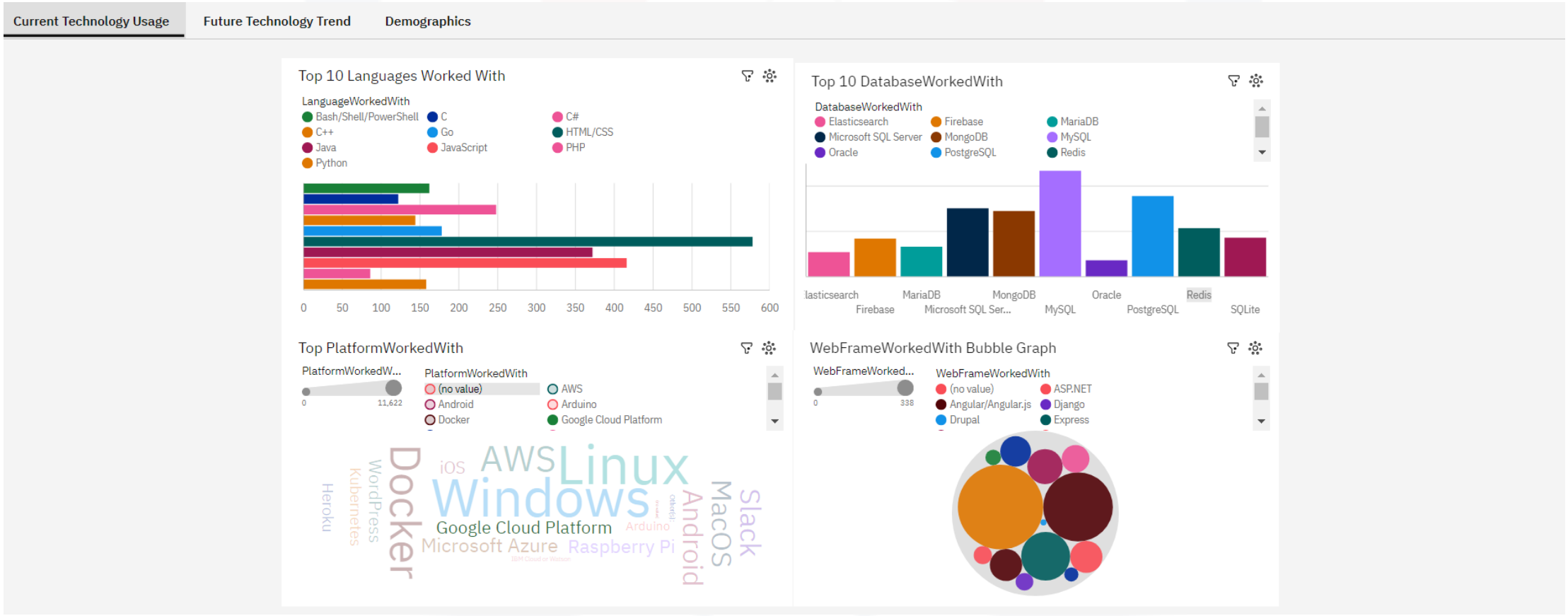
Implications

- **Strategic Focus on NoSQL**: Companies should consider investing in and expanding their use of NoSQL databases like MongoDB and Elasticsearch, as they are increasingly favored by developers, indicating a shift towards more flexible, scalable data storage solutions.
- **Prepare for Distributed Systems**: The rising interest in DynamoDB and Cassandra suggests that organizations should start building expertise in these technologies, preparing their infrastructure for distributed and cloud-based database systems.
- **Reassess RDBMS Reliance**: The declining interest in learning traditional RDBMS like MySQL and Microsoft SQL Server implies that businesses should reassess their reliance on these systems and explore hybrid or NoSQL solutions to stay aligned with evolving industry trends.

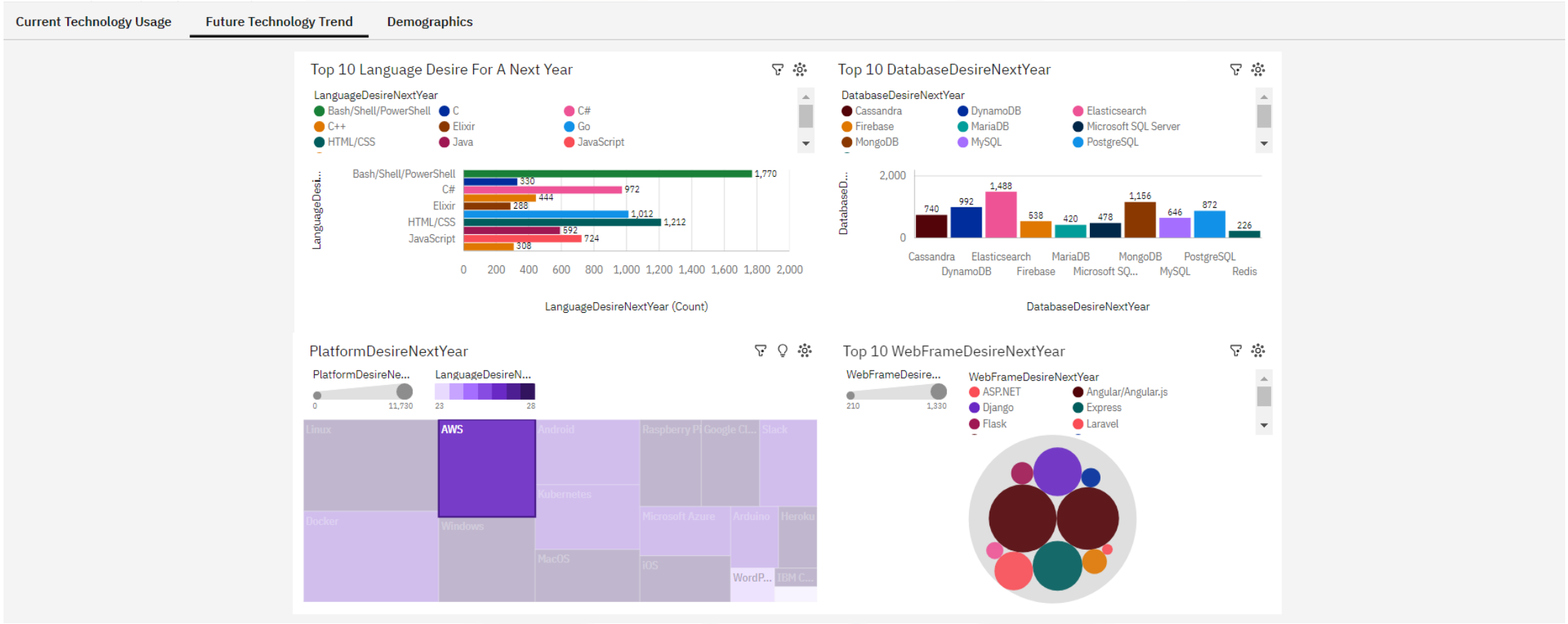
DASHBOARD

- Dashboard: https://us3.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2Fcapstone%2Bproject&action=view&mode=dashboard&subView=model00000191bc727505_00000000
- GitHub Project : https://github.com/Screachail/Data-Analytics-Capstone-Project

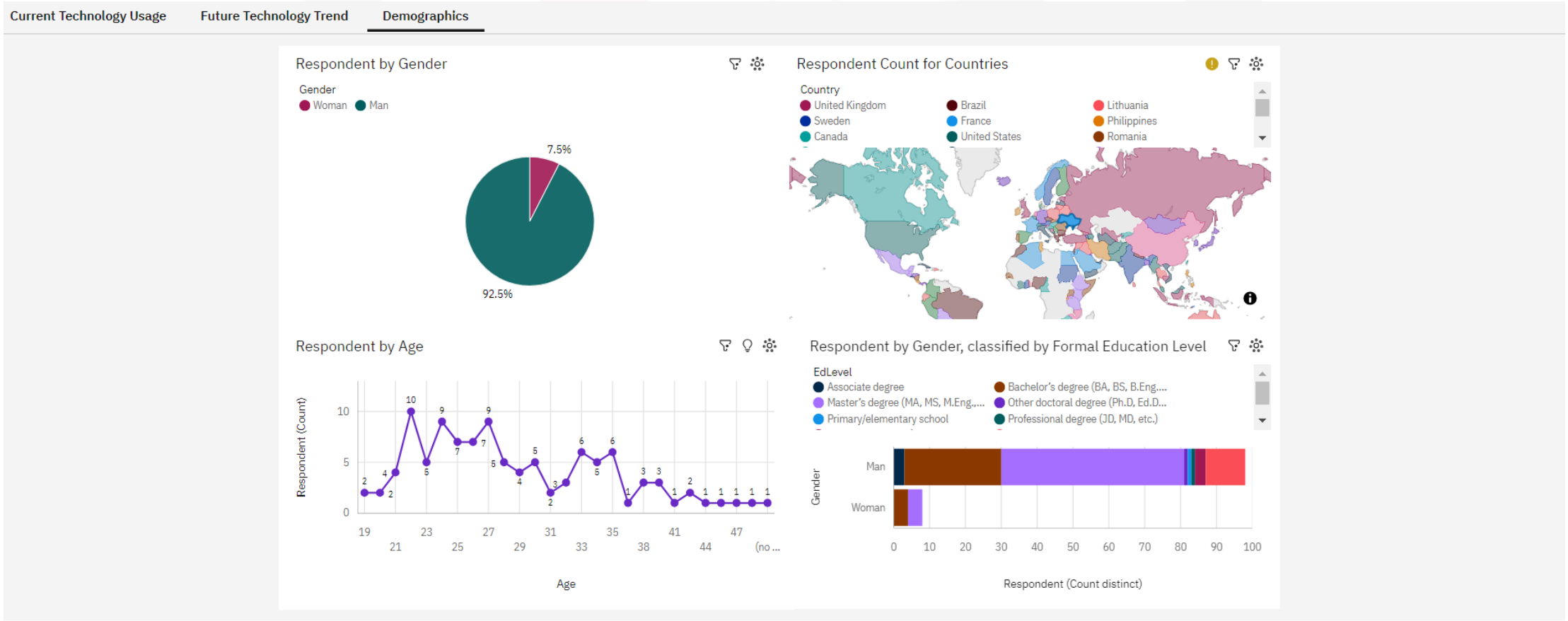
DASHBOARD TAB 1



DASHBOARD TAB 2



DASHBOARD TAB 3



DISCUSSION

- **Skill Demand Shifts**: The project reveals a dynamic shift in the skills landscape, with emerging technologies and languages gaining traction among developers, indicating the need for ongoing skill development and adaptation.
- **Training and Development**: Organizations must prioritize upskilling and reskilling initiatives, focusing on in-demand technologies like Python, Bash/Shell/PowerShell, and emerging databases, to stay competitive and meet future demands.
- **Strategic Workforce Planning**: The trends identified suggest that companies need to align their hiring strategies with the evolving technology preferences to attract and retain top talent, particularly in areas such as cloud computing, big data, and modern programming languages.

DISCUSSION

- **Technological Evolution Impact**: The rise in popularity of newer technologies like Go, Elasticsearch, and DynamoDB implies that traditional systems may see reduced investment, prompting organizations to evaluate and potentially migrate legacy systems.
- **Continuous Monitoring:** As technology trends rapidly evolve, continuous data analysis and monitoring of skill demands are essential for organizations to remain agile and responsive to industry changes.

OVERALL FINDINGS & IMPLICATIONS

Findings

- **Shifting Technology Preferences**: There is a clear trend toward modern and emerging technologies, with languages like Python and Go, and databases like Elasticsearch and MongoDB gaining popularity among developers.
- **Consistency in Core Skills**: Despite the shift towards newer technologies, core skills like HTML/CSS and SQL databases remain essential, indicating their foundational role in the tech industry.
- **Demand for Versatile Skill Sets**: Developers are increasingly interested in learning a diverse range of technologies, suggesting that versatile, multi-skilled professionals are highly valued in the current job market.

Implications

- **Strategic Skill Development**: Organizations must invest in training programs that focus on both emerging and core technologies to ensure their workforce remains competitive and well-rounded.
- **Adapting to Technological Change**: Businesses should consider updating their technology stack to include newer, more in-demand technologies to attract top talent and stay ahead in innovation.
- **Workforce Planning**: HR and recruitment strategies should be aligned with the shifting technology landscape, emphasizing the need for candidates proficient in both established and emerging tools and languages.

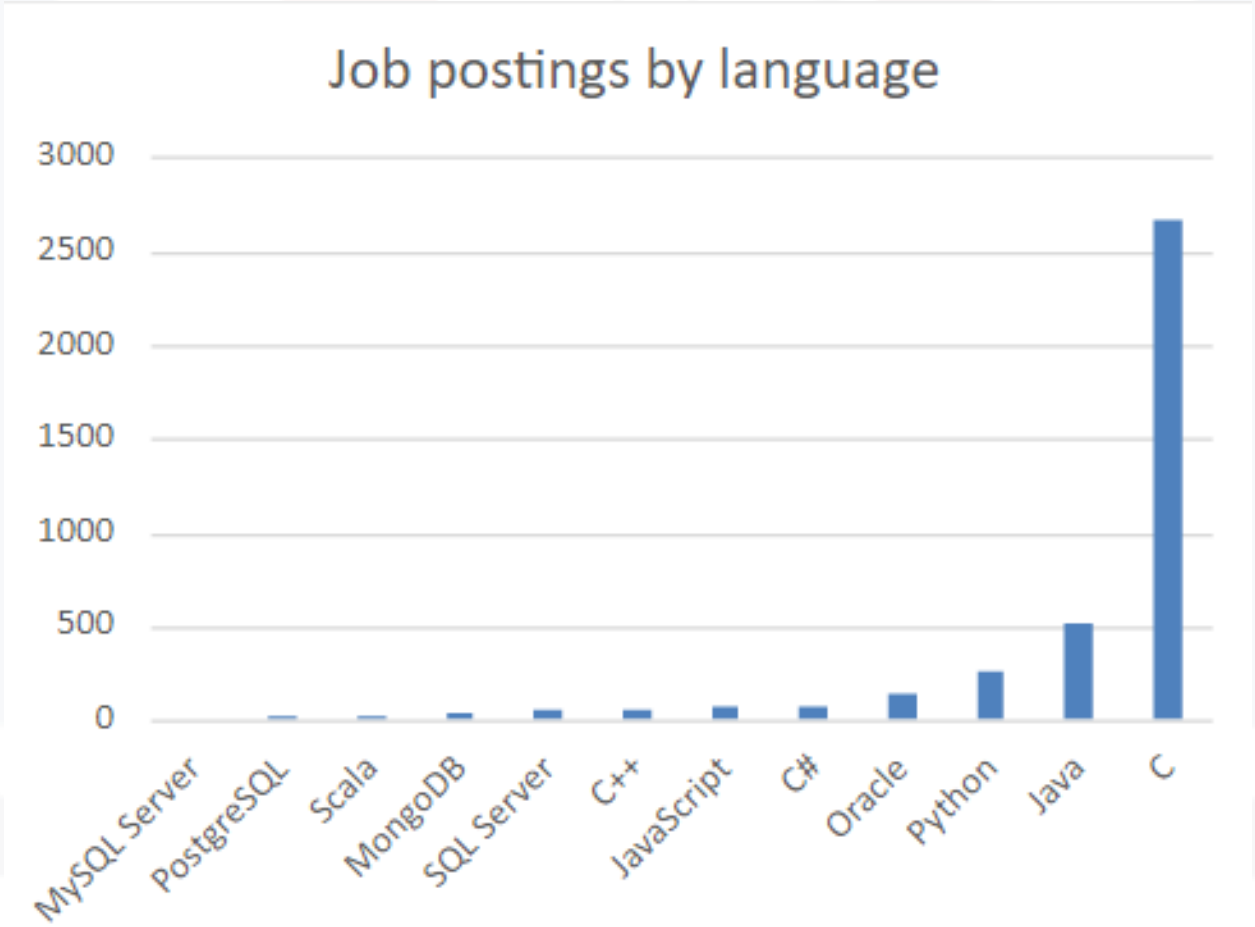
CONCLUSION

- **Evolving Technology Landscape**: The analysis reveals a dynamic shift towards newer programming languages and databases, indicating a rapidly evolving technology landscape that both developers and organizations need to adapt to.
- **Enduring Relevance of Core Technologies**: Despite the rise of new technologies, core languages like HTML/CSS and established databases like MySQL and PostgreSQL continue to hold significant relevance, underscoring their foundational role in software development.
- **Growing Demand for Versatility**: The trend of developers wanting to learn a broad range of technologies highlights the increasing demand for versatile skill sets in the industry, where adaptability and continuous learning are key.
- **Implications for Workforce Development**: Organizations must align their training and recruitment strategies with these trends, focusing on both emerging technologies and essential skills to maintain a competitive edge and foster innovation.

APPENDIX

- Additional charts and Jupyter Notebooks used for this project can be found here: https://github.com/Screachail/Data-Analytics-Capstone-Project

JOB POSTINGS



POPULAR LANGUAGES